FROM TEXT TO MEANING: INFORMATIONAL SEMANTICS OF SHORT SCIENTIFIC TEXTS

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Neslihan Süzen

School of Mathematics and Actuarial Science

University of Leicester

September 2020

Abstract

From Text to Meaning: Informational Semantics of Short Scientific Texts

by Neslihan Suzen

The problems of automatic analysis and representation of human language have been clear since the inception of Natural Language Processing (NLP). Machines can be easily fooled when it comes to interpreting sentences and extracting meanings from texts. Semantically-driven processing needs deep understanding of natural languages by machines, and algorithms relying on word co-occurrence and frequencies can not activate semantically-related concepts/experiences as human brain does.

This thesis presents computational methods for semantic analysis and quantifying the meaning of short scientific texts in a new light. Methods of this research attempt to extract semantic features that are not explicitly expressed in the text, and provide predictions about human cognition. Rather than psychological properties, we describe the situation of use of words for scientific texts by scientifically specific description – subject categories of the text.

First, this thesis investigates Bag of Words model on a corpus of students' answers. Automated scoring systems were created for marking of short answer questions and for providing feedback to students on their answers. Students' marks were predicted by a mathematical model through words selected to transmit information.

Second, we introduced novel techniques for quantifying the meaning for words and then texts. Leicester Scientific Corpus (LSC) and Leicester Scientific Thesaurus (LScT) were built for empirical studies. LSC is a corpus of 1,673,350 scientific texts and LScT is a thesaurus of 5,000 words extracted from the LSC. Methodologies for semantic analysis were developed based on informational representation of the meaning extracted from the occurrence of the word in texts across the scientific categories. Vector representation of words was created in the newly constructed Meaning Space (MS), and utilised in representing text meaning. Feature Vector of Text (FVT) were introduced and created for LSC texts as a vector representation of meaning. This approach obtains superior performance to standard frequency representation in identifying scientific-specific meanings.

Finally, this thesis presents a research in evaluating the impact of scientific articles through their informational semantics. Newly developed approach to meaning have offered a way to predict the scientific impact of papers, and the study details examples of text classification going from 80% success to distinguish highly-cited and less-cited papers.

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from the University of Leicester. I greatly appreciate to the University authorities for their help, and each of very generous offers and decisions made over the four years. I hope someday I will repay some of it by becoming a successful scientist.

Acknowledgement

I would like to express the deepest appreciation to my supervisor, Professor Alexander Nikolaevich Gorban. He has continually and convincingly conveyed a spirit of adventure in regard to research, and an invaluable support, advice and motivation throughout my doctoral research. Indeed, without his guidance and persistent help this thesis wouldn't have been possible. He has been an inspiration to me, and I am sure to many others. His advice, help and life lessons are priceless.

I am extremely grateful to my second supervisor, Dr Evgeny Moiseevich Mirkes. I would like to thank to him for his constant encouragement, patient guidance and advice. His support and constructive critical feedback have been inestimable value throughout my study.

I wish to express my sincere gratitude to Professor Jeremy Levesley for his wonderful collaboration, advice and for always being so supportive. This thesis has been enriched significantly through helpful discussions with him.

A very special thank you to Dr Akinduko Ayodeji Akinwumi for his invaluable advice, especially earlier in my study. I have learnt a lot from him and have become much more confident about my machine learning skills.

I would like to express my eternal appreciation towards my parents, sisters and relatives who have always been there for me. Thank you for your unconditional supports, patience and inspiration throughout my life.

Finally, I am very much indebted to my friends for their invaluable support during my PhD study. Thank you to all of my friends and colleagues in the School of Mathematics and Actuarial Science. It has been great to know all of you during my time here in the school.

Neslihan

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1. Setting the Stage: Preliminaries and Problem Statement

The motivations for this thesis were our awareness of four main problems (and necessities) as a result of our scientific curiosity towards automated text mining for academic publications. The first was that, the existing computational schemes to represent the meanings were not efficient to extract scientific-specific meanings for academic texts in English. To the best of our knowledge, we did not know any semantic-driven representation methodology developed for scientific text. The second was that, when we went deeper in investigation of the 'meaning of meaning', we realised that general-purpose word representation methods did not take the relativeness of the importance of a word across scientific categories into account. The third was that although we knew that several academic dictionaries exist, we did not know any category-specified thesaurus for science, created according to the importance of words within individual categories. Finally, our findings trigged our curiosity to take a step further: investigation the link between the knowledge of semantics of scientific texts and their impact in scientific community. Therefore, our curiosity turns into a comprehensive research on analysing the semantics for a large family of scientific texts.

Understanding the human language and expression in natural languages is one of the ultimate goals of the fields ranging from linguistics, philosophy, psycholinguists to neuroscience, while automatic analysis of texts and understanding natural languages by machines are what Artificial Intelligence and Machine Learning dream about. Human language is ambiguous and system specifically constructed for communication; therefore, there is high level of complexity in representing the speaker's/writer's meaning. Unlike machines, human do not have limitations in interpreting the sentences and semantic binding, where machines can fail when it comes to representation of concepts. Computational methods attempt to predict human behaviours in processing natural languages as human brain does. Therefore, deeper understanding of semantics and creating human-to-machine interaction have required a combined effort from psychology, neuroscientists, philosophers, linguistics and computer scientists in describing the human cognition, creating the schemes for act of communication, and building common-sense knowledge bases for the 'meaning' in texts. Let us now have a quick look at ideas about this deeply discussed topic: *meaning of meaning*.

First of all, we start from the Wittgenstein formulation: "Meaning is use" or, in more detail, "For a large class of cases – though not for all – in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language"[1, §43]. For our world of scientific abstracts, there is a well-defied dominant communicative function: a representative function. In the idealised scheme of the act of communication, two representations of the situation on the blackboards of consciousness exist: the sender's representation (representation 1) of the situation (situation 1) and the receiver's representation (representation 2) of the situation (situation 2). A text related to the first situation is generated by the sender (translation 1). This text is transmitted to the receiver and transformed by receiver into a representation of a situation (translation 2). The sender's and the receiver's representations never coincide, and situations can be real in a real world, impossible in an impossible world and chimeric combined from several possible or imaginary situation. We do not consider any controversy of the reality of the situation. Instead, we consider the chain Representation 1 → Text → Representation 2 and translations between them. Translations depend on much wider context of the communication including experience of the sender and receiver. It is noteworthy that there may be many receivers and senders. One-to-many or even many-to-many communication adds more situations and representations and may also add some less trivial multi-agent structures with additional communication channels.

For our analysis, the language is used to transmit the information about the represented situations, so many other usages of the language, from military orders to psychological manipulations, are disregarded. The act of communication includes just very basic elements and can be elaborated in much more detail.

A very basic scheme is sufficient for our analysis of meaning. Meaning, for our analysis, is hidden in the relationship between the representation of situations on the 'blackboard of the consciousness' and the texts of the messages. The meaning of meaning can be understood if and only if the translation operations are created in the scheme of a communication act. Moreover, understanding can be represented as a reflexive game [2] with different levels (The sender prepares a message taking into account the experience of the receiver, his goals and tools, and guesses that the receiver takes into account the experience of the sender, his goals and tools, and... Analogously, the receiver tries to understand the message taking into account..., etc.)

The relation between the text and the representation of the situation is many-to-many correspondence: each text corresponds to many situations and each situation can have many representing texts. Mel'čuk [3, 4, 5, 6] describes the natural language as "*the meaning to text and text to meaning transformer*". According to this we are able to describe meaning in a special *semantic language*. At this stage, we characterise a situation "behind the text" by a set of attributes. The method of this characterisation can be changed and does not give a unique and exhaustive presentation of it.

Despite the challenges in creating and describing the plausible translation, with remarkable progress of machine translation, applying the modern machine learning tools seems to be attractive idea to analysis and simulation the translation operations. However, there is no generally accepted tools for working directly with representations of situations, and we cannot propose a general solution to this problem. Such a solution, perhaps, is impossible in a finite closed form despite many efforts over decades.

Our goal is more modest. We will provide computational analysis of relations between texts of messages and representations of situations for a large collection of brief scientific texts. Such representations must be standardised, at least in part, and expressed in the form of diagrams, specially organized texts or other means. The simplest and universal approach is to replace the situation representations with the values of some attributes. Many forms of more specific descriptions of situations can be transformed into vectors of attributes. In sentiment analysis, classical examples are provided. We aim to provide another basic example that is specific to scientific texts: a list of scientific subject categories that the text belongs to.

The scheme of simple communication does not contain the representation of situation by a set of attributes and so does not introduce the attributes. This is an additional operation that must be performed: evaluating the values of selected attributes. This operation can be done either on the sender's side, the receiver's side, or by combinations of these approaches. For example, categorisation of a brief scientific text is a result of combined efforts: the authors select the categories by their choice of the journal, of the keywords, or by the pointing the categories directly, then the editors can have their own choice, then databases can finalise the list of subject categories for this text.

The subject categories can be chosen with an understanding of the text by many agents – can be both human or computer system –, and conflicts of understanding are possible. Even famous preprint servers (such as arXiv), moderators can sometimes change the category selected by the authors. This is because the content of the text may differ from its meaning [7], which are often confused (just as understanding the situation behind the text is often confused with recognising the content of the text).

In our analysis of meaning, the starting point is the combination of the text with the list of the subject categories the text belongs to – definition of the attributes of the situation behind the text. The key idea of this approach goes back to the lexical approach of Sir Francis Galton, who selected the personality-descriptive terms and stated the problem of their interrelations for real persons. Following his idea, Thurstone [8] selected sixty adjectives (attributes of a person that are in common use), and asked to the respondents (1300 persons) to select the adjectives that can best describe a person they know well. A person was described by a 60-dimensional Boolean vector. Factor analysis gave five factors. After many years of development and discussions, the modern five-factor personality model became one of the common tools in psychodiagnosis [9, 10].

In classical psycholinguistic studies, similar approach was used. Osgood with co-workers hypothesised 3-dimensional semantic space to quantify connotative meanings in the theory of *Semantic Differential* concerning psychological and behavioural aspects [11, 12]. They used an approach for extraction of three 'coordinates of meaning' from the evaluation of the 'affective meaning' of words (objects) by people. The semantic space was built by, in his words, 'three orthogonal bipolar dimensions': Evaluation (E), Potency (P) and Activity (A). Of course, the researches started from many different scales and these three were extracted by factor analysis.

We can guess that these evaluations of a single object or person were related to some situations with this single object or person, not just to an isolated abstract object. The people evaluated not the abstract 'terms' but the psychologically meaningful situations behind these terms. These situations were the sources of the 'affective meaning' or the personality evaluations.

1.2. On the Stage: Approaches and Contributions in Brief

When it comes to scientific-specific meanings, the 'affective meaning' or psychological properties are not rational to describe the situations behind scientific texts. For our world of scientific texts, we characterise the situation of use by scientifically specific description – the research subject categories of the text. Quantifying the meaning in our research follows the road: Corpus of texts + categories $\rightarrow$ Meaning Space (MS) for words $\rightarrow$ Geometric representation of the meaning of texts.

In our analysis of meanings, the starting point is to combine the text with the list of the subject categories the text belongs to. These categories can intersect: a text can belong to several categories as texts can be assigned to more than one category. The categories evaluate the situation (the research area) related to the text as a whole, not as a results of the combination of words' meaning. This holistic approach defines the general meaning of a word in short scientific texts as the information that the use of this word in texts carries about the categories to which these texts belong. More explicitly, we quantify the meaning by the *Relative Information Gain (RIG)* about the subject categories that the text belongs to, which can be obtained from presence of the the word in the text. This is done via characterisation of the research situations behind the text by binary attributes. Two attributes of text d for a given word w_j and a given category c_k is defined as:

$c_k(d)$: The text d is in the category c_k: Attribute values are Yes ($c_k(d) = 1$) or No ($c_k(d) = 0$);

$w_j(d)$: The word is in the text: Attribute values are Yes ($w_j(d) = 1$) or No ($w_j(d) = 0$).

In this approach, the corpus of scientific texts is a probabilistic sample space (the space of equally probable elementary results, each of which is a random selection of text from the corpus). $RIG(c_k, w_j)$ measures the (normalized) information about the value of $c_k(d)$, which can be extracted from the value $w_j(d)$ (i.e. from observing or not observing the word w_j in the text d) for a text d from the corpus. By this, we identify the importance of the word for the corresponding category in terms of information gained when separating the corresponding category from its complement.

To follow our road, a triad is needed: texts, dictionary and multidimensional evaluation of the situation of use presented by the categories. In this research, short scientific texts are abstracts of research articles or proceeding papers. For the first element of the triad, the whole world of abstracts is narrowed to a sample: 1,673,350 texts from the *Leicester Scientific Corpus (LSC)* [**13**]. The meaning of a word extracted from the corpus is represented by a 252-dimensional vector of RIGs, in which each of texts in LSC is assigned to at least one of these 252 Web of Science (WoS) categories [**14**]. Thus, we use these simple 252 binary attributes for multidimensional evaluation of the text usage situation, where the second element of the triad is *Leicester Scientific Dictionary-Core (LScDC)* [**15**].

Next, a vector space to represent word's meanings has been introduced: *Meaning Space*. In Meaning Space, coordinates correspond to the subject categories. Each word w_j in the dictionary is represented by a vector of $\overrightarrow{RIG_j}$ about the subject categories. These vectors are estimations of the meaning of words as their importance in the research fields. Following to the distributional semantic hypothesis, the hypotheses here are: if words have similar vectors, they tend to have similar meanings, and if texts have a similar distributions of word meanings – similar clouds of word vectors – then they tend to have similar meanings.

Having represented each word in the Meaning Space, they can be used in many text analysis problems including creation of thesaurus. One of techniques to build scientific thesaurus has been introduced and used to construct *Leicester Scientific Thesaurus (LScT)* [**16**]. The LScT contains of the most informative 5,000 words in science, in which the informativeness is measured as the average RIGs of a word across categories.

In this thesis, proposed representation scheme for text meaning is based on the LSC with LScT. The starting point of computational analysis of the meaning is a combination of simple *Bag of Words (BoW)* model with holistic approach to the text meaning: the text is considered as a collection of words, the meaning of the text is hidden in a situation of use, which is evaluated as a whole.

After having the words represented in MS, we return to our starting point, the question of how the meaning can be represented. This scheme creates the *Feature Vector of Text (FVT)* for each text that is created by a set of parameters and analysis in each cloud of words. Different vector representations for text meaning can be created in this framework. For example, in our study, vector representations as a combination of the mean vector, the vector of the first principal component for each text and two centroid vectors obtained by *k-means* (2-means) clustering of words in the text are introduced and analysed for informational representations of semantics.

Informational semantic representations are designed to be pertinent to encompass many text mining applications. In this research, we analyse one of the hardest mathematical challenging tasks: predicting the citation counts of articles.

We then turn to issue of automated evaluation the potential impact of scientific articles through their semantics. On the basis of many researches, citation analyses yield that citation counts are strong indicator of the impact of research [**17**, **18**]. Citation behaviour is, indeed, a very complex and multidimensional phenomenon discussed by many researchers [**19**, **20**]. The findings confirm that not only the content of scientific work, but also other, in part non-scientific, factors play a role in citing behaviour [**21**]. Results in researches suggest that authors have been motivated to cite publications depends on many other factors including time-dependent factors, field-dependent factors, journal dependent factors, author dependent factors and availability of publications. In our research, we did not consider any controversy between these factors; instead, we approach to the problem from another intellectual link to citing.

There is, however, sufficient evidence that the intellectual content of article is an important factor in citing. Here, we study links between citation to a particular work and semantic analysis of the citation context rather than the content concept. This analysis seeks to obtain a better insights into relationships between citing and cited works. The approaches described in this thesis analyse the semantics of texts for the purpose of characterising the cited works. The goal of empirical analysis of citation behaviour has been not only to reveal authors motivations for citing but also to improve the use of citation counts in research evaluation. We aim to understand inter-texts relationship in the selected sample by devising classification models based on an analysis of texts surrounding the citation groups.

To build a model that automatically evaluate the impact of articles, we adopt a work-flow of supervised machine learning with our approaches to informational semantic. We suggest a methodology in classifying highly-cited and less-cited papers in the certain categories selected. This analysis first requires criterion to distinguish between citation counts - definition of binary classifiers. Human has to decide this categorisation of papers. Having labelled classes, our interest is in proposing and performing classifiers and predictive models to discover how much information semantics of texts have in predicting the citation.

Last but not least, it is very important to understand that scientific corpus or thesaurus are not necessarily to appear in the same form. For example, the collection essay texts and students' answers in academic examinations are two other types of such corpus. Specifically, the collection of words in model answers can be considered as *thesaurus of teaching*. Each module, just like category, has its own keywords and academic success in these cases are measured by grades given by teachers.

Before introducing our methodologies, it seems very reasonable to analyse a small corpus of texts to get better insights into semantics of academic texts. Thus, in our research we initially analyse a corpus of student answers for the purpose of measuring keyword-similarity between the student answers and the model answer, and then evaluate the academic success of students through these similarities in short textual responses. The hypothesis is that, parallel to the idea in evaluation of impact of articles, vocabulary that students used to transmit information about the module in their answers are strong indicator for the performance of students. Computational methods are applied by standard BoW model where words are used with their module-specific description, and a new mathematical model is created for prediction of grades. Without getting into the details of complex semantic analysis, this shows a methodology to quantifying category-based (module-specific vocabulary in this case) meaning in short academic texts by keywords. This study draws on ability of computational methods in quantifying the scientific-specific meanings, even without complex semantic models and in-depth knowledge of semantics of answers.

The results are encouraging given the extremely small corpus size of documents for proposed approach of informational semantics. This pilot study solidified our understanding of how to analyse scientific texts, the importance of a large corpus, and how to approach further research to scientific-specific meanings improve and increase the applicability of the results of this pilot study.

Finally, let us wrap everything up. In short, this thesis develops computational techniques, English scientific corpus and dictionaries for one of the most difficult Natural Language Processing(NLP) tasks: quantifying the lexical semantics of scientific texts.

The contributions of this thesis are both methodological and practical. Several previous lines of work in semantic analysis, computational linguistics, NLP as well as machine learning and statistical models are woven together into our novel approach to meaning in scientific texts. Throughout this research, we have analysed a corpus of students' answer and developed a mathematical model to predict grades. We have developed an informational representation of semantics that can be used for quantifying of meaning in scientific texts, and shown that this model obtains superior performance to standard raw-frequency representation in identifying the scientific-specific meanings. We then have used informational semantics in evaluation of scientific impact of articles.

The main application areas of introduced methods in this thesis include but are not limited to NLP, corpus studies, computational linguistics and machine learning. Datasets and software produced in this research are accessible publicly for the benefit of further research on corpus studies, NLP and other related tasks.

1.3. Organisation of the Thesis

This thesis is structured as follows:

Chapter 1. In this chapter, we have already introduced some preliminaries and central problems that we will address to in the thesis.

Chapter 2. Chapter 2 presents an experimental study for automatic grading of short answer questions, and providing useful feedback on their answers to students. Using the corpus of students' answers, we apply standard data mining techniques to evaluate the students' marks through their answers' similarity to the model answer. We measure the similarity between a student answer and the model answer by counting the common words in two Bags of Words. We also study feedback mechanisms constructed based on groups of students who are awarded the same or similar grades. Students' answers are grouped into natural clusters by *k-means* clustering and words in each cluster are compared to show that these clusters are constructed based on how many and which words of the model answer have been used. This approach is determined to be used in automated feedback and grading mechanisms as follows: select one or more prototype mark (feedback) for each group, and assign it to the whole group at once. Finally, a mathematical model is developed to predict marks using the distance (or similarity) between the model answer and the student answer. We discuss the model and approaches with cases in which approaches can be effectively used, cases where grading and providing feedback can be fully automated and cases where human input is required.

Chapter 3. Chapter 3 begins with a review of the principals in corpus and dictionary design as well as standard text and word representation techniques in NLP. We then describe some of widely used and well-known analogue corpora and dictionaries for English to show major differences between them and our newly-created scientific corpus and dictionaries. In this chapter, the LSC, *Leicester Scientific Dictionary (LScD)* and LScDC are created and introduced to be used as the basis of the research in informational semantics and for further use in other tasks. All the pre-processing steps in the creation of the corpus and dictionaries are explained in detail as well as basic statistics, and the organisation of the LSC, LScD and LScDC.

Finally, in order to evaluate the core dictionary LScDC, we compare it with a classical academic word list *New Academic Word List (NAWL)* [22] containing words sampled also from an academic corpus. We analyse two dictionaries to measure the differences in terms of the common words and ranking of words. The comparison of the LScDC and the NAWL is performed with several approaches presented in the chapter.

Chapter 4. Chapter 4 contains a description and analysis of our novel vector space model developed for quantifying the meaning of words and texts. We deeply discuss 'meaning of meaning' from the antiquity definition till modern time in the act of communication. We also discuss the core ideas of the approaches in 'meaning' from different perspectives in fields ranging from psychology to linguistics, philosophy to computer science, with an examination of the areas of application and the different implementations presented previously in literature. We propose an approach to computational analysis of meaning for a large family of texts. This approach is applied to construct the *Meaning Space* based on the LSC and LScDC. We introduce the *Meaning Space*, in which the meaning of a word is represented by a vector of RIGs about the subject categories that the text belongs to. 252 subject categories of WoS are used in construction of vectors of information gains. This representation technique is evaluated by analysing the top-ranked words in each category. For individual categories, RIG-based word ranking is compared with ranking based on raw word frequency in determining the science-specific meaning and importance of a word. The *Word-Category RIG Matrix* for LSC with LScDC is created and described in this chapter, and be prepared to explore quantifying the meaning of a text in the next chapters. We finally create a scientific thesaurus, LScT, in which the most informative words are selected from the LScDC by their average RIGs in categories. LScT contains the most informative 5,000 words in the corpus LSC. These words are considered as the most meaningful words in science. We use the Word-Category RIG Matrix with the LScT in the study of the representation of the meaning of texts.

Chapter 5. In Chapter 5, we hypothesize and test that lexical meaning in science can be represented in a lower dimensional space rather than 252-dimensional space that we described in Chapter 4. We apply *Principal Component Analysis (PCA)* to reduce the dimensionality of the Meaning Space, in which points are 5,000 words of LScT and dimensions are categories. This chapter is meant to serve as an analysis of dimension of the Meaning Space and visualisation of words and categories in the space of PCs. In order to avoid redundant attributes in the data and identify the actual dimension of the space, we explore the Meaning Space by PCA. We begin the study with applying *Double Kaiser Rule* for the selection of important attributes, that is a subset of original set of attributes (categories). We show that none of the attributes is dropped after the first iteration of the Double

Kaiser selection, so all 252 categories will be retained. We then apply PCA and interpret the first five PCs by their coordinates. For each component, categories are divided into three groups defined as the main coordinates of the dimension and being unrelated attributes to the PC: categories that positively and negatively correlated with the corresponding component, and categories having near zero values in the component. We analyse the topics in these groups and visualise both categories and words in PC axes. We also analyse the extreme topic groups at opposite ends of the PCs in order to describe the PCs based on extremely influential categories at both ends (10 categories at both ends). We also introduce another approach based on approximation of vectors for determining the groups of categories which are influential for PCs, and analyse PCs based on this division. Finally, by using three different selection criteria (Double Kaiser, Broken Stick, an empirical method based on multicollinearity control – PCA-CN –), we reduce the dimensionality of the category space to 61, 16 and 13 respectively. Both the 252-dimensional original space and the space of reduced basis will be used in further research.

Chapter 6. Chapter 6 introduces a novel model for quantifying the meaning in texts; that is, extracting informational semantics of texts. The second part of the chapter focuses on showing how much information the semantics of scientific articles have in predicting of citation and developing a quantitative evaluation for scientific impact of articles. We describe the experimental framework used to evaluate the impact of scientific articles through their informational semantics. We begin this chapter with developing a methodology for representation of text meaning. From the holistic point of view as a combination of simple BoW model, we construct text representation using cloud of text's words – that are represented in MS – and the distribution of words' RIGs in each category (dimension). *Feature Vector of Text* is introduced and created for each text as a vector representation of the text meaning. We create FVTs in five different ways by combining the mean vector, the vector of the first principal component for each text and two centroid vectors obtained by k-means clustering of words in the text. Each of FVTs is defined as informational semantics representation and then analysed for evaluating the impact of papers as a binary classification problem: classifying highly-cited (H) papers and less-cited (L) papers in individual categories. Labels in two classes are assigned based on the average citation count in the category. We also differentiate between extremely highly-cited (EH) and extremely less-cited (EL) papers labelled according to lower and upper quartiles of citation counts in the category. The empirical analysis in predicting the citation is done on the basis of the LSC abstracts with citation counts extracted from the WoS website for approximately four years range from 2014. For experiments, we select three categories from three main branches of science: Applied Mathematics, Biology and Management. For both classification problems, we applied classifiers for five FVTs in three different spaces and compared performance of the classification.

Chapter 7. The closing chapter, Chapter 7, contains a summary of the study, discussion of results and suggestions for future work.

1.4. Dissemination of Findings

The dissemination of findings to research community is done via:

(1) Learning Analytics & Knowledge Conference (LAK20), March 23-27, 2020: Frankfurt, GERMANY – Fully virtually. *A note on Automatic Grading of Short Answers and Providing Feedback.*

(2) Süzen, N., Gorban, A. N., Levesley, J., & Mirkes, E. M. (2020). Automatic short answer grading and feedback using text mining methods. *Procedia Computer Science*, 169, 726-743.

(3) Suzen, N., Mirkes, E. M., & Gorban, A. N. (2019). LScDC-new large scientific dictionary. *arXiv preprint* arXiv:1912.06858.

(4) Suzen, N., Mirkes, E. M., & Gorban, A. N. (2020). Informational Space of Meaning for Scientific Texts. *arXiv preprint* arXiv:2004.13717.

(5) Suzen, N., Gorban, A. N., Levesley, J., & Mirkes, E. M. (2020). Principal Components of the Meaning. *arXiv preprint* arXiv:2009.08859.

Datasets. The datasets used in this research are available in the University of Leicester Figshare repository and can be accessed by:

(1) Suzen, Neslihan (2019): LSC (Leicester Scientific Corpus). University of Leicester.Dataset.https://doi.org/10.25392/leicester.data.9449639.v2

(2) Suzen, Neslihan (2019): LScD (Leicester Scientific Dictionary). University of Leicester. Dataset.https://doi.org/10.25392/leicester.data.9746900.v3

(3) Suzen, Neslihan (2019): LScDC (Leicester Scientific Dictionary-Core). University of Leicester. Dataset.https://doi.org/10.25392/leicester.data.9896579.v3

(4) Suzen, Neslihan (2020): LScDC Word Clouds and Tables to Visually Present the Most Informative Words in Subject Categories. University of Leicester. Figure.https://doi.org/10.25392/leicester.data.12191604.v1

(5) Suzen, Neslihan (2020): LScDC Word-Category RIG Matrix. University of Leicester. Dataset. https://doi.org/10.25392/leicester.data.12133431.v2

Codes. The codes produced in this research are published in https://github.com/neslihansuzen.

CHAPTER 2

Automatic Short Answer Grading and Feedback Using Text Mining Methods

Automatic grading is not a new approach but the need to adapt the latest technology to automatic grading has become very important. As the technology has rapidly became more powerful on scoring exams and essays, especially from the 1990s onwards, partially or wholly automated grading systems using computational methods have evolved and have become a major area of research. In particular, the demand of scoring of natural language responses has created a need for tools that can be applied to automatically grade these responses.

In this chapter, we focus on the concept of automatic grading of short answer questions such as are typical in the UK GCSE system, and providing useful feedback on their answers to students. We present experimental results on a dataset provided from *the introductory computer science class* in the University of North Texas. We first apply standard data mining techniques to the corpus of student answers for the purpose of measuring similarity between the student answers and the model answer. This is based on the number of common words. We then evaluate the relation between these similarities and marks awarded by scorers. We consider an approach that groups student answers into clusters. Each cluster would be awarded the same mark, and the same feedback given to each answer in a cluster. In this manner, we demonstrate that clusters indicate the groups of students who are awarded the same or the similar scores. Words in each cluster are compared to show that clusters are constructed based on how many and which words of the model answer have been used. The main novelty in this study is that we design a model to predict marks based on the similarities between the student answers and the model answer.

We argue that computational methods be used to enhance the reliability of human scoring, and not replace it. Humans are required to calibrate the system, and to deal with situations that are challenging. Computational methods can provide insight into which student answers will be found challenging and thus be a place human judgement is required [23].

2.1. Introduction

In the learning process, the assessment of knowledge plays a key role for effective teaching [24]. With the range of assessment methods available, examinations have dominated the assessment of student learning. In particular, knowledge and understanding in academic courses are assessed by end of course examinations combined with coursework. Academic examinations can be performed using many question types from multiple choice questions to free responses. Manual assessment is much more difficult for question types such as short answer and essay questions [25]. Students are required to give *free text responses* for these question types, so each response requires textual understanding and analysis as opposed to grading answers with a single correct answer such as multiple choice tests. Hence, manual scoring takes a considerable amount of time, and provision of meaningful feedback even more so.

Manual scoring of those answers can suffer from inconsistency since the marker must infer meaning from the candidates' own words. Scores on the same answer may vary from marker to marker. However, free text questions are a widely preferred assessment tool, used throughout the learning process, due to their effectiveness on developing cognitive skills of students and also demonstrating knowledge in short texts [26]. Therefore, there is a need to develop tools for mitigating these challenges in assessment. One approach is to create automatic scoring tools and feedback mechanisms that supports markers. This methodology was discussed by several researchers, especially from the 1990s onwards, as computational techniques became applicable in this field. A number of studies have been made to automate short answer grading [27, 28, 29].

Assessment of natural language responses is a challenging task since we can not expect a machine to understand free text answers. However, developments in NLP have made partially or wholly automated scoring of exams possible. Automatic grading has become a popular field among researchers due to its benefits on reducing human mistakes and time spent (see Fig. 2.1).

Automatic assessment can also be considered for pre-assessment of student works by giving them support to improve their work. A student could submit their work to a feedback system and be given information such as: "Answers like this scored this mark. In order to improve this you might ...". By tracking student use of such a system and we also have the possibility of tutor understanding of what is being misunderstood in class, with the opportunity for face to face mitigation.

This research presents both supervised and unsupervised approaches that deal with the automatic marking and feedback for short answers. The proposed models are based on the concept of similarity between the model answer and student answers, and the discovery of the structure in the corpus of student responses. Our initial assumption is that given a set of student answers, marks awarded by scorers are highly dependent on words that the student used which also occur in the model answer. This is because in practice, grades are often awarded based on how similar a student response is to the expected answer (model answer). Using these similarities, we intend to build our model to mark student work.

We will see that this approach is more appropriate for some questions than others, but that questions that the algorithm finds hard to mark are also difficult for humans; the variation of marks of two human markers is higher.

The model calculates the *distance* between the model answer and the student answer using words in the model answer to generate scoring rules. This has the additional benefit of identifying the misconceptions and weaknesses of students for a topic under the assumption that those students who have a lack of knowledge are not able to use all of the words (or synonyms) in the model answer. Depending on the similarities between answers, automatic feedback can be given to make students aware of their level of understanding. Such a system can be also used as a supervised learning process to predict an answer's score and automate marking by using a training set (see A.2 for details about such approaches).

In the learning process, providing feedback to student also plays a key role, as it helps students understand the subjects and improve their learnings and self-awareness. Our second approach focuses on an *automatic feedback mechanism*. We cluster student answers into groups to explore whether or not responses given by students share similar characteristics. This approach can uncover natural groups of answers having similar structures that students frequently use. We find clusters of similar answers, and then to evaluate elements of these cluster using both human and computational means. Thus we will provide teachers with information about the common answers that students give, since students generally answer questions in similar ways.

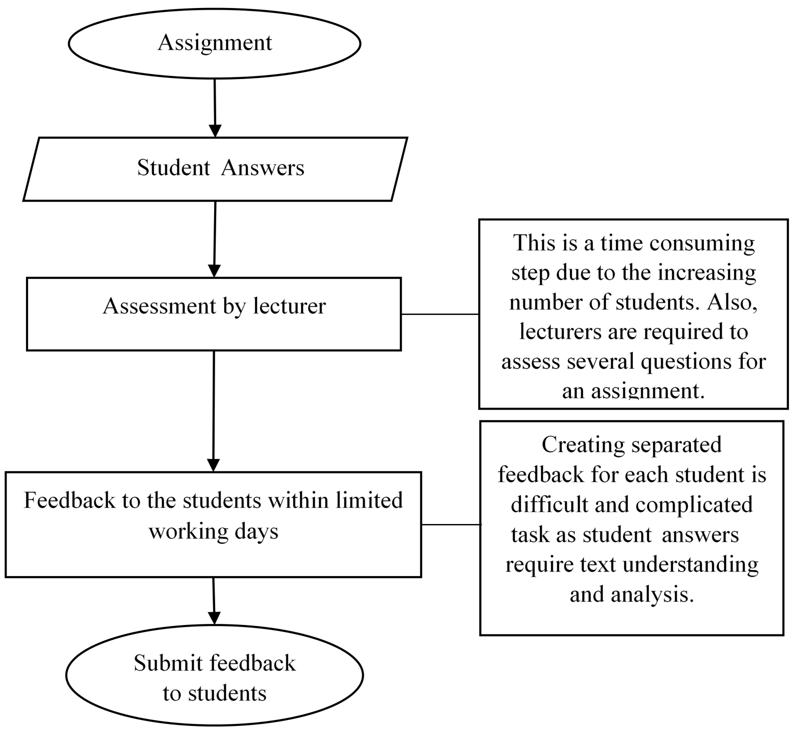


FIGURE 2.1. Steps and challenges in manual assessment

The main advantage of this approach is that a new system can be built that teachers can give a common feedback to students belonging to the same cluster [30]. This can be done by selecting a prototype answer(s) from each cluster. Grouping allows teacher to provide feedback for the prototype answer, and this feedback can be assigned to the entire cluster at once. It could be also used in a feedback system in which students submit answers with historical feedback, and use this feedback to improve their answers for final submission of assignments. Further feedback on student behaviour could be garnered from student use of such a system.

Grouping similar answers together can be implemented in the automatic marking of groups. This approach can significantly cut down on the time for manual marking, and improve consistency in marking and feedback. In particular, once groups of answers are identified, the system assigns a common score to whole group by using human marking. This grouping process will not be 100% reliable and there may be answers which are more difficult to classify. In this case human intervention is necessary. It is not our desire to remove the human from the marking process, just to improve consistency and allow humans to apply judgement in the difficult cases.

Finally, we develop a model to predict marks by using distances between the model answer and the student answers. We hypothesise that marks can be predicted by this distance between student answers and the model answer as marks are highly correlated with this distance. The objective of this approach is to show how distances from model answer can be used to mark student answers. In the results section we will see that the model lies close to the average of the scores of the two markers.

In the following section, we start with a detailed historical analysis on the automatic assessment of short answers. We describe the data set used, and then report an experimental study using supervised and unsupervised learning. We conclude by describing our methodology and results, and future work.

2.2. Related Work

The earliest study of automated grading dates back to the work of Page [31]. In this article, Page introduced his ideas on computational methods for grading student essays, and also on prospective roles of computers for grading. He experimented with automatic evaluation of student essays with 276 essays written by high school students. His idea is based on correlation between basic characteristics of the essays and grades assigned by four teachers [32]. For each essay, overall quality is evaluated by the adding of the ratings by these teachers. To calculate approximated values of these actual ratings by computer, accessible variables to a computer are identified by teachers.

The results show that the computer grades are not distinguishable from human grades. Page states that, in the future, the computer-based judge will be better correlated with each human judge than the other human judges are. He also observes that such successful results from computerized grading may lead to the possibility of automated grading systems for the evaluation of essays. Finally, it is important to stress here that he also outlined the idea of 'giving feedback', suggesting that a computer print-out could suggest that the student correct identified misspellings, syntax mistakes, and the overuse of certain words.

Since his study, automated grading of free text responses has become a popular area with the focus on marking essays rather than short answer questions. However, more significant studies have been done since the start of the 1990s, as computational techniques and software technology have become more powerful [33]. The most well-known essay assessment systems are Project Essay Grade (PEG), e-rater, and Intelligent Essay Assessor [34]. In this study we are concerned with automatically marking of short answer questions, and therefore we present some similar systems for short answer grading in detail, rather than essay marking.

C-rater is a scoring engine designed to grade short content-based answers [27, 35]. The goal of c-rater is to match teacher with students answers in terms of their concepts. It is modelled to identify paraphrases of model answers as correct answers. These paraphrases are built by normalising a variety of responses related to four primary sources of variation among sentences: syntactic variation, pronoun reference, morphological variation and synonyms, and also variation caused by spelling error. Once c-rater matches the concepts found in the student answer with those found in the model answer(s), it assigns scores based on the number of matched concepts. It is reported that c-rater reached 84% agreement with human graders for the scoring of reading comprehension responses. In their own words, it is stated that "if the teacher uses the same question for several classes or over several semesters, then the advantages of the initial effort are worthwhile".

Similar work to ours has been carried out on the grading of short answers in [28, 36, 37, 38, 39, 40]. These grade answers by based on the similarity between the model answer and the student answer. Related studies can be found in [24, 41].

The idea of grouping similar responses has been introduced, in parallel to our approach, in [30]. In this work the approach is referred as *Powergrading* due to its amplification of human effort for scoring. It is designed to group (and subgroup) responses by clustering techniques in order to make scoring partially automated. The proposed approach is based on the idea of using both human and the machine ability to score, under the assumption that groups of similar answers can be quickly marked by a human by considering the whole group at once. They also aim to discover patterns of misunderstanding among students, then to give comprehensive feedback to student answers in the same cluster.

Similar work is described by [42], where the approach relies on the parallel assumption that similar grades can be assigned to groups of similar answers. They use a clustering algorithm to create groups of answers and then assign a single grade to the whole cluster. Specifically, they used 1,668 short answers to 21 questions, with sample solutions and grades assigned by teachers, from a listening comprehension task for German language learners.

In their study, the extraction of features has been done by word n-grams, character n-grams and keywords [43]. The cosine similarity between feature vectors and the centroid of each cluster is calculated, and those items which are the most similar to the centroid (with highest similarity value) are grouped into the cluster. For labelling of items in clusters, three different methods are evaluated for selecting the optimum response to be labelled: random item selection from each cluster, selecting the closest item to the centroid of the cluster, and selection of one item belonging to the majority label of the cluster. When item selection methods are compared, the results show that selection of the closest item to the centroid leads to higher accuracy than random selection methods in general.

2.3. The Data and Features

The data set we use is from the *introductory computer science class* in the University of North Texas[1]. It consists of 29 student answers from ten assignments and two exams. For each question, there is one model answer in the data. The model answers are mostly one sentence, but some of them contain a single word sentence (see example in Table 2.1). In this section we will give these questions our own numbers, which differ from those in the original study. We give further examples of questions from this study in A.1. We show examples of two questions, 1 and 2, for which our scoring model is reliable, and two questions, 3 and 4, for which it is less reliable. This correlates with teacher difficulty in marking. Further study is needed into what sort of questions teachers can reliably score, and which are more challenging.

We define the *model vocabulary (or vocabulary)* for a question as the collection of words in the model answer. Table 2.2 shows an example of model vocabulary and student answers containing words from the vocabulary. The student answers are manually scored by two teachers. Grades are given between 0 and 5, 0 for incorrect answers, 5 for correct answers, and from 1 to 4 for partially correct answers. We consider each grade as an individual label. For our intended approach, we used these labels to evaluate how words and marks are correlated and to see which marks occur in each cluster. There are discrepancies in the grades from the two teachers, and we use both grades for creating our model.

Before starting the analysis of the responses, we initially applied preprocessing steps to remove irrelevant characters (e.g. numbers, punctuation). We also remove

[1]The dataset used in the project was downloaded from the archive hosted at the URL http://lit.csci.unt.edu/index.php/Downloads. It is available at the following link https://github.com/dbbrandt/short_answer_granding_capstone_project/tree/master/data/source_data/ShortAnswerGrading_v2.0 (accessed 6 September 2020)

TABLE 2.1. Examples of questions and model answers with different lengths in the data.

Question 1	What is the role of a prototype program in problem solving?
Model Answer 1	To simulate the behaviour of portions of the desired software product.
Question 2	What is the stack operation corresponding to the enqueue operation in queues?
Model Answer 2	Push

TABLE 2.2. Examples of model vocabulary and student answers.

Model Answer 1	To simulate the behaviour of portions of the desired software product.
Model Vocabulary 1	simulate, behaviour, portion, desire, software, product.
Student Answer	High risk problems are address in the prototype program to make sure that the program is feasible. A prototype may also be used to show a company that the **software** can be possibly programmed.
Student Answer	it **simulates** the **behavior** of **portions** of the **desired** **software product**

so-called *stop words* such as *the, and, it* from both model answer and the student answers in order to improve computational efficiency. Additionally, stemming is performed to combine different versions of words into a root. For instance, *eat, eating, eaten* are all the same word as far as our method is concerned. This preprocessing step significantly reduces the number of words we deal with.

Finally, when applying our clustering algorithm, we performed some user defined pre-processing steps to reduce the high dimension of 163 (163 individual words). As almost all students used words appeared in the question sentence, we treat these words as stop words, and removed them from student answers. In addition, we observed that some words appear in only one or two answers. These words have no discriminative power for clustering and leads to sparse vectors in the algorithm. Sparsity refers to vectors with 0 frequencies in most of their inputs. We also removed those words appears in less than 10% of answers in the corpus. After all preprocessing steps, there are 20 individual words in the collection of student answers, that is, the dimension of the vector space has become 20.

To show the results here, we choose the first question as it represents the average answer length, with one sentence, and also unique answers from each student. Throughout the study, all analyses have been done by using this question, the model answer of the question, and 29 student answers.

2.4. Data Representation and k-means Clustering

A starting point for applying data mining tools to unstructured text data is to transform the text into an appropriate set of data [44, 45]. In other words, text representations of a collection should be converted into numeric vectors (feature vector form) to be able to apply statistical methods on the data.

In our study, we used the BoW model for representation of texts. In this model, each document (answer in our case) is a collection (bag) of words. The idea of the BoW model is to extract unique words from the collection of documents, and to treat these words as individual features. Each document is represented as a vector of word frequencies. Since the frequency of a word increases as the number of appearances of a term increases, this shows us how important a word is for a document. This representation is called *term frequency (TF)*, which represents the relevance of a term to the corresponding document.

In the vector space method, documents are points of a high dimensional space, where each dimension (each feature) corresponds to one word. Each element of a vector indicates the position of a document in a particular dimension. So, distance measures tell us how far two points are in the vector space, i.e., the distance between two documents. In our work on clustering, we perform one of widely applicable distance measure: Euclidean distance (the sum of squares distance).

Suppose we have N features in our vector space and $\mathbf{r_1} = (r_{11}, r_{12}, \cdots, r_{1N})$ and $\mathbf{r_2} = (r_{21}, r_{22}, \cdots, r_{2N})$ are the representations of two documents in our vector space. Then the distance between these documents is

$$d(\mathbf{r_1}, \mathbf{r_2}) = \sqrt{(r_{11} - r_{21})^2 + (r_{12} - r_{22})^2 + \cdots + (r_{1N} - r_{2N})^2}.$$

Note that as longer answers have more words than short answers, the number of non-zero entitles of features and also frequencies may be more compared with shorter answers in the vector representation. This does not mean the longer answer is more relevant. To adjust the effect of length, term frequencies should be normalized. Note that before implementing the clustering algorithm, the data is normalized in order to convert the frequency of terms to a common scale which allows for comparison. Given an answer $\mathbf{r}$, the L_2-normalization is defined as

$$\hat{\mathbf{r}} = \frac{\mathbf{r}}{d(\mathbf{r}, \mathbf{0})},$$

since the length of a vector is its distance from the origin. Given these normalised vectors, we run a classical clustering algorithm, *k-means*, in order to cluster student answers.

The k-means algorithm is one of the most popular clustering methods. The theoretical and algorithmic aspects have been studied by many researchers [46, 47, 48, 49]. The idea behind this method is that all points in a vector space are separated into k clusters, and each cluster is represented by its centroid vector (the sum of the vectors divided by their number). After defining k centroids, each document is assigned to a cluster by using the distance d (for us the Euclidean distance). Then, the centroids are recalculated until we find an optimal set of clusters based on some criterion function [50].

Suppose that we have, after the ith iteration a set of k clusters, and the jth cluster has N_j points in it $\mathbf{r}^i_{j,\ell}, \ell = 1, 2, \cdots, N^i_j$. The distortion function is defined as

$$E_i = \sum_{j=1}^{k} \sum_{\ell=1}^{N^i_j} d(\mathbf{r}_{j\ell}, c_j)^2,$$

where

$$\mathbf{c}_j = \frac{1}{N^i_j} \sum_{\ell=1}^{N^i_j} \mathbf{r}_{j\ell},$$

is the centroid of the jth cluster [51, 52].

The algorithm seeks a partition of data set by optimizing the error criterion. The steps of k-means are as follows [53] :

- Begin with randomly k-partition and calculate centroids for each cluster (initial centroids).
- Assign each data point to nearest cluster (nearest-neighbour rule) by calculating the distance to the nearest centroid.
- Re-calculate the centroids based on the current partition.
- Repeat the second and the third steps until there is no change between two iterations, that is, until the algorithm has converged.

We expect that E_i will decrease as the algorithm proceeds (i increases) until it converges to some minimum value.

2.5. The Marking and Feedback System

Our aim for this research is to design a model of automatic short answer marking and feedback. In this section, we will present an experimental study to demonstrate the processes behind the model creation. We begin with evaluating the dependency of scores and the similarity between the model answer and the student answer. Then, we will present our approach based on the clustering algorithm.

In the data used, not all of students are scored the same grades by the two graders. There are some cases where we observed inconsistency of grading. Fig. 2.2 shows the distribution of scores awarded by the two teachers. The size of points indicates the proportion of grades for the corresponding pairs (Teacher 1 grade, Teacher 2 grade). As can be seen from the figure, there are 2 locations where the points are concentrated. The biggest proportion of scores comes from the pair (5,5) from both teachers, followed by the pair (2,2). This means that the teachers tended to agree with each other in grading for the most part (17 out of 29 students), especially when the answer is very good or very poor. Otherwise, there is inconsistency in scoring for some student answers. The correlation (Pearson correlation) between the two teachers' marks is 0.82 with an error of 0.1 (Mean Squared Error).

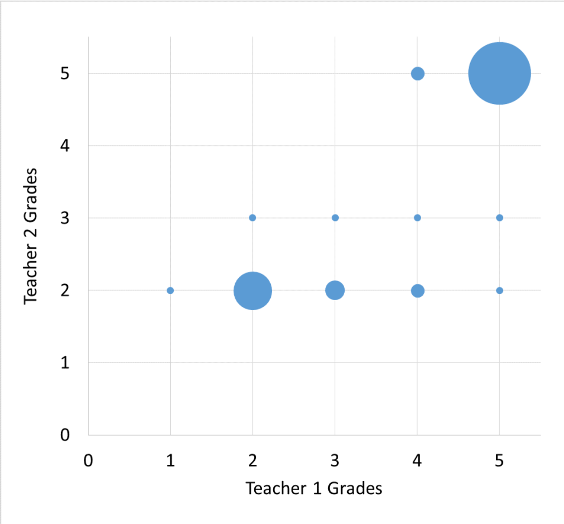


FIGURE 2.2. The distribution of scores awarded by two teachers. Size of points is proportional to number of cases, i.e, the proportion of pairs (teacher 1 grade, teacher 2 grade) in the dataset.

Note that the lowest and the highest grades given by teachers are 2 and 5 for this question, respectively. We suspect that those students graded at the lowest score did not use appropriate terminology for the subject. Similarly, those student graded highest answered used appropriate words. In between, they may use some of words required but the answer is not totally appropriate. We will now investigate why and how scores change depending on the words used.

2.5.1. Unsupervised Learning for Clustering of Student Answers

We now turn to looking at natural clusters of student answers to discover if there are some patterns of answers that students frequently use. If there are some groups in student answers, we can group them to give the same feedback or marks. Recall that we have treated words that appear in the question as stop words, and have removed these words from student answers (most of students include these words in their answers).

In Fig. 2.3, all words that the students used in their answers can be seen with their frequencies. The more a word appears in the collection, the larger the font that is used in the word cloud; these are emphasised by the use of different colors. For instance, the most frequently used word is "program". This is an expected result since this word is contained in the question sentence; most of students started their answer with this word.

To group answers, we use k-means clustering as described in Section 2.4. The optimal number of clusters is determined as three using the Elbow method [54]. In Fig. 2.4, we show these three clusters with the associated grades awarded by the teachers. As can be seen from the tables, two clusters are well separated in terms of

FIGURE 2.3. All words in the collection of student answers. The font and color of words indicate different frequencies of words.

Excellent		
Student No	Teacher 1 Grade	Teacher 2 Grade
2	5	5
3	5	3
4	5	5
10	5	5
11	5	5
12	5	5
17	5	5
21	5	5
22	5	5
25	5	5

Mixed		
Student No	Teacher 1 Grade	Teacher 2 Grade
1	4	3
5	3	3
7	3	2
8	5	5
14	4	5
16	4	5
19	2	2
20	3	2
23	1	2
24	2	3
27	4	2
28	4	2

Weak		
Student No	Teacher 1 Grade	Teacher 2 Grade
6	2	2
9	5	2
13	2	2
15	2	2
18	2	2
26	2	2
29	3	2

FIGURE 2.4. The structure of clusters with original marks awarded by two teachers.

scores contained. We introduce clusters: *Excellent, Mixed* and *Weak* depending on the marks in the cluster. Student marks in the cluster *Excellent* (shown as green) are 5, so we expect that these student answers are similar in terms of usage of the appropriate terminology. Similarly, answers in cluster *Weak* (yellow) are expected to have inappropriate terminology since the marks are 2. However, we cannot identify a scoring rule for the cluster *Mixed* (blue), and further study based on keywords is required for this case. Note that there is also discrepancy in scores graded by two teachers in this cluster, so they are finding it challenging to score responses in this cluster.

We now wish to examine what words are used by students in each group to understand the structure of groups. The frequently used words are displayed by word clouds; see Fig. 2.5. As expected, students in cluster *Excellent* used all of words from the model answer, while there does not exist any words from the model answer in the cluster *Weak*. When considering marks in these clusters, there are

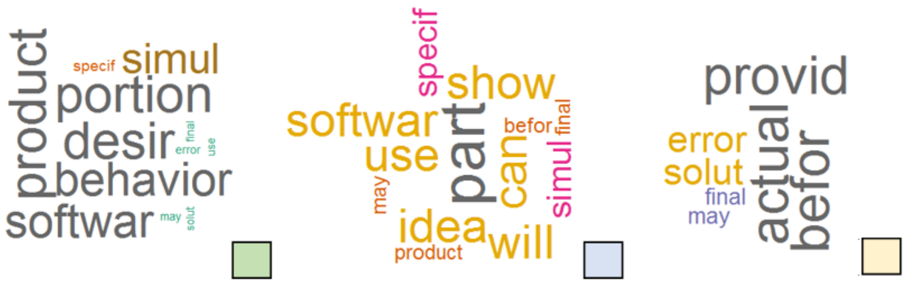


FIGURE 2.5. Frequent words in each cluster. Colors in the bottom right hand corner show clusters (in the Fig. 2.4) that the words belong to.

two marks: 5 and 2, respectively. This means that when students use all or none of the vocabulary words we are able to easily separate clusters for these students.

In addition, there are a few words from the vocabulary in the cluster mixed (software, simulate). We also see that some words such as 'part' and 'final' are synonymous with 'portion' and 'desired' respectively. In this cluster, marks are spread with the highest value 5 and the lowest value 1 producing a range of 4. So, this shows that marks change depending words used in the cluster. We also note that a student has scored 5 from both teachers, but is not in our *Excellent* cluster. This shows that a better knowledge of acceptable vocabulary (synonyms) is needed in order to cluster more effectively.

2.5.2. Similarity between the Model Answer and the Student Answer

Given that we can cluster responses and that the clusters correlate well with marks given, we wish to show how marks depend on the words used by students. We consider this problem as a basic calculation, where each answer is based on one model answer and two labels from two teachers. When we extract the significant words from the model answer we see that they number six (see Table 2.2).

We use the Hamming distance [55] $h(\mathbf{r}, \mathbf{m})$ to measure the distance between student answer $\mathbf{r}$ and model answer $\mathbf{m}$. We count the number of words n that the students use which appear in the model answer. Then $h(\mathbf{r}, \mathbf{m}) = 6 - n$, so that if all the words are used the distance is 0, and if none of the words are used, the distance is 6.

In a more sophisticated implementation we might need to look for synonyms of the words in the model answer, but for this study we are interested only in demonstrating the idea. The results are shown in the Fig. 2.6. We see that scores decrease for both graders in the main, in a regular way as the Hamming distance increases. It is also important to note that the scores are 5 when the distance is 0, and the majority of marks are 2 when the distance is 6. For other cases, we observe that the mark changes depending on the subset of words the student uses in their answer. Therefore, the level of importance of model words varies differently. Some

Student No	Teacher 1 Grade	Teacher 2 Grade	Distance
3	5	3	0
4	5	5	0
10	5	5	0
11	5	5	0
12	5	5	0
17	5	5	0
21	5	5	0
22	5	5	0
25	5	5	0
2	5	5	2
8	5	5	4
14	4	5	4
16	4	5	4

Student No	Teacher 1 Grade	Teacher 2 Grade	Distance
1	4	3	5
5	3	3	5
27	4	2	5
28	4	2	5
7	3	2	6
19	2	2	6
20	3	2	6
23	1	2	6
24	2	3	6
6	2	2	6
9	5	2	6
13	2	2	6
15	2	2	6
18	2	2	6
26	2	2	6
29	3	2	6

FIGURE 2.6. Distances between the model answer and student answers, and original marks for each student. Different colors indicate groups of students with different distances.

of words may contribute more for marking. Teachers need to be able to set the importance of words for scoring for any automated system.

We also computed the Pearson correlation coefficient to check the correlation between the distance and mark given, and found the correlation (Pearson correlation) -0.81 and -0.83 for two teachers with error 0.09 and 0.1, respectively. As expected, there is high correlation between marks given and the distances. In other words, we found that teacher assessment depends highly on how many of the vocabulary words students used in their answers.

2.6. Supervised Learning - A Model to Predict the Mark

In the previous sections we have demonstrated that the vocabulary used by students can be used to cluster their responses, and that the scores given by teachers are strongly related to the particular cluster that the student answer is assigned to. In this section we create and evaluate a model to predict student marks, based on the Hamming distance between the model answer and the student answer. We hypothesise that this distance is a strong indicator of the mark of a student, which suggests the possibility of automated scoring of responses.

The relationship between the distance and the mark is modelled using the following predictor function

$$y = \beta_0 + \beta_1 h^{\beta_2}, \tag{1}$$

where β_0, β_1, β_2 are parameters, and h is the distance between the model answer and the student answer. As shown in the Section 2.5.2, the distance can be 0, 2, 4, 5 or 6 for the question.

TABLE 2.3. Estimation of parameters and the MSE for MM, $i = 1, 2, 3, 4$, for the corresponding questions (see A.1); MSE is the mean square deviation of predicted marks from the average of two markers' grades.

Model	β_0	β_1	β_2	MSE (MM)	MSE (TM)
MM (1)	4.91085	-0.0058	3.42359	0.17440	0.25000
MM (2)	5.00100	-0.00074	4.08469	0.37616	0.36290
MM (3)	5.03413	-0.18210	1.92500	0.88490	0.96667
MM (4)	4.54003	-0.77408	0.40847	1.12573	0.43103

We fitted our model (1) to the data by minimising the Mean Square Error of prediction (MSE) to find the optimal values of parameters [**56, 57**]. As there are inconsistencies in grading by the two teachers, we decided to use the average of their two grades as the actual value of the dependent variable. We call this teacher mark TM. Let us call this mathematical model MM to distinguish it from TM.

The estimates of parameters β_0, β_1, β_2 and the MSE(MM) are presented in the Table 2.3. The value of β_0 represents the overall position of the curve on the y-axis; that is, the maximum mark which can be predicted by the model MM (when the distance is 0). We found a maximum mark of 4.91. The negative sign of β_1 tells us that scores decrease with distance. The value of β_2 indicates how quickly the score decreases with distance from the the base line β_0.

To compare the accuracy of our model with human marking, we calculate the deviation of teacher grades from the average mark of two teachers. This approach is shown as TM (teacher marks) in the Table 2.3. We see that the the teacher marks diverge from their average more than does the mathematical model for this question. This is unsurprising as it suggests the average is more predictable than the individual scores.

Fig. 2.7 shows the observed data, the average marks of two teachers and the mathematical model versus distances from the model answer. In the figure, we see that for this question we have a line that mirrors the average of the teacher scores well. As expected, the mark is decreasing function of distance in general. Both models demonstrate qualitatively the same behaviour. To compare MM with TM model we graph the errors as a function of distance. These are depicted in Fig. 2.7 (b). We see that the accuracy of human marking depends on distance from the model answer. The disagreement between teachers increases significantly with the distance, until we regain agreement and accuracy for poor answers. In this graph, we also look at the difference between these two predictions, and see it is small.

When the distance from the model answer is relatively small the MM performs well, with similar errors to the TM. On the other hand, MM outperforms TM model for big differences between the model answer and student answer (with distance 5 and 6).

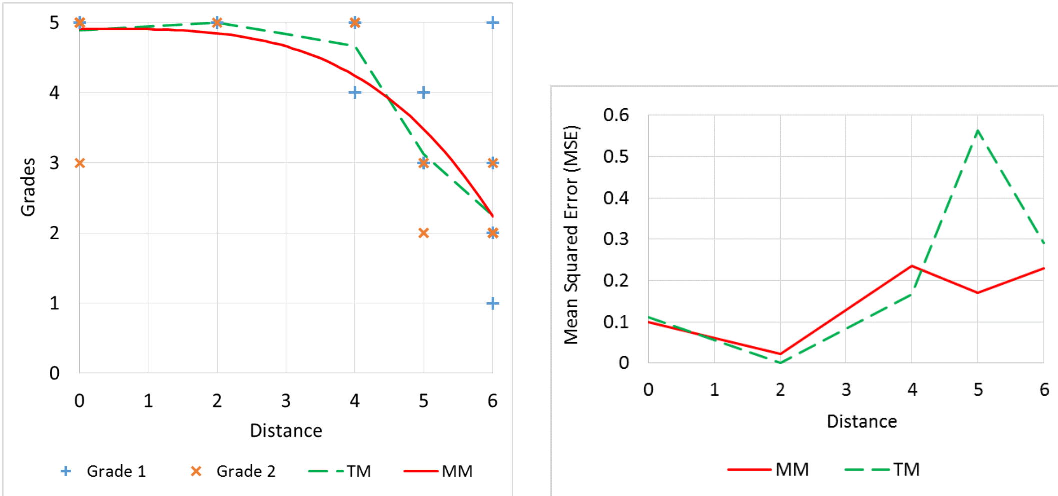


FIGURE 2.7. (a) Distribution of actual marks of two teachers (Grade 1 and Grade 2) with distances from the model answer, the average marks of two teachers (TM) and the predicted marks by the mathematical model (MM); (b) The mean squared errors for each distance from the model answer for MM and TM.

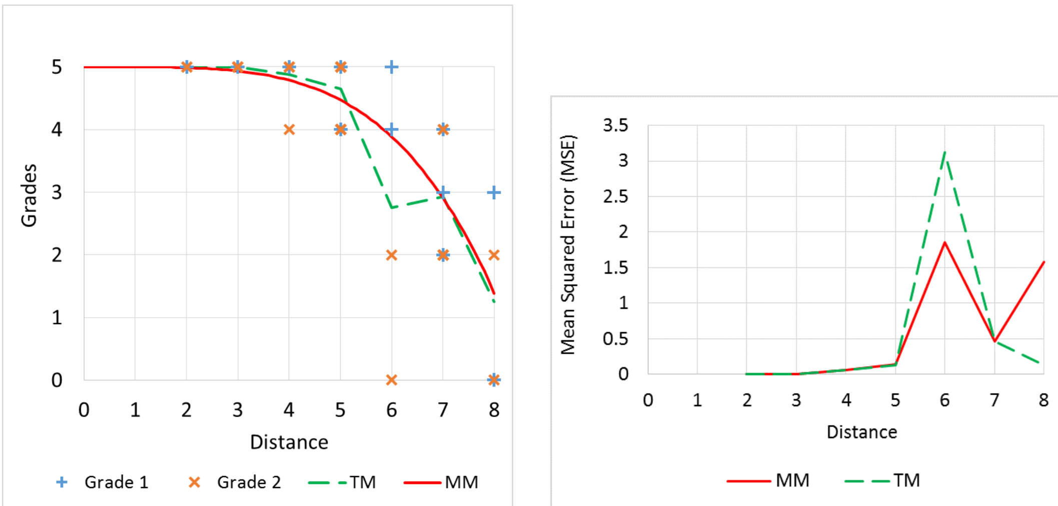
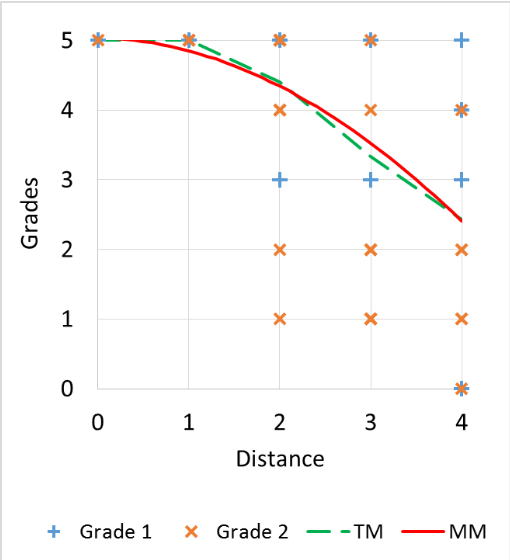


FIGURE 2.8. (a) Distribution of actual marks of two teachers (Grade 1 and Grade 2) with distances from the model answer, the average marks of two teachers (TM) and the predicted marks by the mathematical model (MM) for the second question; (b) The mean squared errors for each distance from the model answer for MM and TM for the second question.

Table 2.3 shows that the accuracy of the prediction of marks by MM is 0.17, and the estimation accuracy of TM is 0.25. This together with the Fig. 2.7 suggests that MM is no worse than human marking.

We have performed a similar analysis on three other questions to demonstrate that some questions are harder to mark than others.



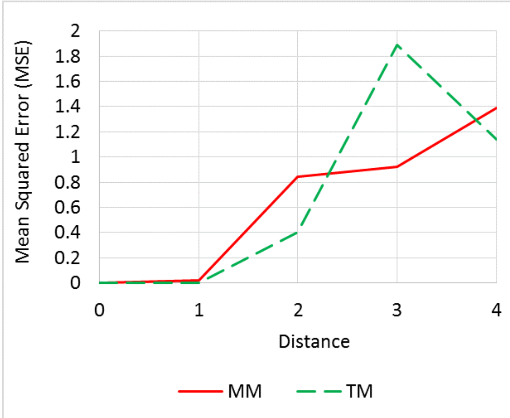
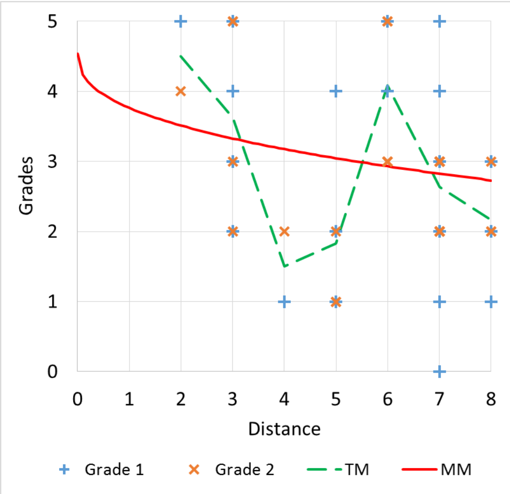


FIGURE 2.9. (a) Distribution of actual marks of two teachers (Grade 1 and Grade 2) with distances from the model answer, the average marks of two teachers (TM) and the predicted marks by the mathematical model (MM) for the third question; (b) The mean squared errors for each distance from the model answer for MM and TM for the third question.

We measure this by the extent to which the model mark deviates from the average of the teacher mark, and the extent to which teacher marks are a function of the distance of the answer from a standard word set. For Question 2, the corresponding results are shown in Fig. 2.8, and show that for this question automated scoring agrees well with teacher scoring.

Question 3 is an example which is harder to assess. In Fig. 2.9 we can see that our prediction is close to TM with a slightly bigger error (MSE) for TM. However, the figure shows that marks vary from 1 to 5 for distances 2, 3 and 4. This contradicts to our assumption that score is a function of distance. Therefore, disagreement of teachers on marking leads to inefficient scoring by the model MM.

Another example is shown in Fig. 2.10. We clearly see the absence of trend in original marks for this question (Question 4). The mark is not a decreasing function of distance. This contradicts to the assumption that as distance from the model answer increases, the mark given decreases. In this case, the model MM does not give reliable results. Another sign that MM cannot be effectively used for such examples is the value of β_2. In MM, we expect $\beta_2 > 1$ and small $|\beta_1|$. With $\beta_2 < 1$ we have a change in the shape of the curve.



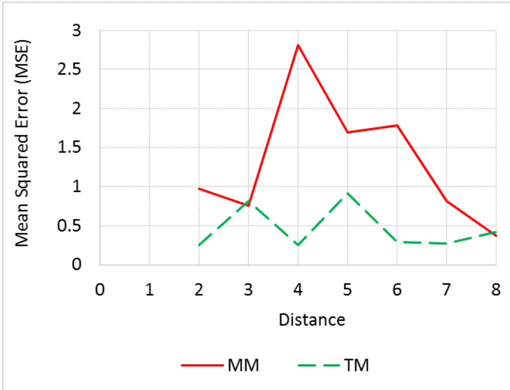


FIGURE 2.10. (a) Distribution of actual marks of two teachers (Grade 1 and Grade 2) with distances from the model answer, the average marks of two teachers (TM) and the predicted marks by the mathematical model (MM) for the fourth question; (b) The mean squared errors for each distance from the model answer for MM and TM for the fourth question.

2.7. Conclusion and Discussion

This study sets out with the aim of developing a model for the automatic grading of short answer questions, and providing useful feedback to students. Experimental studies are presented to show how our approach succeeds with the automatic scoring of short natural language responses. Our methodology works well where a vocabulary for the model answer can be clearly identified. This can be automated in some cases, but may require human input.

Strong correlation of grades and Hamming distance from the model answer are found. We demonstrate correlation of 0.81 and 0.83 between this distance and grades of two human markers. The MSE of linear regression for human marks are 0.09 and 0.1 for the two markers. This suggests that teacher marks can be predicted by distance with high accuracy. We conclude that the number of correctly used words has more influence on marks than semantics or order of words. In particular, if a large number of responses are being graded, it is not unreasonable that a human would move towards pattern recognition via key words rather than "reading for meaning". Note that our system mirrors human scoring, and knows nothing about correct scoring. Hence, identifying words gives an idea about grades and students misunderstanding to teachers. Such an approach allows time saving for scoring, and to provide rapid feedback to students by checking the words used from model vocabulary.

We developed a model to predict marks using the distance between the model answer and the student answer. The proposed model has the form (MM), where parameters were estimated by minimising the mean squared error between the actual mark and the predicted mark. The average marks of two teachers are used as observed marks since there is discrepancy of grades from the two teachers. The error of the prediction by the mathematical model (MM) is 0.17. The estimated accuracy of human grading (TM) is calculated as the mean deviation of actual marks from the average of the marks of the two graders, found as 0.25. MM has a lower deviation from the actual marks than TM.

We also consider a different approach for the automatic grading of short answer questions. This approach searches for natural groups of student answers. For each group, we can select one or more prototype mark (feedback), and assign it to the whole group at once. Clusters are found using k-means clustering. We note that we created clusters using word frequencies, without the use of teacher marks.

The grades inside clusters indicate that our approach effectively identified groups containing the highest (5) and the lowest (1) grades. Analysis of the clusters shows that the first cluster contains only correct answers, the second contains only incorrect answers, and the cluster with mixed grades contains all of the other answers.

Finally, we tested our approach to see whether clusters are characterised by the frequency of words in the cluster and if this has any relation to the scoring rule. We show that there is a strong relationship between the clusters and the model vocabulary in students answers as well as grades. Those students who used all of the model words and those who used none of the model words, belong to different clusters. Grades inside clusters are similar. Such an approach provides teachers an opportunity to give common feedback and also grades to groups of students. The proposed approach also gives us the opportunity to look for common misconceptions in the answers given by students.

In order to improve our results, we need more input from the teachers, and more detailed analysis. In our experiment, we observed spelling mistakes that should be corrected during the feedback process. We can detect again in the scoring process and adapt scores (if the teacher so wishes) related to spelling. We also observed that students used synonyms of the words in the model answer. A technical dictionary of synonyms could be developed (a standard thesaurus would contain too many unrelated words due to context), or teachers could provide acceptable alternative terminology.

Such an automatic scoring system can provide a clear baseline where conversations about assessment and feedback can develop. It is crucial that in this age of improving artificial intelligence, that we use machines to reduce the amount of repetitive straightforward scoring, which the human is poor at performing, and have people engaged in higher level, more valuable assessment and feedback.

The study in this chapter shows a methodology to quantifying module-specific meaning in short academic texts by keywords. This solidified our understanding of how to analyse scientific texts and how to approach further research to scientific-specific meanings.

CHAPTER 3

LScDC – New Large Scientific Dictionary

In this chapter, we present a scientific corpus of abstracts of academic papers in English – Leicester Scientific Corpus (LSC). The LSC contains 1,673,824 abstracts of research articles and proceeding papers indexed by WoS in which publication year is 2014. Each abstract is assigned to at least one of 252 subject categories. Paper metadata include these categories and the number of citations. We then develop scientific dictionaries named Leicester Scientific Dictionary (LScD) and Leicester Scientific Dictionary-Core (LScDC), where words are extracted from the LSC. The LScD is a list of 974,238 unique words (lemmas). The LScDC is a core list (sub-list) of the LScD with 104,223 lemmas. It was created by removing LScD words appearing in no more than 10 texts in the LSC. LScD and LScDC are available online. Both the corpus and dictionaries are developed to be later used for quantification of meaning in academic texts.

Finally, the core list LScDC was analysed by comparing its words and word frequencies with a classic academic word list 'New Academic Word List (NAWL)' containing 963 word families, which is also sampled from an academic corpus. The major sources of the corpus where NAWL is extracted are Cambridge English Corpus (CEC), oral sources and textbooks. We investigate whether two dictionaries are similar in terms of common words and ranking of words. Our comparison leads us to main conclusion: most of words of NAWL (99.6%) are present in the LScDC but two lists differ in word ranking. This difference is measured.

3.1. Introduction

3.1.1. Quantification of Meaning in Academic Texts

The interest of adaptation the modern technologies to text mining is growing fast, along with the awareness of the importance of textual data in almost all industries. The increase in the number of users of the internet and social media platforms makes a huge amount of textual data available that play a crucial role on research and marketing strategies.

The storage of almost all types of data in electronic platforms and the spread of the social networking sites for scientists open up opportunities for researchers to share scientific researches and to access a wide range of publication repositories freely and effectively. According to [**58, 59**], the largest academic social network

ResearchGate has 15+ million researchers registered, with a huge number of publications in multidisciplinary journals. The problem of searching for relevant papers out of an enormous number of publications and extraction the information from texts gradually becomes crucial. Therefore, automated procedure of text processing and extracting the meaning of a text have become important issues.

In NLP and computational linguistics, formal identifying 'which meaning a text includes' is still an open problem. Although in standard text mining applications one may relate the problem to 'topic identification' or 'determining the class of text', we consider this problem more widely. Our goal is not only determining 'which class (classes) a text belongs to' or 'the topic (content) of a text', but also numerical representing the meaning of the text.

This task is different from quantification in classification applications. *Quantification (or text quantification)* in classification is defined as the activity of estimating the prevalence (relative frequency) of each class in a set of unlabelled items [**60**]. The aim here is, given a training set with class labels, to induce a quantifier that takes unlabelled test set, and then accurately estimate the number of cases in each class. Rather than estimating the prevalence of classes (or different meanings) in the corpus or determining the topic of the paper, we intend to represent the meaning in a research text numerically — so-called *quantification of meaning in research texts*. Our assumption is that words and texts have meaning priors and meaning of a text can be extracted, at least, partially, from the information of words for subject categories distributed over the corpus. Given such information, these prior can be exploited firstly for each word and then each research text which is a collection of words. In other words, meaning of a research text is generated when different bits of information are associated with subject categories [**61**].

This approach follows the classical psycholinguistic ideas of measurement of meaning [**62**] but instead of psychologically important sentiments the research categories are used. Each word is represented as a vector of information scores for various categories, and the meaning of the whole text is formalised as a cloud of these vectors for words from the text. For larger texts this approach can take into account correlations between words (co-occurrence of words and combinations of words).

Quantities of meanings of texts can be later used in a number of applications involving categorisation of texts to pre-existing categories, creation of 'natural' categories or more precisely, clustering of similar texts in such a way that texts in the same group have same/similar meanings. The solution to the problem of 'quantification of meanings in text' also impacts on other issues such as prediction of impact of a scientific paper.

Let us consider, for instance, grouping texts based on their contents. Bringing related research texts together gives the community a convenient and easily accessible location where a deep digging becomes possible inside. This provides users many benefits such as learning the hottest topics, the most significant researches and the latest developments in a specific field. Such automated mechanisms have also benefit for editors to help them in associating researches, for instance in the step of evaluating a new submission to determine whether it fits to the journal and standards in the field in terms of content, and more importantly to initiate the peer review process by selecting experts in the field.

In practice, searching and preliminary express understanding of a paper's content is generally done by reading title and abstract rather than reading full-text of the paper. Therefore, it is reasonable to search for relevant papers by searching of relevant abstracts. Natural questions needed to be answered here are: how to automatically process the abstracts, extract the meaning from such relatively short texts and represent the meaning in a text numerically to make them usable for text mining algorithms. These questions are some of our focuses to be answered through our research.

This work is the first stage in the project outlined. In this chapter, we consider the creation of an academic corpus and scientific dictionaries where words are extracted from the corpus. All steps in the creation process will be presented in details. The corpus and dictionaries will be used for quantification of meaning in academic texts in later stages of our research.

3.1.2. Building a Scientific Corpus: Leicester Scientific Corpus (LSC)

One of the key issues in the analysis of meaning of texts is to use a corpus that is built in accordance with the scope of the study. Quantification of meaning of research texts, extracting information of words for subject categories and later prediction of impact of the paper using quantity of texts naturally require well-organised, up-to-date and annotated corpus by subject categories and the information of citations along with the abstracts. For this purpose, we developed a *scientific corpus* where texts are abstracts of research papers or proceeding papers, followed by creation of scientific dictionaries where words are extracted from the corpus. In this chapter, we focus on building of a scientific corpus and dictionary to be used in future work on the quantification of the meaning of research texts. All steps of creation the corpus and the dictionary are presented in the later sections of the chapter.

The LSC is a collection of 1,673,824 English written abstracts of research articles and proceeding papers indexed by WoS [**63**], selected so as to represent the largest variety of abstracts of scientific works published in 2014 [**64**]. Texts within the corpus are distributed across 252 subject categories – with over 298 million words including stop words. No consideration is given to the selection of categories, we extracted all texts regardless of how many texts are included in an individual category. Each document in the corpus includes the text of abstract and the following metadata: title, list of subject categories, list of research areas, and times cited [14, 65, 66]. Documents also have the list of authors with the exception of 119 documents, we did not exclude these documents. We collected documents in July 2018; therefore, the number of citations is from the publication date to July 2018.

Given the LSC, we also intend to create scientific dictionary where words are extracted from the texts of abstracts in LSC to be used on measuring the information of words for subject categories in the process of quantification of meaning of texts. To better represent scientific fields, the variety of disciplines and corpus size are very important criteria in dictionary creation. The more disciplines where texts are collected from, the bigger and comprehensive dictionary can be created. Similarly, the more articles are collected, the more representative set of words for specific fields can be gathered. As we did not exclude any category from the corpus of 1,673,824 texts distributed over 252 categories, we expect a reasonable representativeness of words. In addition, the dynamic nature of languages and changes of words with new discoveries – due to fast changes in science and technology – lead to a need for an up-to-date scientific dictionary. Thus, we created two scientific dictionaries based on a large, multidisciplinary corpus of academic English: LScD and LScDC [67, 68].

3.1.3. Building Scientific Dictionaries: Leicester Scientific Dictionary (LScD) and Leicester Scientific Dictionary-Core (LScDC)

LScD is a list of words extracted from the texts of abstracts in the LSC. The words in the LScD is sorted by the number of documents containing the word in descending order. There are 974,238 unique words (lemmas) in the dictionary. All words in the LScD are in stemmed form and stop words are excluded. The dictionary also contains the number of documents containing each word and the number of appearance of the word in entire corpus. All steps to process the LSC, build LScD and the basic statistics with characteristics of words in the LScD are presented in the later sections of this chapter.

LScDC is a core list (sub-list) of the LScD. The dictionary contains of 104,223 unique words (lemmas). The following decision is taken in the creation of the LScDC: words (in LScD) that appearing in no more than 10 documents (≤ 10) are removed under the assumption that too rare words are not informative for text categorisation and gives almost zero scores in the calculation of information score as they appear in less than 0.01% of documents. 60% of words in LScD appear in only one document. Our casual observation of words indicates that many of such words are non-words or not in an appropriate format to use (e.g. misspelling); therefore, they are likely to be non-informative signals (noise) for algorithms. More information and examples can be found in Section 3.5 and Section 3.6. Removal of such words results in reducing the number of words in applications of text mining algorithms. When the threshold 10 is decided, we consider a cut which is not too small or high to be able to keep a reasonable number of words for analysis, but we paid attention to have a noticeable impact on size of dictionary and results. We did not remove any frequent words in this stage as stop words are already removed in pre-processing steps. The core dictionary is also ordered by the number of documents containing the words and includes the information of the number of appearance of the word in entire corpus.

3.1.4. A Comparison of the LScDC and the New Academic Word List (NAWL)

This study also compares the LScDC and the New Academic Word List (NAWL) [22]. The procedure used to compare involves looking at a classic academic list of words NAWL and investigating whether two lists contain the same words, the number of common words, possible reason of mismatches and whether ratings of matched words are actually the same/similar in two dictionaries. Overall, we intend to see whether there is similarity between two lists.

The reason why we consider the NAWL for comparison lies in two facts. The major reason lies in the way the sampling of the vocabulary, which is similar to ours in terms of being from a general and academic corpus. The second reason is that the AWL and NAWL are classics and landmarks as academic lists in vocabulary and corpus-based lexical studies. Our aim is not to replace the AWL and NAWL, but create a large corpus and scientific dictionary representing research papers from various subject fields without any limitation in disciplines for our goal of discovering and quantifying the meaning of research texts.

The NAWL consists of 963 word families based on a academic corpus of 288 million running words (all words in text without tables' captions, titles and references) [22]. The major categories of the corpus where words are extracted are: the Cambridge English Corpus (CEC), oral sources and textbooks. The largest proportion of tokens came from the CEC, about 86% (over 248 million words). The oral part was taken from the Michigan Corpus of Academic Spoken (MICASE) and the British Academic Spoken English (BASE) corpus. The oral corpora and the corpus of textbooks are divided into four categories: Arts and Humanities, Life and Medical Sciences, Physical Sciences, and Social Sciences. The NAWL covers 92% of its corpus when combined with the New General Service List (NGSL) [69]. In the list, words are listed by headwords of word families where various inflected forms are contained in. However, we observed that some of headwords in the NAWL indicate the same stemmed word in the LScDC. For instance, two distinct headwords 'accumulate' and 'accumulation' in the NAWL matched with 'accumul' in the LScDC. To avoid the effect of this difference on our analysis, we applied stemming on words of the NAWL. We used average statistics of different forms as the final statistics for stemmed word. After stemming, the number unique words (lemmas) decreased to 895 in the NAWL, we take these words into account in the comparison study.

The NAWL however does not contain much specialised technical terms such as chemical terms and species in biology. To illustrate this, let us explain more details. The Coxhead's Academic Word List (AWL) created in 2000 – inspired to create NAWL –, includes 570 word families (from a corpus of 3.5 million words) [70]. According to Coxhead, academic words are supportive of academic text but not central to the topics of the text. The list of AWL contains words that account for approximately 10% of the total words in the collection of academic texts. The AWL and GSL (General Service List) covers approximately 86% of total words in academic corpus. Coverage refers to the number of words (tokens, i.e. all forms of words) in texts which are covered by the list of words. By updating this list with an expended and carefully selected corpus of 288 million words (corpus where NAWL is designed), the coverage was improved to 92% of new corpus when combined with NGSL, with approximately 5% improvement [22, 69]. Not all words extracted from the corpus is included in the list. In the creation of list, a measure considering the distribution of words over disciplines (Dispersion) was taken into account as well as the frequency (Standard Frequency Index). In LScDC, we include all of words appearing in not less than 10 texts in the corpus – distributed over 252 subject categories –, we did not apply any procedure to exclude any other words. This explains the difference between the number of lemmas in two lists – with 895 lemmas of NAWL and 104,223 lemmas in the LScDC.

The NAWL and the LScDC were actually developed from different corpora and the number of words are quite different, but overall the LScDC contains the much lemmas of NAWL (except only 4 words). In this stage, we did not include the New General Service List (NGSL) as our aim is to evaluate only academic words. In comparison, it must be stressed that there is 103,328 more lemmas in LScDC than the NAWL. Adding the NGSL could result in an increase in the number of the same words in the LScDC and NAWL plus NGSL.

In the comparison study, we initially investigate dictionaries to see the coverage of NAWL words by LScDC. We will see that there are 891 words that occur in both the LScDC and the NAWL, which means that the overlap between the LScDC and the NAWL is 99.6%. Four words appearing only in the NAWL are: "ex", "pi", "pardon" and "applaus". This seems to be the result of differences in types and processing of texts in corpora. It is worth to note that corpus of NAWL includes full texts from academic domain, while the LSC includes abstracts of academic texts. This, for instance, may be the reason why "pi" does not appear in LSC as it is commonly used by the symbol π (pi) in math world and not many articles include formulas in abstract. The other two words "pardon" and "applaus" are contained in LScD with low frequencies (5 and 9 respectively); therefore, they are removed in the step of LScDC creation. However, these two words have low rank in the NAWL as well: rank 924 and 956 in the list). Finally, the word “ex” does not occur in the LScDC due to pre-processing steps applied in the creation of the dictionary. We united some prefixes including “ex” with the following words (e.g. ex-president is converted to expresident).

We also present different approaches for comparison to understand what fragment of LScDC contains the NAWL. This is performed by repeatedly searching NAWL words in various subsets of LScDC. Our second focus in dictionary comparison will be the comparison of ranks of words in two dictionaries. In this study, only common words (891 words) are taken into account. Several different methods to compare ranks are considered such as direct comparison of ranks, pairwise comparison of partitions in dictionaries (lists are divided into sub-lists and overlapping words are count in each sub-list), comparison of the top n words, and the comparison of the bottom n words. We will also test similarity of ranks by statistical tests. It is expected to observe that words in two lists are not distributed in the same way as the statistics to order lists are not calculated in the same way. The NAWL considers the dispersion of words over categories, while we simple take the number of documents containing words in the LScDC. All approaches and results are presented in the section of comparison in detail.

In this study, we also consider the reproducibility of dictionaries from the LSC and list of texts from other sources to be used by researchers in many other text mining applications. For this purpose, we made R codes for producing the LScD and the LScDC, and instructions for usage of the code available in [**71**].

3.1.5. The Structure of This Chapter

The chapter is organised as follows. Section 3.2 contains the principles in corpus and dictionary design as well as the text and word representation approaches. In Section 3.3, we describe some of widely used and well-known analogue corpora and dictionaries. Section 3.4 sets out all pre-processing steps in creation process and the structure of LSC. Similarly, Section 3.5 and Section 3.6 present pre-processing steps to build the LScD and the LScDC respectively, and the organisation of dictionaries. In Section 3.7, a study of comparison of the LScDC and the NAWL with several approaches is contained. Finally, Section 3.8 concludes the chapter.

3.2. Methodology

3.2.1. Fundamentals of Corpus Design

In linguistics, a text corpus is defined as a large collection of text and they are used by linguists, lexicographers, experts in NLP and in many other disciplines in order to generate language databases, study general linguistic features, do statistical analysis or learn linguistic rules.

Types of corpora vary depending on how they are sampled and designed for specific research goals. Texts in a corpus are assembled to ensure maximum representativeness of a particular language or language variety. Representativeness refers a sample that includes the complete range of texts in a target population [72]. Target population is closely related to the scope of the research and respectively sampling. Any selection of text is described as a sample; however, representativeness for a sample depends on the definition of the population that sample is intended to represent and methods of selection of the sample from that population. To define the population, the most important two considerations are: boundary of the population (what texts are included) and the range of genres (what text categories are included) within the population. For instance, Lancaster-Oslo/Bergen Corpus (LOB) is defined as the collection of British English texts that all are published in 1961 in the *British National Bibliography Cumulated Subject Index 1960-1964* for books and all 1961 publications in *Willing's Press Guide 1961* for periodicals and newspapers; distributed across 15 text categories (such as general fiction, romance-love story etc.) [73, 74]. The target population for the LOB was written British English texts that all are published in 1961 in United Kingdom (boundary)–distributed across 15 text categories (genres).

The goal of corpus construction is very important for corpus design as it determines the target population. For instance, if the goal of the research is to investigate learners' English, it is reasonable to collect essay of students learning English. However, one who wants to capture the complete range of varieties of English will attempt to collect contemporary British English written texts from a wide variety of different domains. With a given research purpose, a simple broad distinction on corpus types can be done: general corpus and specialised corpus. The criteria for representativeness for these corpora differ from each other by sampling principles. A general corpus contains a broad range of genres with a balance of texts from a wide variety of the language in different domains, while a specialised corpus contains texts from a particular genre or a specific time. For instance, a corpus can be representative for general English language which is an example for general corpus; fiction books or researches in medicine which are examples for specialised corpus.

Some other considerations in sampling decision are the kinds of texts, the number of texts and the length of text samples as well as sampling techniques. Sampling techniques rely on random selection. Basically, selection can be done by a simple random sampling or stratified random sampling. In basic random sampling, texts having equal chance to be selected in a population are randomly selected. In stratified random sampling, the whole population is divided into smaller groups (e.g. genres) and then each subset is sampled using random selection techniques (with proportionality to the subgroups)[75]. In the LOB corpus, for example, the population was first divided into 15 categories; and samples were drawn from each category.

3.2.2. Representation of Texts and Words

In order for an effective text processing to be accomplished, one of the most fundamental tasks is to select the most appropriate text representation technique for a particular application of NLP. The quality of any text mining and NLP techniques is strongly dependent on the text representation. It aims to represent texts to enable them to be used in mathematical computations by the machine.

In general, the most common text representation model in text mining is the Vector Space Model (VSM) [76, 77]. In this model, each text is represented by a numerical vector where its components are taken from the content of the text. Components of the vector denote the features that characterise the text such as words, phrases, paragraphs or a single character etc (*tokens*). Therefore, each text is represented by a collection of words (or words' combinations) and the corpus can be represented by the union of such collections. The most common and simplest way to transform a text into a vector is to represent them by words from the vocabulary of the text collection. Each text in the collection is thus a feature vector in the vector space. In that case, the dimensions of the vector space is equal to the vocabulary size, and the order of the words in each text is ignored.

Having the texts represented by vector of words, list of words can be extracted with various statistics such as weight of words for a given text. This representation of the text by a bunch of words is BoW [44]. In BoW, different word weighting schemes can be used. One simple count for each feature's value (word weight) might be Boolean model. In Boolean model, 1 indicates that the word appears in the text and 0 indicates the word does not appear in the text. This scheme holds only the information about presence or absence of a given word in texts. As an extension of the Boolean model, *TF* (term frequency) shows how many times a word appears in the text [78]. In this scheme, the distribution of the word across the collection is not taken into account. However, some words can be more significance than others in the corpus. In that case, *TF* can be multiplied by *IDF* (inverse document frequency), which is defined as the logarithm of the division of the corpus size by the number of documents containing the word. This scheme is called *TFIDF* [79]. In addition, another scheme is to count the number of appearance of a word in the entire corpus when the corpus is considered as one large document.

Designing the texts by words can be performed in different ways depending on the query. One may want to represent text by all inflected forms of words or stop words. For instance, in the creation of the list of the most widely used conjunctions in a language, removing stop words leads to unreliable results. Therefore, the objection in text representation should be to turn each text into a set of words that supply the task with necessary inputs.

3.2.3. Building a Dictionary from Text Collection

In this study, a dictionary is defined as the set of unique words (*lemmas*) extracted from texts in a corpus. In other words, a dictionary is produced based on corpus data. In corpus linguistics, every dictionary is compiled from a particular corpus and the way to establish of a word list must be defined individually for a given purpose.

Dictionaries differ from one another by the words selected. Several distinctions can be done based on their scopes and purposes. From the overview, the simplest distinction can be observed between general and specialised dictionaries (also refereed as *technical dictionaries*). In specialised dictionaries, words are extracted from a corpus in a single (or multi) specific field(s) and indicate the concepts of the field(s) while general dictionaries contain a complete range of words. Words in specialised dictionaries are called *terms* or *topic-specific words*. In contrast to terms, a word that has a little lexical meaning is called *function word* in linguistics. Some examples of function words are prepositions, pronouns, determiners etc. In English semantic, non-function words (content words) are words that indicate the content or the meaning of the texts such as nouns, verbs and adjectives. In addition, Coxhead used the notion *supportive* for academic words in AWL [70]. She stated that academic words in AWL are supportive of the academic texts. As supportive, she meant words which are not central to the topic of the text. One would consider words that do not indicate any terminology or specialised technical terms in the subject field. ‘establish’ and ‘inherent’, for example, are two of supportive words according to Coxhead. These are words which authors from most or all academic disciplines tend to use them; the majority of them are also used in general English. She excluded all terminologies such as marine species in Oceanography and function names in Mathematics (e.g. Gaussian).

To define words and dictionaries, several other distinctions are applied by lexicographers and experts such as prescriptive or descriptive, dictionaries by language, dictionaries by size, Language-for-Specific-Purposes dictionary (LSP such as medical dictionaries) etc [80]. In this study, rather than consider such distinctions, we focus on building a corpus-based dictionary from scientific abstracts written in English.

Such dictionary may be considered as *scientific dictionary* giving the guidance to scientific writers on such matters as up-to-date, topic-specific and supportive words of academic texts.

In the creation of a scientific dictionary from academic texts, two important criteria are: corpus size and the variety of disciplines where texts are categorised into. The more texts are collected, the more representative set of words for specific fields can be gathered. Similarly, the more disciplines where collected texts belong to, the bigger and more comprehensive dictionary can be created. A large and multidisciplinary dictionary with all supportive and topic-specific words of academic texts can also cover to other corpora and be used for any text analysis tasks on them.

3.3. Related Works

3.3.1. Corpora of English

There are several freely available corpora for NLP tasks. In this section, we begin by listing some of those well-known corpora developed for English. The earliest corpus in electronic form was developed in 1964 at Brown University, which contains written American English published in 1961 [**81**, **82**]. *Brown corpus* includes 500 samples of American English text of published works in the United States in 1961. Each text consists of over 2,000 words sampled from 15 text categories, with totally over one million running words. Although today the size of corpus is considered small when comparing recent corpora, it is still widely seen as a landmark publication as a computer readable and general corpus among linguistic researchers. The corpus is similarly designed as LOB which followed the design and sampling practice of the Brown corpus in order to match the Brown corpus for British English [**73**, **74**]. These two corpus became a model for other national corpora, so-called 'Brown Family' [**83**]. In selecting texts for inclusion in the Brown corpus and the LSC, different considerations applied based on the aim of the design of corpora as well as the differences in size of corpora. Brown corpus is sampled from a wide variety of different types of sources such as novels, news, editorials, reviews and many more; while the LSC is sampled from scientific abstracts and proceeding papers.

British National Corpus (BNC) is a monolingual, general corpus of over 4,000 samples of modern spoken and written British English covering English of the later part of 20th century (from 1960 onwards) [**84**, **85**, **86**]. The latest edition of the BNC is published in 2007. In general, it covers many different styles and varieties of text from various subject fields and genres. The written part of the corpus contains samples from a wide source of text such as: regional and national newspapers, journals, academic books, fiction, letters, school and university essays, other literary text. The spoken component of the corpus is made up of informal conversations recorded by volunteers who were selected from different age, social class and gender, and task-oriented spoken language ranging from formal meetings to radio shows and lectures. The corpus was designed to identify social and generic uses of contemporary British English with 100 million words [82]. The major differences between the BNC and the LSC lie in the size of the corpus, in the aim of design (being to capture the full range of varieties of contemporary language use versus to extract scientific ones), in the definition of the populations and in the sampling of corpora in terms of being mixed corpus (spoken and written English) versus written English.

One other well-known corpus is *Oxford English Corpus (OEC)* which is also used by Oxford lexicographers to construct Oxford English Dictionary (OED), supplied by Oxford University Press [87]. The corpus contains of over 10 billion words of 20th and 21st century English from English-speaking countries: the UK, USA, Ireland, Australia, New Zealand, the Caribbean, Canada, India, Singapore and South Africa. It is one of the largest corpus in the world [87]. The corpus is mainly drawn from the web with all types of English such as academic journals, literary novels, newspapers, magazines, language of blogs, emails and social media [88]. Another Oxford University Press corpus is *Oxford Corpus of Academic English (OCAE)* contains academic journals and textbooks from four main disciplines: physical sciences, life sciences, social sciences, and humanities with 85 million words included [89].

The *SciCorp* is a corpus of 14 English scientific articles sampled from two disciplines: genetics and computational linguistic, released in 2016 [90]. The corpus includes 61,045 tokens. The population of the corpus being compiled from scientific text is similar to the LSC. However, sampling of SciCorp differs from the LSC as being restricted to two disciplines. Apart from sampling principals and the size of corpus, one other difference of SciCorp from the LSC lies in the type of texts: full-text in SciCorp and abstracts of scientific papers in LSC.

The *Reuters-21578 corpus* (Reuters-21578 Text Categorization Collection) is a collection of 21,578 news documents used for text categorisation [91]. It contains news appeared in 1987 with categories. The main differences between the Reuters corpus and the LSC is genre of texts: Reuters corpus contains texts of news while LSC contains abstracts of scientific publications. The LSC is more than 70 times as large as the Reuters corpus.

The *GENIA corpus* is similar to the LSC in terms of the content of texts, both contain the abstracts of scientific papers [92]. The GENIA corpus is built by annotating abstracts with keywords '(MeSH terms) Human', 'Blood Cells' and 'Transcription Factors'. 2,000 abstracts are selected for a research objective in Biological and Clinical domains. LSC was created without research area restriction and contains 700 times more abstracts from different research areas.

The *DBpedia abstract corpus* contains 4,415,993 texts of the introductory section of Wikipedia articles, these sections may not necessarily be scientific writing [93]. As introductory section of Wikipedia articles are not actual abstracts of papers, the average length of documents are different than average length of abstracts: 178 words for the LSC and approximately 524 words for the DBpedia.

3.3.2. English Dictionaries

One question that can not be easily answered in dictionary design is whether there is an exact count of the number of English words. The major reason for this issue lies in the dynamic nature of languages. It is commonly accepted that languages change rapidly with cultural and technological evaluation, and adoption from other languages [94]. For instance, the *Oxford English Dictionary* has recently added 'satoshi' (the smallest unit of a bitcoin), 'yeesh' (expressing exasperation, annoyance, disapproval) and 'simit' (a type of ring-shaped bread roll originating in Turkey) to its database in 2019 [95, 96, 97]. Another consideration on counting the number of words is that what words a dictionary includes. For example, a dictionary would include all technical terms, scientific entries or slang; all of the inflected form of a word (e.g. listen, listening etc.); plurals of words as separate word; or compounds which is made up two words. Therefore, the simple question 'what exactly is a word?' turns out to be surprisingly complicated. Some dictionary-makers agree that different versions of words should be counted only once, while some others consider each form as a separate word [98]. This means that there may be unlimited number of words in writing and spoken English, which do not appear in any dictionary.

Although it is not possible to know exact number of words in English, the estimate has been given roughly one million words (ranging from half a million to over two million) – including names of chemicals and scientific terms– in vocabulary [99, 100]. Many of these words are too rarely used, so it is expected that they do not appear in any English dictionary. One of the most well-known and commonly used dictionary *Oxford English Dictionary (OED)* includes over 600,000 words recorded in 20-volumes [101]. The dictionary provides both present days meaning of the words and the history of words from the across of English speaking countries. In addition to the print edition, the dictionary is available online [102]. Similar to the OED, the *Webster's Third New International Dictionary* contains over 470,000 entries [99, 103]. Another Oxford dictionary so-called *New Oxford Dictionary for Writers and Editors* is built to guide writers, editors, journalists and everyone who works with words [104]. It includes 25,000 words and phrases with providing advice on spelling, capitalisation, specialist words and cultural context such as names, mathematical symbols, chemical elements. The *Oxford Learner's Dictionary of Academic English (OLDAE)* is also supplied by Oxford University Press with over 22,000 words based on the OCAE [105]. The aim of the dictionary is to help students particularly in academic English writing. As an example of specialised dictionary, *Stedman's Medical Dictionary* contains more than 107,000 terms with images (including abbreviations and measurements) in medical references in its 28th edition [106, 107]. It is designed to provide language of medicine, nursing and health profession to medical students, researchers, physicians and many more medical language users. Finally, we paid attention to the work of AWL and NAWL as they are classics as academic word list [69, 70]. The AWL includes 570 word families from the collection of written academic texts distributed across the four main disciplines: arts, commerce, law and science. It covers 86% of the total words in the corpus when combining with GSL. Similarly, NAWL contains 963 word families based on an up-to-date corpus of academic texts. NAWL-NGSL covers 87% of new corpus. The more detailed explanation is given in the Section 3.7.

Although the estimation of number of words in a language is not a easy task and numbers of words in dictionaries vary differently depending on the content of the dictionary, a corpus-based analysis may give a sight to understand the average number of words for a vocabulary. Let us consider Oxford English Corpus and Oxford English Dictionary with base forms of words (lemmas). It is stated in [108] that 25% of all words used in OED is one of lemmas: the, be, to, of, and, a, in, that, have and I. These are the most common 10 lemmas in English. In similar way, the most common 100 and 1,000 lemmas account for 50% and respectively 75% of all words used in OEC. To cover 90% of the corpus, one needs 7,000 lemmas. 95% of the corpus includes approximately 50,000 lemmas which words in between occur very rarely (e.g. only once every several million words). To cover 99% of the corpus, we need a vocabulary of over 1 million lemmas. In that case, many words may appear only once or twice in entire corpus (e.g. specialised technical terms), but lemmas will be representative of the whole corpus. To represent notable part of English, 90-95% of the corpus may be taken as a reasonable number.

3.4. Leicester Scientific Corpus (LSC)

LSC is a collection of abstracts of articles and proceeding papers published in 2014 and indexed by the WoS database [63]. Each document contains the text of abstract and the following metadata: title, list of authors, list of categories, list of research areas, and times cited [14, 65, 66]. The corpus comprises only documents in English. The LSC was collected in July 2018 and contains the number of citations from publication date to July 2018.

We describe a *document* as the text of abstract with metadata listed above. The total number of documents in LSC is **1,673,824** [64]. All documents in LSC have non-empty abstract, title, categories, research areas and times cited in WoS databases. There are 119 documents with empty authors list, we did not exclude these documents.

3.4.1. Corpus Construction

This section describes all steps in order for the LSC to be collected, cleaned and made available to researchers. Data processing consists of four main steps:

3.4.1.1. ***Step 1: Collecting the Data.*** The dataset is downloaded online by exporting documents as tab-delimited files, so all documents are available online. The data are extracted from WoS [**63**]. You may not copy or distribute the data in whole or in part without the written consent of Clarivate Analytics[1]

3.4.1.2. ***Step 2: Cleaning the Data from Documents with Empty Abstract or without Category.*** Not all papers have abstract and categories in the collection. As our research is based on the analysis of abstracts and categories, preliminary detecting and removing inaccurate documents were performed. All documents with empty abstracts and documents without categories are removed.

3.4.1.3. ***Step 3: Identification and Correction of Concatenated Words in Abstracts.*** Traditionally, abstracts are written in a format of executive summary with one paragraph of continuous writing, which is known as *unstructured abstract*. However, especially medicine-related publications use *structured abstracts*. Such type of abstracts are divided into sections with distinct headings such as introduction, aim, objective, method, result, conclusion etc.

Used tool for extracting documents leads to concatenated words of section headings with the first word of the section in abstracts. As a result, some of structured abstracts in the LSC require additional process of correction to split such concatenated words. For instance, we observe words such as "ConclusionHigher" and "ConclusionsRT" etc. in the corpus. The detection and identification of concatenated words cannot be totally automated. Human intervention is needed in the identification of possible headings of sections. We note that we only consider concatenated words captured in headings of sections in medicine-related papers as it is not possible to detect all concatenated words without deep knowledge of research areas. Identification of such words is done by sampling of medicine-related publications. The section headings in such abstracts are listed in Table B.1.

In headings of a section, the words usually start with a capital letter and end with a colon, unless there is typographical error in an electronic material. The words following a heading word (or a colon) also start with a capital letter in structured abstracts. We take these properties into consideration while detecting concatenated words.

All words including headings in the Table B.1 are detected in the entire corpus, and then words are split into two words. For instance, the word "ConclusionHigher" is split into "Conclusion" and "Higher".

[1]Use of the LSC is subject to acceptance of request of the link by email. To access the LSC for research purposes, please email to ns433@le.ac.uk or suzenneslihan@hotmail.com.

3.4.1.4. ***Step 4: Extracting (Sub-setting) the Data Based on Lengths of Abstracts.*** After correction of concatenate words is completed, the lengths of abstracts are calculated. "Length" refers the total number of words in the text, calculated by the same rule as for Microsoft Word "word count"[109]. An abstract is a short text that is written to capture the interest of a reader of the paper. Thus, abstracts briefly describe and summarise the work and the findings usually in one paragraph of words, but very rarely more than a page.

According to APA style manual [110], an abstract should contain between 150 to 250 words. However, word limits vary from journal to journal. For instance, Journal of Vascular Surgery recommends that "Clinical and basic research studies" must include a structured abstract of 400 words or less" [111].

In LSC, the length of abstracts varies from 1 to 3,805. We decided to limit length of abstracts from 30 to 500 words in order to study documents with abstracts of typical length ranges and to avoid the effect of the length to the analysis. Documents containing less than 30 and more than 500 words in abstracts are removed. Fig. 3.1 shows the distribution of lengths over documents of LSC before and after removing documents containing less than 30 and more than 500 words.

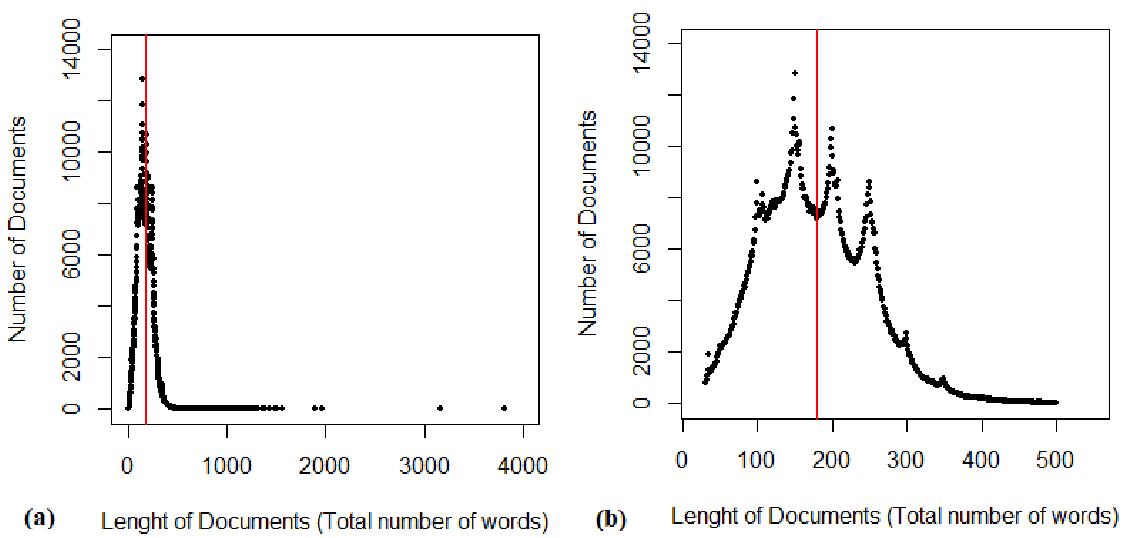


FIGURE 3.1. Length distribution of documents (a) before and (b) after removing documents containing less than 30 and more than 500 words (maximum length was 3,805 before removing). The vertical line shows the average length.

Four main peaks observed on the length graph: at 100 words, 150 words, 200 words and 250 words. The second peak shows the maximum number of documents, where those observed when the number of words in an abstract is 150. This result is expected as typically word limits range from 150 to 250 words for an abstract.

After the process of correction and cleaning, the database contains of raw texts of abstracts with title, list of authors, list of categories, list of research areas, and times cited. The total number of documents is 1,673,824 in LSC (see Table 3.1).

TABLE 3.1. Number of documents before and after cleaning documents with empty abstracts or without category, and after removing too short and too long abstracts.

	# of Documents
Original data	1,727,464
After cleaning documents with empty abstracts or without category	1,681,469
After removing too short and too long abstracts	1,673,824

3.4.2. Organisation of the LSC

In LSC, the information is organised with one record on each line and parts of "List of Authors', "Title", "Abstract", "Categories", "Research Areas", "Total Times cited", and "Times cited in Core Collection" is recorded in separated fields [64]. Table B.2 demonstrates the structure of a document in LSC.

The "Categories" field contains the list of the subject categories where the document is assigned to [14]. Each document in LSC is assigned to at least 1 and at most 6 categories. There are totally 252 categories in the corpus. The full list of categories are presented in [64] (v1) and [112]. It is noteworthy that the study through this chapter is done based on this version (v1) of the LSC; however, another version will be created in the next chapter and the list of categories for the new version are presented in Appendix C.2. The subject categories are the same in both version, only the numbers of texts assigned to categories have been changed.

The "Research Areas" field consists of the list of research areas described as "a subject categorisation scheme" in WoS database [65]. Each category is mapped to one research area in the WoS collection. There are totally 151 research areas in the corpus. The full list of research areas is presented in [64] and [112]. Similar to subject categories, the study through this chapter is done based on this version (v1) of the LSC; however, another version will be created in the next chapter and the list of research areas for the new version are presented in Appendix C.3. The research areas are the same in both version, only the numbers of texts assigned to areas have been changed.

"Total Times Cited" consists the number of times the paper was cited by other items from all databases within WoS platform. A paper can appear in multiple databases indexed in WoS collection. The citation indexes in WoS are: WoS Core Collection, BIOSIS Citation index, Chinese Science Citation Database, Data Citation Index, Russian Science Citation Index and SciELO Citation Index. Duplicate documents across multiple databases is counted only once [66].

"Times Cited in Core Collection" is the total number of times the paper cited by other papers within the WoS Core Collection. The citation indexes in Core Collection are: Science Citation Index Expanded, Social Sciences Citation Index, Arts and Humanities Citation Index, Conference Proceedings Citation Index–Science, Conference Proceedings Citation Index–Social Sciences and Humanities, Book Citation Index–Science, Book Citation Index–Social Sciences and Humanities, Emerging Sources Citation Index [66].

3.5. Leicester Scientific Dictionary (LScD)

This section presents the pre-processing steps for creating an ordered list of words from the LSC [64] and the description of LScD.

LScD is an ordered list of words from texts of abstracts in LSC [67]. The dictionary is sorted by the number of documents containing the word in descending order. The dictionary stores 974,238 unique words, where abbreviations of terminologies and words with number are contained in. All words in the dictionary are in stemmed form of words. The LScD contains the following information: unique words in abstracts in the LSC, number of documents containing each word and number of appearance of each word in the entire corpus.

"The number of documents containing a word" is the number of the documents with the corresponding word. A word that appears multiple times in a document is counted once (binary representation for existence). "Number of appearance of a word in the entire corpus" is defined to be the total number of occurrences of a word in the LSC when the corpus is considered as one large document.

All words obtained after pre-processing steps are included in the LScD. The most frequent 20 words (frequency is calculated by the number of documents containing a word) are presented in Table 3.2 .

Table 3.2. The most frequent 20 words in the LScD

Word	Number of documents containing the word	Word	Number of documents containing the word
use	902,033	also	400,642
result	812,154	present	389,735
studi	723,827	increas	383,676
show	498,705	two	375,586
method	491,586	model	372,911
effect	476,757	signific	370,435
base	446,436	compar	355,381
differ	445,739	paper	346,514
can	441,512	time	344,817
high	402,737	perform	341,547

3.5.1. Processing the LSC and Building the LScD

The main challenge of using text data is that it is mess and not concretely structured. This means that a number of steps is needed to be taken to form the LScD. The initial step of building the dictionary is to convert unstructured text (raw corpus) into structured data. Structured data means highly organised and formatted in a way so the information contained can be easily used by data mining algorithms, mostly numerical data in relational databases [113]. There are different ways to pre-process text data and pre-processing steps should be described for each corpus individually. Decision taken and steps of processing for creation of LScD are described below. All steps can be applied for arbitrary list of texts from any source with changes of parameters and also to LSC to reproduce the dictionary.

3.5.1.1. ***Step 1: Text Pre-processing Steps on the Collection of Abstracts.*** Text pre-processing means to bring the text into a form of analysable for the task. This step is highly important for transferring text from human language to machine analysable format by data mining algorithms. As each task requires different procedures to process the text based on aim of the study, ideal pre-processing procedure of each task should be developed individually. We used standard pre-processing methods in text processing studies such as tokenization, stop word removal, removal of punctuations and special characters, lowercasing, removal of numbers and stemming as well as two non-standard pre-processing steps: uniting prefixes of words and substitution of words. In this section, we present our approaches to pre-process abstracts of the LSC.

(1) **Removing punctuations and special characters:** This is the process of substitution of all non-alphanumeric characters by space. We did not substitute the character "-" in this step, because we need to keep words like "z-score", "non-payment" and "pre-processing" in order not to lose the actual meaning of such words. A processing of uniting prefixes with words are performed later.
(2) **Lowercasing the text data:** Lowercasing is one of the most effective pre-processing step in text mining problems to avoid considering the same words like "Corpus", "corpus" and "CORPUS" differently. Entire collection of texts are converted to lowercase.
(3) **Uniting prefixes of words:** Prefixes are letters placed before a word to create a new word with different meaning. Words containing prefixes joined with character "-" are united as a word. The list of prefixes united for this research are listed in Table B.3. The most of prefixes are extracted from [114]. We also added commonly used prefixes: "e", "extra", "per", "self" and "ultra".
(4) **Substitution of words:** Some of words joined with "-" in the abstracts of the LSC require an additional process of substitution to avoid losing the meaning of the word before removing the character "-". Some examples of

such words are "z-test", "well-known" and "chi-square". These words have been substituted to "ztest", "wellknown" and "chisquare". Identification of such words is done by sampling of abstracts from LSC. The full list of such words and decision taken for substitution are presented in Table B.4.

(5) **Removing the character "-":** All remaining character "-" are replaced by space.

(6) **Removing numbers:** All digits which are not included in a word are replaced by space. All words that contain digits and letters are kept for this study because alphanumeric characters such as chemical formula might be important for our analysis. Some examples of words with digits are "co2", "h2o", "1990s", "zn2" and "21st".

(7) **Stemming:** Stemming is the process of converting inflected words into their word stem. In this process, multiple forms of a specific word are eliminated and words that have the same base in different grammatical forms are mapped to the same stem. As stemming removes suffixes and reduces the number of words in corpus, this step results in uniting several forms of words with similar meaning into one form and also saving memory space and time [**115**]. For instance, the word "listen" is the word stem for "listens", "listened", and "listening". All words in the LScD are stemmed to their word stem by R package [**116**].

(8) **Stop words removal:** In NLP, stop words (including function words) are defined as words that are extreme common but provide little value in a language. Some common stop words in English are "I", "the", "a" etc. Such words appear to be of little informative in documents matching as all documents are likely to include them. There is no universal list of stop words. Stop words must be chosen for a given purpose. In our research, we used "tm" package in R to remove stop words [**117**]. There are 174 English stop words listed in the package. Full list of stop words in tm package can be found in Table B.5.

3.5.1.2. ***Step 2: Extracting Words from Abstracts.*** After pre-processing the abstracts of LSC, there are 1,673,824 processed plain texts for further analysis. All unique words in the processed texts are extracted and listed in the LScD.

3.5.2. Organisation of the LScD

The total number of words in LScD is 974,238. Unique words, the number of documents containing the word and the number of appearance of the word in the entire corpus are recorded on each line in separated fields.

The "Word" field contains unique words from the corpus. All words are in lowercase and their stem form. The list of words is sorted by the number of documents that contain words in descending order.

"Number of documents containing the word" is the number of documents containing the corresponding word in "Word" field. In this content, binary calculation is used: if a word exists in an abstract then there is a count of 1. If the word appears more than once in a document, the count is still 1. Total number of document containing the word is counted as the sum of 1s in the entire corpus.

A word can appear many times in the same document. "Number of appearance of a word in the entire corpus" is computed as the sum of appearance of the word in each document. The field contains how many times a word occurs in the corpus when the corpus is considered as one large document.

3.5.3. Basic Statistics in the LScD

Before moving on creation of a core dictionary LScDC from LScD, we investigated basic statistics of LScD. The Table 3.3 shows the number of the rarest words over documents, where words appear in at most 20 documents. For instance, there are 592,161 words contained in only 1 document in the corpus. This distribution is also presented for all words in the Fig. 3.2. As expected, very few words occur very often, there is a larger number of mid-frequency words and very many words occur very rare in the collection. This is a typical property of text data and the distribution of words in texts [**118**].

TABLE 3.3. The number of documents and the number of words contained in the corresponding number of documents only for those words appearing in at most 20 documents

Number of Documents	**Number of Words Contained in the Corresponding Number of Documents Only**	**Number of Documents**	**Number of Words Contained in the Corresponding Number of Documents Only**
1	592,161	11	5,605
2	118,989	12	4,912
3	54,193	13	4,268
4	32,032	14	3,689
5	21,624	15	3,385
6	15,554	16	2,971
7	11,877	17	2,752
8	9,384	18	2,522
9	7,709	19	2,253
10	6,492	20	2,161

3.5.4. Decision Taken for Rare and Frequent Words

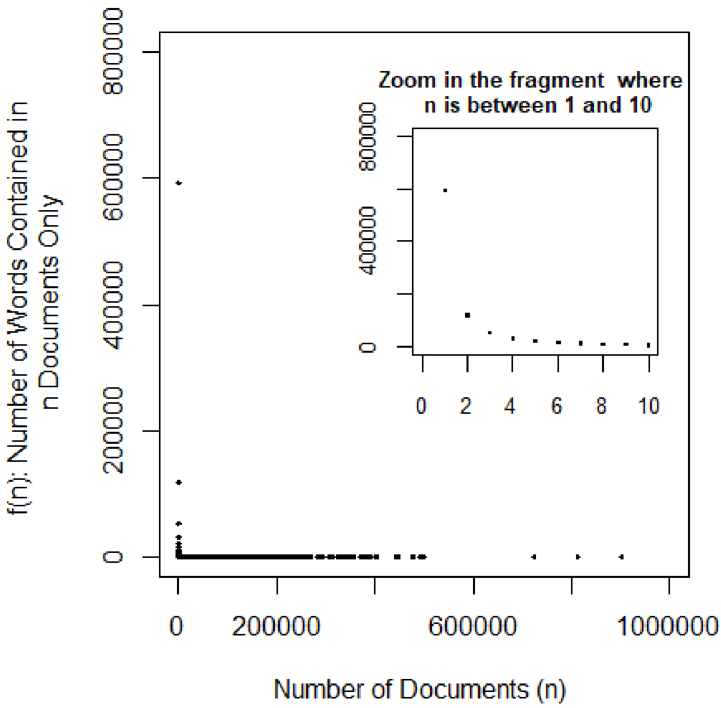


FIGURE 3.2. The number of documents (n) versus the number of words contained in n documents only

3.5.4.1. ***Traditional Approaches to Rare and Frequent Words***

In most studies on text classification and information extraction, it is common to discard rare words in order to improve the performance of methods. The idea of the usability of rare and frequent words for discriminating texts dates back to Luhn's idea [119]. He proposed a model to automatically generate the abstract by extracting the most representative sentences among all the sentences in an article. To select those sentences, a measure of information based on an analysis of words in sentences is used. It is assumed that the word occurrence in an article can be used to compile a set of significant word, and the frequency of such significant words within sentence reflects the "significance of sentence" in the text. According to Luhn, rare and frequent words in a text do not contribute much to the content of the text. Luhn stated that only words between two cut-offs, middle frequency words, can be determined as significant words for the text.

Besides extraction of significant words in an article, such an analysis can be also applied to the collection of documents to extract the most significant words to discriminate articles across the collection. In other words, significant words can be extracted on a corpus basis rather than a per-document basis. Luhn's original idea of counting frequencies can be used to provide weighting to words in order to discriminate documents in a collection. Following to this, it is showed that words appearing in low number of documents and words appearing in high number of documents are not good discriminators across the collection in [120]. They verified

Luhn's conclusion that words with middle frequency are the best discriminator, and words occurring in between 1% and 90% of texts have the highest discriminatory power across the collection. Pruning rare and frequent words is followed by many researchers in text categorisation tasks as it is a common belief that they are not good discriminators of classes [121].

However, as noted in [122] common words (or frequent words) contribute to the text categorisation contrary to a common belief that the removal of frequent words improves the performance of information retrieval methods. In their study, stop words are discarded before evaluation of the performances. They also examined another common assumption that rare words are informative and should not be removed; and concluded that words appearing in less than some pre-determined number of documents (up to 90% or more unique words) can be removed with either an improvement or no loss in the accuracy of text categorisation models. Similar conclusion is obtained for clustering in [123]. They investigate the side effect of vocabulary size to clustering algorithms. The results show that keeping frequent words leads to an improvement on the performance of models in general, while rare words can be removed without loss on the performance. It is also worth to mention that the score that is used to evaluate mutual information is almost 0 when using only words with 1 occurrences in the corpus. Similar results are observed in [124] for their scoring measure where rare words skew the distribution of score defined. A minimum threshold of 10 is used in our study to mitigate this issue.

In information retrieval systems, a common belief among researchers is that both frequent and rare words can be important for a specific field. However, the extraction of words from these two classes should be done by using two criteria not one as their distributions are different in specific areas [125]. Rare words can be topic-specific words and extracted by analysing their co-occurrence in the academic domain. It is stated that a rare word will be a topic-specific word if it is related to a huge number of other possible topic-specific words or words that are considered as informative with a large weight defined [126]. In [127], it is reported that most of rare words that are generally discarded in standard information extraction tasks can be topic-specific words in medical abstracts. They stated that even the frequency of 5 is too high for extraction of informative words in medical abstracts but words appearing only once is needed to be removed in information based statistical models [125, 126, 127].

3.5.4.2. ***Characteristics of Words in the LScD and the Decision Taken***

In practice, word selection strategy is fundamentally important for different text processing tasks since it determines the space of words that can be obtained from the texts and be effectively used for a specific task. Differences in types of corpora must also be considered as a complementary effect in the selection of words. In order for the differences between rare/frequent words' importance in two corpora to be explainable, corpora should be comparable by sampling in the same way. For example, it is natural to expect that frequent words of a topic-specialised corpus are different from frequent words appearing in a general corpus where its texts are from a wide variety of different domains. For information extraction problems, frequent words in a topic-specialised corpus are likely to be extracted as content words while such words can be assessed as non-informative in a general corpus. As mentioned before, 5 occurrences of a word in a topic specific corpus (e.g. medical abstracts) may be too high when compared with a general corpus of the same size. However, one can find that words with 5 occurrences are useless for text categorisation tasks due to its score in probabilistic models (e.g. entropy).

Essentially, rare words fall into two classes: those which are rarely used in the corpus and those which are misspelling. There are several reasons for the first class. It may be because it is used very uniquely like names referring people, places, brands or products. Rare words may also refer infrequent usage or synonyms of words. Similarly, shorten version and abbreviations of words can cause a huge number of rare words. Particularly, those who use corpus containing texts from medical or chemistry domains will tend to see huge number of shorten words, abbreviations and also chemical formulas. The second class of rare words involves words that are misspelled in the writing. Especially, such words is one of the main factor contributing the number of words occurring once in large corpora. As one would expect, a list with all correctly spelled words would not be realistic, especially for a large corpus. In [**128**], it is predicted that 38% of 42,340 words, from a collection of life science abstracts, are misspellings. For both classes of rare words, one needs to be careful about removing them. The decision of cut-off for rare words should be determined individually for each corpus depending on the characteristic of the corpus such as type and size.

Therefore, a natural question arises: what is the optimal cut for rare words in LSC? A simple initial characterisation is taken into account. As mentioned before, LSC is a collection which texts are from 252 different categories. Two expected consequences of this fact are: the identification of informative rare words for text categorisation by using their co-occurrences with other words of the corpus is not reasonable for our case; and it is very likely to observe words occurring only once in the corpus. The first consequence is caused by the fact that two rare words that used in texts from two well-separated categories will tend to be associated with each other due to co-occurrence of these words with the same subset of other words. In the case that the subset of related words has a large weight in terms of containing informative words but one of rare words is actually not informative, the selection of this rare word will be biased on the other one. When considering the large size of

LSC, having a large number of categories has also a side effect: many misspellings and unique names. In fact, approximately 60% of words appear in only 1 document in LSC (Table 3.3). Casual observation of words showed that many of them are non-words or not in an appropriate format to use (e.g. misspellings); therefore, they are likely to be non-informative signals (or noise) for algorithms. Some examples of such cases in LSC for randomly selected rare words are presented in Table 3.4. Our basic assumption is that too rare words are not informative for text categorisation, or not effective in the performance of methods.

In order to mitigate this issue, we set a minimum cut (10) so that words appearing in less than the cut-off will not be included in further analysis. There is no trivial way to decide the optimal cut. We took decision that the threshold which is not too low or high to be able to keep a reasonable number of words for analysis under the assumption that rare words can be relatively informative and they should not be removed aggressively [**122**]. The criteria, removing rare words to improve the performance of information-based text categorisation methods, is taken into account with an attention to have a noticeable impact on size of dictionary and results.

The Fig. 3.3 shows the number of words contained in the corresponding or less number of documents. To explore the fragment where words are very rare, we generate an enlarged view on a fragment in the Fig. 3.4. For instance, there are 592,161 words containing in only 1 document and 711,150 words containing in 2 or 1 documents. We can conclude from the figures and Table 3.3 that 870,015 words out of 974,238 words in LScD are contained in 10 or less than 10 documents, thus, it is reasonable to consider such words as non-informative signals in the corpus of 1,673,824 documents and can be removed from the dictionary for further analysis. If such words are removed from the dictionary, the number of words becomes significantly reduced from 974,238 to 104,223. Note that we did not exclude any frequent words in this stage as stop words are already removed in pre-processing steps.

Fig. 3.5 and 3.6 present the normalised number of words contained in the corresponding (n) or less number of documents. The data are normalised using (maximum-number of words) on y-axis. This means that the plot shows the number of words appearing in more than n documents against these numbers of documents. We observe more or less a negative relationship on a logarithmic scale. However, a remarkable pattern emerges: the plot does not follow a single straight line which fits the data. Instead, the plot starts with a linear behaviour, and a curve is observable in the right tail of the distribution, which follows a Pareto distribution behaviour. Pareto originally purposed that the number of people with incomes higher than a certain limit follows a power law [**129, 130, 131, 132, 133**]. The Pareto principle is also known as 80-20 rule. The rule states that 80% of people's income is held by the top 20% of income recipients in the society. Such characteristic of the distribution is also very typical property for the distribution of words over documents in text data.

TABLE 3.4. Some of rare words in LScD with the number of documents containing them. The last column shows the description of the word provided by checking the papers containing the word in WoS database, and possible reason why it is rare.

Word	**Number of documents containing the word**	**Description of the word and possible reason why it is rare**
luhman	4	An author name
lazerian	5	An author name
goodluck	2	A name (President Goodluck Jonathan)
hansel	5	A name (a name in fairy tale Hansel and Gretel and an author's name)
masculina	8	A marine specie: Appendix Masculina (Latin name)
heterocop	1	A freshwater specie: Heterocope Borealis (Latin name)
lunac18(2)	1	A term in Chemistry
wr3	3	A term in Agriculture (a water regime)
gausian	3	Misspelling- Gaussian
antilmog	1	Misspelling in the database: AntiLMOG (correct writing in the paper is anti-MOG.
acetosa	10	A plant specie Rumex Acetosa (another usage is 'sorreal' appearing in 13 documents)
ansdic	1	An abbreviation for Ammonium Nitrate and Sodium Salt of Dichloroisocyanuric
18cm	10	Non-word
000009sl	1	Non-word (from the expression 'DW=0.000009SL(3.047)')
limite	8	French word
resultan	1	French word
resultadoscon	1	Spanish word with error: ResultadosCon ('resultado' means result in English and appears 90 times in the LSC)

Under Pareto principle, the number of words appearing in more than n documents can be modelled as a power law:

$$N_x = \frac{\beta}{x^\alpha} \tag{2}$$

where N_x is the number of words, x is a certain documents limit and α and β are constants.

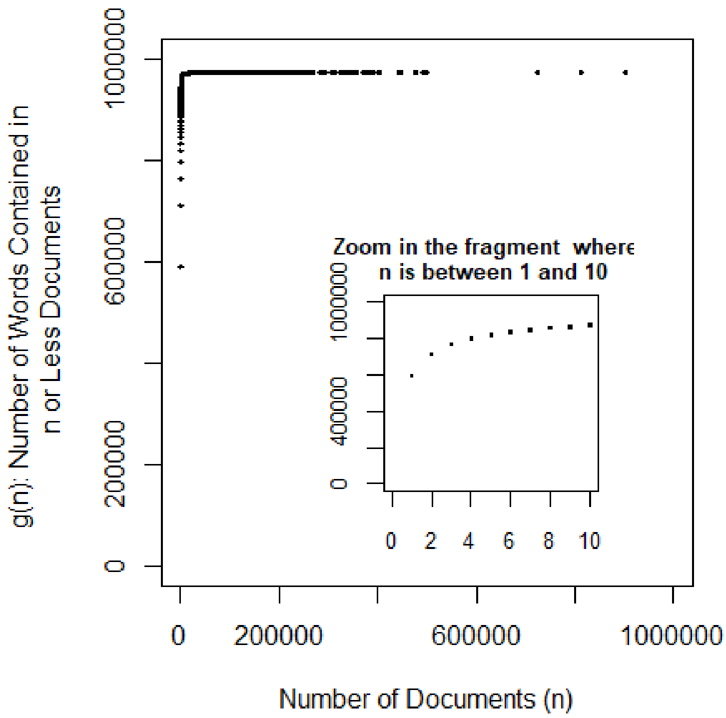


FIGURE 3.3. The number of documents (n) versus the number of LScD words contained in n or less documents in the LSC

A more general description of the Pareto principle is stated by Pareto distribution. Pareto distribution is a two parameter distribution to fit the trend that a large portion of data is held by a small fraction in the tails of distribution (heavy-tailed distribution) [134]. The distribution is characterised by a shape parameter α and a location (scale) parameter x_m. The *tail function* and the cumulative distribution function of a Pareto random variable X are given by [135, 136]:

$$P(X > x) = \begin{cases} (\frac{x_m}{x})^\alpha & x \geq x_m \\ 1 & x < x_m \end{cases}$$

and

$$F(X) = \begin{cases} 1 - (\frac{x_m}{x})^\alpha & x \geq x_m \\ 0 & x < x_m \end{cases}$$

where x_m is the (necessarily positive) minimum value of X (the lower bound of the data). The density function is defined as

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \\ 0 & x < x_m \end{cases}$$

For $0 < \alpha \leq 1$, the distribution is heavy-tailed and the right tail becomes heavier as α decreases .

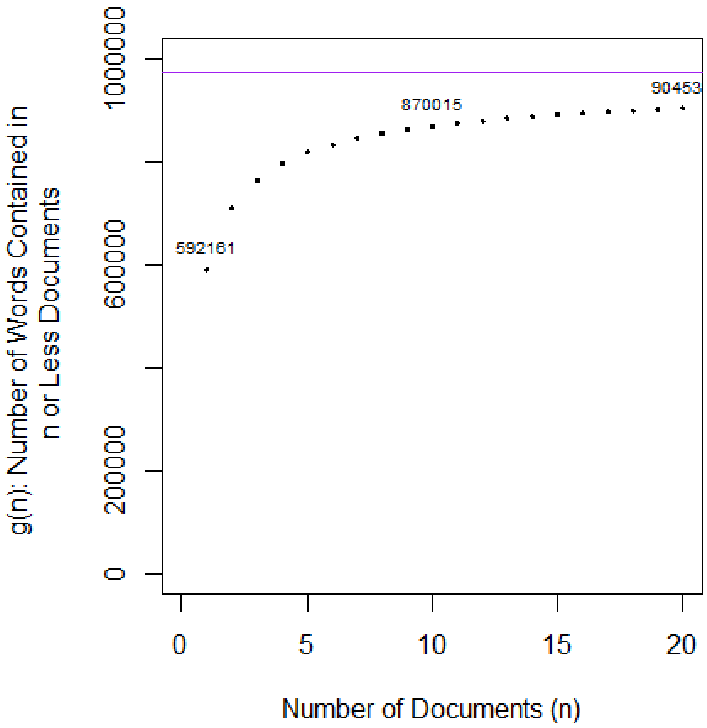


FIGURE 3.4. The number of documents (n) versus the number of LScD words contained in n or less documents in the LSC for those words appearing in at most 20 documents. The horizontal line indicates the number of words in the dictionary (974,238).

In Fig. 3.6, power-law behaviour in the upper tail is well documented. The Pareto distribution (Equation 2) is fitted to the data and resulting graphs are also shown in Fig. 3.5. Table 3.5 presents the estimated parameters and the mean squared error (MSE).

TABLE 3.5. Estimated parameters of Pareto distribution and the Mean Squared Error (MSE) for LScD

α	β	**MSE**
0.5752	388,756	25188

If the logarithm of the number of words appearing in more than a certain number of documents is plotted against the logarithm of these numbers of documents, a straight line (see Fig. 3.6 (b)), where the slope is α, is obtained. α is also known as *Pareto index*.

Due to the characteristic of the data, the log-log plot presents a noisy and diffused behaviour in the upper tail. This is actually because 81 observations fall in the interval 1-100 on y-axis, while there are 5,453 observations lying in the interval 10,000-1,000,000. However, on the logarithmic scale, the size of intervals 1-100 and 10,000-1,000,000 are the same, this leads to a diffusion in the tail. From a heuristic point of view, the plot suggests that there are three subset of words in the collection:

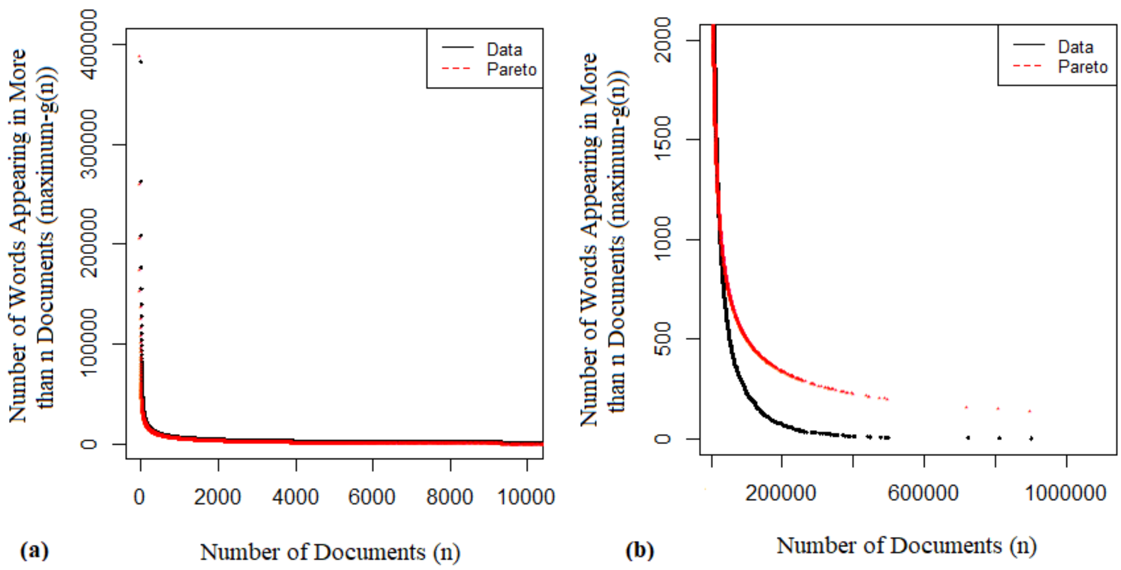


FIGURE 3.5. The number of documents (n) versus the number of LScD words appearing in more than n documents in the LSC for (a) $n < 10,000$ and (b) $n \geq 10,000$. The y-axis is calculated by normalising $g(n)$ to the maximum (maximum- $g(n)$), where $g(n)$ is the number of words contained in n or less documents. The black points are the data; the red points are the fitted Pareto distribution.

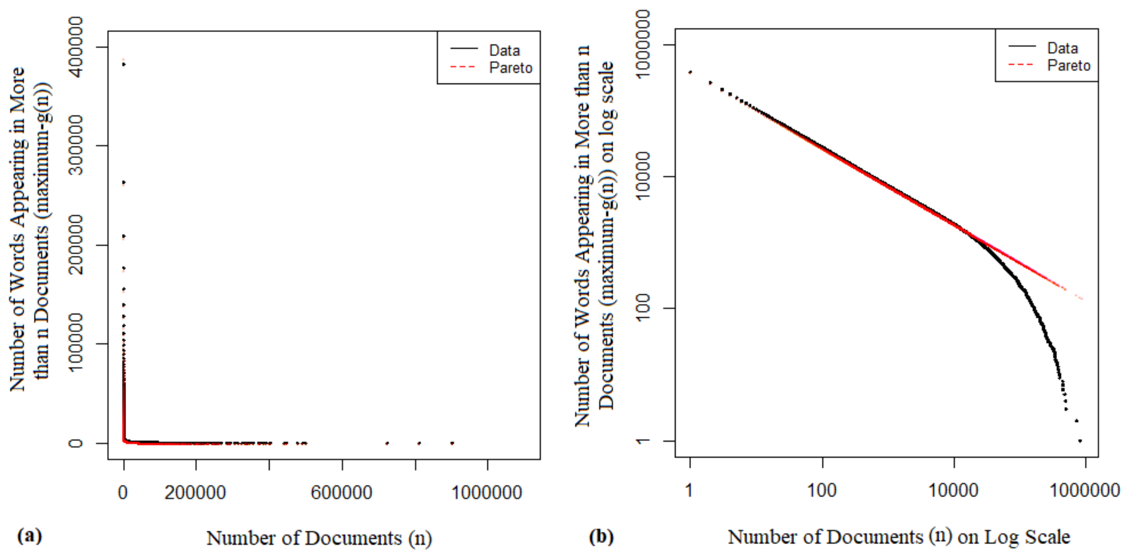


FIGURE 3.6. (a) The number of documents (n) versus the number of LScD words appearing in more than n documents in the LSC for whole data (b) The same plot on logarithmic scales. The y-axis is calculated by normalising $g(n)$ to the maximum (maximum- $g(n)$), where $g(n)$ is the number of words contained in n or less documents. The black points are the data; the red points are the fitted Pareto distribution. In (b), the slope of the line is -0.5752.

too rare words, mid-frequent words and frequent words. A straight down-sloping line covers words the largest part of the list, in which words are not too rare and frequent. It is not actually surprising as words occurring in a few or almost all documents tend to be more evenly diffused across the corpus.

3.6. Leicester Scientific Dictionary-Core (LScDC)

LScDC is an ordered sub-list from existing LScD [68]. There are 104,223 unique words (lemmas) in the LScDC. To build the LScDC, we decided the following process on LScD: removing words that appear in no more than 10 documents (≤ 10). As mentioned before, such words do not contribute much to discrimination of texts as they appear in less than 0.01% of documents. Ignoring these words has the advantages on the reducing the size of words for applications of text mining algorithms. The core dictionary is also sorted by the number of documents as in LScD.

Table 3.6 summarizes the number of words before and after removal. 870,015 words are removed from the LScD, that is, around 89% of words are removed. After removing such words, we also re-check the number of words in each document to affirm that all abstracts have at least 3 words. We note that in this stage "the number of words in an abstract" does not indicate the length of the abstract but the number of unique content words from the LScDC. After removing 870,015 words from the pre-processed abstracts, all documents have at least 3 unique words. None of documents are removed in this stage.

3.6.1. Organisation of the LScDC

In the LScDC, unique stemmed words, the number of documents containing the word and the number of appearance of the word in the entire corpus are recorded on each line in separated fields in the same way as for the LScD [67, 68].

3.6.2. Characteristics of Words in the LScDC

After cleaning words appearing in no more than 10 documents, the distribution of words over documents is presented in Fig. 3.7. As one can expect, we observe the same behaviour here that very few words occur very often, very many words occur very rare in the collection.

The Fig. 3.8 and Fig. 3.9 show the number of words contained in the corresponding or less number of documents with and without rescaling the x-axis. We can conclude that approximately half of words occur in less than 30 documents.

TABLE 3.6. Number of words before and after removing words appearing in no more than 10 documents in the LSC

	Number of Words
LScD	974,238
LScDC	104,223

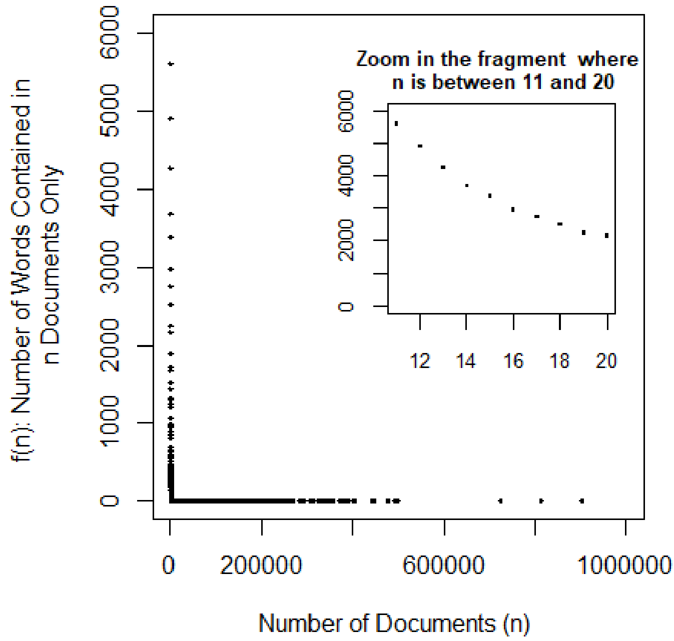


FIGURE 3.7. The number of documents (n) versus the number of LScDC words contained in n documents only after cleaning words appearing in no more than 10 (≤ 10) documents

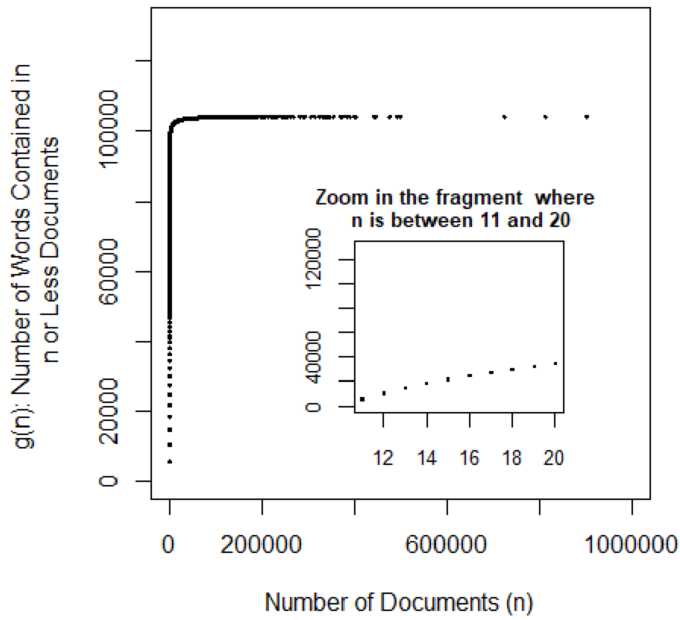


FIGURE 3.8. The number of documents (n) versus the number of LScDC words contained in n or less documents in the LSC after cleaning words appearing in no more than 10 (≤ 10) documents

Fig. 3.10 demonstrates the normalised number of words contained in the corresponding or less number of documents after removing words appearing in no more than 10 documents. The data are normalised using (maximum-number of words) on y-axis as in Fig. 3.6. As expected, noisy behaviour in the lower tail is avoided.

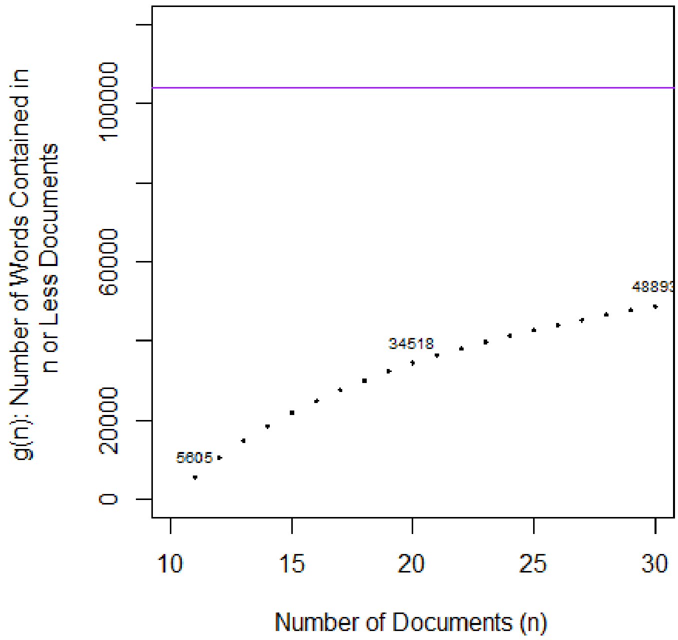


FIGURE 3.9. The number of documents (n) versus the number of LScDC words contained in n or less documents in the LSC (for those words appearing in at most 30 documents) after cleaning words appearing in no more than 10 (≤ 10) documents. The horizontal line indicates the total number of words in the dictionary (104,223).

A downward linear trend is observable at the beginning and a curve is present in the upper tail. From a heuristic point of view, words can be group into two subsets: mid-frequent words and frequent words.

The plots in Fig. 3.10 reveal power-law behaviour (Pareto distribution) in upper tail of documents distribution, but apparently not for the lower tail as expected. The estimated parameters by fitting the power-law (Equation 2) to the data is presented in Table 3.7.

TABLE 3.7. Estimated parameters of Pareto distribution and the Mean Squared Error (MSE) for LScDC

α	β	**MSE**
0.5796	397,707	10737

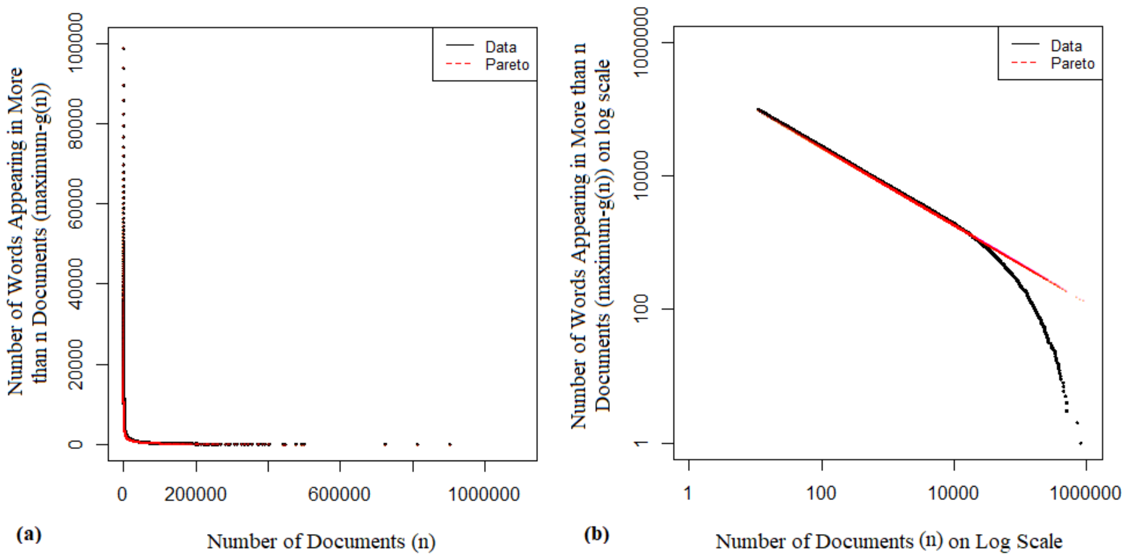


FIGURE 3.10. (a) The number of documents (n) versus the number of LScDC words appearing in more than n documents in the LSC for whole data (b) The same plot on logarithmic scales. The y-axis is calculated by normalising $g(n)$ to the maximum (maximum- $g(n)$), where $g(n)$ is the number of words contained in n or less documents. The black points are the data; the red points are the fitted Pareto distribution. In (b), the slope of the line is -0.5797.

3.7. A Comparison of LScDC and NAWL

This section provides a comprehensive study of comparison of the NAWL and the LScDC. Several different approaches are taken into account based on direct comparison of words and comparison of ranks of words in two dictionaries.

3.7.1. Academic Word List (AWL) and New Academic Word List (NAWL)

Academic word list (AWL) is developed from a written academic corpus with 3.5 million running words [70]. The corpus is gathered from four discipline specific sub-corpora: arts, commerce, law and science with a seven sub-disciplines for each (see Table 3.8). Each sub-corpora has approximately 875,000 running words. The list of words is collected from a total of 414 academic texts in the form of textbooks, articles, book chapters, and laboratory manuals.

A *word family* is defined as the collection of words that appears in various form of the same word (e.g. indicate and indication are in the same family). To select words, three rules are taken into account:

- Specialised occurrence: Academic list does not contain the General Service List (GSL) published by West [137], defined as the first 2,000 frequent words of English.
- Range: The number of appearance of a family member has to be at least 10 in each of main discipline, and 15 or more in 28 sub-disciplines.

TABLE 3.8. Corpus structure of the AWL

Discipline			
Arts	**Commerce**	**Law**	**Science**
Education History Linguistics Philosophy Politics Psychology Sociology	Accounting Economics Finance Industrial relations Management Marketing Public policy	Constitutional Criminal Family and medicolegal International Pure commercial Quasi-commercial Rights and remedies	Biology Chemistry Computer science Geography Geology Mathematics Physics
122 texts	107 texts	72 texts	113 texts
883,214 words	879,547 words	874,723 words	875,846 words

- Frequency: The number of appearance of a family member has to be at least 100 in the academic corpus. Frequency is the secondary criteria for range.

AWL includes 570 word families. It covers 10% of the total words in academic texts. In addition, words in West's GSL and words in AWL together (GSL/AWL) cover approximately 86% of total words in academic corpus. In Coxhead words, "Academic words (e.g. substitute, underline, establish, inherent) are not highly salient in academic texts, as they are supportive of but not central to the topics of the texts in which they occur."

The New Academic Word List (NAWL) is then created by [69] based on an updated and expanded academic corpus of 288 million words with modified lexemes. The corpus, which NAWL is created from, includes Cambridge English Corpus (CEC), oral academic discourse (Michigan Corpus of Academic Spoken MICASE and British Academic Spoken English BASE), and textbooks (Corpus of 100s top-selling academic textbooks). From CEC, the frequency generated word list is used as one group. This made up the largest proportion of total tokens, about 86% (over 248 million words). The oral corpora and the corpus of textbooks are divided into four main categories: Arts and Humanities (AH), Life and Medical Sciences (LS), Physical Sciences (PS), and Social Sciences (SS). The number of tokens for each group in the corpus is presented in Table 3.9. The list is developed by the conjunction with New General Service List (NGSL) as in Coxhead's GSL-AWL. NGSL is also created by [69], which is based on 273 million words from CEC academic.

NAWL contains of 963 word families. While combined GSL/AWL covers approximately 87% of the new corpus, the NAWL covers 92% of the corpus when combined with NGSL. Therefore, NAWL gives an improvement in coverage, with about 5% more coverage [22].

TABLE 3.9. Corpus Structure of the NAWL

Source		# of tokens
Cambridge English Corpus (CEC)		248,666,554
Oral Discourse	Arts and Humanities	803,113
	Life and Medical Sciences	749,610
	Physical Sciences	686,926
	Social Sciences	852,990
Textbooks	Arts and Humanities	6,082,267
	Life and Medical Sciences	16,822,357
	Physical Sciences	4,467,629
	Social Sciences	9,044,779

In the published list of academic words (NAWL), the authors computed the statistics SFI (Standard Frequency Index), U (Estimated Word Frequency per Million) and D (Dispersion) to describe the number of occurrence of the words and the distribution of words in their corpus. To illustrate the information given in the list, we present Table 3.10 that shows 10 words with statistics in the NAWL, ordered by SFI values [138, 139, 140].

TABLE 3.10. Sample words with highest SFIs from NAWL

Word	SFI	U	D
repertoire	72.452	1759	0.5923
obtain	66.519	449	0.7531
distribution	65.665	369	0.6863
parameter	64.369	273	0.6943
aspect	64.190	262	0.9385
dynamic	63.506	224	0.8548
impact	63.491	223	0.9426
domain	63.467	222	0.8276
publish	62.897	195	0.9039
denote	62.571	181	0.7035

D shows the uniformity of frequency of the word in subject categories of NAWL in a 0-1 scale: 0 means that the a word (all forms) appears in a single category, 1 means that frequencies are distributed over all categories proportionally to the total number of words (all inflected forms of words) in a category. U is the estimated frequency per million. It is derived from the frequency of the word in the corpus with an adjustment for D. SFI indicates frequency derived from U in a 0-100 scale. Higher scores of SFI show greater frequency [141]. A word family with SFI=90 occurs once in every 10 tokens (all words with different inflected forms in the corpus); a word with SFI=80 occurs once in every 100 tokens [138, 139].

3.7.2. Difference Between the Principles in Preparation of the LScDC and the NAWL

Both the NAWL and the LScDC are actually made up of academic texts distributed over multiple categories for building academic lists of words. In this manner, two lists seem similar. However, more detailed analysis shows that they differ one another in many respects such as types of texts where words are extracted (e.g. full-text or a part of the text), kind of words included, dictionary size and the statistics used to extract words.

Let us begin with corpora where two dictionaries are created. An obvious difference of corpora lies in the types of texts. As types of texts, we meant the NAWL having extracted from full-texts from academic domains and the LScDC having extracted from abstracts of articles. This is actually an important difference as there is side effect of word limit for an abstract such as the frequency of a word and the vocabulary used. In this case, it is likely to observe changes in the statistics calculated for each word and respectively the ranks of words. The change in statistics may lead to select different words as word selection in NAWL is based on frequency and range. One other difference between two corpora is that NAWL contains oral academic discourse as well as written texts while LSC includes only written academic English. This may have influence on the words listed as spoken and written English are often different in terms of vocabulary used.

It is worth to stress that the calculation of statistics for words in the NAWL and the LScDC are different. In the NAWL, words are selected based on SFI derived from frequency. The dispersion (D) of words over categories is calculated to adjust frequency in SFI calculation. However, in LScDC words are simply sorted by the number of documents containing words. The dispersion of words and SFI are both taken into account to select words in NAWL, not all words appearing in the corpus are included in the NAWL. This difference leads to firstly difference in ranking of common words in both dictionaries, secondly kinds of words and words selected and respectively the size of the dictionaries.

One of the major differences lies in the kind of words. According to the Coxhead, words in AWL are supportive of the academic text but not central to the topics of the text [70]. Words in AWL account for approximately 10% of the total words in the collection of academic texts. The AWL and GSL (general service list) together cover approximately 86% of total words in academic corpus. By updating this list with an expended corpus of 288 million words (NAWL), the coverage was improved to 92% of new corpus when combined with NGSL (New General Service Words), with approximately 5% improvement. By a casual observation of the NAWL, one can see the same property for words in the NAWL. Words in NAWL are not much specialised technical terms such as names of chemicals or names. In contrary, LScDC contains both supportive and topic-specific words such as mathematical terms, chemical elements, names, biological species and many more. As or aim is to quantifying meaning of research texts, we kept such words in LScDC.

Such differences in word selection also effected the size of dictionaries. As expected, the LScDC is much more larger than the NAWL, namely 963 word families in the NAWL and 104,223 lemmas in the LScDC.

3.7.3. Comparison of the LScDC and the NAWL

This section describes a study of comparison of the LScDC [68] and the NAWL. Our primary focus is on obtaining the coverage of NAWL by LScDC, and on analysing how the rank of words in both dictionary are related.

3.7.3.1. *Coverage of the NAWL by the LScDC*

One feature of NAWL is that words are listed by headwords of word families from combination of their derived forms. When comparing with LScDC, headwords with different inflected forms indicate the same stemmed word in LScDC (see Table 3.11). In order to examine the agreement between NAWL and LScDC, we processed stemming to headwords in NAWL. This process returns various forms of each headword into a common root as in LScDC. After stemming, words in NAWL are eliminated, with a decrease number from 963 to 895. Note that as SFIs of two headwords, having actually the same root, are different, we used the average of SFIs for unique stemmed words.

TABLE 3.11. Headwords and inflected forms in the NAWL, and stems of the headwords in the LScDC

Headword in NAWL	**Inflected Forms in NAWL**	**Stemmed Headword in LScDC**
accumulate	accumulates, accumulated, accumulating, accumulatings	accumul
accumulation	accumulations	accumul
acid	acids	acid
acidic	acidics	acid

For purpose of comparison of dictionaries, stemmed words are used. Table 3.12 illustrates the comparison of dictionaries by showing the coverage of the NAWL words by the LScDC. The overlap between the LScDC and the NAWL is 99.6%, with 891 word occurring in both. This means 4 words occurring only in NAWL: "ex", "pi", "pardon" and "applaus". The lower coverage of the dictionary seems to be the result of differences in types and processing of texts in corpora. The corpus of NAWL includes full texts from academic domain [70], while LSC is made up abstracts of texts in LSC.

TABLE 3.12. Coverage of the NAWL by the LScDC

Number of Words in NAWL	Number of Words in NAWL (after stemming)	Coverage of NAWL by LScDC (#)	Coverage of NAWL by LScDC (%)
963	895	891	99.6%

The reason why "pi" does not occur in LScDC lies in the nature of abstracts and also in the usage of this word in articles. It is commonly used by the symbol π (pi) in the math world, and not many articles include formulas in abstracts. Uniting prefixes with the following words is the reason that the word "ex" does not occur in LScDC. For instance, words such as ex-president and ex-wife are converted to expresident and exwife in pre-processing step. The other two words "pardon" and "applaus" are not included in LScDC. However, they occurred in LScD before removing words that appear in no more than 10 documents, with very low occurrences in documents (5 and 9 respectively). Similarly, these two words have low ranks on the NAWL: rank 924 and rank 956 in the list.

We also evaluated different comparison scheme that is focused on a subtly different goal: to give an understanding about what fragment of LScDC contains the NAWL. This analysis performs a search of NAWL words over a specific subset of our rank ordered dictionary, repeatedly searching NAWL words in various subsets of the dictionary. Table 3.13 and Fig. 3.11 show the coverage of NAWL in particular fragments of LScDC. From this perspective, we see that NAWL is covered in the first 89,351 (85.7% of all words) words of LScDC, where the frequency of 89,351th word is 14. Observe that when doubling the number of words from 40,000 to 80,000 there are only 8 more words found in LScDC. This means the majority of NAWL is contained in the first 38.4% of LScDC. The number of documents containing 40,000th and 80,000th words are 16 and 53 in the LSC. It is remarkable that in 10,000 words, the coverage of the NAWL is 90.9%, with a frequency of 572 in LSC. This may be considered that the NAWL is representative of our 10,000 words (9.6% of LScDC). This partly supports that wide range of LScDC is constructed by more specific terminologies of academic disciplines. This is explainable given the variety of texts' categories in corpus, differences in selection methods of words and the fact that abstracts have slightly different writing structure and words.

An alternative view of fragment comparison is to evaluate the last position of the words of a specific fragment of the NAWL in LScDC. Table 3.14 shows the fragment of NAWL and positions in LScDC. Both dictionaries are ranked by their frequencies: SFI for NAWL and the number of documents containing the word for the LScDC. We see that the first half of NAWL words is in approximately the first

Table 3.13. Coverage of the NAWL by the fragments of LScDC. The last column presents words of NAWL which are found between two fragments in LScDC.

Fragment of LScDC		Coverage of NAWL by LScDC		Words added between two fragment
#	%	#	%	Words
1,000	1.0%	231	25.8%	
5,000	4.8%	678	75.8%	
10,000	9.6%	814	90.9%	
15,000	14.4%	845	94.4%	
20,000	19.2%	860	96.1%	
25,000	24.0%	877	98.0%	
30,000	28.8%	879	98.2%	bizarr, terribl
35,000	33.6%	882	98.5%	comma,sneez,jazz
40,000	38.4%	882	98.5%	
45,000	43.2%	884	98.8%	sniff, handout
50,000	48.0%	888	99.2%	unintellig, cheer, footnot, ridicul
55,000	52.8%	888	99.2%	
60,000	57.6%	888	99.2%	
75,000	72.0%	889	99.3%	nasti
80,000	76.8%	890	99.4%	parenthesi
89,351	85.7%	891	99.6%	whoever

14% of LScDC. When the fragment of words in NAWL doubles, the position became around 5 times far from the first word in LScDC. We see that there are 300 words lies between 10,967th and 14,017th words in LScDC (100-400), with an interval of approximately 4,000 words. This interval is around 9,000 for the next 200 words, and followed by an interval of 55,000 for the third 200 words. Thus, we conclude that there are dense regions of LScDC in terms of the coverage of NAWL.

3.7.3.2. *Comparison of Ranks of Words in Two Dictionaries*

Our second approach to compare two lists is based on the order of words. The goal is to examine whether the ranking of words (frequency-based sorting) in dictionaries are actually similar. Note that only common words in both dictionaries (891 words) are taken into account. Words in both lists are descending ordered by their ranks in corpora, which are the number of documents containing the word in

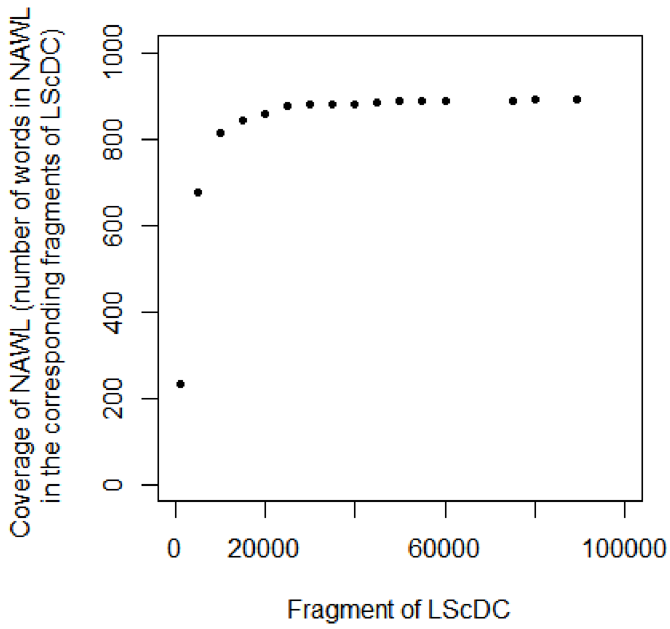


FIGURE 3.11. Coverage of the NAWL by the fragments of LScDC

TABLE 3.14. Last position of the words of a specific fragment of the NAWL in LScDC. Both dictionaries are sorted by their frequencies defined.

Fragment of NAWL	Last position of words of NAWL in LScDC	Fragment of LScD (%)
100	10,967	10.5%
200	10,967	10.5%
300	10,967	10.5%
400	14,017	13.4%
500	17,212	16.5%
600	23,188	22.2%
700	78,492	75.3%
800	78,492	75.3%
891	89,351	85.7%

LScDC and SFI in NAWL. Table 3.15 shows stemmed versions of top 10 words with corresponding statistics in two lists.

From an inspection of order of words, 7 words in the lists is in the same order in both dictionaries when LScDC is restricted by NAWL words. Such words are listed in the Table 3.16. Thus, the direct comparison of order cannot be used.

A new evaluation method is offered that focuses on pairwise comparison of partitions in dictionaries. The word lists are divided into smaller sub-lists, with the same number of intervals. We introduce an analysis that is focused on the overlapping words in intervals by counting the number of words in common. Within intervals,

TABLE 3.15. The top 10 words in stemmed form with corresponding statistics in lists. Blue coloured words are matches in the top 10 of two dictionaries.

Word in NAWL	**SFI in NAWL**	**Word in LScDC**	**The number of documents containing the word in LScDC**
repertoir	72.45	effect	476,757
obtain	66.52	compar	355,381
distribut	65.67	activ	255,630
paramet	64.37	observ	249,965
aspect	64.19	found	234,720
dynam	63.51	import	233,138
impact	63.49	indic	229,775
domain	63.47	demonstr	218,861
publish	62.89	obtain	218,578
denot	62.57	condit	205,643

TABLE 3.16. Words in the same order in both dictionaries. The LScDC is restricted by NAWL words, we ignore other words to compare raking of words in dictionaries.

Word	**Order of word in the lists**	**Number of documents containing the word in LScDC**	**SFI in NAWL**
acut	182	30,876	57.72
decay	368	12,761	55.66
horizon	543	4,897	53.83
portfolio	656	2,299	52.13
kilomet	778	872	49.14
cheat	844	310	46.02
handout	883	51	42.85

the common words are counted, and then the percentage of pairwise intersection of parts (total overlap) are considered to be the agreement of rating between LScDC and NAWL. As would expected, the larger width of intervals (small number of splits) yields the highest agreement of rating. The highest possible width is 891 (only 1 split) as there are 891 words in lists. To find the percentage of total overlap within intervals, the following statistical computation is done:

$$\frac{\sum_i n_i}{N_t}$$

where n_i is the size of intersection in i^{th} interval, and N_t is the total number of words (891). We repeated the same calculation for different widths of intervals, with an increasing sequence 5, 10, 15, ..., 890, 891. For instance, when the width is 5 the lists are divided into 179 intervals: 178 complete interval with 5 words, 1 shorter interval with 1 word. Fig. 3.12 shows the fraction of the intersection in intervals with specified width. Observe that not in all cases lists are divided into equal intervals. For instance, the width 890 of interval means that there are two partitions with 890 and 1 words and so the comparison is not much meaningful in these cases. To avoid unbalanced classes, we consider only those number of intervals where partitions have almost equal widths. Fig. 3.13 and Table 3.17 show the number of intervals selected and the width of intervals for these intervals. When the lists are divided into two intervals, the fraction of overlap is 0.73. Hence, 27% of words of a list do not lie within the same half of the other list. In addition, almost half of words are in different intervals when splitting the lists into 3 intervals, with approximately 300 words in each interval (300 words in two intervals and 291 words in one interval). Our findings raise the possibility that two lists are slightly different in terms of ranking words within lists.

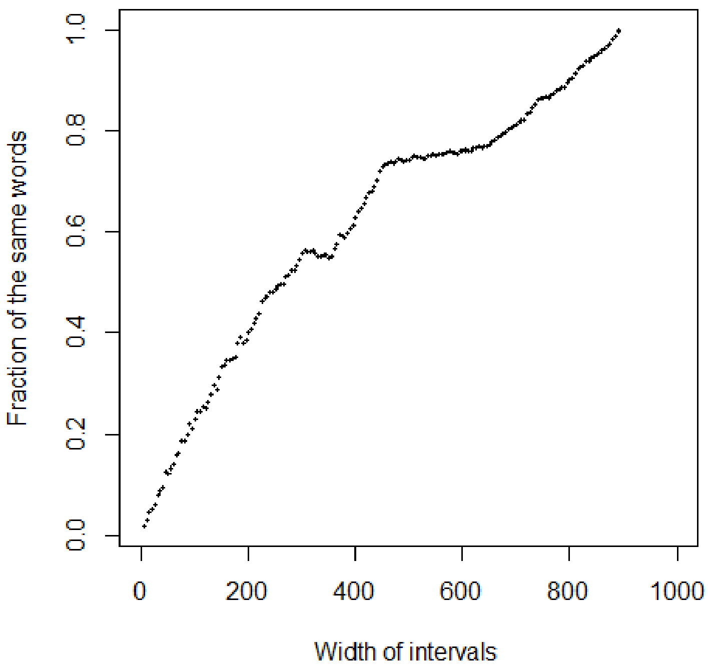


FIGURE 3.12. The fraction of the intersection of words in intervals with specified width

3.7.3.3. *Testing Similarity of Ranks in Two Dictionaries*

The scatter plot suggests a positive correlation between frequencies in the LScDC and SFI values in the NAWL (see Fig. 3.14). In order to test whether there is any or

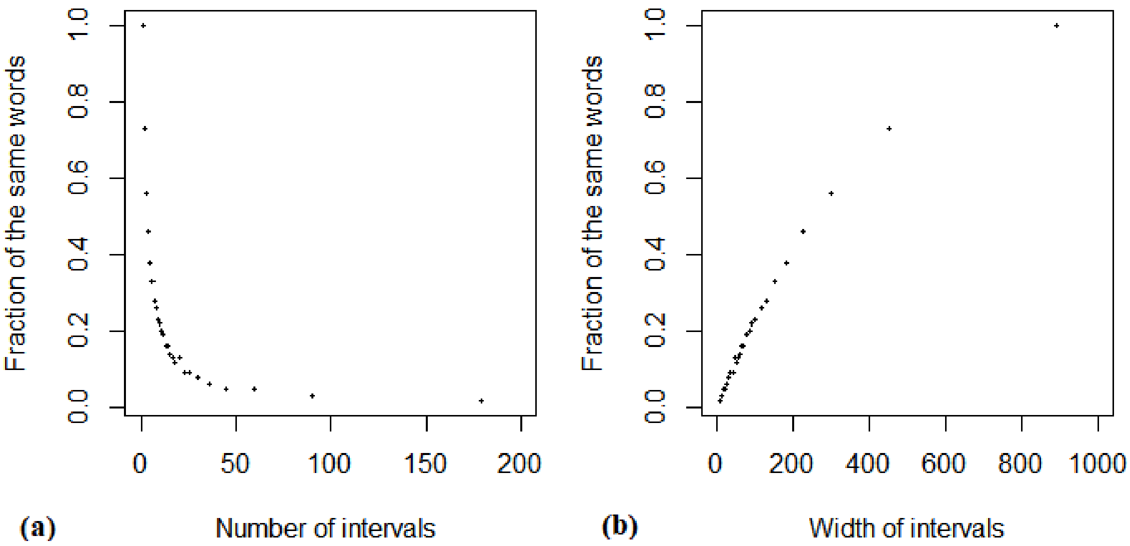


FIGURE 3.13. The fraction of the intersection of words in intervals with number of intervals and specified width. Figures present only those number of intervals and widths where partitions have almost equal widths (e.g. 2 intervals with approximately 450 words in each, 450 in one of intervals and 441 in the other interval).

TABLE 3.17. The percentage of the overlapping of words in intervals with number of intervals and width of intervals

Number of interval	Width of interval	Percentage of overlapping	Number of interval	Width of interval	Percentage of overlapping
179	5	1.8%	13	70	16.2%
90	10	2.9%	12	75	18.5%
60	15	4.5%	11	85	20.0%
45	20	5.1%	10	90	21.9%
36	25	6.2%	9	100	22.9%
30	30	8.0%	8	115	25.5%
26	35	8.6%	7	130	27.9%
23	40	9.3%	6	150	33.2%
20	45	12.6%	5	180	37.9%
18	50	12.2%	4	225	46.4%
17	55	13.2%	3	300	55.7%
15	60	14.1%	2	450	72.8%
14	65	15.8%	1	891	100.0%

no evidence to suggest that linear correlation of ranks is present in two dictionaries, the Spearman's Rank Correlation (SRC) is used. Spearman's correlation coefficient

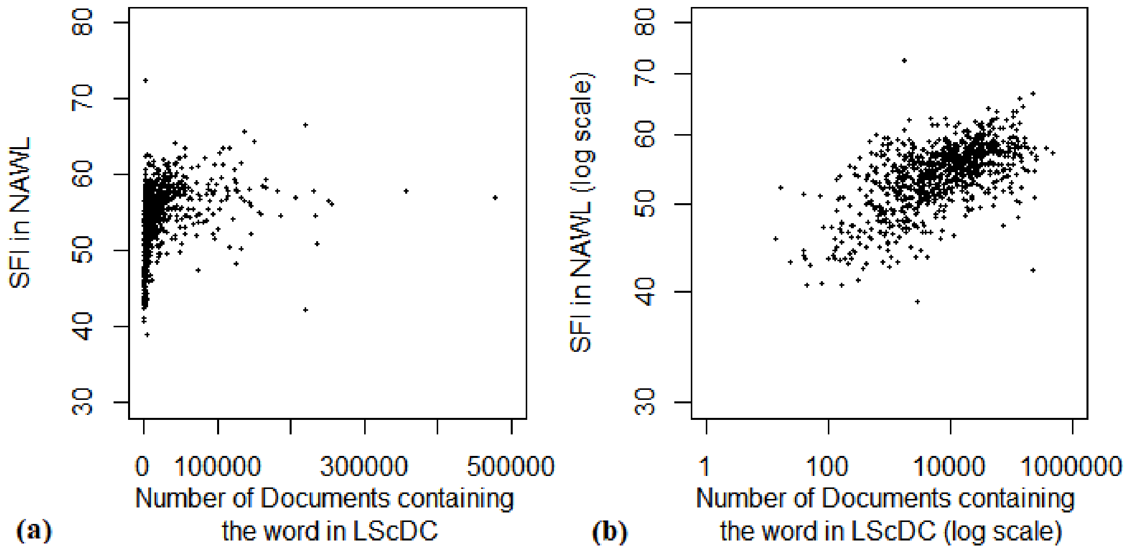


FIGURE 3.14. Relationship between the number of documents containing the word in LScD and SFIs in NAWL. The figure on the right hand side is on logarithmic scale.

is a statistical measure of the strength and direction of a monotonic association between two ranked variables. It is actually equal to Pearson's Correlation Coefficient (PCC) between two variables with ranked-values [142].

For a sample size n, the Spearman's coefficient R_s is computed as:

$$R_s = \frac{1 - 6(\sum d_i^2)}{n^3 - n}$$

where d_i is the difference in the ranks of each variable pair [143].

In this study, the Spearman's correlation is calculated by assigning a rank of 1 to the highest value within each list, 2 to the next highest and so on. Fig. 3.15 presents the relationship between ranks of words in lists. The correlation between words in two lists will be high when words have a similar rank within lists. The calculation of Spearman correlation for this study gives a value of 0.58 which confirms what was found in the comparison of ranks and what was apparent from the graph. There is indeed a moderate positive correlation between two lists, which are monotonically related. We also calculated the Pearson's correlation coefficient with frequencies and logarithmic scaled-frequencies, found 0.30 and 0.61 respectively (see Table 3.18). This is expected results because we did not observe a linear relationship of frequencies, but monotonic in Fig. 3.14. However, the logarithmic scaled-frequencies show a linear relation.

3.7.3.4. *An Alternative Comparison of Ranks of Words in Two Dictionaries*

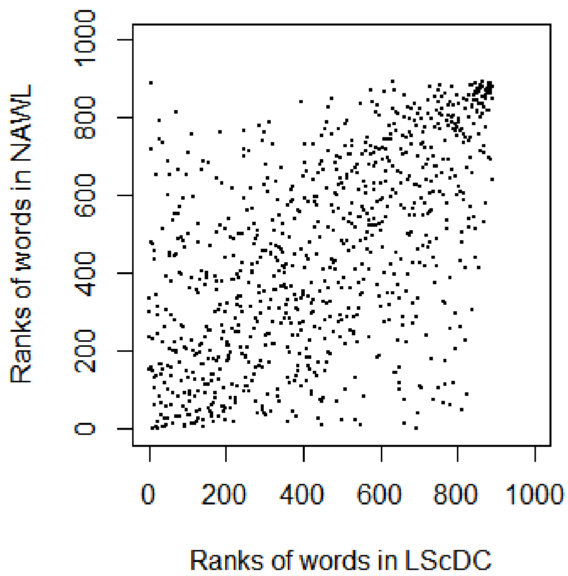


FIGURE 3.15. Relationship between ranks of words in lists (LScDC and NAWL)

TABLE 3.18. Correlation coefficients that measure the relationship between ranks of words in NAWL and LScDC: Pearson's Correlation Coefficient (PCC), Spearman's Rank Correlation (SRC) and PCC for logarithmic scaled frequencies.

Test	Test Statistics
PCC	0.30
SRC	0.58
PCC-log	0.61

Finally, we perform another analysis that is focused on ranks of words, similar to the comparison of ranks by partitioning intervals. Here, common words in both dictionaries (891 words) are used for analysis as in the previous comparison of ranks. The difference in this approach is the creation of intervals. Rather than dividing the whole lists into intervals, we consider the top n words by frequencies presented in dictionaries, where $n = 5, 10, 15, \ldots, 890, 891$. For instance, if $n = 5$ we compare the first 5 words in dictionaries, where words are ordered by the number of documents containing the word for LScDC and SFI for NAWL. Fig. 3.16 shows the number of overlapping of words in top words for specified top n words. Note that in the figures, words are in descending order by their frequencies in both dictionary. We see that there are only 2 common words in the first frequent 20 words of lists. In the top 100 words, this number is 25, which means 25% of words are common. This shows that the widely used words in corpora are slightly different. This may be result of the differences in calculations of statistics for words (the number of documents containing the word and SFI).

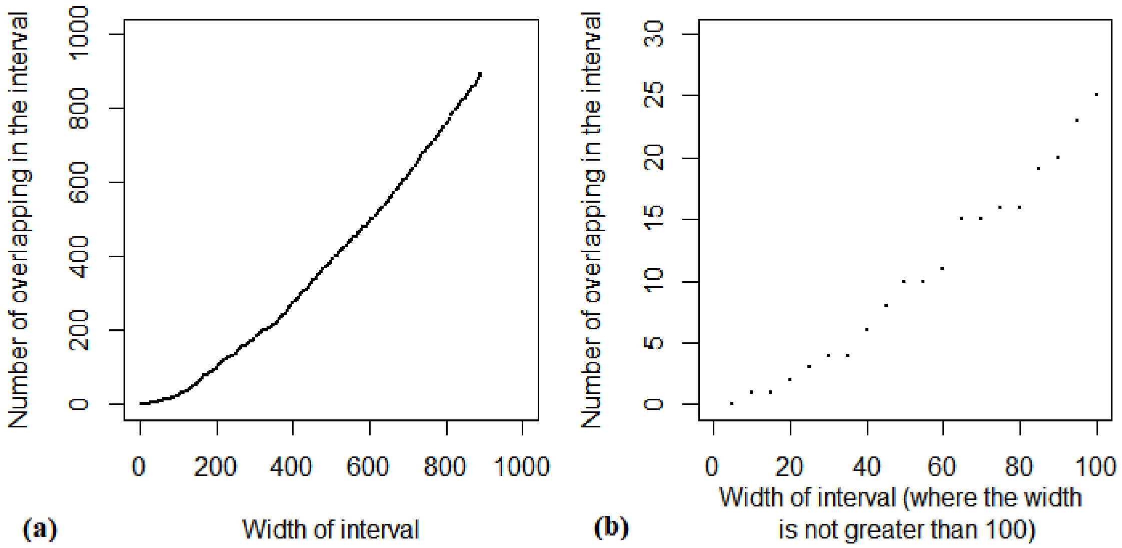


FIGURE 3.16. The number of overlapping of words in the top n words of the lists (width of interval) where $n = 5, 10, 15, \ldots, 890, 891$. Lists are in descending order by their statistics provided (the number of documents containing the word and SFI). The figure on right hand side presents widths until $n = 100$.

We repeated this analysis for ascending order of frequencies. In this case, we consider the bottom n words, where $n = 5, 10, 15, \ldots, 890, 891$. Fig. 3.17 shows the number of overlapping of words for specified bottom n words. We can see that the number of overlapped words for bottom is much more when comparing top words. There are 7 common words in the least frequent 20 words of lists and 50 common words in the bottom 100 words (50% of words). This means that there is an improvement in common words for the least frequent words. Dictionaries are more similar for bottom words.

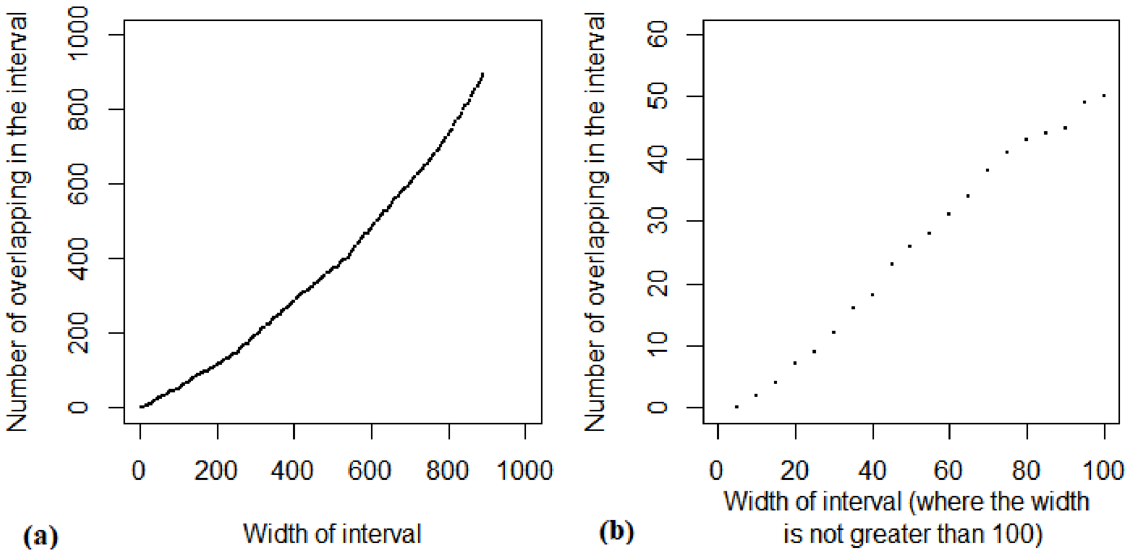


FIGURE 3.17. The number of overlapping of words in the bottom n words of the lists (width of interval) where $n = 5, 10, 15, \ldots, 890, 891$. Lists are in ascending order by their statistics provided (the number of documents containing the word and SFI). The figure on right hand side presents widths until $n = 100$.

3.8. Conclusion and Discussion

In this work, we presented LSC, LScD and LScDC with a description of the methodology and all steps in construction processes. Both the corpus and the dictionaries set out with the aim of quantifying the meaning in research texts in our future work.

LSC is a corpus of abstracts of academic articles and proceeding papers, where all papers are indexed by WoS and published in 2014 in English. It consists of 1,673,824 abstracts with the metadata: title, list of authors, list of subject categories, list of research areas, times cited. In 119 documents, list of authors are not present; however, we did not exclude them. The average length of abstract is 178 words (all words including stop words and different forms of words) with a minimum 30 and a maximum 500 words. Each paper in WoS is assigned to at least one of subject categories and research areas. The number of subject categories that a paper is assigned to vary from 1 to 6 in the LSC.

We then developed the LScD by extracting unique words (excluding stop words and various inflected forms of words) from the LSC. LScD is a scientific dictionary where all words are in stemmed form. It consists of 974,238 words; the number of documents containing each word and the number of appearance of each word in entire corpus are presented with the dictionary. Approximately 60% of words appear in only one document, followed by 72% for one or two documents. We observed that very few words occur very often, large number of mid-frequency words and very many words occur very rare (see Fig. 3.2). This indicates the Pareto's distribution behaviour. Pareto's law originally stated that 'number of people with incomes higher than a certain limit follows a power law'. We can reword the law as 'number of words appearing in more than a certain limit of document (n) follows a power law' (heavy-tailed distribution). The Pareto distribution is fitted the data with the Pareto index 0.5752 (see Fig. 3.5).

LScDC is a core dictionary built by sub-setting the LScD. We decided to remove too rare words under the assumption that they do not contribute to the text categorisation and are likely to have noisy behaviour in the algorithms. Such words also have impact on measuring of meaning in texts by using the probabilistic approaches such as information gain. They are given almost zero score in such approaches. Therefore, we set a cut-off (10) to remove all words (in LScD) appearing in no more than 10 documents in LSC. After removal of words, we obtained the LScDC containing 104,223 unique words. Words in LScDC, similar to the LScD, are associated with the number of documents containing the word and appearance of the word in entire corpus.

Finally, we present a comprehensive analysis of LScDC by comparing with the NAWL. The NAWL is a list of academic words containing 973 word families. Our aim is to investigate how similar two lists are in terms of mainly matched words and the ranking of words. We applied many approaches based on both direct comparison of words and pairwise comparison of partitions of dictionaries in smaller subsets.

Identification of the NAWL words in LScDC shows that out of the 895 word families (after applying stemming to the NAWL words) in NAWL, 891 were found to be included in LScDC, indicating that the LScDC represents almost complete NAWL words. Four words which appear in only NAWL are "pi", "ex", "applaus" and "pardon". These words did not appear in LScDC but in NAWL due to differences in pre-processing and types of texts in corpora.

The ranking positions of many words of NAWL words found in LScDC are slightly different from those in the NAWL itself. We hypothesize that this is due to the difference in calculation of statistics used to order words in lists. In NAWL, the ordering of words is based on both the frequency and the dispersion of words over categories (SFI), while LScDC words are ordered by the number of documents containing the word only. From the plot of frequencies of matched words (SFI against the number of documents containing the word), we observed a monotonic relationship of frequencies (see Fig. 3.14). However, the log-log plot of matched words' statistics suggests a positive correlation between statistics (linear relationship). We tested this similarity of rankings by Spearsman's Rank Correlation (SRC), Pearson's Correlation Coefficient (PCC) with statistics given and PCC with logarithmic scaled-statistics. We found correlation coefficients of 0.58, 0.30 and 0.61 respectively. This was indeed an expected result as it is the same what we observed from plots.

We then perform an analysis on ranking positions of words by checking the overlap the top n words in the LScDC and the NAWL successively where $n = 5, 10, 15, ..., 890, 891$. We report that there are only 2 common words in the top 20 words of dictionaries, followed by 25 in the top 100 words. The same analysis was repeated for the bottom n words and found that there are 7 common words in the least frequent 20 words, followed by 50 common words in the bottom 100 words. From these findings, we conclude that the LScDC and the NAWL are more similar for least frequent words.

LSC is a multidisciplinary academic corpus of abstracts where the subject categories and citations are known. The dictionaries LScD and LScDC are scientific dictionaries where words are extracted from the LSC. This corpus and dictionaries will be used in a comprehensive research in quantification of meaning of research texts. The meaning of each word will be represented by an analysis of information on categories and areas of research that can be extracted from the appearance of this word in the text. Therefore, the next step will be measuring meaning in LSC texts and then using such measures in several data mining applications including prediction of impact of the paper, categorisation of texts to pre-existing categories and clustering of texts into 'natural categories'.

LSC, LScD and LScDC are available online in [**64**, **67**, **68**].

CHAPTER 4

Informational Space of Meaning for Scientific Texts

In NLP, automatic extracting the meaning of texts constitutes an important problem. Our focus is the computational analysis of meaning of short scientific texts (abstracts or brief reports). In this chapter, a vector space model is developed for quantifying the meaning of words and texts. We introduce the *Meaning Space*, in which the meaning of a word is represented by a vector of *Relative Information Gain* (RIG) about the subject categories that the text belongs to, which can be obtained from observing the word in the text.

This new approach is applied to construct the Meaning Space based on LSC and LScDC. The LSC is a scientific corpus of 1,673,350 abstracts and the LScDC is a scientific dictionary which words are extracted from the LSC. Each text in the LSC belongs to at least one of 252 subject categories of WoS. These categories are used in construction of vectors of information gains.

The Meaning Space is described and statistically analysed for the LSC with the LScDC. The usefulness of the proposed representation technique is evaluated through top-ranked words in each category. The most informative n words are ordered. We demonstrated that RIG-based word ranking is much more useful than ranking based on raw word frequency in determining the science-specific meaning and importance of a word. The proposed model based on RIG is shown to have ability to stand out topic-specific words in subject categories. The most informative words are presented for 252 subject categories. The new scientific dictionary and the $103,998 \times 252$ Word-Category RIG Matrix are available online.

Analysis of the Meaning Space provides us with a tool to further explore quantifying the meaning of a text using more complex and context-dependent meaning models that use co-occurrence of words and their combinations.

4.1. Introduction

4.1.1. The Problem and Preliminaries

Automatic analysis of text meaning is one of the main problems in NLP. This work is focused on the computational analysis of the meaning of short scientific texts (abstracts or brief reports). The starting point is a combination of a simple BoW model with the holistic approach to the text meaning: the text is considered as a collection of words, the meaning of the text is hidden in a situation of use, which is evaluated as a whole. A space of meaning for words is created from the analysis of situations of their use and then, after detailed analysis of this space (including dimensionality reduction and clustering) we will return to the texts and introduce more complex models including words co-occurrence analysis, combination of word's meaning etc.

First of all, we have to consider the "meaning of meaning". This is an extremely deeply discussed topic, since antiquity till modern time (see, for example, [**144, 145, 146, 147**]), but the consensus is still on the way. We start from the Wittgenstein formulation: "Meaning is use" or, in more detail, "For a large class of cases – though not for all – in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language"[**1**, §43].

This idea was widely discussed. This chapter aims to propose an approach to computational analysis of meaning for a large family of texts. The texts we work with (abstracts or brief reports) have well defined dominant communicative function: this is the representative function. An elementary basic scheme of the correspondent act of communication is presented in Fig. 4.1. In this scheme, we see two representations of the situation on the blackboard of consciousness: the sender's representation of the situation and the receiver's representation of the situation. Moreover, the represented situations can be different. In fact, they are always different, and special tools are invented and used to make them as close as possible when necessary. The situations are not compulsory real. They can be real, possibly real or imaginary, and even impossible.

It is necessary to stress that the sender's and the receiver's representations never coincide and:

- Do not represent any situation 'in detail' and, therefore, can represent parts (or projections, let us recall the Plato's Cave allegory) of many different situations at the same time;
- Can include internal contradictions and, therefore, can represent nothing possible in reality;
- Can partially represent different situations, that is, can be 'chimeric' combinations of different possible real or imaginary situations.

There are two 'translation' operations in the scheme Fig. 4.1: (i) from the sender's representation of the situation to the text of the message and (ii) from the text of the message to the the receiver's representation. Both these operations depend on much wider context of the communication including experience of the sender and receiver. Of course, the standard scientific communication assumes that there may be many receivers and the sender can be not a single person. This one-to-many or even many-to-many communication adds more situations and representations and may also add some less trivial multi-agent structures with additional communication channels.

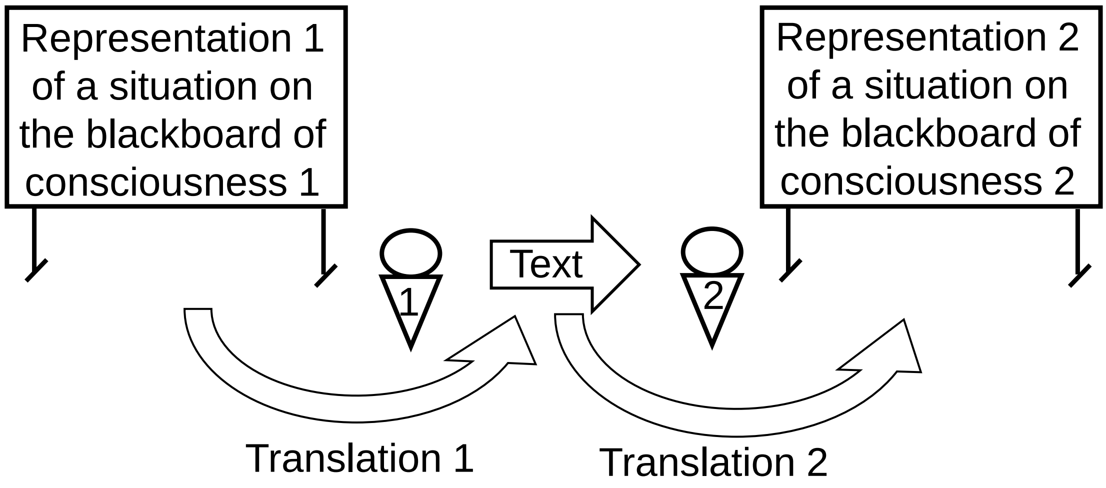


FIGURE 4.1. The idealised scheme of the act of communication. There is a representation of a situation on the sender's 'blackboard of the consciousness' (a representation 1 of a situation 1). A text related to this situation is generated by the sender (translation 1). This text is transmitted to the receiver and transformed by him into a representation of a situation (representation 2 of a situation 2). The situation can be the situation in a real world, the imaginary situation in a possible world, an impossible situation in an impossible world, a chimeric situation combined from several possible or imaginary situation, and so on. We do not study the relations of representation to reality, but only consider the chain: Representation 1 $\to$ Text $\to$ Representation 2.

We consider using the language to transmit information about the represented situations (Fig. 4.1) and neglect many other uses of the language, from military orders to psychological manipulations. The scheme of the act of communication (Fig. 4.1) includes just very basic elements and can be elaborated in much more detail. Here we should refer to the classical works of G.P. Shchedrovitsky [**7, 148**] and J. Habermas [**149, 150**]. For our purposes in this work, the basic scheme (Fig. 4.1) is sufficient.

According to Shchedrovitsky [**7**], at the level of 'simple communication' there is no 'meaning' different from the processes of understanding themselves, which correlate and connect the elements of the text message with each other and with the elements of the situation being restored.

Meaning, for our analysis, is hidden in the relationship between the representation of situations on the 'blackboard of the consciousness' and the texts of the messages. That is, a formal analysis of meaning requires the formalisation of translation operations presented in the scheme of a communication act (Fig. 4.1). Moreover, we can state that we understand the *meaning of meaning* if and only if we can produce such a translation. This translation is context-dependent, the unique experience of the sender and the receiver is involved in this context, so the task of "reproducing the translation" is not fully feasible. Moreover, understanding can be represented as a reflexive game [2] with different levels (The sender prepares a message taking into account the experience of the receiver, his goals and tools, and guesses that the receiver takes into account the experience of the sender, his goals and tools, and... Analogously, the receiver tries to understand the message taking into account..., etc.)

The relation between the text and the representation of the situation cannot be considered as a bijection (both for sender's and receiver's representation). It is many-to-many correspondence: each text corresponds to many situations and each situation can have many representing texts. Moreover, the further consistent formalisation requires the notion of *fuzzy many-to-many correspondence* elaborated for relational databases [151].

According to Mel'čuk [3, 4, 5, 6], the natural language is "*the meaning to text and text to meaning transformer*". He accepted a very strong hypothesis that we are able to describe meaning in a special *semantic language*.

We prefer to be more flexible at this point and characterise a situation "behind the text" by a set of attributes, the method of this characterisation can be changed and does not give a unique and exhaustive presentation of it.

Despite the multiplicity of possible translations, creating of a plausible translation (one of many possible versions) and description of the cloud of such versions of translation can be challenging. This problem resembles the translation problem for natural languages. Now, after impressive progress of machine translation, it seems to be a very attractive idea to apply the modern machine learning tools and encoder-decoder approach [152] to analysis and simulation the translation operations Representation 1 → Text → Representation 2 (Fig. 4.1).

Huge digitized collections of texts exist and are available online. On the contrary, unfortunately, there is no generally accepted common tools for working directly with representations of situations. Various philosophical and logical aspects of this problem were discussed previously by many authors (see, for example, the book 'Representation and reality' [153]).

We do not have a universal toolbox for work with all representations of situations and cannot propose a general solution to this problem. Such a solution, perhaps, is impossible in a finite closed form despite many efforts over decades. Our goal is more modest. We will provide computational analysis of relations between texts of messages and representations of situations for a large collection of brief scientific texts. To do this, these representations must be standardised, at least in part, and expressed in the form of diagrams, specially organized texts or other means.

The simplest approach is to replace the situation representations with the values of some attributes. This approach is not only the simplest, but also quite universal. Many forms of more specific descriptions of situations can be transformed into vectors of attributes. The choice of attributes can be very broad. A classical collection of examples is provided by various version of sentiment analysis. We aim to provide another basic example specific to scientific texts: a list of scientific subject categories that the text belongs to. The list of 252 possible categories is generally accepted and standardised by WoS. Of course, the variety of possible extensions and modifications of the set of attributes characterising the situation is virtually infinite.

Initially, in the act of communication, the situation is not represented by a universally conventional set of attributes. The introduction of attributes is an additional operation external to the communication and is not included in the scheme of simple communication (Fig. 4.1). Moreover, an additional operation must be performed for the selected set of attributes: evaluating their values. This operation can be done either on the sender's side (Fig. 4.2), the receiver's side (Fig. 4.3), or by combinations of these approaches. For example, categorisation of a brief scientific texts is a result of combined efforts: the authors select the categories by their choice of the journal, of the keywords, or by the pointing the categories directly, then the editors can have their own choice, then WoS can finalise the list of subject categories for this text.

For most information services, the choice of subject categories is the result of an understanding of the text by many agents and conflicts of understanding are possible. Even on a famous and very 'liberal' preprint server, arXiv, moderators can sometimes change the category selected by the authors. For example, an author may decide that his paper belong to the category 'condensed matter', whereas the moderator may look through the paper and understand that the main category is not 'condensed matter' research but rather 'nonlinear science' (this was a real life example). This simple example is important because it demonstrates that the content of the text may differ from its meaning: the text contained an explicit reference to 'condensed matter', but this *content* was questioned by the moderator, since in his *understanding* the research refers mainly to nonlinear science, and not to condensed matter. There are important differences between the concepts of 'meaning' and 'content' [7], which are often confused (just as understanding the situation behind the text is often confused with recognising the content of the text).

In the general case, agents who are looking for the meaning of the text can be both humans or computer systems. The latter understand the text in the sense that they define the attributes of the situation behind the text. In our analysis below, the starting point is the combination of the text with the list of the subject categories the text belongs to (1,673,350 abstracts and 252 categories).

The core idea of this approach goes back to the lexical approach of Sir Francis Galton. He selected the personality-descriptive terms and stated the problem of their interrelations for real persons. This work was continued by Thurstone [8]. He selected sixty adjectives (attributes of a person that are in common use). The

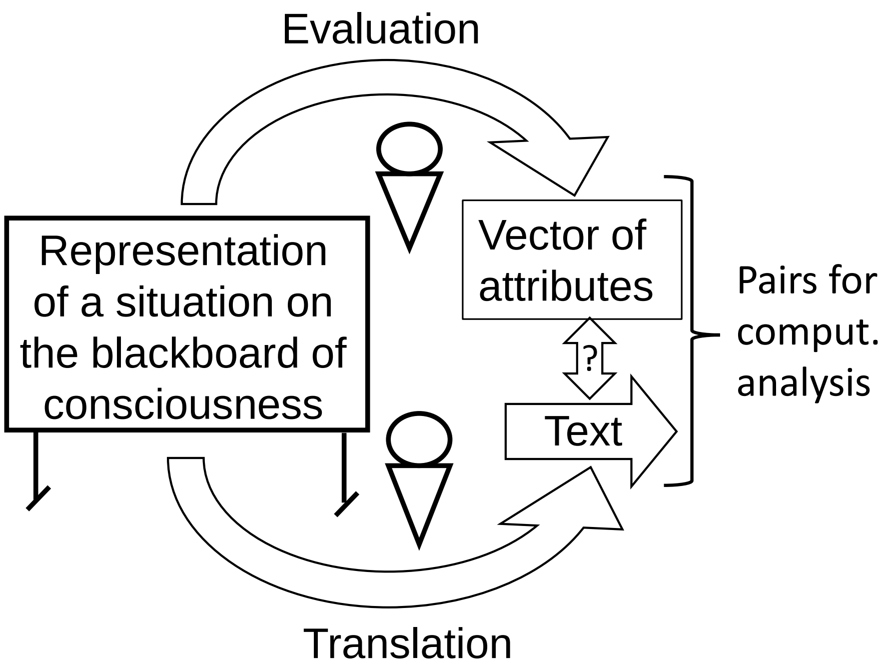


FIGURE 4.2. Parallel (sender's) generation of the learning set for computational analysis and quantification of meaning. One of the main problems is the relationship between the content of texts and the evaluated attributes of the situation.

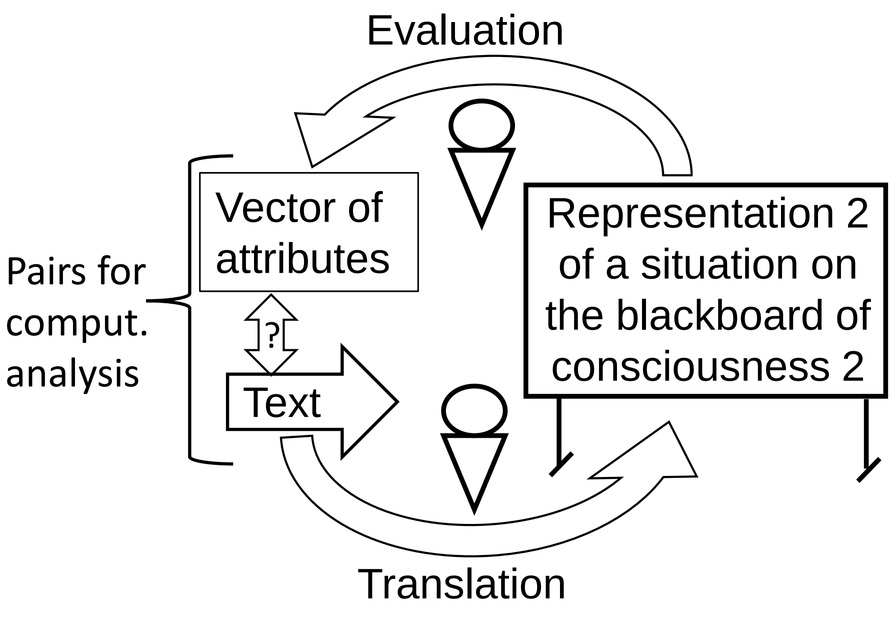


FIGURE 4.3. Antiparallel (receiver's) generation of the learning set for computational analysis and quantification of meaning. One of the main problems is the relationship between the content of texts and the evaluated attributes of the situation.

respondents (1300 persons) were asked to imagine a person they knew well and to select the adjectives that can best describe this person. That is, a person was described by a 60-dimensional Boolean vector. The coordinates correspond to the attributes, the value is 1, if the attribute was selected to characterise the person, and 0 otherwise. Factor analysis gave five factors. After many years of development and discussions, the modern five-factor personality model became one of the common tools in psycho-diagnosis [9, 10].

In psycholinguistics, Osgood with co-workers [11] used a similar approach for creation of the 3D space of meaning by extraction of three 'coordinates of meaning' from the evaluation of the 'affective meaning' of words (objects) by people. These three coordinates are three extracted factors: Evaluation, Potency, and Activity. Of course, the researches started from many different scales and these three were extracted by factor analysis.

Galton, Thurstone, Osgood and their followers asked respondents to evaluate a single object or person. Nevertheless, we can guess that these evaluations were related to some situations with this single object or person, not just to an isolated abstract object. The people evaluated not the abstract 'terms' but the psychologically meaningful situations behind these terms. These situations were the sources of the 'affective meaning' or the personality evaluations. For example, if we evaluate a person as accurate, reliable, and friendly then we have in mind some situations where these properties were demonstrated. The same, if we evaluate a 'dog' as strong, good, and active (or, say, weak, bad and active), we have in mind a dominant situation which we associate with a dog.

The 'affective meaning' or psychological properties do not seem reasonable tools for description of the situations behind scientific texts. In our world of abstracts and brief scientific texts, there is another, scientifically specific description of the situation of use – the categories of the text. There are 252 WoS categories for the LSC, to which the text could belong (see Table C.2). These categories can intersect: a text can belong to several categories. We will use these 252 binary attributes (the text belongs to a given category, or does not belong to it) as a basic description of the situation.

The categories evaluate the situation (the research area) related to the text as a whole, not as a results of the combination of words' meaning. In this holistic approach, we define the general meaning of a word in short scientific texts as the information that the use of this word in texts carries about the categories to which these texts belong. More specifically, that is the RIG about the subject categories that the text belongs to, which can be obtained from observing the word in the text. This RIG is defined for each word and each category. Thus, a meaning of a word is represented by a 252-dimensional vector of RIGs. We create and study this space of meanings.

(1) We intend to analyse the meaning of scientific texts.
(2) We considered the specific world of the texts - the abstracts of research papers.
(3) We narrowed the whole world of abstracts to a sample: 1,673,350 texts from the Leicester Scientific Corpus [13].
(4) We characterize the research situations behind the text by 252 binary attributes – the scientific WoS categories.

Thus, to follow this way, we need a triad: dictionary, texts, and multidimensional evaluation of the situation of use presented by the categories. We prepared the first two elements in the previous chapters and the results are available online [112, 64, 68]. Now, we start to create the space of meaning.

In our case study, we employed very simple attributes for description of the text usage situation, the research subject categories of the text. This list of attributes can be modified and extended. The level of detail of the Meaning Space can vary greatly within the framework of the proposed approach.

4.1.2. Approaches to Meaning of Words

Let us take a quick look at some relevant ideas about quantifying the meaning of words. Quantifying the meanings of words in a metric space might be used to measure the meanings of texts in the same metric as a BoW. A key issue in understanding the meaning of texts is to use a precise metric based on words' meanings. In classical psycholinguistic studies it is common to allocate words in a metric space based on their semantic connotations [154, 155]. Semantic space model is a representation technique where each word is assigned to a point in high dimensional vector space. VSM is one of the most attractive models for researchers since it makes semantics computable [156]. Osgood hypothesised 3-dimensional semantic space to quantify connotative meanings in his theory of *Semantic Differential* concerning psychological and behavioural aspects [11, 12, 157]. The semantic space in his work was built by, in his words, 'three orthogonal bipolar dimensions': Evaluation (E), Potency (P) and Activity (A) where each word is uniquely located on. Following this method of semantic differential, many studies have been attempted by both psychologists and linguists to identify new dimensions of semantic space and to measure the meaning [11, 155, 158, 159]. The structures of semantic spaces are constructed differently by various researchers.

From the perspective of distributional linguistics, a semantic space model is a representation technique for contextual similarity of words by their co-occurrence counts. Distributional hypothesis was introduced by Harris [160] and Distributional Semantic Models (DSM) were then proposed to represent word semantic by distributional vectors [156, 161, 162, 163]. This idea claims that words' similarity can be characterised by their distribution of contexts [164, 165]. The model offers that each word is represented by distribution of its contexts and the distribution of contexts can be learnt from the co-occurrence. The axes in the space are determined by local word co-occurrences and the similarity of a word is measured by its position found by counting co-occurrences to other words in this semantic space [166]. This means that a word's distributional context is represented by a vector of co-occurrences with other context words in a window, where a window can be a certain number of words or lemmas (e.g. words, phrases, sentence, paragraph or document).

Researchers in cognitive studies and information retrieval noted that usage of raw co-occurrence counts is problematic as semantic similarity will have frequency bias [159]. It is proposed that degrees of similarity between word occurrences can be assigned. Different approaches are used to avoid this problem by weighting of elements of the vector. Latent Semantic Analysis (LSA) is one of vector space models in NLP, in particular DSM, for estimating and representing the meaning of word based on statistical computations [167, 168]. In LSA, word senses (or meanings) are approximated in high dimensional space by its effect on the meaning of contexts in which it occurs [169]. Relationship between texts based on their words and relationship between words based on their appearances in texts are analysed simultaneously in order to extract relations of words in terms of their contexts.

LSA has been used for an adequate theory of word meaning by researchers from a wide range of research areas including psychology, philosophy, linguistics, information retrieval and cognitive science [170, 171, 172]. In cognitive science, the focus is to model human memory by activating the meaning potentials by other words in the context under the assumption that cognitive components of meaning of word are linked in a semantic-based network and changes dynamically [173]. It is assumed that human knowledge acquisition actually follows the same process that LSA does: checking events in their internal and external environments and deriving the knowledge from a high dimensional semantic space by a procedure like dimension reduction [174, 175]. Here, the semantic space is used as a basis for all cognitive processing. Although LSA supplies a useful simulation of human cognitive processes, it is argued that LSA knowledge base does not provide a complete modelling of cognition [174]. There are limitations in modification of context and updating the model of semantic dimensions – in this knowledge base – which are characteristic of analytic thinking and dynamic structure of the human cognitive processes [176]. Even if this problem is solved, there are other fundamental semantic problems for LSA such as polysemous words. In LSA, when each word is represented as a single context-free vector in the semantic space, different meanings or senses of a word is not taken into account [167].

This problem matches the task of characterisation of word meaning by its dictionary senses in Word Sense Disambiguation (WSD). WSD is defined as a task to determine the word sense (meaning) by the use of the word in a context in NLP and Machine Learning. In traditional word sense studies, meaning of a word is characterised by mutually disjoint senses covered in dictionaries as the best fit to the its dictionary senses [177]. By both linguistics and psychologists, it has been argued that clear distinctions of senses can be difficult in certain contexts due to fluctuations of meaning in context [171, 173, 178], especially for polysemous words. Hanks (lexicographer) pointed this problem in his paper where he questioned 'Do word meaning exist?' [173] as:

> *"...words have meaning potentials, rather than just meaning. The meaning potential of each word is made up of a number of components, which may be activated cognitively by other words in the context in which it is used. These cognitive components are linked in a network which provides the whole semantic base of the language, with enormous dynamic potential for saying new things and relating the unknown to the known."*

The problem of 'fluctuation of meaning in context' is also important in theories of mental representation of word senses in Psychology. This was very well discussed by Kintsch [167, 174, 179, 180] who stressed the complexity of representation of polygamous words into a single vector in the semantic space in LSA. He questioned 'How is the meaning of words represented in the mind?' and discussed the problem in the aspects of 'mental lexicon' and 'generative lexicon' approaches to the representation of meaning [178]. He came up with the result that both mental lexicon and generative lexicon approaches have limitations in representation of the meanings when word meanings are constructed by their explicit definitions due to multiple senses of words and the flexibility of word meanings. He then discussed the implicit way to define word meaning: relations of the word to other words in the context. According to his research, LSA allows us to modify word meaning by situating the meaning as a vector in high-dimensional semantic space. In this case, the full meaning of the word is not defined, but it is explained in a relational system by only its semantic relationships with other words. He argued that standard composition rule for vectors in LSA does not distinguish the different meanings of a word; therefore, word meanings should be modified according to the different context – where it appears in – by context-sensitive composition algorithms.

Polysemy is one of the characteristics of words in all natural languages. Psycholinguistic studies approach this phenomenon to answer questions of how to represent multiple senses in mental lexicon and how to activate senses during language comprehension [172]. The mental lexicon here can be considered as a mental repertoire containing the list of meanings or senses in the mind. Linguistics proposed several approaches for sense representation in mental lexicon, basically classified as separate sense representation and single core representation. Even though some polysemy studies argue the discreteness of sense storage in mental lexicon [171, 181, 182], the majority of studies suggests that polysemous senses can overlap in their mental representations [177, 183, 184, 185, 186].

Moreover, polysemy of words is one of major focuses in distributional semantics and it is yet to be studied [187, 188, 189]. Some researches in distributional semantics have made modelling the differences of meanings in two occurrences of a word in different contexts possible by developing specialized models for word meaning [190, 191, 192, 193]. Such methods do not approach to word meaning by considering disjoint senses. Alternative models were purposed in which word meaning is not just extracted by pre-defined senses, but from the links between words and their window-based context words. To extract 'contextualised meaning' of a word or a set of words, co-occurrence vectors are constructed and vector operations are used [191, 192, 194, 195]. A probabilistic method that word meaning is modelled as a probability distribution over latent dimensions (senses) was applied by [192, 194]. Contextualized meaning was build as a change in original sense distribution. Cruys, Poibeau and Korhonen then purposed a model in which latent space is used to identify important dimensions for a context and adapt to vector of words constructed by the dependency relations with window-based context words [196].

4.1.3. Quantification of Meaning and Space of Meaning for Scientific Texts

In academic disciplines, the notion of meaning of a word was analysed in many works ranging from psychology to linguistics, philosophy to pedagogy and computer science [197, 198]. Technical innovations in computerised methods and extensive psycholinguistic and neurolinguistic experiments have made investigating word meanings in different perspectives and linking between the language and cognition, and the language in people's mind possible.

There is no unique way to represent meanings that can be used in all theories of lexical semantics from different perspectives. Semantics studies require different semantic representations on the formalism for meaning of word [199]. According to Kintsch, philosophers work with meaning of concepts instead of words, psychologists mostly study concept formation than vocabulary acquisition and linguistics work on meaning of word [174]. But, at this point precise representing and approximating the meaning of concept or specially a text such as sentences, passages or documents are still active problems in NLP and all other disciplines concerning with 'meaning'.

In this research, we specifically focus on meanings in scientific texts. We concern with how meaning can be extracted by analysing the large scientific corpus. Our fundamental assumption is that the meaning of a text can be extracted from the occurrence of its words in texts across the scientific categories. We hypothesize that there is a great connection between the meaning in a text and the vocabulary used in the text; however, we cannot say that each word has the same importance in all research disciplines. In fact, words have scientifically specific meaning in texts based on differences of use in subject categories and these meanings can be estimated from their occurrences in texts within categories. Difference in word meanings for categories correlates with the difference in distribution of words across categories. As they are scientific texts, we consider that occurrence of these words in texts of categories can be used in characterisation of word meaning for science.

Our approach to quantifying the meaning of a word differs from measuring its meaning on the basis of human sensations and feelings, as in psycholinguistic studies. Although measuring the meaning of a word in context by characterization through its dictionary meanings has many important implications in computational linguistics and psycholinguistic research, we do not focus here on dictionary meanings. Rather, we create a model for word representation that allows us to extract the meaning of a word through its importance in various scientific fields without distinguishing its dictionary meanings. We approach the meaning of a word through the predictive power of a corpus analytical procedure under the assumption that the meaning of a word is determined by its use in scientific disciplines. This actually matches the *statistical semantics hypothesis* that 'statistical patterns of human word usage can be utilised to figure out what people mean' [**156**]. We can also reword this as 'statistical patterns of word usage in scientific fields can be used to figure out what a text means'.

In these relations, the meaning of a word is defined as a vector of RIGs from the word to a category. Given such information, meaning can be defined for each word and then for research text [**112**]. A natural way to formalise this is to represent words as vectors and texts as sets of vectors in a specially constructed space. Differences in the distributions of vectors reflect differences in meaning of texts. This technique allowed us to represent each word by a distribution of numerical values over categories and meaning in text through a vector space model, that is, quantifying of meaning.

In many semantic studies, the vector space is obtained by co-occurrence of words as discussed before. There are currently two broad VSMs based on co-occurrence: word-word and word-document where vectors are (normalised) frequency counts and dimensions are contexts (words or documents)[**200**]. Vectors are called *context vectors* in this case, and words are represented by the context vectors. In distributional hypothesis, these vectors are used to compute vector similarity. However, co-occurrence models are plagued with efficiency in real-word applications [**200**]. There are two main problems in the usage of such approaches: first is the dimensionality in contexts vectors and the second is sparse data problem. In the first problem, the dimension of co-occurrence matrix will tend to be extremely big for large data. In the second problem, as the vast majority of words occurs in a very small fraction of set of contexts [118], the majority of the entities of vectors will be zero. Therefore, the co-occurrence matrix will not give reliable results for large data and brief texts. Additional to these two problems, specifically, usage of co-occurrence is not appropriate for the representation of scientific texts due to multidisciplinary researches in the collection [112]. Therefore, we introduce a new vector space to represent word meaning based on words' informational importance in the subject categories.

We begin by creating a space to represent words meaning. The *Meaning Space* is defined as a vector space, in which coordinates correspond to the subject categories. A word is represented by a vector of RIG about the subject categories that the text belongs to, which can be obtained from observing the word in the text. This approach allows us to identify the importance of the word for the corresponding category in terms of information gained when separating the corresponding category from its complement (like, for example, separating texts in category 'algebra' from the text that do not belong to this category).

To define RIGs, we consider the following two attributes of text d for a given word w_j and a given category c_k:

$c_k(d)$: The text d is in the category c_k: Attribute values are Yes ($c_k(d) = 1$) or No ($c_k(d) = 0$);

$w_j(d)$: The word is in the text: Attribute values are Yes ($w_j(d) = 1$) or No ($w_j(d) = 0$).

The corpus is considered as a probabilistic sample space (the space of equally probable elementary results, each of which is a random selection of text from the corpus). RIG measures the (normalized) information about the value of $c_k(d)$, which can be extracted from the value $w_j(d)$ (i.e. from observing the word w_j in the text d) for a text d from the corpus.

As we have a number of word vectors, it is convenient to organise the vectors into a matrix. These vectors are used to construct *Word-Category RIG Matrix*, in which rows correspond to words and columns correspond to categories. Each entry in the matrix corresponds to a pair *(category,word)*. Its value for the pair (c_k, w_j) shows the RIG on the belonging of a text from the corpus to the category c_k from observing the word w_j in this text. Word-Category RIG vectors estimate the meaning of words as their importance in the research fields. Thus, row vectors in the Word-Category RIG Matrix indicate words' scientific meanings.

This approach computes a distributional representation (RIGs) for a word across all research subjects (RIGs in categories). Following to the distributional semantic hypothesis, if words have similar row vectors in the Word-Category RIG Matrix, they tend to have similar meanings. The hypothesis is that if texts have a similar distributions of word meanings – similar clouds of word meanings vectors – then they tend to have similar meanings.

We note that proposed hypothesis does not require an explicit distinguishing between homonymy and polysemy for words; it only requires linking the meaning of words to their importance in categories. With this approach, vocabulary meanings do not directly affect the representation of the meaning of the word. Rather, the meaning of a word is characterized through its measured information content in various scientific subject categories.

In this research, we present the first stage of 'quantifying of meaning': construction of the Meaning Space and representing word meaning as a vector of RIGs for categories in this space. Such an understanding of meaning of words can help analyse the meaning of the texts. Having quantified meaning of words, one can represent all words in a corpus and then texts in the Meaning Space. Specifically, each text in the corpus is a cloud of RIG vectors and the text meaning can be later estimated and constructed by these distributions. Analysis of texts will be focused in the next stage of the research. Text analysis will be the next stage of the research. The earliest (preparatory) stage of the project was presented in [**64, 112**].

The empirical analysis of this research is based on the LSC which includes 1,673,350 texts [**13**] and LScDC of 103,998 words [**15**]. The main hypothesis for construction of the Meaning Space is: meaning is the vector of information gains from the word to the categories assigned to the text. We used 252 categories of WoS.

We evaluated the Meaning Space and representation of word meaning in this space through top-ranked words in each category. We constructed the Word-Category RIG Matrix for the LSC [**16**]. The most informative words in each category are presented. It is shown that the proposed representation technique stands out topic-specific words in categories. We compared this approach with the representation technique where words are represented by vectors of their raw frequencies in categories. Words are ranked by both frequencies and RIGs in categories. We demonstrated that frequencies are not much useful for identifying the most informative words in categories. We concluded that frequency is not much important in this sense.

For each word in the LScDC, the sum and maximum of RIGs in categories are calculated and added at the end of the Word-Category RIG Matrix. Words can be ordered by their informativeness in scientific texts by these two criteria. The most informative n words for scientific texts can be extracted by ordering/sorting words in column of the sum or maximum of RIGs. We compared these two ordering criteria by counting the number of matches in the top n words, where n ranges from 100 to 50,000. We concluded that the majority of the first 100 words do not match, with 28% matched words. The intersection of words reaches to approximately 50% for the top 1,000 words, and then 99% for the top 50,000 words.

Finally, we created a scientific thesaurus in which the most informative words were selected from the LScDC by their average RIGs in categories. The thesaurus was called *Leicester Scientific Thesaurus*. LScT contains the most informative 5,000 words in the corpus LSC. These words are considered as the most meaningful words in science. The full list of words in LScT is available online [**16**].

4.1.4. The Structure of This Chapter

This chapter is organised as follows. In Section 4.2, the Meaning Space is constructed and the representation of words by vectors in the Meaning Space is discussed. Given the representation of words by vectors of RIGs, we look at words ordered by their RIGs in each category. In Section 4.3, we present the first findings of the new representation technique and the anomalies detected in the data by this model. To avoid a possible abnormal appearances of the words in the categories, we apply a further cleaning procedure of the LSC. The latest versions of the LSC, dictionaries LScD and the LScDC are described [**13, 15, 201**]. Finally, we construct the Word-Category RIG Matrix for the LSC [**16**] and discuss the experimental results in this section. In section 4.4, we introduce the LScT, in which there are 5,000 of the LScDC words selected by their average RIGs in categories. In Section 4.5, the conclusion and outlook are summarised.

4.2. Representation of Words by Vectors in the Meaning Space

In this section, we discuss the architecture of our approach to estimating the word meaning in a collection of documents. We assume that the dataset is a large corpus of natural language scientific texts and each text in the corpus belongs to at least one subject category. We hypothesize that words have scientifically specific meaning in categories and the meaning can be estimated by information gains from the word to the category. Before inquiring into the measurement of the meaning, we will mention how to represent each word as a vector of frequencies in categories. We then introduce a new approach to word meaning, in which each word is represented by a vector of RIGs in the *Meaning Space*.

4.2.1. Representation of Words by Vectors of Frequencies in Categories

In this section, we review how to represent a word in a vector space model by using appearances this word in texts belonging to subject categories. A word representation method is defined in order to indicate term absence/presence in texts of categories. Each word is represented by a vector of frequencies in categories. That is, the number of presence of a word is calculated by how frequently this word is observed in texts belonging to the category. Each entry of the vector consists of the number of texts containing the word in the corresponding category.

It is noteworthy that texts in a corpus do not necessarily belong to a single category as they are likely to correspond to multidisciplinary studies, specifically in a corpus of scientific researches. In other words, categories may not be mutually exclusive.

For every word w_j from the dictionary ($j = 1, ..., N$) and every text d_i from the corpus ($i = 1, ..., M$) the indicator $w_j(d_i)$ is defined. If the word w_j occurs in the text d_i (once or more), then $w_j(d_i) = 1$. Otherwise, $w_j(d_i) = 0$.

Let D_k be a set of texts in the category c_k. The frequency of the word w_j in the category c_k is

$$w_{jk} = \sum_{d_i \in D_k} w_j(d_i),$$

This w_{jk} is the number of texts containing the word w_j in the category c_k.

The vector of frequencies is defined for each word w_j from the dictionary. Let us use the notation $\overrightarrow{w_j}$ for it. Coordinates of these vectors are w_{jk}, where index $k = 1, ..., K$ corresponds to the subject categories.

Thus, each word w_j in the corpus is represented by a vector of frequencies w_{jk} denoted by

$$\overrightarrow{w_j} = (w_{j1}, w_{j2}, ..., w_{jK}),$$

where K is the number of categories in the corpus. The collection of vectors, with all words and categories in the entire corpus, can be shown in a table. Each entry w_{jk} of the Table 4.1 corresponds to a word and a category.

TABLE 4.1. The structure of dictionary representation by frequencies w_{jk}

Word \ Category	c_1	c_2	$\cdots$	c_K
w_1	w_{11}	w_{12}	$\cdots$	w_{1K}
w_2	w_{21}	w_{22}	$\cdots$	w_{2K}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
w_N	w_{N1}	w_{N2}	$\cdots$	w_{NK}

The number of documents in the category c_k is $|D_k|$. Importantly,

$$|D_k| \leq \sum_j w_{jk}$$

as each text usually has more than one word, and several different words can belong to the same text. To simplify the notation for further calculations, we now define the set of texts containing the word w_j as D^j. We note that

$$|D^j| \leq \sum_k w_{jk}$$

and equality holds in the case when categories are mutually exclusive.

The number of texts in the categories varies widely, so w_{jk} is expected to increase as the number of texts in a category increases. This does not necessarily mean that a word rarely appearing in a category is less important for this category than for other categories in which the word appears more frequently (see the definition of information gain in the next section). Therefore, direct usage of frequencies may result inappropriate findings in quantification of words' meanings.

Given the collection of vectors, various schemes for normalisation can be performed to adjust the vectors $\overrightarrow{w_j}$ to a common scale. The simplest and the most popular approach for normalisation is transformation to a vector where the sum of the elements is 1, that is normalisation to unite l_1 norm. For the mutually exclusive categories, this normalisation is related to the law of total probability. The objective of this normalisation scheme is to make vectors comparable by rescaling them to the same length in the l_1 norm. For a given vector $\overrightarrow{w_j}$, the normalisation can be performed as

$$P_{jk} = \frac{w_{jk}}{\sum_i w_{ji}}$$

where $\sum_k P_{jk} = 1$. It should be stressed that when categories are not exclusive, $\sum_k w_{jk}$ is not the total number of texts containing the word w_j. In other words, texts containing the word could be counted more than once in the sum.

In similar way, the column vectors can be normalised as

$$Q_{jk} = \frac{w_{jk}}{\sum_i w_{ik}}.$$

However, this representation does not indicate the proportion of exact number of texts in the category.

A reasonable normalisation can also be obtained in two-steps:

(1) Normalize each frequency:

$$w_{jk} \mapsto \frac{w_{jk}}{|D_k|};$$

(2) Normalize the matrix to the unite sum in rows.

As a result, w_{jk} will be transformed into

$$\frac{w_{jk}}{|D_k| \sum_i \frac{w_{ji}}{|D_i|}}.$$

In calculation of RIGs below, the estimation of probabilities are used based on the table of frequencies. For ranking of words in categories, the raw frequencies were also used and compared to RIG-based ranking.

4.2.2. Word Meaning as a Vector of RIGs Extracted for Categories

Having a collection of frequency vectors, it is easy to calculate the vectors of information gains (from observing the word in the text to categories which the text belongs to). These vectors will quantify the meaning the words. The hypothesis here is that the informational content of a word about each category can be measured by comparing the appearance of a word in texts of a given category and its appearance in texts not related to this category (i.e, how the presence/absence of the word in texts can help to separate the category from its set-theoretical complement).

A general concept for computing information is the "Shannon entropy" introduced by Shannon [202]. *Information Gain (IG)* is a common feature selection criterion in machine learning used, in particular, for evaluation of word goodness [122, 203]. The information gain is the measure of the information extracted about one random variable if the value of another random variable is known. It is closely related to the *mutual information*, that measures the statistical dependence between two random variables. The larger value of the gain means the stronger relationship between the variables. The information gain of random variable A with values (or states) $a_1, \dots, a_n$ from the random variable B with values (or states) $b_1, \dots, b_m$ is defined as:

$$\begin{aligned} IG(A,B) = &- \sum_{i=1}^{n} P(A=a_i) \log_2 P(A=a_i) \\ &+ \sum_{j=1}^{m} P(B=b_j) \sum_{i=1}^{n} P(A=a_i|B=b_j) \log_2 P(A=a_i|B=b_j) \end{aligned} \tag{3}$$

where $P(A = a_i)$ is probability of observing the value a_i of the random variable A, $P(B = b_j)$ is probability of observing the value b_j of the random variable B, $P(A = a_i|B = b_j)$ is conditional probability of observing the value a_i of the random variable A given the value b_j of the random variable B. $IG(A, B)$ measures the number of bits of information obtained for prediction of a value of the variable A by knowing the value of the variable B.

In the concept of text categorisation, the information gain measures how important a given word is for category prediction. A larger gain indicates that the probability to find the word in the texts *inside* the category differs considerably from the probability to find it in the text *outside* this category. If the categories are mutually exclusive then we can consider them as values of a categoric feature C of the text with values c_i and define the information gain $IG(C, w)$ from observing a word w in the text about the value of C [122] by the textbook formula:

$$\begin{aligned} IG(C,w) = - \sum_{i=1}^{m} P(c_i) \log_2 P(c_i) + P(w) \sum_{i=1}^{m} P(c_i|w) \log_2 P(c_i|w) \\ + P(\overline{w}) \sum_{i=1}^{m} P(c_i|\overline{w}) \log_2 P(c_i|\overline{w}) \end{aligned} \tag{4}$$

where $\{c_i\}$ is the set of classes in the target space, $P(c_i)$ is the probability of observing the i^{th} class, $P(w)$ is the probability that the term w appears, $P(\overline{w})$ is the probability that w does not appear, $P(c_i|w)$ is the conditional probability of observing the i^{th} class given that the term w appears, and $P(c_i|\overline{w})$ is the conditional probability of observing the i^{th} class given that the term w does not appear.

$IG(C, w)$ measures the number of bits of information obtained for prediction of classes c_i by knowing the presence and absence of a term w in documents of classes.

The quantity $IG(C, w)$ measures the amount of information provided by a word when splitting the documents into classes but only in the case of mutually exclusive classes, that is, each text is assigned to a single class only. On the contrary, the scientific texts belong very often to several categories. The research subject categories are not mutually exclusive and this approach cannot be used directly.

Unlike this approach, we start from measuring how a word is informative for a category in terms of its ability to separate the corresponding category from its set-theoretical complement. We hypothesize that the topic-specific words in categories have larger information gain than other words and such words are expected to have less gain in most other categories. Therefore, we approach to this problem by defining for each subject category c_k a random Boolean variable, or a classification with two states: the text belongs to the category c_k and the text does not belong to the category c_k (this class is denoted as $\overline{c_k}$). The frequencies of words in classes of texts c_k and $\overline{c_k}$ is demonstrated in Table 4.2.

TABLE 4.2. Representation of the word by a pair of frequencies: the number of texts containing the word w_j that belong and does not belong to the category c_k

Word \ Category	c_k	$\overline{c_k}$
w_1	w_{1k}	$\lvert D^1\rvert - w_{1k}$
w_2	w_{2k}	$\lvert D^2\rvert - w_{2k}$
$\vdots$	$\vdots$	$\vdots$
w_N	w_{Nk}	$\lvert D^N\rvert - w_{Nk}$

Since words are obviously not mutually exclusive (one text usually contains several different words) we cannot consider the occurrence of different words as values of a random variable to use (3) directly. To evaluate the information gain of the category c_k from the word w_j it is necessary to introduce for each word w_j a random Boolean variable with two states: w_j denotes the presence of the word in texts of the category c_k and $\overline{w_j}$ denotes the absence of the word w_j in texts of the category c_k. Contingency 2×2 table to calculate information gain of the category c_k from the word w_j is presented in Table 4.3. It used the raw frequencies w_{jk} introduced in previous Subsection.

TABLE 4.3. Contingency table for the category c_k and the word w_j

Word \ Category	c_k	$\overline{c_k}$	Total
w_j	w_{jk}	$\|D^j\| - w_{jk}$	$\|D^j\|$
$\overline{w_j}$	$\|D_k\| - w_{jk}$	$M - \|D_k\| - (\|D^j\| - w_{jk})$	$M - \|D^j\|$
Total	$\|D_k\|$	$M - \|D_k\|$	M

Table 4.3 can be used to calculate two information gains: the word w_j from the category c_k and the category c_k from the word w_j. Both information gains have a meaning for different problems. The goal of this research is to evaluate informativeness of words for category identification and use this informativeness for word ranking and text representations. Therefore, we will consider information gain of the category c_k from the word w_j: $IG(c_k, w_j)$. This information gain evaluates the number of bits extracted from presence/absence of the word w_j in the text for prediction of belonging of this text to the category c_k. One may expect that if a word is a very topic-specific for a category, it appears in texts belonging to this category more frequently than in texts which do not belong to this category; and the major part of texts belonging to this category contains the word.

For each category, c_k, a function is defined on texts that takes the value 1, if the text belongs to the category c_k, and 0 otherwise. For each word, w_j, a function is defined on texts that takes the value 1 if the word w_j belongs to the text, and 0 otherwise. We use for these functions the same notations $c_k$and w_j. Consider the corpus as a probabilistic sample space (the space of equally probable elementary outcomes). For the Boolean random variables, c_k and w_j, the joint probability distribution is defined according to Table 4.3, the entropy and information gains can be defined as follows.

The information gain about the category c_k from the word w_j, $IG(c_k, w_j)$, is the amount of information on belonging of a text from the corpus to the category c_k from observing the word w_j in the text. It can be calculated as [202]:

$$IG(c_k, w_j) = H(c_k) - H(c_k|w_j), \tag{5}$$

where $H(c_k)$ is the Shannon entropy of c_k and $H(c_k|w_j)$ is the conditional entropy of c_k given the observing the word w_j. Entropies $H(c_k)$ and $H(c_k|w_j)$ are

$$H(c_k) = -P(c_k)\log_2 P(c_k) - P(\overline{c_k})\log_2 P(\overline{c_k}), \tag{6}$$

where $P(c_k)$ is the probability that the text belongs to the category c_k, $P(\overline{c_k})$ is the probability that the text does not belong to the category c_k and

$$\begin{aligned} H(c_k|w_j) =& P(w_j)\big(- P(c_k|w_j)\log_2 P(c_k|w_j) - P(\overline{c_k}|w_j)\log_2 P(\overline{c_k}|w_j)\big) \\ &+ P(\overline{w_j})\big(- P(c_k|\overline{w_j})\log_2 P(c_k|\overline{w_j}) - P(\overline{c_k}|\overline{w_j})\log_2 P(\overline{c_k}|\overline{w_j})\big), \end{aligned} \tag{7}$$

where

- $P(w_j)$ is the probability that the word w_j appears in a text from the corpus;
- $P(\overline{w_j})$ is the probability that the word w_j does not appear in a text from the corpus;
- $P(c_k|w_j)$ is the probability that a text belongs to the category c_k under the condition that it contains the word w_j;
- $P(\overline{c_k}|w_j)$ is the probability that a text does not belong to the category c_k under the condition that it contains the word w_j;
- $P(c_k|\overline{w_j})$ is the probability that a text belongs to the category c_k under the condition that it does not contain the word w_j;
- $P(\overline{c_k}|\overline{w_j})$ is the probability that a text does not belong to the category c_k under the condition that it does not contain the word w_j.

All the required probabilities, entropies and relative entropies are evaluated using the contingency table 4.3 as:

$$H(c_k) = -\frac{|D_k|}{M}\log_2\frac{|D_k|}{M} - \frac{M-|D_k|}{M}\log_2\frac{M-|D_k|}{M}, \tag{8}$$

and

$$\begin{aligned} H(c_k|w_j) = \frac{|D^j|}{M}\Big(&- \frac{w_{jk}}{|D^j|}\log_2\frac{w_{jk}}{|D^j|} - \frac{|D^j|-w_{jk}}{|D^j|}\log_2\frac{|D^j|-w_{jk}}{|D^j|}\Big) \\ &+\frac{M-|D^j|}{M}\Big(- \frac{|D_k|-w_{jk}}{M-|D^j|}\log_2\frac{|D_k|-w_{jk}}{M-|D^j|} \\ &-\frac{M-|D_k|-(|D^j|-w_{jk})}{M-|D^j|}\log_2\frac{M-|D_k|-(|D^j|-w_{jk})}{M-|D^j|}\Big). \end{aligned} \tag{9}$$

High value of the informational gain $IG(c_k, w_j)$ (5) does not mean, in general, that the large proportion of information about whether a text belongs to the category c_k can be extracted from observing the word w_j in this text. This proportion depends on the value of the entropy $H(c_k)$ (8). The RIG measures this proportion directly. It provides a normalised measure of the Information Gain with regard to the entropy of c_k. RIG is defined as

$$RIG(c_k|w_j) = \frac{IG(c_k, w_j)}{H(c_k)}. \tag{10}$$

The value of $RIG(c_k|w_j)$ will be 0 when $H(c_k) = H(c_k|w_j)$ and 1 when $H(c_k|w_j) = 0$. In the first case, the presence/absence of the given word w_j does not contain information for the category c_k. So, this word is uninformative. In the second case, using the word in the category provides exactly $H(c_k)$ bits of information. That is, presence or absence of a word resolves exactly the question of belonging the text to the category. $RIG(c_k|w_j)$ can be equal to 1 in two cases:

- All texts with the word w_j belong to the category c_k and all texts without the word w_j do not belong to the category c_k;
- All texts with the word w_j do not belong to the category c_k and all texts without the word w_j belong to the category c_k;

We expect higher $RIG(c_k|w_j)$ for the topic-specific words of the category c_k.

For simplicity, we denote $RIG(c_k|w_j)$ by RIG_{jk}. Given the word w_j, RIG_{jk} is used to form vector $\overrightarrow{RIG_j}$, where each component of the vector corresponds to a category. Therefore, each word is represented by a vector of RIGs. It is obvious that the dimension of vector for each word is the number of categories K (for the WoS subject categories $K = 252$). For the word w_j, this vector is

$$\overrightarrow{RIG_j} = (RIG_{j1}, RIG_{j2}, ..., RIG_{jK}).$$

The set of vectors $\overrightarrow{RIG_j}$ can be used to form the *Word-Category RIG Matrix*, in which each column corresponds to a category c_k and each row corresponds to a word w_j. Each component RIG_{jk} corresponds to a pair (c_k, w_j) and its value is the RIG from the word w_j to the category c_k. The structure of the Word-Category RIG Matrix is demonstrated in Table 4.4.

TABLE 4.4. The structure of the Word-Category RIG Matrix

Word \ Category	c_1	c_2	$\cdots$	c_k
w_1	RIG_{11}	RIG_{12}	$\cdots$	RIG_{1K}
w_2	RIG_{21}	RIG_{22}	$\cdots$	RIG_{2K}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
w_N	RIG_{N1}	RIG_{N2}	$\cdots$	RIG_{NK}

In the Word-Category RIG Matrix, a row vector represents the corresponding word as a vector of RIGs for categories. We defined the *Meaning Space* as the vector space of such vectors $\overrightarrow{RIG_j}$. The dimension of this space is the number of categories and each coordinate is the RIG from a word to this category.

Note that in the Word-Category RIG Matrix, a column vector represents RIGs of all words in an individual category. If we choose an arbitrary category, the words can be ordered by their RIGs from the most informative word to the least informative one. We expect that the topic-specific words will appear at the top of the list.

The words can be ordered by their informativeness in the whole corpus of scientific texts as well as they are ordered in each category. A norm or a more general proximity measure in the Meaning Space is needed to compare the meaningfulness of words across all categories. Two criteria were tested for measuring informativeness of words in the corpus of scientific texts: the sum (l_1 norm) and the maximum (l_∞ norm) of RIGs in categories. For a given word w_j, the sum S_j and the maximum M_j of RIGs are calculated from the Word-Category RIG Matrix as:

(11) $$S_j = \sum_{k=1}^{K} RIG_{jk}$$

and

(12) $$M_j = \max_{k=1,\ldots,K} (RIG_{jk}).$$

The sum S_j is a measure of the average informativeness of a word (this word has the informativeness S_j/K on average), whereas the maximum M_j is a measure of the maximal informativeness of the word across the categories (this word is not more informative than M_j in any category).

Now, the words in the dictionary can be ordered by their S_j or M_j. For each of these ordered lists of words, the most informative (meaningful) n words for scientific texts can be selected based on one of these two criteria. The higher the value of the criterion (S_j or M_j), the more informative the word is.

4.3. Experimental Results

This section describes the experimental details and the analysis done to show the performance of the vector representation method described in Section 4.2. The dataset used in this study is the LSC version 2 [**13**]. The LSC contains a collection of abstracts of research articles and proceeding papers with metadata such as authors, title, categories, research areas and times cited. Each record (text) in the dataset is assigned to at least one of the WoS categories. The LScDC is the collection of unique words appearing in 10 or more documents in the LSC [**15**].

For each word w_j and category c_k, RIG_{jk} is calculated and the Word-Category RIG Matrix for the LSC was formed as described in Section 4.2. In each category, a list of words where words are sorted in descending order by their RIGs can be created. The higher the relative information a word gained in a category, the more important the word is in terms of being topic-specific for the category. Therefore, one could look at the top n words in categories in order to get a good grasp of the representation method. The visualisation of the top words in each category is carried out with the word clouds. Having calculated the frequencies of words in the categories (Table 4.1), we compare the purposed method with the commonly-used approach based on raw frequency.

4.3.1. Discovering Anomalies in the Data by Information Gain-Based Representation Technique

At first, the procedure of word representation was applied to the LSC version 1 [64] with the dictionary [68]. To visualise top words in each categories in a convenient way, we looked at word clouds. The font size of each word in a word cloud is proportional to its RIG in the category. Intuitively, the more informative the word is, the bigger size the word appears in word clouds. For example, from Fig. 4.4, it can be seen that the most informative 6 words for the category 'Acoustics' are 'acoust', 'ultrasound', 'speech', 'nois', 'sound' and 'frequenc'. The majority of papers in Acoustics is expected to include these words which are absent or at least less frequent in many other categories. These words are inferred to be informative for the category 'Acoustics'.

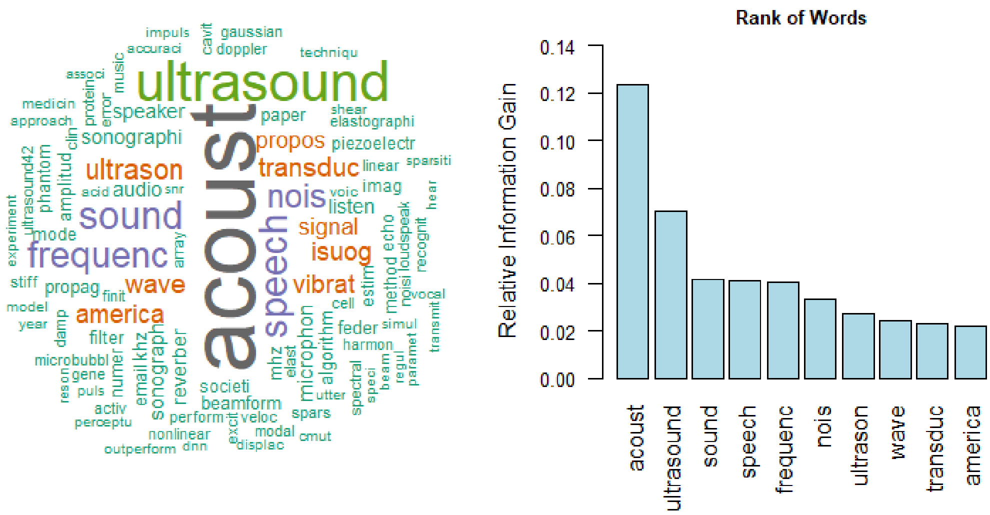


FIGURE 4.4. The most informative 100 words in the category 'Acoustics'. The font size and colour of words indicate different RIGs of the words. The histogram shows RIGs for the top 10 most informative words in the category.

However, this method detected anomalies in some categories. Anomalies here refers to words that do not conform to the expected set of words to be appearing in a subject category. Such words can appear in any category frequently regardless of being a topic-specific word. These words are likely to be potential anomalies generated by inappropriate joints of words, phrases or sentences to the texts of abstracts. As shown in Fig. 4.5, for the category 'Chemistry, Applied', words 'elsevi', 'ltd', 'acid', 'reserv' and 'right' stand out in the word cloud. We see that trends in majority of words in the word cloud agree with each other as being related to the subject. However, 'elsevi', 'ltd', 'reserv' and 'right' seem like more prominent and unusual (non-specific) for Chemistry. In fact, the experiments were preliminary, but we discovered alarms indicating anomalies by our representation technique.

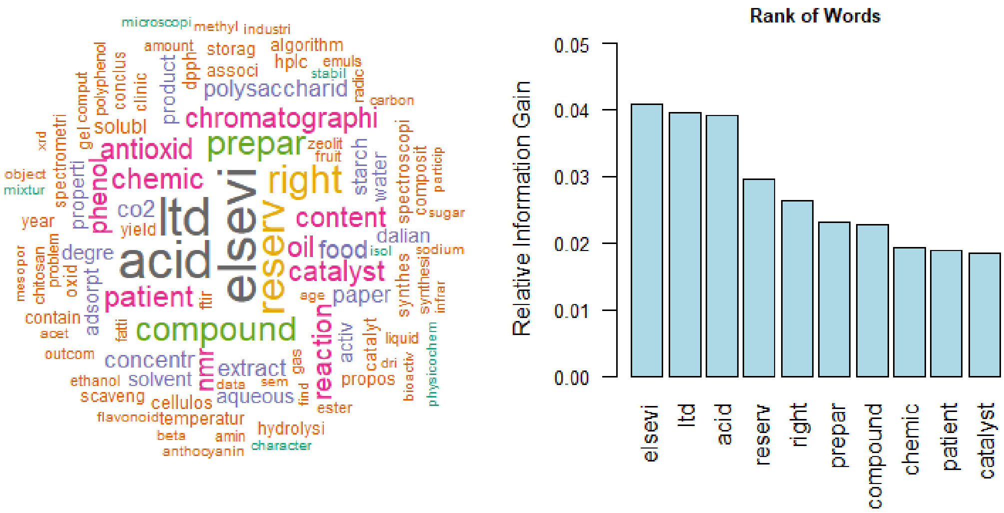


FIGURE 4.5. The most informative 100 words in the category 'Chemistry, Applied' before additional cleaning. The font size and colour of words indicate different RIGs of words. The histogram shows RIGs for the top 10 most informative words in the category.

To understand why these words arose and how they can be avoided, we checked the abstracts containing such words. Our review showed that these words appeared in copyright notices such as "Published by Elsevier ltd." or 'All rights reserved', and they were added at the footer of abstracts. In order to have a comprehensive understanding of their appearance as being informative for only some categories, for instance in Chemistry, we compared distributions of 'elsevier', 'right' and 'reserve' in categories. For each word, categories are ordered by the number of documents containing the word, and the first 20 categories are presented in Fig. 4.6. When we consider the list of categories ordered by the number of documents in the entire corpus, we conclude that not all categories in the list of top categories appear in the charts. This is because usage of copyright notices is much more noticeable in some categories such as Chemistry. For instance, the rank of the category 'Engineering, Electrical & Electronic' is 1 in the corpus; however, one can see that this category has rank 15 for the word 'Elsevier'.

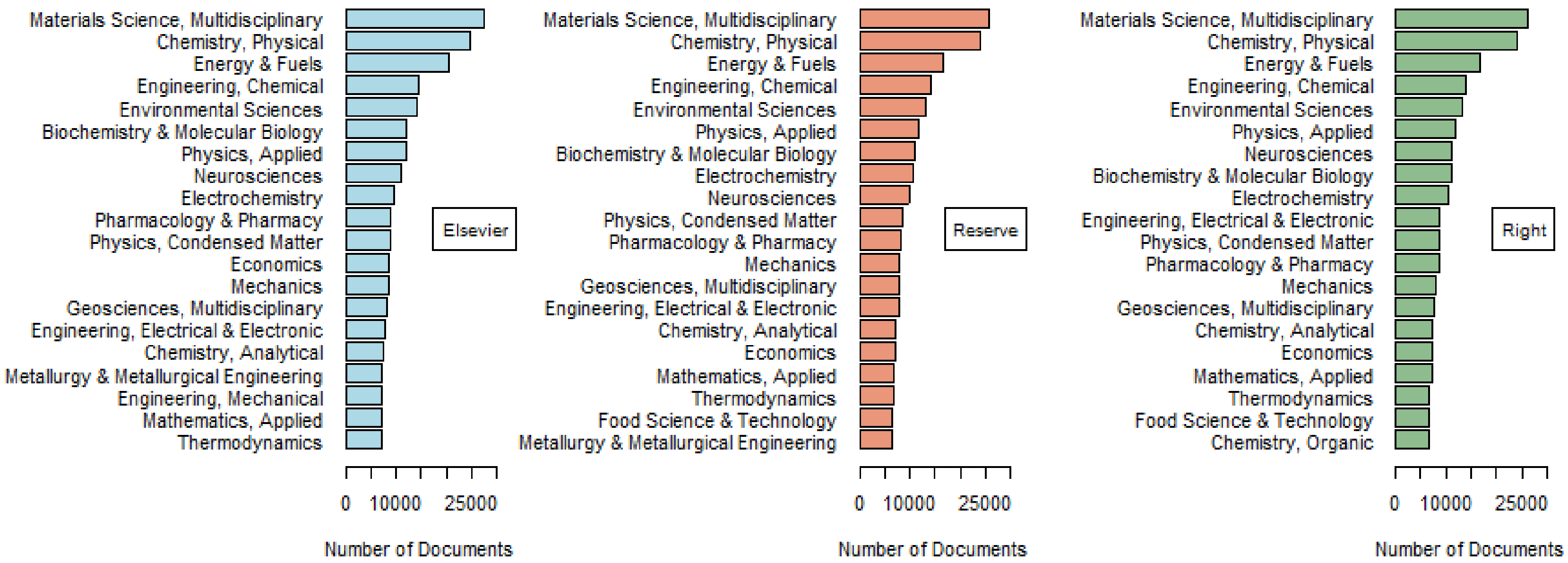


FIGURE 4.6. Top 20 categories that words 'Elsevier', 'Reserve' and 'Right' appear in

To show that not all categories have the same/similar distribution of use of copyright notices, we presented Fig. 4.7 where fractions of documents containing words ‘elsevier’, ‘right’ and ‘reserve’ in four of categories are demonstrated. We can see from the figure that about 40% of texts in ‘Chemistry, Physical’ (rank is 4 in the corpus) contain these three words, while only small fragments of ‘Engineering, Electrical & Electronic’ and ‘Computer Science, Theory & Methods’ collections (rank is 6 in the corpus) include these words. ‘Economics’ (rank is 34 in the corpus) papers include these words in considerable portion. As expected, these words did not appear as the most important words (10 words) for ‘Computer Science, Theory & Methods’ (see Fig. 4.8).

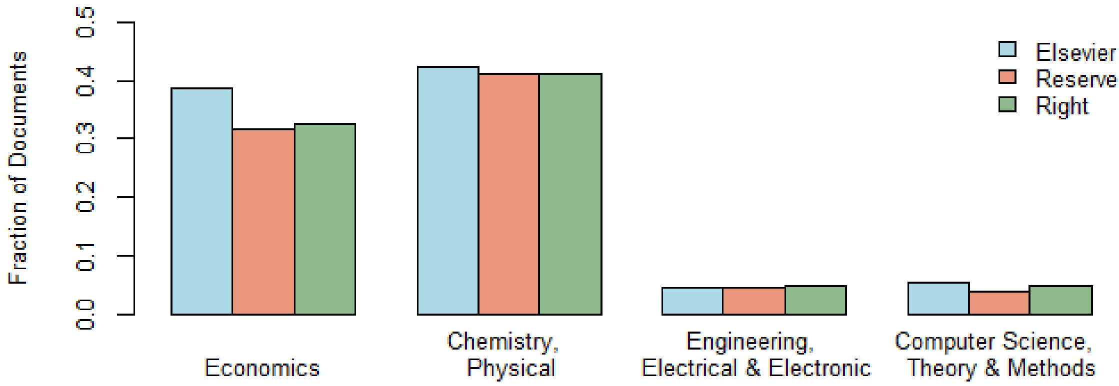


FIGURE 4.7. Fractions of texts containing the words ‘Elsevier’, ‘Right’ and ‘Reserve’ in four categories before additional cleaning.

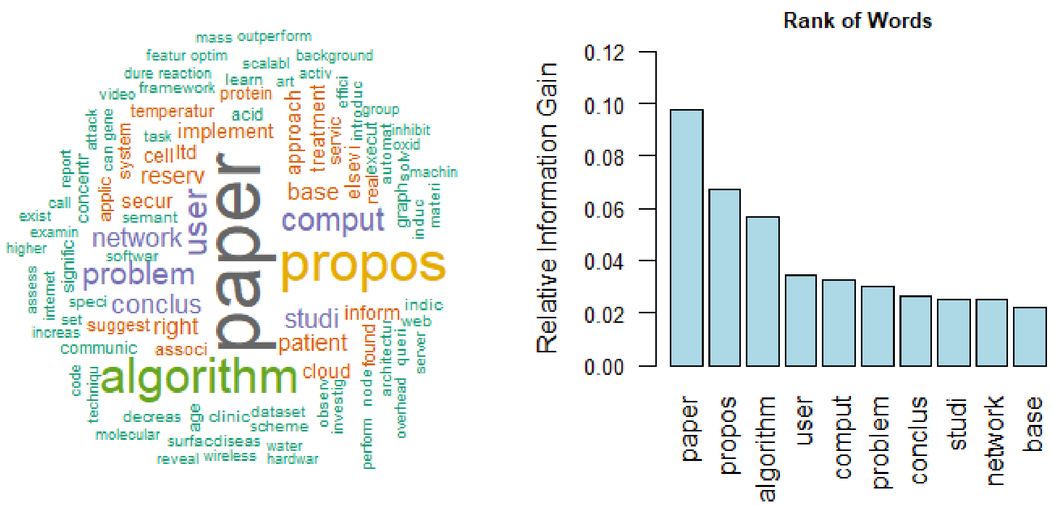


FIGURE 4.8. The most informative 100 words in the category ‘Computer Science, Theory & Methods’. The font size and colour of words indicate different RIGs of words. The histogram shows RIGs for the top 10 most informative words in the category. (Before additional cleaning. However, the additional cleaning does not change noticeably the diagrams for this category.

4.3.2. The Data

This subsection provides the description of procedure of additional cleaning and correction for the LSC and LScDC.

4.3.2.1. *Further Cleaning of the Leicester Scientific Corpus (LSC)*

Many conferences and journals put copyright notices, permission policies or conference names below abstract of papers. Such footers were added to abstracts in many records in WoS database and so in the LSC during processing and storage of the original data (see Table 4.5).

It is really a huge and practically impossible task to find out with the help of human inspection which notifications were added in the texts of 1,673,824 abstracts in [64]. Once a sample of abstracts containing publishing houses names was browsed, we found that there are much more scenarios to consider. Some examples of these scenarios are presented in Table 4.6. As such expressions are more frequent in some categories than in others, a cleaning procedure is needed to avoid possible abnormal appearances of words in categories. A quick look at the scenarios is sufficient to conclude that clearing such sentences or phrases cannot be fully automated. Human intervention is needed to identify and list them to avoid deleting useful information from the data.

TABLE 4.5. An example of abstract with a copyright notice

Title	Neonicotinoid concentrations in arable soils after seed treatment applications in preceding years
Authors	Jones, A; Harrington, P; Turnbull, G
Abstract	Concentrations of the neonicotinoid insecticides clothianidin, thiamethoxam and imidacloprid were determined in arable soils from a variety of locations in England. **...[Truncated]**. As clothianidin and thiamethoxam have largely superseded imidacloprid in the United Kingdom, neonicotinoid levels were lower than suggested by predictions based on imidacloprid alone. (c) 2014 Crown copyright. Pest Management Science (c) 2014 Society of Chemical Industry

Individual notices with different appearances were identified by sampling of abstracts based on keyword search. A keyword search refers to browsing words, phrases or sentences to list different appearances of them in order to delete all identified appearances from abstracts. The position of notices was also taken into account since they appeared either at the beginning (by mistake) or at the end of the text. We used several specially developed procedures successively to clean them. For instance, when removing notices in the form of '(c) Published by Elsevier', we first checked the appearance of 'Crown Copyright (c) Published by Elsevier'. It can also appear

TABLE 4.6. Some examples of notices attached to the abstract

Copyright Notice, Name of Conference, Journal or Publishing House
(c) 2014 Elsevier ltd. All rights reserved.
(c) 2014 Published by Elsevier B.V.
(c) The Authors. Published by Elsevier Inc. All rights reserved.
Crown Copyright (c) 2014 Published by Elsevier B.V. All rights reserved.
(c) 2014 The British Infection Association. Published by Elsevier Ltd. All rights reserved.
(c) Wolters Kluwer Health — Lippincott Williams & Wilkins
(c) 2014 Wiley Periodicals Inc.
(c) Springer-verlag Berlin Heidelberg 2014
(c) The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License.
(c) RSNA.
2014 American Cancer Society.
Pediatr Blood Cancer.
J. med. virol

in the form of 'Published by Elsevier', thus we consider all cases based on empirical study. During cleaning, we removed copyright notices, names of conferences, names of journals, authors' rights, licenses and permission policies identified. To give an insight, Table 4.7 presents the number of document containing some notices before cleaning. These notices were completely removed after cleaning. We note that names of publishing houses could appear inside the text, in this case we did not remove them. More examples of notices that were removed from abstracts can be found in Appendix C.1.

To display the initial result of the cleaning, we present the word cloud and histogram of RIGs for the category 'Chemistry, Applied' in Fig. 4.9. One can see that words 'elsevi', 'ltd', 'acid', 'reserv' and 'right' do not appear in the list of top words as was in the word cloud before cleaning (see Fig. 4.5). Instead, the cloud gives greater prominence to words that are related to specific topics and likely to be more informative for the category. The word 'acid' has been preserved in the list of the most informative words.

4.3.2.2. ***The Latest Version of LSC and Dictionaries***

After detecting and cleaning copyright notices, permission policies and conference names from abstracts, a new version of the LSC was created and made accessible in

Table 4.7. The number of abstracts containing some attached notices before cleaning (These notices were completely removed after cleaning)

Notice	Number of Notices Before Cleaning
Elsevier ltd. All rights reserved	101,994
(c) The Authors	561
All rights reserved	283,041
(c) Springer-verlag Berlin Heidelberg 2014	20

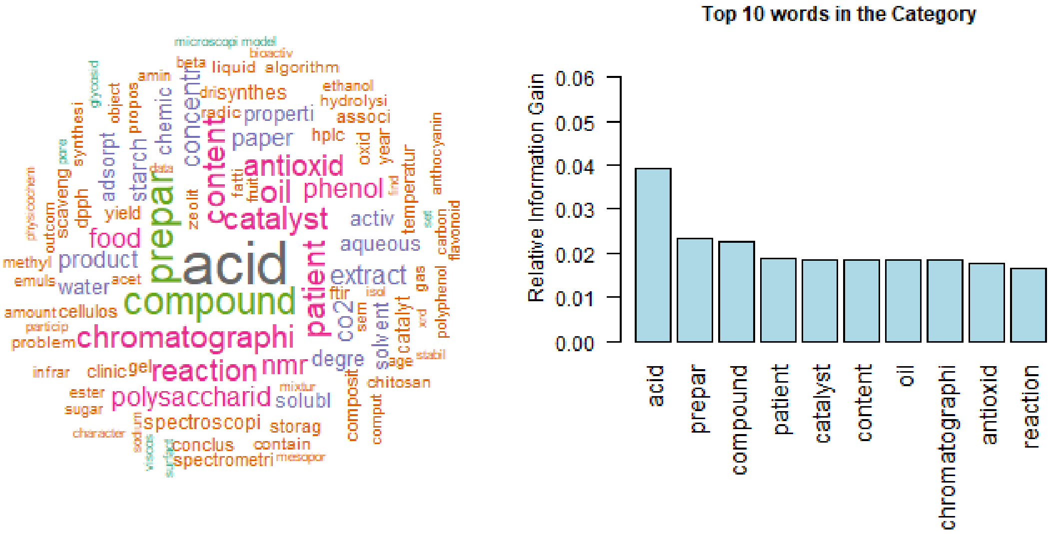


Figure 4.9. The most informative 100 words in the category 'Chemistry, Applied' after cleaning. The font size and colour of words indicate different RIGs of words. The histogram shows RIGs for the top 10 most informative words in the category.

[13]. The cleaning procedure described in previous section leaded to some abstracts having less than our minimum length criteria (30 words). Such abstracts were not contained in the new version; therefore, the remaining 1,673,350 texts were used in this study (474 texts were removed). As was the case for the LSC before cleaning, the latest version of the LSC involved text of abstracts, list of authors, title, list of research areas, list of categories and times cited.

It is noteworthy that, in both versions of the LSC, the number of subject categories is 252. All categories and the number of documents assigned to the corresponding category are presented in Table C.2. Same information for research areas was provided in Table C.3. The distribution of length of abstracts is displayed in Fig. 4.10. There is no noticeable difference between distributions for two versions and the average length of texts is 176 words.

The latest version of the LScD was developed by extracting words from the new version of the LSC [201]. The procedure applied to process the LSC in creation the

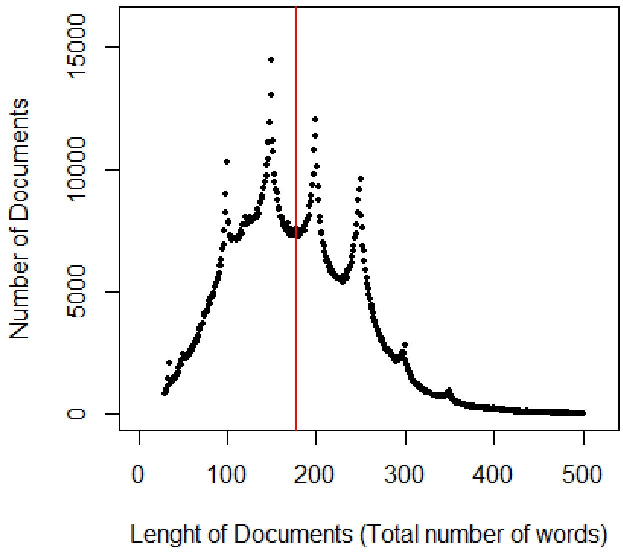


FIGURE 4.10. The number of abstracts with specified length against lengths of abstracts in the latest version of the LSC. The minimum length is 30 and the maximum length is 500 with an average of 176 words.

LScD was the same as described in [112]. The new version of the LScD contains 972,060 unique words with the number of texts that a word appears in. A new version of the core list, LScDC, was created from the LScD by removing words appearing in no more than 10 texts of the LSC [15]. All steps applied were the same as for the previous version of the LScDC and can be found in [112].

Based on the decision to clean copyright notices, we expect that words such as 'Elsevier', 'Reserved', 'Ltd', 'Right' and 'Springer' will not appear frequently in the LSC as they did before. In fact, the number of appearance of these words decreased after cleaning (see Table 4.8).

TABLE 4.8. The number of occurrence of some words appearing in copyright notices before and after cleaning

Word	**Number of Documents Containing the Word Before Cleaning**	**Number of Documents Containing the Word After Cleaning**
Elsevi	314,204	80
Right	306,075	27,279
Reserv	288,193	5,761
Ltd	147,466	830
Springer	296	187
Copyright	21,160	396

The results indicated that some words, for instance 'Right' and 'Reserve', are still relatively frequent in the corpus. This is because these words are specific for some categories. To give an insight, we compared top categories for three words 'Elsevi', 'Reserv' and 'Right' (see Fig. 4.11). The results for the word 'Right' indicate that it is frequently used in medicine related categories such as 'Neuroscience' and 'Surgery', and in social science categories such as 'Law' and 'Political Science'. This is an expected result as it can appear to determine the side of organs such as 'right hippocampus' or 'right hemisphere' in medicine; and the normative rules in such disciplines as law and ethics. 'Elsevier' and 'Reserv' are much more uniformly distributed to the categories when the rank of categories is taken into account. For 'Reserv', one can identify categories related to Biosciences such as 'Ecology', 'Zoology' and 'Environmental Studies'. Specifically, this word occurs to indicate 'nature reserves'.

4.3.3. Words Represented by Vectors of Frequencies in Categories for the LSC

Recall that a representation technique for words was introduced in Section 4.2. The vectors of frequencies in subject categories were obtained for each word. The frequency associated to a category was computed by counting texts containing the word in this category.

Subject categories are used to categorise papers in the WoS collection; however, documents do not necessarily belong to a unique category due to interdisciplinary studies. In other words, categories are not exclusive in WoS and so in the LSC. In the LSC, texts belong to at least 1 and a maximum of 6 categories out of a total of 252 subject categories (see Fig. 4.12). It is noteworthy that our consideration is to count the number of times a word appears in texts of a category rather than analysing exclusivity of categories. Therefore, in this stage, we just looked at the frequency of texts with these words in categories.

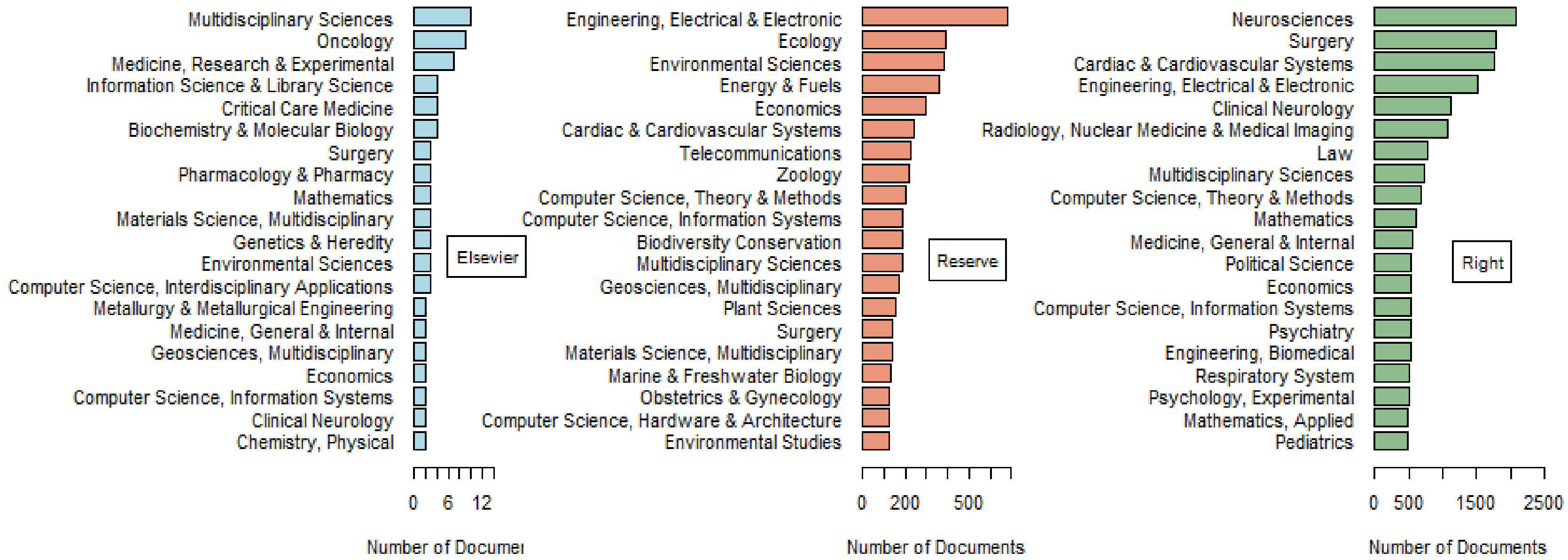


FIGURE 4.11. Top 20 categories that words 'Elsevier', 'Reserve' and 'Right' appear in (after cleaning)

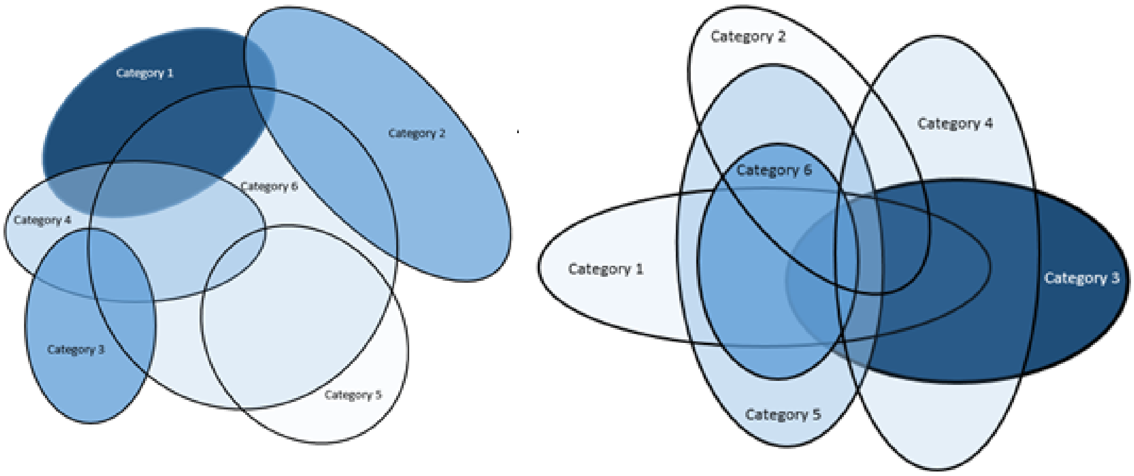


FIGURE 4.12. Two examples of intersection of categories in the LSC. The maximal number of categories that a document belongs to is 6 in the LSC.

The vectors of frequencies in categories are built for 103,998 LScDC words and 252 subject categories. Each row represents a word of the LScDC in 252-dimensional space, that is, each word is represented by a vector of frequencies in 252 categories. For each category, a frequency distribution can be obtained for the set of words. The distribution indicates words used in texts of each category and the most frequently used words can be sorted in categories. To illustrate this, the most frequent 10 words for categories 'Astronomy & Astrophysics', 'Mathematics' and 'Asian Studies' with frequencies are displayed in Table 4.9. A table containing all words and categories are included in [**16**]. One can expect that not all words in the table indicate a topic in the related subject. As an example, words 'use', 'also', 'studi' and 'paper' are frequent words in the LScDC and so in categories. These non-topic specific words occur many times in abstracts without indicating subject specificity. Therefore, using the frequencies of words in categories may not reflect how specific a word is to a category.

4.3.4. Word-Category RIG Matrix for the LSC

On the basis of exploratory work by the frequency table, we concluded that the use of word frequencies in categories does not provide much information about the category. To be specific, we expected that 'use' is not a topic-specific word as it appears in all 252 categories and it is likely to be used in almost all texts. This means that the meaning of a word in the text cannot be directly extracted from the frequency.

Aiming at this result, we must now apply a different perspective to measure the importance of words for categories, with a special attention given to the hypothesis that each word in the LScDC has scientifically specific meaning in categories and the meaning can be extracted from the information of words for 252 subject categories

TABLE 4.9. The most frequent 10 words for categories Astronomy & Astrophysics, Mathematics and Asian Studies

Astronomy & Astrophysics		Mathematics		Asian Studies	
use	11,100	paper	9,408	articl	415
observ	10,237	result	9,074	examin	236
model	9,295	prove	7,743	studi	230
result	9,213	show	6,705	one	226
present	7,810	space	6,224	argu	226
can	7,598	also	6,194	also	216
studi	7,350	studi	6,187	paper	211
also	7,314	use	6,062	polit	196
show	7,191	function	5,904	use	195
similar	6,622	general	5,766	two	192

in the LSC. Thus, as described in Section 4.2, words were represented in a 252-dimensional Meaning Space. RIGs for each word in 252 categories were calculated and vectors of words were formed. We then represented these vectors in the Word-Category RIG Matrix.

For each word in the Word-Category RIG Matrix, the sum S_j and maximum M_j of RIGs in categories were calculated and added at the end of the matrix. The Word-Category RIG Matrix can be found in [16]. One can extract the most informative n words for scientific texts by ordering/sorting the column of words based on their S_j or M_j.

4.3.5. Results

The experimental results presented in this section were obtained using abstracts of academic research papers in the LSC [13]. We used words from the core dictionary LScDC [15].

Having calculated RIGs for each word and created the Word-Category RIG Matrix, we evaluate the representation model by checking words in each category. That is, we consider the list of words with their RIGs in the corresponding category. Those words that have larger RIG are more informative in the category. Being 'more informative' here allows for the interpretation of being 'more specific' to the category's topic.

For each category, words are sorted by their RIGs and the top 100 words are shown in the word clouds. The bigger font size the word in word clouds, the more informative it is. Word clouds for the top 100 most informative words and histograms of RIGs for the top 10 most informative words for each of 252 categories can be found in [204]. The most informative 100 words with their RIGs for each of categories are presented in [204] and [205].

In general, the RIG based method proves to be more sensitive than the frequency based method in identifying topic-specific words of a category. This means that representing words in Meaning Space has the advantage of transforming words to efficient vectors with a benefit of considerably lower dimension than the standard word representation schemes. To illustrate this result, we choose categories 'Biochemistry & Molecular Biology', 'Economics' and 'Mathematics' and compare two word clouds that are formed by using raw frequencies and RIGs in categories (see Figures 4.13, 4.14 and 4.15). It can be seen from the figures that the majority of the most frequent words in all three categories are frequent words in the entire corpus. These words are not topic-specific for categories as they appear in almost all abstracts. The frequent but non-informative words can be considered as *generalised service words of Science* and deserve special analysis.

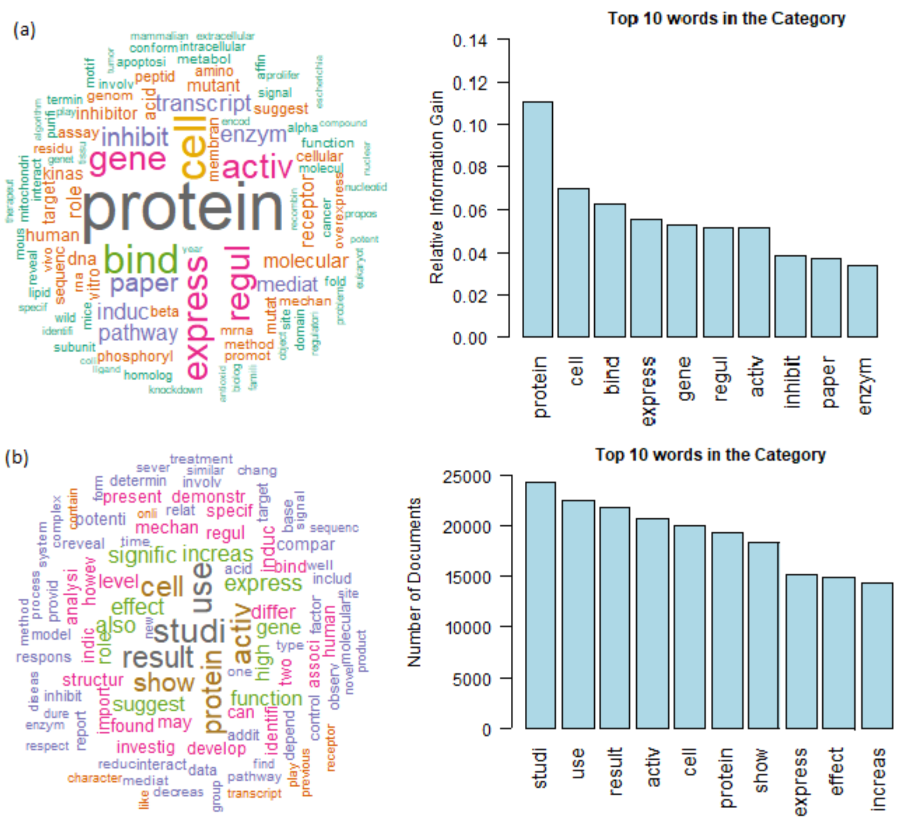


FIGURE 4.13. Category 'Biochemistry & Molecular Biology': word cloud of the top 100 most informative words and the histogram of the the top 10 most informative words. The informativeness is defined by (a) RIG (b) frequency.

This proves that raw frequency is not much important to identify scientifically specific meanings of words. Therefore, by representing words as vector of RIGs, we can avoid such frequency bias. The most informative words in categories for RIG

FIGURE 4.14. Category 'Economics': word cloud of the top 100 most informative words and the histogram of the top 10 most informative words. The informativeness is defined by (a) RIG (b) frequency.

representation are topic-related in the corresponding category. We interpret these results as evidence for the usefulness of the RIG based representation.

Words that are expected to be used together have very close values of RIGs. In 'Health Care Sciences & Services', 'health' and 'care' are top words and RIGs for these words are so close (see Fig. C.1). Another example is 'xrd' and 'difract' in 'Material Science, Ceramics'. 'XRD' is actually abbreviation of 'X-ray diffraction'; therefore, they appear together as 'X-ray diffraction (XRD)' for most of cases in the category (see Fig. C.2).

We can extract some stylistic properties in texts of categories. For instance, in computer science related categories the word 'paper' has the highest RIGs (see Figures C.3, C.4, C.5, C.6, C.7, C.8 and C.9). This may be result of the stylistic features in papers for such categories.

A casual observation indicates that while the most informative words in some categories have similar RIGs, differences in values of RIGs are much more noticeable for the most informative words in some other categories. To give an insight, we present the categories 'Chemistry, Medicinal' and 'Engineering, Chemical' in Fig. 4.16. In 'Chemistry, Medicinal', the word 'compound' can be easily separated from the other words, while in 'Engineering, Chemical', there is a slight decrease in RIGs

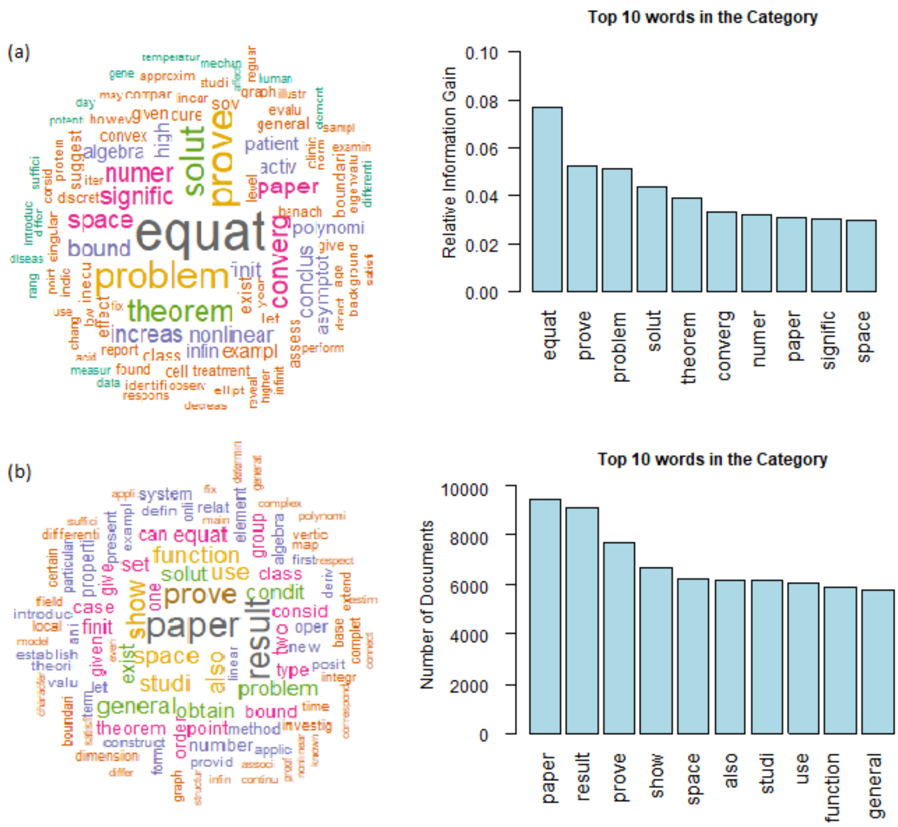


FIGURE 4.15. Category 'Mathematics': word cloud of the top 100 most informative words and the histogram of the top 10 most informative words. The informativeness is defined by (a) RIG (b) frequency.

for the top 10 most informative words. However, in general we did not observe any explicit rule for this property.

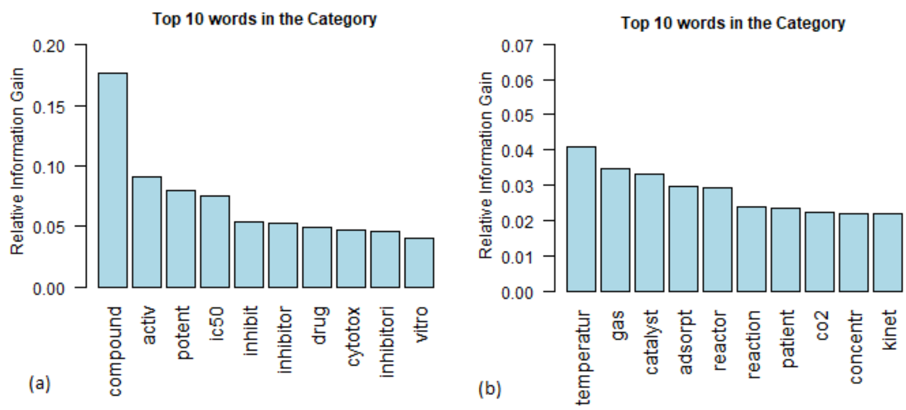


FIGURE 4.16. Histograms of the most informative 10 words in categories (a) Chemistry, Medicinal and (b) Engineering, Chemical

Finally, we formed two lists of words that arranged in descending order based on the sum and maximum of their RIGs in 252 categories. The top 100 words in two lists are displayed by word clouds in Fig. 4.17 and Fig. 4.18. Histograms in the figures show the most informative 10 words in the lists. We found that the most informative 10 words in two lists are completely different, as shown in the figures. From words clouds, one can see that the majority of the first 100 words do not match. We then compared two lists by counting the number of matches in the top n words, where n ranges from 100 to 50,000. The numbers of matched words for different n are presented in Table 4.10. As can be seen, 18% of words match for the top 50 most informative words. This proportion increases to approximately 50% for the top 1,000 words and to 58% for the top 2,000 words. The intersection of lists reaches to approximately 99% for the top 50,000 words. From these results, one can conclude that two lists are different in the top words. When higher number of words is taken into account, lists become more similar in terms of words included. However, the rank of words are not similar. Any of these criteria for selecting the most informative words can be used depending on the task and the information required.

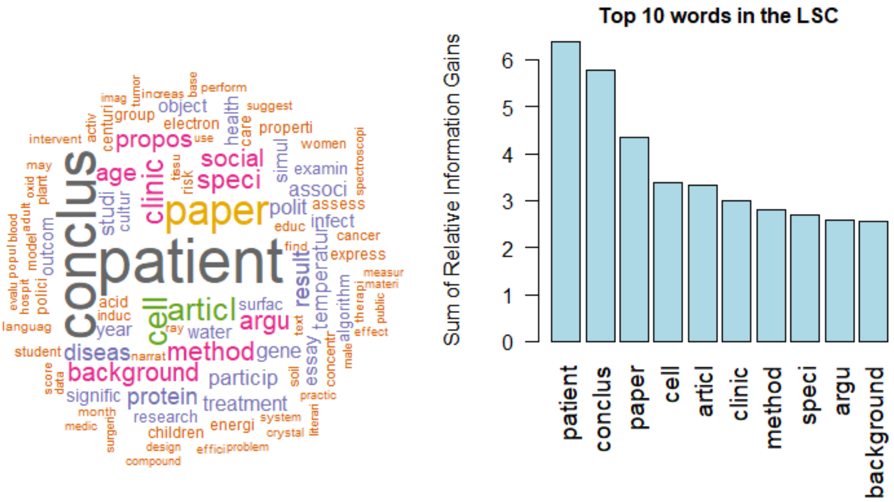


FIGURE 4.17. The most informative 100 words in the LSC. Words are arranged and selected by the sum of their RIGs in 252 categories. The font size and colour of words indicate different sum of RIGs. The histogram shows the sum of RIGs for the top 10 informative words.

The numbers S_j and M_j are differently distributed for words. We observed from the lists that many words have low S_j and M_j. Fig. 4.19 and Fig. 4.20 show the distribution of S_j and M_j for words in the logarithmic scale. Supper-exponential picks near zero RIGs are noticeable for both criteria. We can see that the trend is going down almost linearly beyond the picks. The bottom 10 least informative words in two lists are presented in Table 4.11. One may consider words having almost zero S_j or M_j as less meaningful words for scientific texts.

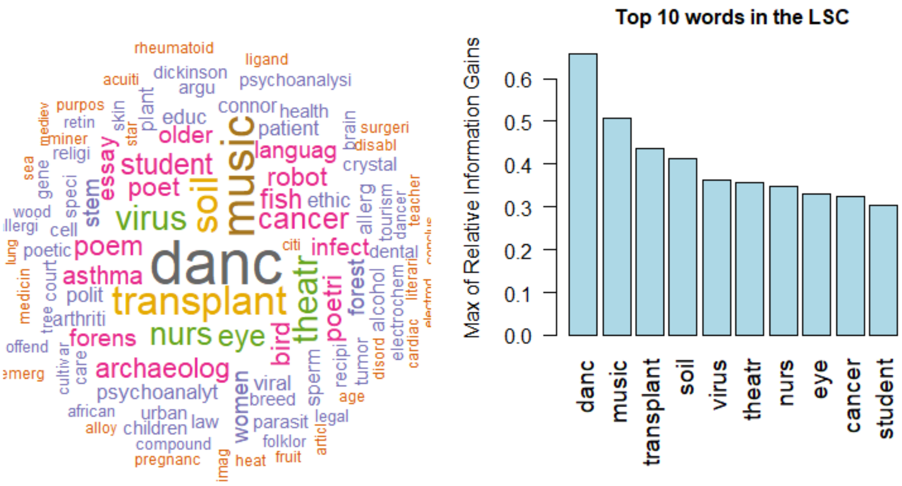


Figure 4.18. The most informative 100 words in the LSC. Words are arranged and selected by the maximal RIG over 252 categories. The font size and colour of words indicate different maximal RIG. The histogram shows the maximal RIG for the top 10 informative words.

Table 4.10. Comparison of words ordered by the maximum M_j RIG and sum S_j of RIGs in categories

Top (n) Words in Two Lists	Number of Matches	Fraction of Matches
10	0	0.000
50	9	0.180
100	28	0.280
500	189	0.378
1,000	498	0.498
2,000	1,168	0.584
5,000	3,412	0.682
10,000	7,492	0.749
50,000	49,542	0.991
103,998	103,998	1.000

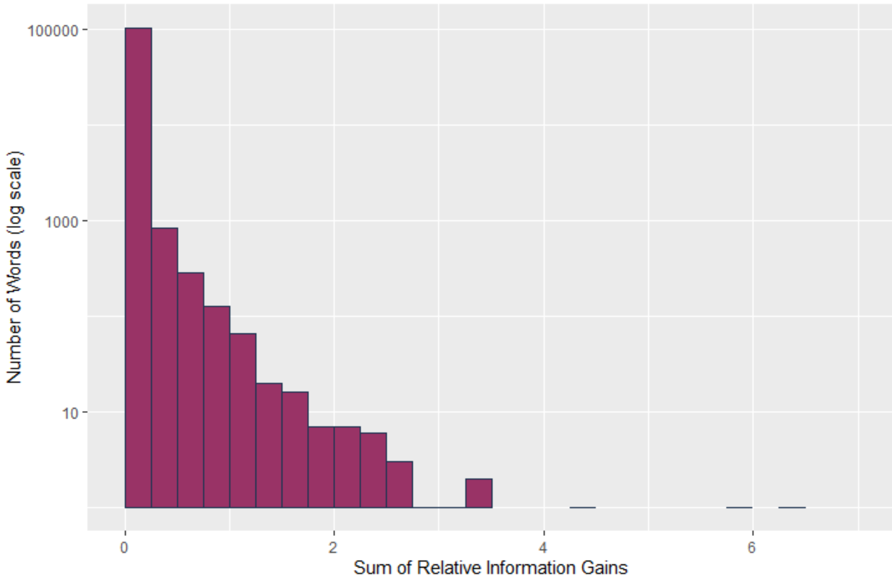


FIGURE 4.19. Histogram of the sum of RIGs for words of the LScDC (logarithmic scale for the y-axis)

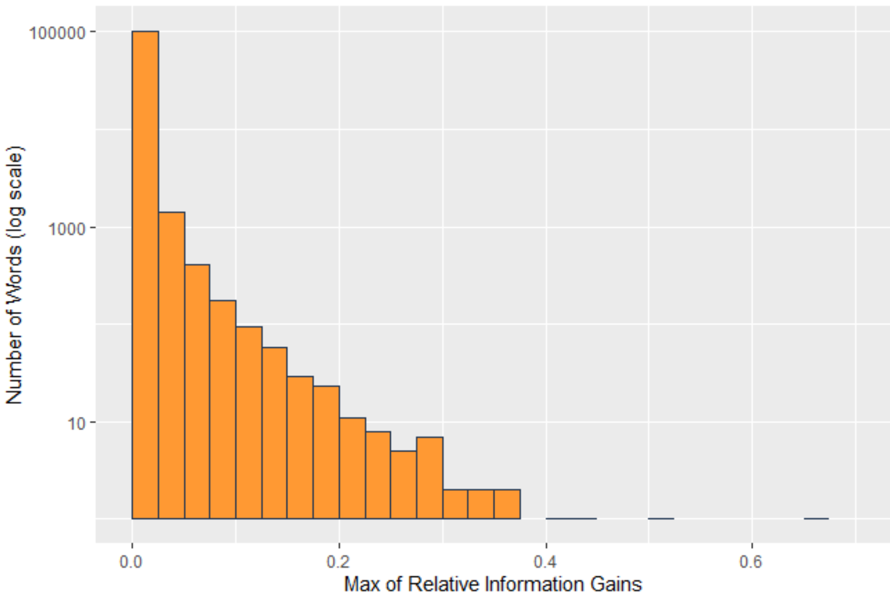


FIGURE 4.20. Histogram of the maximum of RIGs for words of the LScDC (logarithmic scale for the y-axis)

TABLE 4.11. The least informative 10 words that are arranged in ascending order based on the sum S_j and the maximum M_j of RIGs in 252 categories. Words are in stemmed form.

Words in the list where the sum of RIGs is calculated		**Words in the list where the maximum of RIGs is calculated**	
Word	S_j	**Word**	M_j
tgvs	0.000301	vhp	0.0000149
nonisland	0.000302	msmc	0.0000154
antipad	0.000308	scandat	0.0000179
aigan	0.000323	metaloxid	0.0000195
ultrabook	0.000324	ntfs	0.0000195
inzno	0.000324	interg	0.0000196
semiparallel	0.000328	nonfeas	0.0000196
22dbm	0.000328	biperiod	0.0000206
biperiod	0.000329	microl	0.0000206
150mhz	0.000330	multiparallel	0.0000206

4.4. Thesaurus for Science: Leicester Scientific Thesaurus (LScT)

In this section, we introduce a scientific thesaurus of English: LScT. LScT is a list of 5,000 words which are created by arranging words of LScDC in their informativeness in the scientific corpus. The procedure for creation of the thesaurus is described in detail.

Under the assumption that not all words having very low RIGs are informative in categories, we search a cut-off point for RIG to create a list of words that can be considered as relatively meaningful in scientific texts. In other words, we extract meaningful words for science from the LScDC to build a scientific thesaurus. Before moving on the decision taken to determine the number of words for the thesaurus, we recall the notion 'informativeness' and investigate further the criteria of S_j and M_j to arrange words of LScDC in their informativeness.

Having the top 100 words in two lists where words are descending ordered by their S_j and M_j, we see that the criteria of maximum is more likely to stand out some words that are frequently used in specific categories such as categories 'Dance', 'Music', 'Soil Science' and 'Theatre' (see Table 4.12) and are relatively rarely used outside them. Indeed, we expect drastic differences in RIGs of such words for these categories. For instance, one of the most informative word 'dance' is used in 154 categories, but the RIG from this word to the category 'Dance' is very distinguishable from all others (see Table 4.13). This is actually an expected result, since the word 'danc' is likely to be informative for categories related to the performing arts.

TABLE 4.12. Top 10 most informative words that are arranged in descending order based on the sum S_j and the maximum M_j of RIGs in 252 categories. These words are in the stemmed form.

List ordered by the sum S_j of RIGs		**List ordered by the maximum M_j of RIGs**	
Word	S_j	**Word**	M_j
patient	6.382	danc	0.657
conclus	5.766	music	0.508
paper	4.345	transplant	0.435
cell	3.394	soil	0.413
articl	3.336	virus	0.362
clinic	3.003	theatr	0.358
method	2.797	nurs	0.348
speci	2.686	eye	0.329
argu	2.581	cancer	0.324
background	2.562	student	0.302

TABLE 4.13. Five categories with the highest RIGs of the word 'danc'

Category	**RIG**
Dance	0.657
Theatre	0.042
Music	0.024
Folklore	0.009
Literary Reviews	0.008

To compare the meaningfulness of words across all categories, we tested two norms in the Meaning Space, l_1 (S_j or the sum of RIGs) and l_∞ (M_j or the maximal RIG). After a series of trials, we decided to use l_1. This choice cannot be proven formally but the ordering words by M_j leads to some words that are very specific in only one category but stand out in the list of the most informative words on average. The sum can be considered as more appropriate measure for general scientific thesaurus. When creating an LScT, we consider ordering the LScDC words by the sum of their RIGs in categories. The meaningfulness of words was evaluated by the average informativeness of words in the categories. Given the dictionary LScDC, the procedure to create the LScT is:

- Sort the words of the LScDC by their S_j in descending order.
- Take the top 5,000 words.

To find the number of words to be contained in the thesaurus, we initially follow an empirical procedure:

(1) Having arranged list of words in descending order by S_j, take a sub-list of the top m words, denoted by T_m
(2) Create the histogram of S_j for the words in this sub-list
(3) Check the trend in the histogram
(4) Take words when the exponential pick is avoided and the histogram follows roughly linear trend.

We begin with investigating the top 50,000 words in arranged list as it is almost the half of the 103,998 words of the LScDC. As the trend in the histogram for 50,000 words was showing the same behaviour with the histogram of 103,998 words (see Fig. 4.19 and Fig. 4.21 (a)), there was no point to check a number between 50,000 and 103,998. We then decreased the number m to 10,000, 5,000, 2,000, 1,500 and finally 1,000. All histograms are presented in Fig. 4.21. We see a substantial change in the trend of the histogram when we take the subset of 5,000 words. The trend at that point is almost linear. After that, the first bin in the histogram is slightly becoming smaller and finally it disappears for 1,000 words.

In this step, we also checked the minimum of the sum of RIGs in the lists T_m to make sure that the minimal average informativeness in the list (to be selected)

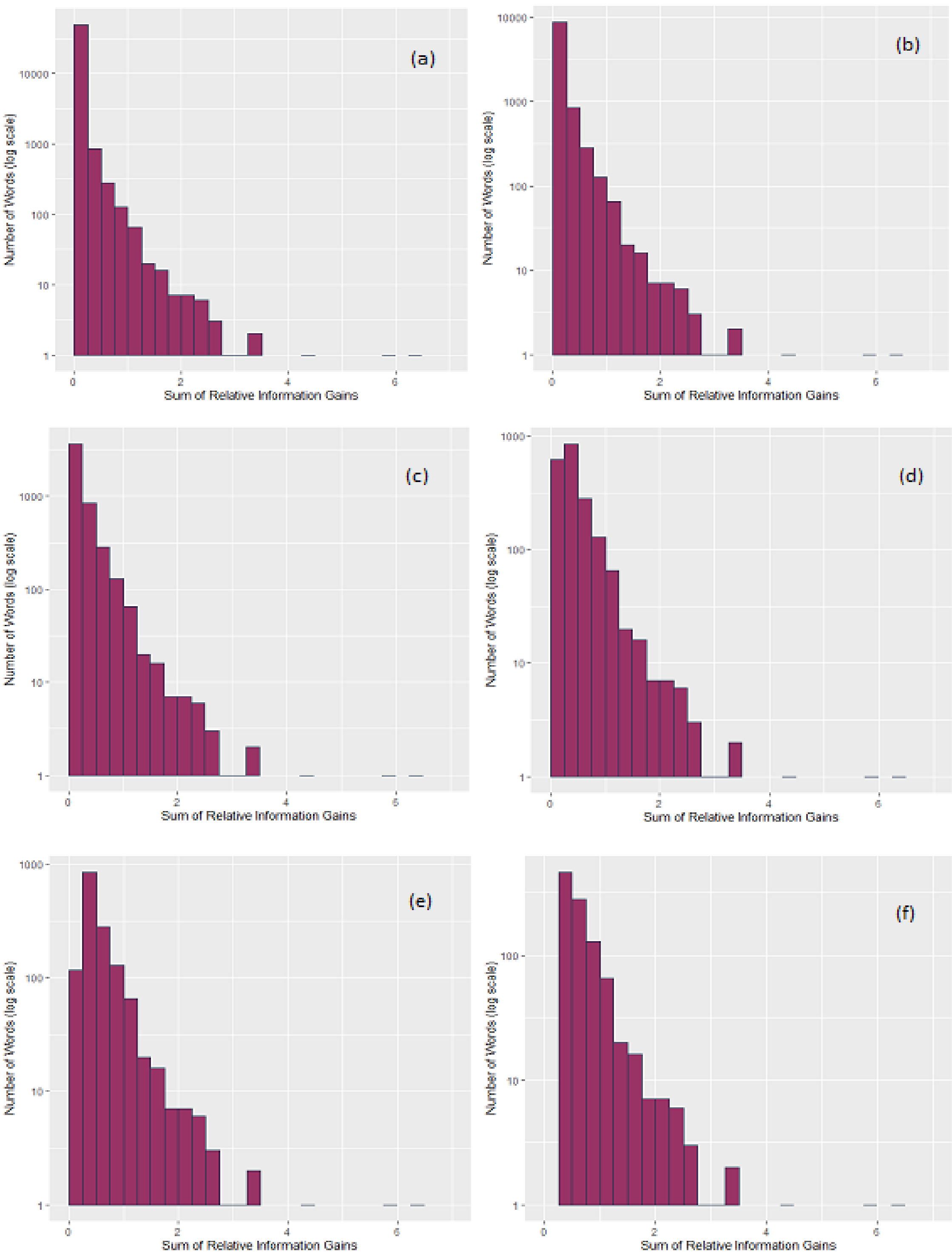


FIGURE 4.21. Histograms of the sums of RIGs for words of the LScDC (logarithmic scale for the y-axis): (a) The top 50,000 words (b) The top 10,000 words (c) The top 5,000 words (d) The top 2,000 words (e) The top 1,500 words (f) The top 1,000 words

is not so close to zero. These values are displayed in the Table 4.14. We can see from the table that the minimal RIG is decreased less than half from 1,000 to 2,000, while it decreased faster (more than halved) from 5,000 to 10,000.

TABLE 4.14. The minimal sum of RIGs in the list of the top m words (T_m)

List	Minimal S_j
$T_{1,000}$	0.3208
$T_{1,500}$	0.2329
$T_{2,000}$	0.1797
$T_{5,000}$	0.0658
$T_{10,000}$	0.0268
$T_{50,000}$	0.0027
$T_{103,998}$	0.0003

Finally, to support our selection of the number of words for the LScT and for evaluation of the result, we consider the following heuristic suggestion: the majority of words in the LScT appears in the list of top informative words in the categories. This does not mean that all informative words in categories should appear in the LScT, but we expect that most of top n words in categories will be included in the LScT. This is a natural result of the statistics (average) used for selection of the LScT. If a word in a specific category is informative with high RIG for only this category, this word may not appear in the LScT as we considered the average informativeness over categories.

We consider the matches of the list T_m with the most informative words in categories defined by the sum of RIGs. For collection $C_{k,n}$ of n most informative words in the category k, we define $X_n = \bigcup_{k=1}^{K} C_{k,n}$. Then we test the coverage of the list T_m by X_n. For each category in the Word-Category RIG matrix (a column in Table 4.4), order in descending by their RIGs. This gives a list of words sorted from the most informative to the least informative for this specific category. Then, individual collections $C_{k,n}$ are formed for each category.

The set $\bigcup C_{k,n}$ was formed with different numbers of words (n). We built the collections containing the most informative 100, 200, 300, 400 and 500 words in each category, and X_n is created by uniting them for each n. The numbers of words and the minimal RIGs of words in X_n are presented in Table 4.15. The minimal RIGs are checked to avoid zero/near-zero RIG in lists. One can see from the table that words in categories are not completely different. For instance, if all $C_{k,100}$ do not intersect then there should be 25,200 words in the list X_{100}, but there are just 6,254 words in this union, which is almost four times less. For other n, the result is similar, and the values of X_n follow almost a linear trend (see Fig. 4.22). That is, the intersection $\bigcap_{k=1}^{K} C_{k,n}$ is not empty. The intersection may be pairwise or q-wise for different q.

The coverage is calculated by counting the number of matches words of the list T_m and words of X_n. Table 4.16 illustrates the numbers of matches when n is 100,

TABLE 4.15. The number of words in the list X_n and the minimal RIGs for X_n

n	**Number of words in the list X_n**	**Minimal RIG**
100	6,254	0.0025
200	10,435	0.0016
300	13,850	0.0012
400	16,910	0.0009
500	19,790	0.0008

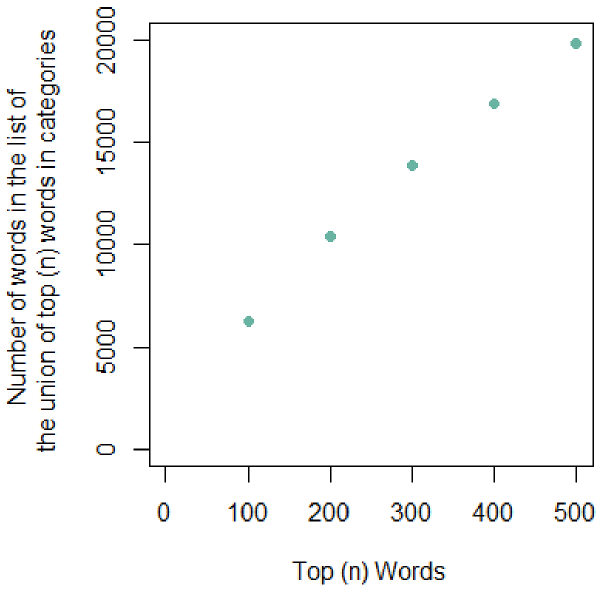


FIGURE 4.22. Number of words in the list of union of the top (n) words in categories (X_n)

200 and 500. Up to top 2,000 words, the words are concordant in the lists T_m and X_n, suggesting that the most informative words are highly consistent. In fact, words in the lists are in agreement for the case where 500 words in each category are considered as informative for categories. Given the list X_{100}, the majority of the words (3,992 words) in the $T_{5,000}$ can be covered by words of X_{100}. This trend changes and goes down when we consider the percentage of words found in 10,000 and 50,000 words. However, in this stage we have to consider the total number of words in X_n. For instance, the number of words in X_{100} is 6,254. In this case, the matches cannot be more than this number for the list $T_{10,000}$. Similar conclusions were obtained by comparing the number of matches for 5,000, 10,000 and 50,000 words for n=200.

We examined various heuristic criteria to evaluate how many words are suitable for inclusion in a thesaurus. Since we want to keep the size of thesaurus reasonable,

TABLE 4.16. Number of matched words of the list T_m and words of X_n. n is the number of words taken from each category to create X_n.

T_m \ X_n	X_{100}	X_{200}	X_{500}
$T_{1,000}$	995	1,000	1,000
$T_{1,500}$	1,469	1,499	1,500
$T_{2,000}$	1,908	1,992	2,000
$T_{5,000}$	3,992	4,674	4,987
$T_{10,000}$	5,631	7,745	9,588
$T_{50,000}$	6,254	10,435	19,754
$T_{103,998}$	6,254	10,435	19,790

and pay attention not to loose many words in case there might be informative words having not very high RIGs, we decided to include these 5,000 words ($T_{5,000}$) in the scientific thesaurus. This thesaurus is called Leicester Scientific Thesaurus. It is published online [**16**].

4.5. Conclusion and Discussion

In this work, we have studied the first stage of 'quantifying of meaning' for scientific texts: constructing the space of meaning. We have introduced the *Meaning Space* for scientific texts based on computational analysis of situations of words' use. The situation of use of the word is described by the absence/presence of the word in the text in scientific subject categories. The meaning of the text is hidden in the situations of usage and should be extracted by evaluating the situation related to the text as a whole.

This research is done based on 1,673,350 texts from the LSC and its 103,998 words listed in the LScDC [**13, 15**]. A text in the LSC belongs to at least one and at most six of 252 WoS categories presented in Table C.2. That is, categories can intersect. The situation of use is described by these 252 binary attributes of the text. These attributes have the form: a text is present (or not present) in a category. The meaning of a word is determined by categorising texts that contain the word and texts that do not. It is represented by the 252-dimensional vector of RIG about the categories that the text belongs to, which can be obtained from observing the word in the text. This representation is demonstrated in Table 4.4. Each text in the LSC can be considered as a cloud of these RIG vectors.

We begin with representing each word as a vector of frequencies in categories (Table 4.1). Components of a vector are the number of texts containing the word and belonging to the corresponding category. Then we moved on to representing the meaning of a word as a RIGs vector about categories.

We consider the corpus (LSC) as a probabilistic sample space (the space of equal probable elementary outcomes). The function is defined on texts that takes the value 1, if the text belongs to the category c_k, and 0 otherwise. Similarly, for each word w_j, a function is defined on texts that takes the value 1 if the word w_j belongs to the text, and 0 otherwise. Both functions can be considered as the random Boolean variables. The information gain $IG(c_k, w_j)$ about the category c_k from the word w_j is calculated by (5), (6) and (7). $IG(c_k, w_j)$ measures the amount of information extracted from observing the word w_j in the text on prediction of belonging of this text to the category c_k. The $RIG(c_k, w_j)$ is calculated by (10) that provides us a normalised measure of Information Gain giving the ability of comparing information gains for different categories.

Vectors of RIGs are denoted by $\overrightarrow{RIG_j}$ for a word w_j. $\overrightarrow{RIG_j}$ vectors for all words are presented in a *Word-Category RIG Matrix* (see the structure in Table 4.4) (available online [**16**]). A column vector of the matrix contains RIGs for all words in an individual category and a row vector represents the corresponding word's meaning as a vector of RIGs for categories. The *Meaning Space* has been described as a 252-dimensional vector space, where vectors are $\overrightarrow{RIG_j}$. Beyond the representation of words, the Word-Category RIG Matrix can be also used for the ordering words in a category from the most informative to the least informative as well as identifying the most informative words in the science for different subjects and their combinations. Ranking of words in a scientific corpus are performed based on two criteria: sum of RIGs (S_j) and maximum of RIGs (M_j) in a row vector. Calculations are done by (11) and (12). Given an ordered list of words, the top n words are considered as the most informative n words in the scientific corpus.

The LSC and LScDC were created and available online [**64, 68, 112**]. The proposed word representation technique was applied to this version of the corpus. The evaluation of the model is done based on checking the most informative words in each category. Word clouds are generated using words in lists for each category (For example, see Fig. 4.4 and Fig. 4.5). The higher RIG a word has, the bigger font size of the word is in the cloud. The clouds demonstrated that our methodology is able to identify topic specific words for categories, and most of the top words are related to the category subjects.

We note, however, that some words that were not expected to be appearing as the most informative words were prominent for some categories (Fig. 4.5). We concluded that words occurring in copyright notices, permission policies and the names of journals and organisations are added at the footer of abstracts in WoS database (see Table 4.5, Table 4.6 and Table C.1). Such joints result in anomalies in the word clouds and our representation technique was able to detect them. A further cleaning on identified phrases, sentences and paragraphs was performed to avoid possible abnormal appearances of words in the lists. This is done by sampling of texts based on keywords search and then deleting them from the texts. After cleaning procedure, new versions of the LSC and the LScDC are created by the same pre-processing steps as for the previous versions and can be found in [13, 15].

Words of LScDC were represented by vectors of RIGs in 252-dimensional Meaning Space as described before. The Word-Category Matrix for the LSC was formed with the collection of all words of the LScDC [16]. The sum S_j and the maximum M_j of RIGs in categories are calculated and added at the end of the matrix. Word clouds with the top 100 words and histograms of the most informative 10 words for each category are presented in [204]. The most informative 100 words for each category with their RIGs can be found in [204] and [205]. The proposed model of RIG-based word representation is analysed through these top ranked words in each category.

We have evaluated the Meaning Space by comparing our approach to traditional frequency-based model. Words in each category were also ranked and ordered by their raw frequencies in categories. It is proven that frequencies is not much important and efficient to represent scientifically specific meanings of words as the most frequent words are not topic related words such as ‘use’, ‘studi’ and ‘result’. Fig. 4.13, Fig. 4.14 and Fig. 4.15 compare two approaches using word clouds for three categories. The word clouds demonstrated that the information gain-based method is capable of standing topic-specific words out. This proves that the frequency is not much important in identifying such words. By representing words in the Meaning Space, we have shown by the human inspection that the top words in categories are topic-related in the corresponding category. It can therefore be viewed as an evidence of the usefulness of the Meaning Space and the representing words in this space.

S_j and M_j have been calculated for the LScDC words and two lists of words are created with words that are in descending orders by their S_j and M_j. The lists enable sorting the most important n words in science. We have compared these lists. The number of matches in the top n words in two lists are counted, where n is ranges from 100 to 50,000 (Table 4.10). The top 10 words in two lists are completely different and only 28% of words match in the first 100 words. This follows approximately 50% for the first 1,000 and 58% for the first 2,000 words. This concludes that two lists are not the same for the top words (Fig. 4.17 and Fig. 4.18), however, both criteria can be used for selection of top n words regarding to task and the information required. Many words in the lists have low S_j and M_j values. The plot of the number of words for S_j and M_j indicate a super exponential picks near zero S_j and M_j (see Fig. 4.19 and Fig. 4.20). The trend beyond the pick is going down almost linearly. Those words with near zero values can be considered as less meaningful words for scientific texts.

Finally, a scientific thesaurus of English, LScT, has been introduced. The thesaurus contains of the most informative 5,000 words from the LScDC. Words in the LScT are selected by their average RIGs in categories. That is, the top 5,000 most informative words in the LScDC, where words are arranged by their S_j are considered as the most meaningful 5,000 words in scientific texts. The full list of words in the LScT with their S_j can be found in [**16**].

The next focus of the research in 'quantifying of meaning' will be extraction of the meaning of text in scientific corpus from the clouds of words in the Meaning Space and study of more complex models in which co-occurrence of words and combination of word's meaning will be used. This, we follow the road: Corpus of texts + categories → Meaning Space for words → Geometric representation of the meaning of texts. The first two technical steps were done: the Corpus of texts was collected and cleaned, and the meaning of words was represented and analysed in the Meaning Space. The next step will be analysis of the meaning of texts.

The analysis of dictionaries is not finalised yet. This work was focused on the most informative words. They are the main *scientific content words*. But, for example, the frequent but non-informative words (like 'use') can be considered as *generalised service words of Science* and deserve special analysis.

It is also very desirable to extend the set of attributes for representation of the situation behind the text (Figures 4.2, 4.3). The first choice, the research subject categories, is simple and natural, but it may be useful to enrich this list of attributes.

CHAPTER 5

Principal Components of the Meaning

In this chapter we argue that (lexical) meaning in science can be represented in a 13 dimension Meaning Space. This space is constructed using principal component analysis (singular decomposition) on the matrix of word-category relative information gains, where the categories are those used by the WoS, and the words are taken from a reduced word set from texts in the WoS. We show that this reduced word set plausibly represents all texts in the corpus, so that the principal component analysis has some objective meaning with respect to the corpus. We argue that 13 dimensions is adequate to describe the meaning of scientific texts, and hypothesise about the qualitative meaning of the principal components.

5.1. Introduction

The purpose of this chapter is to extract the meaning (lexical meaning) of scientific texts by careful analysis of how the words used in documents specify the category the document belongs to [**205**]. The scientific category that a text belongs to is used to evaluate the meaning of the text. Stated differently, the research areas behind the text are characterised by these scientific categories.

In Chapter 4, the meaning space is created for the LSC with 1,673,350 texts and the LScDC of 103,998 words with 252 WoS categories [**13, 15, 112, 205**]. Each word in the LScDC is represented by 252-dimensional vector of RIGs. We have constructed the Word-Category RIG Matrix, where each entry corresponds to a pair (word,category) and its value shows the RIG for a text to belong to a category by observing this word in this text [**16**]. We introduced the LScT where the most informative 5,000 words from the LScDC were included in [**16**]. Later we will use the LScT in the study of the representation of the meaning of texts.

The characterisation of word meaning in a metric space was the first stage in the quantification of meaning in texts. We have built a metric system to allocate meaning to words based on their importance in scientific categories, i.e, to represent each word as a vector in the MS.

Given 252 subject categories, it is unreasonable to expect that every one of these categories is uncorrelated to all others (or distinct from each other). For instance we might expect the categories *Literature* and *Literary Theory & Criticism* to represent words in a very similar way in the Meaning Space. Indeed, subcategories are likely to occur in the data and they are expected to have close values of RIGs for words. Such attributes will measure related information, and so the original 252 dimensional data contain measurements for redundant categories. Although the MS underlying the representation of word meaning has 252 dimensions, we expect that we will be able to represent words with significantly fewer dimensions of the Meaning Space.

An efficient way to represent words would be to map vectors onto a space that is constructed based on combination of original features. Mathematically speaking, we look at a linear transformation from the original set of categories to a new space composed by new components. These new components are called *Components of the Meaning*. Two precise questions to be asked are: how many components of meaning are there and how are these components constructed? Thus, analysis of components (new attributes) based on the original attributes is crucial in understanding the MS. For instance, it is very important to understand which categories contribute the most and which the least to the new dimensions. Also, it is instructive to see if the new dimensions have some real semantic meaning, for instance, in distinguishing between natural and social sciences or experimental and theoretical research.

We narrowed the Word-Category RIG Matrix to words from the LScT; therefore, there are 5,000 rows and 252 columns in the matrix. Here we hypothesize that 252 is not the actual dimension of the Meaning Space. In order to identify the most significant dimensions and construct a new space, we perform Principal Component Analysis (PCA) [**206**]. PCA ensures that words which are represented by similar sets of categories will be nearby points in the lower dimensional space. Through PCA on vectors of RIGs, we map each vector onto a vector with a reduced number of entries.

The PCA of word vectors provides a series of 252 principal components (PC), called as *Principal Components of the Meaning*. Of the 252 PCs produced by PCA, only the first m will resemble the true underlying MS, while the remainder will be mainly represent noise in the data. The most significant m principal components are used to construct new orthogonal axes that span an m-dimensional vector space. Using these m components, every word with 252 dimensions can be mapped onto a word vector with m components. The optimal value of the m can be estimated using several methods [**207, 208, 209**].

It is noteworthy that this analysis can be performed under the following assumptions: (1) the thesaurus is representative of the corpus LSC; (2) each category is represented by reasonable number of words of the LScT, and (3) each word can be represented by the WoS categories. To understand the relationships between the thesaurus and the corpus, and categories and words, we evaluated the LScT in three different ways.

Firstly, we focused on showing how well the LScT represents the texts of the LSC. If there is a text including none of words from the LScT, then this text can not be represented by the LScT. We counted the number of the LScT words in each text and came to the conclusion that all texts of the LSC contains at least 1 word of the LScT, at most 194 words, with an average of 62 words. Indeed, one may think that representing a text with only one word is not the best in quantification of text meaning. However, we will deal with this problem in later research on text representation. We now focus on showing that the LScT is a reasonable selection as a scientific thesaurus.

Secondly, we used a procedure to test the existence of the LScT words which are informative for categories. As we restrict words in the Word-Category Frequency Matrix to the LScT words, it is possible to have column vectors with 0 in all entries. This means that none of words from the LScT is present in the texts of a particular category. For such a category, the LScT would not be an appropriate set of words to represent texts of this category. We can further infer that these words are not a representative set of words for the LSC. This is because we assume that the texts of the LSC selected from the range of 252 subject categories are representative of the population of scientific texts as a whole. Thus none of these categories can be ignored and texts in all categories have to be represented by the selected words. The analysis of each category shows that the LScT is a reasonable selection of words, as the minimum number of words that a category includes in its texts is 733. This means that each category is presented by at least around 15% of the LScT.

Thirdly, we examined how many categories determine a word in the LScT. We expect that some words appear in all or the majority of categories, while some words are present in only a few categories. Even if we do not take the values of RIGs into account, we can still gain an insight into different types of vocabulary: *scientific content words* and *generalised service words of science* [**205**]. One might expect that words appearing in text(s) of all categories are not topic-specific words and words appearing in text(s) of a few categories are likely to be specific for the subjects. Therefore, we looked at the distribution of the number of words that are informative for categories, and found that more than 4,500 words appear in text (s) of at least 100 categories. Approximately 200 words appear in text(s) of all categories and only 4 words appear in text(s) of only 10 categories with a minimum number of 6 categories for a word.

Previous results confirm that the selection of words for the LScT was reasonable and further study of quantification of the meaning of text can be done with the LScT. Thus we can use the RIG matrix based on 5,000 LScT words to create a Meaning Space based on PCA. Firstly, we applied the Double Kaiser rule to attempt to select a subset of the original attributes by ranking them in their importance determined based on some criteria (to be explained later). Using this rule, we aim only to retain attributes which explain the data in some significant way, and to discard 'trivial' attributes. Having discarded some attributes we can repeat the process, and perhaps discard some more. As it turned out, all attributes (categories) were found to be informative by our selection criterion at the first iteration. Therefore, all 252 categories are retained for further use. This means that there is no trivial attribute.

We applied PCA to the data in the 252-dimensional MS and produced 252 PCs. To understand the structure of the space, we visualised the space in two ways.

Firstly, to describe each PC in terms of categories we created charts in which category contributions to the PC (252 coordinates of PC) are shown. These charts are used to evaluate the contributions and identify which categories have the largest influence on each PC. Using this approach we can see the main attributes for a PC and the attributes that do not contribute to that PC. Categories contributing greatly (either positively or negatively) to a PC are used to interpret that dimension. By 'an attribute that is contributed greatly' we mean those attributes having coefficients larger than $1/2\sqrt{252}$ in size (positive or negative). Categories with little influence get scores near zero. Such categories can be interpreted as being unrelated attributes to the PC and this information might be useful. Therefore, all attributes are meaningful in some sense and should be interpreted appropriately.

Secondly, we have projected words onto the space defined by the PCs. This allows us to see a map of how words relate to each other with respect to the principal axes. The words used for similar topics are expected to be located near each other in these plots.

The first PC explains 12.58% of all variation in the data. Coordinates of the vector of the first PC appear to measure the extent to which a category is well-defined by the words in texts in the category. Categories with more precise language have relatively higher entries in this PC, and categories with more nebulous language score lower. We did not observe any explicit distinguishing between the branches of categories. Every word is represented by the vector of RIGs. Projection of this vector on the first PC can be considered as a general *informational value* of the word. This informational value is the first coordinate of the word in the PC basis.

The second PC (PC2) appears to primarily distinguish between categories of two main branches of science. Categories that are related to natural science and engineering have negative associations with the PC2, while those that are related to social/human science have positive associations.

The third, fourth and fifth PCs reflect some of sub-branches of science with the greatest influence on the PCs. Biological sciences, computer science and engineering related categories have a strong positive correlation with PC3, while categories that are related to psychology, medicine-health and applied physics have large negative correlations. Branches of social science, economics, managements, psychology, ethics, education and multidisciplinary social sciences, appear to have strong positive correlations PC4. Literature, medicine-health science have large negative scores. For the fifth PC, categories related to ecological, environmental sciences and geosciences have large correlation.

Next, the words are plotted in the PCA space. PCA scores for words are calculated to determine their location on each PC and the data are shown on plots on planes defined by the first three PCs (PC1-PC2, PC1-PC3, PC2-PC3) and also the volume spanned by the first three PCs together (PC1-PC2-PC3). Words that are represented by similar sets of categories will be located as nearby points in the space. In other words, PCA ensures that words that are close together in PCA space have similarity of meaning. From these plots, we have concluded that some more closely grouped words are similar in that they are used in related academic disciplines. For instance, the words 'argu' and 'polit' are close in the PC1-PC2 plot and they both are expected to be predominantly used in texts assigned to social science categories. Similarly, the words 'health', 'care', and 'particip' are close together in the PC1-PC3 plot, and they are all likely to be used in medicine related texts. Thus, we conclude that in the PC space we are able to cluster words based on their meaning. However, more meaningful clusters can be obtained by using more PCs, as this will allow more separation as of groups of other meaning. With the dimension of the PC space appropriately selected, we expect that vectors of words with similar meanings will form clusters.

Finally, we pick up the first m PCs, which are in descending order of amount of information content in the PC. Three approaches to determining the appropriate number of components to retain were performed; the Kaiser rule, the Broken stick rule and an empirical method based on multicollinearity control (called PCA-CN) were applied to the data [**210, 211, 212, 213**]. We estimated the optimal number of PCs at $m = 61$, $m = 16$ and $m = 13$ by these methods respectively. Too large an m might lead to the isolation of each word vector in the PC space, leading to the possibility of over-fitting related to the curse of dimensionality, especially for a small number of words. Therefore, we can use the 13-dimensional Meaning Space delivered by PCA-CN as the actual dimensions of meaning in future research. Full lists of categories receiving large positive and negative scores on the first 13 PCs are presented in Appendix D.3.

This chapter is organised as follows. In Section 5.2, we test the LScT in a variety of ways to demonstrate that the selection of LScT words is appropriate for the representation of texts. In Section 5.3 we perform a principal component analysis on the Meaning Space. To determine the optimal number of PCs, three methods are applied and the results are discussed. The principal components of meaning are explicitly presented and analysed in some details. Finally, in Section 5.4 we summarise the results of the chapter, discuss their significance, and outline future research directions.

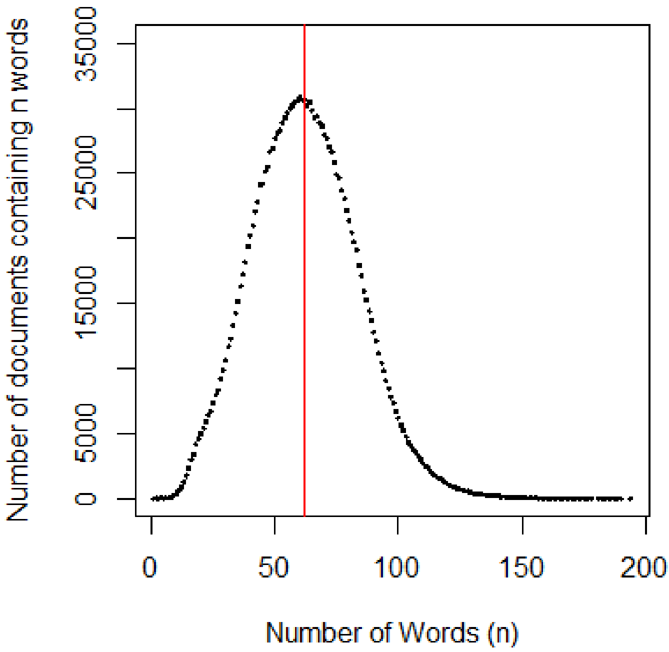


FIGURE 5.1. The number of words (n) from the LScT against the number of documents from the LSC containing n words. The minimum and maximum numbers of words contained in a text are 1 and 194 respectively, with an average of 62. The red line shows the average number of words.

5.2. Basic Statistics in the LScT

In this section we investigate the word-text and word-category relations for the LScT. We ask how many texts in the corpus LSC contain a certain number of words from the LScT (is the LScT representative of the LSC), what proportion of the thesaurus each category is present in, and how many categories each word appears in.

The LScT was created using the LSC with the aim of using a significantly reduced set of words to be representative of scientific texts in the LSC. Representativeness here refers to the extent to which words of the LScT are included in texts of the corpus LSC. Our criterion is that each text from the LSC must contain at least one word from the LScT to make the contents of the thesaurus a reasonable selection. If there is a text that contains no word from the thesaurus, the LScT is not representative for this text. Fig. 5.1 shows the number of texts containing a specified number of words from the LScT. All texts of the LSC include at least 1 word from the LScT, and no text contains more than 194 words, with an average of 62 words. The numbers of texts containing a small number of words are located near to 0 on the x-axis and the graph indicates that very few texts contain a small number of words from the LScT.

In [205], we discussed the relationship between the 252 WoS categories and words from the LScDC, and the subsequent word-category RIG matrix [16]. For each of the 252 categories, the most informative 100 words from the LScDC were listed in descending order of their RIGs. The lists of the most informative words in each category were given in [**205**, **204**]. Clearly, the LScT is a sub-list of the LScDC; however, the two lists the most informative words from the LScDC and LScT from a category may differ from each. Table 5.1 gives such an example for the category 'Mathematics'. We see that some very topic specific words (e.g. isomorph, cohomolog) in the LScDC are not present in the LScT. This is not surprising, as the LScT is representative of the average meaning of words (lexical meaning) in categories. More specific words are not needed if all we wish to do is distinguish categories. For each category, the list of the most informative 100 words from the LScT are given in Appendix D.1.

TABLE 5.1. Uncommon words from the most informative 100 words from the LScT and LScDC for the category 'Mathematics'

Extra Words in the LScT	**Extra Words in the LScDC**
affect	isomorph
higher	denot
cause	cohomolog
research	automorph
examine	semigroup
protein	abelian

For the word-category relation task, our first focus is to see how well each category is represented by the LScT. Table D.2 gives the number of words from the LScT for each category. The number given is the number of words appearing in the texts assigned to the given category. To gain a better insight, we provide Table 5.2 where we list the number of categories containing numbers words from the LScT in a particular range. As we see from the table, 219 categories contains at least 3,000 words of the LScT in their texts. The maximum and the minimum numbers of words that a category includes in its texts are 4,956 and 733 respectively. The top and the bottom five categories, when categories are sorted in descending order by the number of words, are presented in Table 5.3. The bottom five categories are also the five categories with the fewest number of texts. The top five categories in the Table 5.3 appear in the list of top 30 categories containing the most texts [**205**].

Another aspect of evaluating word-category relations is to understand the number of categories with texts containing a particular word. If the number of categories is small then this indicates the word is more specific for these categories and maybe interpreted as a *scientific content word*. We suggest that words appearing in all categories are not topic-specific and may be interpreted as *generalised service words of science* [**205**]. However, further research is necessary to identify these two groups of words.

TABLE 5.2. Intervals of word numbers and the number of categories for each interval. The number of words for each category is calculated by counting the number of words appearing in the texts of the category at least once.

Number of Words (n) from the LScT	**Number of Categories Containing n Words in Their Texts**
$4,000 < n \leq 5,000$	113
$3,000 < n \leq 4,000$	106
$2,000 < n \leq 3,000$	24
$1,000 < n \leq 2,000$	6
$n \leq 1,000$	3

TABLE 5.3. The top and the bottom five categories where categories are sorted in descending order by the number of words

Top 5 categories	**Number of Words**	**Bottom 5 categories**	**Number of Words**
Multidisciplinary Sciences	4956	Literature, Slavic	733
Environmental Sciences	4865	Literary Reviews	953
Engineering, Electrical & Electronic	4864	Poetry	982
Computer Science, Interdisciplinary Applications	4834	Literature, African, Australian, Canadian	1240
Public, Environmental & Occupational Health	4805	Dance	1295

Fig. 5.2 displays the number of categories (n) for which texts in that category contains a word a certain number of times. So a value of 1 against $n = 6$ indicates that there is one word that appears in only 6 categories. Hence, each word from the LScT appears in at least 6 and at most 252 categories. From the figure, we can see that there are approximately 200 words appearing in all categories. 2,570 words appear in more than 200 categories and approximately 2,000 words appear in between 100 and 200 categories. Very few words are located at the very beginning of the graph, and only 4 words appear less than 10 categories. Scores for the numbers of categories that some of the words appear in can be found in Table 5.4. One can see that words appearing in all categories are also the most frequent words in the LSC. Rare words have to be specific to a small number of categories.

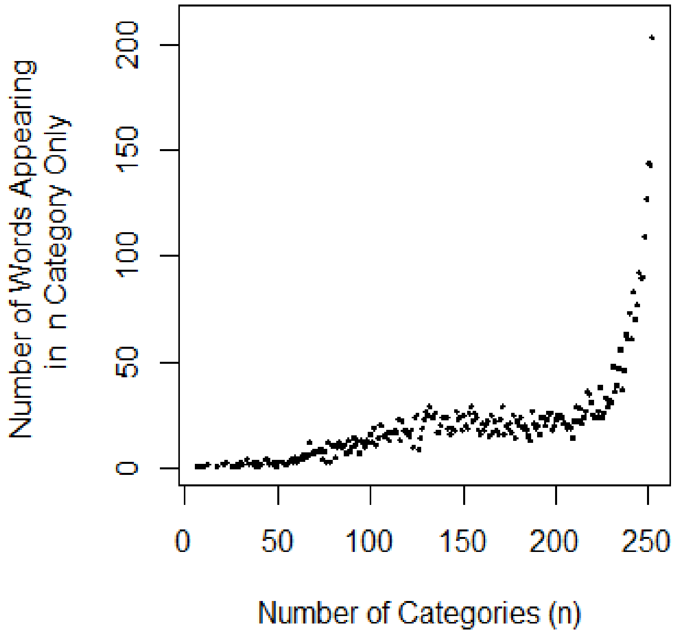


FIGURE 5.2. The number of categories (n) against the number of words appearing in n category. The minimum and maximum numbers of categories where a word appears in are 6 and 252.

TABLE 5.4. Some words of the LScT with the number of categories where each word appears in

Word	**Category Number**	**Word**	**Category Number**	**Word**	**Category Number**
use	252	august	234	antebellum	12
result	252	photographi	200	folklorist	10
studi	252	gravit	166	chaucer	9
show	252	paraffin	145	epigram	8
effect	252	ionospher	56	panegyr	6

5.3. Dimension of the Meaning Space

In this section, our focus is on answering the question: is 252 the actual number of dimensions in the Meaning Space for words of the LScT? We explore the Meaning Space by using Principal Component Analysis (PCA). We discuss the Double Kaiser Rule for the selection of important attributes, that is a subset of original attributes (categories).

The importance of words in each category was organised in the Word-Category RIG Matrix. With 5,000 words and 252 categories, it is difficult to see what information is present in the data. A way to understand the Meaning Space and visualise words' meaning (lexical meaning) in this space is to plot the words in individual categories. We expect that trends in some of plots will be very similar for a certain word, a group of words or all words. This may suggest that groups of words appear to be very close in specific categories. For instance, subcategories are very likely to occur in the corpus and related information will be measured by such attributes. This leads to some redundant attributes that measure the same information in multiple locations in the data.

Words can be similarly represented in two or more categories. If two categories are correlated in the MS, it is possible to represent words in a reduced dimension by using a suitable linear combination of these original attributes. More specifically, if two categories are completely correlated, we would use the sum of two categories as one new attribute. The new attribute can be considered as a representative of the two original attributes. PCA provides a solution to this problem. Linear combination weights (coefficients) are provided by PCA to create the new attribute, which we term a principal component (PC), with the aim of preserving as much variability as possible (the maximum variation in the data). The level of the effectiveness of PCA in explaining the data varies differently with the different sets of PCs. Therefore, in this we investigate the effectiveness of PCA as a technique for determining the actual dimension of the data. Our goal is also to empirically investigate the effectiveness of the RIG based word representation technique using PCs instead of the original attributes.

PCA is a statistical technique that transforms the data into a reduced-dimension represented by linearly uncorrelated attributes (PCs), where PCs are a linear combination of the original attributes [214, 215]. The Kaiser Rule is one of the methods developed to select the optimal number of components [216, 217]. Eigenvalues of the covariance matrix are used to determine the appropriate number by taking components with eigenvalues greater than 1; only components explaining greater data variance than the original attributes should be kept [218].

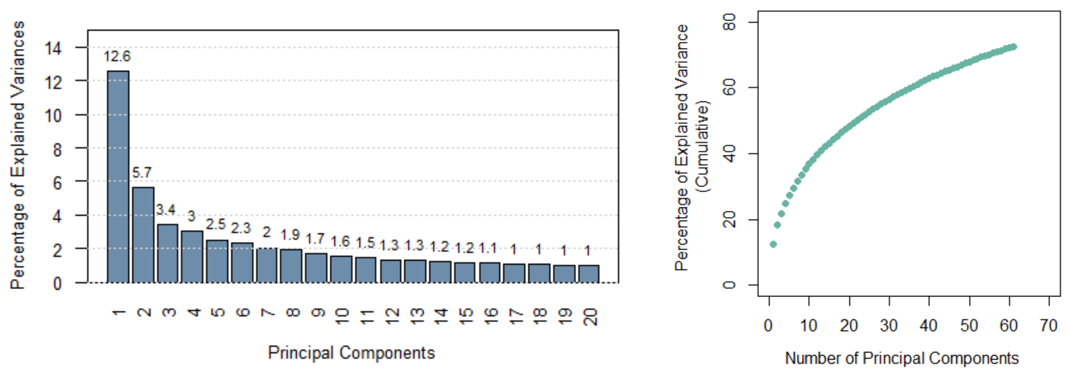
The Double Kaiser Rule is a method for selecting a subset of the original attributes based on their importance in representing the data. It evaluates PCs and the original attributes can be ranked in importance to the PC by the size (be it positive or negative) of their coordinate. The aim is to retain only a subset of the original attributes for further use. The advantage of employing the Double Kaiser Rule is that any remaining, so-called 'trivial', attributes can be discarded iteratively. The iterative algorithm for the Double Kaiser Rule is shown in Algorithm 1 (the trace of a square matrix is the sum of its diagonal elements, and also the sum of its eigenvalues).

PCs were assessed sequentially from the largest eigenvalue to the smallest. All PCs having eigenvalue less than 1 were considered to be trivial (non-significant) by the Kaiser rule. Hence 61 PCs are included as non-trivial, that is, 61 axes summarize the meaningful variation in the entire dataset. These non-trivial PCs are retained as informative at the first stage. The cumulative percentage of variance explained is displayed in Fig. 5.3. The cumulative percentage is approximately 73%, indicating

Algorithm 1 Double Kaiser Rule

Input: The Word-Category RIG Matrix and n (number of categories)
Output: Set of informative categories
1: **repeat**
2: Calculate the covariance Matrix **M** (standardised).
3: Calculate the Kaiser threshold (α) using $\dfrac{\text{trace}(M)}{n}$.
4: α is 1 for correlation matrix.
5: Select informative PCs by the Kaiser rule: Components with eigenvalues above α are informative.
6: Determine the importance of an attribute (β) as the maximum of the absolute values of coordinates in informative PCs for the attribute.
7: Select the informative attributes: (a) Calculate the threshold of importance for an attribute as the root mean squared of values in the coordinate. The threshold of importance for an attribute is $\dfrac{1}{\sqrt{n}}$ for unit vectors. (b) Select attributes which hold $\beta \geq \dfrac{1}{\sqrt{n}}$.
8: Otherwise, the attribute is trivial.
9: Drop the most trivial attribute from the Word-Category RIG Matrix if any.
10: $n = n - 1$
11: **until** There is no trivial attribute.
12: **return** The set of informative attributes (categories).

the variance accounted for by the first 61 components. They explain nearly 73% of the variability in the original 252 attributes, so we can reduce the complexity of the data four times approximately, with only a 27% loss of information.



(a) (b)

FIGURE 5.3. (a) Fraction of variance explained as a function of PCs retained for categories; (b) Cumulative fraction of variance explained as a function of PCs retained for categories (61 PCs)

Following the data reduction via PCA we then restricted the analysis of the informative categories to the non-trivial PCs; these are used to list informative attributes (categories). The importance of an attribute is determined as the maximum of the absolute values in coordinates of informative PCs for this attribute. The threshold

$1/\sqrt{252}$ (threshold of importance) is used in the selection of informative attributes. None of the attributes was dropped after the first iteration of the Double Kaiser selection, so all 252 categories were retained.

To interpret each component, the coefficients (influence) of the linear combination of the original attributes for the first five principal components are examined (see Figures 5.5, 5.6, 5.7, 5.8, 5.9). The coordinates of the attribute divided by the square root of the eigenvalue gives the unit eigenvector, whose components give the cosine of the angle of rotation of the category to the PC. Furthermore, positive values indicate a positive correlation between an attribute and a PC and negative values indicate a negative correlation. Both the magnitude and direction of coefficients for the original attributes are taken into account. The larger the absolute value of the coefficient, the more important the corresponding attribute is in calculating the PC. Positive and negative scores in PCs push the overall score of a word in the Meaning Space to the right or left on the PC axis.

To examine the original attributes in the PCs, we introduce a threshold for categories having near zero values. The threshold used was $1/2\sqrt{252}$, which is half of the threshold of importance in selection of informative attributes. All values between $-1/2\sqrt{252}$ and $1/2\sqrt{252}$ are considered to be negligible so are in the *zero interval*. Hence, the initial attributes are considered as belonging to three groups: (1) positive, (2) negative, and (3) zero. We interpret the categories belonging to positive and negative groups as the main coordinates of the dimension as these categories contribute significantly to that direction. Categories belonging to the 'zero' group are seemed to be unrelated attributes to the PC. However, this information could be also useful. Hence, all categories in the three groups are meaningful and should be interpreted.

Categories in the three groups for each PC can be seen in Figures 5.5, 5.6, 5.7, 5.8, 5.9. The zero interval is shown by a line in the figures. To understand the trend in the first five PCs, the average, maximum, and minimum values for positive, zero, and negative groups are displayed in Fig. 5.4. Similarly, the sum of values in each group is plotted (see Fig. 5.4). The number of categories in each group is presented in Table 5.5.

Mean values in positive and negative groups are close in all PCs with a slight change for maximum and minimum values. We can see that for the second principal component, the difference between means of positive and zero groups is higher than all others. For the sum of values, the highest value belongs to the first PC as there is no negative value for this component. The full list of categories in positive, negative and zero groups for each of five PCs can be found in Appendix D.3.

The list of categories appearing in the zero interval in the first five PCs are presented in Table 5.6. One may expect that as these categories contribute very small positive or negative values to the overall component score, they will not have

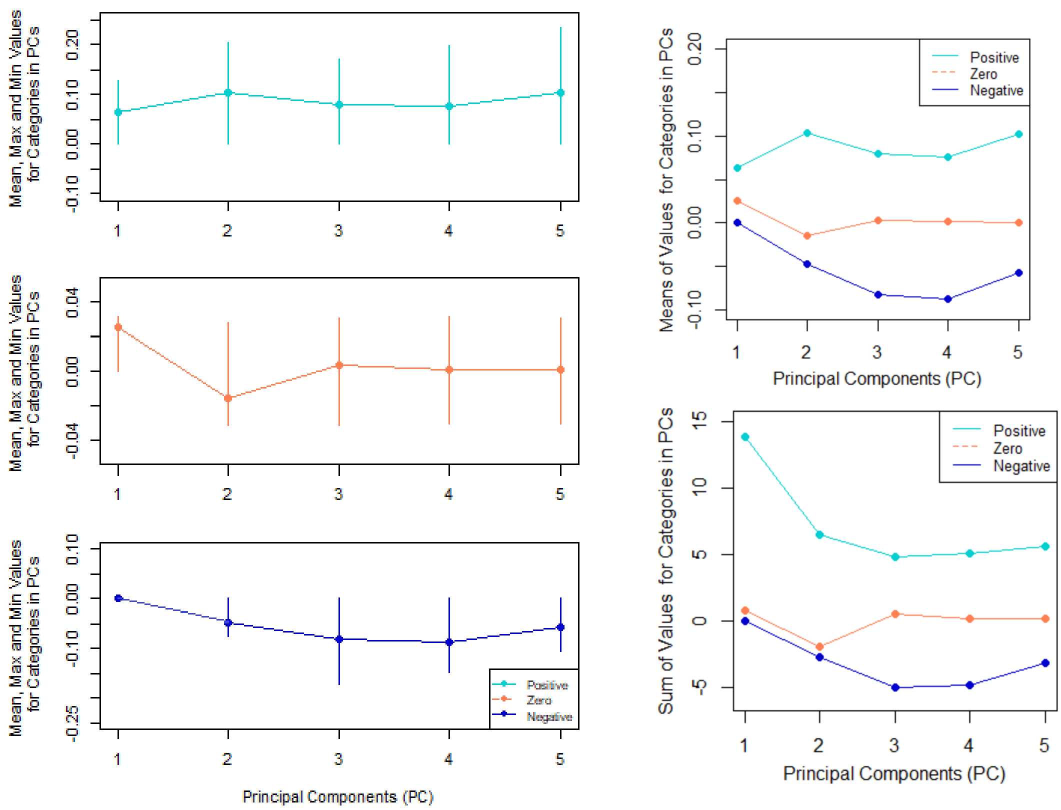


FIGURE 5.4. (Left) The average, maximum, and minimum values (coefficients of the linear combination of the original attributes) in five principal components for the groups of positive, negative and zero. (Top right) The average values in the groups of positive, negative and zero. (Bottom right) The sum of values in the groups of positive, negative and zero.

TABLE 5.5. Number of categories in the groups of positive, zero and negative for the first five principal components

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
Positive	221	63	60	67	55
Zero	31	131	131	129	142
Negative	0	58	61	56	55

a big effect in moving words with high RIG for that category, plotted in the Meaning Space, towards the associated end of the principal axis.

We can see that there are no negative values for the first principal component. The first component primarily measures the magnitude of the contribution of categories to the PC. It is a weighted average of all initial attributes. The most prominent categories are 'Engineering, Multidisciplinary' and 'Engineering, Electrical& Electronic', that is, they strongly influence the component. This component explains 12.58 % of all the variation in the data (see Fig. 5.3 (a) and Fig. 5.5). This means that more than 85% of of the variation still retained in the other PCs.

TABLE 5.6. Categories (initial attributes) appearing in zero interval in the first five components with the number of texts assigned to the category

Categories	**Number of Texts**
Agriculture, Dairy & Animal Science	6,163
Andrology	391
Astronomy & Astrophysics	22,825
Materials Science, Paper & Wood	1,963
Medicine, Legal	1,711

The second component has positive associations with categories related to social sciences and humanities, and negative associations with categories related to engineering and natural sciences (see Fig. 5.6). The plot shows that they are completely oppositely correlated. Hence, this component primarily measures the separation of two main branches of science. The most prominent category in the component is 'Cultural Studies'. This contributes a large positive value to overall component score, that is, it pushes the scores of words with high RIG for 'Cultural Studies' to the right on the axis. The largest negative contribution to the component score is from the category 'Engineering, Multidisciplinary', which is approximately 2.5 times smaller than the contribution of 'Cultural Studies'. In the zero interval, extremely low values are present for attributes such as 'Psychology, Developmental', 'Ergonomics' and 'Medicine, Legal'. Such attributes do not influence the movement of words to the extreme ends of this PC.

The largest positive values on the third component can be interpreted as contrasting the biological science, computer science and engineering related areas with medicine, social care and some other disciplines (Fig. 5.7). We may expect words that are used in biological science, computer science and engineering will go toward the positive side of the axis on the third principal coordinate. The largest negative values suggest a strong effect of psychology, medicine-health and physics related areas.

The other two principal components can be interpreted in the same manner (See Fig. 5.8 and Fig. 5.9). In the fourth component, the most prominent categories with positive values are some of social science branches such as economics, managements, psychology, ethics, education and multidisciplinary social science. It can be seen large negative values for categories related to literature and medicine-health science. The fifth component has large positive associations with ecological, environmental sciences and geosciences.

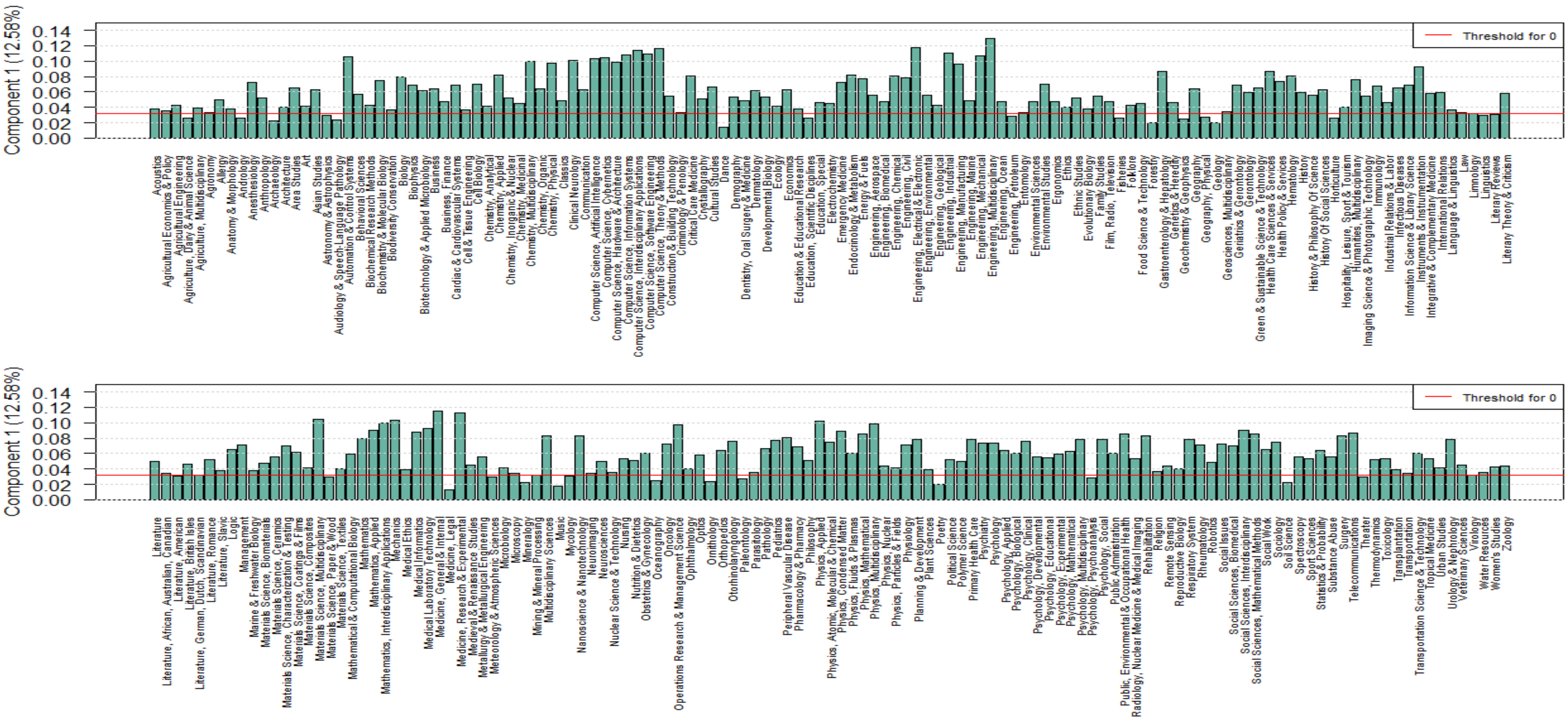


FIGURE 5.5. The first principal component of the LScT. The plot shows the contributions of original attributes (categories) on the first principal component.

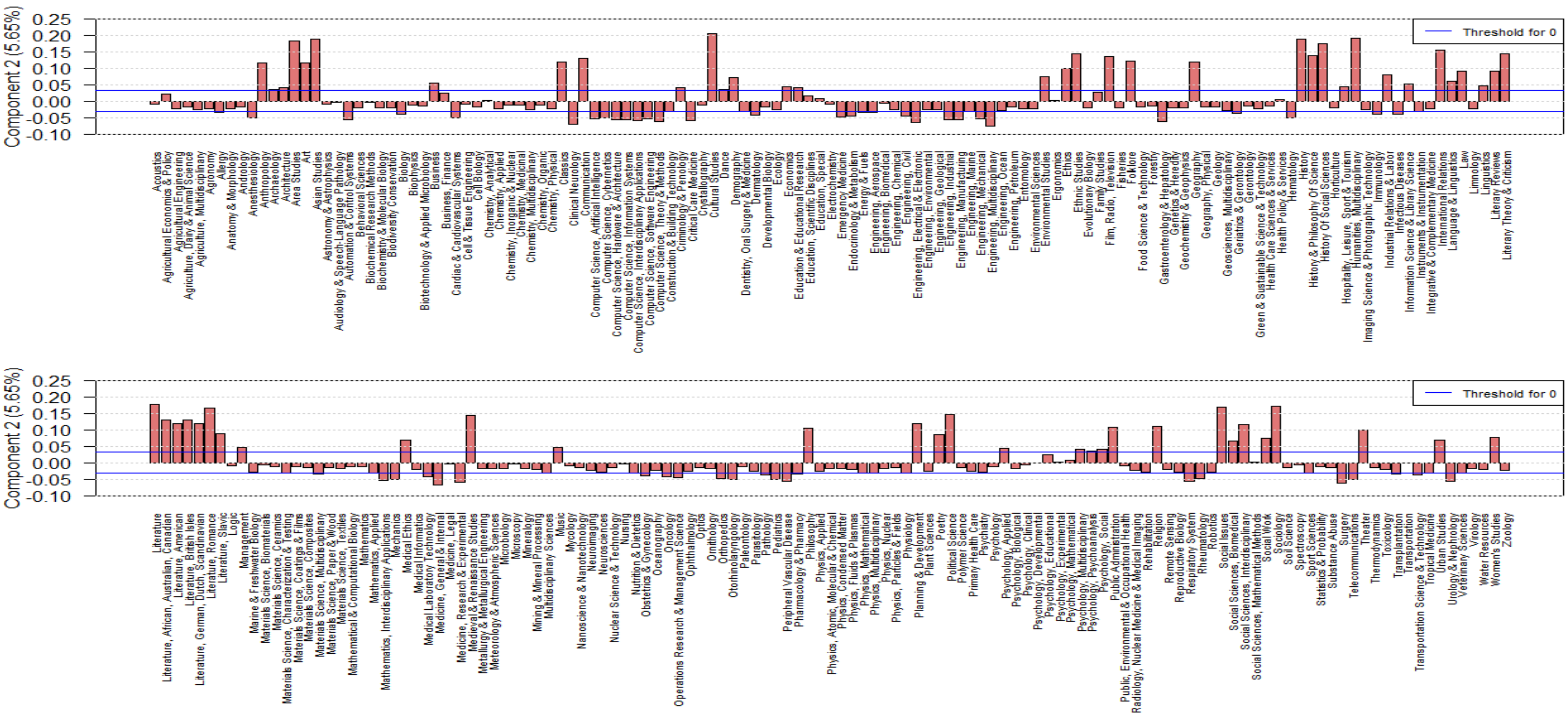


FIGURE 5.6. The second principal component of the LScT. The plot shows the contributions of original attributes (categories) on the second principal component.

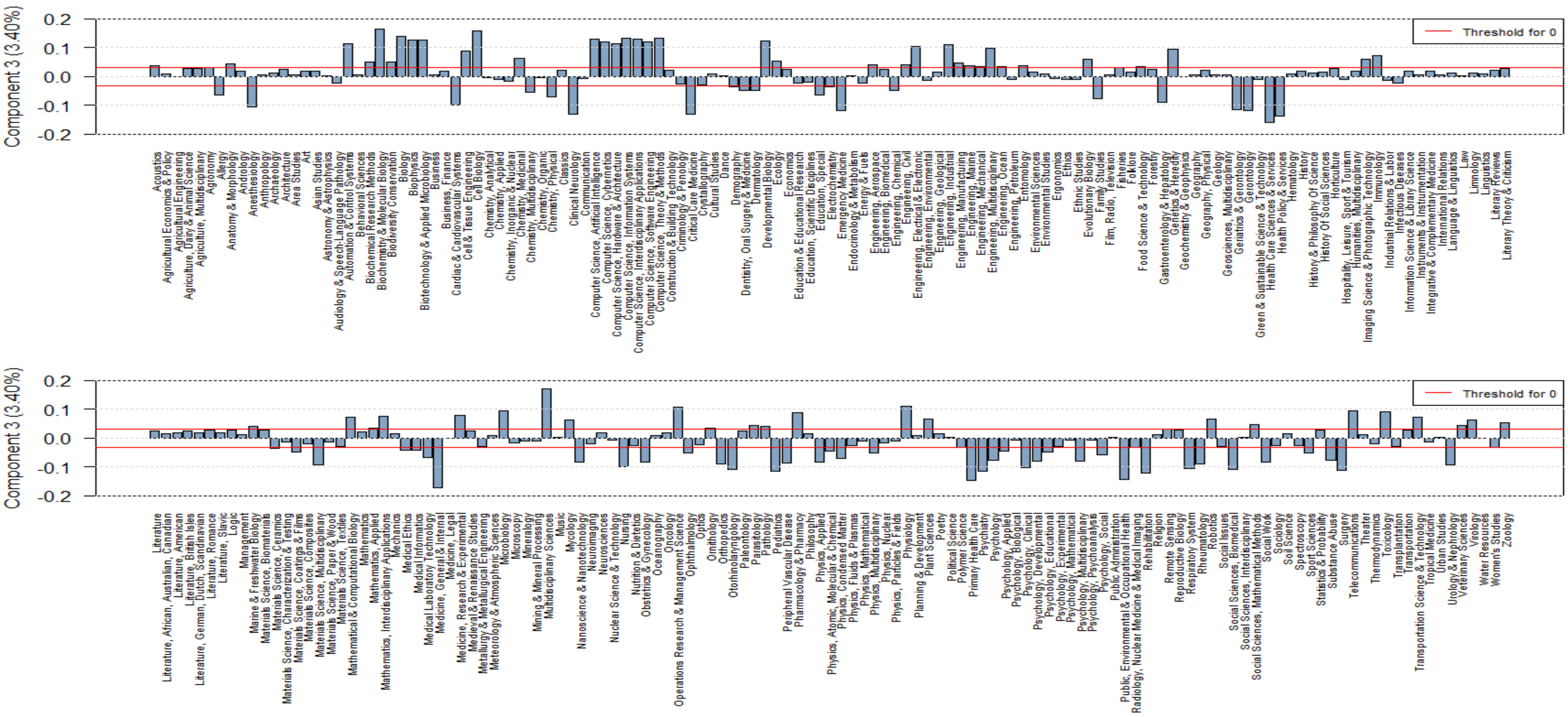


FIGURE 5.7. The third principal component of the LScT. The plot shows the contributions of original attributes (categories) on the third principal component.

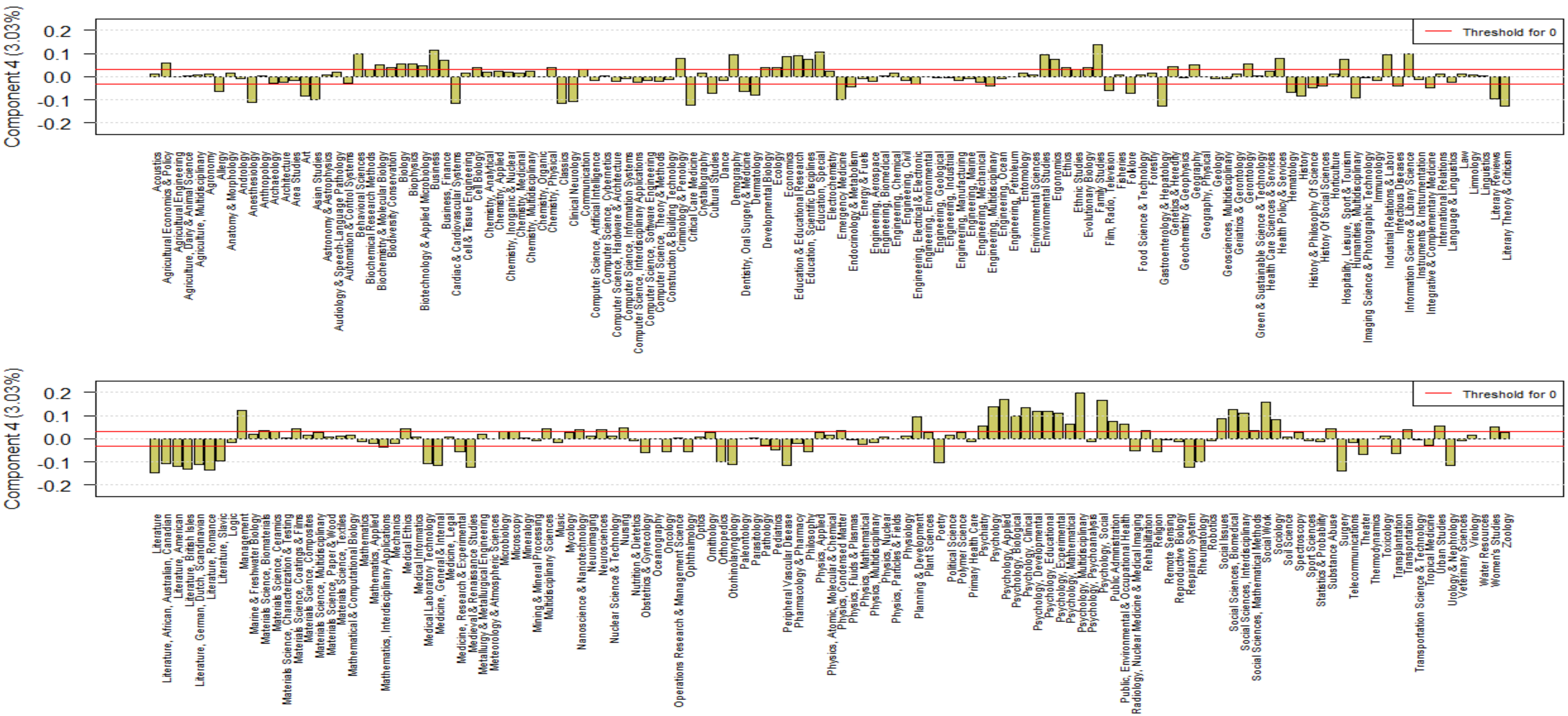


FIGURE 5.8. The fourth principal component of the LScT. The plot shows the contributions of original attributes (categories) on the fourth principal component.

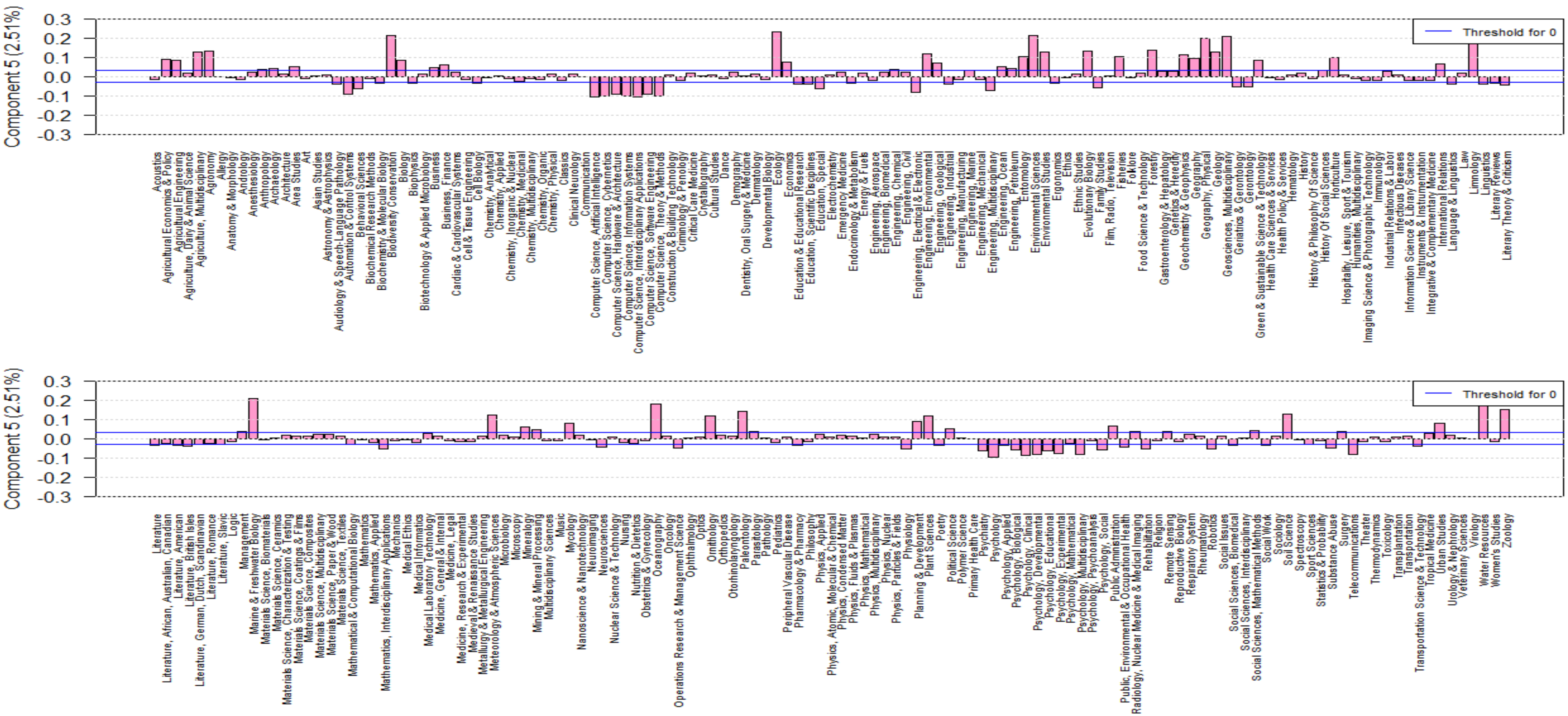


FIGURE 5.9. The fifth principal component of the LScT. The plot shows the contributions of original attributes (categories) on the fifth principal component.

We have seen that individual components are not fully explanatory of variations in the data. For instance, the first component explains only 12.58% of the variation, and it is much less for subsequent PCs. The first two components explain about 18.3% of the variation in the data. To gain insight about correlated attributes in a certain portion of the variation, it is reasonable to plot two PC axes. The Fig. 5.10 (a) shows the results for the first two components. The influence on the variance of the PC axis is low for those categories near to the intercept of zero interval.

The category 'Engineering, Multidisciplinary' has a large positive value to the extreme right on PC1. As we mentioned before, this component measures the magnitude of contribution of the original attributes to the component. For PC2, categories in positive, negative and zero groups are shown in different colors. This indicates that some attributes are correlated in approximately 18% of the variation represented by the first two PCs.

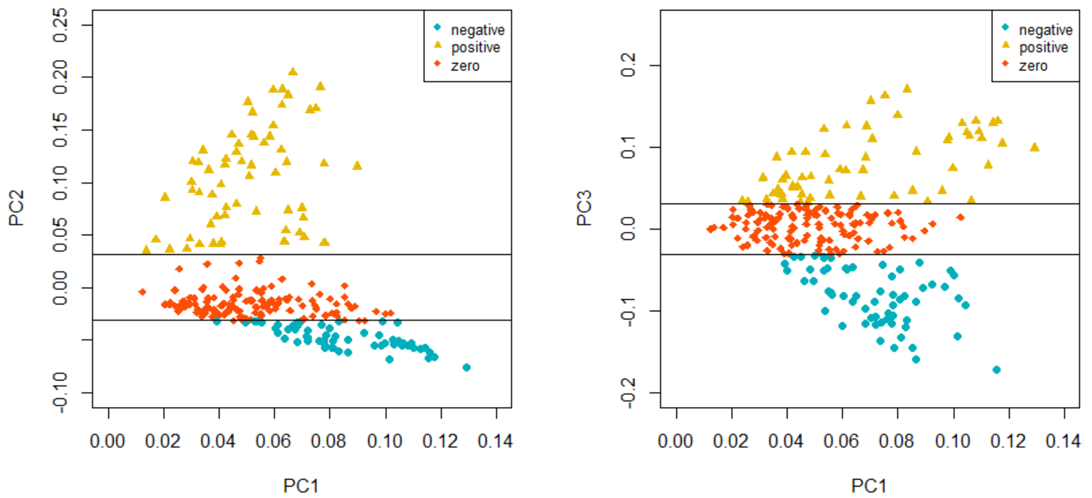
It can be seen that categories in the positive group on PC2 are not at the extreme right on the PC1 axis. Also, it is apparent that attributes in the positive area are not as well correlated (dense) as attributes in the negative and zero areas. The attribute 'Medicine, Legal' appears close to zero in both components, which indicates that these components are not indicative of variations for this attribute.

Fig. 5.10 (b) is a comparison of PC1 with PC3. We note particularly that the density of categories in the positive and negative areas are similar, having almost equal distribution for negative and positive values in the zero interval. Attributes in the zero interval seem to be much more dense denser so more correlated than attributes in the other two groups. However, these two components do not reflect the variation for the zero interval.

The second and the third principal components, which account for 9.1% of the variance in the data, are plotted in Fig. 5.11 (a). If attributes are inversely correlated, they are positioned on opposite sides of the origin in this plot. These two components seem to have correlated attributes (high density) mainly in the zero-zero area. It is noticeable in the plot that the majority of categories have negative and positive values in one of components when they are in zero interval for the other component. This implies that certain attributes are at the extremes of a spread in one PC, and not in the other. Also, there are attributes in the area of negative-positive, negative-negative and positive-negative, but not in positive-positive. This indicates that in a qualitative sense, categories are associated with PCS in very different ways. This result is consistent with the observations from the Fig. 5.11 (b), a three dimensional picture in which we can observe categories stretching out along the PC axes, like the tail and wings of a bird.

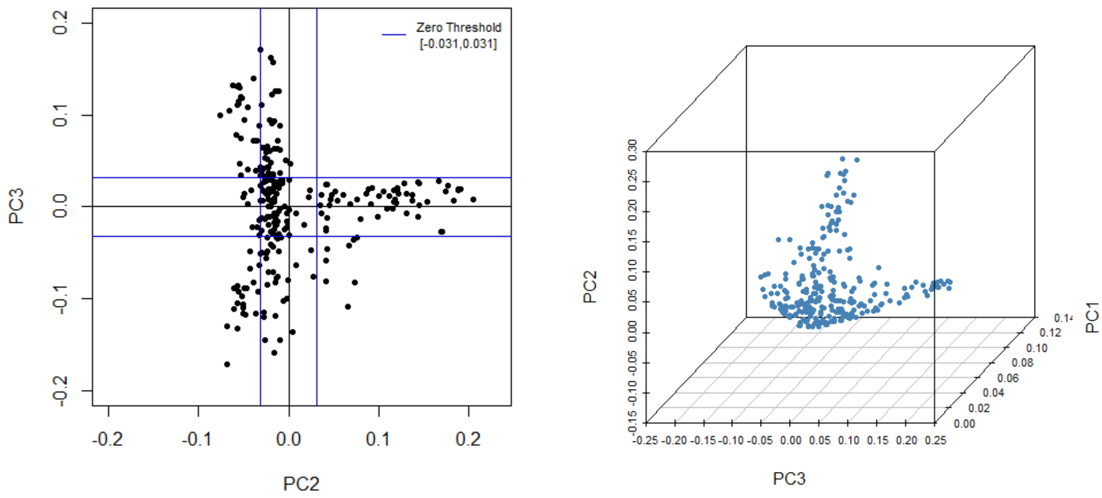
5.3.1. Analysis of Extreme Topic Groups at Opposite Ends of the PCs

As mentioned above, we applied the Double Kaiser rule to evaluate PCs and select a



(a) (b)

FIGURE 5.10. (a) The first principal component versus the second principal component of the original attributes (categories) (b) The first principal component versus the third principal component of the original attributes.



(a) (b)

FIGURE 5.11. (a) The second principal component versus the third principal component of the original attributes (categories) (b) The first three principal components of the original attributes.

subset of original attributes based on their importance for the PCs. The importance of attributes was defined by a 'threshold of importance', which was $1/\sqrt{252}$. All attributes appeared as non-trivial, so we retained all 252 categories. Then, we used the half of the value $1/\sqrt{252}$ to define a threshold for categories having near zero values under the assumption that as those categories having very low contribution to the PC do not have a strong influence on the PC, they are not correlated to the corresponding PC and so not the main coordinates of the dimension. All values of component between $-1/2\sqrt{252}$ and $1/2\sqrt{252}$ were considered into zero interval. So, three groups of categories were defined based on this threshold: (1) positive (2) negative (3) zero. We then analysed PCs based on categories in each groups.

In this section, we analyse the topic groups at opposite ends of the PCs (positive and negative ends) in order to describe the PCs based on extremely influential categories at both ends. As such categories have high contributions in the PC, they are the parts of the trends in PCs and so explain the general trends of PCs. This implies that we consider positive and negative groups introduced before, select the top n categories with the highest component coefficients in each group and describe the grouping of categories in a way that categories at extreme ends can be distinguished from each other somehow and meaningfully described by a classification of research fields in science.

A simple way to explain a PC by groups of categories at the both ends of the PC is by looking at the scientific categorisation of the categories and words used in the texts of grouped categories. In particular, as there exist two sets of categories that are located at opposite ends of the PC, it would be interesting to extract the most influential common words used in each of this extreme topic groups. It is important to note that as common words appear in all selected categories as informative words, they help to describe categories and then PCs.

We implemented a heuristic technique. This approach starts with a search for a set of 10 categories with maximum coefficients at the two ends of the PC. The most informative 150 words are extracted in each of 10 categories, and the common words are listed. Words are analysed by human inspection to understand the meaning behind the opposite ends of the PC. The procedure is repeated for the PC2, PC3, PC4 and PC5. For the first PC1, the sign of coefficients are positive for all categories. High numbers for categories in this PC indicate that that category is well-described by words in the LScT.

Both the top 10 categories and the most influential words in the set of categories are used to extract information for describing the PCs. The description is now more focussed on the big picture of science rather than description based on branches/sub-branches.

The second PC seems to correspond a separation between discourse studies and experimental studies when we consider both the categories and words. For example, it is seen that three of the most informative common words are ‘argu’, ‘polit’ and ‘discours’ for the groups of categories in the positive side and three of the most informative common words are ‘clinic’, ‘treatment’ and ‘therapi’ for the groups of categories in the negative side in the PC. This is the **Nature of Science** dimension.

The third PC reflects two opposite types of research in terms of the requirement of microscopic and macroscopic instruments (see Table 5.7). At the positive end, scientific research mostly required detailed tools to work with the objects. Such tools can be instruments such as the microscope as well as programming tools used in coding. On the negative end, we are talking about human and population scale objects, but still related to humans. So this is the **Human Scale** dimension.

TABLE 5.7. Categories at opposite ends of the PC3

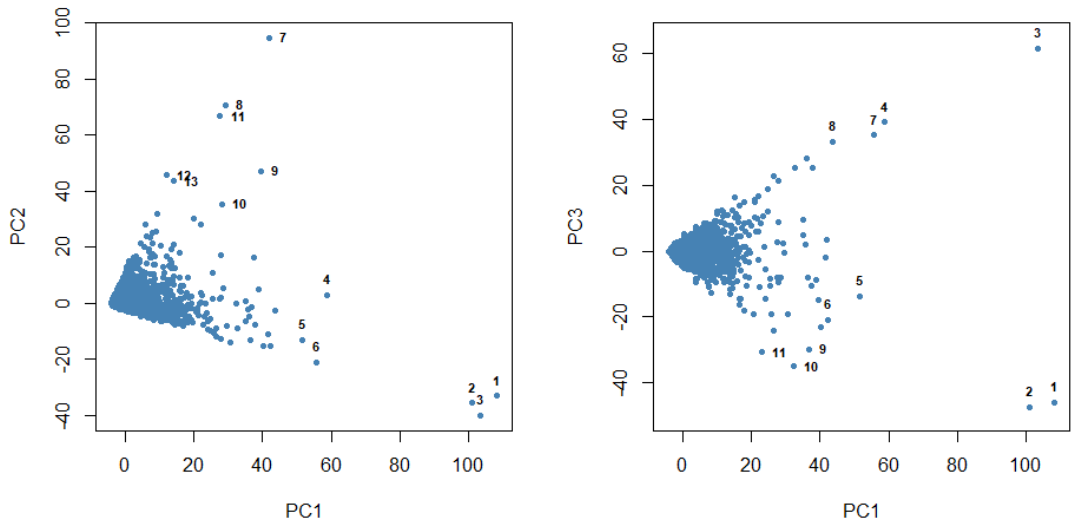
Categories in the positive side of the PC3	**Categories in the negative side of the PC3**
Multidisciplinary Sciences	Medicine, General & Internal
Biochemistry & Molecular Biology	Health Care Sciences & Services
Cell Biology	Primary Health Care
Biology	Public, Environmental & Occupational Health
Computer Science, Information Systems	Health Policy & Services
Computer Science, Theory & Methods	Critical Care Medicine
Computer Science, Interdisciplinary Applications	Clinical Neurology
Computer Science, Artificial Intelligence	Rehabilitation
Biotechnology & Applied Microbiology	Gerontology
Biophysics	Emergency Medicine

The fourth component appears as to describe two classes of science: science of understanding the human condition through experiments and science of understanding the human condition through critical discourse studies. For instance, literary studies in the negative side are prominent and many texts from the literature are literary criticism of works. This is the **Human Condition** dimension.

Finally, the fifth component can be interpreted as contrasting the natural science and the intelligence. Categories related to natural science researches are grouped in the positive extreme side and categories of understanding intelligence are located in the negative extreme side in this PC. 'Intelligence' can be both human intelligence and machine intelligence. For example, the categories 'Computer Science, Artificial Intelligence' and 'Psychology' are two of the top 10 categories. This is the **Inner World/Outer World** dimension.

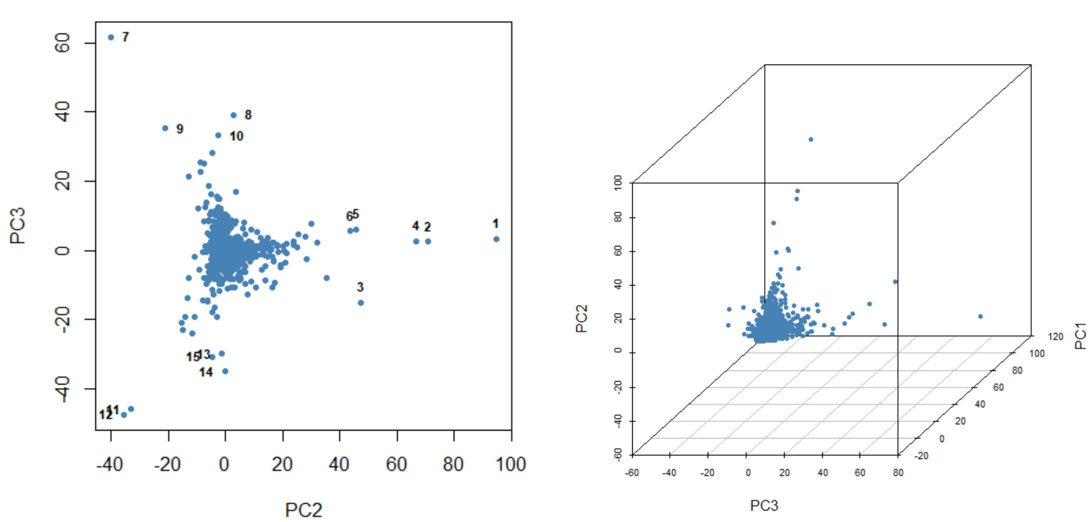
5.3.2. Visualisation of Words in the Space of PCs

We now move onto visualizing words on the principal axes. Fig. 5.12 and 5.13 are visualisations of words in the space of any of the first three components and the first three PCs. PC scores are calculated for each word to determine its location on the PCs. Those words that have negative scores when projected onto a PC will represent categories at that end of the PC. The majority of words are located around the origin in the figures, showing that relatively few words are strong indicators for any category. However, some words seem to be distinctly different from others and grouped together in the plots. This suggests that even a small variation in the data (such as 12.58% and 5.65%) helps to distinguish certain groupings of words.



(a) (b)

FIGURE 5.12. The PCA score plot of words. (a) First and second PC axes with principal component scores (b) First and third PC axes with principal component scores.



(a) (b)

FIGURE 5.13. The PCA score plot of words. (a) Second and third PC axes with principal component scores (b) The first three PC axes with principal component scores.

Words further apart on the PC map are more different from each other than words grouped together. Labelled words in are presented in the Table 5.8. We can see that the words 'patient', 'paper', 'conclus', 'articl' and are located far from other words. It is worth mentioning that these words were the most informative words in the LScT. There are other interesting features of these plots. Words grouped closer together and far from the origin have similarity with respect to the academic disciplines they appear in. For example, the two words 'argu' and 'polit' are far from the origin and close together in Fig. 5.12 (a). They both are expected to be included in texts for social science. Similarly, the words 'essay' and 'centuri' are close in the plot.

TABLE 5.8. Words labelled in Fig. 5.12 and Fig. 5.13.

Label	PC1-PC2	PC1-PC3	PC2-PC3
1	patient	patient	articl
2	conclus	paper	argu
3	paper	conclus	social
4	cell	cell	polit
5	clinic	propos	essay
6	propos	clinic	centuri
7	articl	protein	paper
8	argu	background	cell
9	social	care	propos
10	result	health	protein
11	polit	particip	patient
12	essay		conclus
13	centuri		particip
14			health
15			care

In Fig. 5.12 (b), the words 'cell' and 'protein' are near to each other. A group of words 'health', 'care' and 'particip' are also co-located, which is to be expected as they are likely to be used together in medical texts. The same behaviour can be observed in Fig. 5.13 (a). Words that are likely to be used in social science related articles go towards the positive end of the PC2 axis, and words that are likely to be used in medical texts go towards negative end. Words that are likely to be used in biological science related articles are towards the positive end of the PC3 axis. Words in these three regions are given in Table 5.9. We list words where the scores on

- PC2 are greater than 20 and less than -10;
- PC3 are greater than 20 and less than -20.

We observe a difference of topic-specific words in medicine, social science and biological science related articles. As we expected, words that are used in biological, computer and engineering science related areas are towards the positive end of PC3; see Fig. 5.7.

TABLE 5.9. Words when the scores on PC2 are greater than 20 and less than -10, words when the scores on PC3 are greater than 20 and less than -20 in Fig. 5.12 and Fig. 5.13.

PC2>20	**PC2<-10**	**PC3>20**	**PC3<-20**
article	patient	speci	patient
argu	conclus	paper	conclus
social	paper	cell	age
result	background	protein	background
polit	clinic	propos	particip
essay	propos	gene	health
cultur	age	algorithm	object
centuri	diseas	express	care
polici	associ	problem	
text	year		
public	object		
literari	outcome		
discours	problem		
histori			
contemporari			
govern			
draw			
scholar			
war			
narrat			

5.3.3. Visualisation of Categories into the Word Space

The data can be represented as a set of points in two different spaces: category space and word space. In the category space every point (vector of dimension 252) is a word, represented by its RIGs in 252 categories. In the word space every point (vector of dimension 5,000) is a category, represented by RIGs of 5,000 words. Results on the visualisations in the category space were provided in the previous section. Here we provide results on the visualisation of the data in the word space.

The analysis in this section answers the question of how successfully categories are distinguished into PC space constructed in the word space.

Categories are initially visualised in the PC axes (see Fig. 5.14). The figure demonstrates the data in the first two and three PCs and the obtained bird-like shape. Annotated version of categories in the space of the first three PCs is presented in Fig. 5.15. One can see the bird-like shape of the graphs has a meaningful separation of categories on two wings in terms of the branches of science. The left wing carries medicine related categories, while the right wing carries the literature and social science categories. At the right figure, the bird is coloured: red and green wings, and blue body. Lists of categories for each part of the bird with the projection on the first PC are presented in Table D.4. Colouring indicates the categories of social science (green), categories of medicine (red) and all other categories (blue). They all support the general findings.

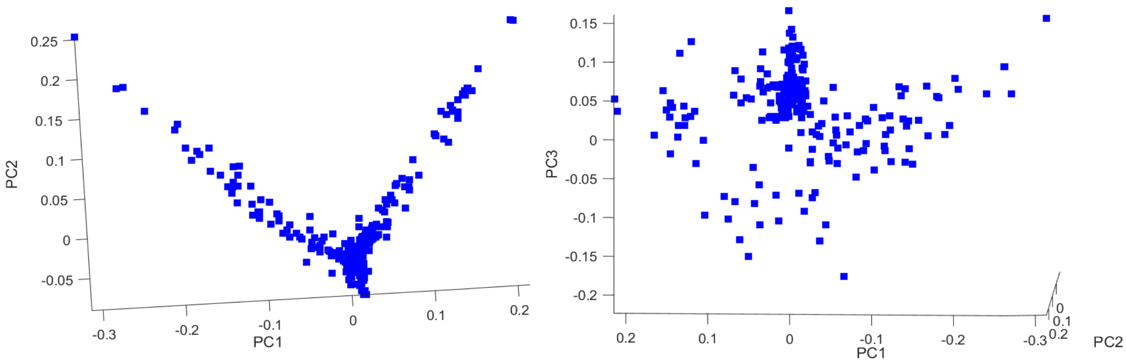


FIGURE 5.14. Visualisation of categories in the first two and three PCs

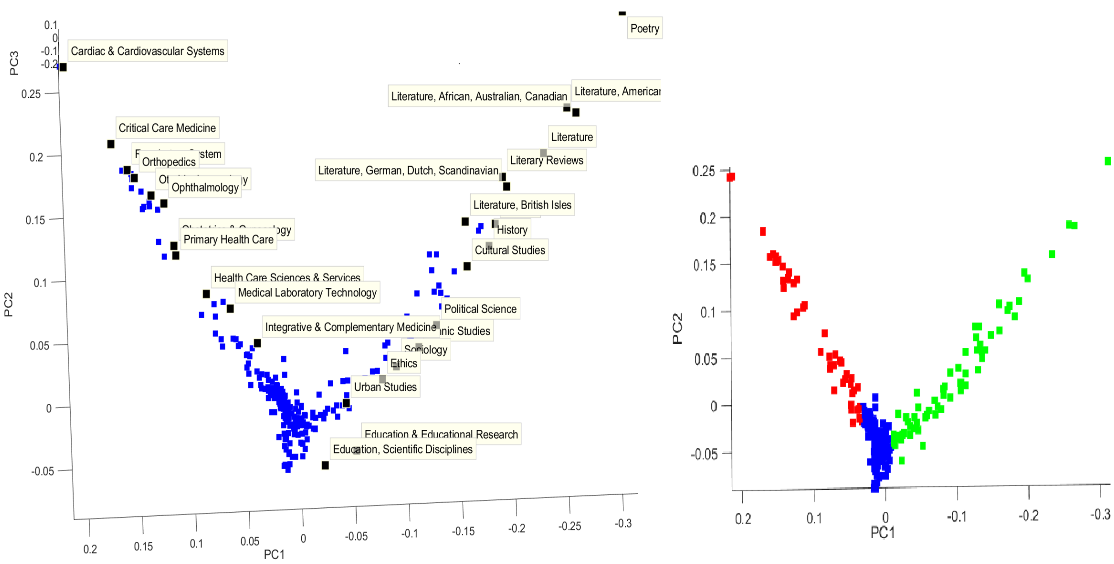


FIGURE 5.15. Visualisation of the annotate version of categories and colouring of the bird-like shape

5.3.4. Identifying Groups of Categories in PCs by Approximation of Vectors

We now introduce another approach for determining the groups of categories which are influential for PCs. We start with introducing the method for determining two such groups of categories in a specific PC.

Suppose we have a given n-dimensional vector of component coefficients in an arbitrary PC, defined as

$$\mathbf{u} = (u_1, u_2, \cdots, u_n),$$

with categories ordered so that $u_1 > u_2 > \cdots > u_n$ (the chance of two being equal is negligible). Our assumption is that two classes of categories can be separated at some point u_k, where categories are correlated to each other within each class. That is, we aim at finding the optimal k in

$$\mathbf{u} = (u_1, u_2, \cdots, u_k, \cdots, u_n),$$

where those categories with coefficients $u_1, u_2, \cdots, u_k$ will belong to the first class and those categories with coefficients $u_{k+1}, u_{k+2}, \cdots, u_n$ will belong to the second class. To find the optimal k, we use a method based on the approximation of vectors.

Let us consider the binarized vector of two classes

$$\mathbf{e} = (1, 1, \cdots, 1, 0, \cdots, 0),$$

in which elements of the first class are transformed to 1 and elements of the second class are transformed to 0. We assume that the best approximate to the given vector $\mathbf{u}$ is a vector aligned in the direction of the vector $\mathbf{e}$, that is, the approximate vector

$$\alpha\mathbf{e} = (\alpha, \alpha, \cdots, \alpha, 0, \cdots, 0)$$

for some $\alpha \in \mathbb{R}$, where the number of elements with the value α is k. The vector $\mathbf{e}$ is a basis vector in the space. The approximation of the vector implies that the Euclidean distance between vectors $\mathbf{u}$ and $\alpha\mathbf{e}$ should be minimum. Therefore, there is an α that minimizes the length of the difference between the approximation $\alpha\mathbf{e}$ and the given $\mathbf{u}$. Our aim is now to find the constant α such that minimizes

$$W_1 = \| \mathbf{u} - \alpha\mathbf{e} \|$$

is minimised.

We can simplify the algebra by minimizing the square of the norm

$$W_1^2 = \| \mathbf{u} - \alpha\mathbf{e} \|^2 .$$

Minimizing W_1^2 implies finding α such that

$$\begin{aligned}\frac{\partial W_1^2}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{k}(u_i-\alpha)^2+\sum_{i=1+k}^{n}u_i^2\right)=\frac{\partial}{\partial \alpha}\sum_{i=1}^{k}(u_i-\alpha)^2\\ &= -2\sum_{i=1}^{k}(u_i-\alpha)=0.\end{aligned}$$

Solving this for α gives

$$\alpha=\frac{1}{k}\sum_{i=1}^{k}u_i.$$

This means that α is the average of values in the first class and the approximate vector is

$$\alpha\mathbf{e}=(\underbrace{\alpha,\cdots,\alpha}_{\text{k times}},\underbrace{0,\cdots,0}_{\text{n-k times}}).$$

To determine the optimal k, we search for the best approximation to the vector of component coefficients among those created for $k=1,\cdots,n-1$. Once the optimal k is found, we will be able to classify categories into two classes.

To identify three groupings of categories in a PC, we first consider the approximate vector

$$\mathbf{v}(\alpha,\beta,k,r)=(\alpha,\cdots,\alpha,0,\cdots 0,\beta,\cdots\beta),$$

where the number of elements with the value α is k, the number of elements with the value β is r and $k+r<n$. Our aim to find the optimal k and r that minimize $W_2=\parallel \mathbf{u}-\mathbf{v}(\alpha,\beta,k,r)\parallel$. This implies

$$\begin{aligned}\frac{\partial W_2^2}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{k}(u_i-\alpha)^2+\sum_{i=1+k}^{n}-ru_i^2+\sum_{i=n-r+1}^{n}(u_i-\beta)^2\right)\\ &= \frac{\partial}{\partial \alpha}\sum_{i=1}^{k}(u_i-\alpha)^2=-2\sum_{i=1}^{k}(u_i-\alpha)=0\end{aligned} \tag{13}$$

and

$$\begin{aligned}\frac{\partial W_2^2}{\partial \beta} &= \frac{\partial}{\partial \beta}\left(\sum_{i=1}^{k}(u_i-\alpha)^2+\sum_{i=1+k}^{n}-ru_i^2+\sum_{i=n-r+1}^{n}(u_i-\beta)^2\right)\\ &= \frac{\partial}{\partial \beta}\sum_{i=n-r+1}^{n}(u_i-\beta)^2=-2\sum_{i=n-r+1}^{n}(u_i-\beta)=0.\end{aligned} \tag{14}$$

We note that finding α and β are independent of each other. Thus we have

$$\alpha=\frac{1}{k}\sum_{i=1}^{k}u_i,\ \text{and}\ \beta=\frac{1}{r}\sum_{i=n-r+1}^{n}u_i.$$

If we want to identify three groupings of categories as above, we will search for the optimal k and r independently. We apply the procedure for two classes to the vector $\mathbf{u}$ in two steps: 1) finding the optimal k in the vector $\mathbf{u}$ where elements u_i are sorted in descending order; 2) finding the optimal r in the vector $\mathbf{u}$ where elements u_i are sorted in ascending order.

5.3.4.1. *Results*

The grouping of categories into PCs is done by using the vector approximation described in the previous section. The numbers of categories in the positive, negative and zero groups are given in Table 5.10. The full lists of categories in each group for PC2, PC3, PC4 and PC5 can be found in Table D.5.

TABLE 5.10. Number of categories in the positive, zero and negative groups identified by vector approximation based approach for the first five principal components

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
Positive	221	46	41	44	37
Zero	31	91	159	160	158
Negative	0	115	52	48	57

The results using the vector approximation based grouping method turn out to show similar patterns to those using the threshold $1/2\sqrt{252}$. In general, there is no clear change of trends in PCs excluding PC2, where psychology related areas were split between the positive and zero group previously. The vector approximation method reduces their importance, placing them in the zero group. This does not conflict with our hypothesis that this PC is related to the difference between discourse and experiment in what we suggest is the Nature of Science dimension. It is certainly not unreasonable to exclude psychology from the group of discourse studies. More categories, where experimental methods of research are used, now appear in the negative category. For instance, the categories 'Developmental Biology' and 'Biochemistry & Molecular Biology' are now included in negative category.

Another interesting finding is that the positive group for PC5 contains only environmental science related areas, and excludes economics and public administration which appear in the tail of the list. Some social science categories such as 'International Relations',' Business', 'Finance' and 'Political Science' were in the positive group, but are categorised into the zero group by the vector approximation based grouping method. Again, this does not conflict with the idea that PC5 is the Inner World/Outer World dimension.

5.3.5. Deciding the Dimension of the Meaning Space

The number of principal components determined by the Kaiser rule was 61. However, the Kaiser rule can underestimate or overestimate the number of PCs to be retained [210]. So, we also tested the Broken-Stick rule to determine the number of PCs [207, 208, 209, 219]. Fig. 5.16 demonstrates the optimal number of components determined by the Broken Stick and the Kaiser rules. The Broken Stick rule suggests that the reduction to only 16 PCs is reasonable.

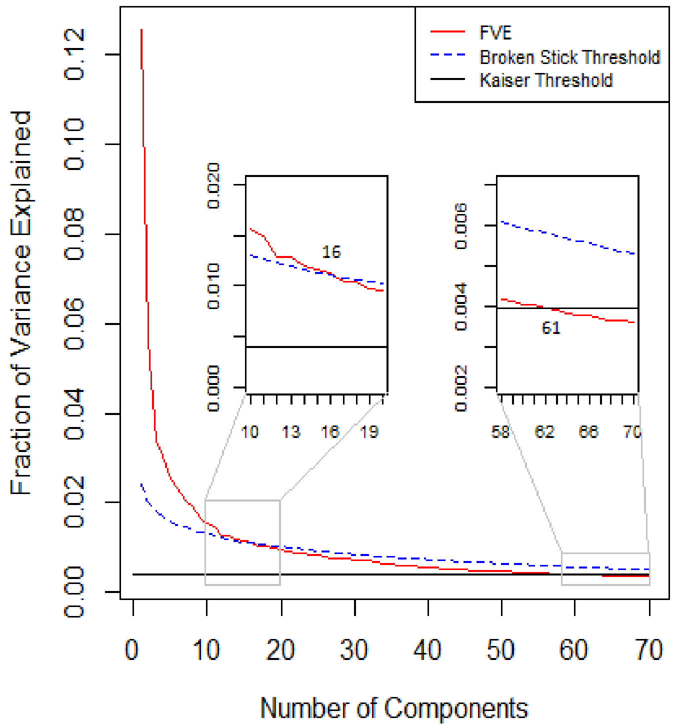


FIGURE 5.16. The number of principal components determined based on the Kaiser rule and the Broken Stick rule

Finally, we compared these two criteria of PC selection with the criterion: the ratio of the maximal and minimal retained eigenvalues ($\lambda_{max}/\lambda_{min}$) should not exceed the selected number (the condition number) [211, 212, 213]. It is described as the *multicollinearity control*. In order to avoid the effects of multicollinearity, the conditional number of the covariance matrix after deleting the minor components should not be too large. That is, k is the number of components to be retained if k is the largest number for which $\lambda_1/\lambda_k < C$, where C is the conditional number. This method is called PCA-CN [213]. In our work, modest collinearity is defined using collinearity with $C < 10$ as in [212]. Therefore, the number of PCs to be retained is 13 by PCA-CN. Table 5.11 shows the number of PCs to be retained (informative PCs) found by three approaches: Kaiser rule, Broken Stick rule and PCA-CN.

Increasing the number of dimensions could lead to even the closest neighbours being too far away in the Meaning Space, especially for a small number of words. This notion is closely related to the problem of curse of dimensionality. To avoid this isolation of word vectors in the space because of over-fitting, we have decided to use this new 13-dimensional Meaning Space in the future study of the quantification

TABLE 5.11. Number of PCs to be retained, found by Kaiser rule, Broken Stick rule and PCA-CN

Method	Number of PCs
Kaiser	61
Broken Stick	16
PCA-CN	13

of meaning of text. The full list of categories in positive and negative groups for each of the first 13 PCs can be found in Appendix D.3.

5.4. Conclusion and Discussion

In this chapter we compute the dimension of meaning; our answer is 13. We also suggest qualitative meanings for the first five of these dimensions:

- PC1 describes how well 252 categories from the WoS are described by the words in texts classified as being in these categories. The first PC coordinate of a word is its general informational value for separating of categories;
- PC2 is the Nature of Science dimension, which categorises topics as either discourse or experiment; The corresponding PC coordinate reflects the difference in the use of the word in these two groups of topics;
- PC3 is the Human Scale dimension, which distinguishes biological science (with some attached disciplines) and medicine (also with some attached disciplines);
- PC4 is the Human Condition dimension, where the distinction is between understanding the human through psychology, social and behavioral sciences, or art (with some admixture of medicine);
- PC5 is the Inner World/Outer World dimension where the human experience of itself (a combination of psychology and computers) is contrasted with the experience of the external world (environment, ecology and related topics).

We welcome fierce debate over the meaning of these dimensions, but giving a qualitative meaning to these is a crucial step to understanding the meaning of meaning.

We arrived at these conclusions by first of all arguing that the reduced word set of 5,000 words in LScT reasonably represent the texts from our corpus; see Section 5.2. Having done this it reasonable to perform a principal component analysis of the word-category matrix, whose entries are the relative information gains for the category with the given word; see Section 5.3. By exploring three different selection criteria (Double Kaiser, Broken Stick, PCA-CN, Section 5.3.5) we reduced the dimensionality of the category space to 61, 16 and 13 respectively. Meaning cannot be so complicated - so we choose the lowest dimensionality. If it turns out that we cannot explain some component of meaning at some time in the future with only 13 dimensions, we can increase the dimension. It remains a challenge to describe all 13 such dimensions in a way that makes some philosophical sense, but we hope that we have opened up this debate in this chapter.

In the future we hope to explore a new categorisation of texts using the positive, negative and zero correlation of each subject category with the influential PCs. Even if we use the first five PCs (with some ternary division on PC1) this would give $3^5 = 243$ categories (eerily close to the 252 categories in WoS).

CHAPTER 6

Semantic Analysis for Automated Evaluation of the Potential Impact of Research Articles

6.1. Introduction

A space of meaning for words was created from the analysis of situations of their use and this space was analysed in Chapter 5 [220]. We now return to texts and introduce a novel, informational text representation model.

In this chapter, we introduce a text representation model based on informational semantics of texts. In this approach, each text is a cloud of words represented in the Meaning Space [205]. In MS, words were represented by their RIGs for categories. We construct text representation using the distribution of words' RIGs in each category (dimension). The information in each text is summarised by a set of parameters such as mean point of a cloud's points (vector of mean). *Feature Vector of Text* is introduced and created for each text as a vector representation of the text. FVTs can be constructed in various ways depending on the task. We create FVTs in five different ways by combining the mean vector, the vector of the first principal component for each text and two centroid vectors obtained by k-means clustering of words in the text. Each of FVTs is defined as informational semantics representation and analysed for binary classification problem.

Given the informational representations of texts, we evaluate the scientific impact of articles through their semantics. Citation count is widely-used measure of impact in scientific world. As well as many other important factors that contribute to the citation counts such as the number of scientific collaboration and high-impact journals [221], one important factor is the context of the paper – usually the most important –. We investigate this fact under the assumption that the informational semantics of texts is an important indicator in research assessment for individual categories.

We hypothesize that research trends and the impact of articles can be portrayed using semantic relations between contexts in texts. Predictive models for citation are employed and tested for assessing scientific impact of articles based on informational representations of texts. The focus of this study lies on showing how much information the semantics of scientific articles have in predicting of citation. Here we also claim that such a linking of semantics of texts and the scientific impact in fact provides the basis for a solution of the problem of analysing and decision-making for scientific articles to be published by journals.

The empirical analysis of the research in predicting the citation is done on the basis of the LSC abstracts with citation counts extracted from the Web of Science website for approximately 4 years range from 2014 [13, 15, 16, 112, 222]. We perform the k-Nearest Neighbour (kNN) and Linear Discriminant Analysis (LDA) classifiers to evaluate the impact and test different FVTs in predicting citation. In classification, we make a very broad separation of citations: highly-cited papers and less-cited papers. Therefore, we deal with binary classification throughout this research. For experiments, three categories are selected from three main branches of science: Applied Mathematics, Biology and Management and the results are discussed for these categories.

Distinguishing between the highly-cited and less-cited papers could be done by several ways. In order to avoid the predominance of highly-cited categories, we use relative thresholds for each category rather than an absolute threshold defined from the whole corpus. Two rules for defining thresholds are used and two classification problems are solved. In the first task, we made a very broad separation between papers based on the average citation in the category. The classification of papers labelled as highly-cited (H) and less-cited (L) is performed. We also differentiate between extremely highly-cited (EH) and extremely less-cited (EL) papers labelled according to lower and upper quartiles of citation counts in the category.

For both classification problems, we applied classifiers for five FVTs in three different spaces and compared performance of the classification. The best classification result is achieved when combining the vectors of the mean and the first principal component for the category Management by LDA, with sensitivity and specificity being greater than 80%. The semantics represented by this vector allowed the identification of extremely highly-cited (EH class) papers with 83.22% accuracy (sensitivity) and the identification of extremely less-cited papers (EL class) with 81.81% accuracy (specificity).

In this study we found that the informational semantics of a paper has important information about the scientific impact of the paper. This fact is sometimes much more clear for some scientific categories than others. Results indicate that development of a quantitative evaluation and predictive model of citation count is possible by the proposed informational semantics based approach.

The rest of the chapter is organised as follows. In Section 6.2, we review standard and widely-used text representation models in the literature. In section a novel text representation method is introduced and the vectors of text for LSC abstracts are constructed. The fundamentals of evaluation of the potential impact of research articles with classification and evaluation methods are discussed in Section 6.4. In Section 6.5, the characteristics of citations in categories for the corpus LSC are analysed. Given constructed FVTs for the LSC, the methods and applications in classification of citation are described and results of classification experiments are discussed in Section 6.6. Finally, the Section 6.7 concludes the chapter and further possible analysis is suggested.

6.2. Classical Text Representation Models

This section reviews commonly used text representation models. The underlying goal of these models is to represent texts in a way that best preserves the original text set in its space. Most of data mining algorithms work based on features extracted from the text, and different representation methods create different set of features. The first thing to be addressed is how to represent a text as an input for data mining algorithms.

One of the simplest and the most commonly used text representation models is BoW. In this model, unique words in a text are extracted and the text is represented by a bunch of words, and syntax, semantics, orders and positions of words are ignored [44]. One of the standard ways in representing texts with words is to represent texts as word-based vectors, VSM, originated by [120, 223]. All words in the corpus define the text space where each dimension is a word. In VSM, given the bag of words extracted from the corpus, the text is represented as a vector, where each dimension corresponds to a word and each entry is the weight of the word in the text. Words are weighted to indicate how important this word is to the text in the corpus.

In the VSM, the most basic weighting scheme is to represent a text as a Boolean vector; the weight is 1 when the word w is present in the text and 0 if it is absent. Using the word count (frequency) is another common type of word weighting. This shows how many times the word occurs in the text d, and is the called *Term Frequency* (TF) representation $\mathrm{TF}(w, d)$. The more a word appears in the text, the more it is considered to be relevant for the text. Each text in the corpus becomes a vector of TFs. In addition to TF, one may take into account the distribution of the word over the corpus. A word that appears in every text is not considered to be useful for distinguishing between texts. This is measured by *Inverse Document Frequency (IDF)*. The IDF is a global measure indicating the importance of the word within the entire corpus. The IDF score for a word w is defined as:

$$IDF(w) = 1 + log\left(\frac{Number\ of\ documents\ in\ the\ corpus}{Number\ of\ documents\ containing\ w}\right) \tag{15}$$

The less frequent a word is the corpus, the higher the IDF is. Term Frequency-Inverse Document Frequency (TFIDF) is a common refinement of TF and IDF. It reflects the importance of a word to the text by its frequency distribution over the corpus. The TFIDF score for a word w in the text d is calculated as follows:

$$TFIDF(w, d) = TF(w, d) \times IDF(w) \tag{16}$$

The importance increases proportionally to TF but is offset by IDF. Words that are common in every text, such as stop words (and, the, an, it for instance) have low TFIDF even though they appear many times in texts. For each text, the TFIDF is calculated for individual words in the corpus. Then, each text is represented as a vector of TFIDFs. Vectors are usually normalised to unit length. The calculation of TFIDF can be done in several different ways, some are described in [44, 224, 225].

In BoW model, every word in the text is an independent potential keyword of the text, and weights are assigned based on the frequency of words in the text and their rarity across the corpus. Words in texts are assumed that independent from each other. The weights of words can be binary, TF or TFIDF, but one of these weighting schemes is not necessarily optimal. They should be experimented to see the best results for a given problem.

As well as single words, features of vectors can be a string, phrases or any concepts characterising the text. In the phrase-based models, a number of phrases are identified and phrases are treated as individual features for texts. The phrases can be formed by several relations between words. Two of the most popular ways to form phrases are: co-occurrence information and linguistic information of words.

With the co-occurrence, two words that occur together are identified by statistical measures and this is used to form the phrases. The lexical co-occurrence of words is also used to construct semantic spaces [156, 226]. In co-occurrence approaches, the construction process is automated and no human judgement of the meaning of words is necessary. This addresses problems with the traditional semantic space approach established by Osgood. The first problem was that human intervention is needed to determine a set of axes that sufficiently represent the meaning of texts. The second problem was a practical problem of gathering information from human judges to determine where words are placed along these axes. This means that human input of size the number of axes multiplied by the number of words is needed. This is a huge problem with large numbers of words.

When using the linguistic information of words, we aim to capture precise syntactic word relations and phrases of two or more words can be formed [227, 228, 229]. Such phrases can be noun phrases or clusters of words such as adjective-noun or adverb-noun. Some alternative models including combination of statistical and syntactic phrases have been proposed in the VSM [230]. Salton analysed syntactic constructions in texts and assigned importance weights to the term phrases identified in order to choose phrases to use in the index for his books[231]. However, both approaches to forming phrases result in long vectors representing the meaning of texts.

In [230], it is suggested that more semantic information can be gathered from phrases as they give an idea about the contexts of the texts. However, Lewis [232] argued that single word representation carries a better statistical quality as frequencies of words in a phrase could be misleading. Even if a word appears several times, the phrase containing the word is likely to appear only once.

6.2.1. Limitations of Traditional Text Representation Schemes for Big Corpora

One inherent limitation of traditional BoW text representation is the sparsity of high dimensional vectors. VSM representation can result in hundreds of thousands of dimension [233, 234]. We usually have a huge number of features (words) with many entries of the vector for a given text being zero. For instance, even if an abstract has 200 unique words, which is approximately the average length of an abstract, it is only a small fraction of the vocabulary from the dictionary for a large text collection as a whole. In a low dimensional space, two points (texts) can seem very close to each other but are far apart separate in high dimension. Increasing the dimensionally degrades the performance of the text processing applications because of the curse of dimensionality [235].

It is hard to apply NLP tasks like text categorization and clustering to sparse vectors due to computational complexity and space limitation. In such a high-dimensional space, distance metrics are usually ineffective as not all features have equal importance for the text data. An appropriate number of dimensions, in other words relevant features, are required to be chosen to gain meaningful results. Global features may result in loss of information; therefore, feature relevance is needed to define locally. To extract local relevance of features, further operations are required for construction of the vector space.

Another consequence of sparsity for BoW is that IDF gives inconsistent results for rare words within a large word set [236]. For high dimensional spaces, the information from TFIDF is not enough to obtain accurate weighting of words. This leads a need to propose and implement a new method of representation of words and texts, the outcome being an improvement in the performance of text mining algorithms.

6.3. An Informational Text Representation Method

This section presents an overview and a formalization of a novel text representation method for the study of extracting 'meaning of text' in scientific corpus. The aim is to introduce the method of text representation, and its applicability to text categorisation and other NLP tasks.

Consider a text space consisting of all texts in a corpus, each contains words that are represented according to their importance extracted from their usage in subject categories. Recall that a word in the Meaning Space is represented as a vector of RIG from word to category, where dimensions are subject categories. A text is a collection of words that are points in the MS. Each text is actually a cloud of these points and each cloud has a distribution of different words in the MS. It then follows that a text is a collection of vectors corresponding to words in the text.

For a text, we expect that RIGs of words in any two categories (two dimensions) are not the same and not equally distributed. The distribution of words' RIGs in each category is unique and characteristic for each text. In a more precise sense, if words of two texts have different degrees of importance for a particular category, distributions of their RIGs in the category will be statistically different. We hypothesize that the central tendency and the statistical dispersion of any two clouds of such words will not be the same in the category given. In other words, for different texts the means and the variances of the distributions of RIGs in the category are statistically different. We expect that two texts with similar words are represented by points that are closely located in the MS, which also means two texts having similar 'meaning'. This formally establishes a new model to represent texts' meanings. For each text, we set vectors summarising the distributions and the information in clouds obtained from RIGs in each category.

Now, each text has a distribution of words' RIGs in each dimension of the MS, defined by sets of parameters such as the mean vector and the vector of standard deviations. To represent a text, we introduce the FVT summarising the information that we maintain about a text by features included. FVT is defined as

$$FVT_i = (\overrightarrow{\mu}, \overrightarrow{\sigma}, \overrightarrow{PC_1}, ..., \overrightarrow{PC_n}, \overrightarrow{c_1}, \overrightarrow{c_2}), \tag{17}$$

where i is the index of the text in the corpus, $\overrightarrow{\mu}$ is the vector of mean values of a cloud's points, $\overrightarrow{\sigma}$ is the vector of standard deviations of a cloud's points, $\overrightarrow{PC_n}$ is the n^{th} principal component, and $\overrightarrow{c_1}$ and $\overrightarrow{c_2}$ are centroids of two clusters (centroid vectors) obtained by k-means clustering of words in the text [**237**]. It is obvious that the length of each vector in the FVT is equal to the number of categories.

We note that the centroids $\overrightarrow{c_1}$ and $\overrightarrow{c_2}$ in 2-means clustering are initialised as the most distant two points in the cloud under the assumption that two distant words should belong to different clusters (Algorithm 2). It is important that we do not make a random initialisation, as this would lead to different representations each time the algorithm is run.

Depending on the task, the FVT can be constructed with a reduced set of features, for instance:

$$\begin{gathered} FVT_i = (\overrightarrow{\mu}) \\ FVT_i = (\overrightarrow{\mu}, \overrightarrow{PC_1}) \\ FVT_i = (\overrightarrow{\mu}, \overrightarrow{c_1}, \overrightarrow{c_2}). \end{gathered} \tag{18}$$

Algorithm 2 Extracting the centroids of two clusters of words for each text by k-means clustering

Input: Text and clouds of words for each text represented in the Meaning Space; the number of clusters k=2
Output: Centroids of two clusters of words for each text
1: Initialize centroids of two clusters as two distant words in the clouds of words
2: Calculate the distance between data points and centroids
3: Form two clusters with data points assigned to nearest centroids
4: Re-calculate centroids based on the current partition
5: Assign each data point to the nearest cluster
6: **repeat**
7: Step 2 to Step 5.
8: **until** there is no change between two iterations, that is, the algorithm is converged.
9: **return** Two clusters and their centroid vectors.

Since each of the vectors in the FVT is of length n, the number of categories in the corpus, if we use all n principle components, plus the four extra categories, we use $n(n+4)$ pieces of information to represent the documents in the corpus. Of course, by choosing fewer vectors in FVT we can significantly reduce this number.

Given the FVT vectors for two texts, it is also possible to compute the similarity between them in a lower dimensional space, which is one of the main problems in typical text mining and NLP tasks, including text categorization, text clustering, concept/entity extraction, sentiment analysis, entity relation modelling and many more.

6.3.1. Representation of LSC Abstracts as Feature Vector of Text (FVT)

In this subsection, we represent the texts of the LSC in the Meaning Space. For a given set of texts $d_i, i = 1, 2, ..., M$, where M is the number of texts (1,673,350) in the LSC and each d_i stands for a text. The problem of text representation is to represent each d_i as a FVT obtained from its words that are points in MS, where dimensions are 252 Web of Science categories.

Words can be represented by two ways in the Meaning Space: their RIGs and the reduced dimension by PCA. In the second case, the dimension is determined as 61 by Kaiser rule, 16 by Broken Stick or 13 by PCA-CN [220]. Therefore, texts will have distribution of words represented in 13, 16 or 61 dimensional space. Besides the original dimension, we will also use this extracted features.

With the words represented in any of these vector spaces, we construct the FVT for each text as described below:

$$\begin{aligned} FVT_{1i} &= (\overrightarrow{\mu}) \\ FVT_{2i} &= (\overrightarrow{PC_1}) \\ FVT_{3i} &= (\overrightarrow{\mu}, \overrightarrow{PC_1}) \\ FVT_{4i} &= (\overrightarrow{c_1}, \overrightarrow{c_2}) \\ FVT_{5i} &= (\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1}). \end{aligned} \tag{19}$$

If words are represented by RIGs vectors in the MS, the dimensions of FVT are 252, 504 or 756. If the space of words is 13-Dimensional reduced basis, the dimensions of FVT can be 13, 26 or 39.

6.4. Evaluation of the Potential Impact of Research Articles through Semantics

The corpus of scientific texts is a dynamic resource is which has changed and improved as new topics in science are introduced. A rapid evolution of research and new priorities among the world's scientists requires the continuous analysis of the corpora of research literature in order to understand the impact of research articles for understanding the development of science. Citation count is a widely used indicator for how much interest an article or topic has attracted interest in the scientific world. The context of an article is often the most important factor in deciding the citation count. It is, indeed, not easy to track the context of texts (or meaning of text) and so quickly monitor and generate trends in a growing volume of publications.

With the construction of the triad - dictionary, texts and representation of texts in the Meaning Space as described in the previous section, we now evaluate the scientific impact of articles via their informational semantics. Predictive models for citation based on this representation will be assessed for articles published in WoS database in 2014 with citation counts in the subsequent (approximately) 4 years. Our focus is to explore how efficiently the semantics of scientific articles can be used to predict the impact of articles as measured by citation.

In this context, we employ two classification algorithms to analyse the predictability of citation: Linear Discriminant Analysis and k-nearest neighbours. In classification of impact, various criteria for distinguishing classes of papers can be used to define citation ranks. We consider this problem as a very broad separation of papers: highly-cited papers and less-cited papers, that is, binary classification of impacts.

An important element in selection of groups of papers is the definition of citation counts for a highly-cited paper. This is controversial in research assessments. Ordinary citation counts for certain fields can be high citation in other fields. Utilising an absolute threshold to distinguish the classes of papers in a corpus might lead to predominance of papers in highly-cited scientific categories [221]. To avoid this disciplinary differences in average citation within categories, one way is to use relative thresholds for each category. In this research, category-based relative thresholds will be defined.

The focus on computational methods for text representation and the impact analysis for such a comprehensive corpus will also provide an important critical trends analysis for communities that dealing with different aspects of corpus studies. Indeed, this analysis of two-classes problem can also be extended to multi-classes of citation counts and prediction of exact citation counts further.

We begin this research by a review of classification tasks and the descriptive analysis of the citations for the corpus LSC.

6.4.1. Methods of Classification

This subsection provides the methodology for classification algorithms used in the evaluation of scientific impacts of articles.

6.4.1.1. *Criterion for evaluating the performance of classification methods*

Data imbalance is encountered in many classification applications. It refers to the unevenly distribution of sample size in the data classes. Particularly, citation distributions are usually unequal to the classes of highly-cited and less-cited papers. The majority of papers are less-cited whereas some papers are extremely highly-cited in certain scientific disciplines such as medicine [221, 238]. This leads to classifiers to be inherently biased toward the class with larger size. Therefore, maximising the measure of performance such as accuracy might be misleading for imbalanced data. In such case, we need to use adequate metrics that take the class distribution into account.

In this work, we use standard metrics to measure the performance of classifiers: sensitivity and specificity [239]. 2×2 confusion matrix showing correct predictions and types of incorrect predictions for binary classification is used to calculate sensitivity and specificity (Table 6.1). The confusion matrix presents the decision made by the classifier in four categories: True positives (TP) referring to correctly labelled as positive, False positives (FP) referring to negative samples incorrectly labelled as positive, True negatives (TN) referring to negatives correctly labelled as negative and False Negatives (FN) referring to positive examples incorrectly labelled as negative [240] .

Sensitivity (True Positive Rate) refers to the probability that a highly-cited paper is classified as highly-cited, calculated as:

TABLE 6.1. Confusion Matrix used to calculate sensitivity and specificity for binary classification

	True Condition		
	Positive	**Negative**	**Total**
Positive	True Positive (TP)	False Positive (FP)	TP+FP
Negative	False Negative (FN)	True Negative (TN)	FN+TN
Total	TP+FN	FP+TN	

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity (True Negative Rate) refers to the probability that a less-cited paper is classified as less-cited, calculated as:

$$Specificity = \frac{TN}{TN + FP}$$

In other words, the sensitivity is the proportion of correctly identified positives and the specificity is the proportion of correctly identified negatives. To evaluate the performance of classifiers, we use the criterion: maximising the sum of sensitivity and specificity. The higher the sum means the better performance of the classifier.

The ROC (receiver operating characteristic) curves are also used to present results for binary classification problems [241]. We plot the ROC curve showing the sensitivity and the corresponding specificity points for every possible cut-off for the set in classification. The x-axis shows the 1-specificity (false positive rate) and the y-axis shows sensitivity (true positive rate) for a chosen cut-off. The diagonal line means that the model predicts at chance.

The area under curve (AUC) is used as a criterion to measure the performance of a binary classifier: how good is the classifier in identifying the correct classes of articles. For a diagonal line, the AUC is 0.5. The higher the AUC, the better the model is at distinguishing between classes. The maximum value of AUC is 1 which indicates a perfect classifier.

6.4.1.2. *k-Nearest Neighbour (kNN)*

kNN is a simple and easy-to-implement classification method. For a test sample to be classified, the set of the closest k-nearest neighbours are found from all the training data according to a distance metric (proximity). Every data point in training set belongs to one class and class labels of this set are known. For the given test point where the class label is unknown, kNN searches for the closest k nearest neighbour based on the distance metric and decides the classification by the majority voting among data points in the neighbourhood [242]. We use Euclidean distance as proximity measure.

The value of k should be defined before applying kNN and the efficacy of the classification is usually dependent on this value chosen. One of the simplest ways to determine the value of k is to run the algorithm several times with different k values and chose the optimal value of k [243, 244]. The optimal value refers to k with the best classification performance.

kNN suffers from imbalance of the data among classes as for many other classification tasks [245]. In the kNN rule, the majority is very likely to have more samples in the set of k-nearest neighbours for a test sample, and so the test sample tends to be classified as the majority class. This results in the higher accuracy for the majority class and the low accuracy for the minority class. An intrinsic method which mitigates against the effect of sample size in classes is to assign weights (inversely proportional to class frequency) to the neighbours.

To improve the classification performance by kNN, we implement weighting described as follows. For a given test point, we form the set of k-nearest neighbours from training texts where the class A has N_A samples and the class B has N_B samples out of N papers. Suppose we have k=t and we obtain s papers from the class A and t-s papers from the class B in the decision set by kNN. By class distribution, each data point in class A has weight $1/P_A$ where $P_A = N_A/N$ and each data point in class B has weight $1/P_B$ where $P_B = N_B/N$. If there are s cases from class A and t-s cases from class B in the decision set, we give class A a score of s/P_A and class B a score of $(t-s)/P_A$. We assign the test point to the class that has a higher score.

6.4.1.3. *Linear Discriminant Analysis (LDA)*

In this research, Fisher's linear discriminant is used for binary classification problem for distinguishing between highly-cited papers and less-cited papers [246]. The means μ_1 and μ_2 of each class for the training set are computed. Then, covariance matrices Σ_1 and Σ_2 for each class are calculated. The direction of the discriminating line is defined as:

$$\omega = (\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2).$$

Projections of each point x_i on the discrimination direction are found by dot product (ω, x_i) [247]. As defined before, the optimal threshold for the LDA is described as the sum of specificity and sensitivity maximisation. This means that the decision criterion of a point being in one of the class becomes a threshold on (ω, x_i) that maximises the sum of sensitivity and specificity. We first apply the LDA to the whole set with two classes and calculate the sensitivity and specificity based on the threshold found.

In order to test the statistical power and the quality of the classifiers, we also test LDA with Leave-One-Out Cross Validation (LOOCV) [248]. LOOCV is performed as follows: (1) Select a single-item as a test point (2) Use all other points as a training set (3) Apply the learning algorithm to training set once for each test point. The performance of the classifier with LOOCV is also evaluated by the sum of sensitivity and specificity.

6.4.1.4. *Supervised PCA*

For a classification problem, supervised PCA searches for a low dimensional linear manifold where the distances between projection points of different sets are maximum and the distances between projection points of the same set are minimum [249, 250]. Given set of points $X = (x^1, ..., x^n), x^i = (x^i_1, ..., x^i_m), x_k = (x^1_k, ..., x^n_k)$ where each row x^i in the matrix X corresponds to one text. Points are centred where $\sum_{i=1}^{n} x^i_k = 0$ for all $k = 1, ..., m$. $L1$ and L_2 are two sets of points where those with label 1 belong to L_1 and those with label label 2 belong to L_2.

We consider the column vectors $V = (v^1, ..., v^p)$ with $(v^i, v^j) = \delta_{ij}$ where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$ (Kronecker delta). Projection of points onto p-dimensional subspace is defined as $P_v(x) = xV$. The squared distance between projections of two points x^i and x^j are calculated as

$$\parallel P_v(x_i) - P_v(x_j) \parallel^2 = \parallel x^i V - x^j V \parallel^2 .$$

For two sets of points with different labels, the averaged squared distance between projections is

$$D_B = \frac{1}{|L_1 L_2|} \sum_{i \in L_1} \sum_{j \in L_2} \parallel P_v(x_i) - P_v(x_j) \parallel^2.$$

The averaged squared distance within a set of points with a class label is computed by

$$D_{w_k} = \frac{2}{|L_k||(L_k - 1)|} \sum_{i,j \in L_k, i>j} \parallel P_v(x_i) - P_v(x_j) \parallel^2.$$

Therefore, the function of interest to be maximised is

$$D_C = D_B - \frac{\alpha}{2}(D_{w_1} + D_{w_2})$$

where α is a parameter that indicates the relative importance of D_B and $D_{w_1} + D_{w_2}$.

6.5. Descriptive Statistics of Citations in the LSC

In this section, descriptive statistics are used to describe the corpus of interest with respect to citation counts in the Web of Science database. They summarise various aspects about the times cited count in the data, providing information about the corpus and details about samples used in further studies.

In WoS, two types of citation are accessible: *Times Cited in WoS Core Collection* and *Total Times Cited*. In the first case, the count shows the total number of times a paper was cited by other papers in WoS Core Collection [**251**]. The total times cited displays the number of times the paper was cited by other items in the WoS products including WoS Core Collection. The counts of total times cited are used in this study. Recall that the dataset consists of 1,673,350 abstracts or proceeding papers that were published in Web of Science database in 2014 with citation counts in approximately 4 years following publication (2014-2018).

Frequency statistics is the first descriptive statistic used to determine the distribution of citations in the corpus. Raw counts of documents for each citation count (times cited) is presented in Fig. 6.1. It should be stressed that as categories are not exclusive, a citation record of a document can appear in multiple categories. However, documents across categories are counted only once in Fig. 6.1 and in all other calculations of descriptive statistics for the corpus. Note that few documents are cited more than 1000: only 64 documents. Another 243 documents are cited from 500 to 1000 times. However, 929,361 documents are cited no more than 5. In the Table 6.2, one sees the highest 5 and the lowest 5 cited counts, supporting the fact that more than 50% of documents cited less than 4 times. The average of citation counts is approximately 9.

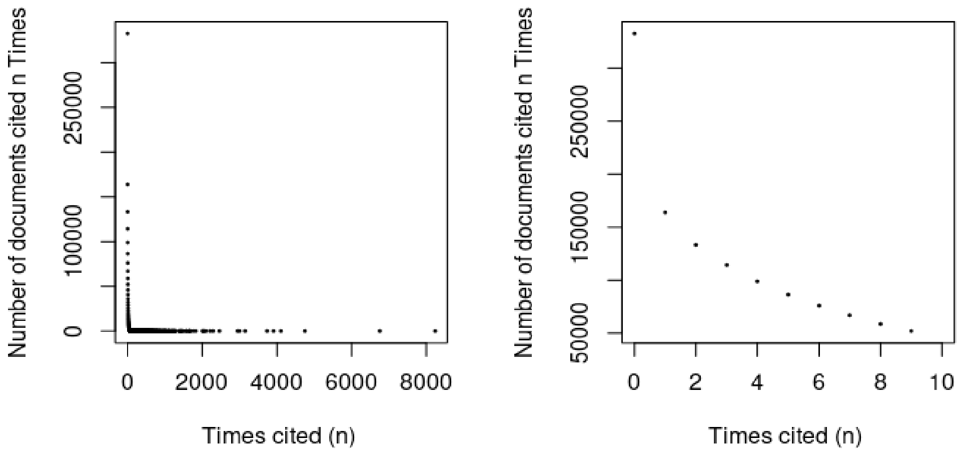


FIGURE 6.1. Graph of the number of documents cited n times in the corpus LSC. The figure on the right hand side shows the number of documents cited up to 10. The average citation in the corpus is approximately 9 (9.38).

It should be mentioned that citation counts may vary differently across scientific categories and so descriptive statistics of citations in individual categories may be characteristic for different branches of science. Also, the number of papers per category may be a factor that correlates with citations. This fact also was suggested

TABLE 6.2. The highest and the lowest five citations in the corpus with the number of documents for the corresponding citations

The highest citations	**Number of Documents**	**The lowest citations**	**Number of Documents**
8,234	1	0	332,610
6,756	1	1	163,907
4,744	1	2	133,280
4,102	1	3	114,309
3,908	1	4	98,843

and analysed in [**252**]. In the study, the most cited 500 papers of each category (236 WoS categories) with citation counts from 2010 to 2014 are compiled. It was stated that the citation counts are the highest in multidisciplinary sciences, general internal medicine, and biochemistry and the lowest in literature, poetry and dance. It was also discussed that the number of papers assigned to a category correlates with the citation counts for the selected top papers in a category. Therefore, a detailed analysis in each category will give a better insight into understanding the factors contributing to the citation and help in the selection of sample as to be representative of the corpus.

A sample of the corpus should be selected in such a way as to be representative of the corpus. An example of corpus sampling is presented as follows. Before sampling, we consider the corpus divided into groups of main branches of the science - by social science (e.g. sociology, law and psychology), physical science (e.g. physics and chemistry), life science (e.g. biology, medicine and physiology), earth science (e.g. geoscience, astronomy) and formal science (e.g. mathematics, computer science and logic). Then the corpus is sampled within each branch (or stratum). A sample of categories from each stratum is collected by ensuring the presence of key categories within the sample. We pay attention that the sample will reflect the characteristic of the corpus, assuring that the all branches will be included and evenly represented in the sample.

Taking into account the considerations above, we can select a sample with the categories presented in Table 6.3. One can see that the sample size of each stratum is similar, containing approximately 30,000 documents. The total number of documents in the sample is 147,901.The counts of documents for each citation number (times cited) for the sample is presented in Fig. 6.2. The figures shows similar trend as for the corpus with a very close value of average citation: 9.43. Detailed explanation and statistics of each stratum and its sub-branches are provided later in this section.

For both the corpus and the sample, two classes of descriptive statistics are further calculated: *location statistics* (mean, median and quantiles) and *dispersion statistics* (standard deviation and standard error). Statistics calculated are:

TABLE 6.3. An example of corpus sampling: Sample structure

Branches and categories		# of documents
Life Science	Biology	9,917
	Medicine, Research & Experimental	19,744
Social Science	Psychology	6,989
	Political Science	5,106
	Management	14,339
	Sociology	4,725
Physical Science	Physics, Particles & Fields	13,203
	Physics, Atomic, Molecular & Chemical	17,010
Earth Science	Geography	3,908
	Geology	2,153
	Astronomy & Astrophysics	22,825
Formal Science	Mathematics, Applied	27,982

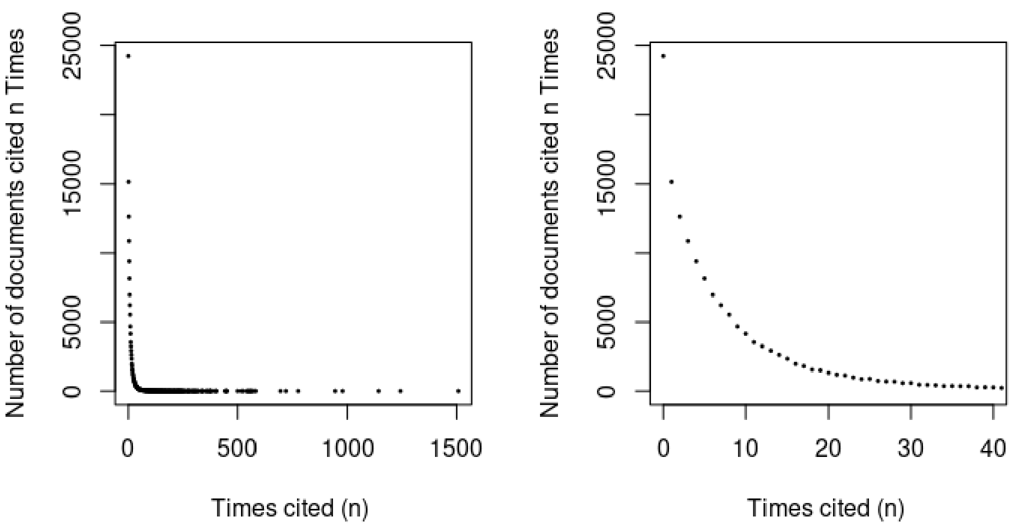


FIGURE 6.2. Graph of the number of documents cited n times in the sample selected. The figure on the right hand side shows the number of documents cited up to 40. The average citation in the sample is approximately 9 (9.43).

- *N:* Number of documents in the corresponding set.
- *Max:* Maximum citation counts in the corresponding set.
- *Min:* Minimum citation counts in the corresponding set.
- *Mean (μ):* The arithmetic average of citation counts in the corresponding set. We note that as the mean is not a robust measure of central tendency, it will be very sensitive to even one aberrant value (very high citation) out of N documents.
- *Q_1:* The first quartile. It indicates the median of the lower half of the data. This means that 25% of documents cited less than Q_1.

- Q_2 *(median):* The middle number in the data. The median is a robust statistic against an outlier and can resist up to 50 % of outliers.
- Q_3*:* The third quartile. It indicates the median of the upper half of the data, which means that 75% of documents cited less than Q_3.
- σ*:* Standard deviation. It is the dispersion of the data. It shows how accurately the mean represents the sample.
- SE*:* Estimated standard error of the mean. It measures how far the sample mean is likely to be from the population mean. That is, it is a measure of how precise is our estimate of the mean.

Descriptive statistics for the corpus are presented in Table 6.4. The average citation in the corpus is approximately 9 and 29% of the documents are cited more than 9 times. Although the maximum number of citation is 8,234, from the upper and lower quartile we can conclude that approximately 49% of the documents are cited 1-9 times in the corpus. The statistics for the sample can be also found in the table. The mean value and the quartiles are very similar to the corpus with a lower maximum citation count.

TABLE 6.4. Descriptive statistics for times cited in the corpus and samples

Set	N	Max	Min	μ	Q_1	Q_2	Q_3	σ	SE
Corpus	1,673,350	8,234	0	9.38	1	4	11	23.33	0.018
Sample	147,901	1,508	0	9.43	1	5	11	18.30	0.047
$Sample_1$ (Life Science)	29,661	978	0	10.95	2	6	13	19.96	0.116
$Sample_2$ (Social Science)	31,159	304	0	7.70	1	4	10	12.83	0.073
$Sample_3$ (Physical Science)	30,213	1,508	0	11.57	2	7	14	21.86	0.126
$Sample_4$ (Earth Science)	28,886	1,243	0	12.44	2	7	15	22.81	0.134
$Sample_5$(Formal Science)	27,982	253	0	4.33	0	2	5	8.03	0.048
Biology	9,917	517	0	10.36	2	6	13	16.09	0.162
Medicine, Research & Experimental	19,744	978	0	11.24	2	6	13	21.63	0.154
Psychology	6,989	300	0	10.56	3	7	14	14.75	0.176
Political Science	5,106	293	0	7.04	1	4	9	10.89	0.152
Management	14,339	304	0	6.92	0	2	8	13.30	0.111
Sociology	4,725	192	0	6.54	1	4	8	9.17	0.133
Physics, Particles & Fields	13,203	1,508	0	11.72	2	6	14	22.81	0.199
Physics, Atomic, Molecular & Chemical	17,010	1,143	0	11.46	3	7	14	21.09	0.162
Geography	3,908	775	0	10.27	2	6	13	18.95	0.303
Geology	2,153	108	0	7.86	2	5	10	9.18	0.198
Astronomy & Astrophysics	22,825	1,243	0	13.24	2	7	16	24.20	0.160
Mathematics, Applied	27,982	253	0	4.33	0	2	5	8.03	0.048

In the LSC, the average citation counts in categories vary differently. Descriptive statistics for each categories are presented in Table E.1 and the Fig. 6.3. As can be seen from the histogram, for more than the half of categories the average citation lies between 5 to 10. Almost 20 categories have documents cited 2 times on average. We suspected that research fields and the number of documents assigned to categories may be two factors for low citation counts. The scope of the category may be a factor that affects the citation counts. Documents of multidisciplinary categories may be cited by researchers from different fields. For instance, the maximum citation in the category ‘Multidisciplinary Sciences’ is 3,004 with a relatively high mean 18.32, first quartile 4 and third quartile 18. This indicated that 50% of documents are cited 4-18 times, which is relatively higher than the corpus.

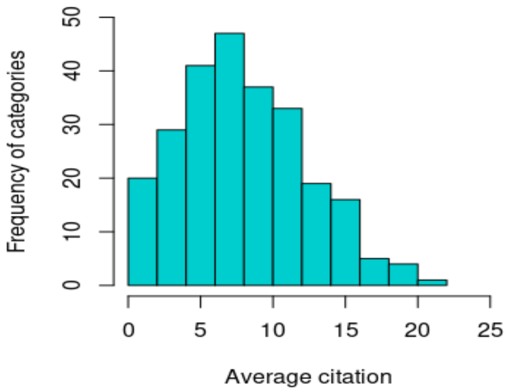


FIGURE 6.3. Frequency of categories versus average citation counts

Ten categories with the maximum citation counts in the corpus are presented with the first, second and third quartile in Table 6.5. We can see categories involving several academic disciplines appear in the list. It should be mentioned that category ‘Computer Science, Interdisciplinary Applications’ has relative small quartiles than other categories. This means that citation counts for this category are low in general even if there exist highly cited documents. This suggests that this fact may be characteristic for some categories. Therefore, it is also important to look at categories low almost 0 citation. We look at categories with very low third quartile value. We presented selected categories in Table 6.6. It can be clearly seen that papers of discourse studies and fine arts are cited a few times in general. This fact is also noticeable for computer science and engineering categories with very high number of documents.

TABLE 6.5. Categories with the maximum citation counts in the corpus

Category	**Max**	Q_1	Q_2	Q_3
Oncology	8,234	4	9	17
Genetics & Heredity	6,756	4	8	17
Computer Science, Interdisciplinary Applications	4,744	0	1	6
Biotechnology & Applied Microbiology	4,744	3	7	14
Biochemical Research Methods	4,744	3	8	14.75
Statistics & Probability	4,744	1	3	7
Mathematical & Computational Biology	4,744	1	4	9
Medicine, General & Internal	3,908	2	4	10
Multidisciplinary Sciences	3,004	4	9	18
Materials Science, Multidisciplinary	2,464	0	4	12
Nanoscience & Nanotechnology	2,464	2	8	20

In general, coverage of the natural sciences and the medicine are much more richer than other branches in the LSC. However, we did not observe any explicit correlation between the average citation count and the number of documents. Fig.

TABLE 6.6. Categories with low citation counts in the corpus

Category	N	**Max**	Q_1	Q_2	Q_3
Literary Theory & Criticism	498	45	0	0	0
Dance	74	8	0	0	0
Literature, Slavic	35	3	0	0	0
Literary Reviews	35	2	0	0	0
Architecture	1,376	145	0	0	1
Humanities, Multidisciplinary	2,559	53	0	0	1
Asian Studies	877	32	0	0	1
Literature	1,608	18	0	0	1
Medieval & Renaissance Studies	485	11	0	0	1
Classics	325	9	0	0	1
Engineering, Electrical & Electronic	174,272	2,028	0	0	3
Telecommunications	40,550	2,028	0	1	3
Computer Science, Information Systems	45,865	715	0	1	3
Computer Science, Hardware & Architecture	18,489	532	0	1	3
Language & Linguistics	5,174	112	0	1	3

6.4 shows the number of documents per category versus the average citation counts in the category with the linear regression line. We can see a slight increasing trend with a wide spread of points around the line. The figure also indicates that two categories with a high documents assigned, can have very different citation counts with different means. For instance, the number of documents of 'Engineering, Electrical & Electronic' is 174,272 and 'Materials Science, Multidisciplinary' is 112,912; however, the average citation for 'Materials Science, Multidisciplinary' is 3 times more than the average citation for 'Engineering, Electrical & Electronic' (Table 6.7). We can also see the same fact for quartiles. This fact is also confirmed in the Fig. 6.5 and Table 6.8. A slight decreasing trend of average citation with the rank of categories is observable in the figure, but we can also see the spread of points. Two closely ranked categories can have very different statistic and two categories with very similar statistics can have very different rank in the corpus. therefore, there may be other factors that contributes to citation counts such as research fields.

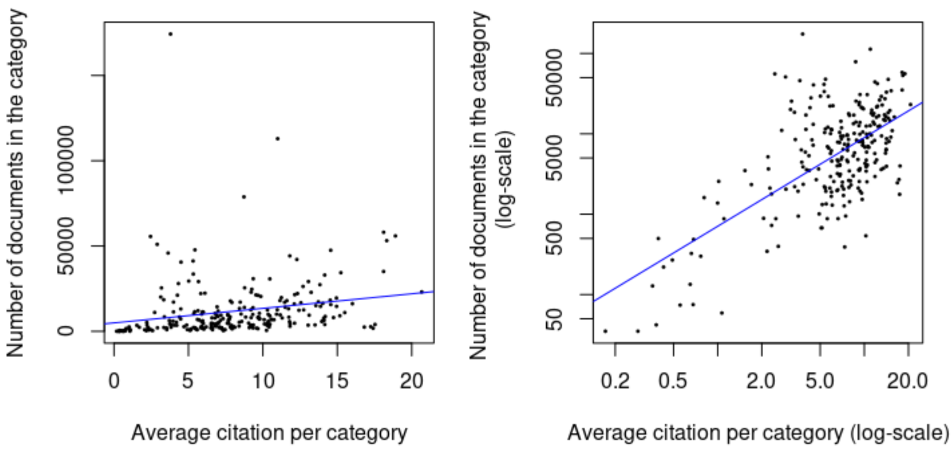


FIGURE 6.4. Number of documents per category versus the average citation counts in the category

TABLE 6.7. Top ranked five categories (by the number of documents)

Category	N	**Max**	**Min**	μ	Q_1	Q_2	Q_3
Engineering, Electrical & Electronic	174,272	2,028	0	3.8	0	0	3
Materials Science, Multidisciplinary	112,912	2,464	0	10.99	0	4	12
Physics, Applied	78,796	2,288	0	8.73	0	3	9
Chemistry, Physical	58,065	2,288	0	18.11	5	10	20
Chemistry, Multidisciplinary	55,907	2,210	0	18.9	3	9	21

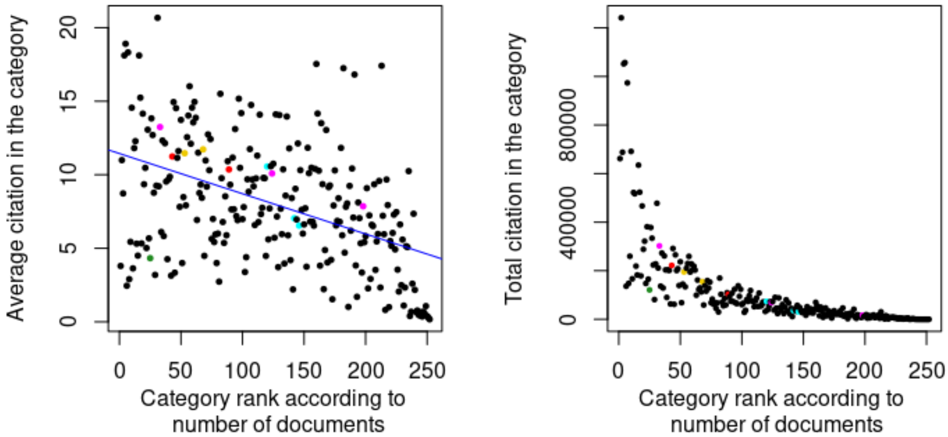


FIGURE 6.5. Average citation per categories versus category rank according to the number of documents. Colours indicate the categories in the sample: red is for life science categories, cyan is for social science categories, yellow is for physical science categories, pink is for earth science categories and green is for formal science category.

TABLE 6.8. Categories with the highest average citation counts

Category	N	**Max**	**Min**	μ	Q_1	Q_2	Q_3
Cell Biology	23,108	978	0	20.67	6	12	24
Chemistry, Multidisciplinary	55,907	2,210	0	18.9	3	9	21
Multidisciplinary Sciences	53,140	3,004	0	18.32	4	9	18
Chemistry, Physical	58,065	2,288	0	18.11	5	10	20
Nanoscience & Nanotechnology	35,050	2,464	0	18.11	2	8	20
Critical Care Medicine	3,982	320	0	17.53	5	11	21
Allergy	1,765	488	0	17.41	4	10	20
Neuroimaging	2,702	527	0	17.24	6	12	22
Cell & Tissue Engineering	2,455	261	0	16.81	4	10	20
Medicine, General & Internal	16,179	3,908	0	16.01	2	4	10

6.6. Citation Classification

In this section, we explore the efficacy of semantic meaning of scientific articles represented in the MS in predicting the impact of articles on basis of the LSC. We employed classification algorithms in order to analyse the predictability of high/less citation of scientific papers.

We apply the methods for binary classification and investigate the binary classification problem for differentiating between two types of impacts of scientific papers:

less-cited (L) and highly-cited (H) papers. Our initial hypothesis is that if the semantic meaning of texts indicate the impact of papers, a classifier should perform significantly better than chance (50%).

Distinguishing between highly-cited and less-cited papers can be done by various definitions. Two basic approaches to identify classes are to use absolute and relative thresholds [221]. Using an absolute number of citations for the entire corpus could lead to the predominance of papers from highly-cited categories due to the differences in the average citation within disciplines. Instead, relative thresholds define the highly-cited and less-cited papers in each scientific category.

In this study, we begin with using relative thresholds for individual categories. Two rules to define thresholds for highly-cited and less-cited classes are used. In the first scheme, papers assigned to a category are divided into H/L classes according to the average citation in the category. A paper is labelled as highly-cited (H) if it has received more than the average citation, less-cited (L) otherwise.

We then conduct the experiment to classify papers according to extremely highly-cited (EH) and less-cited (EL) papers. In this scheme, papers in a certain sample are divided into four partitions by its corresponding quartiles from citations, where papers belong to one of the four partitions, described as:

- P_1: papers cited less than or equal to the lower quartile Q_1
- P_2: papers cited more than Q_1 and less than or equal to the median Q_2
- P_3: papers cited more than Q_2 and less than the upper quartile Q_3
- P_4: papers cited more than or equal to Q_4.

In this experiment, we use articles from the set $P_1 \cup P_4$. This implies that we performed binary classification on the subset of texts with citations more than Q_3 and less than Q_1. The partitions P_1 and P_4 are defied as groups of two extreme citations where papers in the partitions P_1 and P_4 are labelled as extremely less-cited (EL) and extremely highly-cited (EH).

Two classification methods are initially investigated for comparison: Linear Discriminant Analysis with Fisher's discriminant rule and k-nearest neighbour. We follow procedures:

(1) We apply Fisher's LDA in classifying scientific papers by considering the selected sample set of texts as a whole training set.
(2) We apply kNN in classifying scientific papers by considering the selected sample set of texts as a whole training set.
(3) We apply leave-one-out cross validation (LOOCV) in classifying scientific papers for the best LDA separation for demonstration of statistical power of the classifier.

TABLE 6.9. Number of texts assigned to categories selected in each class

	H/L classes			EH/EL classes		
Category	Class H	Class L	Total	Class EH	Class EL	Total
Mathematics, Applied	7,975	20,007	27,982	7,975	7,925	15,900
Biology	3,055	6,862	9,917	2,553	2,772	5,325
Management	4,426	9,913	14,339	3,980	5,257	9,237

We finally extend the study to the supervised PCA defined for classification problem. By supervised PCA, the dimension of FVTs will be reduced and the procedures for classification of papers defined above will be repeated with the reduced dimensionality.

The evaluation of the potential impact of articles through the semantic analysis is made by using different representations of the meaning of texts. Vectors $FVT_{1i}, FVT_{2i}, FVT_{3i}, FVT_{4i}, FVT_{5i}$ in the constructed space are created and compared in classification.

We selected three categories from three different branches of science in Table 6.4 for the experimental study: Mathematics, Applied, Biology and Management. The numbers of texts assigned to each class in categories are presented in Table 6.9.

In kNN, to avoid the impact of the unevenly distribution of sample size in classes, we tested kNN with weights as described in Section 6.4. This modification achieved better classification performance for this imbalance classes, therefore we present only results for weighted kNN. In this study, we applied kNN where the k is 1, 3, 5, 7, 11, 13 and 17 for each vector. We presented only the results for value of k with the best classification performance in each case individually.

6.6.1. The spaces used in classification

In this research, classification methods are compared for different vector representations in 3 different vector spaces.

6.6.1.1. ***Original Space.*** The original space refers to the Meaning Space that is defined in Section 6.3. In this space, each text is represented by the FVTs in various ways such as by only 252-dimensional mean vector or 504-dimensional vector as a combination of the mean vector and PC1 vector. This means that each text has at least 252 dimension in the Meaning Space.

6.6.1.2. ***13-Dimensional Reduced Basis.*** As described in [**220**], words can be represented in PC axes and the optimal number of PC was determined as 13 by the PCN-CN. In order to carry out additional investigations on the basis of 13-dimensional word space, we represented words in 13-Dimensional PC space and then constructed the FVU based on these vectors as described in Section 5. In this case, each text is represented by at least 13-dimensional vectors, for instance, the mean

TABLE 6.10. Dimensions of FVT vectors in the original space and the constructed space after applying Supervised PCA for the category Mathematics, Applied

		H/L classes			EH/EL classes		
Vector	Original Space	Kaiser Rule	Broken Stick	Supervised PCA	Kaiser Rule	Broken Stick	Supervised PCA
$(\vec{\mu})$	252	36	15	14	32	15	14
$(\overrightarrow{PC_1})$	252	25	8	8	24	8	8
$(\vec{\mu}, \overrightarrow{PC_1})$	504	59	22	16	59	22	22
$(\vec{c_1}, \vec{c_2})$	504	41	20	19	41	20	14
$(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$	756	62	24	23	63	24	24

of 13 dimensional coordinates. We compared the performance of the classifiers for each vector in the original space and space of reduced basis.

6.6.1.3. ***The space constructed after Supervised PCA.*** For binary classification, we followed the following procedure: for each vector representation of text,

(1) Apply PCA to the data in the original space.
(2) Find the number of components by Kaiser and Broken Stick rules to be used the initial number of components in supervised PCA.
(3) Initialise the number of components (ncomp) as the minimum number found in previous step (We used Broken Stick results).
(4) Apply Supervised PCA with the components from 1 to ncomp, represent all texts on this new spaces.
(5) Apply LDA with the data on constructed spaces and calculate the sum of sensitivity and specificity.
(6) Take the maximum of the sum of sensitivity and specificity, and identify the number of components for this number. Report the sum of sensitivity and specificity to evaluate the performance of LDA classifier.
(7) Apply kNN with the identified number of dimension by LDA. Calculate the sum of sensitivity and specificity to evaluate the performance of kNN classifier.

The number of components (dimensions of FVT vectors) in the original space and the constructed space after applying Supervised PCA for categories Mathematics, Applied, Biology and Management are presented in Tables 6.10-6.12.

6.6.2. Results

Two classification algorithms were performed to show how much information the semantics of texts have on impact of articles: LDA and kNN. We selected three categories to demonstrate the results of classification and the differences in efficiency of semantics on bibliometric characteristics of articles in different categories.

TABLE 6.11. Dimensions of FVT vectors in the original space and the constructed space after applying Supervised PCA for the category Biology

		H/L classes			EH/EL classes		
Vector	Original Space	Kaiser Rule	Broken Stick	Supervised PCA	Kaiser Rule	Broken Stick	Supervised PCA
$(\overrightarrow{\mu})$	252	35	13	12	36	14	12
$(\overrightarrow{PC_1})$	252	29	10	8	30	11	9
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	504	64	25	25	64	25	25
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	504	50	23	23	49	22	20
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	756	76	30	30	76	30	30

TABLE 6.12. Dimensions of FVT vectors in the original space and the constructed space after applying Supervised PCA for the category Management

		H/L classes			EH/EL classes		
Vector	Original Space	Kaiser Rule	Broken Stick	Supervised PCA	Kaiser Rule	Broken Stick	Supervised PCA
$(\overrightarrow{\mu})$	252	36	13	11	36	13	13
$(\overrightarrow{PC_1})$	252	23	9	9	24	9	9
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	504	62	23	23	61	22	22
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	504	39	16	11	40	16	11
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	756	61	25	25	62	25	25

6.6.2.1. ***Applied Mathematics***

The citation is determined by the nature of the categories individually. There are many factors that contribute the citation counts in individual research fields. Factors like size of the field, annual scientific production and relevance of article for the research activities in a field are very influential on the number of citation.

The category Applied Mathematics is selected as being a sample from Formal Science - one of five main branches of science. It is noteworthy that articles assigned to the category Applied Mathematics usually address issues combining the mathematical methods and specialised knowledge in specific fields such as physics, biology and business. Such categories generally suggest that there is no recurrent common characteristics of articles and there are large differences in citation patterns of papers from different disciplines. In Applied Mathematics, the separation of papers into highly popular fields may not be easily done. Therefore, the prediction of citation by semantics of such papers may be difficult for categories of multidisciplinary publications.

For Applied Mathematics, classification results of two algorithms in different spaces are shown in Tables 6.13-6.17. The first column in the tables demonstrates the type of FVT to represent the text. The other columns present the sensitivity

TABLE 6.13. Results of the citation classifier LDA according to H/L for the category Mathematics, Applied

	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
Vector	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\vec{\mu})$	67.91	49.61	117.52	77.02	29.14	106.16	68.28	44.55	112.83
$(\overrightarrow{PC_1})$	61.66	57.02	118.68	63.42	44.68	108.10	63.49	47.24	110.73
$(\vec{\mu}, \overrightarrow{PC_1})$	58.83	63.73	122.56	62.91	46.47	109.38	59.12	54.37	113.49
$(\vec{c_1}, \vec{c_2})$	67.40	49.53	116.93	49.54	58.89	108.43	57.93	50.82	108.75
$(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$	64.15	58.85	123.00	55.37	54.96	110.33	54.13	59.47	113.61

(%), specificity (%) and the sum of sensitivity and specificity (%) in each experiment. Sensitivity is the proportion of high-cited papers that are correctly identified as high-cited and specificity is the proportion of low-cited papers that are correctly identified as low-cited. The results of LDA and kNN are represented in separate tables. For kNN (weighted kNN), results are presented only for k where classification performance is the best, that is, the sum of sensitivity and specificity is the maximum. In each table, we used red color to highlight the case where the algorithm achieved the highest classification performance according to the criterion used. Also, all procedures are repeated for H/L and EH/EL classes.

For H/L classification, results of the LDA on three spaces are shown in the Table 6.13. Looking at the results in the table, we can see that LDA outperforms in the Original Space for all the vector representations with sensitivity 64.15%, specificity 58.85% and the sum 123%. The best results are obtained in this space followed by the space constructed by supervised PCA and then the space constructed with 13-dimensional reduced basis. We can also see that the vector $(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$ performs better than others. $(\vec{\mu}, \overrightarrow{PC_1})$ has achieved the second best performance in all cases. In general, we can conclude that adding $\overrightarrow{PC_1}$ to vectors $\vec{\mu}$ and $(\vec{c_1}, \vec{c_2})$ improves the classification performances.

The binary classification with classes EH/EL achieves higher performance results for all experiments. This implies that the LDA performs better for distinguishing two extreme citation classes. These classes are also higher sensitivity and specificity values, a clue that proves that the semantics of papers with two extreme citation counts are clearly different. The best performance is still achieved when using the $(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$ in the original space, where sensitivity 67.25%, specificity 62.97% and the sum 130.21%. In general, the results show that performances of LDA have an almost identical trend of rise and decline in all spaces.

In order to carry out additional investigations on the basis of these findings obtained by LDA, we performed LOOCV in H/L classification by LDA in the Original Space only and compared the results (Table 6.15). As the performance in each experiment and general trends and values with and without LOOCV is very similar,

TABLE 6.14. Results of the citation classifier LDA according to EH/EL for the category Mathematics, Applied

	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
Vector	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\vec{\mu})$	69.93	54.69	124.62	74.37	35.51	109.88	69.93	47.31	117.24
$(\overrightarrow{PC_1})$	65.20	58.90	124.11	69.82	39.70	109.52	66.68	47.34	114.03
$(\vec{\mu}, \overrightarrow{PC_1})$	71.56	58.37	129.93	68.60	44.69	113.30	65.08	52.61	117.68
$(\vec{c_1}, \vec{c_2})$	68.66	54.16	122.82	55.94	55.04	110.98	53.98	56.76	110.74
$(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$	67.25	62.97	130.21	54.19	58.03	112.23	64.89	53.56	118.45

TABLE 6.15. Results of the citation classifier LDA with LOOCV in according to H/L for the category Mathematics, Applied (for the Original Space only)

Vector	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\vec{\mu})$	65.24	48.29	113.53
$(\overrightarrow{PC_1})$	58.36	56.35	114.71
$(\vec{\mu}, \overrightarrow{PC_1})$	60.30	57.90	118.20
$(\vec{c_1}, \vec{c_2})$	62.87	48.25	111.12
$(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$	59	56	115.5

TABLE 6.16. The best results of the citation classifier kNN for each vector according to H/L for the category Mathematics, Applied

	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
Vector	k	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	k	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	k	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\vec{\mu})$	11	62.60	48.16	110.76	13	65.53	41.97	107.50	17	68.14	41.06	109.19
$(\overrightarrow{PC_1})$	13	68.46	40.76	109.22	17	63.54	44.16	107.70	17	62.82	44.13	106.95
$(\vec{\mu}, \overrightarrow{PC_1})$	13	68.46	40.77	109.23	11	52.87	55.75	108.62	17	65.54	43.10	108.64
$(\vec{c_1}, \vec{c_2})$	11	57.30	50.07	107.38	13	63.50	43.06	106.56	13	61.87	43.68	105.55
$(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$	11	56.25	53.20	109.45	17	66.78	41.02	107.80	17	65.96	42.77	108.73

we did not present LDA with LOOCV for other spaces and extreme citation classification. In this case, the best performance is achieved for the vector $(\vec{\mu}, \overrightarrow{PC_1})$ with a slight differences from the vector $(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$. This demonstrates us the statistical similarity for two applications.

Tables 6.16 and 6.17 present results of kNN classifier for classes H/L and EH/EL, respectively. Similar to the LDA, we observed that the kNN achieved the highest sum corresponds to the extreme citation classification (EH/EL) with the vector $(\vec{\mu})$ in the space constructed by supervised PCA. The sensitivity reaches to 70.53% and the specificity is 45.94 with a sum of 116.48% in this case. However, there is only a slight difference between values in this space and the Original Space.

TABLE 6.17. The best results of the citation classifier kNN for each vector according to EH/EL for the category Mathematics, Applied

Vector	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	k	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\vec{\mu})$	11	77.83	37.99	115.82	11	68.41	44.57	112.98	17	70.53	45.94	116.48
$(\overrightarrow{PC_1})$	17	73.52	41.22	114.74	17	61.69	49.09	110.78	17	61.87	49.55	111.42
$(\vec{\mu}, \overrightarrow{PC_1})$	17	73.52	41.22	114.74	11	66.55	45.88	112.43	17	67.74	46.97	114.70
$(\vec{c_1}, \vec{c_2})$	13	72.21	39.80	112.01	17	65.92	44.40	110.32	17	60.13	47.77	107.90
$(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$	17	73.63	40.93	114.56	17	66.22	45.31	111.53	17	67.81	46.37	114.18

TABLE 6.18. Results of the citation classifier LDA according to H/L for the category Biology

Vector	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	Sens. (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\vec{\mu})$	74.30	69.21	143.51	64.29	63.48	127.77	70.47	64.33	134.80
$(\overrightarrow{PC_1})$	72.50	63.87	136.38	64.29	52.04	116.33	71.72	55.80	127.52
$(\vec{\mu}, \overrightarrow{PC_1})$	75.38	71.60	146.98	65.66	63.45	129.11	70.77	59.82	130.59
$(\vec{c_1}, \vec{c_2})$	70.70	69.12	139.82	67.53	50.45	117.98	58.89	58.90	117.79
$(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$	78.69	67.06	145.76	68.02	55.09	123.11	67.66	64.16	131.82

For Applied Mathematics, the best results with the sum of sensitivity and specificity being greater than 130% were achieved for the vector $(\vec{c_1}, \vec{c_2}, \overrightarrow{PC_1})$ in the original space by the classifier LDA. This result corresponds to the classification of extreme citations (EH/EL) with 67.25% sensitivity and 62.97% specificity. This means that LDA with selection of the threshold by the sum of sensitivity and specificity maximisation is better for classifying extreme citations.

6.6.2.2. ***Biology***

We selected the category Biology as a sample from Life Science branch (see Table 6.3). The Tables 6.18-6.21 show very similar results obtained for Applied Mathematics. The classification performance is higher when using for both the separation of H/L and EH/EL classes as for Applied Mathematics.

In almost all experiments, results of both classifiers LDA and kNN show that classifiers outperform in the Original Space for all the vector representations, excluding the vector $(\vec{\mu}, \overrightarrow{PC_1})$ when kNN is employed.

For LDA, one can see that the vector $(\vec{\mu}, \overrightarrow{PC_1})$ has the highest sum of sensitivity and specificity for both H/L and EH/EL classification, with 146.98% and 163.05% respectively. The vector $(\vec{\mu})$ has achieved the best performance when using kNN,

TABLE 6.19. Results of the citation classifier LDA according to EH/EL for the category Biology

Vector	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	Sens. (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\overrightarrow{\mu})$	79.91	79.44	159.34	67.18	69.70	136.87	71.05	75.14	146.20
$(\overrightarrow{PC_1})$	75.87	72.37	148.24	69.57	53.03	122.60	71.76	65.95	137.70
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	81.63	81.42	163.05	70.47	68.83	139.30	75.28	67.28	142.56
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	77.63	76.70	154.33	70.07	54.15	124.22	49.94	74.57	124.51
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	81.12	80.66	161.78	67.53	63.28	130.80	71.80	72.04	143.84

TABLE 6.20. The best results of the citation classifier kNN for each vector according to H/L for the category Biology

Vector	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	k	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\overrightarrow{\mu})$	17	70.64	59.72	130.36	17	64.39	61.02	125.40	13	60.88	67.04	127.92
$(\overrightarrow{PC_1})$	17	63.37	59.59	122.96	15	28.84	82.18	111.02	17	58.92	60.65	119.57
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	17	63.34	59.59	122.93	17	64.91	61.99	126.90	17	62.03	60.01	122.04
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	11	65.60	58.36	123.96	17	62.19	59.49	121.68	17	61.11	58.95	120.06
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	17	63.99	59.56	123.55	17	61.73	59.46	121.19	17	60.72	60.71	121.43

TABLE 6.21. The best results of the citation classifier kNN for each vector according to EH/EL for the category Biology

Vector	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
	k	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\overrightarrow{\mu})$	17	82.33	60.61	142.94	17	73.21	64.83	138.03	17	74.19	66.31	140.49
$(\overrightarrow{PC_1})$	17	72.42	61.00	133.43	17	66.51	61.69	128.20	17	66.98	64.54	131.52
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	17	72.42	61.00	133.43	17	72.78	64.25	137.03	17	69.72	61.18	130.91
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	17	71.84	59.63	131.47	17	67.06	63.56	130.62	17	66.94	62.55	129.49
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	13	72.39	61.94	134.33	17	66.98	63.74	130.72	17	70.07	62.16	132.23

where the sums of sensitivity and specificity are 130.36% and 142.94% for H/L and EH/EL classes.

For the category Biology, binary classification results show the classifier giving the highest classification performance was LDA with the text represented by the vector $(\overrightarrow{\mu}, \overrightarrow{PC_1})$ in the Original Space. This result was achieved for extreme citation classification.

6.6.2.3. ***Management***

TABLE 6.22. Results of the citation classifier LDA according to H/L for the category Management

Vector	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	Sens. (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\overrightarrow{\mu})$	79.55	63.93	143.48	70.56	51.25	121.81	76.10	58.19	134.28
$(\overrightarrow{PC_1})$	72.82	65.94	138.76	65.45	49.19	114.64	71.31	56.85	128.16
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	80.05	65.82	145.87	69.20	53.18	122.39	70.20	62.49	132.69
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	79.55	60.96	140.51	48.67	69.12	117.79	53.23	62.47	115.70
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	79.89	65.85	145.74	56.87	61.36	118.23	78.38	54.54	132.92

TABLE 6.23. Results of the citation classifier LDA according to EH/EL for the category Management

Vector	Original Space			13-Dimensional Reduced Basis			Supervised PCA		
	Sens. (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\overrightarrow{\mu})$	85.20	77.46	162.66	72.24	60.17	132.40	80.30	68.86	149.16
$(\overrightarrow{PC_1})$	80.68	74.21	154.88	66.86	54.21	121.07	70.38	70.00	140.38
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	83.22	81.81	165.03	67.24	66.56	133.80	78.97	69.13	148.10
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	81.96	75.82	157.78	51.88	72.00	123.88	64.27	61.50	125.77
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	82.21	82.54	164.75	74.62	52.92	127.54	75.95	73.77	149.72

Finally, we select the category Management from Social Science branch. We tested two classifiers for each vector in three spaces. The behavious of classifiers are very similar across the Management and Biology. In Tables 6.22-6.25, we present a comparison of LDA and kNN based on the maximum sum of sensitivity and specificity for each vector representation and space combination separately. Overall results follow a typical pattern of rise and decline, showing almost the same trend as the category Biology. In general, we observe a significant increase in sensitivity and specificity for extreme citation classification tasks for both LDA and kNN.

According to results in Tables 6.22-6.23, LDA reaches the best classification accuracies in Original Space. By combining the vector representations of $(\overrightarrow{\mu})$ and $(\overrightarrow{PC_1})$, we were able to achieve results that LDA for combined vectors outperforms LDA for other FVTs. The best classification accuracy of LDA is 83.22% sensitivity and 81.81% specificity in binary classification of EH/EL for the vector $(\overrightarrow{\mu}, \overrightarrow{PC_1})$. This is also the best accuracy achieved among the all experiences in categories. However, the differences of performance from the vector $(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$ is not very large, which also support our findings that combined vectors result in an improvement of performance in binary classification of citation.

Tables 6.24 and 6.25 show the performance of kNN on binary classification tasks for H/L and EH/EL. Results in tables indicate that kNN is also sightly better in the Original Space than in other spaces. Papers according to extreme citations are found

TABLE 6.24. The best results of the citation classifier kNN for each vector according to H/L for the category Management

	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
Vector	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\overrightarrow{\mu})$	17	74.40	52.10	126.50	17	65.09	52.89	117.98	17	67.67	57.80	125.47
$(\overrightarrow{PC_1})$	17	71.83	49.12	120.94	17	61.59	55.39	116.98	17	62.34	57.62	119.96
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	17	71.83	49.12	120.94	17	68.28	52.80	121.08	17	66.29	54.72	121.01
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	13	65.50	56.01	121.51	17	63.47	52.85	116.32	17	60.71	56.11	116.82
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	17	71.74	49.08	120.81	17	65.07	51.94	117.01	17	65.14	54.98	120.12

TABLE 6.25. The best results of the citation classifier kNN for each vector according to EH/EL for the category Management

	Original Space				13-Dimensional Reduced Basis				Supervised PCA			
Vector	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)	*k*	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\overrightarrow{\mu})$	17	82.01	60.09	142.10	17	74.62	57.49	132.11	17	78.19	63.31	141.50
$(\overrightarrow{PC_1})$	17	80.10	52.92	133.02	17	71.26	57.43	128.68	17	71.93	60.95	132.88
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	17	80.10	52.94	133.04	17	76.23	57.07	133.30	17	76.26	57.39	133.65
$(\overrightarrow{c_1}, \overrightarrow{c_2})$	7	69.62	63.72	133.35	17	73.39	56.25	129.64	11	72.89	55.79	128.68
$(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$	7	69.55	62.74	132.28	17	74.25	55.26	129.51	11	65.55	68.02	133.58

to be classified more efficiently than for H/L classification by kNN. kNN performed the best when using the vector $(\overrightarrow{\mu})$ in Original Space, achieved the sensitivity 82.01% and specificity 60.09% and the sum 142.10% in EH/EL classification.

In order to illustrate the results of the best performance of classifiers, the histogram of data projection on the first LDA axis and ROC curve (see Figures 6.6 are used. We note that the best performance corresponds to LDA for classes extremely less-cited (EL) and extremely highly-cited (EH) with the vector (μ, PC_1) in the original space of the category Management. From the histogram, we can see the separation of papers in two classes (in proportion) by the optimal cut-off found by maximising the sum of sensitivity and specificity. In the ROC curve, the corresponding AUC achieved in the classification is 90%.

For the vector that gives the best classification results, we applied the LDA with LOOCV to test the statistical power of the classifier LDA, presented in Table 6.26. From the table, the LDA predicted extremely highly-cited papers with 80.43% sensitivity and 78.90% specificity for extremely less-cited papers with slight difference (2.79% for sensitivity and 2.91% for specificity) from LDA without LOOCV.

6.6.2.4. ***Summary of findings from binary classification experiments***

From experiments in three categories, the following overall results are obtained:

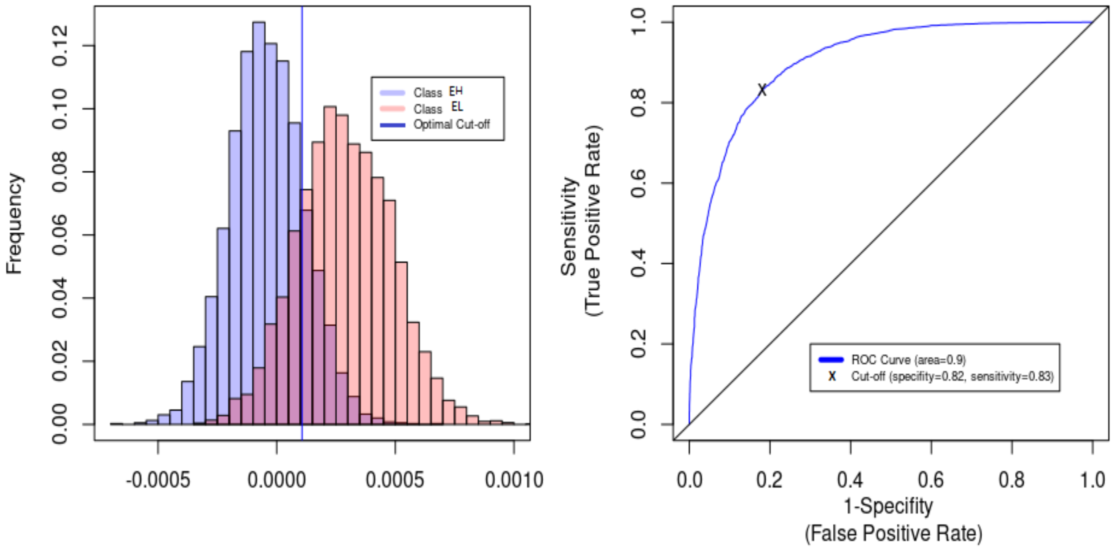


FIGURE 6.6. (Left) Data projection on the first LDA axis (Right) ROC curve for LDA. Both figures correspond to the best performance of classifiers: LDA for classes extremely less-cited (EL) and extremely highly-cited (EH) with the vector (μ, PC_1) in the original space of the category Management.

TABLE 6.26. LOOCV results for the best classification performance among all categories: LDA with LOOCV in according to EH/EL for the category Management. The results correspond to the vector (μ, PC_1) in the Original Space.

Vector	*Sens.* (%)	*Spec.* (%)	*Sum* (%)
$(\overrightarrow{\mu}, \overrightarrow{PC_1})$	80.43	78.90	159.33

(1) In general, the highest sensitivity and specificity are reached by using the vectors $(\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{PC_1})$ (for Applied Mathematics) and $(\overrightarrow{\mu}, \overrightarrow{PC_1})$ (for Biology and Management) for LDA; and the vector $(\overrightarrow{\mu})$ (for all categories experimented) for kNN in almost all spaces in experiments.
(2) We observed that large k values lead to relatively good performance in all classification experiments.
(3) LDA outperforms kNN in both classifications according to H/L and EH/EL classes in all spaces and vectors.
(4) Both classifiers perform better in EH/EL classification in all spaces and with all vectors.
(5) The best results are achieved when we employ LDA classifier for EH/EL classes. This implies that LDA with selection of the threshold by sensitivity+specificity maximisation is better for classifying extreme citations. The LDA reached to sensitivity and specificity greater than 80%. This is a high accuracy for this type of problem.

(6) When comparing spaces, the best results are obtained for the Original Space in almost all experiments, followed by the space constructed by Supervised PCA and then the space constructed with 13-Dimensional reduced basis. We can conclude that representing words in 13-dimensional space leads to lose some information about the citation of papers.

(7) Semantics is more important indicator of citation for some categories than others. For categories Biology and Management the results show an improvement over Applied Mathematics. The poorest results were obtained for the category Applied Mathematics.

(8) Finally, combining ($\overrightarrow{PC_1}$) vector with vectors ($\overrightarrow{\mu}$) and ($\overrightarrow{c_1}, \overrightarrow{c_2}$) results in an improvement of the classifier performance.

6.7. Conclusion and Discussion

In this chapter, two issues in NLP have been identified, solutions have been suggested and proposed approaches have been analysed. First, we introduced a new text representation method which is one of the main problems in all NLP applications. Secondly, we evaluated the impact of scientific articles using this new representation technique. Predictive models are studied for classifying the citation of papers based on their informational semantics and several representation models in different vector spaces are tested in classification experiments. The aim of this study is to assess the discrimination ability of scientific impact of papers through their semantics.

In this approach, each text is a cloud of words represented by RIGs from word to category in the MS. In order to construct text representation, we used the distribution of the corresponding words' RIGs in each dimension. The FVT are created by set of parameters (e.g. mean values of a cloud's points) summarising the information in texts. Five different vector representations as a combination of the mean vector, the vector of the first principal component for each text and two centroid vectors obtained by k-means clustering of words in the text are introduced as informational representations of semantics and analysed for binary classification problem.

With constructed FVTs, we have showed how much information the semantics of texts carry in citation of scientific articles. Classification is performed for two classes problem: classification of highly-cited papers and less-cited papers (classes H/L). We also conduct the experiments to classify papers with two extreme citation counts: extremely highly-cited (EH) and extremely less-cited (EL) papers. Distinguishing between two classes was done by relative thresholds defined for each scientific category individually. Two classification methods were investigated for comparison: LDA and kNN. Five different vector representations were tested with binary classifiers. All experiments were also repeated in three spaces constructed:

the Original Space, 13-Dimensional reduced basis and the space constructed after supervised PCA. Our experiments use three categories selected from three branches of science: Applied Mathematics, Biology and Management. The study underlines the importance of vector representations and the spaces to represent texts in prediction citation, and comparison of categories having different characteristics of citation.

We found that the informational semantics is a very promising approach to developing a quantitative evaluation and predictive model of citation count. Our experiments show that the LDA outperform kNN in all spaces constructed for all categories. The results of binary classification illustrate that the highest performance is achieved for extreme citation prediction by LDA in the Original Space: identification of extremely highly cited papers with 83.22% sensitivity and identification of extremely less-cited papers with 81.81% specificity. LDA outperforms kNN for categories, vectors and spaces experimented.

The prediction of citation in Applied Mathematics is inherently not so easy as Biology and Management since the category is heterogeneous, meaning that it contains of papers from different disciplines and citation patterns are different for these individual fields. That is, Applied Mathematics is not well-separated into popular scientific fields and this also means a mixture of category-based semantics. Categories of Biology and Management are more homogeneous inside in terms of papers assigned and have a clear separation into top popular areas. Therefore, these two categories are expected to reach higher classification scores than Applied Mathematics.

A very important finding of the binary classification experiments is that extreme classes are more discriminative than other, achieving higher sensitivity and specificity in all experiments. This can be explained by the fact that two extremes indicate the agreement among scientific community as 'containing information useful for many researchers' and 'being outside of the interest of the community' and the semantics of texts is an important indicator for scientific impact of a paper.

As a result, informational text representations proved to be efficient in the cases where binary classification was performed in distinguishing the impact of articles. We showed that clues – extracted from the importance of words in categories – about the context of a paper provide important information about citation differentiation in scientific articles. This fact is sometimes much more clear for some scientific categories than others.

Multi-class classification of citation through informational semantic has not yet been assessed. We recommend a further and more in-depth research to compare multi-class classification tasks with a combination of different vector representations. Beside the category-based predictors, we also encourage further research to compare mixture of categories.

CHAPTER 7

Conclusion and Discussion

In this thesis, several issues with NLP have been identified and solution have been proposed with regards to giving new insights in computational methods for quantifying the meanings in texts. We have provided a framework for extracting the lexical meaning from short scientific texts. The developed methods are also suitable for a wide range of corpora as the general framework is reproducible for new text domains.

The chapters 2-6 have detailed an empirical approach to automatic short answer grading and feedback systems, our novel approaches to computational modelling and quantifying the lexical meanings in short scientific texts and creation of the *Meaning Space*, and techniques for automated evaluation of the potential impact through semantics of short scientific (or academic) texts. The empirical studies on both short academic answers and a large corpus of short scientific texts allowed for each of these tasks to be understood and investigated qualitatively and quantitatively. This project has also built a large corpus and thesaurus for science, and created software for corpus analysis, producing dictionaries and semantic analysis of texts. Both datasets and software are published for investigation and application to further works.

In short, the main contributions of the thesis can be summarised as follows:

(1) A new direction in automated scoring and feedback systems based on words: a new mathematical model for predicting academic success of students through BoW that students selected to transmit their knowledge of the module and detailed analysis of the methods using a corpus of short answers.
(2) A new perspective in representation of situation behind brief scientific texts: computational analysis of relations between texts massages and representations of situations for large collection of texts by replacing the situation representation with vectors of attributes – a list of scientific subject categories that the text belongs to.
(3) A new approach to quantify the meaning in short scientific texts: defining the meaning of a word by its scientific-specific meaning that is described by multidimensional evaluation of the situation of use presented by categories – vector of RIG about the subject categories that the text belongs to, which can be obtained from observing the word in the text.

(4) A newly introduced vector space, namely *Meaning Space*, in which coordinates correspond to the subject categories: representation of the meaning of words as a vector of RIGs, and an approach to representing text meaning, in a way that built upon the analysis of words (clouds of words) for each text – the information about the words in each cloud.

(5) A new, large collection of scientific texts LSC – 1,673,350 texts from 252 subject categories of WoS – , scientific dictionaries LScD and LScDC, and thesaurus LScT, that are also used for empirical studies presented in the thesis.

(6) A comprehensive statistical analysis of the Meaning Space for the LSC with LScT, exploring the dimensions of the Meaning Space by PCA – identifying the *principal components of the meaning.*

(7) An approach to evaluating the impact of scientific articles through their semantics – classification of papers according to their citation counts, as highly-cited and less-cited papers, in individual categories by several ways.

In the first part of this thesis, we presented methodologies for automated scoring and providing feedback to students in their answers. The focus was to develop automated systems for machines that can reduce the amount of repetitive straightforward scoring while the artificial intelligence is transforming the world. It is noteworthy that we aimed to use approaches to enhance the reliability of human scoring. Our aim was not replacing the human from marking. Instead, human intervention is needed to calibrate the system, and to deal with situations that are challenging.

Computational methods using the keyword-similarity were used in systems for evaluation of the short textual responses. Academic success of students for the module was predicted through the similarities between the student answers and the model answer. Our approaches with the standard BoW model performed well with the corpus of students answers for problems involving clustering of students answers and predicting the marks. We have shown the strong correlation of grades and vocabulary that students used in their answers. This constitutes an attempt to create a clear and easy-to-understand methodology to quantifying category-based (module-specific vocabulary in this case) meaning in short academic texts, even without complex semantic functions.

The viability of the methods was tested in several examples. Our methodology and mathematical model succeed with the automatic scoring of natural language responses where a vocabulary for the model answer is clearly identified. Modelling of marking was easier for some questions than others, but those that are hard to predict grades were also difficult to mark by human. To further improve computational efficiency of the approaches, more intervention (input) from graders is needed. In our experiment, we did not correct spelling mistakes and did not take synonyms into account. In more sophisticated implementation, spell-checking can be adapted to the system and technical dictionary of synonyms can be created to be used as an acceptable alternative terminology.

In the second part of this thesis, we described approaches to quantifying lexical meaning in scientific texts. Our approach has been directed to meet some of main challenges in extracting meanings from texts. First, it solves the problem of extracting the scientific-specific meanings because proposed models of informational semantics characterise the situation of use by the subject categories of the text. Second, words has good representation for individual categories as well as the entire corpus because the relativeness of importance of a word across scientific categories is taken into account. Third, creation of the space to represent words and texts is automated and reproducible so that it does not require huge amount of intervention from human.

The thesis has introduced an informational space of meaning for short scientific texts. Novel techniques for quantifying the meaning were developed and implemented on the basis of LSC with LScT. For concreteness, we followed the road: Corpus of texts + categories → Meaning Space for words → Geometric representation of the meaning of texts. This involved the representation of words, representation of text in the constructed Meaning Space and detailed analysis of the Meaning Space. In our case study, we employed very simple attributes for description of the text usage situation, the research subject categories of the text. We conclude that the use of informational semantics provides sizeable improvements to represent meaning in scientific texts over classical text representation approaches based on raw frequencies, but how to make best use of it in different NLP tasks remains an open question that deserves further investigation. The list of attributes can be modified and extended. The level of detail of the Meaning Space can vary greatly within the framework of the proposed approach.

Several directions in this research hold promise for making the computational models better representing the meaning. As the next step, it is reasonable to use different combinations of FVTs schemes for improving the representativeness. This does not require any major modification of the approach. Another focus of the research could be the study of more complex models in which co-occurrence of words and combination of word's meaning will be used.

This thesis has also introduced and analysed scientific corpus, dictionaries and thesaurus. In the creation of the thesaurus, we have focused on the most informative words in science, which are main scientific content words. Of course, the analysis of dictionaries has not finalized. For example, the frequent but non-informative words (e.g. 'use') can be considered as generalised service words of Science and deserve special analysis. It is also very desirable to extend the set of attributes for representation of the situation behind the text. The first choice, the research subject categories, is simple and natural, but it may be useful to enrich this list of attributes.

Of course, much of the journey in semantic analysis awaits. We hope that our approaches of informational semantics will be applied to other corpora, and the framework will be extended to other languages where improvements in meaning extraction can be done.

Finally, we used informational semantics extracted by our methodology to evaluate the potential impact of articles. To measure the impact of a paper, we used citation counts which is a very standard and universal metric in scientific community. Automatic prediction of citation counts could provide a powerful new method for evaluating articles to faster identification of promising articles and dissemination of new knowledge in science. Accurate models for predicting citation by semantics can also improve our understanding of the contents that influence citations. This task is however a very hard problem because of the nature and dynamics of citation. There are usually many other factors linked to citation such as authors and journals.

In this research, we build computer models that classify highly-cited and less-cited LSC texts with citation counts extracted from the WoS website for approximately 4 years range from 2014. Therefore, we deal with binary classification throughout this research. Our experiments show that it is indeed feasible to accurately predict the impact using machine learning methods when semantics information is known. The classification models have been tested for three individual categories selected from three main branches of science. The models pave the way for classification of the impact of publications, and the category-based analysis provides better insight into citation behaviour for different subject categories. Having promising results for classification, modelling the citation counts deserves further investigation for other categories and mixture of categories. In addition to this, the models can be tested and validated with new papers cited in the one year forward from the most recent year or the next 4 years from the most recent year of experienced time.

Another way to side step the problem of citation prediction would be multi-class classification of citation through informational semantic as this task has not yet been assessed. We recommend a further and more in-depth research to compare multi-class classification tasks with a combination of different vector representations. In addition to these investigations in classification, direct prediction of the citation is also known to be a very hard mathematical challenge and the classifiers established here can help further studies in this area of research.

The corpus of scientific texts is a dynamic resource due to changes and rapid evaluation of research through new topics and priorities in science. This leads to a continuous analysis of the corpus and impact of articles. In order to update the corpus and models and incorporate a new stream of the data with citation counts for a certain n year, we can shift the evaluation window one year (or m year) forward. To do this, one can drop all articles published in the earliest (m) year and add newly published articles (in WoS) in the next (m) year of the most recent year of time window. This procedure can be automated for any n-year period.

APPENDIX A

Appendix of Chapter 2: Additional Questions and Feedback Processes of the System

A.1. Questions from the University of North Texas Study

The questions in this appendix are a sample of those from the *introductory computer science class* in the University of North Texas. In Questions 1 and 2 we see that the teachers scores are more consistent and this means machine scoring such questions is more reliable. Questions 3 and 4 have more diverse answers and these are harder to score automatically.

TABLE A.1. Example of a question for which we could reliably automate marking (Question 1)

Question 1	What is the role of a prototype program in problem solving?
Model Answer 1	To simulate the behaviour of portions of the desired software product.
Model Vocabulary	simulate, behaviour, portion, desire, software, product.
Student Answer	High risk problems are address in the prototype program to make sure that the program is feasible. A prototype may also be used to show a company that the **software** can be possibly programmed.
Student Answer	it **simulates** the **behavior** of **portions** of the **desired software product**

TABLE A.2. Teacher marks for the answers to Question 1.

	Teacher 1 grade	Teacher 2 grade	Average
1st student	4	3	3.5
2nd student	5	5	5

TABLE A.3. Example of a question for which we could reliably automate marking (Question 2)

Question 2	How does the compiler handle inline functions?
Model Answer 2	It makes a copy of the function code in every place where a function call is made.
Model vocabulary	Make, copy, function, code, every, place, call, made.
Student answer	For inline functions, the compiler creates a **copy** of the **function's code** in **place** so it doesn't have to **make** a function **call** and add to the function call stack
Student answer	It expands a small **function** out... **making** your **code** longer, but also makes it run faster.
Student answer	The compiler can ignore the inline qualifier and typically does so for all but the smallest **functions**.

TABLE A.4. Teacher marks for the answers to Question 2

	Teacher 1 grade	Teacher 2 grade	Average
1st student	5	5	5
2nd student	4	4	4
3rd student	2	2	2

TABLE A.5. Question which is harder to assess (Question 3)

Question 3	How many dimensions need to be specified when passing a multi-dimensional array as an argument to a function?
Model Answer 3	All the dimensions, except the first one.
Model vocabulary	Dimension, except, first, one.
Student answer	All **dimensions except** for the **first one** need to be specified when passing an array to a function, the compiler needs to know how many memory addresses to skip to make it back to the 2nd element in the first dimension. The size of the first dimension does not need to be specified.
Student answer	All **dimensions**, excluding the **first one**.
Student answer	None, just pass the array name.
Student answer	All of the **dimensions** must be specified.
Student answer	At least 2, depending on how many arrays are being used.

TABLE A.6. Teacher marks for the answers to Question 3

	Teacher 1 grade	Teacher 2 grade	Average
1st student	5	5	5
2nd student	5	5	5
3rd student	3	1	2
4th Student	5	2	3.5
5th Student	4	1	2.5

TABLE A.7. Question which is harder to assess (Question 4)

Question 4	What stages in the software life cycle are influenced by the testing stage?
Model Answer 4	The testing stage can influence both the coding stage (phase 5) and the solution refinement stage (phase 7).
Model vocabulary	Test, stage, can, influence, code, phase, solute, refine.
Student answer	The implementation **phase** and the maintenance phase are effected.
Student answer	Elaboration, Construction, and Transition are all affected by **testing**.
Student answer	**Coding** and **refining**.
Student answer	1- specification 2- design 3- risk analysis 4- verification 5- **coding** 6- **testing** 7- **refining** 8- production 9- maintenance.

TABLE A.8. Teacher marks for the answers to Question 4

	Teacher 1 grade	Teacher 2 grade	Average
1st student	5	3	4
2nd student	2	2	2
3rd student	5	5	5
4th Student	4	1	2.5

A.2. Feedback Processes of the System

This section contains algorithms of feedback creation.

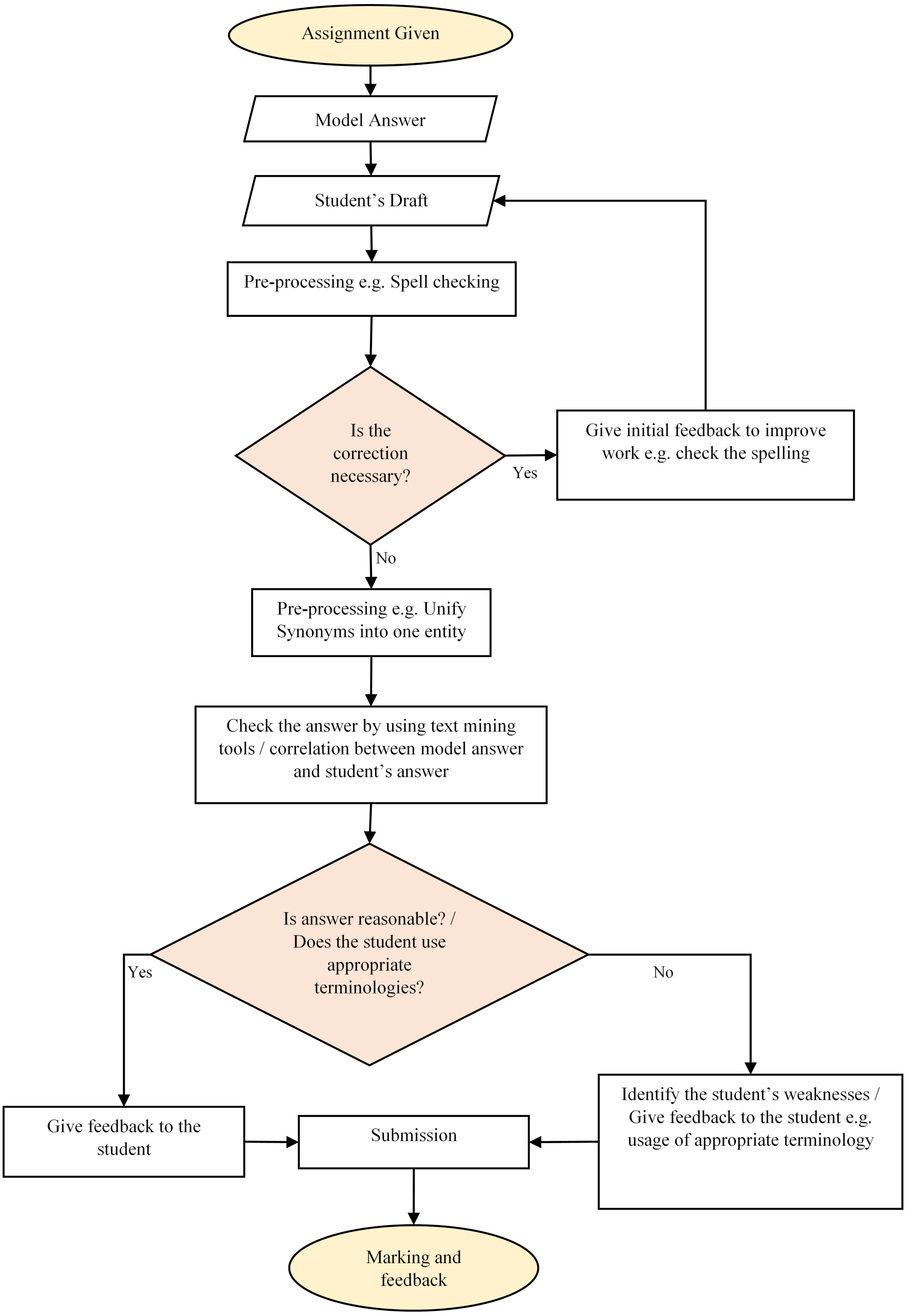


Figure A.1. Feedback process by similarity between the model answer and students' answer

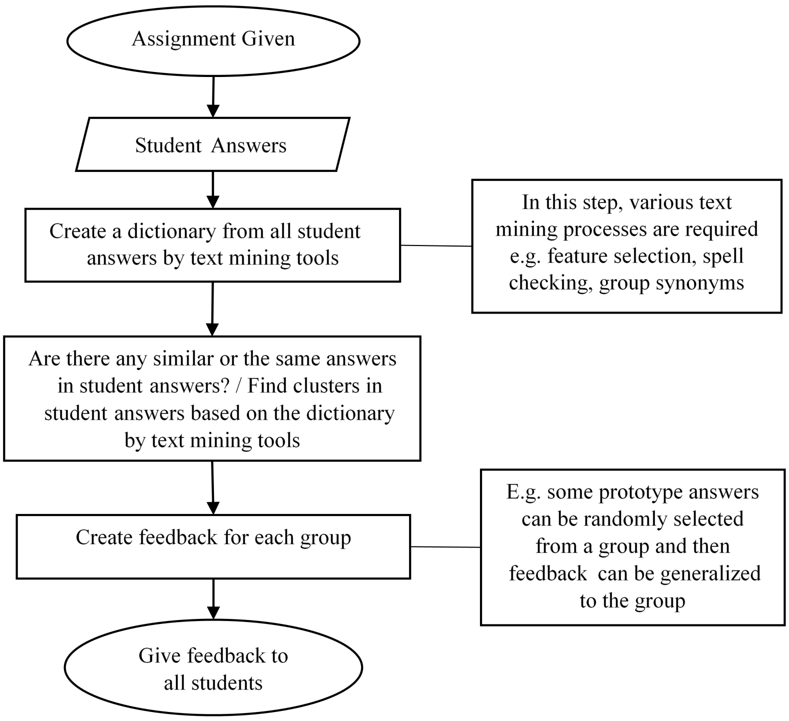


FIGURE A.2. Feedback process to group of students' answers by clustering approach

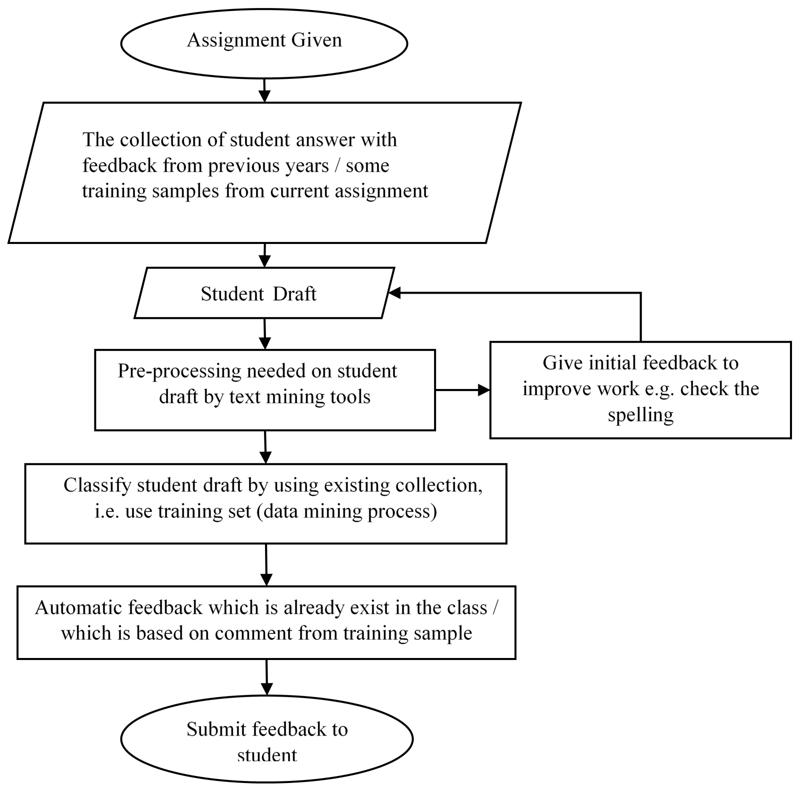


FIGURE A.3. Supervised process for feedback and marking in the case of existing train data

APPENDIX B

Appendix of Chapter 3: Lists of Headings and Stop Words in Pre-processing Steps, and Lists of Categories and Research Areas in the Data

B.1. Table of Headings of Sections in Medical Abstracts

Table B.1. Headings of sections identified in structured abstracts (Headings can be either singular or plural)

Headings of Sections					
Abstract	Aim	Approach	Background	Conclusion	Design
Discussion	Finding	Hypothesis	Introduction	Limitation	Location
Material	Measure	Measurement	Method	Methodology	Objective
Patient	Population	Procedure	Process	Purpose	Rationale
Result	Setting	Subject	Theoretical		
Implication(s) for health and nursing policy					

B.2. An Example of Document Structure in the LSC

TABLE B.2. Structure of a document in the LSC

Authors	**Title**	**Abstract**	**Categories**	**Research Areas**	**Total Times Cited**	**Times Cited in CC**
Cheng, JS; Craft, R; Yu, GQ; Ho, K; Wang, X; Mohan, G; Mangnitsky, S; Ponnusamy, R; Mucke, L	Tau Reduction Diminishes Spatial Learning and Memory Deficits after Mild Repetitive Traumatic Brain Injury in Mice	Objective: Because reduction of the microtubule-associated protein Tau has beneficial effects in mouse models of Alzheimer's disease and epilepsy, we wanted to determine whether this strategy can also improve the outcome of mild traumatic brain injury (TBI). **...(truncated)**	Multidisciplinary Sciences	Science & Technology - Other Topics	24	24

B.3. List of Prefixes

Table B.3. The List of Prefixes

Prefixes						
anti-	ante-	auto-	co-	de-	deca-	di-
dia-	dis-	e-	ex-	extra-	fore-	hemi-
hexa-	hepta-	homo-	hyper-	in-	inter-	im-
ir-	kilo-	micro-	mid-	milli-	mis-	mono-
multi-	non-	octo-	over-	para-	penta-	per-
poly-	post-	pre-	pro-	quadri-	re-	retro-
self-	semi-	sub-	super-	tele-	tetra-	therm-
trans-	tri-	ultra-	un-	under-	uni-	

B.4. List of Substitutes

Table B.4. List of Substitution

Word	**Substitute**
well-known	wellknown
z-test	ztest
z-testing	ztest
z-tests	ztest
z-score	zscore
z-scored	zscored
z-scores	zscore
p-value	pvalue
p-values	pvalue
p-valued	pvalue
p-valuesof	pvalue
chi-square	chisquare
chi-squares	chisquare
chi-squared	chisquared
chi2-test	chisquared

B.5. List of Stop Words in "tm" Package (R package)

TABLE B.5. The List of Stop Words

Stop Words in 'tm' Package							
i	me	my	myself	we	our	ours	own
yours	a	here	he	him	his	himself	she
herself	it	its	itself	they	them	their	theirs
which	who	whom	this	that	these	those	am
was	were	be	been	being	have	has	had
does	did	doing	would	should	could	ought	i'm
she's	it's	we're	they're	i've	you've	we've	they've
he'd	she'd	we'd	they'd	i'll	you'll	he'll	she'll
isn't	aren't	wasn't	weren't	hasn't	haven't	hadn't	doesn't
won't	i'd	shan't	shouldn't	can't	cannot	couldn't	mustn't
who's	what's	here's	there's	when's	where's	why's	how's
the	and	but	if	or	because	as	until
at	by	for	with	about	against	between	into
before	after	above	below	to	from	up	down
on	off	over	under	again	further	then	once
when	where	why	how	all	any	both	each
most	other	some	such	no	nor	not	only
you	your	her	hers	themselves	what	is	are
having	do	you're	he's	wouldn't	you'd	we'll	they'll
don't	didn't	let's	that's	yourself	an	while	of
through	during	in	out	yourselves	there	few	more
so	than	too	very	ourselves	same		

APPENDIX C

Appendix of Chapter 4: Additional Examples of Notices, Lists of Categories and Research Areas, Some Examples of Word Clouds and Histograms

C.1. Some Additional Examples of Notices

Table C.1. Some additional examples of notices attached to the abstract

Copyright Notice; Name of Conference, Journal or Publishing House
(C) 2014 Elsevier B.V. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/3.0/).
Copyright (C) 2014, Taiwan Association of Obstetrics & Gynecology. Published by Elsevier Taiwan LLC. All rights reserved.
(c) 2014 AIP Publishing LLC.
(C) The Author(s) 2014. Published by ECS. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 License (CC BY, http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse of the work in any medium, provided the original work is properly cited. All rights reserved.
(C) 2014 Mosby, Inc. All rights reserved.
(C) 2014 Society of Photo-Optical Instrumentation Engineers (SPIE) & Wilkins
(C) 2014 Elsevier Inc. All rights reserved.
Copyright (C) 2014, Hydrogen Energy Publications, LLC. Published by Elsevier Ltd. All rights reserved
(C) 2014 Optical Society of America
(C) 2014 AACR.
(C) 2014 S. Karger AG, Basel
(C) 2014 American Society of Civil Engineers.
(C) 2014 Wiley Periodicals, Inc. and the American Pharmacists Association.
(C) 2014 Elsevier Ltd and Techna Group S.r.l. All rights reserved.
J. surg. oncol
Developmental Dynamics
Proteins 2014
Bioelectromagnetics
am. J. Hematol

C.2. List of Categories

Table C.2. The list of 252 WoS categories with the number of LSC documents assigned to the corresponding category

No.	Category	Number of Documents
1	Engineering, Electrical & Electronic	174,272
2	Materials Science, Multidisciplinary	112,912
3	Physics, Applied	78,796
4	Chemistry, Physical	58,065
5	Chemistry, Multidisciplinary	55,907
6	Computer Science, Theory & Methods	55,591
7	Multidisciplinary Sciences	53,140
8	Engineering, Mechanical	50,972
9	Optics	47,737
10	Biochemistry & Molecular Biology	47,490
11	Computer Science, Information Systems	45,865
12	Energy & Fuels	44,202
13	Environmental Sciences	42,082
14	Computer Science, Artificial Intelligence	41,210
15	Telecommunications	40,550
16	Nanoscience & Nanotechnology	35,050
17	Oncology	34,339
18	Mechanics	33,545
19	Neurosciences	32,972
20	Surgery	30,805
21	Pharmacology & Pharmacy	30,713
22	Automation & Control Systems	29,427
23	Engineering, Chemical	29,171
24	Computer Science, Interdisciplinary Applications	29,153
25	Mathematics, Applied	27,982
26	Physics, Condensed Matter	27,316
27	Biotechnology & Applied Microbiology	26,286
28	Public, Environmental & Occupational Health	25,493
29	Mathematics	25,450
30	Geosciences, Multidisciplinary	24,644
31	Cell Biology	23,108
32	Physics, Multidisciplinary	22,930
33	Astronomy & Astrophysics	22,825

No.	Category	Number of Documents
34	Economics	22,338
35	Clinical Neurology	22,127
36	Engineering, Civil	22,127
37	Chemistry, Analytical	21,490
38	Plant Sciences	21,321
39	Engineering, Multidisciplinary	21,144
40	Radiology, Nuclear Medicine & Medical Imaging	21,014
41	Food Science & Technology	20,414
42	Education & Educational Research	20,087
43	Medicine, Research & Experimental	19,744
44	Genetics & Heredity	19,512
45	Computer Science, Hardware & Architecture	18,489
46	Immunology	18,270
47	Polymer Science	18,017
48	Chemistry, Organic	17,941
49	Engineering, Biomedical	17,786
50	Microbiology	17,252
51	Computer Science, Software Engineering	17,104
52	Instruments & Instrumentation	17,090
53	Physics, Atomic, Molecular & Chemical	17,010
54	Metallurgy & Metallurgical Engineering	16,898
55	Ecology	16,760
56	Cardiac & Cardiovascular Systems	16,369
57	Medicine, General & Internal	16,179
58	Psychiatry	16,055
59	Electrochemistry	15,663
60	Biochemical Research Methods	15,050
61	Endocrinology & Metabolism	14,622
62	Engineering, Environmental	14,614
63	Management	14,339
64	Chemistry, Applied	14,058
65	Water Resources	13,997
66	Thermodynamics	13,852
67	Pediatrics	13,364
68	Physics, Particles & Fields	13,203
69	Engineering, Manufacturing	13,102
70	Biophysics	12,630

No.	Category	Number of Documents
71	Chemistry, Inorganic & Nuclear	12,591
72	Infectious Diseases	12,521
73	Chemistry, Medicinal	12,456
74	Meteorology & Atmospheric Sciences	12,318
75	Construction & Building Technology	12,078
76	Operations Research & Management Science	11,879
77	Veterinary Sciences	11,502
78	Remote Sensing	11,388
79	Nuclear Science & Technology	11,359
80	Zoology	11,218
81	Social Sciences, Interdisciplinary	11,035
82	Gastroenterology & Hepatology	10,943
83	Orthopedics	10,538
84	Physics, Mathematical	10,426
85	Engineering, Industrial	10,362
86	Marine & Freshwater Biology	10,124
87	Mathematics, Interdisciplinary Applications	10,072
88	Geochemistry & Geophysics	10,023
89	Biology	9,917
90	Obstetrics & Gynecology	9,883
91	Physics, Fluids & Plasmas	9,704
92	Toxicology	9,613
93	Statistics & Probability	9,532
94	Nutrition & Dietetics	9,416
95	Business	9,394
96	Imaging Science & Photographic Technology	9,353
97	Hematology	9,096
98	Physiology	9,009
99	Peripheral Vascular Disease	8,700
100	Agronomy	8,651
101	Dentistry, Oral Surgery & Medicine	8,502
102	Robotics	8,491
103	Transportation Science & Technology	8,411
104	Sport Sciences	8,368
105	Psychology, Multidisciplinary	8,332
106	Urology & Nephrology	8,264
107	Materials Science, Biomaterials	8,040

No.	Category	Number of Documents
108	Mathematical & Computational Biology	8,015
109	Health Care Sciences & Services	7,999
110	Physics, Nuclear	7,876
111	Ophthalmology	7,830
112	Environmental Studies	7,811
113	Rehabilitation	7,791
114	Respiratory System	7,666
115	Oceanography	7,417
116	Spectroscopy	7,388
117	Materials Science, Coatings & Films	7,226
118	Pathology	7,217
119	Business, Finance	7,214
120	Psychology	6,989
121	Acoustics	6,935
122	Crystallography	6,932
123	Psychology, Clinical	6,860
124	Geography, Physical	6,806
125	Psychology, Experimental	6,784
126	Nursing	6,637
127	Green & Sustainable Science & Technology	6,412
128	Agriculture, Multidisciplinary	6,406
129	Education, Scientific Disciplines	6,308
130	Virology	6,270
131	Materials Science, Ceramics	6,222
132	Agriculture, Dairy & Animal Science	6,163
133	Behavioral Sciences	5,922
134	Linguistics	5,921
135	Dermatology	5,793
136	Evolutionary Biology	5,742
137	Entomology	5,704
138	Parasitology	5,683
139	Horticulture	5,338
140	Health Policy & Services	5,318
141	Language & Linguistics	5,174
142	Political Science	5,106
143	Soil Science	4,800
144	Otorhinolaryngology	4,797

No.	Category	Number of Documents
145	Geriatrics & Gerontology	4,742
146	Sociology	4,725
147	Biodiversity Conservation	4,705
148	Fisheries	4,702
149	Engineering, Geological	4,573
150	Information Science & Library Science	4,565
151	Forestry	4,472
152	Engineering, Aerospace	4,435
153	Psychology, Developmental	4,390
154	Materials Science, Composites	4,277
155	Planning & Development	4,115
156	Transplantation	4,105
157	Transportation	4,035
158	Medical Informatics	3,991
159	Reproductive Biology	3,984
160	Critical Care Medicine	3,982
161	Rheumatology	3,942
162	Geography	3,908
163	Materials Science, Characterization & Testing	3,878
164	Agricultural Engineering	3,727
165	Tropical Medicine	3,696
166	Philosophy	3,657
167	Computer Science, Cybernetics	3,652
168	Developmental Biology	3,593
169	Law	3,574
170	Psychology, Social	3,548
171	Psychology, Applied	3,523
172	Social Sciences, Mathematical Methods	3,496
173	History	3,487
174	Integrative & Complementary Medicine	3,453
175	Substance Abuse	3,433
176	Communication	3,200
177	Anthropology	3,149
178	Social Sciences, Biomedical	3,003
179	Hospitality, Leisure, Sport & Tourism	2,998
180	Anesthesiology	2,943
181	International Relations	2,941

No.	Category	Number of Documents
182	Neuroimaging	2,702
183	Mining & Mineral Processing	2,687
184	Emergency Medicine	2,627
185	Medical Laboratory Technology	2,598
186	Humanities, Multidisciplinary	2,559
187	Mineralogy	2,550
188	Materials Science, Textiles	2,548
189	Gerontology	2,531
190	Paleontology	2,503
191	Cell & Tissue Engineering	2,455
192	Engineering, Ocean	2,352
193	Religion	2,335
194	Urban Studies	2,309
195	Family Studies	2,229
196	Public Administration	2,204
197	History & Philosophy Of Science	2,199
198	Geology	2,153
199	Archaeology	2,118
200	Social Work	2,114
201	Psychology, Educational	2,112
202	Engineering, Marine	2,110
203	Audiology & Speech-Language Pathology	2,052
204	Area Studies	2,046
205	Criminology & Penology	2,015
206	Materials Science, Paper & Wood	1,963
207	Limnology	1,941
208	Engineering, Petroleum	1,930
209	Ethics	1,928
210	Anatomy & Morphology	1,889
211	Mycology	1,829
212	Logic	1,786
213	Allergy	1,765
214	Medicine, Legal	1,711
215	Education, Special	1,666
216	Literature	1,608
217	Psychology, Biological	1,527
218	Ergonomics	1,431

No.	Category	Number of Documents
219	Architecture	1,376
220	Women's Studies	1,341
221	Microscopy	1,319
222	Social Issues	1,296
223	Primary Health Care	1,269
224	Ornithology	1,008
225	Demography	948
226	Cultural Studies	945
227	Music	888
228	Agricultural Economics & Policy	880
229	History Of Social Sciences	879
230	Industrial Relations & Labor	879
231	Asian Studies	877
232	Art	725
233	Ethnic Studies	675
234	Medical Ethics	674
235	Psychology, Mathematical	538
236	Literary Theory & Criticism	498
237	Medieval & Renaissance Studies	485
238	Film, Radio, Television	398
239	Andrology	391
240	Psychology, Psychoanalysis	345
241	Classics	325
242	Theater	300
243	Literature, Romance	269
244	Literature, British Isles	220
245	Folklore	134
246	Literature, German, Dutch, Scandinavian	128
247	Literature, American	75
248	Dance	74
249	Literature, African, Australian, Canadian	59
250	Poetry	42
251	Literary Reviews	35
252	Literature, Slavic	35

C.3. List of Research Areas

Table C.3. The list of 151 WoS research areas with the number of LSC documents assigned to the corresponding research area

No.	Research Area	Number of Documents
1	Engineering	328,136
2	Chemistry	162,934
3	Physics	158,438
4	Computer Science	142,633
5	Materials Science	141,754
6	Science & Technology - Other Topics	96,388
7	Environmental Sciences & Ecology	60,657
8	Biochemistry & Molecular Biology	60,027
9	Mathematics	59,525
10	Neurosciences & Neurology	48,680
11	Optics	47,737
12	Energy & Fuels	44,202
13	Business & Economics	40,743
14	Telecommunications	40,550
15	Pharmacology & Pharmacy	38,837
16	Psychology	36,282
17	Oncology	34,339
18	Mechanics	33,545
19	Agriculture	31,191
20	Surgery	30,805
21	Automation & Control Systems	29,427
22	Geology	26,632
23	Biotechnology & Applied Microbiology	26,286
24	Education & Educational Research	25,924
25	Public, Environmental & Occupational Health	25,493
26	Cell Biology	24,145
27	Cardiovascular System & Cardiology	23,396
28	Astronomy & Astrophysics	22,825
29	Plant Sciences	21,321
30	Radiology, Nuclear Medicine & Medical Imaging	21,014
31	Food Science & Technology	20,414
32	General & Internal Medicine	20,409
33	Research & Experimental Medicine	19,744

No.	Research Area	Number of Documents
34	Genetics & Heredity	19,512
35	Immunology	18,270
36	Polymer Science	18,017
37	Microbiology	17,252
38	Instruments & Instrumentation	17,090
39	Metallurgy & Metallurgical Engineering	16,898
40	Social Sciences - Other Topics	16,666
41	Psychiatry	16,055
42	Electrochemistry	15,663
43	Endocrinology & Metabolism	15,013
44	Water Resources	13,997
45	Thermodynamics	13,852
46	Pediatrics	13,365
47	Biophysics	12,630
48	Infectious Diseases	12,521
49	Meteorology & Atmospheric Sciences	12,318
50	Zoology	12,200
51	Construction & Building Technology	12,078
52	Operations Research & Management Science	11,879
53	Marine & Freshwater Biology	11,562
54	Veterinary Sciences	11,502
55	Remote Sensing	11,388
56	Nuclear Science & Technology	11,359
57	Gastroenterology & Hepatology	10,943
58	Orthopedics	10,538
59	Transportation	10,280
60	Health Care Sciences & Services	10,243
61	Geochemistry & Geophysics	10,023
62	Life Sciences & Biomedicine - Other Topics	9,917
63	Obstetrics & Gynecology	9,883
64	Toxicology	9,613
65	Nutrition & Dietetics	9,416
66	Imaging Science & Photographic Technology	9,353
67	Hematology	9,096
68	Physiology	9,009
69	Dentistry, Oral Surgery & Medicine	8,502
70	Government & Law	8,492

No.	Research Area	Number of Documents
71	Robotics	8,491
72	Sport Sciences	8,368
73	Urology & Nephrology	8,264
74	Mathematical & Computational Biology	8,015
75	Ophthalmology	7,830
76	Rehabilitation	7,791
77	Respiratory System	7,666
78	Oceanography	7,417
79	Spectroscopy	7,388
80	Pathology	7,217
81	Linguistics	7,077
82	Acoustics	6,935
83	Crystallography	6,932
84	Physical Geography	6,806
85	Nursing	6,637
86	Virology	6,270
87	Public Administration	6,120
88	Behavioral Sciences	5,922
89	Dermatology	5,793
90	Evolutionary Biology	5,742
91	Entomology	5,704
92	Parasitology	5,683
93	Geriatrics & Gerontology	5,505
94	Otorhinolaryngology	4,797
95	Sociology	4,725
96	Biodiversity & Conservation	4,705
97	Fisheries	4,702
98	Information Science & Library Science	4,565
99	Forestry	4,472
100	Transplantation	4,105
101	Medical Informatics	3,991
102	Reproductive Biology	3,984
103	Rheumatology	3,942
104	Geography	3,908
105	Tropical Medicine	3,696
106	Philosophy	3,657
107	Developmental Biology	3,593

No.	Research Area	Number of Documents
108	Mathematical Methods In Social Sciences	3,496
109	History	3,487
110	Integrative & Complementary Medicine	3,453
111	Substance Abuse	3,433
112	Communication	3,200
113	Arts & Humanities - Other Topics	3,178
114	Anthropology	3,149
115	Biomedical Social Sciences	3,003
116	Anesthesiology	2,943
117	International Relations	2,941
118	Literature	2,735
119	Mining & Mineral Processing	2,687
120	Emergency Medicine	2,627
121	Medical Laboratory Technology	2,598
122	Mineralogy	2,550
123	Paleontology	2,503
124	Religion	2,335
125	Urban Studies	2,309
126	Family Studies	2,229
127	History & Philosophy Of Science	2,199
128	Archaeology	2,118
129	Social Work	2,114
130	Audiology & Speech-Language Pathology	2,052
131	Area Studies	2,046
132	Criminology & Penology	2,015
133	Anatomy & Morphology	1,889
134	Mycology	1,829
135	Allergy	1,765
136	Legal Medicine	1,711
137	Architecture	1,376
138	Women's Studies	1,341
139	Microscopy	1,319
140	Social Issues	1,296
141	Demography	948
142	Cultural Studies	945
143	Music	888
144	Asian Studies	877

No.	Research Area	Number of Documents
145	Art	725
146	Ethnic Studies	675
147	Medical Ethics	674
148	Film, Radio, Television	398
149	Classics	325
150	Theater	300
151	Dance	74

C.4. Some Word Clouds and Histograms for Categories

Word clouds presenting the top 100 words ordered by their RIGs and histograms of RIGs for the first 10 words in the word clouds for 252 categories in LSC.

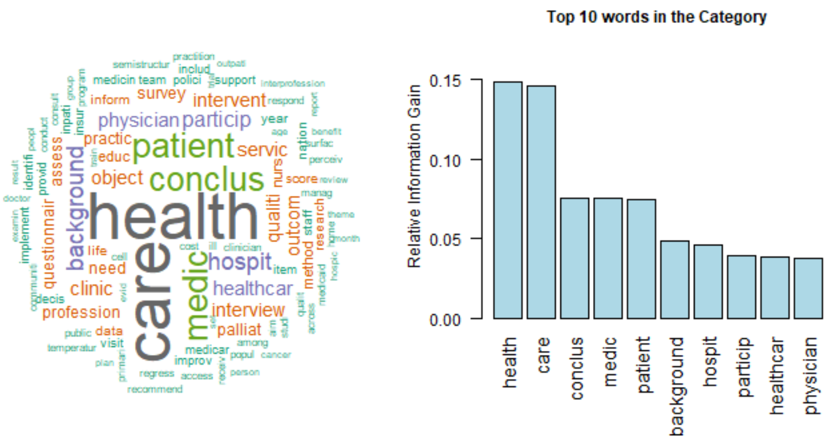


FIGURE C.1. Health Care Sciences & Services

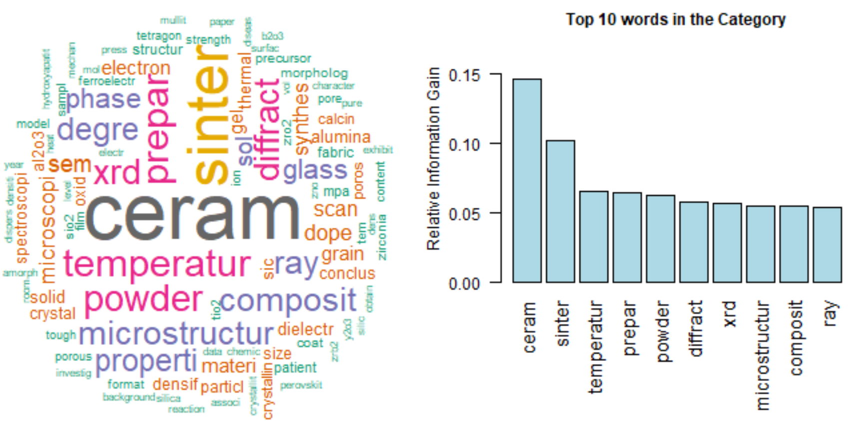


FIGURE C.2. Material Science, Ceramics

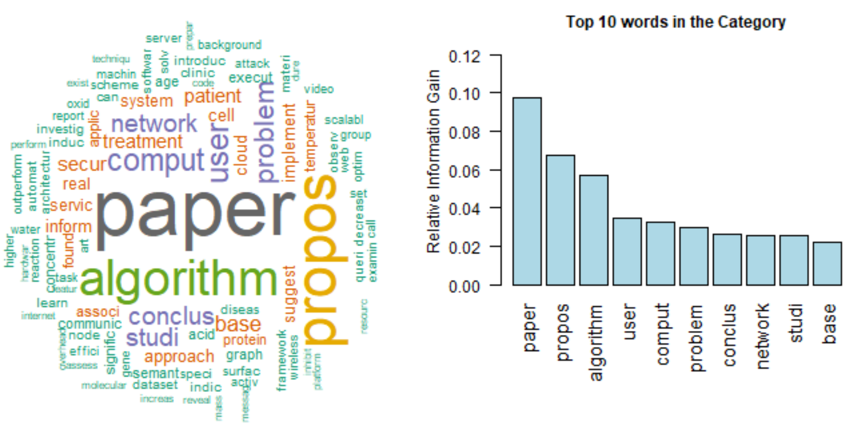


FIGURE C.3. Computer Science, Theory & Methods

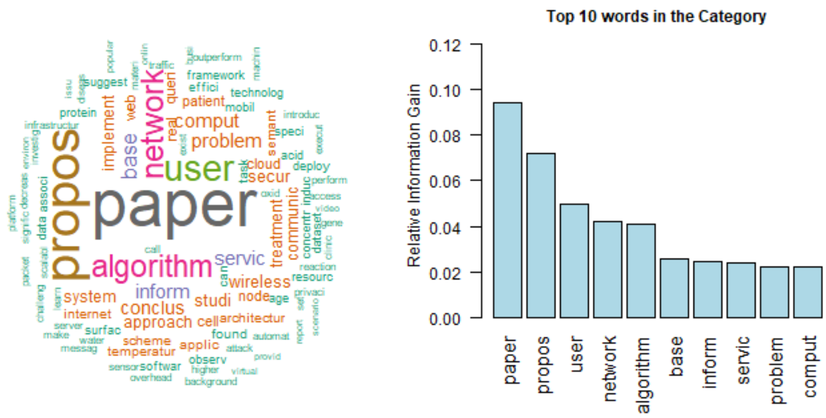


FIGURE C.4. Computer Science, Information Systems

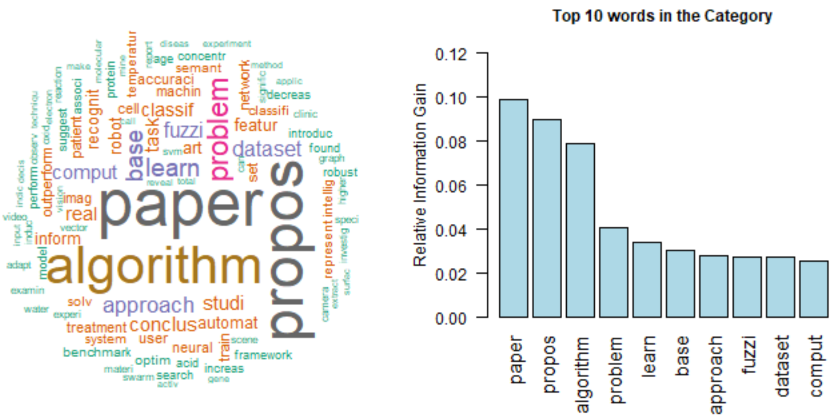


FIGURE C.5. Computer Science, Artificial Intelligence

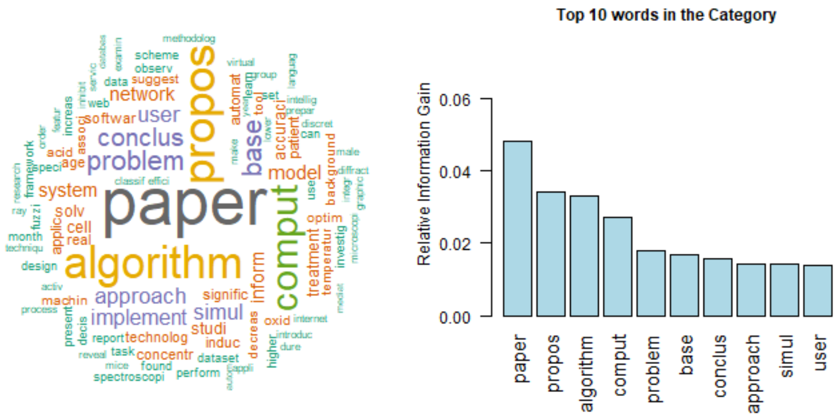


FIGURE C.6. Computer Science, Interdisciplinary Applications

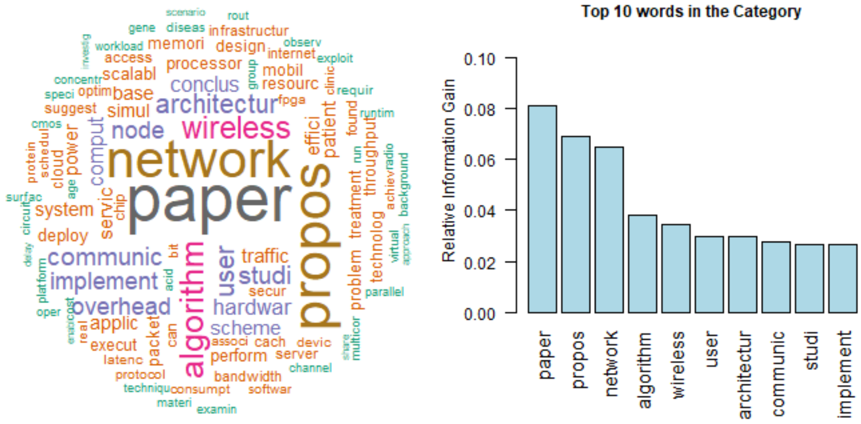


FIGURE C.7. Computer Science, Hardware & Architecture

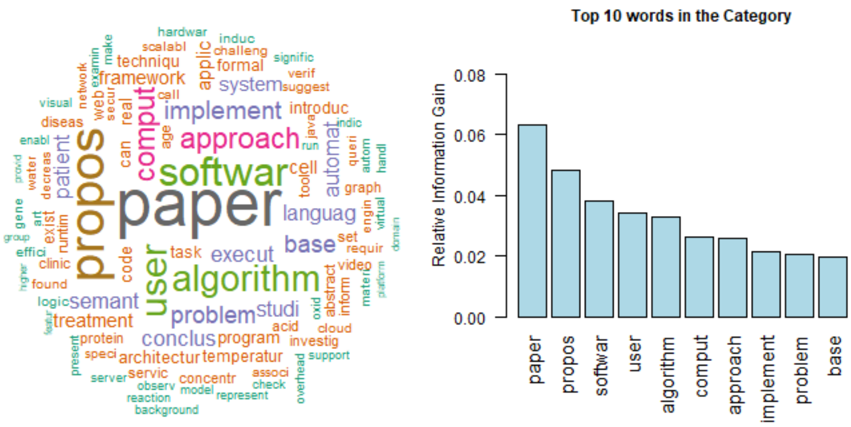


FIGURE C.8. Computer Science, Software Engineering

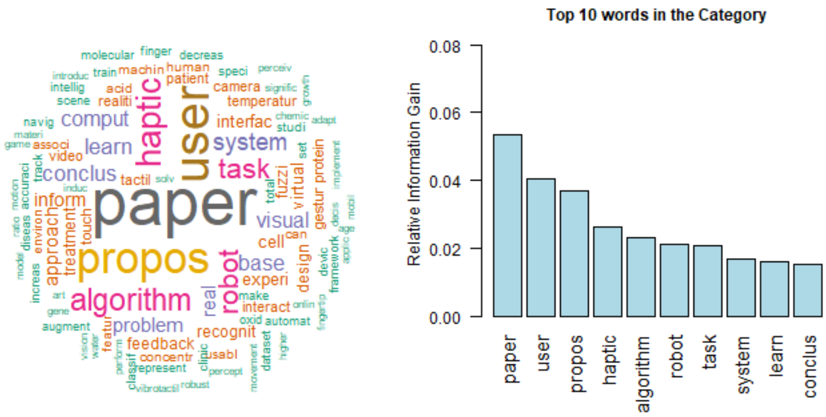


FIGURE C.9. Computer Science, Cybernetics

APPENDIX D

Appendix of Chapter 5: Tables for Principal Component Analysis

D.1. The Most Informative 100 Words of the LScT in Categories with Their RIGs

Table D.1. The list of the top 100 words in the category Acoustics with RIGs

No.	Word	RIG	No.	Word	RIG
1	acoust	9.4×10^{-2}	51	recognit	5.9×10^{-3}
2	ultrasound	6.2×10^{-2}	52	activ	5.7×10^{-3}
3	sound	4.1×10^{-2}	53	modal	5.6×10^{-3}
4	speech	4.1×10^{-2}	54	gaussian	5.5×10^{-3}
5	frequenc	4×10^{-2}	55	simul	5.4×10^{-3}
6	nois	3.3×10^{-2}	56	accuraci	5.3×10^{-3}
7	ultrason	2.7×10^{-2}	57	model	5.2×10^{-3}
8	wave	2.4×10^{-2}	58	experiment	5×10^{-3}
9	transduc	2.3×10^{-2}	59	acid	5×10^{-3}
10	vibrat	2.2×10^{-2}	60	impuls	5×10^{-3}
11	signal	2×10^{-2}	61	techniqu	5×10^{-3}
12	propos	1.8×10^{-2}	62	puls	4.9×10^{-3}
13	audio	1.3×10^{-2}	63	approach	4.9×10^{-3}
14	speaker	1.3×10^{-2}	64	perceptu	4.9×10^{-3}
15	listen	1.2×10^{-2}	65	voic	4.9×10^{-3}
16	paper	1.2×10^{-2}	66	year	4.8×10^{-3}
17	propag	1.2×10^{-2}	67	outperform	4.8×10^{-3}
18	mhz	1.2×10^{-2}	68	beam	4.7×10^{-3}
19	algorithm	1.1×10^{-2}	69	displac	4.6×10^{-3}
20	filter	1.1×10^{-2}	70	regul	4.4×10^{-3}
21	amplitud	1×10^{-2}	71	speci	4.4×10^{-3}
22	khz	1×10^{-2}	72	paramet	4.4×10^{-3}
23	estim	9.9×10^{-3}	73	reson	4.4×10^{-3}
24	imag	9.8×10^{-3}	74	transmit	4.3×10^{-3}
25	numer	8.8×10^{-3}	75	shear	4.3×10^{-3}
26	method	8.7×10^{-3}	76	utter	4.3×10^{-3}
27	phantom	8.5×10^{-3}	77	vocal	4.3×10^{-3}
28	echo	7.9×10^{-3}	78	hear	4.3×10^{-3}
29	mode	7.9×10^{-3}	79	snr	4.2×10^{-3}
30	piezoelectr	7.8×10^{-3}	80	fetal	4.2×10^{-3}
31	elast	7.6×10^{-3}	81	problem	4.2×10^{-3}
32	perform	7.4×10^{-3}	82	associ	4.1×10^{-3}
33	array	7.2×10^{-3}	83	measur	4.1×10^{-3}
34	damp	7.2×10^{-3}	84	automat	4.1×10^{-3}
35	error	7×10^{-3}	85	fetus	4.1×10^{-3}
36	harmon	6.9×10^{-3}	86	inhibit	4.1×10^{-3}
37	excit	6.9×10^{-3}	87	role	4×10^{-3}
38	protein	6.9×10^{-3}	88	auditori	4×10^{-3}
39	doppler	6.8×10^{-3}	89	sourc	4×10^{-3}
40	cell	6.7×10^{-3}	90	motion	4×10^{-3}
41	nonlinear	6.6×10^{-3}	91	molecular	3.9×10^{-3}
42	finit	6.6×10^{-3}	92	comput	3.9×10^{-3}
43	veloc	6.5×10^{-3}	93	pitch	3.8×10^{-3}
44	spars	6.5×10^{-3}	94	treatment	3.8×10^{-3}
45	music	6.1×10^{-3}	95	element	3.8×10^{-3}
46	gene	6.1×10^{-3}	96	vector	3.8×10^{-3}
47	linear	6.1×10^{-3}	97	coeffici	3.7×10^{-3}
48	spectral	6.1×10^{-3}	98	use	3.7×10^{-3}
49	noisi	6×10^{-3}	99	base	3.7×10^{-3}
50	stiff	6×10^{-3}	100	equat	3.7×10^{-3}

Table D.2. The list of the top 100 words in the category Agricultural Economics and Policy with RIGs

No.	Word	RIG	No.	Word	RIG
1	agricultur	1.1×10^{-1}	51	choic	9.3×10^{-3}
2	farmer	8.7×10^{-2}	52	competit	9.3×10^{-3}
3	market	8.6×10^{-2}	53	domest	9.1×10^{-3}
4	farm	8.3×10^{-2}	54	conclus	8.8×10^{-3}
5	price	8.1×10^{-2}	55	clinic	8.8×10^{-3}
6	food	7.2×10^{-2}	56	nutrit	8.6×10^{-3}
7	polici	6×10^{-2}	57	veget	8.6×10^{-3}
8	product	4.8×10^{-2}	58	african	8.5×10^{-3}
9	econom	3.6×10^{-2}	59	grower	8.3×10^{-3}
10	household	3.5×10^{-2}	60	industri	8.3×10^{-3}
11	consum	3.3×10^{-2}	61	govern	8.2×10^{-3}
12	incom	3.3×10^{-2}	62	fertil	8×10^{-3}
13	countri	3.2×10^{-2}	63	urban	8×10^{-3}
14	sector	3.1×10^{-2}	64	irrig	8×10^{-3}
15	crop	2.6×10^{-2}	65	firm	7.8×10^{-3}
16	trade	2.4×10^{-2}	66	busi	7.8×10^{-3}
17	rural	2.3×10^{-2}	67	produc	7.5×10^{-3}
18	commod	2×10^{-2}	68	privat	7.2×10^{-3}
19	export	2×10^{-2}	69	milk	7.2×10^{-3}
20	land	2×10^{-2}	70	meat	7×10^{-3}
21	demand	1.8×10^{-2}	71	beef	6.9×10^{-3}
22	suppli	1.7×10^{-2}	72	secur	6.9×10^{-3}
23	wine	1.6×10^{-2}	73	prefer	6.7×10^{-3}
24	impact	1.6×10^{-2}	74	grape	6.5×10^{-3}
25	find	1.6×10^{-2}	75	sale	6.5×10^{-3}
26	welfar	1.6×10^{-2}	76	decis	6.5×10^{-3}
27	empir	1.5×10^{-2}	77	adopt	6.4×10^{-3}
28	patient	1.4×10^{-2}	78	benefit	6.4×10^{-3}
29	estim	1.4×10^{-2}	79	china	6.3×10^{-3}
30	poverti	1.4×10^{-2}	80	research	6.2×10^{-3}
31	survey	1.3×10^{-2}	81	sell	6.2×10^{-3}
32	purchas	1.3×10^{-2}	82	valuat	6.1×10^{-3}
33	africa	1.3×10^{-2}	83	wheat	6.1×10^{-3}
34	invest	1.3×10^{-2}	84	public	6.1×10^{-3}
35	cost	1.2×10^{-2}	85	detect	6.1×10^{-3}
36	cell	1.1×10^{-2}	86	perform	6.1×10^{-3}
37	econometr	1.1×10^{-2}	87	import	6×10^{-3}
38	livestock	1.1×10^{-2}	88	implic	6×10^{-3}
39	willing	1.1×10^{-2}	89	background	6×10^{-3}
40	dairi	1.1×10^{-2}	90	cereal	5.9×10^{-3}
41	panel	1.1×10^{-2}	91	revenu	5.7×10^{-3}
42	retail	1.1×10^{-2}	92	financi	5.7×10^{-3}
43	pay	1.1×10^{-2}	93	electron	5.7×10^{-3}
44	premium	1.1×10^{-2}	94	climat	5.6×10^{-3}
45	articl	1×10^{-2}	95	maiz	5.6×10^{-3}
46	profit	1×10^{-2}	96	program	5.6×10^{-3}
47	consumpt	9.9×10^{-3}	97	nation	5.5×10^{-3}
48	payment	9.9×10^{-3}	98	livelihood	5.4×10^{-3}
49	economi	9.6×10^{-3}	99	algorithm	5.4×10^{-3}
50	method	9.4×10^{-3}	100	volatil	5.4×10^{-3}

TABLE D.3. The list of the top 100 words in the category Agricultural Engineering with RIGs

No.	Word	RIG	No.	Word	RIG
1	biomass	6.8×10^{-2}	51	corn	1.1×10^{-2}
2	product	6×10^{-2}	52	fuel	1.1×10^{-2}
3	yield	4.1×10^{-2}	53	enzymat	1.1×10^{-2}
4	anaerob	3.1×10^{-2}	54	manur	1×10^{-2}
5	content	2.9×10^{-2}	55	wood	1×10^{-2}
6	ferment	2.9×10^{-2}	56	extract	9.8×10^{-3}
7	crop	2.9×10^{-2}	57	respect	9.8×10^{-3}
8	sludg	2.6×10^{-2}	58	degre	9.7×10^{-3}
9	feedstock	2.6×10^{-2}	59	clinic	9.5×10^{-3}
10	dri	2.5×10^{-2}	60	pyrolysi	9.3×10^{-3}
11	oil	2.4×10^{-2}	61	volatil	9.2×10^{-3}
12	pretreat	2.3×10^{-2}	62	nutrient	9.1×10^{-3}
13	irrig	2.3×10^{-2}	63	bio	8.8×10^{-3}
14	remov	2.1×10^{-2}	64	algal	8.4×10^{-3}
15	lignin	2×10^{-2}	65	phenol	8.4×10^{-3}
16	wastewat	2×10^{-2}	66	wheat	8.3×10^{-3}
17	plant	2×10^{-2}	67	chemic	8.3×10^{-3}
18	water	2×10^{-2}	68	effici	8.1×10^{-3}
19	biodiesel	2×10^{-2}	69	fruit	8×10^{-3}
20	reactor	1.9×10^{-2}	70	total	7.9×10^{-3}
21	cultiv	1.9×10^{-2}	71	condit	7.8×10^{-3}
22	digest	1.8×10^{-2}	72	background	7.8×10^{-3}
23	lignocellulos	1.8×10^{-2}	73	maximum	7.7×10^{-3}
24	biofuel	1.8×10^{-2}	74	higher	7.7×10^{-3}
25	produc	1.8×10^{-2}	75	lipid	7.7×10^{-3}
26	agricultur	1.8×10^{-2}	76	acet	7.5×10^{-3}
27	hydrolysi	1.8×10^{-2}	77	paper	7.4×10^{-3}
28	batch	1.7×10^{-2}	78	seed	7.2×10^{-3}
29	patient	1.7×10^{-2}	79	carbon	6.8×10^{-3}
30	acid	1.7×10^{-2}	80	ammonia	6.8×10^{-3}
31	cellulos	1.7×10^{-2}	81	temperatur	6.8×10^{-3}
32	straw	1.6×10^{-2}	82	rice	6.7×10^{-3}
33	ethanol	1.6×10^{-2}	83	farm	6.4×10^{-3}
34	cod	1.5×10^{-2}	84	amount	6.3×10^{-3}
35	sugar	1.5×10^{-2}	85	evapotranspir	6.3×10^{-3}
36	wast	1.5×10^{-2}	86	industri	6.3×10^{-3}
37	hemicellulos	1.5×10^{-2}	87	composit	6×10^{-3}
38	highest	1.5×10^{-2}	88	pulp	6×10^{-3}
39	nitrogen	1.4×10^{-2}	89	glucos	5.8×10^{-3}
40	methan	1.4×10^{-2}	90	residu	5.7×10^{-3}
41	harvest	1.4×10^{-2}	91	process	5.6×10^{-3}
42	microbi	1.4×10^{-2}	92	raw	5.5×10^{-3}
43	bioreactor	1.3×10^{-2}	93	greenhous	5.5×10^{-3}
44	conclus	1.3×10^{-2}	94	age	5.5×10^{-3}
45	concentr	1.3×10^{-2}	95	outcom	5.4×10^{-3}
46	moistur	1.3×10^{-2}	96	min	5.4×10^{-3}
47	effluent	1.2×10^{-2}	97	propos	5.4×10^{-3}
48	soil	1.2×10^{-2}	98	potenti	5.3×10^{-3}
49	fatti	1.1×10^{-2}	99	obtain	5.2×10^{-3}
50	solid	1.1×10^{-2}	100	hydraul	5.2×10^{-3}

TABLE D.4. The list of the top 100 words in the category Agriculture, Dairy and Animal Science with RIGs

No.	Word	RIG
1	feed	1.1×10^{-1}
2	diet	1.1×10^{-1}
3	cow	1×10^{-1}
4	milk	9.4×10^{-2}
5	fed	9.1×10^{-2}
6	dairi	8.9×10^{-2}
7	anim	7.1×10^{-2}
8	intak	6.9×10^{-2}
9	breed	6.7×10^{-2}
10	holstein	6.4×10^{-2}
11	cattl	6×10^{-2}
12	broiler	5.8×10^{-2}
13	supplement	5.5×10^{-2}
14	dietari	5.5×10^{-2}
15	lactat	5.1×10^{-2}
16	digest	4.9×10^{-2}
17	fat	4.7×10^{-2}
18	carcass	4.6×10^{-2}
19	calv	4.6×10^{-2}
20	weight	4.4×10^{-2}
21	day	4.3×10^{-2}
22	meat	4.3×10^{-2}
23	rumin	4.2×10^{-2}
24	total	4.2×10^{-2}
25	pig	4×10^{-2}
26	slaughter	3.7×10^{-2}
27	dri	3.7×10^{-2}
28	trait	3.7×10^{-2}
29	chicken	3.6×10^{-2}
30	herd	3.5×10^{-2}
31	pen	3.4×10^{-2}
32	bird	3.4×10^{-2}
33	farm	3.3×10^{-2}
34	concentr	3.1×10^{-2}
35	wean	3.1×10^{-2}
36	product	2.9×10^{-2}
37	meal	2.9×10^{-2}
38	corn	2.8×10^{-2}
39	goat	2.8×10^{-2}
40	treatment	2.7×10^{-2}
41	sheep	2.7×10^{-2}
42	beef	2.6×10^{-2}
43	fatti	2.6×10^{-2}
44	per	2.6×10^{-2}
45	random	2.6×10^{-2}
46	daili	2.5×10^{-2}
47	protein	2.5×10^{-2}
48	chick	2.3×10^{-2}
49	matter	2.3×10^{-2}
50	paper	2.2×10^{-2}

No.	Word	RIG
51	period	2.2×10^{-2}
52	affect	2.2×10^{-2}
53	effect	2.1×10^{-2}
54	bodi	2×10^{-2}
55	acid	2×10^{-2}
56	crude	2×10^{-2}
57	studi	2×10^{-2}
58	content	1.9×10^{-2}
59	higher	1.9×10^{-2}
60	reproduct	1.9×10^{-2}
61	soybean	1.9×10^{-2}
62	nutrient	1.9×10^{-2}
63	insemin	1.9×10^{-2}
64	replic	1.8×10^{-2}
65	gain	1.8×10^{-2}
66	increas	1.8×10^{-2}
67	egg	1.8×10^{-2}
68	respect	1.8×10^{-2}
69	group	1.8×10^{-2}
70	percentag	1.7×10^{-2}
71	patient	1.7×10^{-2}
72	greater	1.7×10^{-2}
73	factori	1.7×10^{-2}
74	differ	1.7×10^{-2}
75	assign	1.7×10^{-2}
76	growth	1.7×10^{-2}
77	forag	1.7×10^{-2}
78	hous	1.7×10^{-2}
79	lower	1.7×10^{-2}
80	averag	1.7×10^{-2}
81	propos	1.6×10^{-2}
82	collect	1.6×10^{-2}
83	genet	1.6×10^{-2}
84	pastur	1.6×10^{-2}
85	bovin	1.5×10^{-2}
86	decreas	1.5×10^{-2}
87	ferment	1.5×10^{-2}
88	nutrit	1.5×10^{-2}
89	semen	1.5×10^{-2}
90	litter	1.4×10^{-2}
91	rear	1.4×10^{-2}
92	herit	1.4×10^{-2}
93	dure	1.4×10^{-2}
94	blood	1.3×10^{-2}
95	genotyp	1.3×10^{-2}
96	sampl	1.3×10^{-2}
97	evalu	1.3×10^{-2}
98	graze	1.3×10^{-2}
99	latin	1.3×10^{-2}
100	week	1.2×10^{-2}

Table D.5. The list of the top 100 words in the category Agriculture, Multidisciplinary with RIGs

No.	Word	RIG	No.	Word	RIG
1	crop	6×10^{-2}	51	chromatographi	8.4×10^{-3}
2	plant	4.4×10^{-2}	52	agronom	8.4×10^{-3}
3	agricultur	4×10^{-2}	53	paper	8.3×10^{-3}
4	soil	3.5×10^{-2}	54	day	8.3×10^{-3}
5	product	3.1×10^{-2}	55	weed	8.3×10^{-3}
6	farm	2.9×10^{-2}	56	matter	8.2×10^{-3}
7	wheat	2.4×10^{-2}	57	diet	8.2×10^{-3}
8	cultivar	2.3×10^{-2}	58	digest	8.2×10^{-3}
9	farmer	2.3×10^{-2}	59	cow	8.1×10^{-3}
10	yield	2.3×10^{-2}	60	respect	8×10^{-3}
11	food	2.2×10^{-2}	61	extract	8×10^{-3}
12	dri	2.1×10^{-2}	62	pest	8×10^{-3}
13	content	2×10^{-2}	63	higher	7.9×10^{-3}
14	seed	1.8×10^{-2}	64	leav	7.5×10^{-3}
15	fruit	1.8×10^{-2}	65	seedl	7.5×10^{-3}
16	patient	1.7×10^{-2}	66	clinic	7.4×10^{-3}
17	fertil	1.7×10^{-2}	67	differ	7.3×10^{-3}
18	maiz	1.6×10^{-2}	68	root	7.3×10^{-3}
19	breed	1.6×10^{-2}	69	qualiti	7.2×10^{-3}
20	leaf	1.5×10^{-2}	70	forag	7.1×10^{-3}
21	acid	1.4×10^{-2}	71	water	7.1×10^{-3}
22	highest	1.4×10^{-2}	72	land	7×10^{-3}
23	nutrient	1.3×10^{-2}	73	fed	7×10^{-3}
24	cultiv	1.3×10^{-2}	74	propos	7×10^{-3}
25	pastur	1.3×10^{-2}	75	polyphenol	6.9×10^{-3}
26	total	1.3×10^{-2}	76	genotyp	6.9×10^{-3}
27	feed	1.2×10^{-2}	77	grown	6.9×10^{-3}
28	rice	1.2×10^{-2}	78	greenhous	6.8×10^{-3}
29	season	1.2×10^{-2}	79	anim	6.8×10^{-3}
30	irrig	1.2×10^{-2}	80	intak	6.7×10^{-3}
31	veget	1.2×10^{-2}	81	studi	6.7×10^{-3}
32	nitrogen	1.2×10^{-2}	82	ferment	6.6×10^{-3}
33	livestock	1.1×10^{-2}	83	corn	6.4×10^{-3}
34	phenol	1.1×10^{-2}	84	meat	6.4×10^{-3}
35	growth	1.1×10^{-2}	85	grow	6.4×10^{-3}
36	harvest	1×10^{-2}	86	fresh	6.4×10^{-3}
37	concentr	1×10^{-2}	87	plot	6.3×10^{-3}
38	milk	1×10^{-2}	88	dietari	6.2×10^{-3}
39	grass	9.9×10^{-3}	89	per	6.2×10^{-3}
40	weight	9.9×10^{-3}	90	anthocyanin	6.1×10^{-3}
41	grain	9.9×10^{-3}	91	tillag	6×10^{-3}
42	antioxid	9.8×10^{-3}	92	rumin	6×10^{-3}
43	nutrit	9.7×10^{-3}	93	winter	5.9×10^{-3}
44	dairi	9.6×10^{-3}	94	potato	5.8×10^{-3}
45	soybean	9.5×10^{-3}	95	compound	5.8×10^{-3}
46	graze	9.5×10^{-3}	96	grassland	5.8×10^{-3}
47	biomass	9.3×10^{-3}	97	sheep	5.8×10^{-3}
48	cattl	9.2×10^{-3}	98	fatti	5.7×10^{-3}
49	manur	9.1×10^{-3}	99	flower	5.7×10^{-3}
50	trait	8.5×10^{-3}	100	evalu	5.7×10^{-3}

TABLE D.6. The list of the top 100 words in the category Agronomy with RIGs

No.	Word	RIG
1	plant	1.3×10^{-1}
2	crop	1.2×10^{-1}
3	cultivar	9.3×10^{-2}
4	soil	7.9×10^{-2}
5	yield	6×10^{-2}
6	seed	4.5×10^{-2}
7	leaf	4.4×10^{-2}
8	breed	4.2×10^{-2}
9	wheat	4.2×10^{-2}
10	fruit	3.8×10^{-2}
11	trait	3.8×10^{-2}
12	irrig	3.7×10^{-2}
13	fertil	3.7×10^{-2}
14	season	3.2×10^{-2}
15	product	3.1×10^{-2}
16	harvest	2.8×10^{-2}
17	root	2.8×10^{-2}
18	dri	2.7×10^{-2}
19	agricultur	2.7×10^{-2}
20	cultiv	2.7×10^{-2}
21	genotyp	2.7×10^{-2}
22	agronom	2.5×10^{-2}
23	content	2.5×10^{-2}
24	grown	2.5×10^{-2}
25	maiz	2.4×10^{-2}
26	seedl	2.4×10^{-2}
27	weed	2.4×10^{-2}
28	biomass	2.4×10^{-2}
29	grain	2.3×10^{-2}
30	rice	2.2×10^{-2}
31	germplasm	2.2×10^{-2}
32	flower	2.2×10^{-2}
33	greenhous	2.2×10^{-2}
34	growth	2×10^{-2}
35	genet	1.9×10^{-2}
36	shoot	1.9×10^{-2}
37	patient	1.9×10^{-2}
38	field	1.8×10^{-2}
39	qtl	1.8×10^{-2}
40	veget	1.8×10^{-2}
41	grow	1.8×10^{-2}
42	nutrient	1.8×10^{-2}
43	speci	1.8×10^{-2}
44	leav	1.7×10^{-2}
45	water	1.7×10^{-2}
46	inocul	1.5×10^{-2}
47	paper	1.5×10^{-2}
48	nitrogen	1.5×10^{-2}
49	orchard	1.5×10^{-2}
50	highest	1.4×10^{-2}
51	winter	1.4×10^{-2}
52	plot	1.4×10^{-2}
53	total	1.4×10^{-2}
54	farmer	1.4×10^{-2}
55	pest	1.4×10^{-2}
56	potato	1.3×10^{-2}
57	tree	1.3×10^{-2}
58	germin	1.3×10^{-2}
59	drought	1.3×10^{-2}
60	farm	1.3×10^{-2}
61	propos	1.3×10^{-2}
62	sativa	1.3×10^{-2}
63	postharvest	1.3×10^{-2}
64	loci	1.2×10^{-2}
65	ssr	1.2×10^{-2}
66	marker	1.2×10^{-2}
67	soybean	1.1×10^{-2}
68	grass	1.1×10^{-2}
69	canopi	1.1×10^{-2}
70	conduct	1.1×10^{-2}
71	resist	1×10^{-2}
72	clinic	1×10^{-2}
73	climat	1×10^{-2}
74	matur	1×10^{-2}
75	tillag	1×10^{-2}
76	differ	9.9×10^{-3}
77	manag	9.9×10^{-3}
78	three	9.7×10^{-3}
79	four	9.7×10^{-3}
80	manur	9.6×10^{-3}
81	grower	9.6×10^{-3}
82	qualiti	9.5×10^{-3}
83	tomato	9.4×10^{-3}
84	chromosom	9.2×10^{-3}
85	method	9.1×10^{-3}
86	popul	9×10^{-3}
87	treatment	8.9×10^{-3}
88	spring	8.9×10^{-3}
89	suscept	8.8×10^{-3}
90	fresh	8.7×10^{-3}
91	progeni	8.7×10^{-3}
92	pathogen	8.4×10^{-3}
93	evalu	8.3×10^{-3}
94	evapotranspir	8.2×10^{-3}
95	experi	8.1×10^{-3}
96	increas	8.1×10^{-3}
97	higher	8×10^{-3}
98	matter	8×10^{-3}
99	per	8×10^{-3}
100	sugar	7.8×10^{-3}

Table D.7. The list of the top 100 words in the category Allergy with RIGs

No.	Word	RIG	No.	Word	RIG
1	asthma	2.6×10^{-1}	51	medic	1.7×10^{-2}
2	allerg	2.2×10^{-1}	52	sensit	1.7×10^{-2}
3	allergi	1.7×10^{-1}	53	assess	1.6×10^{-2}
4	allergen	1.4×10^{-1}	54	control	1.6×10^{-2}
5	conclus	1.3×10^{-1}	55	signific	1.6×10^{-2}
6	background	1.3×10^{-1}	56	treatment	1.6×10^{-2}
7	ige	1.3×10^{-1}	57	subject	1.6×10^{-2}
8	patient	8×10^{-2}	58	histori	1.5×10^{-2}
9	atop	8×10^{-2}	59	preval	1.5×10^{-2}
10	rhiniti	7.3×10^{-2}	60	antibodi	1.5×10^{-2}
11	object	7.2×10^{-2}	61	inflammatori	1.5×10^{-2}
12	asthmat	6.8×10^{-2}	62	risk	1.4×10^{-2}
13	method	6.4×10^{-2}	63	placebo	1.4×10^{-2}
14	airway	5.8×10^{-2}	64	diagnos	1.4×10^{-2}
15	children	5.7×10^{-2}	65	healthi	1.4×10^{-2}
16	symptom	5.1×10^{-2}	66	exposur	1.3×10^{-2}
17	dermat	5×10^{-2}	67	test	1.3×10^{-2}
18	prick	4.9×10^{-2}	68	reaction	1.3×10^{-2}
19	clinic	4.9×10^{-2}	69	adult	1.3×10^{-2}
20	eosinophil	4.8×10^{-2}	70	life	1.2×10^{-2}
21	skin	4.8×10^{-2}	71	level	1.2×10^{-2}
22	diseas	4.2×10^{-2}	72	dust	1.2×10^{-2}
23	inhal	4.1×10^{-2}	73	score	1.2×10^{-2}
24	corticosteroid	3.1×10^{-2}	74	respons	1.2×10^{-2}
25	inflamm	3.1×10^{-2}	75	specif	1.2×10^{-2}
26	result	3.1×10^{-2}	76	induc	1.2×10^{-2}
27	immunotherapi	3×10^{-2}	77	simul	1.2×10^{-2}
28	exacerb	2.9×10^{-2}	78	studi	1.1×10^{-2}
29	age	2.6×10^{-2}	79	expiratori	1.1×10^{-2}
30	associ	2.6×10^{-2}	80	elisa	1.1×10^{-2}
31	serum	2.4×10^{-2}	81	propos	1.1×10^{-2}
32	cytokin	2.4×10^{-2}	82	cohort	1.1×10^{-2}
33	immun	2.3×10^{-2}	83	immunoglobulin	1.1×10^{-2}
34	fev1	2.3×10^{-2}	84	cell	1.1×10^{-2}
35	pollen	2.2×10^{-2}	85	phenotyp	1.1×10^{-2}
36	paper	2.2×10^{-2}	86	oral	1.1×10^{-2}
37	bronchial	2.2×10^{-2}	87	igg	1.1×10^{-2}
38	chronic	2.1×10^{-2}	88	mediat	1×10^{-2}
39	questionnair	2.1×10^{-2}	89	baselin	1×10^{-2}
40	respiratori	2.1×10^{-2}	90	odd	1×10^{-2}
41	sought	2.1×10^{-2}	91	persist	1×10^{-2}
42	year	2.1×10^{-2}	92	visit	1×10^{-2}
43	childhood	2.1×10^{-2}	93	group	9.7×10^{-3}
44	mite	2×10^{-2}	94	toler	9.6×10^{-3}
45	lung	2×10^{-2}	95	common	9.4×10^{-3}
46	nasal	1.9×10^{-2}	96	blood	9.3×10^{-3}
47	immunolog	1.9×10^{-2}	97	pulmonari	9×10^{-3}
48	food	1.9×10^{-2}	98	pathogenesi	8.9×10^{-3}
49	sever	1.9×10^{-2}	99	frequent	8.9×10^{-3}
50	diagnosi	1.8×10^{-2}	100	pediatr	8.9×10^{-3}

Table D.8. The list of the top 100 words in the category Anatomy and Morphology with RIGs

No.	Word	RIG
1	anatom	5.8×10^{-2}
2	anatomi	4.8×10^{-2}
3	morpholog	4×10^{-2}
4	cadav	3.8×10^{-2}
5	dissect	3.2×10^{-2}
6	posterior	2.6×10^{-2}
7	neuron	2.5×10^{-2}
8	anterior	2.5×10^{-2}
9	dorsal	2.4×10^{-2}
10	ventral	2.3×10^{-2}
11	medial	2.3×10^{-2}
12	stain	2.3×10^{-2}
13	muscl	2.3×10^{-2}
14	nerv	2.3×10^{-2}
15	tissu	2.2×10^{-2}
16	paper	2.2×10^{-2}
17	cell	2.1×10^{-2}
18	histolog	2.1×10^{-2}
19	adult	1.8×10^{-2}
20	morphometr	1.7×10^{-2}
21	bone	1.7×10^{-2}
22	cranial	1.7×10^{-2}
23	left	1.7×10^{-2}
24	caudal	1.7×10^{-2}
25	later	1.7×10^{-2}
26	immunohistochem	1.6×10^{-2}
27	vertebr	1.5×10^{-2}
28	specimen	1.4×10^{-2}
29	cortex	1.3×10^{-2}
30	male	1.3×10^{-2}
31	express	1.3×10^{-2}
32	development	1.3×10^{-2}
33	epithelium	1.2×10^{-2}
34	right	1.2×10^{-2}
35	ultrastructur	1.2×10^{-2}
36	studi	1.2×10^{-2}
37	skull	1.2×10^{-2}
38	branch	1.2×10^{-2}
39	pattern	1.2×10^{-2}
40	brain	1.1×10^{-2}
41	distal	1.1×10^{-2}
42	inferior	1.1×10^{-2}
43	canal	1.1×10^{-2}
44	boni	1×10^{-2}
45	immunoreact	1×10^{-2}
46	mammal	1×10^{-2}
47	arteri	1×10^{-2}
48	region	9.9×10^{-3}
49	rat	9.9×10^{-3}
50	microscopi	9.9×10^{-3}

No.	Word	RIG
51	section	9.8×10^{-3}
52	bilater	9.6×10^{-3}
53	cartilag	9.3×10^{-3}
54	trunk	9.2×10^{-3}
55	human	8.7×10^{-3}
56	cadaver	8.5×10^{-3}
57	axon	8.5×10^{-3}
58	embryon	8.5×10^{-3}
59	cortic	8.4×10^{-3}
60	ligament	8.4×10^{-3}
61	limb	7.9×10^{-3}
62	femal	7.9×10^{-3}
63	gland	7.8×10^{-3}
64	observ	7.8×10^{-3}
65	nucleus	7.7×10^{-3}
66	anim	7.7×10^{-3}
67	bodi	7.6×10^{-3}
68	basal	7.5×10^{-3}
69	embryo	7.5×10^{-3}
70	temperatur	7.4×10^{-3}
71	mandibular	7.4×10^{-3}
72	postnat	7.3×10^{-3}
73	head	7.3×10^{-3}
74	simul	7.2×10^{-3}
75	immunohistochemistri	7.2×10^{-3}
76	shape	7.2×10^{-3}
77	apic	7×10^{-3}
78	dure	6.9×10^{-3}
79	cytoplasm	6.8×10^{-3}
80	propos	6.8×10^{-3}
81	vessel	6.5×10^{-3}
82	appear	6.5×10^{-3}
83	sensori	6.5×10^{-3}
84	scan	6.5×10^{-3}
85	proxim	6.4×10^{-3}
86	tomographi	6.4×10^{-3}
87	effect	6.3×10^{-3}
88	function	6.3×10^{-3}
89	energi	6.3×10^{-3}
90	design	6.3×10^{-3}
91	variat	6.2×10^{-3}
92	rate	6.2×10^{-3}
93	articular	6.2×10^{-3}
94	found	6.1×10^{-3}
95	speci	6.1×10^{-3}
96	sinus	6.1×10^{-3}
97	side	6×10^{-3}
98	sagitt	6×10^{-3}
99	imag	5.9×10^{-3}
100	aim	5.8×10^{-3}

Table D.9. The list of the top 100 words in the category Andrology with RIGs

No.	Word	RIG	No.	Word	RIG
1	sperm	2.2×10^{-1}	51	cryopreserv	1.2×10^{-2}
2	infertil	1.2×10^{-1}	52	evalu	1.2×10^{-2}
3	testicular	1.2×10^{-1}	53	associ	1.2×10^{-2}
4	semen	1.1×10^{-1}	54	inject	1.2×10^{-2}
5	spermatozoa	1.1×10^{-1}	55	malondialdehyd	1.1×10^{-2}
6	male	9.9×10^{-2}	56	control	1.1×10^{-2}
7	testi	9.7×10^{-2}	57	wistar	1×10^{-2}
8	motil	8×10^{-2}	58	glutathion	1×10^{-2}
9	testosteron	7.9×10^{-2}	59	superoxid	9.8×10^{-3}
10	men	7.6×10^{-2}	60	assay	9.8×10^{-3}
11	spermatogenesi	6.2×10^{-2}	61	express	9.7×10^{-3}
12	fertil	6×10^{-2}	62	fragment	9.6×10^{-3}
13	ejacul	5.4×10^{-2}	63	simul	9.5×10^{-3}
14	erectil	4.9×10^{-2}	64	stimul	9.5×10^{-3}
15	semin	4.7×10^{-2}	65	method	9.5×10^{-3}
16	hormon	4.1×10^{-2}	66	peroxid	9.5×10^{-3}
17	reproduct	3.9×10^{-2}	67	week	9.4×10^{-3}
18	group	3.7×10^{-2}	68	paramet	9.4×10^{-3}
19	acrosom	3.6×10^{-2}	69	administr	9.3×10^{-3}
20	signific	3.4×10^{-2}	70	thaw	9.3×10^{-3}
21	androgen	3.2×10^{-2}	71	compar	9.3×10^{-3}
22	rat	3.1×10^{-2}	72	oocyt	9.2×10^{-3}
23	prostat	2.9×10^{-2}	73	morpholog	9.1×10^{-3}
24	tubul	2.8×10^{-2}	74	concentr	9.1×10^{-3}
25	studi	2.5×10^{-2}	75	percentag	9×10^{-3}
26	dysfunct	2.4×10^{-2}	76	histopatholog	8.9×10^{-3}
27	patient	2.3×10^{-2}	77	analys	8.9×10^{-3}
28	paper	2.3×10^{-2}	78	apoptot	8.8×10^{-3}
29	germ	2.2×10^{-2}	79	histolog	8.8×10^{-3}
30	pregnanc	2.2×10^{-2}	80	catalas	8.8×10^{-3}
31	normal	1.9×10^{-2}	81	intraperiton	8.7×10^{-3}
32	divid	1.9×10^{-2}	82	embryo	8.6×10^{-3}
33	level	1.9×10^{-2}	83	tissu	8.5×10^{-3}
34	decreas	1.7×10^{-2}	84	receptor	8.5×10^{-3}
35	serum	1.6×10^{-2}	85	day	8.5×10^{-3}
36	antioxid	1.6×10^{-2}	86	increas	8.3×10^{-3}
37	treatment	1.6×10^{-2}	87	base	8.3×10^{-3}
38	follicl	1.5×10^{-2}	88	stain	8.3×10^{-3}
39	count	1.5×10^{-2}	89	insemin	8×10^{-3}
40	cell	1.5×10^{-2}	90	dismutas	7.9×10^{-3}
41	treat	1.5×10^{-2}	91	induc	7.8×10^{-3}
42	dna	1.5×10^{-2}	92	vitro	7.8×10^{-3}
43	abnorm	1.4×10^{-2}	93	conclud	7.8×10^{-3}
44	sexual	1.4×10^{-2}	94	receiv	7.8×10^{-3}
45	total	1.4×10^{-2}	95	assess	7.7×10^{-3}
46	chromatin	1.3×10^{-2}	96	order	7.6×10^{-3}
47	ivf	1.3×10^{-2}	97	viabil	7.5×10^{-3}
48	epithelium	1.3×10^{-2}	98	gene	7.5×10^{-3}
49	aim	1.3×10^{-2}	99	dose	7.5×10^{-3}
50	damag	1.3×10^{-2}	100	propos	7.4×10^{-3}

TABLE D.10. The list of the top 100 words in the category Anesthesiology with RIGs

No.	Word	RIG
1	pain	1.4×10^{-1}
2	patient	1.2×10^{-1}
3	anesthesia	9.7×10^{-2}
4	conclus	9×10^{-2}
5	surgeri	8.4×10^{-2}
6	background	8×10^{-2}
7	anesthet	6×10^{-2}
8	postop	5.6×10^{-2}
9	analgesia	5.2×10^{-2}
10	undergo	5.1×10^{-2}
11	analges	4.2×10^{-2}
12	anaesthesia	4×10^{-2}
13	opioid	3.8×10^{-2}
14	random	3.3×10^{-2}
15	receiv	3.2×10^{-2}
16	prospect	3.2×10^{-2}
17	ventil	3.1×10^{-2}
18	hospit	3×10^{-2}
19	spinal	3×10^{-2}
20	care	2.9×10^{-2}
21	periop	2.9×10^{-2}
22	score	2.8×10^{-2}
23	nerv	2.8×10^{-2}
24	outcom	2.7×10^{-2}
25	chronic	2.6×10^{-2}
26	clinic	2.6×10^{-2}
27	elect	2.6×10^{-2}
28	group	2.5×10^{-2}
29	method	2.5×10^{-2}
30	cardiac	2.4×10^{-2}
31	infus	2.4×10^{-2}
32	object	2.4×10^{-2}
33	intub	2.4×10^{-2}
34	paper	2.4×10^{-2}
35	intraop	2.3×10^{-2}
36	assess	2.3×10^{-2}
37	signific	2.2×10^{-2}
38	arteri	2.2×10^{-2}
39	blind	2.2×10^{-2}
40	result	2.1×10^{-2}
41	preoper	2.1×10^{-2}
42	administr	2.1×10^{-2}
43	surgic	2×10^{-2}
44	sedat	2×10^{-2}
45	befor	2×10^{-2}
46	intervent	1.9×10^{-2}
47	dure	1.9×10^{-2}
48	airway	1.9×10^{-2}
49	confid	1.8×10^{-2}
50	interv	1.8×10^{-2}

No.	Word	RIG
51	hour	1.8×10^{-2}
52	trial	1.7×10^{-2}
53	intraven	1.7×10^{-2}
54	incid	1.7×10^{-2}
55	median	1.7×10^{-2}
56	blood	1.7×10^{-2}
57	dose	1.7×10^{-2}
58	complic	1.7×10^{-2}
59	medic	1.7×10^{-2}
60	minut	1.6×10^{-2}
61	injuri	1.6×10^{-2}
62	studi	1.6×10^{-2}
63	min	1.6×10^{-2}
64	record	1.6×10^{-2}
65	administ	1.5×10^{-2}
66	day	1.5×10^{-2}
67	sensori	1.4×10^{-2}
68	associ	1.4×10^{-2}
69	cardiopulmonari	1.4×10^{-2}
70	placebo	1.4×10^{-2}
71	baselin	1.4×10^{-2}
72	compar	1.3×10^{-2}
73	cathet	1.3×10^{-2}
74	acut	1.3×10^{-2}
75	intens	1.3×10^{-2}
76	inject	1.3×10^{-2}
77	hemodynam	1.3×10^{-2}
78	measur	1.3×10^{-2}
79	block	1.3×10^{-2}
80	relief	1.2×10^{-2}
81	durat	1.2×10^{-2}
82	procedur	1.2×10^{-2}
83	propos	1.2×10^{-2}
84	primari	1.2×10^{-2}
85	may	1.1×10^{-2}
86	mean	1.1×10^{-2}
87	particip	1.1×10^{-2}
88	risk	1.1×10^{-2}
89	advers	1.1×10^{-2}
90	pressur	1.1×10^{-2}
91	peripher	1.1×10^{-2}
92	hundr	1.1×10^{-2}
93	bypass	1.1×10^{-2}
94	adult	1.1×10^{-2}
95	respiratori	1×10^{-2}
96	schedul	1×10^{-2}
97	cord	1×10^{-2}
98	whether	1×10^{-2}
99	retrospect	1×10^{-2}
100	salin	1×10^{-2}

Table D.11. The list of the top 100 words in the category Anthropology with RIGs

No.	Word	RIG
1	archaeolog	8.7×10^{-2}
2	anthropolog	4.7×10^{-2}
3	ethnograph	4.5×10^{-2}
4	social	4.3×10^{-2}
5	polit	3.1×10^{-2}
6	cultur	3.1×10^{-2}
7	argu	2.7×10^{-2}
8	articl	2.4×10^{-2}
9	peopl	2.3×10^{-2}
10	centuri	2.1×10^{-2}
11	histor	2.1×10^{-2}
12	neolith	2×10^{-2}
13	site	2×10^{-2}
14	excav	1.9×10^{-2}
15	societi	1.9×10^{-2}
16	fieldwork	1.8×10^{-2}
17	prehistor	1.8×10^{-2}
18	indigen	1.8×10^{-2}
19	assemblag	1.7×10^{-2}
20	context	1.6×10^{-2}
21	modern	1.6×10^{-2}
22	communiti	1.6×10^{-2}
23	subsist	1.6×10^{-2}
24	settlement	1.5×10^{-2}
25	ritual	1.4×10^{-2}
26	contemporari	1.4×10^{-2}
27	draw	1.4×10^{-2}
28	date	1.3×10^{-2}
29	human	1.3×10^{-2}
30	cell	1.3×10^{-2}
31	histori	1.3×10^{-2}
32	explor	1.3×10^{-2}
33	ancient	1.2×10^{-2}
34	discours	1.2×10^{-2}
35	south	1.2×10^{-2}
36	africa	1.2×10^{-2}
37	late	1.2×10^{-2}
38	african	1.1×10^{-2}
39	practic	1.1×10^{-2}
40	bronz	1.1×10^{-2}
41	way	1.1×10^{-2}
42	perform	1.1×10^{-2}
43	ethnographi	1.1×10^{-2}
44	patient	1×10^{-2}
45	burial	1×10^{-2}
46	archaeologist	1×10^{-2}
47	landscap	1×10^{-2}
48	radiocarbon	9.8×10^{-3}
49	stone	9.8×10^{-3}
50	live	9.7×10^{-3}

No.	Word	RIG
51	chronolog	9.6×10^{-3}
52	individu	9.5×10^{-3}
53	world	9.1×10^{-3}
54	earli	9.1×10^{-3}
55	debat	9×10^{-3}
56	effect	9×10^{-3}
57	interpret	9×10^{-3}
58	pleistocen	9×10^{-3}
59	understand	8.7×10^{-3}
60	middl	8.7×10^{-3}
61	moral	8.6×10^{-3}
62	simul	8.5×10^{-3}
63	econom	8.4×10^{-3}
64	christian	8.4×10^{-3}
65	examin	8.4×10^{-3}
66	popul	8.3×10^{-3}
67	cal	8.3×10^{-3}
68	research	8.2×10^{-3}
69	narrat	8.1×10^{-3}
70	northern	8.1×10^{-3}
71	religi	8×10^{-3}
72	holocen	8×10^{-3}
73	scholar	7.9×10^{-3}
74	roman	7.8×10^{-3}
75	engag	7.8×10^{-3}
76	ideolog	7.8×10^{-3}
77	evid	7.5×10^{-3}
78	southern	7.4×10^{-3}
79	effici	7.4×10^{-3}
80	question	7.3×10^{-3}
81	villag	7.3×10^{-3}
82	central	7.3×10^{-3}
83	temperatur	7.2×10^{-3}
84	artefact	7.1×10^{-3}
85	gather	7.1×10^{-3}
86	algorithm	7×10^{-3}
87	hunt	7×10^{-3}
88	past	7×10^{-3}
89	faunal	6.9×10^{-3}
90	north	6.9×10^{-3}
91	primat	6.9×10^{-3}
92	focus	6.8×10^{-3}
93	bone	6.8×10^{-3}
94	document	6.8×10^{-3}
95	isotop	6.6×10^{-3}
96	suggest	6.5×10^{-3}
97	place	6.5×10^{-3}
98	optim	6.5×10^{-3}
99	mediev	6.4×10^{-3}
100	induc	6.3×10^{-3}

TABLE D.12. The list of the top 100 words in the category Archaeology with RIGs

No.	Word	RIG	No.	Word	RIG
1	archaeolog	2.9×10^{-1}	51	eastern	1.3×10^{-2}
2	excav	8.9×10^{-2}	52	cave	1.3×10^{-2}
3	site	7.6×10^{-2}	53	paint	1.3×10^{-2}
4	centuri	7.2×10^{-2}	54	materi	1.2×10^{-2}
5	neolith	5.1×10^{-2}	55	past	1.2×10^{-2}
6	archaeologist	4.9×10^{-2}	56	argu	1.2×10^{-2}
7	date	4.9×10^{-2}	57	domest	1.2×10^{-2}
8	bronz	4.8×10^{-2}	58	ceram	1.2×10^{-2}
9	ancient	4.6×10^{-2}	59	valley	1.2×10^{-2}
10	prehistor	4.2×10^{-2}	60	societi	1.2×10^{-2}
11	roman	4.1×10^{-2}	61	modern	1.1×10^{-2}
12	settlement	4.1×10^{-2}	62	articl	1.1×10^{-2}
13	assemblag	3.7×10^{-2}	63	perform	1.1×10^{-2}
14	late	3.7×10^{-2}	64	document	1.1×10^{-2}
15	monument	3.3×10^{-2}	65	preserv	1.1×10^{-2}
16	cultur	3.2×10^{-2}	66	inscript	1.1×10^{-2}
17	stone	3.1×10^{-2}	67	place	1×10^{-2}
18	histor	2.9×10^{-2}	68	polit	1×10^{-2}
19	chronolog	2.7×10^{-2}	69	iron	1×10^{-2}
20	burial	2.6×10^{-2}	70	town	1×10^{-2}
21	earli	2.6×10^{-2}	71	decor	1×10^{-2}
22	artefact	2.6×10^{-2}	72	record	1×10^{-2}
23	landscap	2.4×10^{-2}	73	conclus	9.8×10^{-3}
24	ritual	2.3×10^{-2}	74	east	9.7×10^{-3}
25	heritag	2.2×10^{-2}	75	europ	9.7×10^{-3}
26	radiocarbon	2.2×10^{-2}	76	reduc	9.7×10^{-3}
27	subsist	2×10^{-2}	77	isotop	9.6×10^{-3}
28	period	1.9×10^{-2}	78	histori	9.5×10^{-3}
29	southern	1.9×10^{-2}	79	templ	9.5×10^{-3}
30	interpret	1.9×10^{-2}	80	citi	9.5×10^{-3}
31	evid	1.8×10^{-2}	81	earliest	9.4×10^{-3}
32	cal	1.8×10^{-2}	82	bone	9.4×10^{-3}
33	effect	1.8×10^{-2}	83	central	9.4×10^{-3}
34	context	1.7×10^{-2}	84	pleistocen	9.2×10^{-3}
35	remain	1.6×10^{-2}	85	gather	9.2×10^{-3}
36	mediev	1.6×10^{-2}	86	age	9.1×10^{-3}
37	artifact	1.6×10^{-2}	87	locat	9×10^{-3}
38	middl	1.6×10^{-2}	88	west	9×10^{-3}
39	occup	1.6×10^{-2}	89	island	9×10^{-3}
40	patient	1.5×10^{-2}	90	19th	8.8×10^{-3}
41	northern	1.5×10^{-2}	91	hunt	8.7×10^{-3}
42	social	1.5×10^{-2}	92	rock	8.5×10^{-3}
43	north	1.4×10^{-2}	93	method	8.4×10^{-3}
44	holocen	1.4×10^{-2}	94	clinic	8.4×10^{-3}
45	cell	1.4×10^{-2}	95	origin	8.4×10^{-3}
46	region	1.3×10^{-2}	96	ethnograph	8.3×10^{-3}
47	museum	1.3×10^{-2}	97	area	8.3×10^{-3}
48	antiqu	1.3×10^{-2}	98	control	8.2×10^{-3}
49	faunal	1.3×10^{-2}	99	typolog	8.1×10^{-3}
50	south	1.3×10^{-2}	100	reconstruct	8.1×10^{-3}

Table D.13. The list of the top 100 words in the category Architecture with RIGs

No.	Word	RIG	No.	Word	RIG
1	architectur	1.3×10^{-1}	51	church	8.5×10^{-3}
2	build	9.6×10^{-2}	52	studio	8.3×10^{-3}
3	architect	6.5×10^{-2}	53	protein	8.3×10^{-3}
4	heritag	6.2×10^{-2}	54	context	8.2×10^{-3}
5	urban	4.8×10^{-2}	55	place	8.2×10^{-3}
6	vernacular	4.5×10^{-2}	56	draw	8.2×10^{-3}
7	hous	4.4×10^{-2}	57	focus	8×10^{-3}
8	centuri	4.2×10^{-2}	58	develop	8×10^{-3}
9	design	3.9×10^{-2}	59	rural	7.9×10^{-3}
10	histor	3.8×10^{-2}	60	practic	7.7×10^{-3}
11	citi	3.7×10^{-2}	61	creat	7.6×10^{-3}
12	project	3.5×10^{-2}	62	treatment	7.6×10^{-3}
13	construct	3.4×10^{-2}	63	archaeolog	7.3×10^{-3}
14	built	2.3×10^{-2}	64	world	7.3×10^{-3}
15	plan	2.3×10^{-2}	65	stone	7.3×10^{-3}
16	landscap	2.1×10^{-2}	66	effect	7×10^{-3}
17	cultur	2.1×10^{-2}	67	restor	7×10^{-3}
18	research	1.9×10^{-2}	68	diseas	7×10^{-3}
19	monument	1.9×10^{-2}	69	tool	7×10^{-3}
20	paper	1.8×10^{-2}	70	environment	7×10^{-3}
21	space	1.8×10^{-2}	71	floor	7×10^{-3}
22	town	1.7×10^{-2}	72	sampl	6.9×10^{-3}
23	contemporari	1.7×10^{-2}	73	histori	6.9×10^{-3}
24	result	1.6×10^{-2}	74	detect	6.8×10^{-3}
25	territori	1.4×10^{-2}	75	decreas	6.7×10^{-3}
26	patient	1.4×10^{-2}	76	idea	6.7×10^{-3}
27	ancient	1.3×10^{-2}	77	new	6.7×10^{-3}
28	modern	1.3×10^{-2}	78	nineteenth	6.6×10^{-3}
29	aesthet	1.3×10^{-2}	79	technolog	6.6×10^{-3}
30	settlement	1.3×10^{-2}	80	induc	6.6×10^{-3}
31	cell	1.2×10^{-2}	81	creation	6.6×10^{-3}
32	tradit	1.2×10^{-2}	82	acid	6.6×10^{-3}
33	typolog	1.2×10^{-2}	83	higher	6.5×10^{-3}
34	sustain	1.1×10^{-2}	84	survey	6.5×10^{-3}
35	twentieth	1×10^{-2}	85	dwell	6.4×10^{-3}
36	conserv	9.7×10^{-3}	86	way	6.3×10^{-3}
37	digit	9.6×10^{-3}	87	social	6.3×10^{-3}
38	preserv	9.5×10^{-3}	88	obtain	6.3×10^{-3}
39	conclus	9.4×10^{-3}	89	high	6.1×10^{-3}
40	show	9.3×10^{-3}	90	gene	6.1×10^{-3}
41	residenti	9.2×10^{-3}	91	wall	6.1×10^{-3}
42	villag	9.1×10^{-3}	92	interior	6×10^{-3}
43	environ	9×10^{-3}	93	itali	5.9×10^{-3}
44	today	8.9×10^{-3}	94	museum	5.9×10^{-3}
45	clinic	8.8×10^{-3}	95	methodolog	5.9×10^{-3}
46	artist	8.7×10^{-3}	96	street	5.8×10^{-3}
47	materi	8.7×10^{-3}	97	area	5.8×10^{-3}
48	rate	8.6×10^{-3}	98	visual	5.8×10^{-3}
49	timber	8.6×10^{-3}	99	concept	5.7×10^{-3}
50	explor	8.6×10^{-3}	100	geometr	5.7×10^{-3}

TABLE D.14. The list of the top 100 words in the category Area Studies with RIGs

No.	Word	RIG
1	polit	1.5×10^{-1}
2	articl	1.1×10^{-1}
3	argu	6.5×10^{-2}
4	govern	5.1×10^{-2}
5	polici	4.4×10^{-2}
6	result	3.7×10^{-2}
7	social	3.7×10^{-2}
8	nation	3.5×10^{-2}
9	econom	3.3×10^{-2}
10	china	3.2×10^{-2}
11	democrat	3.1×10^{-2}
12	parti	2.9×10^{-2}
13	method	2.9×10^{-2}
14	war	2.8×10^{-2}
15	societi	2.7×10^{-2}
16	countri	2.7×10^{-2}
17	democraci	2.7×10^{-2}
18	reform	2.5×10^{-2}
19	discours	2.5×10^{-2}
20	use	2.4×10^{-2}
21	africa	2.3×10^{-2}
22	institut	2.2×10^{-2}
23	state	2.1×10^{-2}
24	centuri	2.1×10^{-2}
25	foreign	2×10^{-2}
26	debat	2×10^{-2}
27	civil	2×10^{-2}
28	elit	2×10^{-2}
29	contemporari	1.9×10^{-2}
30	draw	1.9×10^{-2}
31	actor	1.9×10^{-2}
32	chines	1.9×10^{-2}
33	communist	1.8×10^{-2}
34	histor	1.8×10^{-2}
35	public	1.8×10^{-2}
36	struggl	1.8×10^{-2}
37	ideolog	1.7×10^{-2}
38	conflict	1.7×10^{-2}
39	offici	1.7×10^{-2}
40	scholar	1.7×10^{-2}
41	presid	1.6×10^{-2}
42	economi	1.6×10^{-2}
43	leader	1.6×10^{-2}
44	liber	1.6×10^{-2}
45	african	1.6×10^{-2}
46	narrat	1.5×10^{-2}
47	cultur	1.5×10^{-2}
48	perform	1.5×10^{-2}
49	cell	1.5×10^{-2}
50	nationalist	1.5×10^{-2}
51	elect	1.5×10^{-2}
52	peopl	1.5×10^{-2}
53	patient	1.4×10^{-2}
54	militari	1.4×10^{-2}
55	south	1.4×10^{-2}
56	protest	1.4×10^{-2}
57	religi	1.4×10^{-2}
58	union	1.4×10^{-2}
59	soviet	1.4×10^{-2}
60	islam	1.3×10^{-2}
61	histori	1.3×10^{-2}
62	peac	1.3×10^{-2}
63	engag	1.3×10^{-2}
64	simul	1.3×10^{-2}
65	asia	1.3×10^{-2}
66	question	1.3×10^{-2}
67	world	1.3×10^{-2}
68	transnat	1.3×10^{-2}
69	authoritarian	1.3×10^{-2}
70	context	1.2×10^{-2}
71	seek	1.2×10^{-2}
72	arab	1.2×10^{-2}
73	crisi	1.2×10^{-2}
74	intern	1.2×10^{-2}
75	citizen	1.2×10^{-2}
76	contest	1.2×10^{-2}
77	focus	1.2×10^{-2}
78	author	1.2×10^{-2}
79	obtain	1.2×10^{-2}
80	violenc	1.2×10^{-2}
81	domest	1.2×10^{-2}
82	temperatur	1.2×10^{-2}
83	examin	1.1×10^{-2}
84	attempt	1.1×10^{-2}
85	russia	1.1×10^{-2}
86	high	1.1×10^{-2}
87	elector	1.1×10^{-2}
88	twentieth	1.1×10^{-2}
89	russian	1.1×10^{-2}
90	measur	1.1×10^{-2}
91	conclus	1.1×10^{-2}
92	secur	1.1×10^{-2}
93	turkish	1.1×10^{-2}
94	migrant	1.1×10^{-2}
95	way	1.1×10^{-2}
96	issu	1×10^{-2}
97	studi	1×10^{-2}
98	ident	1×10^{-2}
99	rural	1×10^{-2}
100	ethnic	1×10^{-2}

Table D.15. The list of the top 100 words in the category Art with RIGs

No.	Word	RIG	No.	Word	RIG
1	artist	1.2×10^{-1}	51	creat	1×10^{-2}
2	art	1.1×10^{-1}	52	interpret	9.4×10^{-3}
3	centuri	6.1×10^{-2}	53	question	9.4×10^{-3}
4	paint	6×10^{-2}	54	control	9.3×10^{-3}
5	articl	5.4×10^{-2}	55	rate	9.3×10^{-3}
6	aesthet	4.4×10^{-2}	56	decreas	8.8×10^{-3}
7	artwork	4.3×10^{-2}	57	embodi	8.5×10^{-3}
8	cultur	3.4×10^{-2}	58	modernist	8.4×10^{-3}
9	histor	3.2×10^{-2}	59	imag	8.4×10^{-3}
10	argu	3×10^{-2}	60	idea	8.4×10^{-3}
11	museum	2.7×10^{-2}	61	reduc	8.4×10^{-3}
12	essay	2.6×10^{-2}	62	context	8.3×10^{-3}
13	contemporari	2.6×10^{-2}	63	viewer	8.3×10^{-3}
14	photograph	2.4×10^{-2}	64	simul	8.2×10^{-3}
15	creativ	2.4×10^{-2}	65	world	8×10^{-3}
16	result	2.4×10^{-2}	66	audienc	7.9×10^{-3}
17	work	2.2×10^{-2}	67	genr	7.9×10^{-3}
18	heritag	2.1×10^{-2}	68	artefact	7.9×10^{-3}
19	visual	1.9×10^{-2}	69	sixteenth	7.8×10^{-3}
20	monument	1.8×10^{-2}	70	craft	7.8×10^{-3}
21	painter	1.8×10^{-2}	71	creation	7.8×10^{-3}
22	author	1.7×10^{-2}	72	show	7.8×10^{-3}
23	photographi	1.6×10^{-2}	73	system	7.8×10^{-3}
24	histori	1.6×10^{-2}	74	focus	7.6×10^{-3}
25	philosoph	1.6×10^{-2}	75	argument	7.6×10^{-3}
26	draw	1.6×10^{-2}	76	design	7.4×10^{-3}
27	way	1.5×10^{-2}	77	ancient	7.4×10^{-3}
28	polit	1.5×10^{-2}	78	figur	7.2×10^{-3}
29	patient	1.4×10^{-2}	79	imagin	7.2×10^{-3}
30	nineteenth	1.4×10^{-2}	80	clinic	7.2×10^{-3}
31	church	1.4×10^{-2}	81	think	7.2×10^{-3}
32	practic	1.4×10^{-2}	82	measur	7.1×10^{-3}
33	engag	1.3×10^{-2}	83	claim	7.1×10^{-3}
34	concept	1.3×10^{-2}	84	protein	6.9×10^{-3}
35	modern	1.3×10^{-2}	85	higher	6.9×10^{-3}
36	cell	1.2×10^{-2}	86	low	6.9×10^{-3}
37	decor	1.2×10^{-2}	87	mediev	6.7×10^{-3}
38	increas	1.2×10^{-2}	88	christ	6.7×10^{-3}
39	project	1.2×10^{-2}	89	notion	6.6×10^{-3}
40	depict	1.2×10^{-2}	90	improv	6.6×10^{-3}
41	explor	1.2×10^{-2}	91	induc	6.6×10^{-3}
42	high	1.1×10^{-2}	92	ratio	6.5×10^{-3}
43	method	1.1×10^{-2}	93	style	6.5×10^{-3}
44	twentieth	1.1×10^{-2}	94	roman	6.5×10^{-3}
45	narrat	1.1×10^{-2}	95	diseas	6.3×10^{-3}
46	social	1.1×10^{-2}	96	build	6.3×10^{-3}
47	effect	1×10^{-2}	97	represent	6.3×10^{-3}
48	conclus	1×10^{-2}	98	thing	6.2×10^{-3}
49	music	1×10^{-2}	99	german	6.2×10^{-3}
50	compar	1×10^{-2}	100	conserv	6.2×10^{-3}

Table D.16. The list of the top 100 words in the category Asian Studies with RIGs

No.	Word	RIG	No.	Word	RIG
1	articl	9.8×10^{-2}	51	examin	1.4×10^{-2}
2	argu	6.5×10^{-2}	52	china	1.4×10^{-2}
3	polit	6.3×10^{-2}	53	india	1.4×10^{-2}
4	centuri	6.2×10^{-2}	54	offici	1.4×10^{-2}
5	buddhist	5.2×10^{-2}	55	christian	1.4×10^{-2}
6	result	4.5×10^{-2}	56	cell	1.3×10^{-2}
7	religi	4×10^{-2}	57	patient	1.3×10^{-2}
8	text	3.7×10^{-2}	58	test	1.3×10^{-2}
9	islam	3.6×10^{-2}	59	nineteenth	1.3×10^{-2}
10	scholar	3.4×10^{-2}	60	nationalist	1.3×10^{-2}
11	chines	3.2×10^{-2}	61	intellectu	1.3×10^{-2}
12	ritual	3×10^{-2}	62	asian	1.3×10^{-2}
13	essay	3×10^{-2}	63	east	1.3×10^{-2}
14	korean	2.9×10^{-2}	64	rate	1.2×10^{-2}
15	literari	2.8×10^{-2}	65	obtain	1.2×10^{-2}
16	method	2.8×10^{-2}	66	data	1.2×10^{-2}
17	histor	2.8×10^{-2}	67	attempt	1.2×10^{-2}
18	contemporari	2.8×10^{-2}	68	arab	1.2×10^{-2}
19	histori	2.8×10^{-2}	69	debat	1.2×10^{-2}
20	discours	2.7×10^{-2}	70	write	1.2×10^{-2}
21	modern	2.6×10^{-2}	71	draw	1.2×10^{-2}
22	use	2.5×10^{-2}	72	question	1.2×10^{-2}
23	societi	2.4×10^{-2}	73	effici	1.2×10^{-2}
24	cultur	2.2×10^{-2}	74	reduc	1.2×10^{-2}
25	korea	2.1×10^{-2}	75	genr	1.2×10^{-2}
26	narrat	2.1×10^{-2}	76	way	1.2×10^{-2}
27	effect	2×10^{-2}	77	stori	1.2×10^{-2}
28	tradit	1.9×10^{-2}	78	asia	1.2×10^{-2}
29	coloni	1.8×10^{-2}	79	notion	1.2×10^{-2}
30	japanes	1.8×10^{-2}	80	idea	1.1×10^{-2}
31	war	1.8×10^{-2}	81	govern	1.1×10^{-2}
32	moral	1.8×10^{-2}	82	improv	1.1×10^{-2}
33	author	1.8×10^{-2}	83	indian	1.1×10^{-2}
34	imperi	1.8×10^{-2}	84	doctrin	1.1×10^{-2}
35	ideolog	1.8×10^{-2}	85	south	1.1×10^{-2}
36	high	1.8×10^{-2}	86	evalu	1.1×10^{-2}
37	social	1.7×10^{-2}	87	context	1.1×10^{-2}
38	philosoph	1.7×10^{-2}	88	british	1.1×10^{-2}
39	nation	1.7×10^{-2}	89	read	1.1×10^{-2}
40	late	1.7×10^{-2}	90	world	1.1×10^{-2}
41	languag	1.7×10^{-2}	91	ident	1.1×10^{-2}
42	muslim	1.6×10^{-2}	92	view	1×10^{-2}
43	philosophi	1.6×10^{-2}	93	temperatur	1×10^{-2}
44	twentieth	1.5×10^{-2}	94	templ	1×10^{-2}
45	increas	1.5×10^{-2}	95	elit	1×10^{-2}
46	measur	1.5×10^{-2}	96	court	1×10^{-2}
47	ottoman	1.5×10^{-2}	97	higher	1×10^{-2}
48	inscript	1.5×10^{-2}	98	low	1×10^{-2}
49	model	1.4×10^{-2}	99	royal	1×10^{-2}
50	religion	1.4×10^{-2}	100	simul	1×10^{-2}

Table D.17. The list of the top 100 words in the category Astronomy and Astrophysics with RIGs

No.	Word	RIG	No.	Word	RIG
1	star	1.5×10^{-1}	51	dot	1.6×10^{-2}
2	galaxi	1.1×10^{-1}	52	find	1.6×10^{-2}
3	telescop	9.9×10^{-2}	53	perturb	1.6×10^{-2}
4	stellar	9.8×10^{-2}	54	radiat	1.6×10^{-2}
5	galact	6.6×10^{-2}	55	spacetim	1.6×10^{-2}
6	mass	5.9×10^{-2}	56	magnet	1.5×10^{-2}
7	cosmolog	5.4×10^{-2}	57	densiti	1.5×10^{-2}
8	luminos	5.1×10^{-2}	58	radial	1.5×10^{-2}
9	redshift	5×10^{-2}	59	radius	1.5×10^{-2}
10	observ	4.9×10^{-2}	60	energi	1.5×10^{-2}
11	cosmic	4.5×10^{-2}	61	binari	1.5×10^{-2}
12	gravit	4.5×10^{-2}	62	tev	1.5×10^{-2}
13	accret	3.9×10^{-2}	63	radio	1.5×10^{-2}
14	observatori	3.7×10^{-2}	64	rotat	1.4×10^{-2}
15	circl	3.6×10^{-2}	65	spectrum	1.4×10^{-2}
16	sky	3.4×10^{-2}	66	constrain	1.4×10^{-2}
17	field	3.3×10^{-2}	67	planck	1.4×10^{-2}
18	solar	3.3×10^{-2}	68	clinic	1.4×10^{-2}
19	dark	3.2×10^{-2}	69	momentum	1.4×10^{-2}
20	dwarf	3.1×10^{-2}	70	line	1.4×10^{-2}
21	orbit	3×10^{-2}	71	decay	1.4×10^{-2}
22	massiv	2.8×10^{-2}	72	sigma	1.4×10^{-2}
23	similar	2.7×10^{-2}	73	ray	1.4×10^{-2}
24	scalar	2.6×10^{-2}	74	scale	1.4×10^{-2}
25	spectral	2.6×10^{-2}	75	evalu	1.4×10^{-2}
26	astrophys	2.5×10^{-2}	76	diseas	1.4×10^{-2}
27	graviti	2.5×10^{-2}	77	mission	1.3×10^{-2}
28	flux	2.4×10^{-2}	78	higg	1.3×10^{-2}
29	emiss	2.4×10^{-2}	79	infrar	1.3×10^{-2}
30	evolut	2.4×10^{-2}	80	control	1.3×10^{-2}
31	patient	2.4×10^{-2}	81	optic	1.3×10^{-2}
32	disk	2.4×10^{-2}	82	kev	1.2×10^{-2}
33	dust	2.3×10^{-2}	83	spectroscop	1.2×10^{-2}
34	matter	2.3×10^{-2}	84	survey	1.2×10^{-2}
35	bright	2.2×10^{-2}	85	fit	1.2×10^{-2}
36	veloc	2.2×10^{-2}	86	protein	1.2×10^{-2}
37	neutrino	1.9×10^{-2}	87	region	1.2×10^{-2}
38	hole	1.9×10^{-2}	88	qcd	1.2×10^{-2}
39	quark	1.8×10^{-2}	89	treatment	1.2×10^{-2}
40	cell	1.8×10^{-2}	90	spacecraft	1.1×10^{-2}
41	black	1.8×10^{-2}	91	sourc	1.1×10^{-2}
42	gev	1.8×10^{-2}	92	ionospher	1.1×10^{-2}
43	baryon	1.8×10^{-2}	93	wind	1.1×10^{-2}
44	spectra	1.8×10^{-2}	94	physic	1.1×10^{-2}
45	near	1.7×10^{-2}	95	hadron	1.1×10^{-2}
46	relativist	1.6×10^{-2}	96	gaug	1.1×10^{-2}
47	space	1.6×10^{-2}	97	atmospher	1.1×10^{-2}
48	model	1.6×10^{-2}	98	instrument	1.1×10^{-2}
49	larg	1.6×10^{-2}	99	earth	1.1×10^{-2}
50	angular	1.6×10^{-2}	100	constraint	1.1×10^{-2}

TABLE D.18. The list of the top 100 words in the category Audiology and Speech-Language Pathology with RIGs

No.	Word	RIG
1	speech	1.4×10^{-1}
2	hear	1.3×10^{-1}
3	acoust	8.9×10^{-2}
4	listen	7.9×10^{-2}
5	sound	7.2×10^{-2}
6	auditori	6×10^{-2}
7	languag	5.8×10^{-2}
8	cochlear	5.2×10^{-2}
9	word	4.3×10^{-2}
10	nois	4×10^{-2}
11	sentenc	3.9×10^{-2}
12	ear	3.7×10^{-2}
13	children	3.3×10^{-2}
14	tone	3.1×10^{-2}
15	particip	3.1×10^{-2}
16	vowel	2.9×10^{-2}
17	frequenc	2.9×10^{-2}
18	phonolog	2.6×10^{-2}
19	conson	2.6×10^{-2}
20	purpos	2.5×10^{-2}
21	speaker	2.4×10^{-2}
22	task	2.2×10^{-2}
23	stimuli	2.2×10^{-2}
24	vocal	2.1×10^{-2}
25	syllabl	2×10^{-2}
26	percept	1.9×10^{-2}
27	lexic	1.8×10^{-2}
28	conclus	1.8×10^{-2}
29	impair	1.8×10^{-2}
30	adult	1.8×10^{-2}
31	threshold	1.8×10^{-2}
32	perceptu	1.7×10^{-2}
33	cue	1.7×10^{-2}
34	khz	1.6×10^{-2}
35	measur	1.6×10^{-2}
36	linguist	1.5×10^{-2}
37	english	1.5×10^{-2}
38	phonet	1.5×10^{-2}
39	speak	1.4×10^{-2}
40	age	1.4×10^{-2}
41	normal	1.4×10^{-2}
42	voic	1.3×10^{-2}
43	spoken	1.2×10^{-2}
44	utter	1.2×10^{-2}
45	across	1.2×10^{-2}
46	pathologist	1.2×10^{-2}
47	implant	1.2×10^{-2}
48	amplitud	1.1×10^{-2}
49	recept	1×10^{-2}
50	aid	1×10^{-2}

No.	Word	RIG
51	stimulus	9.5×10^{-3}
52	pitch	9.2×10^{-3}
53	intellig	9.2×10^{-3}
54	result	9×10^{-3}
55	skill	8.9×10^{-3}
56	record	8.9×10^{-3}
57	examin	8.9×10^{-3}
58	tempor	8.8×10^{-3}
59	recognit	8.7×10^{-3}
60	studi	8.7×10^{-3}
61	difficulti	8.6×10^{-3}
62	differ	8.5×10^{-3}
63	verb	8.5×10^{-3}
64	individu	8.4×10^{-3}
65	reader	8.4×10^{-3}
66	evok	8.3×10^{-3}
67	bilingu	8.1×10^{-3}
68	temperatur	8×10^{-3}
69	grammat	7.9×10^{-3}
70	whether	7.7×10^{-3}
71	vocabulari	7.7×10^{-3}
72	score	7.7×10^{-3}
73	preschool	7.7×10^{-3}
74	music	7.7×10^{-3}
75	mask	7.6×10^{-3}
76	test	7.5×10^{-3}
77	read	7.5×10^{-3}
78	typic	7.5×10^{-3}
79	noun	7.2×10^{-3}
80	loss	7.1×10^{-3}
81	subject	7.1×10^{-3}
82	autism	6.9×10^{-3}
83	cell	6.7×10^{-3}
84	group	6.5×10^{-3}
85	abil	6.5×10^{-3}
86	suggest	6.4×10^{-3}
87	disord	6.2×10^{-3}
88	assess	6.2×10^{-3}
89	elicit	6.1×10^{-3}
90	communic	6.1×10^{-3}
91	protein	5.9×10^{-3}
92	object	5.9×10^{-3}
93	wave	5.9×10^{-3}
94	match	5.8×10^{-3}
95	perceiv	5.7×10^{-3}
96	acid	5.7×10^{-3}
97	syntact	5.7×10^{-3}
98	poorer	5.7×10^{-3}
99	concentr	5.6×10^{-3}
100	retest	5.6×10^{-3}

TABLE D.19. The list of the top 100 words in the category Automation and Control Systems with RIGs

No.	Word	RIG	No.	Word	RIG
1	paper	9.7×10^{-2}	51	scheme	1×10^{-2}
2	propos	7×10^{-2}	52	examin	1×10^{-2}
3	system	5.7×10^{-2}	53	trajectori	9.9×10^{-3}
4	control	4.4×10^{-2}	54	year	9.9×10^{-3}
5	robot	3.9×10^{-2}	55	constraint	9.5×10^{-3}
6	problem	3.8×10^{-2}	56	background	9.4×10^{-3}
7	algorithm	3.7×10^{-2}	57	converg	9.3×10^{-3}
8	simul	3.6×10^{-2}	58	motion	9.2×10^{-3}
9	lyapunov	2.6×10^{-2}	59	acid	9.2×10^{-3}
10	nonlinear	2.6×10^{-2}	60	inequ	9.1×10^{-3}
11	design	2.4×10^{-2}	61	assess	9.1×10^{-3}
12	conclus	2.3×10^{-2}	62	network	9×10^{-3}
13	studi	2.2×10^{-2}	63	uncertainti	8.9×10^{-3}
14	feedback	2.2×10^{-2}	64	gene	8.8×10^{-3}
15	loop	1.9×10^{-2}	65	speci	8.8×10^{-3}
16	base	1.9×10^{-2}	66	indic	8.7×10^{-3}
17	dynam	1.8×10^{-2}	67	reveal	8.6×10^{-3}
18	patient	1.7×10^{-2}	68	verifi	8.6×10^{-3}
19	track	1.7×10^{-2}	69	machin	8.6×10^{-3}
20	actuat	1.7×10^{-2}	70	real	8.4×10^{-3}
21	optim	1.7×10^{-2}	71	discret	8.4×10^{-3}
22	guarante	1.6×10^{-2}	72	fault	8.3×10^{-3}
23	linear	1.6×10^{-2}	73	filter	8.2×10^{-3}
24	exampl	1.6×10^{-2}	74	may	8.2×10^{-3}
25	suggest	1.6×10^{-2}	75	activ	8.1×10^{-3}
26	input	1.6×10^{-2}	76	concentr	7.8×10^{-3}
27	output	1.5×10^{-2}	77	implement	7.8×10^{-3}
28	robust	1.5×10^{-2}	78	higher	7.7×10^{-3}
29	solv	1.5×10^{-2}	79	switch	7.5×10^{-3}
30	signific	1.4×10^{-2}	80	decreas	7.5×10^{-3}
31	disturb	1.4×10^{-2}	81	adapt	7.3×10^{-3}
32	illustr	1.4×10^{-2}	82	state	7.3×10^{-3}
33	treatment	1.4×10^{-2}	83	introduc	7.3×10^{-3}
34	stabil	1.3×10^{-2}	84	kalman	7.2×10^{-3}
35	cell	1.3×10^{-2}	85	evid	7.2×10^{-3}
36	found	1.3×10^{-2}	86	can	7.1×10^{-3}
37	model	1.2×10^{-2}	87	asymptot	7.1×10^{-3}
38	associ	1.2×10^{-2}	88	potenti	7.1×10^{-3}
39	increas	1.2×10^{-2}	89	molecular	7×10^{-3}
40	error	1.2×10^{-2}	90	speed	7×10^{-3}
41	report	1.2×10^{-2}	91	numer	7×10^{-3}
42	protein	1.2×10^{-2}	92	comput	7×10^{-3}
43	fuzzi	1.1×10^{-2}	93	final	6.9×10^{-3}
44	age	1.1×10^{-2}	94	month	6.9×10^{-3}
45	clinic	1.1×10^{-2}	95	induc	6.8×10^{-3}
46	group	1.1×10^{-2}	96	role	6.8×10^{-3}
47	diseas	1×10^{-2}	97	perform	6.7×10^{-3}
48	approach	1×10^{-2}	98	given	6.6×10^{-3}
49	sensor	1×10^{-2}	99	outcom	6.5×10^{-3}
50	vehicl	1×10^{-2}	100	compens	6.5×10^{-3}

TABLE D.20. The list of the top 100 words in the category Behavioral Sciences with RIGs

No.	Word	RIG
1	behavior	5.2×10^{-2}
2	anim	4.4×10^{-2}
3	task	4×10^{-2}
4	male	3.7×10^{-2}
5	suggest	3.7×10^{-2}
6	cognit	3.5×10^{-2}
7	brain	3.4×10^{-2}
8	whether	3.1×10^{-2}
9	rat	3×10^{-2}
10	maze	2.9×10^{-2}
11	memori	2.9×10^{-2}
12	stimuli	2.7×10^{-2}
13	femal	2.7×10^{-2}
14	cue	2.7×10^{-2}
15	test	2.7×10^{-2}
16	cortex	2.7×10^{-2}
17	stimulus	2.5×10^{-2}
18	food	2.4×10^{-2}
19	individu	2.4×10^{-2}
20	behaviour	2.4×10^{-2}
21	paper	2.3×10^{-2}
22	may	2.3×10^{-2}
23	mate	2.3×10^{-2}
24	respons	2.2×10^{-2}
25	social	2.2×10^{-2}
26	reward	2.1×10^{-2}
27	learn	2.1×10^{-2}
28	hippocampus	1.9×10^{-2}
29	deficit	1.9×10^{-2}
30	impair	1.9×10^{-2}
31	anxieti	1.9×10^{-2}
32	prefront	1.7×10^{-2}
33	conspecif	1.7×10^{-2}
34	particip	1.7×10^{-2}
35	prefer	1.6×10^{-2}
36	emot	1.6×10^{-2}
37	relat	1.5×10^{-2}
38	disord	1.5×10^{-2}
39	adult	1.5×10^{-2}
40	paradigm	1.5×10^{-2}
41	fear	1.4×10^{-2}
42	associ	1.4×10^{-2}
43	studi	1.4×10^{-2}
44	epilepsi	1.4×10^{-2}
45	predat	1.3×10^{-2}
46	hypothesi	1.3×10^{-2}
47	dure	1.3×10^{-2}
48	neural	1.3×10^{-2}
49	evid	1.2×10^{-2}
50	howev	1.2×10^{-2}

No.	Word	RIG
51	bird	1.2×10^{-2}
52	eat	1.2×10^{-2}
53	forag	1.2×10^{-2}
54	either	1.2×10^{-2}
55	amygdala	1.2×10^{-2}
56	examin	1.2×10^{-2}
57	affect	1.1×10^{-2}
58	locomotor	1.1×10^{-2}
59	find	1.1×10^{-2}
60	frontal	1.1×10^{-2}
61	aggress	1.1×10^{-2}
62	sex	1.1×10^{-2}
63	avers	1.1×10^{-2}
64	seizur	1.1×10^{-2}
65	offspr	1.1×10^{-2}
66	trait	1.1×10^{-2}
67	group	1.1×10^{-2}
68	intak	1×10^{-2}
69	hippocamp	1×10^{-2}
70	respond	1×10^{-2}
71	choic	1×10^{-2}
72	visual	1×10^{-2}
73	method	1×10^{-2}
74	alter	1×10^{-2}
75	depress	1×10^{-2}
76	train	9.8×10^{-3}
77	session	9.7×10^{-3}
78	sexual	9.7×10^{-3}
79	mice	9.4×10^{-3}
80	motor	9.4×10^{-3}
81	previous	9.3×10^{-3}
82	neuron	9.3×10^{-3}
83	discrimin	9.3×10^{-3}
84	effect	9.2×10^{-3}
85	differ	9.2×10^{-3}
86	manipul	9.1×10^{-3}
87	antagonist	9.1×10^{-3}
88	dopamin	9×10^{-3}
89	day	9×10^{-3}
90	healthi	9×10^{-3}
91	rodent	8.9×10^{-3}
92	spent	8.9×10^{-3}
93	receptor	8.8×10^{-3}
94	agonist	8.6×10^{-3}
95	gyrus	8.6×10^{-3}
96	surfac	8.4×10^{-3}
97	later	8.3×10^{-3}
98	like	8.3×10^{-3}
99	latenc	8.2×10^{-3}
100	fmri	8.1×10^{-3}

TABLE D.21. The list of the top 100 words in the category Biochemical Research Methods with RIGs

No.	Word	RIG	No.	Word	RIG
1	chromatographi	5.9×10^{-2}	51	drug	9.4×10^{-3}
2	spectrometri	5.1×10^{-2}	52	analysi	9.4×10^{-3}
3	protein	4.2×10^{-2}	53	quadrupol	9.4×10^{-3}
4	proteom	3.4×10^{-2}	54	molecular	9.2×10^{-3}
5	mass	2.8×10^{-2}	55	molecul	9×10^{-3}
6	detect	2.8×10^{-2}	56	rapid	8.7×10^{-3}
7	biolog	2.8×10^{-2}	57	limit	8.7×10^{-3}
8	quantif	2.7×10^{-2}	58	appli	8.5×10^{-3}
9	liquid	2.7×10^{-2}	59	allow	8.3×10^{-3}
10	column	2.3×10^{-2}	60	tool	8.2×10^{-3}
11	chromatograph	2.3×10^{-2}	61	dna	8.1×10^{-3}
12	tandem	2.2×10^{-2}	62	ion	8×10^{-3}
13	sampl	2×10^{-2}	63	high	8×10^{-3}
14	electrospray	2×10^{-2}	64	target	7.9×10^{-3}
15	analyt	1.7×10^{-2}	65	gene	7.9×10^{-3}
16	assay	1.7×10^{-2}	66	screen	7.7×10^{-3}
17	separ	1.7×10^{-2}	67	rsd	7.5×10^{-3}
18	ioniz	1.7×10^{-2}	68	angstrom	7.4×10^{-3}
19	elut	1.7×10^{-2}	69	identifi	7.4×10^{-3}
20	extract	1.5×10^{-2}	70	capillari	7.4×10^{-3}
21	quantit	1.5×10^{-2}	71	methanol	7.3×10^{-3}
22	peptid	1.4×10^{-2}	72	spectromet	7.3×10^{-3}
23	acid	1.4×10^{-2}	73	standard	7.2×10^{-3}
24	genom	1.4×10^{-2}	74	affin	7.1×10^{-3}
25	resolut	1.4×10^{-2}	75	amino	6.8×10^{-3}
26	label	1.4×10^{-2}	76	phase	6.8×10^{-3}
27	throughput	1.4×10^{-2}	77	year	6.8×10^{-3}
28	use	1.4×10^{-2}	78	solvent	6.7×10^{-3}
29	acetonitril	1.3×10^{-2}	79	novel	6.7×10^{-3}
30	microfluid	1.3×10^{-2}	80	age	6.6×10^{-3}
31	hplc	1.3×10^{-2}	81	min	6.6×10^{-3}
32	sensit	1.3×10^{-2}	82	enzym	6.6×10^{-3}
33	sequenc	1.3×10^{-2}	83	select	6.5×10^{-3}
34	cell	1.3×10^{-2}	84	bioinformat	6.3×10^{-3}
35	develop	1.3×10^{-2}	85	accuraci	6.3×10^{-3}
36	paper	1.2×10^{-2}	86	accur	6.3×10^{-3}
37	fluoresc	1.2×10^{-2}	87	electrophoresi	6.3×10^{-3}
38	identif	1.2×10^{-2}	88	spike	6.1×10^{-3}
39	precis	1.2×10^{-2}	89	discoveri	6×10^{-3}
40	method	1.1×10^{-2}	90	quantifi	5.9×10^{-3}
41	bind	1.1×10^{-2}	91	enabl	5.8×10^{-3}
42	valid	1.1×10^{-2}	92	lod	5.8×10^{-3}
43	compound	1.1×10^{-2}	93	fragment	5.8×10^{-3}
44	esi	1.1×10^{-2}	94	reproduc	5.8×10^{-3}
45	metabolit	1.1×10^{-2}	95	simultan	5.8×10^{-3}
46	human	1.1×10^{-2}	96	enrich	5.8×10^{-3}
47	recoveri	1.1×10^{-2}	97	concentr	5.7×10^{-3}
48	purif	9.7×10^{-3}	98	linear	5.7×10^{-3}
49	purifi	9.5×10^{-3}	99	metabol	5.7×10^{-3}
50	specif	9.5×10^{-3}	100	approach	5.6×10^{-3}

Table D.22. The list of the top 100 words in the category Biochemistry and Molecular Biology with RIGs

No.	Word	RIG	No.	Word	RIG
1	protein	1.1×10^{-1}	51	involv	1.3×10^{-2}
2	cell	7×10^{-2}	52	signal	1.3×10^{-2}
3	bind	6.3×10^{-2}	53	apoptosi	1.2×10^{-2}
4	express	5.5×10^{-2}	54	molecul	1.2×10^{-2}
5	gene	5.3×10^{-2}	55	affin	1.2×10^{-2}
6	regul	5.2×10^{-2}	56	termin	1.2×10^{-2}
7	activ	5.2×10^{-2}	57	domain	1.2×10^{-2}
8	inhibit	3.8×10^{-2}	58	mitochondri	1.1×10^{-2}
9	paper	3.7×10^{-2}	59	cancer	1.1×10^{-2}
10	enzym	3.3×10^{-2}	60	intracellular	1.1×10^{-2}
11	transcript	3.2×10^{-2}	61	purifi	1.1×10^{-2}
12	induc	2.9×10^{-2}	62	reveal	1.1×10^{-2}
13	pathway	2.9×10^{-2}	63	homolog	1.1×10^{-2}
14	mediat	2.8×10^{-2}	64	wild	1.1×10^{-2}
15	role	2.8×10^{-2}	65	subunit	1.1×10^{-2}
16	molecular	2.4×10^{-2}	66	lipid	1.1×10^{-2}
17	receptor	2.4×10^{-2}	67	fold	1.1×10^{-2}
18	dna	2.2×10^{-2}	68	motif	1.1×10^{-2}
19	acid	2.2×10^{-2}	69	mous	1.1×10^{-2}
20	mutant	2.1×10^{-2}	70	conform	1.1×10^{-2}
21	inhibitor	2.1×10^{-2}	71	play	1.1×10^{-2}
22	human	2.1×10^{-2}	72	object	1×10^{-2}
23	kinas	2×10^{-2}	73	prolifer	1×10^{-2}
24	target	1.9×10^{-2}	74	biolog	1×10^{-2}
25	vitro	1.9×10^{-2}	75	knockdown	9.7×10^{-3}
26	mutat	1.8×10^{-2}	76	regulatori	9.5×10^{-3}
27	cellular	1.8×10^{-2}	77	problem	9.3×10^{-3}
28	suggest	1.8×10^{-2}	78	genet	9.3×10^{-3}
29	peptid	1.7×10^{-2}	79	specif	9.3×10^{-3}
30	assay	1.7×10^{-2}	80	encod	9.3×10^{-3}
31	sequenc	1.7×10^{-2}	81	therapeut	9.2×10^{-3}
32	phosphoryl	1.7×10^{-2}	82	eukaryot	9×10^{-3}
33	membran	1.6×10^{-2}	83	tissu	8.8×10^{-3}
34	amino	1.6×10^{-2}	84	propos	8.7×10^{-3}
35	beta	1.6×10^{-2}	85	extracellular	8.7×10^{-3}
36	rna	1.6×10^{-2}	86	identifi	8.6×10^{-3}
37	genom	1.6×10^{-2}	87	famili	8.5×10^{-3}
38	overexpress	1.5×10^{-2}	88	mammalian	8.4×10^{-3}
39	vivo	1.5×10^{-2}	89	nucleotid	8.4×10^{-3}
40	residu	1.4×10^{-2}	90	coli	8.4×10^{-3}
41	mrna	1.4×10^{-2}	91	year	8.3×10^{-3}
42	promot	1.4×10^{-2}	92	recombin	8.2×10^{-3}
43	mechan	1.4×10^{-2}	93	potent	8.2×10^{-3}
44	method	1.4×10^{-2}	94	compound	8.2×10^{-3}
45	function	1.4×10^{-2}	95	ligand	8.1×10^{-3}
46	metabol	1.4×10^{-2}	96	antioxid	7.9×10^{-3}
47	interact	1.3×10^{-2}	97	nuclear	7.9×10^{-3}
48	mice	1.3×10^{-2}	98	escherichia	7.9×10^{-3}
49	alpha	1.3×10^{-2}	99	algorithm	7.8×10^{-3}
50	site	1.3×10^{-2}	100	tumor	7.8×10^{-3}

Table D.23. The list of the top 100 words in the category Biodiversity Conservation with RIGs

No.	Word	RIG	No.	Word	RIG
1	speci	1.8×10^{-1}	51	geograph	1.4×10^{-2}
2	habitat	1.1×10^{-1}	52	south	1.4×10^{-2}
3	conserv	9.6×10^{-2}	53	distribut	1.4×10^{-2}
4	biodivers	6.2×10^{-2}	54	individu	1.4×10^{-2}
5	ecolog	5.9×10^{-2}	55	region	1.4×10^{-2}
6	popul	5.7×10^{-2}	56	cover	1.4×10^{-2}
7	ecosystem	5.5×10^{-2}	57	eastern	1.3×10^{-2}
8	forest	5.1×10^{-2}	58	park	1.3×10^{-2}
9	landscap	4.2×10^{-2}	59	coastal	1.3×10^{-2}
10	area	3.9×10^{-2}	60	genet	1.3×10^{-2}
11	manag	3.9×10^{-2}	61	island	1.3×10^{-2}
12	divers	3.8×10^{-2}	62	fauna	1.2×10^{-2}
13	abund	3.6×10^{-2}	63	breed	1.2×10^{-2}
14	microsatellit	3.6×10^{-2}	64	grassland	1.2×10^{-2}
15	endang	3.4×10^{-2}	65	allel	1.2×10^{-2}
16	wildlif	3.3×10^{-2}	66	protect	1.1×10^{-2}
17	land	2.8×10^{-2}	67	impact	1.1×10^{-2}
18	climat	2.7×10^{-2}	68	pattern	1.1×10^{-2}
19	environment	2.7×10^{-2}	69	natur	1.1×10^{-2}
20	threaten	2.7×10^{-2}	70	may	1.1×10^{-2}
21	declin	2.6×10^{-2}	71	marin	1.1×10^{-2}
22	veget	2.6×10^{-2}	72	agricultur	1.1×10^{-2}
23	bird	2.5×10^{-2}	73	import	1.1×10^{-2}
24	communiti	2.4×10^{-2}	74	paper	1.1×10^{-2}
25	site	2.4×10^{-2}	75	effort	1.1×10^{-2}
26	heterozygos	2.3×10^{-2}	76	assess	1.1×10^{-2}
27	rich	2.2×10^{-2}	77	atlant	1.1×10^{-2}
28	nativ	2.1×10^{-2}	78	invad	1.1×10^{-2}
29	fish	2×10^{-2}	79	tropic	1.1×10^{-2}
30	taxa	1.9×10^{-2}	80	chang	1.1×10^{-2}
31	across	1.9×10^{-2}	81	nest	1×10^{-2}
32	season	1.8×10^{-2}	82	prey	1×10^{-2}
33	patient	1.8×10^{-2}	83	woodland	9.9×10^{-3}
34	endem	1.8×10^{-2}	84	mammal	9.9×10^{-3}
35	north	1.8×10^{-2}	85	clinic	9.9×10^{-3}
36	extinct	1.8×10^{-2}	86	suggest	9.9×10^{-3}
37	assemblag	1.8×10^{-2}	87	america	9.8×10^{-3}
38	plant	1.7×10^{-2}	88	invertebr	9.8×10^{-3}
39	spatial	1.7×10^{-2}	89	within	9.8×10^{-3}
40	anthropogen	1.7×10^{-2}	90	disturb	9.7×10^{-3}
41	predat	1.7×10^{-2}	91	cell	9.7×10^{-3}
42	loci	1.7×10^{-2}	92	taxonom	9.7×10^{-3}
43	tree	1.6×10^{-2}	93	mountain	9.6×10^{-3}
44	northern	1.6×10^{-2}	94	locat	9.6×10^{-3}
45	river	1.6×10^{-2}	95	indic	9.5×10^{-3}
46	invas	1.5×10^{-2}	96	locus	9.4×10^{-3}
47	southern	1.5×10^{-2}	97	forag	9.3×10^{-3}
48	threat	1.5×10^{-2}	98	monitor	9.3×10^{-3}
49	survey	1.5×10^{-2}	99	freshwat	9.3×10^{-3}
50	rang	1.5×10^{-2}	100	annual	9.2×10^{-3}

TABLE D.24. The list of the top 100 words in the category Biology with RIGs

No.	Word	RIG	No.	Word	RIG
1	speci	2.3×10^{-2}	51	conserv	3.8×10^{-3}
2	paper	1.4×10^{-2}	52	pattern	3.7×10^{-3}
3	anim	1.2×10^{-2}	53	transcript	3.7×10^{-3}
4	cell	1.1×10^{-2}	54	patient	3.6×10^{-3}
5	biolog	1.1×10^{-2}	55	vertebr	3.6×10^{-3}
6	protein	1.1×10^{-2}	56	prey	3.6×10^{-3}
7	gene	9.9×10^{-3}	57	mediat	3.6×10^{-3}
8	evolutionari	9×10^{-3}	58	mice	3.5×10^{-3}
9	regul	8.2×10^{-3}	59	pathway	3.5×10^{-3}
10	reproduct	8×10^{-3}	60	induc	3.5×10^{-3}
11	physiolog	7.8×10^{-3}	61	evolut	3.4×10^{-3}
12	insect	7.6×10^{-3}	62	egg	3.4×10^{-3}
13	popul	7.6×10^{-3}	63	femal	3.4×10^{-3}
14	method	7.5×10^{-3}	64	surviv	3.3×10^{-3}
15	suggest	7.2×10^{-3}	65	neuron	3.3×10^{-3}
16	ecolog	7.1×10^{-3}	66	forag	3.3×10^{-3}
17	genet	6.5×10^{-3}	67	perform	3.3×10^{-3}
18	predat	5.9×10^{-3}	68	hypothesi	3.3×10^{-3}
19	express	5.9×10^{-3}	69	organ	3.2×10^{-3}
20	trait	5.7×10^{-3}	70	sperm	3.1×10^{-3}
21	bodi	5.7×10^{-3}	71	base	3×10^{-3}
22	bird	5.6×10^{-3}	72	fabric	3×10^{-3}
23	mate	5.5×10^{-3}	73	embryo	3×10^{-3}
24	human	5.5×10^{-3}	74	offspr	2.9×10^{-3}
25	plant	5×10^{-3}	75	mammalian	2.9×10^{-3}
26	drosophila	5×10^{-3}	76	dose	2.9×10^{-3}
27	fish	5×10^{-3}	77	wild	2.9×10^{-3}
28	activ	4.8×10^{-3}	78	improv	2.8×10^{-3}
29	expos	4.8×10^{-3}	79	dure	2.8×10^{-3}
30	habitat	4.7×10^{-3}	80	environment	2.8×10^{-3}
31	divers	4.7×10^{-3}	81	behaviour	2.8×10^{-3}
32	cryopreserv	4.6×10^{-3}	82	devic	2.8×10^{-3}
33	dna	4.6×10^{-3}	83	mous	2.8×10^{-3}
34	metabol	4.6×10^{-3}	84	receptor	2.8×10^{-3}
35	rhythm	4.5×10^{-3}	85	level	2.7×10^{-3}
36	role	4.4×10^{-3}	86	algorithm	2.7×10^{-3}
37	male	4.4×10^{-3}	87	evolv	2.7×10^{-3}
38	phenotyp	4.4×10^{-3}	88	howev	2.7×10^{-3}
39	tissu	4.3×10^{-3}	89	found	2.6×10^{-3}
40	mammal	4.3×10^{-3}	90	alter	2.6×10^{-3}
41	respons	4.3×10^{-3}	91	phylogenet	2.6×10^{-3}
42	genom	4.2×10^{-3}	92	day	2.6×10^{-3}
43	oper	4.2×10^{-3}	93	abund	2.6×10^{-3}
44	design	4.1×10^{-3}	94	conspecif	2.6×10^{-3}
45	exposur	4.1×10^{-3}	95	radiat	2.6×10^{-3}
46	may	4×10^{-3}	96	studi	2.5×10^{-3}
47	individu	4×10^{-3}	97	problem	2.5×10^{-3}
48	taxa	3.9×10^{-3}	98	mutat	2.5×10^{-3}
49	cellular	3.9×10^{-3}	99	food	2.5×10^{-3}
50	growth	3.9×10^{-3}	100	lineag	2.4×10^{-3}

TABLE D.25. The list of the top 100 words in the category Biophysics with RIGs

No.	Word	RIG	No.	Word	RIG
1	protein	5.6×10^{-2}	51	signal	6.5×10^{-3}
2	bind	4.2×10^{-2}	52	phosphoryl	6.4×10^{-3}
3	cell	3.8×10^{-2}	53	target	6.3×10^{-3}
4	membran	2×10^{-2}	54	fold	6.3×10^{-3}
5	paper	1.9×10^{-2}	55	bilay	6.2×10^{-3}
6	conform	1.6×10^{-2}	56	suggest	5.8×10^{-3}
7	angstrom	1.6×10^{-2}	57	kinas	5.8×10^{-3}
8	molecular	1.6×10^{-2}	58	play	5.7×10^{-3}
9	molecul	1.5×10^{-2}	59	strand	5.6×10^{-3}
10	induc	1.5×10^{-2}	60	coli	5.6×10^{-3}
11	activ	1.5×10^{-2}	61	dimer	5.6×10^{-3}
12	interact	1.4×10^{-2}	62	intracellular	5.5×10^{-3}
13	regul	1.4×10^{-2}	63	escherichia	5.4×10^{-3}
14	enzym	1.4×10^{-2}	64	physiolog	5.4×10^{-3}
15	residu	1.3×10^{-2}	65	extracellular	5.2×10^{-3}
16	fluoresc	1.3×10^{-2}	66	problem	5.2×10^{-3}
17	human	1.3×10^{-2}	67	complex	5.1×10^{-3}
18	inhibit	1.2×10^{-2}	68	resolut	5.1×10^{-3}
19	express	1.2×10^{-2}	69	manag	5.1×10^{-3}
20	mechan	1.1×10^{-2}	70	ligand	5.1×10^{-3}
21	role	1×10^{-2}	71	purifi	5×10^{-3}
22	lipid	1×10^{-2}	72	motif	5×10^{-3}
23	domain	9.3×10^{-3}	73	subunit	5×10^{-3}
24	structur	9.2×10^{-3}	74	atp	5×10^{-3}
25	beta	9.1×10^{-3}	75	transcript	4.9×10^{-3}
26	affin	8.7×10^{-3}	76	oper	4.8×10^{-3}
27	cellular	8.6×10^{-3}	77	year	4.8×10^{-3}
28	vitro	8.6×10^{-3}	78	cancer	4.7×10^{-3}
29	crystal	8.4×10^{-3}	79	catalyz	4.7×10^{-3}
30	acid	8.2×10^{-3}	80	apoptosi	4.7×10^{-3}
31	dna	8.2×10^{-3}	81	immobil	4.7×10^{-3}
32	assay	8.2×10^{-3}	82	dock	4.6×10^{-3}
33	mediat	8.1×10^{-3}	83	mutat	4.5×10^{-3}
34	termin	8.1×10^{-3}	84	biomechan	4.5×10^{-3}
35	vivo	8×10^{-3}	85	homolog	4.5×10^{-3}
36	helic	8×10^{-3}	86	object	4.5×10^{-3}
37	pathway	7.8×10^{-3}	87	drug	4.4×10^{-3}
38	tissu	7.8×10^{-3}	88	social	4.4×10^{-3}
39	peptid	7.8×10^{-3}	89	substrat	4.4×10^{-3}
40	biosensor	7.6×10^{-3}	90	gene	4.4×10^{-3}
41	alpha	7.4×10^{-3}	91	catalyt	4.2×10^{-3}
42	amino	7.4×10^{-3}	92	involv	4.2×10^{-3}
43	hydrophob	7.2×10^{-3}	93	depend	4.1×10^{-3}
44	inhibitor	7.1×10^{-3}	94	transmembran	4×10^{-3}
45	mutant	7.1×10^{-3}	95	label	4×10^{-3}
46	site	6.9×10^{-3}	96	sensit	3.9×10^{-3}
47	overexpress	6.8×10^{-3}	97	reveal	3.9×10^{-3}
48	biolog	6.7×10^{-3}	98	prolifer	3.8×10^{-3}
49	function	6.6×10^{-3}	99	servic	3.8×10^{-3}
50	receptor	6.6×10^{-3}	100	collagen	3.7×10^{-3}

TABLE D.26. The list of the top 100 words in the category Biotechnology and Applied Microbiology with RIGs

No.	Word	RIG	No.	Word	RIG
1	gene	5.4×10^{-2}	51	biofuel	8.2×10^{-3}
2	strain	3×10^{-2}	52	potenti	8.1×10^{-3}
3	genom	3×10^{-2}	53	concentr	8×10^{-3}
4	product	2.9×10^{-2}	54	virus	7.6×10^{-3}
5	ferment	2.9×10^{-2}	55	fold	7.6×10^{-3}
6	cell	2.9×10^{-2}	56	sugar	7.5×10^{-3}
7	protein	2.8×10^{-2}	57	16s	7.3×10^{-3}
8	sequenc	2.7×10^{-2}	58	medium	7.3×10^{-3}
9	enzym	2.6×10^{-2}	59	transcriptom	7.2×10^{-3}
10	express	2.3×10^{-2}	60	hydrolysi	7×10^{-3}
11	paper	2.2×10^{-2}	61	bacterium	6.9×10^{-3}
12	cultur	2×10^{-2}	62	specif	6.9×10^{-3}
13	acid	2×10^{-2}	63	mutant	6.8×10^{-3}
14	biomass	2×10^{-2}	64	propos	6.7×10^{-3}
15	coli	2×10^{-2}	65	human	6.6×10^{-3}
16	bacteria	1.8×10^{-2}	66	rrna	6.6×10^{-3}
17	escherichia	1.8×10^{-2}	67	biosynthesi	6.6×10^{-3}
18	microbi	1.6×10^{-2}	68	amino	6.5×10^{-3}
19	bacteri	1.6×10^{-2}	69	vitro	6.3×10^{-3}
20	produc	1.6×10^{-2}	70	inhibit	6.3×10^{-3}
21	recombin	1.5×10^{-2}	71	feedstock	6.3×10^{-3}
22	isol	1.5×10^{-2}	72	theori	6.2×10^{-3}
23	dna	1.4×10^{-2}	73	pathway	6×10^{-3}
24	yeast	1.4×10^{-2}	74	plasmid	6×10^{-3}
25	assay	1.3×10^{-2}	75	pseudomona	6×10^{-3}
26	pcr	1.2×10^{-2}	76	sludg	5.9×10^{-3}
27	purifi	1.2×10^{-2}	77	seq	5.9×10^{-3}
28	clone	1.2×10^{-2}	78	simul	5.8×10^{-3}
29	growth	1.1×10^{-2}	79	detect	5.8×10^{-3}
30	transcript	1.1×10^{-2}	80	molecular	5.8×10^{-3}
31	biolog	1.1×10^{-2}	81	kda	5.7×10^{-3}
32	bioreactor	1.1×10^{-2}	82	fungal	5.7×10^{-3}
33	activ	1×10^{-2}	83	ethanol	5.6×10^{-3}
34	cultiv	1×10^{-2}	84	stem	5.6×10^{-3}
35	plant	1×10^{-2}	85	substrat	5.5×10^{-3}
36	yield	1×10^{-2}	86	identifi	5.5×10^{-3}
37	batch	1×10^{-2}	87	inocul	5.5×10^{-3}
38	bacillus	9.7×10^{-3}	88	biofilm	5.4×10^{-3}
39	enzymat	9.4×10^{-3}	89	immobil	5.4×10^{-3}
40	microorgan	9.3×10^{-3}	90	aspergillus	5.3×10^{-3}
41	anaerob	9.3×10^{-3}	91	extracellular	5.3×10^{-3}
42	lignocellulos	9.2×10^{-3}	92	contain	5.2×10^{-3}
43	encod	9.1×10^{-3}	93	phenotyp	5.1×10^{-3}
44	pathogen	9.1×10^{-3}	94	speci	5×10^{-3}
45	metabol	8.9×10^{-3}	95	lactobacillus	5×10^{-3}
46	cerevisia	8.8×10^{-3}	96	cellulos	5×10^{-3}
47	saccharomyc	8.6×10^{-3}	97	putat	5×10^{-3}
48	glucos	8.5×10^{-3}	98	bind	4.9×10^{-3}
49	genet	8.5×10^{-3}	99	wild	4.8×10^{-3}
50	rna	8.5×10^{-3}	100	purif	4.8×10^{-3}

TABLE D.27. The list of the top 100 words in the category Business with RIGs

No.	Word	RIG	No.	Word	RIG
1	firm	8.4×10^{-2}	51	surfac	1.2×10^{-2}
2	market	8×10^{-2}	52	advertis	1.2×10^{-2}
3	busi	7.8×10^{-2}	53	focus	1.2×10^{-2}
4	research	6.2×10^{-2}	54	treatment	1.2×10^{-2}
5	compani	5.9×10^{-2}	55	influenc	1.2×10^{-2}
6	manag	4.6×10^{-2}	56	develop	1.2×10^{-2}
7	innov	3.6×10^{-2}	57	govern	1.2×10^{-2}
8	organiz	3.6×10^{-2}	58	sale	1.1×10^{-2}
9	financi	3.5×10^{-2}	59	draw	1.1×10^{-2}
10	corpor	3.5×10^{-2}	60	perceiv	1.1×10^{-2}
11	empir	3.1×10^{-2}	61	retail	1.1×10^{-2}
12	consum	2.9×10^{-2}	62	profit	1.1×10^{-2}
13	econom	2.9×10^{-2}	63	price	1.1×10^{-2}
14	social	2.8×10^{-2}	64	method	1.1×10^{-2}
15	employe	2.7×10^{-2}	65	impact	1.1×10^{-2}
16	strateg	2.6×10^{-2}	66	protein	1.1×10^{-2}
17	brand	2.6×10^{-2}	67	polici	1.1×10^{-2}
18	enterpris	2.6×10^{-2}	68	clinic	1.1×10^{-2}
19	custom	2.6×10^{-2}	69	foreign	1×10^{-2}
20	manageri	2.5×10^{-2}	70	intern	1×10^{-2}
21	find	2.4×10^{-2}	71	methodolog	1×10^{-2}
22	economi	2.3×10^{-2}	72	institut	9.9×10^{-3}
23	industri	2.3×10^{-2}	73	detect	9.9×10^{-3}
24	competit	2.3×10^{-2}	74	conclus	9.8×10^{-3}
25	implic	2.3×10^{-2}	75	anteced	9.8×10^{-3}
26	capit	2.2×10^{-2}	76	intent	9.7×10^{-3}
27	literatur	2.1×10^{-2}	77	survey	9.7×10^{-3}
28	countri	2.1×10^{-2}	78	theoret	9.5×10^{-3}
29	invest	2×10^{-2}	79	diseas	9.5×10^{-3}
30	entrepreneuri	2×10^{-2}	80	financ	9.2×10^{-3}
31	paper	1.9×10^{-2}	81	strategi	9.2×10^{-3}
32	cell	1.9×10^{-2}	82	stakehold	9.1×10^{-3}
33	relationship	1.8×10^{-2}	83	product	9×10^{-3}
34	practic	1.8×10^{-2}	84	paramet	9×10^{-3}
35	sector	1.7×10^{-2}	85	technolog	8.9×10^{-3}
36	entrepreneurship	1.7×10^{-2}	86	observ	8.9×10^{-3}
37	perspect	1.7×10^{-2}	87	acid	8.8×10^{-3}
38	patient	1.7×10^{-2}	88	opportun	8.8×10^{-3}
39	entrepreneur	1.5×10^{-2}	89	investor	8.8×10^{-3}
40	theori	1.5×10^{-2}	90	public	8.6×10^{-3}
41	knowledg	1.5×10^{-2}	91	percept	8.5×10^{-3}
42	servic	1.5×10^{-2}	92	make	8.5×10^{-3}
43	context	1.4×10^{-2}	93	asset	8.4×10^{-3}
44	articl	1.4×10^{-2}	94	crisi	8.4×10^{-3}
45	purchas	1.4×10^{-2}	95	properti	8.3×10^{-3}
46	decis	1.3×10^{-2}	96	import	8.2×10^{-3}
47	author	1.3×10^{-2}	97	gene	8.2×10^{-3}
48	resourc	1.3×10^{-2}	98	issu	8.1×10^{-3}
49	temperatur	1.3×10^{-2}	99	creation	8×10^{-3}
50	conceptu	1.3×10^{-2}	100	examin	8×10^{-3}

Table D.28. The list of the top 100 words in the category Business, Finance with RIGs

No.	Word	RIG
1	market	1.1×10^{-1}
2	financi	1.1×10^{-1}
3	firm	7×10^{-2}
4	price	6.1×10^{-2}
5	investor	5.9×10^{-2}
6	stock	5.4×10^{-2}
7	bank	4.6×10^{-2}
8	asset	4.5×10^{-2}
9	crisi	4.4×10^{-2}
10	return	3.8×10^{-2}
11	capit	3.7×10^{-2}
12	compani	3.5×10^{-2}
13	econom	3.5×10^{-2}
14	invest	3.5×10^{-2}
15	trade	3.3×10^{-2}
16	equiti	3.3×10^{-2}
17	economi	3.2×10^{-2}
18	corpor	3.2×10^{-2}
19	credit	3×10^{-2}
20	financ	3×10^{-2}
21	countri	3×10^{-2}
22	portfolio	3×10^{-2}
23	find	2.9×10^{-2}
24	empir	2.8×10^{-2}
25	debt	2.7×10^{-2}
26	paper	2.5×10^{-2}
27	busi	2.4×10^{-2}
28	volatil	2.4×10^{-2}
29	polici	2.1×10^{-2}
30	earn	2×10^{-2}
31	risk	2×10^{-2}
32	loan	1.9×10^{-2}
33	cell	1.8×10^{-2}
34	fund	1.8×10^{-2}
35	manag	1.8×10^{-2}
36	govern	1.7×10^{-2}
37	patient	1.7×10^{-2}
38	enterpris	1.6×10^{-2}
39	macroeconom	1.6×10^{-2}
40	sector	1.6×10^{-2}
41	premium	1.5×10^{-2}
42	tax	1.5×10^{-2}
43	profit	1.5×10^{-2}
44	monetari	1.4×10^{-2}
45	temperatur	1.3×10^{-2}
46	foreign	1.3×10^{-2}
47	account	1.2×10^{-2}
48	surfac	1.2×10^{-2}
49	privat	1.2×10^{-2}
50	method	1.1×10^{-2}
51	european	1.1×10^{-2}
52	treatment	1.1×10^{-2}
53	clinic	1.1×10^{-2}
54	institut	1.1×10^{-2}
55	incent	1.1×10^{-2}
56	protein	1×10^{-2}
57	conclus	9.9×10^{-3}
58	experiment	9.9×10^{-3}
59	czech	9.9×10^{-3}
60	industri	9.9×10^{-3}
61	impact	9.8×10^{-3}
62	evid	9.5×10^{-3}
63	audit	9.5×10^{-3}
64	public	9.5×10^{-3}
65	valuat	9.1×10^{-3}
66	disclosur	9.1×10^{-3}
67	romania	9×10^{-3}
68	transact	9×10^{-3}
69	diseas	8.9×10^{-3}
70	innov	8.9×10^{-3}
71	romanian	8.9×10^{-3}
72	ownership	8.8×10^{-3}
73	forecast	8.6×10^{-3}
74	research	8.4×10^{-3}
75	acid	8.4×10^{-3}
76	insur	8.3×10^{-3}
77	borrow	8.2×10^{-3}
78	period	8×10^{-3}
79	competit	8×10^{-3}
80	gene	7.7×10^{-3}
81	crise	7.7×10^{-3}
82	intern	7.4×10^{-3}
83	detect	7.4×10^{-3}
84	list	7.4×10^{-3}
85	republ	7.4×10^{-3}
86	speci	7.3×10^{-3}
87	rang	7.3×10^{-3}
88	electron	7.3×10^{-3}
89	sell	7.3×10^{-3}
90	incom	7.1×10^{-3}
91	oxid	7×10^{-3}
92	gdp	6.9×10^{-3}
93	decis	6.8×10^{-3}
94	exchang	6.7×10^{-3}
95	water	6.7×10^{-3}
96	molecular	6.7×10^{-3}
97	background	6.5×10^{-3}
98	materi	6.5×10^{-3}
99	energi	6.5×10^{-3}
100	cost	6.4×10^{-3}

TABLE D.29. The list of the top 100 words in the category Cardiac and Cardiovascular Systems with RIGs

No.	Word	RIG	No.	Word	RIG
1	patient	1.8×10^{-1}	51	hospit	2.8×10^{-2}
2	cardiac	1.6×10^{-1}	52	median	2.7×10^{-2}
3	conclus	1.5×10^{-1}	53	cardiomyopathi	2.7×10^{-2}
4	coronari	1.4×10^{-1}	54	revascular	2.7×10^{-2}
5	heart	1.4×10^{-1}	55	multivari	2.7×10^{-2}
6	ventricular	1.4×10^{-1}	56	confid	2.7×10^{-2}
7	myocardi	1.1×10^{-1}	57	hypertens	2.7×10^{-2}
8	arteri	1.1×10^{-1}	58	month	2.6×10^{-2}
9	left	9.3×10^{-2}	59	dysfunct	2.5×10^{-2}
10	background	8.4×10^{-2}	60	signific	2.5×10^{-2}
11	atrial	7.7×10^{-2}	61	baselin	2.5×10^{-2}
12	infarct	7×10^{-2}	62	angiographi	2.4×10^{-2}
13	aortic	6.9×10^{-2}	63	intervent	2.4×10^{-2}
14	cardiovascular	6.8×10^{-2}	64	result	2.3×10^{-2}
15	mortal	6.2×10^{-2}	65	hazard	2.3×10^{-2}
16	risk	6×10^{-2}	66	stroke	2.3×10^{-2}
17	echocardiographi	5.8×10^{-2}	67	complic	2.2×10^{-2}
18	underw	5.7×10^{-2}	68	vascular	2.2×10^{-2}
19	year	5.5×10^{-2}	69	right	2.2×10^{-2}
20	valv	5.4×10^{-2}	70	event	2.2×10^{-2}
21	method	5.3×10^{-2}	71	therapi	2.2×10^{-2}
22	fibril	4.8×10^{-2}	72	object	2.2×10^{-2}
23	eject	4.7×10^{-2}	73	surgic	2.1×10^{-2}
24	clinic	4.5×10^{-2}	74	prospect	2.1×10^{-2}
25	percutan	4.5×10^{-2}	75	aim	2.1×10^{-2}
26	diseas	4.5×10^{-2}	76	ventricl	2×10^{-2}
27	systol	4.5×10^{-2}	77	assess	2×10^{-2}
28	associ	4.1×10^{-2}	78	group	2×10^{-2}
29	failur	4.1×10^{-2}	79	postop	1.9×10^{-2}
30	pulmonari	4×10^{-2}	80	cohort	1.9×10^{-2}
31	age	4×10^{-2}	81	adjust	1.9×10^{-2}
32	death	4×10^{-2}	82	independ	1.9×10^{-2}
33	outcom	4×10^{-2}	83	cardiopulmonari	1.9×10^{-2}
34	echocardiograph	3.4×10^{-2}	84	ischem	1.9×10^{-2}
35	mitral	3.3×10^{-2}	85	enrol	1.9×10^{-2}
36	follow	3.3×10^{-2}	86	mean	1.8×10^{-2}
37	stenosi	3.3×10^{-2}	87	ecg	1.8×10^{-2}
38	implant	3.3×10^{-2}	88	procedur	1.8×10^{-2}
39	stent	3.2×10^{-2}	89	inhospit	1.8×10^{-2}
40	paper	3.2×10^{-2}	90	atherosclerosi	1.7×10^{-2}
41	diastol	3.2×10^{-2}	91	aorta	1.7×10^{-2}
42	undergo	3.1×10^{-2}	92	blood	1.7×10^{-2}
43	acut	3.1×10^{-2}	93	advers	1.7×10^{-2}
44	regurgit	3×10^{-2}	94	cathet	1.7×10^{-2}
45	consecut	3×10^{-2}	95	day	1.7×10^{-2}
46	bypass	3×10^{-2}	96	ratio	1.7×10^{-2}
47	surgeri	3×10^{-2}	97	syndrom	1.7×10^{-2}
48	interv	2.9×10^{-2}	98	incid	1.6×10^{-2}
49	predictor	2.9×10^{-2}	99	retrospect	1.6×10^{-2}
50	arrhythmia	2.8×10^{-2}	100	propos	1.6×10^{-2}

Table D.30. The list of the top 100 words in the category Cell and Tissue Engineering with RIGs

No.	Word	RIG	No.	Word	RIG
1	stem	2.4×10^{-1}	51	protein	1.8×10^{-2}
2	cell	2×10^{-1}	52	chondrocyt	1.8×10^{-2}
3	mesenchym	1×10^{-1}	53	endotheli	1.8×10^{-2}
4	pluripot	9.3×10^{-2}	54	paper	1.7×10^{-2}
5	differenti	8×10^{-2}	55	adipos	1.7×10^{-2}
6	tissu	7.5×10^{-2}	56	cellular	1.7×10^{-2}
7	progenitor	7.2×10^{-2}	57	promot	1.7×10^{-2}
8	bone	6.9×10^{-2}	58	phenotyp	1.6×10^{-2}
9	embryon	6.6×10^{-2}	59	upregul	1.6×10^{-2}
10	express	6.3×10^{-2}	60	fate	1.5×10^{-2}
11	mscs	6.2×10^{-2}	61	graft	1.5×10^{-2}
12	vitro	6.1×10^{-2}	62	engraft	1.5×10^{-2}
13	cultur	5.9×10^{-2}	63	vascular	1.5×10^{-2}
14	human	5.3×10^{-2}	64	osteoblast	1.4×10^{-2}
15	transplant	5.1×10^{-2}	65	histolog	1.4×10^{-2}
16	regener	5.1×10^{-2}	66	neural	1.4×10^{-2}
17	marrow	5×10^{-2}	67	matrix	1.3×10^{-2}
18	regen	5×10^{-2}	68	neuron	1.3×10^{-2}
19	vivo	4.9×10^{-2}	69	demonstr	1.3×10^{-2}
20	prolifer	4.8×10^{-2}	70	somat	1.3×10^{-2}
21	stromal	4.8×10^{-2}	71	cd34	1.2×10^{-2}
22	scaffold	4.7×10^{-2}	72	medicin	1.2×10^{-2}
23	msc	4.1×10^{-2}	73	week	1.2×10^{-2}
24	lineag	3.8×10^{-2}	74	rat	1.2×10^{-2}
25	selfrenew	3.7×10^{-2}	75	defect	1.2×10^{-2}
26	induc	3.7×10^{-2}	76	cord	1.2×10^{-2}
27	osteogen	3.5×10^{-2}	77	matur	1.2×10^{-2}
28	deriv	3.5×10^{-2}	78	microenviron	1.2×10^{-2}
29	fibroblast	3.3×10^{-2}	79	seed	1.2×10^{-2}
30	collagen	3.3×10^{-2}	80	clinic	1.1×10^{-2}
31	reprogram	3.2×10^{-2}	81	format	1.1×10^{-2}
32	gene	2.8×10^{-2}	82	function	1.1×10^{-2}
33	marker	2.8×10^{-2}	83	biomateri	1.1×10^{-2}
34	mous	2.8×10^{-2}	84	hydrogel	1.1×10^{-2}
35	therapi	2.8×10^{-2}	85	heal	1.1×10^{-2}
36	hematopoiet	2.7×10^{-2}	86	stimul	1.1×10^{-2}
37	engin	2.4×10^{-2}	87	signal	1.1×10^{-2}
38	autolog	2.4×10^{-2}	88	injuri	1.1×10^{-2}
39	repair	2.3×10^{-2}	89	donor	1×10^{-2}
40	cartilag	2.2×10^{-2}	90	secret	1×10^{-2}
41	implant	2.2×10^{-2}	91	isol	1×10^{-2}
42	potenti	2×10^{-2}	92	prolif	1×10^{-2}
43	mice	2×10^{-2}	93	allogen	9.9×10^{-3}
44	therapeut	2×10^{-2}	94	pathway	9.7×10^{-3}
45	growth	1.9×10^{-2}	95	viabil	9.6×10^{-3}
46	transcript	1.9×10^{-2}	96	abil	9.6×10^{-3}
47	cocultur	1.9×10^{-2}	97	mediat	9.5×10^{-3}
48	extracellular	1.8×10^{-2}	98	wnt	9.4×10^{-3}
49	factor	1.8×10^{-2}	99	induct	9.4×10^{-3}
50	regul	1.8×10^{-2}	100	activ	9.4×10^{-3}

Table D.31. The list of the top 100 words in the category Cell Biology with RIGs

No.	Word	RIG	No.	Word	RIG
1	cell	1.7×10^{-1}	51	beta	1.7×10^{-2}
2	express	1.1×10^{-1}	52	factor	1.6×10^{-2}
3	protein	1×10^{-1}	53	extracellular	1.6×10^{-2}
4	regul	9.5×10^{-2}	54	downregul	1.6×10^{-2}
5	induc	6.2×10^{-2}	55	phenotyp	1.6×10^{-2}
6	inhibit	5.6×10^{-2}	56	mutat	1.5×10^{-2}
7	pathway	5.3×10^{-2}	57	embryon	1.5×10^{-2}
8	activ	5.2×10^{-2}	58	involv	1.5×10^{-2}
9	mediat	5.2×10^{-2}	59	rna	1.5×10^{-2}
10	gene	5.1×10^{-2}	60	dna	1.4×10^{-2}
11	transcript	4.6×10^{-2}	61	respons	1.4×10^{-2}
12	role	4.4×10^{-2}	62	intracellular	1.4×10^{-2}
13	kinas	4.1×10^{-2}	63	base	1.4×10^{-2}
14	receptor	3.8×10^{-2}	64	alpha	1.4×10^{-2}
15	prolifer	3.7×10^{-2}	65	chromatin	1.4×10^{-2}
16	signal	3.7×10^{-2}	66	mitochondri	1.4×10^{-2}
17	paper	3.6×10^{-2}	67	fibroblast	1.4×10^{-2}
18	human	3.5×10^{-2}	68	akt	1.3×10^{-2}
19	phosphoryl	3.4×10^{-2}	69	nuclear	1.3×10^{-2}
20	promot	3.4×10^{-2}	70	simul	1.3×10^{-2}
21	mice	3.3×10^{-2}	71	depend	1.3×10^{-2}
22	bind	3.2×10^{-2}	72	play	1.3×10^{-2}
23	tissu	3.2×10^{-2}	73	silenc	1.3×10^{-2}
24	apoptosi	3.1×10^{-2}	74	regulatori	1.3×10^{-2}
25	cellular	3×10^{-2}	75	migrat	1.2×10^{-2}
26	overexpress	3×10^{-2}	76	cytokin	1.2×10^{-2}
27	vivo	3×10^{-2}	77	marker	1.2×10^{-2}
28	vitro	2.9×10^{-2}	78	actin	1.2×10^{-2}
29	tumor	2.9×10^{-2}	79	associ	1.2×10^{-2}
30	cancer	2.9×10^{-2}	80	ubiquitin	1.2×10^{-2}
31	stem	2.9×10^{-2}	81	cultur	1.2×10^{-2}
32	target	2.7×10^{-2}	82	repress	1.2×10^{-2}
33	mous	2.7×10^{-2}	83	progenitor	1.2×10^{-2}
34	inhibitor	2.6×10^{-2}	84	neuron	1.2×10^{-2}
35	knockdown	2.4×10^{-2}	85	suppressor	1.2×10^{-2}
36	mechan	2.4×10^{-2}	86	assay	1.1×10^{-2}
37	function	2.2×10^{-2}	87	alter	1.1×10^{-2}
38	mesenchym	2.1×10^{-2}	88	molecular	1.1×10^{-2}
39	suggest	2×10^{-2}	89	secret	1.1×10^{-2}
40	differenti	2×10^{-2}	90	cytoplasm	1.1×10^{-2}
41	membran	2×10^{-2}	91	inflammatori	1.1×10^{-2}
42	upregul	1.9×10^{-2}	92	rat	1.1×10^{-2}
43	method	1.9×10^{-2}	93	level	1.1×10^{-2}
44	therapeut	1.9×10^{-2}	94	homeostasi	1.1×10^{-2}
45	stimul	1.9×10^{-2}	95	diseas	1.1×10^{-2}
46	mrna	1.9×10^{-2}	96	perform	1.1×10^{-2}
47	suppress	1.8×10^{-2}	97	endotheli	1.1×10^{-2}
48	mutant	1.8×10^{-2}	98	demonstr	1.1×10^{-2}
49	growth	1.8×10^{-2}	99	temperatur	1.1×10^{-2}
50	epitheli	1.8×10^{-2}	100	oncogen	1×10^{-2}

Table D.32. The list of the top 100 words in the category Chemistry, Analytical with RIGs

No.	Word	RIG	No.	Word	RIG
1	detect	8.4×10^{-2}	51	use	1.3×10^{-2}
2	chromatographi	8×10^{-2}	52	microfluid	1.3×10^{-2}
3	spectrometri	6.9×10^{-2}	53	solvent	1.3×10^{-2}
4	sampl	5.3×10^{-2}	54	methanol	1.3×10^{-2}
5	analyt	4.9×10^{-2}	55	esi	1.3×10^{-2}
6	liquid	4.6×10^{-2}	56	gold	1.2×10^{-2}
7	sensit	4×10^{-2}	57	monitor	1.2×10^{-2}
8	limit	3.9×10^{-2}	58	conclus	1.2×10^{-2}
9	chromatograph	3.7×10^{-2}	59	success	1.2×10^{-2}
10	concentr	3.7×10^{-2}	60	develop	1.2×10^{-2}
11	mass	3.6×10^{-2}	61	rapid	1.2×10^{-2}
12	quantif	3.6×10^{-2}	62	quadrupol	1.2×10^{-2}
13	column	3.5×10^{-2}	63	capillari	1.2×10^{-2}
14	linear	3.2×10^{-2}	64	buffer	1.2×10^{-2}
15	recoveri	3.1×10^{-2}	65	reproduc	1.2×10^{-2}
16	rang	3×10^{-2}	66	nanoparticl	1.1×10^{-2}
17	separ	2.8×10^{-2}	67	urin	1.1×10^{-2}
18	extract	2.8×10^{-2}	68	deviat	1.1×10^{-2}
19	acid	2.7×10^{-2}	69	aqueous	1.1×10^{-2}
20	ion	2.6×10^{-2}	70	quantit	1.1×10^{-2}
21	elut	2.5×10^{-2}	71	water	1.1×10^{-2}
22	lod	2.5×10^{-2}	72	metabolit	1.1×10^{-2}
23	electrospray	2.5×10^{-2}	73	probe	1.1×10^{-2}
24	electrod	2.4×10^{-2}	74	simultan	1×10^{-2}
25	ioniz	2.3×10^{-2}	75	patient	1×10^{-2}
26	hplc	2.3×10^{-2}	76	spectromet	1×10^{-2}
27	determin	2.2×10^{-2}	77	reaction	1×10^{-2}
28	acetonitril	2.2×10^{-2}	78	chemic	9.8×10^{-3}
29	tandem	2.2×10^{-2}	79	mixtur	9.7×10^{-3}
30	sensor	2.2×10^{-2}	80	valid	9.6×10^{-3}
31	rsd	2.1×10^{-2}	81	glassi	9.1×10^{-3}
32	compound	2×10^{-2}	82	good	9×10^{-3}
33	electrochem	1.9×10^{-2}	83	high	8.9×10^{-3}
34	fluoresc	1.9×10^{-2}	84	solid	8.8×10^{-3}
35	biosensor	1.8×10^{-2}	85	associ	8.7×10^{-3}
36	appli	1.8×10^{-2}	86	solut	8.7×10^{-3}
37	immobil	1.8×10^{-2}	87	gas	8.5×10^{-3}
38	method	1.8×10^{-2}	88	year	8.5×10^{-3}
39	calibr	1.7×10^{-2}	89	spectroscopi	8.5×10^{-3}
40	prepar	1.7×10^{-2}	90	reagent	8.5×10^{-3}
41	precis	1.6×10^{-2}	91	coupl	8.4×10^{-3}
42	min	1.6×10^{-2}	92	obtain	8.4×10^{-3}
43	select	1.6×10^{-2}	93	analysi	8.3×10^{-3}
44	phase	1.5×10^{-2}	94	modifi	8.3×10^{-3}
45	spike	1.5×10^{-2}	95	carbon	8.2×10^{-3}
46	assay	1.4×10^{-2}	96	desorpt	7.9×10^{-3}
47	standard	1.4×10^{-2}	97	silica	7.9×10^{-3}
48	label	1.4×10^{-2}	98	peak	7.7×10^{-3}
49	simpl	1.4×10^{-2}	99	sens	7.7×10^{-3}
50	voltammetri	1.3×10^{-2}	100	age	7.5×10^{-3}

Table D.33. The list of the top 100 words in the category Chemistry, Applied with RIGs

No.	Word	RIG	No.	Word	RIG
1	acid	3.9×10^{-2}	51	contain	7.5×10^{-3}
2	prepar	2.3×10^{-2}	52	algorithm	7.4×10^{-3}
3	compound	2.3×10^{-2}	53	chitosan	7.3×10^{-3}
4	patient	1.9×10^{-2}	54	problem	7.3×10^{-3}
5	catalyst	1.8×10^{-2}	55	zeolit	7.1×10^{-3}
6	content	1.8×10^{-2}	56	liquid	7×10^{-3}
7	oil	1.8×10^{-2}	57	fatti	6.9×10^{-3}
8	chromatographi	1.8×10^{-2}	58	gas	6.7×10^{-3}
9	antioxid	1.8×10^{-2}	59	synthesi	6.7×10^{-3}
10	reaction	1.7×10^{-2}	60	radic	6.6×10^{-3}
11	phenol	1.6×10^{-2}	61	outcom	6.5×10^{-3}
12	nmr	1.5×10^{-2}	62	polyphenol	6.5×10^{-3}
13	food	1.5×10^{-2}	63	ethanol	6.4×10^{-3}
14	polysaccharid	1.5×10^{-2}	64	amin	6.4×10^{-3}
15	co2	1.4×10^{-2}	65	emuls	6.4×10^{-3}
16	extract	1.4×10^{-2}	66	ester	6.3×10^{-3}
17	concentr	1.4×10^{-2}	67	amount	6.3×10^{-3}
18	paper	1.3×10^{-2}	68	fruit	6.3×10^{-3}
19	starch	1.2×10^{-2}	69	comput	6.2×10^{-3}
20	product	1.2×10^{-2}	70	age	5.9×10^{-3}
21	degre	1.1×10^{-2}	71	acet	5.8×10^{-3}
22	aqueous	1.1×10^{-2}	72	carbon	5.8×10^{-3}
23	activ	1.1×10^{-2}	73	methyl	5.7×10^{-3}
24	chemic	1.1×10^{-2}	74	sem	5.7×10^{-3}
25	adsorpt	1.1×10^{-2}	75	dri	5.6×10^{-3}
26	solvent	1.1×10^{-2}	76	flavonoid	5.6×10^{-3}
27	properti	1.1×10^{-2}	77	object	5.6×10^{-3}
28	water	1×10^{-2}	78	infrar	5.5×10^{-3}
29	solubl	9.9×10^{-3}	79	beta	5.5×10^{-3}
30	spectroscopi	9.8×10^{-3}	80	sugar	5.4×10^{-3}
31	catalyt	9.5×10^{-3}	81	anthocyanin	5.4×10^{-3}
32	synthes	9.2×10^{-3}	82	find	5.3×10^{-3}
33	dpph	8.8×10^{-3}	83	sodium	5.3×10^{-3}
34	cellulos	8.8×10^{-3}	84	particip	5.2×10^{-3}
35	hplc	8.7×10^{-3}	85	mesopor	5.2×10^{-3}
36	oxid	8.7×10^{-3}	86	data	5.2×10^{-3}
37	clinic	8.5×10^{-3}	87	xrd	5.1×10^{-3}
38	temperatur	8.5×10^{-3}	88	bioactiv	5.1×10^{-3}
39	composit	8.4×10^{-3}	89	mixtur	5.1×10^{-3}
40	yield	8.4×10^{-3}	90	isol	5.1×10^{-3}
41	ftir	8.3×10^{-3}	91	character	5×10^{-3}
42	conclus	8.2×10^{-3}	92	physicochem	5×10^{-3}
43	storag	8.2×10^{-3}	93	stabil	5×10^{-3}
44	gel	8.2×10^{-3}	94	microscopi	4.9×10^{-3}
45	associ	8.1×10^{-3}	95	model	4.9×10^{-3}
46	spectrometri	8×10^{-3}	96	set	4.8×10^{-3}
47	scaveng	8×10^{-3}	97	viscos	4.8×10^{-3}
48	year	7.8×10^{-3}	98	surfact	4.8×10^{-3}
49	hydrolysi	7.8×10^{-3}	99	pore	4.7×10^{-3}
50	propos	7.7×10^{-3}	100	glycosid	4.7×10^{-3}

Table D.34. The list of the top 100 words in the category Chemistry, Inorganic and Nuclear with RIGs

No.	Word	RIG	No.	Word	RIG
1	ligand	1.5×10^{-1}	51	supramolecular	1.5×10^{-2}
2	ray	1.1×10^{-1}	52	antiferromagnet	1.5×10^{-2}
3	crystal	9.7×10^{-2}	53	moieti	1.5×10^{-2}
4	complex	9.5×10^{-2}	54	substitut	1.4×10^{-2}
5	coordin	8.4×10^{-2}	55	distort	1.4×10^{-2}
6	synthes	8.2×10^{-2}	56	salt	1.4×10^{-2}
7	diffract	6.8×10^{-2}	57	react	1.4×10^{-2}
8	reaction	6.6×10^{-2}	58	result	1.4×10^{-2}
9	compound	6.1×10^{-2}	59	substitu	1.4×10^{-2}
10	dot	5.8×10^{-2}	60	element	1.4×10^{-2}
11	bis	5.6×10^{-2}	61	mononuclear	1.4×10^{-2}
12	h2o	5.3×10^{-2}	62	model	1.4×10^{-2}
13	structur	5.3×10^{-2}	63	develop	1.3×10^{-2}
14	center	5.1×10^{-2}	64	dimer	1.3×10^{-2}
15	nmr	5.1×10^{-2}	65	object	1.3×10^{-2}
16	bond	5×10^{-2}	66	ring	1.3×10^{-2}
17	character	4.9×10^{-2}	67	yield	1.3×10^{-2}
18	metal	4.6×10^{-2}	68	catalyst	1.3×10^{-2}
19	spectroscopi	3.6×10^{-2}	69	phenyl	1.3×10^{-2}
20	ion	3.4×10^{-2}	70	copper	1.3×10^{-2}
21	atom	3.3×10^{-2}	71	solvent	1.3×10^{-2}
22	anion	3.2×10^{-2}	72	year	1.3×10^{-2}
23	bridg	2.8×10^{-2}	73	carboxyl	1.2×10^{-2}
24	pyridin	2.7×10^{-2}	74	oxid	1.2×10^{-2}
25	dft	2.6×10^{-2}	75	powder	1.2×10^{-2}
26	cation	2.4×10^{-2}	76	methyl	1.2×10^{-2}
27	crystallographi	2.4×10^{-2}	77	associ	1.2×10^{-2}
28	prepar	2.3×10^{-2}	78	molecul	1.2×10^{-2}
29	paper	2.3×10^{-2}	79	steric	1.1×10^{-2}
30	octahedr	2.2×10^{-2}	80	geometri	1.1×10^{-2}
31	hydrogen	2.1×10^{-2}	81	format	1.1×10^{-2}
32	singl	2.1×10^{-2}	82	problem	1.1×10^{-2}
33	conclus	2×10^{-2}	83	crystallograph	1.1×10^{-2}
34	chelat	2×10^{-2}	84	no3	1.1×10^{-2}
35	angstrom	2×10^{-2}	85	donor	1.1×10^{-2}
36	vis	2×10^{-2}	86	use	1.1×10^{-2}
37	afford	1.9×10^{-2}	87	form	1.1×10^{-2}
38	iii	1.9×10^{-2}	88	signific	1.1×10^{-2}
39	patient	1.8×10^{-2}	89	exhibit	1.1×10^{-2}
40	solid	1.8×10^{-2}	90	increas	1.1×10^{-2}
41	spectroscop	1.8×10^{-2}	91	express	1×10^{-2}
42	luminesc	1.7×10^{-2}	92	assess	1×10^{-2}
43	monoclin	1.7×10^{-2}	93	heterocycl	1×10^{-2}
44	acid	1.6×10^{-2}	94	may	1×10^{-2}
45	eta	1.6×10^{-2}	95	background	1×10^{-2}
46	electron	1.6×10^{-2}	96	improv	1×10^{-2}
47	hydrotherm	1.6×10^{-2}	97	chlorid	1×10^{-2}
48	spectra	1.6×10^{-2}	98	aim	9.9×10^{-3}
49	synthesi	1.6×10^{-2}	99	propos	9.9×10^{-3}
50	catalyt	1.5×10^{-2}	100	find	9.9×10^{-3}

TABLE D.35. The list of the top 100 words in the category Chemistry, Medicinal with RIGs

No.	Word	RIG	No.	Word	RIG
1	compound	1.8×10^{-1}	51	bioactiv	1.2×10^{-2}
2	activ	9.1×10^{-2}	52	alkaloid	1.2×10^{-2}
3	potent	8×10^{-2}	53	receptor	1.2×10^{-2}
4	ic50	7.5×10^{-2}	54	scaffold	1.2×10^{-2}
5	inhibit	5.3×10^{-2}	55	discoveri	1.2×10^{-2}
6	inhibitor	5.2×10^{-2}	56	affin	1.2×10^{-2}
7	drug	5×10^{-2}	57	antioxid	1.1×10^{-2}
8	cytotox	4.7×10^{-2}	58	synthesi	1.1×10^{-2}
9	inhibitori	4.6×10^{-2}	59	target	1.1×10^{-2}
10	vitro	4×10^{-2}	60	pharmacokinet	1.1×10^{-2}
11	synthes	3.8×10^{-2}	61	propos	1.1×10^{-2}
12	ethnopharmacolog	3.7×10^{-2}	62	pharmacolog	1.1×10^{-2}
13	cell	3.4×10^{-2}	63	oral	1.1×10^{-2}
14	dock	3.2×10^{-2}	64	substitu	1.1×10^{-2}
15	nmr	3.1×10^{-2}	65	therapeut	1.1×10^{-2}
16	potenc	3×10^{-2}	66	methyl	1.1×10^{-2}
17	isol	3×10^{-2}	67	biolog	1.1×10^{-2}
18	assay	2.9×10^{-2}	68	induc	1.1×10^{-2}
19	paper	2.7×10^{-2}	69	evalu	1×10^{-2}
20	bind	2.6×10^{-2}	70	molecular	1×10^{-2}
21	seri	2.6×10^{-2}	71	sar	1×10^{-2}
22	deriv	2.5×10^{-2}	72	antitumor	1×10^{-2}
23	spectroscop	2.5×10^{-2}	73	alpha	1×10^{-2}
24	medicin	2.4×10^{-2}	74	chemic	1×10^{-2}
25	acid	2.1×10^{-2}	75	protein	9.8×10^{-3}
26	structur	2×10^{-2}	76	dose	9.7×10^{-3}
27	anticanc	2×10^{-2}	77	known	9.6×10^{-3}
28	analogu	2×10^{-2}	78	toxic	9.3×10^{-3}
29	cancer	1.9×10^{-2}	79	metabolit	9.2×10^{-3}
30	moieti	1.8×10^{-2}	80	phenyl	9.1×10^{-3}
31	antiprolif	1.8×10^{-2}	81	glycosid	9.1×10^{-3}
32	agent	1.8×10^{-2}	82	select	9.1×10^{-3}
33	exhibit	1.7×10^{-2}	83	ring	9×10^{-3}
34	human	1.6×10^{-2}	84	mic	8.9×10^{-3}
35	screen	1.5×10^{-2}	85	promis	8.9×10^{-3}
36	elucid	1.5×10^{-2}	86	hplc	8.7×10^{-3}
37	extract	1.5×10^{-2}	87	oper	8.6×10^{-3}
38	vivo	1.4×10^{-2}	88	phytochem	8.5×10^{-3}
39	novel	1.4×10^{-2}	89	flavonoid	8.4×10^{-3}
40	antiinflammatori	1.4×10^{-2}	90	herbal	8.4×10^{-3}
41	molecul	1.3×10^{-2}	91	amino	8.3×10^{-3}
42	beta	1.3×10^{-2}	92	antimicrobi	7.8×10^{-3}
43	antibacteri	1.3×10^{-2}	93	measur	7.6×10^{-3}
44	line	1.3×10^{-2}	94	can	7.4×10^{-3}
45	enzym	1.3×10^{-2}	95	case	7.4×10^{-3}
46	rat	1.3×10^{-2}	96	antagonist	7.3×10^{-3}
47	hydroxi	1.3×10^{-2}	97	apoptosi	7.3×10^{-3}
48	substitut	1.3×10^{-2}	98	potenti	7.1×10^{-3}
49	ligand	1.3×10^{-2}	99	display	7×10^{-3}
50	new	1.3×10^{-2}	100	herb	6.9×10^{-3}

Table D.36. The list of the top 100 words in the category Chemistry, Multidisciplinary with RIGs

No.	Word	RIG	No.	Word	RIG
1	reaction	4×10^{-2}	51	excel	9.6×10^{-3}
2	synthes	3.8×10^{-2}	52	yield	9.5×10^{-3}
3	nanoparticl	2.7×10^{-2}	53	bis	9.5×10^{-3}
4	catalyst	2.5×10^{-2}	54	electrochem	9.3×10^{-3}
5	synthesi	2.5×10^{-2}	55	carbon	9.2×10^{-3}
6	prepar	2.4×10^{-2}	56	nanostructur	9.2×10^{-3}
7	paper	2.4×10^{-2}	57	graphen	9.2×10^{-3}
8	patient	2.2×10^{-2}	58	manag	9.1×10^{-3}
9	conclus	2.1×10^{-2}	59	anion	9.1×10^{-3}
10	bond	2×10^{-2}	60	aim	9×10^{-3}
11	compound	2×10^{-2}	61	format	8.7×10^{-3}
12	acid	1.9×10^{-2}	62	cation	8.6×10^{-3}
13	molecul	1.9×10^{-2}	63	fluoresc	8.5×10^{-3}
14	electron	1.8×10^{-2}	64	algorithm	8.5×10^{-3}
15	solvent	1.8×10^{-2}	65	clinic	8.5×10^{-3}
16	oxid	1.7×10^{-2}	66	model	8.1×10^{-3}
17	spectroscopi	1.7×10^{-2}	67	stabil	8.1×10^{-3}
18	ligand	1.6×10^{-2}	68	supramolecular	8.1×10^{-3}
19	structur	1.6×10^{-2}	69	propos	8×10^{-3}
20	metal	1.6×10^{-2}	70	poli	8×10^{-3}
21	year	1.5×10^{-2}	71	moieti	7.9×10^{-3}
22	catalyt	1.5×10^{-2}	72	outcom	7.8×10^{-3}
23	crystal	1.5×10^{-2}	73	inform	7.7×10^{-3}
24	object	1.5×10^{-2}	74	charg	7.7×10^{-3}
25	aqueous	1.5×10^{-2}	75	may	7.7×10^{-3}
26	nmr	1.4×10^{-2}	76	amin	7.5×10^{-3}
27	chemic	1.4×10^{-2}	77	case	7.4×10^{-3}
28	ray	1.4×10^{-2}	78	herein	7.4×10^{-3}
29	hydrogen	1.4×10^{-2}	79	chemistri	7.4×10^{-3}
30	age	1.3×10^{-2}	80	ionic	7.2×10^{-3}
31	polym	1.3×10^{-2}	81	aryl	7.2×10^{-3}
32	atom	1.3×10^{-2}	82	dft	7.1×10^{-3}
33	ion	1.2×10^{-2}	83	level	6.9×10^{-3}
34	selfassembl	1.2×10^{-2}	84	aromat	6.9×10^{-3}
35	associ	1.2×10^{-2}	85	particip	6.9×10^{-3}
36	catalyz	1.1×10^{-2}	86	consid	6.8×10^{-3}
37	exhibit	1.1×10^{-2}	87	fabric	6.8×10^{-3}
38	surfac	1.1×10^{-2}	88	substitut	6.7×10^{-3}
39	result	1.1×10^{-2}	89	character	6.6×10^{-3}
40	background	1.1×10^{-2}	90	popul	6.6×10^{-3}
41	data	1.1×10^{-2}	91	conjug	6.5×10^{-3}
42	diffract	1.1×10^{-2}	92	alkyl	6.5×10^{-3}
43	assess	1×10^{-2}	93	research	6.4×10^{-3}
44	risk	1×10^{-2}	94	social	6.4×10^{-3}
45	problem	9.9×10^{-3}	95	substitu	6.4×10^{-3}
46	properti	9.9×10^{-3}	96	polymer	6.3×10^{-3}
47	adsorpt	9.8×10^{-3}	97	dye	6.3×10^{-3}
48	molecular	9.8×10^{-3}	98	temperatur	6.3×10^{-3}
49	dot	9.7×10^{-3}	99	tem	6.2×10^{-3}
50	microscopi	9.6×10^{-3}	100	studi	6.1×10^{-3}

TABLE D.37. The list of the top 100 words in the category Chemistry, Organic with RIGs

No.	Word	RIG
1	reaction	1.1×10^{-1}
2	synthesi	8.4×10^{-2}
3	catalyz	6×10^{-2}
4	yield	5.7×10^{-2}
5	compound	5.6×10^{-2}
6	aryl	4.8×10^{-2}
7	synthes	4.7×10^{-2}
8	substitut	4.5×10^{-2}
9	result	3.4×10^{-2}
10	cycliz	3.4×10^{-2}
11	nmr	3.3×10^{-2}
12	paper	3×10^{-2}
13	afford	2.9×10^{-2}
14	catalyst	2.9×10^{-2}
15	deriv	2.9×10^{-2}
16	acid	2.8×10^{-2}
17	ring	2.8×10^{-2}
18	amin	2.7×10^{-2}
19	alkyl	2.5×10^{-2}
20	bond	2.4×10^{-2}
21	moieti	2.4×10^{-2}
22	aldehyd	2.4×10^{-2}
23	heterocycl	2.3×10^{-2}
24	intramolecular	2.3×10^{-2}
25	chiral	2.3×10^{-2}
26	substitu	2.3×10^{-2}
27	enantioselect	2.3×10^{-2}
28	conclus	2.3×10^{-2}
29	pot	2.2×10^{-2}
30	aromat	2.1×10^{-2}
31	alkyn	2×10^{-2}
32	patient	2×10^{-2}
33	nucleophil	1.9×10^{-2}
34	synthet	1.9×10^{-2}
35	ligand	1.9×10^{-2}
36	prepar	1.9×10^{-2}
37	keton	1.9×10^{-2}
38	palladium	1.8×10^{-2}
39	model	1.8×10^{-2}
40	ester	1.7×10^{-2}
41	reagent	1.7×10^{-2}
42	measur	1.7×10^{-2}
43	studi	1.6×10^{-2}
44	increas	1.6×10^{-2}
45	amino	1.6×10^{-2}
46	object	1.5×10^{-2}
47	bis	1.5×10^{-2}
48	catalyt	1.5×10^{-2}
49	data	1.5×10^{-2}
50	year	1.5×10^{-2}
51	perform	1.4×10^{-2}
52	associ	1.4×10^{-2}
53	methyl	1.4×10^{-2}
54	proceed	1.3×10^{-2}
55	amid	1.3×10^{-2}
56	excel	1.3×10^{-2}
57	phenyl	1.3×10^{-2}
58	mild	1.2×10^{-2}
59	age	1.2×10^{-2}
60	assess	1.2×10^{-2}
61	background	1.2×10^{-2}
62	simul	1.2×10^{-2}
63	compar	1.2×10^{-2}
64	differ	1.2×10^{-2}
65	intermedi	1.2×10^{-2}
66	signific	1.2×10^{-2}
67	level	1.2×10^{-2}
68	dure	1.2×10^{-2}
69	may	1.2×10^{-2}
70	problem	1.2×10^{-2}
71	good	1.1×10^{-2}
72	pyridin	1.1×10^{-2}
73	propos	1.1×10^{-2}
74	react	1.1×10^{-2}
75	present	1.1×10^{-2}
76	aim	1.1×10^{-2}
77	find	1.1×10^{-2}
78	analogu	1.1×10^{-2}
79	alpha	1.1×10^{-2}
80	howev	1.1×10^{-2}
81	solvent	1.1×10^{-2}
82	inform	1.1×10^{-2}
83	factor	1×10^{-2}
84	seri	1×10^{-2}
85	carboxyl	1×10^{-2}
86	sampl	1×10^{-2}
87	potent	1×10^{-2}
88	algorithm	1×10^{-2}
89	beta	1×10^{-2}
90	carbonyl	1×10^{-2}
91	consid	9.9×10^{-3}
92	asymmetr	9.9×10^{-3}
93	research	9.8×10^{-3}
94	analyz	9.8×10^{-3}
95	can	9.8×10^{-3}
96	hydroxi	9.7×10^{-3}
97	use	9.7×10^{-3}
98	conjug	9.6×10^{-3}
99	time	9.5×10^{-3}
100	risk	9.5×10^{-3}

Table D.38. The list of the top 100 words in the category Chemistry, Physical with RIGs

No.	Word	RIG	No.	Word	RIG
1	electron	4.6×10^{-2}	51	spectra	1.3×10^{-2}
2	surfac	4.3×10^{-2}	52	exhibit	1.3×10^{-2}
3	catalyst	3.9×10^{-2}	53	photoelectron	1.3×10^{-2}
4	spectroscopi	3.6×10^{-2}	54	phase	1.3×10^{-2}
5	temperatur	3.1×10^{-2}	55	dft	1.3×10^{-2}
6	reaction	3.1×10^{-2}	56	tio2	1.3×10^{-2}
7	hydrogen	3.1×10^{-2}	57	crystal	1.3×10^{-2}
8	atom	2.9×10^{-2}	58	risk	1.3×10^{-2}
9	energi	2.9×10^{-2}	59	xps	1.3×10^{-2}
10	molecul	2.8×10^{-2}	60	cation	1.3×10^{-2}
11	oxid	2.7×10^{-2}	61	dope	1.2×10^{-2}
12	structur	2.7×10^{-2}	62	diseas	1.2×10^{-2}
13	patient	2.7×10^{-2}	63	polym	1.2×10^{-2}
14	charg	2.7×10^{-2}	64	absorpt	1.2×10^{-2}
15	conclus	2.7×10^{-2}	65	kinet	1.2×10^{-2}
16	adsorpt	2.6×10^{-2}	66	solid	1.2×10^{-2}
17	nanoparticl	2.6×10^{-2}	67	assess	1.2×10^{-2}
18	prepar	2.5×10^{-2}	68	composit	1.2×10^{-2}
19	ion	2.5×10^{-2}	69	layer	1.2×10^{-2}
20	properti	2.4×10^{-2}	70	liquid	1.1×10^{-2}
21	catalyt	2.4×10^{-2}	71	thermodynam	1.1×10^{-2}
22	densiti	2.3×10^{-2}	72	particl	1.1×10^{-2}
23	ray	2.3×10^{-2}	73	age	1.1×10^{-2}
24	bond	2.2×10^{-2}	74	raman	1.1×10^{-2}
25	synthes	2.1×10^{-2}	75	mol	1.1×10^{-2}
26	electrochem	2.1×10^{-2}	76	tem	1.1×10^{-2}
27	metal	2×10^{-2}	77	associ	1.1×10^{-2}
28	paper	2×10^{-2}	78	mesopor	1.1×10^{-2}
29	carbon	1.9×10^{-2}	79	water	1×10^{-2}
30	chemic	1.8×10^{-2}	80	aim	1×10^{-2}
31	diffract	1.8×10^{-2}	81	photocatalyt	1×10^{-2}
32	molecular	1.7×10^{-2}	82	nanostructur	1×10^{-2}
33	xrd	1.7×10^{-2}	83	lithium	1×10^{-2}
34	microscopi	1.7×10^{-2}	84	dispers	9.9×10^{-3}
35	year	1.7×10^{-2}	85	oxygen	9.9×10^{-3}
36	format	1.6×10^{-2}	86	transit	9.8×10^{-3}
37	calcul	1.6×10^{-2}	87	crystallin	9.8×10^{-3}
38	stabil	1.6×10^{-2}	88	surfact	9.5×10^{-3}
39	solvent	1.5×10^{-2}	89	thermal	9.5×10^{-3}
40	adsorb	1.5×10^{-2}	90	deposit	9.4×10^{-3}
41	aqueous	1.5×10^{-2}	91	anod	9×10^{-3}
42	electrod	1.5×10^{-2}	92	outcom	9×10^{-3}
43	electrolyt	1.5×10^{-2}	93	nanotub	9×10^{-3}
44	object	1.5×10^{-2}	94	problem	8.9×10^{-3}
45	graphen	1.5×10^{-2}	95	degre	8.9×10^{-3}
46	ionic	1.5×10^{-2}	96	solut	8.9×10^{-3}
47	film	1.5×10^{-2}	97	gene	8.9×10^{-3}
48	background	1.5×10^{-2}	98	character	8.8×10^{-3}
49	clinic	1.4×10^{-2}	99	express	8.8×10^{-3}
50	materi	1.4×10^{-2}	100	manag	8.8×10^{-3}

TABLE D.39. The list of the top 100 words in the category Classics with RIGs

No.	Word	RIG	No.	Word	RIG
1	greek	9.3×10^{-2}	51	measur	1.5×10^{-2}
2	roman	9.3×10^{-2}	52	high	1.5×10^{-2}
3	argu	6.7×10^{-2}	53	charact	1.5×10^{-2}
4	ancient	6.3×10^{-2}	54	argument	1.4×10^{-2}
5	poet	6.2×10^{-2}	55	latin	1.4×10^{-2}
6	literari	5.6×10^{-2}	56	imperi	1.4×10^{-2}
7	poem	5.4×10^{-2}	57	inscript	1.4×10^{-2}
8	result	5.2×10^{-2}	58	intertextu	1.4×10^{-2}
9	text	4.9×10^{-2}	59	higher	1.4×10^{-2}
10	poetri	4.9×10^{-2}	60	low	1.4×10^{-2}
11	poetic	4.7×10^{-2}	61	patient	1.3×10^{-2}
12	articl	4.4×10^{-2}	62	compar	1.3×10^{-2}
13	scholar	4.4×10^{-2}	63	test	1.3×10^{-2}
14	narrat	3.7×10^{-2}	64	evalu	1.3×10^{-2}
15	centuri	3.7×10^{-2}	65	cell	1.3×10^{-2}
16	epic	3.6×10^{-2}	66	imit	1.2×10^{-2}
17	read	3.4×10^{-2}	67	figur	1.2×10^{-2}
18	rome	3.3×10^{-2}	68	rate	1.2×10^{-2}
19	book	3.2×10^{-2}	69	histori	1.2×10^{-2}
20	passag	3×10^{-2}	70	dialogu	1.2×10^{-2}
21	method	2.8×10^{-2}	71	histor	1.2×10^{-2}
22	divin	2.8×10^{-2}	72	hero	1.2×10^{-2}
23	studi	2.7×10^{-2}	73	question	1.2×10^{-2}
24	allus	2.7×10^{-2}	74	philosophi	1.2×10^{-2}
25	interpret	2.5×10^{-2}	75	determin	1.2×10^{-2}
26	polit	2.4×10^{-2}	76	context	1.1×10^{-2}
27	rhetor	2.3×10^{-2}	77	tale	1.1×10^{-2}
28	data	2.3×10^{-2}	78	obtain	1.1×10^{-2}
29	tragedi	2.1×10^{-2}	79	theme	1.1×10^{-2}
30	stori	2.1×10^{-2}	80	improv	1.1×10^{-2}
31	antiqu	2.1×10^{-2}	81	reader	1.1×10^{-2}
32	effect	2.1×10^{-2}	82	panegyr	1.1×10^{-2}
33	myth	1.9×10^{-2}	83	simul	1.1×10^{-2}
34	use	1.9×10^{-2}	84	contemporari	1.1×10^{-2}
35	increas	1.8×10^{-2}	85	satir	1×10^{-2}
36	scholarship	1.8×10^{-2}	86	textual	1×10^{-2}
37	base	1.8×10^{-2}	87	decreas	1×10^{-2}
38	author	1.8×10^{-2}	88	epsilon	1×10^{-2}
39	word	1.8×10^{-2}	89	genr	1×10^{-2}
40	historian	1.8×10^{-2}	90	modern	1×10^{-2}
41	system	1.7×10^{-2}	91	control	1×10^{-2}
42	vers	1.7×10^{-2}	92	observ	1×10^{-2}
43	recept	1.7×10^{-2}	93	whi	9.9×10^{-3}
44	tradit	1.7×10^{-2}	94	perform	9.9×10^{-3}
45	epigram	1.6×10^{-2}	95	speech	9.9×10^{-3}
46	audienc	1.6×10^{-2}	96	paramet	9.8×10^{-3}
47	war	1.6×10^{-2}	97	attempt	9.8×10^{-3}
48	god	1.5×10^{-2}	98	lyric	9.8×10^{-3}
49	philosoph	1.5×10^{-2}	99	sampl	9.8×10^{-3}
50	famous	1.5×10^{-2}	100	theatric	9.6×10^{-3}

Table D.40. The list of the top 100 words in the category Clinical Neurology with RIGs

No.	Word	RIG	No.	Word	RIG
1	patient	1.6×10^{-1}	51	depress	1.8×10^{-2}
2	conclus	9.2×10^{-2}	52	result	1.8×10^{-2}
3	clinic	6.6×10^{-2}	53	spine	1.7×10^{-2}
4	brain	5.9×10^{-2}	54	posterior	1.7×10^{-2}
5	neurolog	4.6×10^{-2}	55	lumbar	1.7×10^{-2}
6	background	4.5×10^{-2}	56	group	1.7×10^{-2}
7	object	4.1×10^{-2}	57	healthi	1.7×10^{-2}
8	symptom	4.1×10^{-2}	58	ischem	1.6×10^{-2}
9	age	3.9×10^{-2}	59	postop	1.6×10^{-2}
10	epilepsi	3.7×10^{-2}	60	injuri	1.6×10^{-2}
11	disord	3.7×10^{-2}	61	cortic	1.6×10^{-2}
12	spinal	3.7×10^{-2}	62	cohort	1.6×10^{-2}
13	cerebr	3.7×10^{-2}	63	medic	1.6×10^{-2}
14	score	3.6×10^{-2}	64	neuropsycholog	1.5×10^{-2}
15	seizur	3.6×10^{-2}	65	headach	1.5×10^{-2}
16	outcom	3.4×10^{-2}	66	dementia	1.5×10^{-2}
17	year	3.2×10^{-2}	67	propos	1.5×10^{-2}
18	paper	3.1×10^{-2}	68	treat	1.5×10^{-2}
19	associ	3.1×10^{-2}	69	anterior	1.4×10^{-2}
20	method	3×10^{-2}	70	abnorm	1.4×10^{-2}
21	stroke	3×10^{-2}	71	acut	1.4×10^{-2}
22	diseas	2.9×10^{-2}	72	prospect	1.4×10^{-2}
23	treatment	2.9×10^{-2}	73	report	1.4×10^{-2}
24	mri	2.7×10^{-2}	74	cord	1.4×10^{-2}
25	intracrani	2.7×10^{-2}	75	risk	1.4×10^{-2}
26	underw	2.7×10^{-2}	76	review	1.4×10^{-2}
27	cognit	2.7×10^{-2}	77	baselin	1.4×10^{-2}
28	assess	2.7×10^{-2}	78	aneurysm	1.3×10^{-2}
29	pain	2.6×10^{-2}	79	consecut	1.3×10^{-2}
30	month	2.6×10^{-2}	80	imag	1.3×10^{-2}
31	onset	2.5×10^{-2}	81	includ	1.3×10^{-2}
32	retrospect	2.4×10^{-2}	82	diagnos	1.3×10^{-2}
33	sclerosi	2.4×10^{-2}	83	arteri	1.3×10^{-2}
34	surgeri	2.4×10^{-2}	84	simul	1.3×10^{-2}
35	signific	2.3×10^{-2}	85	alzheim	1.3×10^{-2}
36	impair	2.2×10^{-2}	86	temperatur	1.3×10^{-2}
37	deficit	2.1×10^{-2}	87	cerebrospin	1.2×10^{-2}
38	motor	2.1×10^{-2}	88	cortex	1.2×10^{-2}
39	surgic	2.1×10^{-2}	89	subject	1.2×10^{-2}
40	disabl	2.1×10^{-2}	90	particip	1.2×10^{-2}
41	follow	2.1×10^{-2}	91	bilater	1.1×10^{-2}
42	sleep	2×10^{-2}	92	traumat	1.1×10^{-2}
43	studi	2×10^{-2}	93	therapi	1.1×10^{-2}
44	hemorrhag	2×10^{-2}	94	frontal	1.1×10^{-2}
45	nerv	1.9×10^{-2}	95	patholog	1.1×10^{-2}
46	parkinson	1.9×10^{-2}	96	complic	1.1×10^{-2}
47	diagnosi	1.9×10^{-2}	97	adult	1.1×10^{-2}
48	lesion	1.8×10^{-2}	98	eeg	1×10^{-2}
49	may	1.8×10^{-2}	99	energi	1×10^{-2}
50	syndrom	1.8×10^{-2}	100	solut	1×10^{-2}

Table D.41. The list of the top 100 words in the category Communication with RIGs

No.	Word	RIG	No.	Word	RIG
1	media	1×10^{-1}	51	essay	1.1×10^{-2}
2	news	6.7×10^{-2}	52	simul	1.1×10^{-2}
3	social	6×10^{-2}	53	campaign	1.1×10^{-2}
4	communic	5.6×10^{-2}	54	talk	1.1×10^{-2}
5	articl	5.5×10^{-2}	55	paramet	1.1×10^{-2}
6	polit	5.2×10^{-2}	56	surfac	1×10^{-2}
7	journalist	4.7×10^{-2}	57	genr	1×10^{-2}
8	discours	4.6×10^{-2}	58	digit	1×10^{-2}
9	public	4.3×10^{-2}	59	entertain	1×10^{-2}
10	audienc	3.5×10^{-2}	60	perform	1×10^{-2}
11	televis	3.4×10^{-2}	61	question	1×10^{-2}
12	onlin	3.1×10^{-2}	62	conclus	1×10^{-2}
13	argu	3×10^{-2}	63	text	1×10^{-2}
14	newspap	2.9×10^{-2}	64	obtain	9.9×10^{-3}
15	method	2.3×10^{-2}	65	seek	9.9×10^{-3}
16	examin	2.3×10^{-2}	66	energi	9.8×10^{-3}
17	research	2.2×10^{-2}	67	societi	9.7×10^{-3}
18	engag	2.2×10^{-2}	68	debat	9.4×10^{-3}
19	advertis	2.1×10^{-2}	69	protein	9.3×10^{-3}
20	narrat	2×10^{-2}	70	theori	9.2×10^{-3}
21	journal	1.9×10^{-2}	71	high	9.2×10^{-3}
22	discurs	1.8×10^{-2}	72	inform	9.2×10^{-3}
23	draw	1.8×10^{-2}	73	market	9.1×10^{-3}
24	citizen	1.7×10^{-2}	74	perspect	9.1×10^{-3}
25	messag	1.6×10^{-2}	75	brand	9.1×10^{-3}
26	internet	1.6×10^{-2}	76	corpor	9×10^{-3}
27	cultur	1.6×10^{-2}	77	contemporari	8.9×10^{-3}
28	stori	1.6×10^{-2}	78	system	8.9×10^{-3}
29	peopl	1.5×10^{-2}	79	author	8.8×10^{-3}
30	attitud	1.5×10^{-2}	80	american	8.7×10^{-3}
31	rhetor	1.5×10^{-2}	81	patient	8.6×10^{-3}
32	cell	1.5×10^{-2}	82	parti	8.6×10^{-3}
33	interview	1.5×10^{-2}	83	democrat	8.6×10^{-3}
34	context	1.5×10^{-2}	84	person	8.5×10^{-3}
35	nation	1.4×10^{-2}	85	detect	8.4×10^{-3}
36	practic	1.4×10^{-2}	86	relationship	8.3×10^{-3}
37	implic	1.4×10^{-2}	87	ratio	8.3×10^{-3}
38	scholar	1.4×10^{-2}	88	viewer	8.2×10^{-3}
39	perceiv	1.3×10^{-2}	89	frame	8.2×10^{-3}
40	result	1.3×10^{-2}	90	argument	8×10^{-3}
41	way	1.3×10^{-2}	91	properti	8×10^{-3}
42	discuss	1.3×10^{-2}	92	issu	7.9×10^{-3}
43	explor	1.3×10^{-2}	93	view	7.8×10^{-3}
44	survey	1.3×10^{-2}	94	water	7.8×10^{-3}
45	focus	1.2×10^{-2}	95	find	7.8×10^{-3}
46	temperatur	1.2×10^{-2}	96	democraci	7.7×10^{-3}
47	percept	1.2×10^{-2}	97	understand	7.7×10^{-3}
48	broadcast	1.2×10^{-2}	98	content	7.6×10^{-3}
49	particip	1.2×10^{-2}	99	conceptu	7.6×10^{-3}
50	ideolog	1.2×10^{-2}	100	agenda	7.6×10^{-3}

TABLE D.42. The list of the top 100 words in the category Computer Science, Artificial Intelligence with RIGs

No.	Word	RIG
1	paper	9.9×10^{-2}
2	propos	8.9×10^{-2}
3	algorithm	7.9×10^{-2}
4	problem	4×10^{-2}
5	learn	3.4×10^{-2}
6	base	3×10^{-2}
7	approach	2.8×10^{-2}
8	fuzzi	2.7×10^{-2}
9	dataset	2.7×10^{-2}
10	comput	2.5×10^{-2}
11	studi	2.5×10^{-2}
12	conclus	2.4×10^{-2}
13	real	2.2×10^{-2}
14	classif	2.2×10^{-2}
15	task	2.2×10^{-2}
16	art	2×10^{-2}
17	robot	2×10^{-2}
18	featur	1.9×10^{-2}
19	recognit	1.9×10^{-2}
20	automat	1.7×10^{-2}
21	inform	1.7×10^{-2}
22	imag	1.6×10^{-2}
23	cell	1.6×10^{-2}
24	set	1.6×10^{-2}
25	network	1.6×10^{-2}
26	outperform	1.6×10^{-2}
27	machin	1.5×10^{-2}
28	user	1.5×10^{-2}
29	treatment	1.5×10^{-2}
30	neural	1.5×10^{-2}
31	patient	1.4×10^{-2}
32	accuraci	1.4×10^{-2}
33	train	1.4×10^{-2}
34	classifi	1.4×10^{-2}
35	solv	1.4×10^{-2}
36	represent	1.3×10^{-2}
37	semant	1.3×10^{-2}
38	intellig	1.3×10^{-2}
39	temperatur	1.3×10^{-2}
40	system	1.2×10^{-2}
41	benchmark	1.2×10^{-2}
42	optim	1.2×10^{-2}
43	search	1.1×10^{-2}
44	increas	1.1×10^{-2}
45	introduc	1.1×10^{-2}
46	framework	1.1×10^{-2}
47	suggest	1.1×10^{-2}
48	protein	1.1×10^{-2}
49	found	1.1×10^{-2}
50	perform	1.1×10^{-2}

No.	Word	RIG
51	acid	1.1×10^{-2}
52	decreas	1×10^{-2}
53	robust	1×10^{-2}
54	concentr	1×10^{-2}
55	associ	1×10^{-2}
56	age	9.9×10^{-3}
57	model	9.9×10^{-3}
58	signific	9.5×10^{-3}
59	can	9.5×10^{-3}
60	vector	9.4×10^{-3}
61	induc	9.4×10^{-3}
62	investig	9.3×10^{-3}
63	scene	9×10^{-3}
64	experi	9×10^{-3}
65	materi	8.9×10^{-3}
66	speci	8.7×10^{-3}
67	clinic	8.6×10^{-3}
68	report	8.5×10^{-3}
69	oxid	8.4×10^{-3}
70	video	8.4×10^{-3}
71	graph	8.2×10^{-3}
72	higher	7.9×10^{-3}
73	adapt	7.8×10^{-3}
74	camera	7.8×10^{-3}
75	swarm	7.7×10^{-3}
76	reaction	7.7×10^{-3}
77	observ	7.7×10^{-3}
78	svm	7.7×10^{-3}
79	examin	7.6×10^{-3}
80	water	7.5×10^{-3}
81	decis	7.4×10^{-3}
82	vision	7.4×10^{-3}
83	molecular	7.4×10^{-3}
84	applic	7.2×10^{-3}
85	method	7.2×10^{-3}
86	experiment	7.2×10^{-3}
87	extract	7.2×10^{-3}
88	diseas	7.1×10^{-3}
89	reveal	7.1×10^{-3}
90	activ	7.1×10^{-3}
91	call	7×10^{-3}
92	electron	6.9×10^{-3}
93	gene	6.9×10^{-3}
94	total	6.8×10^{-3}
95	make	6.7×10^{-3}
96	techniqu	6.7×10^{-3}
97	input	6.7×10^{-3}
98	mine	6.6×10^{-3}
99	surfac	6.6×10^{-3}
100	indic	6.5×10^{-3}

Table D.43. The list of the top 100 words in the category Computer Science, Cybernetics with RIGs

No.	Word	RIG
1	paper	5.4×10^{-2}
2	user	4.1×10^{-2}
3	propos	3.7×10^{-2}
4	algorithm	2.3×10^{-2}
5	robot	2.1×10^{-2}
6	task	2.1×10^{-2}
7	system	1.7×10^{-2}
8	learn	1.6×10^{-2}
9	conclus	1.5×10^{-2}
10	base	1.5×10^{-2}
11	comput	1.4×10^{-2}
12	visual	1.4×10^{-2}
13	problem	1.4×10^{-2}
14	real	1.3×10^{-2}
15	inform	1.2×10^{-2}
16	approach	1.2×10^{-2}
17	experi	1.1×10^{-2}
18	virtual	1.1×10^{-2}
19	cell	1.1×10^{-2}
20	interfac	1×10^{-2}
21	feedback	9.8×10^{-3}
22	recognit	9.7×10^{-3}
23	design	9.7×10^{-3}
24	treatment	9.5×10^{-3}
25	fuzzi	9×10^{-3}
26	temperatur	8.8×10^{-3}
27	interact	8.8×10^{-3}
28	camera	8.7×10^{-3}
29	patient	8.3×10^{-3}
30	gestur	8.1×10^{-3}
31	video	8×10^{-3}
32	realiti	7.8×10^{-3}
33	protein	7.8×10^{-3}
34	can	7.6×10^{-3}
35	featur	7.5×10^{-3}
36	environ	7.4×10^{-3}
37	acid	7.4×10^{-3}
38	concentr	7.4×10^{-3}
39	associ	7.3×10^{-3}
40	human	7×10^{-3}
41	machin	6.9×10^{-3}
42	usabl	6.9×10^{-3}
43	studi	6.6×10^{-3}
44	track	6.4×10^{-3}
45	accuraci	6.2×10^{-3}
46	automat	6.2×10^{-3}
47	diseas	6.2×10^{-3}
48	oxid	6.1×10^{-3}
49	represent	6×10^{-3}
50	make	5.9×10^{-3}
51	clinic	5.8×10^{-3}
52	speci	5.8×10^{-3}
53	set	5.8×10^{-3}
54	increas	5.8×10^{-3}
55	devic	5.8×10^{-3}
56	decreas	5.8×10^{-3}
57	framework	5.6×10^{-3}
58	dataset	5.5×10^{-3}
59	scene	5.5×10^{-3}
60	train	5.4×10^{-3}
61	molecular	5.4×10^{-3}
62	intellig	5.4×10^{-3}
63	total	5.4×10^{-3}
64	classif	5.4×10^{-3}
65	navig	5.3×10^{-3}
66	gene	5.2×10^{-3}
67	perform	5.2×10^{-3}
68	materi	5.2×10^{-3}
69	vision	5×10^{-3}
70	solv	5×10^{-3}
71	ratio	5×10^{-3}
72	motion	4.9×10^{-3}
73	signific	4.8×10^{-3}
74	onlin	4.8×10^{-3}
75	decis	4.7×10^{-3}
76	model	4.7×10^{-3}
77	implement	4.6×10^{-3}
78	art	4.6×10^{-3}
79	percept	4.6×10^{-3}
80	chemic	4.5×10^{-3}
81	higher	4.5×10^{-3}
82	introduc	4.5×10^{-3}
83	induc	4.5×10^{-3}
84	robust	4.5×10^{-3}
85	game	4.4×10^{-3}
86	water	4.4×10^{-3}
87	adapt	4.3×10^{-3}
88	age	4.3×10^{-3}
89	perceiv	4.3×10^{-3}
90	mobil	4.3×10^{-3}
91	movement	4.3×10^{-3}
92	applic	4.2×10^{-3}
93	growth	4.1×10^{-3}
94	prepar	4.1×10^{-3}
95	present	4.1×10^{-3}
96	ray	4×10^{-3}
97	report	4×10^{-3}
98	imag	3.9×10^{-3}
99	scenario	3.9×10^{-3}
100	electron	3.9×10^{-3}

Table D.44. The list of the top 100 words in the category Computer Science, Hardware and Architecture with RIGs

No.	Word	RIG
1	paper	8.1×10^{-2}
2	propos	6.9×10^{-2}
3	network	6.5×10^{-2}
4	algorithm	3.8×10^{-2}
5	wireless	3.5×10^{-2}
6	user	3×10^{-2}
7	architectur	3×10^{-2}
8	communic	2.7×10^{-2}
9	studi	2.7×10^{-2}
10	implement	2.6×10^{-2}
11	node	2.6×10^{-2}
12	overhead	2.5×10^{-2}
13	hardwar	2.5×10^{-2}
14	conclus	2.2×10^{-2}
15	comput	2.1×10^{-2}
16	scheme	2×10^{-2}
17	servic	1.9×10^{-2}
18	traffic	1.9×10^{-2}
19	effici	1.9×10^{-2}
20	simul	1.9×10^{-2}
21	system	1.8×10^{-2}
22	applic	1.8×10^{-2}
23	base	1.8×10^{-2}
24	patient	1.7×10^{-2}
25	power	1.7×10^{-2}
26	packet	1.7×10^{-2}
27	processor	1.7×10^{-2}
28	perform	1.7×10^{-2}
29	resourc	1.6×10^{-2}
30	deploy	1.6×10^{-2}
31	problem	1.5×10^{-2}
32	scalabl	1.5×10^{-2}
33	treatment	1.5×10^{-2}
34	mobil	1.5×10^{-2}
35	design	1.5×10^{-2}
36	technolog	1.4×10^{-2}
37	memori	1.4×10^{-2}
38	throughput	1.4×10^{-2}
39	cloud	1.4×10^{-2}
40	execut	1.4×10^{-2}
41	bandwidth	1.3×10^{-2}
42	secur	1.3×10^{-2}
43	suggest	1.3×10^{-2}
44	server	1.3×10^{-2}
45	access	1.3×10^{-2}
46	can	1.3×10^{-2}
47	found	1.2×10^{-2}
48	optim	1.2×10^{-2}
49	chip	1.2×10^{-2}
50	internet	1.2×10^{-2}

No.	Word	RIG
51	infrastructur	1.2×10^{-2}
52	consumpt	1.2×10^{-2}
53	fpga	1.2×10^{-2}
54	clinic	1.2×10^{-2}
55	devic	1.2×10^{-2}
56	achiev	1.1×10^{-2}
57	schedul	1.1×10^{-2}
58	protein	1.1×10^{-2}
59	associ	1.1×10^{-2}
60	bit	1.1×10^{-2}
61	real	1.1×10^{-2}
62	protocol	1.1×10^{-2}
63	latenc	1.1×10^{-2}
64	softwar	1.1×10^{-2}
65	requir	9.9×10^{-3}
66	diseas	9.9×10^{-3}
67	circuit	9.9×10^{-3}
68	acid	9.7×10^{-3}
69	surfac	9.6×10^{-3}
70	workload	9.4×10^{-3}
71	cost	9.4×10^{-3}
72	concentr	9.4×10^{-3}
73	age	9.3×10^{-3}
74	techniqu	9.1×10^{-3}
75	platform	9×10^{-3}
76	speci	8.8×10^{-3}
77	oper	8.7×10^{-3}
78	materi	8.7×10^{-3}
79	group	8.7×10^{-3}
80	run	8.6×10^{-3}
81	virtual	8.5×10^{-3}
82	channel	8.5×10^{-3}
83	observ	8.4×10^{-3}
84	background	8.3×10^{-3}
85	gene	8.2×10^{-3}
86	parallel	8.2×10^{-3}
87	cmos	8.2×10^{-3}
88	examin	8.1×10^{-3}
89	radio	8.1×10^{-3}
90	exploit	8×10^{-3}
91	rout	8×10^{-3}
92	scenario	7.9×10^{-3}
93	investig	7.9×10^{-3}
94	delay	7.9×10^{-3}
95	share	7.8×10^{-3}
96	approach	7.8×10^{-3}
97	enabl	7.8×10^{-3}
98	indic	7.7×10^{-3}
99	assess	7.6×10^{-3}
100	reaction	7.6×10^{-3}

TABLE D.45. The list of the top 100 words in the category Computer Science, Information Systems with RIGs

No.	Word	RIG
1	paper	9.4×10^{-2}
2	propos	7.2×10^{-2}
3	user	5×10^{-2}
4	network	4.2×10^{-2}
5	algorithm	4.1×10^{-2}
6	base	2.5×10^{-2}
7	inform	2.4×10^{-2}
8	servic	2.4×10^{-2}
9	problem	2.2×10^{-2}
10	comput	2.2×10^{-2}
11	conclus	2×10^{-2}
12	studi	1.9×10^{-2}
13	wireless	1.9×10^{-2}
14	secur	1.9×10^{-2}
15	implement	1.7×10^{-2}
16	system	1.7×10^{-2}
17	communic	1.6×10^{-2}
18	approach	1.6×10^{-2}
19	treatment	1.6×10^{-2}
20	cloud	1.5×10^{-2}
21	cell	1.5×10^{-2}
22	node	1.4×10^{-2}
23	patient	1.4×10^{-2}
24	real	1.4×10^{-2}
25	queri	1.4×10^{-2}
26	web	1.4×10^{-2}
27	internet	1.3×10^{-2}
28	scheme	1.3×10^{-2}
29	temperatur	1.3×10^{-2}
30	applic	1.3×10^{-2}
31	semant	1.2×10^{-2}
32	architectur	1.2×10^{-2}
33	can	1.1×10^{-2}
34	found	1.1×10^{-2}
35	surfac	1.1×10^{-2}
36	concentr	1.1×10^{-2}
37	mobil	1.1×10^{-2}
38	associ	1.1×10^{-2}
39	induc	1.1×10^{-2}
40	acid	1.1×10^{-2}
41	effici	1.1×10^{-2}
42	protein	1×10^{-2}
43	technolog	1×10^{-2}
44	framework	1×10^{-2}
45	observ	1×10^{-2}
46	deploy	1×10^{-2}
47	task	1×10^{-2}
48	data	1×10^{-2}
49	age	1×10^{-2}
50	dataset	1×10^{-2}

No.	Word	RIG
51	resourc	9.7×10^{-3}
52	speci	9.7×10^{-3}
53	suggest	9.6×10^{-3}
54	softwar	9.6×10^{-3}
55	scalabl	9.3×10^{-3}
56	attack	9.2×10^{-3}
57	materi	9.2×10^{-3}
58	investig	9.1×10^{-3}
59	learn	9.1×10^{-3}
60	automat	9.1×10^{-3}
61	exist	8.9×10^{-3}
62	decreas	8.8×10^{-3}
63	infrastructur	8.8×10^{-3}
64	oxid	8.7×10^{-3}
65	access	8.7×10^{-3}
66	traffic	8.7×10^{-3}
67	privaci	8.7×10^{-3}
68	clinic	8.6×10^{-3}
69	diseas	8.4×10^{-3}
70	water	8.3×10^{-3}
71	server	8.3×10^{-3}
72	signific	8.3×10^{-3}
73	gene	8.2×10^{-3}
74	packet	8.1×10^{-3}
75	reaction	8.1×10^{-3}
76	challeng	8×10^{-3}
77	platform	7.9×10^{-3}
78	introduc	7.9×10^{-3}
79	outperform	7.9×10^{-3}
80	perform	7.8×10^{-3}
81	execut	7.8×10^{-3}
82	set	7.8×10^{-3}
83	machin	7.7×10^{-3}
84	overhead	7.6×10^{-3}
85	call	7.5×10^{-3}
86	messag	7.5×10^{-3}
87	busi	7.4×10^{-3}
88	sensor	7.4×10^{-3}
89	background	7.3×10^{-3}
90	make	7.2×10^{-3}
91	higher	7.1×10^{-3}
92	popular	7×10^{-3}
93	virtual	7×10^{-3}
94	provid	6.9×10^{-3}
95	video	6.9×10^{-3}
96	report	6.9×10^{-3}
97	onlin	6.9×10^{-3}
98	environ	6.9×10^{-3}
99	issu	6.9×10^{-3}
100	scenario	6.9×10^{-3}

Table D.46. The list of the top 100 words in the category Computer Science, Interdisciplinary Applications with RIGs

No.	Word	RIG
1	paper	4.8×10^{-2}
2	propos	3.4×10^{-2}
3	algorithm	3.3×10^{-2}
4	comput	2.7×10^{-2}
5	problem	1.8×10^{-2}
6	base	1.7×10^{-2}
7	conclus	1.6×10^{-2}
8	approach	1.4×10^{-2}
9	simul	1.4×10^{-2}
10	user	1.4×10^{-2}
11	implement	1.3×10^{-2}
12	model	1.1×10^{-2}
13	system	1.1×10^{-2}
14	network	1.1×10^{-2}
15	inform	1×10^{-2}
16	solv	8.7×10^{-3}
17	treatment	8.6×10^{-3}
18	cell	8.4×10^{-3}
19	accuraci	8.3×10^{-3}
20	studi	8.2×10^{-3}
21	softwar	7.9×10^{-3}
22	signific	7.6×10^{-3}
23	age	7.4×10^{-3}
24	automat	7.2×10^{-3}
25	applic	7.2×10^{-3}
26	patient	7.1×10^{-3}
27	real	7.1×10^{-3}
28	technolog	7×10^{-3}
29	concentr	7×10^{-3}
30	induc	6.9×10^{-3}
31	decreas	6.8×10^{-3}
32	suggest	6.7×10^{-3}
33	associ	6.5×10^{-3}
34	machin	6.4×10^{-3}
35	optim	6.4×10^{-3}
36	temperatur	6.4×10^{-3}
37	oxid	6.3×10^{-3}
38	tool	6.3×10^{-3}
39	background	6.2×10^{-3}
40	acid	6×10^{-3}
41	can	5.8×10^{-3}
42	web	5.8×10^{-3}
43	observ	5.7×10^{-3}
44	design	5.7×10^{-3}
45	dataset	5.7×10^{-3}
46	set	5.6×10^{-3}
47	learn	5.4×10^{-3}
48	increas	5.4×10^{-3}
49	speci	5.4×10^{-3}
50	fuzzi	5.3×10^{-3}

No.	Word	RIG
51	report	5.3×10^{-3}
52	found	5.2×10^{-3}
53	data	5.1×10^{-3}
54	scheme	5.1×10^{-3}
55	present	5.1×10^{-3}
56	month	5.1×10^{-3}
57	investig	5×10^{-3}
58	use	5×10^{-3}
59	framework	5×10^{-3}
60	task	5×10^{-3}
61	higher	4.8×10^{-3}
62	decis	4.8×10^{-3}
63	perform	4.8×10^{-3}
64	spectroscopi	4.7×10^{-3}
65	prepar	4.6×10^{-3}
66	introduc	4.5×10^{-3}
67	classif	4.5×10^{-3}
68	make	4.5×10^{-3}
69	microscopi	4.4×10^{-3}
70	group	4.3×10^{-3}
71	activ	4.2×10^{-3}
72	integr	4.2×10^{-3}
73	effici	4.2×10^{-3}
74	inhibit	4.1×10^{-3}
75	techniqu	4.1×10^{-3}
76	featur	4.1×10^{-3}
77	lower	4.1×10^{-3}
78	intellig	4×10^{-3}
79	discret	3.9×10^{-3}
80	virtual	3.9×10^{-3}
81	servic	3.9×10^{-3}
82	order	3.9×10^{-3}
83	examin	3.9×10^{-3}
84	appli	3.8×10^{-3}
85	methodolog	3.8×10^{-3}
86	mice	3.8×10^{-3}
87	graphic	3.8×10^{-3}
88	internet	3.8×10^{-3}
89	ray	3.8×10^{-3}
90	reveal	3.8×10^{-3}
91	process	3.7×10^{-3}
92	research	3.7×10^{-3}
93	diffract	3.6×10^{-3}
94	dure	3.6×10^{-3}
95	male	3.6×10^{-3}
96	languag	3.6×10^{-3}
97	mediat	3.6×10^{-3}
98	autom	3.6×10^{-3}
99	year	3.6×10^{-3}
100	databas	3.6×10^{-3}

Table D.47. The list of the top 100 words in the category Computer Science, Software Engineering with RIGs

No.	Word	RIG	No.	Word	RIG
1	paper	6.3×10^{-2}	51	call	9×10^{-3}
2	propos	4.8×10^{-2}	52	suggest	8.9×10^{-3}
3	softwar	3.8×10^{-2}	53	associ	8.8×10^{-3}
4	user	3.4×10^{-2}	54	decreas	8.7×10^{-3}
5	algorithm	3.3×10^{-2}	55	requir	8.7×10^{-3}
6	comput	2.6×10^{-2}	56	network	8.5×10^{-3}
7	approach	2.6×10^{-2}	57	secur	8.4×10^{-3}
8	implement	2.2×10^{-2}	58	speci	8.3×10^{-3}
9	problem	2×10^{-2}	59	engin	8.2×10^{-3}
10	base	2×10^{-2}	60	water	8.2×10^{-3}
11	execut	1.9×10^{-2}	61	challeng	8.2×10^{-3}
12	semant	1.9×10^{-2}	62	cloud	8.1×10^{-3}
13	conclus	1.8×10^{-2}	63	verif	8.1×10^{-3}
14	languag	1.7×10^{-2}	64	scalabl	8×10^{-3}
15	automat	1.7×10^{-2}	65	logic	7.9×10^{-3}
16	studi	1.7×10^{-2}	66	art	7.9×10^{-3}
17	patient	1.6×10^{-2}	67	gene	7.8×10^{-3}
18	system	1.6×10^{-2}	68	induc	7.8×10^{-3}
19	applic	1.5×10^{-2}	69	effici	7.7×10^{-3}
20	framework	1.5×10^{-2}	70	materi	7.7×10^{-3}
21	cell	1.5×10^{-2}	71	hardwar	7.7×10^{-3}
22	treatment	1.4×10^{-2}	72	oxid	7.6×10^{-3}
23	architectur	1.3×10^{-2}	73	visual	7.5×10^{-3}
24	code	1.3×10^{-2}	74	autom	7.5×10^{-3}
25	real	1.3×10^{-2}	75	overhead	7.5×10^{-3}
26	can	1.3×10^{-2}	76	observ	7.3×10^{-3}
27	program	1.2×10^{-2}	77	model	7.3×10^{-3}
28	temperatur	1.2×10^{-2}	78	signific	7.2×10^{-3}
29	task	1.2×10^{-2}	79	virtual	7.1×10^{-3}
30	introduc	1.1×10^{-2}	80	present	7.1×10^{-3}
31	web	1.1×10^{-2}	81	handl	7×10^{-3}
32	formal	1.1×10^{-2}	82	run	6.9×10^{-3}
33	exist	1.1×10^{-2}	83	represent	6.8×10^{-3}
34	techniqu	1.1×10^{-2}	84	enabl	6.8×10^{-3}
35	clinic	1×10^{-2}	85	examin	6.7×10^{-3}
36	abstract	1×10^{-2}	86	reaction	6.6×10^{-3}
37	protein	1×10^{-2}	87	indic	6.6×10^{-3}
38	tool	1×10^{-2}	88	background	6.5×10^{-3}
39	servic	9.9×10^{-3}	89	make	6.5×10^{-3}
40	concentr	9.6×10^{-3}	90	support	6.4×10^{-3}
41	investig	9.6×10^{-3}	91	server	6.4×10^{-3}
42	video	9.5×10^{-3}	92	group	6.2×10^{-3}
43	acid	9.5×10^{-3}	93	platform	6.2×10^{-3}
44	set	9.4×10^{-3}	94	higher	6.1×10^{-3}
45	graph	9.3×10^{-3}	95	domain	6×10^{-3}
46	age	9.3×10^{-3}	96	provid	6×10^{-3}
47	inform	9.3×10^{-3}	97	featur	5.9×10^{-3}
48	diseas	9.2×10^{-3}	98	activ	5.9×10^{-3}
49	queri	9×10^{-3}	99	metric	5.9×10^{-3}
50	found	9×10^{-3}	100	increas	5.9×10^{-3}

TABLE D.48. The list of the top 100 words in the category Computer Science, Theory and Methods with RIGs

No.	Word	RIG	No.	Word	RIG
1	paper	9.8×10^{-2}	51	dataset	9.7×10^{-3}
2	propos	6.7×10^{-2}	52	surfac	9.7×10^{-3}
3	algorithm	5.7×10^{-2}	53	diseas	9.6×10^{-3}
4	user	3.4×10^{-2}	54	introduc	9.6×10^{-3}
5	comput	3.3×10^{-2}	55	queri	9.5×10^{-3}
6	problem	3×10^{-2}	56	indic	9.5×10^{-3}
7	conclus	2.7×10^{-2}	57	web	9.3×10^{-3}
8	network	2.5×10^{-2}	58	materi	9.3×10^{-3}
9	studi	2.5×10^{-2}	59	solv	9×10^{-3}
10	base	2.2×10^{-2}	60	effici	8.9×10^{-3}
11	patient	1.9×10^{-2}	61	gene	8.9×10^{-3}
12	secur	1.8×10^{-2}	62	art	8.8×10^{-3}
13	treatment	1.7×10^{-2}	63	attack	8.8×10^{-3}
14	approach	1.6×10^{-2}	64	water	8.8×10^{-3}
15	implement	1.6×10^{-2}	65	activ	8.8×10^{-3}
16	cell	1.6×10^{-2}	66	set	8.7×10^{-3}
17	cloud	1.5×10^{-2}	67	scalabl	8.7×10^{-3}
18	inform	1.5×10^{-2}	68	reaction	8.6×10^{-3}
19	suggest	1.3×10^{-2}	69	wireless	8.6×10^{-3}
20	system	1.3×10^{-2}	70	report	8.5×10^{-3}
21	real	1.3×10^{-2}	71	machin	8.5×10^{-3}
22	servic	1.3×10^{-2}	72	examin	8.3×10^{-3}
23	protein	1.3×10^{-2}	73	oxid	8.2×10^{-3}
24	found	1.3×10^{-2}	74	optim	8.2×10^{-3}
25	temperatur	1.2×10^{-2}	75	background	8.2×10^{-3}
26	applic	1.2×10^{-2}	76	outperform	8.1×10^{-3}
27	associ	1.2×10^{-2}	77	call	7.9×10^{-3}
28	communic	1.2×10^{-2}	78	group	7.9×10^{-3}
29	graph	1.2×10^{-2}	79	video	7.8×10^{-3}
30	acid	1.2×10^{-2}	80	server	7.8×10^{-3}
31	concentr	1.2×10^{-2}	81	higher	7.7×10^{-3}
32	execut	1.1×10^{-2}	82	dure	7.5×10^{-3}
33	signific	1.1×10^{-2}	83	hardwar	7.4×10^{-3}
34	age	1.1×10^{-2}	84	internet	7.4×10^{-3}
35	semant	1.1×10^{-2}	85	assess	7.3×10^{-3}
36	observ	1.1×10^{-2}	86	molecular	7.3×10^{-3}
37	clinic	1.1×10^{-2}	87	overhead	7.3×10^{-3}
38	architectur	1.1×10^{-2}	88	increas	7.3×10^{-3}
39	softwar	1.1×10^{-2}	89	code	7.3×10^{-3}
40	scheme	1.1×10^{-2}	90	techniqu	7.2×10^{-3}
41	task	1.1×10^{-2}	91	reveal	7.2×10^{-3}
42	investig	1.1×10^{-2}	92	inhibit	7.2×10^{-3}
43	can	1×10^{-2}	93	featur	7.2×10^{-3}
44	induc	1×10^{-2}	94	perform	7.1×10^{-3}
45	learn	1×10^{-2}	95	exist	6.9×10^{-3}
46	node	1×10^{-2}	96	mass	6.7×10^{-3}
47	automat	1×10^{-2}	97	platform	6.7×10^{-3}
48	framework	9.9×10^{-3}	98	messag	6.7×10^{-3}
49	speci	9.9×10^{-3}	99	prepar	6.7×10^{-3}
50	decreas	9.8×10^{-3}	100	resourc	6.7×10^{-3}

Table D.49. The list of the top 100 words in the category Construction and Building Technology with RIGs

No.	Word	RIG
1	concret	1.2×10^{-1}
2	build	5.8×10^{-2}
3	cement	4.4×10^{-2}
4	strength	3.4×10^{-2}
5	reinforc	3.2×10^{-2}
6	steel	3×10^{-2}
7	load	3×10^{-2}
8	asphalt	2.7×10^{-2}
9	paper	2.6×10^{-2}
10	mortar	2.3×10^{-2}
11	compress	2.3×10^{-2}
12	pavement	2.2×10^{-2}
13	patient	2×10^{-2}
14	construct	1.9×10^{-2}
15	test	1.7×10^{-2}
16	crack	1.7×10^{-2}
17	indoor	1.6×10^{-2}
18	binder	1.5×10^{-2}
19	seismic	1.5×10^{-2}
20	shear	1.5×10^{-2}
21	specimen	1.5×10^{-2}
22	flexur	1.4×10^{-2}
23	conclus	1.4×10^{-2}
24	air	1.4×10^{-2}
25	heat	1.4×10^{-2}
26	cell	1.4×10^{-2}
27	stiff	1.3×10^{-2}
28	design	1.2×10^{-2}
29	durabl	1.2×10^{-2}
30	comfort	1.2×10^{-2}
31	clinic	1.2×10^{-2}
32	energi	1.1×10^{-2}
33	materi	1.1×10^{-2}
34	beam	1.1×10^{-2}
35	protein	1.1×10^{-2}
36	ventil	1.1×10^{-2}
37	ductil	1×10^{-2}
38	ash	1×10^{-2}
39	diseas	9.9×10^{-3}
40	thermal	9.8×10^{-3}
41	earthquak	9.7×10^{-3}
42	group	9.7×10^{-3}
43	element	9.6×10^{-3}
44	bridg	9.5×10^{-3}
45	experiment	9.4×10^{-3}
46	structur	8.6×10^{-3}
47	associ	8.5×10^{-3}
48	gene	8.4×10^{-3}
49	residenti	8.4×10^{-3}
50	tensil	8.4×10^{-3}

No.	Word	RIG
51	save	8.2×10^{-3}
52	hydrat	8×10^{-3}
53	engin	8×10^{-3}
54	column	7.9×10^{-3}
55	background	7.7×10^{-3}
56	finit	7.6×10^{-3}
57	floor	7.6×10^{-3}
58	treatment	7.5×10^{-3}
59	cool	7.5×10^{-3}
60	aggreg	7.4×10^{-3}
61	displac	7.2×10^{-3}
62	slab	7.1×10^{-3}
63	simul	7×10^{-3}
64	bend	6.9×10^{-3}
65	cure	6.9×10^{-3}
66	hous	6.7×10^{-3}
67	failur	6.7×10^{-3}
68	wall	6.6×10^{-3}
69	perform	6.6×10^{-3}
70	modulus	6.6×10^{-3}
71	model	6.6×10^{-3}
72	capac	6.5×10^{-3}
73	deform	6.5×10^{-3}
74	activ	6.5×10^{-3}
75	harden	6.2×10^{-3}
76	numer	6.2×10^{-3}
77	damag	6.2×10^{-3}
78	suggest	6.1×10^{-3}
79	tissu	5.9×10^{-3}
80	cancer	5.7×10^{-3}
81	instal	5.7×10^{-3}
82	mixtur	5.6×10^{-3}
83	humid	5.6×10^{-3}
84	express	5.6×10^{-3}
85	frame	5.5×10^{-3}
86	carri	5.4×10^{-3}
87	fli	5.4×10^{-3}
88	research	5.4×10^{-3}
89	elast	5.4×10^{-3}
90	molecular	5.4×10^{-3}
91	therapi	5.3×10^{-3}
92	speci	5.3×10^{-3}
93	report	5.3×10^{-3}
94	consumpt	5.2×10^{-3}
95	popul	5.1×10^{-3}
96	deflect	5.1×10^{-3}
97	vibrat	5.1×10^{-3}
98	acid	5×10^{-3}
99	water	5×10^{-3}
100	blood	5×10^{-3}

TABLE D.50. The list of the top 100 words in the category Criminology and Penology with RIGs

No.	Word	RIG	No.	Word	RIG
1	offend	1.7×10^{-1}	51	survey	1.2×10^{-2}
2	crime	1.4×10^{-1}	52	selfreport	1.2×10^{-2}
3	crimin	1.3×10^{-1}	53	temperatur	1.2×10^{-2}
4	violenc	9.5×10^{-2}	54	nation	1.1×10^{-2}
5	victim	8.9×10^{-2}	55	aggress	1.1×10^{-2}
6	polic	8.2×10^{-2}	56	punish	1.1×10^{-2}
7	justic	5.1×10^{-2}	57	juvenil	1.1×10^{-2}
8	research	4.4×10^{-2}	58	child	1.1×10^{-2}
9	prison	4.3×10^{-2}	59	sex	1.1×10^{-2}
10	sexual	4.2×10^{-2}	60	surfac	1.1×10^{-2}
11	violent	4.2×10^{-2}	61	whether	1.1×10^{-2}
12	examin	4.1×10^{-2}	62	particip	1×10^{-2}
13	abus	3.3×10^{-2}	63	explor	1×10^{-2}
14	perpetr	3.1×10^{-2}	64	simul	1×10^{-2}
15	implic	3.1×10^{-2}	65	draw	9.8×10^{-3}
16	delinqu	3×10^{-2}	66	peer	9.7×10^{-3}
17	articl	2.9×10^{-2}	67	femal	9.4×10^{-3}
18	legal	2.6×10^{-2}	68	public	9.1×10^{-3}
19	assault	2.6×10^{-2}	69	induc	9.1×10^{-3}
20	find	2.5×10^{-2}	70	paramet	9.1×10^{-3}
21	convict	2.5×10^{-2}	71	sanction	9×10^{-3}
22	social	2.4×10^{-2}	72	intervent	9×10^{-3}
23	incarcer	2.4×10^{-2}	73	energi	9×10^{-3}
24	court	2.2×10^{-2}	74	legisl	8.9×10^{-3}
25	youth	2.1×10^{-2}	75	emot	8.8×10^{-3}
26	antisoci	2×10^{-2}	76	murder	8.7×10^{-3}
27	law	2×10^{-2}	77	empir	8.7×10^{-3}
28	relationship	1.9×10^{-2}	78	percept	8.4×10^{-3}
29	offic	1.8×10^{-2}	79	question	8.4×10^{-3}
30	forens	1.7×10^{-2}	80	suggest	8.3×10^{-3}
31	interview	1.7×10^{-2}	81	among	8.2×10^{-3}
32	intim	1.7×10^{-2}	82	protein	8×10^{-3}
33	polici	1.6×10^{-2}	83	effici	8×10^{-3}
34	discuss	1.6×10^{-2}	84	perceiv	7.9×10^{-3}
35	gender	1.6×10^{-2}	85	interperson	7.9×10^{-3}
36	commit	1.6×10^{-2}	86	materi	7.9×10^{-3}
37	risk	1.6×10^{-2}	87	offici	7.9×10^{-3}
38	ipv	1.5×10^{-2}	88	obtain	7.7×10^{-3}
39	cell	1.5×10^{-2}	89	literatur	7.6×10^{-3}
40	adolesc	1.5×10^{-2}	90	wit	7.6×10^{-3}
41	partner	1.4×10^{-2}	91	across	7.6×10^{-3}
42	sentenc	1.4×10^{-2}	92	futur	7.6×10^{-3}
43	male	1.3×10^{-2}	93	argu	7.6×10^{-3}
44	perform	1.3×10^{-2}	94	behavior	7.6×10^{-3}
45	mental	1.3×10^{-2}	95	practic	7.1×10^{-3}
46	enforc	1.3×10^{-2}	96	method	7.1×10^{-3}
47	communiti	1.2×10^{-2}	97	context	7.1×10^{-3}
48	person	1.2×10^{-2}	98	men	7.1×10^{-3}
49	psycholog	1.2×10^{-2}	99	studi	7.1×10^{-3}
50	women	1.2×10^{-2}	100	show	6.9×10^{-3}

Table D.51. The list of the top 100 words in the category Critical Care Medicine with RIGs

No.	Word	RIG
1	patient	1.6×10^{-1}
2	conclus	1.3×10^{-1}
3	icu	9.9×10^{-2}
4	injuri	9.4×10^{-2}
5	hospit	8.4×10^{-2}
6	care	8×10^{-2}
7	mortal	6.7×10^{-2}
8	trauma	6.7×10^{-2}
9	ventil	6.5×10^{-2}
10	outcom	6.4×10^{-2}
11	admiss	5.6×10^{-2}
12	resuscit	5.5×10^{-2}
13	prospect	4.9×10^{-2}
14	score	4.6×10^{-2}
15	admit	4.4×10^{-2}
16	acut	4.4×10^{-2}
17	ill	3.9×10^{-2}
18	clinic	3.9×10^{-2}
19	retrospect	3.8×10^{-2}
20	method	3.8×10^{-2}
21	cardiac	3.7×10^{-2}
22	intervent	3.7×10^{-2}
23	associ	3.6×10^{-2}
24	background	3.6×10^{-2}
25	sepsi	3.5×10^{-2}
26	hour	3.4×10^{-2}
27	respiratori	3.3×10^{-2}
28	pulmonari	3.2×10^{-2}
29	cohort	3.2×10^{-2}
30	day	3.2×10^{-2}
31	unit	3×10^{-2}
32	result	3×10^{-2}
33	stay	2.9×10^{-2}
34	traumat	2.9×10^{-2}
35	receiv	2.8×10^{-2}
36	introduct	2.7×10^{-2}
37	arrest	2.7×10^{-2}
38	septic	2.6×10^{-2}
39	paper	2.6×10^{-2}
40	lung	2.4×10^{-2}
41	medic	2.4×10^{-2}
42	age	2.4×10^{-2}
43	blood	2.3×10^{-2}
44	median	2.3×10^{-2}
45	object	2.3×10^{-2}
46	intens	2.2×10^{-2}
47	risk	2.2×10^{-2}
48	year	2.2×10^{-2}
49	signific	2.2×10^{-2}
50	assess	2.2×10^{-2}

No.	Word	RIG
51	studi	2.1×10^{-2}
52	injur	2×10^{-2}
53	sever	2×10^{-2}
54	shock	2×10^{-2}
55	death	2×10^{-2}
56	odd	2×10^{-2}
57	inhospit	2×10^{-2}
58	adult	2×10^{-2}
59	surviv	1.9×10^{-2}
60	critic	1.8×10^{-2}
61	includ	1.7×10^{-2}
62	measur	1.7×10^{-2}
63	airway	1.7×10^{-2}
64	surgic	1.7×10^{-2}
65	group	1.7×10^{-2}
66	hemodynam	1.7×10^{-2}
67	neurolog	1.7×10^{-2}
68	hemorrhag	1.7×10^{-2}
69	rational	1.6×10^{-2}
70	survivor	1.6×10^{-2}
71	arteri	1.6×10^{-2}
72	intub	1.6×10^{-2}
73	cardiopulmonari	1.6×10^{-2}
74	logist	1.6×10^{-2}
75	chest	1.6×10^{-2}
76	follow	1.6×10^{-2}
77	discharg	1.5×10^{-2}
78	burn	1.5×10^{-2}
79	trial	1.5×10^{-2}
80	regress	1.5×10^{-2}
81	therapi	1.5×10^{-2}
82	pediatr	1.5×10^{-2}
83	complic	1.5×10^{-2}
84	total	1.5×10^{-2}
85	multivari	1.4×10^{-2}
86	compar	1.4×10^{-2}
87	surgeri	1.4×10^{-2}
88	interquartil	1.4×10^{-2}
89	consecut	1.4×10^{-2}
90	month	1.4×10^{-2}
91	pneumonia	1.3×10^{-2}
92	dure	1.3×10^{-2}
93	adjust	1.3×10^{-2}
94	review	1.3×10^{-2}
95	failur	1.3×10^{-2}
96	none	1.3×10^{-2}
97	enrol	1.3×10^{-2}
98	confid	1.3×10^{-2}
99	tertiari	1.3×10^{-2}
100	interv	1.3×10^{-2}

TABLE D.52. The list of the top 100 words in the category Crystallography with RIGs

No.	Word	RIG	No.	Word	RIG
1	crystal	1.9×10^{-1}	51	spectroscopi	1.1×10^{-2}
2	diffract	7.8×10^{-2}	52	assess	1.1×10^{-2}
3	ray	7.5×10^{-2}	53	pack	1.1×10^{-2}
4	angstrom	7.1×10^{-2}	54	solid	1.1×10^{-2}
5	structur	6.4×10^{-2}	55	cation	1.1×10^{-2}
6	dot	5.2×10^{-2}	56	bridg	1.1×10^{-2}
7	bond	4.1×10^{-2}	57	degre	1×10^{-2}
8	ligand	4×10^{-2}	58	octahedr	1×10^{-2}
9	h2o	3.4×10^{-2}	59	unit	1×10^{-2}
10	synthes	3.3×10^{-2}	60	dimer	1×10^{-2}
11	monoclin	3.2×10^{-2}	61	stack	1×10^{-2}
12	molecul	3.2×10^{-2}	62	test	1×10^{-2}
13	center	3.1×10^{-2}	63	background	1×10^{-2}
14	compound	2.8×10^{-2}	64	molecular	1×10^{-2}
15	atom	2.7×10^{-2}	65	electron	1×10^{-2}
16	hydrogen	2.7×10^{-2}	66	polym	1×10^{-2}
17	coordin	2.6×10^{-2}	67	signific	1×10^{-2}
18	supramolecular	2.2×10^{-2}	68	luminesc	9.7×10^{-3}
19	crystallograph	2.1×10^{-2}	69	clinic	9.5×10^{-3}
20	grown	2.1×10^{-2}	70	nmr	9.5×10^{-3}
21	conclus	1.9×10^{-2}	71	layer	9.3×10^{-3}
22	patient	1.8×10^{-2}	72	resolut	9.3×10^{-3}
23	bis	1.8×10^{-2}	73	dimension	9.3×10^{-3}
24	singl	1.8×10^{-2}	74	age	9.3×10^{-3}
25	intermolecular	1.7×10^{-2}	75	interact	9.2×10^{-3}
26	crystallin	1.6×10^{-2}	76	aim	9.2×10^{-3}
27	paper	1.6×10^{-2}	77	reaction	9.2×10^{-3}
28	anion	1.6×10^{-2}	78	properti	8.9×10^{-3}
29	form	1.4×10^{-2}	79	growth	8.9×10^{-3}
30	hydrotherm	1.4×10^{-2}	80	risk	8.9×10^{-3}
31	complex	1.4×10^{-2}	81	acid	8.7×10^{-3}
32	character	1.4×10^{-2}	82	associ	8.7×10^{-3}
33	solvent	1.4×10^{-2}	83	evalu	8.5×10^{-3}
34	temperatur	1.3×10^{-2}	84	develop	8.5×10^{-3}
35	epitaxi	1.3×10^{-2}	85	level	8.4×10^{-3}
36	powder	1.3×10^{-2}	86	exhibit	8.4×10^{-3}
37	result	1.3×10^{-2}	87	ring	8.2×10^{-3}
38	sic	1.3×10^{-2}	88	format	8.1×10^{-3}
39	orthorhomb	1.2×10^{-2}	89	research	8×10^{-3}
40	titl	1.2×10^{-2}	90	phase	8×10^{-3}
41	year	1.2×10^{-2}	91	nucleat	7.9×10^{-3}
42	crystallographi	1.2×10^{-2}	92	conform	7.9×10^{-3}
43	ion	1.2×10^{-2}	93	carboxyl	7.9×10^{-3}
44	beta	1.2×10^{-2}	94	asymmetr	7.7×10^{-3}
45	metal	1.1×10^{-2}	95	lattic	7.4×10^{-3}
46	object	1.1×10^{-2}	96	find	7.3×10^{-3}
47	pyridin	1.1×10^{-2}	97	problem	7.1×10^{-3}
48	photoluminesc	1.1×10^{-2}	98	dft	7×10^{-3}
49	space	1.1×10^{-2}	99	diseas	6.9×10^{-3}
50	chain	1.1×10^{-2}	100	microscopi	6.9×10^{-3}

Table D.53. The list of the top 100 words in the category Cultural Studies with RIGs

No.	Word	RIG	No.	Word	RIG
1	articl	9.7×10^{-2}	51	patient	1.4×10^{-2}
2	polit	8.6×10^{-2}	52	literari	1.4×10^{-2}
3	cultur	8.1×10^{-2}	53	obtain	1.4×10^{-2}
4	argu	7.1×10^{-2}	54	cinema	1.4×10^{-2}
5	essay	5.7×10^{-2}	55	everyday	1.3×10^{-2}
6	discours	4.8×10^{-2}	56	focus	1.3×10^{-2}
7	social	4.7×10^{-2}	57	struggl	1.3×10^{-2}
8	contemporari	4.3×10^{-2}	58	compar	1.3×10^{-2}
9	result	4.1×10^{-2}	59	seek	1.3×10^{-2}
10	narrat	3.4×10^{-2}	60	represent	1.3×10^{-2}
11	draw	3.1×10^{-2}	61	discurs	1.3×10^{-2}
12	media	2.7×10^{-2}	62	text	1.2×10^{-2}
13	way	2.5×10^{-2}	63	fiction	1.2×10^{-2}
14	method	2.4×10^{-2}	64	evalu	1.2×10^{-2}
15	ident	2.3×10^{-2}	65	high	1.2×10^{-2}
16	histor	2.3×10^{-2}	66	understand	1.2×10^{-2}
17	public	2.2×10^{-2}	67	emerg	1.2×10^{-2}
18	societi	2.2×10^{-2}	68	conclus	1.2×10^{-2}
19	context	2.2×10^{-2}	69	themselv	1.2×10^{-2}
20	nation	2.1×10^{-2}	70	activist	1.2×10^{-2}
21	centuri	2×10^{-2}	71	ethnograph	1.2×10^{-2}
22	imagin	1.9×10^{-2}	72	aesthet	1.2×10^{-2}
23	practic	1.9×10^{-2}	73	data	1.2×10^{-2}
24	notion	1.9×10^{-2}	74	read	1.1×10^{-2}
25	ideolog	1.8×10^{-2}	75	author	1.1×10^{-2}
26	explor	1.8×10^{-2}	76	popular	1.1×10^{-2}
27	artist	1.7×10^{-2}	77	figur	1.1×10^{-2}
28	critiqu	1.7×10^{-2}	78	cell	1.1×10^{-2}
29	audienc	1.7×10^{-2}	79	examin	1.1×10^{-2}
30	claim	1.7×10^{-2}	80	capit	1.1×10^{-2}
31	transnat	1.7×10^{-2}	81	coloni	1.1×10^{-2}
32	world	1.7×10^{-2}	82	contest	1.1×10^{-2}
33	neoliber	1.6×10^{-2}	83	determin	1.1×10^{-2}
34	look	1.6×10^{-2}	84	simul	1.1×10^{-2}
35	peopl	1.6×10^{-2}	85	filmmak	1.1×10^{-2}
36	engag	1.6×10^{-2}	86	test	1.1×10^{-2}
37	histori	1.6×10^{-2}	87	rate	1.1×10^{-2}
38	war	1.6×10^{-2}	88	concept	1.1×10^{-2}
39	stori	1.6×10^{-2}	89	improv	1.1×10^{-2}
40	postcoloni	1.5×10^{-2}	90	write	1×10^{-2}
41	articul	1.5×10^{-2}	91	decreas	1×10^{-2}
42	celebr	1.5×10^{-2}	92	particular	1×10^{-2}
43	televis	1.5×10^{-2}	93	think	1×10^{-2}
44	scholar	1.5×10^{-2}	94	model	1×10^{-2}
45	debat	1.5×10^{-2}	95	econom	1×10^{-2}
46	use	1.5×10^{-2}	96	reduc	1×10^{-2}
47	question	1.5×10^{-2}	97	rhetor	9.8×10^{-3}
48	effect	1.5×10^{-2}	98	modern	9.7×10^{-3}
49	creativ	1.5×10^{-2}	99	low	9.6×10^{-3}
50	measur	1.4×10^{-2}	100	citizenship	9.5×10^{-3}

TABLE D.54. The list of the top 100 words in the category Dance with RIGs

No.	Word	RIG
1	danc	6.6×10^{-1}
2	dancer	1.7×10^{-1}
3	choreograph	1.4×10^{-1}
4	festiv	1.2×10^{-1}
5	music	9.3×10^{-2}
6	ballet	8.7×10^{-2}
7	folk	7.9×10^{-2}
8	place	6.2×10^{-2}
9	contemporari	6.2×10^{-2}
10	choreographi	4.5×10^{-2}
11	cultur	4.4×10^{-2}
12	tradit	4.4×10^{-2}
13	creativ	3.9×10^{-2}
14	ethnograph	3.7×10^{-2}
15	result	3×10^{-2}
16	essay	3×10^{-2}
17	fieldwork	2.9×10^{-2}
18	celebr	2.9×10^{-2}
19	musician	2.8×10^{-2}
20	audienc	2.8×10^{-2}
21	genr	2.8×10^{-2}
22	villag	2.7×10^{-2}
23	communiti	2.6×10^{-2}
24	context	2.5×10^{-2}
25	ritual	2.5×10^{-2}
26	artist	2.5×10^{-2}
27	teacher	2.4×10^{-2}
28	creat	2.4×10^{-2}
29	postmodern	2.3×10^{-2}
30	aesthet	2.2×10^{-2}
31	effect	2.1×10^{-2}
32	movement	2.1×10^{-2}
33	art	2×10^{-2}
34	practic	2×10^{-2}
35	theatr	2×10^{-2}
36	form	1.9×10^{-2}
37	question	1.8×10^{-2}
38	social	1.8×10^{-2}
39	interview	1.8×10^{-2}
40	diaspora	1.7×10^{-2}
41	discours	1.7×10^{-2}
42	measur	1.7×10^{-2}
43	brought	1.7×10^{-2}
44	organis	1.7×10^{-2}
45	ident	1.7×10^{-2}
46	high	1.7×10^{-2}
47	discuss	1.6×10^{-2}
48	today	1.6×10^{-2}
49	held	1.6×10^{-2}
50	style	1.5×10^{-2}

No.	Word	RIG
51	societi	1.5×10^{-2}
52	increas	1.5×10^{-2}
53	citi	1.5×10^{-2}
54	represent	1.5×10^{-2}
55	argu	1.5×10^{-2}
56	show	1.5×10^{-2}
57	ethnographi	1.5×10^{-2}
58	participatori	1.5×10^{-2}
59	repertoir	1.5×10^{-2}
60	1930s	1.5×10^{-2}
61	educ	1.4×10^{-2}
62	event	1.4×10^{-2}
63	look	1.4×10^{-2}
64	articl	1.4×10^{-2}
65	test	1.4×10^{-2}
66	improv	1.3×10^{-2}
67	multicultur	1.3×10^{-2}
68	explor	1.3×10^{-2}
69	master	1.3×10^{-2}
70	obtain	1.3×10^{-2}
71	system	1.2×10^{-2}
72	realiti	1.2×10^{-2}
73	becam	1.2×10^{-2}
74	model	1.2×10^{-2}
75	student	1.2×10^{-2}
76	book	1.2×10^{-2}
77	low	1.2×10^{-2}
78	centuri	1.2×10^{-2}
79	negoti	1.2×10^{-2}
80	reduc	1.2×10^{-2}
81	patient	1.1×10^{-2}
82	entertain	1.1×10^{-2}
83	peopl	1.1×10^{-2}
84	public	1.1×10^{-2}
85	nineteenth	1.1×10^{-2}
86	inhabit	1.1×10^{-2}
87	speak	1.1×10^{-2}
88	focus	1.1×10^{-2}
89	cell	1.1×10^{-2}
90	modern	1.1×10^{-2}
91	exil	1.1×10^{-2}
92	instructor	1.1×10^{-2}
93	way	1.1×10^{-2}
94	franc	1×10^{-2}
95	someth	1×10^{-2}
96	sixteenth	1×10^{-2}
97	tourist	1×10^{-2}
98	jew	9.9×10^{-3}
99	turkish	9.7×10^{-3}
100	spectacl	9.7×10^{-3}

Table D.55. The list of the top 100 words in the category Demography with RIGs

No.	Word	RIG
1	migrant	9.2×10^{-2}
2	immigr	8.4×10^{-2}
3	countri	6.6×10^{-2}
4	migrat	5.9×10^{-2}
5	marriag	4.7×10^{-2}
6	survey	4.1×10^{-2}
7	women	3.9×10^{-2}
8	econom	3.9×10^{-2}
9	social	3.8×10^{-2}
10	demograph	3.8×10^{-2}
11	fertil	3.7×10^{-2}
12	famili	3.7×10^{-2}
13	nation	3.3×10^{-2}
14	marri	3.3×10^{-2}
15	polici	3.2×10^{-2}
16	articl	3.2×10^{-2}
17	educ	3.1×10^{-2}
18	birth	2.9×10^{-2}
19	household	2.9×10^{-2}
20	children	2.7×10^{-2}
21	examin	2.5×10^{-2}
22	labour	2.4×10^{-2}
23	census	2.2×10^{-2}
24	popul	2.2×10^{-2}
25	data	2.2×10^{-2}
26	find	2.1×10^{-2}
27	context	2.1×10^{-2}
28	contracept	2.1×10^{-2}
29	age	1.9×10^{-2}
30	union	1.9×10^{-2}
31	argu	1.9×10^{-2}
32	born	1.9×10^{-2}
33	transnat	1.9×10^{-2}
34	refuge	1.8×10^{-2}
35	among	1.8×10^{-2}
36	incom	1.7×10^{-2}
37	european	1.7×10^{-2}
38	socioeconom	1.7×10^{-2}
39	child	1.6×10^{-2}
40	marit	1.6×10^{-2}
41	rural	1.6×10^{-2}
42	men	1.5×10^{-2}
43	gender	1.5×10^{-2}
44	live	1.5×10^{-2}
45	draw	1.4×10^{-2}
46	polit	1.4×10^{-2}
47	ethnic	1.4×10^{-2}
48	socio	1.3×10^{-2}
49	declin	1.3×10^{-2}
50	cell	1.3×10^{-2}

No.	Word	RIG
51	capit	1.3×10^{-2}
52	mother	1.3×10^{-2}
53	urban	1.2×10^{-2}
54	parent	1.2×10^{-2}
55	labor	1.2×10^{-2}
56	emigr	1.2×10^{-2}
57	health	1.2×10^{-2}
58	across	1.2×10^{-2}
59	research	1.2×10^{-2}
60	resid	1.2×10^{-2}
61	life	1.2×10^{-2}
62	relationship	1.2×10^{-2}
63	patient	1.1×10^{-2}
64	longitudin	1.1×10^{-2}
65	intern	1.1×10^{-2}
66	individu	1.1×10^{-2}
67	peopl	1.1×10^{-2}
68	temperatur	1.1×10^{-2}
69	young	1×10^{-2}
70	legal	1×10^{-2}
71	societi	1×10^{-2}
72	empir	1×10^{-2}
73	status	1×10^{-2}
74	perform	9.9×10^{-3}
75	subsaharan	9.9×10^{-3}
76	mortal	9.9×10^{-3}
77	europ	9.6×10^{-3}
78	abroad	9.3×10^{-3}
79	market	9.3×10^{-3}
80	properti	9.1×10^{-3}
81	focus	9.1×10^{-3}
82	foreign	9×10^{-3}
83	surfac	8.7×10^{-3}
84	destin	8.6×10^{-3}
85	less	8.5×10^{-3}
86	partner	8.4×10^{-3}
87	decad	8.4×10^{-3}
88	experiment	8.3×10^{-3}
89	husband	8.3×10^{-3}
90	worker	8.1×10^{-3}
91	father	8×10^{-3}
92	detect	7.8×10^{-3}
93	spous	7.8×10^{-3}
94	africa	7.8×10^{-3}
95	electron	7.7×10^{-3}
96	sex	7.7×10^{-3}
97	interview	7.6×10^{-3}
98	propos	7.5×10^{-3}
99	energi	7.5×10^{-3}
100	regress	7.5×10^{-3}

TABLE D.56. The list of the top 100 words in the category Dentistry, Oral Surgery and Medicine with RIGs

No.	Word	RIG
1	dental	1.9×10^{-1}
2	teeth	1.1×10^{-1}
3	conclus	1×10^{-1}
4	tooth	8×10^{-2}
5	periodont	7.4×10^{-2}
6	maxillari	7.2×10^{-2}
7	mandibular	7.2×10^{-2}
8	oral	6.5×10^{-2}
9	implant	5.3×10^{-2}
10	bone	5.3×10^{-2}
11	materi	4.8×10^{-2}
12	patient	4.7×10^{-2}
13	dentin	4.6×10^{-2}
14	mandibl	4×10^{-2}
15	signific	4×10^{-2}
16	clinic	3.9×10^{-2}
17	method	3.7×10^{-2}
18	molar	3.6×10^{-2}
19	group	3.4×10^{-2}
20	purpos	3.4×10^{-2}
21	resin	3.2×10^{-2}
22	aim	3.2×10^{-2}
23	statist	3.1×10^{-2}
24	object	3.1×10^{-2}
25	evalu	3.1×10^{-2}
26	enamel	2.8×10^{-2}
27	radiograph	2.8×10^{-2}
28	restor	2.7×10^{-2}
29	treatment	2.5×10^{-2}
30	anova	2.4×10^{-2}
31	canal	2.4×10^{-2}
32	result	2.4×10^{-2}
33	palat	2.3×10^{-2}
34	alveolar	2.2×10^{-2}
35	paper	2.1×10^{-2}
36	studi	2.1×10^{-2}
37	crown	2.1×10^{-2}
38	specimen	2×10^{-2}
39	occlus	1.8×10^{-2}
40	mouth	1.7×10^{-2}
41	surgeri	1.6×10^{-2}
42	heal	1.6×10^{-2}
43	surgic	1.6×10^{-2}
44	root	1.6×10^{-2}
45	tissu	1.6×10^{-2}
46	test	1.5×10^{-2}
47	facial	1.5×10^{-2}
48	divid	1.5×10^{-2}
49	assess	1.4×10^{-2}
50	treat	1.4×10^{-2}

No.	Word	RIG
51	placement	1.4×10^{-2}
52	propos	1.4×10^{-2}
53	pulp	1.4×10^{-2}
54	year	1.4×10^{-2}
55	apic	1.3×10^{-2}
56	follow	1.3×10^{-2}
57	month	1.3×10^{-2}
58	anterior	1.2×10^{-2}
59	plaqu	1.2×10^{-2}
60	titanium	1.2×10^{-2}
61	cone	1.2×10^{-2}
62	cement	1.2×10^{-2}
63	tomographi	1.1×10^{-2}
64	age	1.1×10^{-2}
65	prosthesi	1.1×10^{-2}
66	fractur	1.1×10^{-2}
67	posterior	1×10^{-2}
68	introduct	1×10^{-2}
69	differ	9.9×10^{-3}
70	mean	9.9×10^{-3}
71	postop	9.8×10^{-3}
72	lesion	9.5×10^{-3}
73	flap	9.2×10^{-3}
74	compar	9.2×10^{-3}
75	adhes	9×10^{-3}
76	subject	9×10^{-3}
77	neck	8.7×10^{-3}
78	soft	8.7×10^{-3}
79	histolog	8.4×10^{-3}
80	squamous	8.4×10^{-3}
81	ceram	8.3×10^{-3}
82	fluorid	8.3×10^{-3}
83	random	8.2×10^{-3}
84	graft	7.9×10^{-3}
85	place	7.8×10^{-3}
86	record	7.8×10^{-3}
87	screw	7.8×10^{-3}
88	temperatur	7.7×10^{-3}
89	examin	7.7×10^{-3}
90	statement	7.5×10^{-3}
91	bleed	7.3×10^{-3}
92	procedur	7.3×10^{-3}
93	effici	7.1×10^{-3}
94	remov	7.1×10^{-3}
95	algorithm	7×10^{-3}
96	fill	7×10^{-3}
97	thirti	6.9×10^{-3}
98	old	6.8×10^{-3}
99	befor	6.8×10^{-3}
100	scan	6.8×10^{-3}

TABLE D.57. The list of the top 100 words in the category Dermatology with RIGs

No.	Word	RIG	No.	Word	RIG
1	skin	1.8×10^{-1}	51	advers	1×10^{-2}
2	patient	9.5×10^{-2}	52	week	1×10^{-2}
3	cutan	8.7×10^{-2}	53	model	1×10^{-2}
4	dermatolog	6.4×10^{-2}	54	simul	1×10^{-2}
5	lesion	5.5×10^{-2}	55	age	9.5×10^{-3}
6	psoriasi	5.4×10^{-2}	56	disord	9.5×10^{-3}
7	clinic	5.2×10^{-2}	57	may	9.4×10^{-3}
8	background	5×10^{-2}	58	month	9.3×10^{-3}
9	conclus	4.8×10^{-2}	59	review	9.2×10^{-3}
10	dermat	4.3×10^{-2}	60	tissu	9.2×10^{-3}
11	treatment	4×10^{-2}	61	base	8.9×10^{-3}
12	object	3.9×10^{-2}	62	associ	8.7×10^{-3}
13	melanoma	3.8×10^{-2}	63	temperatur	8.6×10^{-3}
14	wound	3×10^{-2}	64	carcinoma	8.6×10^{-3}
15	diseas	3×10^{-2}	65	pathogenesi	8.5×10^{-3}
16	keratinocyt	2.8×10^{-2}	66	corticosteroid	8.5×10^{-3}
17	dermal	2.7×10^{-2}	67	method	8.4×10^{-3}
18	epiderm	2.6×10^{-2}	68	sever	7.9×10^{-3}
19	paper	2.6×10^{-2}	69	structur	7.9×10^{-3}
20	therapi	2.4×10^{-2}	70	squamous	7.8×10^{-3}
21	treat	2.3×10^{-2}	71	manifest	7.7×10^{-3}
22	biopsi	2.3×10^{-2}	72	allerg	7.7×10^{-3}
23	ulcer	2.2×10^{-2}	73	erupt	7.6×10^{-3}
24	histopatholog	2.2×10^{-2}	74	follicl	7.6×10^{-3}
25	heal	2.1×10^{-2}	75	therapeut	7.5×10^{-3}
26	case	2.1×10^{-2}	76	frequent	7.4×10^{-3}
27	diagnosi	2.1×10^{-2}	77	subcutan	7.2×10^{-3}
28	hair	2.1×10^{-2}	78	infect	7×10^{-3}
29	burn	2×10^{-2}	79	autoimmun	6.9×10^{-3}
30	scar	1.9×10^{-2}	80	infiltr	6.8×10^{-3}
31	atop	1.9×10^{-2}	81	effici	6.7×10^{-3}
32	topic	1.8×10^{-2}	82	collagen	6.6×10^{-3}
33	rare	1.8×10^{-2}	83	medic	6.6×10^{-3}
34	report	1.8×10^{-2}	84	tumor	6.6×10^{-3}
35	efficaci	1.7×10^{-2}	85	recurr	6.6×10^{-3}
36	histolog	1.4×10^{-2}	86	evalu	6.6×10^{-3}
37	old	1.4×10^{-2}	87	energi	6.5×10^{-3}
38	year	1.4×10^{-2}	88	physician	6.5×10^{-3}
39	inflammatori	1.4×10^{-2}	89	process	6.3×10^{-3}
40	chronic	1.4×10^{-2}	90	immunohistochem	6.2×10^{-3}
41	common	1.4×10^{-2}	91	experiment	6.2×10^{-3}
42	retrospect	1.3×10^{-2}	92	score	6.2×10^{-3}
43	pigment	1.2×10^{-2}	93	oral	6.2×10^{-3}
44	facial	1.2×10^{-2}	94	syndrom	6.2×10^{-3}
45	cell	1.2×10^{-2}	95	assess	6.2×10^{-3}
46	plaqu	1.1×10^{-2}	96	includ	6.1×10^{-3}
47	propos	1.1×10^{-2}	97	basal	6.1×10^{-3}
48	malign	1.1×10^{-2}	98	woman	6×10^{-3}
49	excis	1.1×10^{-2}	99	neoplasm	6×10^{-3}
50	diagnos	1.1×10^{-2}	100	signific	6×10^{-3}

TABLE D.58. The list of the top 100 words in the category Developmental Biology with RIGs

No.	Word	RIG	No.	Word	RIG
1	embryo	8.7×10^{-2}	51	neural	1.4×10^{-2}
2	cell	8.7×10^{-2}	52	mrna	1.4×10^{-2}
3	express	8×10^{-2}	53	regulatori	1.4×10^{-2}
4	embryon	6.6×10^{-2}	54	suggest	1.4×10^{-2}
5	regul	6.3×10^{-2}	55	epithelium	1.3×10^{-2}
6	gene	5.6×10^{-2}	56	activ	1.3×10^{-2}
7	development	5.1×10^{-2}	57	pattern	1.3×10^{-2}
8	transcript	4.9×10^{-2}	58	epitheli	1.3×10^{-2}
9	protein	3.9×10^{-2}	59	factor	1.2×10^{-2}
10	mous	3.5×10^{-2}	60	promot	1.2×10^{-2}
11	dure	2.9×10^{-2}	61	cellular	1.2×10^{-2}
12	role	2.9×10^{-2}	62	repress	1.2×10^{-2}
13	zebrafish	2.9×10^{-2}	63	growth	1.2×10^{-2}
14	develop	2.9×10^{-2}	64	vivo	1.2×10^{-2}
15	differenti	2.9×10^{-2}	65	adult	1.2×10^{-2}
16	progenitor	2.7×10^{-2}	66	format	1.2×10^{-2}
17	drosophila	2.6×10^{-2}	67	essenti	1.2×10^{-2}
18	vertebr	2.5×10^{-2}	68	genet	1.2×10^{-2}
19	paper	2.5×10^{-2}	69	simul	1.1×10^{-2}
20	earli	2.4×10^{-2}	70	sperm	1.1×10^{-2}
21	tissu	2.3×10^{-2}	71	birth	1.1×10^{-2}
22	signal	2.3×10^{-2}	72	bind	1.1×10^{-2}
23	placent	2.3×10^{-2}	73	induc	1.1×10^{-2}
24	placenta	2.2×10^{-2}	74	mesenchym	1.1×10^{-2}
25	oocyt	2.1×10^{-2}	75	base	1.1×10^{-2}
26	prolifer	2.1×10^{-2}	76	cultur	1.1×10^{-2}
27	lineag	2.1×10^{-2}	77	rescu	1.1×10^{-2}
28	matern	2×10^{-2}	78	transgen	1×10^{-2}
29	stem	2×10^{-2}	79	method	1×10^{-2}
30	fetal	2×10^{-2}	80	mechan	1×10^{-2}
31	mutant	1.9×10^{-2}	81	fertil	1×10^{-2}
32	matur	1.9×10^{-2}	82	follicl	1×10^{-2}
33	pathway	1.9×10^{-2}	83	nervous	1×10^{-2}
34	germ	1.8×10^{-2}	84	anim	1×10^{-2}
35	fate	1.8×10^{-2}	85	chromatin	1×10^{-2}
36	mammalian	1.8×10^{-2}	86	normal	9.8×10^{-3}
37	neuron	1.8×10^{-2}	87	disrupt	9.7×10^{-3}
38	defect	1.7×10^{-2}	88	dorsal	9.6×10^{-3}
39	blastocyst	1.7×10^{-2}	89	genom	9.4×10^{-3}
40	pregnanc	1.7×10^{-2}	90	ovari	9.4×10^{-3}
41	function	1.7×10^{-2}	91	migrat	9.3×10^{-3}
42	mediat	1.6×10^{-2}	92	gestat	9.3×10^{-3}
43	specif	1.6×10^{-2}	93	knockdown	9.3×10^{-3}
44	phenotyp	1.6×10^{-2}	94	axon	9.3×10^{-3}
45	postnat	1.6×10^{-2}	95	conserv	9.3×10^{-3}
46	mice	1.5×10^{-2}	96	design	9.2×10^{-3}
47	stage	1.5×10^{-2}	97	reproduct	9.1×10^{-3}
48	wnt	1.5×10^{-2}	98	understood	8.7×10^{-3}
49	vitro	1.5×10^{-2}	99	rna	8.7×10^{-3}
50	receptor	1.5×10^{-2}	100	testi	8.6×10^{-3}

TABLE D.59. The list of the top 100 words in the category Ecology with RIGs

No.	Word	RIG
1	speci	1.6×10^{-1}
2	habitat	9.1×10^{-2}
3	ecolog	8.5×10^{-2}
4	ecosystem	7.9×10^{-2}
5	forest	5.2×10^{-2}
6	plant	4.8×10^{-2}
7	communiti	4.5×10^{-2}
8	popul	4.5×10^{-2}
9	abund	4.3×10^{-2}
10	climat	3.7×10^{-2}
11	divers	3.7×10^{-2}
12	biodivers	3.6×10^{-2}
13	predat	3.5×10^{-2}
14	landscap	3.5×10^{-2}
15	environment	3.4×10^{-2}
16	veget	3.3×10^{-2}
17	trait	3.1×10^{-2}
18	soil	3.1×10^{-2}
19	season	3×10^{-2}
20	conserv	2.9×10^{-2}
21	across	2.6×10^{-2}
22	tree	2.4×10^{-2}
23	suggest	2.4×10^{-2}
24	spatial	2.3×10^{-2}
25	pattern	2.3×10^{-2}
26	area	2.3×10^{-2}
27	land	2.3×10^{-2}
28	bird	2.2×10^{-2}
29	reproduct	2.2×10^{-2}
30	variat	2.2×10^{-2}
31	grassland	2.1×10^{-2}
32	may	2.1×10^{-2}
33	patient	2.1×10^{-2}
34	site	2.1×10^{-2}
35	biomass	2.1×10^{-2}
36	chang	2×10^{-2}
37	prey	2×10^{-2}
38	nutrient	1.9×10^{-2}
39	forag	1.9×10^{-2}
40	taxa	1.9×10^{-2}
41	rich	1.8×10^{-2}
42	fish	1.8×10^{-2}
43	declin	1.8×10^{-2}
44	herbivor	1.8×10^{-2}
45	manag	1.8×10^{-2}
46	paper	1.8×10^{-2}
47	evolutionari	1.7×10^{-2}
48	nativ	1.7×10^{-2}
49	import	1.7×10^{-2}
50	assemblag	1.6×10^{-2}

No.	Word	RIG
51	individu	1.6×10^{-2}
52	trophic	1.6×10^{-2}
53	natur	1.6×10^{-2}
54	domin	1.5×10^{-2}
55	north	1.5×10^{-2}
56	nich	1.4×10^{-2}
57	within	1.4×10^{-2}
58	marin	1.3×10^{-2}
59	affect	1.3×10^{-2}
60	understand	1.3×10^{-2}
61	water	1.3×10^{-2}
62	tropic	1.3×10^{-2}
63	annual	1.3×10^{-2}
64	resourc	1.3×10^{-2}
65	wetland	1.3×10^{-2}
66	anthropogen	1.3×10^{-2}
67	wildlif	1.3×10^{-2}
68	geograph	1.3×10^{-2}
69	cover	1.3×10^{-2}
70	river	1.3×10^{-2}
71	northern	1.3×10^{-2}
72	influenc	1.2×10^{-2}
73	shrub	1.2×10^{-2}
74	howev	1.2×10^{-2}
75	southern	1.2×10^{-2}
76	breed	1.2×10^{-2}
77	food	1.2×10^{-2}
78	summer	1.2×10^{-2}
79	clinic	1.2×10^{-2}
80	among	1.2×10^{-2}
81	canopi	1.2×10^{-2}
82	plot	1.2×10^{-2}
83	aquat	1.2×10^{-2}
84	graze	1.2×10^{-2}
85	insect	1.1×10^{-2}
86	relat	1.1×10^{-2}
87	gradient	1.1×10^{-2}
88	terrestri	1.1×10^{-2}
89	mate	1.1×10^{-2}
90	variabl	1.1×10^{-2}
91	agricultur	1.1×10^{-2}
92	scale	1.1×10^{-2}
93	biotic	1.1×10^{-2}
94	temper	1.1×10^{-2}
95	method	1.1×10^{-2}
96	woodi	1.1×10^{-2}
97	invertebr	1.1×10^{-2}
98	differ	1.1×10^{-2}
99	grass	1.1×10^{-2}
100	genet	1×10^{-2}

TABLE D.60. The list of the top 100 words in the category Economics with RIGs

No.	Word	RIG
1	market	9.1×10^{-2}
2	econom	6.4×10^{-2}
3	price	6×10^{-2}
4	polici	5.8×10^{-2}
5	economi	5.1×10^{-2}
6	countri	5.1×10^{-2}
7	financi	4.8×10^{-2}
8	firm	4.4×10^{-2}
9	empir	3.4×10^{-2}
10	capit	3.2×10^{-2}
11	invest	3.2×10^{-2}
12	trade	3.1×10^{-2}
13	sector	2.7×10^{-2}
14	find	2.6×10^{-2}
15	incom	2.6×10^{-2}
16	tax	2.4×10^{-2}
17	paper	2.4×10^{-2}
18	crisi	2.3×10^{-2}
19	cell	2.1×10^{-2}
20	macroeconom	2×10^{-2}
21	asset	2×10^{-2}
22	busi	1.9×10^{-2}
23	monetari	1.9×10^{-2}
24	bank	1.9×10^{-2}
25	wage	1.9×10^{-2}
26	welfar	1.8×10^{-2}
27	stock	1.7×10^{-2}
28	govern	1.7×10^{-2}
29	method	1.7×10^{-2}
30	patient	1.6×10^{-2}
31	incent	1.6×10^{-2}
32	household	1.6×10^{-2}
33	competit	1.6×10^{-2}
34	financ	1.6×10^{-2}
35	debt	1.5×10^{-2}
36	credit	1.5×10^{-2}
37	surfac	1.5×10^{-2}
38	investor	1.5×10^{-2}
39	privat	1.5×10^{-2}
40	temperatur	1.4×10^{-2}
41	gdp	1.4×10^{-2}
42	panel	1.4×10^{-2}
43	labor	1.4×10^{-2}
44	profit	1.4×10^{-2}
45	compani	1.3×10^{-2}
46	impact	1.3×10^{-2}
47	econometr	1.3×10^{-2}
48	return	1.3×10^{-2}
49	foreign	1.3×10^{-2}
50	fiscal	1.3×10^{-2}

No.	Word	RIG
51	conclus	1.3×10^{-2}
52	cost	1.2×10^{-2}
53	protein	1.2×10^{-2}
54	articl	1.2×10^{-2}
55	public	1.2×10^{-2}
56	european	1.2×10^{-2}
57	decis	1.2×10^{-2}
58	equilibrium	1.1×10^{-2}
59	clinic	1.1×10^{-2}
60	unemploy	1.1×10^{-2}
61	labour	1.1×10^{-2}
62	equiti	1.1×10^{-2}
63	export	1×10^{-2}
64	portfolio	1×10^{-2}
65	acid	1×10^{-2}
66	social	1×10^{-2}
67	estim	9.8×10^{-3}
68	detect	9.7×10^{-3}
69	corpor	9.7×10^{-3}
70	polit	9.6×10^{-3}
71	fund	9.5×10^{-3}
72	volatil	9.3×10^{-3}
73	industri	9.1×10^{-3}
74	shock	9.1×10^{-3}
75	gene	9×10^{-3}
76	union	8.9×10^{-3}
77	electron	8.8×10^{-3}
78	inflat	8.7×10^{-3}
79	experiment	8.7×10^{-3}
80	domest	8.6×10^{-3}
81	premium	8.5×10^{-3}
82	loan	8.5×10^{-3}
83	institut	8.4×10^{-3}
84	demand	8.4×10^{-3}
85	treatment	8.4×10^{-3}
86	diseas	8.4×10^{-3}
87	molecular	8.3×10^{-3}
88	model	8.2×10^{-3}
89	choic	8.2×10^{-3}
90	pay	8.2×10^{-3}
91	imag	8.1×10^{-3}
92	background	8.1×10^{-3}
93	earn	8×10^{-3}
94	oxid	7.9×10^{-3}
95	phase	7.8×10^{-3}
96	literatur	7.7×10^{-3}
97	materi	7.7×10^{-3}
98	evid	7.7×10^{-3}
99	reform	7.7×10^{-3}
100	express	7.6×10^{-3}

TABLE D.61. The list of the top 100 words in the category Education and Educational Research with RIGs

No.	Word	RIG
1	student	2.7×10^{-1}
2	educ	2.2×10^{-1}
3	teacher	1.6×10^{-1}
4	learn	1.4×10^{-1}
5	teach	1.4×10^{-1}
6	school	1.2×10^{-1}
7	univers	6.5×10^{-2}
8	classroom	6.3×10^{-2}
9	skill	5.7×10^{-2}
10	research	5.5×10^{-2}
11	cours	5.1×10^{-2}
12	learner	4.8×10^{-2}
13	academ	4.6×10^{-2}
14	curriculum	4.3×10^{-2}
15	pedagog	3.6×10^{-2}
16	profession	3.5×10^{-2}
17	practic	3.5×10^{-2}
18	social	3.3×10^{-2}
19	languag	3.3×10^{-2}
20	knowledg	3.3×10^{-2}
21	instruct	3.2×10^{-2}
22	scienc	3.1×10^{-2}
23	colleg	2.9×10^{-2}
24	english	2.8×10^{-2}
25	particip	2.8×10^{-2}
26	undergradu	2.6×10^{-2}
27	faculti	2.5×10^{-2}
28	engag	2.4×10^{-2}
29	graduat	2.4×10^{-2}
30	interview	2.3×10^{-2}
31	develop	2.2×10^{-2}
32	think	2.1×10^{-2}
33	context	2.1×10^{-2}
34	train	2.1×10^{-2}
35	institut	2.1×10^{-2}
36	focus	2×10^{-2}
37	cell	2×10^{-2}
38	pedagogi	1.9×10^{-2}
39	articl	1.9×10^{-2}
40	way	1.9×10^{-2}
41	program	1.8×10^{-2}
42	compet	1.7×10^{-2}
43	project	1.7×10^{-2}
44	taught	1.7×10^{-2}
45	question	1.7×10^{-2}
46	lectur	1.6×10^{-2}
47	instructor	1.6×10^{-2}
48	collabor	1.6×10^{-2}
49	questionnair	1.6×10^{-2}
50	discuss	1.6×10^{-2}

No.	Word	RIG
51	concept	1.6×10^{-2}
52	attitud	1.5×10^{-2}
53	semest	1.5×10^{-2}
54	literaci	1.5×10^{-2}
55	technolog	1.5×10^{-2}
56	temperatur	1.5×10^{-2}
57	lesson	1.4×10^{-2}
58	patient	1.4×10^{-2}
59	opportun	1.4×10^{-2}
60	motiv	1.4×10^{-2}
61	creativ	1.4×10^{-2}
62	person	1.4×10^{-2}
63	career	1.3×10^{-2}
64	explor	1.3×10^{-2}
65	perspect	1.3×10^{-2}
66	qualit	1.3×10^{-2}
67	societi	1.3×10^{-2}
68	need	1.3×10^{-2}
69	understand	1.3×10^{-2}
70	percept	1.2×10^{-2}
71	innov	1.2×10^{-2}
72	disciplin	1.2×10^{-2}
73	surfac	1.2×10^{-2}
74	write	1.2×10^{-2}
75	onlin	1.2×10^{-2}
76	communic	1.2×10^{-2}
77	reform	1.2×10^{-2}
78	induc	1.2×10^{-2}
79	protein	1.2×10^{-2}
80	curricula	1.2×10^{-2}
81	survey	1.2×10^{-2}
82	properti	1.1×10^{-2}
83	draw	1.1×10^{-2}
84	cultur	1.1×10^{-2}
85	children	1.1×10^{-2}
86	polici	1.1×10^{-2}
87	grade	1.1×10^{-2}
88	paramet	1×10^{-2}
89	experi	1×10^{-2}
90	find	9.9×10^{-3}
91	peer	9.9×10^{-3}
92	nation	9.9×10^{-3}
93	peopl	9.7×10^{-3}
94	acid	9.7×10^{-3}
95	work	9.6×10^{-3}
96	issu	9.6×10^{-3}
97	treatment	9.6×10^{-3}
98	vocat	9.5×10^{-3}
99	creat	9.5×10^{-3}
100	decreas	9.4×10^{-3}

TABLE D.62. The list of the top 100 words in the category Education, Scientific Disciplines with RIGs

No.	Word	RIG
1	student	3×10^{-1}
2	educ	1.4×10^{-1}
3	learn	1.3×10^{-1}
4	cours	1.1×10^{-1}
5	teach	1×10^{-1}
6	engin	8.8×10^{-2}
7	undergradu	8.1×10^{-2}
8	skill	6.5×10^{-2}
9	curriculum	6×10^{-2}
10	univers	5.5×10^{-2}
11	faculti	5.2×10^{-2}
12	school	5.1×10^{-2}
13	graduat	4.9×10^{-2}
14	program	4.7×10^{-2}
15	scienc	4.6×10^{-2}
16	profession	3.7×10^{-2}
17	classroom	3.7×10^{-2}
18	instructor	3.7×10^{-2}
19	semest	3.3×10^{-2}
20	academ	3.2×10^{-2}
21	project	3.1×10^{-2}
22	lectur	3.1×10^{-2}
23	introductori	3.1×10^{-2}
24	colleg	3×10^{-2}
25	practic	2.8×10^{-2}
26	instruct	2.8×10^{-2}
27	knowledg	2.8×10^{-2}
28	teacher	2.6×10^{-2}
29	curricula	2.5×10^{-2}
30	taught	2.4×10^{-2}
31	experi	2.4×10^{-2}
32	particip	2.4×10^{-2}
33	engag	2.3×10^{-2}
34	team	2.3×10^{-2}
35	survey	2.3×10^{-2}
36	career	2.2×10^{-2}
37	train	2.1×10^{-2}
38	develop	2.1×10^{-2}
39	research	2.1×10^{-2}
40	design	1.9×10^{-2}
41	concept	1.9×10^{-2}
42	implement	1.8×10^{-2}
43	institut	1.8×10^{-2}
44	disciplin	1.8×10^{-2}
45	collabor	1.8×10^{-2}
46	understand	1.7×10^{-2}
47	topic	1.7×10^{-2}
48	pedagog	1.7×10^{-2}
49	opportun	1.6×10^{-2}
50	medic	1.6×10^{-2}

No.	Word	RIG
51	technolog	1.6×10^{-2}
52	think	1.6×10^{-2}
53	compet	1.5×10^{-2}
54	learner	1.5×10^{-2}
55	question	1.5×10^{-2}
56	discuss	1.5×10^{-2}
57	year	1.4×10^{-2}
58	focus	1.4×10^{-2}
59	laboratori	1.4×10^{-2}
60	tool	1.3×10^{-2}
61	need	1.3×10^{-2}
62	senior	1.3×10^{-2}
63	assess	1.3×10^{-2}
64	goal	1.3×10^{-2}
65	onlin	1.3×10^{-2}
66	chemistri	1.2×10^{-2}
67	help	1.2×10^{-2}
68	interview	1.2×10^{-2}
69	feedback	1.1×10^{-2}
70	describ	1.1×10^{-2}
71	work	1.1×10^{-2}
72	motiv	1×10^{-2}
73	peer	1×10^{-2}
74	paper	1×10^{-2}
75	percept	1×10^{-2}
76	lesson	9.9×10^{-3}
77	pedagogi	9.9×10^{-3}
78	cell	9.9×10^{-3}
79	workshop	9.8×10^{-3}
80	creat	9.6×10^{-3}
81	technic	9.4×10^{-3}
82	encourag	9.3×10^{-3}
83	will	9.3×10^{-3}
84	class	9.1×10^{-3}
85	challeng	9×10^{-3}
86	profess	9×10^{-3}
87	attitud	9×10^{-3}
88	inquiri	8.8×10^{-3}
89	qualit	8.8×10^{-3}
90	induc	8.7×10^{-3}
91	provid	8.5×10^{-3}
92	innov	8.4×10^{-3}
93	communic	8.2×10^{-3}
94	session	8.2×10^{-3}
95	theme	7.9×10^{-3}
96	medicin	7.8×10^{-3}
97	complet	7.8×10^{-3}
98	nurs	7.7×10^{-3}
99	author	7.6×10^{-3}
100	address	7.5×10^{-3}

Table D.63. The list of the top 100 words in the category Education, Special with RIGs

No.	Word	RIG
1	disabl	1.6×10^{-1}
2	children	1.1×10^{-1}
3	autism	8.8×10^{-2}
4	intellectu	8.7×10^{-2}
5	asd	6.2×10^{-2}
6	particip	5.5×10^{-2}
7	student	5.5×10^{-2}
8	intervent	5.2×10^{-2}
9	skill	5×10^{-2}
10	development	4.5×10^{-2}
11	school	4.3×10^{-2}
12	disord	4.1×10^{-2}
13	educ	3.7×10^{-2}
14	teacher	3.3×10^{-2}
15	instruct	3.1×10^{-2}
16	parent	3×10^{-2}
17	child	2.7×10^{-2}
18	research	2.7×10^{-2}
19	social	2.6×10^{-2}
20	languag	2.6×10^{-2}
21	age	2.6×10^{-2}
22	peopl	2.6×10^{-2}
23	difficulti	2.6×10^{-2}
24	learn	2.5×10^{-2}
25	read	2.4×10^{-2}
26	classroom	2.3×10^{-2}
27	impair	2.3×10^{-2}
28	spectrum	2.1×10^{-2}
29	peer	1.9×10^{-2}
30	implic	1.9×10^{-2}
31	emot	1.9×10^{-2}
32	examin	1.8×10^{-2}
33	preschool	1.7×10^{-2}
34	word	1.7×10^{-2}
35	individu	1.7×10^{-2}
36	deficit	1.7×10^{-2}
37	studi	1.7×10^{-2}
38	group	1.6×10^{-2}
39	cognit	1.6×10^{-2}
40	support	1.5×10^{-2}
41	score	1.4×10^{-2}
42	cell	1.4×10^{-2}
43	adult	1.4×10^{-2}
44	assess	1.4×10^{-2}
45	task	1.3×10^{-2}
46	behavior	1.3×10^{-2}
47	discuss	1.3×10^{-2}
48	typic	1.3×10^{-2}
49	across	1.3×10^{-2}
50	special	1.3×10^{-2}

No.	Word	RIG
51	need	1.2×10^{-2}
52	syndrom	1.2×10^{-2}
53	hear	1.2×10^{-2}
54	phonolog	1.2×10^{-2}
55	young	1.2×10^{-2}
56	literaci	1.2×10^{-2}
57	temperatur	1.1×10^{-2}
58	find	1.1×10^{-2}
59	abil	1×10^{-2}
60	practic	1×10^{-2}
61	vocabulari	1×10^{-2}
62	fluenci	1×10^{-2}
63	nonverb	1×10^{-2}
64	caregiv	1×10^{-2}
65	interview	1×10^{-2}
66	access	9.8×10^{-3}
67	verbal	9.8×10^{-3}
68	year	9.7×10^{-3}
69	simul	9.7×10^{-3}
70	palsi	9.3×10^{-3}
71	teach	9.1×10^{-3}
72	taught	9.1×10^{-3}
73	person	9×10^{-3}
74	reader	8.7×10^{-3}
75	academ	8.5×10^{-3}
76	energi	8.5×10^{-3}
77	speech	8.4×10^{-3}
78	train	8.2×10^{-3}
79	elementari	8×10^{-3}
80	concentr	8×10^{-3}
81	program	8×10^{-3}
82	mental	7.8×10^{-3}
83	surfac	7.8×10^{-3}
84	item	7.7×10^{-3}
85	profession	7.6×10^{-3}
86	staff	7.5×10^{-3}
87	grade	7.4×10^{-3}
88	adolesc	7.4×10^{-3}
89	boy	7.4×10^{-3}
90	questionnair	7.3×10^{-3}
91	propos	7.3×10^{-3}
92	properti	7.2×10^{-3}
93	induc	7.2×10^{-3}
94	famili	7.1×10^{-3}
95	session	7.1×10^{-3}
96	develop	7.1×10^{-3}
97	protein	6.9×10^{-3}
98	mechan	6.9×10^{-3}
99	servic	6.8×10^{-3}
100	visual	6.8×10^{-3}

Table D.64. The list of the top 100 words in the category Electrochemistry with RIGs

No.	Word	RIG
1	electrochem	1.9×10^{-1}
2	electrod	1.5×10^{-1}
3	electrolyt	7.8×10^{-2}
4	voltammetri	6.7×10^{-2}
5	anod	5.8×10^{-2}
6	cathod	5.6×10^{-2}
7	oxid	5.1×10^{-2}
8	carbon	4.7×10^{-2}
9	fuel	4.1×10^{-2}
10	cyclic	4×10^{-2}
11	batteri	3.7×10^{-2}
12	imped	3.5×10^{-2}
13	lithium	3.5×10^{-2}
14	ion	3.4×10^{-2}
15	hydrogen	3.3×10^{-2}
16	surfac	3.2×10^{-2}
17	prepar	3.1×10^{-2}
18	spectroscopi	3×10^{-2}
19	charg	3×10^{-2}
20	electron	3×10^{-2}
21	electrocatalyt	2.8×10^{-2}
22	mah	2.8×10^{-2}
23	sensor	2.7×10^{-2}
24	film	2.4×10^{-2}
25	discharg	2.4×10^{-2}
26	reaction	2.3×10^{-2}
27	cycl	2.3×10^{-2}
28	catalyst	2.2×10^{-2}
29	conclus	2.2×10^{-2}
30	nanoparticl	2.1×10^{-2}
31	fabric	1.9×10^{-2}
32	glassi	1.9×10^{-2}
33	electrodeposit	1.8×10^{-2}
34	layer	1.8×10^{-2}
35	deposit	1.8×10^{-2}
36	exhibit	1.8×10^{-2}
37	synthes	1.8×10^{-2}
38	stabil	1.8×10^{-2}
39	patient	1.8×10^{-2}
40	sem	1.7×10^{-2}
41	redox	1.7×10^{-2}
42	biosensor	1.7×10^{-2}
43	microscopi	1.7×10^{-2}
44	graphen	1.6×10^{-2}
45	xrd	1.6×10^{-2}
46	scan	1.6×10^{-2}
47	densiti	1.6×10^{-2}
48	capac	1.5×10^{-2}
49	cell	1.5×10^{-2}
50	excel	1.5×10^{-2}

No.	Word	RIG
51	composit	1.5×10^{-2}
52	coat	1.5×10^{-2}
53	metal	1.4×10^{-2}
54	materi	1.4×10^{-2}
55	nanotub	1.4×10^{-2}
56	catalyt	1.4×10^{-2}
57	ray	1.4×10^{-2}
58	temperatur	1.4×10^{-2}
59	graphit	1.3×10^{-2}
60	sensit	1.3×10^{-2}
61	current	1.3×10^{-2}
62	ionic	1.3×10^{-2}
63	capacit	1.3×10^{-2}
64	oxygen	1.3×10^{-2}
65	solut	1.3×10^{-2}
66	porous	1.2×10^{-2}
67	year	1.2×10^{-2}
68	diffract	1.2×10^{-2}
69	concentr	1.2×10^{-2}
70	perform	1.2×10^{-2}
71	detect	1.2×10^{-2}
72	dope	1.1×10^{-2}
73	immobil	1.1×10^{-2}
74	associ	1.1×10^{-2}
75	membran	1.1×10^{-2}
76	xps	1.1×10^{-2}
77	modifi	1×10^{-2}
78	corros	1×10^{-2}
79	solid	1×10^{-2}
80	photoelectron	9.8×10^{-3}
81	high	9.8×10^{-3}
82	diffus	9.7×10^{-3}
83	polym	9.6×10^{-3}
84	platinum	9.3×10^{-3}
85	tem	9.2×10^{-3}
86	proton	9.2×10^{-3}
87	mol	9.1×10^{-3}
88	background	9.1×10^{-3}
89	adsorpt	9.1×10^{-3}
90	object	9.1×10^{-3}
91	transfer	9×10^{-3}
92	character	8.7×10^{-3}
93	conduct	8.5×10^{-3}
94	risk	8.5×10^{-3}
95	gold	8.5×10^{-3}
96	chemic	8.5×10^{-3}
97	aqueous	8.5×10^{-3}
98	electr	8.4×10^{-3}
99	voltag	8.3×10^{-3}
100	degre	8.3×10^{-3}

TABLE D.65. The list of the top 100 words in the category Emergency Medicine with RIGs

No.	Word	RIG	No.	Word	RIG
1	emerg	1.5×10^{-1}	51	visit	1.6×10^{-2}
2	depart	1.5×10^{-1}	52	old	1.6×10^{-2}
3	patient	1.4×10^{-1}	53	median	1.6×10^{-2}
4	conclus	1×10^{-1}	54	adult	1.5×10^{-2}
5	hospit	8×10^{-2}	55	month	1.5×10^{-2}
6	trauma	7.8×10^{-2}	56	interv	1.5×10^{-2}
7	resuscit	7.4×10^{-2}	57	treatment	1.5×10^{-2}
8	injuri	7.1×10^{-2}	58	surgic	1.5×10^{-2}
9	care	6.1×10^{-2}	59	januari	1.5×10^{-2}
10	physician	5.1×10^{-2}	60	inhospit	1.5×10^{-2}
11	prehospit	4.8×10^{-2}	61	report	1.4×10^{-2}
12	medic	4.6×10^{-2}	62	intervent	1.4×10^{-2}
13	outcom	4.5×10^{-2}	63	blunt	1.3×10^{-2}
14	object	4.4×10^{-2}	64	record	1.3×10^{-2}
15	retrospect	4.3×10^{-2}	65	medicin	1.3×10^{-2}
16	arrest	4.2×10^{-2}	66	propos	1.3×10^{-2}
17	method	4.1×10^{-2}	67	treat	1.3×10^{-2}
18	cardiac	3.9×10^{-2}	68	associ	1.2×10^{-2}
19	year	3.9×10^{-2}	69	abdomin	1.2×10^{-2}
20	admiss	3.7×10^{-2}	70	arriv	1.2×10^{-2}
21	acut	3.6×10^{-2}	71	icu	1.2×10^{-2}
22	admit	3.3×10^{-2}	72	hour	1.2×10^{-2}
23	clinic	3.2×10^{-2}	73	children	1.2×10^{-2}
24	chest	3×10^{-2}	74	aim	1.2×10^{-2}
25	background	3×10^{-2}	75	common	1.1×10^{-2}
26	prospect	2.9×10^{-2}	76	minut	1.1×10^{-2}
27	cardiopulmonari	2.5×10^{-2}	77	cohort	1.1×10^{-2}
28	score	2.4×10^{-2}	78	decemb	1.1×10^{-2}
29	pain	2.4×10^{-2}	79	death	1.1×10^{-2}
30	review	2.4×10^{-2}	80	enrol	1.1×10^{-2}
31	fractur	2.4×10^{-2}	81	properti	1.1×10^{-2}
32	mortal	2.3×10^{-2}	82	iqr	1.1×10^{-2}
33	pediatr	2.3×10^{-2}	83	total	1×10^{-2}
34	introduct	2.3×10^{-2}	84	studi	1×10^{-2}
35	age	2.3×10^{-2}	85	group	1×10^{-2}
36	paper	2.2×10^{-2}	86	diagnos	1×10^{-2}
37	discharg	2.2×10^{-2}	87	day	1×10^{-2}
38	traumat	2×10^{-2}	88	return	1×10^{-2}
39	case	1.9×10^{-2}	89	assess	9.9×10^{-3}
40	diagnosi	1.9×10^{-2}	90	burn	9.9×10^{-3}
41	result	1.8×10^{-2}	91	demograph	9.9×10^{-3}
42	neurolog	1.8×10^{-2}	92	initi	9.9×10^{-3}
43	injur	1.7×10^{-2}	93	registri	9.8×10^{-3}
44	manag	1.7×10^{-2}	94	risk	9.7×10^{-3}
45	complic	1.7×10^{-2}	95	rate	9.6×10^{-3}
46	includ	1.7×10^{-2}	96	period	9.6×10^{-3}
47	intub	1.7×10^{-2}	97	primari	9.6×10^{-3}
48	stay	1.6×10^{-2}	98	sign	9.5×10^{-3}
49	confid	1.6×10^{-2}	99	servic	9.3×10^{-3}
50	receiv	1.6×10^{-2}	100	intraven	9.3×10^{-3}

TABLE D.66. The list of the top 100 words in the category Endocrinology and Metabolism with RIGs

No.	Word	RIG
1	diabet	1.2×10^{-1}
2	insulin	8.7×10^{-2}
3	glucos	6.7×10^{-2}
4	conclus	6.1×10^{-2}
5	obes	5.3×10^{-2}
6	hormon	5.3×10^{-2}
7	metabol	5.1×10^{-2}
8	associ	4×10^{-2}
9	age	3.8×10^{-2}
10	patient	3.6×10^{-2}
11	bmi	3.4×10^{-2}
12	level	3.3×10^{-2}
13	receptor	3.2×10^{-2}
14	paper	3.2×10^{-2}
15	serum	3×10^{-2}
16	thyroid	3×10^{-2}
17	increas	2.9×10^{-2}
18	mellitus	2.7×10^{-2}
19	fat	2.6×10^{-2}
20	blood	2.5×10^{-2}
21	adipos	2.5×10^{-2}
22	bone	2.4×10^{-2}
23	bodi	2.3×10^{-2}
24	risk	2.3×10^{-2}
25	women	2.3×10^{-2}
26	express	2.3×10^{-2}
27	diseas	2.2×10^{-2}
28	secret	2.1×10^{-2}
29	signific	2.1×10^{-2}
30	regul	2.1×10^{-2}
31	treatment	2×10^{-2}
32	aim	2×10^{-2}
33	mice	2×10^{-2}
34	pituitari	1.9×10^{-2}
35	studi	1.9×10^{-2}
36	protein	1.8×10^{-2}
37	type	1.8×10^{-2}
38	object	1.8×10^{-2}
39	induc	1.8×10^{-2}
40	plasma	1.7×10^{-2}
41	year	1.7×10^{-2}
42	diet	1.7×10^{-2}
43	subject	1.6×10^{-2}
44	stimul	1.6×10^{-2}
45	homeostasi	1.6×10^{-2}
46	baselin	1.6×10^{-2}
47	cardiovascular	1.6×10^{-2}
48	male	1.6×10^{-2}
49	clinic	1.5×10^{-2}
50	rat	1.5×10^{-2}
51	gene	1.5×10^{-2}
52	decreas	1.5×10^{-2}
53	cell	1.5×10^{-2}
54	syndrom	1.5×10^{-2}
55	assess	1.5×10^{-2}
56	propos	1.5×10^{-2}
57	impair	1.4×10^{-2}
58	week	1.4×10^{-2}
59	defici	1.4×10^{-2}
60	lipid	1.4×10^{-2}
61	control	1.4×10^{-2}
62	tissu	1.4×10^{-2}
63	beta	1.4×10^{-2}
64	adult	1.4×10^{-2}
65	simul	1.4×10^{-2}
66	may	1.4×10^{-2}
67	normal	1.3×10^{-2}
68	weight	1.3×10^{-2}
69	whether	1.3×10^{-2}
70	adjust	1.3×10^{-2}
71	sex	1.3×10^{-2}
72	group	1.3×10^{-2}
73	elev	1.3×10^{-2}
74	mass	1.3×10^{-2}
75	mrna	1.3×10^{-2}
76	cholesterol	1.3×10^{-2}
77	factor	1.3×10^{-2}
78	men	1.3×10^{-2}
79	endocrin	1.2×10^{-2}
80	triglycerid	1.2×10^{-2}
81	suggest	1.2×10^{-2}
82	marker	1.2×10^{-2}
83	cortisol	1.2×10^{-2}
84	testosteron	1.2×10^{-2}
85	intak	1.2×10^{-2}
86	healthi	1.2×10^{-2}
87	mediat	1.2×10^{-2}
88	femal	1.1×10^{-2}
89	peptid	1.1×10^{-2}
90	waist	1.1×10^{-2}
91	skelet	1.1×10^{-2}
92	estradiol	1.1×10^{-2}
93	alter	1.1×10^{-2}
94	role	1.1×10^{-2}
95	activ	1.1×10^{-2}
96	temperatur	1.1×10^{-2}
97	index	1.1×10^{-2}
98	androgen	1.1×10^{-2}
99	overweight	1.1×10^{-2}
100	cohort	1.1×10^{-2}

TABLE D.67. The list of the top 100 words in the category Energy and Fuels with RIGs

No.	Word	RIG	No.	Word	RIG
1	power	5.2×10^{-2}	51	express	9.1×10^{-3}
2	energi	5.2×10^{-2}	52	reactor	9.1×10^{-3}
3	fuel	5.1×10^{-2}	53	suppli	9.1×10^{-3}
4	solar	4.3×10^{-2}	54	age	8.9×10^{-3}
5	electr	2.9×10^{-2}	55	human	8.8×10^{-3}
6	heat	2.9×10^{-2}	56	cathod	8.8×10^{-3}
7	gas	2.7×10^{-2}	57	capac	8.7×10^{-3}
8	patient	2.6×10^{-2}	58	gene	8.5×10^{-3}
9	voltag	2.5×10^{-2}	59	product	8.3×10^{-3}
10	oper	2.3×10^{-2}	60	electrod	8.2×10^{-3}
11	photovolta	2.2×10^{-2}	61	biodiesel	8.1×10^{-3}
12	temperatur	2.2×10^{-2}	62	output	8×10^{-3}
13	renew	2.2×10^{-2}	63	cancer	8×10^{-3}
14	co2	2.2×10^{-2}	64	instal	7.9×10^{-3}
15	effici	2.1×10^{-2}	65	flow	7.9×10^{-3}
16	conclus	2.1×10^{-2}	66	emiss	7.8×10^{-3}
17	turbin	2×10^{-2}	67	anod	7.8×10^{-3}
18	wind	1.9×10^{-2}	68	optim	7.8×10^{-3}
19	grid	1.8×10^{-2}	69	feedstock	7.7×10^{-3}
20	thermal	1.8×10^{-2}	70	methan	7.7×10^{-3}
21	combust	1.8×10^{-2}	71	tissu	7.5×10^{-3}
22	storag	1.8×10^{-2}	72	consumpt	7.3×10^{-3}
23	carbon	1.6×10^{-2}	73	demand	7.2×10^{-3}
24	batteri	1.6×10^{-2}	74	pressur	6.9×10^{-3}
25	clinic	1.6×10^{-2}	75	condit	6.8×10^{-3}
26	paper	1.5×10^{-2}	76	collector	6.8×10^{-3}
27	convert	1.5×10^{-2}	77	outcom	6.7×10^{-3}
28	diseas	1.4×10^{-2}	78	therapi	6.6×10^{-3}
29	oil	1.4×10^{-2}	79	invert	6.6×10^{-3}
30	load	1.4×10^{-2}	80	lithium	6.6×10^{-3}
31	coal	1.4×10^{-2}	81	cool	6.6×10^{-3}
32	simul	1.3×10^{-2}	82	econom	6.6×10^{-3}
33	cost	1.3×10^{-2}	83	technolog	6.5×10^{-3}
34	biomass	1.2×10^{-2}	84	blood	6.3×10^{-3}
35	perform	1.2×10^{-2}	85	charg	6.3×10^{-3}
36	air	1.2×10^{-2}	86	current	6.2×10^{-3}
37	steam	1.1×10^{-2}	87	pyrolysi	6.2×10^{-3}
38	hydrogen	1.1×10^{-2}	88	infect	6.1×10^{-3}
39	convers	1.1×10^{-2}	89	drug	6×10^{-3}
40	background	1.1×10^{-2}	90	fossil	6×10^{-3}
41	system	1.1×10^{-2}	91	popul	5.9×10^{-3}
42	associ	1.1×10^{-2}	92	reservoir	5.9×10^{-3}
43	generat	1×10^{-2}	93	degre	5.9×10^{-3}
44	electrochem	1×10^{-2}	94	suggest	5.9×10^{-3}
45	group	1×10^{-2}	95	male	5.8×10^{-3}
46	protein	1×10^{-2}	96	adult	5.8×10^{-3}
47	cycl	9.4×10^{-3}	97	wast	5.7×10^{-3}
48	catalyst	9.4×10^{-3}	98	may	5.7×10^{-3}
49	electrolyt	9.3×10^{-3}	99	tumor	5.7×10^{-3}
50	diesel	9.2×10^{-3}	100	circuit	5.6×10^{-3}

TABLE D.68. The list of the top 100 words in the category Engineering, Aerospace with RIGs

No.	Word	RIG	No.	Word	RIG
1	spacecraft	6.7×10^{-2}	51	cell	9×10^{-3}
2	aircraft	6.2×10^{-2}	52	group	8.9×10^{-3}
3	flight	6.2×10^{-2}	53	protein	8.9×10^{-3}
4	mission	5.8×10^{-2}	54	turbul	8.7×10^{-3}
5	orbit	4×10^{-2}	55	model	8.6×10^{-3}
6	aerodynam	3.6×10^{-2}	56	air	8.5×10^{-3}
7	satellit	3.2×10^{-2}	57	jet	8.5×10^{-3}
8	paper	3×10^{-2}	58	requir	8.3×10^{-3}
9	simul	2.5×10^{-2}	59	pressur	8.2×10^{-3}
10	space	2.5×10^{-2}	60	sensor	8.1×10^{-3}
11	earth	2.3×10^{-2}	61	suggest	7.9×10^{-3}
12	design	2.2×10^{-2}	62	nonlinear	7.6×10^{-3}
13	wing	2.2×10^{-2}	63	problem	7.5×10^{-3}
14	propuls	2.1×10^{-2}	64	present	7.5×10^{-3}
15	thrust	2×10^{-2}	65	fli	7.4×10^{-3}
16	launch	1.8×10^{-2}	66	stoke	7.3×10^{-3}
17	trajectori	1.8×10^{-2}	67	radar	7.1×10^{-3}
18	patient	1.8×10^{-2}	68	gene	7×10^{-3}
19	system	1.7×10^{-2}	69	age	7×10^{-3}
20	vehicl	1.7×10^{-2}	70	ground	6.9×10^{-3}
21	numer	1.5×10^{-2}	71	acid	6.7×10^{-3}
22	engin	1.5×10^{-2}	72	wind	6.7×10^{-3}
23	conclus	1.4×10^{-2}	73	background	6.5×10^{-3}
24	navig	1.3×10^{-2}	74	vortex	6.5×10^{-3}
25	flow	1.3×10^{-2}	75	propos	6.4×10^{-3}
26	drag	1.2×10^{-2}	76	error	6.3×10^{-3}
27	dynam	1.2×10^{-2}	77	fuel	6.1×10^{-3}
28	angl	1.2×10^{-2}	78	speed	5.8×10^{-3}
29	unsteadi	1.2×10^{-2}	79	kalman	5.6×10^{-3}
30	unman	1.1×10^{-2}	80	studi	5.5×10^{-3}
31	oper	1.1×10^{-2}	81	solver	5.5×10^{-3}
32	attitud	1.1×10^{-2}	82	optim	5.5×10^{-3}
33	reynold	1.1×10^{-2}	83	cancer	5.4×10^{-3}
34	perform	1.1×10^{-2}	84	aerial	5.4×10^{-3}
35	treatment	1.1×10^{-2}	85	guidanc	5.3×10^{-3}
36	capabl	1.1×10^{-2}	86	report	5.2×10^{-3}
37	blade	1.1×10^{-2}	87	tunnel	5.1×10^{-3}
38	clinic	1×10^{-2}	88	concentr	5.1×10^{-3}
39	comput	1×10^{-2}	89	tissu	5.1×10^{-3}
40	altitud	1×10^{-2}	90	base	5×10^{-3}
41	configur	1×10^{-2}	91	equat	4.9×10^{-3}
42	diseas	9.9×10^{-3}	92	constraint	4.9×10^{-3}
43	motion	9.6×10^{-3}	93	hardwar	4.9×10^{-3}
44	track	9.5×10^{-3}	94	accuraci	4.9×10^{-3}
45	veloc	9.3×10^{-3}	95	approach	4.8×10^{-3}
46	rotor	9.3×10^{-3}	96	architectur	4.8×10^{-3}
47	navier	9.2×10^{-3}	97	concept	4.8×10^{-3}
48	algorithm	9.1×10^{-3}	98	associ	4.8×10^{-3}
49	actuat	9×10^{-3}	99	autonom	4.8×10^{-3}
50	lift	9×10^{-3}	100	pitch	4.8×10^{-3}

TABLE D.69. The list of the top 100 words in the category Engineering, Biomedical with RIGs

No.	Word	RIG	No.	Word	RIG
1	implant	3×10^{-2}	51	automat	5.1×10^{-3}
2	tissu	2.9×10^{-2}	52	hydroxyapatit	5.1×10^{-3}
3	bone	2.2×10^{-2}	53	heal	5×10^{-3}
4	imag	1.9×10^{-2}	54	engin	5×10^{-3}
5	scaffold	1.6×10^{-2}	55	flexion	4.9×10^{-3}
6	biomechan	1.4×10^{-2}	56	poli	4.9×10^{-3}
7	vivo	1.2×10^{-2}	57	cartilag	4.9×10^{-3}
8	biomateri	1.1×10^{-2}	58	mesenchym	4.8×10^{-3}
9	collagen	1.1×10^{-2}	59	cardiac	4.8×10^{-3}
10	phantom	1.1×10^{-2}	60	anatom	4.7×10^{-3}
11	gait	1×10^{-2}	61	blood	4.6×10^{-3}
12	use	1×10^{-2}	62	osteoblast	4.5×10^{-3}
13	clinic	9.3×10^{-3}	63	mri	4.4×10^{-3}
14	biocompat	9.2×10^{-3}	64	ankl	4.4×10^{-3}
15	vitro	8.8×10^{-3}	65	adhes	4.4×10^{-3}
16	segment	8.7×10^{-3}	66	mechan	4.3×10^{-3}
17	devic	8.6×10^{-3}	67	speci	4.3×10^{-3}
18	motion	8.5×10^{-3}	68	accur	4.3×10^{-3}
19	hydrogel	8.3×10^{-3}	69	evalu	4.3×10^{-3}
20	muscl	8.1×10^{-3}	70	stem	4.2×10^{-3}
21	comput	8.1×10^{-3}	71	applic	4.2×10^{-3}
22	subject	7.9×10^{-3}	72	prosthesi	4.1×10^{-3}
23	human	7.7×10^{-3}	73	patient	4.1×10^{-3}
24	accuraci	7.5×10^{-3}	74	classif	4.1×10^{-3}
25	movement	7.4×10^{-3}	75	test	4.1×10^{-3}
26	eeg	7.1×10^{-3}	76	propos	4×10^{-3}
27	kinemat	7.1×10^{-3}	77	postur	3.9×10^{-3}
28	robot	7.1×10^{-3}	78	algorithm	3.9×10^{-3}
29	heart	6.8×10^{-3}	79	compar	3.8×10^{-3}
30	regener	6.8×10^{-3}	80	hemodynam	3.8×10^{-3}
31	signal	6.7×10^{-3}	81	dental	3.8×10^{-3}
32	knee	6.7×10^{-3}	82	plant	3.8×10^{-3}
33	brain	6.5×10^{-3}	83	temperatur	3.8×10^{-3}
34	stiff	6.4×10^{-3}	84	method	3.7×10^{-3}
35	joint	6.3×10^{-3}	85	hip	3.7×10^{-3}
36	ecg	6.2×10^{-3}	86	regen	3.7×10^{-3}
37	cell	6.1×10^{-3}	87	ventricular	3.6×10^{-3}
38	noninvas	5.9×10^{-3}	88	perform	3.6×10^{-3}
39	rehabilit	5.9×10^{-3}	89	attach	3.6×10^{-3}
40	biomed	5.8×10^{-3}	90	ultrasound	3.6×10^{-3}
41	walk	5.8×10^{-3}	91	neural	3.6×10^{-3}
42	tomographi	5.7×10^{-3}	92	registr	3.6×10^{-3}
43	load	5.7×10^{-3}	93	base	3.5×10^{-3}
44	limb	5.7×10^{-3}	94	promis	3.5×10^{-3}
45	forc	5.7×10^{-3}	95	modulus	3.5×10^{-3}
46	physiolog	5.6×10^{-3}	96	can	3.5×10^{-3}
47	osteogen	5.3×10^{-3}	97	vessel	3.5×10^{-3}
48	reconstruct	5.3×10^{-3}	98	therapi	3.4×10^{-3}
49	deliveri	5.2×10^{-3}	99	prolifer	3.4×10^{-3}
50	healthi	5.2×10^{-3}	100	measur	3.4×10^{-3}

Table D.70. The list of the top 100 words in the category Engineering, Chemical with RIGs

No.	Word	RIG
1	temperatur	4.1×10^{-2}
2	gas	3.5×10^{-2}
3	catalyst	3.3×10^{-2}
4	adsorpt	3×10^{-2}
5	reactor	2.9×10^{-2}
6	reaction	2.4×10^{-2}
7	patient	2.4×10^{-2}
8	co2	2.2×10^{-2}
9	concentr	2.2×10^{-2}
10	kinet	2.2×10^{-2}
11	process	2.1×10^{-2}
12	fuel	2.1×10^{-2}
13	water	2.1×10^{-2}
14	pressur	2×10^{-2}
15	oil	2×10^{-2}
16	carbon	2×10^{-2}
17	liquid	1.9×10^{-2}
18	particl	1.9×10^{-2}
19	conclus	1.8×10^{-2}
20	mixtur	1.8×10^{-2}
21	remov	1.7×10^{-2}
22	chemic	1.7×10^{-2}
23	aqueous	1.7×10^{-2}
24	catalyt	1.7×10^{-2}
25	combust	1.7×10^{-2}
26	product	1.6×10^{-2}
27	bed	1.6×10^{-2}
28	prepar	1.6×10^{-2}
29	batch	1.5×10^{-2}
30	adsorb	1.5×10^{-2}
31	isotherm	1.5×10^{-2}
32	degre	1.4×10^{-2}
33	surfac	1.4×10^{-2}
34	solut	1.4×10^{-2}
35	experiment	1.3×10^{-2}
36	clinic	1.3×10^{-2}
37	acid	1.3×10^{-2}
38	wastewat	1.2×10^{-2}
39	diseas	1.2×10^{-2}
40	solvent	1.2×10^{-2}
41	heat	1.2×10^{-2}
42	langmuir	1.2×10^{-2}
43	oxid	1.2×10^{-2}
44	flow	1.2×10^{-2}
45	solid	1.2×10^{-2}
46	condit	1.1×10^{-2}
47	convers	1.1×10^{-2}
48	coal	1.1×10^{-2}
49	membran	1.1×10^{-2}
50	industri	1×10^{-2}

No.	Word	RIG
51	associ	1×10^{-2}
52	biomass	9.9×10^{-3}
53	methan	9.8×10^{-3}
54	equilibrium	9.7×10^{-3}
55	composit	9.7×10^{-3}
56	xrd	9.6×10^{-3}
57	obtain	9.6×10^{-3}
58	pore	9.3×10^{-3}
59	hydrogen	9.2×10^{-3}
60	background	9.2×10^{-3}
61	age	9.2×10^{-3}
62	energi	9×10^{-3}
63	year	8.9×10^{-3}
64	permeat	8.9×10^{-3}
65	thermodynam	8.8×10^{-3}
66	sem	8.6×10^{-3}
67	effici	8.5×10^{-3}
68	hydrocarbon	8.5×10^{-3}
69	investig	8×10^{-3}
70	properti	8×10^{-3}
71	ethanol	7.9×10^{-3}
72	gene	7.7×10^{-3}
73	paramet	7.6×10^{-3}
74	thermal	7.6×10^{-3}
75	viscos	7.5×10^{-3}
76	pyrolysi	7.5×10^{-3}
77	steam	7.3×10^{-3}
78	separ	7.3×10^{-3}
79	wast	7.3×10^{-3}
80	spectroscopi	7.3×10^{-3}
81	optimum	7.2×10^{-3}
82	synthes	7.2×10^{-3}
83	biodiesel	7.1×10^{-3}
84	molar	7.1×10^{-3}
85	outcom	7.1×10^{-3}
86	pseudo	7.1×10^{-3}
87	signal	7×10^{-3}
88	flame	6.9×10^{-3}
89	oper	6.9×10^{-3}
90	may	6.9×10^{-3}
91	mpa	6.8×10^{-3}
92	human	6.8×10^{-3}
93	methanol	6.8×10^{-3}
94	air	6.7×10^{-3}
95	feed	6.7×10^{-3}
96	oxygen	6.6×10^{-3}
97	sorption	6.6×10^{-3}
98	phase	6.6×10^{-3}
99	diesel	6.6×10^{-3}
100	feedstock	6.6×10^{-3}

TABLE D.71. The list of the top 100 words in the category Engineering, Civil with RIGs

No.	Word	RIG
1	concret	7.2×10^{-2}
2	load	3.2×10^{-2}
3	paper	3.1×10^{-2}
4	steel	2.7×10^{-2}
5	build	2.6×10^{-2}
6	reinforc	2.5×10^{-2}
7	seismic	2.3×10^{-2}
8	patient	2.3×10^{-2}
9	pavement	2.3×10^{-2}
10	asphalt	2.1×10^{-2}
11	model	2.1×10^{-2}
12	bridg	1.8×10^{-2}
13	strength	1.8×10^{-2}
14	shear	1.7×10^{-2}
15	earthquak	1.6×10^{-2}
16	conclus	1.5×10^{-2}
17	numer	1.5×10^{-2}
18	cell	1.5×10^{-2}
19	cement	1.5×10^{-2}
20	construct	1.5×10^{-2}
21	finit	1.5×10^{-2}
22	stiff	1.5×10^{-2}
23	clinic	1.3×10^{-2}
24	element	1.3×10^{-2}
25	water	1.3×10^{-2}
26	crack	1.3×10^{-2}
27	test	1.2×10^{-2}
28	protein	1.2×10^{-2}
29	compress	1.2×10^{-2}
30	displac	1.2×10^{-2}
31	traffic	1.2×10^{-2}
32	diseas	1.2×10^{-2}
33	design	1.1×10^{-2}
34	structur	1.1×10^{-2}
35	highway	1.1×10^{-2}
36	engin	1.1×10^{-2}
37	flexur	1.1×10^{-2}
38	simul	1×10^{-2}
39	group	1×10^{-2}
40	road	1×10^{-2}
41	beam	1×10^{-2}
42	deform	9.4×10^{-3}
43	gene	9.2×10^{-3}
44	mortar	9.2×10^{-3}
45	soil	9.2×10^{-3}
46	damag	8.9×10^{-3}
47	background	8.9×10^{-3}
48	activ	8.7×10^{-3}
49	column	8.7×10^{-3}
50	river	8.3×10^{-3}
51	buckl	8.2×10^{-3}
52	nonlinear	8×10^{-3}
53	ductil	8×10^{-3}
54	treatment	8×10^{-3}
55	failur	7.9×10^{-3}
56	specimen	7.9×10^{-3}
57	hydraul	7.7×10^{-3}
58	railway	7.4×10^{-3}
59	hydrolog	7.3×10^{-3}
60	capac	7.2×10^{-3}
61	vibrat	7.2×10^{-3}
62	binder	7.1×10^{-3}
63	bend	7.1×10^{-3}
64	associ	6.8×10^{-3}
65	cancer	6.7×10^{-3}
66	elast	6.7×10^{-3}
67	tissu	6.6×10^{-3}
68	flow	6.5×10^{-3}
69	experiment	6.5×10^{-3}
70	molecular	6.4×10^{-3}
71	speci	6.4×10^{-3}
72	acid	6.3×10^{-3}
73	project	6.2×10^{-3}
74	base	6.2×10^{-3}
75	wall	6.2×10^{-3}
76	propos	6×10^{-3}
77	therapi	6×10^{-3}
78	report	5.9×10^{-3}
79	deflect	5.8×10^{-3}
80	electron	5.8×10^{-3}
81	slab	5.8×10^{-3}
82	forc	5.7×10^{-3}
83	tunnel	5.6×10^{-3}
84	blood	5.5×10^{-3}
85	frame	5.5×10^{-3}
86	carri	5.5×10^{-3}
87	stress	5.5×10^{-3}
88	mediat	5.4×10^{-3}
89	paramet	5.4×10^{-3}
90	span	5.4×10^{-3}
91	predict	5.4×10^{-3}
92	suggest	5.3×10^{-3}
93	vehicl	5.3×10^{-3}
94	rail	5.2×10^{-3}
95	tumor	5.2×10^{-3}
96	research	5.2×10^{-3}
97	drug	5.2×10^{-3}
98	infect	5.1×10^{-3}
99	inhibit	5.1×10^{-3}
100	pile	5.1×10^{-3}

TABLE D.72. The list of the top 100 words in the category Engineering, Electrical and Electronic with RIGs

No.	Word	RIG
1	paper	1.1×10^{-1}
2	propos	9.2×10^{-2}
3	power	4.2×10^{-2}
4	algorithm	4.1×10^{-2}
5	studi	4.1×10^{-2}
6	conclus	3.9×10^{-2}
7	simul	3.8×10^{-2}
8	voltag	3.1×10^{-2}
9	patient	2.8×10^{-2}
10	network	2.5×10^{-2}
11	system	2.4×10^{-2}
12	antenna	2.4×10^{-2}
13	circuit	2.4×10^{-2}
14	suggest	2.3×10^{-2}
15	treatment	2.3×10^{-2}
16	associ	2.2×10^{-2}
17	wireless	2.1×10^{-2}
18	base	2.1×10^{-2}
19	protein	1.9×10^{-2}
20	scheme	1.9×10^{-2}
21	frequenc	1.8×10^{-2}
22	group	1.8×10^{-2}
23	output	1.7×10^{-2}
24	design	1.7×10^{-2}
25	bandwidth	1.7×10^{-2}
26	clinic	1.7×10^{-2}
27	ghz	1.6×10^{-2}
28	signific	1.6×10^{-2}
29	devic	1.6×10^{-2}
30	age	1.6×10^{-2}
31	diseas	1.6×10^{-2}
32	oper	1.6×10^{-2}
33	nois	1.6×10^{-2}
34	gene	1.5×10^{-2}
35	implement	1.5×10^{-2}
36	cmos	1.5×10^{-2}
37	acid	1.4×10^{-2}
38	perform	1.4×10^{-2}
39	speci	1.4×10^{-2}
40	achiev	1.4×10^{-2}
41	examin	1.4×10^{-2}
42	found	1.4×10^{-2}
43	filter	1.4×10^{-2}
44	communic	1.3×10^{-2}
45	sensor	1.3×10^{-2}
46	background	1.3×10^{-2}
47	may	1.2×10^{-2}
48	signal	1.2×10^{-2}
49	switch	1.2×10^{-2}
50	problem	1.2×10^{-2}
51	optim	1.2×10^{-2}
52	user	1.2×10^{-2}
53	year	1.2×10^{-2}
54	electr	1.2×10^{-2}
55	indic	1.1×10^{-2}
56	applic	1.1×10^{-2}
57	assess	1.1×10^{-2}
58	channel	1.1×10^{-2}
59	bit	1.1×10^{-2}
60	cell	1.1×10^{-2}
61	input	1.1×10^{-2}
62	effici	1.1×10^{-2}
63	activ	1.1×10^{-2}
64	technolog	1.1×10^{-2}
65	error	1.1×10^{-2}
66	popul	1.1×10^{-2}
67	evid	1×10^{-2}
68	outcom	1×10^{-2}
69	role	1×10^{-2}
70	relat	1×10^{-2}
71	risk	1×10^{-2}
72	reaction	1×10^{-2}
73	reveal	1×10^{-2}
74	follow	9.9×10^{-3}
75	concentr	9.9×10^{-3}
76	inhibit	9.8×10^{-3}
77	convert	9.8×10^{-3}
78	comput	9.7×10^{-3}
79	month	9.6×10^{-3}
80	molecular	9.5×10^{-3}
81	mediat	9.2×10^{-3}
82	techniqu	9.2×10^{-3}
83	can	9.2×10^{-3}
84	observ	9.1×10^{-3}
85	grid	9×10^{-3}
86	treat	8.8×10^{-3}
87	cancer	8.6×10^{-3}
88	induc	8.6×10^{-3}
89	present	8.5×10^{-3}
90	decreas	8.4×10^{-3}
91	express	8.4×10^{-3}
92	report	8.4×10^{-3}
93	infect	8.3×10^{-3}
94	hardwar	8.2×10^{-3}
95	verifi	8.2×10^{-3}
96	increas	8.2×10^{-3}
97	pathway	8.2×10^{-3}
98	therapi	8.2×10^{-3}
99	radio	8×10^{-3}
100	transmiss	8×10^{-3}

TABLE D.73. The list of the top 100 words in the category Engineering, Environmental with RIGs

No.	Word	RIG	No.	Word	RIG
1	water	4.4×10^{-2}	51	bed	7.4×10^{-3}
2	wastewat	4.2×10^{-2}	52	sediment	7.4×10^{-3}
3	remov	3.5×10^{-2}	53	sorption	7.4×10^{-3}
4	concentr	3.1×10^{-2}	54	condit	6.8×10^{-3}
5	wast	3×10^{-2}	55	groundwat	6.8×10^{-3}
6	environment	3×10^{-2}	56	reduct	6.8×10^{-3}
7	pollut	2.7×10^{-2}	57	flow	6.7×10^{-3}
8	sludg	2×10^{-2}	58	biodegrad	6.7×10^{-3}
9	patient	2×10^{-2}	59	oxygen	6.6×10^{-3}
10	effluent	1.8×10^{-2}	60	ozon	6.6×10^{-3}
11	carbon	1.8×10^{-2}	61	toxic	6.6×10^{-3}
12	adsorpt	1.8×10^{-2}	62	temperatur	6.3×10^{-3}
13	contamin	1.6×10^{-2}	63	express	6.2×10^{-3}
14	reactor	1.4×10^{-2}	64	capac	6.2×10^{-3}
15	co2	1.3×10^{-2}	65	asphalt	6.2×10^{-3}
16	product	1.3×10^{-2}	66	sustain	6.1×10^{-3}
17	chemic	1.3×10^{-2}	67	metal	6×10^{-3}
18	recycl	1.3×10^{-2}	68	impact	6×10^{-3}
19	batch	1.2×10^{-2}	69	reaction	5.8×10^{-3}
20	organ	1.2×10^{-2}	70	aquat	5.8×10^{-3}
21	oxid	1.2×10^{-2}	71	greenhous	5.8×10^{-3}
22	gas	1.2×10^{-2}	72	isotherm	5.8×10^{-3}
23	clinic	1.2×10^{-2}	73	fuel	5.8×10^{-3}
24	process	1.2×10^{-2}	74	amount	5.7×10^{-3}
25	adsorb	1.1×10^{-2}	75	pseudo	5.7×10^{-3}
26	cod	1.1×10^{-2}	76	pavement	5.7×10^{-3}
27	soil	1.1×10^{-2}	77	matter	5.6×10^{-3}
28	plant	1.1×10^{-2}	78	urban	5.6×10^{-3}
29	municip	1.1×10^{-2}	79	signal	5.5×10^{-3}
30	anaerob	1.1×10^{-2}	80	tio2	5.4×10^{-3}
31	emiss	1.1×10^{-2}	81	catalyt	5.3×10^{-3}
32	conclus	1.1×10^{-2}	82	therapi	5.3×10^{-3}
33	river	1×10^{-2}	83	miner	5.2×10^{-3}
34	kinet	1×10^{-2}	84	dioxid	5.2×10^{-3}
35	catalyst	9.7×10^{-3}	85	cycl	5.2×10^{-3}
36	degrad	9.6×10^{-3}	86	sourc	5.1×10^{-3}
37	dissolv	9.2×10^{-3}	87	phosphorus	5.1×10^{-3}
38	industri	9×10^{-3}	88	reus	4.9×10^{-3}
39	hydraul	8.7×10^{-3}	89	bioreactor	4.9×10^{-3}
40	nitrogen	8.6×10^{-3}	90	indic	4.9×10^{-3}
41	diseas	8.5×10^{-3}	91	ammonia	4.9×10^{-3}
42	biomass	8.4×10^{-3}	92	wetland	4.8×10^{-3}
43	aqueous	8.3×10^{-3}	93	particul	4.8×10^{-3}
44	surfac	8.3×10^{-3}	94	methan	4.8×10^{-3}
45	effici	8.1×10^{-3}	95	leach	4.8×10^{-3}
46	air	8×10^{-3}	96	age	4.7×10^{-3}
47	photocatalyt	8×10^{-3}	97	protein	4.6×10^{-3}
48	langmuir	7.6×10^{-3}	98	demand	4.6×10^{-3}
49	microbi	7.6×10^{-3}	99	iron	4.6×10^{-3}
50	solid	7.5×10^{-3}	100	photocatalyst	4.6×10^{-3}

TABLE D.74. The list of the top 100 words in the category Engineering, Geological with RIGs

No.	Word	RIG
1	soil	1.2×10^{-1}
2	seismic	6.5×10^{-2}
3	rock	6.3×10^{-2}
4	shear	5.1×10^{-2}
5	earthquak	5×10^{-2}
6	geotechn	4.8×10^{-2}
7	clay	3.9×10^{-2}
8	pile	3.7×10^{-2}
9	stress	3.6×10^{-2}
10	load	3.5×10^{-2}
11	sand	3.3×10^{-2}
12	displac	3.3×10^{-2}
13	ground	3×10^{-2}
14	slope	3×10^{-2}
15	deform	2.8×10^{-2}
16	excav	2.7×10^{-2}
17	landslid	2.7×10^{-2}
18	strength	2.7×10^{-2}
19	foundat	2.6×10^{-2}
20	numer	2.5×10^{-2}
21	test	2.5×10^{-2}
22	stiff	2.3×10^{-2}
23	failur	2.3×10^{-2}
24	settlement	2.2×10^{-2}
25	suction	2.2×10^{-2}
26	finit	2.1×10^{-2}
27	geolog	2.1×10^{-2}
28	compress	2×10^{-2}
29	element	2×10^{-2}
30	concret	1.9×10^{-2}
31	pressur	1.9×10^{-2}
32	pore	1.9×10^{-2}
33	patient	1.8×10^{-2}
34	unsatur	1.8×10^{-2}
35	underground	1.8×10^{-2}
36	tunnel	1.7×10^{-2}
37	model	1.7×10^{-2}
38	strain	1.7×10^{-2}
39	paper	1.6×10^{-2}
40	engin	1.6×10^{-2}
41	consolid	1.5×10^{-2}
42	reinforc	1.5×10^{-2}
43	elast	1.5×10^{-2}
44	depth	1.5×10^{-2}
45	horizont	1.5×10^{-2}
46	satur	1.4×10^{-2}
47	vertic	1.3×10^{-2}
48	water	1.3×10^{-2}
49	hydraul	1.3×10^{-2}
50	plastic	1.3×10^{-2}

No.	Word	RIG
51	laboratori	1.3×10^{-2}
52	damag	1.3×10^{-2}
53	conclus	1.2×10^{-2}
54	friction	1.1×10^{-2}
55	paramet	1.1×10^{-2}
56	behaviour	1.1×10^{-2}
57	wall	1.1×10^{-2}
58	build	1.1×10^{-2}
59	clinic	1.1×10^{-2}
60	shallow	1×10^{-2}
61	protein	9.8×10^{-3}
62	cell	9.7×10^{-3}
63	modulus	9.3×10^{-3}
64	diseas	9.3×10^{-3}
65	construct	9.3×10^{-3}
66	drill	9.2×10^{-3}
67	fractur	9.2×10^{-3}
68	bear	9×10^{-3}
69	collaps	8.9×10^{-3}
70	permeabl	8.9×10^{-3}
71	borehol	8.7×10^{-3}
72	dam	8.6×10^{-3}
73	crack	8.5×10^{-3}
74	motion	8.4×10^{-3}
75	specimen	8.4×10^{-3}
76	behavior	8.3×10^{-3}
77	curv	8.3×10^{-3}
78	slide	8.1×10^{-3}
79	instal	8×10^{-3}
80	mine	8×10^{-3}
81	nonlinear	8×10^{-3}
82	simul	7.8×10^{-3}
83	treatment	7.8×10^{-3}
84	group	7.7×10^{-3}
85	soft	7.4×10^{-3}
86	background	7.4×10^{-3}
87	column	7.3×10^{-3}
88	gene	7.2×10^{-3}
89	condit	7.1×10^{-3}
90	static	7×10^{-3}
91	influenc	7×10^{-3}
92	cyclic	7×10^{-3}
93	mechan	7×10^{-3}
94	activ	6.8×10^{-3}
95	penetr	6.7×10^{-3}
96	zone	6.7×10^{-3}
97	damp	6.7×10^{-3}
98	predict	6.5×10^{-3}
99	uniaxi	6.5×10^{-3}
100	parametr	6.5×10^{-3}

Table D.75. The list of the top 100 words in the category Engineering, Industrial with RIGs

No.	Word	RIG	No.	Word	RIG
1	paper	4.7×10^{-2}	51	develop	5.6×10^{-3}
2	manufactur	2.7×10^{-2}	52	approach	5.6×10^{-3}
3	industri	2.1×10^{-2}	53	tool	5.3×10^{-3}
4	propos	1.7×10^{-2}	54	report	5.3×10^{-3}
5	design	1.6×10^{-2}	55	innov	5.2×10^{-3}
6	patient	1.6×10^{-2}	56	safeti	5.2×10^{-3}
7	compani	1.5×10^{-2}	57	algorithm	5.1×10^{-3}
8	machin	1.4×10^{-2}	58	age	5.1×10^{-3}
9	cost	1.3×10^{-2}	59	background	5.1×10^{-3}
10	product	1.2×10^{-2}	60	implement	5×10^{-3}
11	model	1.2×10^{-2}	61	integr	5×10^{-3}
12	engin	1.2×10^{-2}	62	associ	4.9×10^{-3}
13	custom	1.1×10^{-2}	63	automot	4.8×10^{-3}
14	process	1.1×10^{-2}	64	order	4.8×10^{-3}
15	conclus	1.1×10^{-2}	65	enterpris	4.7×10^{-3}
16	protein	1.1×10^{-2}	66	therapi	4.6×10^{-3}
17	problem	1.1×10^{-2}	67	inhibit	4.5×10^{-3}
18	decis	1.1×10^{-2}	68	tissu	4.5×10^{-3}
19	cell	1×10^{-2}	69	perform	4.5×10^{-3}
20	oper	9.8×10^{-3}	70	tumor	4.4×10^{-3}
21	clinic	9.7×10^{-3}	71	induc	4.4×10^{-3}
22	diseas	9.2×10^{-3}	72	profit	4.3×10^{-3}
23	manag	9.1×10^{-3}	73	load	4.3×10^{-3}
24	demand	8.9×10^{-3}	74	requir	4.3×10^{-3}
25	simul	8.9×10^{-3}	75	ergonom	4.2×10^{-3}
26	treatment	8.8×10^{-3}	76	assay	4.2×10^{-3}
27	system	8.2×10^{-3}	77	job	4.2×10^{-3}
28	suppli	8.1×10^{-3}	78	applic	4.2×10^{-3}
29	gene	8.1×10^{-3}	79	infect	4.2×10^{-3}
30	supplier	8×10^{-3}	80	drug	4.2×10^{-3}
31	speci	7.6×10^{-3}	81	project	4.1×10^{-3}
32	schedul	7.5×10^{-3}	82	oxid	4.1×10^{-3}
33	base	7.4×10^{-3}	83	receptor	4.1×10^{-3}
34	technolog	7.2×10^{-3}	84	inventori	4.1×10^{-3}
35	optim	7×10^{-3}	85	exhibit	4.1×10^{-3}
36	group	6.8×10^{-3}	86	practic	4.1×10^{-3}
37	research	6.7×10^{-3}	87	pathway	4.1×10^{-3}
38	acid	6.6×10^{-3}	88	reaction	4×10^{-3}
39	heurist	6.3×10^{-3}	89	make	4×10^{-3}
40	firm	6.3×10^{-3}	90	popul	4×10^{-3}
41	concentr	6.3×10^{-3}	91	month	3.9×10^{-3}
42	molecular	6.2×10^{-3}	92	market	3.9×10^{-3}
43	plan	6.2×10^{-3}	93	integ	3.9×10^{-3}
44	busi	6.1×10^{-3}	94	illustr	3.8×10^{-3}
45	solv	6.1×10^{-3}	95	can	3.8×10^{-3}
46	observ	6×10^{-3}	96	softwar	3.8×10^{-3}
47	express	5.9×10^{-3}	97	vitro	3.8×10^{-3}
48	methodolog	5.8×10^{-3}	98	bind	3.7×10^{-3}
49	cancer	5.8×10^{-3}	99	molecul	3.7×10^{-3}
50	exampl	5.7×10^{-3}	100	signific	3.7×10^{-3}

TABLE D.76. The list of the top 100 words in the category Engineering, Manufacturing with RIGs

No.	Word	RIG	No.	Word	RIG
1	manufactur	4.6×10^{-2}	51	strength	6.1×10^{-3}
2	machin	3.9×10^{-2}	52	treatment	6.1×10^{-3}
3	paper	3×10^{-2}	53	popul	6.1×10^{-3}
4	process	2.7×10^{-2}	54	custom	5.9×10^{-3}
5	cut	1.7×10^{-2}	55	cancer	5.9×10^{-3}
6	patient	1.7×10^{-2}	56	compani	5.7×10^{-3}
7	weld	1.6×10^{-2}	57	aluminum	5.6×10^{-3}
8	conclus	1.6×10^{-2}	58	deform	5.6×10^{-3}
9	industri	1.5×10^{-2}	59	supplier	5.4×10^{-3}
10	steel	1.5×10^{-2}	60	signific	5.4×10^{-3}
11	tool	1.4×10^{-2}	61	month	5.4×10^{-3}
12	alloy	1.3×10^{-2}	62	carbid	5.3×10^{-3}
13	speed	1.2×10^{-2}	63	experiment	5.3×10^{-3}
14	design	1.2×10^{-2}	64	automot	5.2×10^{-3}
15	materi	1.2×10^{-2}	65	softwar	4.9×10^{-3}
16	wear	1.1×10^{-2}	66	element	4.8×10^{-3}
17	clinic	1.1×10^{-2}	67	therapi	4.8×10^{-3}
18	protein	1×10^{-2}	68	suppli	4.7×10^{-3}
19	associ	1×10^{-2}	69	infect	4.6×10^{-3}
20	product	1×10^{-2}	70	qualiti	4.5×10^{-3}
21	cell	1×10^{-2}	71	pathway	4.4×10^{-3}
22	diseas	1×10^{-2}	72	research	4.3×10^{-3}
23	background	9.6×10^{-3}	73	part	4.3×10^{-3}
24	gene	8.6×10^{-3}	74	evid	4.3×10^{-3}
25	group	8.6×10^{-3}	75	molecular	4.3×10^{-3}
26	mill	8.4×10^{-3}	76	blood	4.2×10^{-3}
27	suggest	8.4×10^{-3}	77	tumor	4.2×10^{-3}
28	paramet	8.4×10^{-3}	78	mediat	4.2×10^{-3}
29	activ	8.3×10^{-3}	79	receptor	4.2×10^{-3}
30	rough	8.2×10^{-3}	80	particip	4.2×10^{-3}
31	technolog	8.2×10^{-3}	81	adult	4.2×10^{-3}
32	optim	7.9×10^{-3}	82	base	4.2×10^{-3}
33	age	7.8×10^{-3}	83	inhibit	4.1×10^{-3}
34	tensil	7.7×10^{-3}	84	fabric	4×10^{-3}
35	simul	7.4×10^{-3}	85	assay	4×10^{-3}
36	speci	7.3×10^{-3}	86	day	4×10^{-3}
37	model	7×10^{-3}	87	demand	4×10^{-3}
38	friction	6.9×10^{-3}	88	problem	3.9×10^{-3}
39	express	6.9×10^{-3}	89	heat	3.9×10^{-3}
40	report	6.8×10^{-3}	90	bind	3.9×10^{-3}
41	may	6.7×10^{-3}	91	male	3.9×10^{-3}
42	hard	6.6×10^{-3}	92	sheet	3.9×10^{-3}
43	cost	6.6×10^{-3}	93	metal	3.9×10^{-3}
44	propos	6.6×10^{-3}	94	induc	3.9×10^{-3}
45	microstructur	6.5×10^{-3}	95	week	3.8×10^{-3}
46	finit	6.4×10^{-3}	96	drug	3.8×10^{-3}
47	engin	6.3×10^{-3}	97	outcom	3.8×10^{-3}
48	year	6.2×10^{-3}	98	acid	3.7×10^{-3}
49	forc	6.1×10^{-3}	99	laser	3.7×10^{-3}
50	surfac	6.1×10^{-3}	100	carri	3.7×10^{-3}

Table D.77. The list of the top 100 words in the category Engineering, Marine with RIGs

No.	Word	RIG	No.	Word	RIG
1	ship	1.1×10^{-1}	51	cylind	9.4×10^{-3}
2	underwat	7.7×10^{-2}	52	reynold	9.2×10^{-3}
3	sea	6.1×10^{-2}	53	treatment	9.1×10^{-3}
4	offshor	5.7×10^{-2}	54	carri	9×10^{-3}
5	wave	4.9×10^{-2}	55	diseas	8.9×10^{-3}
6	hull	4.9×10^{-2}	56	system	8.6×10^{-3}
7	paper	3.6×10^{-2}	57	seafloor	8.3×10^{-3}
8	ocean	3.3×10^{-2}	58	fatigu	8.3×10^{-3}
9	vessel	3×10^{-2}	59	navier	8.3×10^{-3}
10	hydrodynam	3×10^{-2}	60	protein	8.1×10^{-3}
11	numer	2.8×10^{-2}	61	lift	7.9×10^{-3}
12	marin	2.7×10^{-2}	62	age	7.8×10^{-3}
13	simul	2.6×10^{-2}	63	coastal	7.8×10^{-3}
14	motion	2.4×10^{-2}	64	ice	7.8×10^{-3}
15	water	2.4×10^{-2}	65	blade	7.8×10^{-3}
16	float	2.2×10^{-2}	66	pressur	7.8×10^{-3}
17	model	2×10^{-2}	67	submerg	7.7×10^{-3}
18	tank	2×10^{-2}	68	valid	7.6×10^{-3}
19	load	1.9×10^{-2}	69	vertic	7.5×10^{-3}
20	cfd	1.9×10^{-2}	70	height	7.5×10^{-3}
21	turbin	1.9×10^{-2}	71	stoke	7.4×10^{-3}
22	oper	1.9×10^{-2}	72	acid	7.4×10^{-3}
23	acoust	1.9×10^{-2}	73	suggest	7.4×10^{-3}
24	riser	1.9×10^{-2}	74	bottom	7.3×10^{-3}
25	moor	1.9×10^{-2}	75	deep	7.2×10^{-3}
26	design	1.7×10^{-2}	76	gene	6.9×10^{-3}
27	vehicl	1.7×10^{-2}	77	damp	6.8×10^{-3}
28	patient	1.6×10^{-2}	78	perform	6.8×10^{-3}
29	propuls	1.6×10^{-2}	79	dynam	6.8×10^{-3}
30	forc	1.6×10^{-2}	80	oil	6.7×10^{-3}
31	navig	1.5×10^{-2}	81	shallow	6.5×10^{-3}
32	drag	1.4×10^{-2}	82	predict	6.5×10^{-3}
33	speed	1.4×10^{-2}	83	comput	6.5×10^{-3}
34	autonom	1.4×10^{-2}	84	condit	6.4×10^{-3}
35	wind	1.4×10^{-2}	85	group	6.3×10^{-3}
36	instal	1.3×10^{-2}	86	estim	6.3×10^{-3}
37	vortex	1.2×10^{-2}	87	turbul	6×10^{-3}
38	flow	1.2×10^{-2}	88	equip	6×10^{-3}
39	veloc	1.1×10^{-2}	89	activ	5.9×10^{-3}
40	depth	1.1×10^{-2}	90	experiment	5.9×10^{-3}
41	present	1.1×10^{-2}	91	appli	5.9×10^{-3}
42	conclus	1.1×10^{-2}	92	environ	5.8×10^{-3}
43	cell	1×10^{-2}	93	molecular	5.8×10^{-3}
44	fluid	1×10^{-2}	94	unsteadi	5.8×10^{-3}
45	engin	1×10^{-2}	95	vibrat	5.7×10^{-3}
46	platform	1×10^{-2}	96	popul	5.6×10^{-3}
47	clinic	9.8×10^{-3}	97	horizont	5.5×10^{-3}
48	nonlinear	9.6×10^{-3}	98	angl	5.4×10^{-3}
49	maritim	9.6×10^{-3}	99	background	5.4×10^{-3}
50	finit	9.6×10^{-3}	100	consid	5.3×10^{-3}

Table D.78. The list of the top 100 words in the category Engineering, Mechanical with RIGs

No.	Word	RIG
1	paper	3.4×10^{-2}
2	simul	2.6×10^{-2}
3	flow	2.6×10^{-2}
4	heat	2.4×10^{-2}
5	numer	2.4×10^{-2}
6	patient	2.4×10^{-2}
7	load	2.4×10^{-2}
8	turbin	2.2×10^{-2}
9	conclus	2.1×10^{-2}
10	vibrat	2×10^{-2}
11	finit	1.9×10^{-2}
12	design	1.9×10^{-2}
13	pressur	1.9×10^{-2}
14	experiment	1.7×10^{-2}
15	model	1.6×10^{-2}
16	steel	1.6×10^{-2}
17	engin	1.6×10^{-2}
18	speed	1.5×10^{-2}
19	fluid	1.5×10^{-2}
20	clinic	1.5×10^{-2}
21	veloc	1.5×10^{-2}
22	group	1.4×10^{-2}
23	fatigu	1.4×10^{-2}
24	reynold	1.4×10^{-2}
25	cfd	1.4×10^{-2}
26	blade	1.4×10^{-2}
27	friction	1.4×10^{-2}
28	pipe	1.4×10^{-2}
29	forc	1.4×10^{-2}
30	crack	1.4×10^{-2}
31	protein	1.3×10^{-2}
32	diseas	1.3×10^{-2}
33	cell	1.3×10^{-2}
34	activ	1.2×10^{-2}
35	suggest	1.2×10^{-2}
36	associ	1.2×10^{-2}
37	element	1.1×10^{-2}
38	gene	1.1×10^{-2}
39	dynam	1.1×10^{-2}
40	paramet	1.1×10^{-2}
41	background	1.1×10^{-2}
42	inlet	1.1×10^{-2}
43	turbul	1×10^{-2}
44	age	1×10^{-2}
45	cool	9.4×10^{-3}
46	stress	9.4×10^{-3}
47	thermal	9.2×10^{-3}
48	treatment	9.2×10^{-3}
49	rotor	9.1×10^{-3}
50	motion	9×10^{-3}
51	equat	8.8×10^{-3}
52	stiff	8.5×10^{-3}
53	oper	8.5×10^{-3}
54	system	8.5×10^{-3}
55	condit	8.4×10^{-3}
56	machin	8.4×10^{-3}
57	report	8.3×10^{-3}
58	popul	8.2×10^{-3}
59	wear	8×10^{-3}
60	damp	7.9×10^{-3}
61	nonlinear	7.8×10^{-3}
62	cancer	7.8×10^{-3}
63	weld	7.6×10^{-3}
64	unsteadi	7.6×10^{-3}
65	aerodynam	7.6×10^{-3}
66	temperatur	7.6×10^{-3}
67	wall	7.6×10^{-3}
68	deform	7.4×10^{-3}
69	plate	7.4×10^{-3}
70	air	7.3×10^{-3}
71	manufactur	7.2×10^{-3}
72	hydraul	7.2×10^{-3}
73	cylind	7.2×10^{-3}
74	transfer	7.1×10^{-3}
75	displac	7×10^{-3}
76	fuel	6.9×10^{-3}
77	ansi	6.9×10^{-3}
78	axial	6.9×10^{-3}
79	speci	6.9×10^{-3}
80	outcom	6.8×10^{-3}
81	mediat	6.8×10^{-3}
82	year	6.8×10^{-3}
83	acid	6.8×10^{-3}
84	solv	6.8×10^{-3}
85	particip	6.7×10^{-3}
86	gas	6.5×10^{-3}
87	may	6.5×10^{-3}
88	base	6.5×10^{-3}
89	therapi	6.5×10^{-3}
90	express	6.4×10^{-3}
91	actuat	6.4×10^{-3}
92	month	6.4×10^{-3}
93	boundari	6.4×10^{-3}
94	day	6.2×10^{-3}
95	infect	6.2×10^{-3}
96	coeffici	6.1×10^{-3}
97	geometri	6.1×10^{-3}
98	combust	6.1×10^{-3}
99	carri	6×10^{-3}
100	evid	6×10^{-3}

TABLE D.79. The list of the top 100 words in the category Engineering, Multidisciplinary with RIGs

No.	Word	RIG	No.	Word	RIG
1	paper	4.2×10^{-2}	51	month	4.8×10^{-3}
2	engin	1.9×10^{-2}	52	molecular	4.7×10^{-3}
3	patient	1.8×10^{-2}	53	express	4.7×10^{-3}
4	propos	1.8×10^{-2}	54	inhibit	4.6×10^{-3}
5	problem	1.7×10^{-2}	55	load	4.6×10^{-3}
6	conclus	1.7×10^{-2}	56	infect	4.5×10^{-3}
7	numer	1.3×10^{-2}	57	assay	4.5×10^{-3}
8	solv	1.3×10^{-2}	58	tissu	4.5×10^{-3}
9	protein	1.1×10^{-2}	59	drug	4.4×10^{-3}
10	cell	1.1×10^{-2}	60	adult	4.3×10^{-3}
11	algorithm	1.1×10^{-2}	61	receptor	4.3×10^{-3}
12	clinic	1.1×10^{-2}	62	accuraci	4.3×10^{-3}
13	simul	1×10^{-2}	63	matlab	4.3×10^{-3}
14	associ	1×10^{-2}	64	tumor	4.2×10^{-3}
15	diseas	9.1×10^{-3}	65	flame	4.2×10^{-3}
16	design	9×10^{-3}	66	industri	4.2×10^{-3}
17	student	8.8×10^{-3}	67	fuzzi	4.1×10^{-3}
18	gene	8.8×10^{-3}	68	potenti	4.1×10^{-3}
19	finit	8.7×10^{-3}	69	present	4×10^{-3}
20	suggest	8.7×10^{-3}	70	blood	4×10^{-3}
21	treatment	8.6×10^{-3}	71	softwar	4×10^{-3}
22	age	8.6×10^{-3}	72	order	3.9×10^{-3}
23	base	8.3×10^{-3}	73	evid	3.9×10^{-3}
24	system	8.2×10^{-3}	74	popul	3.9×10^{-3}
25	exampl	7.9×10^{-3}	75	bind	3.9×10^{-3}
26	signific	7.2×10^{-3}	76	vitro	3.9×10^{-3}
27	group	7.1×10^{-3}	77	solut	3.8×10^{-3}
28	background	6.6×10^{-3}	78	mice	3.8×10^{-3}
29	comput	5.9×10^{-3}	79	therapeut	3.7×10^{-3}
30	activ	5.9×10^{-3}	80	pathway	3.6×10^{-3}
31	cancer	5.9×10^{-3}	81	cours	3.6×10^{-3}
32	machin	5.9×10^{-3}	82	beta	3.5×10^{-3}
33	report	5.8×10^{-3}	83	observ	3.5×10^{-3}
34	nonlinear	5.7×10^{-3}	84	day	3.5×10^{-3}
35	may	5.7×10^{-3}	85	learn	3.4×10^{-3}
36	studi	5.6×10^{-3}	86	male	3.4×10^{-3}
37	acid	5.5×10^{-3}	87	process	3.4×10^{-3}
38	model	5.5×10^{-3}	88	molecul	3.4×10^{-3}
39	element	5.3×10^{-3}	89	undergradu	3.4×10^{-3}
40	induc	5.3×10^{-3}	90	dna	3.4×10^{-3}
41	optim	5.3×10^{-3}	91	dose	3.4×10^{-3}
42	therapi	5.2×10^{-3}	92	remain	3.4×10^{-3}
43	implement	5.2×10^{-3}	93	verifi	3.4×10^{-3}
44	mediat	5×10^{-3}	94	surviv	3.3×10^{-3}
45	equat	5×10^{-3}	95	faculti	3.3×10^{-3}
46	technolog	5×10^{-3}	96	disord	3.3×10^{-3}
47	mathemat	4.9×10^{-3}	97	examin	3.3×10^{-3}
48	speed	4.8×10^{-3}	98	underw	3.3×10^{-3}
49	paramet	4.8×10^{-3}	99	retrospect	3.2×10^{-3}
50	speci	4.8×10^{-3}	100	discret	3.2×10^{-3}

TABLE D.80. The list of the top 100 words in the category Engineering, Ocean with RIGs

No.	Word	RIG
1	wave	9.1×10^{-2}
2	offshor	8.6×10^{-2}
3	sea	6.1×10^{-2}
4	ship	5.2×10^{-2}
5	hydrodynam	4.4×10^{-2}
6	underwat	4.3×10^{-2}
7	numer	4.1×10^{-2}
8	riser	3.8×10^{-2}
9	ocean	3.8×10^{-2}
10	float	3.6×10^{-2}
11	water	3.6×10^{-2}
12	moor	3.2×10^{-2}
13	motion	3.1×10^{-2}
14	paper	3.1×10^{-2}
15	simul	2.9×10^{-2}
16	model	2.9×10^{-2}
17	load	2.8×10^{-2}
18	wind	2.8×10^{-2}
19	hull	2.7×10^{-2}
20	forc	2.4×10^{-2}
21	depth	2.3×10^{-2}
22	marin	2×10^{-2}
23	veloc	1.9×10^{-2}
24	pipe	1.8×10^{-2}
25	drag	1.8×10^{-2}
26	vessel	1.7×10^{-2}
27	coastal	1.7×10^{-2}
28	patient	1.7×10^{-2}
29	horizont	1.7×10^{-2}
30	height	1.6×10^{-2}
31	flow	1.6×10^{-2}
32	tank	1.6×10^{-2}
33	instal	1.6×10^{-2}
34	vertic	1.5×10^{-2}
35	pipelin	1.5×10^{-2}
36	turbin	1.5×10^{-2}
37	submerg	1.4×10^{-2}
38	cfd	1.4×10^{-2}
39	vortex	1.4×10^{-2}
40	fluid	1.4×10^{-2}
41	finit	1.3×10^{-2}
42	nonlinear	1.3×10^{-2}
43	acoust	1.3×10^{-2}
44	design	1.3×10^{-2}
45	fatigu	1.2×10^{-2}
46	pressur	1.2×10^{-2}
47	shallow	1.1×10^{-2}
48	platform	1.1×10^{-2}
49	bottom	1.1×10^{-2}
50	predict	1.1×10^{-2}

No.	Word	RIG
51	present	1.1×10^{-2}
52	reynold	1.1×10^{-2}
53	equat	1×10^{-2}
54	cylind	1×10^{-2}
55	conclus	1×10^{-2}
56	tension	1×10^{-2}
57	condit	1×10^{-2}
58	cell	1×10^{-2}
59	valid	1×10^{-2}
60	oper	9.9×10^{-3}
61	clinic	9.9×10^{-3}
62	beach	9.6×10^{-3}
63	navier	9.6×10^{-3}
64	dynam	9.6×10^{-3}
65	stoke	9.4×10^{-3}
66	damp	9.3×10^{-3}
67	speed	9.2×10^{-3}
68	deep	9.1×10^{-3}
69	diseas	9.1×10^{-3}
70	coast	9.1×10^{-3}
71	test	8.9×10^{-3}
72	carri	8.8×10^{-3}
73	propag	8.4×10^{-3}
74	oil	8.4×10^{-3}
75	turbul	8.3×10^{-3}
76	protein	8.3×10^{-3}
77	ice	8.2×10^{-3}
78	drill	8.2×10^{-3}
79	estim	8.1×10^{-3}
80	consid	8.1×10^{-3}
81	appli	8.1×10^{-3}
82	shore	7.8×10^{-3}
83	boundari	7.8×10^{-3}
84	comparison	7.7×10^{-3}
85	storm	7.6×10^{-3}
86	bend	7.5×10^{-3}
87	age	7.5×10^{-3}
88	activ	7.4×10^{-3}
89	acid	7.2×10^{-3}
90	frequenc	7.2×10^{-3}
91	experiment	7.1×10^{-3}
92	gene	7×10^{-3}
93	treatment	6.9×10^{-3}
94	tidal	6.8×10^{-3}
95	group	6.6×10^{-3}
96	lift	6.6×10^{-3}
97	simplifi	6.3×10^{-3}
98	accur	6.3×10^{-3}
99	calcul	6.2×10^{-3}
100	steel	6.2×10^{-3}

Table D.81. The list of the top 100 words in the category Engineering, Petroleum with RIGs

No.	Word	RIG
1	reservoir	1.4×10^{-1}
2	oil	1.3×10^{-1}
3	gas	6.2×10^{-2}
4	permeabl	4.5×10^{-2}
5	drill	4.5×10^{-2}
6	rock	4.2×10^{-2}
7	pressur	4.1×10^{-2}
8	fluid	3.7×10^{-2}
9	shale	3.3×10^{-2}
10	hydrocarbon	2.9×10^{-2}
11	fractur	2.8×10^{-2}
12	petroleum	2.6×10^{-2}
13	sandston	2.3×10^{-2}
14	format	2.2×10^{-2}
15	geolog	2.2×10^{-2}
16	flow	2.2×10^{-2}
17	recoveri	2.1×10^{-2}
18	viscos	2×10^{-2}
19	product	1.9×10^{-2}
20	basin	1.8×10^{-2}
21	inject	1.8×10^{-2}
22	pore	1.6×10^{-2}
23	patient	1.6×10^{-2}
24	water	1.6×10^{-2}
25	crude	1.6×10^{-2}
26	poros	1.5×10^{-2}
27	well	1.5×10^{-2}
28	flood	1.4×10^{-2}
29	horizont	1.3×10^{-2}
30	borehol	1.2×10^{-2}
31	sand	1.2×10^{-2}
32	satur	1.1×10^{-2}
33	conclus	1.1×10^{-2}
34	hydraul	1.1×10^{-2}
35	co2	1.1×10^{-2}
36	field	1×10^{-2}
37	coal	1×10^{-2}
38	log	9.7×10^{-3}
39	clinic	9.7×10^{-3}
40	temperatur	9.3×10^{-3}
41	diseas	9.2×10^{-3}
42	carbon	8.9×10^{-3}
43	offshor	8.9×10^{-3}
44	pipe	8.5×10^{-3}
45	sedimentari	8.1×10^{-3}
46	litholog	8×10^{-3}
47	zone	8×10^{-3}
48	protein	8×10^{-3}
49	faci	7.9×10^{-3}
50	miner	7.5×10^{-3}

No.	Word	RIG
51	heavi	7.5×10^{-3}
52	cell	7.5×10^{-3}
53	porous	7.4×10^{-3}
54	core	7×10^{-3}
55	background	7×10^{-3}
56	displac	6.7×10^{-3}
57	drainag	6.7×10^{-3}
58	human	6.5×10^{-3}
59	seismic	6.5×10^{-3}
60	surfact	6.5×10^{-3}
61	properti	6.3×10^{-3}
62	gene	6.2×10^{-3}
63	condit	5.9×10^{-3}
64	capillari	5.8×10^{-3}
65	bed	5.8×10^{-3}
66	simul	5.8×10^{-3}
67	graviti	5.7×10^{-3}
68	rheolog	5.7×10^{-3}
69	industri	5.5×10^{-3}
70	deposit	5.4×10^{-3}
71	volum	5.3×10^{-3}
72	particip	5.3×10^{-3}
73	methan	5.3×10^{-3}
74	paramet	5.1×10^{-3}
75	cancer	5×10^{-3}
76	solut	4.9×10^{-3}
77	popul	4.9×10^{-3}
78	cement	4.9×10^{-3}
79	tecton	4.9×10^{-3}
80	ore	4.9×10^{-3}
81	model	4.8×10^{-3}
82	steam	4.8×10^{-3}
83	level	4.8×10^{-3}
84	tissu	4.6×10^{-3}
85	express	4.6×10^{-3}
86	geochem	4.6×10^{-3}
87	numer	4.5×10^{-3}
88	diesel	4.4×10^{-3}
89	catalyst	4.4×10^{-3}
90	crack	4.4×10^{-3}
91	fault	4.3×10^{-3}
92	therapi	4.2×10^{-3}
93	strata	4.2×10^{-3}
94	group	4.1×10^{-3}
95	vertic	4.1×10^{-3}
96	blood	4.1×10^{-3}
97	distribut	4.1×10^{-3}
98	outcom	4×10^{-3}
99	health	4×10^{-3}
100	subsurfac	4×10^{-3}

TABLE D.82. The list of the top 100 words in the category Entomology with RIGs

No.	Word	RIG
1	insect	1.3×10^{-1}
2	pest	1.3×10^{-1}
3	speci	1.3×10^{-1}
4	larva	8.1×10^{-2}
5	egg	5.2×10^{-2}
6	larval	5.1×10^{-2}
7	beetl	4.8×10^{-2}
8	adult	4.4×10^{-2}
9	parasitoid	4.4×10^{-2}
10	instar	4.4×10^{-2}
11	femal	4.2×10^{-2}
12	insecticid	4.2×10^{-2}
13	oviposit	4.1×10^{-2}
14	host	4×10^{-2}
15	plant	3.9×10^{-2}
16	infest	3.3×10^{-2}
17	rear	3×10^{-2}
18	mite	3×10^{-2}
19	genus	2.9×10^{-2}
20	feed	2.7×10^{-2}
21	fli	2.6×10^{-2}
22	predat	2.4×10^{-2}
23	crop	2.4×10^{-2}
24	popul	2.3×10^{-2}
25	fecund	2.3×10^{-2}
26	paper	2×10^{-2}
27	enemi	2×10^{-2}
28	male	1.9×10^{-2}
29	method	1.9×10^{-2}
30	biolog	1.8×10^{-2}
31	patient	1.8×10^{-2}
32	nov	1.8×10^{-2}
33	mosquito	1.8×10^{-2}
34	habitat	1.7×10^{-2}
35	mate	1.7×10^{-2}
36	reproduct	1.6×10^{-2}
37	laboratori	1.6×10^{-2}
38	parasit	1.6×10^{-2}
39	genera	1.5×10^{-2}
40	abund	1.5×10^{-2}
41	trap	1.5×10^{-2}
42	bioassay	1.4×10^{-2}
43	leaf	1.4×10^{-2}
44	coloni	1.4×10^{-2}
45	fruit	1.3×10^{-2}
46	herbivor	1.3×10^{-2}
47	arthropod	1.3×10^{-2}
48	model	1.3×10^{-2}
49	prey	1.2×10^{-2}
50	collect	1.2×10^{-2}

No.	Word	RIG
51	aed	1.2×10^{-2}
52	propos	1.2×10^{-2}
53	record	1.1×10^{-2}
54	forest	1.1×10^{-2}
55	tree	1×10^{-2}
56	immatur	1×10^{-2}
57	mortal	1×10^{-2}
58	season	1×10^{-2}
59	orchard	9.8×10^{-3}
60	clinic	9.8×10^{-3}
61	pesticid	9.7×10^{-3}
62	forag	9.6×10^{-3}
63	simul	9.6×10^{-3}
64	wing	9.3×10^{-3}
65	hatch	9×10^{-3}
66	development	9×10^{-3}
67	ecolog	8.8×10^{-3}
68	leav	8.8×10^{-3}
69	nativ	8.7×10^{-3}
70	sex	8.6×10^{-3}
71	brazil	8.6×10^{-3}
72	fauna	8.4×10^{-3}
73	spp	8.3×10^{-3}
74	divers	8.1×10^{-3}
75	surviv	8×10^{-3}
76	properti	8×10^{-3}
77	comput	8×10^{-3}
78	america	7.9×10^{-3}
79	control	7.8×10^{-3}
80	algorithm	7.6×10^{-3}
81	stage	7.5×10^{-3}
82	food	7.5×10^{-3}
83	taxonom	7.4×10^{-3}
84	perform	7.4×10^{-3}
85	taxa	7.3×10^{-3}
86	fed	7.3×10^{-3}
87	flower	7.3×10^{-3}
88	problem	7.2×10^{-3}
89	system	7.1×10^{-3}
90	suscept	7×10^{-3}
91	process	7×10^{-3}
92	nest	6.9×10^{-3}
93	energi	6.9×10^{-3}
94	design	6.8×10^{-3}
95	belong	6.7×10^{-3}
96	oper	6.4×10^{-3}
97	field	6.4×10^{-3}
98	north	6.4×10^{-3}
99	southern	6.4×10^{-3}
100	phylogenet	6.4×10^{-3}

Table D.83. The list of the top 100 words in the category Environmental Sciences with RIGs

No.	Word	RIG
1	water	5.2×10^{-2}
2	pollut	5.1×10^{-2}
3	concentr	5.1×10^{-2}
4	environment	4.4×10^{-2}
5	soil	4.4×10^{-2}
6	contamin	3.6×10^{-2}
7	ecosystem	2.5×10^{-2}
8	river	2.4×10^{-2}
9	wastewat	2.4×10^{-2}
10	climat	2.4×10^{-2}
11	land	2.2×10^{-2}
12	area	2.1×10^{-2}
13	patient	2×10^{-2}
14	sediment	2×10^{-2}
15	organ	1.9×10^{-2}
16	agricultur	1.9×10^{-2}
17	plant	1.9×10^{-2}
18	carbon	1.7×10^{-2}
19	exposur	1.6×10^{-2}
20	wast	1.6×10^{-2}
21	season	1.6×10^{-2}
22	urban	1.6×10^{-2}
23	groundwat	1.6×10^{-2}
24	aquat	1.6×10^{-2}
25	toxic	1.5×10^{-2}
26	biomass	1.5×10^{-2}
27	remov	1.4×10^{-2}
28	anthropogen	1.4×10^{-2}
29	ecolog	1.4×10^{-2}
30	emiss	1.3×10^{-2}
31	impact	1.3×10^{-2}
32	sourc	1.3×10^{-2}
33	effluent	1.3×10^{-2}
34	indic	1.3×10^{-2}
35	air	1.3×10^{-2}
36	site	1.2×10^{-2}
37	particul	1.2×10^{-2}
38	dissolv	1.2×10^{-2}
39	veget	1.2×10^{-2}
40	matter	1.2×10^{-2}
41	lake	1.1×10^{-2}
42	sludg	1.1×10^{-2}
43	clinic	1.1×10^{-2}
44	conclus	1.1×10^{-2}
45	speci	1.1×10^{-2}
46	china	1.1×10^{-2}
47	chemic	1.1×10^{-2}
48	nitrogen	1.1×10^{-2}
49	hydrolog	1×10^{-2}
50	assess	1×10^{-2}
51	product	1×10^{-2}
52	coastal	1×10^{-2}
53	forest	9.8×10^{-3}
54	total	9.7×10^{-3}
55	annual	9.5×10^{-3}
56	nutrient	9.4×10^{-3}
57	wetland	9.3×10^{-3}
58	summer	9.2×10^{-3}
59	environ	9×10^{-3}
60	atmospher	8.9×10^{-3}
61	spatial	8.5×10^{-3}
62	manag	8.5×10^{-3}
63	pesticid	8.4×10^{-3}
64	microbi	8.3×10^{-3}
65	runoff	8×10^{-3}
66	citi	7.9×10^{-3}
67	monitor	7.8×10^{-3}
68	hydrocarbon	7.8×10^{-3}
69	sustain	7.8×10^{-3}
70	sampl	7.8×10^{-3}
71	watersh	7.6×10^{-3}
72	aerosol	7.5×10^{-3}
73	municip	7.4×10^{-3}
74	phosphorus	7.4×10^{-3}
75	studi	7.2×10^{-3}
76	highest	7.2×10^{-3}
77	catchment	7.2×10^{-3}
78	habitat	7.2×10^{-3}
79	heavi	7.1×10^{-3}
80	winter	7.1×10^{-3}
81	basin	7×10^{-3}
82	industri	7×10^{-3}
83	increas	6.9×10^{-3}
84	greenhous	6.7×10^{-3}
85	degrad	6.7×10^{-3}
86	rainfal	6.6×10^{-3}
87	collect	6.6×10^{-3}
88	fish	6.5×10^{-3}
89	potenti	6.4×10^{-3}
90	metal	6.4×10^{-3}
91	region	6.4×10^{-3}
92	communiti	6.3×10^{-3}
93	chang	6.2×10^{-3}
94	econom	6.2×10^{-3}
95	dri	6.2×10^{-3}
96	ozon	6.1×10^{-3}
97	precipit	6.1×10^{-3}
98	estim	6×10^{-3}
99	nitrat	6×10^{-3}
100	freshwat	5.9×10^{-3}

TABLE D.84. The list of the top 100 words in the category Environmental Studies with RIGs

No.	Word	RIG
1	polici	8.7×10^{-2}
2	econom	5.6×10^{-2}
3	urban	4.6×10^{-2}
4	environment	4.5×10^{-2}
5	sustain	4.1×10^{-2}
6	govern	4.1×10^{-2}
7	climat	3.8×10^{-2}
8	social	3.6×10^{-2}
9	land	3.5×10^{-2}
10	citi	3.2×10^{-2}
11	sector	2.6×10^{-2}
12	plan	2.5×10^{-2}
13	manag	2.4×10^{-2}
14	market	2.4×10^{-2}
15	polit	2.4×10^{-2}
16	countri	2.3×10^{-2}
17	impact	2.1×10^{-2}
18	resourc	2.1×10^{-2}
19	price	1.9×10^{-2}
20	ecolog	1.9×10^{-2}
21	paper	1.9×10^{-2}
22	patient	1.9×10^{-2}
23	economi	1.8×10^{-2}
24	stakehold	1.8×10^{-2}
25	household	1.8×10^{-2}
26	public	1.7×10^{-2}
27	ecosystem	1.7×10^{-2}
28	develop	1.6×10^{-2}
29	cell	1.6×10^{-2}
30	communiti	1.6×10^{-2}
31	hous	1.5×10^{-2}
32	nation	1.5×10^{-2}
33	method	1.5×10^{-2}
34	agricultur	1.5×10^{-2}
35	landscap	1.5×10^{-2}
36	chang	1.5×10^{-2}
37	actor	1.4×10^{-2}
38	research	1.4×10^{-2}
39	area	1.4×10^{-2}
40	argu	1.4×10^{-2}
41	institut	1.3×10^{-2}
42	livelihood	1.3×10^{-2}
43	tourism	1.3×10^{-2}
44	focus	1.3×10^{-2}
45	invest	1.3×10^{-2}
46	local	1.3×10^{-2}
47	decis	1.3×10^{-2}
48	rural	1.3×10^{-2}
49	fisheri	1.2×10^{-2}
50	global	1.2×10^{-2}

No.	Word	RIG
51	project	1.1×10^{-2}
52	framework	1.1×10^{-2}
53	clinic	1.1×10^{-2}
54	build	1.1×10^{-2}
55	context	1.1×10^{-2}
56	futur	1.1×10^{-2}
57	conclus	1.1×10^{-2}
58	issu	1.1×10^{-2}
59	empir	1.1×10^{-2}
60	region	1.1×10^{-2}
61	spatial	1.1×10^{-2}
62	trade	1.1×10^{-2}
63	socio	1×10^{-2}
64	privat	1×10^{-2}
65	interview	1×10^{-2}
66	industri	1×10^{-2}
67	capit	1×10^{-2}
68	maker	1×10^{-2}
69	residenti	9.9×10^{-3}
70	benefit	9.7×10^{-3}
71	incom	9.7×10^{-3}
72	conserv	9.7×10^{-3}
73	peopl	9.6×10^{-3}
74	emiss	9.4×10^{-3}
75	servic	9.4×10^{-3}
76	draw	9.3×10^{-3}
77	protein	9.3×10^{-3}
78	incent	9.2×10^{-3}
79	perspect	9.2×10^{-3}
80	china	9.2×10^{-3}
81	forest	9.1×10^{-3}
82	natur	9×10^{-3}
83	farmer	8.9×10^{-3}
84	mitig	8.7×10^{-3}
85	treatment	8.7×10^{-3}
86	survey	8.6×10^{-3}
87	obtain	8.6×10^{-3}
88	explor	8.5×10^{-3}
89	biodivers	8.5×10^{-3}
90	articl	8.5×10^{-3}
91	demand	8.5×10^{-3}
92	practic	8.4×10^{-3}
93	infrastructur	8×10^{-3}
94	electron	8×10^{-3}
95	diseas	8×10^{-3}
96	address	7.9×10^{-3}
97	find	7.7×10^{-3}
98	acid	7.6×10^{-3}
99	detect	7.6×10^{-3}
100	experiment	7.6×10^{-3}

TABLE D.85. The list of the top 100 words in the category Ergonomics with RIGs

No.	Word	RIG	No.	Word	RIG
1	crash	5.3×10^{-2}	51	comfort	7.2×10^{-3}
2	driver	4.9×10^{-2}	52	acid	7.1×10^{-3}
3	task	4.5×10^{-2}	53	practic	7.1×10^{-3}
4	safeti	4.3×10^{-2}	54	questionnair	6.9×10^{-3}
5	ergonom	4×10^{-2}	55	passeng	6.9×10^{-3}
6	road	3.9×10^{-2}	56	train	6.9×10^{-3}
7	user	3.4×10^{-2}	57	employe	6.8×10^{-3}
8	particip	3.3×10^{-2}	58	organis	6.5×10^{-3}
9	worker	3×10^{-2}	59	violat	6.5×10^{-3}
10	traffic	2.8×10^{-2}	60	speci	6.5×10^{-3}
11	practition	2.6×10^{-2}	61	visual	6.5×10^{-3}
12	accid	2.4×10^{-2}	62	polic	6.4×10^{-3}
13	vehicl	2.3×10^{-2}	63	selfreport	6.3×10^{-3}
14	drive	2.2×10^{-2}	64	environ	6.2×10^{-3}
15	injuri	1.8×10^{-2}	65	gene	6.1×10^{-3}
16	perceiv	1.8×10^{-2}	66	treatment	6.1×10^{-3}
17	postur	1.7×10^{-2}	67	patient	6×10^{-3}
18	research	1.7×10^{-2}	68	fatigu	5.9×10^{-3}
19	summari	1.6×10^{-2}	69	situat	5.9×10^{-3}
20	musculoskelet	1.6×10^{-2}	70	hand	5.9×10^{-3}
21	studi	1.6×10^{-2}	71	energi	5.8×10^{-3}
22	design	1.5×10^{-2}	72	back	5.8×10^{-3}
23	risk	1.5×10^{-2}	73	interfac	5.7×10^{-3}
24	occup	1.4×10^{-2}	74	decis	5.7×10^{-3}
25	fatal	1.3×10^{-2}	75	lift	5.4×10^{-3}
26	workplac	1.3×10^{-2}	76	assess	5.3×10^{-3}
27	lane	1.3×10^{-2}	77	result	5.3×10^{-3}
28	work	1.2×10^{-2}	78	technolog	5.3×10^{-3}
29	industri	1.2×10^{-2}	79	need	5.2×10^{-3}
30	usabl	1.2×10^{-2}	80	movement	5.2×10^{-3}
31	car	1.1×10^{-2}	81	peopl	5.1×10^{-3}
32	cognit	1.1×10^{-2}	82	support	5.1×10^{-3}
33	cell	1×10^{-2}	83	object	5.1×10^{-3}
34	relev	9.7×10^{-3}	84	awar	5×10^{-3}
35	person	9.7×10^{-3}	85	differ	5×10^{-3}
36	percept	9.5×10^{-3}	86	help	5×10^{-3}
37	pedestrian	9.4×10^{-3}	87	concentr	4.9×10^{-3}
38	job	9.2×10^{-3}	88	wear	4.9×10^{-3}
39	factor	9.1×10^{-3}	89	biomechan	4.9×10^{-3}
40	inform	9×10^{-3}	90	provid	4.9×10^{-3}
41	shoulder	8.9×10^{-3}	91	find	4.9×10^{-3}
42	workload	8.9×10^{-3}	92	character	4.8×10^{-3}
43	subject	8.2×10^{-3}	93	affect	4.8×10^{-3}
44	speed	8.2×10^{-3}	94	effort	4.8×10^{-3}
45	protein	7.9×10^{-3}	95	examin	4.8×10^{-3}
46	collis	7.5×10^{-3}	96	molecular	4.8×10^{-3}
47	highway	7.5×10^{-3}	97	virtual	4.8×10^{-3}
48	behaviour	7.5×10^{-3}	98	perform	4.7×10^{-3}
49	manual	7.3×10^{-3}	99	attitud	4.7×10^{-3}
50	experi	7.3×10^{-3}	100	growth	4.7×10^{-3}

Table D.86. The list of the top 100 words in the category Ethics with RIGs

No.	Word	RIG
1	ethic	2.2×10^{-1}
2	moral	1×10^{-1}
3	argu	8.2×10^{-2}
4	philosoph	3.4×10^{-2}
5	argument	3.4×10^{-2}
6	bioethic	3.2×10^{-2}
7	social	3.1×10^{-2}
8	issu	2.5×10^{-2}
9	question	2.4×10^{-2}
10	consent	2.4×10^{-2}
11	articl	2.3×10^{-2}
12	legal	2.3×10^{-2}
13	debat	2.2×10^{-2}
14	claim	2.2×10^{-2}
15	concern	2.2×10^{-2}
16	view	2.2×10^{-2}
17	autonomi	2.2×10^{-2}
18	research	2.1×10^{-2}
19	practic	2.1×10^{-2}
20	corpor	1.9×10^{-2}
21	discuss	1.8×10^{-2}
22	result	1.8×10^{-2}
23	public	1.8×10^{-2}
24	justifi	1.8×10^{-2}
25	decis	1.8×10^{-2}
26	profession	1.8×10^{-2}
27	oblig	1.8×10^{-2}
28	way	1.7×10^{-2}
29	justic	1.7×10^{-2}
30	philosophi	1.7×10^{-2}
31	right	1.6×10^{-2}
32	reason	1.6×10^{-2}
33	defend	1.6×10^{-2}
34	harm	1.5×10^{-2}
35	make	1.5×10^{-2}
36	normat	1.5×10^{-2}
37	perspect	1.4×10^{-2}
38	justif	1.4×10^{-2}
39	medic	1.4×10^{-2}
40	polici	1.3×10^{-2}
41	scienc	1.3×10^{-2}
42	person	1.3×10^{-2}
43	principl	1.3×10^{-2}
44	concept	1.3×10^{-2}
45	societi	1.3×10^{-2}
46	rais	1.2×10^{-2}
47	busi	1.2×10^{-2}
48	think	1.2×10^{-2}
49	method	1.2×10^{-2}
50	institut	1.1×10^{-2}
51	stakehold	1.1×10^{-2}
52	organiz	1.1×10^{-2}
53	dilemma	1.1×10^{-2}
54	temperatur	1.1×10^{-2}
55	health	1.1×10^{-2}
56	draw	1.1×10^{-2}
57	duti	1.1×10^{-2}
58	polit	1.1×10^{-2}
59	cell	1.1×10^{-2}
60	context	1.1×10^{-2}
61	compar	1×10^{-2}
62	high	1×10^{-2}
63	whi	1×10^{-2}
64	peopl	1×10^{-2}
65	attitud	1×10^{-2}
66	paramet	1×10^{-2}
67	care	9.9×10^{-3}
68	conceptu	9.9×10^{-3}
69	surfac	9.9×10^{-3}
70	observ	9.8×10^{-3}
71	simul	9.4×10^{-3}
72	empir	9.4×10^{-3}
73	implic	9.3×10^{-3}
74	physician	8.9×10^{-3}
75	scientif	8.9×10^{-3}
76	address	8.8×10^{-3}
77	theori	8.7×10^{-3}
78	virtu	8.7×10^{-3}
79	engag	8.6×10^{-3}
80	essay	8.5×10^{-3}
81	conclud	8.5×10^{-3}
82	employe	8.4×10^{-3}
83	measur	8.4×10^{-3}
84	scholar	8.2×10^{-3}
85	individu	8.1×10^{-3}
86	will	8×10^{-3}
87	account	8×10^{-3}
88	protein	8×10^{-3}
89	studi	7.8×10^{-3}
90	focus	7.8×10^{-3}
91	commit	7.8×10^{-3}
92	judgment	7.7×10^{-3}
93	law	7.6×10^{-3}
94	ratio	7.6×10^{-3}
95	understand	7.6×10^{-3}
96	perform	7.4×10^{-3}
97	undermin	7.4×10^{-3}
98	offer	7.3×10^{-3}
99	idea	7.2×10^{-3}
100	human	7.1×10^{-3}

Table D.87. The list of the top 100 words in the category Ethnic Studies with RIGs

No.	Word	RIG
1	ethnic	1×10^{-1}
2	racial	9×10^{-2}
3	polit	8.7×10^{-2}
4	immigr	8.5×10^{-2}
5	articl	7.4×10^{-2}
6	argu	5.7×10^{-2}
7	migrant	5.6×10^{-2}
8	cultur	5.1×10^{-2}
9	social	5×10^{-2}
10	american	4.7×10^{-2}
11	nation	4.6×10^{-2}
12	ident	4.4×10^{-2}
13	race	4.4×10^{-2}
14	racism	4.2×10^{-2}
15	multicultur	3.3×10^{-2}
16	white	3.1×10^{-2}
17	examin	3×10^{-2}
18	polici	3×10^{-2}
19	black	3×10^{-2}
20	refuge	2.9×10^{-2}
21	minor	2.9×10^{-2}
22	draw	2.9×10^{-2}
23	discours	2.8×10^{-2}
24	societi	2.6×10^{-2}
25	transnat	2.4×10^{-2}
26	method	2.4×10^{-2}
27	asian	2.2×10^{-2}
28	context	2×10^{-2}
29	latino	2×10^{-2}
30	communiti	2×10^{-2}
31	contemporari	2×10^{-2}
32	citizenship	1.9×10^{-2}
33	african	1.9×10^{-2}
34	countri	1.9×10^{-2}
35	peopl	1.8×10^{-2}
36	muslim	1.8×10^{-2}
37	interview	1.8×10^{-2}
38	result	1.7×10^{-2}
39	ethnograph	1.6×10^{-2}
40	attitud	1.6×10^{-2}
41	debat	1.6×10^{-2}
42	migrat	1.6×10^{-2}
43	narrat	1.6×10^{-2}
44	religi	1.5×10^{-2}
45	explor	1.5×10^{-2}
46	usa	1.5×10^{-2}
47	educ	1.4×10^{-2}
48	prejudic	1.4×10^{-2}
49	european	1.4×10^{-2}
50	live	1.4×10^{-2}

No.	Word	RIG
51	focus	1.3×10^{-2}
52	research	1.3×10^{-2}
53	ideolog	1.3×10^{-2}
54	divers	1.3×10^{-2}
55	conflict	1.3×10^{-2}
56	europ	1.3×10^{-2}
57	public	1.3×10^{-2}
58	cell	1.2×10^{-2}
59	question	1.2×10^{-2}
60	actor	1.2×10^{-2}
61	legal	1.2×10^{-2}
62	seek	1.2×10^{-2}
63	simul	1.1×10^{-2}
64	liber	1.1×10^{-2}
65	mainstream	1.1×10^{-2}
66	scholar	1.1×10^{-2}
67	way	1.1×10^{-2}
68	implic	1.1×10^{-2}
69	youth	1.1×10^{-2}
70	perform	1.1×10^{-2}
71	contest	1.1×10^{-2}
72	use	1.1×10^{-2}
73	struggl	1×10^{-2}
74	school	1×10^{-2}
75	econom	1×10^{-2}
76	diaspora	1×10^{-2}
77	right	1×10^{-2}
78	temperatur	1×10^{-2}
79	among	1×10^{-2}
80	obtain	9.9×10^{-3}
81	gender	9.9×10^{-3}
82	perceiv	9.9×10^{-3}
83	british	9.9×10^{-3}
84	group	9.8×10^{-3}
85	discurs	9.6×10^{-3}
86	understand	9.6×10^{-3}
87	state	9.6×10^{-3}
88	surfac	9.4×10^{-3}
89	themselv	9.4×10^{-3}
90	histor	9.3×10^{-3}
91	nationalist	9.3×10^{-3}
92	war	9.2×10^{-3}
93	analys	9.2×10^{-3}
94	stereotyp	9.2×10^{-3}
95	oppress	9.1×10^{-3}
96	justic	9×10^{-3}
97	citizen	8.8×10^{-3}
98	engag	8.7×10^{-3}
99	detect	8.6×10^{-3}
100	women	8.5×10^{-3}

Table D.88. The list of the top 100 words in the category Evolutionary Biology with RIGs

No.	Word	RIG	No.	Word	RIG
1	speci	1.5×10^{-1}	51	individu	1.8×10^{-2}
2	evolutionari	1×10^{-1}	52	patient	1.8×10^{-2}
3	phylogenet	8×10^{-2}	53	dna	1.8×10^{-2}
4	genet	6.9×10^{-2}	54	sexual	1.7×10^{-2}
5	popul	6.1×10^{-2}	55	nuclear	1.7×10^{-2}
6	evolut	6×10^{-2}	56	evid	1.6×10^{-2}
7	diverg	5.9×10^{-2}	57	plant	1.6×10^{-2}
8	lineag	5.6×10^{-2}	58	taxon	1.5×10^{-2}
9	clade	5×10^{-2}	59	polymorph	1.5×10^{-2}
10	phylogeni	4.9×10^{-2}	60	conserv	1.5×10^{-2}
11	taxa	4.9×10^{-2}	61	predat	1.4×10^{-2}
12	divers	4.8×10^{-2}	62	biogeograph	1.4×10^{-2}
13	trait	4.6×10^{-2}	63	marker	1.4×10^{-2}
14	gene	4.4×10^{-2}	64	pleistocen	1.4×10^{-2}
15	ecolog	4.3×10^{-2}	65	molecular	1.4×10^{-2}
16	sequenc	4×10^{-2}	66	allel	1.3×10^{-2}
17	genom	3.8×10^{-2}	67	charact	1.3×10^{-2}
18	genus	3.4×10^{-2}	68	may	1.3×10^{-2}
19	suggest	3.2×10^{-2}	69	nich	1.3×10^{-2}
20	variat	3.2×10^{-2}	70	endem	1.3×10^{-2}
21	loci	3.2×10^{-2}	71	bird	1.3×10^{-2}
22	taxonom	3.1×10^{-2}	72	hypothes	1.2×10^{-2}
23	pattern	2.9×10^{-2}	73	mammal	1.2×10^{-2}
24	evolv	2.7×10^{-2}	74	fossil	1.2×10^{-2}
25	reproduct	2.7×10^{-2}	75	extant	1.2×10^{-2}
26	histori	2.7×10^{-2}	76	extinct	1.2×10^{-2}
27	diversif	2.7×10^{-2}	77	genotyp	1.2×10^{-2}
28	habitat	2.6×10^{-2}	78	relationship	1.2×10^{-2}
29	ancestr	2.6×10^{-2}	79	interspecif	1.2×10^{-2}
30	geograph	2.6×10^{-2}	80	method	1.2×10^{-2}
31	mate	2.5×10^{-2}	81	distinct	1.2×10^{-2}
32	morpholog	2.5×10^{-2}	82	insect	1.2×10^{-2}
33	mitochondri	2.4×10^{-2}	83	ancient	1.1×10^{-2}
34	microsatellit	2.4×10^{-2}	84	femal	1.1×10^{-2}
35	paper	2.4×10^{-2}	85	male	1.1×10^{-2}
36	sister	2.4×10^{-2}	86	fit	1.1×10^{-2}
37	infer	2.3×10^{-2}	87	offspr	1.1×10^{-2}
38	within	2.2×10^{-2}	88	anim	1.1×10^{-2}
39	among	2.2×10^{-2}	89	vertebr	1.1×10^{-2}
40	genera	2.2×10^{-2}	90	haplotyp	1.1×10^{-2}
41	adapt	2.1×10^{-2}	91	origin	1×10^{-2}
42	analys	2.1×10^{-2}	92	north	1×10^{-2}
43	across	2.1×10^{-2}	93	support	1×10^{-2}
44	speciat	2.1×10^{-2}	94	strong	1×10^{-2}
45	ancestor	2×10^{-2}	95	fish	1×10^{-2}
46	phenotyp	1.9×10^{-2}	96	climat	9.3×10^{-3}
47	hypothesi	1.9×10^{-2}	97	host	9.3×10^{-3}
48	select	1.9×10^{-2}	98	famili	9.2×10^{-3}
49	bayesian	1.9×10^{-2}	99	region	9.2×10^{-3}
50	tree	1.8×10^{-2}	100	relat	9.1×10^{-3}

Table D.89. The list of the top 100 words in the category Family Studies with RIGs

No.	Word	RIG	No.	Word	RIG
1	child	1.3×10^{-1}	51	perform	1.6×10^{-2}
2	famili	1×10^{-1}	52	survey	1.6×10^{-2}
3	parent	9.4×10^{-2}	53	discuss	1.5×10^{-2}
4	children	8.6×10^{-2}	54	live	1.4×10^{-2}
5	abus	6.5×10^{-2}	55	depress	1.4×10^{-2}
6	violenc	5.9×10^{-2}	56	cell	1.4×10^{-2}
7	adolesc	5.9×10^{-2}	57	report	1.4×10^{-2}
8	sexual	5.3×10^{-2}	58	associ	1.4×10^{-2}
9	youth	5.2×10^{-2}	59	explor	1.3×10^{-2}
10	examin	5.1×10^{-2}	60	engag	1.3×10^{-2}
11	relationship	4.8×10^{-2}	61	method	1.3×10^{-2}
12	maltreat	4.7×10^{-2}	62	selfreport	1.3×10^{-2}
13	partner	4.6×10^{-2}	63	adult	1.3×10^{-2}
14	mother	4.4×10^{-2}	64	girl	1.3×10^{-2}
15	victim	4.4×10^{-2}	65	peer	1.3×10^{-2}
16	social	3.8×10^{-2}	66	articl	1.2×10^{-2}
17	research	3.2×10^{-2}	67	men	1.2×10^{-2}
18	emot	3×10^{-2}	68	perceiv	1.2×10^{-2}
19	father	2.9×10^{-2}	69	romant	1.2×10^{-2}
20	particip	2.9×10^{-2}	70	spous	1.2×10^{-2}
21	implic	2.8×10^{-2}	71	surfac	1.2×10^{-2}
22	interview	2.8×10^{-2}	72	neglect	1.2×10^{-2}
23	women	2.7×10^{-2}	73	communiti	1.2×10^{-2}
24	intervent	2.7×10^{-2}	74	temperatur	1.2×10^{-2}
25	childhood	2.5×10^{-2}	75	simul	1.2×10^{-2}
26	perpetr	2.5×10^{-2}	76	practic	1.2×10^{-2}
27	caregiv	2.3×10^{-2}	77	longitudin	1.2×10^{-2}
28	health	2.2×10^{-2}	78	adulthood	1.2×10^{-2}
29	marriag	2.2×10^{-2}	79	percept	1.2×10^{-2}
30	intim	2.2×10^{-2}	80	dyad	1.1×10^{-2}
31	gender	2.2×10^{-2}	81	trauma	1.1×10^{-2}
32	find	2.2×10^{-2}	82	conflict	1.1×10^{-2}
33	psycholog	2.1×10^{-2}	83	effici	1.1×10^{-2}
34	welfar	2.1×10^{-2}	84	paramet	1.1×10^{-2}
35	school	2×10^{-2}	85	offend	1.1×10^{-2}
36	ipv	2×10^{-2}	86	context	1.1×10^{-2}
37	mental	1.9×10^{-2}	87	symptom	1.1×10^{-2}
38	age	1.9×10^{-2}	88	violent	1.1×10^{-2}
39	behavior	1.9×10^{-2}	89	outcom	1.1×10^{-2}
40	marit	1.9×10^{-2}	90	home	1×10^{-2}
41	among	1.8×10^{-2}	91	assault	1×10^{-2}
42	support	1.8×10^{-2}	92	energi	1×10^{-2}
43	care	1.8×10^{-2}	93	qualit	1×10^{-2}
44	young	1.8×10^{-2}	94	educ	1×10^{-2}
45	foster	1.7×10^{-2}	95	matern	1×10^{-2}
46	servic	1.7×10^{-2}	96	aggress	1×10^{-2}
47	studi	1.7×10^{-2}	97	american	1×10^{-2}
48	risk	1.7×10^{-2}	98	ethnic	1×10^{-2}
49	experienc	1.7×10^{-2}	99	relat	9.9×10^{-3}
50	marri	1.7×10^{-2}	100	induc	9.8×10^{-3}

Table D.90. The list of the top 100 words in the category Film, Radio, Television with RIGs

No.	Word	RIG
1	televis	1.3×10^{-1}
2	articl	1.2×10^{-1}
3	cinema	9.3×10^{-2}
4	film	8.8×10^{-2}
5	audienc	6.8×10^{-2}
6	media	6×10^{-2}
7	argu	5.9×10^{-2}
8	cultur	4.8×10^{-2}
9	narrat	4.6×10^{-2}
10	cinemat	4.3×10^{-2}
11	genr	4.2×10^{-2}
12	polit	3.8×10^{-2}
13	broadcast	3.3×10^{-2}
14	filmmak	3.3×10^{-2}
15	discours	3.2×10^{-2}
16	viewer	3×10^{-2}
17	contemporari	3×10^{-2}
18	result	2.9×10^{-2}
19	british	2.7×10^{-2}
20	social	2.6×10^{-2}
21	essay	2.4×10^{-2}
22	drama	2.4×10^{-2}
23	aesthet	2.3×10^{-2}
24	explor	2.3×10^{-2}
25	draw	2.3×10^{-2}
26	method	2.3×10^{-2}
27	transnat	2.3×10^{-2}
28	stori	2.3×10^{-2}
29	studio	2.2×10^{-2}
30	entertain	2.1×10^{-2}
31	documentari	2.1×10^{-2}
32	text	2×10^{-2}
33	charact	1.9×10^{-2}
34	director	1.9×10^{-2}
35	portray	1.8×10^{-2}
36	news	1.8×10^{-2}
37	popular	1.8×10^{-2}
38	american	1.7×10^{-2}
39	writer	1.7×10^{-2}
40	imagin	1.7×10^{-2}
41	way	1.7×10^{-2}
42	artist	1.7×10^{-2}
43	style	1.6×10^{-2}
44	comedi	1.6×10^{-2}
45	theatr	1.6×10^{-2}
46	examin	1.6×10^{-2}
47	fiction	1.6×10^{-2}
48	represent	1.5×10^{-2}
49	ideolog	1.5×10^{-2}
50	creativ	1.4×10^{-2}
51	engag	1.4×10^{-2}
52	1970s	1.4×10^{-2}
53	nation	1.4×10^{-2}
54	digit	1.4×10^{-2}
55	patient	1.3×10^{-2}
56	journalist	1.3×10^{-2}
57	improv	1.3×10^{-2}
58	measur	1.2×10^{-2}
59	onlin	1.2×10^{-2}
60	fan	1.2×10^{-2}
61	war	1.2×10^{-2}
62	conclus	1.2×10^{-2}
63	1960s	1.2×10^{-2}
64	public	1.2×10^{-2}
65	recept	1.2×10^{-2}
66	realism	1.2×10^{-2}
67	cell	1.2×10^{-2}
68	celebr	1.2×10^{-2}
69	1980s	1.1×10^{-2}
70	histori	1.1×10^{-2}
71	context	1.1×10^{-2}
72	script	1.1×10^{-2}
73	obtain	1.1×10^{-2}
74	screen	1.1×10^{-2}
75	articul	1.1×10^{-2}
76	critiqu	1.1×10^{-2}
77	spectat	1.1×10^{-2}
78	literari	1×10^{-2}
79	compar	1×10^{-2}
80	discuss	1×10^{-2}
81	theatric	1×10^{-2}
82	theme	1×10^{-2}
83	applic	1×10^{-2}
84	industri	1×10^{-2}
85	higher	1×10^{-2}
86	focus	9.8×10^{-3}
87	histor	9.7×10^{-3}
88	evalu	9.6×10^{-3}
89	effect	9.6×10^{-3}
90	temperatur	9.5×10^{-3}
91	debat	9.5×10^{-3}
92	observ	9.4×10^{-3}
93	discurs	9.3×10^{-3}
94	1950s	9.1×10^{-3}
95	offer	9.1×10^{-3}
96	data	9.1×10^{-3}
97	world	9×10^{-3}
98	corpor	8.9×10^{-3}
99	archiv	8.9×10^{-3}
100	use	8.9×10^{-3}

Table D.91. The list of the top 100 words in the category Fisheries with RIGs

No.	Word	RIG	No.	Word	RIG
1	fish	3×10^{-1}	51	prey	1.3×10^{-2}
2	fisheri	7.4×10^{-2}	52	caught	1.3×10^{-2}
3	speci	6.5×10^{-2}	53	infect	1.2×10^{-2}
4	juvenil	6.3×10^{-2}	54	innat	1.2×10^{-2}
5	aquacultur	6.2×10^{-2}	55	gulf	1.2×10^{-2}
6	spawn	5.2×10^{-2}	56	cdna	1.2×10^{-2}
7	stock	5×10^{-2}	57	signific	1.2×10^{-2}
8	feed	4.5×10^{-2}	58	matur	1.1×10^{-2}
9	salmon	4.3×10^{-2}	59	pathogen	1.1×10^{-2}
10	diet	3.9×10^{-2}	60	suggest	1.1×10^{-2}
11	fed	3.7×10^{-2}	61	pond	1.1×10^{-2}
12	trout	3.5×10^{-2}	62	commerci	1.1×10^{-2}
13	sea	3.3×10^{-2}	63	supplement	1.1×10^{-2}
14	catch	3.1×10^{-2}	64	protein	1×10^{-2}
15	atlant	3.1×10^{-2}	65	gonad	1×10^{-2}
16	gill	2.8×10^{-2}	66	coastal	1×10^{-2}
17	growth	2.8×10^{-2}	67	pelag	9.8×10^{-3}
18	rear	2.7×10^{-2}	68	conserv	9.8×10^{-3}
19	larva	2.6×10^{-2}	69	total	9.5×10^{-3}
20	river	2.6×10^{-2}	70	bodi	9.4×10^{-3}
21	shrimp	2.6×10^{-2}	71	highest	8.9×10^{-3}
22	marin	2.5×10^{-2}	72	north	8.8×10^{-3}
23	immun	2.5×10^{-2}	73	intestin	8.8×10^{-3}
24	habitat	2.4×10^{-2}	74	conclus	8.7×10^{-3}
25	dietari	2.3×10^{-2}	75	manag	8.7×10^{-3}
26	vibrio	2.3×10^{-2}	76	tag	8.4×10^{-3}
27	freshwat	2.1×10^{-2}	77	estuari	8.4×10^{-3}
28	abund	2.1×10^{-2}	78	dure	8.4×10^{-3}
29	farm	2×10^{-2}	79	ecosystem	8.3×10^{-3}
30	paper	1.9×10^{-2}	80	bay	8.3×10^{-3}
31	mortal	1.9×10^{-2}	81	amino	8.2×10^{-3}
32	larval	1.9×10^{-2}	82	predat	8.1×10^{-3}
33	water	1.9×10^{-2}	83	spleen	8.1×10^{-3}
34	length	1.8×10^{-2}	84	cultur	8.1×10^{-3}
35	patient	1.8×10^{-2}	85	bacteri	7.9×10^{-3}
36	surviv	1.7×10^{-2}	86	liver	7.8×10^{-3}
37	pacif	1.7×10^{-2}	87	period	7.8×10^{-3}
38	popul	1.6×10^{-2}	88	summer	7.7×10^{-3}
39	hatch	1.6×10^{-2}	89	biomass	7.7×10^{-3}
40	weight	1.5×10^{-2}	90	meal	7.6×10^{-3}
41	coast	1.4×10^{-2}	91	femal	7.6×10^{-3}
42	egg	1.4×10^{-2}	92	tissu	7.5×10^{-3}
43	season	1.4×10^{-2}	93	lipid	7.5×10^{-3}
44	lake	1.4×10^{-2}	94	ocean	7.4×10^{-3}
45	indic	1.4×10^{-2}	95	digest	7.3×10^{-3}
46	tank	1.4×10^{-2}	96	gene	7.3×10^{-3}
47	method	1.3×10^{-2}	97	trophic	7.2×10^{-3}
48	propos	1.3×10^{-2}	98	declin	7.2×10^{-3}
49	reproduct	1.3×10^{-2}	99	muscl	7.1×10^{-3}
50	day	1.3×10^{-2}	100	food	7.1×10^{-3}

TABLE D.92. The list of the top 100 words in the category Folklore with RIGs

No.	Word	RIG	No.	Word	RIG
1	folklor	1.7×10^{-1}	51	measur	1.5×10^{-2}
2	articl	1.1×10^{-1}	52	anthropolog	1.5×10^{-2}
3	folk	9×10^{-2}	53	written	1.5×10^{-2}
4	folklorist	8.2×10^{-2}	54	test	1.4×10^{-2}
5	centuri	6.5×10^{-2}	55	way	1.4×10^{-2}
6	cultur	6.1×10^{-2}	56	improv	1.4×10^{-2}
7	ritual	5.3×10^{-2}	57	increas	1.4×10^{-2}
8	tradit	5.3×10^{-2}	58	represent	1.4×10^{-2}
9	legend	4.6×10^{-2}	59	high	1.4×10^{-2}
10	narrat	4.3×10^{-2}	60	author	1.4×10^{-2}
11	result	4.2×10^{-2}	61	method	1.3×10^{-2}
12	indigen	3.5×10^{-2}	62	collect	1.3×10^{-2}
13	song	3.2×10^{-2}	63	aesthet	1.3×10^{-2}
14	tale	3×10^{-2}	64	typolog	1.3×10^{-2}
15	symbol	3×10^{-2}	65	world	1.2×10^{-2}
16	villag	2.9×10^{-2}	66	higher	1.2×10^{-2}
17	polit	2.8×10^{-2}	67	low	1.2×10^{-2}
18	religi	2.8×10^{-2}	68	ethnograph	1.2×10^{-2}
19	stori	2.6×10^{-2}	69	reduc	1.2×10^{-2}
20	nineteenth	2.6×10^{-2}	70	religion	1.2×10^{-2}
21	festiv	2.6×10^{-2}	71	patient	1.2×10^{-2}
22	social	2.6×10^{-2}	72	use	1.2×10^{-2}
23	heritag	2.5×10^{-2}	73	artist	1.2×10^{-2}
24	histor	2.5×10^{-2}	74	19th	1.2×10^{-2}
25	scholar	2.3×10^{-2}	75	cell	1.2×10^{-2}
26	music	2.2×10^{-2}	76	british	1.2×10^{-2}
27	text	2.2×10^{-2}	77	concept	1.2×10^{-2}
28	fieldwork	2.1×10^{-2}	78	communiti	1.2×10^{-2}
29	societi	2.1×10^{-2}	79	whi	1.1×10^{-2}
30	argu	2.1×10^{-2}	80	ancient	1.1×10^{-2}
31	peopl	2.1×10^{-2}	81	south	1.1×10^{-2}
32	archiv	2×10^{-2}	82	system	1.1×10^{-2}
33	effect	2×10^{-2}	83	magic	1.1×10^{-2}
34	popular	2×10^{-2}	84	themselv	1.1×10^{-2}
35	genr	2×10^{-2}	85	obtain	1.1×10^{-2}
36	contemporari	1.9×10^{-2}	86	collector	1×10^{-2}
37	twentieth	1.9×10^{-2}	87	eighteenth	1×10^{-2}
38	model	1.9×10^{-2}	88	sing	1×10^{-2}
39	offici	1.8×10^{-2}	89	context	1×10^{-2}
40	belief	1.8×10^{-2}	90	past	1×10^{-2}
41	ident	1.7×10^{-2}	91	literari	1×10^{-2}
42	celebr	1.7×10^{-2}	92	treatment	9.8×10^{-3}
43	modern	1.7×10^{-2}	93	compar	9.8×10^{-3}
44	discuss	1.7×10^{-2}	94	interpret	9.8×10^{-3}
45	essay	1.6×10^{-2}	95	simul	9.7×10^{-3}
46	20th	1.6×10^{-2}	96	scholarship	9.7×10^{-3}
47	mediev	1.5×10^{-2}	97	rate	9.7×10^{-3}
48	vernacular	1.5×10^{-2}	98	effici	9.6×10^{-3}
49	today	1.5×10^{-2}	99	discours	9.6×10^{-3}
50	histori	1.5×10^{-2}	100	danc	9.6×10^{-3}

Table D.93. The list of the top 100 words in the category Food Science and Technology with RIGs

No.	Word	RIG	No.	Word	RIG
1	food	8×10^{-2}	51	raw	1.1×10^{-2}
2	acid	5.4×10^{-2}	52	respect	1.1×10^{-2}
3	content	4.3×10^{-2}	53	grape	1.1×10^{-2}
4	antioxid	4×10^{-2}	54	moistur	1×10^{-2}
5	milk	3.8×10^{-2}	55	lipid	1×10^{-2}
6	product	3.6×10^{-2}	56	min	1×10^{-2}
7	extract	3.2×10^{-2}	57	contamin	9.9×10^{-3}
8	chromatographi	2.9×10^{-2}	58	flavor	9.7×10^{-3}
9	phenol	2.9×10^{-2}	59	activ	9.7×10^{-3}
10	concentr	2.9×10^{-2}	60	liquid	9.6×10^{-3}
11	compound	2.6×10^{-2}	61	commerci	9.6×10^{-3}
12	sensori	2.4×10^{-2}	62	water	9.6×10^{-3}
13	sampl	2.4×10^{-2}	63	qualiti	9.6×10^{-3}
14	fruit	2.3×10^{-2}	64	shelf	9.5×10^{-3}
15	ferment	2.2×10^{-2}	65	isol	9.4×10^{-3}
16	oil	2.2×10^{-2}	66	bacteria	9.4×10^{-3}
17	meat	2.2×10^{-2}	67	highest	9.4×10^{-3}
18	hplc	2×10^{-2}	68	textur	9.2×10^{-3}
19	storag	1.9×10^{-2}	69	enzym	9.2×10^{-3}
20	fatti	1.8×10^{-2}	70	digest	9.1×10^{-3}
21	dri	1.8×10^{-2}	71	salmonella	9×10^{-3}
22	dairi	1.8×10^{-2}	72	lactic	8.9×10^{-3}
23	polyphenol	1.7×10^{-2}	73	contain	8.8×10^{-3}
24	fat	1.7×10^{-2}	74	diet	8.8×10^{-3}
25	starch	1.7×10^{-2}	75	volatil	8.8×10^{-3}
26	patient	1.7×10^{-2}	76	detect	8.7×10^{-3}
27	paper	1.7×10^{-2}	77	assay	8.7×10^{-3}
28	cook	1.6×10^{-2}	78	cultivar	8.6×10^{-3}
29	dpph	1.5×10^{-2}	79	produc	8.6×10^{-3}
30	wine	1.4×10^{-2}	80	physicochem	8.5×10^{-3}
31	spectrometri	1.4×10^{-2}	81	tast	8.4×10^{-3}
32	fresh	1.4×10^{-2}	82	colour	8.1×10^{-3}
33	cfu	1.3×10^{-2}	83	beef	7.8×10^{-3}
34	degre	1.3×10^{-2}	84	valu	7.8×10^{-3}
35	nutrit	1.3×10^{-2}	85	antimicrobi	7.8×10^{-3}
36	consum	1.3×10^{-2}	86	ripen	7.7×10^{-3}
37	scaveng	1.3×10^{-2}	87	propos	7.6×10^{-3}
38	protein	1.3×10^{-2}	88	microbi	7.5×10^{-3}
39	total	1.3×10^{-2}	89	composit	7.4×10^{-3}
40	sugar	1.2×10^{-2}	90	effect	7.2×10^{-3}
41	flavonoid	1.2×10^{-2}	91	decreas	7.2×10^{-3}
42	determin	1.2×10^{-2}	92	radic	6.9×10^{-3}
43	lactobacillus	1.2×10^{-2}	93	rice	6.9×10^{-3}
44	cow	1.2×10^{-2}	94	comput	6.9×10^{-3}
45	dietari	1.2×10^{-2}	95	yeast	6.8×10^{-3}
46	bioactiv	1.1×10^{-2}	96	recoveri	6.8×10^{-3}
47	ingredi	1.1×10^{-2}	97	store	6.7×10^{-3}
48	wheat	1.1×10^{-2}	98	microbiolog	6.6×10^{-3}
49	solubl	1.1×10^{-2}	99	acet	6.6×10^{-3}
50	anthocyanin	1.1×10^{-2}	100	soybean	6.5×10^{-3}

TABLE D.94. The list of the top 100 words in the category Forestry with RIGs

No.	Word	RIG
1	forest	2.3×10^{-1}
2	tree	1.7×10^{-1}
3	speci	8.5×10^{-2}
4	wood	8×10^{-2}
5	stand	7.2×10^{-2}
6	pine	6.4×10^{-2}
7	pinus	6×10^{-2}
8	plant	5×10^{-2}
9	plantat	4.5×10^{-2}
10	canopi	4.5×10^{-2}
11	spruce	4.1×10^{-2}
12	timber	4×10^{-2}
13	soil	4×10^{-2}
14	ecosystem	3.8×10^{-2}
15	plot	3.8×10^{-2}
16	forestri	3.7×10^{-2}
17	climat	3.6×10^{-2}
18	veget	3.6×10^{-2}
19	seedl	3.4×10^{-2}
20	area	3.3×10^{-2}
21	manag	3×10^{-2}
22	growth	2.9×10^{-2}
23	height	2.9×10^{-2}
24	fire	2.7×10^{-2}
25	oak	2.6×10^{-2}
26	woodi	2.6×10^{-2}
27	biomass	2.5×10^{-2}
28	land	2.4×10^{-2}
29	site	2.3×10^{-2}
30	harvest	2.3×10^{-2}
31	leaf	2.2×10^{-2}
32	stem	2×10^{-2}
33	ecolog	2×10^{-2}
34	year	1.9×10^{-2}
35	landscap	1.8×10^{-2}
36	season	1.8×10^{-2}
37	patient	1.8×10^{-2}
38	regener	1.8×10^{-2}
39	drought	1.8×10^{-2}
40	diamet	1.8×10^{-2}
41	annual	1.7×10^{-2}
42	grow	1.7×10^{-2}
43	habitat	1.5×10^{-2}
44	cover	1.5×10^{-2}
45	biodivers	1.5×10^{-2}
46	mountain	1.5×10^{-2}
47	crown	1.5×10^{-2}
48	shrub	1.4×10^{-2}
49	moistur	1.3×10^{-2}
50	bark	1.3×10^{-2}

No.	Word	RIG
51	disturb	1.3×10^{-2}
52	divers	1.3×10^{-2}
53	dri	1.3×10^{-2}
54	densiti	1.2×10^{-2}
55	communiti	1.2×10^{-2}
56	boreal	1.2×10^{-2}
57	burn	1.2×10^{-2}
58	northern	1.2×10^{-2}
59	variabl	1.2×10^{-2}
60	root	1.2×10^{-2}
61	seed	1.1×10^{-2}
62	domin	1.1×10^{-2}
63	southern	1.1×10^{-2}
64	inventori	1.1×10^{-2}
65	tropic	1.1×10^{-2}
66	clinic	1.1×10^{-2}
67	product	1.1×10^{-2}
68	environment	1.1×10^{-2}
69	basal	1.1×10^{-2}
70	differ	1×10^{-2}
71	natur	1×10^{-2}
72	conserv	1×10^{-2}
73	temper	1×10^{-2}
74	trait	1×10^{-2}
75	stock	1×10^{-2}
76	chang	9.8×10^{-3}
77	mediterranean	9.8×10^{-3}
78	nutrient	9.6×10^{-3}
79	woodland	9.4×10^{-3}
80	nativ	9.3×10^{-3}
81	locat	9.1×10^{-3}
82	north	8.9×10^{-3}
83	litter	8.6×10^{-3}
84	variat	8.6×10^{-3}
85	increas	8.6×10^{-3}
86	spatial	8.6×10^{-3}
87	abund	8.5×10^{-3}
88	beetl	8.5×10^{-3}
89	declin	8.3×10^{-3}
90	across	8.1×10^{-3}
91	grassland	8.1×10^{-3}
92	carbon	8.1×10^{-3}
93	eastern	7.7×10^{-3}
94	influenc	7.7×10^{-3}
95	indic	7.5×10^{-3}
96	log	7.5×10^{-3}
97	region	7.5×10^{-3}
98	propos	7.5×10^{-3}
99	old	7.4×10^{-3}
100	estim	7.2×10^{-3}

TABLE D.95. The list of the top 100 words in the category Gastroenterology and Hepatology with RIGs

No.	Word	RIG	No.	Word	RIG
1	patient	1.8×10^{-1}	51	surgic	2.2×10^{-2}
2	conclus	1.2×10^{-1}	52	histolog	2.1×10^{-2}
3	liver	9.8×10^{-2}	53	mucos	2×10^{-2}
4	hepat	9×10^{-2}	54	recurr	2×10^{-2}
5	aim	8.4×10^{-2}	55	bile	2×10^{-2}
6	background	7×10^{-2}	56	infect	2×10^{-2}
7	diseas	6.7×10^{-2}	57	diagnos	1.9×10^{-2}
8	endoscop	6.7×10^{-2}	58	postop	1.9×10^{-2}
9	bowel	6.6×10^{-2}	59	prospect	1.9×10^{-2}
10	resect	4.7×10^{-2}	60	abdomin	1.9×10^{-2}
11	cirrhosi	4.6×10^{-2}	61	multivari	1.9×10^{-2}
12	clinic	4.2×10^{-2}	62	hospit	1.8×10^{-2}
13	crohn	4.2×10^{-2}	63	age	1.8×10^{-2}
14	gastric	4×10^{-2}	64	fibrosi	1.8×10^{-2}
15	hepatocellular	3.9×10^{-2}	65	hbv	1.7×10^{-2}
16	underw	3.9×10^{-2}	66	rectal	1.7×10^{-2}
17	pancreat	3.8×10^{-2}	67	transplant	1.7×10^{-2}
18	coliti	3.7×10^{-2}	68	virus	1.7×10^{-2}
19	endoscopi	3.7×10^{-2}	69	serum	1.6×10^{-2}
20	cancer	3.7×10^{-2}	70	result	1.6×10^{-2}
21	ulcer	3.5×10^{-2}	71	treat	1.6×10^{-2}
22	treatment	3.4×10^{-2}	72	includ	1.6×10^{-2}
23	therapi	3.4×10^{-2}	73	factor	1.6×10^{-2}
24	gastrointestin	3.3×10^{-2}	74	lesion	1.6×10^{-2}
25	colorect	3.2×10^{-2}	75	bleed	1.5×10^{-2}
26	carcinoma	3.2×10^{-2}	76	cohort	1.5×10^{-2}
27	retrospect	3.1×10^{-2}	77	mortal	1.5×10^{-2}
28	paper	3.1×10^{-2}	78	structur	1.5×10^{-2}
29	chronic	3.1×10^{-2}	79	group	1.5×10^{-2}
30	associ	2.9×10^{-2}	80	score	1.5×10^{-2}
31	hcc	2.9×10^{-2}	81	propos	1.5×10^{-2}
32	esophag	2.9×10^{-2}	82	duct	1.4×10^{-2}
33	surgeri	2.8×10^{-2}	83	inflamm	1.4×10^{-2}
34	outcom	2.7×10^{-2}	84	follow	1.4×10^{-2}
35	complic	2.7×10^{-2}	85	assess	1.4×10^{-2}
36	median	2.7×10^{-2}	86	symptom	1.4×10^{-2}
37	method	2.6×10^{-2}	87	consecut	1.4×10^{-2}
38	intestin	2.6×10^{-2}	88	enrol	1.4×10^{-2}
39	diagnosi	2.6×10^{-2}	89	adenocarcinoma	1.3×10^{-2}
40	tumor	2.5×10^{-2}	90	laparoscop	1.3×10^{-2}
41	biopsi	2.5×10^{-2}	91	receiv	1.3×10^{-2}
42	inflammatori	2.5×10^{-2}	92	mucosa	1.3×10^{-2}
43	risk	2.4×10^{-2}	93	simul	1.3×10^{-2}
44	hcv	2.4×10^{-2}	94	virolog	1.3×10^{-2}
45	colon	2.4×10^{-2}	95	efficaci	1.3×10^{-2}
46	biliari	2.3×10^{-2}	96	evalu	1.3×10^{-2}
47	surviv	2.3×10^{-2}	97	temperatur	1.2×10^{-2}
48	signific	2.3×10^{-2}	98	rate	1.2×10^{-2}
49	year	2.3×10^{-2}	99	incid	1.2×10^{-2}
50	review	2.2×10^{-2}	100	studi	1.2×10^{-2}

TABLE D.96. The list of the top 100 words in the category Genetics and Heredity with RIGs

No.	Word	RIG
1	gene	1.8×10^{-1}
2	genom	1.3×10^{-1}
3	genet	1.2×10^{-1}
4	sequenc	7.6×10^{-2}
5	chromosom	5.9×10^{-2}
6	allel	5.5×10^{-2}
7	mutat	5.4×10^{-2}
8	polymorph	5.1×10^{-2}
9	phenotyp	4.9×10^{-2}
10	loci	4.8×10^{-2}
11	dna	4.6×10^{-2}
12	genotyp	4.3×10^{-2}
13	express	4.2×10^{-2}
14	transcript	3.8×10^{-2}
15	protein	3.7×10^{-2}
16	identifi	3.6×10^{-2}
17	nucleotid	3.3×10^{-2}
18	paper	2.9×10^{-2}
19	trait	2.9×10^{-2}
20	popul	2.8×10^{-2}
21	locus	2.7×10^{-2}
22	marker	2.6×10^{-2}
23	famili	2.6×10^{-2}
24	associ	2.5×10^{-2}
25	variant	2.5×10^{-2}
26	snps	2.4×10^{-2}
27	speci	2.2×10^{-2}
28	encod	2.2×10^{-2}
29	regul	2.2×10^{-2}
30	snp	2.2×10^{-2}
31	evolutionari	2.1×10^{-2}
32	rna	2×10^{-2}
33	delet	2×10^{-2}
34	microsatellit	1.9×10^{-2}
35	transcriptom	1.9×10^{-2}
36	autosom	1.9×10^{-2}
37	cell	1.8×10^{-2}
38	phylogenet	1.8×10^{-2}
39	suggest	1.8×10^{-2}
40	molecular	1.7×10^{-2}
41	divers	1.6×10^{-2}
42	conserv	1.6×10^{-2}
43	breed	1.6×10^{-2}
44	exon	1.6×10^{-2}
45	pathway	1.5×10^{-2}
46	diseas	1.5×10^{-2}
47	region	1.5×10^{-2}
48	haplotyp	1.5×10^{-2}
49	lineag	1.5×10^{-2}
50	syndrom	1.4×10^{-2}

No.	Word	RIG
51	mitochondri	1.4×10^{-2}
52	inherit	1.4×10^{-2}
53	role	1.3×10^{-2}
54	qtl	1.3×10^{-2}
55	mutant	1.3×10^{-2}
56	heterozygos	1.3×10^{-2}
57	diverg	1.3×10^{-2}
58	human	1.2×10^{-2}
59	seq	1.2×10^{-2}
60	involv	1.2×10^{-2}
61	putat	1.2×10^{-2}
62	individu	1.2×10^{-2}
63	copi	1.1×10^{-2}
64	background	1.1×10^{-2}
65	reveal	1.1×10^{-2}
66	homolog	1.1×10^{-2}
67	pcr	1.1×10^{-2}
68	linkag	1.1×10^{-2}
69	novo	1.1×10^{-2}
70	recess	1.1×10^{-2}
71	heterozyg	1.1×10^{-2}
72	annot	1.1×10^{-2}
73	wild	1.1×10^{-2}
74	differenti	1×10^{-2}
75	development	1×10^{-2}
76	homozyg	1×10^{-2}
77	variat	1×10^{-2}
78	drosophila	1×10^{-2}
79	epigenet	1×10^{-2}
80	chromatin	9.9×10^{-3}
81	specif	9.7×10^{-3}
82	solut	9.5×10^{-3}
83	regulatori	9.4×10^{-3}
84	analysi	9.1×10^{-3}
85	previous	9×10^{-3}
86	evolut	8.9×10^{-3}
87	metabol	8.8×10^{-3}
88	candid	8.8×10^{-3}
89	disord	8.7×10^{-3}
90	surfac	8.5×10^{-3}
91	phylogeni	8.2×10^{-3}
92	recombin	8.2×10^{-3}
93	microarray	8.2×10^{-3}
94	ancestr	8.2×10^{-3}
95	herit	8.1×10^{-3}
96	wide	8×10^{-3}
97	bind	7.8×10^{-3}
98	splice	7.8×10^{-3}
99	plant	7.7×10^{-3}
100	mous	7.7×10^{-3}

TABLE D.97. The list of the top 100 words in the category Geochemistry and Geophysics with RIGs

No.	Word	RIG	No.	Word	RIG
1	seismic	9.4×10^{-2}	51	radar	1.7×10^{-2}
2	rock	8.5×10^{-2}	52	northern	1.6×10^{-2}
3	mantl	6.9×10^{-2}	53	constrain	1.6×10^{-2}
4	crust	5.2×10^{-2}	54	model	1.6×10^{-2}
5	crustal	4.9×10^{-2}	55	orogen	1.6×10^{-2}
6	earthquak	4.7×10^{-2}	56	composit	1.6×10^{-2}
7	zone	4.6×10^{-2}	57	conclus	1.5×10^{-2}
8	isotop	4.5×10^{-2}	58	belt	1.5×10^{-2}
9	subduct	4.2×10^{-2}	59	upper	1.5×10^{-2}
10	tecton	4.1×10^{-2}	60	near	1.5×10^{-2}
11	geolog	4×10^{-2}	61	emplac	1.5×10^{-2}
12	lithospher	4×10^{-2}	62	estim	1.5×10^{-2}
13	earth	3.9×10^{-2}	63	fluid	1.5×10^{-2}
14	magma	3.8×10^{-2}	64	variat	1.4×10^{-2}
15	miner	3.8×10^{-2}	65	subsurfac	1.4×10^{-2}
16	geochem	3.8×10^{-2}	66	station	1.4×10^{-2}
17	magmat	3.7×10^{-2}	67	deform	1.4×10^{-2}
18	fault	3.5×10^{-2}	68	observ	1.4×10^{-2}
19	ocean	3.4×10^{-2}	69	igneous	1.4×10^{-2}
20	depth	3.3×10^{-2}	70	event	1.4×10^{-2}
21	volcan	3.2×10^{-2}	71	erupt	1.4×10^{-2}
22	data	2.9×10^{-2}	72	record	1.4×10^{-2}
23	basin	2.9×10^{-2}	73	eastern	1.4×10^{-2}
24	sediment	2.8×10^{-2}	74	locat	1.4×10^{-2}
25	basalt	2.7×10^{-2}	75	plagioclas	1.3×10^{-2}
26	melt	2.5×10^{-2}	76	resolut	1.3×10^{-2}
27	beneath	2.5×10^{-2}	77	silic	1.3×10^{-2}
28	continent	2.5×10^{-2}	78	surfac	1.3×10^{-2}
29	geophys	2.3×10^{-2}	79	granit	1.3×10^{-2}
30	interpret	2.2×10^{-2}	80	mineralog	1.3×10^{-2}
31	sedimentari	2.1×10^{-2}	81	deep	1.3×10^{-2}
32	slip	2.1×10^{-2}	82	litholog	1.3×10^{-2}
33	olivin	2.1×10^{-2}	83	enrich	1.2×10^{-2}
34	shallow	2.1×10^{-2}	84	east	1.2×10^{-2}
35	patient	2.1×10^{-2}	85	occur	1.2×10^{-2}
36	metamorph	2.1×10^{-2}	86	clinic	1.2×10^{-2}
37	sourc	2×10^{-2}	87	reservoir	1.2×10^{-2}
38	veloc	2×10^{-2}	88	intrus	1.2×10^{-2}
39	along	1.9×10^{-2}	89	ree	1.2×10^{-2}
40	synthet	1.9×10^{-2}	90	craton	1.2×10^{-2}
41	similar	1.9×10^{-2}	91	south	1.2×10^{-2}
42	wave	1.9×10^{-2}	92	satellit	1.2×10^{-2}
43	region	1.9×10^{-2}	93	dip	1.2×10^{-2}
44	anomali	1.8×10^{-2}	94	slab	1.2×10^{-2}
45	mafic	1.8×10^{-2}	95	ridg	1.2×10^{-2}
46	southern	1.8×10^{-2}	96	thrust	1.2×10^{-2}
47	trace	1.8×10^{-2}	97	vertic	1.1×10^{-2}
48	zircon	1.7×10^{-2}	98	uplift	1.1×10^{-2}
49	north	1.7×10^{-2}	99	reflect	1.1×10^{-2}
50	invers	1.7×10^{-2}	100	sea	1.1×10^{-2}

TABLE D.98. The list of the top 100 words in the category Geography with RIGs

No.	Word	RIG	No.	Word	RIG
1	geographi	7.9×10^{-2}	51	climat	1.2×10^{-2}
2	urban	7.6×10^{-2}	52	method	1.2×10^{-2}
3	citi	5.7×10^{-2}	53	agricultur	1.2×10^{-2}
4	polit	5.3×10^{-2}	54	gis	1.2×10^{-2}
5	geograph	5×10^{-2}	55	metropolitan	1.2×10^{-2}
6	spatial	4.7×10^{-2}	56	contest	1.2×10^{-2}
7	argu	4.2×10^{-2}	57	countri	1.2×10^{-2}
8	polici	4.2×10^{-2}	58	interview	1.2×10^{-2}
9	social	4.1×10^{-2}	59	labour	1.2×10^{-2}
10	econom	3.9×10^{-2}	60	migrant	1.2×10^{-2}
11	articl	3.6×10^{-2}	61	perspect	1.1×10^{-2}
12	land	3.3×10^{-2}	62	conclus	1.1×10^{-2}
13	draw	3.2×10^{-2}	63	treatment	1.1×10^{-2}
14	landscap	2.9×10^{-2}	64	debat	1.1×10^{-2}
15	govern	2.7×10^{-2}	65	narrat	1.1×10^{-2}
16	economi	2×10^{-2}	66	examin	1.1×10^{-2}
17	place	2×10^{-2}	67	environment	1.1×10^{-2}
18	rural	2×10^{-2}	68	sector	1.1×10^{-2}
19	research	2×10^{-2}	69	livelihood	1×10^{-2}
20	area	1.9×10^{-2}	70	clinic	1×10^{-2}
21	explor	1.9×10^{-2}	71	world	1×10^{-2}
22	actor	1.9×10^{-2}	72	cultur	1×10^{-2}
23	territori	1.8×10^{-2}	73	result	1×10^{-2}
24	region	1.8×10^{-2}	74	develop	9.9×10^{-3}
25	space	1.8×10^{-2}	75	particular	9.9×10^{-3}
26	patient	1.7×10^{-2}	76	emerg	9.7×10^{-3}
27	communiti	1.7×10^{-2}	77	ecolog	9.7×10^{-3}
28	context	1.7×10^{-2}	78	question	9.6×10^{-3}
29	focus	1.7×10^{-2}	79	conceptu	9.6×10^{-3}
30	public	1.7×10^{-2}	80	concept	9.3×10^{-3}
31	discours	1.6×10^{-2}	81	residenti	9.3×10^{-3}
32	paper	1.6×10^{-2}	82	protein	9.1×10^{-3}
33	socio	1.6×10^{-2}	83	map	9×10^{-3}
34	nation	1.6×10^{-2}	84	within	8.9×10^{-3}
35	plan	1.6×10^{-2}	85	resid	8.9×10^{-3}
36	understand	1.6×10^{-2}	86	natur	8.8×10^{-3}
37	engag	1.5×10^{-2}	87	across	8.8×10^{-3}
38	capit	1.5×10^{-2}	88	transnat	8.7×10^{-3}
39	way	1.5×10^{-2}	89	attent	8.5×10^{-3}
40	market	1.5×10^{-2}	90	sustain	8.4×10^{-3}
41	local	1.4×10^{-2}	91	farmer	8.1×10^{-3}
42	global	1.4×10^{-2}	92	negoti	7.9×10^{-3}
43	peopl	1.4×10^{-2}	93	experiment	7.9×10^{-3}
44	cell	1.4×10^{-2}	94	project	7.8×10^{-3}
45	practic	1.4×10^{-2}	95	embodi	7.8×10^{-3}
46	contemporari	1.4×10^{-2}	96	opportun	7.8×10^{-3}
47	neoliber	1.3×10^{-2}	97	electron	7.7×10^{-3}
48	empir	1.3×10^{-2}	98	highlight	7.7×10^{-3}
49	histor	1.3×10^{-2}	99	privat	7.6×10^{-3}
50	institut	1.2×10^{-2}	100	live	7.6×10^{-3}

TABLE D.99. The list of the top 100 words in the category Geography, Physical with RIGs

No.	Word	RIG
1	climat	7.2×10^{-2}
2	sediment	6.2×10^{-2}
3	holocen	5.1×10^{-2}
4	area	5.1×10^{-2}
5	ice	4.5×10^{-2}
6	glacial	4.2×10^{-2}
7	spatial	4×10^{-2}
8	land	3.9×10^{-2}
9	glacier	3.6×10^{-2}
10	pleistocen	3.4×10^{-2}
11	veget	3.3×10^{-2}
12	sea	3.3×10^{-2}
13	river	3.1×10^{-2}
14	region	3.1×10^{-2}
15	landscap	3×10^{-2}
16	basin	2.7×10^{-2}
17	remot	2.6×10^{-2}
18	mountain	2.6×10^{-2}
19	date	2.6×10^{-2}
20	map	2.6×10^{-2}
21	data	2.5×10^{-2}
22	cover	2.5×10^{-2}
23	satellit	2.5×10^{-2}
24	forest	2.4×10^{-2}
25	lake	2.4×10^{-2}
26	southern	2.4×10^{-2}
27	geomorpholog	2.4×10^{-2}
28	sar	2.4×10^{-2}
29	northern	2.4×10^{-2}
30	eros	2.4×10^{-2}
31	resolut	2.3×10^{-2}
32	locat	2.3×10^{-2}
33	cal	2.3×10^{-2}
34	late	2.2×10^{-2}
35	north	2.2×10^{-2}
36	reconstruct	2.2×10^{-2}
37	chang	2.2×10^{-2}
38	slope	2.1×10^{-2}
39	coastal	2.1×10^{-2}
40	record	2.1×10^{-2}
41	chronolog	1.9×10^{-2}
42	gis	1.9×10^{-2}
43	assemblag	1.9×10^{-2}
44	patient	1.8×10^{-2}
45	fluvial	1.8×10^{-2}
46	deposit	1.8×10^{-2}
47	imageri	1.7×10^{-2}
48	radar	1.7×10^{-2}
49	warm	1.7×10^{-2}
50	ocean	1.7×10^{-2}

No.	Word	RIG
51	site	1.7×10^{-2}
52	sedimentari	1.7×10^{-2}
53	radiocarbon	1.7×10^{-2}
54	valley	1.7×10^{-2}
55	topograph	1.6×10^{-2}
56	eastern	1.6×10^{-2}
57	south	1.6×10^{-2}
58	geograph	1.6×10^{-2}
59	period	1.6×10^{-2}
60	stratigraph	1.5×10^{-2}
61	season	1.5×10^{-2}
62	east	1.5×10^{-2}
63	coast	1.5×10^{-2}
64	dure	1.5×10^{-2}
65	last	1.5×10^{-2}
66	hydrolog	1.5×10^{-2}
67	marin	1.5×10^{-2}
68	summer	1.4×10^{-2}
69	soil	1.4×10^{-2}
70	terrestri	1.4×10^{-2}
71	tempor	1.3×10^{-2}
72	flood	1.3×10^{-2}
73	water	1.3×10^{-2}
74	scale	1.3×10^{-2}
75	sens	1.3×10^{-2}
76	proxi	1.3×10^{-2}
77	archaeolog	1.3×10^{-2}
78	pattern	1.3×10^{-2}
79	isotop	1.3×10^{-2}
80	fossil	1.3×10^{-2}
81	precipit	1.3×10^{-2}
82	retreat	1.2×10^{-2}
83	urban	1.2×10^{-2}
84	pollen	1.2×10^{-2}
85	monsoon	1.2×10^{-2}
86	ecolog	1.2×10^{-2}
87	west	1.2×10^{-2}
88	clinic	1.2×10^{-2}
89	snow	1.2×10^{-2}
90	ecosystem	1.1×10^{-2}
91	island	1.1×10^{-2}
92	treatment	1.1×10^{-2}
93	elev	1.1×10^{-2}
94	past	1.1×10^{-2}
95	geolog	1.1×10^{-2}
96	along	1.1×10^{-2}
97	cell	1.1×10^{-2}
98	landsat	1.1×10^{-2}
99	zone	1.1×10^{-2}
100	middl	1.1×10^{-2}

Table D.100. The list of the top 100 words in the category Geology with RIGs

No.	Word	RIG	No.	Word	RIG
1	rock	1.2×10^{-1}	51	quartz	2.5×10^{-2}
2	deposit	9.3×10^{-2}	52	outcrop	2.4×10^{-2}
3	sediment	8.6×10^{-2}	53	granit	2.4×10^{-2}
4	basin	8.2×10^{-2}	54	strata	2.4×10^{-2}
5	sedimentari	7.4×10^{-2}	55	preserv	2.4×10^{-2}
6	stratigraph	6.9×10^{-2}	56	miocen	2.4×10^{-2}
7	late	6.4×10^{-2}	57	south	2.3×10^{-2}
8	faci	6.1×10^{-2}	58	magma	2.3×10^{-2}
9	tecton	6×10^{-2}	59	domin	2.3×10^{-2}
10	cretac	6×10^{-2}	60	crustal	2.3×10^{-2}
11	zone	5.5×10^{-2}	61	ore	2.3×10^{-2}
12	format	4.9×10^{-2}	62	carbon	2.3×10^{-2}
13	geolog	4.8×10^{-2}	63	sedimentolog	2.3×10^{-2}
14	isotop	4.8×10^{-2}	64	mineralog	2.3×10^{-2}
15	zircon	4.7×10^{-2}	65	emplac	2.2×10^{-2}
16	upper	4.6×10^{-2}	66	permian	2.2×10^{-2}
17	assemblag	4.5×10^{-2}	67	dure	2.1×10^{-2}
18	miner	4.3×10^{-2}	68	litholog	2.1×10^{-2}
19	geochem	4.1×10^{-2}	69	stratigraphi	2.1×10^{-2}
20	volcan	4.1×10^{-2}	70	method	2×10^{-2}
21	magmat	4×10^{-2}	71	geochemistri	2×10^{-2}
22	marin	3.6×10^{-2}	72	grain	2×10^{-2}
23	sandston	3.6×10^{-2}	73	ocean	2×10^{-2}
24	metamorph	3.5×10^{-2}	74	uplift	1.9×10^{-2}
25	jurass	3.4×10^{-2}	75	form	1.9×10^{-2}
26	middl	3.3×10^{-2}	76	date	1.9×10^{-2}
27	continent	3.3×10^{-2}	77	fauna	1.9×10^{-2}
28	shallow	3.3×10^{-2}	78	eocen	1.9×10^{-2}
29	crust	3.2×10^{-2}	79	cambrian	1.9×10^{-2}
30	earli	3.2×10^{-2}	80	occur	1.8×10^{-2}
31	north	3.2×10^{-2}	81	western	1.8×10^{-2}
32	fossil	3.2×10^{-2}	82	central	1.8×10^{-2}
33	southern	3.2×10^{-2}	83	rich	1.8×10^{-2}
34	limeston	3.1×10^{-2}	84	basalt	1.8×10^{-2}
35	northern	3×10^{-2}	85	pluton	1.8×10^{-2}
36	interpret	3×10^{-2}	86	transgress	1.8×10^{-2}
37	belt	3×10^{-2}	87	along	1.7×10^{-2}
38	sea	3×10^{-2}	88	sand	1.7×10^{-2}
39	orogen	2.9×10^{-2}	89	geochronolog	1.7×10^{-2}
40	eastern	2.9×10^{-2}	90	climat	1.7×10^{-2}
41	subduct	2.9×10^{-2}	91	water	1.7×10^{-2}
42	part	2.9×10^{-2}	92	area	1.7×10^{-2}
43	margin	2.8×10^{-2}	93	repres	1.7×10^{-2}
44	ordovician	2.8×10^{-2}	94	fault	1.7×10^{-2}
45	record	2.7×10^{-2}	95	igneous	1.7×10^{-2}
46	triassic	2.6×10^{-2}	96	river	1.7×10^{-2}
47	bed	2.6×10^{-2}	97	clay	1.7×10^{-2}
48	calcit	2.6×10^{-2}	98	evolut	1.7×10^{-2}
49	fluvial	2.5×10^{-2}	99	eros	1.6×10^{-2}
50	mantl	2.5×10^{-2}	100	patient	1.6×10^{-2}

Table D.101. The list of the top 100 words in the category Geosciences, Multidisciplinary with RIGs

No.	Word	RIG
1	rock	5.8×10^{-2}
2	sediment	5.7×10^{-2}
3	basin	5.4×10^{-2}
4	climat	4.1×10^{-2}
5	geolog	3.7×10^{-2}
6	area	3.5×10^{-2}
7	water	3.4×10^{-2}
8	soil	3.4×10^{-2}
9	sea	3.3×10^{-2}
10	zone	3.2×10^{-2}
11	tecton	3.1×10^{-2}
12	river	3.1×10^{-2}
13	sedimentari	3×10^{-2}
14	ocean	2.7×10^{-2}
15	region	2.7×10^{-2}
16	southern	2.5×10^{-2}
17	deposit	2.5×10^{-2}
18	seismic	2.5×10^{-2}
19	volcan	2.4×10^{-2}
20	patient	2.4×10^{-2}
21	geochem	2.4×10^{-2}
22	north	2.4×10^{-2}
23	northern	2.4×10^{-2}
24	hydrolog	2.3×10^{-2}
25	ice	2.3×10^{-2}
26	land	2.3×10^{-2}
27	groundwat	2.2×10^{-2}
28	holocen	2.2×10^{-2}
29	isotop	2.2×10^{-2}
30	stratigraph	2.1×10^{-2}
31	depth	2.1×10^{-2}
32	shallow	2.1×10^{-2}
33	late	2.1×10^{-2}
34	continent	2×10^{-2}
35	slope	2×10^{-2}
36	eastern	1.9×10^{-2}
37	spatial	1.9×10^{-2}
38	dure	1.8×10^{-2}
39	glacial	1.8×10^{-2}
40	conclus	1.8×10^{-2}
41	data	1.7×10^{-2}
42	south	1.7×10^{-2}
43	locat	1.7×10^{-2}
44	eros	1.7×10^{-2}
45	magmat	1.6×10^{-2}
46	earthquak	1.6×10^{-2}
47	miner	1.5×10^{-2}
48	upper	1.5×10^{-2}
49	east	1.5×10^{-2}
50	mountain	1.5×10^{-2}
51	flood	1.5×10^{-2}
52	lake	1.5×10^{-2}
53	marin	1.5×10^{-2}
54	faci	1.5×10^{-2}
55	aquif	1.5×10^{-2}
56	clinic	1.4×10^{-2}
57	event	1.4×10^{-2}
58	assemblag	1.4×10^{-2}
59	along	1.4×10^{-2}
60	precipit	1.4×10^{-2}
61	rainfal	1.4×10^{-2}
62	sand	1.4×10^{-2}
63	earth	1.4×10^{-2}
64	fault	1.4×10^{-2}
65	record	1.4×10^{-2}
66	season	1.4×10^{-2}
67	coastal	1.4×10^{-2}
68	catchment	1.4×10^{-2}
69	belt	1.3×10^{-2}
70	cell	1.3×10^{-2}
71	pleistocen	1.3×10^{-2}
72	satellit	1.3×10^{-2}
73	crust	1.3×10^{-2}
74	date	1.3×10^{-2}
75	reservoir	1.3×10^{-2}
76	surfac	1.3×10^{-2}
77	crustal	1.3×10^{-2}
78	clay	1.2×10^{-2}
79	interpret	1.2×10^{-2}
80	landslid	1.2×10^{-2}
81	domin	1.2×10^{-2}
82	weather	1.2×10^{-2}
83	period	1.2×10^{-2}
84	zircon	1.2×10^{-2}
85	glacier	1.2×10^{-2}
86	litholog	1.2×10^{-2}
87	diseas	1.2×10^{-2}
88	cover	1.2×10^{-2}
89	indic	1.2×10^{-2}
90	part	1.2×10^{-2}
91	metamorph	1.2×10^{-2}
92	fluvial	1.1×10^{-2}
93	warm	1.1×10^{-2}
94	granit	1.1×10^{-2}
95	site	1.1×10^{-2}
96	atmospher	1.1×10^{-2}
97	cretac	1.1×10^{-2}
98	geomorpholog	1.1×10^{-2}
99	deep	1.1×10^{-2}
100	west	1.1×10^{-2}

Table D.102. The list of the top 100 words in the category Geriatrics and Gerontology with RIGs

No.	Word	RIG
1	older	1.7×10^{-1}
2	age	1.6×10^{-1}
3	elder	1×10^{-1}
4	dementia	7.5×10^{-2}
5	cognit	6.9×10^{-2}
6	adult	6.5×10^{-2}
7	conclus	6×10^{-2}
8	particip	5.7×10^{-2}
9	geriatr	5.2×10^{-2}
10	alzheim	5×10^{-2}
11	associ	4.8×10^{-2}
12	year	4.4×10^{-2}
13	health	4×10^{-2}
14	dwell	4×10^{-2}
15	impair	3.7×10^{-2}
16	diseas	3.3×10^{-2}
17	declin	3.3×10^{-2}
18	depress	3.1×10^{-2}
19	assess	3×10^{-2}
20	care	2.9×10^{-2}
21	object	2.8×10^{-2}
22	score	2.7×10^{-2}
23	risk	2.6×10^{-2}
24	peopl	2.5×10^{-2}
25	mental	2.5×10^{-2}
26	live	2.5×10^{-2}
27	home	2.4×10^{-2}
28	life	2.3×10^{-2}
29	intervent	2.3×10^{-2}
30	studi	2.2×10^{-2}
31	paper	2.2×10^{-2}
32	patient	2.2×10^{-2}
33	nurs	2.1×10^{-2}
34	women	2.1×10^{-2}
35	mini	2.1×10^{-2}
36	communiti	2.1×10^{-2}
37	baselin	2×10^{-2}
38	old	2×10^{-2}
39	individu	2×10^{-2}
40	regress	2×10^{-2}
41	examin	2×10^{-2}
42	status	1.9×10^{-2}
43	symptom	1.9×10^{-2}
44	men	1.9×10^{-2}
45	signific	1.9×10^{-2}
46	cohort	1.8×10^{-2}
47	popul	1.8×10^{-2}
48	background	1.8×10^{-2}
49	group	1.8×10^{-2}
50	clinic	1.8×10^{-2}

No.	Word	RIG
51	medic	1.8×10^{-2}
52	person	1.7×10^{-2}
53	caregiv	1.6×10^{-2}
54	adjust	1.5×10^{-2}
55	daili	1.5×10^{-2}
56	comorbid	1.5×10^{-2}
57	result	1.5×10^{-2}
58	memori	1.4×10^{-2}
59	relat	1.4×10^{-2}
60	resid	1.4×10^{-2}
61	physic	1.4×10^{-2}
62	function	1.3×10^{-2}
63	month	1.3×10^{-2}
64	measur	1.3×10^{-2}
65	outcom	1.3×10^{-2}
66	healthi	1.3×10^{-2}
67	hospit	1.3×10^{-2}
68	younger	1.2×10^{-2}
69	confid	1.2×10^{-2}
70	mild	1.2×10^{-2}
71	propos	1.2×10^{-2}
72	preval	1.2×10^{-2}
73	simul	1.2×10^{-2}
74	fall	1.2×10^{-2}
75	subject	1.2×10^{-2}
76	whether	1.2×10^{-2}
77	section	1.1×10^{-2}
78	may	1.1×10^{-2}
79	aim	1.1×10^{-2}
80	neuropsycholog	1.1×10^{-2}
81	young	1.1×10^{-2}
82	longitudin	1.1×10^{-2}
83	factor	1.1×10^{-2}
84	temperatur	1.1×10^{-2}
85	includ	1×10^{-2}
86	sex	1×10^{-2}
87	walk	1×10^{-2}
88	brain	1×10^{-2}
89	odd	1×10^{-2}
90	logist	1×10^{-2}
91	interview	9.9×10^{-3}
92	amyloid	9.9×10^{-3}
93	method	9.6×10^{-3}
94	muscl	9.5×10^{-3}
95	questionnair	9.5×10^{-3}
96	among	9.4×10^{-3}
97	selfreport	9.3×10^{-3}
98	disabl	9.2×10^{-3}
99	independ	9×10^{-3}
100	increas	8.8×10^{-3}

TABLE D.103. The list of the top 100 words in the category Gerontology with RIGs

No.	Word	RIG
1	older	2.5×10^{-1}
2	age	1.4×10^{-1}
3	adult	1.1×10^{-1}
4	dementia	8.5×10^{-2}
5	particip	8.4×10^{-2}
6	elder	7.3×10^{-2}
7	health	7.3×10^{-2}
8	cognit	6×10^{-2}
9	care	5.7×10^{-2}
10	dwell	4.9×10^{-2}
11	geriatr	4.8×10^{-2}
12	conclus	4.8×10^{-2}
13	depress	4.7×10^{-2}
14	life	4.5×10^{-2}
15	home	4.2×10^{-2}
16	communiti	3.9×10^{-2}
17	peopl	3.8×10^{-2}
18	person	3.8×10^{-2}
19	live	3.8×10^{-2}
20	object	3.7×10^{-2}
21	caregiv	3.4×10^{-2}
22	examin	3.4×10^{-2}
23	mental	3.3×10^{-2}
24	nurs	3.3×10^{-2}
25	associ	3.1×10^{-2}
26	interview	3.1×10^{-2}
27	year	3.1×10^{-2}
28	individu	2.9×10^{-2}
29	social	2.8×10^{-2}
30	intervent	2.8×10^{-2}
31	assess	2.7×10^{-2}
32	score	2.5×10^{-2}
33	resid	2.4×10^{-2}
34	symptom	2.4×10^{-2}
35	studi	2.4×10^{-2}
36	regress	2.3×10^{-2}
37	impair	2.3×10^{-2}
38	physic	2.2×10^{-2}
39	declin	2.1×10^{-2}
40	retir	2×10^{-2}
41	medic	2×10^{-2}
42	baselin	1.9×10^{-2}
43	status	1.9×10^{-2}
44	women	1.8×10^{-2}
45	alzheim	1.8×10^{-2}
46	longitudin	1.8×10^{-2}
47	men	1.7×10^{-2}
48	educ	1.7×10^{-2}
49	mini	1.7×10^{-2}
50	risk	1.7×10^{-2}

No.	Word	RIG
51	daili	1.7×10^{-2}
52	among	1.6×10^{-2}
53	cohort	1.6×10^{-2}
54	method	1.6×10^{-2}
55	group	1.6×10^{-2}
56	adjust	1.6×10^{-2}
57	result	1.6×10^{-2}
58	popul	1.5×10^{-2}
59	outcom	1.5×10^{-2}
60	selfreport	1.5×10^{-2}
61	younger	1.4×10^{-2}
62	survey	1.4×10^{-2}
63	measur	1.4×10^{-2}
64	confid	1.4×10^{-2}
65	propos	1.3×10^{-2}
66	memori	1.3×10^{-2}
67	comorbid	1.3×10^{-2}
68	paper	1.2×10^{-2}
69	logist	1.2×10^{-2}
70	old	1.2×10^{-2}
71	disabl	1.2×10^{-2}
72	relationship	1.2×10^{-2}
73	greater	1.2×10^{-2}
74	demograph	1.2×10^{-2}
75	preval	1.1×10^{-2}
76	psycholog	1.1×10^{-2}
77	predictor	1.1×10^{-2}
78	questionnair	1.1×10^{-2}
79	staff	1.1×10^{-2}
80	aim	1.1×10^{-2}
81	whether	1.1×10^{-2}
82	clinic	1.1×10^{-2}
83	gender	1.1×10^{-2}
84	surfac	1.1×10^{-2}
85	simul	1.1×10^{-2}
86	temperatur	1×10^{-2}
87	section	1×10^{-2}
88	item	1×10^{-2}
89	odd	1×10^{-2}
90	servic	1×10^{-2}
91	cell	1×10^{-2}
92	hospit	1×10^{-2}
93	need	9.9×10^{-3}
94	support	9.7×10^{-3}
95	relat	9.7×10^{-3}
96	perceiv	9.6×10^{-3}
97	sociodemograph	9.3×10^{-3}
98	spous	9.3×10^{-3}
99	famili	9.1×10^{-3}
100	factor	9.1×10^{-3}

Table D.104. The list of the top 100 words in the category Green and Sustainable Science and Technology with RIGs

No.	Word	RIG	No.	Word	RIG
1	energi	3.9×10^{-2}	51	system	7.2×10^{-3}
2	sustain	3.7×10^{-2}	52	oper	7.2×10^{-3}
3	environment	3.3×10^{-2}	53	yield	6.8×10^{-3}
4	renew	2.7×10^{-2}	54	develop	6.7×10^{-3}
5	co2	2.3×10^{-2}	55	gene	6.7×10^{-3}
6	product	2.3×10^{-2}	56	plant	6.5×10^{-3}
7	econom	2.2×10^{-2}	57	biodiesel	6.5×10^{-3}
8	fuel	2.2×10^{-2}	58	climat	6.5×10^{-3}
9	patient	1.9×10^{-2}	59	solvent	6.4×10^{-3}
10	wast	1.8×10^{-2}	60	age	6.3×10^{-3}
11	turbin	1.7×10^{-2}	61	express	6.3×10^{-3}
12	carbon	1.6×10^{-2}	62	catalyt	6.1×10^{-3}
13	catalyst	1.6×10^{-2}	63	save	6×10^{-3}
14	solar	1.6×10^{-2}	64	temperatur	5.8×10^{-3}
15	wind	1.5×10^{-2}	65	coal	5.6×10^{-3}
16	biomass	1.5×10^{-2}	66	biofuel	5.5×10^{-3}
17	electr	1.5×10^{-2}	67	protein	5.5×10^{-3}
18	effici	1.5×10^{-2}	68	reaction	5.5×10^{-3}
19	emiss	1.4×10^{-2}	69	cancer	5.5×10^{-3}
20	conclus	1.4×10^{-2}	70	instal	5.4×10^{-3}
21	industri	1.4×10^{-2}	71	price	5.4×10^{-3}
22	power	1.4×10^{-2}	72	pollut	5.4×10^{-3}
23	consumpt	1.2×10^{-2}	73	agricultur	5.3×10^{-3}
24	cost	1.2×10^{-2}	74	dioxid	5.3×10^{-3}
25	gas	1.2×10^{-2}	75	diesel	5.3×10^{-3}
26	demand	1.1×10^{-2}	76	process	5.3×10^{-3}
27	recycl	1.1×10^{-2}	77	produc	5.1×10^{-3}
28	greenhous	1.1×10^{-2}	78	therapi	5×10^{-3}
29	clinic	1.1×10^{-2}	79	lignin	5×10^{-3}
30	photovolta	1.1×10^{-2}	80	wastewat	4.9×10^{-3}
31	water	1.1×10^{-2}	81	paper	4.9×10^{-3}
32	polici	8.8×10^{-3}	82	build	4.9×10^{-3}
33	diseas	8.8×10^{-3}	83	capac	4.8×10^{-3}
34	impact	8.6×10^{-3}	84	detect	4.8×10^{-3}
35	green	8.5×10^{-3}	85	imag	4.8×10^{-3}
36	fossil	8.2×10^{-3}	86	tissu	4.6×10^{-3}
37	background	8.2×10^{-3}	87	sourc	4.6×10^{-3}
38	convers	8.2×10^{-3}	88	convert	4.5×10^{-3}
39	heat	8.1×10^{-3}	89	stakehold	4.5×10^{-3}
40	feedstock	8×10^{-3}	90	invest	4.5×10^{-3}
41	cycl	8×10^{-3}	91	cellulos	4.4×10^{-3}
42	suppli	7.8×10^{-3}	92	blood	4.4×10^{-3}
43	technolog	7.8×10^{-3}	93	optim	4.3×10^{-3}
44	resourc	7.6×10^{-3}	94	induc	4.3×10^{-3}
45	storag	7.5×10^{-3}	95	infect	4.3×10^{-3}
46	generat	7.3×10^{-3}	96	scenario	4.3×10^{-3}
47	farm	7.3×10^{-3}	97	aqueous	4.3×10^{-3}
48	sector	7.3×10^{-3}	98	signal	4.2×10^{-3}
49	oil	7.2×10^{-3}	99	china	4.2×10^{-3}
50	grid	7.2×10^{-3}	100	tumor	4.2×10^{-3}

Table D.105. The list of the top 100 words in the category Health Care Sciences and Services with RIGs

No.	Word	RIG	No.	Word	RIG
1	health	1.5×10^{-1}	51	among	1.4×10^{-2}
2	care	1.5×10^{-1}	52	public	1.4×10^{-2}
3	conclus	7.5×10^{-2}	53	manag	1.4×10^{-2}
4	medic	7.5×10^{-2}	54	clinician	1.4×10^{-2}
5	patient	7.5×10^{-2}	55	program	1.4×10^{-2}
6	background	4.8×10^{-2}	56	semistructur	1.4×10^{-2}
7	hospit	4.6×10^{-2}	57	studi	1.4×10^{-2}
8	particip	3.9×10^{-2}	58	conduct	1.3×10^{-2}
9	healthcar	3.8×10^{-2}	59	qualit	1.3×10^{-2}
10	physician	3.7×10^{-2}	60	respond	1.3×10^{-2}
11	object	3.7×10^{-2}	61	primari	1.3×10^{-2}
12	clinic	3.7×10^{-2}	62	receiv	1.3×10^{-2}
13	servic	3.7×10^{-2}	63	regress	1.3×10^{-2}
14	interview	3.4×10^{-2}	64	cost	1.3×10^{-2}
15	qualiti	3.2×10^{-2}	65	ill	1.3×10^{-2}
16	intervent	3.2×10^{-2}	66	access	1.3×10^{-2}
17	outcom	3.2×10^{-2}	67	popul	1.2×10^{-2}
18	survey	3.1×10^{-2}	68	home	1.2×10^{-2}
19	assess	2.7×10^{-2}	69	aim	1.2×10^{-2}
20	profession	2.7×10^{-2}	70	practition	1.2×10^{-2}
21	palliat	2.5×10^{-2}	71	perceiv	1.2×10^{-2}
22	practic	2.5×10^{-2}	72	medicaid	1.2×10^{-2}
23	need	2.5×10^{-2}	73	cell	1.2×10^{-2}
24	questionnair	2.5×10^{-2}	74	communiti	1.2×10^{-2}
25	method	2.4×10^{-2}	75	temperatur	1.2×10^{-2}
26	nurs	2.3×10^{-2}	76	across	1.2×10^{-2}
27	educ	2.3×10^{-2}	77	consult	1.1×10^{-2}
28	life	2×10^{-2}	78	surfac	1.1×10^{-2}
29	inform	2×10^{-2}	79	recommend	1.1×10^{-2}
30	data	1.9×10^{-2}	80	outpati	1.1×10^{-2}
31	research	1.9×10^{-2}	81	set	1.1×10^{-2}
32	score	1.9×10^{-2}	82	peopl	1.1×10^{-2}
33	nation	1.8×10^{-2}	83	plan	1.1×10^{-2}
34	year	1.8×10^{-2}	84	doctor	1.1×10^{-2}
35	medicar	1.7×10^{-2}	85	age	1.1×10^{-2}
36	staff	1.7×10^{-2}	86	train	1.1×10^{-2}
37	support	1.7×10^{-2}	87	theme	1×10^{-2}
38	insur	1.6×10^{-2}	88	person	1×10^{-2}
39	provid	1.6×10^{-2}	89	review	1×10^{-2}
40	polici	1.6×10^{-2}	90	evid	1×10^{-2}
41	visit	1.6×10^{-2}	91	cancer	1×10^{-2}
42	team	1.6×10^{-2}	92	benefit	9.9×10^{-3}
43	item	1.6×10^{-2}	93	month	9.8×10^{-3}
44	medicin	1.6×10^{-2}	94	examin	9.8×10^{-3}
45	improv	1.6×10^{-2}	95	group	9.8×10^{-3}
46	includ	1.5×10^{-2}	96	report	9.7×10^{-3}
47	inpati	1.5×10^{-2}	97	result	9.7×10^{-3}
48	identifi	1.5×10^{-2}	98	trial	9.6×10^{-3}
49	decis	1.5×10^{-2}	99	satisfact	9.5×10^{-3}
50	implement	1.5×10^{-2}	100	impact	9.4×10^{-3}

Table D.106. The list of the top 100 words in the category Health Policy and Services with RIGs

No.	Word	RIG
1	health	1.9×10^{-1}
2	care	1.4×10^{-1}
3	servic	5.1×10^{-2}
4	interview	4.2×10^{-2}
5	medic	4×10^{-2}
6	patient	4×10^{-2}
7	hospit	3.7×10^{-2}
8	survey	3.7×10^{-2}
9	particip	3.5×10^{-2}
10	healthcar	3.2×10^{-2}
11	mental	3.2×10^{-2}
12	polici	3.2×10^{-2}
13	qualiti	3.1×10^{-2}
14	intervent	3.1×10^{-2}
15	insur	2.6×10^{-2}
16	outcom	2.6×10^{-2}
17	nation	2.5×10^{-2}
18	need	2.4×10^{-2}
19	among	2.4×10^{-2}
20	conclus	2.4×10^{-2}
21	physician	2.3×10^{-2}
22	practic	2.2×10^{-2}
23	public	2.2×10^{-2}
24	object	2.2×10^{-2}
25	communiti	2.1×10^{-2}
26	profession	2.1×10^{-2}
27	assess	2×10^{-2}
28	questionnair	2×10^{-2}
29	medicar	1.9×10^{-2}
30	peopl	1.9×10^{-2}
31	medicaid	1.9×10^{-2}
32	clinic	1.9×10^{-2}
33	hiv	1.8×10^{-2}
34	nurs	1.8×10^{-2}
35	educ	1.8×10^{-2}
36	life	1.8×10^{-2}
37	popul	1.8×10^{-2}
38	examin	1.7×10^{-2}
39	program	1.7×10^{-2}
40	staff	1.7×10^{-2}
41	social	1.7×10^{-2}
42	research	1.6×10^{-2}
43	year	1.6×10^{-2}
44	regress	1.6×10^{-2}
45	data	1.6×10^{-2}
46	incom	1.5×10^{-2}
47	background	1.5×10^{-2}
48	access	1.5×10^{-2}
49	implement	1.5×10^{-2}
50	ill	1.5×10^{-2}

No.	Word	RIG
51	perceiv	1.5×10^{-2}
52	individu	1.4×10^{-2}
53	support	1.4×10^{-2}
54	visit	1.4×10^{-2}
55	inform	1.4×10^{-2}
56	age	1.3×10^{-2}
57	provid	1.3×10^{-2}
58	item	1.3×10^{-2}
59	score	1.3×10^{-2}
60	countri	1.3×10^{-2}
61	cost	1.3×10^{-2}
62	evid	1.3×10^{-2}
63	qualit	1.3×10^{-2}
64	decis	1.3×10^{-2}
65	cell	1.2×10^{-2}
66	temperatur	1.2×10^{-2}
67	surfac	1.2×10^{-2}
68	fund	1.2×10^{-2}
69	engag	1.2×10^{-2}
70	inpati	1.2×10^{-2}
71	includ	1.2×10^{-2}
72	person	1.2×10^{-2}
73	improv	1.1×10^{-2}
74	across	1.1×10^{-2}
75	respond	1.1×10^{-2}
76	studi	1.1×10^{-2}
77	adult	1.1×10^{-2}
78	conduct	1.1×10^{-2}
79	maker	1.1×10^{-2}
80	older	1.1×10^{-2}
81	semistructur	1.1×10^{-2}
82	impact	1.1×10^{-2}
83	status	1×10^{-2}
84	percent	1×10^{-2}
85	practition	1×10^{-2}
86	associ	1×10^{-2}
87	demograph	1×10^{-2}
88	find	1×10^{-2}
89	team	1×10^{-2}
90	outpati	9.9×10^{-3}
91	expenditur	9.9×10^{-3}
92	payment	9.9×10^{-3}
93	address	9.9×10^{-3}
94	live	9.8×10^{-3}
95	satisfact	9.8×10^{-3}
96	identifi	9.8×10^{-3}
97	report	9.8×10^{-3}
98	receiv	9.7×10^{-3}
99	energi	9.5×10^{-3}
100	plan	9.4×10^{-3}

Table D.107. The list of the top 100 words in the category Hematology with RIGs

No.	Word	RIG
1	patient	1×10^{-1}
2	cell	8.1×10^{-2}
3	leukemia	7.8×10^{-2}
4	hematopoiet	5.9×10^{-2}
5	blood	5.8×10^{-2}
6	transplant	5.2×10^{-2}
7	platelet	5.1×10^{-2}
8	acut	4.6×10^{-2}
9	stem	4.6×10^{-2}
10	myeloid	4.5×10^{-2}
11	marrow	4.3×10^{-2}
12	diseas	4.3×10^{-2}
13	surviv	4×10^{-2}
14	therapi	4×10^{-2}
15	allogen	3.6×10^{-2}
16	lymphoma	3.5×10^{-2}
17	transfus	3.4×10^{-2}
18	paper	3.3×10^{-2}
19	relaps	3.3×10^{-2}
20	clinic	3×10^{-2}
21	treatment	2.8×10^{-2}
22	median	2.6×10^{-2}
23	risk	2.5×10^{-2}
24	hematolog	2.5×10^{-2}
25	thrombosi	2.4×10^{-2}
26	bone	2.4×10^{-2}
27	bleed	2.3×10^{-2}
28	conclus	2.2×10^{-2}
29	associ	2.1×10^{-2}
30	anemia	2×10^{-2}
31	factor	2×10^{-2}
32	donor	2×10^{-2}
33	gvhd	2×10^{-2}
34	remiss	1.9×10^{-2}
35	express	1.9×10^{-2}
36	chemotherapi	1.9×10^{-2}
37	inhibitor	1.8×10^{-2}
38	regimen	1.8×10^{-2}
39	chronic	1.8×10^{-2}
40	coagul	1.7×10^{-2}
41	anticoagul	1.7×10^{-2}
42	thromboembol	1.7×10^{-2}
43	treat	1.7×10^{-2}
44	mutat	1.7×10^{-2}
45	dose	1.7×10^{-2}
46	progenitor	1.6×10^{-2}
47	lymphocyt	1.6×10^{-2}
48	malign	1.6×10^{-2}
49	therapeut	1.6×10^{-2}
50	mice	1.6×10^{-2}
51	prognost	1.5×10^{-2}
52	plasma	1.5×10^{-2}
53	venous	1.5×10^{-2}
54	propos	1.4×10^{-2}
55	outcom	1.4×10^{-2}
56	autolog	1.4×10^{-2}
57	engraft	1.4×10^{-2}
58	receiv	1.4×10^{-2}
59	year	1.3×10^{-2}
60	syndrom	1.3×10^{-2}
61	count	1.2×10^{-2}
62	vivo	1.2×10^{-2}
63	cd34	1.2×10^{-2}
64	gene	1.2×10^{-2}
65	endotheli	1.2×10^{-2}
66	diagnosi	1.2×10^{-2}
67	structur	1.2×10^{-2}
68	induc	1.2×10^{-2}
69	simul	1.2×10^{-2}
70	overal	1.2×10^{-2}
71	defici	1.1×10^{-2}
72	neutrophil	1.1×10^{-2}
73	mortal	1.1×10^{-2}
74	hemoglobin	1.1×10^{-2}
75	receptor	1.1×10^{-2}
76	refractori	1.1×10^{-2}
77	signific	1.1×10^{-2}
78	background	1.1×10^{-2}
79	pediatr	1.1×10^{-2}
80	temperatur	1.1×10^{-2}
81	month	1.1×10^{-2}
82	peripher	1.1×10^{-2}
83	progress	1.1×10^{-2}
84	antigen	1.1×10^{-2}
85	antibodi	1.1×10^{-2}
86	incid	1×10^{-2}
87	immun	1×10^{-2}
88	protein	1×10^{-2}
89	retrospect	1×10^{-2}
90	age	1×10^{-2}
91	recipi	9.9×10^{-3}
92	versus	9.9×10^{-3}
93	activ	9.7×10^{-3}
94	arteri	9.7×10^{-3}
95	vitro	9.7×10^{-3}
96	diagnos	9.5×10^{-3}
97	cohort	9.5×10^{-3}
98	inhibit	9.4×10^{-3}
99	energi	9.4×10^{-3}
100	mediat	9.3×10^{-3}

TABLE D.108. The list of the top 100 words in the category History with RIGs

No.	Word	RIG	No.	Word	RIG
1	articl	1.5×10^{-1}	51	archiv	1.6×10^{-2}
2	centuri	1.1×10^{-1}	52	militari	1.6×10^{-2}
3	polit	9.9×10^{-2}	53	author	1.6×10^{-2}
4	war	8.8×10^{-2}	54	german	1.6×10^{-2}
5	argu	6×10^{-2}	55	patient	1.5×10^{-2}
6	historian	5.7×10^{-2}	56	measur	1.5×10^{-2}
7	histori	5.4×10^{-2}	57	evalu	1.5×10^{-2}
8	result	4.8×10^{-2}	58	data	1.5×10^{-2}
9	histor	4.6×10^{-2}	59	earli	1.5×10^{-2}
10	british	4.5×10^{-2}	60	struggl	1.5×10^{-2}
11	nineteenth	4.4×10^{-2}	61	obtain	1.5×10^{-2}
12	twentieth	3.5×10^{-2}	62	govern	1.5×10^{-2}
13	nation	3.3×10^{-2}	63	studi	1.5×10^{-2}
14	method	3.1×10^{-2}	64	1930s	1.4×10^{-2}
15	essay	3.1×10^{-2}	65	model	1.4×10^{-2}
16	britain	3×10^{-2}	66	era	1.4×10^{-2}
17	imperi	2.9×10^{-2}	67	ottoman	1.4×10^{-2}
18	modern	2.7×10^{-2}	68	econom	1.4×10^{-2}
19	cultur	2.6×10^{-2}	69	europ	1.4×10^{-2}
20	scholar	2.6×10^{-2}	70	cathol	1.4×10^{-2}
21	social	2.5×10^{-2}	71	nationalist	1.4×10^{-2}
22	world	2.5×10^{-2}	72	communist	1.4×10^{-2}
23	use	2.5×10^{-2}	73	idea	1.4×10^{-2}
24	debat	2.4×10^{-2}	74	1970s	1.4×10^{-2}
25	religi	2.4×10^{-2}	75	elit	1.4×10^{-2}
26	narrat	2.3×10^{-2}	76	civil	1.4×10^{-2}
27	historiographi	2.3×10^{-2}	77	question	1.3×10^{-2}
28	offici	2.2×10^{-2}	78	claim	1.3×10^{-2}
29	postwar	2.2×10^{-2}	79	simul	1.3×10^{-2}
30	soviet	2.2×10^{-2}	80	1960s	1.3×10^{-2}
31	societi	2.2×10^{-2}	81	way	1.3×10^{-2}
32	effect	2.1×10^{-2}	82	context	1.3×10^{-2}
33	ideolog	2×10^{-2}	83	stori	1.3×10^{-2}
34	public	1.9×10^{-2}	84	compar	1.3×10^{-2}
35	coloni	1.8×10^{-2}	85	european	1.3×10^{-2}
36	christian	1.8×10^{-2}	86	union	1.3×10^{-2}
37	contemporari	1.8×10^{-2}	87	book	1.3×10^{-2}
38	discours	1.8×10^{-2}	88	seventeenth	1.3×10^{-2}
39	perform	1.8×10^{-2}	89	reform	1.3×10^{-2}
40	high	1.7×10^{-2}	90	religion	1.2×10^{-2}
41	eighteenth	1.7×10^{-2}	91	test	1.2×10^{-2}
42	american	1.7×10^{-2}	92	movement	1.2×10^{-2}
43	becam	1.7×10^{-2}	93	protest	1.2×10^{-2}
44	late	1.7×10^{-2}	94	improv	1.2×10^{-2}
45	french	1.7×10^{-2}	95	low	1.2×10^{-2}
46	write	1.7×10^{-2}	96	labour	1.2×10^{-2}
47	church	1.7×10^{-2}	97	came	1.2×10^{-2}
48	jewish	1.6×10^{-2}	98	reduc	1.2×10^{-2}
49	cell	1.6×10^{-2}	99	ident	1.2×10^{-2}
50	revolut	1.6×10^{-2}	100	decreas	1.2×10^{-2}

TABLE D.109. The list of the top 100 words in the category History and Philosophy Of Science with RIGs

No.	Word	RIG
1	scienc	7.9×10^{-2}
2	argu	7.5×10^{-2}
3	scientif	6.2×10^{-2}
4	centuri	4.3×10^{-2}
5	philosoph	3.6×10^{-2}
6	ethic	3.1×10^{-2}
7	result	3×10^{-2}
8	argument	2.9×10^{-2}
9	scientist	2.8×10^{-2}
10	articl	2.5×10^{-2}
11	claim	2.3×10^{-2}
12	philosophi	2.3×10^{-2}
13	epistem	2.2×10^{-2}
14	view	2×10^{-2}
15	histori	2×10^{-2}
16	theori	1.9×10^{-2}
17	essay	1.9×10^{-2}
18	nineteenth	1.8×10^{-2}
19	debat	1.7×10^{-2}
20	histor	1.7×10^{-2}
21	method	1.6×10^{-2}
22	way	1.6×10^{-2}
23	perform	1.6×10^{-2}
24	concept	1.6×10^{-2}
25	twentieth	1.6×10^{-2}
26	effect	1.6×10^{-2}
27	epistemolog	1.5×10^{-2}
28	world	1.5×10^{-2}
29	war	1.5×10^{-2}
30	social	1.5×10^{-2}
31	idea	1.5×10^{-2}
32	high	1.4×10^{-2}
33	public	1.4×10^{-2}
34	question	1.4×10^{-2}
35	notion	1.4×10^{-2}
36	historian	1.3×10^{-2}
37	use	1.2×10^{-2}
38	practic	1.2×10^{-2}
39	defend	1.2×10^{-2}
40	rate	1.2×10^{-2}
41	concern	1.1×10^{-2}
42	compar	1.1×10^{-2}
43	explan	1.1×10^{-2}
44	modern	1.1×10^{-2}
45	natur	1.1×10^{-2}
46	knowledg	1.1×10^{-2}
47	studi	1.1×10^{-2}
48	low	1.1×10^{-2}
49	moral	1.1×10^{-2}
50	attempt	1.1×10^{-2}

No.	Word	RIG
51	think	1.1×10^{-2}
52	cell	1.1×10^{-2}
53	belief	1×10^{-2}
54	societi	1×10^{-2}
55	polit	1×10^{-2}
56	context	9.6×10^{-3}
57	discuss	9.6×10^{-3}
58	decreas	9.6×10^{-3}
59	british	9.6×10^{-3}
60	higher	9.4×10^{-3}
61	draw	9.4×10^{-3}
62	obtain	9.3×10^{-3}
63	measur	9.2×10^{-3}
64	contemporari	9.1×10^{-3}
65	temperatur	9.1×10^{-3}
66	increas	9×10^{-3}
67	research	9×10^{-3}
68	control	9×10^{-3}
69	effici	8.9×10^{-3}
70	becam	8.9×10^{-3}
71	sampl	8.9×10^{-3}
72	lower	8.9×10^{-3}
73	scholar	8.9×10^{-3}
74	conceptu	8.8×10^{-3}
75	univers	8.7×10^{-3}
76	simul	8.6×10^{-3}
77	book	8.6×10^{-3}
78	john	8.4×10^{-3}
79	data	8.3×10^{-3}
80	induc	8×10^{-3}
81	paramet	8×10^{-3}
82	author	8×10^{-3}
83	understand	7.9×10^{-3}
84	discours	7.7×10^{-3}
85	reduc	7.7×10^{-3}
86	eighteenth	7.6×10^{-3}
87	1950s	7.5×10^{-3}
88	improv	7.5×10^{-3}
89	evalu	7.4×10^{-3}
90	account	7.4×10^{-3}
91	mathemat	7.3×10^{-3}
92	stori	7.3×10^{-3}
93	issu	7.3×10^{-3}
94	surfac	7.2×10^{-3}
95	investig	7.2×10^{-3}
96	particular	7.2×10^{-3}
97	concentr	7.2×10^{-3}
98	write	7.2×10^{-3}
99	focus	7.1×10^{-3}
100	test	7.1×10^{-3}

Table D.110. The list of the top 100 words in the category History Of Social Sciences with RIGs

No.	Word	RIG	No.	Word	RIG
1	articl	9.6×10^{-2}	51	firm	1.3×10^{-2}
2	centuri	9.3×10^{-2}	52	german	1.2×10^{-2}
3	histor	5.4×10^{-2}	53	govern	1.2×10^{-2}
4	nineteenth	5×10^{-2}	54	1950s	1.2×10^{-2}
5	twentieth	4.9×10^{-2}	55	scholarship	1.2×10^{-2}
6	argu	4.2×10^{-2}	56	evalu	1.2×10^{-2}
7	polit	4×10^{-2}	57	draw	1.2×10^{-2}
8	histori	4×10^{-2}	58	discours	1.1×10^{-2}
9	war	3.5×10^{-2}	59	cultur	1.1×10^{-2}
10	historian	3.1×10^{-2}	60	way	1.1×10^{-2}
11	result	3×10^{-2}	61	patient	1.1×10^{-2}
12	econom	2.8×10^{-2}	62	simul	1.1×10^{-2}
13	social	2.6×10^{-2}	63	book	1.1×10^{-2}
14	becam	2.2×10^{-2}	64	obtain	1.1×10^{-2}
15	educ	2.1×10^{-2}	65	historiographi	1.1×10^{-2}
16	scholar	2.1×10^{-2}	66	effect	1.1×10^{-2}
17	british	2.1×10^{-2}	67	focus	1.1×10^{-2}
18	institut	2×10^{-2}	68	context	1.1×10^{-2}
19	britain	1.9×10^{-2}	69	write	1.1×10^{-2}
20	market	1.9×10^{-2}	70	detect	1×10^{-2}
21	1970s	1.9×10^{-2}	71	temperatur	1×10^{-2}
22	1930s	1.8×10^{-2}	72	paramet	1×10^{-2}
23	earli	1.8×10^{-2}	73	high	9.9×10^{-3}
24	method	1.7×10^{-2}	74	intern	9.8×10^{-3}
25	debat	1.7×10^{-2}	75	intellectu	9.7×10^{-3}
26	nation	1.7×10^{-2}	76	eighteenth	9.6×10^{-3}
27	busi	1.7×10^{-2}	77	trade	9.6×10^{-3}
28	economi	1.7×10^{-2}	78	1960s	9.5×10^{-3}
29	american	1.6×10^{-2}	79	surfac	9.4×10^{-3}
30	modern	1.6×10^{-2}	80	urban	9.3×10^{-3}
31	world	1.5×10^{-2}	81	offici	9.3×10^{-3}
32	1920s	1.5×10^{-2}	82	question	9.3×10^{-3}
33	idea	1.5×10^{-2}	83	period	9.1×10^{-3}
34	capit	1.5×10^{-2}	84	explor	9×10^{-3}
35	und	1.5×10^{-2}	85	geographi	9×10^{-3}
36	perform	1.4×10^{-2}	86	view	8.9×10^{-3}
37	citi	1.4×10^{-2}	87	industri	8.9×10^{-3}
38	cell	1.4×10^{-2}	88	narrat	8.9×10^{-3}
39	archiv	1.4×10^{-2}	89	measur	8.8×10^{-3}
40	reform	1.4×10^{-2}	90	transnat	8.8×10^{-3}
41	revolut	1.4×10^{-2}	91	polici	8.7×10^{-3}
42	school	1.4×10^{-2}	92	census	8.4×10^{-3}
43	societi	1.3×10^{-2}	93	labor	8.3×10^{-3}
44	public	1.3×10^{-2}	94	compani	8.3×10^{-3}
45	use	1.3×10^{-2}	95	essay	8.3×10^{-3}
46	contemporari	1.3×10^{-2}	96	perspect	8.2×10^{-3}
47	late	1.3×10^{-2}	97	wage	8.1×10^{-3}
48	postwar	1.3×10^{-2}	98	studi	8.1×10^{-3}
49	countri	1.3×10^{-2}	99	organis	8×10^{-3}
50	labour	1.3×10^{-2}	100	coloni	8×10^{-3}

TABLE D.111. The list of the top 100 words in the category Horticulture with RIGs

No.	Word	RIG
1	cultivar	1.7×10^{-1}
2	fruit	1.5×10^{-1}
3	plant	1.3×10^{-1}
4	crop	7.7×10^{-2}
5	breed	5.8×10^{-2}
6	flower	5.1×10^{-2}
7	leaf	5×10^{-2}
8	orchard	4.3×10^{-2}
9	harvest	4.2×10^{-2}
10	trait	3.8×10^{-2}
11	cultiv	3.6×10^{-2}
12	shoot	3.5×10^{-2}
13	grown	3.5×10^{-2}
14	tree	3.4×10^{-2}
15	greenhous	3.1×10^{-2}
16	product	3.1×10^{-2}
17	grape	3×10^{-2}
18	seedl	3×10^{-2}
19	genotyp	3×10^{-2}
20	genet	3×10^{-2}
21	yield	2.9×10^{-2}
22	irrig	2.9×10^{-2}
23	germplasm	2.9×10^{-2}
24	postharvest	2.9×10^{-2}
25	grower	2.8×10^{-2}
26	seed	2.7×10^{-2}
27	leav	2.7×10^{-2}
28	content	2.6×10^{-2}
29	qualiti	2.6×10^{-2}
30	root	2.5×10^{-2}
31	soil	2.5×10^{-2}
32	fresh	2.4×10^{-2}
33	qtl	2.4×10^{-2}
34	season	2.3×10^{-2}
35	appl	2.3×10^{-2}
36	tomato	2.3×10^{-2}
37	marker	2.3×10^{-2}
38	veget	2.3×10^{-2}
39	growth	2.3×10^{-2}
40	commerci	2.2×10^{-2}
41	ssr	2.1×10^{-2}
42	grow	2.1×10^{-2}
43	bud	2.1×10^{-2}
44	ripen	2×10^{-2}
45	sweet	1.9×10^{-2}
46	patient	1.8×10^{-2}
47	loci	1.8×10^{-2}
48	dri	1.8×10^{-2}
49	anthocyanin	1.7×10^{-2}
50	fertil	1.6×10^{-2}

No.	Word	RIG
51	speci	1.6×10^{-2}
52	matur	1.5×10^{-2}
53	inocul	1.5×10^{-2}
54	gene	1.5×10^{-2}
55	paper	1.5×10^{-2}
56	chromosom	1.4×10^{-2}
57	progeni	1.4×10^{-2}
58	highest	1.4×10^{-2}
59	wine	1.4×10^{-2}
60	day	1.4×10^{-2}
61	total	1.4×10^{-2}
62	agronom	1.3×10^{-2}
63	ornament	1.3×10^{-2}
64	germin	1.3×10^{-2}
65	nutrient	1.3×10^{-2}
66	sugar	1.3×10^{-2}
67	acid	1.2×10^{-2}
68	propos	1.2×10^{-2}
69	weight	1.2×10^{-2}
70	pathogen	1.2×10^{-2}
71	phenol	1.2×10^{-2}
72	solubl	1.1×10^{-2}
73	per	1.1×10^{-2}
74	model	1.1×10^{-2}
75	polymorph	1.1×10^{-2}
76	locus	1.1×10^{-2}
77	allel	1.1×10^{-2}
78	produc	1.1×10^{-2}
79	treatment	1.1×10^{-2}
80	genom	1×10^{-2}
81	canopi	1×10^{-2}
82	three	1×10^{-2}
83	compost	1×10^{-2}
84	pest	1×10^{-2}
85	chlorophyl	9.9×10^{-3}
86	storag	9.9×10^{-3}
87	pollin	9.8×10^{-3}
88	clinic	9.8×10^{-3}
89	ascorb	9.8×10^{-3}
90	control	9.6×10^{-3}
91	agricultur	9.6×10^{-3}
92	resist	9.6×10^{-3}
93	clone	9.4×10^{-3}
94	linkag	9.4×10^{-3}
95	four	9.3×10^{-3}
96	water	9.2×10^{-3}
97	market	9.1×10^{-3}
98	select	9×10^{-3}
99	winter	9×10^{-3}
100	bloom	9×10^{-3}

Table D.112. The list of the top 100 words in the category Hospitality, Leisure, Sport and Tourism with RIGs

No.	Word	RIG	No.	Word	RIG
1	tourism	1.8×10^{-1}	51	economi	1×10^{-2}
2	sport	1×10^{-1}	52	nation	1×10^{-2}
3	tourist	9.8×10^{-2}	53	motiv	1×10^{-2}
4	destin	5.1×10^{-2}	54	author	1×10^{-2}
5	hotel	4.7×10^{-2}	55	tour	1×10^{-2}
6	research	3.9×10^{-2}	56	explor	1×10^{-2}
7	athlet	3.5×10^{-2}	57	simul	9.9×10^{-3}
8	market	3.2×10^{-2}	58	survey	9.8×10^{-3}
9	social	3.2×10^{-2}	59	servic	9.8×10^{-3}
10	visitor	2.7×10^{-2}	60	empir	9.7×10^{-3}
11	footbal	2.2×10^{-2}	61	influenc	9.5×10^{-3}
12	travel	2.2×10^{-2}	62	focus	9.5×10^{-3}
13	perceiv	2.1×10^{-2}	63	engag	9×10^{-3}
14	implic	1.9×10^{-2}	64	peopl	9×10^{-3}
15	leisur	1.9×10^{-2}	65	temperatur	9×10^{-3}
16	percept	1.8×10^{-2}	66	attract	9×10^{-3}
17	satisfact	1.8×10^{-2}	67	surfac	8.9×10^{-3}
18	interview	1.8×10^{-2}	68	detect	8.9×10^{-3}
19	econom	1.8×10^{-2}	69	protein	8.9×10^{-3}
20	coach	1.8×10^{-2}	70	clinic	8.7×10^{-3}
21	industri	1.8×10^{-2}	71	sector	8.6×10^{-3}
22	examin	1.7×10^{-2}	72	loyalti	8.6×10^{-3}
23	brand	1.6×10^{-2}	73	treatment	8.5×10^{-3}
24	patient	1.6×10^{-2}	74	develop	8.4×10^{-3}
25	manag	1.5×10^{-2}	75	paramet	8.3×10^{-3}
26	cell	1.5×10^{-2}	76	properti	8.1×10^{-3}
27	find	1.4×10^{-2}	77	literatur	8×10^{-3}
28	practic	1.4×10^{-2}	78	intern	7.9×10^{-3}
29	competit	1.4×10^{-2}	79	profession	7.7×10^{-3}
30	intent	1.4×10^{-2}	80	psycholog	7.5×10^{-3}
31	cultur	1.3×10^{-2}	81	draw	7.3×10^{-3}
32	particip	1.3×10^{-2}	82	elit	7.3×10^{-3}
33	custom	1.3×10^{-2}	83	manageri	7.3×10^{-3}
34	sustain	1.2×10^{-2}	84	gene	7.3×10^{-3}
35	countri	1.2×10^{-2}	85	govern	7.2×10^{-3}
36	relationship	1.2×10^{-2}	86	discuss	7.1×10^{-3}
37	busi	1.2×10^{-2}	87	acid	7.1×10^{-3}
38	malaysia	1.2×10^{-2}	88	algorithm	7.1×10^{-3}
39	educ	1.2×10^{-2}	89	world	7.1×10^{-3}
40	articl	1.2×10^{-2}	90	system	7.1×10^{-3}
41	purpos	1.2×10^{-2}	91	understand	7.1×10^{-3}
42	player	1.2×10^{-2}	92	visit	7×10^{-3}
43	recreat	1.2×10^{-2}	93	diseas	6.9×10^{-3}
44	attitud	1.2×10^{-2}	94	high	6.8×10^{-3}
45	team	1.1×10^{-2}	95	energi	6.8×10^{-3}
46	heritag	1.1×10^{-2}	96	impact	6.6×10^{-3}
47	questionnair	1.1×10^{-2}	97	electron	6.6×10^{-3}
48	employe	1.1×10^{-2}	98	induc	6.6×10^{-3}
49	perspect	1.1×10^{-2}	99	signal	6.5×10^{-3}
50	game	1×10^{-2}	100	context	6.5×10^{-3}

Table D.113. The list of the top 100 words in the category Humanities, Multidisciplinary with RIGs

No.	Word	RIG
1	centuri	4.8×10^{-2}
2	articl	4.8×10^{-2}
3	cultur	3.6×10^{-2}
4	essay	3.4×10^{-2}
5	result	3.3×10^{-2}
6	polit	3.1×10^{-2}
7	artist	2.8×10^{-2}
8	argu	2.3×10^{-2}
9	social	2.3×10^{-2}
10	contemporari	2.1×10^{-2}
11	histor	2.1×10^{-2}
12	modern	2×10^{-2}
13	art	1.9×10^{-2}
14	text	1.9×10^{-2}
15	semiot	1.9×10^{-2}
16	aesthet	1.8×10^{-2}
17	teach	1.8×10^{-2}
18	heritag	1.7×10^{-2}
19	societi	1.7×10^{-2}
20	way	1.6×10^{-2}
21	nineteenth	1.5×10^{-2}
22	author	1.5×10^{-2}
23	creativ	1.5×10^{-2}
24	literari	1.5×10^{-2}
25	narrat	1.5×10^{-2}
26	colleg	1.5×10^{-2}
27	museum	1.5×10^{-2}
28	use	1.5×10^{-2}
29	cell	1.4×10^{-2}
30	reform	1.4×10^{-2}
31	world	1.4×10^{-2}
32	peopl	1.4×10^{-2}
33	ideolog	1.3×10^{-2}
34	patient	1.3×10^{-2}
35	educ	1.3×10^{-2}
36	write	1.3×10^{-2}
37	music	1.3×10^{-2}
38	explor	1.2×10^{-2}
39	histori	1.2×10^{-2}
40	high	1.2×10^{-2}
41	languag	1.2×10^{-2}
42	method	1.2×10^{-2}
43	talent	1.2×10^{-2}
44	war	1.2×10^{-2}
45	discours	1.1×10^{-2}
46	idea	1.1×10^{-2}
47	eighteenth	1.1×10^{-2}
48	context	1.1×10^{-2}
49	tradit	1.1×10^{-2}
50	stori	1.1×10^{-2}
51	english	1.1×10^{-2}
52	show	1.1×10^{-2}
53	student	1.1×10^{-2}
54	compar	1.1×10^{-2}
55	measur	1×10^{-2}
56	philosoph	1×10^{-2}
57	decreas	1×10^{-2}
58	scholar	1×10^{-2}
59	obtain	9.9×10^{-3}
60	nation	9.8×10^{-3}
61	conclus	9.8×10^{-3}
62	data	9.7×10^{-3}
63	simul	9.6×10^{-3}
64	temperatur	9.4×10^{-3}
65	think	9.4×10^{-3}
66	public	9.4×10^{-3}
67	draw	9.2×10^{-3}
68	test	9.2×10^{-3}
69	twentieth	9.2×10^{-3}
70	put	9.1×10^{-3}
71	writer	9.1×10^{-3}
72	view	9×10^{-3}
73	situat	8.9×10^{-3}
74	rate	8.8×10^{-3}
75	induc	8.6×10^{-3}
76	increas	8.6×10^{-3}
77	poem	8.6×10^{-3}
78	effect	8.6×10^{-3}
79	detect	8.5×10^{-3}
80	philosophi	8.4×10^{-3}
81	religi	8.4×10^{-3}
82	paramet	8.4×10^{-3}
83	low	8.3×10^{-3}
84	fiction	8.2×10^{-3}
85	practic	8.2×10^{-3}
86	concept	8.2×10^{-3}
87	audienc	8.2×10^{-3}
88	observ	8.1×10^{-3}
89	perform	8×10^{-3}
90	paint	8×10^{-3}
91	clinic	7.9×10^{-3}
92	read	7.9×10^{-3}
93	charact	7.8×10^{-3}
94	protein	7.8×10^{-3}
95	control	7.8×10^{-3}
96	investig	7.7×10^{-3}
97	reduc	7.7×10^{-3}
98	perspect	7.7×10^{-3}
99	univers	7.6×10^{-3}
100	spiritu	7.6×10^{-3}

Table D.114. The list of the top 100 words in the category Imaging Science and Photographic Technology with RIGs

No.	Word	RIG	No.	Word	RIG
1	imag	1.6×10^{-1}	51	extract	1.2×10^{-2}
2	propos	5.6×10^{-2}	52	studi	1.2×10^{-2}
3	algorithm	5.3×10^{-2}	53	textur	1.1×10^{-2}
4	pixel	4.6×10^{-2}	54	spars	1.1×10^{-2}
5	radar	3.6×10^{-2}	55	tempor	1.1×10^{-2}
6	resolut	3.6×10^{-2}	56	robust	1.1×10^{-2}
7	sar	3.6×10^{-2}	57	filter	1.1×10^{-2}
8	satellit	3.6×10^{-2}	58	classifi	1.1×10^{-2}
9	accuraci	3.3×10^{-2}	59	treatment	1.1×10^{-2}
10	remot	3.2×10^{-2}	60	use	1×10^{-2}
11	spatial	3.2×10^{-2}	61	protein	1×10^{-2}
12	scene	3×10^{-2}	62	techniqu	1×10^{-2}
13	map	2.8×10^{-2}	63	forest	1×10^{-2}
14	imageri	2.8×10^{-2}	64	letter	9.8×10^{-3}
15	apertur	2.8×10^{-2}	65	cell	9.8×10^{-3}
16	paper	2.8×10^{-2}	66	registr	9.8×10^{-3}
17	classif	2.5×10^{-2}	67	detect	9.8×10^{-3}
18	video	2.5×10^{-2}	68	acquir	9.7×10^{-3}
19	hyperspectr	2.5×10^{-2}	69	sensor	9.2×10^{-3}
20	dataset	2.4×10^{-2}	70	lidar	9.2×10^{-3}
21	base	2.4×10^{-2}	71	cover	9.1×10^{-3}
22	art	2.3×10^{-2}	72	band	8.8×10^{-3}
23	sens	2.1×10^{-2}	73	acid	8.5×10^{-3}
24	segment	2.1×10^{-2}	74	vector	8.5×10^{-3}
25	camera	2.1×10^{-2}	75	color	8.3×10^{-3}
26	estim	2.1×10^{-2}	76	digit	8.1×10^{-3}
27	synthet	2.1×10^{-2}	77	area	8.1×10^{-3}
28	data	2×10^{-2}	78	group	8×10^{-3}
29	modi	2×10^{-2}	79	represent	7.9×10^{-3}
30	conclus	2×10^{-2}	80	age	7.8×10^{-3}
31	approach	1.9×10^{-2}	81	perform	7.8×10^{-3}
32	featur	1.9×10^{-2}	82	gene	7.7×10^{-3}
33	land	1.6×10^{-2}	83	patient	7.6×10^{-3}
34	automat	1.6×10^{-2}	84	exploit	7.5×10^{-3}
35	inform	1.5×10^{-2}	85	reaction	7.4×10^{-3}
36	error	1.5×10^{-2}	86	problem	7.4×10^{-3}
37	reconstruct	1.5×10^{-2}	87	can	7.2×10^{-3}
38	airborn	1.4×10^{-2}	88	motion	7.1×10^{-3}
39	retriev	1.4×10^{-2}	89	control	7×10^{-3}
40	landsat	1.4×10^{-2}	90	region	7×10^{-3}
41	nois	1.4×10^{-2}	91	track	7×10^{-3}
42	comput	1.4×10^{-2}	92	mechan	6.9×10^{-3}
43	spectral	1.4×10^{-2}	93	geometr	6.9×10^{-3}
44	method	1.4×10^{-2}	94	framework	6.8×10^{-3}
45	veget	1.3×10^{-2}	95	model	6.7×10^{-3}
46	outperform	1.3×10^{-2}	96	set	6.7×10^{-3}
47	visual	1.2×10^{-2}	97	vision	6.7×10^{-3}
48	ground	1.2×10^{-2}	98	captur	6.6×10^{-3}
49	accur	1.2×10^{-2}	99	frame	6.6×10^{-3}
50	real	1.2×10^{-2}	100	associ	6.6×10^{-3}

TABLE D.115. The list of the top 100 words in the category Immunology with RIGs

No.	Word	RIG	No.	Word	RIG
1	immun	1.4×10^{-1}	51	igg	1.7×10^{-2}
2	infect	9.3×10^{-2}	52	interferon	1.7×10^{-2}
3	cell	8.2×10^{-2}	53	immunogen	1.7×10^{-2}
4	vaccin	6.7×10^{-2}	54	dendrit	1.7×10^{-2}
5	cytokin	6.4×10^{-2}	55	asthma	1.7×10^{-2}
6	cd4	5.2×10^{-2}	56	monocyt	1.6×10^{-2}
7	diseas	4.6×10^{-2}	57	gene	1.6×10^{-2}
8	antibodi	4.5×10^{-2}	58	protect	1.6×10^{-2}
9	express	4.4×10^{-2}	59	lps	1.6×10^{-2}
10	mice	4.4×10^{-2}	60	propos	1.6×10^{-2}
11	antigen	4.4×10^{-2}	61	inhibit	1.5×10^{-2}
12	ifn	4.1×10^{-2}	62	host	1.5×10^{-2}
13	inflammatori	4.1×10^{-2}	63	secret	1.5×10^{-2}
14	respons	3.9×10^{-2}	64	peripher	1.5×10^{-2}
15	transplant	3.8×10^{-2}	65	effector	1.5×10^{-2}
16	paper	3.5×10^{-2}	66	simul	1.5×10^{-2}
17	virus	3.4×10^{-2}	67	alpha	1.5×10^{-2}
18	hiv	3.3×10^{-2}	68	regul	1.5×10^{-2}
19	induc	3.2×10^{-2}	69	ige	1.4×10^{-2}
20	inflamm	3.2×10^{-2}	70	immunosuppress	1.4×10^{-2}
21	patient	3×10^{-2}	71	assay	1.4×10^{-2}
22	cd8	2.9×10^{-2}	72	therapi	1.4×10^{-2}
23	receptor	2.8×10^{-2}	73	chronic	1.4×10^{-2}
24	innat	2.8×10^{-2}	74	neutrophil	1.3×10^{-2}
25	interleukin	2.7×10^{-2}	75	allergi	1.3×10^{-2}
26	tnf	2.7×10^{-2}	76	acut	1.3×10^{-2}
27	mediat	2.6×10^{-2}	77	chemokin	1.3×10^{-2}
28	macrophag	2.5×10^{-2}	78	elisa	1.3×10^{-2}
29	lymphocyt	2.4×10^{-2}	79	immunodefici	1.3×10^{-2}
30	autoimmun	2.4×10^{-2}	80	donor	1.3×10^{-2}
31	pathogen	2.3×10^{-2}	81	influenza	1.3×10^{-2}
32	associ	2.3×10^{-2}	82	mous	1.3×10^{-2}
33	recipi	2.3×10^{-2}	83	bacteri	1.2×10^{-2}
34	protein	2.3×10^{-2}	84	vivo	1.2×10^{-2}
35	proinflammatori	2.3×10^{-2}	85	vitro	1.2×10^{-2}
36	allerg	2.1×10^{-2}	86	regulatori	1.2×10^{-2}
37	human	2.1×10^{-2}	87	immunoglobulin	1.2×10^{-2}
38	blood	2.1×10^{-2}	88	pathogenesi	1.2×10^{-2}
39	viral	2×10^{-2}	89	murin	1.2×10^{-2}
40	activ	1.9×10^{-2}	90	allergen	1.2×10^{-2}
41	conclus	1.9×10^{-2}	91	specif	1.2×10^{-2}
42	background	1.9×10^{-2}	92	level	1.2×10^{-2}
43	serum	1.9×10^{-2}	93	defici	1.2×10^{-2}
44	stimul	1.8×10^{-2}	94	hematopoiet	1.2×10^{-2}
45	antiretrovir	1.8×10^{-2}	95	titer	1.1×10^{-2}
46	immunolog	1.8×10^{-2}	96	immunotherapi	1.1×10^{-2}
47	gamma	1.8×10^{-2}	97	signific	1.1×10^{-2}
48	role	1.8×10^{-2}	98	dose	1.1×10^{-2}
49	clinic	1.8×10^{-2}	99	energi	1.1×10^{-2}
50	hla	1.7×10^{-2}	100	beta	1.1×10^{-2}

TABLE D.116. The list of the top 100 words in the category Industrial Relations and Labor with RIGs

No.	Word	RIG	No.	Word	RIG
1	worker	1×10^{-1}	51	economi	1.3×10^{-2}
2	employe	8.9×10^{-2}	52	crisi	1.2×10^{-2}
3	labour	7.6×10^{-2}	53	resourc	1.2×10^{-2}
4	job	7.4×10^{-2}	54	context	1.2×10^{-2}
5	union	5.9×10^{-2}	55	profession	1.2×10^{-2}
6	wage	5.7×10^{-2}	56	satisfact	1.2×10^{-2}
7	articl	5.3×10^{-2}	57	leadership	1.1×10^{-2}
8	employ	4.7×10^{-2}	58	earn	1.1×10^{-2}
9	workplac	4×10^{-2}	59	gender	1.1×10^{-2}
10	labor	4×10^{-2}	60	career	1.1×10^{-2}
11	find	3.7×10^{-2}	61	simul	1.1×10^{-2}
12	workforc	3.7×10^{-2}	62	manageri	1×10^{-2}
13	organiz	3.1×10^{-2}	63	australia	1×10^{-2}
14	bargain	2.9×10^{-2}	64	educ	1×10^{-2}
15	market	2.8×10^{-2}	65	privat	1×10^{-2}
16	survey	2.7×10^{-2}	66	focus	9.9×10^{-3}
17	work	2.7×10^{-2}	67	institut	9.8×10^{-3}
18	firm	2.6×10^{-2}	68	temperatur	9.7×10^{-3}
19	methodolog	2.6×10^{-2}	69	british	9.6×10^{-3}
20	implic	2.5×10^{-2}	70	surfac	9.4×10^{-3}
21	sector	2.4×10^{-2}	71	servic	9.3×10^{-3}
22	polici	2.4×10^{-2}	72	trade	9.3×10^{-3}
23	practic	2.4×10^{-2}	73	commit	9.2×10^{-3}
24	organis	2.3×10^{-2}	74	draw	9.2×10^{-3}
25	author	2.2×10^{-2}	75	collect	8.9×10^{-3}
26	social	2.2×10^{-2}	76	health	8.9×10^{-3}
27	skill	2×10^{-2}	77	approach	8.9×10^{-3}
28	research	2×10^{-2}	78	paramet	8.8×10^{-3}
29	countri	2×10^{-2}	79	reform	8.8×10^{-3}
30	examin	1.8×10^{-2}	80	incom	8.7×10^{-3}
31	origin	1.8×10^{-2}	81	detect	8.7×10^{-3}
32	polit	1.8×10^{-2}	82	explor	8.7×10^{-3}
33	empir	1.7×10^{-2}	83	supervisor	8.5×10^{-3}
34	occup	1.6×10^{-2}	84	properti	8.5×10^{-3}
35	industri	1.6×10^{-2}	85	experiment	8.5×10^{-3}
36	econom	1.6×10^{-2}	86	migrant	8.4×10^{-3}
37	manag	1.6×10^{-2}	87	relat	8.4×10^{-3}
38	pay	1.5×10^{-2}	88	data	8.3×10^{-3}
39	purpos	1.4×10^{-2}	89	method	8.3×10^{-3}
40	capit	1.4×10^{-2}	90	compani	8.3×10^{-3}
41	relationship	1.4×10^{-2}	91	patient	8.1×10^{-3}
42	nation	1.4×10^{-2}	92	perceiv	7.9×10^{-3}
43	interview	1.4×10^{-2}	93	protein	7.9×10^{-3}
44	unemploy	1.4×10^{-2}	94	agenc	7.9×10^{-3}
45	cell	1.3×10^{-2}	95	immigr	7.8×10^{-3}
46	govern	1.3×10^{-2}	96	literatur	7.8×10^{-3}
47	und	1.3×10^{-2}	97	impact	7.6×10^{-3}
48	australian	1.3×10^{-2}	98	obtain	7.6×10^{-3}
49	argu	1.3×10^{-2}	99	paid	7.4×10^{-3}
50	public	1.3×10^{-2}	100	women	7.3×10^{-3}

TABLE D.117. The list of the top 100 words in the category Infectious Diseases with RIGs

No.	Word	RIG
1	infect	2.3×10^{-1}
2	hiv	8.7×10^{-2}
3	virus	5.6×10^{-2}
4	conclus	5×10^{-2}
5	antiretrovir	4.9×10^{-2}
6	patient	4.6×10^{-2}
7	background	4.6×10^{-2}
8	vaccin	3.9×10^{-2}
9	hospit	3.6×10^{-2}
10	clinic	3.5×10^{-2}
11	diseas	3.4×10^{-2}
12	pathogen	3.4×10^{-2}
13	isol	3.2×10^{-2}
14	preval	3.1×10^{-2}
15	paper	3×10^{-2}
16	antibiot	2.9×10^{-2}
17	associ	2.9×10^{-2}
18	epidemiolog	2.9×10^{-2}
19	viral	2.8×10^{-2}
20	malaria	2.7×10^{-2}
21	surveil	2.7×10^{-2}
22	pneumonia	2.6×10^{-2}
23	resist	2.6×10^{-2}
24	among	2.6×10^{-2}
25	risk	2.6×10^{-2}
26	immun	2.5×10^{-2}
27	tuberculosi	2.4×10^{-2}
28	fever	2.4×10^{-2}
29	therapi	2.4×10^{-2}
30	strain	2.4×10^{-2}
31	cd4	2.4×10^{-2}
32	outbreak	2.3×10^{-2}
33	suscept	2.3×10^{-2}
34	pcr	2.1×10^{-2}
35	health	2.1×10^{-2}
36	year	2×10^{-2}
37	influenza	1.9×10^{-2}
38	antimicrobi	1.9×10^{-2}
39	infecti	1.9×10^{-2}
40	endem	1.7×10^{-2}
41	staphylococcus	1.7×10^{-2}
42	plasmodium	1.7×10^{-2}
43	antibodi	1.7×10^{-2}
44	prevent	1.7×10^{-2}
45	children	1.6×10^{-2}
46	aureus	1.6×10^{-2}
47	coinfect	1.6×10^{-2}
48	virolog	1.6×10^{-2}
49	immunodefici	1.6×10^{-2}
50	parasit	1.6×10^{-2}
51	propos	1.6×10^{-2}
52	bacteri	1.6×10^{-2}
53	popul	1.5×10^{-2}
54	genotyp	1.5×10^{-2}
55	mortal	1.5×10^{-2}
56	age	1.5×10^{-2}
57	care	1.5×10^{-2}
58	treatment	1.5×10^{-2}
59	blood	1.5×10^{-2}
60	human	1.5×10^{-2}
61	incid	1.4×10^{-2}
62	cohort	1.4×10^{-2}
63	drug	1.4×10^{-2}
64	antigen	1.4×10^{-2}
65	falciparum	1.4×10^{-2}
66	caus	1.4×10^{-2}
67	epidem	1.3×10^{-2}
68	case	1.3×10^{-2}
69	transmiss	1.3×10^{-2}
70	posit	1.3×10^{-2}
71	regimen	1.3×10^{-2}
72	identifi	1.3×10^{-2}
73	virul	1.3×10^{-2}
74	confid	1.3×10^{-2}
75	host	1.2×10^{-2}
76	countri	1.2×10^{-2}
77	multidrug	1.2×10^{-2}
78	collect	1.2×10^{-2}
79	process	1.2×10^{-2}
80	detect	1.2×10^{-2}
81	assay	1.2×10^{-2}
82	energi	1.2×10^{-2}
83	serolog	1.2×10^{-2}
84	africa	1.2×10^{-2}
85	respiratori	1.2×10^{-2}
86	diagnosi	1.1×10^{-2}
87	median	1.1×10^{-2}
88	month	1.1×10^{-2}
89	serotyp	1.1×10^{-2}
90	mycobacterium	1.1×10^{-2}
91	count	1×10^{-2}
92	properti	1×10^{-2}
93	simul	1×10^{-2}
94	interv	1×10^{-2}
95	hepat	1×10^{-2}
96	microbiolog	1×10^{-2}
97	structur	1×10^{-2}
98	sequenc	1×10^{-2}
99	adult	9.9×10^{-3}
100	mosquito	9.6×10^{-3}

Table D.118. The list of the top 100 words in the category Information Science and Library Science with RIGs

No.	Word	RIG	No.	Word	RIG
1	research	5.7×10^{-2}	51	collect	9.1×10^{-3}
2	inform	5.5×10^{-2}	52	protein	9.1×10^{-3}
3	librari	5.2×10^{-2}	53	explor	9×10^{-3}
4	citat	3.7×10^{-2}	54	induc	9×10^{-3}
5	user	3.2×10^{-2}	55	purpos	9×10^{-3}
6	journal	2.6×10^{-2}	56	data	8.9×10^{-3}
7	public	2.5×10^{-2}	57	focus	8.8×10^{-3}
8	social	2.5×10^{-2}	58	seek	8.8×10^{-3}
9	scienc	2.4×10^{-2}	59	profession	8.5×10^{-3}
10	web	2.3×10^{-2}	60	understand	8.3×10^{-3}
11	academ	2.3×10^{-2}	61	perspect	8.3×10^{-3}
12	paper	2.1×10^{-2}	62	archiv	8.2×10^{-3}
13	onlin	2×10^{-2}	63	discuss	7.9×10^{-3}
14	servic	2×10^{-2}	64	issu	7.9×10^{-3}
15	knowledg	1.9×10^{-2}	65	conceptu	7.8×10^{-3}
16	technolog	1.8×10^{-2}	66	adopt	7.8×10^{-3}
17	digit	1.7×10^{-2}	67	network	7.8×10^{-3}
18	author	1.7×10^{-2}	68	innov	7.7×10^{-3}
19	articl	1.7×10^{-2}	69	languag	7.7×10^{-3}
20	find	1.6×10^{-2}	70	websit	7.6×10^{-3}
21	scholar	1.5×10^{-2}	71	treatment	7.6×10^{-3}
22	methodolog	1.5×10^{-2}	72	search	7.4×10^{-3}
23	practic	1.5×10^{-2}	73	acid	7.4×10^{-3}
24	organiz	1.5×10^{-2}	74	support	7.3×10^{-3}
25	implic	1.5×10^{-2}	75	origin	7.3×10^{-3}
26	manag	1.5×10^{-2}	76	concentr	7.3×10^{-3}
27	collabor	1.4×10^{-2}	77	project	7.2×10^{-3}
28	institut	1.4×10^{-2}	78	book	7.1×10^{-3}
29	access	1.4×10^{-2}	79	literaci	7.1×10^{-3}
30	share	1.3×10^{-2}	80	framework	7×10^{-3}
31	cell	1.2×10^{-2}	81	energi	7×10^{-3}
32	survey	1.2×10^{-2}	82	decreas	6.8×10^{-3}
33	univers	1.2×10^{-2}	83	qualit	6.8×10^{-3}
34	approach	1.2×10^{-2}	84	retriev	6.7×10^{-3}
35	temperatur	1.2×10^{-2}	85	perceiv	6.5×10^{-3}
36	text	1.2×10^{-2}	86	water	6.4×10^{-3}
37	publish	1.1×10^{-2}	87	semant	6.4×10^{-3}
38	busi	1.1×10^{-2}	88	need	6.4×10^{-3}
39	communic	1.1×10^{-2}	89	strateg	6.4×10^{-3}
40	scientif	1.1×10^{-2}	90	decis	6.4×10^{-3}
41	internet	1×10^{-2}	91	databas	6.3×10^{-3}
42	document	1×10^{-2}	92	learn	6.2×10^{-3}
43	context	9.9×10^{-3}	93	topic	6.2×10^{-3}
44	resourc	9.7×10^{-3}	94	oxid	6.2×10^{-3}
45	disciplin	9.6×10^{-3}	95	countri	6.1×10^{-3}
46	literatur	9.6×10^{-3}	96	gene	6.1×10^{-3}
47	empir	9.6×10^{-3}	97	control	6×10^{-3}
48	interview	9.3×10^{-3}	98	ratio	5.9×10^{-3}
49	surfac	9.3×10^{-3}	99	educ	5.9×10^{-3}
50	provid	9.2×10^{-3}	100	impact	5.8×10^{-3}

TABLE D.119. The list of the top 100 words in the category Instruments and Instrumentation with RIGs

No.	Word	RIG
1	sensor	5.1×10^{-2}
2	detector	3.1×10^{-2}
3	beam	2.1×10^{-2}
4	measur	2.1×10^{-2}
5	conclus	1.9×10^{-2}
6	patient	1.6×10^{-2}
7	telescop	1.5×10^{-2}
8	paper	1.4×10^{-2}
9	design	1.3×10^{-2}
10	resolut	1.3×10^{-2}
11	system	1.3×10^{-2}
12	propos	1.3×10^{-2}
13	optic	1.3×10^{-2}
14	sens	1.2×10^{-2}
15	fabric	1.2×10^{-2}
16	calibr	1.2×10^{-2}
17	studi	1.2×10^{-2}
18	instrument	1.1×10^{-2}
19	devic	1.1×10^{-2}
20	detect	1.1×10^{-2}
21	suggest	1.1×10^{-2}
22	associ	1.1×10^{-2}
23	experiment	1×10^{-2}
24	treatment	9.6×10^{-3}
25	group	9.5×10^{-3}
26	signific	9.4×10^{-3}
27	kev	9.4×10^{-3}
28	actuat	9.3×10^{-3}
29	age	9.1×10^{-3}
30	signal	8.9×10^{-3}
31	base	8.8×10^{-3}
32	diseas	8.7×10^{-3}
33	prototyp	8.6×10^{-3}
34	sensit	8.4×10^{-3}
35	nois	8×10^{-3}
36	oper	8×10^{-3}
37	gene	7.4×10^{-3}
38	silicon	7.4×10^{-3}
39	mev	7.1×10^{-3}
40	may	7.1×10^{-3}
41	voltag	7×10^{-3}
42	protein	7×10^{-3}
43	clinic	7×10^{-3}
44	accuraci	6.8×10^{-3}
45	error	6.6×10^{-3}
46	scintil	6.6×10^{-3}
47	role	6.5×10^{-3}
48	array	6.5×10^{-3}
49	linear	6.5×10^{-3}
50	perform	6.5×10^{-3}
51	year	6.4×10^{-3}
52	techniqu	6.4×10^{-3}
53	find	6.4×10^{-3}
54	camera	6.3×10^{-3}
55	electrod	6.3×10^{-3}
56	risk	6.3×10^{-3}
57	simul	6.3×10^{-3}
58	examin	6.2×10^{-3}
59	popul	6.2×10^{-3}
60	present	6.2×10^{-3}
61	express	6.2×10^{-3}
62	ion	6.1×10^{-3}
63	outcom	6×10^{-3}
64	applic	6×10^{-3}
65	capabl	5.9×10^{-3}
66	mirror	5.9×10^{-3}
67	cell	5.8×10^{-3}
68	spectromet	5.8×10^{-3}
69	frequenc	5.7×10^{-3}
70	track	5.6×10^{-3}
71	imag	5.6×10^{-3}
72	precis	5.5×10^{-3}
73	pixel	5.4×10^{-3}
74	assess	5.4×10^{-3}
75	particip	5.3×10^{-3}
76	output	5.3×10^{-3}
77	rang	5.2×10^{-3}
78	ray	5.2×10^{-3}
79	evid	5.1×10^{-3}
80	laser	5.1×10^{-3}
81	energi	5×10^{-3}
82	puls	5×10^{-3}
83	piezoelectr	5×10^{-3}
84	monitor	4.9×10^{-3}
85	wavelength	4.8×10^{-3}
86	whether	4.8×10^{-3}
87	observatori	4.7×10^{-3}
88	day	4.7×10^{-3}
89	instal	4.7×10^{-3}
90	setup	4.6×10^{-3}
91	field	4.6×10^{-3}
92	increas	4.6×10^{-3}
93	filter	4.6×10^{-3}
94	treat	4.5×10^{-3}
95	among	4.5×10^{-3}
96	month	4.5×10^{-3}
97	adult	4.5×10^{-3}
98	electron	4.5×10^{-3}
99	interferomet	4.3×10^{-3}
100	neutron	4.3×10^{-3}

TABLE D.120. The list of the top 100 words in the category Integrative and Complementary Medicine with RIGs

No.	Word	RIG	No.	Word	RIG
1	medicin	1.6×10^{-1}	51	drug	1.7×10^{-2}
2	ethnopharmacolog	1.4×10^{-1}	52	serum	1.6×10^{-2}
3	conclus	9.6×10^{-2}	53	cytotox	1.6×10^{-2}
4	treatment	6.6×10^{-2}	54	western	1.5×10^{-2}
5	extract	6.6×10^{-2}	55	aim	1.5×10^{-2}
6	herbal	6.2×10^{-2}	56	express	1.5×10^{-2}
7	tradit	6×10^{-2}	57	day	1.5×10^{-2}
8	chines	5.3×10^{-2}	58	materi	1.5×10^{-2}
9	rat	5.2×10^{-2}	59	liver	1.4×10^{-2}
10	activ	5×10^{-2}	60	investig	1.4×10^{-2}
11	effect	5×10^{-2}	61	hplc	1.4×10^{-2}
12	inhibit	4.9×10^{-2}	62	tnf	1.4×10^{-2}
13	treat	4.6×10^{-2}	63	ethanol	1.3×10^{-2}
14	induc	4.3×10^{-2}	64	mtt	1.3×10^{-2}
15	herb	3.9×10^{-2}	65	inflammatori	1.3×10^{-2}
16	dose	3.6×10^{-2}	66	bioactiv	1.3×10^{-2}
17	assay	3.6×10^{-2}	67	level	1.3×10^{-2}
18	method	3.6×10^{-2}	68	propos	1.3×10^{-2}
19	signific	3.1×10^{-2}	69	leav	1.2×10^{-2}
20	oral	2.9×10^{-2}	70	beta	1.2×10^{-2}
21	antioxid	2.9×10^{-2}	71	apoptosi	1.2×10^{-2}
22	studi	2.9×10^{-2}	72	isol	1.2×10^{-2}
23	relev	2.9×10^{-2}	73	alkaloid	1.2×10^{-2}
24	antiinflammatori	2.7×10^{-2}	74	blood	1.2×10^{-2}
25	result	2.7×10^{-2}	75	efficaci	1.2×10^{-2}
26	plant	2.6×10^{-2}	76	inflamm	1.1×10^{-2}
27	administr	2.3×10^{-2}	77	spragu	1.1×10^{-2}
28	compound	2.3×10^{-2}	78	alpha	1.1×10^{-2}
29	inhibitori	2.2×10^{-2}	79	symptom	1.1×10^{-2}
30	background	2.2×10^{-2}	80	bark	1.1×10^{-2}
31	group	2.2×10^{-2}	81	dawley	1.1×10^{-2}
32	mice	2×10^{-2}	82	divid	1.1×10^{-2}
33	week	2×10^{-2}	83	vivo	1.1×10^{-2}
34	ic50	1.9×10^{-2}	84	reduc	1.1×10^{-2}
35	evalu	1.9×10^{-2}	85	stimul	1.1×10^{-2}
36	flavonoid	1.9×10^{-2}	86	sham	1.1×10^{-2}
37	decreas	1.8×10^{-2}	87	amelior	1.1×10^{-2}
38	pharmacolog	1.8×10^{-2}	88	methanol	1.1×10^{-2}
39	cell	1.8×10^{-2}	89	nitric	1×10^{-2}
40	pain	1.8×10^{-2}	90	potent	1×10^{-2}
41	administ	1.8×10^{-2}	91	therapi	1×10^{-2}
42	phytochem	1.8×10^{-2}	92	dpph	1×10^{-2}
43	paper	1.7×10^{-2}	93	clinic	1×10^{-2}
44	vitro	1.7×10^{-2}	94	chromatographi	1×10^{-2}
45	constitu	1.7×10^{-2}	95	diabet	9.8×10^{-3}
46	diseas	1.7×10^{-2}	96	crude	9.8×10^{-3}
47	blot	1.7×10^{-2}	97	anticanc	9.7×10^{-3}
48	folk	1.7×10^{-2}	98	sod	9.7×10^{-3}
49	random	1.7×10^{-2}	99	control	9.7×10^{-3}
50	therapeut	1.7×10^{-2}	100	potenti	9.4×10^{-3}

TABLE D.121. The list of the top 100 words in the category International Relations with RIGs

No.	Word	RIG	No.	Word	RIG
1	polit	1.2×10^{-1}	51	cooper	1.4×10^{-2}
2	articl	1.1×10^{-1}	52	social	1.4×10^{-2}
3	polici	7.2×10^{-2}	53	civil	1.4×10^{-2}
4	argu	7×10^{-2}	54	china	1.4×10^{-2}
5	intern	6.3×10^{-2}	55	threat	1.4×10^{-2}
6	govern	4.8×10^{-2}	56	perform	1.3×10^{-2}
7	war	4.8×10^{-2}	57	question	1.3×10^{-2}
8	countri	4.5×10^{-2}	58	transnat	1.3×10^{-2}
9	econom	3.7×10^{-2}	59	law	1.3×10^{-2}
10	foreign	3.6×10^{-2}	60	public	1.3×10^{-2}
11	actor	3.4×10^{-2}	61	engag	1.3×10^{-2}
12	nation	3.2×10^{-2}	62	agenda	1.2×10^{-2}
13	trade	3.2×10^{-2}	63	violenc	1.2×10^{-2}
14	secur	3.1×10^{-2}	64	obtain	1.2×10^{-2}
15	state	3.1×10^{-2}	65	focus	1.2×10^{-2}
16	domest	2.9×10^{-2}	66	reform	1.1×10^{-2}
17	european	2.8×10^{-2}	67	issu	1.1×10^{-2}
18	method	2.8×10^{-2}	68	invest	1.1×10^{-2}
19	union	2.7×10^{-2}	69	temperatur	1.1×10^{-2}
20	conflict	2.7×10^{-2}	70	commit	1.1×10^{-2}
21	world	2.6×10^{-2}	71	sector	1.1×10^{-2}
22	peac	2.5×10^{-2}	72	regim	1.1×10^{-2}
23	democrat	2.5×10^{-2}	73	claim	1.1×10^{-2}
24	militari	2.5×10^{-2}	74	discours	1.1×10^{-2}
25	institut	2.4×10^{-2}	75	studi	1×10^{-2}
26	global	2.2×10^{-2}	76	use	1×10^{-2}
27	result	2.2×10^{-2}	77	europ	1×10^{-2}
28	treati	2.2×10^{-2}	78	asia	1×10^{-2}
29	democraci	2.2×10^{-2}	79	legitimaci	1×10^{-2}
30	crisi	2×10^{-2}	80	seek	1×10^{-2}
31	legal	1.9×10^{-2}	81	leader	1×10^{-2}
32	economi	1.8×10^{-2}	82	compar	9.9×10^{-3}
33	argument	1.8×10^{-2}	83	conclus	9.9×10^{-3}
34	scholar	1.8×10^{-2}	84	financi	9.8×10^{-3}
35	terror	1.7×10^{-2}	85	surfac	9.8×10^{-3}
36	market	1.7×10^{-2}	86	histor	9.6×10^{-3}
37	debat	1.7×10^{-2}	87	member	9.5×10^{-3}
38	parti	1.7×10^{-2}	88	presid	9.4×10^{-3}
39	fisheri	1.7×10^{-2}	89	paramet	9.4×10^{-3}
40	negoti	1.7×10^{-2}	90	clinic	9.3×10^{-3}
41	disput	1.6×10^{-2}	91	simul	9.2×10^{-3}
42	weapon	1.6×10^{-2}	92	examin	9×10^{-3}
43	empir	1.6×10^{-2}	93	experiment	9×10^{-3}
44	liber	1.6×10^{-2}	94	protein	8.6×10^{-3}
45	whi	1.5×10^{-2}	95	emerg	8.6×10^{-3}
46	cell	1.5×10^{-2}	96	marin	8.5×10^{-3}
47	patient	1.5×10^{-2}	97	scholarship	8.5×10^{-3}
48	strateg	1.4×10^{-2}	98	export	8.4×10^{-3}
49	draw	1.4×10^{-2}	99	societi	8.4×10^{-3}
50	right	1.4×10^{-2}	100	territori	8.4×10^{-3}

Table D.122. The list of the top 100 words in the category Language and Linguistics with RIGs

No.	Word	RIG
1	languag	2.2×10^{-1}
2	english	1.3×10^{-1}
3	linguist	1.1×10^{-1}
4	corpus	7.7×10^{-2}
5	text	5.8×10^{-2}
6	speaker	5.4×10^{-2}
7	lexic	5.3×10^{-2}
8	word	5.1×10^{-2}
9	verb	4.7×10^{-2}
10	semant	4.2×10^{-2}
11	discours	4×10^{-2}
12	learner	3.9×10^{-2}
13	syntact	3.8×10^{-2}
14	speech	3.5×10^{-2}
15	corpora	3.5×10^{-2}
16	argu	3×10^{-2}
17	articl	3×10^{-2}
18	grammat	3×10^{-2}
19	grammar	2.9×10^{-2}
20	sentenc	2.8×10^{-2}
21	annot	2.7×10^{-2}
22	translat	2.7×10^{-2}
23	multilingu	2.6×10^{-2}
24	spoken	2.3×10^{-2}
25	student	2.2×10^{-2}
26	noun	2.2×10^{-2}
27	pragmat	2.1×10^{-2}
28	context	2×10^{-2}
29	phonolog	2×10^{-2}
30	teacher	1.9×10^{-2}
31	speak	1.9×10^{-2}
32	claus	1.8×10^{-2}
33	paper	1.8×10^{-2}
34	bilingu	1.8×10^{-2}
35	spanish	1.8×10^{-2}
36	narrat	1.8×10^{-2}
37	cell	1.7×10^{-2}
38	write	1.7×10^{-2}
39	phrase	1.7×10^{-2}
40	teach	1.6×10^{-2}
41	learn	1.6×10^{-2}
42	german	1.5×10^{-2}
43	written	1.5×10^{-2}
44	patient	1.5×10^{-2}
45	method	1.5×10^{-2}
46	dialect	1.4×10^{-2}
47	genr	1.4×10^{-2}
48	profici	1.4×10^{-2}
49	literari	1.4×10^{-2}
50	syntax	1.4×10^{-2}
51	vowel	1.3×10^{-2}
52	foreign	1.3×10^{-2}
53	phonet	1.3×10^{-2}
54	task	1.3×10^{-2}
55	argument	1.3×10^{-2}
56	increas	1.3×10^{-2}
57	vocabulari	1.2×10^{-2}
58	metaphor	1.2×10^{-2}
59	temperatur	1.2×10^{-2}
60	read	1.2×10^{-2}
61	classroom	1.2×10^{-2}
62	french	1.2×10^{-2}
63	question	1.2×10^{-2}
64	focus	1.2×10^{-2}
65	research	1.1×10^{-2}
66	energi	1.1×10^{-2}
67	writer	1.1×10^{-2}
68	decreas	1.1×10^{-2}
69	way	1.1×10^{-2}
70	utter	1.1×10^{-2}
71	verbal	1.1×10^{-2}
72	simul	1.1×10^{-2}
73	high	1.1×10^{-2}
74	result	1×10^{-2}
75	conclus	1×10^{-2}
76	protein	9.9×10^{-3}
77	interpret	9.8×10^{-3}
78	claim	9.4×10^{-3}
79	draw	9.3×10^{-3}
80	cultur	9×10^{-3}
81	reduc	9×10^{-3}
82	communic	8.9×10^{-3}
83	clinic	8.9×10^{-3}
84	instruct	8.8×10^{-3}
85	predic	8.7×10^{-3}
86	concentr	8.7×10^{-3}
87	ratio	8.6×10^{-3}
88	discuss	8.6×10^{-3}
89	diseas	8.6×10^{-3}
90	measur	8.5×10^{-3}
91	acid	8.4×10^{-3}
92	control	8.3×10^{-3}
93	automat	8.3×10^{-3}
94	perspect	8.2×10^{-3}
95	effect	8.2×10^{-3}
96	water	8.1×10^{-3}
97	surfac	8×10^{-3}
98	social	8×10^{-3}
99	explor	8×10^{-3}
100	low	8×10^{-3}

Table D.123. The list of the top 100 words in the category Law with RIGs

No.	Word	RIG	No.	Word	RIG
1	law	1.9×10^{-1}	51	case	1.6×10^{-2}
2	court	1.9×10^{-1}	52	oblig	1.6×10^{-2}
3	legal	1.7×10^{-1}	53	union	1.5×10^{-2}
4	articl	1.2×10^{-1}	54	privat	1.5×10^{-2}
5	crimin	6.6×10^{-2}	55	cell	1.5×10^{-2}
6	argu	6.6×10^{-2}	56	concern	1.5×10^{-2}
7	judici	6.2×10^{-2}	57	commit	1.5×10^{-2}
8	suprem	6.2×10^{-2}	58	use	1.4×10^{-2}
9	right	5.5×10^{-2}	59	judgment	1.4×10^{-2}
10	legisl	5.2×10^{-2}	60	enact	1.4×10^{-2}
11	justic	5.1×10^{-2}	61	principl	1.4×10^{-2}
12	doctrin	5×10^{-2}	62	perform	1.4×10^{-2}
13	rule	4.2×10^{-2}	63	note	1.4×10^{-2}
14	feder	3.8×10^{-2}	64	violat	1.3×10^{-2}
15	govern	3.5×10^{-2}	65	substant	1.3×10^{-2}
16	jurisdict	3.4×10^{-2}	66	regulatori	1.3×10^{-2}
17	polit	3.3×10^{-2}	67	market	1.3×10^{-2}
18	crime	3.1×10^{-2}	68	conclud	1.3×10^{-2}
19	enforc	2.9×10^{-2}	69	practic	1.2×10^{-2}
20	polici	2.9×10^{-2}	70	conflict	1.2×10^{-2}
21	constitut	2.9×10^{-2}	71	make	1.2×10^{-2}
22	state	2.8×10^{-2}	72	econom	1.2×10^{-2}
23	decis	2.7×10^{-2}	73	seek	1.2×10^{-2}
24	parti	2.6×10^{-2}	74	claus	1.2×10^{-2}
25	civil	2.5×10^{-2}	75	argument	1.2×10^{-2}
26	claim	2.4×10^{-2}	76	commiss	1.2×10^{-2}
27	result	2.4×10^{-2}	77	countri	1.2×10^{-2}
28	reform	2.4×10^{-2}	78	social	1.2×10^{-2}
29	nation	2.4×10^{-2}	79	interpret	1.2×10^{-2}
30	treati	2.4×10^{-2}	80	regul	1.2×10^{-2}
31	intern	2.3×10^{-2}	81	simul	1.2×10^{-2}
32	question	2.3×10^{-2}	82	whether	1.2×10^{-2}
33	public	2.3×10^{-2}	83	trade	1.2×10^{-2}
34	method	2.3×10^{-2}	84	temperatur	1.2×10^{-2}
35	defend	2.2×10^{-2}	85	will	1.2×10^{-2}
36	scholar	2.2×10^{-2}	86	convict	1.1×10^{-2}
37	european	2.1×10^{-2}	87	person	1.1×10^{-2}
38	disput	2.1×10^{-2}	88	surfac	1.1×10^{-2}
39	protect	2.1×10^{-2}	89	polic	1.1×10^{-2}
40	provis	2×10^{-2}	90	corpor	1.1×10^{-2}
41	judg	2×10^{-2}	91	agenc	1.1×10^{-2}
42	act	1.9×10^{-2}	92	citizen	1.1×10^{-2}
43	offend	1.9×10^{-2}	93	legitimaci	1.1×10^{-2}
44	institut	1.9×10^{-2}	94	scholarship	1.1×10^{-2}
45	amend	1.9×10^{-2}	95	paramet	1×10^{-2}
46	author	1.8×10^{-2}	96	context	1×10^{-2}
47	issu	1.8×10^{-2}	97	victim	1×10^{-2}
48	debat	1.8×10^{-2}	98	reason	1×10^{-2}
49	studi	1.7×10^{-2}	99	draw	1×10^{-2}
50	normat	1.7×10^{-2}	100	way	1×10^{-2}

Table D.124. The list of the top 100 words in the category Limnology with RIGs

No.	Word	RIG
1	water	1×10^{-1}
2	lake	1×10^{-1}
3	river	5.8×10^{-2}
4	hydrolog	4.5×10^{-2}
5	sediment	4.5×10^{-2}
6	phytoplankton	3.5×10^{-2}
7	aquat	3.3×10^{-2}
8	ecosystem	3.3×10^{-2}
9	season	2.8×10^{-2}
10	nutrient	2.8×10^{-2}
11	catchment	2.8×10^{-2}
12	dissolv	2.6×10^{-2}
13	speci	2.6×10^{-2}
14	habitat	2.5×10^{-2}
15	freshwat	2.5×10^{-2}
16	abund	2.4×10^{-2}
17	benthic	2.4×10^{-2}
18	fish	2.3×10^{-2}
19	climat	2.2×10^{-2}
20	ecolog	2.2×10^{-2}
21	groundwat	2.2×10^{-2}
22	basin	2.2×10^{-2}
23	summer	2.1×10^{-2}
24	flow	2.1×10^{-2}
25	eutroph	2×10^{-2}
26	zooplankton	2×10^{-2}
27	spatial	1.9×10^{-2}
28	phosphorus	1.9×10^{-2}
29	domin	1.9×10^{-2}
30	variabl	1.9×10^{-2}
31	communiti	1.8×10^{-2}
32	stream	1.8×10^{-2}
33	trophic	1.8×10^{-2}
34	environment	1.8×10^{-2}
35	assemblag	1.7×10^{-2}
36	depth	1.6×10^{-2}
37	patient	1.6×10^{-2}
38	runoff	1.6×10^{-2}
39	aquif	1.6×10^{-2}
40	watersh	1.5×10^{-2}
41	algal	1.5×10^{-2}
42	diatom	1.5×10^{-2}
43	shallow	1.4×10^{-2}
44	conclus	1.4×10^{-2}
45	bloom	1.4×10^{-2}
46	biomass	1.4×10^{-2}
47	hydraul	1.4×10^{-2}
48	wetland	1.3×10^{-2}
49	bay	1.3×10^{-2}
50	chlorophyl	1.3×10^{-2}
51	taxa	1.3×10^{-2}
52	alga	1.2×10^{-2}
53	flood	1.2×10^{-2}
54	rainfal	1.2×10^{-2}
55	chang	1.2×10^{-2}
56	reservoir	1.2×10^{-2}
57	tempor	1.2×10^{-2}
58	concentr	1.2×10^{-2}
59	precipit	1.1×10^{-2}
60	annual	1.1×10^{-2}
61	spring	1.1×10^{-2}
62	column	1.1×10^{-2}
63	invertebr	1.1×10^{-2}
64	winter	1.1×10^{-2}
65	condit	1.1×10^{-2}
66	period	1×10^{-2}
67	coastal	1×10^{-2}
68	nitrogen	1×10^{-2}
69	estuari	1×10^{-2}
70	scale	1×10^{-2}
71	subsurfac	1×10^{-2}
72	flux	9.7×10^{-3}
73	clinic	9.7×10^{-3}
74	soil	9.6×10^{-3}
75	veget	9.5×10^{-3}
76	plankton	9.5×10^{-3}
77	dure	9.4×10^{-3}
78	site	9.2×10^{-3}
79	pelag	9.2×10^{-3}
80	organ	9.1×10^{-3}
81	distribut	9×10^{-3}
82	estim	9×10^{-3}
83	indic	8.9×10^{-3}
84	anthropogen	8.9×10^{-3}
85	variat	8.7×10^{-3}
86	zone	8.6×10^{-3}
87	predat	8.5×10^{-3}
88	area	8.4×10^{-3}
89	north	8.2×10^{-3}
90	sea	8.1×10^{-3}
91	prey	7.7×10^{-3}
92	influenc	7.6×10^{-3}
93	wastewat	7.6×10^{-3}
94	transport	7.6×10^{-3}
95	nitrat	7.6×10^{-3}
96	land	7.6×10^{-3}
97	relat	7.4×10^{-3}
98	discharg	7.4×10^{-3}
99	diseas	7.4×10^{-3}
100	salin	7.2×10^{-3}

Table D.125. The list of the top 100 words in the category Linguistics with RIGs

No.	Word	RIG	No.	Word	RIG
1	languag	2.6×10^{-1}	51	teacher	1.3×10^{-2}
2	english	1.3×10^{-1}	52	research	1.3×10^{-2}
3	linguist	1.2×10^{-1}	53	verbal	1.3×10^{-2}
4	speaker	7.2×10^{-2}	54	french	1.3×10^{-2}
5	word	7.1×10^{-2}	55	written	1.3×10^{-2}
6	corpus	6.8×10^{-2}	56	write	1.3×10^{-2}
7	lexic	6.7×10^{-2}	57	classroom	1.2×10^{-2}
8	speech	6.6×10^{-2}	58	temperatur	1.2×10^{-2}
9	verb	5.2×10^{-2}	59	foreign	1.2×10^{-2}
10	semant	4.3×10^{-2}	60	conson	1.2×10^{-2}
11	text	4.2×10^{-2}	61	patient	1.2×10^{-2}
12	syntact	4.1×10^{-2}	62	narrat	1.1×10^{-2}
13	sentenc	3.9×10^{-2}	63	argument	1.1×10^{-2}
14	learner	3.8×10^{-2}	64	energi	1.1×10^{-2}
15	grammat	3.6×10^{-2}	65	genr	1.1×10^{-2}
16	discours	3.6×10^{-2}	66	nativ	1×10^{-2}
17	phonolog	3.3×10^{-2}	67	interpret	1×10^{-2}
18	corpora	3×10^{-2}	68	question	1×10^{-2}
19	bilingu	2.8×10^{-2}	69	protein	1×10^{-2}
20	spoken	2.8×10^{-2}	70	focus	9.9×10^{-3}
21	noun	2.7×10^{-2}	71	teach	9.9×10^{-3}
22	grammar	2.5×10^{-2}	72	simul	9.8×10^{-3}
23	speak	2.5×10^{-2}	73	discuss	9.7×10^{-3}
24	task	2.5×10^{-2}	74	communic	9.7×10^{-3}
25	multilingu	2.4×10^{-2}	75	paper	9.7×10^{-3}
26	articl	2.4×10^{-2}	76	draw	9.1×10^{-3}
27	argu	2.4×10^{-2}	77	instruct	9.1×10^{-3}
28	annot	2.3×10^{-2}	78	acquisit	9×10^{-3}
29	spanish	2.3×10^{-2}	79	increas	9×10^{-3}
30	context	2.2×10^{-2}	80	decreas	9×10^{-3}
31	learn	2.1×10^{-2}	81	concentr	8.9×10^{-3}
32	vowel	1.9×10^{-2}	82	metaphor	8.8×10^{-3}
33	claus	1.9×10^{-2}	83	high	8.8×10^{-3}
34	phonet	1.8×10^{-2}	84	social	8.6×10^{-3}
35	pragmat	1.8×10^{-2}	85	reader	8.4×10^{-3}
36	phrase	1.8×10^{-2}	86	method	8.4×10^{-3}
37	profici	1.8×10^{-2}	87	water	8.3×10^{-3}
38	cell	1.8×10^{-2}	88	school	8.3×10^{-3}
39	vocabulari	1.6×10^{-2}	89	acid	8.3×10^{-3}
40	student	1.6×10^{-2}	90	chines	8.2×10^{-3}
41	translat	1.6×10^{-2}	91	examin	8.1×10^{-3}
42	utter	1.5×10^{-2}	92	way	8×10^{-3}
43	children	1.4×10^{-2}	93	voic	8×10^{-3}
44	syntax	1.4×10^{-2}	94	cognit	7.9×10^{-3}
45	particip	1.4×10^{-2}	95	explor	7.9×10^{-3}
46	syllabl	1.4×10^{-2}	96	surfac	7.8×10^{-3}
47	german	1.3×10^{-2}	97	predic	7.7×10^{-3}
48	dialect	1.3×10^{-2}	98	resourc	7.7×10^{-3}
49	read	1.3×10^{-2}	99	automat	7.6×10^{-3}
50	listen	1.3×10^{-2}	100	claim	7.5×10^{-3}

TABLE D.126. The list of the top 100 words in the category Literary Reviews with RIGs

No.	Word	RIG
1	essay	2.3×10^{-1}
2	connor	2.1×10^{-1}
3	steven	1.3×10^{-1}
4	irish	1.2×10^{-1}
5	writer	7.8×10^{-2}
6	poem	6.3×10^{-2}
7	polit	5.6×10^{-2}
8	cultur	5.5×10^{-2}
9	ireland	4.9×10^{-2}
10	theatr	4.6×10^{-2}
11	narrat	4.6×10^{-2}
12	result	4.6×10^{-2}
13	stori	3.9×10^{-2}
14	intellectu	3.9×10^{-2}
15	imagin	3.5×10^{-2}
16	poetri	3.5×10^{-2}
17	beckett	3.4×10^{-2}
18	write	3.2×10^{-2}
19	argu	3.2×10^{-2}
20	adorno	3.1×10^{-2}
21	studi	3.1×10^{-2}
22	book	3.1×10^{-2}
23	ulyss	3.1×10^{-2}
24	english	3×10^{-2}
25	fiction	3×10^{-2}
26	playwright	3×10^{-2}
27	method	3×10^{-2}
28	work	2.9×10^{-2}
29	celebr	2.9×10^{-2}
30	magic	2.9×10^{-2}
31	oeuvr	2.8×10^{-2}
32	refus	2.7×10^{-2}
33	literari	2.7×10^{-2}
34	lyric	2.6×10^{-2}
35	london	2.5×10^{-2}
36	press	2.5×10^{-2}
37	prose	2.5×10^{-2}
38	theme	2.4×10^{-2}
39	vers	2.4×10^{-2}
40	obliqu	2.3×10^{-2}
41	read	2.3×10^{-2}
42	think	2.3×10^{-2}
43	jean	2.3×10^{-2}
44	york	2.2×10^{-2}
45	struggl	2.2×10^{-2}
46	drama	2.2×10^{-2}
47	model	2.1×10^{-2}
48	poetic	2.1×10^{-2}
49	heroin	2.1×10^{-2}
50	wrote	2.1×10^{-2}

No.	Word	RIG
51	claim	2.1×10^{-2}
52	british	2×10^{-2}
53	compar	2×10^{-2}
54	sexual	2×10^{-2}
55	assault	2×10^{-2}
56	tour	2×10^{-2}
57	paper	2×10^{-2}
58	effect	2×10^{-2}
59	jewish	2×10^{-2}
60	languag	2×10^{-2}
61	transcend	2×10^{-2}
62	reader	1.9×10^{-2}
63	system	1.9×10^{-2}
64	piec	1.8×10^{-2}
65	studio	1.8×10^{-2}
66	investig	1.8×10^{-2}
67	troubl	1.7×10^{-2}
68	world	1.7×10^{-2}
69	rememb	1.7×10^{-2}
70	entertain	1.7×10^{-2}
71	marriag	1.7×10^{-2}
72	life	1.7×10^{-2}
73	genr	1.7×10^{-2}
74	shelley	1.7×10^{-2}
75	public	1.7×10^{-2}
76	love	1.6×10^{-2}
77	metaphor	1.6×10^{-2}
78	measur	1.6×10^{-2}
79	retreat	1.6×10^{-2}
80	tell	1.6×10^{-2}
81	radic	1.6×10^{-2}
82	john	1.6×10^{-2}
83	show	1.6×10^{-2}
84	scholarship	1.5×10^{-2}
85	bront	1.5×10^{-2}
86	high	1.5×10^{-2}
87	figur	1.5×10^{-2}
88	famous	1.5×10^{-2}
89	artist	1.5×10^{-2}
90	articl	1.4×10^{-2}
91	propos	1.4×10^{-2}
92	art	1.4×10^{-2}
93	increas	1.4×10^{-2}
94	joyc	1.4×10^{-2}
95	inquiri	1.4×10^{-2}
96	antholog	1.4×10^{-2}
97	philosoph	1.4×10^{-2}
98	tradit	1.4×10^{-2}
99	control	1.4×10^{-2}
100	coloni	1.4×10^{-2}

Table D.127. The list of the top 100 words in the category Literary Theory and Criticism with RIGs

No.	Word	RIG	No.	Word	RIG
1	literari	6.6×10^{-2}	51	higher	1.2×10^{-2}
2	english	5.6×10^{-2}	52	histor	1.2×10^{-2}
3	languag	5×10^{-2}	53	tri	1.2×10^{-2}
4	essay	4.3×10^{-2}	54	attempt	1.2×10^{-2}
5	fiction	4.1×10^{-2}	55	model	1.2×10^{-2}
6	poetri	4×10^{-2}	56	ident	1.2×10^{-2}
7	result	4×10^{-2}	57	obtain	1.2×10^{-2}
8	text	3.9×10^{-2}	58	control	1.2×10^{-2}
9	writer	3.7×10^{-2}	59	philosoph	1.2×10^{-2}
10	cultur	3.7×10^{-2}	60	patient	1.2×10^{-2}
11	narrat	3.7×10^{-2}	61	protagonist	1.2×10^{-2}
12	write	3.6×10^{-2}	62	view	1.2×10^{-2}
13	stori	3.5×10^{-2}	63	author	1.2×10^{-2}
14	reader	3.5×10^{-2}	64	data	1.1×10^{-2}
15	read	2.8×10^{-2}	65	foreign	1.1×10^{-2}
16	discours	2.8×10^{-2}	66	system	1.1×10^{-2}
17	poem	2.7×10^{-2}	67	cell	1.1×10^{-2}
18	polit	2.6×10^{-2}	68	postcoloni	1.1×10^{-2}
19	poet	2.5×10^{-2}	69	art	1.1×10^{-2}
20	translat	2.3×10^{-2}	70	decreas	1.1×10^{-2}
21	centuri	2.2×10^{-2}	71	low	1.1×10^{-2}
22	charact	2.2×10^{-2}	72	tale	1.1×10^{-2}
23	linguist	2.1×10^{-2}	73	perspect	1.1×10^{-2}
24	argu	2×10^{-2}	74	reduc	1.1×10^{-2}
25	book	1.9×10^{-2}	75	compar	1.1×10^{-2}
26	use	1.9×10^{-2}	76	novelist	1.1×10^{-2}
27	measur	1.9×10^{-2}	77	improv	1×10^{-2}
28	poetic	1.8×10^{-2}	78	british	1×10^{-2}
29	method	1.8×10^{-2}	79	beckett	1×10^{-2}
30	stylist	1.7×10^{-2}	80	theori	1×10^{-2}
31	aesthet	1.7×10^{-2}	81	war	1×10^{-2}
32	modern	1.7×10^{-2}	82	theme	1×10^{-2}
33	contemporari	1.7×10^{-2}	83	embodi	1×10^{-2}
34	american	1.6×10^{-2}	84	treatment	1×10^{-2}
35	artist	1.6×10^{-2}	85	turkish	1×10^{-2}
36	ideolog	1.6×10^{-2}	86	religi	1×10^{-2}
37	romanian	1.6×10^{-2}	87	literatur	1×10^{-2}
38	world	1.6×10^{-2}	88	genr	1×10^{-2}
39	increas	1.5×10^{-2}	89	realiti	9.9×10^{-3}
40	high	1.5×10^{-2}	90	societi	9.9×10^{-3}
41	imagin	1.5×10^{-2}	91	christian	9.8×10^{-3}
42	teach	1.5×10^{-2}	92	temperatur	9.8×10^{-3}
43	way	1.5×10^{-2}	93	lower	9.6×10^{-3}
44	theolog	1.4×10^{-2}	94	divin	9.5×10^{-3}
45	chines	1.4×10^{-2}	95	william	9.4×10^{-3}
46	articl	1.4×10^{-2}	96	simul	9.2×10^{-3}
47	perform	1.3×10^{-2}	97	paramet	9.2×10^{-3}
48	rate	1.3×10^{-2}	98	written	9.2×10^{-3}
49	symbol	1.3×10^{-2}	99	histori	9.2×10^{-3}
50	effect	1.2×10^{-2}	100	tragic	9.2×10^{-3}

Table D.128. The list of the top 100 words in the category Literature with RIGs

No.	Word	RIG	No.	Word	RIG
1	literari	1.2×10^{-1}	51	engag	1.7×10^{-2}
2	essay	1×10^{-1}	52	nineteenth	1.7×10^{-2}
3	narrat	9.1×10^{-2}	53	high	1.7×10^{-2}
4	articl	8.6×10^{-2}	54	increas	1.6×10^{-2}
5	fiction	8×10^{-2}	55	literatur	1.6×10^{-2}
6	text	7.9×10^{-2}	56	measur	1.6×10^{-2}
7	argu	7.6×10^{-2}	57	scholar	1.6×10^{-2}
8	writer	6.1×10^{-2}	58	shakespear	1.6×10^{-2}
9	write	5.9×10^{-2}	59	american	1.6×10^{-2}
10	read	5.9×10^{-2}	60	textual	1.6×10^{-2}
11	result	5.2×10^{-2}	61	draw	1.6×10^{-2}
12	centuri	4.7×10^{-2}	62	linguist	1.6×10^{-2}
13	poem	4.5×10^{-2}	63	figur	1.6×10^{-2}
14	poetri	4.2×10^{-2}	64	prose	1.5×10^{-2}
15	reader	4.1×10^{-2}	65	patient	1.5×10^{-2}
16	polit	4×10^{-2}	66	studi	1.5×10^{-2}
17	stori	3.8×10^{-2}	67	novel	1.5×10^{-2}
18	cultur	3.7×10^{-2}	68	ideolog	1.4×10^{-2}
19	contemporari	3.5×10^{-2}	69	written	1.4×10^{-2}
20	rhetor	3.4×10^{-2}	70	scholarship	1.4×10^{-2}
21	genr	3.4×10^{-2}	71	audienc	1.4×10^{-2}
22	poetic	3.3×10^{-2}	72	cell	1.4×10^{-2}
23	discours	3.3×10^{-2}	73	represent	1.4×10^{-2}
24	poet	3.1×10^{-2}	74	rate	1.4×10^{-2}
25	english	3.1×10^{-2}	75	ethic	1.4×10^{-2}
26	charact	2.6×10^{-2}	76	improv	1.3×10^{-2}
27	postcoloni	2.6×10^{-2}	77	john	1.3×10^{-2}
28	book	2.5×10^{-2}	78	obtain	1.3×10^{-2}
29	histor	2.5×10^{-2}	79	social	1.3×10^{-2}
30	protagonist	2.4×10^{-2}	80	william	1.3×10^{-2}
31	method	2.4×10^{-2}	81	ident	1.3×10^{-2}
32	way	2.4×10^{-2}	82	reduc	1.2×10^{-2}
33	imagin	2.3×10^{-2}	83	evalu	1.2×10^{-2}
34	aesthet	2.3×10^{-2}	84	novelist	1.2×10^{-2}
35	languag	2.2×10^{-2}	85	notion	1.2×10^{-2}
36	author	2.2×10^{-2}	86	compar	1.2×10^{-2}
37	translat	2.2×10^{-2}	87	modernist	1.2×10^{-2}
38	explor	2.1×10^{-2}	88	question	1.2×10^{-2}
39	use	2.1×10^{-2}	89	higher	1.2×10^{-2}
40	war	2×10^{-2}	90	manuscript	1.1×10^{-2}
41	histori	2×10^{-2}	91	tale	1.1×10^{-2}
42	critiqu	1.9×10^{-2}	92	control	1.1×10^{-2}
43	modern	1.9×10^{-2}	93	system	1.1×10^{-2}
44	artist	1.9×10^{-2}	94	struggl	1.1×10^{-2}
45	world	1.8×10^{-2}	95	decreas	1.1×10^{-2}
46	effect	1.8×10^{-2}	96	art	1.1×10^{-2}
47	critic	1.8×10^{-2}	97	temperatur	1.1×10^{-2}
48	love	1.8×10^{-2}	98	effici	1.1×10^{-2}
49	data	1.8×10^{-2}	99	trope	1.1×10^{-2}
50	work	1.7×10^{-2}	100	context	1.1×10^{-2}

TABLE D.129. The list of the top 100 words in the category Literature, African, Australian, Canadian with RIGs

No.	Word	RIG
1	essay	1.8×10^{-1}
2	african	1.8×10^{-1}
3	literari	1.6×10^{-1}
4	postcoloni	1.5×10^{-1}
5	narrat	1.3×10^{-1}
6	writer	9.2×10^{-2}
7	read	8×10^{-2}
8	novel	7.6×10^{-2}
9	fiction	7.4×10^{-2}
10	diaspora	7.2×10^{-2}
11	argu	6.8×10^{-2}
12	articl	6.6×10^{-2}
13	protagonist	6.5×10^{-2}
14	africa	6.2×10^{-2}
15	discours	6×10^{-2}
16	text	5.9×10^{-2}
17	violenc	5.3×10^{-2}
18	coloni	5×10^{-2}
19	black	4.8×10^{-2}
20	novelist	4.7×10^{-2}
21	result	4.6×10^{-2}
22	cultur	4.6×10^{-2}
23	notion	4.2×10^{-2}
24	imagin	3.8×10^{-2}
25	inscrib	3.7×10^{-2}
26	modern	3.7×10^{-2}
27	contemporari	3.5×10^{-2}
28	subvers	3.2×10^{-2}
29	write	3.2×10^{-2}
30	societi	3×10^{-2}
31	histor	3×10^{-2}
32	atlant	3×10^{-2}
33	racial	2.9×10^{-2}
34	genr	2.9×10^{-2}
35	storytel	2.9×10^{-2}
36	overlook	2.9×10^{-2}
37	craft	2.9×10^{-2}
38	world	2.8×10^{-2}
39	portray	2.8×10^{-2}
40	symbol	2.8×10^{-2}
41	poetic	2.7×10^{-2}
42	poetri	2.7×10^{-2}
43	poem	2.7×10^{-2}
44	stori	2.6×10^{-2}
45	way	2.6×10^{-2}
46	critiqu	2.6×10^{-2}
47	literatur	2.5×10^{-2}
48	slaveri	2.5×10^{-2}
49	method	2.5×10^{-2}
50	agenc	2.4×10^{-2}
51	intertextu	2.4×10^{-2}
52	nation	2.4×10^{-2}
53	diaspor	2.4×10^{-2}
54	tradit	2.3×10^{-2}
55	ideolog	2.3×10^{-2}
56	languag	2.3×10^{-2}
57	trope	2.3×10^{-2}
58	question	2.3×10^{-2}
59	discurs	2.2×10^{-2}
60	transnat	2.2×10^{-2}
61	women	2.2×10^{-2}
62	base	2.1×10^{-2}
63	articul	2.1×10^{-2}
64	stylist	2.1×10^{-2}
65	race	2.1×10^{-2}
66	war	2.1×10^{-2}
67	centuri	2.1×10^{-2}
68	context	2×10^{-2}
69	english	1.9×10^{-2}
70	solidar	1.9×10^{-2}
71	author	1.9×10^{-2}
72	histori	1.9×10^{-2}
73	data	1.9×10^{-2}
74	west	1.9×10^{-2}
75	use	1.9×10^{-2}
76	charact	1.9×10^{-2}
77	high	1.9×10^{-2}
78	polit	1.8×10^{-2}
79	particular	1.8×10^{-2}
80	explor	1.8×10^{-2}
81	intersect	1.8×10^{-2}
82	represent	1.8×10^{-2}
83	indian	1.7×10^{-2}
84	american	1.7×10^{-2}
85	attempt	1.7×10^{-2}
86	trauma	1.6×10^{-2}
87	written	1.6×10^{-2}
88	global	1.6×10^{-2}
89	neglect	1.6×10^{-2}
90	effect	1.6×10^{-2}
91	slave	1.6×10^{-2}
92	haunt	1.6×10^{-2}
93	book	1.6×10^{-2}
94	figur	1.6×10^{-2}
95	satir	1.5×10^{-2}
96	work	1.5×10^{-2}
97	journey	1.5×10^{-2}
98	comedi	1.5×10^{-2}
99	critic	1.5×10^{-2}
100	speak	1.5×10^{-2}

TABLE D.130. The list of the top 100 words in the category Literature, American with RIGs

No.	Word	RIG
1	essay	2.6×10^{-1}
2	american	1.2×10^{-1}
3	antebellum	9.9×10^{-2}
4	argu	9.2×10^{-2}
5	dickinson	8.9×10^{-2}
6	poem	8.5×10^{-2}
7	literari	8.3×10^{-2}
8	narrat	8.2×10^{-2}
9	melvill	8×10^{-2}
10	poe	6.6×10^{-2}
11	slaveri	6.1×10^{-2}
12	stori	6×10^{-2}
13	text	5.7×10^{-2}
14	fiction	5.6×10^{-2}
15	jewish	5.2×10^{-2}
16	tale	5×10^{-2}
17	writer	5×10^{-2}
18	centuri	4.6×10^{-2}
19	satir	4.3×10^{-2}
20	letter	4.3×10^{-2}
21	poet	4.3×10^{-2}
22	write	4.1×10^{-2}
23	america	4×10^{-2}
24	scholar	3.9×10^{-2}
25	result	3.8×10^{-2}
26	engag	3.8×10^{-2}
27	racial	3.7×10^{-2}
28	critiqu	3.6×10^{-2}
29	histor	3.5×10^{-2}
30	poetic	3.4×10^{-2}
31	poetri	3.4×10^{-2}
32	rhetor	3.4×10^{-2}
33	read	3.3×10^{-2}
34	nineteenth	3.3×10^{-2}
35	emili	3.3×10^{-2}
36	reader	3.3×10^{-2}
37	trope	3.2×10^{-2}
38	cultur	3.1×10^{-2}
39	discours	3×10^{-2}
40	jew	3×10^{-2}
41	contemporari	2.9×10^{-2}
42	commentari	2.9×10^{-2}
43	anglo	2.9×10^{-2}
44	william	2.8×10^{-2}
45	archiv	2.8×10^{-2}
46	peter	2.7×10^{-2}
47	tell	2.6×10^{-2}
48	studi	2.6×10^{-2}
49	prose	2.5×10^{-2}
50	african	2.4×10^{-2}
51	war	2.4×10^{-2}
52	represent	2.4×10^{-2}
53	man	2.4×10^{-2}
54	slave	2.3×10^{-2}
55	biographi	2.3×10^{-2}
56	charact	2.3×10^{-2}
57	york	2.2×10^{-2}
58	formal	2.2×10^{-2}
59	genr	2.2×10^{-2}
60	articul	2×10^{-2}
61	method	2×10^{-2}
62	artist	1.9×10^{-2}
63	effect	1.9×10^{-2}
64	figur	1.8×10^{-2}
65	lyric	1.8×10^{-2}
66	audienc	1.8×10^{-2}
67	subvers	1.8×10^{-2}
68	publish	1.7×10^{-2}
69	newspap	1.7×10^{-2}
70	editor	1.7×10^{-2}
71	compar	1.7×10^{-2}
72	base	1.7×10^{-2}
73	way	1.6×10^{-2}
74	negoti	1.6×10^{-2}
75	portray	1.6×10^{-2}
76	violenc	1.6×10^{-2}
77	increas	1.5×10^{-2}
78	work	1.5×10^{-2}
79	articl	1.5×10^{-2}
80	edit	1.5×10^{-2}
81	autobiograph	1.5×10^{-2}
82	race	1.5×10^{-2}
83	protagonist	1.5×10^{-2}
84	data	1.5×10^{-2}
85	critic	1.5×10^{-2}
86	textual	1.5×10^{-2}
87	scholarship	1.5×10^{-2}
88	magazin	1.5×10^{-2}
89	voic	1.5×10^{-2}
90	wrote	1.5×10^{-2}
91	evalu	1.4×10^{-2}
92	chapter	1.4×10^{-2}
93	civil	1.4×10^{-2}
94	world	1.4×10^{-2}
95	offer	1.4×10^{-2}
96	antholog	1.4×10^{-2}
97	ident	1.4×10^{-2}
98	test	1.4×10^{-2}
99	novel	1.4×10^{-2}
100	improv	1.3×10^{-2}

TABLE D.131. The list of the top 100 words in the category Literature, British Isles with RIGs

No.	Word	RIG	No.	Word	RIG
1	essay	1.1×10^{-1}	51	high	1.8×10^{-2}
2	shakespear	9.9×10^{-2}	52	richard	1.8×10^{-2}
3	bront	8.4×10^{-2}	53	work	1.8×10^{-2}
4	joyc	6.5×10^{-2}	54	romanc	1.8×10^{-2}
5	argu	5.7×10^{-2}	55	modern	1.8×10^{-2}
6	poem	5.7×10^{-2}	56	twentieth	1.7×10^{-2}
7	result	5.3×10^{-2}	57	walter	1.7×10^{-2}
8	narrat	4.5×10^{-2}	58	reader	1.7×10^{-2}
9	english	4.5×10^{-2}	59	data	1.6×10^{-2}
10	chaucer	4.5×10^{-2}	60	poetri	1.6×10^{-2}
11	articl	4.4×10^{-2}	61	book	1.6×10^{-2}
12	text	4.3×10^{-2}	62	renaiss	1.6×10^{-2}
13	read	3.5×10^{-2}	63	histor	1.6×10^{-2}
14	john	3.4×10^{-2}	64	edit	1.6×10^{-2}
15	literari	3.2×10^{-2}	65	writer	1.6×10^{-2}
16	centuri	3.2×10^{-2}	66	explor	1.6×10^{-2}
17	theatr	3.2×10^{-2}	67	audienc	1.5×10^{-2}
18	ulyss	3.1×10^{-2}	68	improv	1.5×10^{-2}
19	poet	3×10^{-2}	69	obtain	1.4×10^{-2}
20	tale	3×10^{-2}	70	signific	1.4×10^{-2}
21	method	2.7×10^{-2}	71	compar	1.4×10^{-2}
22	studi	2.7×10^{-2}	72	measur	1.4×10^{-2}
23	jame	2.7×10^{-2}	73	emili	1.4×10^{-2}
24	write	2.6×10^{-2}	74	show	1.3×10^{-2}
25	artist	2.6×10^{-2}	75	celebr	1.3×10^{-2}
26	samuel	2.6×10^{-2}	76	way	1.3×10^{-2}
27	theatric	2.6×10^{-2}	77	king	1.3×10^{-2}
28	charact	2.5×10^{-2}	78	genr	1.3×10^{-2}
29	beckett	2.5×10^{-2}	79	persona	1.3×10^{-2}
30	play	2.4×10^{-2}	80	cultur	1.3×10^{-2}
31	polit	2.3×10^{-2}	81	allus	1.2×10^{-2}
32	mediev	2.3×10^{-2}	82	depict	1.2×10^{-2}
33	effect	2.3×10^{-2}	83	increas	1.2×10^{-2}
34	fiction	2.2×10^{-2}	84	cell	1.2×10^{-2}
35	drama	2.2×10^{-2}	85	investig	1.2×10^{-2}
36	scholar	2.2×10^{-2}	86	eighteenth	1.2×10^{-2}
37	henri	2.2×10^{-2}	87	report	1.2×10^{-2}
38	poetic	2.1×10^{-2}	88	languag	1.2×10^{-2}
39	use	2.1×10^{-2}	89	conclus	1.2×10^{-2}
40	aesthet	2.1×10^{-2}	90	test	1.1×10^{-2}
41	imagin	2.1×10^{-2}	91	determin	1.1×10^{-2}
42	thoma	2×10^{-2}	92	vers	1.1×10^{-2}
43	system	2×10^{-2}	93	rate	1.1×10^{-2}
44	england	2×10^{-2}	94	evalu	1.1×10^{-2}
45	contemporari	2×10^{-2}	95	sampl	1.1×10^{-2}
46	prose	2×10^{-2}	96	love	1.1×10^{-2}
47	victorian	2×10^{-2}	97	level	1.1×10^{-2}
48	figur	2×10^{-2}	98	satir	1.1×10^{-2}
49	base	1.9×10^{-2}	99	patient	1.1×10^{-2}
50	moral	1.9×10^{-2}	100	late	1.1×10^{-2}

Table D.132. The list of the top 100 words in the category Literature, German, Dutch, Scandinavian with RIGs

No.	Word	RIG	No.	Word	RIG
1	german	1.4×10^{-1}	51	question	2.1×10^{-2}
2	literari	1.1×10^{-1}	52	contemporari	2.1×10^{-2}
3	von	9.5×10^{-2}	53	data	2.1×10^{-2}
4	und	9.4×10^{-2}	54	imagin	2.1×10^{-2}
5	der	8.4×10^{-2}	55	war	2×10^{-2}
6	essay	8.3×10^{-2}	56	effect	2×10^{-2}
7	ein	8.1×10^{-2}	57	languag	1.9×10^{-2}
8	write	7.5×10^{-2}	58	engag	1.9×10^{-2}
9	die	6.9×10^{-2}	59	literatur	1.9×10^{-2}
10	das	6.9×10^{-2}	60	critiqu	1.9×10^{-2}
11	text	6.8×10^{-2}	61	represent	1.9×10^{-2}
12	narrat	6.8×10^{-2}	62	insist	1.8×10^{-2}
13	articl	6.7×10^{-2}	63	peter	1.8×10^{-2}
14	auf	6.6×10^{-2}	64	islam	1.8×10^{-2}
15	germani	6.6×10^{-2}	65	theme	1.8×10^{-2}
16	wird	5.9×10^{-2}	66	studi	1.8×10^{-2}
17	des	5.4×10^{-2}	67	genr	1.8×10^{-2}
18	protagonist	5.4×10^{-2}	68	ambival	1.7×10^{-2}
19	aesthet	5.1×10^{-2}	69	philosophi	1.6×10^{-2}
20	read	4.9×10^{-2}	70	embodi	1.6×10^{-2}
21	writer	4.5×10^{-2}	71	mediev	1.6×10^{-2}
22	poetic	4.4×10^{-2}	72	recept	1.6×10^{-2}
23	postwar	3.9×10^{-2}	73	artist	1.6×10^{-2}
24	poetri	3.9×10^{-2}	74	victim	1.5×10^{-2}
25	argu	3.9×10^{-2}	75	work	1.5×10^{-2}
26	result	3.8×10^{-2}	76	increas	1.5×10^{-2}
27	jean	3.7×10^{-2}	77	translat	1.5×10^{-2}
28	poet	3.7×10^{-2}	78	reflect	1.5×10^{-2}
29	discours	3.6×10^{-2}	79	world	1.5×10^{-2}
30	use	3.2×10^{-2}	80	modern	1.5×10^{-2}
31	dem	3.2×10^{-2}	81	secular	1.5×10^{-2}
32	centuri	3.1×10^{-2}	82	base	1.5×10^{-2}
33	book	2.9×10^{-2}	83	intertextu	1.4×10^{-2}
34	cultur	2.8×10^{-2}	84	nineteenth	1.4×10^{-2}
35	benjamin	2.7×10^{-2}	85	idea	1.4×10^{-2}
36	prose	2.7×10^{-2}	86	love	1.4×10^{-2}
37	poem	2.6×10^{-2}	87	publish	1.4×10^{-2}
38	polit	2.6×10^{-2}	88	notion	1.4×10^{-2}
39	johann	2.6×10^{-2}	89	test	1.4×10^{-2}
40	depict	2.5×10^{-2}	90	romant	1.4×10^{-2}
41	histor	2.4×10^{-2}	91	ideolog	1.4×10^{-2}
42	berlin	2.4×10^{-2}	92	myth	1.4×10^{-2}
43	fiction	2.3×10^{-2}	93	exil	1.4×10^{-2}
44	muslim	2.3×10^{-2}	94	ident	1.4×10^{-2}
45	method	2.3×10^{-2}	95	trope	1.3×10^{-2}
46	stori	2.2×10^{-2}	96	obtain	1.3×10^{-2}
47	walter	2.2×10^{-2}	97	scholarship	1.3×10^{-2}
48	reader	2.2×10^{-2}	98	feminist	1.3×10^{-2}
49	charact	2.2×10^{-2}	99	paul	1.3×10^{-2}
50	author	2.1×10^{-2}	100	measur	1.3×10^{-2}

TABLE D.133. The list of the top 100 words in the category Literature, Romance with RIGs

No.	Word	RIG	No.	Word	RIG
1	articl	9.4×10^{-2}	51	des	1.6×10^{-2}
2	literari	7.6×10^{-2}	52	histor	1.6×10^{-2}
3	spanish	6.9×10^{-2}	53	linguist	1.6×10^{-2}
4	narrat	6.7×10^{-2}	54	theatr	1.6×10^{-2}
5	essay	6×10^{-2}	55	data	1.5×10^{-2}
6	centuri	5.6×10^{-2}	56	author	1.5×10^{-2}
7	war	5.5×10^{-2}	57	ident	1.5×10^{-2}
8	writer	5.3×10^{-2}	58	effect	1.5×10^{-2}
9	text	4.7×10^{-2}	59	eighteenth	1.5×10^{-2}
10	argu	4.5×10^{-2}	60	cinema	1.4×10^{-2}
11	french	4.5×10^{-2}	61	dialect	1.4×10^{-2}
12	write	4×10^{-2}	62	notion	1.4×10^{-2}
13	result	4×10^{-2}	63	jean	1.4×10^{-2}
14	spain	3.8×10^{-2}	64	explor	1.4×10^{-2}
15	artist	3.4×10^{-2}	65	represent	1.4×10^{-2}
16	polit	3.1×10^{-2}	66	novel	1.4×10^{-2}
17	poetic	3.1×10^{-2}	67	ideolog	1.4×10^{-2}
18	fiction	3.1×10^{-2}	68	franc	1.4×10^{-2}
19	cultur	2.9×10^{-2}	69	test	1.4×10^{-2}
20	civil	2.8×10^{-2}	70	improv	1.3×10^{-2}
21	method	2.8×10^{-2}	71	work	1.3×10^{-2}
22	discours	2.7×10^{-2}	72	recept	1.3×10^{-2}
23	poem	2.6×10^{-2}	73	measur	1.3×10^{-2}
24	poet	2.6×10^{-2}	74	patient	1.3×10^{-2}
25	protagonist	2.6×10^{-2}	75	critiqu	1.3×10^{-2}
26	novelist	2.6×10^{-2}	76	obtain	1.3×10^{-2}
27	stori	2.5×10^{-2}	77	compar	1.2×10^{-2}
28	italian	2.4×10^{-2}	78	base	1.2×10^{-2}
29	genr	2.3×10^{-2}	79	photograph	1.2×10^{-2}
30	poetri	2.3×10^{-2}	80	way	1.2×10^{-2}
31	use	2.1×10^{-2}	81	determin	1.2×10^{-2}
32	que	2.1×10^{-2}	82	son	1.2×10^{-2}
33	exil	2.1×10^{-2}	83	attempt	1.2×10^{-2}
34	question	2×10^{-2}	84	creativ	1.2×10^{-2}
35	nineteenth	1.9×10^{-2}	85	increas	1.2×10^{-2}
36	reader	1.9×10^{-2}	86	trope	1.2×10^{-2}
37	aesthet	1.9×10^{-2}	87	higher	1.2×10^{-2}
38	portray	1.9×10^{-2}	88	low	1.2×10^{-2}
39	contemporari	1.8×10^{-2}	89	sixteenth	1.1×10^{-2}
40	read	1.8×10^{-2}	90	oeuvr	1.1×10^{-2}
41	written	1.8×10^{-2}	91	cathol	1.1×10^{-2}
42	metaphor	1.8×10^{-2}	92	theater	1.1×10^{-2}
43	languag	1.7×10^{-2}	93	offer	1.1×10^{-2}
44	modern	1.7×10^{-2}	94	cell	1.1×10^{-2}
45	high	1.7×10^{-2}	95	itali	1.1×10^{-2}
46	figur	1.7×10^{-2}	96	philosoph	1.1×10^{-2}
47	histori	1.7×10^{-2}	97	effici	1×10^{-2}
48	romanc	1.7×10^{-2}	98	decreas	1×10^{-2}
49	charact	1.6×10^{-2}	99	societi	1×10^{-2}
50	system	1.6×10^{-2}	100	persona	1×10^{-2}

Table D.134. The list of the top 100 words in the category Literature, Slavic with RIGs

No.	Word	RIG	No.	Word	RIG
1	russian	1.4×10^{-1}	51	manifest	1.8×10^{-2}
2	panegyr	8.5×10^{-2}	52	discours	1.8×10^{-2}
3	text	8.3×10^{-2}	53	literatur	1.8×10^{-2}
4	poetic	8.1×10^{-2}	54	sentiment	1.7×10^{-2}
5	poetri	8×10^{-2}	55	scholar	1.7×10^{-2}
6	catherin	7.2×10^{-2}	56	thought	1.7×10^{-2}
7	literari	6.4×10^{-2}	57	rhetor	1.7×10^{-2}
8	soviet	6.1×10^{-2}	58	polit	1.7×10^{-2}
9	poet	5.2×10^{-2}	59	studi	1.7×10^{-2}
10	insist	4.9×10^{-2}	60	foreground	1.7×10^{-2}
11	use	4.6×10^{-2}	61	entitl	1.6×10^{-2}
12	centuri	4.3×10^{-2}	62	realiti	1.6×10^{-2}
13	reign	4.2×10^{-2}	63	measur	1.6×10^{-2}
14	essay	4×10^{-2}	64	explor	1.6×10^{-2}
15	genr	4×10^{-2}	65	writer	1.6×10^{-2}
16	work	3.9×10^{-2}	66	write	1.6×10^{-2}
17	postmodern	3.8×10^{-2}	67	high	1.5×10^{-2}
18	result	3.8×10^{-2}	68	effect	1.4×10^{-2}
19	artist	3.6×10^{-2}	69	propos	1.4×10^{-2}
20	cultur	3.5×10^{-2}	70	activ	1.4×10^{-2}
21	eighteenth	3.4×10^{-2}	71	person	1.4×10^{-2}
22	method	3×10^{-2}	72	russia	1.4×10^{-2}
23	iconographi	2.9×10^{-2}	73	philosoph	1.4×10^{-2}
24	articl	2.9×10^{-2}	74	control	1.4×10^{-2}
25	visual	2.8×10^{-2}	75	observ	1.4×10^{-2}
26	christian	2.8×10^{-2}	76	themselv	1.4×10^{-2}
27	ironi	2.7×10^{-2}	77	exemplifi	1.3×10^{-2}
28	mystic	2.7×10^{-2}	78	ideolog	1.3×10^{-2}
29	mytholog	2.7×10^{-2}	79	conclus	1.3×10^{-2}
30	can	2.6×10^{-2}	80	confront	1.3×10^{-2}
31	memoir	2.5×10^{-2}	81	test	1.3×10^{-2}
32	attempt	2.3×10^{-2}	82	margaret	1.3×10^{-2}
33	historiographi	2.3×10^{-2}	83	improv	1.3×10^{-2}
34	cinema	2.3×10^{-2}	84	determin	1.3×10^{-2}
35	increas	2.2×10^{-2}	85	indic	1.3×10^{-2}
36	protagonist	2.2×10^{-2}	86	deleuz	1.2×10^{-2}
37	peter	2.1×10^{-2}	87	oeuvr	1.2×10^{-2}
38	cathol	2.1×10^{-2}	88	univers	1.2×10^{-2}
39	existenti	2×10^{-2}	89	associ	1.2×10^{-2}
40	metaphys	2×10^{-2}	90	themat	1.2×10^{-2}
41	art	2×10^{-2}	91	tone	1.2×10^{-2}
42	myth	2×10^{-2}	92	satir	1.2×10^{-2}
43	idea	2×10^{-2}	93	book	1.2×10^{-2}
44	imperi	2×10^{-2}	94	obtain	1.2×10^{-2}
45	enlighten	1.9×10^{-2}	95	influenti	1.2×10^{-2}
46	investig	1.8×10^{-2}	96	emphas	1.2×10^{-2}
47	conceptu	1.8×10^{-2}	97	design	1.2×10^{-2}
48	data	1.8×10^{-2}	98	painter	1.2×10^{-2}
49	differ	1.8×10^{-2}	99	cult	1.1×10^{-2}
50	deal	1.8×10^{-2}	100	verbal	1.1×10^{-2}

Table D.135. The list of the top 100 words in the category Logic with RIGs

No.	Word	RIG
1	logic	1.1×10^{-1}
2	semant	4.5×10^{-2}
3	formal	4.1×10^{-2}
4	languag	3.5×10^{-2}
5	proof	3.2×10^{-2}
6	prove	3.1×10^{-2}
7	notion	2.9×10^{-2}
8	proposit	2.6×10^{-2}
9	algebra	2.4×10^{-2}
10	theorem	2.2×10^{-2}
11	set	2×10^{-2}
12	problem	1.9×10^{-2}
13	defin	1.8×10^{-2}
14	introduc	1.8×10^{-2}
15	effect	1.7×10^{-2}
16	decid	1.7×10^{-2}
17	high	1.6×10^{-2}
18	finit	1.6×10^{-2}
19	increas	1.6×10^{-2}
20	satisfi	1.6×10^{-2}
21	paper	1.6×10^{-2}
22	patient	1.5×10^{-2}
23	formula	1.5×10^{-2}
24	class	1.5×10^{-2}
25	theori	1.5×10^{-2}
26	abstract	1.5×10^{-2}
27	signific	1.5×10^{-2}
28	predic	1.5×10^{-2}
29	infinit	1.4×10^{-2}
30	studi	1.4×10^{-2}
31	bound	1.4×10^{-2}
32	activ	1.4×10^{-2}
33	polynomi	1.4×10^{-2}
34	verif	1.4×10^{-2}
35	call	1.3×10^{-2}
36	equival	1.3×10^{-2}
37	cell	1.3×10^{-2}
38	conclus	1.2×10^{-2}
39	program	1.2×10^{-2}
40	give	1.2×10^{-2}
41	general	1.2×10^{-2}
42	comput	1.1×10^{-2}
43	rate	1.1×10^{-2}
44	rule	1.1×10^{-2}
45	temperatur	1.1×10^{-2}
46	graph	1.1×10^{-2}
47	indic	1×10^{-2}
48	surfac	1×10^{-2}
49	extend	1×10^{-2}
50	dure	1×10^{-2}

No.	Word	RIG
51	reason	1×10^{-2}
52	answer	1×10^{-2}
53	truth	9.9×10^{-3}
54	found	9.8×10^{-3}
55	complet	9.6×10^{-3}
56	clinic	9.6×10^{-3}
57	compar	9.5×10^{-3}
58	constraint	9.5×10^{-3}
59	decreas	9.4×10^{-3}
60	materi	9.1×10^{-3}
61	modal	9.1×10^{-3}
62	algorithm	9×10^{-3}
63	classic	9×10^{-3}
64	method	9×10^{-3}
65	instanc	8.8×10^{-3}
66	age	8.8×10^{-3}
67	higher	8.8×10^{-3}
68	reveal	8.7×10^{-3}
69	diseas	8.7×10^{-3}
70	given	8.7×10^{-3}
71	extens	8.6×10^{-3}
72	correl	8.6×10^{-3}
73	observ	8.6×10^{-3}
74	suggest	8.6×10^{-3}
75	argument	8.5×10^{-3}
76	measur	8.5×10^{-3}
77	probabilist	8.4×10^{-3}
78	low	8.4×10^{-3}
79	respons	8.2×10^{-3}
80	sound	8×10^{-3}
81	result	7.9×10^{-3}
82	protein	7.8×10^{-3}
83	syntact	7.8×10^{-3}
84	treatment	7.8×10^{-3}
85	sampl	7.8×10^{-3}
86	framework	7.7×10^{-3}
87	tree	7.4×10^{-3}
88	potenti	7.3×10^{-3}
89	acid	7.3×10^{-3}
90	everi	7.2×10^{-3}
91	control	7.2×10^{-3}
92	influenc	7.2×10^{-3}
93	concentr	7.2×10^{-3}
94	water	7.1×10^{-3}
95	year	7.1×10^{-3}
96	energi	7×10^{-3}
97	sentenc	6.9×10^{-3}
98	restrict	6.7×10^{-3}
99	region	6.6×10^{-3}
100	specifi	6.6×10^{-3}

TABLE D.136. The list of the top 100 words in the category Management with RIGs

No.	Word	RIG	No.	Word	RIG
1	firm	6.7×10^{-2}	51	entrepreneurship	1.2×10^{-2}
2	manag	6.3×10^{-2}	52	treatment	1.2×10^{-2}
3	research	6.2×10^{-2}	53	protein	1.1×10^{-2}
4	busi	5.6×10^{-2}	54	focus	1.1×10^{-2}
5	organiz	5.5×10^{-2}	55	conceptu	1.1×10^{-2}
6	compani	5.2×10^{-2}	56	profit	1.1×10^{-2}
7	market	4.6×10^{-2}	57	leader	1.1×10^{-2}
8	employe	4.4×10^{-2}	58	govern	1.1×10^{-2}
9	innov	3.7×10^{-2}	59	detect	1×10^{-2}
10	empir	3.1×10^{-2}	60	entrepreneur	1×10^{-2}
11	paper	3×10^{-2}	61	draw	1×10^{-2}
12	implic	2.8×10^{-2}	62	survey	1×10^{-2}
13	enterpris	2.8×10^{-2}	63	make	9.7×10^{-3}
14	strateg	2.7×10^{-2}	64	acid	9.7×10^{-3}
15	financi	2.6×10^{-2}	65	conclus	9.7×10^{-3}
16	practic	2.6×10^{-2}	66	articl	9.6×10^{-3}
17	literatur	2.4×10^{-2}	67	clinic	9.5×10^{-3}
18	corpor	2.4×10^{-2}	68	diseas	9.5×10^{-3}
19	find	2.4×10^{-2}	69	purpos	9.3×10^{-3}
20	social	2.3×10^{-2}	70	technolog	9.2×10^{-3}
21	competit	2.3×10^{-2}	71	theoret	9.1×10^{-3}
22	industri	2.3×10^{-2}	72	induc	8.9×10^{-3}
23	manageri	2.2×10^{-2}	73	stakehold	8.9×10^{-3}
24	econom	2.1×10^{-2}	74	observ	8.8×10^{-3}
25	custom	2×10^{-2}	75	supplier	8.8×10^{-3}
26	cell	1.9×10^{-2}	76	perceiv	8.7×10^{-3}
27	relationship	1.9×10^{-2}	77	gene	8.7×10^{-3}
28	economi	1.8×10^{-2}	78	intern	8.5×10^{-3}
29	decis	1.8×10^{-2}	79	sale	8.5×10^{-3}
30	methodolog	1.7×10^{-2}	80	price	8.5×10^{-3}
31	perspect	1.7×10^{-2}	81	explor	8.3×10^{-3}
32	capit	1.7×10^{-2}	82	team	8.3×10^{-3}
33	knowledg	1.6×10^{-2}	83	institut	8.2×10^{-3}
34	servic	1.6×10^{-2}	84	impact	8.2×10^{-3}
35	job	1.5×10^{-2}	85	method	8.2×10^{-3}
36	invest	1.5×10^{-2}	86	asset	8.1×10^{-3}
37	resourc	1.5×10^{-2}	87	framework	8.1×10^{-3}
38	countri	1.5×10^{-2}	88	author	8×10^{-3}
39	develop	1.5×10^{-2}	89	speci	8×10^{-3}
40	theori	1.5×10^{-2}	90	strategi	7.9×10^{-3}
41	leadership	1.5×10^{-2}	91	anteced	7.9×10^{-3}
42	patient	1.5×10^{-2}	92	china	7.9×10^{-3}
43	sector	1.4×10^{-2}	93	oxid	7.9×10^{-3}
44	temperatur	1.4×10^{-2}	94	experiment	7.8×10^{-3}
45	entrepreneuri	1.3×10^{-2}	95	satisfact	7.7×10^{-3}
46	organ	1.3×10^{-2}	96	import	7.7×10^{-3}
47	surfac	1.3×10^{-2}	97	issu	7.6×10^{-3}
48	context	1.3×10^{-2}	98	rang	7.6×10^{-3}
49	polici	1.2×10^{-2}	99	interview	7.4×10^{-3}
50	organis	1.2×10^{-2}	100	water	7.4×10^{-3}

Table D.137. The list of the top 100 words in the category Marine and Freshwater Biology with RIGs

No.	Word	RIG	No.	Word	RIG
1	fish	1.2×10^{-1}	51	invertebr	1.7×10^{-2}
2	speci	1.1×10^{-1}	52	popul	1.7×10^{-2}
3	sea	6.5×10^{-2}	53	gulf	1.7×10^{-2}
4	marin	6×10^{-2}	54	diatom	1.6×10^{-2}
5	habitat	5.1×10^{-2}	55	pacif	1.6×10^{-2}
6	water	4.7×10^{-2}	56	larval	1.6×10^{-2}
7	abund	4.6×10^{-2}	57	reproduct	1.5×10^{-2}
8	coastal	3.8×10^{-2}	58	mediterranean	1.5×10^{-2}
9	benthic	3.6×10^{-2}	59	chlorophyl	1.5×10^{-2}
10	ecosystem	3.6×10^{-2}	60	plankton	1.5×10^{-2}
11	river	3.2×10^{-2}	61	north	1.4×10^{-2}
12	atlant	3.2×10^{-2}	62	growth	1.4×10^{-2}
13	fisheri	3.1×10^{-2}	63	conclus	1.4×10^{-2}
14	ecolog	3×10^{-2}	64	island	1.4×10^{-2}
15	juvenil	3×10^{-2}	65	diet	1.4×10^{-2}
16	lake	3×10^{-2}	66	southern	1.4×10^{-2}
17	coast	2.9×10^{-2}	67	food	1.3×10^{-2}
18	freshwat	2.8×10^{-2}	68	salmon	1.3×10^{-2}
19	season	2.7×10^{-2}	69	indic	1.3×10^{-2}
20	sediment	2.7×10^{-2}	70	salin	1.3×10^{-2}
21	bay	2.6×10^{-2}	71	gill	1.3×10^{-2}
22	phytoplankton	2.6×10^{-2}	72	trout	1.3×10^{-2}
23	estuari	2.5×10^{-2}	73	suggest	1.3×10^{-2}
24	aquat	2.4×10^{-2}	74	stock	1.3×10^{-2}
25	trophic	2.4×10^{-2}	75	shallow	1.2×10^{-2}
26	alga	2.4×10^{-2}	76	concentr	1.2×10^{-2}
27	spawn	2.3×10^{-2}	77	domin	1.2×10^{-2}
28	ocean	2.3×10^{-2}	78	divers	1.2×10^{-2}
29	biomass	2.2×10^{-2}	79	spring	1.2×10^{-2}
30	reef	2.2×10^{-2}	80	organ	1.2×10^{-2}
31	prey	2.2×10^{-2}	81	shrimp	1.2×10^{-2}
32	nutrient	2.2×10^{-2}	82	northern	1.2×10^{-2}
33	communiti	2.1×10^{-2}	83	area	1.1×10^{-2}
34	assemblag	2.1×10^{-2}	84	eutroph	1.1×10^{-2}
35	algal	2.1×10^{-2}	85	bivalv	1.1×10^{-2}
36	environment	2.1×10^{-2}	86	pattern	1.1×10^{-2}
37	predat	2.1×10^{-2}	87	anthropogen	1.1×10^{-2}
38	summer	2×10^{-2}	88	collect	1.1×10^{-2}
39	feed	1.9×10^{-2}	89	winter	1.1×10^{-2}
40	zooplankton	1.9×10^{-2}	90	south	1.1×10^{-2}
41	paper	1.9×10^{-2}	91	genus	1×10^{-2}
42	taxa	1.9×10^{-2}	92	site	1×10^{-2}
43	method	1.9×10^{-2}	93	propos	1×10^{-2}
44	estuarin	1.9×10^{-2}	94	shore	1×10^{-2}
45	aquacultur	1.9×10^{-2}	95	egg	1×10^{-2}
46	patient	1.8×10^{-2}	96	dissolv	1×10^{-2}
47	bloom	1.8×10^{-2}	97	taxonom	1×10^{-2}
48	larva	1.8×10^{-2}	98	seawat	1×10^{-2}
49	pelag	1.7×10^{-2}	99	catch	9.9×10^{-3}
50	coral	1.7×10^{-2}	100	spatial	9.9×10^{-3}

Table D.138. The list of the top 100 words in the category Materials Science, Biomaterials with RIGs

No.	Word	RIG	No.	Word	RIG
1	cell	5.6×10^{-2}	51	matrix	1.3×10^{-2}
2	biocompat	5.4×10^{-2}	52	stem	1.3×10^{-2}
3	scaffold	5.2×10^{-2}	53	mesenchym	1.3×10^{-2}
4	vitro	4.7×10^{-2}	54	hydrophil	1.2×10^{-2}
5	poli	4.6×10^{-2}	55	composit	1.2×10^{-2}
6	tissu	4.1×10^{-2}	56	modulus	1.2×10^{-2}
7	biomateri	3.8×10^{-2}	57	porous	1.2×10^{-2}
8	implant	3.7×10^{-2}	58	cultur	1.2×10^{-2}
9	deliveri	3.5×10^{-2}	59	calcium	1.2×10^{-2}
10	bone	3.5×10^{-2}	60	hydrophob	1.2×10^{-2}
11	releas	3.4×10^{-2}	61	scan	1.2×10^{-2}
12	hydrogel	3.4×10^{-2}	62	fibroblast	1.2×10^{-2}
13	adhes	3.1×10^{-2}	63	copolym	1.2×10^{-2}
14	surfac	3.1×10^{-2}	64	enhanc	1.2×10^{-2}
15	vivo	3×10^{-2}	65	ethylen	1.1×10^{-2}
16	nanoparticl	2.9×10^{-2}	66	potenti	1.1×10^{-2}
17	hydroxyapatit	2.7×10^{-2}	67	mechan	1.1×10^{-2}
18	properti	2.5×10^{-2}	68	human	1.1×10^{-2}
19	prepar	2.4×10^{-2}	69	conjug	1×10^{-2}
20	cytotox	2.4×10^{-2}	70	selfassembl	1×10^{-2}
21	drug	2.4×10^{-2}	71	attach	1×10^{-2}
22	collagen	2.3×10^{-2}	72	exhibit	1×10^{-2}
23	polym	2.2×10^{-2}	73	biolog	1×10^{-2}
24	prolifer	2.1×10^{-2}	74	synthes	1×10^{-2}
25	microscopi	2.1×10^{-2}	75	sem	1×10^{-2}
26	biomed	2×10^{-2}	76	fluoresc	9.7×10^{-3}
27	coat	2×10^{-2}	77	extracellular	9.6×10^{-3}
28	regener	2×10^{-2}	78	titanium	9.5×10^{-3}
29	bioactiv	2×10^{-2}	79	gel	9.4×10^{-3}
30	engin	2×10^{-2}	80	methacryl	9.3×10^{-3}
31	load	1.9×10^{-2}	81	spectroscopi	8.8×10^{-3}
32	glycol	1.9×10^{-2}	82	immobil	8.6×10^{-3}
33	chitosan	1.9×10^{-2}	83	assay	8.6×10^{-3}
34	encapsul	1.8×10^{-2}	84	dental	8.3×10^{-3}
35	osteoblast	1.8×10^{-2}	85	lactic	8.2×10^{-3}
36	biodegrad	1.8×10^{-2}	86	apatit	8.2×10^{-3}
37	polymer	1.7×10^{-2}	87	antibacteri	8.2×10^{-3}
38	materi	1.7×10^{-2}	88	electrospin	8.2×10^{-3}
39	osteogen	1.6×10^{-2}	89	modifi	8.1×10^{-3}
40	phosphat	1.6×10^{-2}	90	strength	8×10^{-3}
41	viabil	1.6×10^{-2}	91	heal	8×10^{-3}
42	promis	1.6×10^{-2}	92	anticanc	7.9×10^{-3}
43	fabric	1.5×10^{-2}	93	paper	7.9×10^{-3}
44	peg	1.4×10^{-2}	94	format	7.9×10^{-3}
45	acid	1.4×10^{-2}	95	background	7.7×10^{-3}
46	graft	1.4×10^{-2}	96	carrier	7.6×10^{-3}
47	crosslink	1.3×10^{-2}	97	regen	7.5×10^{-3}
48	morpholog	1.3×10^{-2}	98	onto	7.5×10^{-3}
49	applic	1.3×10^{-2}	99	uptak	7.5×10^{-3}
50	cellular	1.3×10^{-2}	100	infrar	7.4×10^{-3}

TABLE D.139. The list of the top 100 words in the category Materials Science, Ceramics with RIGs

No.	Word	RIG
1	ceram	1.5×10^{-1}
2	sinter	1×10^{-1}
3	temperatur	6.5×10^{-2}
4	prepar	6.4×10^{-2}
5	powder	6.3×10^{-2}
6	diffract	5.8×10^{-2}
7	xrd	5.7×10^{-2}
8	microstructur	5.5×10^{-2}
9	composit	5.4×10^{-2}
10	ray	5.4×10^{-2}
11	degre	5.3×10^{-2}
12	properti	5×10^{-2}
13	glass	4.4×10^{-2}
14	phase	4.4×10^{-2}
15	sol	3.7×10^{-2}
16	sem	3.6×10^{-2}
17	microscopi	3.2×10^{-2}
18	dope	3.2×10^{-2}
19	synthes	3.1×10^{-2}
20	scan	3×10^{-2}
21	materi	2.9×10^{-2}
22	electron	2.9×10^{-2}
23	grain	2.8×10^{-2}
24	gel	2.6×10^{-2}
25	thermal	2.4×10^{-2}
26	sic	2.3×10^{-2}
27	dielectr	2.3×10^{-2}
28	solid	2.2×10^{-2}
29	alumina	2.2×10^{-2}
30	crystal	2.2×10^{-2}
31	al2o3	2.1×10^{-2}
32	spectroscopi	2×10^{-2}
33	particl	2×10^{-2}
34	size	1.9×10^{-2}
35	oxid	1.9×10^{-2}
36	conclus	1.9×10^{-2}
37	poros	1.8×10^{-2}
38	crystallin	1.8×10^{-2}
39	calcin	1.8×10^{-2}
40	structur	1.8×10^{-2}
41	fabric	1.8×10^{-2}
42	mpa	1.8×10^{-2}
43	patient	1.7×10^{-2}
44	coat	1.6×10^{-2}
45	morpholog	1.6×10^{-2}
46	sio2	1.6×10^{-2}
47	precursor	1.5×10^{-2}
48	strength	1.4×10^{-2}
49	tetragon	1.3×10^{-2}
50	tough	1.3×10^{-2}
51	ferroelectr	1.3×10^{-2}
52	sampl	1.3×10^{-2}
53	porous	1.3×10^{-2}
54	tio2	1.2×10^{-2}
55	film	1.2×10^{-2}
56	ion	1.2×10^{-2}
57	pore	1.2×10^{-2}
58	content	1.2×10^{-2}
59	tem	1.2×10^{-2}
60	format	1.1×10^{-2}
61	model	1.1×10^{-2}
62	amorph	1.1×10^{-2}
63	silica	1.1×10^{-2}
64	data	1.1×10^{-2}
65	heat	1.1×10^{-2}
66	background	1.1×10^{-2}
67	surfac	1.1×10^{-2}
68	zno	1.1×10^{-2}
69	year	1.1×10^{-2}
70	crystallit	1.1×10^{-2}
71	reaction	1×10^{-2}
72	obtain	1×10^{-2}
73	room	1×10^{-2}
74	perovskit	1×10^{-2}
75	press	1×10^{-2}
76	electr	1×10^{-2}
77	densiti	1×10^{-2}
78	hydroxyapatit	1×10^{-2}
79	diseas	1×10^{-2}
80	associ	1×10^{-2}
81	character	1×10^{-2}
82	paper	1×10^{-2}
83	chemic	9.8×10^{-3}
84	exhibit	9.7×10^{-3}
85	investig	9.7×10^{-3}
86	level	9.6×10^{-3}
87	pure	9.6×10^{-3}
88	mol	9.6×10^{-3}
89	mechan	9.5×10^{-3}
90	dispers	9.5×10^{-3}
91	silic	9.5×10^{-3}
92	dens	9.4×10^{-3}
93	find	9.4×10^{-3}
94	express	9.2×10^{-3}
95	deposit	9.2×10^{-3}
96	assess	9.2×10^{-3}
97	risk	9.1×10^{-3}
98	raman	8.9×10^{-3}
99	spinel	8.9×10^{-3}
100	hydrotherm	8.8×10^{-3}

TABLE D.140. The list of the top 100 words in the category Materials Science, Characterization and Testing with RIGs

No.	Word	RIG
1	steel	3.5×10^{-2}
2	crack	3×10^{-2}
3	materi	2.8×10^{-2}
4	stress	2.3×10^{-2}
5	load	2.2×10^{-2}
6	specimen	2.1×10^{-2}
7	tensil	2.1×10^{-2}
8	strain	1.8×10^{-2}
9	test	1.7×10^{-2}
10	weld	1.7×10^{-2}
11	strength	1.7×10^{-2}
12	deform	1.6×10^{-2}
13	patient	1.6×10^{-2}
14	microstructur	1.5×10^{-2}
15	corros	1.5×10^{-2}
16	fatigu	1.5×10^{-2}
17	nondestruct	1.5×10^{-2}
18	pipelin	1.4×10^{-2}
19	pipe	1.4×10^{-2}
20	concret	1.4×10^{-2}
21	conclus	1.3×10^{-2}
22	fractur	1.3×10^{-2}
23	alloy	1.3×10^{-2}
24	reinforc	1.1×10^{-2}
25	finit	1.1×10^{-2}
26	failur	1×10^{-2}
27	compress	1×10^{-2}
28	properti	1×10^{-2}
29	modulus	9.9×10^{-3}
30	plastic	9.6×10^{-3}
31	element	9.6×10^{-3}
32	elast	9.3×10^{-3}
33	experiment	9.1×10^{-3}
34	diseas	8.9×10^{-3}
35	clinic	8.9×10^{-3}
36	composit	8.7×10^{-3}
37	temperatur	8.6×10^{-3}
38	mechan	8.5×10^{-3}
39	pavement	8.3×10^{-3}
40	defect	8.2×10^{-3}
41	protein	8.2×10^{-3}
42	tough	8.1×10^{-3}
43	shear	8.1×10^{-3}
44	ultrason	7.7×10^{-3}
45	bend	7.5×10^{-3}
46	background	7.2×10^{-3}
47	thick	7.2×10^{-3}
48	gene	7×10^{-3}
49	surfac	6.8×10^{-3}
50	activ	6.6×10^{-3}
51	uniaxi	6.6×10^{-3}
52	asphalt	6.4×10^{-3}
53	diffract	6.1×10^{-3}
54	damag	6.1×10^{-3}
55	group	6×10^{-3}
56	ductil	6×10^{-3}
57	displac	5.9×10^{-3}
58	popul	5.9×10^{-3}
59	plate	5.9×10^{-3}
60	harden	5.5×10^{-3}
61	cell	5.4×10^{-3}
62	associ	5.4×10^{-3}
63	suggest	5.4×10^{-3}
64	grain	5.3×10^{-3}
65	austenit	5.1×10^{-3}
66	structur	5.1×10^{-3}
67	human	5.1×10^{-3}
68	express	5.1×10^{-3}
69	speci	4.9×10^{-3}
70	aluminum	4.9×10^{-3}
71	thermal	4.7×10^{-3}
72	particip	4.7×10^{-3}
73	cancer	4.7×10^{-3}
74	brittl	4.7×10^{-3}
75	paramet	4.6×10^{-3}
76	lamin	4.6×10^{-3}
77	heat	4.5×10^{-3}
78	stainless	4.4×10^{-3}
79	outcom	4.3×10^{-3}
80	fiber	4.3×10^{-3}
81	therapi	4.3×10^{-3}
82	epoxi	4.2×10^{-3}
83	powder	4.1×10^{-3}
84	among	4.1×10^{-3}
85	may	4.1×10^{-3}
86	stiff	4.1×10^{-3}
87	manufactur	4.1×10^{-3}
88	tension	4.1×10^{-3}
89	mediat	4×10^{-3}
90	scan	4×10^{-3}
91	find	4×10^{-3}
92	deflect	4×10^{-3}
93	year	4×10^{-3}
94	propag	3.9×10^{-3}
95	wear	3.7×10^{-3}
96	friction	3.7×10^{-3}
97	carri	3.7×10^{-3}
98	beam	3.7×10^{-3}
99	tumor	3.7×10^{-3}
100	coat	3.7×10^{-3}

TABLE D.141. The list of the top 100 words in the category Materials Science, Coatings and Films with RIGs

No.	Word	RIG	No.	Word	RIG
1	film	1.1×10^{-1}	51	rough	1.5×10^{-2}
2	deposit	9.4×10^{-2}	52	wear	1.4×10^{-2}
3	coat	8.8×10^{-2}	53	steel	1.4×10^{-2}
4	surfac	7.4×10^{-2}	54	carbon	1.3×10^{-2}
5	substrat	6.5×10^{-2}	55	adhes	1.3×10^{-2}
6	layer	5.1×10^{-2}	56	thermal	1.3×10^{-2}
7	thin	4.7×10^{-2}	57	hard	1.3×10^{-2}
8	spectroscopi	4.7×10^{-2}	58	electrodeposit	1.3×10^{-2}
9	electron	4.6×10^{-2}	59	format	1.2×10^{-2}
10	microscopi	4.2×10^{-2}	60	silicon	1.2×10^{-2}
11	ray	4×10^{-2}	61	vapor	1.2×10^{-2}
12	sputter	3.9×10^{-2}	62	glass	1.2×10^{-2}
13	oxid	3.7×10^{-2}	63	optic	1.2×10^{-2}
14	electrochem	3.6×10^{-2}	64	investig	1.2×10^{-2}
15	properti	3.5×10^{-2}	65	titanium	1.2×10^{-2}
16	scan	3.2×10^{-2}	66	dispers	1.2×10^{-2}
17	diffract	3×10^{-2}	67	tin	1.1×10^{-2}
18	magnetron	2.9×10^{-2}	68	character	1.1×10^{-2}
19	sem	2.9×10^{-2}	69	plasma	1.1×10^{-2}
20	morpholog	2.8×10^{-2}	70	zno	1.1×10^{-2}
21	prepar	2.8×10^{-2}	71	materi	1.1×10^{-2}
22	temperatur	2.7×10^{-2}	72	exhibit	1.1×10^{-2}
23	corros	2.7×10^{-2}	73	nano	1.1×10^{-2}
24	xrd	2.5×10^{-2}	74	raman	1.1×10^{-2}
25	photoelectron	2.5×10^{-2}	75	associ	1.1×10^{-2}
26	composit	2.5×10^{-2}	76	year	1×10^{-2}
27	thick	2.4×10^{-2}	77	afm	1×10^{-2}
28	atom	2.4×10^{-2}	78	densiti	1×10^{-2}
29	resist	2.4×10^{-2}	79	adsorpt	1×10^{-2}
30	microstructur	2.3×10^{-2}	80	diseas	1×10^{-2}
31	xps	2.2×10^{-2}	81	oxygen	1×10^{-2}
32	anneal	2.2×10^{-2}	82	clinic	1×10^{-2}
33	electrolyt	2.2×10^{-2}	83	contact	1×10^{-2}
34	electrod	2.1×10^{-2}	84	grown	1×10^{-2}
35	chemic	2.1×10^{-2}	85	imped	9.8×10^{-3}
36	conclus	1.9×10^{-2}	86	object	9.7×10^{-3}
37	amorph	1.9×10^{-2}	87	nanoparticl	9.6×10^{-3}
38	alloy	1.9×10^{-2}	88	background	9.5×10^{-3}
39	patient	1.9×10^{-2}	89	tem	9.5×10^{-3}
40	fabric	1.8×10^{-2}	90	energi	9.1×10^{-3}
41	structur	1.8×10^{-2}	91	photocatalyt	9×10^{-3}
42	spray	1.8×10^{-2}	92	phase	9×10^{-3}
43	degre	1.7×10^{-2}	93	particl	9×10^{-3}
44	tio2	1.7×10^{-2}	94	synthes	8.7×10^{-3}
45	dope	1.7×10^{-2}	95	data	8.7×10^{-3}
46	metal	1.7×10^{-2}	96	express	8.7×10^{-3}
47	crystallin	1.7×10^{-2}	97	interfac	8.5×10^{-3}
48	anod	1.6×10^{-2}	98	electr	8.5×10^{-3}
49	ion	1.6×10^{-2}	99	onto	8.4×10^{-3}
50	cathod	1.5×10^{-2}	100	diamond	8.3×10^{-3}

TABLE D.142. The list of the top 100 words in the category Materials Science, Composites with RIGs

No.	Word	RIG	No.	Word	RIG
1	composit	1.4×10^{-1}	51	buckl	1.5×10^{-2}
2	reinforc	1×10^{-1}	52	prepar	1.4×10^{-2}
3	strength	7.3×10^{-2}	53	plastic	1.3×10^{-2}
4	fiber	6.6×10^{-2}	54	damag	1.3×10^{-2}
5	lamin	6.6×10^{-2}	55	experiment	1.3×10^{-2}
6	tensil	6.2×10^{-2}	56	investig	1.2×10^{-2}
7	properti	5.7×10^{-2}	57	mold	1.2×10^{-2}
8	epoxi	5.1×10^{-2}	58	temperatur	1.2×10^{-2}
9	load	4.5×10^{-2}	59	tough	1.2×10^{-2}
10	matrix	4.2×10^{-2}	60	transvers	1.1×10^{-2}
11	modulus	4.1×10^{-2}	61	microscopi	1.1×10^{-2}
12	materi	3.9×10^{-2}	62	adhes	1.1×10^{-2}
13	polym	3.8×10^{-2}	63	displac	1.1×10^{-2}
14	shear	3.8×10^{-2}	64	cement	1.1×10^{-2}
15	mechan	3.7×10^{-2}	65	dispers	1.1×10^{-2}
16	resin	3.5×10^{-2}	66	behavior	1×10^{-2}
17	fibr	3.5×10^{-2}	67	sem	1×10^{-2}
18	carbon	3×10^{-2}	68	cure	1×10^{-2}
19	flexur	3×10^{-2}	69	bond	1×10^{-2}
20	concret	2.6×10^{-2}	70	background	1×10^{-2}
21	nanocomposit	2.4×10^{-2}	71	diseas	9.9×10^{-3}
22	deform	2.4×10^{-2}	72	beam	9.8×10^{-3}
23	glass	2.3×10^{-2}	73	structur	9.7×10^{-3}
24	compress	2.2×10^{-2}	74	clinic	9.6×10^{-3}
25	thermal	2.2×10^{-2}	75	associ	9.6×10^{-3}
26	crack	2.2×10^{-2}	76	steel	9.5×10^{-3}
27	filler	2.1×10^{-2}	77	inplan	9×10^{-3}
28	stiff	2.1×10^{-2}	78	year	8.5×10^{-3}
29	elast	2.1×10^{-2}	79	ductil	8.5×10^{-3}
30	specimen	2×10^{-2}	80	manufactur	8.4×10^{-3}
31	finit	2×10^{-2}	81	static	8.4×10^{-3}
32	polypropylen	1.9×10^{-2}	82	deflect	8.4×10^{-3}
33	stress	1.8×10^{-2}	83	electron	8.3×10^{-3}
34	interfaci	1.8×10^{-2}	84	human	8.2×10^{-3}
35	patient	1.8×10^{-2}	85	protein	8.1×10^{-3}
36	bend	1.7×10^{-2}	86	multiwal	8×10^{-3}
37	strain	1.7×10^{-2}	87	melt	7.9×10^{-3}
38	microstructur	1.7×10^{-2}	88	activ	7.9×10^{-3}
39	thick	1.7×10^{-2}	89	surfac	7.7×10^{-3}
40	fractur	1.6×10^{-2}	90	yarn	7.7×10^{-3}
41	plate	1.6×10^{-2}	91	numer	7.7×10^{-3}
42	sandwich	1.6×10^{-2}	92	blend	7.6×10^{-3}
43	woven	1.6×10^{-2}	93	thermogravimetr	7.5×10^{-3}
44	test	1.6×10^{-2}	94	layer	7.3×10^{-3}
45	nanotub	1.6×10^{-2}	95	risk	7.2×10^{-3}
46	scan	1.5×10^{-2}	96	resist	7.2×10^{-3}
47	conclus	1.5×10^{-2}	97	gene	7.1×10^{-3}
48	element	1.5×10^{-2}	98	foam	7×10^{-3}
49	failur	1.5×10^{-2}	99	effect	7×10^{-3}
50	fabric	1.5×10^{-2}	100	morpholog	6.9×10^{-3}

TABLE D.143. The list of the top 100 words in the category Materials Science, Multidisciplinary with RIGs

No.	Word	RIG	No.	Word	RIG
1	properti	3.5×10^{-2}	51	data	1.1×10^{-2}
2	electron	3.4×10^{-2}	52	outcom	1.1×10^{-2}
3	alloy	3.3×10^{-2}	53	morpholog	1.1×10^{-2}
4	materi	3.3×10^{-2}	54	crystallin	1.1×10^{-2}
5	temperatur	3.3×10^{-2}	55	powder	1.1×10^{-2}
6	patient	3.2×10^{-2}	56	energi	1.1×10^{-2}
7	conclus	3.1×10^{-2}	57	electrod	1×10^{-2}
8	microstructur	2.9×10^{-2}	58	crack	1×10^{-2}
9	film	2.8×10^{-2}	59	popul	1×10^{-2}
10	surfac	2.7×10^{-2}	60	identifi	1×10^{-2}
11	diffract	2.4×10^{-2}	61	graphen	1×10^{-2}
12	microscopi	2.3×10^{-2}	62	degre	1×10^{-2}
13	fabric	2.2×10^{-2}	63	object	1×10^{-2}
14	ray	2.1×10^{-2}	64	particip	1×10^{-2}
15	prepar	2.1×10^{-2}	65	mechan	1×10^{-2}
16	structur	2.1×10^{-2}	66	level	9.9×10^{-3}
17	layer	1.9×10^{-2}	67	thick	9.9×10^{-3}
18	nanoparticl	1.9×10^{-2}	68	phase	9.8×10^{-3}
19	associ	1.9×10^{-2}	69	atom	9.8×10^{-3}
20	metal	1.8×10^{-2}	70	electr	9.7×10^{-3}
21	composit	1.7×10^{-2}	71	particl	9.4×10^{-3}
22	clinic	1.7×10^{-2}	72	protein	9.3×10^{-3}
23	background	1.6×10^{-2}	73	carbon	9.2×10^{-3}
24	steel	1.6×10^{-2}	74	polym	9×10^{-3}
25	diseas	1.6×10^{-2}	75	age	8.9×10^{-3}
26	year	1.6×10^{-2}	76	deform	8.9×10^{-3}
27	dope	1.6×10^{-2}	77	exhibit	8.7×10^{-3}
28	xrd	1.5×10^{-2}	78	whether	8.6×10^{-3}
29	spectroscopi	1.5×10^{-2}	79	ion	8.6×10^{-3}
30	scan	1.4×10^{-2}	80	human	8.6×10^{-3}
31	crystal	1.4×10^{-2}	81	group	8.6×10^{-3}
32	thin	1.4×10^{-2}	82	glass	8.5×10^{-3}
33	deposit	1.4×10^{-2}	83	aim	8.4×10^{-3}
34	tensil	1.3×10^{-2}	84	includ	8.3×10^{-3}
35	thermal	1.3×10^{-2}	85	densiti	8.3×10^{-3}
36	coat	1.3×10^{-2}	86	charg	8.3×10^{-3}
37	grain	1.3×10^{-2}	87	zno	8.2×10^{-3}
38	synthes	1.3×10^{-2}	88	chemic	8.2×10^{-3}
39	sem	1.3×10^{-2}	89	tem	8.2×10^{-3}
40	oxid	1.2×10^{-2}	90	nanotub	8.1×10^{-3}
41	express	1.2×10^{-2}	91	photoluminesc	8×10^{-3}
42	anneal	1.2×10^{-2}	92	size	8×10^{-3}
43	assess	1.2×10^{-2}	93	room	7.9×10^{-3}
44	may	1.2×10^{-2}	94	month	7.9×10^{-3}
45	nanostructur	1.2×10^{-2}	95	health	7.9×10^{-3}
46	strength	1.2×10^{-2}	96	optic	7.6×10^{-3}
47	electrochem	1.2×10^{-2}	97	interfac	7.6×10^{-3}
48	risk	1.2×10^{-2}	98	sinter	7.6×10^{-3}
49	gene	1.2×10^{-2}	99	suggest	7.6×10^{-3}
50	substrat	1.2×10^{-2}	100	score	7.6×10^{-3}

TABLE D.144. The list of the top 100 words in the category Materials Science, Paper and Wood with RIGs

No.	Word	RIG
1	wood	1.7×10^{-1}
2	cellulos	1.2×10^{-1}
3	pulp	7.7×10^{-2}
4	lignin	5.9×10^{-2}
5	strength	3.5×10^{-2}
6	fiber	3.5×10^{-2}
7	timber	3.4×10^{-2}
8	properti	3.3×10^{-2}
9	hemicellulos	2.5×10^{-2}
10	pine	2.4×10^{-2}
11	content	2.4×10^{-2}
12	modulus	2.3×10^{-2}
13	tensil	2.2×10^{-2}
14	spruce	2.1×10^{-2}
15	degre	2×10^{-2}
16	chemic	1.9×10^{-2}
17	dri	1.9×10^{-2}
18	moistur	1.8×10^{-2}
19	pinus	1.6×10^{-2}
20	patient	1.6×10^{-2}
21	mill	1.6×10^{-2}
22	product	1.6×10^{-2}
23	materi	1.5×10^{-2}
24	water	1.5×10^{-2}
25	lignocellulos	1.5×10^{-2}
26	fourier	1.4×10^{-2}
27	composit	1.3×10^{-2}
28	hydrolysi	1.3×10^{-2}
29	conclus	1.3×10^{-2}
30	fibr	1.3×10^{-2}
31	swell	1.3×10^{-2}
32	bond	1.2×10^{-2}
33	infrar	1.2×10^{-2}
34	pretreat	1.2×10^{-2}
35	ftir	1.1×10^{-2}
36	scan	1.1×10^{-2}
37	crystallin	1.1×10^{-2}
38	temperatur	1.1×10^{-2}
39	prepar	1.1×10^{-2}
40	naoh	1.1×10^{-2}
41	board	1.1×10^{-2}
42	biomass	1.1×10^{-2}
43	surfac	1×10^{-2}
44	cotton	1×10^{-2}
45	microscopi	9.9×10^{-3}
46	bend	9.6×10^{-3}
47	associ	9.1×10^{-3}
48	spectroscopi	9.1×10^{-3}
49	raw	8.9×10^{-3}
50	background	8.8×10^{-3}
51	alkali	8.7×10^{-3}
52	clinic	8.7×10^{-3}
53	sem	8.5×10^{-3}
54	lamin	8.4×10^{-3}
55	yield	8.4×10^{-3}
56	enzymat	8.3×10^{-3}
57	mechan	8.2×10^{-3}
58	elast	8.2×10^{-3}
59	industri	8×10^{-3}
60	phenol	7.8×10^{-3}
61	impregn	7.8×10^{-3}
62	thermogravimetr	7.3×10^{-3}
63	diseas	7.2×10^{-3}
64	oak	7×10^{-3}
65	shear	6.8×10^{-3}
66	straw	6.8×10^{-3}
67	mpa	6.8×10^{-3}
68	acid	6.8×10^{-3}
69	bark	6.7×10^{-3}
70	decreas	6.6×10^{-3}
71	sugar	6.6×10^{-3}
72	resin	6.6×10^{-3}
73	remov	6.5×10^{-3}
74	alkalin	6.5×10^{-3}
75	thermal	6.4×10^{-3}
76	acet	6.4×10^{-3}
77	adhes	6.3×10^{-3}
78	adsorpt	6.3×10^{-3}
79	wall	6.3×10^{-3}
80	propos	6.2×10^{-3}
81	solvent	6.2×10^{-3}
82	sodium	6.2×10^{-3}
83	amount	5.9×10^{-3}
84	specimen	5.9×10^{-3}
85	sampl	5.9×10^{-3}
86	viscos	5.9×10^{-3}
87	urea	5.8×10^{-3}
88	human	5.8×10^{-3}
89	hydrophob	5.8×10^{-3}
90	signal	5.7×10^{-3}
91	valu	5.7×10^{-3}
92	dissolv	5.7×10^{-3}
93	compress	5.7×10^{-3}
94	treat	5.6×10^{-3}
95	ethanol	5.6×10^{-3}
96	risk	5.6×10^{-3}
97	manufactur	5.5×10^{-3}
98	aqueous	5.4×10^{-3}
99	ash	5.4×10^{-3}
100	press	5.4×10^{-3}

TABLE D.145. The list of the top 100 words in the category Materials Science, Textiles with RIGs

No.	Word	RIG	No.	Word	RIG
1	yarn	7.4×10^{-2}	51	wood	6.9×10^{-3}
2	fabric	7.2×10^{-2}	52	break	6.6×10^{-3}
3	textil	6.9×10^{-2}	53	reinforc	6.5×10^{-3}
4	cotton	6×10^{-2}	54	mechan	6.5×10^{-3}
5	cellulos	5.8×10^{-2}	55	express	6.4×10^{-3}
6	properti	5.2×10^{-2}	56	moistur	6.4×10^{-3}
7	fiber	5.1×10^{-2}	57	chitosan	6.3×10^{-3}
8	dye	3.5×10^{-2}	58	antibacteri	6.3×10^{-3}
9	woven	3×10^{-2}	59	solvent	6.3×10^{-3}
10	tensil	2.8×10^{-2}	60	gene	6.2×10^{-3}
11	fibr	2.2×10^{-2}	61	process	6.1×10^{-3}
12	strength	2×10^{-2}	62	colour	5.8×10^{-3}
13	conclus	1.5×10^{-2}	63	poli	5.8×10^{-3}
14	patient	1.4×10^{-2}	64	aqueous	5.8×10^{-3}
15	scan	1.4×10^{-2}	65	temperatur	5.7×10^{-3}
16	prepar	1.4×10^{-2}	66	water	5.6×10^{-3}
17	modulus	1.2×10^{-2}	67	permeabl	5.6×10^{-3}
18	associ	1.2×10^{-2}	68	identifi	5.6×10^{-3}
19	electron	1.2×10^{-2}	69	risk	5.6×10^{-3}
20	absorpt	1.2×10^{-2}	70	degre	5.5×10^{-3}
21	materi	1.1×10^{-2}	71	manufactur	5.1×10^{-3}
22	color	1.1×10^{-2}	72	suggest	5.1×10^{-3}
23	surfac	1×10^{-2}	73	produc	5×10^{-3}
24	infrar	1×10^{-2}	74	coat	5×10^{-3}
25	sem	1×10^{-2}	75	outcom	5×10^{-3}
26	composit	1×10^{-2}	76	polymer	4.9×10^{-3}
27	microscopi	1×10^{-2}	77	dri	4.9×10^{-3}
28	thermal	9.9×10^{-3}	78	popul	4.7×10^{-3}
29	nanofib	9.9×10^{-3}	79	blue	4.7×10^{-3}
30	spectroscopi	9.5×10^{-3}	80	age	4.7×10^{-3}
31	fourier	9.4×10^{-3}	81	hemicellulos	4.6×10^{-3}
32	electrospin	9.3×10^{-3}	82	spin	4.6×10^{-3}
33	polym	9.2×10^{-3}	83	hydrophil	4.6×10^{-3}
34	crystallin	9×10^{-3}	84	thermogravimetr	4.5×10^{-3}
35	comfort	8.9×10^{-3}	85	control	4.4×10^{-3}
36	polypropylen	8.7×10^{-3}	86	ray	4.4×10^{-3}
37	synthes	8.4×10^{-3}	87	character	4.4×10^{-3}
38	background	8.4×10^{-3}	88	solut	4.4×10^{-3}
39	blend	8.3×10^{-3}	89	fluoresc	4.3×10^{-3}
40	elong	8×10^{-3}	90	data	4.3×10^{-3}
41	chemic	8×10^{-3}	91	receiv	4.2×10^{-3}
42	clinic	7.8×10^{-3}	92	microscop	4.1×10^{-3}
43	diseas	7.7×10^{-3}	93	case	4.1×10^{-3}
44	morpholog	7.5×10^{-3}	94	wet	4×10^{-3}
45	viscos	7.3×10^{-3}	95	naoh	4×10^{-3}
46	ftir	7.2×10^{-3}	96	dispers	4×10^{-3}
47	structur	7.2×10^{-3}	97	whether	4×10^{-3}
48	pulp	7×10^{-3}	98	crosslink	3.9×10^{-3}
49	year	7×10^{-3}	99	alkali	3.9×10^{-3}
50	may	6.9×10^{-3}	100	made	3.9×10^{-3}

Table D.146. The list of the top 100 words in the category Mathematical and Computational Biology with RIGs

No.	Word	RIG
1	model	2.9×10^{-2}
2	comput	2.7×10^{-2}
3	dataset	2.7×10^{-2}
4	genom	2.6×10^{-2}
5	biolog	2.5×10^{-2}
6	data	2.2×10^{-2}
7	algorithm	1.9×10^{-2}
8	sequenc	1.8×10^{-2}
9	propos	1.7×10^{-2}
10	approach	1.6×10^{-2}
11	motiv	1.6×10^{-2}
12	annot	1.4×10^{-2}
13	simul	1.4×10^{-2}
14	predict	1.3×10^{-2}
15	infer	1.3×10^{-2}
16	set	1.2×10^{-2}
17	gene	1.2×10^{-2}
18	can	1.2×10^{-2}
19	base	1.2×10^{-2}
20	inform	1.1×10^{-2}
21	mathemat	1.1×10^{-2}
22	featur	1×10^{-2}
23	tool	1×10^{-2}
24	bioinformat	1×10^{-2}
25	accuraci	1×10^{-2}
26	network	9.6×10^{-3}
27	statist	9.5×10^{-3}
28	databas	9.4×10^{-3}
29	avail	9.1×10^{-3}
30	temperatur	8.8×10^{-3}
31	bayesian	8.4×10^{-3}
32	estim	8.1×10^{-3}
33	classif	8.1×10^{-3}
34	appli	7.9×10^{-3}
35	dynam	7.7×10^{-3}
36	framework	7.2×10^{-3}
37	use	6.9×10^{-3}
38	method	6.9×10^{-3}
39	ecg	6.8×10^{-3}
40	evolutionari	6.4×10^{-3}
41	exist	6.4×10^{-3}
42	number	6.4×10^{-3}
43	mani	6.3×10^{-3}
44	accur	6.2×10^{-3}
45	protein	6.2×10^{-3}
46	interact	6.2×10^{-3}
47	allow	6.2×10^{-3}
48	provid	6.2×10^{-3}
49	materi	6.1×10^{-3}
50	classifi	6.1×10^{-3}
51	novel	6×10^{-3}
52	analysi	5.9×10^{-3}
53	complex	5.9×10^{-3}
54	robust	5.8×10^{-3}
55	illustr	5.7×10^{-3}
56	throughput	5.7×10^{-3}
57	genet	5.5×10^{-3}
58	valid	5.5×10^{-3}
59	outperform	5.5×10^{-3}
60	water	5.4×10^{-3}
61	seq	5.3×10^{-3}
62	discoveri	5.2×10^{-3}
63	automat	5.2×10^{-3}
64	regulatori	5.1×10^{-3}
65	machin	5.1×10^{-3}
66	softwar	5×10^{-3}
67	often	4.9×10^{-3}
68	electron	4.9×10^{-3}
69	align	4.8×10^{-3}
70	multipl	4.8×10^{-3}
71	experiment	4.7×10^{-3}
72	ontolog	4.6×10^{-3}
73	likelihood	4.6×10^{-3}
74	graph	4.5×10^{-3}
75	identif	4.5×10^{-3}
76	markov	4.5×10^{-3}
77	oxid	4.5×10^{-3}
78	biomed	4.5×10^{-3}
79	error	4.4×10^{-3}
80	identifi	4.4×10^{-3}
81	diseas	4.4×10^{-3}
82	web	4.4×10^{-3}
83	prepar	4.3×10^{-3}
84	covari	4.3×10^{-3}
85	call	4.3×10^{-3}
86	metal	4.3×10^{-3}
87	problem	4.3×10^{-3}
88	larg	4.2×10^{-3}
89	challeng	4.1×10^{-3}
90	svm	4.1×10^{-3}
91	function	4.1×10^{-3}
92	thermal	4.1×10^{-3}
93	manual	3.9×10^{-3}
94	cluster	3.9×10^{-3}
95	specif	3.9×10^{-3}
96	graphic	3.8×10^{-3}
97	surfac	3.8×10^{-3}
98	real	3.8×10^{-3}
99	microarray	3.8×10^{-3}
100	nucleotid	3.8×10^{-3}

Table D.147. The list of the top 100 words in the category Mathematics with RIGs

No.	Word	RIG	No.	Word	RIG
1	prove	1×10^{-1}	51	design	1.8×10^{-2}
2	theorem	7.4×10^{-2}	52	mechan	1.8×10^{-2}
3	let	7×10^{-2}	53	respons	1.8×10^{-2}
4	algebra	6.5×10^{-2}	54	exist	1.7×10^{-2}
5	space	4.9×10^{-2}	55	bar	1.7×10^{-2}
6	effect	4.2×10^{-2}	56	year	1.7×10^{-2}
7	high	4.1×10^{-2}	57	treatment	1.7×10^{-2}
8	signific	4.1×10^{-2}	58	satisfi	1.7×10^{-2}
9	increas	4×10^{-2}	59	control	1.7×10^{-2}
10	use	3.7×10^{-2}	60	chang	1.6×10^{-2}
11	bound	3.5×10^{-2}	61	may	1.6×10^{-2}
12	finit	3.4×10^{-2}	62	demonstr	1.6×10^{-2}
13	infin	3.4×10^{-2}	63	asymptot	1.6×10^{-2}
14	equat	3.4×10^{-2}	64	invari	1.6×10^{-2}
15	perform	3.3×10^{-2}	65	experiment	1.5×10^{-2}
16	compar	3.2×10^{-2}	66	model	1.5×10^{-2}
17	give	3.1×10^{-2}	67	method	1.5×10^{-2}
18	conjectur	3×10^{-2}	68	hilbert	1.5×10^{-2}
19	activ	2.9×10^{-2}	69	sampl	1.5×10^{-2}
20	manifold	2.9×10^{-2}	70	studi	1.5×10^{-2}
21	class	2.7×10^{-2}	71	convex	1.5×10^{-2}
22	dure	2.5×10^{-2}	72	given	1.5×10^{-2}
23	banach	2.5×10^{-2}	73	solut	1.5×10^{-2}
24	conclus	2.5×10^{-2}	74	singular	1.5×10^{-2}
25	patient	2.4×10^{-2}	75	import	1.5×10^{-2}
26	polynomi	2.4×10^{-2}	76	vertic	1.5×10^{-2}
27	graph	2.4×10^{-2}	77	commut	1.5×10^{-2}
28	evalu	2.3×10^{-2}	78	decreas	1.5×10^{-2}
29	differ	2.3×10^{-2}	79	temperatur	1.5×10^{-2}
30	data	2.2×10^{-2}	80	reveal	1.5×10^{-2}
31	level	2.2×10^{-2}	81	clinic	1.4×10^{-2}
32	proof	2.2×10^{-2}	82	identifi	1.4×10^{-2}
33	general	2.2×10^{-2}	83	detect	1.4×10^{-2}
34	howev	2.2×10^{-2}	84	age	1.4×10^{-2}
35	observ	2.1×10^{-2}	85	everi	1.4×10^{-2}
36	suggest	2.1×10^{-2}	86	experi	1.4×10^{-2}
37	test	2.1×10^{-2}	87	materi	1.4×10^{-2}
38	develop	2.1×10^{-2}	88	paper	1.4×10^{-2}
39	found	2×10^{-2}	89	subset	1.3×10^{-2}
40	integ	2×10^{-2}	90	compact	1.3×10^{-2}
41	inequ	2×10^{-2}	91	problem	1.3×10^{-2}
42	indic	2×10^{-2}	92	background	1.3×10^{-2}
43	assess	2×10^{-2}	93	enhanc	1.3×10^{-2}
44	analysi	2×10^{-2}	94	specif	1.3×10^{-2}
45	report	1.9×10^{-2}	95	affect	1.3×10^{-2}
46	infinit	1.9×10^{-2}	96	higher	1.3×10^{-2}
47	low	1.9×10^{-2}	97	caus	1.3×10^{-2}
48	base	1.9×10^{-2}	98	research	1.3×10^{-2}
49	cell	1.9×10^{-2}	99	examin	1.3×10^{-2}
50	process	1.9×10^{-2}	100	protein	1.3×10^{-2}

Table D.148. The list of the top 100 words in the category Mathematics, Applied with RIGs

No.	Word	RIG	No.	Word	RIG
1	equat	7.7×10^{-2}	51	graph	1.3×10^{-2}
2	prove	5.3×10^{-2}	52	found	1.3×10^{-2}
3	problem	5.1×10^{-2}	53	howev	1.3×10^{-2}
4	solut	4.4×10^{-2}	54	observ	1.3×10^{-2}
5	theorem	3.9×10^{-2}	55	age	1.3×10^{-2}
6	converg	3.3×10^{-2}	56	studi	1.2×10^{-2}
7	numer	3.2×10^{-2}	57	protein	1.2×10^{-2}
8	paper	3.1×10^{-2}	58	background	1.2×10^{-2}
9	signific	3×10^{-2}	59	low	1.2×10^{-2}
10	space	3×10^{-2}	60	iter	1.2×10^{-2}
11	increas	2.8×10^{-2}	61	norm	1.1×10^{-2}
12	bound	2.8×10^{-2}	62	eigenvalu	1.1×10^{-2}
13	nonlinear	2.5×10^{-2}	63	ellipt	1.1×10^{-2}
14	finit	2.5×10^{-2}	64	approxim	1.1×10^{-2}
15	conclus	2.4×10^{-2}	65	use	1.1×10^{-2}
16	activ	2.3×10^{-2}	66	identifi	1.1×10^{-2}
17	high	2.3×10^{-2}	67	respons	1.1×10^{-2}
18	patient	2.3×10^{-2}	68	chang	1.1×10^{-2}
19	asymptot	2.2×10^{-2}	69	higher	1.1×10^{-2}
20	infin	2.1×10^{-2}	70	decreas	1.1×10^{-2}
21	polynomi	2×10^{-2}	71	may	1.1×10^{-2}
22	algebra	2×10^{-2}	72	consid	1×10^{-2}
23	exampl	1.9×10^{-2}	73	acid	1×10^{-2}
24	exist	1.8×10^{-2}	74	infinit	1×10^{-2}
25	inequ	1.8×10^{-2}	75	perform	1×10^{-2}
26	boundari	1.8×10^{-2}	76	reveal	1×10^{-2}
27	dure	1.8×10^{-2}	77	point	1×10^{-2}
28	class	1.8×10^{-2}	78	fix	1×10^{-2}
29	given	1.8×10^{-2}	79	examin	9.9×10^{-3}
30	general	1.7×10^{-2}	80	sampl	9.7×10^{-3}
31	effect	1.7×10^{-2}	81	satisfi	9.7×10^{-3}
32	solv	1.7×10^{-2}	82	regular	9.7×10^{-3}
33	suggest	1.7×10^{-2}	83	detect	9.7×10^{-3}
34	let	1.7×10^{-2}	84	illustr	9.6×10^{-3}
35	convex	1.6×10^{-2}	85	mechan	9.6×10^{-3}
36	assess	1.5×10^{-2}	86	human	9.5×10^{-3}
37	cell	1.5×10^{-2}	87	element	9.2×10^{-3}
38	level	1.5×10^{-2}	88	measur	9.2×10^{-3}
39	year	1.4×10^{-2}	89	temperatur	9.2×10^{-3}
40	report	1.4×10^{-2}	90	gene	9.1×10^{-3}
41	banach	1.4×10^{-2}	91	data	9.1×10^{-3}
42	linear	1.4×10^{-2}	92	differ	9×10^{-3}
43	discret	1.4×10^{-2}	93	potenti	9×10^{-3}
44	treatment	1.4×10^{-2}	94	day	9×10^{-3}
45	evalu	1.4×10^{-2}	95	diseas	8.9×10^{-3}
46	compar	1.4×10^{-2}	96	rang	8.9×10^{-3}
47	clinic	1.3×10^{-2}	97	differenti	8.9×10^{-3}
48	singular	1.3×10^{-2}	98	introduc	8.9×10^{-3}
49	give	1.3×10^{-2}	99	suffici	8.8×10^{-3}
50	indic	1.3×10^{-2}	100	affect	8.7×10^{-3}

Table D.149. The list of the top 100 words in the category Mathematics, Interdisciplinary Applications with RIGs

No.	Word	RIG	No.	Word	RIG
1	numer	4.4×10^{-2}	51	suggest	6.6×10^{-3}
2	problem	3.8×10^{-2}	52	scheme	6.6×10^{-3}
3	paper	3.1×10^{-2}	53	appli	6.5×10^{-3}
4	equat	2.9×10^{-2}	54	dimension	6.4×10^{-3}
5	propos	2.7×10^{-2}	55	acid	6.4×10^{-3}
6	exampl	2.5×10^{-2}	56	mesh	6.4×10^{-3}
7	model	2.5×10^{-2}	57	diseas	6.4×10^{-3}
8	solv	2.2×10^{-2}	58	obtain	6.3×10^{-3}
9	finit	2.1×10^{-2}	59	accuraci	6.2×10^{-3}
10	nonlinear	1.8×10^{-2}	60	price	6.2×10^{-3}
11	conclus	1.6×10^{-2}	61	introduc	6.1×10^{-3}
12	patient	1.6×10^{-2}	62	group	6×10^{-3}
13	discret	1.6×10^{-2}	63	base	6×10^{-3}
14	algorithm	1.5×10^{-2}	64	error	6×10^{-3}
15	illustr	1.5×10^{-2}	65	associ	5.8×10^{-3}
16	simul	1.5×10^{-2}	66	element	5.8×10^{-3}
17	solut	1.4×10^{-2}	67	estim	5.7×10^{-3}
18	comput	1.3×10^{-2}	68	function	5.5×10^{-3}
19	converg	1.3×10^{-2}	69	general	5.5×10^{-3}
20	stochast	1.3×10^{-2}	70	prepar	5.4×10^{-3}
21	asymptot	1.2×10^{-2}	71	gene	5.4×10^{-3}
22	activ	1.1×10^{-2}	72	studi	5.2×10^{-3}
23	signific	1.1×10^{-2}	73	iter	5.1×10^{-3}
24	linear	1.1×10^{-2}	74	high	5.1×10^{-3}
25	boundari	1.1×10^{-2}	75	day	5×10^{-3}
26	increas	1×10^{-2}	76	polynomi	4.7×10^{-3}
27	lyapunov	1×10^{-2}	77	explicit	4.7×10^{-3}
28	formul	9.8×10^{-3}	78	singular	4.6×10^{-3}
29	consid	9.6×10^{-3}	79	analyt	4.6×10^{-3}
30	approach	9.2×10^{-3}	80	exist	4.6×10^{-3}
31	dynam	9.2×10^{-3}	81	equilibrium	4.5×10^{-3}
32	optim	8.9×10^{-3}	82	variabl	4.5×10^{-3}
33	clinic	8.8×10^{-3}	83	theorem	4.5×10^{-3}
34	cell	8.3×10^{-3}	84	final	4.5×10^{-3}
35	background	8.2×10^{-3}	85	robust	4.4×10^{-3}
36	paramet	8.1×10^{-3}	86	condit	4.4×10^{-3}
37	age	8.1×10^{-3}	87	ray	4.4×10^{-3}
38	given	8×10^{-3}	88	bound	4.4×10^{-3}
39	theori	7.9×10^{-3}	89	found	4.4×10^{-3}
40	treatment	7.7×10^{-3}	90	higher	4.3×10^{-3}
41	approxim	7.7×10^{-3}	91	content	4.2×10^{-3}
42	report	7.7×10^{-3}	92	inhibit	4.2×10^{-3}
43	method	7.6×10^{-3}	93	construct	4.2×10^{-3}
44	mathemat	7.3×10^{-3}	94	space	4.2×10^{-3}
45	order	7.2×10^{-3}	95	oxid	4.2×10^{-3}
46	protein	7.2×10^{-3}	96	therapi	4.1×10^{-3}
47	dure	7.1×10^{-3}	97	set	4.1×10^{-3}
48	year	6.8×10^{-3}	98	uncertainti	4.1×10^{-3}
49	bifurc	6.7×10^{-3}	99	low	4.1×10^{-3}
50	deriv	6.6×10^{-3}	100	matrix	4.1×10^{-3}

Table D.150. The list of the top 100 words in the category Mechanics with RIGs

No.	Word	RIG
1	numer	4.5×10^{-2}
2	flow	3.1×10^{-2}
3	equat	2.8×10^{-2}
4	finit	2.8×10^{-2}
5	patient	2.3×10^{-2}
6	simul	2.2×10^{-2}
7	paper	2.1×10^{-2}
8	fluid	2×10^{-2}
9	veloc	2×10^{-2}
10	reynold	2×10^{-2}
11	conclus	1.9×10^{-2}
12	boundari	1.8×10^{-2}
13	heat	1.8×10^{-2}
14	elast	1.7×10^{-2}
15	turbul	1.6×10^{-2}
16	shear	1.5×10^{-2}
17	model	1.4×10^{-2}
18	activ	1.4×10^{-2}
19	clinic	1.4×10^{-2}
20	deform	1.4×10^{-2}
21	solv	1.4×10^{-2}
22	load	1.3×10^{-2}
23	group	1.3×10^{-2}
24	nonlinear	1.3×10^{-2}
25	element	1.3×10^{-2}
26	stress	1.2×10^{-2}
27	protein	1.2×10^{-2}
28	diseas	1.2×10^{-2}
29	crack	1.2×10^{-2}
30	vibrat	1.1×10^{-2}
31	suggest	1.1×10^{-2}
32	age	1.1×10^{-2}
33	paramet	1.1×10^{-2}
34	forc	1.1×10^{-2}
35	associ	1.1×10^{-2}
36	displac	1.1×10^{-2}
37	experiment	1×10^{-2}
38	background	1×10^{-2}
39	gene	1×10^{-2}
40	convect	9.9×10^{-3}
41	year	9.7×10^{-3}
42	cell	9.7×10^{-3}
43	problem	9.6×10^{-3}
44	wall	9.4×10^{-3}
45	treatment	8.9×10^{-3}
46	report	8.8×10^{-3}
47	plate	8.8×10^{-3}
48	dynam	8.8×10^{-3}
49	may	8.4×10^{-3}
50	dimension	8×10^{-3}

No.	Word	RIG
51	pressur	7.9×10^{-3}
52	signific	7.8×10^{-3}
53	vortex	7.8×10^{-3}
54	cylind	7.7×10^{-3}
55	popul	7.5×10^{-3}
56	risk	7.5×10^{-3}
57	motion	7.4×10^{-3}
58	particip	7.4×10^{-3}
59	vortic	7.3×10^{-3}
60	level	7.2×10^{-3}
61	friction	7.1×10^{-3}
62	cancer	7×10^{-3}
63	outcom	6.8×10^{-3}
64	speed	6.6×10^{-3}
65	viscous	6.6×10^{-3}
66	obtain	6.5×10^{-3}
67	navier	6.5×10^{-3}
68	speci	6.4×10^{-3}
69	solut	6.4×10^{-3}
70	human	6.3×10^{-3}
71	stoke	6.2×10^{-3}
72	month	6.1×10^{-3}
73	therapi	6×10^{-3}
74	formul	6×10^{-3}
75	day	5.9×10^{-3}
76	assess	5.9×10^{-3}
77	thermal	5.9×10^{-3}
78	steadi	5.9×10^{-3}
79	unsteadi	5.9×10^{-3}
80	among	5.9×10^{-3}
81	mediat	5.8×10^{-3}
82	acid	5.8×10^{-3}
83	health	5.7×10^{-3}
84	materi	5.7×10^{-3}
85	stiff	5.6×10^{-3}
86	identifi	5.6×10^{-3}
87	howev	5.6×10^{-3}
88	analyt	5.5×10^{-3}
89	steel	5.5×10^{-3}
90	laminar	5.5×10^{-3}
91	sampl	5.5×10^{-3}
92	engin	5.4×10^{-3}
93	evid	5.3×10^{-3}
94	potenti	5.3×10^{-3}
95	adult	5.3×10^{-3}
96	pathway	5.3×10^{-3}
97	law	5.2×10^{-3}
98	male	5.2×10^{-3}
99	tissu	5.1×10^{-3}
100	infect	5.1×10^{-3}

TABLE D.151. The list of the top 100 words in the category Medical Ethics with RIGs

No.	Word	RIG
1	ethic	2.3×10^{-1}
2	moral	6.9×10^{-2}
3	bioethic	6.5×10^{-2}
4	consent	6×10^{-2}
5	medic	5.5×10^{-2}
6	argu	5.2×10^{-2}
7	health	4.2×10^{-2}
8	autonomi	3.6×10^{-2}
9	physician	3.2×10^{-2}
10	research	3.1×10^{-2}
11	legal	3×10^{-2}
12	profession	3×10^{-2}
13	debat	2.8×10^{-2}
14	decis	2.7×10^{-2}
15	public	2.7×10^{-2}
16	care	2.6×10^{-2}
17	issu	2.6×10^{-2}
18	justifi	2.5×10^{-2}
19	argument	2.5×10^{-2}
20	concern	2.4×10^{-2}
21	oblig	2.1×10^{-2}
22	practic	2×10^{-2}
23	discuss	2×10^{-2}
24	justif	2×10^{-2}
25	doctor	1.9×10^{-2}
26	harm	1.9×10^{-2}
27	question	1.9×10^{-2}
28	polici	1.7×10^{-2}
29	right	1.7×10^{-2}
30	committe	1.7×10^{-2}
31	patient	1.6×10^{-2}
32	articl	1.6×10^{-2}
33	particip	1.6×10^{-2}
34	inform	1.5×10^{-2}
35	interview	1.5×10^{-2}
36	rais	1.5×10^{-2}
37	medicin	1.4×10^{-2}
38	view	1.4×10^{-2}
39	controversi	1.4×10^{-2}
40	clinic	1.4×10^{-2}
41	reason	1.4×10^{-2}
42	person	1.4×10^{-2}
43	make	1.3×10^{-2}
44	claim	1.3×10^{-2}
45	attitud	1.3×10^{-2}
46	philosoph	1.3×10^{-2}
47	healthcar	1.3×10^{-2}
48	institut	1.3×10^{-2}
49	dilemma	1.2×10^{-2}
50	scientif	1.1×10^{-2}

No.	Word	RIG
51	principl	1.1×10^{-2}
52	whether	1.1×10^{-2}
53	donat	1.1×10^{-2}
54	result	1.1×10^{-2}
55	benefit	1.1×10^{-2}
56	accept	1.1×10^{-2}
57	social	1×10^{-2}
58	need	1×10^{-2}
59	justic	1×10^{-2}
60	regard	1×10^{-2}
61	life	1×10^{-2}
62	ask	9.9×10^{-3}
63	parent	9.9×10^{-3}
64	disclosur	9.8×10^{-3}
65	model	9.8×10^{-3}
66	paramet	9.8×10^{-3}
67	defend	9.8×10^{-3}
68	guidelin	9.7×10^{-3}
69	way	9.6×10^{-3}
70	biomed	9.5×10^{-3}
71	countri	9.4×10^{-3}
72	law	9.3×10^{-3}
73	author	9.3×10^{-3}
74	measur	9.2×10^{-3}
75	temperatur	9.2×10^{-3}
76	show	9.1×10^{-3}
77	refus	9×10^{-3}
78	simul	8.9×10^{-3}
79	peopl	8.7×10^{-3}
80	duti	8.6×10^{-3}
81	individu	8.6×10^{-3}
82	will	8.5×10^{-3}
83	conflict	8.4×10^{-3}
84	undermin	8.4×10^{-3}
85	context	8.3×10^{-3}
86	surfac	8.2×10^{-3}
87	scholar	8.1×10^{-3}
88	ill	8.1×10^{-3}
89	case	8×10^{-3}
90	scienc	7.8×10^{-3}
91	whi	7.8×10^{-3}
92	religi	7.6×10^{-3}
93	interest	7.6×10^{-3}
94	compar	7.5×10^{-3}
95	energi	7.5×10^{-3}
96	request	7.5×10^{-3}
97	concept	7.4×10^{-3}
98	respond	7.4×10^{-3}
99	high	7.3×10^{-3}
100	risk	7.3×10^{-3}

Table D.152. The list of the top 100 words in the category Medical Informatics with RIGs

No.	Word	RIG	No.	Word	RIG
1	health	4.1×10^{-2}	51	cell	6.5×10^{-3}
2	medic	3.6×10^{-2}	52	servic	6.3×10^{-3}
3	inform	3.1×10^{-2}	53	person	6.3×10^{-3}
4	data	2.5×10^{-2}	54	communic	6.2×10^{-3}
5	care	2.3×10^{-2}	55	concentr	6.1×10^{-3}
6	clinic	2.3×10^{-2}	56	nurs	6×10^{-3}
7	patient	2.1×10^{-2}	57	background	6×10^{-3}
8	healthcar	2.1×10^{-2}	58	train	5.9×10^{-3}
9	method	2×10^{-2}	59	evalu	5.8×10^{-3}
10	object	1.9×10^{-2}	60	standard	5.8×10^{-3}
11	user	1.8×10^{-2}	61	ecg	5.8×10^{-3}
12	base	1.5×10^{-2}	62	manag	5.8×10^{-3}
13	record	1.4×10^{-2}	63	usabl	5.8×10^{-3}
14	propos	1.4×10^{-2}	64	improv	5.8×10^{-3}
15	technolog	1.3×10^{-2}	65	oxid	5.7×10^{-3}
16	use	1.3×10^{-2}	66	privaci	5.7×10^{-3}
17	decis	1.3×10^{-2}	67	medicin	5.7×10^{-3}
18	hospit	1.2×10^{-2}	68	practic	5.7×10^{-3}
19	support	1.1×10^{-2}	69	learn	5.6×10^{-3}
20	implement	1.1×10^{-2}	70	speci	5.6×10^{-3}
21	research	1.1×10^{-2}	71	manual	5.6×10^{-3}
22	conclus	1.1×10^{-2}	72	surfac	5.5×10^{-3}
23	paper	1.1×10^{-2}	73	profession	5.5×10^{-3}
24	tool	1.1×10^{-2}	74	avail	5.5×10^{-3}
25	physician	1×10^{-2}	75	biomed	5.4×10^{-3}
26	classif	1×10^{-2}	76	task	5.3×10^{-3}
27	provid	9.8×10^{-3}	77	help	5.2×10^{-3}
28	system	9.8×10^{-3}	78	doctor	5.2×10^{-3}
29	web	9.1×10^{-3}	79	can	5.1×10^{-3}
30	approach	8.9×10^{-3}	80	knowledg	5.1×10^{-3}
31	comput	8.5×10^{-3}	81	intervent	5.1×10^{-3}
32	dataset	8.3×10^{-3}	82	acid	5.1×10^{-3}
33	clinician	8.3×10^{-3}	83	valid	5.1×10^{-3}
34	algorithm	8.2×10^{-3}	84	featur	5×10^{-3}
35	temperatur	8.1×10^{-3}	85	error	5×10^{-3}
36	automat	8×10^{-3}	86	machin	5×10^{-3}
37	expert	7.8×10^{-3}	87	induc	4.9×10^{-3}
38	accuraci	7.6×10^{-3}	88	framework	4.9×10^{-3}
39	develop	7.4×10^{-3}	89	text	4.8×10^{-3}
40	classifi	7.4×10^{-3}	90	make	4.8×10^{-3}
41	trial	7.3×10^{-3}	91	statist	4.8×10^{-3}
42	set	7.3×10^{-3}	92	challeng	4.8×10^{-3}
43	autom	7.2×10^{-3}	93	adopt	4.8×10^{-3}
44	onlin	7.2×10^{-3}	94	random	4.6×10^{-3}
45	rational	7.2×10^{-3}	95	qualiti	4.5×10^{-3}
46	need	7.1×10^{-3}	96	carbon	4.4×10^{-3}
47	databas	7×10^{-3}	97	public	4.4×10^{-3}
48	internet	7×10^{-3}	98	websit	4.4×10^{-3}
49	access	6.8×10^{-3}	99	real	4.2×10^{-3}
50	water	6.7×10^{-3}	100	particip	4.2×10^{-3}

Table D.153. The list of the top 100 words in the category Medical Laboratory Technology with RIGs

No.	Word	RIG
1	conclus	1×10^{-1}
2	background	6.6×10^{-2}
3	patient	6×10^{-2}
4	method	4.6×10^{-2}
5	clinic	4.2×10^{-2}
6	serum	4.1×10^{-2}
7	assay	3.8×10^{-2}
8	diagnosi	3.3×10^{-2}
9	cytolog	3.3×10^{-2}
10	diagnost	3.2×10^{-2}
11	blood	3×10^{-2}
12	laboratori	2.8×10^{-2}
13	sampl	2.4×10^{-2}
14	result	2.3×10^{-2}
15	diseas	2.1×10^{-2}
16	immunoassay	2×10^{-2}
17	concentr	1.9×10^{-2}
18	plasma	1.9×10^{-2}
19	tumor	1.7×10^{-2}
20	biomark	1.7×10^{-2}
21	aspir	1.6×10^{-2}
22	object	1.6×10^{-2}
23	healthi	1.6×10^{-2}
24	paper	1.6×10^{-2}
25	carcinoma	1.6×10^{-2}
26	marker	1.5×10^{-2}
27	detect	1.5×10^{-2}
28	context	1.5×10^{-2}
29	specimen	1.4×10^{-2}
30	needl	1.4×10^{-2}
31	cell	1.3×10^{-2}
32	evalu	1.3×10^{-2}
33	routin	1.3×10^{-2}
34	signific	1.2×10^{-2}
35	pathologist	1.2×10^{-2}
36	stain	1.2×10^{-2}
37	neoplasm	1.2×10^{-2}
38	test	1.2×10^{-2}
39	diagnos	1.2×10^{-2}
40	chromatographi	1.2×10^{-2}
41	malign	1.2×10^{-2}
42	pcr	1.1×10^{-2}
43	immunohistochem	1.1×10^{-2}
44	correl	1.1×10^{-2}
45	elisa	1.1×10^{-2}
46	sensit	9.9×10^{-3}
47	biopsi	9.7×10^{-3}
48	urin	9.3×10^{-3}
49	smear	9.2×10^{-3}
50	structur	9.1×10^{-3}

No.	Word	RIG
51	patholog	8.9×10^{-3}
52	creatinin	8.8×10^{-3}
53	respect	8.3×10^{-3}
54	energi	8.2×10^{-3}
55	histolog	8.2×10^{-3}
56	posit	8.1×10^{-3}
57	cancer	7.8×10^{-3}
58	case	7.8×10^{-3}
59	refer	7.7×10^{-3}
60	aim	7.7×10^{-3}
61	adenocarcinoma	7.3×10^{-3}
62	model	7.2×10^{-3}
63	screen	7.2×10^{-3}
64	analyt	7.1×10^{-3}
65	renal	7×10^{-3}
66	spectrometri	6.9×10^{-3}
67	determin	6.9×10^{-3}
68	level	6.9×10^{-3}
69	protein	6.8×10^{-3}
70	benign	6.8×10^{-3}
71	quantif	6.7×10^{-3}
72	negat	6.7×10^{-3}
73	specif	6.6×10^{-3}
74	compar	6.6×10^{-3}
75	prognost	6.5×10^{-3}
76	cholesterol	6.5×10^{-3}
77	kit	6.5×10^{-3}
78	hemoglobin	6.5×10^{-3}
79	simul	6.4×10^{-3}
80	subject	6.4×10^{-3}
81	tandem	6.4×10^{-3}
82	medicin	6.4×10^{-3}
83	surfac	6.4×10^{-3}
84	tissu	6.3×10^{-3}
85	antibodi	6.3×10^{-3}
86	rare	6.2×10^{-3}
87	assess	6.2×10^{-3}
88	valu	6.1×10^{-3}
89	autom	6.1×10^{-3}
90	diabet	5.9×10^{-3}
91	studi	5.9×10^{-3}
92	lesion	5.9×10^{-3}
93	total	5.9×10^{-3}
94	genotyp	5.9×10^{-3}
95	kidney	5.8×10^{-3}
96	measur	5.7×10^{-3}
97	immunohistochemistri	5.7×10^{-3}
98	may	5.7×10^{-3}
99	urinari	5.6×10^{-3}
100	immunosorb	5.6×10^{-3}

TABLE D.154. The list of the top 100 words in the category Medicine, General and Internal with RIGs

No.	Word	RIG	No.	Word	RIG
1	conclus	1.4×10^{-1}	51	popul	1.3×10^{-2}
2	patient	1.3×10^{-1}	52	acut	1.3×10^{-2}
3	object	5.9×10^{-2}	53	questionnair	1.3×10^{-2}
4	background	5.3×10^{-2}	54	review	1.3×10^{-2}
5	year	5×10^{-2}	55	blood	1.2×10^{-2}
6	hospit	5×10^{-2}	56	simul	1.2×10^{-2}
7	clinic	4.9×10^{-2}	57	old	1.2×10^{-2}
8	care	4.8×10^{-2}	58	male	1.2×10^{-2}
9	medic	4.6×10^{-2}	59	depart	1.2×10^{-2}
10	age	4.4×10^{-2}	60	random	1.2×10^{-2}
11	health	4.3×10^{-2}	61	adult	1.2×10^{-2}
12	outcom	4.1×10^{-2}	62	temperatur	1.2×10^{-2}
13	method	3.9×10^{-2}	63	decemb	1.2×10^{-2}
14	diseas	3.3×10^{-2}	64	introduct	1.2×10^{-2}
15	risk	3.3×10^{-2}	65	enrol	1.1×10^{-2}
16	particip	3×10^{-2}	66	report	1.1×10^{-2}
17	paper	2.6×10^{-2}	67	confid	1.1×10^{-2}
18	result	2.5×10^{-2}	68	admiss	1.1×10^{-2}
19	primari	2.4×10^{-2}	69	properti	1.1×10^{-2}
20	associ	2.4×10^{-2}	70	healthcar	1.1×10^{-2}
21	treatment	2.4×10^{-2}	71	cardiovascular	1.1×10^{-2}
22	group	2.3×10^{-2}	72	femal	1.1×10^{-2}
23	diagnosi	2.2×10^{-2}	73	adjust	1.1×10^{-2}
24	month	2.2×10^{-2}	74	regress	1.1×10^{-2}
25	physician	2.2×10^{-2}	75	structur	1.1×10^{-2}
26	aim	2×10^{-2}	76	week	1.1×10^{-2}
27	intervent	2×10^{-2}	77	day	1.1×10^{-2}
28	includ	2×10^{-2}	78	nation	1×10^{-2}
29	assess	1.9×10^{-2}	79	admit	1×10^{-2}
30	diagnos	1.9×10^{-2}	80	prevent	1×10^{-2}
31	studi	1.9×10^{-2}	81	odd	1×10^{-2}
32	symptom	1.8×10^{-2}	82	incid	1×10^{-2}
33	receiv	1.8×10^{-2}	83	death	1×10^{-2}
34	retrospect	1.7×10^{-2}	84	treat	1×10^{-2}
35	women	1.7×10^{-2}	85	advers	9.7×10^{-3}
36	propos	1.7×10^{-2}	86	case	9.6×10^{-3}
37	mortal	1.7×10^{-2}	87	medicin	9.5×10^{-3}
38	among	1.7×10^{-2}	88	section	9.5×10^{-3}
39	signific	1.7×10^{-2}	89	surgeri	9.3×10^{-3}
40	therapi	1.7×10^{-2}	90	prospect	9.3×10^{-3}
41	chronic	1.6×10^{-2}	91	common	9.3×10^{-3}
42	score	1.6×10^{-2}	92	complic	9.3×10^{-3}
43	diabet	1.6×10^{-2}	93	men	9.2×10^{-3}
44	preval	1.6×10^{-2}	94	hypertens	9.1×10^{-3}
45	pain	1.5×10^{-2}	95	interv	9.1×10^{-3}
46	trial	1.5×10^{-2}	96	baselin	9.1×10^{-3}
47	total	1.5×10^{-2}	97	older	9×10^{-3}
48	januari	1.4×10^{-2}	98	mean	9×10^{-3}
49	cohort	1.4×10^{-2}	99	demograph	9×10^{-3}
50	follow	1.4×10^{-2}	100	heart	8.8×10^{-3}

Table D.155. The list of the top 100 words in the category Medicine, Legal with RIGs

No.	Word	RIG	No.	Word	RIG
1	forens	2.5×10^{-1}	51	human	6.8×10^{-3}
2	autopsi	8.6×10^{-2}	52	function	6.8×10^{-3}
3	postmortem	6.2×10^{-2}	53	year	6.3×10^{-3}
4	death	4.4×10^{-2}	54	murder	6.2×10^{-3}
5	case	3.6×10^{-2}	55	assault	6.2×10^{-3}
6	toxicolog	2.6×10^{-2}	56	hair	6.2×10^{-3}
7	victim	2.6×10^{-2}	57	medicin	6.1×10^{-3}
8	legal	2.5×10^{-2}	58	exposur	5.9×10^{-3}
9	crime	2.3×10^{-2}	59	popul	5.9×10^{-3}
10	sampl	2.2×10^{-2}	60	weapon	5.8×10^{-3}
11	crimin	2×10^{-2}	61	autosom	5.8×10^{-3}
12	suicid	1.8×10^{-2}	62	repeat	5.8×10^{-3}
13	dna	1.8×10^{-2}	63	structur	5.8×10^{-3}
14	fatal	1.7×10^{-2}	64	skull	5.7×10^{-3}
15	male	1.5×10^{-2}	65	trauma	5.7×10^{-3}
16	suspect	1.5×10^{-2}	66	properti	5.6×10^{-3}
17	identif	1.4×10^{-2}	67	age	5.6×10^{-3}
18	injuri	1.3×10^{-2}	68	multiplex	5.5×10^{-3}
19	substanc	1.3×10^{-2}	69	spectrometri	5.5×10^{-3}
20	blood	1.2×10^{-2}	70	lethal	5.5×10^{-3}
21	scene	1.2×10^{-2}	71	blunt	5.4×10^{-3}
22	abus	1.2×10^{-2}	72	bone	5.4×10^{-3}
23	polic	1.2×10^{-2}	73	perpetr	5.4×10^{-3}
24	tandem	1.1×10^{-2}	74	cocain	5.3×10^{-3}
25	femal	1.1×10^{-2}	75	interpret	5.3×10^{-3}
26	anthropolog	1.1×10^{-2}	76	propos	5.3×10^{-3}
27	pathologist	1.1×10^{-2}	77	accid	5.2×10^{-3}
28	kit	1.1×10^{-2}	78	illicit	5.1×10^{-3}
29	skelet	9.8×10^{-3}	79	dynam	5.1×10^{-3}
30	court	9.5×10^{-3}	80	chromatographi	5.1×10^{-3}
31	drug	9.3×10^{-3}	81	paper	5×10^{-3}
32	allel	9.3×10^{-3}	82	alleg	5×10^{-3}
33	discrimin	9×10^{-3}	83	laboratori	5×10^{-3}
34	deceas	8.6×10^{-3}	84	amplif	5×10^{-3}
35	loci	8.4×10^{-3}	85	traumat	5×10^{-3}
36	urin	8.4×10^{-3}	86	rare	5×10^{-3}
37	die	8.3×10^{-3}	87	marker	4.8×10^{-3}
38	old	8.1×10^{-3}	88	estim	4.8×10^{-3}
39	bodi	8.1×10^{-3}	89	pcr	4.7×10^{-3}
40	individu	8×10^{-3}	90	man	4.7×10^{-3}
41	report	7.9×10^{-3}	91	interact	4.6×10^{-3}
42	sex	7.8×10^{-3}	92	energi	4.5×10^{-3}
43	profil	7.6×10^{-3}	93	model	4.5×10^{-3}
44	cadav	7.5×10^{-3}	94	inhal	4.3×10^{-3}
45	detect	7.4×10^{-3}	95	assess	4.3×10^{-3}
46	medic	7.3×10^{-3}	96	genet	4.2×10^{-3}
47	examin	7.3×10^{-3}	97	sexual	4.2×10^{-3}
48	evid	7.2×10^{-3}	98	stain	4.1×10^{-3}
49	toxic	7.1×10^{-3}	99	ingest	4.1×10^{-3}
50	caus	6.8×10^{-3}	100	activ	4.1×10^{-3}

Table D.156. The list of the top 100 words in the category Medicine, Research and Experimental with RIGs

No.	Word	RIG
1	cell	5.3×10^{-2}
2	express	4×10^{-2}
3	patient	3.9×10^{-2}
4	diseas	3.1×10^{-2}
5	treatment	2.8×10^{-2}
6	clinic	2.7×10^{-2}
7	paper	2.7×10^{-2}
8	protein	2.5×10^{-2}
9	conclus	2.5×10^{-2}
10	induc	2.3×10^{-2}
11	mice	2.3×10^{-2}
12	therapeut	2.2×10^{-2}
13	therapi	2.1×10^{-2}
14	studi	2.1×10^{-2}
15	tissu	2×10^{-2}
16	tumor	2×10^{-2}
17	gene	1.9×10^{-2}
18	signific	1.9×10^{-2}
19	immun	1.8×10^{-2}
20	cancer	1.8×10^{-2}
21	inhibit	1.8×10^{-2}
22	human	1.7×10^{-2}
23	vaccin	1.7×10^{-2}
24	blood	1.7×10^{-2}
25	rat	1.5×10^{-2}
26	prolifer	1.5×10^{-2}
27	vitro	1.5×10^{-2}
28	receptor	1.5×10^{-2}
29	associ	1.4×10^{-2}
30	treat	1.4×10^{-2}
31	vivo	1.3×10^{-2}
32	level	1.3×10^{-2}
33	aim	1.3×10^{-2}
34	may	1.3×10^{-2}
35	group	1.2×10^{-2}
36	apoptosi	1.2×10^{-2}
37	blot	1.2×10^{-2}
38	serum	1.2×10^{-2}
39	antibodi	1.2×10^{-2}
40	dose	1.2×10^{-2}
41	efficaci	1.2×10^{-2}
42	assay	1.2×10^{-2}
43	activ	1.1×10^{-2}
44	regul	1.1×10^{-2}
45	propos	1×10^{-2}
46	inflammatori	1×10^{-2}
47	antigen	1×10^{-2}
48	pathway	1×10^{-2}
49	mediat	9.9×10^{-3}
50	cytokin	9.8×10^{-3}
51	trial	9.7×10^{-3}
52	stem	9.7×10^{-3}
53	background	9.5×10^{-3}
54	mrna	9.4×10^{-3}
55	temperatur	9.3×10^{-3}
56	target	9.2×10^{-3}
57	mous	9.1×10^{-3}
58	upregul	9×10^{-3}
59	week	8.9×10^{-3}
60	chronic	8.8×10^{-3}
61	mesenchym	8.8×10^{-3}
62	endotheli	8.7×10^{-3}
63	drug	8.4×10^{-3}
64	downregul	8.3×10^{-3}
65	marker	8.1×10^{-3}
66	liver	7.9×10^{-3}
67	injuri	7.9×10^{-3}
68	inflamm	7.9×10^{-3}
69	bone	7.8×10^{-3}
70	administr	7.8×10^{-3}
71	diabet	7.7×10^{-3}
72	control	7.7×10^{-3}
73	inhibitor	7.7×10^{-3}
74	factor	7.7×10^{-3}
75	increas	7.7×10^{-3}
76	anim	7.6×10^{-3}
77	simul	7.4×10^{-3}
78	energi	7.3×10^{-3}
79	progress	7.1×10^{-3}
80	vascular	7×10^{-3}
81	polymeras	7×10^{-3}
82	surviv	6.9×10^{-3}
83	interleukin	6.8×10^{-3}
84	inject	6.8×10^{-3}
85	stimul	6.7×10^{-3}
86	infect	6.7×10^{-3}
87	normal	6.7×10^{-3}
88	kinas	6.7×10^{-3}
89	transplant	6.7×10^{-3}
90	beta	6.7×10^{-3}
91	carcinoma	6.6×10^{-3}
92	pcr	6.6×10^{-3}
93	day	6.5×10^{-3}
94	prevent	6.5×10^{-3}
95	cellular	6.4×10^{-3}
96	marrow	6.4×10^{-3}
97	follow	6.3×10^{-3}
98	pathogenesi	6.3×10^{-3}
99	effect	6.3×10^{-3}
100	stain	6.3×10^{-3}

TABLE D.157. The list of the top 100 words in the category Medieval and Renaissance Studies with RIGs

No.	Word	RIG	No.	Word	RIG
1	mediev	1.5×10^{-1}	51	chronicl	1.7×10^{-2}
2	centuri	1.5×10^{-1}	52	data	1.7×10^{-2}
3	articl	7.8×10^{-2}	53	tale	1.7×10^{-2}
4	text	7.5×10^{-2}	54	author	1.6×10^{-2}
5	argu	6.4×10^{-2}	55	base	1.6×10^{-2}
6	essay	6.4×10^{-2}	56	england	1.6×10^{-2}
7	result	4.9×10^{-2}	57	court	1.6×10^{-2}
8	thirteenth	4.5×10^{-2}	58	christ	1.5×10^{-2}
9	christian	4.5×10^{-2}	59	jew	1.5×10^{-2}
10	histor	3.7×10^{-2}	60	high	1.5×10^{-2}
11	polit	3.6×10^{-2}	61	measur	1.5×10^{-2}
12	read	3.5×10^{-2}	62	anglo	1.4×10^{-2}
13	english	3.4×10^{-2}	63	holi	1.4×10^{-2}
14	scholar	3.3×10^{-2}	64	cultur	1.4×10^{-2}
15	church	3.3×10^{-2}	65	tradit	1.4×10^{-2}
16	renaiss	3.2×10^{-2}	66	cell	1.3×10^{-2}
17	narrat	3.2×10^{-2}	67	system	1.3×10^{-2}
18	manuscript	3.2×10^{-2}	68	rate	1.3×10^{-2}
19	fifteenth	3.1×10^{-2}	69	artist	1.3×10^{-2}
20	religi	3.1×10^{-2}	70	discours	1.3×10^{-2}
21	write	3.1×10^{-2}	71	improv	1.3×10^{-2}
22	earli	2.7×10^{-2}	72	obtain	1.2×10^{-2}
23	poem	2.7×10^{-2}	73	henri	1.2×10^{-2}
24	literari	2.7×10^{-2}	74	poet	1.2×10^{-2}
25	roman	2.6×10^{-2}	75	perform	1.2×10^{-2}
26	king	2.6×10^{-2}	76	romanc	1.2×10^{-2}
27	john	2.6×10^{-2}	77	god	1.2×10^{-2}
28	contemporari	2.5×10^{-2}	78	test	1.2×10^{-2}
29	sixteenth	2.5×10^{-2}	79	compar	1.2×10^{-2}
30	scholarship	2.5×10^{-2}	80	audienc	1.2×10^{-2}
31	modern	2.4×10^{-2}	81	patient	1.2×10^{-2}
32	chaucer	2.4×10^{-2}	82	model	1.2×10^{-2}
33	late	2.3×10^{-2}	83	depict	1.2×10^{-2}
34	figur	2.3×10^{-2}	84	middl	1.2×10^{-2}
35	seventeenth	2.3×10^{-2}	85	intellectu	1.2×10^{-2}
36	use	2.3×10^{-2}	86	poetri	1.2×10^{-2}
37	latin	2.2×10^{-2}	87	writer	1.1×10^{-2}
38	rhetor	2.2×10^{-2}	88	low	1.1×10^{-2}
39	thoma	2.2×10^{-2}	89	reduc	1.1×10^{-2}
40	method	2.1×10^{-2}	90	way	1.1×10^{-2}
41	theolog	2.1×10^{-2}	91	studi	1.1×10^{-2}
42	histori	2.1×10^{-2}	92	effici	1.1×10^{-2}
43	written	2.1×10^{-2}	93	royal	1.1×10^{-2}
44	book	2×10^{-2}	94	decreas	1.1×10^{-2}
45	saint	2×10^{-2}	95	william	1.1×10^{-2}
46	historian	2×10^{-2}	96	philosoph	1.1×10^{-2}
47	effect	2×10^{-2}	97	poetic	1.1×10^{-2}
48	textual	1.9×10^{-2}	98	jewish	1.1×10^{-2}
49	reader	1.9×10^{-2}	99	muslim	1.1×10^{-2}
50	increas	1.8×10^{-2}	100	vernacular	1.1×10^{-2}

Table D.158. The list of the top 100 words in the category Metallurgy and Metallurgical Engineering with RIGs

No.	Word	RIG	No.	Word	RIG
1	alloy	1.5×10^{-1}	51	clinic	1.2×10^{-2}
2	microstructur	1×10^{-1}	52	format	1.2×10^{-2}
3	steel	7.1×10^{-2}	53	particl	1.2×10^{-2}
4	temperatur	5.8×10^{-2}	54	stainless	1.2×10^{-2}
5	grain	5.7×10^{-2}	55	diseas	1.2×10^{-2}
6	phase	3.7×10^{-2}	56	microhard	1.2×10^{-2}
7	cast	3.4×10^{-2}	57	background	1.2×10^{-2}
8	tensil	3.4×10^{-2}	58	process	1.2×10^{-2}
9	corros	3.2×10^{-2}	59	furnac	1.2×10^{-2}
10	properti	3.1×10^{-2}	60	refin	1.1×10^{-2}
11	diffract	3×10^{-2}	61	protein	1.1×10^{-2}
12	austenit	2.9×10^{-2}	62	boundari	1.1×10^{-2}
13	strength	2.9×10^{-2}	63	thermal	1.1×10^{-2}
14	deform	2.8×10^{-2}	64	year	1.1×10^{-2}
15	electron	2.6×10^{-2}	65	associ	1.1×10^{-2}
16	degre	2.6×10^{-2}	66	stress	1×10^{-2}
17	composit	2.5×10^{-2}	67	group	1×10^{-2}
18	metal	2.4×10^{-2}	68	size	1×10^{-2}
19	microscopi	2.4×10^{-2}	69	iron	1×10^{-2}
20	martensit	2.3×10^{-2}	70	resist	1×10^{-2}
21	patient	2.2×10^{-2}	71	nucleat	9.9×10^{-3}
22	ductil	2.2×10^{-2}	72	specimen	9.8×10^{-3}
23	mechan	2.2×10^{-2}	73	surfac	9.6×10^{-3}
24	disloc	2.1×10^{-2}	74	human	9.5×10^{-3}
25	precipit	2.1×10^{-2}	75	molten	9.5×10^{-3}
26	conclus	2.1×10^{-2}	76	coat	9.5×10^{-3}
27	powder	2.1×10^{-2}	77	solid	9.4×10^{-3}
28	ray	2×10^{-2}	78	magnesium	9.4×10^{-3}
29	heat	1.9×10^{-2}	79	room	9.3×10^{-3}
30	scan	1.9×10^{-2}	80	gene	9.1×10^{-3}
31	harden	1.9×10^{-2}	81	may	9×10^{-3}
32	ferrit	1.9×10^{-2}	82	elong	8.9×10^{-3}
33	weld	1.8×10^{-2}	83	titanium	8.8×10^{-3}
34	sinter	1.8×10^{-2}	84	behavior	8.8×10^{-3}
35	mpa	1.7×10^{-2}	85	crystal	8.8×10^{-3}
36	sem	1.7×10^{-2}	86	object	8.6×10^{-3}
37	melt	1.7×10^{-2}	87	express	8.6×10^{-3}
38	anneal	1.7×10^{-2}	88	isotherm	8.3×10^{-3}
39	hard	1.7×10^{-2}	89	nickel	8.1×10^{-3}
40	materi	1.7×10^{-2}	90	cool	8.1×10^{-3}
41	investig	1.6×10^{-2}	91	structur	8.1×10^{-3}
42	roll	1.6×10^{-2}	92	al2o3	8×10^{-3}
43	aluminum	1.6×10^{-2}	93	risk	8×10^{-3}
44	crack	1.5×10^{-2}	94	compress	7.9×10^{-3}
45	strain	1.5×10^{-2}	95	tough	7.8×10^{-3}
46	xrd	1.5×10^{-2}	96	tem	7.7×10^{-3}
47	carbid	1.4×10^{-2}	97	level	7.6×10^{-3}
48	plastic	1.4×10^{-2}	98	content	7.3×10^{-3}
49	hot	1.3×10^{-2}	99	matrix	7.3×10^{-3}
50	fractur	1.3×10^{-2}	100	find	7.2×10^{-3}

Table D.159. The list of the top 100 words in the category Meteorology and Atmospheric Sciences with RIGs

No.	Word	RIG	No.	Word	RIG
1	atmospher	1.2×10^{-1}	51	vertic	2×10^{-2}
2	climat	9.5×10^{-2}	52	atlant	2×10^{-2}
3	aerosol	6.7×10^{-2}	53	patient	2×10^{-2}
4	season	5.8×10^{-2}	54	scale	2×10^{-2}
5	precipit	5.4×10^{-2}	55	resolut	2×10^{-2}
6	meteorolog	5.2×10^{-2}	56	ensembl	2×10^{-2}
7	wind	5.1×10^{-2}	57	radiat	1.8×10^{-2}
8	region	5×10^{-2}	58	pollut	1.7×10^{-2}
9	ocean	5×10^{-2}	59	event	1.7×10^{-2}
10	weather	5×10^{-2}	60	altitud	1.7×10^{-2}
11	summer	4.9×10^{-2}	61	anthropogen	1.7×10^{-2}
12	air	4.6×10^{-2}	62	averag	1.7×10^{-2}
13	tropospher	4.6×10^{-2}	63	estim	1.7×10^{-2}
14	warm	4.6×10^{-2}	64	rain	1.6×10^{-2}
15	tropic	4.4×10^{-2}	65	eastern	1.6×10^{-2}
16	sea	4.4×10^{-2}	66	impact	1.6×10^{-2}
17	forecast	4.2×10^{-2}	67	diurnal	1.6×10^{-2}
18	winter	3.9×10^{-2}	68	east	1.6×10^{-2}
19	satellit	3.7×10^{-2}	69	simul	1.6×10^{-2}
20	rainfal	3.7×10^{-2}	70	water	1.6×10^{-2}
21	cloud	3.6×10^{-2}	71	data	1.6×10^{-2}
22	observ	3.5×10^{-2}	72	conclus	1.6×10^{-2}
23	convect	3.4×10^{-2}	73	variat	1.6×10^{-2}
24	climatolog	3.3×10^{-2}	74	humid	1.6×10^{-2}
25	global	3.1×10^{-2}	75	south	1.5×10^{-2}
26	north	3.1×10^{-2}	76	assimil	1.4×10^{-2}
27	circul	2.9×10^{-2}	77	eddi	1.4×10^{-2}
28	model	2.9×10^{-2}	78	area	1.4×10^{-2}
29	cyclon	2.9×10^{-2}	79	hemispher	1.4×10^{-2}
30	station	2.9×10^{-2}	80	uncertainti	1.4×10^{-2}
31	pacif	2.7×10^{-2}	81	chang	1.4×10^{-2}
32	period	2.6×10^{-2}	82	horizont	1.3×10^{-2}
33	temperatur	2.6×10^{-2}	83	advect	1.3×10^{-2}
34	southern	2.5×10^{-2}	84	moistur	1.3×10^{-2}
35	variabl	2.5×10^{-2}	85	height	1.3×10^{-2}
36	monsoon	2.5×10^{-2}	86	tempor	1.3×10^{-2}
37	latitud	2.4×10^{-2}	87	parameter	1.3×10^{-2}
38	surfac	2.4×10^{-2}	88	mean	1.3×10^{-2}
39	flux	2.3×10^{-2}	89	concentr	1.3×10^{-2}
40	spatial	2.3×10^{-2}	90	particul	1.3×10^{-2}
41	ozon	2.2×10^{-2}	91	radar	1.3×10^{-2}
42	northern	2.2×10^{-2}	92	trend	1.3×10^{-2}
43	storm	2.2×10^{-2}	93	clinic	1.2×10^{-2}
44	emiss	2.2×10^{-2}	94	bias	1.2×10^{-2}
45	interannu	2.2×10^{-2}	95	near	1.2×10^{-2}
46	ice	2.2×10^{-2}	96	larg	1.2×10^{-2}
47	anomali	2.1×10^{-2}	97	spring	1.2×10^{-2}
48	dure	2.1×10^{-2}	98	daili	1.2×10^{-2}
49	annual	2.1×10^{-2}	99	transport	1.2×10^{-2}
50	land	2×10^{-2}	100	sourc	1.1×10^{-2}

TABLE D.160. The list of the top 100 words in the category Microbiology with RIGs

No.	Word	RIG
1	strain	1.1×10^{-1}
2	infect	9.7×10^{-2}
3	gene	8.8×10^{-2}
4	isol	8.7×10^{-2}
5	bacteri	8×10^{-2}
6	bacteria	7.4×10^{-2}
7	pathogen	6.4×10^{-2}
8	rrna	6.2×10^{-2}
9	16s	6×10^{-2}
10	sequenc	5.6×10^{-2}
11	virul	3.7×10^{-2}
12	gram	3.7×10^{-2}
13	antibiot	3.4×10^{-2}
14	paper	3.4×10^{-2}
15	pcr	3.4×10^{-2}
16	phylogenet	3.2×10^{-2}
17	coli	3.2×10^{-2}
18	escherichia	3.2×10^{-2}
19	resist	3.2×10^{-2}
20	speci	3.1×10^{-2}
21	bacterium	3×10^{-2}
22	host	3×10^{-2}
23	genus	2.9×10^{-2}
24	dna	2.9×10^{-2}
25	microbi	2.8×10^{-2}
26	genom	2.7×10^{-2}
27	antimicrobi	2.7×10^{-2}
28	virus	2.5×10^{-2}
29	mutant	2.5×10^{-2}
30	cell	2.3×10^{-2}
31	nov	2.2×10^{-2}
32	cultur	2.1×10^{-2}
33	staphylococcus	2.1×10^{-2}
34	biofilm	2.1×10^{-2}
35	encod	2×10^{-2}
36	protein	2×10^{-2}
37	phenotyp	2×10^{-2}
38	pseudomona	2×10^{-2}
39	suscept	1.9×10^{-2}
40	assay	1.9×10^{-2}
41	mic	1.8×10^{-2}
42	aureus	1.7×10^{-2}
43	microorgan	1.7×10^{-2}
44	acid	1.7×10^{-2}
45	plasmid	1.6×10^{-2}
46	belong	1.6×10^{-2}
47	pneumonia	1.5×10^{-2}
48	identifi	1.5×10^{-2}
49	viral	1.4×10^{-2}
50	type	1.4×10^{-2}

No.	Word	RIG
51	clone	1.4×10^{-2}
52	growth	1.4×10^{-2}
53	mycobacterium	1.4×10^{-2}
54	enzym	1.4×10^{-2}
55	divers	1.4×10^{-2}
56	immun	1.3×10^{-2}
57	bacillus	1.3×10^{-2}
58	detect	1.3×10^{-2}
59	fungal	1.3×10^{-2}
60	aeruginosa	1.2×10^{-2}
61	multidrug	1.2×10^{-2}
62	spp	1.2×10^{-2}
63	yeast	1.2×10^{-2}
64	vaccin	1.2×10^{-2}
65	fatti	1.2×10^{-2}
66	aerob	1.2×10^{-2}
67	salmonella	1.2×10^{-2}
68	colon	1.2×10^{-2}
69	wild	1.1×10^{-2}
70	genotyp	1.1×10^{-2}
71	human	1.1×10^{-2}
72	streptococcus	1.1×10^{-2}
73	tuberculosi	1.1×10^{-2}
74	anaerob	1.1×10^{-2}
75	major	1.1×10^{-2}
76	genet	1.1×10^{-2}
77	communiti	1.1×10^{-2}
78	simul	1.1×10^{-2}
79	activ	1×10^{-2}
80	taxonom	1×10^{-2}
81	ferment	1×10^{-2}
82	negat	9.9×10^{-3}
83	delet	9.8×10^{-3}
84	fungi	9.8×10^{-3}
85	cfu	9.7×10^{-3}
86	agar	9.6×10^{-3}
87	candida	9.6×10^{-3}
88	transcript	9.5×10^{-3}
89	lactobacillus	9.4×10^{-3}
90	presenc	9.4×10^{-3}
91	produc	9.3×10^{-3}
92	express	9.2×10^{-3}
93	comput	9.2×10^{-3}
94	vitro	9.1×10^{-3}
95	rod	9.1×10^{-3}
96	rna	9×10^{-3}
97	microbiolog	8.8×10^{-3}
98	toxin	8.5×10^{-3}
99	caus	8.4×10^{-3}
100	serotyp	8.3×10^{-3}

Table D.161. The list of the top 100 words in the category Microscopy with RIGs

No.	Word	RIG
1	microscopi	1.1×10^{-1}
2	electron	7.7×10^{-2}
3	microscop	6.4×10^{-2}
4	imag	6.3×10^{-2}
5	scan	5.1×10^{-2}
6	transmiss	3.4×10^{-2}
7	resolut	3.2×10^{-2}
8	ultrastructur	2.7×10^{-2}
9	beam	2.5×10^{-2}
10	specimen	2.1×10^{-2}
11	atom	1.8×10^{-2}
12	sem	1.7×10^{-2}
13	tem	1.6×10^{-2}
14	optic	1.5×10^{-2}
15	morpholog	1.4×10^{-2}
16	fluoresc	1.3×10^{-2}
17	cell	1.2×10^{-2}
18	reconstruct	1.2×10^{-2}
19	light	1.1×10^{-2}
20	aberr	1.1×10^{-2}
21	sampl	1.1×10^{-2}
22	probe	1×10^{-2}
23	thick	9.1×10^{-3}
24	afm	8.8×10^{-3}
25	techniqu	8.8×10^{-3}
26	confoc	8.7×10^{-3}
27	tomographi	8.7×10^{-3}
28	laser	8.6×10^{-3}
29	tissu	8.5×10^{-3}
30	scatter	8.4×10^{-3}
31	surfac	8.4×10^{-3}
32	detector	7.5×10^{-3}
33	structur	7.5×10^{-3}
34	section	7.3×10^{-3}
35	stem	6.8×10^{-3}
36	thin	6.5×10^{-3}
37	cytoplasm	6.3×10^{-3}
38	contrast	6.2×10^{-3}
39	dark	6.2×10^{-3}
40	conclus	6.1×10^{-3}
41	stain	6×10^{-3}
42	patient	5.9×10^{-3}
43	illumin	5.6×10^{-3}
44	spatial	5.4×10^{-3}
45	len	5.4×10^{-3}
46	angl	5.2×10^{-3}
47	materi	5.1×10^{-3}
48	shape	4.9×10^{-3}
49	membran	4.7×10^{-3}
50	diffract	4.7×10^{-3}
51	situ	4.7×10^{-3}
52	tip	4.6×10^{-3}
53	nuclei	4.5×10^{-3}
54	backscatt	4.5×10^{-3}
55	collagen	4.5×10^{-3}
56	cellular	4.5×10^{-3}
57	fiber	4.5×10^{-3}
58	biolog	4.3×10^{-3}
59	quantit	4.3×10^{-3}
60	spectroscopi	4.1×10^{-3}
61	outcom	4.1×10^{-3}
62	rate	4×10^{-3}
63	paper	3.9×10^{-3}
64	nanoscal	3.9×10^{-3}
65	model	3.9×10^{-3}
66	correct	3.9×10^{-3}
67	focal	3.8×10^{-3}
68	allow	3.8×10^{-3}
69	immunoreact	3.7×10^{-3}
70	year	3.7×10^{-3}
71	dentin	3.7×10^{-3}
72	risk	3.6×10^{-3}
73	manag	3.6×10^{-3}
74	particip	3.6×10^{-3}
75	axi	3.5×10^{-3}
76	dimension	3.5×10^{-3}
77	microstructur	3.4×10^{-3}
78	rough	3.4×10^{-3}
79	age	3.4×10^{-3}
80	use	3.4×10^{-3}
81	vesicl	3.4×10^{-3}
82	nanoparticl	3.3×10^{-3}
83	observ	3.3×10^{-3}
84	label	3.2×10^{-3}
85	damag	3.2×10^{-3}
86	layer	3.1×10^{-3}
87	enabl	3.1×10^{-3}
88	social	3.1×10^{-3}
89	acquisit	3.1×10^{-3}
90	predict	3×10^{-3}
91	epitheli	3×10^{-3}
92	field	3×10^{-3}
93	coher	2.9×10^{-3}
94	topographi	2.9×10^{-3}
95	health	2.9×10^{-3}
96	instrument	2.9×10^{-3}
97	photon	2.9×10^{-3}
98	ray	2.8×10^{-3}
99	enamel	2.8×10^{-3}
100	assess	2.8×10^{-3}

Table D.162. The list of the top 100 words in the category Mineralogy with RIGs

No.	Word	RIG	No.	Word	RIG
1	miner	1.5×10^{-1}	51	angstrom	2.2×10^{-2}
2	rock	1×10^{-1}	52	assemblag	2.2×10^{-2}
3	magmat	6.6×10^{-2}	53	element	2.1×10^{-2}
4	mantl	6.5×10^{-2}	54	fluid	2.1×10^{-2}
5	magma	6.4×10^{-2}	55	calcit	2.1×10^{-2}
6	ore	6.3×10^{-2}	56	content	2.1×10^{-2}
7	crystal	5.5×10^{-2}	57	orogen	2×10^{-2}
8	melt	5.3×10^{-2}	58	trace	2×10^{-2}
9	zircon	4.8×10^{-2}	59	host	1.9×10^{-2}
10	composit	4.8×10^{-2}	60	occur	1.8×10^{-2}
11	geochem	4.7×10^{-2}	61	format	1.8×10^{-2}
12	mineralog	4.5×10^{-2}	62	h2o	1.8×10^{-2}
13	quartz	4.5×10^{-2}	63	geolog	1.7×10^{-2}
14	mafic	4.2×10^{-2}	64	gpa	1.7×10^{-2}
15	metamorph	4×10^{-2}	65	earth	1.7×10^{-2}
16	crustal	4×10^{-2}	66	grain	1.7×10^{-2}
17	clay	4×10^{-2}	67	continent	1.7×10^{-2}
18	isotop	3.9×10^{-2}	68	geochronolog	1.7×10^{-2}
19	crust	3.8×10^{-2}	69	craton	1.7×10^{-2}
20	subduct	3.8×10^{-2}	70	mgo	1.6×10^{-2}
21	plagioclas	3.6×10^{-2}	71	apatit	1.6×10^{-2}
22	granit	3.6×10^{-2}	72	similar	1.6×10^{-2}
23	basalt	3.4×10^{-2}	73	tecton	1.6×10^{-2}
24	zone	3.3×10^{-2}	74	sio2	1.6×10^{-2}
25	ree	3.1×10^{-2}	75	patient	1.6×10^{-2}
26	olivin	3.1×10^{-2}	76	cation	1.5×10^{-2}
27	rich	3.1×10^{-2}	77	alkalin	1.5×10^{-2}
28	silic	3.1×10^{-2}	78	geochemistri	1.5×10^{-2}
29	bear	3×10^{-2}	79	interlay	1.5×10^{-2}
30	diffract	2.9×10^{-2}	80	epsilon	1.5×10^{-2}
31	igneous	2.8×10^{-2}	81	textur	1.5×10^{-2}
32	degre	2.8×10^{-2}	82	deplet	1.4×10^{-2}
33	garnet	2.7×10^{-2}	83	iron	1.4×10^{-2}
34	intrus	2.7×10^{-2}	84	mine	1.4×10^{-2}
35	feldspar	2.6×10^{-2}	85	dissolut	1.4×10^{-2}
36	deposit	2.6×10^{-2}	86	phase	1.4×10^{-2}
37	form	2.5×10^{-2}	87	arc	1.3×10^{-2}
38	emplac	2.5×10^{-2}	88	late	1.3×10^{-2}
39	pyrit	2.4×10^{-2}	89	erupt	1.3×10^{-2}
40	enrich	2.4×10^{-2}	90	refin	1.3×10^{-2}
41	pluton	2.4×10^{-2}	91	precipit	1.3×10^{-2}
42	sulfid	2.4×10^{-2}	92	spinel	1.2×10^{-2}
43	flotat	2.4×10^{-2}	93	sedimentari	1.2×10^{-2}
44	volcan	2.4×10^{-2}	94	conclus	1.2×10^{-2}
45	hydrotherm	2.3×10^{-2}	95	indic	1.2×10^{-2}
46	temperatur	2.3×10^{-2}	96	cretac	1.2×10^{-2}
47	ray	2.3×10^{-2}	97	fraction	1.2×10^{-2}
48	chemic	2.2×10^{-2}	98	contain	1.2×10^{-2}
49	belt	2.2×10^{-2}	99	leach	1.2×10^{-2}
50	lithospher	2.2×10^{-2}	100	vein	1.1×10^{-2}

TABLE D.163. The list of the top 100 words in the category Mining and Mineral Processing with RIGs

No.	Word	RIG
1	mine	9.4×10^{-2}
2	coal	7.2×10^{-2}
3	ore	6.2×10^{-2}
4	rock	5.5×10^{-2}
5	miner	4.4×10^{-2}
6	flotat	3.8×10^{-2}
7	geolog	2.8×10^{-2}
8	underground	2.4×10^{-2}
9	patient	1.7×10^{-2}
10	seismic	1.3×10^{-2}
11	leach	1.3×10^{-2}
12	conclus	1.1×10^{-2}
13	drill	1.1×10^{-2}
14	mineralog	1.1×10^{-2}
15	process	1.1×10^{-2}
16	deposit	1.1×10^{-2}
17	geophys	1.1×10^{-2}
18	pyrit	1×10^{-2}
19	sulfid	1×10^{-2}
20	clinic	9.8×10^{-3}
21	excav	9.7×10^{-3}
22	iron	9.5×10^{-3}
23	fractur	9.3×10^{-3}
24	zone	9.3×10^{-3}
25	borehol	8.6×10^{-3}
26	protein	8.3×10^{-3}
27	diseas	8.3×10^{-3}
28	quartz	8.1×10^{-3}
29	deform	8×10^{-3}
30	copper	7.5×10^{-3}
31	gas	7.4×10^{-3}
32	reservoir	7.2×10^{-3}
33	fine	7.1×10^{-3}
34	particl	7×10^{-3}
35	cell	6.7×10^{-3}
36	area	6.7×10^{-3}
37	gene	6.6×10^{-3}
38	grain	6.2×10^{-3}
39	alloy	6.2×10^{-3}
40	geotechn	6.1×10^{-3}
41	industri	6×10^{-3}
42	recoveri	6×10^{-3}
43	background	6×10^{-3}
44	depth	5.9×10^{-3}
45	sandston	5.8×10^{-3}
46	clay	5.6×10^{-3}
47	stress	5.5×10^{-3}
48	concentr	5.4×10^{-3}
49	particip	5.4×10^{-3}
50	tecton	5.2×10^{-3}
51	granit	5.1×10^{-3}
52	strength	5.1×10^{-3}
53	pressur	5.1×10^{-3}
54	microstructur	5.1×10^{-3}
55	cancer	5×10^{-3}
56	shear	4.9×10^{-3}
57	laboratori	4.9×10^{-3}
58	condit	4.9×10^{-3}
59	group	4.7×10^{-3}
60	metal	4.7×10^{-3}
61	wast	4.7×10^{-3}
62	safeti	4.6×10^{-3}
63	poros	4.6×10^{-3}
64	surfac	4.6×10^{-3}
65	collector	4.5×10^{-3}
66	cave	4.5×10^{-3}
67	bear	4.5×10^{-3}
68	precipit	4.4×10^{-3}
69	shale	4.4×10^{-3}
70	tissu	4.4×10^{-3}
71	horizont	4.3×10^{-3}
72	veloc	4.3×10^{-3}
73	element	4.3×10^{-3}
74	limeston	4.2×10^{-3}
75	furnac	4.2×10^{-3}
76	basin	4.2×10^{-3}
77	report	4.1×10^{-3}
78	express	4.1×10^{-3}
79	therapi	4.1×10^{-3}
80	steel	4.1×10^{-3}
81	water	4.1×10^{-3}
82	hydraul	4.1×10^{-3}
83	infect	4.1×10^{-3}
84	locat	4.1×10^{-3}
85	materi	4×10^{-3}
86	blood	4×10^{-3}
87	strata	4×10^{-3}
88	human	4×10^{-3}
89	popul	4×10^{-3}
90	articl	4×10^{-3}
91	paramet	3.9×10^{-3}
92	calcit	3.9×10^{-3}
93	crack	3.9×10^{-3}
94	main	3.8×10^{-3}
95	adult	3.8×10^{-3}
96	mediat	3.8×10^{-3}
97	grade	3.8×10^{-3}
98	compress	3.7×10^{-3}
99	tumor	3.7×10^{-3}
100	drug	3.7×10^{-3}

Table D.164. The list of the top 100 words in the category Multidisciplinary Sciences with RIGs

No.	Word	RIG
1	gene	2.4×10^{-2}
2	cell	2.3×10^{-2}
3	protein	2.2×10^{-2}
4	express	2.1×10^{-2}
5	regul	1.6×10^{-2}
6	paper	1.4×10^{-2}
7	suggest	1.1×10^{-2}
8	genom	1.1×10^{-2}
9	mice	1.1×10^{-2}
10	human	1×10^{-2}
11	pathway	9.5×10^{-3}
12	transcript	9.3×10^{-3}
13	induc	9.2×10^{-3}
14	role	9×10^{-3}
15	mediat	8.9×10^{-3}
16	associ	8.5×10^{-3}
17	diseas	7.7×10^{-3}
18	genet	7.4×10^{-3}
19	receptor	7.3×10^{-3}
20	mous	7.2×10^{-3}
21	bind	7.2×10^{-3}
22	activ	7.1×10^{-3}
23	sequenc	6.7×10^{-3}
24	howev	6.5×10^{-3}
25	vivo	6.5×10^{-3}
26	phenotyp	6.4×10^{-3}
27	identifi	6.4×10^{-3}
28	inhibit	6.4×10^{-3}
29	infect	6.3×10^{-3}
30	respons	6.2×10^{-3}
31	mutat	6.1×10^{-3}
32	cellular	5.9×10^{-3}
33	mutant	5.9×10^{-3}
34	rna	5.7×10^{-3}
35	tissu	5.6×10^{-3}
36	speci	5.6×10^{-3}
37	dna	5.6×10^{-3}
38	immun	5.6×10^{-3}
39	may	5.6×10^{-3}
40	specif	5.5×10^{-3}
41	target	5.3×10^{-3}
42	anim	5.2×10^{-3}
43	vitro	5.2×10^{-3}
44	popul	5.1×10^{-3}
45	molecular	4.9×10^{-3}
46	remain	4.9×10^{-3}
47	function	4.8×10^{-3}
48	alter	4.6×10^{-3}
49	biolog	4.6×10^{-3}
50	find	4.6×10^{-3}
51	mechan	4.5×10^{-3}
52	neuron	4.4×10^{-3}
53	wild	4.4×10^{-3}
54	pathogen	4.3×10^{-3}
55	metabol	4.3×10^{-3}
56	demonstr	4.3×10^{-3}
57	reveal	4.2×10^{-3}
58	signal	4.2×10^{-3}
59	level	4.2×10^{-3}
60	involv	4×10^{-3}
61	assay	4×10^{-3}
62	signific	3.9×10^{-3}
63	virus	3.9×10^{-3}
64	play	3.8×10^{-3}
65	encod	3.8×10^{-3}
66	known	3.8×10^{-3}
67	cancer	3.7×10^{-3}
68	mrna	3.7×10^{-3}
69	studi	3.7×10^{-3}
70	kinas	3.6×10^{-3}
71	marker	3.6×10^{-3}
72	potenti	3.6×10^{-3}
73	prolifer	3.6×10^{-3}
74	transcriptom	3.6×10^{-3}
75	bacteri	3.5×10^{-3}
76	promot	3.5×10^{-3}
77	pcr	3.4×10^{-3}
78	divers	3.4×10^{-3}
79	import	3.4×10^{-3}
80	differenti	3.4×10^{-3}
81	overexpress	3.3×10^{-3}
82	viral	3.3×10^{-3}
83	stimul	3.3×10^{-3}
84	factor	3.2×10^{-3}
85	therapeut	3.2×10^{-3}
86	beta	3.2×10^{-3}
87	previous	3.2×10^{-3}
88	articl	3.1×10^{-3}
89	nucleotid	3.1×10^{-3}
90	phosphoryl	3.1×10^{-3}
91	cytokin	3.1×10^{-3}
92	mammalian	3.1×10^{-3}
93	unknown	3×10^{-3}
94	lineag	3×10^{-3}
95	conserv	3×10^{-3}
96	physiolog	3×10^{-3}
97	regulatori	3×10^{-3}
98	host	3×10^{-3}
99	like	3×10^{-3}
100	brain	2.9×10^{-3}

Table D.165. The list of the top 100 words in the category Music with RIGs

No.	Word	RIG
1	music	5.1×10^{-1}
2	musician	9.8×10^{-2}
3	articl	5.5×10^{-2}
4	artist	4.7×10^{-2}
5	song	3.7×10^{-2}
6	listen	3.7×10^{-2}
7	centuri	3.4×10^{-2}
8	compos	3.3×10^{-2}
9	genr	3×10^{-2}
10	sound	2.8×10^{-2}
11	sing	2.7×10^{-2}
12	instrument	2.7×10^{-2}
13	creativ	2.7×10^{-2}
14	piec	2.4×10^{-2}
15	cultur	2.2×10^{-2}
16	style	2.1×10^{-2}
17	danc	2.1×10^{-2}
18	aesthet	2×10^{-2}
19	explor	1.9×10^{-2}
20	vocal	1.8×10^{-2}
21	audienc	1.7×10^{-2}
22	work	1.7×10^{-2}
23	student	1.6×10^{-2}
24	teacher	1.5×10^{-2}
25	voic	1.5×10^{-2}
26	school	1.5×10^{-2}
27	contemporari	1.4×10^{-2}
28	dancer	1.4×10^{-2}
29	practic	1.4×10^{-2}
30	art	1.4×10^{-2}
31	popular	1.4×10^{-2}
32	string	1.4×10^{-2}
33	argu	1.4×10^{-2}
34	educ	1.3×10^{-2}
35	particip	1.3×10^{-2}
36	text	1.3×10^{-2}
37	percept	1.3×10^{-2}
38	cell	1.2×10^{-2}
39	result	1.2×10^{-2}
40	stylist	1.2×10^{-2}
41	essay	1.2×10^{-2}
42	pitch	1.2×10^{-2}
43	patient	1.1×10^{-2}
44	lyric	1.1×10^{-2}
45	method	1.1×10^{-2}
46	idea	1.1×10^{-2}
47	draw	1×10^{-2}
48	audio	1×10^{-2}
49	tradit	1×10^{-2}
50	engag	1×10^{-2}
51	emot	1×10^{-2}
52	teach	1×10^{-2}
53	discours	1×10^{-2}
54	scholar	1×10^{-2}
55	manuscript	1×10^{-2}
56	way	1×10^{-2}
57	theme	9.9×10^{-3}
58	histor	9.8×10^{-3}
59	temperatur	9.7×10^{-3}
60	eighteenth	9.5×10^{-3}
61	narrat	9.5×10^{-3}
62	context	9.4×10^{-3}
63	creat	9.4×10^{-3}
64	increas	9.4×10^{-3}
65	gestur	9.3×10^{-3}
66	question	9.3×10^{-3}
67	profession	9.2×10^{-3}
68	career	9.1×10^{-3}
69	folk	9.1×10^{-3}
70	twentieth	8.8×10^{-3}
71	ratio	8.6×10^{-3}
72	nineteenth	8.4×10^{-3}
73	write	8.4×10^{-3}
74	repertoir	8.1×10^{-3}
75	effect	8.1×10^{-3}
76	social	8.1×10^{-3}
77	sacr	8.1×10^{-3}
78	interview	8.1×10^{-3}
79	notion	7.9×10^{-3}
80	simul	7.9×10^{-3}
81	low	7.8×10^{-3}
82	examin	7.8×10^{-3}
83	offer	7.6×10^{-3}
84	effici	7.5×10^{-3}
85	languag	7.5×10^{-3}
86	obtain	7.4×10^{-3}
87	tone	7.4×10^{-3}
88	decreas	7.3×10^{-3}
89	recept	7.1×10^{-3}
90	movement	7.1×10^{-3}
91	note	7.1×10^{-3}
92	written	7×10^{-3}
93	polit	6.9×10^{-3}
94	research	6.9×10^{-3}
95	perceiv	6.8×10^{-3}
96	pari	6.8×10^{-3}
97	high	6.8×10^{-3}
98	diseas	6.8×10^{-3}
99	protein	6.7×10^{-3}
100	acid	6.7×10^{-3}

Table D.166. The list of the top 100 words in the category Mycology with RIGs

No.	Word	RIG
1	speci	1.2×10^{-1}
2	fungal	1.1×10^{-1}
3	fungi	1.1×10^{-1}
4	phylogenet	6.4×10^{-2}
5	fungus	6.2×10^{-2}
6	genus	4.9×10^{-2}
7	isol	4.7×10^{-2}
8	yeast	4.7×10^{-2}
9	sequenc	4.3×10^{-2}
10	pathogen	4.1×10^{-2}
11	candida	3.9×10^{-2}
12	morpholog	3.9×10^{-2}
13	antifung	3.6×10^{-2}
14	spacer	3.6×10^{-2}
15	transcrib	3.6×10^{-2}
16	aspergillus	3.1×10^{-2}
17	strain	2.8×10^{-2}
18	albican	2.7×10^{-2}
19	clade	2.7×10^{-2}
20	spore	2.7×10^{-2}
21	gene	2.7×10^{-2}
22	rdna	2.6×10^{-2}
23	saccharomyc	2.6×10^{-2}
24	cerevisia	2.5×10^{-2}
25	nov	2.5×10^{-2}
26	genera	2.4×10^{-2}
27	taxa	2.2×10^{-2}
28	infect	2.2×10^{-2}
29	host	2.1×10^{-2}
30	cultur	2.1×10^{-2}
31	describ	2×10^{-2}
32	paper	1.9×10^{-2}
33	ribosom	1.9×10^{-2}
34	molecular	1.7×10^{-2}
35	phylogeni	1.7×10^{-2}
36	virul	1.6×10^{-2}
37	plant	1.6×10^{-2}
38	spp	1.6×10^{-2}
39	mutant	1.5×10^{-2}
40	belong	1.4×10^{-2}
41	taxonom	1.4×10^{-2}
42	dna	1.4×10^{-2}
43	subunit	1.2×10^{-2}
44	agar	1.2×10^{-2}
45	new	1.2×10^{-2}
46	growth	1.1×10^{-2}
47	fusarium	1.1×10^{-2}
48	measur	1.1×10^{-2}
49	method	1.1×10^{-2}
50	delet	1.1×10^{-2}

No.	Word	RIG
51	simul	1.1×10^{-2}
52	filament	1.1×10^{-2}
53	charact	1×10^{-2}
54	model	1×10^{-2}
55	brown	9.8×10^{-3}
56	tree	9.6×10^{-3}
57	collect	9.6×10^{-3}
58	brazil	8.8×10^{-3}
59	forest	8.7×10^{-3}
60	intern	8.3×10^{-3}
61	identifi	8.1×10^{-3}
62	coloni	8.1×10^{-3}
63	distinct	8×10^{-3}
64	taxon	8×10^{-3}
65	conclus	7.6×10^{-3}
66	specimen	7.5×10^{-3}
67	var	7.3×10^{-3}
68	inocul	7.2×10^{-3}
69	encod	7.2×10^{-3}
70	divers	7.1×10^{-3}
71	perform	7×10^{-3}
72	result	7×10^{-3}
73	ecolog	6.9×10^{-3}
74	energi	6.9×10^{-3}
75	identif	6.7×10^{-3}
76	algorithm	6.6×10^{-3}
77	colon	6.6×10^{-3}
78	lineag	6.5×10^{-3}
79	fruit	6.4×10^{-3}
80	analys	6.4×10^{-3}
81	genom	6.4×10^{-3}
82	caus	6.3×10^{-3}
83	mate	6.3×10^{-3}
84	problem	6.2×10^{-3}
85	solut	6.2×10^{-3}
86	illustr	6.1×10^{-3}
87	phenotyp	6×10^{-3}
88	cultiv	5.9×10^{-3}
89	eukaryot	5.9×10^{-3}
90	known	5.9×10^{-3}
91	biosynthesi	5.9×10^{-3}
92	experiment	5.9×10^{-3}
93	wall	5.9×10^{-3}
94	comput	5.8×10^{-3}
95	protein	5.8×10^{-3}
96	behavior	5.8×10^{-3}
97	enzym	5.7×10^{-3}
98	wild	5.6×10^{-3}
99	calcul	5.4×10^{-3}
100	soil	5.4×10^{-3}

TABLE D.167. The list of the top 100 words in the category Nanoscience and Nanotechnology with RIGs

No.	Word	RIG
1	nanoparticl	5.2×10^{-2}
2	electron	4.2×10^{-2}
3	fabric	4.2×10^{-2}
4	surfac	3.9×10^{-2}
5	film	2.9×10^{-2}
6	devic	2.6×10^{-2}
7	nanostructur	2.5×10^{-2}
8	layer	2.2×10^{-2}
9	microscopi	2.2×10^{-2}
10	metal	2.2×10^{-2}
11	graphen	2.2×10^{-2}
12	properti	2.2×10^{-2}
13	conclus	2.1×10^{-2}
14	patient	2×10^{-2}
15	materi	1.9×10^{-2}
16	electrod	1.8×10^{-2}
17	substrat	1.8×10^{-2}
18	oxid	1.8×10^{-2}
19	spectroscopi	1.8×10^{-2}
20	structur	1.7×10^{-2}
21	nanotub	1.6×10^{-2}
22	nanowir	1.6×10^{-2}
23	prepar	1.6×10^{-2}
24	plasmon	1.6×10^{-2}
25	atom	1.5×10^{-2}
26	charg	1.5×10^{-2}
27	deposit	1.5×10^{-2}
28	data	1.4×10^{-2}
29	optic	1.4×10^{-2}
30	year	1.4×10^{-2}
31	synthes	1.4×10^{-2}
32	silicon	1.4×10^{-2}
33	thin	1.4×10^{-2}
34	coat	1.3×10^{-2}
35	temperatur	1.3×10^{-2}
36	gold	1.2×10^{-2}
37	size	1.2×10^{-2}
38	nanoscal	1.2×10^{-2}
39	nano	1.2×10^{-2}
40	electrochem	1.2×10^{-2}
41	polym	1.2×10^{-2}
42	applic	1.1×10^{-2}
43	diffract	1.1×10^{-2}
44	associ	1.1×10^{-2}
45	exhibit	1.1×10^{-2}
46	ray	1.1×10^{-2}
47	electr	1.1×10^{-2}
48	chemic	1.1×10^{-2}
49	object	1.1×10^{-2}
50	dope	1.1×10^{-2}

No.	Word	RIG
51	microfluid	1.1×10^{-2}
52	semiconductor	1.1×10^{-2}
53	enhanc	1.1×10^{-2}
54	high	1×10^{-2}
55	carbon	1×10^{-2}
56	background	1×10^{-2}
57	tio2	9.8×10^{-3}
58	assess	9.8×10^{-3}
59	risk	9.8×10^{-3}
60	particl	9.6×10^{-3}
61	selfassembl	9.6×10^{-3}
62	ion	9.2×10^{-3}
63	densiti	9×10^{-3}
64	nps	9×10^{-3}
65	energi	8.9×10^{-3}
66	scan	8.9×10^{-3}
67	quantum	8.9×10^{-3}
68	zno	8.9×10^{-3}
69	age	8.9×10^{-3}
70	raman	8.8×10^{-3}
71	nanomateri	8.8×10^{-3}
72	nanocryst	8.8×10^{-3}
73	etch	8.8×10^{-3}
74	adsorpt	8.7×10^{-3}
75	morpholog	8.6×10^{-3}
76	light	8.6×10^{-3}
77	absorpt	8.2×10^{-3}
78	molecul	8.2×10^{-3}
79	microstructur	8.1×10^{-3}
80	photoluminesc	8.1×10^{-3}
81	tem	8.1×10^{-3}
82	anneal	8.1×10^{-3}
83	poli	8×10^{-3}
84	thick	8×10^{-3}
85	interfac	7.9×10^{-3}
86	nanorod	7.9×10^{-3}
87	aim	7.9×10^{-3}
88	voltag	7.8×10^{-3}
89	identifi	7.7×10^{-3}
90	diamet	7.6×10^{-3}
91	transistor	7.6×10^{-3}
92	outcom	7.6×10^{-3}
93	particip	7.6×10^{-3}
94	nanocomposit	7.5×10^{-3}
95	monolay	7.4×10^{-3}
96	mesopor	7.4×10^{-3}
97	promis	7.3×10^{-3}
98	crystallin	7.2×10^{-3}
99	shell	7.1×10^{-3}
100	excel	7×10^{-3}

TABLE D.168. The list of the top 100 words in the category Neuroimaging with RIGs

No.	Word	RIG
1	brain	1.8×10^{-1}
2	fmri	1.2×10^{-1}
3	imag	1.1×10^{-1}
4	cortex	1×10^{-1}
5	cortic	8×10^{-2}
6	frontal	5.7×10^{-2}
7	mri	5.3×10^{-2}
8	neuroimag	5.2×10^{-2}
9	pariet	4.8×10^{-2}
10	region	4.8×10^{-2}
11	voxel	4.8×10^{-2}
12	prefront	4.6×10^{-2}
13	healthi	4.6×10^{-2}
14	reson	4.5×10^{-2}
15	cerebr	4.4×10^{-2}
16	magnet	4.3×10^{-2}
17	neural	4.2×10^{-2}
18	gyrus	4.1×10^{-2}
19	patient	4.1×10^{-2}
20	anterior	4×10^{-2}
21	cingul	4×10^{-2}
22	cognit	3.9×10^{-2}
23	tempor	3.9×10^{-2}
24	task	3.7×10^{-2}
25	matter	3.6×10^{-2}
26	bold	3.5×10^{-2}
27	posterior	3.4×10^{-2}
28	left	3.4×10^{-2}
29	right	3.1×10^{-2}
30	function	3×10^{-2}
31	subject	2.9×10^{-2}
32	inferior	2.8×10^{-2}
33	intracrani	2.8×10^{-2}
34	rest	2.8×10^{-2}
35	connect	2.7×10^{-2}
36	correl	2.7×10^{-2}
37	bilater	2.6×10^{-2}
38	eeg	2.6×10^{-2}
39	occipit	2.5×10^{-2}
40	gray	2.5×10^{-2}
41	white	2.4×10^{-2}
42	subcort	2.3×10^{-2}
43	visual	2.2×10^{-2}
44	anatom	2.2×10^{-2}
45	motor	2.1×10^{-2}
46	aneurysm	2×10^{-2}
47	medial	2×10^{-2}
48	arteri	1.9×10^{-2}
49	endovascular	1.9×10^{-2}
50	volum	1.9×10^{-2}
51	insula	1.8×10^{-2}
52	diffus	1.8×10^{-2}
53	abnorm	1.8×10^{-2}
54	hemispher	1.7×10^{-2}
55	lobe	1.7×10^{-2}
56	paper	1.7×10^{-2}
57	relat	1.7×10^{-2}
58	tensor	1.7×10^{-2}
59	network	1.6×10^{-2}
60	disord	1.6×10^{-2}
61	purpos	1.6×10^{-2}
62	amygdala	1.5×10^{-2}
63	stimuli	1.5×10^{-2}
64	stimulus	1.5×10^{-2}
65	clinic	1.5×10^{-2}
66	ventral	1.5×10^{-2}
67	anisotropi	1.5×10^{-2}
68	underw	1.5×10^{-2}
69	embol	1.5×10^{-2}
70	spatial	1.4×10^{-2}
71	find	1.4×10^{-2}
72	area	1.4×10^{-2}
73	associ	1.4×10^{-2}
74	deficit	1.3×10^{-2}
75	age	1.3×10^{-2}
76	auditori	1.3×10^{-2}
77	activ	1.3×10^{-2}
78	stroke	1.3×10^{-2}
79	background	1.3×10^{-2}
80	superior	1.2×10^{-2}
81	hippocampus	1.2×10^{-2}
82	map	1.2×10^{-2}
83	occlus	1.2×10^{-2}
84	lesion	1.2×10^{-2}
85	contrast	1.2×10^{-2}
86	pattern	1.1×10^{-2}
87	tract	1.1×10^{-2}
88	across	1.1×10^{-2}
89	hemodynam	1.1×10^{-2}
90	suggest	1.1×10^{-2}
91	schizophrenia	1.1×10^{-2}
92	studi	1.1×10^{-2}
93	may	1.1×10^{-2}
94	evok	1.1×10^{-2}
95	temperatur	1.1×10^{-2}
96	impair	1.1×10^{-2}
97	stent	1×10^{-2}
98	neuron	1×10^{-2}
99	alzheim	1×10^{-2}
100	angiographi	1×10^{-2}

TABLE D.169. The list of the top 100 words in the category Neurosciences with RIGs

No.	Word	RIG
1	brain	1.5×10^{-1}
2	neuron	1.1×10^{-1}
3	cortex	8.4×10^{-2}
4	cortic	5.2×10^{-2}
5	rat	4.8×10^{-2}
6	cognit	4.4×10^{-2}
7	neural	4×10^{-2}
8	suggest	3.8×10^{-2}
9	hippocampus	3.7×10^{-2}
10	synapt	3.4×10^{-2}
11	motor	3.4×10^{-2}
12	impair	3.3×10^{-2}
13	receptor	3.2×10^{-2}
14	hippocamp	3.2×10^{-2}
15	disord	3.1×10^{-2}
16	prefront	3.1×10^{-2}
17	activ	3×10^{-2}
18	deficit	3×10^{-2}
19	alzheim	2.9×10^{-2}
20	paper	2.9×10^{-2}
21	task	2.8×10^{-2}
22	cerebr	2.7×10^{-2}
23	memori	2.6×10^{-2}
24	stimuli	2.5×10^{-2}
25	mice	2.4×10^{-2}
26	healthi	2.4×10^{-2}
27	evok	2.3×10^{-2}
28	fmri	2.3×10^{-2}
29	frontal	2.3×10^{-2}
30	stimulus	2.3×10^{-2}
31	induc	2.2×10^{-2}
32	function	2.1×10^{-2}
33	associ	2.1×10^{-2}
34	sensori	2.1×10^{-2}
35	gyrus	2×10^{-2}
36	may	2×10^{-2}
37	stimul	2×10^{-2}
38	whether	2×10^{-2}
39	diseas	1.9×10^{-2}
40	alter	1.9×10^{-2}
41	axon	1.9×10^{-2}
42	respons	1.9×10^{-2}
43	parkinson	1.8×10^{-2}
44	nervous	1.8×10^{-2}
45	dorsal	1.7×10^{-2}
46	anim	1.7×10^{-2}
47	nucleus	1.7×10^{-2}
48	express	1.7×10^{-2}
49	subject	1.7×10^{-2}
50	medial	1.7×10^{-2}

No.	Word	RIG
51	dopamin	1.7×10^{-2}
52	pariet	1.7×10^{-2}
53	neurolog	1.7×10^{-2}
54	neuroprotect	1.6×10^{-2}
55	adult	1.6×10^{-2}
56	auditori	1.6×10^{-2}
57	patient	1.6×10^{-2}
58	amygdala	1.6×10^{-2}
59	role	1.6×10^{-2}
60	neurodegen	1.5×10^{-2}
61	tempor	1.5×10^{-2}
62	striatum	1.5×10^{-2}
63	nerv	1.5×10^{-2}
64	behavior	1.5×10^{-2}
65	spinal	1.5×10^{-2}
66	cingul	1.5×10^{-2}
67	ventral	1.5×10^{-2}
68	modul	1.4×10^{-2}
69	control	1.4×10^{-2}
70	eeg	1.4×10^{-2}
71	antagonist	1.4×10^{-2}
72	mediat	1.4×10^{-2}
73	find	1.4×10^{-2}
74	sclerosi	1.4×10^{-2}
75	neuroimag	1.4×10^{-2}
76	glutam	1.3×10^{-2}
77	particip	1.3×10^{-2}
78	stroke	1.3×10^{-2}
79	involv	1.3×10^{-2}
80	electrophysiolog	1.3×10^{-2}
81	temperatur	1.3×10^{-2}
82	later	1.3×10^{-2}
83	injuri	1.3×10^{-2}
84	visual	1.3×10^{-2}
85	mous	1.3×10^{-2}
86	signific	1.3×10^{-2}
87	studi	1.3×10^{-2}
88	onset	1.2×10^{-2}
89	depress	1.2×10^{-2}
90	amyloid	1.2×10^{-2}
91	anterior	1.2×10^{-2}
92	evid	1.2×10^{-2}
93	maze	1.2×10^{-2}
94	dysfunct	1.2×10^{-2}
95	increas	1.2×10^{-2}
96	inhibit	1.1×10^{-2}
97	agonist	1.1×10^{-2}
98	bilater	1.1×10^{-2}
99	glial	1.1×10^{-2}
100	regul	1.1×10^{-2}

Table D.170. The list of the top 100 words in the category Nuclear Science and Technology with RIGs

No.	Word	RIG	No.	Word	RIG
1	reactor	1.1×10^{-1}	51	gene	7.6×10^{-3}
2	nuclear	7.1×10^{-2}	52	particl	7.5×10^{-3}
3	neutron	7.1×10^{-2}	53	associ	7.2×10^{-3}
4	detector	4.4×10^{-2}	54	acceler	7.2×10^{-3}
5	irradi	4.3×10^{-2}	55	proton	7×10^{-3}
6	fuel	3.3×10^{-2}	56	express	6.9×10^{-3}
7	radiat	3.3×10^{-2}	57	fast	6.9×10^{-3}
8	coolant	3.3×10^{-2}	58	find	6.9×10^{-3}
9	accid	3.1×10^{-2}	59	pressur	6.8×10^{-3}
10	code	3×10^{-2}	60	phantom	6.6×10^{-3}
11	mev	2.9×10^{-2}	61	gas	6.5×10^{-3}
12	radioact	2.7×10^{-2}	62	suggest	6.1×10^{-3}
13	gamma	2.4×10^{-2}	63	fusion	6.1×10^{-3}
14	cool	2.3×10^{-2}	64	steel	6.1×10^{-3}
15	scintil	2.2×10^{-2}	65	pipe	6.1×10^{-3}
16	energi	2.1×10^{-2}	66	sourc	6×10^{-3}
17	mont	2.1×10^{-2}	67	decay	5.9×10^{-3}
18	carlo	2×10^{-2}	68	flow	5.9×10^{-3}
19	beam	2×10^{-2}	69	age	5.8×10^{-3}
20	fission	2×10^{-2}	70	isotop	5.7×10^{-3}
21	dose	1.8×10^{-2}	71	vessel	5.6×10^{-3}
22	thermal	1.7×10^{-2}	72	cfd	5.5×10^{-3}
23	calcul	1.6×10^{-2}	73	group	5.3×10^{-3}
24	heat	1.6×10^{-2}	74	design	5.3×10^{-3}
25	safeti	1.6×10^{-2}	75	molten	5.2×10^{-3}
26	kev	1.6×10^{-2}	76	iter	5.1×10^{-3}
27	ray	1.5×10^{-2}	77	resolut	5.1×10^{-3}
28	facil	1.3×10^{-2}	78	geometri	5×10^{-3}
29	power	1.3×10^{-2}	79	assembl	5×10^{-3}
30	temperatur	1.3×10^{-2}	80	instal	4.9×10^{-3}
31	conclus	1.2×10^{-2}	81	pet	4.9×10^{-3}
32	core	1.2×10^{-2}	82	transient	4.9×10^{-3}
33	hydraul	1.2×10^{-2}	83	clinic	4.6×10^{-3}
34	simul	1.1×10^{-2}	84	carri	4.6×10^{-3}
35	steam	1.1×10^{-2}	85	infect	4.5×10^{-3}
36	helium	1×10^{-2}	86	outcom	4.5×10^{-3}
37	patient	1×10^{-2}	87	wast	4.5×10^{-3}
38	ion	9.9×10^{-3}	88	perform	4.5×10^{-3}
39	materi	9.6×10^{-3}	89	spent	4.4×10^{-3}
40	flux	9.1×10^{-3}	90	cell	4.4×10^{-3}
41	plant	8.9×10^{-3}	91	electron	4.4×10^{-3}
42	diseas	8.9×10^{-3}	92	particip	4.3×10^{-3}
43	rod	8.7×10^{-3}	93	signific	4.3×10^{-3}
44	water	8.6×10^{-3}	94	experi	4.2×10^{-3}
45	oper	8.5×10^{-3}	95	social	4.1×10^{-3}
46	protein	8×10^{-3}	96	radiolog	4×10^{-3}
47	photon	8×10^{-3}	97	agreement	3.9×10^{-3}
48	experiment	7.8×10^{-3}	98	mediat	3.9×10^{-3}
49	measur	7.7×10^{-3}	99	wall	3.9×10^{-3}
50	tube	7.6×10^{-3}	100	liquid	3.8×10^{-3}

Table D.171. The list of the top 100 words in the category Nursing with RIGs

No.	Word	RIG
1	nurs	3.5×10^{-1}
2	care	1.5×10^{-1}
3	health	9.1×10^{-2}
4	educ	6.4×10^{-2}
5	practic	6.1×10^{-2}
6	particip	5.6×10^{-2}
7	interview	5.5×10^{-2}
8	hospit	5.1×10^{-2}
9	profession	4.9×10^{-2}
10	patient	4.9×10^{-2}
11	conclus	4.6×10^{-2}
12	intervent	4.3×10^{-2}
13	theme	3.8×10^{-2}
14	qualit	3.5×10^{-2}
15	clinic	3.3×10^{-2}
16	descript	3.2×10^{-2}
17	staff	3.2×10^{-2}
18	need	3.2×10^{-2}
19	purpos	3×10^{-2}
20	healthcar	3×10^{-2}
21	background	3×10^{-2}
22	questionnair	2.7×10^{-2}
23	support	2.6×10^{-2}
24	object	2.4×10^{-2}
25	student	2.4×10^{-2}
26	medic	2.4×10^{-2}
27	women	2.2×10^{-2}
28	semistructur	2.1×10^{-2}
29	aim	2.1×10^{-2}
30	percept	2.1×10^{-2}
31	explor	2×10^{-2}
32	research	2×10^{-2}
33	design	1.9×10^{-2}
34	skill	1.9×10^{-2}
35	ill	1.9×10^{-2}
36	outcom	1.8×10^{-2}
37	knowledg	1.8×10^{-2}
38	implic	1.8×10^{-2}
39	survey	1.8×10^{-2}
40	team	1.7×10^{-2}
41	manag	1.7×10^{-2}
42	perceiv	1.7×10^{-2}
43	mental	1.7×10^{-2}
44	life	1.6×10^{-2}
45	unit	1.6×10^{-2}
46	studi	1.6×10^{-2}
47	experienc	1.5×10^{-2}
48	famili	1.5×10^{-2}
49	find	1.5×10^{-2}
50	method	1.4×10^{-2}

No.	Word	RIG
51	qualiti	1.4×10^{-2}
52	satisfact	1.4×10^{-2}
53	person	1.4×10^{-2}
54	caregiv	1.4×10^{-2}
55	score	1.4×10^{-2}
56	experi	1.3×10^{-2}
57	propos	1.3×10^{-2}
58	themat	1.3×10^{-2}
59	practition	1.3×10^{-2}
60	cell	1.3×10^{-2}
61	home	1.3×10^{-2}
62	program	1.3×10^{-2}
63	collect	1.3×10^{-2}
64	show	1.2×10^{-2}
65	provid	1.2×10^{-2}
66	identifi	1.2×10^{-2}
67	symptom	1.2×10^{-2}
68	recommend	1.2×10^{-2}
69	feel	1.2×10^{-2}
70	articl	1.2×10^{-2}
71	recruit	1.1×10^{-2}
72	section	1.1×10^{-2}
73	emot	1.1×10^{-2}
74	pain	1.1×10^{-2}
75	paper	1.1×10^{-2}
76	teach	1.1×10^{-2}
77	temperatur	1.1×10^{-2}
78	profess	1.1×10^{-2}
79	attitud	1.1×10^{-2}
80	older	1.1×10^{-2}
81	learn	1×10^{-2}
82	surfac	1×10^{-2}
83	mother	1×10^{-2}
84	social	1×10^{-2}
85	depress	9.9×10^{-3}
86	compet	9.9×10^{-3}
87	understand	9.6×10^{-3}
88	distress	9.5×10^{-3}
89	paramet	9.5×10^{-3}
90	focus	9.3×10^{-3}
91	servic	9.2×10^{-3}
92	infant	9.2×10^{-3}
93	live	9.2×10^{-3}
94	psycholog	9.1×10^{-3}
95	assess	9.1×10^{-3}
96	conduct	9×10^{-3}
97	anxieti	9×10^{-3}
98	data	8.9×10^{-3}
99	set	8.9×10^{-3}
100	evid	8.8×10^{-3}

TABLE D.172. The list of the top 100 words in the category Nutrition and Dietetics with RIGs

No.	Word	RIG	No.	Word	RIG
1	intak	1×10^{-1}	51	signific	1.6×10^{-2}
2	dietari	1×10^{-1}	52	index	1.6×10^{-2}
3	food	9.4×10^{-2}	53	waist	1.6×10^{-2}
4	diet	8.5×10^{-2}	54	circumfer	1.6×10^{-2}
5	nutrit	8×10^{-2}	55	status	1.5×10^{-2}
6	obes	6.3×10^{-2}	56	aim	1.5×10^{-2}
7	fat	6×10^{-2}	57	triglycerid	1.5×10^{-2}
8	conclus	4.5×10^{-2}	58	increas	1.5×10^{-2}
9	bmi	3.8×10^{-2}	59	day	1.4×10^{-2}
10	age	3.8×10^{-2}	60	milk	1.4×10^{-2}
11	weight	3.5×10^{-2}	61	baselin	1.3×10^{-2}
12	bodi	3.5×10^{-2}	62	carbohydr	1.3×10^{-2}
13	eat	3.5×10^{-2}	63	questionnair	1.3×10^{-2}
14	supplement	3.3×10^{-2}	64	men	1.3×10^{-2}
15	object	3.3×10^{-2}	65	year	1.3×10^{-2}
16	fatti	3.2×10^{-2}	66	adult	1.3×10^{-2}
17	overweight	3.1×10^{-2}	67	nutrient	1.3×10^{-2}
18	vitamin	2.7×10^{-2}	68	lifestyl	1.3×10^{-2}
19	consumpt	2.6×10^{-2}	69	propos	1.3×10^{-2}
20	lipid	2.6×10^{-2}	70	blood	1.3×10^{-2}
21	metabol	2.5×10^{-2}	71	higher	1.3×10^{-2}
22	health	2.5×10^{-2}	72	veget	1.3×10^{-2}
23	consum	2.4×10^{-2}	73	lower	1.3×10^{-2}
24	acid	2.3×10^{-2}	74	protein	1.2×10^{-2}
25	women	2.3×10^{-2}	75	diabet	1.2×10^{-2}
26	particip	2.3×10^{-2}	76	regress	1.2×10^{-2}
27	associ	2.2×10^{-2}	77	sugar	1.1×10^{-2}
28	meal	2.2×10^{-2}	78	level	1.1×10^{-2}
29	healthi	2.2×10^{-2}	79	adjust	1.1×10^{-2}
30	paper	2.1×10^{-2}	80	random	1.1×10^{-2}
31	children	2.1×10^{-2}	81	simul	1.1×10^{-2}
32	total	2.1×10^{-2}	82	cardiovascular	1.1×10^{-2}
33	cholesterol	2×10^{-2}	83	plasma	1×10^{-2}
34	glucos	2×10^{-2}	84	content	1×10^{-2}
35	week	2×10^{-2}	85	lipoprotein	1×10^{-2}
36	studi	2×10^{-2}	86	trial	1×10^{-2}
37	insulin	2×10^{-2}	87	concentr	9.9×10^{-3}
38	assess	2×10^{-2}	88	hdl	9.6×10^{-3}
39	fruit	1.9×10^{-2}	89	infant	9.4×10^{-3}
40	antioxid	1.9×10^{-2}	90	effect	9.2×10^{-3}
41	adipos	1.8×10^{-2}	91	decreas	8.9×10^{-3}
42	risk	1.8×10^{-2}	92	daili	8.9×10^{-3}
43	subject	1.7×10^{-2}	93	system	8.9×10^{-3}
44	serum	1.7×10^{-2}	94	oil	8.9×10^{-3}
45	group	1.7×10^{-2}	95	meat	8.9×10^{-3}
46	intervent	1.7×10^{-2}	96	preval	8.6×10^{-3}
47	background	1.7×10^{-2}	97	phenol	8.2×10^{-3}
48	fed	1.6×10^{-2}	98	liver	8.1×10^{-3}
49	anthropometr	1.6×10^{-2}	99	placebo	8×10^{-3}
50	mass	1.6×10^{-2}	100	may	8×10^{-3}

TABLE D.173. The list of the top 100 words in the category Obstetrics and Gynecology with RIGs

No.	Word	RIG	No.	Word	RIG
1	women	2.1×10^{-1}	51	prenat	2×10^{-2}
2	pregnanc	1.6×10^{-1}	52	placenta	1.9×10^{-2}
3	conclus	1.3×10^{-1}	53	intervent	1.9×10^{-2}
4	gestat	1×10^{-1}	54	woman	1.9×10^{-2}
5	object	8.9×10^{-2}	55	method	1.9×10^{-2}
6	matern	7.9×10^{-2}	56	ultrasound	1.9×10^{-2}
7	birth	7.3×10^{-2}	57	care	1.9×10^{-2}
8	outcom	5.9×10^{-2}	58	contracept	1.9×10^{-2}
9	fetal	5.9×10^{-2}	59	underw	1.9×10^{-2}
10	vagin	4.8×10^{-2}	60	hormon	1.8×10^{-2}
11	obstetr	4.8×10^{-2}	61	year	1.8×10^{-2}
12	ovarian	4.7×10^{-2}	62	complic	1.8×10^{-2}
13	uterin	4.7×10^{-2}	63	interv	1.7×10^{-2}
14	pregnant	4.6×10^{-2}	64	surgeri	1.6×10^{-2}
15	patient	4.5×10^{-2}	65	confid	1.6×10^{-2}
16	deliveri	4.4×10^{-2}	66	propos	1.6×10^{-2}
17	age	4.3×10^{-2}	67	ovari	1.6×10^{-2}
18	preterm	4.2×10^{-2}	68	odd	1.5×10^{-2}
19	neonat	4.2×10^{-2}	69	cancer	1.5×10^{-2}
20	risk	3.9×10^{-2}	70	oocyt	1.5×10^{-2}
21	week	3.7×10^{-2}	71	diagnosi	1.5×10^{-2}
22	singleton	3.7×10^{-2}	72	regress	1.4×10^{-2}
23	gynecolog	3.5×10^{-2}	73	assess	1.4×10^{-2}
24	retrospect	3.4×10^{-2}	74	diagnos	1.4×10^{-2}
25	pelvic	3.2×10^{-2}	75	newborn	1.4×10^{-2}
26	endometri	3.1×10^{-2}	76	medic	1.4×10^{-2}
27	trimest	3×10^{-2}	77	postmenopaus	1.4×10^{-2}
28	infant	3×10^{-2}	78	logist	1.4×10^{-2}
29	preeclampsia	3×10^{-2}	79	compar	1.4×10^{-2}
30	cohort	3×10^{-2}	80	health	1.4×10^{-2}
31	hospit	2.9×10^{-2}	81	month	1.4×10^{-2}
32	paper	2.9×10^{-2}	82	surgic	1.3×10^{-2}
33	clinic	2.8×10^{-2}	83	embryo	1.3×10^{-2}
34	fetus	2.8×10^{-2}	84	median	1.3×10^{-2}
35	placent	2.7×10^{-2}	85	includ	1.3×10^{-2}
36	intrauterin	2.6×10^{-2}	86	simul	1.3×10^{-2}
37	infertil	2.6×10^{-2}	87	abnorm	1.2×10^{-2}
38	associ	2.6×10^{-2}	88	case	1.2×10^{-2}
39	ivf	2.5×10^{-2}	89	design	1.2×10^{-2}
40	group	2.5×10^{-2}	90	syndrom	1.2×10^{-2}
41	postpartum	2.4×10^{-2}	91	labor	1.2×10^{-2}
42	studi	2.4×10^{-2}	92	properti	1.2×10^{-2}
43	perinat	2.3×10^{-2}	93	rate	1.2×10^{-2}
44	prospect	2.3×10^{-2}	94	recurr	1.2×10^{-2}
45	result	2.3×10^{-2}	95	laparoscop	1.2×10^{-2}
46	cervic	2.2×10^{-2}	96	aim	1.1×10^{-2}
47	fertil	2.2×10^{-2}	97	treatment	1.1×10^{-2}
48	signific	2.1×10^{-2}	98	temperatur	1.1×10^{-2}
49	mother	2.1×10^{-2}	99	incontin	1.1×10^{-2}
50	reproduct	2.1×10^{-2}	100	sperm	1.1×10^{-2}

TABLE D.174. The list of the top 100 words in the category Oceanography with RIGs

No.	Word	RIG
1	sea	1.5×10^{-1}
2	ocean	1.1×10^{-1}
3	coastal	6.7×10^{-2}
4	water	6.3×10^{-2}
5	marin	5.5×10^{-2}
6	sediment	4.3×10^{-2}
7	coast	4.3×10^{-2}
8	depth	3.8×10^{-2}
9	atlant	3.4×10^{-2}
10	phytoplankton	3.3×10^{-2}
11	season	3.3×10^{-2}
12	shelf	3.2×10^{-2}
13	offshor	3.1×10^{-2}
14	summer	2.9×10^{-2}
15	bay	2.8×10^{-2}
16	fish	2.8×10^{-2}
17	bottom	2.6×10^{-2}
18	salin	2.5×10^{-2}
19	north	2.5×10^{-2}
20	tidal	2.5×10^{-2}
21	wind	2.5×10^{-2}
22	pacif	2.4×10^{-2}
23	ecosystem	2.3×10^{-2}
24	fisheri	2.3×10^{-2}
25	underwat	2.3×10^{-2}
26	southern	2.3×10^{-2}
27	benthic	2.2×10^{-2}
28	estuari	2.2×10^{-2}
29	abund	2.2×10^{-2}
30	tide	2.1×10^{-2}
31	shallow	2.1×10^{-2}
32	gulf	2.1×10^{-2}
33	winter	2×10^{-2}
34	speci	2×10^{-2}
35	wave	1.9×10^{-2}
36	deep	1.9×10^{-2}
37	patient	1.9×10^{-2}
38	south	1.9×10^{-2}
39	chlorophyl	1.9×10^{-2}
40	surfac	1.9×10^{-2}
41	vertic	1.9×10^{-2}
42	eddi	1.8×10^{-2}
43	spatial	1.8×10^{-2}
44	upwel	1.8×10^{-2}
45	river	1.7×10^{-2}
46	spring	1.7×10^{-2}
47	bloom	1.7×10^{-2}
48	along	1.7×10^{-2}
49	northern	1.7×10^{-2}
50	domin	1.7×10^{-2}

No.	Word	RIG
51	area	1.6×10^{-2}
52	basin	1.6×10^{-2}
53	continent	1.6×10^{-2}
54	conclus	1.5×10^{-2}
55	variabl	1.5×10^{-2}
56	region	1.5×10^{-2}
57	interannu	1.5×10^{-2}
58	pelag	1.5×10^{-2}
59	reef	1.5×10^{-2}
60	estuarin	1.5×10^{-2}
61	habitat	1.5×10^{-2}
62	zooplankton	1.5×10^{-2}
63	eastern	1.4×10^{-2}
64	east	1.4×10^{-2}
65	circul	1.4×10^{-2}
66	ice	1.4×10^{-2}
67	climat	1.4×10^{-2}
68	plankton	1.4×10^{-2}
69	trophic	1.4×10^{-2}
70	nutrient	1.4×10^{-2}
71	satellit	1.4×10^{-2}
72	shore	1.4×10^{-2}
73	dure	1.3×10^{-2}
74	period	1.3×10^{-2}
75	zone	1.3×10^{-2}
76	advect	1.3×10^{-2}
77	flux	1.3×10^{-2}
78	dissolv	1.3×10^{-2}
79	island	1.3×10^{-2}
80	coral	1.3×10^{-2}
81	warm	1.3×10^{-2}
82	moor	1.2×10^{-2}
83	slope	1.2×10^{-2}
84	arctic	1.2×10^{-2}
85	freshwat	1.2×10^{-2}
86	seawat	1.2×10^{-2}
87	hydrodynam	1.2×10^{-2}
88	biomass	1.2×10^{-2}
89	variat	1.2×10^{-2}
90	seafloor	1.2×10^{-2}
91	observ	1.2×10^{-2}
92	clinic	1.2×10^{-2}
93	station	1.2×10^{-2}
94	mediterranean	1.1×10^{-2}
95	ecolog	1.1×10^{-2}
96	atmospher	1.1×10^{-2}
97	diatom	1.1×10^{-2}
98	tempor	1.1×10^{-2}
99	near	1.1×10^{-2}
100	assemblag	1.1×10^{-2}

Table D.175. The list of the top 100 words in the category Oncology with RIGs

No.	Word	RIG	No.	Word	RIG
1	cancer	3.2×10^{-1}	51	nonsmal	2.3×10^{-2}
2	tumor	1.9×10^{-1}	52	multivari	2.3×10^{-2}
3	patient	1.4×10^{-1}	53	prostat	2.3×10^{-2}
4	surviv	1×10^{-1}	54	squamous	2.3×10^{-2}
5	cell	1×10^{-1}	55	risk	2.3×10^{-2}
6	carcinoma	8×10^{-2}	56	clinicopatholog	2.3×10^{-2}
7	chemotherapi	7.2×10^{-2}	57	immunohistochemistri	2.3×10^{-2}
8	breast	7.1×10^{-2}	58	kinas	2.2×10^{-2}
9	express	6.6×10^{-2}	59	growth	2.2×10^{-2}
10	treatment	6.4×10^{-2}	60	dose	2.2×10^{-2}
11	conclus	5.8×10^{-2}	61	adenocarcinoma	2.2×10^{-2}
12	metastasi	5.8×10^{-2}	62	pathway	2.2×10^{-2}
13	therapi	5.8×10^{-2}	63	resect	2.1×10^{-2}
14	prognost	5.2×10^{-2}	64	oncolog	2.1×10^{-2}
15	malign	5.2×10^{-2}	65	nsclc	2.1×10^{-2}
16	clinic	4.9×10^{-2}	66	studi	2×10^{-2}
17	metastat	4.7×10^{-2}	67	month	2×10^{-2}
18	progress	4.4×10^{-2}	68	downregul	2×10^{-2}
19	prognosi	4.4×10^{-2}	69	primari	1.9×10^{-2}
20	associ	4.2×10^{-2}	70	oncogen	1.9×10^{-2}
21	invas	4×10^{-2}	71	diagnosi	1.9×10^{-2}
22	paper	4×10^{-2}	72	adjuv	1.9×10^{-2}
23	median	4×10^{-2}	73	histolog	1.9×10^{-2}
24	lung	3.7×10^{-2}	74	may	1.9×10^{-2}
25	prolifer	3.7×10^{-2}	75	leukemia	1.8×10^{-2}
26	overal	3.5×10^{-2}	76	propos	1.8×10^{-2}
27	inhibit	3.4×10^{-2}	77	upregul	1.8×10^{-2}
28	treat	3.3×10^{-2}	78	human	1.8×10^{-2}
29	signific	3.3×10^{-2}	79	receiv	1.8×10^{-2}
30	apoptosi	3.3×10^{-2}	80	cox	1.8×10^{-2}
31	radiotherapi	3.2×10^{-2}	81	antitumor	1.7×10^{-2}
32	diseas	3.1×10^{-2}	82	poor	1.7×10^{-2}
33	overexpress	3.1×10^{-2}	83	regul	1.7×10^{-2}
34	therapeut	3×10^{-2}	84	assay	1.7×10^{-2}
35	target	3×10^{-2}	85	blot	1.7×10^{-2}
36	tissu	3×10^{-2}	86	biomark	1.7×10^{-2}
37	grade	3×10^{-2}	87	structur	1.6×10^{-2}
38	lymph	2.9×10^{-2}	88	outcom	1.6×10^{-2}
39	background	2.9×10^{-2}	89	marker	1.6×10^{-2}
40	recurr	2.9×10^{-2}	90	vitro	1.6×10^{-2}
41	colorect	2.8×10^{-2}	91	diagnos	1.6×10^{-2}
42	inhibitor	2.8×10^{-2}	92	suppressor	1.6×10^{-2}
43	gene	2.8×10^{-2}	93	status	1.6×10^{-2}
44	metastas	2.7×10^{-2}	94	retrospect	1.6×10^{-2}
45	protein	2.6×10^{-2}	95	tumour	1.6×10^{-2}
46	line	2.5×10^{-2}	96	year	1.6×10^{-2}
47	xenograft	2.4×10^{-2}	97	epitheli	1.6×10^{-2}
48	stage	2.4×10^{-2}	98	vivo	1.5×10^{-2}
49	factor	2.4×10^{-2}	99	lymphoma	1.5×10^{-2}
50	receptor	2.4×10^{-2}	100	temperatur	1.5×10^{-2}

TABLE D.176. The list of the top 100 words in the category Operations Research and Management Science with RIGs

No.	Word	RIG	No.	Word	RIG
1	problem	6.4×10^{-2}	51	temperatur	8.3×10^{-3}
2	paper	5×10^{-2}	52	concentr	8.2×10^{-3}
3	optim	3.7×10^{-2}	53	base	8×10^{-3}
4	propos	3.1×10^{-2}	54	background	7.7×10^{-3}
5	algorithm	3×10^{-2}	55	convex	7.5×10^{-3}
6	solv	2.8×10^{-2}	56	polici	7.4×10^{-3}
7	decis	2.6×10^{-2}	57	speci	7.4×10^{-3}
8	heurist	2.1×10^{-2}	58	observ	7.3×10^{-3}
9	cost	1.9×10^{-2}	59	make	7.2×10^{-3}
10	model	1.9×10^{-2}	60	gene	7.2×10^{-3}
11	integ	1.7×10^{-2}	61	search	7.2×10^{-3}
12	custom	1.6×10^{-2}	62	uncertainti	7.2×10^{-3}
13	constraint	1.5×10^{-2}	63	oper	7.1×10^{-3}
14	conclus	1.5×10^{-2}	64	surfac	7×10^{-3}
15	schedul	1.5×10^{-2}	65	bound	6.9×10^{-3}
16	cell	1.4×10^{-2}	66	servic	6.8×10^{-3}
17	solut	1.4×10^{-2}	67	oxid	6.8×10^{-3}
18	demand	1.4×10^{-2}	68	report	6.8×10^{-3}
19	instanc	1.4×10^{-2}	69	competit	6.7×10^{-3}
20	price	1.4×10^{-2}	70	multiobject	6.7×10^{-3}
21	patient	1.3×10^{-2}	71	age	6.6×10^{-3}
22	compani	1.3×10^{-2}	72	literatur	6.6×10^{-3}
23	suppli	1.3×10^{-2}	73	molecular	6.5×10^{-3}
24	manufactur	1.3×10^{-2}	74	busi	6.4×10^{-3}
25	program	1.2×10^{-2}	75	induc	6.3×10^{-3}
26	stochast	1.2×10^{-2}	76	found	6.3×10^{-3}
27	supplier	1.2×10^{-2}	77	introduc	6.3×10^{-3}
28	profit	1.2×10^{-2}	78	activ	6.2×10^{-3}
29	comput	1.2×10^{-2}	79	product	6.2×10^{-3}
30	exampl	1.2×10^{-2}	80	signific	6.1×10^{-3}
31	market	1.2×10^{-2}	81	real	6.1×10^{-3}
32	consid	1.1×10^{-2}	82	chain	6.1×10^{-3}
33	formul	1.1×10^{-2}	83	reaction	6×10^{-3}
34	approach	1.1×10^{-2}	84	group	5.9×10^{-3}
35	manag	1.1×10^{-2}	85	network	5.9×10^{-3}
36	set	1.1×10^{-2}	86	given	5.8×10^{-3}
37	inventori	1.1×10^{-2}	87	job	5.8×10^{-3}
38	firm	1×10^{-2}	88	system	5.7×10^{-3}
39	illustr	1×10^{-2}	89	machin	5.6×10^{-3}
40	treatment	1×10^{-2}	90	inhibit	5.5×10^{-3}
41	protein	1×10^{-2}	91	order	5.4×10^{-3}
42	minim	1×10^{-2}	92	tissu	5.4×10^{-3}
43	numer	9.8×10^{-3}	93	fuzzi	5.4×10^{-3}
44	enterpris	9.6×10^{-3}	94	ray	5.4×10^{-3}
45	retail	9.4×10^{-3}	95	high	5.3×10^{-3}
46	clinic	8.8×10^{-3}	96	converg	5.3×10^{-3}
47	diseas	8.7×10^{-3}	97	decreas	5.3×10^{-3}
48	acid	8.5×10^{-3}	98	water	5.2×10^{-3}
49	industri	8.5×10^{-3}	99	game	5.1×10^{-3}
50	plan	8.4×10^{-3}	100	alloc	5.1×10^{-3}

TABLE D.177. The list of the top 100 words in the category Ophthalmology with RIGs

No.	Word	RIG	No.	Word	RIG
1	eye	3.3×10^{-1}	51	measur	1.9×10^{-2}
2	retin	1.7×10^{-1}	52	epithelium	1.8×10^{-2}
3	purpos	1.6×10^{-1}	53	tear	1.8×10^{-2}
4	acuiti	1.5×10^{-1}	54	evalu	1.6×10^{-2}
5	conclus	1.5×10^{-1}	55	central	1.6×10^{-2}
6	corneal	1.3×10^{-1}	56	consecut	1.6×10^{-2}
7	macular	1.2×10^{-1}	57	propos	1.5×10^{-2}
8	visual	1.1×10^{-1}	58	surgic	1.5×10^{-2}
9	ocular	1.1×10^{-1}	59	bilater	1.4×10^{-2}
10	intraocular	1×10^{-1}	60	treatment	1.4×10^{-2}
11	patient	8.6×10^{-2}	61	subject	1.4×10^{-2}
12	glaucoma	7.9×10^{-2}	62	ganglion	1.4×10^{-2}
13	method	6.4×10^{-2}	63	examin	1.4×10^{-2}
14	oct	6×10^{-2}	64	inject	1.3×10^{-2}
15	cataract	5.9×10^{-2}	65	chamber	1.3×10^{-2}
16	coher	5.6×10^{-2}	66	group	1.2×10^{-2}
17	retina	5.4×10^{-2}	67	treat	1.2×10^{-2}
18	tomographi	5.1×10^{-2}	68	epitheli	1.2×10^{-2}
19	cornea	4.9×10^{-2}	69	particip	1.2×10^{-2}
20	choroid	4.8×10^{-2}	70	fellow	1.1×10^{-2}
21	optic	4.2×10^{-2}	71	includ	1.1×10^{-2}
22	len	4.1×10^{-2}	72	baselin	1.1×10^{-2}
23	month	4×10^{-2}	73	visit	1.1×10^{-2}
24	vision	4×10^{-2}	74	temperatur	1.1×10^{-2}
25	surgeri	3.9×10^{-2}	75	diseas	1×10^{-2}
26	age	3.8×10^{-2}	76	implant	1×10^{-2}
27	result	3.8×10^{-2}	77	normal	9.9×10^{-3}
28	anterior	3.7×10^{-2}	78	review	9.9×10^{-3}
29	retrospect	3.5×10^{-2}	79	statist	9.9×10^{-3}
30	postop	3.3×10^{-2}	80	imag	9.8×10^{-3}
31	mean	3.2×10^{-2}	81	complic	9.7×10^{-3}
32	underw	3.1×10^{-2}	82	spectral	9.6×10^{-3}
33	paper	3×10^{-2}	83	medic	9.6×10^{-3}
34	degener	2.9×10^{-2}	84	laser	9.6×10^{-3}
35	refract	2.9×10^{-2}	85	angiographi	9.5×10^{-3}
36	thick	2.7×10^{-2}	86	diabet	9.5×10^{-3}
37	clinic	2.6×10^{-2}	87	week	9.4×10^{-3}
38	signific	2.6×10^{-2}	88	segment	9.4×10^{-3}
39	outcom	2.5×10^{-2}	89	disc	9.4×10^{-3}
40	year	2.4×10^{-2}	90	seri	9.3×10^{-3}
41	nerv	2.2×10^{-2}	91	old	9.2×10^{-3}
42	follow	2.2×10^{-2}	92	unilater	9×10^{-3}
43	posterior	2.2×10^{-2}	93	associ	8.8×10^{-3}
44	correct	2.1×10^{-2}	94	spectacl	8.7×10^{-3}
45	endotheli	2.1×10^{-2}	95	simul	8.6×10^{-3}
46	edema	2×10^{-2}	96	energi	8.6×10^{-3}
47	preoper	2×10^{-2}	97	compar	8.3×10^{-3}
48	prospect	2×10^{-2}	98	photographi	8.2×10^{-3}
49	pigment	1.9×10^{-2}	99	inferior	8×10^{-3}
50	case	1.9×10^{-2}	100	photograph	7.9×10^{-3}

Table D.178. The list of the top 100 words in the category Optics with RIGs

No.	Word	RIG	No.	Word	RIG
1	optic	1.5×10^{-1}	51	risk	9.7×10^{-3}
2	laser	7.6×10^{-2}	52	band	9.5×10^{-3}
3	wavelength	5.4×10^{-2}	53	telescop	9.4×10^{-3}
4	photon	4.2×10^{-2}	54	assess	9.3×10^{-3}
5	fiber	3.8×10^{-2}	55	silicon	9.2×10^{-3}
6	light	2.9×10^{-2}	56	aim	9.1×10^{-3}
7	beam	2.8×10^{-2}	57	propag	8.9×10^{-3}
8	conclus	2.5×10^{-2}	58	may	8.9×10^{-3}
9	grate	2.4×10^{-2}	59	diseas	8.9×10^{-3}
10	puls	2.4×10^{-2}	60	emit	8.7×10^{-3}
11	waveguid	2.3×10^{-2}	61	sensor	8.6×10^{-3}
12	imag	2.3×10^{-2}	62	frequenc	8.5×10^{-3}
13	studi	2.2×10^{-2}	63	illumin	8.4×10^{-3}
14	spectral	2×10^{-2}	64	crystal	8.3×10^{-3}
15	mode	1.9×10^{-2}	65	detector	8.3×10^{-3}
16	patient	1.8×10^{-2}	66	identifi	8.3×10^{-3}
17	refract	1.8×10^{-2}	67	infrar	8.1×10^{-3}
18	pump	1.6×10^{-2}	68	intens	8×10^{-3}
19	quantum	1.6×10^{-2}	69	examin	8×10^{-3}
20	coher	1.6×10^{-2}	70	apertur	7.9×10^{-3}
21	wave	1.6×10^{-2}	71	len	7.8×10^{-3}
22	polar	1.5×10^{-2}	72	femtosecond	7.7×10^{-3}
23	associ	1.5×10^{-2}	73	multiplex	7.7×10^{-3}
24	suggest	1.4×10^{-2}	74	outcom	7.6×10^{-3}
25	signific	1.3×10^{-2}	75	clinic	7.6×10^{-3}
26	bandwidth	1.3×10^{-2}	76	follow	7.5×10^{-3}
27	age	1.3×10^{-2}	77	dope	7.5×10^{-3}
28	mirror	1.2×10^{-2}	78	particip	7.4×10^{-3}
29	interferomet	1.2×10^{-2}	79	field	7.3×10^{-3}
30	fabric	1.2×10^{-2}	80	power	7.3×10^{-3}
31	caviti	1.2×10^{-2}	81	regul	7.3×10^{-3}
32	plasmon	1.2×10^{-2}	82	realiz	7.2×10^{-3}
33	diod	1.2×10^{-2}	83	pixel	7.1×10^{-3}
34	treatment	1.1×10^{-2}	84	ghz	7×10^{-3}
35	experiment	1.1×10^{-2}	85	among	7×10^{-3}
36	demonstr	1.1×10^{-2}	86	camera	7×10^{-3}
37	excit	1.1×10^{-2}	87	array	7×10^{-3}
38	tunabl	1.1×10^{-2}	88	dure	7×10^{-3}
39	gene	1.1×10^{-2}	89	transmiss	6.9×10^{-3}
40	reson	1.1×10^{-2}	90	role	6.8×10^{-3}
41	year	1×10^{-2}	91	speci	6.8×10^{-3}
42	group	1×10^{-2}	92	nonlinear	6.8×10^{-3}
43	absorpt	1×10^{-2}	93	radiat	6.7×10^{-3}
44	nois	1×10^{-2}	94	express	6.6×10^{-3}
45	protein	1×10^{-2}	95	broadband	6.6×10^{-3}
46	devic	1×10^{-2}	96	day	6.6×10^{-3}
47	scatter	1×10^{-2}	97	acid	6.5×10^{-3}
48	resolut	9.8×10^{-3}	98	achiev	6.5×10^{-3}
49	modul	9.7×10^{-3}	99	singl	6.3×10^{-3}
50	activ	9.7×10^{-3}	100	filter	6.3×10^{-3}

Table D.179. The list of the top 100 words in the category Ornithology with RIGs

No.	Word	RIG
1	bird	2.8×10^{-1}
2	breed	1.9×10^{-1}
3	speci	1.5×10^{-1}
4	nest	1.3×10^{-1}
5	habitat	1.3×10^{-1}
6	nestl	7×10^{-2}
7	season	5.8×10^{-2}
8	popul	5.6×10^{-2}
9	brood	5.4×10^{-2}
10	avian	5.2×10^{-2}
11	forag	5.2×10^{-2}
12	chick	4.3×10^{-2}
13	migratori	4.2×10^{-2}
14	winter	3.9×10^{-2}
15	conserv	3.9×10^{-2}
16	clutch	3.8×10^{-2}
17	prey	3.6×10^{-2}
18	reproduct	3.6×10^{-2}
19	egg	3.5×10^{-2}
20	site	3.4×10^{-2}
21	predat	3.2×10^{-2}
22	area	3.2×10^{-2}
23	territori	3.2×10^{-2}
24	ecolog	3.2×10^{-2}
25	abund	3.1×10^{-2}
26	forest	3×10^{-2}
27	male	2.9×10^{-2}
28	individu	2.9×10^{-2}
29	island	2.7×10^{-2}
30	femal	2.7×10^{-2}
31	north	2.6×10^{-2}
32	hatch	2.6×10^{-2}
33	food	2.6×10^{-2}
34	endang	2.5×10^{-2}
35	northern	2.5×10^{-2}
36	wing	2.4×10^{-2}
37	may	2.4×10^{-2}
38	dure	2.3×10^{-2}
39	suggest	2.2×10^{-2}
40	declin	2.2×10^{-2}
41	south	2.1×10^{-2}
42	coloni	2×10^{-2}
43	migrat	2×10^{-2}
44	migrant	2×10^{-2}
45	capsul	1.9×10^{-2}
46	adult	1.8×10^{-2}
47	record	1.8×10^{-2}
48	juvenil	1.8×10^{-2}
49	eastern	1.7×10^{-2}
50	paper	1.7×10^{-2}
51	landscap	1.7×10^{-2}
52	southern	1.7×10^{-2}
53	endem	1.6×10^{-2}
54	variat	1.5×10^{-2}
55	patient	1.5×10^{-2}
56	mate	1.5×10^{-2}
57	sex	1.5×10^{-2}
58	distanc	1.5×10^{-2}
59	size	1.4×10^{-2}
60	period	1.4×10^{-2}
61	year	1.4×10^{-2}
62	america	1.4×10^{-2}
63	wetland	1.4×10^{-2}
64	grassland	1.3×10^{-2}
65	lay	1.3×10^{-2}
66	annual	1.3×10^{-2}
67	black	1.3×10^{-2}
68	vocal	1.3×10^{-2}
69	surviv	1.3×10^{-2}
70	threaten	1.3×10^{-2}
71	tree	1.3×10^{-2}
72	success	1.3×10^{-2}
73	veget	1.3×10^{-2}
74	coast	1.2×10^{-2}
75	song	1.2×10^{-2}
76	pair	1.2×10^{-2}
77	survey	1.1×10^{-2}
78	flight	1.1×10^{-2}
79	pattern	1.1×10^{-2}
80	bodi	1.1×10^{-2}
81	geograph	1.1×10^{-2}
82	feed	1.1×10^{-2}
83	argentina	1.1×10^{-2}
84	hypothesi	1×10^{-2}
85	probabl	1×10^{-2}
86	applic	1×10^{-2}
87	autumn	1×10^{-2}
88	africa	1×10^{-2}
89	ground	1×10^{-2}
90	white	9.9×10^{-3}
91	mark	9.8×10^{-3}
92	spring	9.8×10^{-3}
93	littl	9.7×10^{-3}
94	wildlif	9.5×10^{-3}
95	west	9.5×10^{-3}
96	cell	9.4×10^{-3}
97	within	9.3×10^{-3}
98	parent	9.3×10^{-3}
99	method	9.2×10^{-3}
100	lake	9.2×10^{-3}

TABLE D.180. The list of the top 100 words in the category Orthopedics with RIGs

No.	Word	RIG	No.	Word	RIG
1	patient	1.3×10^{-1}	51	total	2.8×10^{-2}
2	knee	1.2×10^{-1}	52	implant	2.8×10^{-2}
3	arthroplasti	1×10^{-1}	53	articular	2.7×10^{-2}
4	surgeri	8.5×10^{-2}	54	biomechan	2.7×10^{-2}
5	radiograph	8.4×10^{-2}	55	studi	2.7×10^{-2}
6	postop	7.9×10^{-2}	56	method	2.7×10^{-2}
7	pain	7.8×10^{-2}	57	later	2.6×10^{-2}
8	hip	7.4×10^{-2}	58	distal	2.6×10^{-2}
9	conclus	6.6×10^{-2}	59	foot	2.5×10^{-2}
10	surgic	6.6×10^{-2}	60	group	2.5×10^{-2}
11	fractur	6.2×10^{-2}	61	signific	2.5×10^{-2}
12	femor	6.2×10^{-2}	62	treat	2.5×10^{-2}
13	background	6.1×10^{-2}	63	radiolog	2.3×10^{-2}
14	score	5.8×10^{-2}	64	review	2.2×10^{-2}
15	outcom	5.8×10^{-2}	65	treatment	2.2×10^{-2}
16	fixat	5.6×10^{-2}	66	cartilag	2.2×10^{-2}
17	clinic	5.5×10^{-2}	67	sagitt	2.2×10^{-2}
18	bone	5.1×10^{-2}	68	femur	2.2×10^{-2}
19	preoper	5.1×10^{-2}	69	assess	2.1×10^{-2}
20	anterior	5×10^{-2}	70	trauma	2.1×10^{-2}
21	year	5×10^{-2}	71	prospect	2×10^{-2}
22	flexion	4.7×10^{-2}	72	evalu	2×10^{-2}
23	osteoarthr	4.7×10^{-2}	73	motion	1.9×10^{-2}
24	joint	4.6×10^{-2}	74	limb	1.9×10^{-2}
25	ligament	4.5×10^{-2}	75	procedur	1.9×10^{-2}
26	injuri	4.4×10^{-2}	76	cadaver	1.8×10^{-2}
27	surgeon	4.4×10^{-2}	77	elbow	1.8×10^{-2}
28	underw	4.3×10^{-2}	78	reconstruct	1.8×10^{-2}
29	tibial	4.2×10^{-2}	79	proxim	1.8×10^{-2}
30	purpos	4.2×10^{-2}	80	heal	1.8×10^{-2}
31	follow	4×10^{-2}	81	summari	1.6×10^{-2}
32	posterior	4×10^{-2}	82	rotat	1.6×10^{-2}
33	month	4×10^{-2}	83	disloc	1.6×10^{-2}
34	retrospect	4×10^{-2}	84	consecut	1.6×10^{-2}
35	medial	3.9×10^{-2}	85	anatom	1.6×10^{-2}
36	shoulder	3.8×10^{-2}	86	tear	1.6×10^{-2}
37	complic	3.7×10^{-2}	87	disabl	1.6×10^{-2}
38	mean	3.6×10^{-2}	88	degen	1.6×10^{-2}
39	spine	3.6×10^{-2}	89	result	1.5×10^{-2}
40	tendon	3.5×10^{-2}	90	compar	1.5×10^{-2}
41	age	3.5×10^{-2}	91	vas	1.5×10^{-2}
42	orthopaed	3.4×10^{-2}	92	intraop	1.5×10^{-2}
43	ankl	3.4×10^{-2}	93	propos	1.4×10^{-2}
44	arthroscop	3.1×10^{-2}	94	repair	1.4×10^{-2}
45	lumbar	3.1×10^{-2}	95	fusion	1.4×10^{-2}
46	revis	3.1×10^{-2}	96	prosthesi	1.4×10^{-2}
47	cruciat	3×10^{-2}	97	evid	1.3×10^{-2}
48	screw	2.9×10^{-2}	98	gait	1.3×10^{-2}
49	paper	2.9×10^{-2}	99	case	1.3×10^{-2}
50	spinal	2.9×10^{-2}	100	cohort	1.3×10^{-2}

Table D.181. The list of the top 100 words in the category Otorhinolaryngology with RIGs

No.	Word	RIG
1	conclus	1.4×10^{-1}
2	hear	1.3×10^{-1}
3	patient	1.2×10^{-1}
4	object	1.1×10^{-1}
5	ear	8.2×10^{-2}
6	surgeri	6.6×10^{-2}
7	cochlear	6.2×10^{-2}
8	retrospect	5.4×10^{-2}
9	nasal	5.2×10^{-2}
10	surgic	5.1×10^{-2}
11	neck	4.7×10^{-2}
12	postop	4.5×10^{-2}
13	auditori	4.3×10^{-2}
14	underw	4.1×10^{-2}
15	laryng	4×10^{-2}
16	endoscop	3.9×10^{-2}
17	tertiari	3.8×10^{-2}
18	result	3.7×10^{-2}
19	method	3.6×10^{-2}
20	outcom	3.4×10^{-2}
21	sinus	3.4×10^{-2}
22	speech	3.3×10^{-2}
23	year	3×10^{-2}
24	review	2.9×10^{-2}
25	clinic	2.8×10^{-2}
26	studi	2.8×10^{-2}
27	vocal	2.8×10^{-2}
28	head	2.6×10^{-2}
29	preoper	2.6×10^{-2}
30	canal	2.5×10^{-2}
31	month	2.4×10^{-2}
32	subject	2.4×10^{-2}
33	paper	2.4×10^{-2}
34	hypothesi	2.3×10^{-2}
35	signific	2.3×10^{-2}
36	age	2.3×10^{-2}
37	recurr	2.2×10^{-2}
38	prospect	2.2×10^{-2}
39	tone	2.2×10^{-2}
40	design	2.2×10^{-2}
41	implant	2×10^{-2}
42	case	2×10^{-2}
43	score	2×10^{-2}
44	symptom	1.9×10^{-2}
45	referr	1.9×10^{-2}
46	voic	1.9×10^{-2}
47	airway	1.8×10^{-2}
48	obstruct	1.8×10^{-2}
49	bilater	1.8×10^{-2}
50	children	1.8×10^{-2}

No.	Word	RIG
51	unilater	1.8×10^{-2}
52	group	1.8×10^{-2}
53	treatment	1.8×10^{-2}
54	evalu	1.7×10^{-2}
55	squamous	1.7×10^{-2}
56	carcinoma	1.7×10^{-2}
57	nerv	1.7×10^{-2}
58	follow	1.6×10^{-2}
59	complic	1.6×10^{-2}
60	bone	1.5×10^{-2}
61	diagnosi	1.5×10^{-2}
62	pediatr	1.4×10^{-2}
63	medic	1.4×10^{-2}
64	chart	1.4×10^{-2}
65	hospit	1.4×10^{-2}
66	loss	1.4×10^{-2}
67	diseas	1.3×10^{-2}
68	chronic	1.3×10^{-2}
69	facial	1.3×10^{-2}
70	treat	1.3×10^{-2}
71	listen	1.2×10^{-2}
72	record	1.2×10^{-2}
73	assess	1.2×10^{-2}
74	threshold	1.2×10^{-2}
75	propos	1.2×10^{-2}
76	diagnos	1.2×10^{-2}
77	resect	1.2×10^{-2}
78	mean	1.1×10^{-2}
79	rhiniti	1.1×10^{-2}
80	middl	1.1×10^{-2}
81	background	1×10^{-2}
82	acoust	1×10^{-2}
83	adult	1×10^{-2}
84	normal	1×10^{-2}
85	temperatur	1×10^{-2}
86	dissect	1×10^{-2}
87	intervent	1×10^{-2}
88	statist	9.9×10^{-3}
89	flap	9.9×10^{-3}
90	surgeon	9.9×10^{-3}
91	undergo	9.7×10^{-3}
92	endoscopi	9.7×10^{-3}
93	procedur	9.2×10^{-3}
94	cohort	9.1×10^{-3}
95	sleep	9.1×10^{-3}
96	center	9.1×10^{-3}
97	patholog	9×10^{-3}
98	sound	8.8×10^{-3}
99	intraop	8.7×10^{-3}
100	mucosa	8.7×10^{-3}

TABLE D.182. The list of the top 100 words in the category Paleontology with RIGs

No.	Word	RIG
1	fossil	1.4×10^{-1}
2	late	9.1×10^{-2}
3	assemblag	8.7×10^{-2}
4	cretac	7.7×10^{-2}
5	taxa	7.3×10^{-2}
6	speci	7×10^{-2}
7	sediment	6.3×10^{-2}
8	record	6×10^{-2}
9	stratigraph	6×10^{-2}
10	fauna	6×10^{-2}
11	middl	5.8×10^{-2}
12	basin	5.6×10^{-2}
13	genus	5.3×10^{-2}
14	upper	5.3×10^{-2}
15	nov	5.3×10^{-2}
16	earli	5.2×10^{-2}
17	preserv	5.1×10^{-2}
18	marin	4.7×10^{-2}
19	miocen	4.4×10^{-2}
20	specimen	4.4×10^{-2}
21	deposit	4.3×10^{-2}
22	jurass	4.2×10^{-2}
23	format	3.7×10^{-2}
24	gen	3.6×10^{-2}
25	sea	3.6×10^{-2}
26	genera	3.6×10^{-2}
27	eocen	3.5×10^{-2}
28	taxon	3.3×10^{-2}
29	north	3.3×10^{-2}
30	southern	3.3×10^{-2}
31	cambrian	3.3×10^{-2}
32	shallow	3.1×10^{-2}
33	south	3×10^{-2}
34	climat	2.9×10^{-2}
35	repres	2.9×10^{-2}
36	ordovician	2.8×10^{-2}
37	triassic	2.8×10^{-2}
38	taxonom	2.8×10^{-2}
39	sedimentari	2.7×10^{-2}
40	northern	2.7×10^{-2}
41	morpholog	2.7×10^{-2}
42	extant	2.7×10^{-2}
43	isotop	2.7×10^{-2}
44	describ	2.7×10^{-2}
45	extinct	2.6×10^{-2}
46	benthic	2.5×10^{-2}
47	ocean	2.5×10^{-2}
48	faunal	2.4×10^{-2}
49	method	2.4×10^{-2}
50	holocen	2.4×10^{-2}

No.	Word	RIG
51	earliest	2.3×10^{-2}
52	divers	2.3×10^{-2}
53	abund	2.3×10^{-2}
54	effect	2.2×10^{-2}
55	limeston	2.2×10^{-2}
56	faci	2.2×10^{-2}
57	permian	2.1×10^{-2}
58	pleistocen	2.1×10^{-2}
59	new	2.1×10^{-2}
60	oldest	2.1×10^{-2}
61	interpret	2.1×10^{-2}
62	eastern	2.1×10^{-2}
63	calcar	2.1×10^{-2}
64	reconstruct	2.1×10^{-2}
65	proxi	2×10^{-2}
66	strata	1.9×10^{-2}
67	use	1.9×10^{-2}
68	zone	1.9×10^{-2}
69	teeth	1.9×10^{-2}
70	occurr	1.8×10^{-2}
71	bivalv	1.8×10^{-2}
72	date	1.8×10^{-2}
73	western	1.8×10^{-2}
74	warm	1.7×10^{-2}
75	bed	1.7×10^{-2}
76	pollen	1.7×10^{-2}
77	skeleton	1.7×10^{-2}
78	sedimentolog	1.7×10^{-2}
79	patient	1.6×10^{-2}
80	shale	1.6×10^{-2}
81	terrestri	1.6×10^{-2}
82	part	1.6×10^{-2}
83	clade	1.6×10^{-2}
84	perform	1.6×10^{-2}
85	charact	1.5×10^{-2}
86	domin	1.5×10^{-2}
87	atlant	1.5×10^{-2}
88	glacial	1.5×10^{-2}
89	phylogenet	1.5×10^{-2}
90	margin	1.5×10^{-2}
91	carbon	1.5×10^{-2}
92	skull	1.5×10^{-2}
93	dure	1.4×10^{-2}
94	known	1.4×10^{-2}
95	america	1.4×10^{-2}
96	result	1.4×10^{-2}
97	belong	1.3×10^{-2}
98	continent	1.3×10^{-2}
99	section	1.3×10^{-2}
100	provinc	1.3×10^{-2}

Table D.183. The list of the top 100 words in the category Parasitology with RIGs

No.	Word	RIG	No.	Word	RIG
1	parasit	2×10^{-1}	51	caus	1.4×10^{-2}
2	infect	1.8×10^{-1}	52	born	1.4×10^{-2}
3	host	8.2×10^{-2}	53	vaccin	1.4×10^{-2}
4	malaria	7.6×10^{-2}	54	assay	1.4×10^{-2}
5	speci	5.4×10^{-2}	55	dengu	1.4×10^{-2}
6	plasmodium	5.2×10^{-2}	56	cattl	1.4×10^{-2}
7	mosquito	4.9×10^{-2}	57	identifi	1.3×10^{-2}
8	tick	3.9×10^{-2}	58	virul	1.2×10^{-2}
9	endem	3.6×10^{-2}	59	serolog	1.2×10^{-2}
10	pcr	3.6×10^{-2}	60	isol	1.2×10^{-2}
11	falciparum	3.6×10^{-2}	61	infest	1.2×10^{-2}
12	pathogen	3.5×10^{-2}	62	mice	1.2×10^{-2}
13	nematod	3.4×10^{-2}	63	protein	1.1×10^{-2}
14	larva	3.2×10^{-2}	64	phylogenet	1.1×10^{-2}
15	diseas	3×10^{-2}	65	viral	1.1×10^{-2}
16	egg	3×10^{-2}	66	molecular	1.1×10^{-2}
17	preval	3×10^{-2}	67	rdna	1.1×10^{-2}
18	leishmania	3×10^{-2}	68	elisa	1×10^{-2}
19	background	2.8×10^{-2}	69	infecti	1×10^{-2}
20	immun	2.5×10^{-2}	70	femal	1×10^{-2}
21	vector	2.5×10^{-2}	71	princip	1×10^{-2}
22	leishmaniasi	2.5×10^{-2}	72	health	1×10^{-2}
23	human	2.4×10^{-2}	73	sampl	9.9×10^{-3}
24	collect	2.4×10^{-2}	74	geograph	9.9×10^{-3}
25	spp	2.3×10^{-2}	75	simul	9.8×10^{-3}
26	dog	2.3×10^{-2}	76	studi	9.8×10^{-3}
27	anophel	2.2×10^{-2}	77	cyst	9.8×10^{-3}
28	anim	2.2×10^{-2}	78	sheep	9.5×10^{-3}
29	antibodi	2.1×10^{-2}	79	africa	9.5×10^{-3}
30	paper	2.1×10^{-2}	80	ribosom	9.4×10^{-3}
31	blood	2.1×10^{-2}	81	wild	9.4×10^{-3}
32	aed	2.1×10^{-2}	82	propos	9.3×10^{-3}
33	antigen	2×10^{-2}	83	drug	9.2×10^{-3}
34	epidemiolog	2×10^{-2}	84	strain	9.1×10^{-3}
35	sequenc	2×10^{-2}	85	coinfect	9×10^{-3}
36	virus	1.9×10^{-2}	86	transmit	8.9×10^{-3}
37	zoonot	1.9×10^{-2}	87	igg	8.9×10^{-3}
38	transmiss	1.9×10^{-2}	88	sera	8.7×10^{-3}
39	gene	1.9×10^{-2}	89	control	8.7×10^{-3}
40	larval	1.8×10^{-2}	90	found	8.6×10^{-3}
41	insecticid	1.7×10^{-2}	91	canin	8.6×10^{-3}
42	dna	1.6×10^{-2}	92	region	8.6×10^{-3}
43	detect	1.6×10^{-2}	93	stage	8.3×10^{-3}
44	conclus	1.6×10^{-2}	94	rrna	8.2×10^{-3}
45	genus	1.6×10^{-2}	95	report	8.2×10^{-3}
46	fever	1.5×10^{-2}	96	smear	8.2×10^{-3}
47	intestin	1.5×10^{-2}	97	outbreak	8.1×10^{-3}
48	adult	1.5×10^{-2}	98	genet	8.1×10^{-3}
49	popul	1.5×10^{-2}	99	presenc	7.9×10^{-3}
50	brazil	1.5×10^{-2}	100	insect	7.9×10^{-3}

Table D.184. The list of the top 100 words in the category Pathology with RIGs

No.	Word	RIG
1	tumor	8.7×10^{-2}
2	immunohistochem	7.9×10^{-2}
3	cell	7.7×10^{-2}
4	carcinoma	7.2×10^{-2}
5	immunohistochemistri	6.2×10^{-2}
6	express	5.9×10^{-2}
7	histolog	5.6×10^{-2}
8	patient	5.5×10^{-2}
9	tissu	5.1×10^{-2}
10	stain	5.1×10^{-2}
11	case	4.7×10^{-2}
12	patholog	4.2×10^{-2}
13	malign	4.1×10^{-2}
14	cancer	4×10^{-2}
15	lesion	3.9×10^{-2}
16	diagnosi	3.7×10^{-2}
17	neoplasm	3.6×10^{-2}
18	clinicopatholog	3.5×10^{-2}
19	clinic	3.4×10^{-2}
20	biopsi	3.2×10^{-2}
21	prognost	3.1×10^{-2}
22	marker	2.9×10^{-2}
23	rare	2.8×10^{-2}
24	diseas	2.8×10^{-2}
25	paper	2.6×10^{-2}
26	histopatholog	2.6×10^{-2}
27	prognosi	2.6×10^{-2}
28	grade	2.5×10^{-2}
29	pathologist	2.4×10^{-2}
30	associ	2.4×10^{-2}
31	cytolog	2.4×10^{-2}
32	invas	2.3×10^{-2}
33	protein	2.3×10^{-2}
34	adenocarcinoma	2.3×10^{-2}
35	lymph	2.2×10^{-2}
36	diagnost	2.1×10^{-2}
37	epitheli	2.1×10^{-2}
38	metastasi	2.1×10^{-2}
39	gene	2.1×10^{-2}
40	specimen	2.1×10^{-2}
41	benign	2.1×10^{-2}
42	cytoplasm	2×10^{-2}
43	posit	1.9×10^{-2}
44	paraffin	1.9×10^{-2}
45	mutat	1.9×10^{-2}
46	surviv	1.8×10^{-2}
47	squamous	1.8×10^{-2}
48	differenti	1.8×10^{-2}
49	diagnos	1.7×10^{-2}
50	progress	1.7×10^{-2}

No.	Word	RIG
51	tumour	1.7×10^{-2}
52	prolifer	1.6×10^{-2}
53	old	1.6×10^{-2}
54	immunostain	1.6×10^{-2}
55	conclus	1.6×10^{-2}
56	lymphoma	1.5×10^{-2}
57	pcr	1.5×10^{-2}
58	may	1.5×10^{-2}
59	featur	1.4×10^{-2}
60	aggress	1.4×10^{-2}
61	lung	1.4×10^{-2}
62	autopsi	1.4×10^{-2}
63	infiltr	1.3×10^{-2}
64	metastat	1.3×10^{-2}
65	pathogenesi	1.3×10^{-2}
66	overexpress	1.3×10^{-2}
67	negat	1.3×10^{-2}
68	simul	1.2×10^{-2}
69	detect	1.2×10^{-2}
70	resect	1.2×10^{-2}
71	metastas	1.2×10^{-2}
72	subtyp	1.2×10^{-2}
73	atyp	1.2×10^{-2}
74	correl	1.2×10^{-2}
75	signific	1.2×10^{-2}
76	report	1.1×10^{-2}
77	gland	1.1×10^{-2}
78	breast	1.1×10^{-2}
79	microarray	1.1×10^{-2}
80	receptor	1.1×10^{-2}
81	immunoreact	1.1×10^{-2}
82	propos	1.1×10^{-2}
83	needl	1.1×10^{-2}
84	recurr	1.1×10^{-2}
85	aspir	1×10^{-2}
86	temperatur	1×10^{-2}
87	stromal	1×10^{-2}
88	inflammatori	1×10^{-2}
89	energi	1×10^{-2}
90	antibodi	1×10^{-2}
91	normal	9.8×10^{-3}
92	therapeut	9.6×10^{-3}
93	biomark	9.5×10^{-3}
94	human	9.4×10^{-3}
95	poor	8.9×10^{-3}
96	node	8.7×10^{-3}
97	hyperplasia	8.6×10^{-3}
98	epithelium	8.6×10^{-3}
99	nuclear	8.6×10^{-3}
100	morpholog	8.5×10^{-3}

TABLE D.185. The list of the top 100 words in the category Pediatrics with RIGs

No.	Word	RIG
1	children	2×10^{-1}
2	pediatr	1.1×10^{-1}
3	conclus	1×10^{-1}
4	age	9.1×10^{-2}
5	infant	8.7×10^{-2}
6	year	6.2×10^{-2}
7	patient	6.2×10^{-2}
8	child	4.9×10^{-2}
9	neonat	4.8×10^{-2}
10	clinic	3.9×10^{-2}
11	birth	3.8×10^{-2}
12	month	3.7×10^{-2}
13	boy	3.6×10^{-2}
14	adolesc	3.4×10^{-2}
15	girl	3.4×10^{-2}
16	object	3.3×10^{-2}
17	preterm	3×10^{-2}
18	gestat	3×10^{-2}
19	outcom	3×10^{-2}
20	old	2.9×10^{-2}
21	paper	2.9×10^{-2}
22	hospit	2.9×10^{-2}
23	diagnosi	2.9×10^{-2}
24	retrospect	2.9×10^{-2}
25	parent	2.9×10^{-2}
26	background	2.9×10^{-2}
27	care	2.8×10^{-2}
28	congenit	2.8×10^{-2}
29	childhood	2.7×10^{-2}
30	associ	2.4×10^{-2}
31	risk	2.4×10^{-2}
32	syndrom	2.3×10^{-2}
33	diagnos	2×10^{-2}
34	newborn	2×10^{-2}
35	review	1.9×10^{-2}
36	symptom	1.8×10^{-2}
37	propos	1.8×10^{-2}
38	method	1.8×10^{-2}
39	medic	1.8×10^{-2}
40	report	1.8×10^{-2}
41	cohort	1.8×10^{-2}
42	week	1.7×10^{-2}
43	intervent	1.7×10^{-2}
44	mother	1.7×10^{-2}
45	diseas	1.7×10^{-2}
46	median	1.6×10^{-2}
47	rare	1.6×10^{-2}
48	case	1.5×10^{-2}
49	disord	1.5×10^{-2}
50	group	1.4×10^{-2}

No.	Word	RIG
51	paediatr	1.4×10^{-2}
52	follow	1.4×10^{-2}
53	born	1.4×10^{-2}
54	treatment	1.4×10^{-2}
55	matern	1.3×10^{-2}
56	properti	1.3×10^{-2}
57	assess	1.3×10^{-2}
58	complic	1.3×10^{-2}
59	prospect	1.3×10^{-2}
60	respiratori	1.3×10^{-2}
61	score	1.3×10^{-2}
62	simul	1.3×10^{-2}
63	receiv	1.3×10^{-2}
64	surgic	1.3×10^{-2}
65	earli	1.2×10^{-2}
66	health	1.2×10^{-2}
67	process	1.2×10^{-2}
68	abnorm	1.2×10^{-2}
69	includ	1.2×10^{-2}
70	underw	1.2×10^{-2}
71	result	1.2×10^{-2}
72	aim	1.2×10^{-2}
73	therapi	1.1×10^{-2}
74	structur	1.1×10^{-2}
75	temperatur	1.1×10^{-2}
76	effici	1.1×10^{-2}
77	signific	1.1×10^{-2}
78	prematur	1.1×10^{-2}
79	day	1×10^{-2}
80	surgeri	1×10^{-2}
81	infect	1×10^{-2}
82	enrol	1×10^{-2}
83	applic	1×10^{-2}
84	experiment	1×10^{-2}
85	model	9.9×10^{-3}
86	surfac	9.8×10^{-3}
87	acut	9.6×10^{-3}
88	preval	9.5×10^{-3}
89	common	9.3×10^{-3}
90	neurolog	9.3×10^{-3}
91	young	9×10^{-3}
92	studi	8.9×10^{-3}
93	male	8.6×10^{-3}
94	malform	8.6×10^{-3}
95	weight	8.4×10^{-3}
96	morbid	8.4×10^{-3}
97	treat	8.4×10^{-3}
98	dynam	8.3×10^{-3}
99	show	8.2×10^{-3}
100	may	8.1×10^{-3}

Table D.186. The list of the top 100 words in the category Peripheral Vascular Disease with RIGs

No.	Word	RIG	No.	Word	RIG
1	arteri	1.2×10^{-1}	51	stenosi	2×10^{-2}
2	conclus	1.1×10^{-1}	52	prospect	1.9×10^{-2}
3	patient	1.1×10^{-1}	53	adjust	1.9×10^{-2}
4	stroke	7.3×10^{-2}	54	embol	1.9×10^{-2}
5	hypertens	7.3×10^{-2}	55	underw	1.9×10^{-2}
6	blood	6.2×10^{-2}	56	treat	1.9×10^{-2}
7	vascular	6.2×10^{-2}	57	renal	1.9×10^{-2}
8	cardiovascular	5.7×10^{-2}	58	atherosclerot	1.8×10^{-2}
9	risk	5×10^{-2}	59	baselin	1.8×10^{-2}
10	ischem	5×10^{-2}	60	cerebr	1.8×10^{-2}
11	endovascular	4.7×10^{-2}	61	platelet	1.8×10^{-2}
12	aortic	4.7×10^{-2}	62	ventricular	1.8×10^{-2}
13	associ	4.2×10^{-2}	63	angiographi	1.8×10^{-2}
14	carotid	4×10^{-2}	64	diabet	1.8×10^{-2}
15	diseas	3.7×10^{-2}	65	complic	1.8×10^{-2}
16	aneurysm	3.7×10^{-2}	66	multivari	1.8×10^{-2}
17	coronari	3.5×10^{-2}	67	angiotensin	1.7×10^{-2}
18	age	3.5×10^{-2}	68	thromboembol	1.7×10^{-2}
19	systol	3.5×10^{-2}	69	odd	1.7×10^{-2}
20	background	3.4×10^{-2}	70	bleed	1.7×10^{-2}
21	infarct	3.3×10^{-2}	71	assess	1.7×10^{-2}
22	venous	3.2×10^{-2}	72	median	1.7×10^{-2}
23	thrombosi	3.1×10^{-2}	73	therapi	1.6×10^{-2}
24	paper	3.1×10^{-2}	74	men	1.6×10^{-2}
25	confid	3×10^{-2}	75	result	1.6×10^{-2}
26	year	2.9×10^{-2}	76	month	1.6×10^{-2}
27	clinic	2.9×10^{-2}	77	follow	1.6×10^{-2}
28	heart	2.9×10^{-2}	78	propos	1.6×10^{-2}
29	acut	2.9×10^{-2}	79	ratio	1.6×10^{-2}
30	stent	2.8×10^{-2}	80	day	1.6×10^{-2}
31	interv	2.8×10^{-2}	81	revascular	1.6×10^{-2}
32	atherosclerosi	2.7×10^{-2}	82	cohort	1.5×10^{-2}
33	pressur	2.7×10^{-2}	83	lipoprotein	1.5×10^{-2}
34	cardiac	2.7×10^{-2}	84	increas	1.5×10^{-2}
35	method	2.6×10^{-2}	85	death	1.5×10^{-2}
36	endotheli	2.6×10^{-2}	86	event	1.4×10^{-2}
37	mortal	2.6×10^{-2}	87	dysfunct	1.4×10^{-2}
38	myocardi	2.3×10^{-2}	88	regress	1.4×10^{-2}
39	outcom	2.3×10^{-2}	89	left	1.4×10^{-2}
40	diastol	2.3×10^{-2}	90	hospit	1.4×10^{-2}
41	vein	2.2×10^{-2}	91	symptomat	1.4×10^{-2}
42	signific	2.2×10^{-2}	92	independ	1.4×10^{-2}
43	occlus	2.2×10^{-2}	93	aorta	1.4×10^{-2}
44	treatment	2.1×10^{-2}	94	consecut	1.3×10^{-2}
45	hemorrhag	2.1×10^{-2}	95	predictor	1.3×10^{-2}
46	ischemia	2.1×10^{-2}	96	intervent	1.3×10^{-2}
47	factor	2×10^{-2}	97	bypass	1.3×10^{-2}
48	object	2×10^{-2}	98	versus	1.2×10^{-2}
49	anticoagul	2×10^{-2}	99	cholesterol	1.2×10^{-2}
50	vessel	2×10^{-2}	100	group	1.2×10^{-2}

Table D.187. The list of the top 100 words in the category Pharmacology and Pharmacy with RIGs

No.	Word	RIG	No.	Word	RIG
1	drug	1.1×10^{-1}	51	agonist	1.2×10^{-2}
2	dose	5.5×10^{-2}	52	diseas	1.2×10^{-2}
3	rat	4.3×10^{-2}	53	blood	1.1×10^{-2}
4	pharmacokinet	4.3×10^{-2}	54	mediat	1.1×10^{-2}
5	treatment	4.2×10^{-2}	55	potenti	1.1×10^{-2}
6	inhibit	4×10^{-2}	56	acid	1.1×10^{-2}
7	activ	3.5×10^{-2}	57	evalu	1.1×10^{-2}
8	administr	3.5×10^{-2}	58	deliveri	1.1×10^{-2}
9	induc	3.1×10^{-2}	59	liver	1×10^{-2}
10	vitro	3.1×10^{-2}	60	metabol	1×10^{-2}
11	paper	3×10^{-2}	61	auc	9.9×10^{-3}
12	inhibitor	3×10^{-2}	62	beta	9.8×10^{-3}
13	oral	2.8×10^{-2}	63	formul	9.7×10^{-3}
14	therapeut	2.7×10^{-2}	64	isol	9.7×10^{-3}
15	receptor	2.6×10^{-2}	65	cancer	9.6×10^{-3}
16	cell	2.4×10^{-2}	66	pharmaceut	9.5×10^{-3}
17	effect	2.2×10^{-2}	67	intraven	9.4×10^{-3}
18	compound	2.1×10^{-2}	68	advers	9.3×10^{-3}
19	treat	2.1×10^{-2}	69	enzym	9.3×10^{-3}
20	vivo	2×10^{-2}	70	anticanc	9.3×10^{-3}
21	assay	1.9×10^{-2}	71	metabolit	9.2×10^{-3}
22	pharmacolog	1.9×10^{-2}	72	serum	9.1×10^{-3}
23	administ	1.9×10^{-2}	73	antioxid	9×10^{-3}
24	mice	1.9×10^{-2}	74	exposur	9×10^{-3}
25	studi	1.8×10^{-2}	75	decreas	8.7×10^{-3}
26	efficaci	1.8×10^{-2}	76	aim	8.7×10^{-3}
27	clinic	1.8×10^{-2}	77	apoptosi	8.6×10^{-3}
28	therapi	1.8×10^{-2}	78	target	8.5×10^{-3}
29	concentr	1.8×10^{-2}	79	chromatographi	8.5×10^{-3}
30	medicin	1.7×10^{-2}	80	pathway	8.4×10^{-3}
31	toxic	1.6×10^{-2}	81	anim	8.4×10^{-3}
32	potent	1.6×10^{-2}	82	hplc	8.2×10^{-3}
33	agent	1.6×10^{-2}	83	inflammatori	8.2×10^{-3}
34	cytotox	1.6×10^{-2}	84	algorithm	7.9×10^{-3}
35	antagonist	1.5×10^{-2}	85	increas	7.9×10^{-3}
36	releas	1.5×10^{-2}	86	alpha	7.9×10^{-3}
37	inhibitori	1.5×10^{-2}	87	clearanc	7.8×10^{-3}
38	ethnopharmacolog	1.5×10^{-2}	88	inject	7.7×10^{-3}
39	protein	1.4×10^{-2}	89	energi	7.7×10^{-3}
40	conclus	1.4×10^{-2}	90	simul	7.6×10^{-3}
41	signific	1.3×10^{-2}	91	action	7.5×10^{-3}
42	human	1.3×10^{-2}	92	week	7.4×10^{-3}
43	ic50	1.3×10^{-2}	93	lipid	7.3×10^{-3}
44	antiinflammatori	1.3×10^{-2}	94	chronic	7.3×10^{-3}
45	propos	1.3×10^{-2}	95	kinas	7.2×10^{-3}
46	day	1.3×10^{-2}	96	mic	7.2×10^{-3}
47	express	1.2×10^{-2}	97	attenu	7.1×10^{-3}
48	plasma	1.2×10^{-2}	98	suggest	7×10^{-3}
49	patient	1.2×10^{-2}	99	pretreat	6.9×10^{-3}
50	bioavail	1.2×10^{-2}	100	comput	6.7×10^{-3}

Table D.188. The list of the top 100 words in the category Philosophy with RIGs

No.	Word	RIG
1	argu	1.9×10^{-1}
2	philosoph	1.1×10^{-1}
3	argument	7.8×10^{-2}
4	philosophi	7.4×10^{-2}
5	claim	6.9×10^{-2}
6	moral	5.1×10^{-2}
7	result	4.7×10^{-2}
8	defend	4.5×10^{-2}
9	metaphys	4.4×10^{-2}
10	view	4.3×10^{-2}
11	epistem	4.2×10^{-2}
12	studi	4×10^{-2}
13	notion	3.6×10^{-2}
14	truth	3.4×10^{-2}
15	epistemolog	3.4×10^{-2}
16	use	3.3×10^{-2}
17	theori	3.3×10^{-2}
18	account	3.2×10^{-2}
19	essay	3.2×10^{-2}
20	concept	3.1×10^{-2}
21	logic	2.9×10^{-2}
22	ethic	2.9×10^{-2}
23	way	2.8×10^{-2}
24	question	2.7×10^{-2}
25	think	2.7×10^{-2}
26	method	2.7×10^{-2}
27	high	2.5×10^{-2}
28	reason	2.4×10^{-2}
29	belief	2.4×10^{-2}
30	effect	2.4×10^{-2}
31	thesi	2.3×10^{-2}
32	increas	2.1×10^{-2}
33	articl	2.1×10^{-2}
34	idea	2×10^{-2}
35	critiqu	2×10^{-2}
36	thought	2×10^{-2}
37	perform	1.9×10^{-2}
38	proposit	1.9×10^{-2}
39	data	1.8×10^{-2}
40	normat	1.8×10^{-2}
41	measur	1.8×10^{-2}
42	contemporari	1.8×10^{-2}
43	intuit	1.7×10^{-2}
44	thing	1.7×10^{-2}
45	someth	1.7×10^{-2}
46	virtu	1.7×10^{-2}
47	compar	1.7×10^{-2}
48	phenomenolog	1.7×10^{-2}
49	debat	1.6×10^{-2}
50	rate	1.6×10^{-2}

No.	Word	RIG
51	semant	1.6×10^{-2}
52	ontolog	1.5×10^{-2}
53	say	1.5×10^{-2}
54	cell	1.5×10^{-2}
55	interpret	1.5×10^{-2}
56	kind	1.5×10^{-2}
57	control	1.5×10^{-2}
58	appeal	1.5×10^{-2}
59	low	1.4×10^{-2}
60	reject	1.4×10^{-2}
61	justif	1.4×10^{-2}
62	mind	1.4×10^{-2}
63	paper	1.4×10^{-2}
64	test	1.4×10^{-2}
65	sampl	1.4×10^{-2}
66	patient	1.4×10^{-2}
67	fact	1.4×10^{-2}
68	attempt	1.4×10^{-2}
69	call	1.3×10^{-2}
70	polit	1.3×10^{-2}
71	improv	1.3×10^{-2}
72	doe	1.3×10^{-2}
73	mere	1.3×10^{-2}
74	conscious	1.3×10^{-2}
75	scienc	1.3×10^{-2}
76	whi	1.2×10^{-2}
77	temperatur	1.2×10^{-2}
78	decreas	1.2×10^{-2}
79	ration	1.2×10^{-2}
80	commit	1.2×10^{-2}
81	obtain	1.2×10^{-2}
82	god	1.2×10^{-2}
83	distinct	1.2×10^{-2}
84	world	1.2×10^{-2}
85	effici	1.2×10^{-2}
86	offer	1.1×10^{-2}
87	dure	1.1×10^{-2}
88	certain	1.1×10^{-2}
89	seem	1.1×10^{-2}
90	signific	1.1×10^{-2}
91	higher	1.1×10^{-2}
92	principl	1.1×10^{-2}
93	observ	1.1×10^{-2}
94	true	1.1×10^{-2}
95	paramet	1.1×10^{-2}
96	indic	1.1×10^{-2}
97	investig	1.1×10^{-2}
98	explan	1.1×10^{-2}
99	report	1.1×10^{-2}
100	doctrin	1.1×10^{-2}

TABLE D.189. The list of the top 100 words in the category Physics, Applied with RIGs

No.	Word	RIG
1	film	4.5×10^{-2}
2	optic	3.7×10^{-2}
3	electron	3.1×10^{-2}
4	conclus	3.1×10^{-2}
5	fabric	3×10^{-2}
6	patient	2.7×10^{-2}
7	layer	2.6×10^{-2}
8	devic	2.5×10^{-2}
9	thin	2.5×10^{-2}
10	laser	2.5×10^{-2}
11	temperatur	2.4×10^{-2}
12	substrat	2.2×10^{-2}
13	deposit	2.1×10^{-2}
14	dope	2.1×10^{-2}
15	electr	1.8×10^{-2}
16	magnet	1.7×10^{-2}
17	associ	1.7×10^{-2}
18	silicon	1.7×10^{-2}
19	surfac	1.7×10^{-2}
20	properti	1.6×10^{-2}
21	year	1.5×10^{-2}
22	beam	1.5×10^{-2}
23	assess	1.5×10^{-2}
24	band	1.4×10^{-2}
25	voltag	1.4×10^{-2}
26	puls	1.4×10^{-2}
27	ray	1.4×10^{-2}
28	wavelength	1.4×10^{-2}
29	semiconductor	1.4×10^{-2}
30	clinic	1.4×10^{-2}
31	diseas	1.4×10^{-2}
32	crystal	1.4×10^{-2}
33	dielectr	1.4×10^{-2}
34	metal	1.3×10^{-2}
35	field	1.3×10^{-2}
36	age	1.3×10^{-2}
37	diffract	1.3×10^{-2}
38	thick	1.3×10^{-2}
39	risk	1.3×10^{-2}
40	studi	1.2×10^{-2}
41	anneal	1.2×10^{-2}
42	group	1.2×10^{-2}
43	structur	1.2×10^{-2}
44	atom	1.2×10^{-2}
45	epitaxi	1.2×10^{-2}
46	aim	1.2×10^{-2}
47	background	1.2×10^{-2}
48	identifi	1.1×10^{-2}
49	spectroscopi	1.1×10^{-2}
50	gene	1.1×10^{-2}

No.	Word	RIG
51	may	1.1×10^{-2}
52	transistor	1.1×10^{-2}
53	grown	1.1×10^{-2}
54	microscopi	1.1×10^{-2}
55	signific	1.1×10^{-2}
56	object	1.1×10^{-2}
57	quantum	1×10^{-2}
58	protein	1×10^{-2}
59	express	1×10^{-2}
60	outcom	1×10^{-2}
61	materi	1×10^{-2}
62	nanoparticl	1×10^{-2}
63	diod	1×10^{-2}
64	thermal	1×10^{-2}
65	sputter	1×10^{-2}
66	photon	1×10^{-2}
67	data	9.8×10^{-3}
68	energi	9.7×10^{-3}
69	particip	9.7×10^{-3}
70	nanowir	9.5×10^{-3}
71	coat	9.5×10^{-3}
72	light	9.4×10^{-3}
73	popul	9.2×10^{-3}
74	superconduct	9×10^{-3}
75	zno	8.9×10^{-3}
76	photoluminesc	8.9×10^{-3}
77	graphen	8.9×10^{-3}
78	human	8.8×10^{-3}
79	densiti	8.8×10^{-3}
80	nanostructur	8.7×10^{-3}
81	gan	8.7×10^{-3}
82	evalu	8.7×10^{-3}
83	suggest	8.7×10^{-3}
84	among	8.5×10^{-3}
85	whether	8.5×10^{-3}
86	carrier	8.3×10^{-3}
87	health	8.3×10^{-3}
88	charg	8.2×10^{-3}
89	absorpt	8.2×10^{-3}
90	crystallin	8.2×10^{-3}
91	manag	8.2×10^{-3}
92	day	8×10^{-3}
93	level	7.9×10^{-3}
94	room	7.8×10^{-3}
95	regul	7.7×10^{-3}
96	month	7.5×10^{-3}
97	includ	7.5×10^{-3}
98	polar	7.4×10^{-3}
99	plasmon	7.4×10^{-3}
100	examin	7.3×10^{-3}

Table D.190. The list of the top 100 words in the category Physics, Atomic, Molecular and Chemical with RIGs

No.	Word	RIG	No.	Word	RIG
1	atom	5.2×10^{-2}	51	assess	8.4×10^{-3}
2	energi	5.1×10^{-2}	52	depend	8.4×10^{-3}
3	quantum	4.9×10^{-2}	53	gene	8.3×10^{-3}
4	calcul	4.5×10^{-2}	54	polar	8.3×10^{-3}
5	molecul	4.4×10^{-2}	55	solvent	8.2×10^{-3}
6	electron	4.1×10^{-2}	56	ground	8.2×10^{-3}
7	state	3.5×10^{-2}	57	aim	8.1×10^{-3}
8	molecular	3.3×10^{-2}	58	human	8.1×10^{-3}
9	excit	3.1×10^{-2}	59	manag	8×10^{-3}
10	theori	2.8×10^{-2}	60	function	7.8×10^{-3}
11	densiti	2.5×10^{-2}	61	energet	7.7×10^{-3}
12	initio	2.5×10^{-2}	62	ionic	7.6×10^{-3}
13	bond	2.4×10^{-2}	63	chemic	7.6×10^{-3}
14	ion	2.2×10^{-2}	64	cluster	7.6×10^{-3}
15	spectra	2.1×10^{-2}	65	relax	7.5×10^{-3}
16	patient	2×10^{-2}	66	reaction	7.5×10^{-3}
17	transit	2×10^{-2}	67	entangl	7.4×10^{-3}
18	phys	2×10^{-2}	68	transfer	7.3×10^{-3}
19	conclus	1.9×10^{-2}	69	research	7.3×10^{-3}
20	interact	1.8×10^{-2}	70	thermodynam	7.1×10^{-3}
21	charg	1.8×10^{-2}	71	cation	7.1×10^{-3}
22	dft	1.7×10^{-2}	72	perturb	7.1×10^{-3}
23	vibrat	1.7×10^{-2}	73	laser	7×10^{-3}
24	paper	1.6×10^{-2}	74	health	6.8×10^{-3}
25	dynam	1.6×10^{-2}	75	scatter	6.8×10^{-3}
26	experiment	1.6×10^{-2}	76	temperatur	6.7×10^{-3}
27	spin	1.6×10^{-2}	77	intermolecular	6.7×10^{-3}
28	orbit	1.5×10^{-2}	78	anion	6.7×10^{-3}
29	hydrogen	1.5×10^{-2}	79	day	6.5×10^{-3}
30	spectroscopi	1.5×10^{-2}	80	dure	6.4×10^{-3}
31	ioniz	1.4×10^{-2}	81	associ	6.4×10^{-3}
32	dissoci	1.4×10^{-2}	82	properti	6.4×10^{-3}
33	theoret	1.3×10^{-2}	83	signific	6.3×10^{-3}
34	dipol	1.3×10^{-2}	84	phase	6.3×10^{-3}
35	coupl	1.3×10^{-2}	85	test	6.3×10^{-3}
36	year	1.2×10^{-2}	86	trap	6.1×10^{-3}
37	agreement	1.2×10^{-2}	87	symmetri	6.1×10^{-3}
38	photon	1.1×10^{-2}	88	evalu	6.1×10^{-3}
39	absorpt	1.1×10^{-2}	89	dimer	6×10^{-3}
40	clinic	1.1×10^{-2}	90	outcom	5.9×10^{-3}
41	age	1.1×10^{-2}	91	kinet	5.9×10^{-3}
42	object	1.1×10^{-2}	92	design	5.9×10^{-3}
43	diseas	1×10^{-2}	93	surfac	5.8×10^{-3}
44	risk	1×10^{-2}	94	puls	5.8×10^{-3}
45	structur	1×10^{-2}	95	decay	5.8×10^{-3}
46	background	9.7×10^{-3}	96	rotat	5.7×10^{-3}
47	hamiltonian	9.4×10^{-3}	97	month	5.7×10^{-3}
48	proton	9×10^{-3}	98	simul	5.6×10^{-3}
49	b3lyp	8.4×10^{-3}	99	collis	5.6×10^{-3}
50	reson	8.4×10^{-3}	100	control	5.6×10^{-3}

TABLE D.191. The list of the top 100 words in the category Physics, Condensed Matter with RIGs

No.	Word	RIG	No.	Word	RIG
1	electron	6.4×10^{-2}	51	clinic	1.2×10^{-2}
2	temperatur	4.6×10^{-2}	52	insul	1.2×10^{-2}
3	film	4.6×10^{-2}	53	aim	1.2×10^{-2}
4	spin	4.1×10^{-2}	54	bulk	1.2×10^{-2}
5	magnet	4×10^{-2}	55	diseas	1.2×10^{-2}
6	properti	3.4×10^{-2}	56	calcul	1.2×10^{-2}
7	dope	3.2×10^{-2}	57	spectra	1.1×10^{-2}
8	structur	2.8×10^{-2}	58	risk	1.1×10^{-2}
9	diffract	2.8×10^{-2}	59	room	1.1×10^{-2}
10	ray	2.6×10^{-2}	60	crystallin	1.1×10^{-2}
11	quantum	2.6×10^{-2}	61	nanoparticl	1.1×10^{-2}
12	lattic	2.5×10^{-2}	62	scan	1.1×10^{-2}
13	conclus	2.4×10^{-2}	63	zno	1.1×10^{-2}
14	atom	2.4×10^{-2}	64	age	1.1×10^{-2}
15	patient	2.3×10^{-2}	65	paper	1.1×10^{-2}
16	thin	2.3×10^{-2}	66	materi	1.1×10^{-2}
17	ferromagnet	2.2×10^{-2}	67	oxid	1.1×10^{-2}
18	crystal	2.2×10^{-2}	68	photoluminesc	1.1×10^{-2}
19	band	2.2×10^{-2}	69	background	1×10^{-2}
20	deposit	2.2×10^{-2}	70	ion	1×10^{-2}
21	layer	2.2×10^{-2}	71	associ	1×10^{-2}
22	spectroscopi	2.2×10^{-2}	72	polar	1×10^{-2}
23	energi	2.1×10^{-2}	73	exhibit	1×10^{-2}
24	densiti	2.1×10^{-2}	74	grown	9.8×10^{-3}
25	transit	2.1×10^{-2}	75	sputter	9.8×10^{-3}
26	surfac	2.1×10^{-2}	76	symmetri	9.5×10^{-3}
27	phase	2×10^{-2}	77	impur	9.3×10^{-3}
28	substrat	1.9×10^{-2}	78	epitaxi	9.3×10^{-3}
29	superconduct	1.8×10^{-2}	79	synthes	9.2×10^{-3}
30	microscopi	1.8×10^{-2}	80	data	9.2×10^{-3}
31	electr	1.8×10^{-2}	81	human	9.2×10^{-3}
32	metal	1.7×10^{-2}	82	nanowir	9.2×10^{-3}
33	charg	1.7×10^{-2}	83	coat	9×10^{-3}
34	optic	1.6×10^{-2}	84	evalu	9×10^{-3}
35	year	1.6×10^{-2}	85	nanostructur	9×10^{-3}
36	xrd	1.5×10^{-2}	86	thermal	8.9×10^{-3}
37	field	1.5×10^{-2}	87	manag	8.8×10^{-3}
38	gap	1.5×10^{-2}	88	gene	8.7×10^{-3}
39	antiferromagnet	1.5×10^{-2}	89	raman	8.7×10^{-3}
40	state	1.4×10^{-2}	90	devic	8.6×10^{-3}
41	phonon	1.4×10^{-2}	91	carrier	8.5×10^{-3}
42	anneal	1.4×10^{-2}	92	outcom	8.4×10^{-3}
43	assess	1.4×10^{-2}	93	research	8.4×10^{-3}
44	graphen	1.4×10^{-2}	94	thick	8.3×10^{-3}
45	object	1.4×10^{-2}	95	ferroelectr	8.3×10^{-3}
46	semiconductor	1.3×10^{-2}	96	depend	8.3×10^{-3}
47	fabric	1.3×10^{-2}	97	vacanc	8.3×10^{-3}
48	fermi	1.3×10^{-2}	98	signific	8.2×10^{-3}
49	prepar	1.3×10^{-2}	99	absorpt	8.1×10^{-3}
50	dielectr	1.2×10^{-2}	100	orbit	8.1×10^{-3}

Table D.192. The list of the top 100 words in the category Physics, Fluids and Plasmas with RIGs

No.	Word	RIG
1	plasma	8.4×10^{-2}
2	flow	5.1×10^{-2}
3	fluid	3.7×10^{-2}
4	numer	3.7×10^{-2}
5	turbul	3.7×10^{-2}
6	veloc	3.6×10^{-2}
7	reynold	2.8×10^{-2}
8	equat	2.8×10^{-2}
9	particl	2.7×10^{-2}
10	wave	2.6×10^{-2}
11	instabl	2.5×10^{-2}
12	simul	2.3×10^{-2}
13	field	2.1×10^{-2}
14	regim	2×10^{-2}
15	vortic	2×10^{-2}
16	discharg	2×10^{-2}
17	patient	1.8×10^{-2}
18	conclus	1.8×10^{-2}
19	phys	1.8×10^{-2}
20	dimension	1.8×10^{-2}
21	densiti	1.8×10^{-2}
22	boundari	1.6×10^{-2}
23	dynam	1.5×10^{-2}
24	wall	1.5×10^{-2}
25	pressur	1.5×10^{-2}
26	mode	1.5×10^{-2}
27	vortex	1.5×10^{-2}
28	magnet	1.4×10^{-2}
29	jet	1.4×10^{-2}
30	nonlinear	1.3×10^{-2}
31	perturb	1.3×10^{-2}
32	ion	1.3×10^{-2}
33	shear	1.3×10^{-2}
34	amplitud	1.2×10^{-2}
35	flux	1.2×10^{-2}
36	confin	1.2×10^{-2}
37	stoke	1.2×10^{-2}
38	hydrodynam	1.2×10^{-2}
39	year	1.2×10^{-2}
40	oscil	1.2×10^{-2}
41	experiment	1.1×10^{-2}
42	viscous	1.1×10^{-2}
43	group	1×10^{-2}
44	gas	1×10^{-2}
45	fluctuat	1×10^{-2}
46	clinic	1×10^{-2}
47	puls	1×10^{-2}
48	heat	1×10^{-2}
49	age	1×10^{-2}
50	steadi	9.7×10^{-3}
51	motion	9.6×10^{-3}
52	energi	9.4×10^{-3}
53	analyt	9.2×10^{-3}
54	navier	9.2×10^{-3}
55	forc	9.2×10^{-3}
56	transit	9×10^{-3}
57	convect	8.9×10^{-3}
58	propag	8.9×10^{-3}
59	activ	8.7×10^{-3}
60	scale	8.7×10^{-3}
61	electron	8.6×10^{-3}
62	viscos	8.5×10^{-3}
63	rotat	8.3×10^{-3}
64	inerti	8.2×10^{-3}
65	risk	8.1×10^{-3}
66	diseas	8.1×10^{-3}
67	coupl	8.1×10^{-3}
68	assess	8×10^{-3}
69	evalu	8×10^{-3}
70	paramet	8×10^{-3}
71	object	8×10^{-3}
72	driven	7.9×10^{-3}
73	agreement	7.8×10^{-3}
74	momentum	7.8×10^{-3}
75	microfluid	7.6×10^{-3}
76	laminar	7.5×10^{-3}
77	radial	7.5×10^{-3}
78	dissip	7.5×10^{-3}
79	equilibrium	7.4×10^{-3}
80	model	7.4×10^{-3}
81	diffus	7.4×10^{-3}
82	kinet	7.3×10^{-3}
83	finit	7.3×10^{-3}
84	number	7.3×10^{-3}
85	layer	7.1×10^{-3}
86	level	7×10^{-3}
87	human	6.7×10^{-3}
88	gene	6.6×10^{-3}
89	frequenc	6.5×10^{-3}
90	aim	6.5×10^{-3}
91	collis	6.5×10^{-3}
92	transport	6.4×10^{-3}
93	gradient	6.4×10^{-3}
94	evolut	6.3×10^{-3}
95	shown	6.3×10^{-3}
96	liquid	6.2×10^{-3}
97	manag	6.2×10^{-3}
98	law	6.2×10^{-3}
99	health	6.1×10^{-3}
100	treatment	6.1×10^{-3}

Table D.193. The list of the top 100 words in the category Physics, Mathematical with RIGs

No.	Word	RIG
1	equat	6×10^{-2}
2	numer	3.2×10^{-2}
3	quantum	3×10^{-2}
4	dimension	2.6×10^{-2}
5	conclus	2×10^{-2}
6	patient	1.9×10^{-2}
7	algebra	1.8×10^{-2}
8	signific	1.8×10^{-2}
9	hamiltonian	1.8×10^{-2}
10	finit	1.7×10^{-2}
11	space	1.6×10^{-2}
12	solut	1.6×10^{-2}
13	phys	1.6×10^{-2}
14	dynam	1.5×10^{-2}
15	theori	1.5×10^{-2}
16	activ	1.4×10^{-2}
17	general	1.4×10^{-2}
18	nonlinear	1.4×10^{-2}
19	lattic	1.3×10^{-2}
20	symmetri	1.2×10^{-2}
21	assess	1.2×10^{-2}
22	explicit	1.2×10^{-2}
23	exact	1.2×10^{-2}
24	year	1.2×10^{-2}
25	asymptot	1.2×10^{-2}
26	boundari	1.2×10^{-2}
27	increas	1.2×10^{-2}
28	discret	1.1×10^{-2}
29	clinic	1.1×10^{-2}
30	dure	1.1×10^{-2}
31	schroding	1.1×10^{-2}
32	arbitrari	1.1×10^{-2}
33	particl	1.1×10^{-2}
34	problem	1.1×10^{-2}
35	invari	1×10^{-2}
36	classic	1×10^{-2}
37	infinit	1×10^{-2}
38	age	1×10^{-2}
39	analyt	9.9×10^{-3}
40	expon	9.7×10^{-3}
41	evalu	9.7×10^{-3}
42	treatment	9.2×10^{-3}
43	wave	9.2×10^{-3}
44	law	9.1×10^{-3}
45	risk	8.8×10^{-3}
46	diseas	8.7×10^{-3}
47	prove	8.7×10^{-3}
48	eigenvalu	8.6×10^{-3}
49	manifold	8.5×10^{-3}
50	oscil	8.4×10^{-3}

No.	Word	RIG
51	acid	8.2×10^{-3}
52	entangl	8.2×10^{-3}
53	perturb	8.1×10^{-3}
54	state	8×10^{-3}
55	comput	8×10^{-3}
56	high	7.8×10^{-3}
57	aim	7.7×10^{-3}
58	field	7.7×10^{-3}
59	dimens	7.7×10^{-3}
60	human	7.6×10^{-3}
61	solv	7.6×10^{-3}
62	theorem	7.5×10^{-3}
63	system	7.4×10^{-3}
64	approxim	7.4×10^{-3}
65	suggest	7.4×10^{-3}
66	converg	7.2×10^{-3}
67	cell	7.2×10^{-3}
68	day	7.2×10^{-3}
69	stochast	7.1×10^{-3}
70	design	7×10^{-3}
71	report	7×10^{-3}
72	indic	7×10^{-3}
73	compar	7×10^{-3}
74	model	6.9×10^{-3}
75	gene	6.9×10^{-3}
76	scheme	6.9×10^{-3}
77	research	6.8×10^{-3}
78	health	6.8×10^{-3}
79	bifurc	6.8×10^{-3}
80	formula	6.8×10^{-3}
81	protein	6.8×10^{-3}
82	exampl	6.7×10^{-3}
83	consid	6.7×10^{-3}
84	entropi	6.6×10^{-3}
85	object	6.5×10^{-3}
86	diffus	6.5×10^{-3}
87	singular	6.5×10^{-3}
88	polynomi	6.5×10^{-3}
89	deriv	6.4×10^{-3}
90	coupl	6.3×10^{-3}
91	regim	6.3×10^{-3}
92	level	6.3×10^{-3}
93	improv	6.3×10^{-3}
94	background	6.2×10^{-3}
95	month	6.1×10^{-3}
96	correspond	6.1×10^{-3}
97	sampl	6.1×10^{-3}
98	fluctuat	6.1×10^{-3}
99	data	6×10^{-3}
100	factor	6×10^{-3}

Table D.194. The list of the top 100 words in the category Physics, Multidisciplinary with RIGs

No.	Word	RIG	No.	Word	RIG
1	quantum	4.8×10^{-2}	51	dynam	7.3×10^{-3}
2	spin	2.2×10^{-2}	52	phase	7.3×10^{-3}
3	patient	2.1×10^{-2}	53	decay	7.2×10^{-3}
4	conclus	2.1×10^{-2}	54	manag	7.2×10^{-3}
5	state	1.7×10^{-2}	55	oscil	7.1×10^{-3}
6	energi	1.7×10^{-2}	56	group	7×10^{-3}
7	equat	1.6×10^{-2}	57	dirac	7×10^{-3}
8	wave	1.5×10^{-2}	58	howev	6.9×10^{-3}
9	signific	1.5×10^{-2}	59	boson	6.9×10^{-3}
10	field	1.5×10^{-2}	60	day	6.7×10^{-3}
11	assess	1.4×10^{-2}	61	theoret	6.6×10^{-3}
12	symmetri	1.3×10^{-2}	62	charg	6.5×10^{-3}
13	entangl	1.3×10^{-2}	63	relativist	6.4×10^{-3}
14	theori	1.3×10^{-2}	64	superconduct	6.4×10^{-3}
15	clinic	1.2×10^{-2}	65	result	6.4×10^{-3}
16	evalu	1.2×10^{-2}	66	densiti	6.4×10^{-3}
17	transit	1.2×10^{-2}	67	outcom	6.4×10^{-3}
18	activ	1.2×10^{-2}	68	cosmolog	6.3×10^{-3}
19	diseas	1.1×10^{-2}	69	particip	6.3×10^{-3}
20	age	1.1×10^{-2}	70	month	6.2×10^{-3}
21	lattic	1.1×10^{-2}	71	perform	6.2×10^{-3}
22	year	1.1×10^{-2}	72	gravit	6.1×10^{-3}
23	magnet	1.1×10^{-2}	73	suggest	6.1×10^{-3}
24	particl	1×10^{-2}	74	acid	6×10^{-3}
25	atom	1×10^{-2}	75	regul	6×10^{-3}
26	momentum	9.9×10^{-3}	76	fermi	6×10^{-3}
27	risk	9.9×10^{-3}	77	physic	6×10^{-3}
28	electron	9.8×10^{-3}	78	identifi	6×10^{-3}
29	aim	9.6×10^{-3}	79	spacetim	5.9×10^{-3}
30	treatment	9.5×10^{-3}	80	design	5.9×10^{-3}
31	dimension	8.9×10^{-3}	81	schroding	5.9×10^{-3}
32	hamiltonian	8.7×10^{-3}	82	exact	5.9×10^{-3}
33	develop	8.7×10^{-3}	83	increas	5.8×10^{-3}
34	gene	8.6×10^{-3}	84	level	5.7×10^{-3}
35	object	8.6×10^{-3}	85	studi	5.7×10^{-3}
36	human	8.4×10^{-3}	86	interact	5.6×10^{-3}
37	photon	8.2×10^{-3}	87	entropi	5.6×10^{-3}
38	dure	8.2×10^{-3}	88	neutron	5.6×10^{-3}
39	scatter	8×10^{-3}	89	research	5.6×10^{-3}
40	associ	8×10^{-3}	90	examin	5.6×10^{-3}
41	fermion	8×10^{-3}	91	among	5.5×10^{-3}
42	improv	8×10^{-3}	92	orbit	5.4×10^{-3}
43	protein	7.9×10^{-3}	93	background	5.4×10^{-3}
44	coupl	7.9×10^{-3}	94	adult	5.3×10^{-3}
45	einstein	7.8×10^{-3}	95	compar	5.2×10^{-3}
46	test	7.7×10^{-3}	96	major	5.2×10^{-3}
47	excit	7.7×10^{-3}	97	cancer	5.2×10^{-3}
48	cell	7.6×10^{-3}	98	includ	5.2×10^{-3}
49	calcul	7.5×10^{-3}	99	may	5.1×10^{-3}
50	health	7.5×10^{-3}	100	method	5.1×10^{-3}

TABLE D.195. The list of the top 100 words in the category Physics, Nuclear with RIGs

No.	Word	RIG
1	neutron	7.8×10^{-2}
2	mev	7.1×10^{-2}
3	energi	6.3×10^{-2}
4	nucleon	5.2×10^{-2}
5	nuclear	4.8×10^{-2}
6	nuclei	4.4×10^{-2}
7	beam	4.2×10^{-2}
8	proton	4.2×10^{-2}
9	collis	4×10^{-2}
10	gev	3.9×10^{-2}
11	hadron	3.9×10^{-2}
12	detector	3.8×10^{-2}
13	quark	3.6×10^{-2}
14	decay	3.4×10^{-2}
15	lhc	3.1×10^{-2}
16	heavi	2.8×10^{-2}
17	momentum	2.6×10^{-2}
18	meson	2.6×10^{-2}
19	particl	2.4×10^{-2}
20	gamma	2.3×10^{-2}
21	qcd	2.3×10^{-2}
22	kev	2.2×10^{-2}
23	calcul	2.2×10^{-2}
24	ion	2.1×10^{-2}
25	collid	2.1×10^{-2}
26	relativist	2×10^{-2}
27	gluon	1.9×10^{-2}
28	section	1.8×10^{-2}
29	tev	1.8×10^{-2}
30	scatter	1.8×10^{-2}
31	patient	1.8×10^{-2}
32	fission	1.7×10^{-2}
33	nucleus	1.7×10^{-2}
34	charg	1.7×10^{-2}
35	isotop	1.6×10^{-2}
36	excit	1.6×10^{-2}
37	neutrino	1.5×10^{-2}
38	mass	1.4×10^{-2}
39	baryon	1.4×10^{-2}
40	cross	1.4×10^{-2}
41	conclus	1.4×10^{-2}
42	symmetri	1.4×10^{-2}
43	matter	1.3×10^{-2}
44	scintil	1.2×10^{-2}
45	transvers	1.2×10^{-2}
46	signific	1.2×10^{-2}
47	angular	1.1×10^{-2}
48	cell	1.1×10^{-2}
49	acceler	1.1×10^{-2}
50	control	1.1×10^{-2}
51	boson	1.1×10^{-2}
52	assess	1×10^{-2}
53	spin	1×10^{-2}
54	age	9.8×10^{-3}
55	diseas	9.8×10^{-3}
56	clinic	9.7×10^{-3}
57	state	9.7×10^{-3}
58	scalar	9.5×10^{-3}
59	increas	9.5×10^{-3}
60	protein	9.4×10^{-3}
61	radiat	9.4×10^{-3}
62	reaction	9.4×10^{-3}
63	experiment	9.4×10^{-3}
64	photon	9.3×10^{-3}
65	facil	9.2×10^{-3}
66	radioact	9×10^{-3}
67	ray	8.7×10^{-3}
68	risk	8.4×10^{-3}
69	human	8.4×10^{-3}
70	associ	8.4×10^{-3}
71	higg	8.3×10^{-3}
72	lepton	8.2×10^{-3}
73	coulomb	8.2×10^{-3}
74	treatment	8.2×10^{-3}
75	gene	8.1×10^{-3}
76	shell	8.1×10^{-3}
77	measur	8.1×10^{-3}
78	astrophys	7.9×10^{-3}
79	object	7.4×10^{-3}
80	activ	7.4×10^{-3}
81	carlo	7.3×10^{-3}
82	quadrupol	7.3×10^{-3}
83	group	7.2×10^{-3}
84	mont	7.2×10^{-3}
85	gaug	7.2×10^{-3}
86	spectromet	6.9×10^{-3}
87	acid	6.9×10^{-3}
88	examin	6.8×10^{-3}
89	discuss	6.8×10^{-3}
90	physic	6.5×10^{-3}
91	suggest	6.5×10^{-3}
92	chiral	6.5×10^{-3}
93	pariti	6.4×10^{-3}
94	howev	6.3×10^{-3}
95	agreement	6.3×10^{-3}
96	cosmic	6.2×10^{-3}
97	health	6.1×10^{-3}
98	bar	5.9×10^{-3}
99	experi	5.9×10^{-3}
100	flavor	5.9×10^{-3}

TABLE D.196. The list of the top 100 words in the category Physics, Particles and Fields with RIGs

No.	Word	RIG	No.	Word	RIG
1	scalar	7.3×10^{-2}	51	proton	1.8×10^{-2}
2	quark	6.4×10^{-2}	52	nucleon	1.8×10^{-2}
3	higg	5.9×10^{-2}	53	hole	1.8×10^{-2}
4	gaug	5.9×10^{-2}	54	neutron	1.7×10^{-2}
5	lhc	5.9×10^{-2}	55	relativist	1.7×10^{-2}
6	cosmolog	5.3×10^{-2}	56	space	1.6×10^{-2}
7	gev	5.2×10^{-2}	57	cell	1.6×10^{-2}
8	boson	5.2×10^{-2}	58	inflat	1.6×10^{-2}
9	decay	4.8×10^{-2}	59	mev	1.6×10^{-2}
10	neutrino	4.5×10^{-2}	60	activ	1.5×10^{-2}
11	qcd	4.4×10^{-2}	61	control	1.5×10^{-2}
12	tev	4.2×10^{-2}	62	method	1.5×10^{-2}
13	hadron	4.2×10^{-2}	63	bar	1.4×10^{-2}
14	symmetri	4×10^{-2}	64	spin	1.4×10^{-2}
15	lepton	4×10^{-2}	65	assess	1.4×10^{-2}
16	theori	3.9×10^{-2}	66	constraint	1.4×10^{-2}
17	mass	3.7×10^{-2}	67	vacuum	1.4×10^{-2}
18	spacetim	3.7×10^{-2}	68	luminos	1.3×10^{-2}
19	supersymmetr	3.6×10^{-2}	69	equat	1.3×10^{-2}
20	matter	3.6×10^{-2}	70	signific	1.3×10^{-2}
21	graviti	3.5×10^{-2}	71	heavi	1.3×10^{-2}
22	gravit	3.5×10^{-2}	72	gamma	1.2×10^{-2}
23	collid	3.3×10^{-2}	73	physic	1.2×10^{-2}
24	energi	3.2×10^{-2}	74	chiral	1.2×10^{-2}
25	dark	3.1×10^{-2}	75	phenomenolog	1.2×10^{-2}
26	momentum	3×10^{-2}	76	diseas	1.2×10^{-2}
27	meson	2.9×10^{-2}	77	clinic	1.2×10^{-2}
28	cosmic	2.8×10^{-2}	78	transvers	1.2×10^{-2}
29	fermion	2.8×10^{-2}	79	protein	1.1×10^{-2}
30	field	2.8×10^{-2}	80	dirac	1.1×10^{-2}
31	detector	2.8×10^{-2}	81	horizon	1.1×10^{-2}
32	perturb	2.6×10^{-2}	82	model	1.1×10^{-2}
33	violat	2.4×10^{-2}	83	sigma	1.1×10^{-2}
34	tensor	2.3×10^{-2}	84	loop	1.1×10^{-2}
35	gluon	2.3×10^{-2}	85	human	1.1×10^{-2}
36	coupl	2.3×10^{-2}	86	age	1.1×10^{-2}
37	baryon	2.2×10^{-2}	87	univers	1.1×10^{-2}
38	einstein	2.1×10^{-2}	88	angular	1.1×10^{-2}
39	quantum	2.1×10^{-2}	89	treatment	1×10^{-2}
40	flavor	2.1×10^{-2}	90	break	1×10^{-2}
41	particl	2.1×10^{-2}	91	risk	1×10^{-2}
42	patient	2.1×10^{-2}	92	scatter	1×10^{-2}
43	charg	1.9×10^{-2}	93	photon	1×10^{-2}
44	invari	1.9×10^{-2}	94	discuss	1×10^{-2}
45	planck	1.9×10^{-2}	95	massiv	9.9×10^{-3}
46	collis	1.9×10^{-2}	96	evalu	9.8×10^{-3}
47	conclus	1.9×10^{-2}	97	astrophys	9.8×10^{-3}
48	increas	1.9×10^{-2}	98	sector	9.6×10^{-3}
49	string	1.9×10^{-2}	99	beam	9.6×10^{-3}
50	black	1.9×10^{-2}	100	correct	9.6×10^{-3}

TABLE D.197. The list of the top 100 words in the category Physiology with RIGs

No.	Word	RIG	No.	Word	RIG
1	muscl	4.6×10^{-2}	51	phosphoryl	1.1×10^{-2}
2	activ	3.8×10^{-2}	52	vascular	1.1×10^{-2}
3	induc	3.7×10^{-2}	53	pathway	1.1×10^{-2}
4	rat	3.6×10^{-2}	54	studi	1×10^{-2}
5	express	3×10^{-2}	55	base	1×10^{-2}
6	increas	3×10^{-2}	56	inhibitor	1×10^{-2}
7	paper	2.7×10^{-2}	57	signal	1×10^{-2}
8	receptor	2.5×10^{-2}	58	kinas	1×10^{-2}
9	regul	2.4×10^{-2}	59	glucos	9.9×10^{-3}
10	respons	2.4×10^{-2}	60	propos	9.7×10^{-3}
11	physiolog	2.4×10^{-2}	61	mechan	9.5×10^{-3}
12	protein	2.4×10^{-2}	62	insulin	9.1×10^{-3}
13	stimul	2.4×10^{-2}	63	control	9.1×10^{-3}
14	exercis	2.3×10^{-2}	64	intracellular	9×10^{-3}
15	decreas	2.1×10^{-2}	65	impair	9×10^{-3}
16	inhibit	2.1×10^{-2}	66	alpha	8.8×10^{-3}
17	mediat	2.1×10^{-2}	67	hypothes	8.7×10^{-3}
18	mice	2×10^{-2}	68	reduc	8.6×10^{-3}
19	cell	1.9×10^{-2}	69	agonist	8.6×10^{-3}
20	min	1.9×10^{-2}	70	mous	8.5×10^{-3}
21	suggest	1.9×10^{-2}	71	subject	8.3×10^{-3}
22	heart	1.8×10^{-2}	72	membran	8.3×10^{-3}
23	dure	1.8×10^{-2}	73	elev	8.2×10^{-3}
24	blood	1.8×10^{-2}	74	injuri	8.2×10^{-3}
25	alter	1.7×10^{-2}	75	acut	8.1×10^{-3}
26	metabol	1.6×10^{-2}	76	oxygen	7.8×10^{-3}
27	ca2	1.6×10^{-2}	77	antagonist	7.8×10^{-3}
28	neuron	1.6×10^{-2}	78	function	7.8×10^{-3}
29	arteri	1.4×10^{-2}	79	howev	7.7×10^{-3}
30	level	1.4×10^{-2}	80	abolish	7.6×10^{-3}
31	signific	1.4×10^{-2}	81	diet	7.6×10^{-3}
32	role	1.3×10^{-2}	82	renal	7.6×10^{-3}
33	tissu	1.3×10^{-2}	83	record	7.5×10^{-3}
34	chang	1.3×10^{-2}	84	extracellular	7.5×10^{-3}
35	mrna	1.3×10^{-2}	85	clamp	7.4×10^{-3}
36	male	1.3×10^{-2}	86	problem	7.4×10^{-3}
37	skelet	1.3×10^{-2}	87	plasma	7.3×10^{-3}
38	day	1.3×10^{-2}	88	nitric	7.3×10^{-3}
39	rest	1.3×10^{-2}	89	kidney	7.2×10^{-3}
40	bodi	1.2×10^{-2}	90	upregul	7.2×10^{-3}
41	contract	1.2×10^{-2}	91	hypothesi	7.2×10^{-3}
42	whether	1.2×10^{-2}	92	stress	7.2×10^{-3}
43	healthi	1.2×10^{-2}	93	vivo	7.2×10^{-3}
44	cardiac	1.2×10^{-2}	94	blot	7.1×10^{-3}
45	evok	1.2×10^{-2}	95	modul	7.1×10^{-3}
46	anim	1.1×10^{-2}	96	materi	7.1×10^{-3}
47	attenu	1.1×10^{-2}	97	stimulus	7×10^{-3}
48	effect	1.1×10^{-2}	98	stimuli	7×10^{-3}
49	hypoxia	1.1×10^{-2}	99	endotheli	7×10^{-3}
50	may	1.1×10^{-2}	100	cardiovascular	7×10^{-3}

TABLE D.198. The list of the top 100 words in the category Planning and Development with RIGs

No.	Word	RIG	No.	Word	RIG
1	polici	6.7×10^{-2}	51	opportun	1.1×10^{-2}
2	econom	5×10^{-2}	52	agenda	1.1×10^{-2}
3	govern	4×10^{-2}	53	fund	1.1×10^{-2}
4	countri	4×10^{-2}	54	result	1.1×10^{-2}
5	market	3.6×10^{-2}	55	find	1.1×10^{-2}
6	articl	3.4×10^{-2}	56	temperatur	1×10^{-2}
7	social	3.4×10^{-2}	57	trade	1×10^{-2}
8	polit	3.3×10^{-2}	58	labour	9.9×10^{-3}
9	urban	3.1×10^{-2}	59	hous	9.8×10^{-3}
10	economi	3×10^{-2}	60	reform	9.5×10^{-3}
11	sector	2.9×10^{-2}	61	project	9.5×10^{-3}
12	rural	2.5×10^{-2}	62	clinic	9.4×10^{-3}
13	argu	2.5×10^{-2}	63	issu	9.2×10^{-3}
14	citi	2.3×10^{-2}	64	surfac	9.1×10^{-3}
15	develop	2.3×10^{-2}	65	livelihood	9×10^{-3}
16	innov	2.2×10^{-2}	66	literatur	9×10^{-3}
17	institut	2.1×10^{-2}	67	manag	8.9×10^{-3}
18	household	2.1×10^{-2}	68	obtain	8.8×10^{-3}
19	capit	2.1×10^{-2}	69	local	8.8×10^{-3}
20	public	2.1×10^{-2}	70	compani	8.8×10^{-3}
21	research	2.1×10^{-2}	71	intern	8.7×10^{-3}
22	actor	1.9×10^{-2}	72	paramet	8.7×10^{-3}
23	incom	1.9×10^{-2}	73	organis	8.6×10^{-3}
24	invest	1.8×10^{-2}	74	perspect	8.5×10^{-3}
25	paper	1.8×10^{-2}	75	practic	8.5×10^{-3}
26	busi	1.7×10^{-2}	76	societi	8.5×10^{-3}
27	poverti	1.7×10^{-2}	77	enterpris	8.5×10^{-3}
28	empir	1.7×10^{-2}	78	protein	8.3×10^{-3}
29	agricultur	1.6×10^{-2}	79	detect	8.3×10^{-3}
30	sustain	1.6×10^{-2}	80	explor	8.3×10^{-3}
31	cell	1.6×10^{-2}	81	experiment	8.2×10^{-3}
32	firm	1.6×10^{-2}	82	global	8.2×10^{-3}
33	patient	1.6×10^{-2}	83	financ	8.1×10^{-3}
34	method	1.5×10^{-2}	84	foreign	8×10^{-3}
35	nation	1.4×10^{-2}	85	conclus	7.9×10^{-3}
36	communiti	1.3×10^{-2}	86	servic	7.9×10^{-3}
37	financi	1.3×10^{-2}	87	way	7.8×10^{-3}
38	draw	1.3×10^{-2}	88	discours	7.8×10^{-3}
39	privat	1.3×10^{-2}	89	world	7.7×10^{-3}
40	land	1.3×10^{-2}	90	peopl	7.6×10^{-3}
41	impact	1.3×10^{-2}	91	simul	7.6×10^{-3}
42	focus	1.2×10^{-2}	92	territori	7.5×10^{-3}
43	context	1.2×10^{-2}	93	treatment	7.4×10^{-3}
44	africa	1.2×10^{-2}	94	framework	7.4×10^{-3}
45	farmer	1.2×10^{-2}	95	knowledg	7.4×10^{-3}
46	plan	1.2×10^{-2}	96	interview	7.2×10^{-3}
47	resourc	1.2×10^{-2}	97	socio	7.2×10^{-3}
48	strateg	1.1×10^{-2}	98	emerg	7.1×10^{-3}
49	industri	1.1×10^{-2}	99	stakehold	7.1×10^{-3}
50	debat	1.1×10^{-2}	100	survey	7.1×10^{-3}

Table D.199. The list of the top 100 words in the category Plant Sciences with RIGs

No.	Word	RIG	No.	Word	RIG
1	plant	2.1×10^{-1}	51	sequenc	1.3×10^{-2}
2	speci	8.4×10^{-2}	52	genom	1.3×10^{-2}
3	leaf	6.2×10^{-2}	53	content	1.3×10^{-2}
4	arabidopsi	5.7×10^{-2}	54	cultiv	1.3×10^{-2}
5	cultivar	4.7×10^{-2}	55	pathogen	1.3×10^{-2}
6	leav	4.6×10^{-2}	56	express	1.3×10^{-2}
7	root	4.2×10^{-2}	57	germplasm	1.2×10^{-2}
8	gene	4×10^{-2}	58	inocul	1.2×10^{-2}
9	flower	4×10^{-2}	59	regul	1.2×10^{-2}
10	seed	3.8×10^{-2}	60	sativa	1.2×10^{-2}
11	seedl	3.7×10^{-2}	61	tomato	1.1×10^{-2}
12	crop	3.6×10^{-2}	62	nutrient	1.1×10^{-2}
13	thaliana	3.4×10^{-2}	63	protein	1.1×10^{-2}
14	shoot	3.1×10^{-2}	64	loci	1×10^{-2}
15	fruit	2.8×10^{-2}	65	phenotyp	1×10^{-2}
16	trait	2.6×10^{-2}	66	simul	1×10^{-2}
17	soil	2.5×10^{-2}	67	propos	1×10^{-2}
18	paper	2.5×10^{-2}	68	pollin	1×10^{-2}
19	growth	2.4×10^{-2}	69	model	1×10^{-2}
20	germin	2.1×10^{-2}	70	biomass	9.9×10^{-3}
21	photosynthet	2.1×10^{-2}	71	dri	9.7×10^{-3}
22	ethnopharmacolog	2.1×10^{-2}	72	morpholog	9.7×10^{-3}
23	patient	2×10^{-2}	73	taxonom	9.5×10^{-3}
24	breed	1.9×10^{-2}	74	marker	9.5×10^{-3}
25	isol	1.9×10^{-2}	75	weed	9.5×10^{-3}
26	genet	1.9×10^{-2}	76	antioxid	9.2×10^{-3}
27	accumul	1.9×10^{-2}	77	forest	9.1×10^{-3}
28	genus	1.9×10^{-2}	78	product	9.1×10^{-3}
29	rice	1.8×10^{-2}	79	comput	9.1×10^{-3}
30	veget	1.7×10^{-2}	80	maiz	9×10^{-3}
31	grown	1.7×10^{-2}	81	qtl	9×10^{-3}
32	tree	1.7×10^{-2}	82	greenhous	8.9×10^{-3}
33	photosynthesi	1.6×10^{-2}	83	fertil	8.7×10^{-3}
34	transgen	1.6×10^{-2}	84	genera	8.6×10^{-3}
35	drought	1.6×10^{-2}	85	stem	8.5×10^{-3}
36	wheat	1.6×10^{-2}	86	algorithm	8.5×10^{-3}
37	mutant	1.5×10^{-2}	87	chromosom	8.5×10^{-3}
38	transcript	1.5×10^{-2}	88	oper	8.4×10^{-3}
39	wild	1.5×10^{-2}	89	medicin	8.4×10^{-3}
40	toler	1.5×10^{-2}	90	respons	8.3×10^{-3}
41	pollen	1.5×10^{-2}	91	fungal	8.3×10^{-3}
42	biosynthesi	1.5×10^{-2}	92	enzym	8.2×10^{-3}
43	phylogenet	1.4×10^{-2}	93	clinic	8.2×10^{-3}
44	abiot	1.4×10^{-2}	94	compound	8.2×10^{-3}
45	chlorophyl	1.4×10^{-2}	95	grow	8.2×10^{-3}
46	stress	1.4×10^{-2}	96	physiolog	8.1×10^{-3}
47	acid	1.4×10^{-2}	97	conserv	7.8×10^{-3}
48	taxa	1.3×10^{-2}	98	method	7.7×10^{-3}
49	divers	1.3×10^{-2}	99	perform	7.6×10^{-3}
50	genotyp	1.3×10^{-2}	100	molecular	7.5×10^{-3}

TABLE D.200. The list of the top 100 words in the category Poetry with RIGs

No.	Word	RIG
1	poem	2.9×10^{-1}
2	poet	2.9×10^{-1}
3	poetri	2.8×10^{-1}
4	essay	2.3×10^{-1}
5	poetic	2×10^{-1}
6	dickinson	1.8×10^{-1}
7	argu	9.3×10^{-2}
8	emili	7.9×10^{-2}
9	literari	7.5×10^{-2}
10	text	7.4×10^{-2}
11	lyric	6.3×10^{-2}
12	antholog	5.7×10^{-2}
13	centuri	5.5×10^{-2}
14	epigram	5.1×10^{-2}
15	read	5×10^{-2}
16	shelley	4.8×10^{-2}
17	result	4.8×10^{-2}
18	vers	4.4×10^{-2}
19	write	4.1×10^{-2}
20	languag	3.7×10^{-2}
21	persona	3.4×10^{-2}
22	song	3.4×10^{-2}
23	interpret	3.4×10^{-2}
24	studi	3.3×10^{-2}
25	ancient	3.2×10^{-2}
26	greek	3.1×10^{-2}
27	john	3.1×10^{-2}
28	voic	3.1×10^{-2}
29	wine	3×10^{-2}
30	enact	2.9×10^{-2}
31	jame	2.6×10^{-2}
32	book	2.4×10^{-2}
33	moment	2.4×10^{-2}
34	edit	2.4×10^{-2}
35	corpus	2.4×10^{-2}
36	stori	2.3×10^{-2}
37	author	2.3×10^{-2}
38	shakespear	2.2×10^{-2}
39	histor	2.2×10^{-2}
40	american	2.2×10^{-2}
41	imit	2.2×10^{-2}
42	portray	2.2×10^{-2}
43	cultur	2.2×10^{-2}
44	theatric	2.1×10^{-2}
45	compar	2.1×10^{-2}
46	letter	2×10^{-2}
47	artist	2×10^{-2}
48	pen	2×10^{-2}
49	wife	2×10^{-2}
50	use	1.9×10^{-2}
51	republican	1.9×10^{-2}
52	audienc	1.8×10^{-2}
53	manuscript	1.8×10^{-2}
54	charact	1.8×10^{-2}
55	histori	1.8×10^{-2}
56	polit	1.8×10^{-2}
57	rather	1.8×10^{-2}
58	explor	1.8×10^{-2}
59	aesthet	1.7×10^{-2}
60	origin	1.7×10^{-2}
61	speaker	1.7×10^{-2}
62	method	1.7×10^{-2}
63	recept	1.7×10^{-2}
64	form	1.6×10^{-2}
65	measur	1.6×10^{-2}
66	music	1.6×10^{-2}
67	war	1.6×10^{-2}
68	engag	1.6×10^{-2}
69	way	1.6×10^{-2}
70	translat	1.6×10^{-2}
71	tension	1.6×10^{-2}
72	part	1.6×10^{-2}
73	reader	1.5×10^{-2}
74	life	1.5×10^{-2}
75	increas	1.5×10^{-2}
76	ritual	1.5×10^{-2}
77	focus	1.5×10^{-2}
78	dialect	1.5×10^{-2}
79	fiction	1.5×10^{-2}
80	level	1.4×10^{-2}
81	represent	1.4×10^{-2}
82	newspap	1.4×10^{-2}
83	journey	1.4×10^{-2}
84	evalu	1.4×10^{-2}
85	william	1.4×10^{-2}
86	possibl	1.3×10^{-2}
87	tradit	1.3×10^{-2}
88	genr	1.3×10^{-2}
89	scholar	1.3×10^{-2}
90	love	1.3×10^{-2}
91	metaphor	1.3×10^{-2}
92	conclus	1.3×10^{-2}
93	later	1.3×10^{-2}
94	dramat	1.3×10^{-2}
95	scene	1.3×10^{-2}
96	view	1.3×10^{-2}
97	tell	1.3×10^{-2}
98	writer	1.3×10^{-2}
99	theme	1.2×10^{-2}
100	someth	1.2×10^{-2}

Table D.201. The list of the top 100 words in the category Political Science with RIGs

No.	Word	RIG	No.	Word	RIG
1	polit	2.3×10^{-1}	51	obtain	1.3×10^{-2}
2	articl	1.2×10^{-1}	52	claim	1.3×10^{-2}
3	parti	8.1×10^{-2}	53	authoritarian	1.3×10^{-2}
4	argu	7.5×10^{-2}	54	foreign	1.3×10^{-2}
5	polici	7.4×10^{-2}	55	issu	1.3×10^{-2}
6	elector	6.7×10^{-2}	56	temperatur	1.3×10^{-2}
7	govern	6.5×10^{-2}	57	legal	1.2×10^{-2}
8	democrat	6.3×10^{-2}	58	legitimaci	1.2×10^{-2}
9	democraci	6.2×10^{-2}	59	economi	1.2×10^{-2}
10	elect	6.1×10^{-2}	60	perform	1.2×10^{-2}
11	voter	4.6×10^{-2}	61	violenc	1.2×10^{-2}
12	vote	4.6×10^{-2}	62	surfac	1.2×10^{-2}
13	countri	3.7×10^{-2}	63	peac	1.2×10^{-2}
14	nation	3.6×10^{-2}	64	justic	1.2×10^{-2}
15	institut	3.5×10^{-2}	65	societi	1.2×10^{-2}
16	citizen	3.5×10^{-2}	66	rule	1.2×10^{-2}
17	public	3.4×10^{-2}	67	campaign	1.1×10^{-2}
18	actor	3.1×10^{-2}	68	protest	1.1×10^{-2}
19	social	3×10^{-2}	69	negoti	1.1×10^{-2}
20	european	3×10^{-2}	70	engag	1.1×10^{-2}
21	econom	2.9×10^{-2}	71	simul	1.1×10^{-2}
22	union	2.6×10^{-2}	72	contest	1.1×10^{-2}
23	method	2.6×10^{-2}	73	militari	1.1×10^{-2}
24	state	2.5×10^{-2}	74	paramet	1.1×10^{-2}
25	ideolog	2.5×10^{-2}	75	transnat	1×10^{-2}
26	liber	2.4×10^{-2}	76	clinic	1×10^{-2}
27	empir	2.4×10^{-2}	77	explain	1×10^{-2}
28	argument	2.3×10^{-2}	78	seek	1×10^{-2}
29	war	2.2×10^{-2}	79	survey	1×10^{-2}
30	scholar	2.2×10^{-2}	80	discours	1×10^{-2}
31	legisl	2.2×10^{-2}	81	member	1×10^{-2}
32	conflict	2.1×10^{-2}	82	opinion	1×10^{-2}
33	result	2×10^{-2}	83	leader	1×10^{-2}
34	debat	2×10^{-2}	84	protein	9.9×10^{-3}
35	reform	1.9×10^{-2}	85	examin	9.8×10^{-3}
36	question	1.8×10^{-2}	86	detect	9.8×10^{-3}
37	intern	1.8×10^{-2}	87	focus	9.8×10^{-3}
38	draw	1.8×10^{-2}	88	regim	9.7×10^{-3}
39	right	1.7×10^{-2}	89	strateg	9.6×10^{-3}
40	patient	1.7×10^{-2}	90	literatur	9.4×10^{-3}
41	crisi	1.7×10^{-2}	91	diseas	9.4×10^{-3}
42	whi	1.6×10^{-2}	92	europ	9.4×10^{-3}
43	cell	1.6×10^{-2}	93	treatment	9.3×10^{-3}
44	domest	1.5×10^{-2}	94	american	9.3×10^{-3}
45	politician	1.5×10^{-2}	95	secur	9.2×10^{-3}
46	agenda	1.5×10^{-2}	96	high	9×10^{-3}
47	find	1.4×10^{-2}	97	world	9×10^{-3}
48	elit	1.4×10^{-2}	98	ratio	8.9×10^{-3}
49	civil	1.4×10^{-2}	99	feder	8.9×10^{-3}
50	presid	1.4×10^{-2}	100	rather	8.9×10^{-3}

Table D.202. The list of the top 100 words in the category Polymer Science with RIGs

No.	Word	RIG	No.	Word	RIG
1	polym	1.4×10^{-1}	51	character	1.6×10^{-2}
2	poli	1.1×10^{-1}	52	electron	1.6×10^{-2}
3	polymer	9.3×10^{-2}	53	gel	1.6×10^{-2}
4	copolym	7.4×10^{-2}	54	structur	1.6×10^{-2}
5	prepar	6.1×10^{-2}	55	stabil	1.6×10^{-2}
6	properti	5.5×10^{-2}	56	glass	1.6×10^{-2}
7	monom	4.3×10^{-2}	57	polyethylen	1.5×10^{-2}
8	synthes	3.6×10^{-2}	58	hydrophob	1.5×10^{-2}
9	chain	3.6×10^{-2}	59	glycol	1.5×10^{-2}
10	thermal	3.5×10^{-2}	60	hydrogel	1.5×10^{-2}
11	blend	3.2×10^{-2}	61	graft	1.5×10^{-2}
12	tensil	3.1×10^{-2}	62	aqueous	1.5×10^{-2}
13	methacryl	3×10^{-2}	63	swell	1.5×10^{-2}
14	ethylen	2.9×10^{-2}	64	matrix	1.5×10^{-2}
15	composit	2.9×10^{-2}	65	filler	1.4×10^{-2}
16	scan	2.8×10^{-2}	66	polystyren	1.4×10^{-2}
17	microscopi	2.7×10^{-2}	67	surfac	1.4×10^{-2}
18	crosslink	2.7×10^{-2}	68	paper	1.4×10^{-2}
19	spectroscopi	2.6×10^{-2}	69	fiber	1.4×10^{-2}
20	temperatur	2.6×10^{-2}	70	radic	1.4×10^{-2}
21	morpholog	2.6×10^{-2}	71	weight	1.4×10^{-2}
22	calorimetri	2.6×10^{-2}	72	permeat	1.4×10^{-2}
23	ftir	2.5×10^{-2}	73	viscos	1.4×10^{-2}
24	nanocomposit	2.5×10^{-2}	74	materi	1.4×10^{-2}
25	modulus	2.4×10^{-2}	75	water	1.4×10^{-2}
26	melt	2.3×10^{-2}	76	year	1.3×10^{-2}
27	fourier	2.2×10^{-2}	77	membran	1.3×10^{-2}
28	dsc	2.2×10^{-2}	78	background	1.3×10^{-2}
29	cellulos	2.1×10^{-2}	79	solut	1.3×10^{-2}
30	nmr	2.1×10^{-2}	80	chitosan	1.2×10^{-2}
31	infrar	2.1×10^{-2}	81	content	1.2×10^{-2}
32	molecular	2.1×10^{-2}	82	resin	1.2×10^{-2}
33	patient	2.1×10^{-2}	83	epoxi	1.2×10^{-2}
34	conclus	2×10^{-2}	84	reaction	1.2×10^{-2}
35	film	2×10^{-2}	85	exhibit	1.2×10^{-2}
36	thermogravimetr	2×10^{-2}	86	nanoparticl	1.2×10^{-2}
37	solvent	2×10^{-2}	87	selfassembl	1.1×10^{-2}
38	acid	1.9×10^{-2}	88	ray	1.1×10^{-2}
39	acryl	1.9×10^{-2}	89	associ	1.1×10^{-2}
40	rheolog	1.9×10^{-2}	90	data	1.1×10^{-2}
41	degre	1.9×10^{-2}	91	solubl	1.1×10^{-2}
42	strength	1.9×10^{-2}	92	backbon	1.1×10^{-2}
43	mechan	1.8×10^{-2}	93	chemic	1×10^{-2}
44	dispers	1.8×10^{-2}	94	transit	1×10^{-2}
45	tga	1.8×10^{-2}	95	fabric	1×10^{-2}
46	sem	1.7×10^{-2}	96	risk	1×10^{-2}
47	crystallin	1.7×10^{-2}	97	behavior	1×10^{-2}
48	polypropylen	1.7×10^{-2}	98	via	9.9×10^{-3}
49	vinyl	1.7×10^{-2}	99	micell	9.9×10^{-3}
50	hydrophil	1.6×10^{-2}	100	cure	9.8×10^{-3}

Table D.203. The list of the top 100 words in the category Primary Health Care with RIGs

No.	Word	RIG
1	care	1.8×10^{-1}
2	background	1.1×10^{-1}
3	patient	1.1×10^{-1}
4	primari	1.1×10^{-1}
5	health	9.2×10^{-2}
6	conclus	7.9×10^{-2}
7	practic	7.5×10^{-2}
8	physician	7.4×10^{-2}
9	gps	7.2×10^{-2}
10	practition	7.1×10^{-2}
11	medic	6.4×10^{-2}
12	consult	4.4×10^{-2}
13	particip	4×10^{-2}
14	interview	3.9×10^{-2}
15	general	3.4×10^{-2}
16	clinic	3.4×10^{-2}
17	method	3.2×10^{-2}
18	famili	3.1×10^{-2}
19	chronic	3.1×10^{-2}
20	manag	2.8×10^{-2}
21	outcom	2.7×10^{-2}
22	referr	2.6×10^{-2}
23	year	2.6×10^{-2}
24	object	2.5×10^{-2}
25	profession	2.5×10^{-2}
26	guidelin	2.5×10^{-2}
27	medicin	2.5×10^{-2}
28	educ	2.4×10^{-2}
29	intervent	2.4×10^{-2}
30	doctor	2.4×10^{-2}
31	aim	2.4×10^{-2}
32	semistructur	2.3×10^{-2}
33	servic	2.3×10^{-2}
34	need	2.3×10^{-2}
35	qualit	2.2×10^{-2}
36	recommend	2.2×10^{-2}
37	risk	2.2×10^{-2}
38	prescrib	2.2×10^{-2}
39	diabet	2.2×10^{-2}
40	age	2.1×10^{-2}
41	visit	2.1×10^{-2}
42	set	2.1×10^{-2}
43	clinician	2.1×10^{-2}
44	questionnair	2×10^{-2}
45	symptom	2×10^{-2}
46	diseas	2×10^{-2}
47	survey	1.9×10^{-2}
48	nurs	1.8×10^{-2}
49	includ	1.8×10^{-2}
50	prescript	1.7×10^{-2}

No.	Word	RIG
51	common	1.7×10^{-2}
52	themat	1.7×10^{-2}
53	diagnosi	1.6×10^{-2}
54	communiti	1.6×10^{-2}
55	specialist	1.5×10^{-2}
56	assess	1.5×10^{-2}
57	team	1.5×10^{-2}
58	perceiv	1.5×10^{-2}
59	healthcar	1.4×10^{-2}
60	advic	1.4×10^{-2}
61	ill	1.4×10^{-2}
62	show	1.3×10^{-2}
63	copd	1.3×10^{-2}
64	propos	1.3×10^{-2}
65	staff	1.3×10^{-2}
66	prevent	1.3×10^{-2}
67	nation	1.3×10^{-2}
68	result	1.3×10^{-2}
69	train	1.3×10^{-2}
70	diagnos	1.3×10^{-2}
71	depress	1.3×10^{-2}
72	australia	1.3×10^{-2}
73	theme	1.3×10^{-2}
74	support	1.2×10^{-2}
75	peopl	1.2×10^{-2}
76	receiv	1.2×10^{-2}
77	explor	1.2×10^{-2}
78	preval	1.2×10^{-2}
79	screen	1.2×10^{-2}
80	popul	1.2×10^{-2}
81	australian	1.2×10^{-2}
82	attend	1.1×10^{-2}
83	group	1.1×10^{-2}
84	identifi	1.1×10^{-2}
85	obstruct	1.1×10^{-2}
86	treatment	1.1×10^{-2}
87	paper	1.1×10^{-2}
88	pain	1.1×10^{-2}
89	cell	1.1×10^{-2}
90	simul	1.1×10^{-2}
91	properti	1×10^{-2}
92	barrier	1×10^{-2}
93	qualiti	1×10^{-2}
94	transcrib	1×10^{-2}
95	month	1×10^{-2}
96	evid	9.9×10^{-3}
97	improv	9.9×10^{-3}
98	counsel	9.8×10^{-3}
99	routin	9.8×10^{-3}
100	adult	9.7×10^{-3}

Table D.204. The list of the top 100 words in the category Psychiatry with RIGs

No.	Word	RIG
1	disord	1.6×10^{-1}
2	depress	1.2×10^{-1}
3	symptom	9.1×10^{-2}
4	mental	8.7×10^{-2}
5	psychiatr	8.4×10^{-2}
6	schizophrenia	7.7×10^{-2}
7	conclus	5.7×10^{-2}
8	anxieti	5.5×10^{-2}
9	particip	5.3×10^{-2}
10	cognit	5.1×10^{-2}
11	associ	4.6×10^{-2}
12	assess	4×10^{-2}
13	patient	4×10^{-2}
14	clinic	3.8×10^{-2}
15	health	3.6×10^{-2}
16	score	3.5×10^{-2}
17	examin	3.3×10^{-2}
18	psychosi	3.3×10^{-2}
19	age	3.2×10^{-2}
20	suicid	3.2×10^{-2}
21	ill	3.1×10^{-2}
22	treatment	3×10^{-2}
23	emot	3×10^{-2}
24	mood	3×10^{-2}
25	dsm	3×10^{-2}
26	studi	3×10^{-2}
27	deficit	2.9×10^{-2}
28	adolesc	2.9×10^{-2}
29	psychot	2.9×10^{-2}
30	bipolar	2.7×10^{-2}
31	background	2.7×10^{-2}
32	interview	2.7×10^{-2}
33	intervent	2.6×10^{-2}
34	psychopatholog	2.5×10^{-2}
35	risk	2.5×10^{-2}
36	individu	2.5×10^{-2}
37	person	2.5×10^{-2}
38	social	2.4×10^{-2}
39	adult	2.4×10^{-2}
40	selfreport	2.4×10^{-2}
41	questionnair	2.3×10^{-2}
42	psycholog	2.3×10^{-2}
43	ptsd	2.3×10^{-2}
44	posttraumat	2.3×10^{-2}
45	comorbid	2.3×10^{-2}
46	healthi	2.2×10^{-2}
47	group	2.2×10^{-2}
48	impair	2.2×10^{-2}
49	antidepress	2.2×10^{-2}
50	may	2.1×10^{-2}
51	medic	2.1×10^{-2}
52	paper	2×10^{-2}
53	abus	2×10^{-2}
54	signific	2×10^{-2}
55	object	1.9×10^{-2}
56	adhd	1.9×10^{-2}
57	find	1.8×10^{-2}
58	year	1.8×10^{-2}
59	hyperact	1.8×10^{-2}
60	behavior	1.7×10^{-2}
61	episod	1.7×10^{-2}
62	whether	1.7×10^{-2}
63	brain	1.7×10^{-2}
64	baselin	1.7×10^{-2}
65	outcom	1.6×10^{-2}
66	children	1.5×10^{-2}
67	suggest	1.5×10^{-2}
68	control	1.5×10^{-2}
69	aim	1.5×10^{-2}
70	prefront	1.5×10^{-2}
71	relat	1.5×10^{-2}
72	sampl	1.5×10^{-2}
73	method	1.5×10^{-2}
74	inventori	1.5×10^{-2}
75	simul	1.4×10^{-2}
76	scale	1.4×10^{-2}
77	distress	1.4×10^{-2}
78	temperatur	1.4×10^{-2}
79	outpati	1.4×10^{-2}
80	relationship	1.4×10^{-2}
81	regress	1.4×10^{-2}
82	among	1.4×10^{-2}
83	diagnos	1.4×10^{-2}
84	subject	1.3×10^{-2}
85	alcohol	1.3×10^{-2}
86	surfac	1.3×10^{-2}
87	addict	1.3×10^{-2}
88	substanc	1.3×10^{-2}
89	cortex	1.3×10^{-2}
90	care	1.3×10^{-2}
91	peopl	1.3×10^{-2}
92	preval	1.3×10^{-2}
93	negat	1.3×10^{-2}
94	gender	1.3×10^{-2}
95	propos	1.2×10^{-2}
96	psychosoci	1.2×10^{-2}
97	parent	1.2×10^{-2}
98	child	1.2×10^{-2}
99	drug	1.2×10^{-2}
100	childhood	1.2×10^{-2}

TABLE D.205. The list of the top 100 words in the category Psychology with RIGs

No.	Word	RIG	No.	Word	RIG
1	particip	6.9×10^{-2}	51	temperatur	1.2×10^{-2}
2	task	5.9×10^{-2}	52	paradigm	1.2×10^{-2}
3	cognit	5.5×10^{-2}	53	conclus	1.1×10^{-2}
4	memori	3.1×10^{-2}	54	score	1.1×10^{-2}
5	emot	2.9×10^{-2}	55	auditori	1.1×10^{-2}
6	disord	2.9×10^{-2}	56	research	1.1×10^{-2}
7	stimuli	2.8×10^{-2}	57	item	1×10^{-2}
8	examin	2.4×10^{-2}	58	subject	1×10^{-2}
9	brain	2.4×10^{-2}	59	learn	1×10^{-2}
10	stimulus	2.4×10^{-2}	60	selfreport	1×10^{-2}
11	depress	2.3×10^{-2}	61	fmri	1×10^{-2}
12	whether	2.3×10^{-2}	62	may	1×10^{-2}
13	neuropsycholog	2.1×10^{-2}	63	older	9.8×10^{-3}
14	adult	2.1×10^{-2}	64	evid	9.8×10^{-3}
15	suggest	2×10^{-2}	65	group	9.7×10^{-3}
16	cortex	2×10^{-2}	66	result	9.7×10^{-3}
17	age	1.9×10^{-2}	67	test	9.4×10^{-3}
18	mental	1.9×10^{-2}	68	pariet	9.3×10^{-3}
19	psycholog	1.9×10^{-2}	69	questionnair	9.2×10^{-3}
20	perceptu	1.9×10^{-2}	70	motor	9.1×10^{-3}
21	visual	1.9×10^{-2}	71	cingul	9×10^{-3}
22	attent	1.9×10^{-2}	72	relationship	8.8×10^{-3}
23	individu	1.8×10^{-2}	73	energi	8.7×10^{-3}
24	symptom	1.8×10^{-2}	74	execut	8.5×10^{-3}
25	associ	1.8×10^{-2}	75	eeg	8.4×10^{-3}
26	anxieti	1.8×10^{-2}	76	erp	8.3×10^{-3}
27	object	1.7×10^{-2}	77	experi	8.2×10^{-3}
28	relat	1.7×10^{-2}	78	trial	8.2×10^{-3}
29	studi	1.7×10^{-2}	79	session	8×10^{-3}
30	find	1.7×10^{-2}	80	negat	7.9×10^{-3}
31	impair	1.6×10^{-2}	81	control	7.8×10^{-3}
32	deficit	1.6×10^{-2}	82	left	7.7×10^{-3}
33	healthi	1.6×10^{-2}	83	cortic	7.7×10^{-3}
34	social	1.6×10^{-2}	84	greater	7.6×10^{-3}
35	percept	1.5×10^{-2}	85	amygdala	7.6×10^{-3}
36	assess	1.5×10^{-2}	86	fear	7.6×10^{-3}
37	paper	1.5×10^{-2}	87	electron	7.6×10^{-3}
38	person	1.5×10^{-2}	88	support	7.6×10^{-3}
39	cue	1.4×10^{-2}	89	intervent	7.3×10^{-3}
40	prefront	1.4×10^{-2}	90	judgment	7.3×10^{-3}
41	behavior	1.4×10^{-2}	91	psychiatr	7.3×10^{-3}
42	verbal	1.3×10^{-2}	92	surfac	7.3×10^{-3}
43	adolesc	1.3×10^{-2}	93	peopl	7.1×10^{-3}
44	cell	1.3×10^{-2}	94	dure	7×10^{-3}
45	neural	1.3×10^{-2}	95	respons	7×10^{-3}
46	perceiv	1.3×10^{-2}	96	acid	6.9×10^{-3}
47	across	1.3×10^{-2}	97	gyrus	6.9×10^{-3}
48	children	1.2×10^{-2}	98	tempor	6.9×10^{-3}
49	word	1.2×10^{-2}	99	eat	6.8×10^{-3}
50	frontal	1.2×10^{-2}	100	skill	6.8×10^{-3}

Table D.206. The list of the top 100 words in the category Psychology, Applied with RIGs

No.	Word	RIG
1	employe	6.3×10^{-2}
2	research	5.8×10^{-2}
3	organiz	5.1×10^{-2}
4	job	4.3×10^{-2}
5	implic	3.9×10^{-2}
6	psycholog	3.4×10^{-2}
7	examin	3.4×10^{-2}
8	perceiv	3.2×10^{-2}
9	relationship	3.2×10^{-2}
10	particip	3.1×10^{-2}
11	career	3.1×10^{-2}
12	person	2.6×10^{-2}
13	find	2.5×10^{-2}
14	social	2.3×10^{-2}
15	percept	2.1×10^{-2}
16	practic	2×10^{-2}
17	theori	1.9×10^{-2}
18	discuss	1.9×10^{-2}
19	leadership	1.8×10^{-2}
20	individu	1.8×10^{-2}
21	studi	1.7×10^{-2}
22	satisfact	1.7×10^{-2}
23	support	1.6×10^{-2}
24	task	1.6×10^{-2}
25	emot	1.5×10^{-2}
26	moder	1.5×10^{-2}
27	cell	1.5×10^{-2}
28	supervisor	1.5×10^{-2}
29	practition	1.5×10^{-2}
30	workplac	1.5×10^{-2}
31	work	1.4×10^{-2}
32	student	1.4×10^{-2}
33	leader	1.4×10^{-2}
34	relat	1.4×10^{-2}
35	sport	1.3×10^{-2}
36	team	1.3×10^{-2}
37	cognit	1.3×10^{-2}
38	athlet	1.3×10^{-2}
39	engag	1.3×10^{-2}
40	behavior	1.2×10^{-2}
41	worker	1.2×10^{-2}
42	manag	1.2×10^{-2}
43	futur	1.2×10^{-2}
44	interview	1.2×10^{-2}
45	posit	1.1×10^{-2}
46	motiv	1.1×10^{-2}
47	hypothes	1.1×10^{-2}
48	selfefficaci	1.1×10^{-2}
49	literatur	1.1×10^{-2}
50	counsel	1×10^{-2}

No.	Word	RIG
51	conceptu	1×10^{-2}
52	anteced	1×10^{-2}
53	patient	1×10^{-2}
54	coach	9.6×10^{-3}
55	perspect	9.6×10^{-3}
56	negat	9.3×10^{-3}
57	offend	9.1×10^{-3}
58	temperatur	9×10^{-3}
59	context	9×10^{-3}
60	selfreport	9×10^{-3}
61	skill	8.9×10^{-3}
62	empir	8.7×10^{-3}
63	commit	8.6×10^{-3}
64	profession	8.6×10^{-3}
65	survey	8.5×10^{-3}
66	intent	8.5×10^{-3}
67	victim	8.4×10^{-3}
68	across	8.4×10^{-3}
69	colleg	8.3×10^{-3}
70	gender	8.3×10^{-3}
71	protein	8.3×10^{-3}
72	decis	8.3×10^{-3}
73	peopl	8.2×10^{-3}
74	mediat	8.2×10^{-3}
75	attitud	8.2×10^{-3}
76	focus	8.2×10^{-3}
77	surfac	8.1×10^{-3}
78	knowledg	7.9×10^{-3}
79	violenc	7.9×10^{-3}
80	interperson	7.9×10^{-3}
81	questionnair	7.9×10^{-3}
82	driver	7.8×10^{-3}
83	articl	7.7×10^{-3}
84	understand	7.7×10^{-3}
85	acid	7.7×10^{-3}
86	outcom	7.6×10^{-3}
87	energi	7.5×10^{-3}
88	ergonom	7.5×10^{-3}
89	train	7.4×10^{-3}
90	purpos	7.3×10^{-3}
91	justic	7.3×10^{-3}
92	mental	7.2×10^{-3}
93	compani	7.1×10^{-3}
94	speci	7×10^{-3}
95	resourc	7×10^{-3}
96	materi	6.9×10^{-3}
97	conflict	6.9×10^{-3}
98	paramet	6.8×10^{-3}
99	organ	6.8×10^{-3}
100	busi	6.7×10^{-3}

TABLE D.207. The list of the top 100 words in the category Psychology, Biological with RIGs

No.	Word	RIG
1	task	4.8×10^{-2}
2	stimulus	4.6×10^{-2}
3	stimuli	4.2×10^{-2}
4	cue	3.6×10^{-2}
5	behavior	3.5×10^{-2}
6	suggest	3×10^{-2}
7	respons	2.9×10^{-2}
8	particip	2.8×10^{-2}
9	cognit	2.6×10^{-2}
10	whether	2.3×10^{-2}
11	erp	2.1×10^{-2}
12	anim	2.1×10^{-2}
13	male	2×10^{-2}
14	method	1.9×10^{-2}
15	individu	1.9×10^{-2}
16	rat	1.8×10^{-2}
17	food	1.7×10^{-2}
18	femal	1.7×10^{-2}
19	paper	1.7×10^{-2}
20	relat	1.7×10^{-2}
21	emot	1.7×10^{-2}
22	eeg	1.7×10^{-2}
23	learn	1.6×10^{-2}
24	mate	1.6×10^{-2}
25	maze	1.6×10^{-2}
26	test	1.4×10^{-2}
27	respond	1.4×10^{-2}
28	adult	1.4×10^{-2}
29	dure	1.4×10^{-2}
30	memori	1.4×10^{-2}
31	reward	1.3×10^{-2}
32	paradigm	1.3×10^{-2}
33	social	1.3×10^{-2}
34	brain	1.3×10^{-2}
35	trial	1.3×10^{-2}
36	attent	1.2×10^{-2}
37	may	1.2×10^{-2}
38	behaviour	1.2×10^{-2}
39	session	1.2×10^{-2}
40	choic	1.2×10^{-2}
41	experi	1.2×10^{-2}
42	hypothesi	1.2×10^{-2}
43	manipul	1.2×10^{-2}
44	examin	1.2×10^{-2}
45	anxieti	1.1×10^{-2}
46	discrimin	1.1×10^{-2}
47	either	1×10^{-2}
48	prefer	1×10^{-2}
49	differ	1×10^{-2}
50	physiolog	1×10^{-2}
51	conspecif	1×10^{-2}
52	entitl	1×10^{-2}
53	cell	9.8×10^{-3}
54	evid	9.7×10^{-3}
55	train	9.7×10^{-3}
56	reinforc	9.6×10^{-3}
57	elicit	9.6×10^{-3}
58	cortisol	9.5×10^{-3}
59	auditori	9.3×10^{-3}
60	avers	9.3×10^{-3}
61	associ	9.3×10^{-3}
62	intak	9.1×10^{-3}
63	latenc	8.9×10^{-3}
64	bird	8.9×10^{-3}
65	find	8.9×10^{-3}
66	applic	8.4×10^{-3}
67	forag	8.4×10^{-3}
68	trait	8.4×10^{-3}
69	amplitud	8.4×10^{-3}
70	subject	8.4×10^{-3}
71	predat	8.2×10^{-3}
72	spent	8.2×10^{-3}
73	psycholog	8.1×10^{-3}
74	surfac	8.1×10^{-3}
75	visual	8.1×10^{-3}
76	arous	7.9×10^{-3}
77	condit	7.6×10^{-3}
78	effect	7.5×10^{-3}
79	electron	7.1×10^{-3}
80	negat	7×10^{-3}
81	aggress	7×10^{-3}
82	propos	6.9×10^{-3}
83	howev	6.9×10^{-3}
84	greater	6.9×10^{-3}
85	perceptu	6.8×10^{-3}
86	across	6.7×10^{-3}
87	affect	6.5×10^{-3}
88	sex	6.5×10^{-3}
89	conting	6.4×10^{-3}
90	materi	6.4×10^{-3}
91	word	6.3×10^{-3}
92	base	6.3×10^{-3}
93	indic	6.2×10^{-3}
94	studi	6.1×10^{-3}
95	stressor	6×10^{-3}
96	cortex	6×10^{-3}
97	tempor	5.9×10^{-3}
98	healthi	5.9×10^{-3}
99	sexual	5.9×10^{-3}
100	conclus	5.8×10^{-3}

Table D.208. The list of the top 100 words in the category Psychology, Clinical with RIGs

No.	Word	RIG
1	disord	8.4×10^{-2}
2	particip	7.4×10^{-2}
3	depress	6.8×10^{-2}
4	symptom	6.7×10^{-2}
5	anxieti	5.7×10^{-2}
6	cognit	5.2×10^{-2}
7	examin	4.9×10^{-2}
8	intervent	4.4×10^{-2}
9	emot	4×10^{-2}
10	psycholog	3.9×10^{-2}
11	mental	3.7×10^{-2}
12	selfreport	3.6×10^{-2}
13	adolesc	3.5×10^{-2}
14	posttraumat	3.3×10^{-2}
15	behavior	3.2×10^{-2}
16	ptsd	3.1×10^{-2}
17	assess	3×10^{-2}
18	studi	2.8×10^{-2}
19	research	2.8×10^{-2}
20	associ	2.8×10^{-2}
21	social	2.6×10^{-2}
22	person	2.6×10^{-2}
23	health	2.6×10^{-2}
24	relationship	2.6×10^{-2}
25	individu	2.4×10^{-2}
26	sexual	2.3×10^{-2}
27	questionnair	2.3×10^{-2}
28	abus	2.3×10^{-2}
29	eat	2.2×10^{-2}
30	complet	2.2×10^{-2}
31	clinic	2.2×10^{-2}
32	age	2.2×10^{-2}
33	find	2.2×10^{-2}
34	sampl	2.1×10^{-2}
35	treatment	2×10^{-2}
36	adult	2×10^{-2}
37	parent	2×10^{-2}
38	session	2×10^{-2}
39	interview	2×10^{-2}
40	child	2×10^{-2}
41	distress	1.9×10^{-2}
42	outcom	1.9×10^{-2}
43	score	1.9×10^{-2}
44	therapist	1.9×10^{-2}
45	children	1.9×10^{-2}
46	women	1.9×10^{-2}
47	psychopatholog	1.9×10^{-2}
48	psychotherapi	1.8×10^{-2}
49	psychiatr	1.8×10^{-2}
50	implic	1.8×10^{-2}

No.	Word	RIG
51	dsm	1.8×10^{-2}
52	among	1.7×10^{-2}
53	paper	1.7×10^{-2}
54	youth	1.7×10^{-2}
55	trauma	1.6×10^{-2}
56	cell	1.6×10^{-2}
57	relat	1.6×10^{-2}
58	substanc	1.5×10^{-2}
59	negat	1.5×10^{-2}
60	femal	1.5×10^{-2}
61	moder	1.5×10^{-2}
62	whether	1.5×10^{-2}
63	therapi	1.5×10^{-2}
64	suggest	1.5×10^{-2}
65	interperson	1.4×10^{-2}
66	conclus	1.4×10^{-2}
67	psychometr	1.4×10^{-2}
68	inventori	1.3×10^{-2}
69	temperatur	1.3×10^{-2}
70	group	1.3×10^{-2}
71	baselin	1.3×10^{-2}
72	surfac	1.3×10^{-2}
73	across	1.3×10^{-2}
74	risk	1.3×10^{-2}
75	year	1.2×10^{-2}
76	may	1.2×10^{-2}
77	support	1.2×10^{-2}
78	alcohol	1.2×10^{-2}
79	partner	1.2×10^{-2}
80	mood	1.2×10^{-2}
81	gender	1.2×10^{-2}
82	perceiv	1.2×10^{-2}
83	subscal	1.2×10^{-2}
84	simul	1.2×10^{-2}
85	student	1.2×10^{-2}
86	item	1.2×10^{-2}
87	childhood	1.1×10^{-2}
88	predictor	1.1×10^{-2}
89	system	1.1×10^{-2}
90	engag	1.1×10^{-2}
91	posttreat	1.1×10^{-2}
92	psychosoci	1.1×10^{-2}
93	object	1×10^{-2}
94	clinician	1×10^{-2}
95	violenc	1×10^{-2}
96	stress	1×10^{-2}
97	traumat	1×10^{-2}
98	confirmatori	1×10^{-2}
99	longitudin	1×10^{-2}
100	comorbid	1×10^{-2}

Table D.209. The list of the top 100 words in the category Psychology, Developmental with RIGs

No.	Word	RIG
1	children	1.8×10^{-1}
2	adolesc	1.2×10^{-1}
3	parent	9.7×10^{-2}
4	age	8.7×10^{-2}
5	child	8.7×10^{-2}
6	autism	8.3×10^{-2}
7	examin	5.6×10^{-2}
8	youth	5.6×10^{-2}
9	asd	5.4×10^{-2}
10	social	5×10^{-2}
11	year	4.9×10^{-2}
12	development	4.9×10^{-2}
13	school	4.8×10^{-2}
14	disord	4.7×10^{-2}
15	particip	4.3×10^{-2}
16	behavior	4.1×10^{-2}
17	peer	4×10^{-2}
18	mother	4×10^{-2}
19	emot	3.7×10^{-2}
20	old	3.5×10^{-2}
21	preschool	3.5×10^{-2}
22	childhood	3.2×10^{-2}
23	girl	3.2×10^{-2}
24	boy	3.1×10^{-2}
25	longitudin	3.1×10^{-2}
26	cognit	3×10^{-2}
27	intervent	3×10^{-2}
28	symptom	2.9×10^{-2}
29	young	2.8×10^{-2}
30	find	2.6×10^{-2}
31	adult	2.6×10^{-2}
32	earli	2.6×10^{-2}
33	skill	2.5×10^{-2}
34	famili	2.4×10^{-2}
35	depress	2.4×10^{-2}
36	gender	2.3×10^{-2}
37	adhd	2.3×10^{-2}
38	infant	2.3×10^{-2}
39	studi	2.2×10^{-2}
40	paper	2.2×10^{-2}
41	spectrum	2.1×10^{-2}
42	associ	2.1×10^{-2}
43	research	2×10^{-2}
44	task	2×10^{-2}
45	matern	1.9×10^{-2}
46	across	1.9×10^{-2}
47	implic	1.9×10^{-2}
48	mental	1.9×10^{-2}
49	selfreport	1.9×10^{-2}
50	teacher	1.8×10^{-2}

No.	Word	RIG
51	assess	1.8×10^{-2}
52	whether	1.8×10^{-2}
53	relat	1.7×10^{-2}
54	hyperact	1.6×10^{-2}
55	anxieti	1.6×10^{-2}
56	relationship	1.6×10^{-2}
57	cell	1.6×10^{-2}
58	languag	1.5×10^{-2}
59	month	1.5×10^{-2}
60	suggest	1.4×10^{-2}
61	deficit	1.4×10^{-2}
62	attent	1.4×10^{-2}
63	adulthood	1.4×10^{-2}
64	father	1.4×10^{-2}
65	caregiv	1.4×10^{-2}
66	support	1.3×10^{-2}
67	verbal	1.3×10^{-2}
68	engag	1.3×10^{-2}
69	temperatur	1.3×10^{-2}
70	moder	1.3×10^{-2}
71	psycholog	1.2×10^{-2}
72	predict	1.2×10^{-2}
73	risk	1.1×10^{-2}
74	psychopatholog	1.1×10^{-2}
75	group	1.1×10^{-2}
76	older	1.1×10^{-2}
77	questionnair	1.1×10^{-2}
78	sampl	1.1×10^{-2}
79	simul	1.1×10^{-2}
80	health	1.1×10^{-2}
81	surfac	1.1×10^{-2}
82	energi	1.1×10^{-2}
83	system	1.1×10^{-2}
84	individu	1.1×10^{-2}
85	ethnic	1×10^{-2}
86	propos	1×10^{-2}
87	dyad	1×10^{-2}
88	difficulti	9.8×10^{-3}
89	student	9.5×10^{-3}
90	femal	9.5×10^{-3}
91	discuss	9.4×10^{-3}
92	educ	9.4×10^{-3}
93	outcom	9.3×10^{-3}
94	among	9.3×10^{-3}
95	report	9.2×10^{-3}
96	applic	9.2×10^{-3}
97	victim	9.1×10^{-3}
98	score	9×10^{-3}
99	paramet	9×10^{-3}
100	interview	8.7×10^{-3}

Table D.210. The list of the top 100 words in the category Psychology, Educational with RIGs

No.	Word	RIG	No.	Word	RIG
1	student	1.4×10^{-1}	51	english	1.2×10^{-2}
2	school	1.2×10^{-1}	52	fluenci	1.2×10^{-2}
3	teacher	7.3×10^{-2}	53	text	1.2×10^{-2}
4	children	6.2×10^{-2}	54	boy	1.2×10^{-2}
5	educ	5.5×10^{-2}	55	temperatur	1.2×10^{-2}
6	learn	5×10^{-2}	56	task	1.1×10^{-2}
7	read	4.3×10^{-2}	57	girl	1.1×10^{-2}
8	grade	4.3×10^{-2}	58	compet	1.1×10^{-2}
9	academ	4×10^{-2}	59	confirmatori	1.1×10^{-2}
10	skill	3.9×10^{-2}	60	perceiv	1.1×10^{-2}
11	classroom	3.8×10^{-2}	61	knowledg	1.1×10^{-2}
12	research	3.8×10^{-2}	62	abil	1.1×10^{-2}
13	implic	3.5×10^{-2}	63	math	1.1×10^{-2}
14	particip	3.3×10^{-2}	64	system	1×10^{-2}
15	examin	3.1×10^{-2}	65	creativ	1×10^{-2}
16	instruct	3×10^{-2}	66	gender	1×10^{-2}
17	discuss	2.3×10^{-2}	67	assess	1×10^{-2}
18	studi	2.3×10^{-2}	68	longitudin	9.8×10^{-3}
19	cognit	2.3×10^{-2}	69	energi	9.7×10^{-3}
20	literaci	2.2×10^{-2}	70	reader	9.6×10^{-3}
21	languag	2.1×10^{-2}	71	predictor	9.5×10^{-3}
22	find	2.1×10^{-2}	72	profession	9.4×10^{-3}
23	social	1.9×10^{-2}	73	person	9.3×10^{-3}
24	preschool	1.9×10^{-2}	74	psychometr	9.2×10^{-3}
25	emot	1.9×10^{-2}	75	relationship	9×10^{-3}
26	peer	1.8×10^{-2}	76	context	8.9×10^{-3}
27	word	1.8×10^{-2}	77	method	8.8×10^{-3}
28	psycholog	1.8×10^{-2}	78	protein	8.8×10^{-3}
29	intervent	1.6×10^{-2}	79	age	8.8×10^{-3}
30	support	1.6×10^{-2}	80	undergradu	8.7×10^{-3}
31	practic	1.6×10^{-2}	81	percept	8.6×10^{-3}
32	adolesc	1.6×10^{-2}	82	latent	8.5×10^{-3}
33	item	1.5×10^{-2}	83	relat	8.3×10^{-3}
34	cell	1.5×10^{-2}	84	belief	8.3×10^{-3}
35	motiv	1.5×10^{-2}	85	think	8.3×10^{-3}
36	phonolog	1.5×10^{-2}	86	questionnair	8.2×10^{-3}
37	learner	1.5×10^{-2}	87	varianc	8.2×10^{-3}
38	teach	1.5×10^{-2}	88	youth	8.1×10^{-3}
39	psychologist	1.5×10^{-2}	89	articl	8.1×10^{-3}
40	child	1.4×10^{-2}	90	comprehens	8×10^{-3}
41	parent	1.4×10^{-2}	91	selfreport	7.9×10^{-3}
42	score	1.4×10^{-2}	92	surfac	7.9×10^{-3}
43	vocabulari	1.4×10^{-2}	93	write	7.8×10^{-3}
44	patient	1.4×10^{-2}	94	year	7.7×10^{-3}
45	elementari	1.3×10^{-2}	95	analys	7.7×10^{-3}
46	verbal	1.3×10^{-2}	96	concentr	7.7×10^{-3}
47	engag	1.3×10^{-2}	97	conceptu	7.7×10^{-3}
48	selfefficaci	1.3×10^{-2}	98	water	7.7×10^{-3}
49	across	1.3×10^{-2}	99	multilevel	7.4×10^{-3}
50	paper	1.3×10^{-2}	100	whether	7.3×10^{-3}

Table D.211. The list of the top 100 words in the category Psychology, Experimental with RIGs

No.	Word	RIG	No.	Word	RIG
1	task	9.7×10^{-2}	51	inform	1.2×10^{-2}
2	particip	8×10^{-2}	52	elicit	1.2×10^{-2}
3	cognit	5×10^{-2}	53	peopl	1.2×10^{-2}
4	stimuli	4.8×10^{-2}	54	research	1.1×10^{-2}
5	memori	4.5×10^{-2}	55	psycholog	1.1×10^{-2}
6	word	3.9×10^{-2}	56	left	1.1×10^{-2}
7	stimulus	3.9×10^{-2}	57	rememb	1.1×10^{-2}
8	visual	3.5×10^{-2}	58	adult	1.1×10^{-2}
9	experi	3.3×10^{-2}	59	listen	1.1×10^{-2}
10	perceptu	3.3×10^{-2}	60	hypothesi	1.1×10^{-2}
11	whether	3.2×10^{-2}	61	movement	1.1×10^{-2}
12	suggest	3.2×10^{-2}	62	respons	1.1×10^{-2}
13	cue	3.1×10^{-2}	63	context	1×10^{-2}
14	emot	2.9×10^{-2}	64	conclus	1×10^{-2}
15	method	2.5×10^{-2}	65	tempor	1×10^{-2}
16	percept	2.3×10^{-2}	66	frontal	1×10^{-2}
17	paradigm	2.2×10^{-2}	67	read	9.9×10^{-3}
18	attent	2.2×10^{-2}	68	fmri	9.6×10^{-3}
19	lexic	2.2×10^{-2}	69	energi	9.6×10^{-3}
20	learn	2.2×10^{-2}	70	speech	9.5×10^{-3}
21	find	2.2×10^{-2}	71	motor	9.5×10^{-3}
22	judgment	2×10^{-2}	72	right	9.4×10^{-3}
23	manipul	1.9×10^{-2}	73	linguist	9.2×10^{-3}
24	erp	1.9×10^{-2}	74	treatment	9.2×10^{-3}
25	represent	1.8×10^{-2}	75	prefront	9.2×10^{-3}
26	languag	1.8×10^{-2}	76	instruct	9.2×10^{-3}
27	examin	1.8×10^{-2}	77	recognit	9.1×10^{-3}
28	process	1.7×10^{-2}	78	speaker	9.1×10^{-3}
29	auditori	1.6×10^{-2}	79	pictur	9.1×10^{-3}
30	paper	1.5×10^{-2}	80	implic	9.1×10^{-3}
31	item	1.5×10^{-2}	81	pariet	9×10^{-3}
32	cell	1.5×10^{-2}	82	eye	8.9×10^{-3}
33	neural	1.5×10^{-2}	83	trial	8.8×10^{-3}
34	cortex	1.5×10^{-2}	84	prime	8.7×10^{-3}
35	ask	1.4×10^{-2}	85	concentr	8.7×10^{-3}
36	semant	1.4×10^{-2}	86	protein	8.5×10^{-3}
37	perceiv	1.4×10^{-2}	87	bias	8.4×10^{-3}
38	children	1.4×10^{-2}	88	behavior	8.2×10^{-3}
39	recal	1.4×10^{-2}	89	acid	8.2×10^{-3}
40	brain	1.4×10^{-2}	90	gyrus	8.1×10^{-3}
41	sentenc	1.4×10^{-2}	91	applic	8.1×10^{-3}
42	relat	1.4×10^{-2}	92	test	8×10^{-3}
43	temperatur	1.3×10^{-2}	93	familiar	8×10^{-3}
44	across	1.3×10^{-2}	94	music	7.8×10^{-3}
45	individu	1.2×10^{-2}	95	view	7.7×10^{-3}
46	social	1.2×10^{-2}	96	noun	7.6×10^{-3}
47	phonolog	1.2×10^{-2}	97	electron	7.6×10^{-3}
48	verbal	1.2×10^{-2}	98	person	7.5×10^{-3}
49	evid	1.2×10^{-2}	99	spoken	7.3×10^{-3}
50	either	1.2×10^{-2}	100	abil	7.3×10^{-3}

TABLE D.212. The list of the top 100 words in the category Psychology, Mathematical with RIGs

No.	Word	RIG	No.	Word	RIG
1	item	8.1×10^{-2}	51	experi	7.4×10^{-3}
2	latent	3.2×10^{-2}	52	manipul	7.2×10^{-3}
3	stimulus	2.3×10^{-2}	53	english	7.2×10^{-3}
4	task	2.2×10^{-2}	54	concentr	7.1×10^{-3}
5	cognit	2.1×10^{-2}	55	individu	7.1×10^{-3}
6	visual	1.9×10^{-2}	56	varianc	6.8×10^{-3}
7	particip	1.9×10^{-2}	57	variabl	6.7×10^{-3}
8	articl	1.9×10^{-2}	58	inform	6.7×10^{-3}
9	estim	1.8×10^{-2}	59	surfac	6.6×10^{-3}
10	research	1.7×10^{-2}	60	acid	6.3×10^{-3}
11	word	1.7×10^{-2}	61	electron	6.3×10^{-3}
12	respons	1.6×10^{-2}	62	diseas	6.3×10^{-3}
13	psycholog	1.5×10^{-2}	63	present	6.2×10^{-3}
14	simul	1.5×10^{-2}	64	abil	6.1×10^{-3}
15	empir	1.4×10^{-2}	65	statist	6×10^{-3}
16	bias	1.4×10^{-2}	66	languag	6×10^{-3}
17	procedur	1.3×10^{-2}	67	two	6×10^{-3}
18	lexic	1.3×10^{-2}	68	categori	5.9×10^{-3}
19	stimuli	1.3×10^{-2}	69	gene	5.9×10^{-3}
20	test	1.3×10^{-2}	70	set	5.7×10^{-3}
21	perceptu	1.2×10^{-2}	71	instruct	5.6×10^{-3}
22	across	1.2×10^{-2}	72	psychometr	5.6×10^{-3}
23	likelihood	1.2×10^{-2}	73	assum	5.5×10^{-3}
24	cell	1.2×10^{-2}	74	discuss	5.5×10^{-3}
25	memori	1.1×10^{-2}	75	mass	5.4×10^{-3}
26	model	1.1×10^{-2}	76	learn	5.3×10^{-3}
27	covari	1.1×10^{-2}	77	semant	5.3×10^{-3}
28	error	1.1×10^{-2}	78	water	5.2×10^{-3}
29	paradigm	1×10^{-2}	79	often	5.1×10^{-3}
30	temperatur	9.8×10^{-3}	80	whether	5.1×10^{-3}
31	patient	9.6×10^{-3}	81	verbal	5.1×10^{-3}
32	score	9.5×10^{-3}	82	oxid	5×10^{-3}
33	data	9.4×10^{-3}	83	molecular	4.9×10^{-3}
34	theori	9.2×10^{-3}	84	behavior	4.9×10^{-3}
35	multilevel	9×10^{-3}	85	infer	4.8×10^{-3}
36	bayesian	8.8×10^{-3}	86	size	4.8×10^{-3}
37	illustr	8.6×10^{-3}	87	scienc	4.7×10^{-3}
38	journal	8.4×10^{-3}	88	reliabl	4.7×10^{-3}
39	energi	8.3×10^{-3}	89	context	4.7×10^{-3}
40	fit	8.1×10^{-3}	90	activ	4.7×10^{-3}
41	exampl	8×10^{-3}	91	percept	4.6×10^{-3}
42	assumpt	8×10^{-3}	92	trial	4.6×10^{-3}
43	conclus	7.9×10^{-3}	93	high	4.5×10^{-3}
44	hierarch	7.8×10^{-3}	94	attent	4.5×10^{-3}
45	judgment	7.8×10^{-3}	95	paper	4.5×10^{-3}
46	general	7.7×10^{-3}	96	cue	4.5×10^{-3}
47	carlo	7.5×10^{-3}	97	account	4.5×10^{-3}
48	protein	7.4×10^{-3}	98	read	4.5×10^{-3}
49	mont	7.4×10^{-3}	99	familiar	4.5×10^{-3}
50	student	7.4×10^{-3}	100	chemic	4.4×10^{-3}

Table D.213. The list of the top 100 words in the category Psychology, Multidisciplinary with RIGs

No.	Word	RIG	No.	Word	RIG
1	psycholog	6.1×10^{-2}	51	experi	9.5×10^{-3}
2	particip	5.1×10^{-2}	52	motiv	9.1×10^{-3}
3	social	4.3×10^{-2}	53	interview	9.1×10^{-3}
4	research	3.6×10^{-2}	54	memori	8.9×10^{-3}
5	emot	3.5×10^{-2}	55	child	8.8×10^{-3}
6	cognit	3.1×10^{-2}	56	life	8.8×10^{-3}
7	person	3×10^{-2}	57	belief	8.8×10^{-3}
8	student	2.7×10^{-2}	58	interperson	8.6×10^{-3}
9	examin	2.2×10^{-2}	59	item	8.5×10^{-3}
10	perceiv	2×10^{-2}	60	men	8.1×10^{-3}
11	individu	1.9×10^{-2}	61	suicid	8×10^{-3}
12	mental	1.9×10^{-2}	62	concentr	8×10^{-3}
13	relationship	1.9×10^{-2}	63	simul	8×10^{-3}
14	peopl	1.8×10^{-2}	64	paramet	8×10^{-3}
15	behavior	1.6×10^{-2}	65	protein	7.9×10^{-3}
16	percept	1.6×10^{-2}	66	psychometr	7.9×10^{-3}
17	find	1.5×10^{-2}	67	acid	7.9×10^{-3}
18	task	1.5×10^{-2}	68	parent	7.7×10^{-3}
19	implic	1.5×10^{-2}	69	posit	7.7×10^{-3}
20	questionnair	1.5×10^{-2}	70	undergradu	7.7×10^{-3}
21	cell	1.5×10^{-2}	71	satisfact	7.7×10^{-3}
22	studi	1.4×10^{-2}	72	sexual	7.5×10^{-3}
23	children	1.3×10^{-2}	73	skill	7.5×10^{-3}
24	temperatur	1.2×10^{-2}	74	stimuli	7.4×10^{-3}
25	depress	1.2×10^{-2}	75	method	7.4×10^{-3}
26	anxieti	1.2×10^{-2}	76	onlin	7.3×10^{-3}
27	support	1.2×10^{-2}	77	distress	7.2×10^{-3}
28	relat	1.2×10^{-2}	78	cope	7.2×10^{-3}
29	whether	1.2×10^{-2}	79	water	7.1×10^{-3}
30	health	1.2×10^{-2}	80	explor	7×10^{-3}
31	psychologist	1.1×10^{-2}	81	system	6.8×10^{-3}
32	surfac	1.1×10^{-2}	82	effici	6.8×10^{-3}
33	discuss	1.1×10^{-2}	83	word	6.8×10^{-3}
34	engag	1.1×10^{-2}	84	algorithm	6.6×10^{-3}
35	attitud	1.1×10^{-2}	85	ask	6.6×10^{-3}
36	adolesc	1.1×10^{-2}	86	survey	6.6×10^{-3}
37	school	1.1×10^{-2}	87	attent	6.5×10^{-3}
38	suggest	1.1×10^{-2}	88	electron	6.5×10^{-3}
39	feel	1.1×10^{-2}	89	densiti	6.3×10^{-3}
40	selfreport	1.1×10^{-2}	90	oxid	6.3×10^{-3}
41	adult	1.1×10^{-2}	91	resili	6.2×10^{-3}
42	age	1.1×10^{-2}	92	across	6.2×10^{-3}
43	context	1×10^{-2}	93	among	6.2×10^{-3}
44	educ	1×10^{-2}	94	associ	6.2×10^{-3}
45	energi	1×10^{-2}	95	selfesteem	6.2×10^{-3}
46	gender	1×10^{-2}	96	calcul	6.1×10^{-3}
47	intervent	1×10^{-2}	97	colleg	6.1×10^{-3}
48	learn	9.9×10^{-3}	98	evid	6×10^{-3}
49	women	9.8×10^{-3}	99	assess	6×10^{-3}
50	negat	9.6×10^{-3}	100	chemic	5.9×10^{-3}

TABLE D.214. The list of the top 100 words in the category Psychology, Psychoanalysis with RIGs

No.	Word	RIG
1	psychoanalyt	2.3×10^{-1}
2	psychoanalysi	1.9×10^{-1}
3	freud	1.2×10^{-1}
4	analyst	1.1×10^{-1}
5	unconsci	6×10^{-2}
6	psychoanalyst	5.8×10^{-2}
7	psychic	5.6×10^{-2}
8	psychotherapi	4.2×10^{-2}
9	therapist	4×10^{-2}
10	author	4×10^{-2}
11	result	3.8×10^{-2}
12	psycholog	3.6×10^{-2}
13	trauma	3.4×10^{-2}
14	clinic	2.9×10^{-2}
15	theori	2.9×10^{-2}
16	vignett	2.9×10^{-2}
17	self	2.9×10^{-2}
18	patient	2.5×10^{-2}
19	emot	2.5×10^{-2}
20	think	2.4×10^{-2}
21	person	2.3×10^{-2}
22	concept	2.3×10^{-2}
23	method	2.3×10^{-2}
24	contemporari	2.2×10^{-2}
25	development	2.2×10^{-2}
26	traumat	2.2×10^{-2}
27	mind	2.1×10^{-2}
28	articl	2.1×10^{-2}
29	mental	2.1×10^{-2}
30	dream	2×10^{-2}
31	fantasi	2×10^{-2}
32	work	1.9×10^{-2}
33	analyt	1.9×10^{-2}
34	interperson	1.8×10^{-2}
35	perform	1.8×10^{-2}
36	studi	1.8×10^{-2}
37	argu	1.8×10^{-2}
38	discuss	1.7×10^{-2}
39	mourn	1.7×10^{-2}
40	perspect	1.7×10^{-2}
41	feel	1.6×10^{-2}
42	experi	1.6×10^{-2}
43	show	1.5×10^{-2}
44	use	1.5×10^{-2}
45	idea	1.5×10^{-2}
46	enact	1.5×10^{-2}
47	realiti	1.5×10^{-2}
48	metaphor	1.4×10^{-2}
49	increas	1.4×10^{-2}
50	therapeut	1.4×10^{-2}

No.	Word	RIG
51	understand	1.3×10^{-2}
52	child	1.3×10^{-2}
53	relationship	1.3×10^{-2}
54	effect	1.3×10^{-2}
55	attempt	1.3×10^{-2}
56	cell	1.3×10^{-2}
57	sexual	1.2×10^{-2}
58	anxieti	1.2×10^{-2}
59	high	1.2×10^{-2}
60	obtain	1.2×10^{-2}
61	conflict	1.2×10^{-2}
62	explor	1.2×10^{-2}
63	conscious	1.2×10^{-2}
64	question	1.2×10^{-2}
65	illustr	1.2×10^{-2}
66	attach	1.2×10^{-2}
67	way	1.1×10^{-2}
68	jew	1.1×10^{-2}
69	write	1.1×10^{-2}
70	simul	1.1×10^{-2}
71	essay	1.1×10^{-2}
72	effici	1.1×10^{-2}
73	notion	1.1×10^{-2}
74	commentari	1×10^{-2}
75	psychopatholog	1×10^{-2}
76	dialogu	1×10^{-2}
77	investig	1×10^{-2}
78	data	9.9×10^{-3}
79	view	9.9×10^{-3}
80	symbol	9.8×10^{-3}
81	conceptu	9.6×10^{-3}
82	low	9.4×10^{-3}
83	temperatur	9.4×10^{-3}
84	childhood	9.4×10^{-3}
85	interpret	9.4×10^{-3}
86	neurosci	9.3×10^{-3}
87	life	9.2×10^{-3}
88	test	9.2×10^{-3}
89	relat	9×10^{-3}
90	base	8.9×10^{-3}
91	struggl	8.9×10^{-3}
92	narrat	8.9×10^{-3}
93	autobiograph	8.9×10^{-3}
94	distribut	8.9×10^{-3}
95	adolesc	8.7×10^{-3}
96	confus	8.6×10^{-3}
97	experiment	8.6×10^{-3}
98	paramet	8.5×10^{-3}
99	defens	8.3×10^{-3}
100	dyad	8.2×10^{-3}

Table D.215. The list of the top 100 words in the category Psychology, Social with RIGs

No.	Word	RIG
1	social	6.6×10^{-2}
2	particip	5.9×10^{-2}
3	person	5.3×10^{-2}
4	research	4.7×10^{-2}
5	examin	4.4×10^{-2}
6	psycholog	4.2×10^{-2}
7	perceiv	4.1×10^{-2}
8	emot	3.9×10^{-2}
9	peopl	3.7×10^{-2}
10	relationship	3.3×10^{-2}
11	implic	3.3×10^{-2}
12	individu	3×10^{-2}
13	method	2.7×10^{-2}
14	find	2.7×10^{-2}
15	trait	2.5×10^{-2}
16	discuss	2.3×10^{-2}
17	interperson	2.3×10^{-2}
18	negat	2.1×10^{-2}
19	behavior	2.1×10^{-2}
20	percept	2×10^{-2}
21	attitud	2×10^{-2}
22	motiv	2×10^{-2}
23	student	1.9×10^{-2}
24	mediat	1.9×10^{-2}
25	studi	1.8×10^{-2}
26	selfreport	1.8×10^{-2}
27	predict	1.7×10^{-2}
28	gender	1.7×10^{-2}
29	cell	1.6×10^{-2}
30	romant	1.6×10^{-2}
31	paper	1.6×10^{-2}
32	feel	1.6×10^{-2}
33	whether	1.5×10^{-2}
34	moder	1.5×10^{-2}
35	belief	1.5×10^{-2}
36	partner	1.5×10^{-2}
37	selfesteem	1.4×10^{-2}
38	posit	1.4×10^{-2}
39	undergradu	1.4×10^{-2}
40	stereotyp	1.3×10^{-2}
41	cognit	1.3×10^{-2}
42	relat	1.3×10^{-2}
43	support	1.3×10^{-2}
44	theori	1.3×10^{-2}
45	hypothes	1.2×10^{-2}
46	prejudic	1.2×10^{-2}
47	adolesc	1.2×10^{-2}
48	conclus	1.2×10^{-2}
49	men	1.2×10^{-2}
50	system	1.2×10^{-2}

No.	Word	RIG
51	moral	1.2×10^{-2}
52	temperatur	1.2×10^{-2}
53	experi	1.2×10^{-2}
54	self	1.2×10^{-2}
55	endors	1.1×10^{-2}
56	paramet	1.1×10^{-2}
57	victim	1.1×10^{-2}
58	across	1.1×10^{-2}
59	cultur	1.1×10^{-2}
60	engag	1.1×10^{-2}
61	surfac	1.1×10^{-2}
62	sexual	1.1×10^{-2}
63	context	1.1×10^{-2}
64	suggest	1.1×10^{-2}
65	satisfact	1×10^{-2}
66	judgment	1×10^{-2}
67	effici	1×10^{-2}
68	member	1×10^{-2}
69	abus	9.8×10^{-3}
70	mental	9.8×10^{-3}
71	energi	9.7×10^{-3}
72	obtain	9.6×10^{-3}
73	anxieti	9.5×10^{-3}
74	women	9.4×10^{-3}
75	child	9.4×10^{-3}
76	concentr	9.3×10^{-3}
77	maltreat	9.1×10^{-3}
78	simul	8.6×10^{-3}
79	protein	8.6×10^{-3}
80	questionnair	8.4×10^{-3}
81	water	8.4×10^{-3}
82	tendenc	8.1×10^{-3}
83	prime	8×10^{-3}
84	sampl	8×10^{-3}
85	acid	8×10^{-3}
86	manipul	8×10^{-3}
87	properti	7.8×10^{-3}
88	parent	7.8×10^{-3}
89	colleg	7.7×10^{-3}
90	patient	7.7×10^{-3}
91	calcul	7.6×10^{-3}
92	threat	7.6×10^{-3}
93	associ	7.5×10^{-3}
94	american	7.4×10^{-3}
95	applic	7.4×10^{-3}
96	depress	7.4×10^{-3}
97	perform	7.2×10^{-3}
98	phase	7.2×10^{-3}
99	techniqu	7.2×10^{-3}
100	violenc	7.1×10^{-3}

Table D.216. The list of the top 100 words in the category Public Administration with RIGs

No.	Word	RIG	No.	Word	RIG
1	govern	1.3×10^{-1}	51	make	9.8×10^{-3}
2	public	1.1×10^{-1}	52	author	9.7×10^{-3}
3	polici	9.7×10^{-2}	53	issu	9.6×10^{-3}
4	articl	5.3×10^{-2}	54	government	9.6×10^{-3}
5	polit	4.7×10^{-2}	55	civil	9.2×10^{-3}
6	reform	3.7×10^{-2}	56	explor	9.1×10^{-3}
7	sector	3.7×10^{-2}	57	paramet	8.9×10^{-3}
8	social	3.6×10^{-2}	58	protein	8.8×10^{-3}
9	servic	3.4×10^{-2}	59	conclus	8.7×10^{-3}
10	institut	2.8×10^{-2}	60	literatur	8.7×10^{-3}
11	administr	2.6×10^{-2}	61	municip	8.6×10^{-3}
12	manag	2.6×10^{-2}	62	organ	8.4×10^{-3}
13	argu	2.4×10^{-2}	63	strateg	8.4×10^{-3}
14	citizen	2.4×10^{-2}	64	resourc	8.3×10^{-3}
15	countri	2.4×10^{-2}	65	debat	8.2×10^{-3}
16	actor	2.1×10^{-2}	66	find	8.2×10^{-3}
17	econom	2×10^{-2}	67	feder	8.1×10^{-3}
18	organiz	1.8×10^{-2}	68	obtain	8.1×10^{-3}
19	nation	1.8×10^{-2}	69	legitimaci	8×10^{-3}
20	agenc	1.8×10^{-2}	70	observ	8×10^{-3}
21	method	1.7×10^{-2}	71	scholar	7.9×10^{-3}
22	financi	1.7×10^{-2}	72	detect	7.9×10^{-3}
23	china	1.7×10^{-2}	73	democraci	7.8×10^{-3}
24	empir	1.6×10^{-2}	74	treatment	7.8×10^{-3}
25	welfar	1.5×10^{-2}	75	clinic	7.7×10^{-3}
26	market	1.5×10^{-2}	76	induc	7.6×10^{-3}
27	research	1.5×10^{-2}	77	offici	7.6×10^{-3}
28	privat	1.5×10^{-2}	78	council	7.5×10^{-3}
29	cell	1.5×10^{-2}	79	acid	7.5×10^{-3}
30	result	1.3×10^{-2}	80	use	7.4×10^{-3}
31	patient	1.3×10^{-2}	81	simul	7.4×10^{-3}
32	tax	1.3×10^{-2}	82	practic	7.3×10^{-3}
33	democrat	1.2×10^{-2}	83	way	7.3×10^{-3}
34	crisi	1.2×10^{-2}	84	citi	7.2×10^{-3}
35	union	1.2×10^{-2}	85	stakehold	7.2×10^{-3}
36	perspect	1.2×10^{-2}	86	decis	7.1×10^{-3}
37	innov	1.2×10^{-2}	87	experiment	7.1×10^{-3}
38	employe	1.2×10^{-2}	88	properti	7×10^{-3}
39	fiscal	1.1×10^{-2}	89	incom	7×10^{-3}
40	draw	1.1×10^{-2}	90	provis	7×10^{-3}
41	european	1.1×10^{-2}	91	gene	7×10^{-3}
42	financ	1.1×10^{-2}	92	survey	7×10^{-3}
43	surfac	1.1×10^{-2}	93	examin	7×10^{-3}
44	fund	1.1×10^{-2}	94	state	6.9×10^{-3}
45	societi	1×10^{-2}	95	policymak	6.9×10^{-3}
46	temperatur	1×10^{-2}	96	ratio	6.9×10^{-3}
47	economi	1×10^{-2}	97	capit	6.8×10^{-3}
48	agenda	1×10^{-2}	98	high	6.7×10^{-3}
49	bureaucrat	1×10^{-2}	99	manageri	6.7×10^{-3}
50	focus	9.9×10^{-3}	100	legal	6.6×10^{-3}

Table D.217. The list of the top 100 words in the category Public, Environmental and Occupational Health with RIGs

No.	Word	RIG	No.	Word	RIG
1	health	1.5×10^{-1}	51	mental	1.5×10^{-2}
2	conclus	6.6×10^{-2}	52	sex	1.4×10^{-2}
3	risk	5×10^{-2}	53	peopl	1.4×10^{-2}
4	among	4.6×10^{-2}	54	countri	1.4×10^{-2}
5	age	4.4×10^{-2}	55	socioeconom	1.4×10^{-2}
6	survey	4.3×10^{-2}	56	live	1.4×10^{-2}
7	particip	4.3×10^{-2}	57	school	1.4×10^{-2}
8	women	3.7×10^{-2}	58	factor	1.4×10^{-2}
9	regress	3.6×10^{-2}	59	person	1.4×10^{-2}
10	associ	3.5×10^{-2}	60	rural	1.4×10^{-2}
11	intervent	3.5×10^{-2}	61	polici	1.3×10^{-2}
12	year	3.5×10^{-2}	62	outcom	1.3×10^{-2}
13	care	3.3×10^{-2}	63	individu	1.3×10^{-2}
14	popul	3.2×10^{-2}	64	hospit	1.2×10^{-2}
15	background	3×10^{-2}	65	cohort	1.2×10^{-2}
16	interview	2.9×10^{-2}	66	sexual	1.2×10^{-2}
17	preval	2.7×10^{-2}	67	adolesc	1.2×10^{-2}
18	public	2.7×10^{-2}	68	resid	1.2×10^{-2}
19	object	2.7×10^{-2}	69	section	1.2×10^{-2}
20	nation	2.7×10^{-2}	70	gender	1.2×10^{-2}
21	logist	2.5×10^{-2}	71	epidemiolog	1.2×10^{-2}
22	questionnair	2.5×10^{-2}	72	child	1.1×10^{-2}
23	educ	2.5×10^{-2}	73	propos	1.1×10^{-2}
24	assess	2.3×10^{-2}	74	paper	1.1×10^{-2}
25	exposur	2.3×10^{-2}	75	report	1.1×10^{-2}
26	prevent	2.2×10^{-2}	76	household	1.1×10^{-2}
27	communiti	2.2×10^{-2}	77	older	1.1×10^{-2}
28	worker	2.2×10^{-2}	78	interv	1.1×10^{-2}
29	examin	2.1×10^{-2}	79	program	1.1×10^{-2}
30	smoke	2.1×10^{-2}	80	tobacco	1.1×10^{-2}
31	method	2.1×10^{-2}	81	ethnic	1.1×10^{-2}
32	adjust	2×10^{-2}	82	obes	1×10^{-2}
33	children	2×10^{-2}	83	perceiv	1×10^{-2}
34	studi	1.9×10^{-2}	84	includ	1×10^{-2}
35	hiv	1.9×10^{-2}	85	smoker	1×10^{-2}
36	men	1.8×10^{-2}	86	group	1×10^{-2}
37	data	1.8×10^{-2}	87	healthcar	1×10^{-2}
38	odd	1.8×10^{-2}	88	sampl	9.8×10^{-3}
39	demograph	1.7×10^{-2}	89	collect	9.7×10^{-3}
40	status	1.7×10^{-2}	90	properti	9.7×10^{-3}
41	social	1.7×10^{-2}	91	birth	9.6×10^{-3}
42	medic	1.6×10^{-2}	92	cross	9.6×10^{-3}
43	need	1.6×10^{-2}	93	surveil	9.5×10^{-3}
44	confid	1.6×10^{-2}	94	infect	9.2×10^{-3}
45	selfreport	1.6×10^{-2}	95	youth	9.1×10^{-3}
46	adult	1.5×10^{-2}	96	diseas	9×10^{-3}
47	occup	1.5×10^{-2}	97	multivari	9×10^{-3}
48	incom	1.5×10^{-2}	98	result	9×10^{-3}
49	conduct	1.5×10^{-2}	99	identifi	9×10^{-3}
50	servic	1.5×10^{-2}	100	research	8.9×10^{-3}

Table D.218. The list of the top 100 words in the category Radiology, Nuclear Medicine and Medical Imaging with RIGs

No.	Word	RIG	No.	Word	RIG
1	imag	1.6×10^{-1}	51	accuraci	1.2×10^{-2}
2	patient	6.2×10^{-2}	52	compar	1.1×10^{-2}
3	mri	5.4×10^{-2}	53	perform	1.1×10^{-2}
4	purpos	4.9×10^{-2}	54	measur	1.1×10^{-2}
5	phantom	4.9×10^{-2}	55	standard	1.1×10^{-2}
6	tomographi	4.7×10^{-2}	56	resolut	1.1×10^{-2}
7	conclus	3.6×10^{-2}	57	left	1.1×10^{-2}
8	dose	3.4×10^{-2}	58	fmri	1×10^{-2}
9	pet	3.3×10^{-2}	59	detect	1×10^{-2}
10	radiologist	3.2×10^{-2}	60	use	1×10^{-2}
11	radiat	3.1×10^{-2}	61	artifact	1×10^{-2}
12	method	3.1×10^{-2}	62	paper	1×10^{-2}
13	clinic	3×10^{-2}	63	detector	9.8×10^{-3}
14	lesion	2.7×10^{-2}	64	breast	9.8×10^{-3}
15	reson	2.6×10^{-2}	65	prostat	9.7×10^{-3}
16	tissu	2.4×10^{-2}	66	beam	9.6×10^{-3}
17	brain	2.3×10^{-2}	67	registr	9.4×10^{-3}
18	ultrasound	2.3×10^{-2}	68	review	9.2×10^{-3}
19	volum	2.2×10^{-2}	69	vivo	9.1×10^{-3}
20	radiotherapi	2.2×10^{-2}	70	patholog	9×10^{-3}
21	underw	2.2×10^{-2}	71	malign	9×10^{-3}
22	magnet	2.1×10^{-2}	72	approv	8.9×10^{-3}
23	contrast	2×10^{-2}	73	visual	8.8×10^{-3}
24	tumor	2×10^{-2}	74	lung	8.7×10^{-3}
25	evalu	1.9×10^{-2}	75	uptak	8.5×10^{-3}
26	diagnost	1.9×10^{-2}	76	correl	8.5×10^{-3}
27	arteri	1.9×10^{-2}	77	median	8.4×10^{-3}
28	scan	1.8×10^{-2}	78	cortic	8.2×10^{-3}
29	cancer	1.8×10^{-2}	79	plan	8.1×10^{-3}
30	radiolog	1.7×10^{-2}	80	irradi	7.9×10^{-3}
31	scanner	1.7×10^{-2}	81	accur	7.9×10^{-3}
32	retrospect	1.7×10^{-2}	82	normal	7.6×10^{-3}
33	materi	1.6×10^{-2}	83	valu	7.6×10^{-3}
34	voxel	1.6×10^{-2}	84	anatomi	7.6×10^{-3}
35	anatom	1.6×10^{-2}	85	result	7.5×10^{-3}
36	acquir	1.6×10^{-2}	86	cortex	7.5×10^{-3}
37	assess	1.5×10^{-2}	87	temperatur	7.4×10^{-3}
38	reconstruct	1.5×10^{-2}	88	cardiac	7.4×10^{-3}
39	acquisit	1.5×10^{-2}	89	sensit	7.4×10^{-3}
40	angiographi	1.4×10^{-2}	90	nois	7.4×10^{-3}
41	segment	1.3×10^{-2}	91	vessel	7.3×10^{-3}
42	diagnosi	1.3×10^{-2}	92	signific	7.3×10^{-3}
43	mean	1.3×10^{-2}	93	benign	7.2×10^{-3}
44	echo	1.3×10^{-2}	94	diffus	7.2×10^{-3}
45	perfus	1.3×10^{-2}	95	consecut	7.2×10^{-3}
46	noninvas	1.3×10^{-2}	96	healthi	7.1×10^{-3}
47	comput	1.2×10^{-2}	97	liver	7.1×10^{-3}
48	positron	1.2×10^{-2}	98	guid	7.1×10^{-3}
49	therapi	1.2×10^{-2}	99	feasibl	7×10^{-3}
50	modal	1.2×10^{-2}	100	quantit	7×10^{-3}

Table D.219. The list of the top 100 words in the category Rehabilitation with RIGs

No.	Word	RIG	No.	Word	RIG
1	particip	9×10^{-2}	51	asd	1.4×10^{-2}
2	rehabilit	8.2×10^{-2}	52	interview	1.4×10^{-2}
3	disabl	8×10^{-2}	53	cell	1.3×10^{-2}
4	conclus	7.1×10^{-2}	54	disord	1.3×10^{-2}
5	intervent	5.7×10^{-2}	55	examin	1.3×10^{-2}
6	purpos	4×10^{-2}	56	knee	1.3×10^{-2}
7	children	3.6×10^{-2}	57	research	1.3×10^{-2}
8	object	3.5×10^{-2}	58	speech	1.3×10^{-2}
9	outcom	3.1×10^{-2}	59	paper	1.2×10^{-2}
10	score	3.1×10^{-2}	60	item	1.2×10^{-2}
11	age	3×10^{-2}	61	design	1.2×10^{-2}
12	assess	2.9×10^{-2}	62	movement	1.2×10^{-2}
13	therapist	2.8×10^{-2}	63	therapi	1.2×10^{-2}
14	studi	2.7×10^{-2}	64	task	1.2×10^{-2}
15	exercis	2.7×10^{-2}	65	temperatur	1.2×10^{-2}
16	pain	2.6×10^{-2}	66	cognit	1.2×10^{-2}
17	group	2.5×10^{-2}	67	cord	1.1×10^{-2}
18	injuri	2.5×10^{-2}	68	postur	1.1×10^{-2}
19	muscl	2.4×10^{-2}	69	occup	1.1×10^{-2}
20	peopl	2.3×10^{-2}	70	palsi	1.1×10^{-2}
21	impair	2.3×10^{-2}	71	educ	1.1×10^{-2}
22	stroke	2.3×10^{-2}	72	background	1.1×10^{-2}
23	subject	2.2×10^{-2}	73	difficulti	1×10^{-2}
24	walk	2.2×10^{-2}	74	life	1×10^{-2}
25	session	2.2×10^{-2}	75	scale	1×10^{-2}
26	motor	2.2×10^{-2}	76	aim	1×10^{-2}
27	measur	2.2×10^{-2}	77	shoulder	1×10^{-2}
28	autism	2.1×10^{-2}	78	support	1×10^{-2}
29	patient	2.1×10^{-2}	79	symptom	1×10^{-2}
30	individu	2×10^{-2}	80	month	1×10^{-2}
31	clinic	2×10^{-2}	81	healthi	1×10^{-2}
32	gait	2×10^{-2}	82	trial	9.9×10^{-3}
33	intellectu	2×10^{-2}	83	propos	9.8×10^{-3}
34	train	1.9×10^{-2}	84	care	9.7×10^{-3}
35	limb	1.9×10^{-2}	85	ankl	9.5×10^{-3}
36	questionnair	1.9×10^{-2}	86	need	9.3×10^{-3}
37	skill	1.9×10^{-2}	87	development	9.3×10^{-3}
38	week	1.7×10^{-2}	88	simul	9.2×10^{-3}
39	person	1.7×10^{-2}	89	complet	9×10^{-3}
40	result	1.7×10^{-2}	90	inpati	9×10^{-3}
41	spinal	1.6×10^{-2}	91	child	9×10^{-3}
42	signific	1.6×10^{-2}	92	social	8.8×10^{-3}
43	adult	1.6×10^{-2}	93	mental	8.7×10^{-3}
44	method	1.6×10^{-2}	94	program	8.7×10^{-3}
45	traumat	1.6×10^{-2}	95	clinician	8.6×10^{-3}
46	physic	1.5×10^{-2}	96	deficit	8.4×10^{-3}
47	health	1.5×10^{-2}	97	profession	8.3×10^{-3}
48	year	1.4×10^{-2}	98	moder	8.2×10^{-3}
49	languag	1.4×10^{-2}	99	function	8.2×10^{-3}
50	flexion	1.4×10^{-2}	100	baselin	8×10^{-3}

Table D.220. The list of the top 100 words in the category Religion with RIGs

No.	Word	RIG	No.	Word	RIG
1	religi	1.8×10^{-1}	51	discours	1.4×10^{-2}
2	christian	1.3×10^{-1}	52	roman	1.4×10^{-2}
3	religion	1.2×10^{-1}	53	cell	1.4×10^{-2}
4	theolog	1.1×10^{-1}	54	interpret	1.4×10^{-2}
5	god	9.7×10^{-2}	55	philosoph	1.4×10^{-2}
6	church	7.9×10^{-2}	56	world	1.4×10^{-2}
7	articl	7.7×10^{-2}	57	ritual	1.4×10^{-2}
8	spiritu	6×10^{-2}	58	scholarship	1.4×10^{-2}
9	argu	5.6×10^{-2}	59	read	1.3×10^{-2}
10	islam	4.7×10^{-2}	60	context	1.3×10^{-2}
11	biblic	4.6×10^{-2}	61	critiqu	1.3×10^{-2}
12	faith	4.5×10^{-2}	62	increas	1.3×10^{-2}
13	divin	4.1×10^{-2}	63	system	1.3×10^{-2}
14	result	3.9×10^{-2}	64	low	1.3×10^{-2}
15	scholar	3.7×10^{-2}	65	obtain	1.3×10^{-2}
16	muslim	3.6×10^{-2}	66	peopl	1.3×10^{-2}
17	christ	3.5×10^{-2}	67	simul	1.2×10^{-2}
18	text	3.3×10^{-2}	68	spirit	1.2×10^{-2}
19	essay	3×10^{-2}	69	literari	1.2×10^{-2}
20	secular	3×10^{-2}	70	view	1.2×10^{-2}
21	centuri	2.9×10^{-2}	71	love	1.2×10^{-2}
22	cathol	2.8×10^{-2}	72	ancient	1.1×10^{-2}
23	contemporari	2.6×10^{-2}	73	improv	1.1×10^{-2}
24	method	2.4×10^{-2}	74	compar	1.1×10^{-2}
25	moral	2.4×10^{-2}	75	temperatur	1.1×10^{-2}
26	doctrin	2.3×10^{-2}	76	understand	1.1×10^{-2}
27	belief	2.2×10^{-2}	77	debat	1.1×10^{-2}
28	jewish	2.2×10^{-2}	78	life	1.1×10^{-2}
29	tradit	2.1×10^{-2}	79	stori	1.1×10^{-2}
30	book	2.1×10^{-2}	80	sacr	1.1×10^{-2}
31	narrat	2.1×10^{-2}	81	perspect	1.1×10^{-2}
32	question	2×10^{-2}	82	communiti	1×10^{-2}
33	polit	2×10^{-2}	83	paramet	1×10^{-2}
34	use	1.9×10^{-2}	84	cultur	1×10^{-2}
35	author	1.8×10^{-2}	85	draw	1×10^{-2}
36	john	1.8×10^{-2}	86	rate	1×10^{-2}
37	buddhist	1.7×10^{-2}	87	effici	1×10^{-2}
38	protest	1.7×10^{-2}	88	hermeneut	1×10^{-2}
39	social	1.7×10^{-2}	89	induc	9.9×10^{-3}
40	way	1.7×10^{-2}	90	reduc	9.9×10^{-3}
41	effect	1.7×10^{-2}	91	modern	9.8×10^{-3}
42	societi	1.6×10^{-2}	92	person	9.7×10^{-3}
43	ethic	1.6×10^{-2}	93	show	9.7×10^{-3}
44	perform	1.6×10^{-2}	94	detect	9.7×10^{-3}
45	high	1.6×10^{-2}	95	surfac	9.7×10^{-3}
46	argument	1.6×10^{-2}	96	mystic	9.5×10^{-3}
47	histor	1.5×10^{-2}	97	histori	9.4×10^{-3}
48	claim	1.5×10^{-2}	98	dialogu	9.3×10^{-3}
49	paul	1.5×10^{-2}	99	control	9.3×10^{-3}
50	holi	1.5×10^{-2}	100	decreas	9.2×10^{-3}

Table D.221. The list of the top 100 words in the category Remote Sensing with RIGs

No.	Word	RIG
1	satellit	9.6×10^{-2}
2	remot	8×10^{-2}
3	imag	6.8×10^{-2}
4	sens	6.2×10^{-2}
5	sar	5.8×10^{-2}
6	radar	5.5×10^{-2}
7	resolut	5.3×10^{-2}
8	imageri	4.5×10^{-2}
9	land	4.5×10^{-2}
10	spatial	4.3×10^{-2}
11	data	4.3×10^{-2}
12	modi	4.3×10^{-2}
13	sensor	4.2×10^{-2}
14	accuraci	3.9×10^{-2}
15	landsat	3.6×10^{-2}
16	hyperspectr	3.6×10^{-2}
17	apertur	3.6×10^{-2}
18	map	3.6×10^{-2}
19	veget	3.4×10^{-2}
20	pixel	3.1×10^{-2}
21	area	2.7×10^{-2}
22	algorithm	2.7×10^{-2}
23	airborn	2.6×10^{-2}
24	spectral	2.6×10^{-2}
25	ground	2.4×10^{-2}
26	cover	2.3×10^{-2}
27	band	2.2×10^{-2}
28	classif	2.2×10^{-2}
29	estim	2.2×10^{-2}
30	retriev	2.1×10^{-2}
31	paper	2×10^{-2}
32	synthet	2×10^{-2}
33	lidar	2×10^{-2}
34	error	1.9×10^{-2}
35	forest	1.9×10^{-2}
36	base	1.9×10^{-2}
37	patient	1.9×10^{-2}
38	monitor	1.8×10^{-2}
39	conclus	1.8×10^{-2}
40	earth	1.7×10^{-2}
41	mission	1.6×10^{-2}
42	inform	1.5×10^{-2}
43	gps	1.4×10^{-2}
44	tempor	1.4×10^{-2}
45	acquir	1.3×10^{-2}
46	use	1.3×10^{-2}
47	treatment	1.3×10^{-2}
48	propos	1.2×10^{-2}
49	backscatt	1.2×10^{-2}
50	cloud	1.2×10^{-2}
51	calibr	1.1×10^{-2}
52	aerial	1.1×10^{-2}
53	clinic	1.1×10^{-2}
54	cell	1.1×10^{-2}
55	canopi	1.1×10^{-2}
56	navig	1×10^{-2}
57	global	1×10^{-2}
58	atmospher	1×10^{-2}
59	dem	1×10^{-2}
60	scene	1×10^{-2}
61	accur	1×10^{-2}
62	terrain	9.9×10^{-3}
63	ocean	9.6×10^{-3}
64	soil	9.4×10^{-3}
65	gis	9.3×10^{-3}
66	group	9.2×10^{-3}
67	detect	9.2×10^{-3}
68	protein	9.1×10^{-3}
69	valid	9.1×10^{-3}
70	dataset	9×10^{-3}
71	age	8.9×10^{-3}
72	surfac	8.8×10^{-3}
73	season	8.3×10^{-3}
74	gene	8.3×10^{-3}
75	digit	8.3×10^{-3}
76	urban	8.2×10^{-3}
77	sea	8.1×10^{-3}
78	region	7.9×10^{-3}
79	measur	7.8×10^{-3}
80	reflect	7.7×10^{-3}
81	deriv	7.7×10^{-3}
82	extract	7.5×10^{-3}
83	meteorolog	7.3×10^{-3}
84	applic	7.3×10^{-3}
85	letter	7.2×10^{-3}
86	diseas	7.1×10^{-3}
87	height	7×10^{-3}
88	classifi	6.9×10^{-3}
89	associ	6.9×10^{-3}
90	acid	6.8×10^{-3}
91	reaction	6.8×10^{-3}
92	squar	6.8×10^{-3}
93	moistur	6.7×10^{-3}
94	agricultur	6.7×10^{-3}
95	filter	6.6×10^{-3}
96	scatter	6.6×10^{-3}
97	snow	6.5×10^{-3}
98	nois	6.5×10^{-3}
99	acquisit	6.5×10^{-3}
100	climat	6.3×10^{-3}

Table D.222. The list of the top 100 words in the category Reproductive Biology with RIGs

No.	Word	RIG	No.	Word	RIG
1	pregnanc	1.1×10^{-1}	51	group	1.8×10^{-2}
2	sperm	9.3×10^{-2}	52	embryon	1.8×10^{-2}
3	oocyt	8.8×10^{-2}	53	testi	1.7×10^{-2}
4	fertil	8.7×10^{-2}	54	femal	1.6×10^{-2}
5	embryo	8.5×10^{-2}	55	dure	1.6×10^{-2}
6	reproduct	8.1×10^{-2}	56	ejacul	1.5×10^{-2}
7	ovarian	7.6×10^{-2}	57	intrauterin	1.5×10^{-2}
8	ivf	6.7×10^{-2}	58	germ	1.5×10^{-2}
9	infertil	6.1×10^{-2}	59	intervent	1.5×10^{-2}
10	follicl	5.9×10^{-2}	60	preeclampsia	1.5×10^{-2}
11	women	5.7×10^{-2}	61	treatment	1.4×10^{-2}
12	blastocyst	5.5×10^{-2}	62	thaw	1.4×10^{-2}
13	semen	4.8×10^{-2}	63	implant	1.4×10^{-2}
14	ovari	4.5×10^{-2}	64	bovin	1.4×10^{-2}
15	hormon	3.8×10^{-2}	65	patient	1.4×10^{-2}
16	progesteron	3.7×10^{-2}	66	regul	1.4×10^{-2}
17	uterin	3.7×10^{-2}	67	serum	1.3×10^{-2}
18	placent	3.6×10^{-2}	68	stimul	1.3×10^{-2}
19	spermatozoa	3.5×10^{-2}	69	cow	1.3×10^{-2}
20	placenta	3.4×10^{-2}	70	simul	1.3×10^{-2}
21	matern	3.4×10^{-2}	71	male	1.2×10^{-2}
22	ovul	3.3×10^{-2}	72	control	1.2×10^{-2}
23	gestat	3.3×10^{-2}	73	receptor	1.2×10^{-2}
24	express	3.2×10^{-2}	74	human	1.2×10^{-2}
25	fetal	3×10^{-2}	75	development	1.2×10^{-2}
26	follicular	3×10^{-2}	76	associ	1.2×10^{-2}
27	birth	3×10^{-2}	77	normal	1.1×10^{-2}
28	insemin	3×10^{-2}	78	cultur	1.1×10^{-2}
29	conclus	3×10^{-2}	79	undergo	1.1×10^{-2}
30	cell	2.9×10^{-2}	80	vagin	1.1×10^{-2}
31	endometri	2.9×10^{-2}	81	propos	1.1×10^{-2}
32	vitro	2.9×10^{-2}	82	testosteron	1.1×10^{-2}
33	paper	2.7×10^{-2}	83	earli	1.1×10^{-2}
34	matur	2.6×10^{-2}	84	compar	1.1×10^{-2}
35	day	2.5×10^{-2}	85	pcr	1.1×10^{-2}
36	studi	2.5×10^{-2}	86	fetus	1.1×10^{-2}
37	outcom	2.5×10^{-2}	87	increas	1.1×10^{-2}
38	motil	2.5×10^{-2}	88	tissu	1.1×10^{-2}
39	pregnant	2.4×10^{-2}	89	frozen	1.1×10^{-2}
40	cryopreserv	2.3×10^{-2}	90	abnorm	1.1×10^{-2}
41	signific	2.1×10^{-2}	91	trimest	1.1×10^{-2}
42	mrna	2.1×10^{-2}	92	may	1.1×10^{-2}
43	object	2×10^{-2}	93	offspr	1×10^{-2}
44	estradiol	1.9×10^{-2}	94	level	1×10^{-2}
45	testicular	1.9×10^{-2}	95	higher	1×10^{-2}
46	spermatogenesi	1.8×10^{-2}	96	stage	1×10^{-2}
47	protein	1.8×10^{-2}	97	estrogen	9.9×10^{-3}
48	gene	1.8×10^{-2}	98	none	9.8×10^{-3}
49	acrosom	1.8×10^{-2}	99	anim	9.4×10^{-3}
50	cycl	1.8×10^{-2}	100	retrospect	9.4×10^{-3}

TABLE D.223. The list of the top 100 words in the category Respiratory System with RIGs

No.	Word	RIG
1	lung	1.5×10^{-1}
2	pulmonari	1.5×10^{-1}
3	patient	1.4×10^{-1}
4	conclus	1.1×10^{-1}
5	copd	6.9×10^{-2}
6	obstruct	6.1×10^{-2}
7	respiratori	5.3×10^{-2}
8	diseas	5.2×10^{-2}
9	airway	5.1×10^{-2}
10	background	4.9×10^{-2}
11	mortal	4.4×10^{-2}
12	fev1	4.1×10^{-2}
13	asthma	4×10^{-2}
14	clinic	4×10^{-2}
15	method	3.9×10^{-2}
16	year	3.8×10^{-2}
17	chronic	3.4×10^{-2}
18	object	3.4×10^{-2}
19	surgeri	3.3×10^{-2}
20	expiratori	3.3×10^{-2}
21	outcom	3.3×10^{-2}
22	underw	3.3×10^{-2}
23	age	3.1×10^{-2}
24	hospit	3.1×10^{-2}
25	thorac	3.1×10^{-2}
26	arteri	3×10^{-2}
27	associ	3×10^{-2}
28	cardiac	2.8×10^{-2}
29	ventil	2.7×10^{-2}
30	aortic	2.7×10^{-2}
31	paper	2.6×10^{-2}
32	postop	2.6×10^{-2}
33	surviv	2.6×10^{-2}
34	valv	2.6×10^{-2}
35	median	2.5×10^{-2}
36	risk	2.5×10^{-2}
37	surgic	2.5×10^{-2}
38	bypass	2.4×10^{-2}
39	heart	2.2×10^{-2}
40	chest	2.2×10^{-2}
41	nonsmal	2.2×10^{-2}
42	fibrosi	2.2×10^{-2}
43	retrospect	2.1×10^{-2}
44	result	2×10^{-2}
45	month	2×10^{-2}
46	ventricular	2×10^{-2}
47	exacerb	2×10^{-2}
48	signific	2×10^{-2}
49	cardiopulmonari	1.9×10^{-2}
50	inhal	1.9×10^{-2}

No.	Word	RIG
51	nsclc	1.8×10^{-2}
52	breath	1.8×10^{-2}
53	preoper	1.8×10^{-2}
54	treatment	1.7×10^{-2}
55	resect	1.7×10^{-2}
56	alveolar	1.7×10^{-2}
57	group	1.7×10^{-2}
58	tuberculosi	1.7×10^{-2}
59	smoke	1.6×10^{-2}
60	cohort	1.6×10^{-2}
61	therapi	1.6×10^{-2}
62	undergo	1.6×10^{-2}
63	day	1.6×10^{-2}
64	morbid	1.6×10^{-2}
65	left	1.6×10^{-2}
66	multivari	1.6×10^{-2}
67	follow	1.6×10^{-2}
68	bronchial	1.5×10^{-2}
69	care	1.5×10^{-2}
70	propos	1.5×10^{-2}
71	assess	1.5×10^{-2}
72	death	1.5×10^{-2}
73	acut	1.4×10^{-2}
74	prospect	1.4×10^{-2}
75	coronari	1.3×10^{-2}
76	diagnosi	1.3×10^{-2}
77	smoker	1.3×10^{-2}
78	regurgit	1.3×10^{-2}
79	mitral	1.3×10^{-2}
80	pneumonia	1.3×10^{-2}
81	inflamm	1.3×10^{-2}
82	reoper	1.2×10^{-2}
83	receiv	1.2×10^{-2}
84	consecut	1.2×10^{-2}
85	score	1.2×10^{-2}
86	cystic	1.2×10^{-2}
87	complic	1.2×10^{-2}
88	baselin	1.2×10^{-2}
89	lavag	1.2×10^{-2}
90	includ	1.2×10^{-2}
91	stay	1.2×10^{-2}
92	medic	1.1×10^{-2}
93	diagnos	1.1×10^{-2}
94	symptom	1.1×10^{-2}
95	repair	1.1×10^{-2}
96	predictor	1.1×10^{-2}
97	review	1.1×10^{-2}
98	periop	1.1×10^{-2}
99	asthmat	1×10^{-2}
100	regress	1×10^{-2}

Table D.224. The list of the top 100 words in the category Rheumatology with RIGs

No.	Word	RIG
1	arthriti	2.1×10^{-1}
2	patient	1.5×10^{-1}
3	rheumatoid	1.5×10^{-1}
4	conclus	1.3×10^{-1}
5	diseas	1.1×10^{-1}
6	object	1×10^{-1}
7	rheumatolog	7.8×10^{-2}
8	lupus	7.6×10^{-2}
9	osteoarthr	7.4×10^{-2}
10	erythematosus	6.8×10^{-2}
11	sle	6.6×10^{-2}
12	joint	6.5×10^{-2}
13	clinic	6.4×10^{-2}
14	method	5.8×10^{-2}
15	score	5.6×10^{-2}
16	pain	5×10^{-2}
17	knee	4.4×10^{-2}
18	assess	3.8×10^{-2}
19	inflammatori	3.7×10^{-2}
20	age	3.5×10^{-2}
21	associ	3.4×10^{-2}
22	year	3.3×10^{-2}
23	cartilag	3.1×10^{-2}
24	treatment	3.1×10^{-2}
25	antibodi	3×10^{-2}
26	cohort	2.9×10^{-2}
27	result	2.8×10^{-2}
28	autoimmun	2.5×10^{-2}
29	radiograph	2.5×10^{-2}
30	baselin	2.4×10^{-2}
31	paper	2.2×10^{-2}
32	therapi	2.2×10^{-2}
33	remiss	2.1×10^{-2}
34	criteria	2.1×10^{-2}
35	healthi	2.1×10^{-2}
36	signific	2×10^{-2}
37	inflamm	2×10^{-2}
38	index	2×10^{-2}
39	diagnosi	1.8×10^{-2}
40	studi	1.8×10^{-2}
41	bone	1.8×10^{-2}
42	articular	1.8×10^{-2}
43	hip	1.7×10^{-2}
44	month	1.7×10^{-2}
45	serum	1.7×10^{-2}
46	activ	1.7×10^{-2}
47	musculoskelet	1.7×10^{-2}
48	syndrom	1.7×10^{-2}
49	symptom	1.7×10^{-2}
50	questionnair	1.7×10^{-2}
51	outcom	1.7×10^{-2}
52	factor	1.7×10^{-2}
53	crp	1.6×10^{-2}
54	disabl	1.6×10^{-2}
55	durat	1.6×10^{-2}
56	necrosi	1.6×10^{-2}
57	risk	1.5×10^{-2}
58	aim	1.5×10^{-2}
59	sclerosi	1.5×10^{-2}
60	includ	1.4×10^{-2}
61	femal	1.4×10^{-2}
62	group	1.4×10^{-2}
63	tnf	1.4×10^{-2}
64	diagnos	1.4×10^{-2}
65	sex	1.4×10^{-2}
66	idiopath	1.3×10^{-2}
67	erythrocyt	1.3×10^{-2}
68	week	1.3×10^{-2}
69	interleukin	1.3×10^{-2}
70	evalu	1.3×10^{-2}
71	pathogenesi	1.3×10^{-2}
72	propos	1.3×10^{-2}
73	regress	1.3×10^{-2}
74	mean	1.3×10^{-2}
75	juvenil	1.2×10^{-2}
76	manifest	1.2×10^{-2}
77	compar	1.2×10^{-2}
78	preval	1.2×10^{-2}
79	trial	1.2×10^{-2}
80	treat	1.2×10^{-2}
81	corticosteroid	1.2×10^{-2}
82	follow	1.2×10^{-2}
83	control	1.2×10^{-2}
84	onset	1.2×10^{-2}
85	temperatur	1.1×10^{-2}
86	colleg	1.1×10^{-2}
87	simul	1.1×10^{-2}
88	chronic	1.1×10^{-2}
89	chondrocyt	1.1×10^{-2}
90	health	1.1×10^{-2}
91	higher	1.1×10^{-2}
92	total	1×10^{-2}
93	correl	1×10^{-2}
94	peripher	9.9×10^{-3}
95	american	9.8×10^{-3}
96	physician	9.7×10^{-3}
97	immunosuppress	9.7×10^{-3}
98	elisa	9.7×10^{-3}
99	cytokin	9.6×10^{-3}
100	enrol	9.5×10^{-3}

TABLE D.225. The list of the top 100 words in the category Robotics with RIGs

No.	Word	RIG	No.	Word	RIG
1	robot	2.8×10^{-1}	51	wheel	9.5×10^{-3}
2	paper	7.6×10^{-2}	52	suggest	9.4×10^{-3}
3	motion	5.1×10^{-2}	53	associ	9.3×10^{-3}
4	propos	4.2×10^{-2}	54	unman	9.2×10^{-3}
5	actuat	3.4×10^{-2}	55	walk	9.2×10^{-3}
6	manipul	3.1×10^{-2}	56	path	9.1×10^{-3}
7	algorithm	3×10^{-2}	57	leg	9×10^{-3}
8	kinemat	2.9×10^{-2}	58	increas	8.9×10^{-3}
9	task	2.9×10^{-2}	59	temperatur	8.4×10^{-3}
10	sensor	2.5×10^{-2}	60	age	8.3×10^{-3}
11	trajectori	2.5×10^{-2}	61	perform	8.3×10^{-3}
12	system	2.3×10^{-2}	62	found	8.3×10^{-3}
13	dof	2.3×10^{-2}	63	comput	8.3×10^{-3}
14	autonom	2.2×10^{-2}	64	acid	8.3×10^{-3}
15	control	2.2×10^{-2}	65	inspir	8.2×10^{-3}
16	studi	2×10^{-2}	66	platform	8.1×10^{-3}
17	environ	2×10^{-2}	67	capabl	8×10^{-3}
18	navig	1.8×10^{-2}	68	human	7.9×10^{-3}
19	conclus	1.8×10^{-2}	69	gene	7.8×10^{-3}
20	camera	1.8×10^{-2}	70	visual	7.8×10^{-3}
21	track	1.7×10^{-2}	71	constraint	7.7×10^{-3}
22	simul	1.6×10^{-2}	72	desir	7.6×10^{-3}
23	real	1.5×10^{-2}	73	accuraci	7.6×10^{-3}
24	joint	1.5×10^{-2}	74	experiment	7.5×10^{-3}
25	design	1.5×10^{-2}	75	pose	7.5×10^{-3}
26	vehicl	1.5×10^{-2}	76	can	7.5×10^{-3}
27	freedom	1.5×10^{-2}	77	indic	7.4×10^{-3}
28	present	1.4×10^{-2}	78	group	7.3×10^{-3}
29	approach	1.4×10^{-2}	79	year	7.1×10^{-3}
30	mobil	1.3×10^{-2}	80	error	7.1×10^{-3}
31	forc	1.3×10^{-2}	81	diseas	7.1×10^{-3}
32	torqu	1.3×10^{-2}	82	investig	7×10^{-3}
33	base	1.2×10^{-2}	83	terrain	7×10^{-3}
34	movement	1.2×10^{-2}	84	aerial	6.9×10^{-3}
35	vision	1.2×10^{-2}	85	loop	6.8×10^{-3}
36	robust	1.2×10^{-2}	86	reveal	6.7×10^{-3}
37	problem	1.1×10^{-2}	87	prototyp	6.7×10^{-3}
38	signific	1.1×10^{-2}	88	higher	6.7×10^{-3}
39	gait	1.1×10^{-2}	89	planner	6.7×10^{-3}
40	plan	1.1×10^{-2}	90	speci	6.6×10^{-3}
41	arm	1.1×10^{-2}	91	learn	6.4×10^{-3}
42	feedback	1.1×10^{-2}	92	concentr	6.4×10^{-3}
43	dynam	1.1×10^{-2}	93	factor	6.3×10^{-3}
44	move	1×10^{-2}	94	background	6.3×10^{-3}
45	experi	1×10^{-2}	95	map	6.1×10^{-3}
46	implement	1×10^{-2}	96	report	6×10^{-3}
47	cell	1×10^{-2}	97	decreas	6×10^{-3}
48	motor	9.8×10^{-3}	98	patient	6×10^{-3}
49	treatment	9.6×10^{-3}	99	verifi	6×10^{-3}
50	protein	9.6×10^{-3}	100	total	6×10^{-3}

TABLE D.226. The list of the top 100 words in the category Social Issues with RIGs

No.	Word	RIG
1	articl	5.5×10^{-2}
2	argu	5.3×10^{-2}
3	ethic	4.8×10^{-2}
4	social	4.2×10^{-2}
5	polici	3.9×10^{-2}
6	polit	2.9×10^{-2}
7	moral	2.8×10^{-2}
8	welfar	2.6×10^{-2}
9	public	2.1×10^{-2}
10	peopl	1.9×10^{-2}
11	govern	1.8×10^{-2}
12	health	1.8×10^{-2}
13	bioethic	1.8×10^{-2}
14	author	1.6×10^{-2}
15	legal	1.6×10^{-2}
16	result	1.6×10^{-2}
17	question	1.5×10^{-2}
18	debat	1.5×10^{-2}
19	right	1.5×10^{-2}
20	argument	1.4×10^{-2}
21	claim	1.4×10^{-2}
22	research	1.4×10^{-2}
23	countri	1.4×10^{-2}
24	reform	1.4×10^{-2}
25	draw	1.3×10^{-2}
26	interview	1.3×10^{-2}
27	perform	1.3×10^{-2}
28	profession	1.2×10^{-2}
29	religi	1.2×10^{-2}
30	nation	1.2×10^{-2}
31	societi	1.2×10^{-2}
32	engag	1.1×10^{-2}
33	autonomi	1.1×10^{-2}
34	simul	1.1×10^{-2}
35	concern	1.1×10^{-2}
36	cell	1.1×10^{-2}
37	temperatur	1.1×10^{-2}
38	explor	1.1×10^{-2}
39	focus	1.1×10^{-2}
40	practic	1×10^{-2}
41	worker	1×10^{-2}
42	examin	1×10^{-2}
43	particip	1×10^{-2}
44	care	1×10^{-2}
45	attitud	9.9×10^{-3}
46	discuss	9.9×10^{-3}
47	method	9.8×10^{-3}
48	justic	9.8×10^{-3}
49	religion	9.6×10^{-3}
50	issu	9.6×10^{-3}
51	paramet	9.5×10^{-3}
52	oblig	9.4×10^{-3}
53	surfac	9.4×10^{-3}
54	individu	9.3×10^{-3}
55	way	9.2×10^{-3}
56	make	9.1×10^{-3}
57	weapon	9×10^{-3}
58	decis	8.9×10^{-3}
59	consent	8.8×10^{-3}
60	person	8.6×10^{-3}
61	discours	8.5×10^{-3}
62	justif	8.5×10^{-3}
63	organis	8.3×10^{-3}
64	context	8.3×10^{-3}
65	implic	8.1×10^{-3}
66	institut	8×10^{-3}
67	fund	8×10^{-3}
68	observ	7.9×10^{-3}
69	ideolog	7.9×10^{-3}
70	state	7.8×10^{-3}
71	gender	7.8×10^{-3}
72	medic	7.8×10^{-3}
73	sector	7.7×10^{-3}
74	obtain	7.6×10^{-3}
75	servic	7.4×10^{-3}
76	support	7.4×10^{-3}
77	law	7.4×10^{-3}
78	protein	7.3×10^{-3}
79	women	7.3×10^{-3}
80	liber	7.2×10^{-3}
81	seek	7.2×10^{-3}
82	scholar	7.2×10^{-3}
83	normat	7.1×10^{-3}
84	undermin	7.1×10^{-3}
85	high	7.1×10^{-3}
86	harm	7×10^{-3}
87	effici	7×10^{-3}
88	detect	7×10^{-3}
89	view	7×10^{-3}
90	provis	6.9×10^{-3}
91	theme	6.9×10^{-3}
92	crisi	6.8×10^{-3}
93	electron	6.8×10^{-3}
94	parti	6.7×10^{-3}
95	experiment	6.5×10^{-3}
96	agenda	6.5×10^{-3}
97	disabl	6.3×10^{-3}
98	perceiv	6.3×10^{-3}
99	actor	6.2×10^{-3}
100	properti	6.2×10^{-3}

Table D.227. The list of the top 100 words in the category Social Sciences, Biomedical with RIGs

No.	Word	RIG	No.	Word	RIG
1	health	9.8×10^{-2}	51	status	1.2×10^{-2}
2	hiv	7.6×10^{-2}	52	theme	1.2×10^{-2}
3	social	5×10^{-2}	53	simul	1.2×10^{-2}
4	interview	4.9×10^{-2}	54	healthcar	1.1×10^{-2}
5	care	4.3×10^{-2}	55	relationship	1.1×10^{-2}
6	ethic	4.3×10^{-2}	56	survivor	1.1×10^{-2}
7	particip	3.4×10^{-2}	57	communiti	1.1×10^{-2}
8	women	3.2×10^{-2}	58	nation	1.1×10^{-2}
9	sexual	3×10^{-2}	59	consent	1.1×10^{-2}
10	medic	3×10^{-2}	60	autonomi	1.1×10^{-2}
11	moral	2.7×10^{-2}	61	decis	1.1×10^{-2}
12	intervent	2.5×10^{-2}	62	temperatur	1.1×10^{-2}
13	men	2.4×10^{-2}	63	child	1.1×10^{-2}
14	peopl	2.3×10^{-2}	64	age	1×10^{-2}
15	sex	2.3×10^{-2}	65	focus	1×10^{-2}
16	risk	2.3×10^{-2}	66	understand	1×10^{-2}
17	live	2.2×10^{-2}	67	semistructur	1×10^{-2}
18	individu	2.2×10^{-2}	68	perform	1×10^{-2}
19	among	2.2×10^{-2}	69	socioeconom	1×10^{-2}
20	argu	2.1×10^{-2}	70	surfac	1×10^{-2}
21	partner	2×10^{-2}	71	prevent	1×10^{-2}
22	research	1.9×10^{-2}	72	associ	9.7×10^{-3}
23	survey	1.9×10^{-2}	73	support	9.7×10^{-3}
24	public	1.9×10^{-2}	74	physician	9.6×10^{-3}
25	articl	1.8×10^{-2}	75	paramet	9.6×10^{-3}
26	explor	1.7×10^{-2}	76	attitud	9.6×10^{-3}
27	draw	1.7×10^{-2}	77	incom	9.5×10^{-3}
28	examin	1.7×10^{-2}	78	discours	9.4×10^{-3}
29	stigma	1.6×10^{-2}	79	energi	9.4×10^{-3}
30	profession	1.6×10^{-2}	80	popul	9.3×10^{-3}
31	person	1.6×10^{-2}	81	debat	9.3×10^{-3}
32	qualit	1.6×10^{-2}	82	respond	9.3×10^{-3}
33	polici	1.6×10^{-2}	83	parent	9.2×10^{-3}
34	counsel	1.6×10^{-2}	84	seek	9.2×10^{-3}
35	mental	1.5×10^{-2}	85	psychosoci	9.2×10^{-3}
36	antiretrovir	1.5×10^{-2}	86	properti	9.1×10^{-3}
37	ill	1.5×10^{-2}	87	experienc	9.1×10^{-3}
38	bioethic	1.5×10^{-2}	88	children	9×10^{-3}
39	ethnograph	1.5×10^{-2}	89	emot	8.9×10^{-3}
40	life	1.5×10^{-2}	90	question	8.9×10^{-3}
41	indepth	1.4×10^{-2}	91	doctor	8.8×10^{-3}
42	need	1.4×10^{-2}	92	narrat	8.7×10^{-3}
43	famili	1.4×10^{-2}	93	psycholog	8.7×10^{-3}
44	context	1.4×10^{-2}	94	address	8.5×10^{-3}
45	educ	1.4×10^{-2}	95	inform	8.4×10^{-3}
46	engag	1.3×10^{-2}	96	regress	8.4×10^{-3}
47	practic	1.3×10^{-2}	97	distress	8.4×10^{-3}
48	concern	1.3×10^{-2}	98	implic	8.3×10^{-3}
49	servic	1.3×10^{-2}	99	find	8.3×10^{-3}
50	perceiv	1.2×10^{-2}	100	countri	8.3×10^{-3}

TABLE D.228. The list of the top 100 words in the category Social Sciences, Interdisciplinary with RIGs

No.	Word	RIG
1	social	4.3×10^{-2}
2	educ	3.7×10^{-2}
3	student	3.3×10^{-2}
4	teach	3.1×10^{-2}
5	colleg	3×10^{-2}
6	research	2.9×10^{-2}
7	peopl	2.2×10^{-2}
8	practic	2.1×10^{-2}
9	societi	1.9×10^{-2}
10	articl	1.9×10^{-2}
11	talent	1.8×10^{-2}
12	reform	1.8×10^{-2}
13	cell	1.8×10^{-2}
14	china	1.8×10^{-2}
15	polit	1.7×10^{-2}
16	cultur	1.7×10^{-2}
17	econom	1.6×10^{-2}
18	put	1.5×10^{-2}
19	result	1.4×10^{-2}
20	teacher	1.4×10^{-2}
21	countri	1.3×10^{-2}
22	enterpris	1.3×10^{-2}
23	chines	1.3×10^{-2}
24	univers	1.3×10^{-2}
25	temperatur	1.3×10^{-2}
26	polici	1.2×10^{-2}
27	innov	1.2×10^{-2}
28	perspect	1.2×10^{-2}
29	explor	1.1×10^{-2}
30	school	1.1×10^{-2}
31	economi	1.1×10^{-2}
32	surfac	1.1×10^{-2}
33	situat	1.1×10^{-2}
34	paper	1.1×10^{-2}
35	nation	1.1×10^{-2}
36	develop	1.1×10^{-2}
37	protein	1.1×10^{-2}
38	english	1.1×10^{-2}
39	think	1×10^{-2}
40	induc	1×10^{-2}
41	interview	9.9×10^{-3}
42	learn	9.6×10^{-3}
43	patient	9.5×10^{-3}
44	way	9.4×10^{-3}
45	market	9.3×10^{-3}
46	vocat	9.3×10^{-3}
47	make	9.2×10^{-3}
48	detect	9×10^{-3}
49	paramet	8.8×10^{-3}
50	profession	8.8×10^{-3}

No.	Word	RIG
51	person	8.7×10^{-3}
52	acid	8.7×10^{-3}
53	modern	8.5×10^{-3}
54	treatment	8.4×10^{-3}
55	focus	8.2×10^{-3}
56	high	8.1×10^{-3}
57	ideolog	8.1×10^{-3}
58	govern	8×10^{-3}
59	perform	8×10^{-3}
60	busi	8×10^{-3}
61	forward	8×10^{-3}
62	observ	8×10^{-3}
63	gene	7.9×10^{-3}
64	psycholog	7.9×10^{-3}
65	aspect	7.8×10^{-3}
66	discuss	7.8×10^{-3}
67	compar	7.7×10^{-3}
68	properti	7.7×10^{-3}
69	obtain	7.6×10^{-3}
70	survey	7.6×10^{-3}
71	concept	7.5×10^{-3}
72	manag	7.5×10^{-3}
73	life	7.5×10^{-3}
74	becom	7.5×10^{-3}
75	author	7.5×10^{-3}
76	languag	7.4×10^{-3}
77	simul	7.4×10^{-3}
78	decreas	7.4×10^{-3}
79	speci	7.3×10^{-3}
80	public	7.3×10^{-3}
81	oxid	7.2×10^{-3}
82	show	7.2×10^{-3}
83	foreign	7.2×10^{-3}
84	use	7.1×10^{-3}
85	cultiv	7.1×10^{-3}
86	theori	7.1×10^{-3}
87	molecular	7×10^{-3}
88	concentr	7×10^{-3}
89	issu	6.9×10^{-3}
90	rang	6.8×10^{-3}
91	knowledg	6.8×10^{-3}
92	particip	6.7×10^{-3}
93	curriculum	6.7×10^{-3}
94	experiment	6.7×10^{-3}
95	relationship	6.7×10^{-3}
96	densiti	6.6×10^{-3}
97	implic	6.5×10^{-3}
98	energi	6.5×10^{-3}
99	promot	6.5×10^{-3}
100	ratio	6.4×10^{-3}

TABLE D.229. The list of the top 100 words in the category Social Sciences, Mathematical Methods with RIGs

No.	Word	RIG
1	price	3.3×10^{-2}
2	market	2.9×10^{-2}
3	model	2.6×10^{-2}
4	estim	2.4×10^{-2}
5	financi	2.3×10^{-2}
6	empir	2.3×10^{-2}
7	paper	1.9×10^{-2}
8	econom	1.7×10^{-2}
9	asset	1.5×10^{-2}
10	cell	1.4×10^{-2}
11	stochast	1.4×10^{-2}
12	czech	1.4×10^{-2}
13	portfolio	1.3×10^{-2}
14	patient	1.3×10^{-2}
15	insur	1.2×10^{-2}
16	economi	1.2×10^{-2}
17	conclus	1.1×10^{-2}
18	stock	1.1×10^{-2}
19	asymptot	1.1×10^{-2}
20	nonparametr	1.1×10^{-2}
21	econometr	1×10^{-2}
22	forecast	1×10^{-2}
23	articl	1×10^{-2}
24	carlo	9.3×10^{-3}
25	high	9.2×10^{-3}
26	mont	9.2×10^{-3}
27	temperatur	9.1×10^{-3}
28	surfac	9×10^{-3}
29	protein	8.9×10^{-3}
30	invest	8.8×10^{-3}
31	volatil	8.7×10^{-3}
32	macroeconom	8.5×10^{-3}
33	bayesian	8.3×10^{-3}
34	capit	8.3×10^{-3}
35	polici	8.1×10^{-3}
36	clinic	8×10^{-3}
37	risk	8×10^{-3}
38	background	7.9×10^{-3}
39	illustr	7.8×10^{-3}
40	compani	7.7×10^{-3}
41	republ	7.7×10^{-3}
42	variabl	7.6×10^{-3}
43	return	7.5×10^{-3}
44	data	7.4×10^{-3}
45	electron	7.4×10^{-3}
46	acid	7.3×10^{-3}
47	activ	7.2×10^{-3}
48	pedestrian	7.1×10^{-3}
49	countri	7×10^{-3}
50	assumpt	7×10^{-3}
51	likelihood	6.9×10^{-3}
52	materi	6.9×10^{-3}
53	decis	6.8×10^{-3}
54	diseas	6.7×10^{-3}
55	statist	6.7×10^{-3}
56	crisi	6.7×10^{-3}
57	trade	6.7×10^{-3}
58	premium	6.6×10^{-3}
59	investor	6.5×10^{-3}
60	gene	6.5×10^{-3}
61	probabl	6.5×10^{-3}
62	panel	6.4×10^{-3}
63	speci	6.3×10^{-3}
64	molecular	6.2×10^{-3}
65	game	6.1×10^{-3}
66	general	6×10^{-3}
67	problem	5.9×10^{-3}
68	firm	5.9×10^{-3}
69	energi	5.9×10^{-3}
70	found	5.8×10^{-3}
71	choic	5.6×10^{-3}
72	avers	5.6×10^{-3}
73	treatment	5.6×10^{-3}
74	varianc	5.6×10^{-3}
75	exampl	5.5×10^{-3}
76	oxid	5.5×10^{-3}
77	set	5.5×10^{-3}
78	markov	5.5×10^{-3}
79	prefer	5.4×10^{-3}
80	low	5.4×10^{-3}
81	phase	5.3×10^{-3}
82	signific	5.2×10^{-3}
83	imag	5.2×10^{-3}
84	report	5.2×10^{-3}
85	induc	5.1×10^{-3}
86	financ	5.1×10^{-3}
87	studi	5×10^{-3}
88	tissu	5×10^{-3}
89	credit	5×10^{-3}
90	control	4.9×10^{-3}
91	gdp	4.8×10^{-3}
92	monetari	4.8×10^{-3}
93	latent	4.7×10^{-3}
94	ray	4.7×10^{-3}
95	profit	4.7×10^{-3}
96	optic	4.7×10^{-3}
97	mechan	4.7×10^{-3}
98	concentr	4.6×10^{-3}
99	resist	4.6×10^{-3}
100	incom	4.6×10^{-3}

Table D.230. The list of the top 100 words in the category Social Work with RIGs

No.	Word	RIG	No.	Word	RIG
1	social	1.4×10^{-1}	51	experienc	1.5×10^{-2}
2	child	9.8×10^{-2}	52	cell	1.5×10^{-2}
3	children	6.8×10^{-2}	53	psycholog	1.4×10^{-2}
4	worker	6×10^{-2}	54	mother	1.4×10^{-2}
5	welfar	5.5×10^{-2}	55	context	1.4×10^{-2}
6	servic	5.1×10^{-2}	56	perform	1.3×10^{-2}
7	famili	5.1×10^{-2}	57	perceiv	1.3×10^{-2}
8	abus	4.9×10^{-2}	58	focus	1.3×10^{-2}
9	parent	4.5×10^{-2}	59	method	1.3×10^{-2}
10	articl	4.2×10^{-2}	60	draw	1.3×10^{-2}
11	maltreat	4.1×10^{-2}	61	perspect	1.2×10^{-2}
12	research	4.1×10^{-2}	62	poverti	1.2×10^{-2}
13	youth	4.1×10^{-2}	63	program	1.2×10^{-2}
14	practic	4×10^{-2}	64	profess	1.2×10^{-2}
15	polici	3.4×10^{-2}	65	young	1.2×10^{-2}
16	care	3.4×10^{-2}	66	simul	1.2×10^{-2}
17	interview	3.3×10^{-2}	67	outcom	1.2×10^{-2}
18	health	3.1×10^{-2}	68	temperatur	1.2×10^{-2}
19	mental	3.1×10^{-2}	69	address	1.2×10^{-2}
20	implic	3×10^{-2}	70	neglect	1.1×10^{-2}
21	communiti	2.9×10^{-2}	71	risk	1.1×10^{-2}
22	examin	2.8×10^{-2}	72	paramet	1.1×10^{-2}
23	intervent	2.8×10^{-2}	73	incom	1.1×10^{-2}
24	find	2.7×10^{-2}	74	properti	1.1×10^{-2}
25	support	2.7×10^{-2}	75	surfac	1.1×10^{-2}
26	work	2.5×10^{-2}	76	experi	1.1×10^{-2}
27	educ	2.4×10^{-2}	77	survey	1.1×10^{-2}
28	violenc	2.4×10^{-2}	78	nation	1.1×10^{-2}
29	adolesc	2.4×10^{-2}	79	percept	1.1×10^{-2}
30	profession	2.3×10^{-2}	80	theme	1×10^{-2}
31	particip	2.2×10^{-2}	81	trauma	1×10^{-2}
32	discuss	2.2×10^{-2}	82	among	1×10^{-2}
33	relationship	2.1×10^{-2}	83	energi	1×10^{-2}
34	victim	2.1×10^{-2}	84	person	9.8×10^{-3}
35	caregiv	2×10^{-2}	85	peer	9.6×10^{-3}
36	explor	2×10^{-2}	86	student	9.4×10^{-3}
37	foster	2×10^{-2}	87	induc	9.2×10^{-3}
38	practition	1.9×10^{-2}	88	psychosoci	9.2×10^{-3}
39	engag	1.9×10^{-2}	89	provis	9.1×10^{-3}
40	school	1.9×10^{-2}	90	organis	9×10^{-3}
41	client	1.9×10^{-2}	91	gender	8.9×10^{-3}
42	qualit	1.9×10^{-2}	92	empower	8.6×10^{-3}
43	need	1.8×10^{-2}	93	challeng	8.5×10^{-3}
44	emot	1.8×10^{-2}	94	protein	8.4×10^{-3}
45	sexual	1.8×10^{-2}	95	show	8.3×10^{-3}
46	agenc	1.8×10^{-2}	96	seek	8.3×10^{-3}
47	peopl	1.7×10^{-2}	97	indepth	8.2×10^{-3}
48	live	1.6×10^{-2}	98	result	8.2×10^{-3}
49	home	1.6×10^{-2}	99	protect	8.2×10^{-3}
50	childhood	1.5×10^{-2}	100	issu	8.2×10^{-3}

Table D.231. The list of the top 100 words in the category Sociology with RIGs

No.	Word	RIG	No.	Word	RIG
1	social	1.2×10^{-1}	51	communiti	1.2×10^{-2}
2	articl	8.1×10^{-2}	52	perform	1.1×10^{-2}
3	polit	5.5×10^{-2}	53	actor	1.1×10^{-2}
4	sociolog	5.5×10^{-2}	54	migrant	1.1×10^{-2}
5	argu	4.3×10^{-2}	55	obtain	1.1×10^{-2}
6	draw	3.5×10^{-2}	56	patient	1.1×10^{-2}
7	cultur	2.7×10^{-2}	57	simul	1.1×10^{-2}
8	research	2.7×10^{-2}	58	surfac	1.1×10^{-2}
9	nation	2.6×10^{-2}	59	understand	1.1×10^{-2}
10	examin	2.5×10^{-2}	60	practic	1×10^{-2}
11	societi	2.5×10^{-2}	61	individu	1×10^{-2}
12	interview	2.3×10^{-2}	62	attitud	9.9×10^{-3}
13	peopl	2.2×10^{-2}	63	theori	9.8×10^{-3}
14	survey	2.2×10^{-2}	64	concept	9.7×10^{-3}
15	discours	2.1×10^{-2}	65	market	9.6×10^{-3}
16	religi	2.1×10^{-2}	66	school	9.6×10^{-3}
17	econom	2×10^{-2}	67	paramet	9.6×10^{-3}
18	polici	1.9×10^{-2}	68	live	9.5×10^{-3}
19	gender	1.8×10^{-2}	69	american	9.4×10^{-3}
20	context	1.8×10^{-2}	70	particip	9.3×10^{-3}
21	immigr	1.7×10^{-2}	71	protest	9.2×10^{-3}
22	ethnograph	1.7×10^{-2}	72	claim	9.1×10^{-3}
23	method	1.7×10^{-2}	73	narrat	9×10^{-3}
24	racial	1.7×10^{-2}	74	conceptu	8.9×10^{-3}
25	public	1.7×10^{-2}	75	socio	8.9×10^{-3}
26	cell	1.6×10^{-2}	76	detect	8.8×10^{-3}
27	educ	1.5×10^{-2}	77	protein	8.8×10^{-3}
28	focus	1.5×10^{-2}	78	women	8.7×10^{-3}
29	contemporari	1.5×10^{-2}	79	author	8.6×10^{-3}
30	way	1.5×10^{-2}	80	moral	8.6×10^{-3}
31	engag	1.5×10^{-2}	81	govern	8.4×10^{-3}
32	countri	1.5×10^{-2}	82	incom	8.3×10^{-3}
33	inequ	1.5×10^{-2}	83	implic	8.3×10^{-3}
34	find	1.4×10^{-2}	84	youth	8.2×10^{-3}
35	question	1.4×10^{-2}	85	induc	8.1×10^{-3}
36	scholar	1.4×10^{-2}	86	person	8.1×10^{-3}
37	ident	1.4×10^{-2}	87	energi	8×10^{-3}
38	religion	1.3×10^{-2}	88	ratio	7.9×10^{-3}
39	capit	1.3×10^{-2}	89	effici	7.9×10^{-3}
40	ethnic	1.3×10^{-2}	90	electron	7.9×10^{-3}
41	empir	1.3×10^{-2}	91	liber	7.8×10^{-3}
42	result	1.3×10^{-2}	92	discurs	7.8×10^{-3}
43	explor	1.3×10^{-2}	93	whi	7.8×10^{-3}
44	institut	1.3×10^{-2}	94	qualit	7.8×10^{-3}
45	ideolog	1.3×10^{-2}	95	histor	7.8×10^{-3}
46	temperatur	1.2×10^{-2}	96	algorithm	7.8×10^{-3}
47	perspect	1.2×10^{-2}	97	discuss	7.8×10^{-3}
48	debat	1.2×10^{-2}	98	labour	7.8×10^{-3}
49	relationship	1.2×10^{-2}	99	function	7.7×10^{-3}
50	life	1.2×10^{-2}	100	labor	7.6×10^{-3}

TABLE D.232. The list of the top 100 words in the category Soil Science with RIGs

No.	Word	RIG	No.	Word	RIG
1	soil	4.1×10^{-1}	51	greenhous	1.4×10^{-2}
2	plant	6.5×10^{-2}	52	decreas	1.3×10^{-2}
3	crop	6.5×10^{-2}	53	fungi	1.3×10^{-2}
4	fertil	5.5×10^{-2}	54	shoot	1.3×10^{-2}
5	organ	5.4×10^{-2}	55	maiz	1.3×10^{-2}
6	content	4.7×10^{-2}	56	experi	1.3×10^{-2}
7	microbi	4.7×10^{-2}	57	unsatur	1.2×10^{-2}
8	agricultur	4.2×10^{-2}	58	avail	1.2×10^{-2}
9	nutrient	4.2×10^{-2}	59	irrig	1.2×10^{-2}
10	water	4.2×10^{-2}	60	suction	1.2×10^{-2}
11	biomass	4.1×10^{-2}	61	site	1.2×10^{-2}
12	carbon	3.9×10^{-2}	62	accumul	1.2×10^{-2}
13	nitrogen	3.9×10^{-2}	63	rainfal	1.2×10^{-2}
14	clay	3.8×10^{-2}	64	communiti	1.2×10^{-2}
15	miner	3.6×10^{-2}	65	indic	1.1×10^{-2}
16	root	3.5×10^{-2}	66	sampl	1.1×10^{-2}
17	amend	3.4×10^{-2}	67	manag	1.1×10^{-2}
18	tillag	3.3×10^{-2}	68	semiarid	1.1×10^{-2}
19	forest	3.2×10^{-2}	69	leaf	1.1×10^{-2}
20	matter	3.2×10^{-2}	70	area	1.1×10^{-2}
21	land	3×10^{-2}	71	cover	1.1×10^{-2}
22	ecosystem	2.6×10^{-2}	72	yield	1.1×10^{-2}
23	dri	2.5×10^{-2}	73	leach	1.1×10^{-2}
24	depth	2.5×10^{-2}	74	clinic	1.1×10^{-2}
25	veget	2.4×10^{-2}	75	rice	1.1×10^{-2}
26	phosphorus	2.4×10^{-2}	76	effect	1×10^{-2}
27	moistur	2.2×10^{-2}	77	wet	1×10^{-2}
28	manur	2.1×10^{-2}	78	fraction	1×10^{-2}
29	litter	2×10^{-2}	79	signific	1×10^{-2}
30	sandi	1.9×10^{-2}	80	bulk	1×10^{-2}
31	increas	1.9×10^{-2}	81	textur	1×10^{-2}
32	wheat	1.9×10^{-2}	82	straw	1×10^{-2}
33	patient	1.8×10^{-2}	83	grown	1×10^{-2}
34	season	1.8×10^{-2}	84	slope	1×10^{-2}
35	plot	1.8×10^{-2}	85	environment	1×10^{-2}
36	concentr	1.8×10^{-2}	86	satur	9.9×10^{-3}
37	grassland	1.7×10^{-2}	87	greater	9.8×10^{-3}
38	incub	1.7×10^{-2}	88	chemic	9.6×10^{-3}
39	eros	1.7×10^{-2}	89	no3	9.6×10^{-3}
40	field	1.6×10^{-2}	90	nh4	9.6×10^{-3}
41	total	1.6×10^{-2}	91	product	9.5×10^{-3}
42	differ	1.6×10^{-2}	92	growth	9.2×10^{-3}
43	sand	1.6×10^{-2}	93	studi	9.1×10^{-3}
44	horizon	1.5×10^{-2}	94	grass	9.1×10^{-3}
45	cultiv	1.5×10^{-2}	95	chang	9×10^{-3}
46	climat	1.5×10^{-2}	96	cell	9×10^{-3}
47	uptak	1.4×10^{-2}	97	influenc	9×10^{-3}
48	affect	1.4×10^{-2}	98	nitrat	8.9×10^{-3}
49	respir	1.4×10^{-2}	99	conduct	8.9×10^{-3}
50	compost	1.4×10^{-2}	100	co2	8.9×10^{-3}

TABLE D.233. The list of the top 100 words in the category Spectroscopy with RIGs

No.	Word	RIG	No.	Word	RIG
1	spectra	6.9×10^{-2}	51	proton	8.9×10^{-3}
2	spectrometri	4.4×10^{-2}	52	molecular	8.9×10^{-3}
3	spectroscopi	3.9×10^{-2}	53	concentr	8.8×10^{-3}
4	raman	3.3×10^{-2}	54	year	8.6×10^{-3}
5	ion	3.2×10^{-2}	55	lod	8.5×10^{-3}
6	spectral	2.6×10^{-2}	56	paper	8.4×10^{-3}
7	nmr	2.4×10^{-2}	57	linear	8.4×10^{-3}
8	ioniz	2.3×10^{-2}	58	sensit	8.3×10^{-3}
9	sampl	2.3×10^{-2}	59	method	8.3×10^{-3}
10	spectromet	2.3×10^{-2}	60	optic	8×10^{-3}
11	infrar	2.2×10^{-2}	61	tandem	8×10^{-3}
12	absorpt	2.1×10^{-2}	62	coupl	8×10^{-3}
13	mass	2.1×10^{-2}	63	recoveri	8×10^{-3}
14	detect	2×10^{-2}	64	icp	7.8×10^{-3}
15	vibrat	2×10^{-2}	65	limit	7.7×10^{-3}
16	b3lyp	1.8×10^{-2}	66	ray	7.6×10^{-3}
17	chromatographi	1.8×10^{-2}	67	gas	7.4×10^{-3}
18	compound	1.8×10^{-2}	68	analysi	7.2×10^{-3}
19	molecul	1.7×10^{-2}	69	solid	7.1×10^{-3}
20	analyt	1.6×10^{-2}	70	character	7.1×10^{-3}
21	electrospray	1.6×10^{-2}	71	pyrolysi	7.1×10^{-3}
22	dft	1.5×10^{-2}	72	rsd	7×10^{-3}
23	fluoresc	1.5×10^{-2}	73	emiss	7×10^{-3}
24	chemic	1.4×10^{-2}	74	techniqu	7×10^{-3}
25	spectroscop	1.4×10^{-2}	75	deviat	6.8×10^{-3}
26	band	1.3×10^{-2}	76	hplc	6.7×10^{-3}
27	calcul	1.3×10^{-2}	77	extract	6.6×10^{-3}
28	determin	1.3×10^{-2}	78	peak	6.6×10^{-3}
29	excit	1.2×10^{-2}	79	conclus	6.5×10^{-3}
30	laser	1.2×10^{-2}	80	isotop	6.4×10^{-3}
31	bond	1.2×10^{-2}	81	associ	6.4×10^{-3}
32	patient	1.2×10^{-2}	82	wavelength	6.4×10^{-3}
33	vis	1.2×10^{-2}	83	dissoci	6.4×10^{-3}
34	liquid	1.1×10^{-2}	84	ftir	6.3×10^{-3}
35	rational	1.1×10^{-2}	85	hydrogen	6.3×10^{-3}
36	fourier	1.1×10^{-2}	86	risk	6.3×10^{-3}
37	electron	1.1×10^{-2}	87	resolut	6.3×10^{-3}
38	reson	1.1×10^{-2}	88	synthes	6.3×10^{-3}
39	atom	1.1×10^{-2}	89	use	6.2×10^{-3}
40	calibr	1.1×10^{-2}	90	age	6.2×10^{-3}
41	rang	1.1×10^{-2}	91	energi	6.2×10^{-3}
42	assign	1.1×10^{-2}	92	control	6.1×10^{-3}
43	intens	1.1×10^{-2}	93	manag	6.1×10^{-3}
44	esi	1×10^{-2}	94	crystal	6.1×10^{-3}
45	obtain	9.5×10^{-3}	95	shift	6.1×10^{-3}
46	solvent	9.4×10^{-3}	96	fragment	5.8×10^{-3}
47	quadrupol	9.4×10^{-3}	97	outcom	5.8×10^{-3}
48	spectrum	9.3×10^{-3}	98	appli	5.7×10^{-3}
49	quantif	9.1×10^{-3}	99	object	5.7×10^{-3}
50	acid	9×10^{-3}	100	prepar	5.7×10^{-3}

Table D.234. The list of the top 100 words in the category Sport Sciences with RIGs

No.	Word	RIG
1	athlet	9.2×10^{-2}
2	exercis	7.9×10^{-2}
3	knee	7.8×10^{-2}
4	sport	7.7×10^{-2}
5	particip	5.4×10^{-2}
6	train	4.9×10^{-2}
7	muscl	4.9×10^{-2}
8	purpos	4.1×10^{-2}
9	player	4×10^{-2}
10	conclus	4×10^{-2}
11	flexion	3.5×10^{-2}
12	injuri	3.3×10^{-2}
13	dure	3.2×10^{-2}
14	age	3.2×10^{-2}
15	male	3.1×10^{-2}
16	studi	2.9×10^{-2}
17	signific	2.8×10^{-2}
18	leg	2.8×10^{-2}
19	measur	2.8×10^{-2}
20	bodi	2.7×10^{-2}
21	anterior	2.7×10^{-2}
22	shoulder	2.6×10^{-2}
23	ligament	2.4×10^{-2}
24	elit	2.4×10^{-2}
25	cruciat	2.4×10^{-2}
26	coach	2.3×10^{-2}
27	ankl	2.3×10^{-2}
28	session	2.3×10^{-2}
29	men	2.3×10^{-2}
30	outcom	2.2×10^{-2}
31	year	2.2×10^{-2}
32	medial	2.2×10^{-2}
33	hip	2.1×10^{-2}
34	rehabilit	2.1×10^{-2}
35	walk	2.1×10^{-2}
36	group	2.1×10^{-2}
37	min	2.1×10^{-2}
38	assess	2.1×10^{-2}
39	pain	2×10^{-2}
40	physic	2×10^{-2}
41	paper	2×10^{-2}
42	limb	2×10^{-2}
43	subject	2×10^{-2}
44	greater	2×10^{-2}
45	perform	2×10^{-2}
46	arthroscop	2×10^{-2}
47	kinemat	1.9×10^{-2}
48	score	1.9×10^{-2}
49	intervent	1.8×10^{-2}
50	healthi	1.8×10^{-2}
51	mean	1.8×10^{-2}
52	arthroplasti	1.8×10^{-2}
53	biomechan	1.8×10^{-2}
54	joint	1.7×10^{-2}
55	week	1.7×10^{-2}
56	maxim	1.7×10^{-2}
57	tibial	1.7×10^{-2}
58	tendon	1.7×10^{-2}
59	gait	1.7×10^{-2}
60	footbal	1.6×10^{-2}
61	strength	1.6×10^{-2}
62	jump	1.5×10^{-2}
63	endur	1.5×10^{-2}
64	femor	1.5×10^{-2}
65	aerob	1.4×10^{-2}
66	peak	1.4×10^{-2}
67	motion	1.3×10^{-2}
68	befor	1.3×10^{-2}
69	lower	1.3×10^{-2}
70	elbow	1.3×10^{-2}
71	differ	1.3×10^{-2}
72	rest	1.3×10^{-2}
73	test	1.2×10^{-2}
74	movement	1.2×10^{-2}
75	trial	1.2×10^{-2}
76	femal	1.2×10^{-2}
77	compar	1.2×10^{-2}
78	propos	1.2×10^{-2}
79	osteoarthr	1.2×10^{-2}
80	follow	1.2×10^{-2}
81	postur	1.2×10^{-2}
82	height	1.1×10^{-2}
83	minut	1.1×10^{-2}
84	foot	1.1×10^{-2}
85	posterior	1.1×10^{-2}
86	object	1.1×10^{-2}
87	voluntari	1.1×10^{-2}
88	aim	1.1×10^{-2}
89	young	1.1×10^{-2}
90	examin	1.1×10^{-2}
91	patient	1.1×10^{-2}
92	stanc	1.1×10^{-2}
93	twenti	1.1×10^{-2}
94	run	1.1×10^{-2}
95	later	1.1×10^{-2}
96	team	1×10^{-2}
97	complet	1×10^{-2}
98	repetit	1×10^{-2}
99	total	9.9×10^{-3}
100	rotat	9.7×10^{-3}

Table D.235. The list of the top 100 words in the category Statistics and Probability with RIGs

No.	Word	RIG	No.	Word	RIG
1	estim	5.3×10^{-2}	51	procedur	8.4×10^{-3}
2	asymptot	4.9×10^{-2}	52	energi	8.3×10^{-3}
3	illustr	3.3×10^{-2}	53	report	8.3×10^{-3}
4	likelihood	3.3×10^{-2}	54	given	8.2×10^{-3}
5	random	2.9×10^{-2}	55	squar	8.2×10^{-3}
6	bayesian	2.8×10^{-2}	56	electron	8.2×10^{-3}
7	nonparametr	2.7×10^{-2}	57	surfac	8.1×10^{-3}
8	distribut	2.7×10^{-2}	58	temperatur	8×10^{-3}
9	covari	2.7×10^{-2}	59	patient	8×10^{-3}
10	model	2.6×10^{-2}	60	found	8×10^{-3}
11	statist	2.4×10^{-2}	61	parametr	7.9×10^{-3}
12	simul	2.1×10^{-2}	62	theorem	7.9×10^{-3}
13	infer	2.1×10^{-2}	63	function	7.8×10^{-3}
14	markov	2×10^{-2}	64	low	7.7×10^{-3}
15	data	2×10^{-2}	65	prove	7.7×10^{-3}
16	stochast	1.9×10^{-2}	66	decreas	7.5×10^{-3}
17	propos	1.9×10^{-2}	67	algorithm	7.5×10^{-3}
18	probabl	1.9×10^{-2}	68	system	7.4×10^{-3}
19	motiv	1.8×10^{-2}	69	linear	7.3×10^{-3}
20	general	1.6×10^{-2}	70	infin	7.3×10^{-3}
21	conclus	1.6×10^{-2}	71	higher	7.2×10^{-3}
22	variabl	1.5×10^{-2}	72	mechan	7.1×10^{-3}
23	exampl	1.5×10^{-2}	73	high	7×10^{-3}
24	carlo	1.5×10^{-2}	74	comput	6.9×10^{-3}
25	mont	1.5×10^{-2}	75	assum	6.8×10^{-3}
26	activ	1.4×10^{-2}	76	multivari	6.3×10^{-3}
27	regress	1.4×10^{-2}	77	moment	6.3×10^{-3}
28	varianc	1.3×10^{-2}	78	enhanc	6.2×10^{-3}
29	assumpt	1.3×10^{-2}	79	oxid	6.2×10^{-3}
30	set	1.3×10^{-2}	80	kernel	6.1×10^{-3}
31	converg	1.2×10^{-2}	81	prepar	6×10^{-3}
32	signific	1.2×10^{-2}	82	deriv	5.9×10^{-3}
33	increas	1.2×10^{-2}	83	introduc	5.8×10^{-3}
34	error	1.2×10^{-2}	84	acid	5.6×10^{-3}
35	articl	1.1×10^{-2}	85	appli	5.5×10^{-3}
36	consid	1.1×10^{-2}	86	induc	5.5×10^{-3}
37	gaussian	1.1×10^{-2}	87	bound	5.3×10^{-3}
38	dure	1.1×10^{-2}	88	thermal	5.3×10^{-3}
39	approach	1×10^{-2}	89	posterior	5.3×10^{-3}
40	class	9.6×10^{-3}	90	day	5.3×10^{-3}
41	finit	9.5×10^{-3}	91	age	5.3×10^{-3}
42	background	9.1×10^{-3}	92	optic	5.3×10^{-3}
43	sampl	9×10^{-3}	93	year	5.2×10^{-3}
44	problem	9×10^{-3}	94	bias	5.1×10^{-3}
45	paramet	9×10^{-3}	95	layer	5.1×10^{-3}
46	cell	8.9×10^{-3}	96	water	5×10^{-3}
47	exponenti	8.7×10^{-3}	97	fit	4.9×10^{-3}
48	empir	8.5×10^{-3}	98	explicit	4.9×10^{-3}
49	real	8.5×10^{-3}	99	ray	4.8×10^{-3}
50	dataset	8.5×10^{-3}	100	resist	4.8×10^{-3}

TABLE D.236. The list of the top 100 words in the category Substance Abuse with RIGs

No.	Word	RIG	No.	Word	RIG
1	alcohol	2.1×10^{-1}	51	male	1.9×10^{-2}
2	drink	1.1×10^{-1}	52	youth	1.9×10^{-2}
3	substanc	9.2×10^{-2}	53	report	1.9×10^{-2}
4	conclus	7.5×10^{-2}	54	social	1.9×10^{-2}
5	addict	6.7×10^{-2}	55	sampl	1.8×10^{-2}
6	drug	6.7×10^{-2}	56	demograph	1.8×10^{-2}
7	smoke	6.5×10^{-2}	57	motiv	1.8×10^{-2}
8	particip	6.2×10^{-2}	58	year	1.7×10^{-2}
9	abus	6.1×10^{-2}	59	may	1.7×10^{-2}
10	abstin	5.8×10^{-2}	60	use	1.7×10^{-2}
11	smoker	5.3×10^{-2}	61	logist	1.6×10^{-2}
12	background	4.7×10^{-2}	62	outcom	1.6×10^{-2}
13	cigarett	4.6×10^{-2}	63	past	1.6×10^{-2}
14	drinker	4.6×10^{-2}	64	whether	1.5×10^{-2}
15	among	4.6×10^{-2}	65	gender	1.5×10^{-2}
16	tobacco	4.4×10^{-2}	66	nation	1.5×10^{-2}
17	examin	4.3×10^{-2}	67	student	1.5×10^{-2}
18	associ	4.2×10^{-2}	68	depend	1.4×10^{-2}
19	cocain	3.8×10^{-2}	69	men	1.4×10^{-2}
20	intervent	3.8×10^{-2}	70	femal	1.4×10^{-2}
21	adolesc	3.8×10^{-2}	71	research	1.4×10^{-2}
22	cessat	3.7×10^{-2}	72	women	1.4×10^{-2}
23	disord	3.6×10^{-2}	73	heavi	1.4×10^{-2}
24	nicotin	3.5×10^{-2}	74	greater	1.4×10^{-2}
25	risk	3.3×10^{-2}	75	recruit	1.4×10^{-2}
26	illicit	3.2×10^{-2}	76	aim	1.3×10^{-2}
27	opioid	3.1×10^{-2}	77	propos	1.3×10^{-2}
28	health	3.1×10^{-2}	78	ethanol	1.3×10^{-2}
29	gambl	3.1×10^{-2}	79	young	1.3×10^{-2}
30	consumpt	3×10^{-2}	80	baselin	1.3×10^{-2}
31	survey	2.9×10^{-2}	81	mental	1.3×10^{-2}
32	interview	2.8×10^{-2}	82	colleg	1.3×10^{-2}
33	treatment	2.6×10^{-2}	83	odd	1.3×10^{-2}
34	behavior	2.6×10^{-2}	84	suggest	1.3×10^{-2}
35	age	2.5×10^{-2}	85	day	1.3×10^{-2}
36	selfreport	2.5×10^{-2}	86	paper	1.3×10^{-2}
37	find	2.4×10^{-2}	87	prevent	1.2×10^{-2}
38	heroin	2.4×10^{-2}	88	withdraw	1.2×10^{-2}
39	assess	2.4×10^{-2}	89	riski	1.2×10^{-2}
40	studi	2.3×10^{-2}	90	person	1.2×10^{-2}
41	adult	2.3×10^{-2}	91	negat	1.2×10^{-2}
42	preval	2.2×10^{-2}	92	complet	1.2×10^{-2}
43	individu	2.2×10^{-2}	93	psychiatr	1.2×10^{-2}
44	method	2.2×10^{-2}	94	result	1.2×10^{-2}
45	regress	2.1×10^{-2}	95	temperatur	1.2×10^{-2}
46	harm	2×10^{-2}	96	questionnair	1.1×10^{-2}
47	seek	2×10^{-2}	97	relaps	1.1×10^{-2}
48	user	2×10^{-2}	98	surfac	1.1×10^{-2}
49	relat	2×10^{-2}	99	group	1.1×10^{-2}
50	month	2×10^{-2}	100	relationship	1.1×10^{-2}

TABLE D.237. The list of the top 100 words in the category Surgery with RIGs

No.	Word	RIG	No.	Word	RIG
1	patient	2.2×10^{-1}	51	pain	2×10^{-2}
2	surgeri	1.5×10^{-1}	52	tumor	2×10^{-2}
3	surgic	1.3×10^{-1}	53	total	2×10^{-2}
4	postop	1.2×10^{-1}	54	result	2×10^{-2}
5	conclus	1.2×10^{-1}	55	januari	2×10^{-2}
6	underw	1×10^{-1}	56	aneurysm	1.9×10^{-2}
7	complic	8×10^{-2}	57	wound	1.9×10^{-2}
8	outcom	7.1×10^{-2}	58	rate	1.9×10^{-2}
9	retrospect	6.9×10^{-2}	59	endoscop	1.8×10^{-2}
10	preoper	6.4×10^{-2}	60	evalu	1.8×10^{-2}
11	resect	6.2×10^{-2}	61	endovascular	1.8×10^{-2}
12	background	6.2×10^{-2}	62	flap	1.8×10^{-2}
13	year	6×10^{-2}	63	abdomin	1.8×10^{-2}
14	surgeon	5.7×10^{-2}	64	oper	1.8×10^{-2}
15	month	4.9×10^{-2}	65	anterior	1.7×10^{-2}
16	clinic	4.6×10^{-2}	66	propos	1.7×10^{-2}
17	follow	4.3×10^{-2}	67	trauma	1.7×10^{-2}
18	procedur	4.2×10^{-2}	68	dissect	1.7×10^{-2}
19	method	3.8×10^{-2}	69	multivari	1.7×10^{-2}
20	review	3.8×10^{-2}	70	object	1.6×10^{-2}
21	hospit	3.5×10^{-2}	71	bypass	1.6×10^{-2}
22	paper	3.5×10^{-2}	72	lesion	1.6×10^{-2}
23	laparoscop	3.4×10^{-2}	73	recipi	1.6×10^{-2}
24	graft	3.4×10^{-2}	74	reoper	1.6×10^{-2}
25	treatment	3.2×10^{-2}	75	aortic	1.6×10^{-2}
26	undergo	3.2×10^{-2}	76	associ	1.5×10^{-2}
27	morbid	3.1×10^{-2}	77	includ	1.5×10^{-2}
28	mortal	3.1×10^{-2}	78	reconstruct	1.5×10^{-2}
29	recurr	3.1×10^{-2}	79	incid	1.4×10^{-2}
30	intraop	3.1×10^{-2}	80	primari	1.4×10^{-2}
31	median	3.1×10^{-2}	81	anatom	1.4×10^{-2}
32	treat	3×10^{-2}	82	distal	1.4×10^{-2}
33	age	3×10^{-2}	83	diseas	1.4×10^{-2}
34	transplant	2.9×10^{-2}	84	implant	1.4×10^{-2}
35	surviv	2.8×10^{-2}	85	aim	1.3×10^{-2}
36	group	2.6×10^{-2}	86	medic	1.3×10^{-2}
37	arteri	2.6×10^{-2}	87	temperatur	1.3×10^{-2}
38	prospect	2.6×10^{-2}	88	process	1.3×10^{-2}
39	injuri	2.6×10^{-2}	89	perform	1.3×10^{-2}
40	safe	2.5×10^{-2}	90	properti	1.3×10^{-2}
41	stay	2.5×10^{-2}	91	invas	1.3×10^{-2}
42	case	2.5×10^{-2}	92	overal	1.3×10^{-2}
43	score	2.4×10^{-2}	93	posterior	1.3×10^{-2}
44	signific	2.3×10^{-2}	94	incis	1.3×10^{-2}
45	consecut	2.3×10^{-2}	95	studi	1.3×10^{-2}
46	risk	2.3×10^{-2}	96	cohort	1.2×10^{-2}
47	repair	2.3×10^{-2}	97	institut	1.2×10^{-2}
48	mean	2.2×10^{-2}	98	diagnosi	1.2×10^{-2}
49	periop	2.1×10^{-2}	99	histolog	1.2×10^{-2}
50	day	2.1×10^{-2}	100	left	1.2×10^{-2}

Table D.238. The list of the top 100 words in the category Telecommunications with RIGs

No.	Word	RIG
1	network	9.9×10^{-2}
2	propos	9.3×10^{-2}
3	paper	8.7×10^{-2}
4	wireless	8×10^{-2}
5	antenna	6.5×10^{-2}
6	communic	4.7×10^{-2}
7	simul	4.5×10^{-2}
8	user	4.3×10^{-2}
9	scheme	4.3×10^{-2}
10	algorithm	4.1×10^{-2}
11	channel	3.9×10^{-2}
12	radio	3.2×10^{-2}
13	bandwidth	3.1×10^{-2}
14	studi	3.1×10^{-2}
15	node	3×10^{-2}
16	packet	2.7×10^{-2}
17	conclus	2.6×10^{-2}
18	transmiss	2.5×10^{-2}
19	transmit	2.5×10^{-2}
20	traffic	2.4×10^{-2}
21	perform	2.2×10^{-2}
22	mobil	2.2×10^{-2}
23	interfer	2.1×10^{-2}
24	patient	2.1×10^{-2}
25	relay	2×10^{-2}
26	base	2×10^{-2}
27	servic	1.9×10^{-2}
28	mimo	1.9×10^{-2}
29	throughput	1.9×10^{-2}
30	ghz	1.8×10^{-2}
31	treatment	1.8×10^{-2}
32	signal	1.8×10^{-2}
33	frequenc	1.8×10^{-2}
34	effici	1.7×10^{-2}
35	system	1.7×10^{-2}
36	achiev	1.7×10^{-2}
37	suggest	1.7×10^{-2}
38	deploy	1.7×10^{-2}
39	fade	1.7×10^{-2}
40	power	1.6×10^{-2}
41	sensor	1.5×10^{-2}
42	problem	1.5×10^{-2}
43	bit	1.5×10^{-2}
44	access	1.5×10^{-2}
45	alloc	1.5×10^{-2}
46	lte	1.4×10^{-2}
47	associ	1.4×10^{-2}
48	scenario	1.4×10^{-2}
49	protocol	1.4×10^{-2}
50	nois	1.4×10^{-2}
51	found	1.4×10^{-2}
52	clinic	1.4×10^{-2}
53	protein	1.3×10^{-2}
54	age	1.3×10^{-2}
55	optim	1.3×10^{-2}
56	error	1.3×10^{-2}
57	qos	1.2×10^{-2}
58	slot	1.2×10^{-2}
59	secur	1.2×10^{-2}
60	diseas	1.2×10^{-2}
61	acid	1.1×10^{-2}
62	receiv	1.1×10^{-2}
63	gain	1.1×10^{-2}
64	station	1.1×10^{-2}
65	can	1.1×10^{-2}
66	internet	1.1×10^{-2}
67	decod	1.1×10^{-2}
68	gene	1.1×10^{-2}
69	concentr	1.1×10^{-2}
70	transmitt	1.1×10^{-2}
71	delay	1×10^{-2}
72	design	1×10^{-2}
73	divis	1×10^{-2}
74	implement	1×10^{-2}
75	band	1×10^{-2}
76	temperatur	1×10^{-2}
77	speci	1×10^{-2}
78	dure	1×10^{-2}
79	overhead	1×10^{-2}
80	multiplex	9.9×10^{-3}
81	radar	9.9×10^{-3}
82	background	9.8×10^{-3}
83	architectur	9.5×10^{-3}
84	observ	9.5×10^{-3}
85	reaction	9.4×10^{-3}
86	group	9.4×10^{-3}
87	examin	9.3×10^{-3}
88	orthogon	9.3×10^{-3}
89	activ	9.3×10^{-3}
90	assess	9.3×10^{-3}
91	induc	9.2×10^{-3}
92	multipl	9.1×10^{-3}
93	indic	9.1×10^{-3}
94	resourc	9×10^{-3}
95	comput	8.9×10^{-3}
96	applic	8.8×10^{-3}
97	outperform	8.6×10^{-3}
98	report	8.6×10^{-3}
99	maxim	8.3×10^{-3}
100	code	8.2×10^{-3}

Table D.239. The list of the top 100 words in the category Theater with RIGs

No.	Word	RIG	No.	Word	RIG
1	theatr	3.6×10^{-1}	51	lectur	1.8×10^{-2}
2	theatric	1.3×10^{-1}	52	engag	1.8×10^{-2}
3	drama	1.2×10^{-1}	53	compani	1.8×10^{-2}
4	articl	1×10^{-1}	54	queer	1.7×10^{-2}
5	audienc	8.9×10^{-2}	55	effect	1.7×10^{-2}
6	artist	8.5×10^{-2}	56	book	1.7×10^{-2}
7	contemporari	5.9×10^{-2}	57	measur	1.7×10^{-2}
8	spectat	5.8×10^{-2}	58	enact	1.6×10^{-2}
9	shakespear	5.3×10^{-2}	59	critiqu	1.6×10^{-2}
10	essay	5.1×10^{-2}	60	write	1.6×10^{-2}
11	playwright	5×10^{-2}	61	social	1.6×10^{-2}
12	play	4.9×10^{-2}	62	text	1.6×10^{-2}
13	result	4.8×10^{-2}	63	high	1.5×10^{-2}
14	argu	4.7×10^{-2}	64	celebr	1.5×10^{-2}
15	danc	4.2×10^{-2}	65	choreograph	1.5×10^{-2}
16	cultur	4.2×10^{-2}	66	obtain	1.5×10^{-2}
17	aesthet	4.1×10^{-2}	67	data	1.4×10^{-2}
18	professor	4×10^{-2}	68	act	1.4×10^{-2}
19	polit	4×10^{-2}	69	project	1.4×10^{-2}
20	perform	3.9×10^{-2}	70	nation	1.4×10^{-2}
21	art	3.8×10^{-2}	71	increas	1.4×10^{-2}
22	music	3.6×10^{-2}	72	show	1.4×10^{-2}
23	royal	3.3×10^{-2}	73	realiti	1.4×10^{-2}
24	embodi	3.1×10^{-2}	74	piec	1.4×10^{-2}
25	actor	2.9×10^{-2}	75	satir	1.3×10^{-2}
26	beckett	2.9×10^{-2}	76	evalu	1.3×10^{-2}
27	univers	2.9×10^{-2}	77	poetic	1.3×10^{-2}
28	work	2.8×10^{-2}	78	theater	1.3×10^{-2}
29	stage	2.8×10^{-2}	79	charact	1.3×10^{-2}
30	british	2.7×10^{-2}	80	patient	1.3×10^{-2}
31	way	2.7×10^{-2}	81	popular	1.3×10^{-2}
32	entertain	2.7×10^{-2}	82	examin	1.3×10^{-2}
33	world	2.6×10^{-2}	83	compar	1.3×10^{-2}
34	draw	2.6×10^{-2}	84	conclus	1.3×10^{-2}
35	narrat	2.6×10^{-2}	85	model	1.3×10^{-2}
36	festiv	2.5×10^{-2}	86	test	1.3×10^{-2}
37	director	2.4×10^{-2}	87	cell	1.3×10^{-2}
38	method	2.4×10^{-2}	88	practic	1.3×10^{-2}
39	london	2.4×10^{-2}	89	fiction	1.3×10^{-2}
40	explor	2.2×10^{-2}	90	improv	1.3×10^{-2}
41	histor	2.2×10^{-2}	91	dancer	1.2×10^{-2}
42	modern	2.1×10^{-2}	92	style	1.2×10^{-2}
43	war	2.1×10^{-2}	93	system	1.2×10^{-2}
44	samuel	2.1×10^{-2}	94	choreographi	1.2×10^{-2}
45	centuri	2.1×10^{-2}	95	rate	1.2×10^{-2}
46	histori	2×10^{-2}	96	tragedi	1.2×10^{-2}
47	notion	2×10^{-2}	97	represent	1.2×10^{-2}
48	use	1.9×10^{-2}	98	read	1.1×10^{-2}
49	genr	1.9×10^{-2}	99	thing	1.1×10^{-2}
50	product	1.8×10^{-2}	100	seek	1.1×10^{-2}

TABLE D.240. The list of the top 100 words in the category Thermodynamics with RIGs

No.	Word	RIG	No.	Word	RIG
1	heat	1.6×10^{-1}	51	obtain	9.5×10^{-3}
2	temperatur	8.8×10^{-2}	52	protein	9.5×10^{-3}
3	thermal	7.3×10^{-2}	53	isotherm	9.5×10^{-3}
4	flow	6×10^{-2}	54	drop	9.5×10^{-3}
5	transfer	5.1×10^{-2}	55	condens	9.4×10^{-3}
6	cool	4.3×10^{-2}	56	hot	9×10^{-3}
7	fluid	4.1×10^{-2}	57	perform	8.9×10^{-3}
8	pressur	4×10^{-2}	58	oper	8.8×10^{-3}
9	experiment	3.7×10^{-2}	59	steadi	8.8×10^{-3}
10	convect	3.5×10^{-2}	60	pump	8.7×10^{-3}
11	refriger	2.9×10^{-2}	61	gene	8.6×10^{-3}
12	energi	2.6×10^{-2}	62	phase	8.6×10^{-3}
13	air	2.6×10^{-2}	63	exchang	8.6×10^{-3}
14	reynold	2.6×10^{-2}	64	volum	8.4×10^{-3}
15	liquid	2.6×10^{-2}	65	calorimetri	8.3×10^{-3}
16	combust	2.5×10^{-2}	66	boundari	8.1×10^{-3}
17	gas	2.5×10^{-2}	67	investig	8×10^{-3}
18	numer	2.4×10^{-2}	68	agreement	7.9×10^{-3}
19	equat	2.2×10^{-2}	69	jet	7.9×10^{-3}
20	inlet	2.2×10^{-2}	70	conduct	7.8×10^{-3}
21	thermodynam	2.1×10^{-2}	71	human	7.8×10^{-3}
22	patient	2×10^{-2}	72	group	7.7×10^{-3}
23	fuel	2×10^{-2}	73	diamet	7.6×10^{-3}
24	coeffici	2×10^{-2}	74	treatment	7.6×10^{-3}
25	turbul	1.8×10^{-2}	75	work	7.5×10^{-3}
26	conclus	1.8×10^{-2}	76	cylind	7.5×10^{-3}
27	veloc	1.8×10^{-2}	77	unsteadi	7.4×10^{-3}
28	wall	1.8×10^{-2}	78	power	7.3×10^{-3}
29	enthalpi	1.7×10^{-2}	79	co2	7.2×10^{-3}
30	tube	1.6×10^{-2}	80	associ	7.2×10^{-3}
31	laminar	1.6×10^{-2}	81	steam	7×10^{-3}
32	flux	1.5×10^{-2}	82	suggest	7×10^{-3}
33	cfd	1.5×10^{-2}	83	viscos	7×10^{-3}
34	mixtur	1.5×10^{-2}	84	risk	6.9×10^{-3}
35	vapor	1.4×10^{-2}	85	predict	6.7×10^{-3}
36	evapor	1.4×10^{-2}	86	system	6.7×10^{-3}
37	paramet	1.4×10^{-2}	87	rang	6.6×10^{-3}
38	water	1.3×10^{-2}	88	molar	6.5×10^{-3}
39	model	1.3×10^{-2}	89	diesel	6.5×10^{-3}
40	turbin	1.3×10^{-2}	90	may	6.4×10^{-3}
41	flame	1.3×10^{-2}	91	engin	6.4×10^{-3}
42	clinic	1.2×10^{-2}	92	solid	6.4×10^{-3}
43	simul	1.2×10^{-2}	93	popul	6.3×10^{-3}
44	diseas	1.1×10^{-2}	94	detect	6.3×10^{-3}
45	background	1.1×10^{-2}	95	constant	6.2×10^{-3}
46	condit	1×10^{-2}	96	coolant	6.2×10^{-3}
47	pipe	9.9×10^{-3}	97	dsc	6.1×10^{-3}
48	calcul	9.8×10^{-3}	98	cancer	6.1×10^{-3}
49	mass	9.6×10^{-3}	99	binari	6.1×10^{-3}
50	age	9.6×10^{-3}	100	equilibrium	6×10^{-3}

TABLE D.241. The list of the top 100 words in the category Toxicology with RIGs

No.	Word	RIG	No.	Word	RIG
1	exposur	1.1×10^{-1}	51	environment	1.2×10^{-2}
2	toxic	9.9×10^{-2}	52	toxin	1.2×10^{-2}
3	concentr	5.3×10^{-2}	53	superoxid	1.1×10^{-2}
4	expos	5×10^{-2}	54	biomark	1.1×10^{-2}
5	dose	5×10^{-2}	55	propos	1.1×10^{-2}
6	induc	4.6×10^{-2}	56	oral	1.1×10^{-2}
7	rat	3.9×10^{-2}	57	treat	1.1×10^{-2}
8	toxicolog	2.8×10^{-2}	58	stress	1.1×10^{-2}
9	liver	2.8×10^{-2}	59	pathway	1.1×10^{-2}
10	cell	2.7×10^{-2}	60	assess	1.1×10^{-2}
11	level	2.7×10^{-2}	61	speci	1.1×10^{-2}
12	contamin	2.5×10^{-2}	62	blood	1.1×10^{-2}
13	human	2.5×10^{-2}	63	caus	1.1×10^{-2}
14	paper	2.4×10^{-2}	64	dismutas	1.1×10^{-2}
15	glutathion	2.3×10^{-2}	65	peroxid	1.1×10^{-2}
16	activ	2.3×10^{-2}	66	catalas	1.1×10^{-2}
17	effect	2.2×10^{-2}	67	food	1.1×10^{-2}
18	assay	2.2×10^{-2}	68	tissu	1×10^{-2}
19	damag	2.1×10^{-2}	69	accumul	1×10^{-2}
20	studi	2.1×10^{-2}	70	histopatholog	1×10^{-2}
21	day	2×10^{-2}	71	administr	1×10^{-2}
22	antioxid	2×10^{-2}	72	chemic	9.7×10^{-3}
23	inhibit	1.9×10^{-2}	73	lipid	9.6×10^{-3}
24	anim	1.8×10^{-2}	74	vivo	9.5×10^{-3}
25	increas	1.7×10^{-2}	75	caspas	9.2×10^{-3}
26	oxid	1.7×10^{-2}	76	indic	9.1×10^{-3}
27	dna	1.6×10^{-2}	77	male	9.1×10^{-3}
28	vitro	1.6×10^{-2}	78	acut	9.1×10^{-3}
29	metabolit	1.6×10^{-2}	79	reactiv	9×10^{-3}
30	decreas	1.6×10^{-2}	80	serum	8.9×10^{-3}
31	enzym	1.6×10^{-2}	81	respons	8.9×10^{-3}
32	protein	1.5×10^{-2}	82	observ	8.8×10^{-3}
33	cytotox	1.5×10^{-2}	83	kidney	8.8×10^{-3}
34	signific	1.5×10^{-2}	84	inhal	8.8×10^{-3}
35	apoptosi	1.4×10^{-2}	85	endocrin	8.8×10^{-3}
36	aquat	1.4×10^{-2}	86	may	8.7×10^{-3}
37	metabol	1.4×10^{-2}	87	cytochrom	8.5×10^{-3}
38	express	1.4×10^{-2}	88	inhibitor	8.2×10^{-3}
39	advers	1.4×10^{-2}	89	administ	8.2×10^{-3}
40	ros	1.4×10^{-2}	90	transferas	8.2×10^{-3}
41	pollut	1.3×10^{-2}	91	induct	8.1×10^{-3}
42	alter	1.3×10^{-2}	92	organ	8.1×10^{-3}
43	mice	1.3×10^{-2}	93	urin	7.9×10^{-3}
44	pesticid	1.3×10^{-2}	94	mrna	7.8×10^{-3}
45	fish	1.3×10^{-2}	95	risk	7.8×10^{-3}
46	drug	1.3×10^{-2}	96	sod	7.7×10^{-3}
47	treatment	1.2×10^{-2}	97	gene	7.7×10^{-3}
48	potenti	1.2×10^{-2}	98	chromatographi	7.5×10^{-3}
49	suggest	1.2×10^{-2}	99	substanc	7.5×10^{-3}
50	compound	1.2×10^{-2}	100	apoptot	7.4×10^{-3}

Table D.242. The list of the top 100 words in the category Transplantation with RIGs

No.	Word	RIG
1	transplant	4.4×10^{-1}
2	recipi	2×10^{-1}
3	graft	1.3×10^{-1}
4	donor	1.2×10^{-1}
5	patient	1.2×10^{-1}
6	kidney	1×10^{-1}
7	renal	7.3×10^{-2}
8	allograft	6.6×10^{-2}
9	surviv	6.5×10^{-2}
10	liver	6.1×10^{-2}
11	immunosuppress	6×10^{-2}
12	reject	6×10^{-2}
13	posttranspl	5.8×10^{-2}
14	conclus	4.5×10^{-2}
15	allogen	4.5×10^{-2}
16	hematopoiet	4.2×10^{-2}
17	acut	4×10^{-2}
18	outcom	3.5×10^{-2}
19	year	3.2×10^{-2}
20	retrospect	3.2×10^{-2}
21	therapi	3.2×10^{-2}
22	risk	3.2×10^{-2}
23	mortal	3.2×10^{-2}
24	diseas	3.1×10^{-2}
25	stem	3.1×10^{-2}
26	month	2.9×10^{-2}
27	gvhd	2.9×10^{-2}
28	background	2.9×10^{-2}
29	blood	2.9×10^{-2}
30	paper	2.9×10^{-2}
31	deceas	2.9×10^{-2}
32	receiv	2.9×10^{-2}
33	underw	2.8×10^{-2}
34	complic	2.7×10^{-2}
35	median	2.6×10^{-2}
36	cell	2.5×10^{-2}
37	day	2.4×10^{-2}
38	creatinin	2.4×10^{-2}
39	clinic	2.3×10^{-2}
40	donat	2.3×10^{-2}
41	chronic	2.2×10^{-2}
42	hla	2.2×10^{-2}
43	death	2.2×10^{-2}
44	biopsi	2.1×10^{-2}
45	heart	2×10^{-2}
46	incid	1.9×10^{-2}
47	failur	1.9×10^{-2}
48	associ	1.9×10^{-2}
49	dialysi	1.9×10^{-2}
50	hepat	1.8×10^{-2}

No.	Word	RIG
51	regimen	1.8×10^{-2}
52	signific	1.8×10^{-2}
53	versus	1.8×10^{-2}
54	glomerular	1.7×10^{-2}
55	multivari	1.7×10^{-2}
56	follow	1.7×10^{-2}
57	engraft	1.6×10^{-2}
58	ventricular	1.6×10^{-2}
59	infect	1.6×10^{-2}
60	serum	1.5×10^{-2}
61	age	1.5×10^{-2}
62	befor	1.5×10^{-2}
63	pediatr	1.5×10^{-2}
64	group	1.5×10^{-2}
65	antibodi	1.4×10^{-2}
66	marrow	1.4×10^{-2}
67	organ	1.3×10^{-2}
68	cardiac	1.3×10^{-2}
69	treatment	1.3×10^{-2}
70	live	1.3×10^{-2}
71	propos	1.3×10^{-2}
72	cohort	1.3×10^{-2}
73	postop	1.2×10^{-2}
74	factor	1.2×10^{-2}
75	structur	1.2×10^{-2}
76	ischemia	1.2×10^{-2}
77	earli	1.1×10^{-2}
78	autolog	1.1×10^{-2}
79	treat	1.1×10^{-2}
80	filtrat	1.1×10^{-2}
81	relaps	1.1×10^{-2}
82	prophylaxi	1.1×10^{-2}
83	center	1.1×10^{-2}
84	undergo	1.1×10^{-2}
85	adult	1.1×10^{-2}
86	end	1.1×10^{-2}
87	dysfunct	1.1×10^{-2}
88	may	1×10^{-2}
89	rate	1×10^{-2}
90	left	1×10^{-2}
91	dose	1×10^{-2}
92	injuri	1×10^{-2}
93	remain	9.8×10^{-3}
94	lung	9.7×10^{-3}
95	temperatur	9.6×10^{-3}
96	die	9.6×10^{-3}
97	reperfus	9.6×10^{-3}
98	morbid	9.3×10^{-3}
99	arteri	9.2×10^{-3}
100	order	9.1×10^{-3}

TABLE D.243. The list of the top 100 words in the category Transportation with RIGs

No.	Word	RIG
1	traffic	1.1×10^{-1}
2	transport	9.3×10^{-2}
3	travel	9.1×10^{-2}
4	road	8.8×10^{-2}
5	vehicl	7.8×10^{-2}
6	driver	6×10^{-2}
7	passeng	5×10^{-2}
8	crash	4.7×10^{-2}
9	car	4.1×10^{-2}
10	citi	3.6×10^{-2}
11	trip	3.5×10^{-2}
12	urban	3.4×10^{-2}
13	lane	3.1×10^{-2}
14	safeti	3×10^{-2}
15	pedestrian	3×10^{-2}
16	polici	2.8×10^{-2}
17	congest	2.7×10^{-2}
18	highway	2.6×10^{-2}
19	drive	2.6×10^{-2}
20	rail	2.4×10^{-2}
21	plan	2.3×10^{-2}
22	rout	2.3×10^{-2}
23	paper	2.3×10^{-2}
24	infrastructur	2.2×10^{-2}
25	cost	2.1×10^{-2}
26	servic	2.1×10^{-2}
27	public	1.9×10^{-2}
28	bus	1.9×10^{-2}
29	choic	1.9×10^{-2}
30	model	1.9×10^{-2}
31	demand	1.8×10^{-2}
32	speed	1.8×10^{-2}
33	impact	1.7×10^{-2}
34	accid	1.7×10^{-2}
35	pavement	1.7×10^{-2}
36	research	1.6×10^{-2}
37	commut	1.6×10^{-2}
38	decis	1.5×10^{-2}
39	patient	1.4×10^{-2}
40	network	1.3×10^{-2}
41	price	1.3×10^{-2}
42	survey	1.3×10^{-2}
43	metropolitan	1.3×10^{-2}
44	fatal	1.2×10^{-2}
45	destin	1.2×10^{-2}
46	data	1.2×10^{-2}
47	estim	1.2×10^{-2}
48	cell	1.1×10^{-2}
49	user	1.1×10^{-2}
50	intersect	1.1×10^{-2}

No.	Word	RIG
51	oper	9.9×10^{-3}
52	econom	9.7×10^{-3}
53	protein	9.6×10^{-3}
54	agenc	9.5×10^{-3}
55	methodolog	9.4×10^{-3}
56	benefit	9.4×10^{-3}
57	clinic	9.2×10^{-3}
58	invest	9.2×10^{-3}
59	street	9.1×10^{-3}
60	develop	9.1×10^{-3}
61	market	9×10^{-3}
62	maker	8.8×10^{-3}
63	scenario	8.8×10^{-3}
64	ship	8.6×10^{-3}
65	asphalt	8.4×10^{-3}
66	railway	8.2×10^{-3}
67	privat	8.2×10^{-3}
68	planner	8.1×10^{-3}
69	diseas	7.8×10^{-3}
70	locat	7.6×10^{-3}
71	treatment	7.6×10^{-3}
72	gene	7.6×10^{-3}
73	empir	7.4×10^{-3}
74	household	7.3×10^{-3}
75	acid	7.3×10^{-3}
76	implement	7.2×10^{-3}
77	injuri	6.8×10^{-3}
78	walk	6.6×10^{-3}
79	speci	6.6×10^{-3}
80	variabl	6.5×10^{-3}
81	transit	6.5×10^{-3}
82	sustain	6.4×10^{-3}
83	time	6.3×10^{-3}
84	ownership	6.2×10^{-3}
85	area	6.2×10^{-3}
86	project	6.2×10^{-3}
87	distanc	6.2×10^{-3}
88	park	6.1×10^{-3}
89	electron	6.1×10^{-3}
90	temperatur	6.1×10^{-3}
91	logist	6.1×10^{-3}
92	compani	6.1×10^{-3}
93	occup	6×10^{-3}
94	molecular	6×10^{-3}
95	behavior	5.9×10^{-3}
96	access	5.9×10^{-3}
97	mobil	5.9×10^{-3}
98	need	5.8×10^{-3}
99	focus	5.7×10^{-3}
100	revenu	5.7×10^{-3}

TABLE D.244. The list of the top 100 words in the category Transportation Science and Technology with RIGs

No.	Word	RIG
1	vehicl	1.3×10^{-1}
2	traffic	8.4×10^{-2}
3	paper	6.4×10^{-2}
4	road	5.8×10^{-2}
5	travel	4×10^{-2}
6	transport	3.3×10^{-2}
7	propos	2.9×10^{-2}
8	car	2.9×10^{-2}
9	driver	2.7×10^{-2}
10	passeng	2.7×10^{-2}
11	network	2.7×10^{-2}
12	simul	2.7×10^{-2}
13	drive	2.5×10^{-2}
14	lane	2.3×10^{-2}
15	rail	2.3×10^{-2}
16	speed	2×10^{-2}
17	patient	1.9×10^{-2}
18	system	1.9×10^{-2}
19	highway	1.9×10^{-2}
20	congest	1.8×10^{-2}
21	railway	1.7×10^{-2}
22	urban	1.7×10^{-2}
23	model	1.6×10^{-2}
24	pavement	1.5×10^{-2}
25	pedestrian	1.5×10^{-2}
26	trip	1.5×10^{-2}
27	infrastructur	1.5×10^{-2}
28	safeti	1.4×10^{-2}
29	algorithm	1.4×10^{-2}
30	user	1.3×10^{-2}
31	scenario	1.3×10^{-2}
32	conclus	1.2×10^{-2}
33	rout	1.2×10^{-2}
34	base	1.2×10^{-2}
35	bus	1.2×10^{-2}
36	clinic	1.1×10^{-2}
37	citi	1.1×10^{-2}
38	oper	1.1×10^{-2}
39	optim	1.1×10^{-2}
40	treatment	1.1×10^{-2}
41	protein	1.1×10^{-2}
42	real	1.1×10^{-2}
43	cost	1.1×10^{-2}
44	electr	1.1×10^{-2}
45	diseas	1.1×10^{-2}
46	wheel	1×10^{-2}
47	servic	1×10^{-2}
48	batteri	9.7×10^{-3}
49	demand	9.6×10^{-3}
50	plan	9.1×10^{-3}

No.	Word	RIG
51	scheme	9×10^{-3}
52	asphalt	9×10^{-3}
53	intersect	8.9×10^{-3}
54	crash	8.8×10^{-3}
55	perform	8.6×10^{-3}
56	fuel	8.6×10^{-3}
57	station	8.5×10^{-3}
58	problem	8.5×10^{-3}
59	gene	8.4×10^{-3}
60	destin	8.1×10^{-3}
61	acid	7.8×10^{-3}
62	power	7.6×10^{-3}
63	effici	7.4×10^{-3}
64	background	7.4×10^{-3}
65	design	7.4×10^{-3}
66	speci	7.3×10^{-3}
67	wireless	7.1×10^{-3}
68	estim	6.9×10^{-3}
69	molecular	6.7×10^{-3}
70	induc	6.7×10^{-3}
71	communic	6.6×10^{-3}
72	can	6.4×10^{-3}
73	concentr	6.2×10^{-3}
74	relay	6.1×10^{-3}
75	decis	6×10^{-3}
76	cancer	6×10^{-3}
77	accid	5.9×10^{-3}
78	group	5.9×10^{-3}
79	motor	5.9×10^{-3}
80	associ	5.9×10^{-3}
81	track	5.9×10^{-3}
82	tissu	5.9×10^{-3}
83	time	5.7×10^{-3}
84	intellig	5.5×10^{-3}
85	implement	5.4×10^{-3}
86	mobil	5.4×10^{-3}
87	consid	5.4×10^{-3}
88	report	5.3×10^{-3}
89	torqu	5.2×10^{-3}
90	improv	5.2×10^{-3}
91	automot	5.2×10^{-3}
92	therapi	5.1×10^{-3}
93	schedul	5.1×10^{-3}
94	suggest	5×10^{-3}
95	engin	4.9×10^{-3}
96	inhibit	4.9×10^{-3}
97	age	4.9×10^{-3}
98	strategi	4.8×10^{-3}
99	capac	4.8×10^{-3}
100	drug	4.8×10^{-3}

Table D.245. The list of the top 100 words in the category Tropical Medicine with RIGs

No.	Word	RIG
1	infect	1.3×10^{-1}
2	malaria	1.2×10^{-1}
3	parasit	9.4×10^{-2}
4	conclus	7.7×10^{-2}
5	mosquito	6.8×10^{-2}
6	background	6.8×10^{-2}
7	plasmodium	6.5×10^{-2}
8	endem	5.6×10^{-2}
9	falciparum	5×10^{-2}
10	diseas	3.8×10^{-2}
11	preval	3.6×10^{-2}
12	vector	3.5×10^{-2}
13	pcr	3.5×10^{-2}
14	anophel	3.2×10^{-2}
15	leishmaniasi	3.1×10^{-2}
16	epidemiolog	3×10^{-2}
17	fever	2.9×10^{-2}
18	dengu	2.9×10^{-2}
19	aed	2.8×10^{-2}
20	leishmania	2.6×10^{-2}
21	insecticid	2.5×10^{-2}
22	health	2.5×10^{-2}
23	transmiss	2.5×10^{-2}
24	blood	2.4×10^{-2}
25	popul	2.1×10^{-2}
26	human	2.1×10^{-2}
27	host	2.1×10^{-2}
28	speci	2×10^{-2}
29	virus	2×10^{-2}
30	collect	2×10^{-2}
31	method	2×10^{-2}
32	princip	1.9×10^{-2}
33	paper	1.8×10^{-2}
34	africa	1.8×10^{-2}
35	brazil	1.8×10^{-2}
36	tick	1.7×10^{-2}
37	detect	1.7×10^{-2}
38	surveil	1.6×10^{-2}
39	pathogen	1.5×10^{-2}
40	countri	1.5×10^{-2}
41	area	1.5×10^{-2}
42	antibodi	1.5×10^{-2}
43	antigen	1.5×10^{-2}
44	assay	1.5×10^{-2}
45	outbreak	1.4×10^{-2}
46	district	1.4×10^{-2}
47	born	1.4×10^{-2}
48	serolog	1.4×10^{-2}
49	children	1.4×10^{-2}
50	villag	1.4×10^{-2}

No.	Word	RIG
51	zoonot	1.3×10^{-2}
52	larva	1.3×10^{-2}
53	vaccin	1.3×10^{-2}
54	clinic	1.3×10^{-2}
55	control	1.2×10^{-2}
56	spp	1.2×10^{-2}
57	caus	1.2×10^{-2}
58	propos	1.2×10^{-2}
59	dog	1.2×10^{-2}
60	among	1.2×10^{-2}
61	elisa	1.1×10^{-2}
62	isol	1.1×10^{-2}
63	drug	1.1×10^{-2}
64	infecti	1.1×10^{-2}
65	smear	1.1×10^{-2}
66	tropic	1.1×10^{-2}
67	sampl	1.1×10^{-2}
68	household	1.1×10^{-2}
69	studi	1×10^{-2}
70	geograph	1×10^{-2}
71	signific	1×10^{-2}
72	immun	1×10^{-2}
73	polymeras	1×10^{-2}
74	region	1×10^{-2}
75	rural	1×10^{-2}
76	year	1×10^{-2}
77	egg	1×10^{-2}
78	coinfect	9.9×10^{-3}
79	larval	9.9×10^{-3}
80	diagnosi	9.7×10^{-3}
81	identifi	9.6×10^{-3}
82	mortal	9.5×10^{-3}
83	femal	9.5×10^{-3}
84	adult	9.3×10^{-3}
85	dna	9.2×10^{-3}
86	transmit	9.2×10^{-3}
87	risk	9.1×10^{-3}
88	treatment	9×10^{-3}
89	simul	8.9×10^{-3}
90	viscer	8.9×10^{-3}
91	posit	8.9×10^{-3}
92	subsaharan	8.8×10^{-3}
93	provinc	8.7×10^{-3}
94	case	8.6×10^{-3}
95	insect	8.5×10^{-3}
96	programm	8.5×10^{-3}
97	energi	8.3×10^{-3}
98	associ	8.2×10^{-3}
99	burden	8.2×10^{-3}
100	sequenc	8.1×10^{-3}

Table D.246. The list of the top 100 words in the category Urban Studies with RIGs

No.	Word	RIG	No.	Word	RIG
1	urban	2×10^{-1}	51	place	1.1×10^{-2}
2	citi	1.5×10^{-1}	52	infrastructur	1.1×10^{-2}
3	hous	7.2×10^{-2}	53	district	1.1×10^{-2}
4	polici	6.1×10^{-2}	54	settlement	1.1×10^{-2}
5	plan	4.4×10^{-2}	55	examin	1.1×10^{-2}
6	econom	4.2×10^{-2}	56	park	1.1×10^{-2}
7	govern	3.8×10^{-2}	57	draw	1×10^{-2}
8	land	3.8×10^{-2}	58	rural	1×10^{-2}
9	social	3.5×10^{-2}	59	project	9.6×10^{-3}
10	metropolitan	3.4×10^{-2}	60	conclus	9.6×10^{-3}
11	polit	3.1×10^{-2}	61	clinic	9.5×10^{-3}
12	articl	2.8×10^{-2}	62	town	9.5×10^{-3}
13	residenti	2.7×10^{-2}	63	interview	9.1×10^{-3}
14	area	2.7×10^{-2}	64	paper	9×10^{-3}
15	market	2.6×10^{-2}	65	find	8.8×10^{-3}
16	neighborhood	2.6×10^{-2}	66	obtain	8.7×10^{-3}
17	public	2.6×10^{-2}	67	impact	8.6×10^{-3}
18	local	2.5×10^{-2}	68	histor	8.6×10^{-3}
19	spatial	2.5×10^{-2}	69	protein	8.5×10^{-3}
20	landscap	2.4×10^{-2}	70	innov	8.4×10^{-3}
21	communiti	2.3×10^{-2}	71	perform	8.3×10^{-3}
22	resid	2.3×10^{-2}	72	invest	8.2×10^{-3}
23	capit	2.1×10^{-2}	73	home	8.1×10^{-3}
24	planner	2×10^{-2}	74	nation	8.1×10^{-3}
25	household	1.9×10^{-2}	75	green	8×10^{-3}
26	neighbourhood	1.8×10^{-2}	76	region	7.9×10^{-3}
27	estat	1.8×10^{-2}	77	census	7.9×10^{-3}
28	argu	1.8×10^{-2}	78	ecolog	7.8×10^{-3}
29	sustain	1.7×10^{-2}	79	countri	7.8×10^{-3}
30	research	1.7×10^{-2}	80	treatment	7.7×10^{-3}
31	develop	1.7×10^{-2}	81	servic	7.6×10^{-3}
32	territori	1.6×10^{-2}	82	electron	7.6×10^{-3}
33	incom	1.6×10^{-2}	83	geographi	7.5×10^{-3}
34	municip	1.5×10^{-2}	84	peopl	7.4×10^{-3}
35	economi	1.5×10^{-2}	85	decad	7.3×10^{-3}
36	patient	1.4×10^{-2}	86	poverti	7.3×10^{-3}
37	method	1.4×10^{-2}	87	experiment	7.3×10^{-3}
38	cell	1.4×10^{-2}	88	live	7.2×10^{-3}
39	privat	1.4×10^{-2}	89	acid	7.2×10^{-3}
40	build	1.3×10^{-2}	90	strateg	7.2×10^{-3}
41	price	1.3×10^{-2}	91	creat	7.1×10^{-3}
42	focus	1.3×10^{-2}	92	china	7.1×10^{-3}
43	space	1.3×10^{-2}	93	environment	7.1×10^{-3}
44	empir	1.3×10^{-2}	94	locat	7.1×10^{-3}
45	sector	1.2×10^{-2}	95	perspect	7.1×10^{-3}
46	institut	1.2×10^{-2}	96	citizen	7.1×10^{-3}
47	socio	1.2×10^{-2}	97	practic	7×10^{-3}
48	actor	1.2×10^{-2}	98	neoliber	6.9×10^{-3}
49	context	1.2×10^{-2}	99	cultur	6.9×10^{-3}
50	street	1.2×10^{-2}	100	european	6.9×10^{-3}

Table D.247. The list of the top 100 words in the category Urology and Nephrology with RIGs

No.	Word	RIG
1	patient	1.5×10^{-1}
2	renal	1.5×10^{-1}
3	kidney	1.2×10^{-1}
4	conclus	9.9×10^{-2}
5	urinari	8.7×10^{-2}
6	prostat	7.4×10^{-2}
7	bladder	6.6×10^{-2}
8	dialysi	5.9×10^{-2}
9	outcom	5.3×10^{-2}
10	glomerular	4.9×10^{-2}
11	hemodialysi	4.7×10^{-2}
12	diseas	4.5×10^{-2}
13	ckd	4.4×10^{-2}
14	underw	4×10^{-2}
15	prostatectomi	3.9×10^{-2}
16	year	3.8×10^{-2}
17	month	3.8×10^{-2}
18	age	3.6×10^{-2}
19	clinic	3.5×10^{-2}
20	men	3.4×10^{-2}
21	median	3.4×10^{-2}
22	associ	3.3×10^{-2}
23	risk	3.2×10^{-2}
24	creatinin	3.2×10^{-2}
25	treatment	3.1×10^{-2}
26	retrospect	3×10^{-2}
27	paper	3×10^{-2}
28	incontin	2.9×10^{-2}
29	urin	2.8×10^{-2}
30	tract	2.7×10^{-2}
31	cancer	2.7×10^{-2}
32	serum	2.7×10^{-2}
33	signific	2.6×10^{-2}
34	method	2.6×10^{-2}
35	surgeri	2.6×10^{-2}
36	complic	2.6×10^{-2}
37	cohort	2.5×10^{-2}
38	chronic	2.5×10^{-2}
39	multivari	2.5×10^{-2}
40	radic	2.5×10^{-2}
41	postop	2.5×10^{-2}
42	surgic	2.5×10^{-2}
43	score	2.4×10^{-2}
44	follow	2.4×10^{-2}
45	therapi	2.4×10^{-2}
46	object	2.3×10^{-2}
47	filtrat	2.3×10^{-2}
48	biopsi	2.2×10^{-2}
49	erectil	2.1×10^{-2}
50	prospect	2.1×10^{-2}
51	treat	2×10^{-2}
52	nephropathi	2×10^{-2}
53	dysfunct	2×10^{-2}
54	group	1.8×10^{-2}
55	stone	1.8×10^{-2}
56	regress	1.8×10^{-2}
57	recurr	1.8×10^{-2}
58	pelvic	1.7×10^{-2}
59	rate	1.7×10^{-2}
60	preoper	1.7×10^{-2}
61	propos	1.6×10^{-2}
62	stage	1.6×10^{-2}
63	result	1.6×10^{-2}
64	review	1.6×10^{-2}
65	assess	1.6×10^{-2}
66	cox	1.6×10^{-2}
67	egfr	1.6×10^{-2}
68	injuri	1.5×10^{-2}
69	surviv	1.5×10^{-2}
70	hospit	1.5×10^{-2}
71	introduct	1.5×10^{-2}
72	undergo	1.5×10^{-2}
73	patholog	1.5×10^{-2}
74	background	1.5×10^{-2}
75	includ	1.5×10^{-2}
76	symptom	1.5×10^{-2}
77	predictor	1.5×10^{-2}
78	baselin	1.5×10^{-2}
79	void	1.5×10^{-2}
80	laparoscop	1.5×10^{-2}
81	mortal	1.4×10^{-2}
82	diabet	1.4×10^{-2}
83	tubular	1.4×10^{-2}
84	evalu	1.4×10^{-2}
85	blood	1.4×10^{-2}
86	total	1.3×10^{-2}
87	mean	1.3×10^{-2}
88	periton	1.3×10^{-2}
89	receiv	1.2×10^{-2}
90	hazard	1.2×10^{-2}
91	structur	1.2×10^{-2}
92	carcinoma	1.1×10^{-2}
93	periop	1.1×10^{-2}
94	transplant	1.1×10^{-2}
95	temperatur	1.1×10^{-2}
96	simul	1.1×10^{-2}
97	grade	1.1×10^{-2}
98	progress	1.1×10^{-2}
99	confid	1.1×10^{-2}
100	male	1.1×10^{-2}

Table D.248. The list of the top 100 words in the category Veterinary Sciences with RIGs

No.	Word	RIG
1	dog	1×10^{-1}
2	anim	9.5×10^{-2}
3	infect	4.8×10^{-2}
4	hors	4.7×10^{-2}
5	veterinari	4.4×10^{-2}
6	canin	3.8×10^{-2}
7	cattl	3.6×10^{-2}
8	breed	3×10^{-2}
9	farm	2.7×10^{-2}
10	cat	2.7×10^{-2}
11	paper	2.6×10^{-2}
12	day	2.5×10^{-2}
13	herd	2.4×10^{-2}
14	cow	2.4×10^{-2}
15	clinic	2.3×10^{-2}
16	diseas	2.3×10^{-2}
17	pcr	2.1×10^{-2}
18	virus	2.1×10^{-2}
19	pig	2.1×10^{-2}
20	pathogen	2×10^{-2}
21	sheep	2×10^{-2}
22	blood	2×10^{-2}
23	sign	1.9×10^{-2}
24	dairi	1.9×10^{-2}
25	calv	1.9×10^{-2}
26	propos	1.8×10^{-2}
27	bovin	1.6×10^{-2}
28	sampl	1.6×10^{-2}
29	collect	1.6×10^{-2}
30	studi	1.5×10^{-2}
31	antibodi	1.4×10^{-2}
32	serum	1.4×10^{-2}
33	immun	1.4×10^{-2}
34	vaccin	1.4×10^{-2}
35	isol	1.3×10^{-2}
36	old	1.3×10^{-2}
37	speci	1.3×10^{-2}
38	chicken	1.3×10^{-2}
39	goat	1.3×10^{-2}
40	rumin	1.3×10^{-2}
41	parasit	1.2×10^{-2}
42	feed	1.2×10^{-2}
43	histopatholog	1.2×10^{-2}
44	milk	1.2×10^{-2}
45	simul	1.1×10^{-2}
46	preval	1.1×10^{-2}
47	outbreak	1.1×10^{-2}
48	detect	1.1×10^{-2}
49	slaughter	1.1×10^{-2}
50	signific	1.1×10^{-2}

No.	Word	RIG
51	bird	1.1×10^{-2}
52	owner	1×10^{-2}
53	lesion	1×10^{-2}
54	livestock	1×10^{-2}
55	infecti	9.9×10^{-3}
56	group	9.9×10^{-3}
57	evalu	9.8×10^{-3}
58	diet	9.8×10^{-3}
59	spp	9.6×10^{-3}
60	domest	9.6×10^{-3}
61	healthi	9.5×10^{-3}
62	zoonot	9.5×10^{-3}
63	week	9.4×10^{-3}
64	fed	9.4×10^{-3}
65	egg	9.4×10^{-3}
66	tissu	9.4×10^{-3}
67	model	9.2×10^{-3}
68	intestin	9.1×10^{-3}
69	avian	8.9×10^{-3}
70	respiratori	8.6×10^{-3}
71	holstein	8.3×10^{-3}
72	process	8.3×10^{-3}
73	welfar	8.2×10^{-3}
74	structur	8.1×10^{-3}
75	treatment	8×10^{-3}
76	properti	8×10^{-3}
77	algorithm	8×10^{-3}
78	fish	7.9×10^{-3}
79	serolog	7.9×10^{-3}
80	caus	7.8×10^{-3}
81	broiler	7.7×10^{-3}
82	elisa	7.7×10^{-3}
83	femal	7.7×10^{-3}
84	histolog	7.6×10^{-3}
85	porcin	7.6×10^{-3}
86	lactat	7.5×10^{-3}
87	insemin	7.5×10^{-3}
88	object	7.3×10^{-3}
89	spleen	7.3×10^{-3}
90	strain	7.1×10^{-3}
91	semen	7.1×10^{-3}
92	viral	7×10^{-3}
93	supplement	7×10^{-3}
94	reproduct	7×10^{-3}
95	power	6.9×10^{-3}
96	gene	6.9×10^{-3}
97	virul	6.8×10^{-3}
98	concentr	6.7×10^{-3}
99	polymeras	6.7×10^{-3}
100	fecal	6.7×10^{-3}

Table D.249. The list of the top 100 words in the category Virology with RIGs

No.	Word	RIG
1	virus	3.6×10^{-1}
2	infect	2.6×10^{-1}
3	viral	2.3×10^{-1}
4	replic	8.7×10^{-2}
5	hiv	8.7×10^{-2}
6	host	6.6×10^{-2}
7	protein	6.5×10^{-2}
8	rna	6.3×10^{-2}
9	cell	6.1×10^{-2}
10	antivir	5.8×10^{-2}
11	genom	5.5×10^{-2}
12	immun	5.1×10^{-2}
13	vaccin	4.6×10^{-2}
14	pathogen	4.5×10^{-2}
15	influenza	4.2×10^{-2}
16	sequenc	4.1×10^{-2}
17	human	4×10^{-2}
18	gene	3.6×10^{-2}
19	strain	3.2×10^{-2}
20	cd4	3.2×10^{-2}
21	hepat	3.1×10^{-2}
22	antiretrovir	3.1×10^{-2}
23	interferon	3.1×10^{-2}
24	antibodi	2.9×10^{-2}
25	infecti	2.9×10^{-2}
26	paper	2.8×10^{-2}
27	express	2.7×10^{-2}
28	pcr	2.7×10^{-2}
29	antigen	2.6×10^{-2}
30	assay	2.6×10^{-2}
31	mutat	2.6×10^{-2}
32	hcv	2.4×10^{-2}
33	ifn	2.3×10^{-2}
34	diseas	2.3×10^{-2}
35	dna	2.2×10^{-2}
36	phylogenet	2.2×10^{-2}
37	recombin	2.2×10^{-2}
38	genotyp	2.1×10^{-2}
39	isol	2.1×10^{-2}
40	virul	2×10^{-2}
41	virolog	2×10^{-2}
42	mutant	2×10^{-2}
43	bind	1.9×10^{-2}
44	encod	1.9×10^{-2}
45	transcript	1.9×10^{-2}
46	cellular	1.8×10^{-2}
47	glycoprotein	1.8×10^{-2}
48	innat	1.7×10^{-2}
49	amino	1.7×10^{-2}
50	envelop	1.7×10^{-2}
51	entri	1.6×10^{-2}
52	nucleotid	1.6×10^{-2}
53	immunodefici	1.6×10^{-2}
54	titer	1.6×10^{-2}
55	inhibit	1.6×10^{-2}
56	mediat	1.6×10^{-2}
57	epitop	1.6×10^{-2}
58	hbv	1.6×10^{-2}
59	respiratori	1.6×10^{-2}
60	associ	1.5×10^{-2}
61	specif	1.5×10^{-2}
62	mice	1.5×10^{-2}
63	avian	1.4×10^{-2}
64	detect	1.4×10^{-2}
65	genet	1.4×10^{-2}
66	polymeras	1.3×10^{-2}
67	respons	1.3×10^{-2}
68	caus	1.3×10^{-2}
69	suggest	1.3×10^{-2}
70	target	1.3×10^{-2}
71	cd8	1.3×10^{-2}
72	pathogenesi	1.3×10^{-2}
73	coinfect	1.3×10^{-2}
74	wild	1.3×10^{-2}
75	identifi	1.2×10^{-2}
76	strand	1.2×10^{-2}
77	induc	1.2×10^{-2}
78	simul	1.2×10^{-2}
79	inhibitor	1.2×10^{-2}
80	vitro	1.1×10^{-2}
81	therapi	1.1×10^{-2}
82	outbreak	1.1×10^{-2}
83	fever	1.1×10^{-2}
84	subtyp	1.1×10^{-2}
85	receptor	1.1×10^{-2}
86	copi	1×10^{-2}
87	role	1×10^{-2}
88	import	1×10^{-2}
89	inocul	9.7×10^{-3}
90	variant	9.7×10^{-3}
91	neutral	9.7×10^{-3}
92	clone	9.2×10^{-3}
93	cytokin	9.1×10^{-3}
94	dengu	9.1×10^{-3}
95	epidem	8.9×10^{-3}
96	energi	8.9×10^{-3}
97	conserv	8.9×10^{-3}
98	revers	8.7×10^{-3}
99	delet	8.7×10^{-3}
100	circul	8.7×10^{-3}

TABLE D.250. The list of the top 100 words in the category Water Resources with RIGs

No.	Word	RIG	No.	Word	RIG
1	water	1.4×10^{-1}	51	eros	1.1×10^{-2}
2	hydrolog	6.9×10^{-2}	52	coastal	1.1×10^{-2}
3	river	6.8×10^{-2}	53	storm	1×10^{-2}
4	groundwat	6×10^{-2}	54	ecosystem	1×10^{-2}
5	soil	3.8×10^{-2}	55	estim	1×10^{-2}
6	rainfal	3.8×10^{-2}	56	impact	1×10^{-2}
7	runoff	3.8×10^{-2}	57	resourc	1×10^{-2}
8	flow	3.7×10^{-2}	58	region	9.8×10^{-3}
9	flood	3.5×10^{-2}	59	calibr	9.8×10^{-3}
10	catchment	3.4×10^{-2}	60	indic	9.4×10^{-3}
11	hydraul	3.4×10^{-2}	61	bed	9.3×10^{-3}
12	aquif	3.4×10^{-2}	62	condit	9.1×10^{-3}
13	basin	3.4×10^{-2}	63	slope	9.1×10^{-3}
14	wastewat	3.1×10^{-2}	64	period	9×10^{-3}
15	climat	2.7×10^{-2}	65	flux	9×10^{-3}
16	sediment	2.7×10^{-2}	66	suspend	8.9×10^{-3}
17	area	2.4×10^{-2}	67	sand	8.9×10^{-3}
18	irrig	2.2×10^{-2}	68	variabl	8.9×10^{-3}
19	watersh	2.2×10^{-2}	69	surfac	8.7×10^{-3}
20	patient	2.1×10^{-2}	70	locat	8.7×10^{-3}
21	remov	1.9×10^{-2}	71	suppli	8.7×10^{-3}
22	precipit	1.9×10^{-2}	72	subsurfac	8.5×10^{-3}
23	land	1.9×10^{-2}	73	station	8.3×10^{-3}
24	discharg	1.9×10^{-2}	74	diseas	8.3×10^{-3}
25	concentr	1.8×10^{-2}	75	drought	8.3×10^{-3}
26	season	1.8×10^{-2}	76	shallow	8.2×10^{-3}
27	manag	1.8×10^{-2}	77	distribut	8.1×10^{-3}
28	pollut	1.7×10^{-2}	78	veget	8.1×10^{-3}
29	model	1.7×10^{-2}	79	weather	7.8×10^{-3}
30	conclus	1.6×10^{-2}	80	background	7.8×10^{-3}
31	evapotranspir	1.6×10^{-2}	81	semiarid	7.8×10^{-3}
32	agricultur	1.6×10^{-2}	82	meteorolog	7.5×10^{-3}
33	contamin	1.5×10^{-2}	83	cell	7.5×10^{-3}
34	effluent	1.5×10^{-2}	84	batch	7.5×10^{-3}
35	spatial	1.4×10^{-2}	85	tempor	7.4×10^{-3}
36	annual	1.3×10^{-2}	86	downstream	7.4×10^{-3}
37	depth	1.3×10^{-2}	87	cod	7.3×10^{-3}
38	scale	1.3×10^{-2}	88	sourc	7.2×10^{-3}
39	drainag	1.3×10^{-2}	89	uncertainti	7.2×10^{-3}
40	clinic	1.3×10^{-2}	90	studi	7×10^{-3}
41	stream	1.2×10^{-2}	91	wetland	6.9×10^{-3}
42	reservoir	1.1×10^{-2}	92	plant	6.9×10^{-3}
43	dissolv	1.1×10^{-2}	93	rain	6.8×10^{-3}
44	dam	1.1×10^{-2}	94	use	6.8×10^{-3}
45	sludg	1.1×10^{-2}	95	transport	6.7×10^{-3}
46	environment	1.1×10^{-2}	96	upstream	6.7×10^{-3}
47	urban	1.1×10^{-2}	97	event	6.7×10^{-3}
48	lake	1.1×10^{-2}	98	adsorpt	6.6×10^{-3}
49	zone	1.1×10^{-2}	99	anthropogen	6.6×10^{-3}
50	drink	1.1×10^{-2}	100	demand	6.6×10^{-3}

Table D.251. The list of the top 100 words in the category Women's Studies with RIGs

No.	Word	RIG
1	women	2.4×10^{-1}
2	gender	1.3×10^{-1}
3	feminist	1.2×10^{-1}
4	sexual	5.3×10^{-2}
5	violenc	5.1×10^{-2}
6	articl	5×10^{-2}
7	social	4.9×10^{-2}
8	men	3.7×10^{-2}
9	femal	3.6×10^{-2}
10	femin	3.4×10^{-2}
11	argu	3.3×10^{-2}
12	polit	3.3×10^{-2}
13	masculin	2.8×10^{-2}
14	interview	2.6×10^{-2}
15	discours	2.3×10^{-2}
16	health	2.3×10^{-2}
17	draw	2.1×10^{-2}
18	intim	2.1×10^{-2}
19	explor	2.1×10^{-2}
20	partner	2×10^{-2}
21	victim	2×10^{-2}
22	examin	1.9×10^{-2}
23	narrat	1.9×10^{-2}
24	particip	1.8×10^{-2}
25	woman	1.7×10^{-2}
26	feminin	1.7×10^{-2}
27	cultur	1.7×10^{-2}
28	research	1.5×10^{-2}
29	result	1.5×10^{-2}
30	engag	1.5×10^{-2}
31	famili	1.5×10^{-2}
32	stereotyp	1.5×10^{-2}
33	ipv	1.4×10^{-2}
34	cell	1.4×10^{-2}
35	societi	1.3×10^{-2}
36	nation	1.3×10^{-2}
37	ident	1.3×10^{-2}
38	oppress	1.3×10^{-2}
39	critiqu	1.2×10^{-2}
40	mother	1.2×10^{-2}
41	context	1.2×10^{-2}
42	queer	1.2×10^{-2}
43	male	1.2×10^{-2}
44	heterosexu	1.2×10^{-2}
45	negoti	1.2×10^{-2}
46	discurs	1.2×10^{-2}
47	sex	1.2×10^{-2}
48	educ	1.2×10^{-2}
49	race	1.2×10^{-2}
50	live	1.2×10^{-2}
51	intersect	1.2×10^{-2}
52	care	1.1×10^{-2}
53	relationship	1.1×10^{-2}
54	way	1.1×10^{-2}
55	temperatur	1.1×10^{-2}
56	domest	1.1×10^{-2}
57	simul	1.1×10^{-2}
58	empower	1.1×10^{-2}
59	polici	1.1×10^{-2}
60	practic	1.1×10^{-2}
61	american	1.1×10^{-2}
62	activist	1×10^{-2}
63	effici	1×10^{-2}
64	among	1×10^{-2}
65	obtain	1×10^{-2}
66	seek	1×10^{-2}
67	scholar	1×10^{-2}
68	transnat	1×10^{-2}
69	perform	9.9×10^{-3}
70	norm	9.9×10^{-3}
71	perpetr	9.8×10^{-3}
72	focus	9.8×10^{-3}
73	lesbian	9.7×10^{-3}
74	paramet	9.6×10^{-3}
75	public	9.5×10^{-3}
76	properti	9.5×10^{-3}
77	pregnanc	9.3×10^{-3}
78	racial	9.3×10^{-3}
79	ideolog	9.2×10^{-3}
80	surfac	9.2×10^{-3}
81	patriarch	9.1×10^{-3}
82	psycholog	9×10^{-3}
83	survey	8.9×10^{-3}
84	qualit	8.9×10^{-3}
85	marriag	8.8×10^{-3}
86	debat	8.8×10^{-3}
87	labor	8.8×10^{-3}
88	abus	8.8×10^{-3}
89	show	8.7×10^{-3}
90	communiti	8.6×10^{-3}
91	work	8.6×10^{-3}
92	inequ	8.6×10^{-3}
93	energi	8.4×10^{-3}
94	right	8.3×10^{-3}
95	reproduct	8.3×10^{-3}
96	question	8.1×10^{-3}
97	ethnic	8.1×10^{-3}
98	system	8.1×10^{-3}
99	propos	8.1×10^{-3}
100	hegemon	8.1×10^{-3}

Table D.252. The list of the top 100 words in the category Zoology with RIGs

No.	Word	RIG
1	speci	1.9×10^{-1}
2	genus	7.2×10^{-2}
3	nov	4×10^{-2}
4	habitat	3.9×10^{-2}
5	femal	3.9×10^{-2}
6	male	3.8×10^{-2}
7	describ	2.6×10^{-2}
8	anim	2.5×10^{-2}
9	predat	2.4×10^{-2}
10	genera	2.3×10^{-2}
11	method	2.2×10^{-2}
12	reproduct	2.1×10^{-2}
13	morpholog	2.1×10^{-2}
14	record	2×10^{-2}
15	bird	1.9×10^{-2}
16	paper	1.9×10^{-2}
17	taxonom	1.8×10^{-2}
18	charact	1.8×10^{-2}
19	patient	1.8×10^{-2}
20	fish	1.8×10^{-2}
21	ecolog	1.8×10^{-2}
22	taxa	1.7×10^{-2}
23	specimen	1.7×10^{-2}
24	forag	1.7×10^{-2}
25	popul	1.7×10^{-2}
26	phylogenet	1.7×10^{-2}
27	mate	1.7×10^{-2}
28	mammal	1.6×10^{-2}
29	prey	1.5×10^{-2}
30	egg	1.4×10^{-2}
31	bodi	1.4×10^{-2}
32	fauna	1.4×10^{-2}
33	conspecif	1.3×10^{-2}
34	larva	1.3×10^{-2}
35	dorsal	1.3×10^{-2}
36	endem	1.3×10^{-2}
37	new	1.3×10^{-2}
38	breed	1.2×10^{-2}
39	primat	1.2×10^{-2}
40	season	1.2×10^{-2}
41	juvenil	1.2×10^{-2}
42	forest	1.2×10^{-2}
43	adult	1.1×10^{-2}
44	individu	1.1×10^{-2}
45	known	1.1×10^{-2}
46	brazil	1.1×10^{-2}
47	southern	1.1×10^{-2}
48	island	1.1×10^{-2}
49	collect	1×10^{-2}
50	abund	1×10^{-2}

No.	Word	RIG
51	conserv	1×10^{-2}
52	applic	1×10^{-2}
53	insect	1×10^{-2}
54	simul	9.9×10^{-3}
55	sex	9.9×10^{-3}
56	south	9.8×10^{-3}
57	perform	9.7×10^{-3}
58	pattern	9.7×10^{-3}
59	nest	9.6×10^{-3}
60	suggest	9.4×10^{-3}
61	gen	9.4×10^{-3}
62	clade	9.3×10^{-3}
63	sexual	9.2×10^{-3}
64	food	9×10^{-3}
65	ventral	9×10^{-3}
66	mitochondri	8.9×10^{-3}
67	trait	8.9×10^{-3}
68	properti	8.7×10^{-3}
69	improv	8.7×10^{-3}
70	solut	8.5×10^{-3}
71	eastern	8.5×10^{-3}
72	larval	8.4×10^{-3}
73	algorithm	8.2×10^{-3}
74	northern	8.2×10^{-3}
75	divers	8.2×10^{-3}
76	nematod	8×10^{-3}
77	propos	8×10^{-3}
78	distinguish	8×10^{-3}
79	taxon	7.9×10^{-3}
80	morphometr	7.6×10^{-3}
81	wildlif	7.6×10^{-3}
82	behaviour	7.6×10^{-3}
83	feed	7.5×10^{-3}
84	north	7.4×10^{-3}
85	offspr	7.3×10^{-3}
86	model	7.2×10^{-3}
87	inhabit	7.2×10^{-3}
88	comput	7.1×10^{-3}
89	geograph	7.1×10^{-3}
90	oper	7.1×10^{-3}
91	vertebr	7.1×10^{-3}
92	evolutionari	7×10^{-3}
93	conclus	6.9×10^{-3}
94	freshwat	6.8×10^{-3}
95	parasit	6.7×10^{-3}
96	wild	6.7×10^{-3}
97	use	6.7×10^{-3}
98	design	6.4×10^{-3}
99	hatch	6.3×10^{-3}
100	endang	6.2×10^{-3}

D.2. Number of Words for Each Category

TABLE D.253. Number of words for each category

No.	Category	Number of Words
1	Acoustics	4198
2	Agricultural Economics & Policy	2788
3	Agricultural Engineering	3727
4	Agriculture, Dairy & Animal Science	4080
5	Agriculture, Multidisciplinary	4359
6	Agronomy	4180
7	Allergy	3466
8	Anatomy & Morphology	3849
9	Andrology	2456
10	Anesthesiology	3695
11	Anthropology	4104
12	Archaeology	3689
13	Architecture	3243
14	Area Studies	3193
15	Art	3071
16	Asian Studies	2841
17	Astronomy & Astrophysics	3768
18	Audiology & Speech-Language Pathology	3467
19	Automation & Control Systems	4365
20	Behavioral Sciences	4195
21	Biochemical Research Methods	4582
22	Biochemistry & Molecular Biology	4755
23	Biodiversity Conservation	3915
24	Biology	4735
25	Biophysics	4501
26	Biotechnology & Applied Microbiology	4737
27	Business	3649
28	Business, Finance	3392
29	Cardiac & Cardiovascular Systems	4259
30	Cell & Tissue Engineering	3686
31	Cell Biology	4525
32	Chemistry, Analytical	4622
33	Chemistry, Applied	4348
34	Chemistry, Inorganic & Nuclear	3825

No.	Category	Number of Words
35	Chemistry, Medicinal	4292
36	Chemistry, Multidisciplinary	4694
37	Chemistry, Organic	3904
38	Chemistry, Physical	4480
39	Classics	2114
40	Clinical Neurology	4483
41	Communication	3323
42	Computer Science, Artificial Intelligence	4600
43	Computer Science, Cybernetics	3518
44	Computer Science, Hardware & Architecture	3983
45	Computer Science, Information Systems	4656
46	Computer Science, Interdisciplinary Applications	4834
47	Computer Science, Software Engineering	4100
48	Computer Science, Theory & Methods	4662
49	Construction & Building Technology	4060
50	Criminology & Penology	3055
51	Critical Care Medicine	3972
52	Crystallography	3593
53	Cultural Studies	3018
54	Dance	1295
55	Demography	2719
56	Dentistry, Oral Surgery & Medicine	4228
57	Dermatology	4150
58	Developmental Biology	3904
59	Ecology	4476
60	Economics	4027
61	Education & Educational Research	4218
62	Education, Scientific Disciplines	4299
63	Education, Special	2885
64	Electrochemistry	4126
65	Emergency Medicine	3596
66	Endocrinology & Metabolism	4436
67	Energy & Fuels	4473
68	Engineering, Aerospace	3525
69	Engineering, Biomedical	4571
70	Engineering, Chemical	4379
71	Engineering, Civil	4266

No.	Category	Number of Words
72	Engineering, Electrical & Electronic	4864
73	Engineering, Environmental	4439
74	Engineering, Geological	3345
75	Engineering, Industrial	4040
76	Engineering, Manufacturing	4155
77	Engineering, Marine	3027
78	Engineering, Mechanical	4577
79	Engineering, Multidisciplinary	4504
80	Engineering, Ocean	3028
81	Engineering, Petroleum	3164
82	Entomology	3874
83	Environmental Sciences	4865
84	Environmental Studies	3963
85	Ergonomics	3099
86	Ethics	3235
87	Ethnic Studies	2626
88	Evolutionary Biology	4123
89	Family Studies	2938
90	Film, Radio, Television	2457
91	Fisheries	3926
92	Folklore	1675
93	Food Science & Technology	4569
94	Forestry	3864
95	Gastroenterology & Hepatology	4138
96	Genetics & Heredity	4610
97	Geochemistry & Geophysics	3869
98	Geography	3691
99	Geography, Physical	3964
100	Geology	3302
101	Geosciences, Multidisciplinary	4395
102	Geriatrics & Gerontology	4012
103	Gerontology	3511
104	Green & Sustainable Science & Technology	4018
105	Health Care Sciences & Services	4119
106	Health Policy & Services	3716
107	Hematology	4082
108	History	3497

No.	Category	Number of Words
109	History & Philosophy Of Science	3698
110	History Of Social Sciences	2879
111	Horticulture	3864
112	Hospitality, Leisure, Sport & Tourism	3383
113	Humanities, Multidisciplinary	3579
114	Imaging Science & Photographic Technology	4090
115	Immunology	4421
116	Industrial Relations & Labor	2676
117	Infectious Diseases	4358
118	Information Science & Library Science	3672
119	Instruments & Instrumentation	4420
120	Integrative & Complementary Medicine	4060
121	International Relations	3260
122	Language & Linguistics	3456
123	Law	3415
124	Limnology	3436
125	Linguistics	3593
126	Literary Reviews	953
127	Literary Theory & Criticism	2473
128	Literature	3140
129	Literature, African, Australian, Canadian	1240
130	Literature, American	1380
131	Literature, British Isles	1895
132	Literature, German, Dutch, Scandinavian	1671
133	Literature, Romance	2145
134	Literature, Slavic	733
135	Logic	2450
136	Management	3860
137	Marine & Freshwater Biology	4328
138	Materials Science, Biomaterials	4165
139	Materials Science, Ceramics	3301
140	Materials Science, Characterization & Testing	3620
141	Materials Science, Coatings & Films	3336
142	Materials Science, Composites	3129
143	Materials Science, Multidisciplinary	4798
144	Materials Science, Paper & Wood	3144
145	Materials Science, Textiles	3353

No.	Category	Number of Words
146	Mathematical & Computational Biology	4412
147	Mathematics	3341
148	Mathematics, Applied	4221
149	Mathematics, Interdisciplinary Applications	4241
150	Mechanics	4458
151	Medical Ethics	2795
152	Medical Informatics	4057
153	Medical Laboratory Technology	3799
154	Medicine, General & Internal	4535
155	Medicine, Legal	3926
156	Medicine, Research & Experimental	4639
157	Medieval & Renaissance Studies	2445
158	Metallurgy & Metallurgical Engineering	3820
159	Meteorology & Atmospheric Sciences	4188
160	Microbiology	4480
161	Microscopy	3567
162	Mineralogy	3354
163	Mining & Mineral Processing	3468
164	Multidisciplinary Sciences	4956
165	Music	2933
166	Mycology	3441
167	Nanoscience & Nanotechnology	4470
168	Neuroimaging	3389
169	Neurosciences	4652
170	Nuclear Science & Technology	4283
171	Nursing	3891
172	Nutrition & Dietetics	4460
173	Obstetrics & Gynecology	4175
174	Oceanography	4053
175	Oncology	4461
176	Operations Research & Management Science	3940
177	Ophthalmology	4172
178	Optics	4674
179	Ornithology	2771
180	Orthopedics	4126
181	Otorhinolaryngology	3962
182	Paleontology	3388

No.	Category	Number of Words
183	Parasitology	4141
184	Pathology	4161
185	Pediatrics	4369
186	Peripheral Vascular Disease	4071
187	Pharmacology & Pharmacy	4690
188	Philosophy	3258
189	Physics, Applied	4659
190	Physics, Atomic, Molecular & Chemical	4041
191	Physics, Condensed Matter	4065
192	Physics, Fluids & Plasmas	3895
193	Physics, Mathematical	3743
194	Physics, Multidisciplinary	4284
195	Physics, Nuclear	3478
196	Physics, Particles & Fields	3248
197	Physiology	4480
198	Planning & Development	3594
199	Plant Sciences	4531
200	Poetry	982
201	Political Science	3282
202	Polymer Science	4052
203	Primary Health Care	3134
204	Psychiatry	4296
205	Psychology	4029
206	Psychology, Applied	3412
207	Psychology, Biological	3396
208	Psychology, Clinical	3700
209	Psychology, Developmental	3379
210	Psychology, Educational	2853
211	Psychology, Experimental	3680
212	Psychology, Mathematical	2222
213	Psychology, Multidisciplinary	3998
214	Psychology, Psychoanalysis	2119
215	Psychology, Social	3188
216	Public Administration	2958
217	Public, Environmental & Occupational Health	4805
218	Radiology, Nuclear Medicine & Medical Imaging	4635
219	Rehabilitation	4025

No.	Category	Number of Words
220	Religion	3317
221	Remote Sensing	4168
222	Reproductive Biology	3898
223	Respiratory System	4190
224	Rheumatology	3779
225	Robotics	3852
226	Social Issues	3064
227	Social Sciences, Biomedical	3655
228	Social Sciences, Interdisciplinary	4170
229	Social Sciences, Mathematical Methods	3339
230	Social Work	2947
231	Sociology	3715
232	Soil Science	3760
233	Spectroscopy	4267
234	Sport Sciences	4204
235	Statistics & Probability	4152
236	Substance Abuse	3717
237	Surgery	4548
238	Telecommunications	4282
239	Theater	2336
240	Thermodynamics	3997
241	Toxicology	4476
242	Transplantation	3852
243	Transportation	3420
244	Transportation Science & Technology	3578
245	Tropical Medicine	4054
246	Urban Studies	3385
247	Urology & Nephrology	4102
248	Veterinary Sciences	4446
249	Virology	3953
250	Water Resources	4384
251	Women's Studies	3145
252	Zoology	4471

D.3. List of Original Attributes (Categories) in Groups of Positive, Negative and Zero on the Principal Components

Table D.254. List of original attributes (categories) in group of positive on the 1^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC1)		
No.	**Attribute**	**Component Coefficient**
1	Engineering, Multidisciplinary	0.1294
2	Engineering, Electrical & Electronic	0.1177
3	Computer Science, Theory & Methods	0.1161
4	Medicine, General & Internal	0.1155
5	Computer Science, Interdisciplinary Applications	0.1145
6	Medicine, Research & Experimental	0.1125
7	Engineering, Industrial	0.1102
8	Computer Science, Software Engineering	0.1092
9	Computer Science, Information Systems	0.1081
10	Engineering, Mechanical	0.1065
11	Automation & Control Systems	0.1058
12	Computer Science, Cybernetics	0.1049
13	Materials Science, Multidisciplinary	0.1045
14	Computer Science, Artificial Intelligence	0.1033
15	Mechanics	0.1027
16	Physics, Applied	0.1020
17	Clinical Neurology	0.1014
18	Chemistry, Multidisciplinary	0.1001
19	Mathematics, Interdisciplinary Applications	0.0999
20	Physics, Multidisciplinary	0.0990
21	Computer Science, Hardware & Architecture	0.0983
22	Operations Research & Management Science	0.0979
23	Chemistry, Physical	0.0968
24	Engineering, Manufacturing	0.0959
25	Instruments & Instrumentation	0.0925
26	Medical Laboratory Technology	0.0924
27	Mathematics, Applied	0.0905
28	Social Sciences, Interdisciplinary	0.0897
29	Physics, Condensed Matter	0.0891

No.	Attribute	Component Coefficient
30	Medical Informatics	0.0879
31	Telecommunications	0.0867
32	Health Care Sciences & Services	0.0867
33	Gastroenterology & Hepatology	0.0864
34	Physics, Mathematical	0.0854
35	Public, Environmental & Occupational Health	0.0854
36	Social Sciences, Mathematical Methods	0.0853
37	Multidisciplinary Sciences	0.0834
38	Surgery	0.0832
39	Rehabilitation	0.0827
40	Nanoscience & Nanotechnology	0.0825
41	Chemistry, Applied	0.0825
42	Endocrinology & Metabolism	0.0820
43	Hematology	0.0810
44	Critical Care Medicine	0.0809
45	Engineering, Chemical	0.0808
46	Peripheral Vascular Disease	0.0808
47	Mathematics	0.0801
48	Biology	0.0798
49	Engineering, Civil	0.0786
50	Primary Health Care	0.0786
51	Urology & Nephrology	0.0784
52	Psychology, Multidisciplinary	0.0782
53	Respiratory System	0.0781
54	Planning & Development	0.0779
55	Psychology, Social	0.0778
56	Pediatrics	0.0776
57	Energy & Fuels	0.0771
58	Humanities, Multidisciplinary	0.0765
59	Psychology, Clinical	0.0764
60	Otorhinolaryngology	0.0760
61	Biochemistry & Molecular Biology	0.0752
62	Sociology	0.0749
63	Physics, Atomic, Molecular & Chemical	0.0746
64	Psychiatry	0.0739
65	Psychology	0.0737
66	Health Policy & Services	0.0735

No.	Attribute	Component Coefficient
67	Social Issues	0.0727
68	Anesthesiology	0.0719
69	Emergency Medicine	0.0719
70	Oncology	0.0718
71	Rheumatology	0.0716
72	Physiology	0.0707
73	Management	0.0706
74	Social Sciences, Biomedical	0.0704
75	Cell Biology	0.0701
76	Environmental Studies	0.0697
77	Materials Science, Characterization & Testing	0.0695
78	Information Science & Library Science	0.0691
79	Biophysics	0.0687
80	Cardiac & Cardiovascular Systems	0.0684
81	Geriatrics & Gerontology	0.0684
82	Pharmacology & Pharmacy	0.0684
83	Immunology	0.0674
84	Pathology	0.0666
85	Cultural Studies	0.0665
86	Green & Sustainable Science & Technology	0.0657
87	Logic	0.0655
88	Social Work	0.0649
89	Area Studies	0.0648
90	Infectious Diseases	0.0647
91	Statistics & Probability	0.0646
92	Geography	0.0643
93	Chemistry, Organic	0.0641
94	Business	0.0639
95	Psychology, Applied	0.0637
96	Orthopedics	0.0635
97	Economics	0.0634
98	Psychology, Mathematical	0.0628
99	Asian Studies	0.0627
100	History Of Social Sciences	0.0626
101	Communication	0.0623
102	Materials Science, Coatings & Films	0.0617
103	Biotechnology & Applied Microbiology	0.0614

No.	Attribute	Component Coefficient
104	Dermatology	0.0611
105	Transportation Science & Technology	0.0611
106	Psychology, Biological	0.0607
107	Physics, Fluids & Plasmas	0.0604
108	Public Administration	0.0604
109	Obstetrics & Gynecology	0.0602
110	Gerontology	0.0597
111	History	0.0595
112	International Relations	0.0594
113	Mathematical & Computational Biology	0.0590
114	Psychology, Experimental	0.0590
115	Integrative & Complementary Medicine	0.0584
116	Literary Theory & Criticism	0.0582
117	Optics	0.0577
118	Behavioral Sciences	0.0572
119	Psychology, Developmental	0.0561
120	History & Philosophy Of Science	0.0561
121	Metallurgy & Metallurgical Engineering	0.0558
122	Substance Abuse	0.0556
123	Materials Science, Ceramics	0.0556
124	Engineering, Aerospace	0.0554
125	Engineering, Environmental	0.0554
126	Spectroscopy	0.0554
127	Family Studies	0.0551
128	Imaging Science & Photographic Technology	0.0551
129	Psychology, Educational	0.0546
130	Construction & Building Technology	0.0540
131	Nursing	0.0537
132	Toxicology	0.0536
133	Demography	0.0535
134	Developmental Biology	0.0533
135	Sport Sciences	0.0532
136	Radiology, Nuclear Medicine & Medical Imaging	0.0532
137	Tropical Medicine	0.0529
138	Ethnic Studies	0.0524
139	Chemistry, Inorganic & Nuclear	0.0521
140	Literature, Romance	0.0519

No.	Attribute	Component Coefficient
141	Thermodynamics	0.0517
142	Political Science	0.0517
143	Anthropology	0.0516
144	Nutrition & Dietetics	0.0515
145	Crystallography	0.0513
146	Philosophy	0.0510
147	Literature	0.0503
148	Polymer Science	0.0498
149	Neurosciences	0.0497
150	Allergy	0.0494
151	Robotics	0.0485
152	Engineering, Marine	0.0483
153	Dentistry, Oral Surgery & Medicine	0.0482
154	Classics	0.0481
155	Engineering, Ocean	0.0476
156	Environmental Sciences	0.0474
157	Business, Finance	0.0473
158	Ergonomics	0.0473
159	Materials Science, Biomaterials	0.0471
160	Film, Radio, Television	0.0470
161	Engineering, Biomedical	0.0469
162	Genetics & Heredity	0.0465
163	Industrial Relations & Labor	0.0463
164	Literature, British Isles	0.0463
165	Education, Special	0.0460
166	Chemistry, Medicinal	0.0453
167	Veterinary Sciences	0.0452
168	Electrochemistry	0.0451
169	Food Science & Technology	0.0448
170	Medieval & Renaissance Studies	0.0445
171	Remote Sensing	0.0437
172	Zoology	0.0437
173	Physics, Nuclear	0.0436
174	Women's Studies	0.0426
175	Biochemical Research Methods	0.0426
176	Agricultural Engineering	0.0425
177	Folklore	0.0423

No.	Attribute	Component Coefficient
178	Engineering, Geological	0.0422
179	Urban Studies	0.0420
180	Art	0.0420
181	Ecology	0.0419
182	Materials Science, Composites	0.0418
183	Microbiology	0.0415
184	Physics, Particles & Fields	0.0414
185	Chemistry, Analytical	0.0414
186	Materials Science, Textiles	0.0408
187	Hospitality, Leisure, Sport & Tourism	0.0407
188	Ethics	0.0406
189	Architecture	0.0403
190	Reproductive Biology	0.0400
191	Ophthalmology	0.0399
192	Plant Sciences	0.0395
193	Medical Ethics	0.0390
194	Transplantation	0.0389
195	Agriculture, Multidisciplinary	0.0386
196	Acoustics	0.0382
197	Evolutionary Biology	0.0382
198	Anatomy & Morphology	0.0380
199	Education & Educational Research	0.0378
200	Marine & Freshwater Biology	0.0377
201	Literature, Slavic	0.0374
202	Language & Linguistics	0.0369
203	Biodiversity Conservation	0.0366
204	Cell & Tissue Engineering	0.0362
205	Religion	0.0362
206	Agricultural Economics & Policy	0.0359
207	Nuclear Science & Technology	0.0357
208	Water Resources	0.0352
209	Parasitology	0.0351
210	Neuroimaging	0.0348
211	Transportation	0.0344
212	Literature, African, Australian, Canadian	0.0342
213	Microscopy	0.0340
214	Geosciences, Multidisciplinary	0.0339

No.	Attribute	Component Coefficient
215	Agronomy	0.0329
216	Criminology & Penology	0.0329
217	Law	0.0327
218	Entomology	0.0327
219	Mining & Mineral Processing	0.0325
220	Literature, German, Dutch, Scandinavian	0.0324
221	Limnology	0.0320

Table D.255. List of original attributes (categories) in group of zero on the 1^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Zero (PC1)		
No.	Attribute	Component Coefficient
1	Virology	0.0314
2	Mycology	0.0311
3	Literature, American	0.0306
4	Literary Reviews	0.0302
5	Theater	0.0301
6	Astronomy & Astrophysics	0.0296
7	Meteorology & Atmospheric Sciences	0.0294
8	Linguistics	0.0294
9	Materials Science, Paper & Wood	0.0291
10	Psychology, Psychoanalysis	0.0282
11	Engineering, Petroleum	0.0278
12	Geography, Physical	0.0277
13	Paleontology	0.0271
14	Agriculture, Dairy & Animal Science	0.0266
15	Horticulture	0.0266
16	Andrology	0.0260
17	Fisheries	0.0259
18	Education, Scientific Disciplines	0.0258
19	Oceanography	0.0253
20	Geochemistry & Geophysics	0.0251
21	Audiology & Speech-Language Pathology	0.0242
22	Ornithology	0.0239

No.	Attribute	Component Coefficient
23	Soil Science	0.0225
24	Mineralogy	0.0223
25	Archaeology	0.0220
26	Forestry	0.0206
27	Geology	0.0206
28	Poetry	0.0204
29	Music	0.0172
30	Dance	0.0136
31	Medicine, Legal	0.0125

Table D.256. List of original attributes (categories) in group of positive on the 2^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC2)		
No.	**Attribute**	**Component Coefficient**
1	Cultural Studies	0.2044
2	Humanities, Multidisciplinary	0.1905
3	Asian Studies	0.1887
4	History	0.1877
5	Area Studies	0.1831
6	Literature	0.1765
7	History Of Social Sciences	0.1736
8	Sociology	0.1704
9	Social Issues	0.1688
10	Literature, Romance	0.1666
11	International Relations	0.1541
12	Political Science	0.1452
13	Medieval & Renaissance Studies	0.1451
14	Literary Theory & Criticism	0.1437
15	Ethnic Studies	0.1436
16	History & Philosophy Of Science	0.1376
17	Film, Radio, Television	0.1367
18	Communication	0.1311
19	Literature, African, Australian, Canadian	0.1304
20	Literature, British Isles	0.1286

No.	Attribute	Component Coefficient
21	Folklore	0.1222
22	Classics	0.1200
23	Literature, American	0.1199
24	Geography	0.1196
25	Literature, German, Dutch, Scandinavian	0.1192
26	Planning & Development	0.1180
27	Art	0.1165
28	Anthropology	0.1162
29	Social Sciences, Interdisciplinary	0.1150
30	Religion	0.1118
31	Public Administration	0.1090
32	Philosophy	0.1058
33	Theater	0.1000
34	Ethics	0.0982
35	Literary Reviews	0.0924
36	Law	0.0905
37	Literature, Slavic	0.0884
38	Poetry	0.0856
39	Industrial Relations & Labor	0.0797
40	Women's Studies	0.0762
41	Environmental Studies	0.0755
42	Social Work	0.0733
43	Demography	0.0722
44	Urban Studies	0.0684
45	Medical Ethics	0.0675
46	Social Sciences, Biomedical	0.0661
47	Language & Linguistics	0.0603
48	Business	0.0544
49	Information Science & Library Science	0.0521
50	Management	0.0475
51	Linguistics	0.0462
52	Music	0.0457
53	Economics	0.0433
54	Psychology, Applied	0.0430
55	Hospitality, Leisure, Sport & Tourism	0.0429
56	Psychology, Social	0.0420
57	Psychology, Multidisciplinary	0.0418

No.	Attribute	Component Coefficient
58	Education & Educational Research	0.0418
59	Criminology & Penology	0.0412
60	Architecture	0.0411
61	Psychology, Psychoanalysis	0.0363
62	Archaeology	0.0362
63	Dance	0.0349

TABLE D.257. List of original attributes (categories) in group of zero on the 2^{th} principal component. Categories with positive component coefficients are sorted by values of their component coefficients in descending order. Categories with negative component coefficients are sorted by absolute values of their component coefficients in descending order. That is, categories in two directions are sorted from the greatest contribution to lowest contribution to the component.

Zero (PC2)		
No.	**Attribute**	**Component Coefficient**
1	Family Studies	0.0279
2	Psychology, Educational	0.0255
3	Business, Finance	0.0236
4	Agricultural Economics & Policy	0.0226
5	Education, Scientific Disciplines	0.0175
6	Education, Special	0.0087
7	Psychology, Mathematical	0.0083
8	Health Policy & Services	0.0038
9	Social Sciences, Mathematical Methods	0.0014
10	Chemistry, Analytical	0.0012
11	Psychology, Experimental	0.0010
12	Transportation	0.0003
13	Veterinary Sciences	-0.0314
14	Instruments & Instrumentation	-0.0314
15	Mathematics, Applied	-0.0313
16	Physics, Mathematical	-0.0305
17	Dentistry, Oral Surgery & Medicine	-0.0302
18	Nutrition & Dietetics	-0.0302
19	Physiology	-0.0300
20	Engineering, Marine	-0.0299
21	Neurosciences	-0.0292

No.	Attribute	Component Coefficient
22	Robotics	-0.0286
23	Geosciences, Multidisciplinary	-0.0282
24	Rehabilitation	-0.0276
25	Engineering, Ocean	-0.0275
26	Psychiatry	-0.0273
27	Marine & Freshwater Biology	-0.0272
28	Reproductive Biology	-0.0270
29	Engineering, Environmental	-0.0267
30	Primary Health Care	-0.0266
31	Ecology	-0.0259
32	Engineering, Geological	-0.0258
33	Agriculture, Multidisciplinary	-0.0250
34	Chemistry, Multidisciplinary	-0.0248
35	Ophthalmology	-0.0247
36	Imaging Science & Photographic Technology	-0.0245
37	Plant Sciences	-0.0242
38	Engineering, Chemical	-0.0242
39	Parasitology	-0.0241
40	Physics, Applied	-0.0240
41	Radiology, Nuclear Medicine & Medical Imaging	-0.0239
42	Chemistry, Applied	-0.0237
43	Limnology	-0.0235
44	Zoology	-0.0230
45	Agricultural Engineering	-0.0228
46	Integrative & Complementary Medicine	-0.0227
47	Oceanography	-0.0227
48	Chemistry, Physical	-0.0225
49	Environmental Sciences	-0.0225
50	Anatomy & Morphology	-0.0224
51	Entomology	-0.0223
52	Agronomy	-0.0222
53	Neuroimaging	-0.0220
54	Green & Sustainable Science & Technology	-0.0217
55	Genetics & Heredity	-0.0212
56	Evolutionary Biology	-0.0208
57	Remote Sensing	-0.0207
58	Mining & Mineral Processing	-0.0207

No.	Attribute	Component Coefficient
59	Physics, Fluids & Plasmas	-0.0202
60	Water Resources	-0.0200
61	Biochemistry & Molecular Biology	-0.0200
62	Behavioral Sciences	-0.0199
63	Geochemistry & Geophysics	-0.0197
64	Biodiversity Conservation	-0.0196
65	Medical Informatics	-0.0196
66	Horticulture	-0.0194
67	Fisheries	-0.0192
68	Toxicology	-0.0192
69	Developmental Biology	-0.0182
70	Geography, Physical	-0.0179
71	Metallurgy & Metallurgical Engineering	-0.0178
72	Materials Science, Textiles	-0.0177
73	Physics, Condensed Matter	-0.0175
74	Cell Biology	-0.0171
75	Agriculture, Dairy & Animal Science	-0.0171
76	Andrology	-0.0170
77	Ornithology	-0.0169
78	Physics, Atomic, Molecular & Chemical	-0.0169
79	Food Science & Technology	-0.0168
80	Engineering, Petroleum	-0.0165
81	Psychology, Biological	-0.0164
82	Geology	-0.0163
83	Mineralogy	-0.0163
84	Meteorology & Atmospheric Sciences	-0.0162
85	Microbiology	-0.0161
86	Virology	-0.0161
87	Physics, Nuclear	-0.0161
88	Optics	-0.0156
89	Forestry	-0.0156
90	Health Care Sciences & Services	-0.0156
91	Physics, Particles & Fields	-0.0153
92	Nanoscience & Nanotechnology	-0.0151
93	Gerontology	-0.0149
94	Materials Science, Composites	-0.0148
95	Biotechnology & Applied Microbiology	-0.0144

No.	Attribute	Component Coefficient
96	Materials Science, Paper & Wood	-0.0143
97	Soil Science	-0.0141
98	Nuclear Science & Technology	-0.0133
99	Thermodynamics	-0.0133
100	Polymer Science	-0.0132
101	Substance Abuse	-0.0130
102	Paleontology	-0.0125
103	Chemistry, Inorganic & Nuclear	-0.0124
104	Psychology	-0.0123
105	Materials Science, Ceramics	-0.0123
106	Statistics & Probability	-0.0122
107	Chemistry, Medicinal	-0.0122
108	Mathematical & Computational Biology	-0.0117
109	Crystallography	-0.0114
110	Biophysics	-0.0109
111	Materials Science, Coatings & Films	-0.0109
112	Mathematics	-0.0107
113	Chemistry, Organic	-0.0103
114	Mycology	-0.0101
115	Cell & Tissue Engineering	-0.0099
116	Astronomy & Astrophysics	-0.0095
117	Acoustics	-0.0093
118	Public, Environmental & Occupational Health	-0.0091
119	Electrochemistry	-0.0087
120	Logic	-0.0087
121	Engineering, Biomedical	-0.0066
122	Materials Science, Biomaterials	-0.0066
123	Psychology, Clinical	-0.0064
124	Spectroscopy	-0.0047
125	Medicine, Legal	-0.0041
126	Biochemical Research Methods	-0.0028
127	Nursing	-0.0025
128	Audiology & Speech-Language Pathology	-0.0025
129	Microscopy	-0.0019
130	Ergonomics	-0.0014
131	Psychology, Developmental	-0.0011

Table D.258. List of original attributes (categories) in group of negative on the 2^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC2)		
No.	**Attribute**	**Component Coefficient**
1	Engineering, Multidisciplinary	-0.0759
2	Clinical Neurology	-0.0686
3	Medicine, General & Internal	-0.0677
4	Engineering, Electrical & Electronic	-0.0655
5	Computer Science, Theory & Methods	-0.0621
6	Gastroenterology & Hepatology	-0.0612
7	Surgery	-0.0605
8	Medicine, Research & Experimental	-0.0579
9	Critical Care Medicine	-0.0574
10	Computer Science, Interdisciplinary Applications	-0.0573
11	Peripheral Vascular Disease	-0.0573
12	Respiratory System	-0.0570
13	Engineering, Industrial	-0.0570
14	Automation & Control Systems	-0.0556
15	Computer Science, Information Systems	-0.0555
16	Urology & Nephrology	-0.0554
17	Computer Science, Hardware & Architecture	-0.0550
18	Engineering, Manufacturing	-0.0546
19	Computer Science, Artificial Intelligence	-0.0540
20	Engineering, Mechanical	-0.0532
21	Mathematics, Interdisciplinary Applications	-0.0524
22	Computer Science, Software Engineering	-0.0524
23	Cardiac & Cardiovascular Systems	-0.0516
24	Computer Science, Cybernetics	-0.0512
25	Anesthesiology	-0.0510
26	Hematology	-0.0506
27	Otorhinolaryngology	-0.0503
28	Pediatrics	-0.0502
29	Mechanics	-0.0497
30	Telecommunications	-0.0493
31	Orthopedics	-0.0486

No.	Attribute	Component Coefficient
32	Emergency Medicine	-0.0477
33	Rheumatology	-0.0464
34	Endocrinology & Metabolism	-0.0458
35	Engineering, Civil	-0.0449
36	Operations Research & Management Science	-0.0447
37	Dermatology	-0.0429
38	Medical Laboratory Technology	-0.0426
39	Oncology	-0.0412
40	Infectious Diseases	-0.0403
41	Immunology	-0.0392
42	Obstetrics & Gynecology	-0.0391
43	Biology	-0.0386
44	Pathology	-0.0365
45	Geriatrics & Gerontology	-0.0360
46	Transportation Science & Technology	-0.0352
47	Energy & Fuels	-0.0348
48	Materials Science, Multidisciplinary	-0.0334
49	Allergy	-0.0332
50	Pharmacology & Pharmacy	-0.0328
51	Engineering, Aerospace	-0.0328
52	Transplantation	-0.0326
53	Materials Science, Characterization & Testing	-0.0323
54	Sport Sciences	-0.0317
55	Tropical Medicine	-0.0317
56	Construction & Building Technology	-0.0317
57	Physics, Multidisciplinary	-0.0316
58	Multidisciplinary Sciences	-0.0316

TABLE D.259. List of original attributes (categories) in group of positive on the 3^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC3)		
No.	**Attribute**	**Component Coefficient**
1	Multidisciplinary Sciences	0.1711
2	Biochemistry & Molecular Biology	0.1628

No.	Attribute	Component Coefficient
3	Cell Biology	0.1570
4	Biology	0.1392
5	Computer Science, Information Systems	0.1324
6	Computer Science, Theory & Methods	0.1320
7	Computer Science, Interdisciplinary Applications	0.1303
8	Computer Science, Artificial Intelligence	0.1296
9	Biotechnology & Applied Microbiology	0.1262
10	Biophysics	0.1261
11	Developmental Biology	0.1226
12	Computer Science, Software Engineering	0.1199
13	Computer Science, Cybernetics	0.1183
14	Automation & Control Systems	0.1140
15	Computer Science, Hardware & Architecture	0.1122
16	Engineering, Industrial	0.1109
17	Physiology	0.1104
18	Operations Research & Management Science	0.1079
19	Engineering, Electrical & Electronic	0.1044
20	Engineering, Multidisciplinary	0.0991
21	Telecommunications	0.0950
22	Genetics & Heredity	0.0941
23	Microbiology	0.0938
24	Toxicology	0.0914
25	Cell & Tissue Engineering	0.0880
26	Pharmacology & Pharmacy	0.0880
27	Medicine, Research & Experimental	0.0777
28	Mathematics, Interdisciplinary Applications	0.0746
29	Mathematical & Computational Biology	0.0726
30	Immunology	0.0723
31	Transportation Science & Technology	0.0718
32	Plant Sciences	0.0652
33	Robotics	0.0647
34	Virology	0.0634
35	Chemistry, Medicinal	0.0630
36	Mycology	0.0614
37	Evolutionary Biology	0.0601
38	Imaging Science & Photographic Technology	0.0596
39	Ecology	0.0526

No.	Attribute	Component Coefficient
40	Zoology	0.0521
41	Biochemical Research Methods	0.0506
42	Biodiversity Conservation	0.0485
43	Engineering, Manufacturing	0.0469
44	Social Sciences, Mathematical Methods	0.0467
45	Parasitology	0.0435
46	Anatomy & Morphology	0.0430
47	Veterinary Sciences	0.0428
48	Marine & Freshwater Biology	0.0418
49	Engineering, Aerospace	0.0412
50	Engineering, Civil	0.0406
51	Pathology	0.0399
52	Engineering, Marine	0.0373
53	Acoustics	0.0369
54	Entomology	0.0364
55	Ornithology	0.0350
56	Engineering, Mechanical	0.0342
57	Food Science & Technology	0.0339
58	Mathematics, Applied	0.0337
59	Engineering, Ocean	0.0329
60	Fisheries	0.0325

TABLE D.260. List of original attributes (categories) in group of zero on the 3^{th} principal component. Categories with positive component coefficients are sorted by values of their component coefficients in descending order. Categories with negative component coefficients are sorted by absolute values of their component coefficients in descending order. That is, categories in two directions are sorted from the greatest contribution to lowest contribution to the component.

Zero (PC3)		
No.	Attribute	Component Coefficient
1	Remote Sensing	0.0305
2	Agronomy	0.0303
3	Horticulture	0.0295
4	Logic	0.0293
5	Materials Science, Biomaterials	0.0292
6	Transportation	0.0290

No.	Attribute	Component Coefficient
7	Agriculture, Dairy & Animal Science	0.0283
8	Literature, Romance	0.0279
9	Agriculture, Multidisciplinary	0.0276
10	Statistics & Probability	0.0273
11	Literary Theory & Criticism	0.0273
12	Reproductive Biology	0.0267
13	Literature, British Isles	0.0259
14	Engineering, Biomedical	0.0258
15	Medieval & Renaissance Studies	0.0255
16	Architecture	0.0248
17	Economics	0.0244
18	Forestry	0.0242
19	Paleontology	0.0236
20	Literature	0.0232
21	Construction & Building Technology	0.0227
22	Literary Reviews	0.0209
23	Classics	0.0209
24	Mathematics	0.0203
25	Geography, Physical	0.0202
26	Humanities, Multidisciplinary	0.0198
27	Literature, American	0.0198
28	Integrative & Complementary Medicine	0.0196
29	Literature, Slavic	0.0191
30	Asian Studies	0.0188
31	Neurosciences	0.0184
32	History	0.0181
33	Oncology	0.0181
34	Literature, German, Dutch, Scandinavian	0.0177
35	Art	0.0176
36	Business, Finance	0.0175
37	Information Science & Library Science	0.0173
38	Andrology	0.0173
39	History Of Social Sciences	0.0167
40	Philosophy	0.0167
41	Engineering, Geological	0.0159
42	Poetry	0.0149
43	Mechanics	0.0145

No.	Attribute	Component Coefficient
44	Folklore	0.0142
45	Soil Science	0.0141
46	Environmental Sciences	0.0140
47	Literature, African, Australian, Canadian	0.0137
48	History & Philosophy Of Science	0.0136
49	Limnology	0.0135
50	Archaeology	0.0131
51	Language & Linguistics	0.0131
52	Management	0.0125
53	Religion	0.0117
54	Theater	0.0116
55	Hematology	0.0099
56	Agricultural Economics & Policy	0.0099
57	Meteorology & Atmospheric Sciences	0.0096
58	Oceanography	0.0083
59	Cultural Studies	0.0082
60	Linguistics	0.0081
61	Planning & Development	0.0078
62	Environmental Studies	0.0076
63	Area Studies	0.0072
64	Instruments & Instrumentation	0.0066
65	Behavioral Sciences	0.0066
66	Business	0.0066
67	Geosciences, Multidisciplinary	0.0064
68	Film, Radio, Television	0.0063
69	Anthropology	0.0061
70	Geology	0.0053
71	International Relations	0.0053
72	Geography	0.0047
73	Endocrinology & Metabolism	0.0032
74	Urban Studies	0.0032
75	Law	0.0028
76	Dance	0.0022
77	Music	0.0022
78	Political Science	0.0019
79	Agricultural Engineering	0.0009
80	Geochemistry & Geophysics	0.0006

No.	Attribute	Component Coefficient
81	Medicine, Legal	0.0002
82	Water Resources	0.0002
83	Transplantation	-0.0307
84	Psychology, Experimental	-0.0304
85	Metallurgy & Metallurgical Engineering	-0.0300
86	Materials Science, Textiles	-0.0294
87	Crystallography	-0.0293
88	Social Issues	-0.0276
89	Sociology	-0.0273
90	Physics, Fluids & Plasmas	-0.0268
91	Criminology & Penology	-0.0262
92	Nutrition & Dietetics	-0.0258
93	Spectroscopy	-0.0246
94	Optics	-0.0242
95	Education & Educational Research	-0.0239
96	Infectious Diseases	-0.0219
97	Energy & Fuels	-0.0218
98	Audiology & Speech-Language Pathology	-0.0216
99	Materials Science, Composites	-0.0202
100	Education, Scientific Disciplines	-0.0197
101	Thermodynamics	-0.0195
102	Neuroimaging	-0.0191
103	Physics, Nuclear	-0.0166
104	Microscopy	-0.0163
105	Chemistry, Inorganic & Nuclear	-0.0158
106	Materials Science, Characterization & Testing	-0.0145
107	Engineering, Environmental	-0.0136
108	Industrial Relations & Labor	-0.0136
109	Materials Science, Paper & Wood	-0.0135
110	Tropical Medicine	-0.0134
111	Hospitality, Leisure, Sport & Tourism	-0.0116
112	Physics, Mathematical	-0.0116
113	Physics, Particles & Fields	-0.0115
114	Green & Sustainable Science & Technology	-0.0113
115	Ethics	-0.0112
116	Ethnic Studies	-0.0110
117	Mineralogy	-0.0105

No.	Attribute	Component Coefficient
118	Chemistry, Applied	-0.0104
119	Mining & Mineral Processing	-0.0102
120	Engineering, Petroleum	-0.0092
121	Nuclear Science & Technology	-0.0085
122	Psychology, Mathematical	-0.0074
123	Ergonomics	-0.0073
124	Communication	-0.0071
125	Psychology, Psychoanalysis	-0.0070
126	Psychology, Biological	-0.0064
127	Chemistry, Organic	-0.0046
128	Chemistry, Analytical	-0.0032
129	Public Administration	-0.0020
130	Astronomy & Astrophysics	-0.0019
131	Social Sciences, Interdisciplinary	-0.0017

Table D.261. List of original attributes (categories) in group of negative on the 3^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC3)		
No.	Attribute	Component Coefficient
1	Medicine, General & Internal	-0.1713
2	Health Care Sciences & Services	-0.1592
3	Primary Health Care	-0.1449
4	Public, Environmental & Occupational Health	-0.1448
5	Health Policy & Services	-0.1363
6	Critical Care Medicine	-0.1320
7	Clinical Neurology	-0.1303
8	Rehabilitation	-0.1201
9	Gerontology	-0.1180
10	Emergency Medicine	-0.1167
11	Geriatrics & Gerontology	-0.1161
12	Pediatrics	-0.1160
13	Psychiatry	-0.1143
14	Surgery	-0.1109

No.	Attribute	Component Coefficient
15	Otorhinolaryngology	-0.1095
16	Social Sciences, Biomedical	-0.1085
17	Anesthesiology	-0.1066
18	Respiratory System	-0.1051
19	Psychology, Clinical	-0.1026
20	Nursing	-0.1003
21	Cardiac & Cardiovascular Systems	-0.0977
22	Urology & Nephrology	-0.0928
23	Materials Science, Multidisciplinary	-0.0923
24	Rheumatology	-0.0885
25	Orthopedics	-0.0883
26	Gastroenterology & Hepatology	-0.0880
27	Peripheral Vascular Disease	-0.0865
28	Physics, Applied	-0.0844
29	Obstetrics & Gynecology	-0.0822
30	Social Work	-0.0822
31	Nanoscience & Nanotechnology	-0.0817
32	Psychology, Multidisciplinary	-0.0805
33	Psychology, Developmental	-0.0801
34	Substance Abuse	-0.0768
35	Family Studies	-0.0766
36	Psychology	-0.0763
37	Physics, Condensed Matter	-0.0713
38	Chemistry, Physical	-0.0708
39	Medical Laboratory Technology	-0.0677
40	Education, Special	-0.0629
41	Allergy	-0.0629
42	Psychology, Social	-0.0580
43	Chemistry, Multidisciplinary	-0.0559
44	Physics, Multidisciplinary	-0.0513
45	Ophthalmology	-0.0508
46	Sport Sciences	-0.0506
47	Dentistry, Oral Surgery & Medicine	-0.0495
48	Materials Science, Coatings & Films	-0.0489
49	Engineering, Chemical	-0.0485
50	Dermatology	-0.0482
51	Psychology, Educational	-0.0472

No.	Attribute	Component Coefficient
52	Psychology, Applied	-0.0464
53	Physics, Atomic, Molecular & Chemical	-0.0439
54	Medical Ethics	-0.0419
55	Medical Informatics	-0.0412
56	Demography	-0.0353
57	Materials Science, Ceramics	-0.0344
58	Electrochemistry	-0.0339
59	Women's Studies	-0.0339
60	Radiology, Nuclear Medicine & Medical Imaging	-0.0336
61	Polymer Science	-0.0329

Table D.262. List of original attributes (categories) in group of positive on the 4^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC4)		
No.	Attribute	Component Coefficient
1	Psychology, Multidisciplinary	0.1975
2	Psychology, Applied	0.1690
3	Psychology, Social	0.1638
4	Social Work	0.1593
5	Psychology	0.1373
6	Family Studies	0.1363
7	Psychology, Clinical	0.1331
8	Social Sciences, Biomedical	0.1265
9	Management	0.1221
10	Psychology, Developmental	0.1195
11	Psychology, Educational	0.1174
12	Business	0.1161
13	Psychology, Experimental	0.1103
14	Social Sciences, Interdisciplinary	0.1088
15	Education, Special	0.1052
16	Behavioral Sciences	0.0976
17	Information Science & Library Science	0.0974
18	Psychology, Biological	0.0963
19	Planning & Development	0.0947

No.	**Attribute**	**Component Coefficient**
20	Industrial Relations & Labor	0.0946
21	Demography	0.0924
22	Environmental Studies	0.0923
23	Education & Educational Research	0.0910
24	Social Issues	0.0861
25	Economics	0.0855
26	Sociology	0.0827
27	Health Policy & Services	0.0786
28	Criminology & Penology	0.0782
29	Ergonomics	0.0760
30	Education, Scientific Disciplines	0.0757
31	Public Administration	0.0742
32	Hospitality, Leisure, Sport & Tourism	0.0742
33	Business, Finance	0.0700
34	Psychology, Mathematical	0.0637
35	Public, Environmental & Occupational Health	0.0611
36	Agricultural Economics & Policy	0.0600
37	Biophysics	0.0559
38	Urban Studies	0.0558
39	Gerontology	0.0552
40	Psychiatry	0.0550
41	Biology	0.0543
42	Biochemistry & Molecular Biology	0.0524
43	Women's Studies	0.0510
44	Geography	0.0491
45	Nursing	0.0465
46	Biotechnology & Applied Microbiology	0.0459
47	Multidisciplinary Sciences	0.0439
48	Substance Abuse	0.0421
49	Materials Science, Coatings & Films	0.0418
50	Medical Ethics	0.0418
51	Genetics & Heredity	0.0411
52	Nanoscience & Nanotechnology	0.0405
53	Cell Biology	0.0404
54	Evolutionary Biology	0.0400
55	Developmental Biology	0.0399
56	Ecology	0.0396

No.	Attribute	Component Coefficient
57	Ethics	0.0390
58	Chemistry, Physical	0.0386
59	Neurosciences	0.0382
60	Biodiversity Conservation	0.0374
61	Transportation	0.0370
62	Materials Science, Biomaterials	0.0355
63	Physics, Condensed Matter	0.0344
64	Rehabilitation	0.0341
65	Social Sciences, Mathematical Methods	0.0338
66	Materials Science, Ceramics	0.0321
67	Microbiology	0.0318

TABLE D.263. List of original attributes (categories) in group of zero on the 4^{th} principal component. Categories with positive component coefficients are sorted by values of their component coefficients in descending order. Categories with negative component coefficients are sorted by absolute values of their component coefficients in descending order. That is, categories in two directions are sorted from the greatest contribution to lowest contribution to the component.

Zero (PC4)		
No.	**Attribute**	**Component Coefficient**
1	Biochemical Research Methods	0.0313
2	Communication	0.0309
3	Ethnic Studies	0.0307
4	Microscopy	0.0300
5	Spectroscopy	0.0284
6	Zoology	0.0280
7	Polymer Science	0.0279
8	Mycology	0.0272
9	Plant Sciences	0.0268
10	Ornithology	0.0261
11	Materials Science, Multidisciplinary	0.0259
12	Physics, Applied	0.0253
13	Electrochemistry	0.0249
14	Chemistry, Multidisciplinary	0.0223
15	Health Care Sciences & Services	0.0218
16	Chemistry, Applied	0.0216

No.	Attribute	Component Coefficient
17	Audiology & Speech-Language Pathology	0.0200
18	Metallurgy & Metallurgical Engineering	0.0199
19	Chemistry, Inorganic & Nuclear	0.0186
20	Chemistry, Analytical	0.0179
21	Marine & Freshwater Biology	0.0171
22	Mathematical & Computational Biology	0.0167
23	Chemistry, Medicinal	0.0167
24	Entomology	0.0164
25	Cell & Tissue Engineering	0.0163
26	Anatomy & Morphology	0.0161
27	Forestry	0.0159
28	Materials Science, Composites	0.0156
29	Engineering, Chemical	0.0153
30	Crystallography	0.0152
31	Physics, Atomic, Molecular & Chemical	0.0148
32	Virology	0.0140
33	Political Science	0.0134
34	International Relations	0.0130
35	Neuroimaging	0.0129
36	Physiology	0.0123
37	Agronomy	0.0120
38	Geriatrics & Gerontology	0.0111
39	Law	0.0110
40	Horticulture	0.0109
41	Toxicology	0.0106
42	Nuclear Science & Technology	0.0096
43	Acoustics	0.0094
44	Materials Science, Textiles	0.0093
45	Medical Informatics	0.0083
46	Optics	0.0082
47	Fisheries	0.0082
48	Astronomy & Astrophysics	0.0082
49	Agriculture, Multidisciplinary	0.0079
50	Food Science & Technology	0.0079
51	Environmental Sciences	0.0078
52	Soil Science	0.0078
53	Limnology	0.0070

No.	Attribute	Component Coefficient
54	Physics, Nuclear	0.0063
55	Materials Science, Paper & Wood	0.0060
56	Medicine, Legal	0.0058
57	Mineralogy	0.0050
58	Green & Sustainable Science & Technology	0.0045
59	Parasitology	0.0044
60	Engineering, Biomedical	0.0043
61	Operations Research & Management Science	0.0042
62	Materials Science, Characterization & Testing	0.0038
63	Anthropology	0.0029
64	Geography, Physical	0.0012
65	Oceanography	0.0011
66	Thermodynamics	0.0011
67	Physics, Particles & Fields	0.0010
68	Agricultural Engineering	0.0000
69	Pathology	-0.0305
70	Tropical Medicine	-0.0293
71	Automation & Control Systems	-0.0289
72	Archaeology	-0.0279
73	Language & Linguistics	-0.0262
74	Physics, Mathematical	-0.0259
75	Computer Science, Interdisciplinary Applications	-0.0242
76	Engineering, Mechanical	-0.0228
77	Architecture	-0.0227
78	Mechanics	-0.0219
79	Computer Science, Theory & Methods	-0.0217
80	Mathematics, Applied	-0.0196
81	Pharmacology & Pharmacy	-0.0191
82	Computer Science, Hardware & Architecture	-0.0191
83	Engineering, Aerospace	-0.0187
84	Area Studies	-0.0186
85	Computer Science, Software Engineering	-0.0182
86	Engineering, Manufacturing	-0.0181
87	Engineering, Civil	-0.0178
88	Computer Science, Artificial Intelligence	-0.0178
89	Telecommunications	-0.0175
90	Immunology	-0.0174

No.	Attribute	Component Coefficient
91	Physics, Multidisciplinary	-0.0171
92	Music	-0.0167
93	Dance	-0.0165
94	Logic	-0.0164
95	Reproductive Biology	-0.0131
96	Primary Health Care	-0.0129
97	Mathematics	-0.0128
98	Statistics & Probability	-0.0127
99	Psychology, Psychoanalysis	-0.0124
100	Instruments & Instrumentation	-0.0117
101	Construction & Building Technology	-0.0112
102	Sport Sciences	-0.0106
103	Energy & Fuels	-0.0106
104	Computer Science, Information Systems	-0.0104
105	Geosciences, Multidisciplinary	-0.0100
106	Engineering, Ocean	-0.0100
107	Engineering, Marine	-0.0100
108	Veterinary Sciences	-0.0091
109	Nutrition & Dietetics	-0.0083
110	Robotics	-0.0082
111	Andrology	-0.0080
112	Geology	-0.0076
113	Mining & Mineral Processing	-0.0071
114	Remote Sensing	-0.0066
115	Engineering, Industrial	-0.0066
116	Physics, Fluids & Plasmas	-0.0062
117	Imaging Science & Photographic Technology	-0.0061
118	Geochemistry & Geophysics	-0.0059
119	Transportation Science & Technology	-0.0047
120	Engineering, Geological	-0.0030
121	Agriculture, Dairy & Animal Science	-0.0027
122	Linguistics	-0.0025
123	Computer Science, Cybernetics	-0.0020
124	Meteorology & Atmospheric Sciences	-0.0019
125	Chemistry, Organic	-0.0019
126	Paleontology	-0.0016
127	Engineering, Environmental	-0.0015

No.	Attribute	Component Coefficient
128	Engineering, Petroleum	-0.0015
129	Water Resources	-0.0011

TABLE D.264. List of original attributes (categories) in group of negative on the 4^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC4)		
No.	Attribute	Component Coefficient
1	Literature	-0.1482
2	Surgery	-0.1389
3	Literature, Romance	-0.1364
4	Literature, British Isles	-0.1318
5	Literary Theory & Criticism	-0.1295
6	Gastroenterology & Hepatology	-0.1279
7	Medieval & Renaissance Studies	-0.1242
8	Critical Care Medicine	-0.1241
9	Respiratory System	-0.1227
10	Literature, American	-0.1203
11	Medicine, General & Internal	-0.1171
12	Cardiac & Cardiovascular Systems	-0.1162
13	Peripheral Vascular Disease	-0.1155
14	Urology & Nephrology	-0.1152
15	Classics	-0.1145
16	Otorhinolaryngology	-0.1128
17	Literature, German, Dutch, Scandinavian	-0.1118
18	Anesthesiology	-0.1118
19	Medical Laboratory Technology	-0.1088
20	Literature, African, Australian, Canadian	-0.1088
21	Clinical Neurology	-0.1087
22	Poetry	-0.1041
23	Asian Studies	-0.1015
24	Rheumatology	-0.0990
25	Orthopedics	-0.0990
26	Emergency Medicine	-0.0988

No.	Attribute	Component Coefficient
27	Literature, Slavic	-0.0953
28	Literary Reviews	-0.0949
29	Humanities, Multidisciplinary	-0.0907
30	Art	-0.0857
31	History	-0.0835
32	Dermatology	-0.0807
33	Folklore	-0.0704
34	Cultural Studies	-0.0703
35	Hematology	-0.0687
36	Theater	-0.0670
37	Dentistry, Oral Surgery & Medicine	-0.0656
38	Allergy	-0.0646
39	Transplantation	-0.0628
40	Film, Radio, Television	-0.0614
41	Obstetrics & Gynecology	-0.0606
42	Oncology	-0.0573
43	Ophthalmology	-0.0564
44	Medicine, Research & Experimental	-0.0563
45	Philosophy	-0.0560
46	Religion	-0.0559
47	Radiology, Nuclear Medicine & Medical Imaging	-0.0507
48	History & Philosophy Of Science	-0.0495
49	Integrative & Complementary Medicine	-0.0472
50	Pediatrics	-0.0467
51	Endocrinology & Metabolism	-0.0448
52	Engineering, Multidisciplinary	-0.0419
53	History Of Social Sciences	-0.0415
54	Infectious Diseases	-0.0387
55	Mathematics, Interdisciplinary Applications	-0.0366
56	Engineering, Electrical & Electronic	-0.0319

TABLE D.265. List of original attributes (categories) in group of positive on the 5^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC5)		
No.	**Attribute**	**Component Coefficient**
1	Ecology	0.2341
2	Biodiversity Conservation	0.2128
3	Environmental Sciences	0.2122
4	Marine & Freshwater Biology	0.2103
5	Geosciences, Multidisciplinary	0.2101
6	Limnology	0.2056
7	Geography, Physical	0.1976
8	Oceanography	0.1783
9	Water Resources	0.1724
10	Zoology	0.1493
11	Paleontology	0.1397
12	Forestry	0.1384
13	Agronomy	0.1314
14	Evolutionary Biology	0.1306
15	Geology	0.1298
16	Environmental Studies	0.1263
17	Agriculture, Multidisciplinary	0.1263
18	Soil Science	0.1261
19	Meteorology & Atmospheric Sciences	0.1226
20	Plant Sciences	0.1203
21	Ornithology	0.1189
22	Engineering, Environmental	0.1185
23	Geochemistry & Geophysics	0.1127
24	Fisheries	0.1028
25	Entomology	0.1019
26	Horticulture	0.0994
27	Geography	0.0923
28	Planning & Development	0.0888
29	Agricultural Economics & Policy	0.0876
30	Green & Sustainable Science & Technology	0.0835
31	Biology	0.0835
32	Agricultural Engineering	0.0830

No.	Attribute	Component Coefficient
33	Urban Studies	0.0820
34	Mycology	0.0784
35	Economics	0.0761
36	Engineering, Geological	0.0698
37	Public Administration	0.0661
38	International Relations	0.0635
39	Mineralogy	0.0610
40	Business, Finance	0.0595
41	Political Science	0.0518
42	Engineering, Ocean	0.0501
43	Area Studies	0.0501
44	Mining & Mineral Processing	0.0475
45	Business	0.0442
46	Social Sciences, Mathematical Methods	0.0441
47	Engineering, Petroleum	0.0433
48	Archaeology	0.0395
49	Remote Sensing	0.0393
50	Surgery	0.0389
51	Engineering, Chemical	0.0385
52	Anthropology	0.0379
53	Radiology, Nuclear Medicine & Medical Imaging	0.0369
54	Parasitology	0.0357
55	Management	0.0351

TABLE D.266. List of original attributes (categories) in group of zero on the 5^{th} principal component. Categories with positive component coefficients are sorted by values of their component coefficients in descending order. Categories with negative component coefficients are sorted by absolute values of their component coefficients in descending order. That is, categories in two directions are sorted from the greatest contribution to lowest contribution to the component.

Zero (PC5)		
No.	Attribute	Component Coefficient
1	Engineering, Marine	0.0307
2	History Of Social Sciences	0.0301
3	Genetics & Heredity	0.0294
4	Medical Laboratory Technology	0.0290

No.	Attribute	Component Coefficient
5	Gastroenterology & Hepatology	0.0286
6	Industrial Relations & Labor	0.0258
7	Tropical Medicine	0.0257
8	Physics, Multidisciplinary	0.0251
9	Physics, Applied	0.0245
10	Engineering, Biomedical	0.0243
11	Emergency Medicine	0.0236
12	Materials Science, Multidisciplinary	0.0232
13	Demography	0.0227
14	Cardiac & Cardiovascular Systems	0.0226
15	Engineering, Civil	0.0215
16	Respiratory System	0.0210
17	Anesthesiology	0.0209
18	Materials Science, Paper & Wood	0.0209
19	Microbiology	0.0198
20	Critical Care Medicine	0.0194
21	Orthopedics	0.0193
22	Food Science & Technology	0.0183
23	Physics, Condensed Matter	0.0182
24	Agriculture, Dairy & Animal Science	0.0179
25	History	0.0172
26	Materials Science, Characterization & Testing	0.0172
27	Energy & Fuels	0.0170
28	Law	0.0170
29	Nanoscience & Nanotechnology	0.0169
30	Urology & Nephrology	0.0168
31	Dermatology	0.0154
32	Rheumatology	0.0153
33	Ethnic Studies	0.0152
34	Otorhinolaryngology	0.0150
35	Materials Science, Composites	0.0147
36	Sociology	0.0146
37	Transportation	0.0143
38	Materials Science, Coatings & Films	0.0140
39	Architecture	0.0139
40	Physics, Fluids & Plasmas	0.0137
41	Biotechnology & Applied Microbiology	0.0136

No.	Attribute	Component Coefficient
42	Social Issues	0.0128
43	Oncology	0.0127
44	Metallurgy & Metallurgical Engineering	0.0127
45	Clinical Neurology	0.0122
46	Materials Science, Textiles	0.0122
47	Chemistry, Physical	0.0120
48	Medicine, General & Internal	0.0113
49	Astronomy & Astrophysics	0.0106
50	Microscopy	0.0102
51	Optics	0.0100
52	Physics, Atomic, Molecular & Chemical	0.0098
53	Construction & Building Technology	0.0096
54	Transplantation	0.0096
55	Peripheral Vascular Disease	0.0085
56	Electrochemistry	0.0085
57	Nuclear Science & Technology	0.0084
58	Physics, Particles & Fields	0.0083
59	Cultural Studies	0.0079
60	Hospitality, Leisure, Sport & Tourism	0.0077
61	Thermodynamics	0.0072
62	Hematology	0.0071
63	Infectious Diseases	0.0071
64	Physics, Nuclear	0.0069
65	Physics, Mathematical	0.0058
66	Dentistry, Oral Surgery & Medicine	0.0052
67	Veterinary Sciences	0.0049
68	Asian Studies	0.0047
69	Materials Science, Ceramics	0.0045
70	Pathology	0.0045
71	Social Sciences, Interdisciplinary	0.0041
72	Ophthalmology	0.0036
73	Polymer Science	0.0031
74	Chemistry, Applied	0.0018
75	Allergy	0.0014
76	Virology	0.0007
77	Literature, German, Dutch, Scandinavian	-0.0301
78	Sport Sciences	-0.0292

No.	Attribute	Component Coefficient
79	Literature, Slavic	-0.0280
80	Mathematical & Computational Biology	-0.0271
81	Psychology, Mathematical	-0.0270
82	Chemistry, Medicinal	-0.0259
83	Nutrition & Dietetics	-0.0243
84	Literature, African, Australian, Canadian	-0.0243
85	Literature, Romance	-0.0239
86	Classics	-0.0222
87	Nursing	-0.0221
88	Imaging Science & Photographic Technology	-0.0208
89	Pediatrics	-0.0208
90	Criminology & Penology	-0.0207
91	Information Science & Library Science	-0.0199
92	Instruments & Instrumentation	-0.0199
93	Mathematics, Applied	-0.0195
94	Engineering, Aerospace	-0.0188
95	Medical Informatics	-0.0182
96	Integrative & Complementary Medicine	-0.0177
97	Immunology	-0.0176
98	Developmental Biology	-0.0175
99	Women's Studies	-0.0175
100	Engineering, Mechanical	-0.0172
101	Chemistry, Organic	-0.0172
102	Health Policy & Services	-0.0170
103	Engineering, Manufacturing	-0.0170
104	Toxicology	-0.0166
105	Medieval & Renaissance Studies	-0.0158
106	Reproductive Biology	-0.0157
107	Logic	-0.0153
108	Medicine, Research & Experimental	-0.0144
109	Acoustics	-0.0142
110	Cell & Tissue Engineering	-0.0141
111	Andrology	-0.0140
112	Theater	-0.0136
113	Philosophy	-0.0131
114	Music	-0.0121
115	Multidisciplinary Sciences	-0.0118

No.	Attribute	Component Coefficient
116	Biochemical Research Methods	-0.0117
117	Chemistry, Inorganic & Nuclear	-0.0116
118	Psychology, Psychoanalysis	-0.0112
119	History & Philosophy Of Science	-0.0111
120	Statistics & Probability	-0.0107
121	Chemistry, Multidisciplinary	-0.0106
122	Mechanics	-0.0102
123	Art	-0.0100
124	Humanities, Multidisciplinary	-0.0099
125	Religion	-0.0090
126	Obstetrics & Gynecology	-0.0090
127	Dance	-0.0089
128	Medicine, Legal	-0.0086
129	Medical Ethics	-0.0079
130	Neuroimaging	-0.0078
131	Anatomy & Morphology	-0.0067
132	Spectroscopy	-0.0054
133	Ethics	-0.0054
134	Mathematics	-0.0053
135	Folklore	-0.0049
136	Health Care Sciences & Services	-0.0048
137	Materials Science, Biomaterials	-0.0040
138	Chemistry, Analytical	-0.0037
139	Crystallography	-0.0032
140	Film, Radio, Television	-0.0031
141	Primary Health Care	-0.0020
142	Communication	-0.0002

Table D.267. List of original attributes (categories) in group of negative on the 5^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC5)		
No.	Attribute	Component Coefficient
1	Computer Science, Interdisciplinary Applications	-0.1071

No.	Attribute	Component Coefficient
2	Computer Science, Artificial Intelligence	-0.1066
3	Computer Science, Information Systems	-0.1022
4	Computer Science, Theory & Methods	-0.1016
5	Computer Science, Cybernetics	-0.0991
6	Psychology	-0.0978
7	Computer Science, Software Engineering	-0.0934
8	Computer Science, Hardware & Architecture	-0.0922
9	Automation & Control Systems	-0.0921
10	Psychology, Clinical	-0.0865
11	Psychology, Multidisciplinary	-0.0830
12	Engineering, Electrical & Electronic	-0.0817
13	Psychology, Developmental	-0.0813
14	Telecommunications	-0.0807
15	Psychology, Experimental	-0.0757
16	Engineering, Multidisciplinary	-0.0717
17	Psychology, Educational	-0.0651
18	Education, Special	-0.0650
19	Psychiatry	-0.0617
20	Behavioral Sciences	-0.0608
21	Family Studies	-0.0602
22	Psychology, Social	-0.0597
23	Psychology, Biological	-0.0590
24	Gerontology	-0.0549
25	Robotics	-0.0536
26	Physiology	-0.0532
27	Geriatrics & Gerontology	-0.0524
28	Mathematics, Interdisciplinary Applications	-0.0515
29	Rehabilitation	-0.0511
30	Operations Research & Management Science	-0.0500
31	Substance Abuse	-0.0480
32	Neurosciences	-0.0445
33	Literary Theory & Criticism	-0.0419
34	Public, Environmental & Occupational Health	-0.0417
35	Education, Scientific Disciplines	-0.0396
36	Audiology & Speech-Language Pathology	-0.0395
37	Linguistics	-0.0393
38	Literature, British Isles	-0.0379

No.	Attribute	Component Coefficient
39	Education & Educational Research	-0.0378
40	Engineering, Industrial	-0.0377
41	Transportation Science & Technology	-0.0375
42	Language & Linguistics	-0.0375
43	Social Work	-0.0366
44	Literature, American	-0.0363
45	Ergonomics	-0.0354
46	Social Sciences, Biomedical	-0.0349
47	Psychology, Applied	-0.0346
48	Literature	-0.0341
49	Biophysics	-0.0339
50	Pharmacology & Pharmacy	-0.0328
51	Biochemistry & Molecular Biology	-0.0328
52	Poetry	-0.0326
53	Cell Biology	-0.0325
54	Literary Reviews	-0.0322
55	Endocrinology & Metabolism	-0.0319

TABLE D.268. List of original attributes (categories) in group of positive on the 6^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC6)		
No.	Attribute	Component Coefficient
1	Psychology	0.1652
2	Ecology	0.1549
3	Marine & Freshwater Biology	0.1460
4	Psychology, Experimental	0.1365
5	Zoology	0.1343
6	Psychology, Clinical	0.1327
7	Psychology, Multidisciplinary	0.1308
8	Behavioral Sciences	0.1294
9	Biodiversity Conservation	0.1275
10	Psychology, Biological	0.1245
11	Literary Theory & Criticism	0.1176
12	Psychology, Developmental	0.1138

No.	Attribute	Component Coefficient
13	Literature	0.1104
14	Limnology	0.1090
15	Evolutionary Biology	0.1075
16	Education, Special	0.1025
17	Literature, American	0.1016
18	Literature, British Isles	0.1007
19	Ornithology	0.0991
20	Psychiatry	0.0973
21	Psychology, Educational	0.0941
22	Psychology, Social	0.0936
23	Literature, Romance	0.0931
24	Poetry	0.0926
25	Entomology	0.0908
26	Oceanography	0.0896
27	Paleontology	0.0882
28	Literature, African, Australian, Canadian	0.0876
29	Literature, German, Dutch, Scandinavian	0.0854
30	Rehabilitation	0.0845
31	Plant Sciences	0.0845
32	Geosciences, Multidisciplinary	0.0830
33	Literary Reviews	0.0802
34	Geography, Physical	0.0798
35	Classics	0.0791
36	Literature, Slavic	0.0788
37	Gerontology	0.0784
38	Agronomy	0.0776
39	Linguistics	0.0775
40	Medieval & Renaissance Studies	0.0773
41	Family Studies	0.0767
42	Forestry	0.0760
43	Language & Linguistics	0.0759
44	Environmental Sciences	0.0743
45	Fisheries	0.0724
46	Geriatrics & Gerontology	0.0687
47	Biology	0.0684
48	Mycology	0.0674
49	Audiology & Speech-Language Pathology	0.0669

No.	Attribute	Component Coefficient
50	Geology	0.0669
51	Soil Science	0.0646
52	Water Resources	0.0637
53	Agriculture, Multidisciplinary	0.0622
54	Psychology, Mathematical	0.0588
55	Horticulture	0.0570
56	Substance Abuse	0.0561
57	Neurosciences	0.0502
58	Meteorology & Atmospheric Sciences	0.0474
59	Geochemistry & Geophysics	0.0430
60	Humanities, Multidisciplinary	0.0410
61	Neuroimaging	0.0393
62	Art	0.0392
63	Computer Science, Artificial Intelligence	0.0387
64	Education & Educational Research	0.0373
65	Theater	0.0372
66	Education, Scientific Disciplines	0.0369
67	Asian Studies	0.0360
68	Public, Environmental & Occupational Health	0.0350
69	Folklore	0.0349
70	Computer Science, Cybernetics	0.0345
71	Nutrition & Dietetics	0.0342
72	Sport Sciences	0.0337
73	Agricultural Engineering	0.0326
74	Religion	0.0321
75	Social Sciences, Biomedical	0.0317
76	Ergonomics	0.0315

Table D.269. List of original attributes (categories) in group of negative on the 6^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC6)		
No.	**Attribute**	**Component Coefficient**
1	Economics	-0.2028

No.	Attribute	Component Coefficient
2	Planning & Development	-0.1995
3	Business, Finance	-0.1672
4	Environmental Studies	-0.1599
5	Public Administration	-0.1541
6	Business	-0.1538
7	Management	-0.1420
8	Social Sciences, Mathematical Methods	-0.1284
9	International Relations	-0.1257
10	Urban Studies	-0.1174
11	Agricultural Economics & Policy	-0.1164
12	Geography	-0.1120
13	Political Science	-0.1068
14	Medicine, Research & Experimental	-0.0884
15	Industrial Relations & Labor	-0.0882
16	Area Studies	-0.0802
17	Hematology	-0.0761
18	Cell Biology	-0.0703
19	Oncology	-0.0697
20	Pathology	-0.0685
21	Gastroenterology & Hepatology	-0.0677
22	Immunology	-0.0616
23	Surgery	-0.0608
24	Biophysics	-0.0593
25	Transportation	-0.0585
26	Biochemistry & Molecular Biology	-0.0574
27	Social Issues	-0.0573
28	Medical Laboratory Technology	-0.0561
29	Law	-0.0556
30	Hospitality, Leisure, Sport & Tourism	-0.0527
31	Sociology	-0.0514
32	Respiratory System	-0.0511
33	Green & Sustainable Science & Technology	-0.0507
34	Peripheral Vascular Disease	-0.0499
35	Social Sciences, Interdisciplinary	-0.0496
36	Demography	-0.0492
37	Cardiac & Cardiovascular Systems	-0.0490
38	Urology & Nephrology	-0.0485

No.	Attribute	Component Coefficient
39	Cell & Tissue Engineering	-0.0485
40	Dermatology	-0.0465
41	Engineering, Industrial	-0.0463
42	Operations Research & Management Science	-0.0454
43	Transplantation	-0.0443
44	Pharmacology & Pharmacy	-0.0434
45	Emergency Medicine	-0.0400
46	Developmental Biology	-0.0395
47	Critical Care Medicine	-0.0379
48	History Of Social Sciences	-0.0377
49	Information Science & Library Science	-0.0366
50	Materials Science, Biomaterials	-0.0359
51	Multidisciplinary Sciences	-0.0357
52	Anesthesiology	-0.0351
53	Rheumatology	-0.0324
54	Orthopedics	-0.0318

Table D.270. List of original attributes (categories) in group of positive on the 7^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC7)		
No.	**Attribute**	**Component Coefficient**
1	Cell Biology	0.2339
2	Multidisciplinary Sciences	0.2313
3	Biochemistry & Molecular Biology	0.1986
4	Developmental Biology	0.1826
5	Medicine, Research & Experimental	0.1784
6	Cell & Tissue Engineering	0.1582
7	Immunology	0.1435
8	Biophysics	0.1386
9	Genetics & Heredity	0.1238
10	Physiology	0.1219
11	Biotechnology & Applied Microbiology	0.1180
12	Virology	0.1080
13	Pathology	0.1068

No.	Attribute	Component Coefficient
14	Materials Science, Characterization & Testing	0.1001
15	Oncology	0.0976
16	Biology	0.0958
17	Mechanics	0.0957
18	Hematology	0.0953
19	Physics, Multidisciplinary	0.0948
20	Microbiology	0.0912
21	Engineering, Mechanical	0.0858
22	Pharmacology & Pharmacy	0.0793
23	Neurosciences	0.0788
24	Construction & Building Technology	0.0680
25	Physics, Mathematical	0.0661
26	Endocrinology & Metabolism	0.0657
27	Engineering, Civil	0.0641
28	Parasitology	0.0605
29	Physics, Fluids & Plasmas	0.0598
30	Toxicology	0.0595
31	Physics, Nuclear	0.0590
32	Reproductive Biology	0.0562
33	Thermodynamics	0.0553
34	Infectious Diseases	0.0536
35	Physics, Particles & Fields	0.0531
36	Energy & Fuels	0.0481
37	Psychology	0.0481
38	Psychology, Multidisciplinary	0.0475
39	Behavioral Sciences	0.0469
40	Materials Science, Biomaterials	0.0467
41	Engineering, Geological	0.0456
42	Integrative & Complementary Medicine	0.0454
43	Nuclear Science & Technology	0.0452
44	Mining & Mineral Processing	0.0444
45	Psychology, Clinical	0.0431
46	Education, Special	0.0404
47	Materials Science, Composites	0.0401
48	Psychology, Experimental	0.0384
49	Psychology, Educational	0.0377
50	Optics	0.0374

No.	Attribute	Component Coefficient
51	Tropical Medicine	0.0371
52	Psychology, Social	0.0369
53	Rehabilitation	0.0366
54	Psychology, Biological	0.0351
55	Biochemical Research Methods	0.0337
56	Psychology, Developmental	0.0335
57	Medical Laboratory Technology	0.0334
58	Literary Theory & Criticism	0.0333
59	Literature	0.0329
60	Anatomy & Morphology	0.0328
61	Astronomy & Astrophysics	0.0318
62	Engineering, Petroleum	0.0317

TABLE D.271. List of original attributes (categories) in group of negative on the 7^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC7)		
No.	**Attribute**	**Component Coefficient**
1	Computer Science, Information Systems	-0.1967
2	Computer Science, Artificial Intelligence	-0.1877
3	Computer Science, Interdisciplinary Applications	-0.1855
4	Computer Science, Theory & Methods	-0.1831
5	Computer Science, Hardware & Architecture	-0.1729
6	Computer Science, Software Engineering	-0.1705
7	Computer Science, Cybernetics	-0.1646
8	Telecommunications	-0.1446
9	Operations Research & Management Science	-0.1274
10	Automation & Control Systems	-0.1236
11	Chemistry, Multidisciplinary	-0.1137
12	Engineering, Electrical & Electronic	-0.1082
13	Chemistry, Applied	-0.1059
14	Medical Informatics	-0.1012
15	Imaging Science & Photographic Technology	-0.0976
16	Mathematical & Computational Biology	-0.0928

No.	Attribute	Component Coefficient
17	Transportation Science & Technology	-0.0897
18	Robotics	-0.0799
19	Chemistry, Organic	-0.0770
20	Chemistry, Inorganic & Nuclear	-0.0674
21	Chemistry, Physical	-0.0668
22	Planning & Development	-0.0667
23	Environmental Studies	-0.0663
24	Engineering, Industrial	-0.0655
25	Food Science & Technology	-0.0622
26	Economics	-0.0594
27	Information Science & Library Science	-0.0579
28	Agricultural Economics & Policy	-0.0576
29	Agriculture, Multidisciplinary	-0.0565
30	Transportation	-0.0524
31	Business, Finance	-0.0523
32	Polymer Science	-0.0521
33	Management	-0.0504
34	Business	-0.0504
35	Crystallography	-0.0500
36	Nanoscience & Nanotechnology	-0.0494
37	Mathematics, Interdisciplinary Applications	-0.0485
38	Public Administration	-0.0471
39	Remote Sensing	-0.0471
40	Materials Science, Coatings & Films	-0.0458
41	Geography	-0.0440
42	Urban Studies	-0.0429
43	Statistics & Probability	-0.0421
44	Marine & Freshwater Biology	-0.0407
45	Environmental Sciences	-0.0403
46	Social Sciences, Mathematical Methods	-0.0398
47	International Relations	-0.0394
48	Agronomy	-0.0376
49	Materials Science, Ceramics	-0.0373
50	Chemistry, Analytical	-0.0342
51	Spectroscopy	-0.0341
52	Limnology	-0.0315

TABLE D.272. List of original attributes (categories) in group of positive on the 8^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC8)		
No.	**Attribute**	**Component Coefficient**
1	Materials Science, Coatings & Films	0.2693
2	Chemistry, Physical	0.2641
3	Materials Science, Multidisciplinary	0.2636
4	Nanoscience & Nanotechnology	0.2627
5	Physics, Condensed Matter	0.2369
6	Materials Science, Ceramics	0.2214
7	Chemistry, Multidisciplinary	0.1869
8	Physics, Applied	0.1779
9	Polymer Science	0.1733
10	Electrochemistry	0.1727
11	Metallurgy & Metallurgical Engineering	0.1637
12	Chemistry, Inorganic & Nuclear	0.1479
13	Crystallography	0.1437
14	Materials Science, Textiles	0.1260
15	Engineering, Chemical	0.1238
16	Materials Science, Composites	0.1237
17	Microscopy	0.1209
18	Materials Science, Biomaterials	0.1104
19	Spectroscopy	0.1057
20	Chemistry, Applied	0.1039
21	Physics, Atomic, Molecular & Chemical	0.0962
22	Materials Science, Characterization & Testing	0.0801
23	Mineralogy	0.0796
24	Materials Science, Paper & Wood	0.0723
25	Literary Theory & Criticism	0.0643
26	Chemistry, Analytical	0.0635
27	Mathematical & Computational Biology	0.0634
28	Statistics & Probability	0.0580
29	Chemistry, Organic	0.0536
30	Literature	0.0493
31	Language & Linguistics	0.0456
32	Linguistics	0.0438

No.	Attribute	Component Coefficient
33	Psychology, Mathematical	0.0435
34	Social Sciences, Mathematical Methods	0.0432
35	Poetry	0.0431
36	Literature, British Isles	0.0427
37	Classics	0.0412
38	Literature, Romance	0.0387
39	Literature, American	0.0379
40	Medieval & Renaissance Studies	0.0368
41	Literature, German, Dutch, Scandinavian	0.0360
42	Logic	0.0358
43	Literature, Slavic	0.0326
44	Genetics & Heredity	0.0321
45	Thermodynamics	0.0315

Table D.273. List of original attributes (categories) in group of negative on the 8^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC8)		
No.	Attribute	Component Coefficient
1	Social Work	-0.0908
2	Health Policy & Services	-0.0897
3	Engineering, Electrical & Electronic	-0.0865
4	Health Care Sciences & Services	-0.0851
5	Primary Health Care	-0.0836
6	Engineering, Multidisciplinary	-0.0835
7	Social Sciences, Biomedical	-0.0830
8	Family Studies	-0.0761
9	Public, Environmental & Occupational Health	-0.0750
10	Medicine, General & Internal	-0.0734
11	Automation & Control Systems	-0.0666
12	Critical Care Medicine	-0.0656
13	Gerontology	-0.0653
14	Psychology, Clinical	-0.0650
15	Nursing	-0.0650

No.	Attribute	Component Coefficient
16	Planning & Development	-0.0648
17	Psychiatry	-0.0644
18	Environmental Studies	-0.0627
19	Geography	-0.0613
20	Sociology	-0.0582
21	Social Issues	-0.0577
22	Pediatrics	-0.0576
23	Emergency Medicine	-0.0573
24	Psychology, Developmental	-0.0572
25	Engineering, Industrial	-0.0567
26	Geriatrics & Gerontology	-0.0552
27	Public Administration	-0.0544
28	Anesthesiology	-0.0542
29	Clinical Neurology	-0.0538
30	Surgery	-0.0529
31	Respiratory System	-0.0516
32	Demography	-0.0512
33	International Relations	-0.0507
34	Political Science	-0.0502
35	Urban Studies	-0.0501
36	Marine & Freshwater Biology	-0.0499
37	Computer Science, Hardware & Architecture	-0.0487
38	Ethnic Studies	-0.0482
39	Telecommunications	-0.0478
40	Substance Abuse	-0.0477
41	Peripheral Vascular Disease	-0.0472
42	Otorhinolaryngology	-0.0455
43	Rehabilitation	-0.0452
44	Area Studies	-0.0449
45	Ecology	-0.0448
46	Cardiac & Cardiovascular Systems	-0.0445
47	Urology & Nephrology	-0.0444
48	Orthopedics	-0.0443
49	Psychology, Social	-0.0437
50	Engineering, Aerospace	-0.0431
51	Industrial Relations & Labor	-0.0430
52	Zoology	-0.0416

No.	Attribute	Component Coefficient
53	Engineering, Mechanical	-0.0406
54	Gastroenterology & Hepatology	-0.0404
55	Rheumatology	-0.0403
56	Medical Ethics	-0.0402
57	Women's Studies	-0.0395
58	Obstetrics & Gynecology	-0.0394
59	Entomology	-0.0384
60	Biodiversity Conservation	-0.0383
61	Allergy	-0.0371
62	Mechanics	-0.0363
63	Criminology & Penology	-0.0324
64	Dermatology	-0.0321
65	Mathematics, Interdisciplinary Applications	-0.0318
66	Transportation Science & Technology	-0.0316

Table D.274. List of original attributes (categories) in group of positive on the 9^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC9)		
No.	Attribute	Component Coefficient
1	Geosciences, Multidisciplinary	0.2756
2	Geochemistry & Geophysics	0.2152
3	Geology	0.2057
4	Water Resources	0.1901
5	Geography, Physical	0.1767
6	Oceanography	0.1379
7	Mineralogy	0.1360
8	Meteorology & Atmospheric Sciences	0.1345
9	Environmental Sciences	0.1321
10	Engineering, Environmental	0.1310
11	Engineering, Ocean	0.1294
12	Engineering, Geological	0.1213
13	Limnology	0.1206
14	Mining & Mineral Processing	0.1080
15	Engineering, Marine	0.1064

No.	Attribute	Component Coefficient
16	Engineering, Petroleum	0.1002
17	Engineering, Civil	0.0783
18	Remote Sensing	0.0767
19	Psychology, Experimental	0.0657
20	Engineering, Chemical	0.0616
21	Psychology, Educational	0.0609
22	Psychology	0.0596
23	Pharmacology & Pharmacy	0.0595
24	Psychology, Multidisciplinary	0.0536
25	Green & Sustainable Science & Technology	0.0533
26	Thermodynamics	0.0526
27	Paleontology	0.0518
28	Psychology, Mathematical	0.0516
29	Neurosciences	0.0512
30	Soil Science	0.0512
31	Engineering, Mechanical	0.0474
32	Archaeology	0.0463
33	Physics, Fluids & Plasmas	0.0462
34	Imaging Science & Photographic Technology	0.0459
35	Neuroimaging	0.0452
36	Education & Educational Research	0.0444
37	Education, Special	0.0443
38	Construction & Building Technology	0.0423
39	Mechanics	0.0408
40	Chemistry, Medicinal	0.0405
41	Cell Biology	0.0385
42	Psychology, Social	0.0378
43	Engineering, Aerospace	0.0378
44	Linguistics	0.0378
45	Psychology, Developmental	0.0367
46	Education, Scientific Disciplines	0.0355
47	Medicine, Research & Experimental	0.0350
48	Oncology	0.0347
49	Biophysics	0.0343
50	Language & Linguistics	0.0340
51	Engineering, Biomedical	0.0333
52	Psychology, Clinical	0.0330

No.	Attribute	Component Coefficient
53	Audiology & Speech-Language Pathology	0.0323
54	Chemistry, Applied	0.0319

TABLE D.275. List of original attributes (categories) in group of negative on the 9^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC9)		
No.	**Attribute**	**Component Coefficient**
1	Zoology	-0.2252
2	Evolutionary Biology	-0.2095
3	Biology	-0.1865
4	Entomology	-0.1826
5	Mycology	-0.1805
6	Biodiversity Conservation	-0.1613
7	Parasitology	-0.1577
8	Plant Sciences	-0.1575
9	Ecology	-0.1550
10	Ornithology	-0.1430
11	Tropical Medicine	-0.1297
12	Microbiology	-0.1226
13	Infectious Diseases	-0.1096
14	Genetics & Heredity	-0.1053
15	Veterinary Sciences	-0.0989
16	Horticulture	-0.0970
17	Physics, Applied	-0.0831
18	Agronomy	-0.0778
19	Materials Science, Multidisciplinary	-0.0767
20	Physics, Condensed Matter	-0.0765
21	Virology	-0.0741
22	Fisheries	-0.0715
23	Forestry	-0.0705
24	Nanoscience & Nanotechnology	-0.0695
25	Public, Environmental & Occupational Health	-0.0650
26	Agriculture, Dairy & Animal Science	-0.0564

No.	Attribute	Component Coefficient
27	Health Policy & Services	-0.0562
28	Marine & Freshwater Biology	-0.0560
29	Health Care Sciences & Services	-0.0548
30	Multidisciplinary Sciences	-0.0548
31	Biotechnology & Applied Microbiology	-0.0511
32	Agriculture, Multidisciplinary	-0.0504
33	Primary Health Care	-0.0499
34	Materials Science, Coatings & Films	-0.0495
35	Social Sciences, Biomedical	-0.0446
36	Medicine, General & Internal	-0.0419
37	Chemistry, Physical	-0.0400
38	Materials Science, Ceramics	-0.0399
39	Optics	-0.0342
40	Emergency Medicine	-0.0328
41	Critical Care Medicine	-0.0326
42	Obstetrics & Gynecology	-0.0325
43	Physics, Atomic, Molecular & Chemical	-0.0322

Table D.276. List of original attributes (categories) in group of positive on the 10^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC10)		
No.	Attribute	Component Coefficient
1	Agriculture, Multidisciplinary	0.3288
2	Agronomy	0.2928
3	Agricultural Engineering	0.2720
4	Horticulture	0.2497
5	Food Science & Technology	0.2079
6	Plant Sciences	0.1762
7	Soil Science	0.1735
8	Engineering, Environmental	0.1523
9	Chemistry, Applied	0.1429
10	Green & Sustainable Science & Technology	0.1343
11	Agricultural Economics & Policy	0.1335
12	Environmental Sciences	0.1178

No.	Attribute	Component Coefficient
13	Agriculture, Dairy & Animal Science	0.1068
14	Engineering, Chemical	0.1004
15	Biotechnology & Applied Microbiology	0.0938
16	Psychology, Experimental	0.0799
17	Nutrition & Dietetics	0.0674
18	Chemistry, Analytical	0.0659
19	Psychology	0.0653
20	Linguistics	0.0646
21	Language & Linguistics	0.0625
22	Energy & Fuels	0.0605
23	Neurosciences	0.0563
24	Psychology, Biological	0.0553
25	Psychology, Educational	0.0552
26	Neuroimaging	0.0549
27	Audiology & Speech-Language Pathology	0.0536
28	Behavioral Sciences	0.0536
29	Materials Science, Paper & Wood	0.0522
30	Economics	0.0507
31	Psychology, Mathematical	0.0503
32	Business	0.0500
33	Business, Finance	0.0493
34	Forestry	0.0469
35	Integrative & Complementary Medicine	0.0442
36	Literary Theory & Criticism	0.0438
37	Clinical Neurology	0.0431
38	Management	0.0424
39	Biochemical Research Methods	0.0421
40	Water Resources	0.0401
41	Toxicology	0.0381
42	Education, Special	0.0377
43	Education & Educational Research	0.0372
44	Physics, Particles & Fields	0.0369
45	Surgery	0.0353
46	Orthopedics	0.0352
47	Chemistry, Medicinal	0.0344
48	Psychology, Multidisciplinary	0.0342
49	Social Sciences, Mathematical Methods	0.0320

Table D.277. List of original attributes (categories) in group of negative on the 10^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC10)		
No.	**Attribute**	**Component Coefficient**
1	Geology	-0.1763
2	Paleontology	-0.1562
3	Geosciences, Multidisciplinary	-0.1480
4	Geochemistry & Geophysics	-0.1418
5	Geography, Physical	-0.1236
6	Oceanography	-0.1195
7	Marine & Freshwater Biology	-0.1040
8	Zoology	-0.1021
9	Mineralogy	-0.0980
10	Social Sciences, Biomedical	-0.0953
11	Public, Environmental & Occupational Health	-0.0943
12	Health Policy & Services	-0.0936
13	Biodiversity Conservation	-0.0857
14	Evolutionary Biology	-0.0799
15	Parasitology	-0.0781
16	Health Care Sciences & Services	-0.0781
17	Tropical Medicine	-0.0745
18	Medical Informatics	-0.0702
19	Primary Health Care	-0.0666
20	Anthropology	-0.0656
21	Infectious Diseases	-0.0639
22	Materials Science, Multidisciplinary	-0.0628
23	Materials Science, Coatings & Films	-0.0601
24	Archaeology	-0.0590
25	Engineering, Ocean	-0.0574
26	Engineering, Marine	-0.0572
27	Meteorology & Atmospheric Sciences	-0.0553
28	Ornithology	-0.0535
29	Medical Ethics	-0.0534
30	Nursing	-0.0531
31	Nanoscience & Nanotechnology	-0.0512

No.	Attribute	Component Coefficient
32	Fisheries	-0.0488
33	Social Issues	-0.0479
34	Materials Science, Characterization & Testing	-0.0475
35	Metallurgy & Metallurgical Engineering	-0.0471
36	Biology	-0.0466
37	Computer Science, Interdisciplinary Applications	-0.0465
38	Virology	-0.0442
39	Ecology	-0.0442
40	Mining & Mineral Processing	-0.0435
41	Computer Science, Information Systems	-0.0422
42	Physics, Condensed Matter	-0.0420
43	Women's Studies	-0.0396
44	Computer Science, Hardware & Architecture	-0.0387
45	Automation & Control Systems	-0.0384
46	Mycology	-0.0378
47	Materials Science, Ceramics	-0.0377
48	Remote Sensing	-0.0375
49	Physics, Applied	-0.0373
50	Limnology	-0.0372
51	Materials Science, Composites	-0.0364
52	Computer Science, Artificial Intelligence	-0.0360
53	Gerontology	-0.0359
54	Computer Science, Theory & Methods	-0.0358
55	Mathematical & Computational Biology	-0.0354
56	Imaging Science & Photographic Technology	-0.0348
57	Ethics	-0.0346
58	Social Work	-0.0335
59	Immunology	-0.0335
60	Multidisciplinary Sciences	-0.0325
61	Engineering, Mechanical	-0.0321
62	Telecommunications	-0.0318

Table D.278. List of original attributes (categories) in group of positive on the 11^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC11)		
No.	**Attribute**	**Component Coefficient**
1	Health Policy & Services	0.2125
2	Health Care Sciences & Services	0.1977
3	Social Sciences, Biomedical	0.1841
4	Primary Health Care	0.1571
5	Medical Informatics	0.1571
6	Nursing	0.1566
7	Public, Environmental & Occupational Health	0.1558
8	Biotechnology & Applied Microbiology	0.1341
9	Medical Ethics	0.1316
10	Agricultural Engineering	0.1183
11	Food Science & Technology	0.1102
12	Agriculture, Multidisciplinary	0.1097
13	Agronomy	0.1074
14	Chemistry, Applied	0.0981
15	Biochemical Research Methods	0.0952
16	Horticulture	0.0934
17	Infectious Diseases	0.0841
18	Microbiology	0.0840
19	Ethics	0.0834
20	Soil Science	0.0807
21	Biochemistry & Molecular Biology	0.0770
22	Social Work	0.0751
23	Virology	0.0717
24	Plant Sciences	0.0710
25	Chemistry, Analytical	0.0668
26	Engineering, Environmental	0.0653
27	Biophysics	0.0650
28	Engineering, Chemical	0.0619
29	Chemistry, Medicinal	0.0611
30	Education, Scientific Disciplines	0.0608
31	Social Issues	0.0579
32	Water Resources	0.0562

No.	Attribute	Component Coefficient
33	Engineering, Geological	0.0554
34	Mathematical & Computational Biology	0.0535
35	Women's Studies	0.0534
36	Environmental Sciences	0.0529
37	Education & Educational Research	0.0522
38	Family Studies	0.0517
39	Genetics & Heredity	0.0504
40	Tropical Medicine	0.0485
41	Engineering, Ocean	0.0484
42	Engineering, Petroleum	0.0480
43	Physics, Mathematical	0.0466
44	Gerontology	0.0458
45	Physics, Fluids & Plasmas	0.0443
46	Information Science & Library Science	0.0431
47	Mechanics	0.0405
48	Parasitology	0.0403
49	Chemistry, Organic	0.0394
50	Spectroscopy	0.0384
51	Medicine, General & Internal	0.0383
52	Cell Biology	0.0382
53	Engineering, Marine	0.0373
54	Geochemistry & Geophysics	0.0357
55	Mineralogy	0.0352
56	Materials Science, Paper & Wood	0.0349
57	Geosciences, Multidisciplinary	0.0331
58	Immunology	0.0331
59	History & Philosophy Of Science	0.0328

Table D.279. List of original attributes (categories) in group of negative on the 11^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC11)		
No.	**Attribute**	**Component Coefficient**
1	Behavioral Sciences	-0.2509

No.	Attribute	Component Coefficient
2	Psychology, Biological	-0.2311
3	Neurosciences	-0.2232
4	Neuroimaging	-0.2048
5	Psychology	-0.1966
6	Psychology, Experimental	-0.1948
7	Clinical Neurology	-0.1311
8	Economics	-0.1061
9	Anatomy & Morphology	-0.1047
10	Zoology	-0.1036
11	Radiology, Nuclear Medicine & Medical Imaging	-0.0971
12	Business, Finance	-0.0961
13	Orthopedics	-0.0820
14	Planning & Development	-0.0799
15	Audiology & Speech-Language Pathology	-0.0784
16	Psychology, Mathematical	-0.0784
17	Ornithology	-0.0760
18	Biodiversity Conservation	-0.0759
19	Surgery	-0.0743
20	Sport Sciences	-0.0740
21	Physiology	-0.0737
22	Ecology	-0.0734
23	Otorhinolaryngology	-0.0721
24	International Relations	-0.0711
25	Environmental Studies	-0.0660
26	Psychiatry	-0.0636
27	Peripheral Vascular Disease	-0.0635
28	Business	-0.0634
29	Political Science	-0.0606
30	Cardiac & Cardiovascular Systems	-0.0597
31	Engineering, Biomedical	-0.0593
32	Ophthalmology	-0.0587
33	Biology	-0.0584
34	Social Sciences, Mathematical Methods	-0.0581
35	Evolutionary Biology	-0.0577
36	Geography	-0.0573
37	Marine & Freshwater Biology	-0.0563
38	Urban Studies	-0.0559

No.	Attribute	Component Coefficient
39	Area Studies	-0.0553
40	Anesthesiology	-0.0499
41	Urology & Nephrology	-0.0498
42	Management	-0.0498
43	Public Administration	-0.0481
44	Dentistry, Oral Surgery & Medicine	-0.0448
45	Entomology	-0.0435
46	Gastroenterology & Hepatology	-0.0426
47	Respiratory System	-0.0413
48	Psychology, Clinical	-0.0412
49	History Of Social Sciences	-0.0396
50	Fisheries	-0.0376
51	Microscopy	-0.0374
52	Rheumatology	-0.0371
53	Psychology, Social	-0.0345
54	Agricultural Economics & Policy	-0.0338
55	Physics, Applied	-0.0335
56	Rehabilitation	-0.0326
57	Psychology, Multidisciplinary	-0.0323

TABLE D.280. List of original attributes (categories) in group of positive on the 12^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC12)		
No.	Attribute	Component Coefficient
1	Education & Educational Research	0.2070
2	Education, Scientific Disciplines	0.1871
3	Psychology, Educational	0.1697
4	Language & Linguistics	0.1529
5	Linguistics	0.1479
6	Management	0.1459
7	Business	0.1344
8	Information Science & Library Science	0.1208
9	Medical Informatics	0.1173
10	Cell Biology	0.1147

No.	Attribute	Component Coefficient
11	Social Sciences, Interdisciplinary	0.1125
12	Literary Theory & Criticism	0.1077
13	Health Care Sciences & Services	0.1038
14	Cell & Tissue Engineering	0.1037
15	Biophysics	0.1027
16	Nursing	0.0925
17	Psychology, Applied	0.0886
18	Primary Health Care	0.0879
19	Biochemistry & Molecular Biology	0.0863
20	Business, Finance	0.0845
21	Poetry	0.0836
22	Developmental Biology	0.0833
23	Education, Special	0.0817
24	Hematology	0.0757
25	Medicine, Research & Experimental	0.0723
26	Oncology	0.0689
27	Pathology	0.0689
28	Emergency Medicine	0.0686
29	Hospitality, Leisure, Sport & Tourism	0.0666
30	Health Policy & Services	0.0648
31	Geography, Physical	0.0629
32	Multidisciplinary Sciences	0.0624
33	Literature	0.0619
34	Surgery	0.0613
35	Economics	0.0593
36	Audiology & Speech-Language Pathology	0.0589
37	Literature, American	0.0585
38	Critical Care Medicine	0.0582
39	Biodiversity Conservation	0.0554
40	Ergonomics	0.0552
41	Otorhinolaryngology	0.0545
42	Literature, Slavic	0.0537
43	Radiology, Nuclear Medicine & Medical Imaging	0.0533
44	Literature, German, Dutch, Scandinavian	0.0519
45	Literature, British Isles	0.0504
46	Literary Reviews	0.0503
47	Limnology	0.0482

No.	**Attribute**	**Component Coefficient**
48	Anesthesiology	0.0469
49	Remote Sensing	0.0441
50	Materials Science, Biomaterials	0.0434
51	Respiratory System	0.0430
52	Literature, African, Australian, Canadian	0.0430
53	Ecology	0.0430
54	Rehabilitation	0.0419
55	Environmental Sciences	0.0416
56	Imaging Science & Photographic Technology	0.0414
57	Biology	0.0409
58	Marine & Freshwater Biology	0.0405
59	Gastroenterology & Hepatology	0.0403
60	Orthopedics	0.0402
61	Computer Science, Software Engineering	0.0399
62	Transplantation	0.0391
63	Humanities, Multidisciplinary	0.0376
64	Computer Science, Information Systems	0.0374
65	Medicine, General & Internal	0.0372
66	Geosciences, Multidisciplinary	0.0353
67	Social Sciences, Mathematical Methods	0.0347
68	Paleontology	0.0347
69	Cardiac & Cardiovascular Systems	0.0343
70	Transportation	0.0334
71	Literature, Romance	0.0330
72	Environmental Studies	0.0327
73	Industrial Relations & Labor	0.0325
74	Computer Science, Cybernetics	0.0321

TABLE D.281. List of original attributes (categories) in group of negative on the 12^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC12)		
No.	**Attribute**	**Component Coefficient**
1	Mechanics	-0.2128

No.	Attribute	Component Coefficient
2	Engineering, Mechanical	-0.1907
3	Engineering, Civil	-0.1439
4	Materials Science, Characterization & Testing	-0.1426
5	Physics, Fluids & Plasmas	-0.1424
6	Engineering, Ocean	-0.1421
7	Engineering, Geological	-0.1305
8	Construction & Building Technology	-0.1292
9	Mathematics, Interdisciplinary Applications	-0.1253
10	Physics, Mathematical	-0.1246
11	Political Science	-0.1235
12	Thermodynamics	-0.1196
13	Engineering, Marine	-0.1195
14	Area Studies	-0.1177
15	Materials Science, Composites	-0.1168
16	International Relations	-0.1058
17	Behavioral Sciences	-0.1044
18	Mathematics, Applied	-0.1025
19	Ethnic Studies	-0.0948
20	History	-0.0830
21	Nutrition & Dietetics	-0.0824
22	Social Issues	-0.0811
23	Law	-0.0787
24	Women's Studies	-0.0749
25	Engineering, Aerospace	-0.0746
26	Psychology, Biological	-0.0733
27	Food Science & Technology	-0.0720
28	Veterinary Sciences	-0.0709
29	Sociology	-0.0701
30	Agriculture, Dairy & Animal Science	-0.0694
31	Psychiatry	-0.0664
32	Demography	-0.0653
33	Agriculture, Multidisciplinary	-0.0627
34	Anthropology	-0.0606
35	Geriatrics & Gerontology	-0.0600
36	Mathematics	-0.0592
37	Psychology, Clinical	-0.0585
38	Metallurgy & Metallurgical Engineering	-0.0583

No.	Attribute	Component Coefficient
39	Cultural Studies	-0.0581
40	Psychology	-0.0562
41	Statistics & Probability	-0.0557
42	Neurosciences	-0.0544
43	Automation & Control Systems	-0.0526
44	Asian Studies	-0.0496
45	Substance Abuse	-0.0484
46	Engineering, Manufacturing	-0.0479
47	Parasitology	-0.0477
48	Infectious Diseases	-0.0476
49	Tropical Medicine	-0.0462
50	Engineering, Petroleum	-0.0460
51	Agronomy	-0.0434
52	Engineering, Multidisciplinary	-0.0431
53	Criminology & Penology	-0.0423
54	History Of Social Sciences	-0.0418
55	Agricultural Engineering	-0.0413
56	Chemistry, Applied	-0.0413
57	Horticulture	-0.0411
58	Gerontology	-0.0405
59	Energy & Fuels	-0.0391
60	Philosophy	-0.0389
61	Engineering, Chemical	-0.0370
62	Family Studies	-0.0350
63	Materials Science, Paper & Wood	-0.0349
64	Public Administration	-0.0335

Table D.282. List of original attributes (categories) in group of positive on the 13^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC13)		
No.	Attribute	Component Coefficient
1	Parasitology	0.2090
2	Infectious Diseases	0.2047
3	Education & Educational Research	0.1851

No.	Attribute	Component Coefficient
4	Tropical Medicine	0.1850
5	Psychology, Educational	0.1814
6	Education, Scientific Disciplines	0.1586
7	Microbiology	0.1520
8	Virology	0.1503
9	Engineering, Civil	0.1463
10	Engineering, Mechanical	0.1452
11	Construction & Building Technology	0.1349
12	Linguistics	0.1344
13	Language & Linguistics	0.1336
14	Materials Science, Characterization & Testing	0.1277
15	Mechanics	0.1191
16	Management	0.1163
17	Business	0.1144
18	Engineering, Ocean	0.1127
19	Social Sciences, Interdisciplinary	0.1088
20	Thermodynamics	0.1070
21	Engineering, Marine	0.1009
22	Education, Special	0.0994
23	Veterinary Sciences	0.0991
24	Materials Science, Composites	0.0938
25	Engineering, Geological	0.0900
26	Mycology	0.0868
27	Psychology, Applied	0.0771
28	Physics, Fluids & Plasmas	0.0768
29	Business, Finance	0.0760
30	Immunology	0.0735
31	Literary Theory & Criticism	0.0735
32	Engineering, Industrial	0.0734
33	Engineering, Manufacturing	0.0698
34	Energy & Fuels	0.0642
35	Engineering, Multidisciplinary	0.0636
36	Information Science & Library Science	0.0613
37	Economics	0.0603
38	Audiology & Speech-Language Pathology	0.0598
39	Gastroenterology & Hepatology	0.0564
40	Hospitality, Leisure, Sport & Tourism	0.0547

No.	Attribute	Component Coefficient
41	Psychology, Multidisciplinary	0.0538
42	Metallurgy & Metallurgical Engineering	0.0503
43	Transportation Science & Technology	0.0496
44	Engineering, Aerospace	0.0476
45	Psychology, Developmental	0.0465
46	Green & Sustainable Science & Technology	0.0445
47	Transportation	0.0431
48	Engineering, Petroleum	0.0425
49	Psychology, Social	0.0425
50	Psychology, Mathematical	0.0416
51	Zoology	0.0400
52	Entomology	0.0398
53	Poetry	0.0396
54	Social Sciences, Mathematical Methods	0.0391
55	Dermatology	0.0390
56	Engineering, Chemical	0.0379
57	Medical Laboratory Technology	0.0373
58	Humanities, Multidisciplinary	0.0368
59	Literature	0.0336
60	Industrial Relations & Labor	0.0325
61	Mining & Mineral Processing	0.0322
62	Acoustics	0.0319

TABLE D.283. List of original attributes (categories) in group of negative on the 13^{th} principal component. Categories are sorted by absolute values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC13)		
No.	Attribute	Component Coefficient
1	Physiology	-0.1588
2	Cell Biology	-0.1469
3	Developmental Biology	-0.1407
4	Biochemistry & Molecular Biology	-0.1145
5	Neurosciences	-0.1113
6	Cell & Tissue Engineering	-0.1080

No.	Attribute	Component Coefficient
7	Health Policy & Services	-0.1061
8	Biophysics	-0.1059
9	Reproductive Biology	-0.0992
10	Endocrinology & Metabolism	-0.0916
11	Gerontology	-0.0904
12	Public, Environmental & Occupational Health	-0.0890
13	Health Care Sciences & Services	-0.0880
14	Toxicology	-0.0871
15	Social Sciences, Biomedical	-0.0869
16	Geriatrics & Gerontology	-0.0847
17	Medical Ethics	-0.0817
18	Neuroimaging	-0.0798
19	Anatomy & Morphology	-0.0783
20	Imaging Science & Photographic Technology	-0.0770
21	Integrative & Complementary Medicine	-0.0759
22	Remote Sensing	-0.0745
23	Multidisciplinary Sciences	-0.0734
24	Nutrition & Dietetics	-0.0732
25	Pharmacology & Pharmacy	-0.0722
26	Medical Informatics	-0.0719
27	Geography, Physical	-0.0717
28	Social Issues	-0.0713
29	Chemistry, Medicinal	-0.0676
30	Materials Science, Biomaterials	-0.0666
31	Engineering, Biomedical	-0.0651
32	Ethics	-0.0633
33	Andrology	-0.0625
34	Biology	-0.0616
35	Obstetrics & Gynecology	-0.0616
36	Medicine, Research & Experimental	-0.0615
37	Primary Health Care	-0.0614
38	Behavioral Sciences	-0.0605
39	Agriculture, Multidisciplinary	-0.0547
40	Political Science	-0.0530
41	Mathematics	-0.0527
42	Area Studies	-0.0521
43	Sport Sciences	-0.0501

No.	Attribute	Component Coefficient
44	Chemistry, Multidisciplinary	-0.0500
45	Agronomy	-0.0494
46	International Relations	-0.0485
47	Philosophy	-0.0477
48	Women's Studies	-0.0465
49	Law	-0.0462
50	History	-0.0454
51	Physics, Applied	-0.0450
52	Anthropology	-0.0450
53	Chemistry, Organic	-0.0443
54	Physics, Multidisciplinary	-0.0436
55	Nursing	-0.0430
56	Nanoscience & Nanotechnology	-0.0418
57	Microscopy	-0.0406
58	Optics	-0.0398
59	Horticulture	-0.0397
60	Crystallography	-0.0393
61	Chemistry, Inorganic & Nuclear	-0.0392
62	Psychiatry	-0.0387
63	Mathematics, Applied	-0.0385
64	Psychology, Biological	-0.0380
65	Environmental Sciences	-0.0376
66	History & Philosophy Of Science	-0.0363
67	Radiology, Nuclear Medicine & Medical Imaging	-0.0362
68	Soil Science	-0.0360
69	Geosciences, Multidisciplinary	-0.0345
70	Agriculture, Dairy & Animal Science	-0.0339
71	Physics, Condensed Matter	-0.0339

D.4. List of Categories on the Left Wing, Right Wing and the Body of the Bird Shape, Obtained in the First Three PCs, with Projection onto the First Principal Component

Table D.284. List of categories on the left wing (green) with projection onto the first principal component. Categories are sorted by absolute values of their component score in descending order.

Left (Green) Wing		
No.	**Category**	**PC1**
1	Poetry	-0.3080
2	Literature, African, Australian, Canadian	-0.2620
3	Literature, American	-0.2530
4	Literature	-0.2300
5	Literature, German, Dutch, Scandinavian	-0.1950
6	Literary Reviews	-0.1910
7	Theater	-0.1830
8	History	-0.1770
9	Literature, Romance	-0.1700
10	Medieval & Renaissance Studies	-0.1670
11	Cultural Studies	-0.1560
12	Literature, British Isles	-0.1550
13	Asian Studies	-0.1430
14	Dance	-0.1360
15	Area Studies	-0.1310
16	Religion	-0.1300
17	Film, Radio, Television	-0.1280
18	Classics	-0.1270
19	Political Science	-0.1250
20	Folklore	-0.1240
21	Literary Theory & Criticism	-0.1220
22	Literature, Slavic	-0.1200
23	Ethnic Studies	-0.1080
24	Philosophy	-0.1080
25	History Of Social Sciences	-0.1080
26	Music	-0.1010
27	Law	-0.1000
28	International Relations	-0.0990
29	Humanities, Multidisciplinary	-0.0970
30	Art	-0.0860

No.	Category	PC1
31	Sociology	-0.0860
32	Language & Linguistics	-0.0790
33	Communication	-0.0770
34	History & Philosophy Of Science	-0.0760
35	Ethics	-0.0730
36	Linguistics	-0.0680
37	Social Issues	-0.0650
38	Women's Studies	-0.0640
39	Geography	-0.0560
40	Anthropology	-0.0530
41	Public Administration	-0.0510
42	Education & Educational Research	-0.0470
43	Industrial Relations & Labor	-0.0400
44	Medical Ethics	-0.0400
45	Psychology, Psychoanalysis	-0.0390
46	Planning & Development	-0.0390
47	Urban Studies	-0.0380
48	Archaeology	-0.0350
49	Social Sciences, Interdisciplinary	-0.0350
50	Criminology & Penology	-0.0300
51	Demography	-0.0290
52	Social Work	-0.0260
53	Architecture	-0.0240
54	Environmental Studies	-0.0200
55	Hospitality, Leisure, Sport & Tourism	-0.0180
56	Education, Scientific Disciplines	-0.0170
57	Information Science & Library Science	-0.0160
58	Business	-0.0140
59	Social Sciences, Biomedical	-0.0130
60	Management	-0.0090
61	Psychology, Educational	-0.0090
62	Psychology, Applied	-0.0080
63	Psychology, Multidisciplinary	-0.0060
64	Psychology, Social	-0.0060
65	Economics	-0.0060

TABLE D.285. List of categories on the body (blue) with projection onto the first principal component. Categories are sorted by absolute values of their component score in descending order.

Body		
No.	**Category**	**PC1**
1	Agricultural Economics & Policy	-0.0020
2	Mathematical & Computational Biology	0.0000
3	Biochemical Research Methods	0.0000
4	Biotechnology & Applied Microbiology	0.0000
5	Family Studies	0.0000
6	Biophysics	0.0010
7	Psychology, Mathematical	0.0010
8	Business, Finance	0.0020
9	Mycology	0.0020
10	Acoustics	0.0030
11	Transportation	0.0040
12	Engineering, Biomedical	0.0040
13	Multidisciplinary Sciences	0.0040
14	Medicine, Legal	0.0040
15	Biochemistry & Molecular Biology	0.0040
16	Chemistry, Analytical	0.0040
17	Audiology & Speech-Language Pathology	0.0040
18	Biology	0.0050
19	Developmental Biology	0.0050
20	Microbiology	0.0050
21	Psychology, Experimental	0.0050
22	Ergonomics	0.0050
23	Microscopy	0.0050
24	Materials Science, Biomaterials	0.0060
25	Cell Biology	0.0060
26	Toxicology	0.0070
27	Social Sciences, Mathematical Methods	0.0070
28	Chemistry, Medicinal	0.0070
29	Education, Special	0.0080
30	Statistics & Probability	0.0080
31	Logic	0.0080
32	Cell & Tissue Engineering	0.0090
33	Food Science & Technology	0.0090
34	Spectroscopy	0.0100

No.	Category	PC1
35	Psychology, Biological	0.0100
36	Anatomy & Morphology	0.0110
37	Behavioral Sciences	0.0110
38	Physiology	0.0110
39	Genetics & Heredity	0.0120
40	Paleontology	0.0120
41	Agriculture, Multidisciplinary	0.0120
42	Geography, Physical	0.0120
43	Nuclear Science & Technology	0.0130
44	Chemistry, Applied	0.0140
45	Evolutionary Biology	0.0150
46	Environmental Sciences	0.0150
47	Geology	0.0160
48	Zoology	0.0160
49	Agricultural Engineering	0.0160
50	Mining & Mineral Processing	0.0160
51	Agriculture, Dairy & Animal Science	0.0160
52	Astronomy & Astrophysics	0.0160
53	Engineering, Environmental	0.0160
54	Computer Science, Cybernetics	0.0170
55	Engineering, Petroleum	0.0170
56	Imaging Science & Photographic Technology	0.0170
57	Biodiversity Conservation	0.0170
58	Meteorology & Atmospheric Sciences	0.0170
59	Plant Sciences	0.0170
60	Materials Science, Paper & Wood	0.0170
61	Computer Science, Interdisciplinary Applications	0.0170
62	Limnology	0.0170
63	Green & Sustainable Science & Technology	0.0170
64	Geochemistry & Geophysics	0.0170
65	Materials Science, Characterization & Testing	0.0170
66	Forestry	0.0180
67	Engineering, Ocean	0.0180
68	Materials Science, Textiles	0.0180
69	Medical Informatics	0.0190
70	Mineralogy	0.0190
71	Chemistry, Organic	0.0190
72	Ornithology	0.0190

No.	Category	PC1
73	Mathematics	0.0190
74	Water Resources	0.0190
75	Horticulture	0.0190
76	Engineering, Marine	0.0190
77	Agronomy	0.0190
78	Remote Sensing	0.0190
79	Instruments & Instrumentation	0.0200
80	Engineering, Industrial	0.0200
81	Psychology	0.0200
82	Soil Science	0.0200
83	Engineering, Geological	0.0200
84	Physics, Nuclear	0.0200
85	Engineering, Aerospace	0.0200
86	Materials Science, Composites	0.0210
87	Entomology	0.0210
88	Fisheries	0.0210
89	Physics, Fluids & Plasmas	0.0210
90	Veterinary Sciences	0.0210
91	Operations Research & Management Science	0.0210
92	Oceanography	0.0210
93	Marine & Freshwater Biology	0.0210
94	Thermodynamics	0.0210
95	Ecology	0.0210
96	Geosciences, Multidisciplinary	0.0220
97	Construction & Building Technology	0.0220
98	Pharmacology & Pharmacy	0.0220
99	Electrochemistry	0.0220
100	Physics, Mathematical	0.0220
101	Physics, Particles & Fields	0.0220
102	Mathematics, Interdisciplinary Applications	0.0230
103	Virology	0.0230
104	Nanoscience & Nanotechnology	0.0230
105	Computer Science, Software Engineering	0.0240
106	Transportation Science & Technology	0.0240
107	Materials Science, Coatings & Films	0.0240
108	Robotics	0.0240
109	Engineering, Manufacturing	0.0240
110	Engineering, Chemical	0.0240

No.	Category	PC1
111	Materials Science, Ceramics	0.0240
112	Engineering, Civil	0.0240
113	Physics, Atomic, Molecular & Chemical	0.0240
114	Engineering, Multidisciplinary	0.0250
115	Polymer Science	0.0260
116	Crystallography	0.0260
117	Optics	0.0260
118	Physics, Multidisciplinary	0.0260
119	Parasitology	0.0260
120	Chemistry, Multidisciplinary	0.0260
121	Chemistry, Inorganic & Nuclear	0.0270
122	Mathematics, Applied	0.0280
123	Metallurgy & Metallurgical Engineering	0.0280
124	Andrology	0.0280
125	Computer Science, Information Systems	0.0290
126	Mechanics	0.0290
127	Neurosciences	0.0290
128	Engineering, Mechanical	0.0300
129	Computer Science, Artificial Intelligence	0.0300
130	Physics, Condensed Matter	0.0300
131	Energy & Fuels	0.0310
132	Psychology, Developmental	0.0320
133	Computer Science, Hardware & Architecture	0.0330
134	Chemistry, Physical	0.0340
135	Psychology, Clinical	0.0340
136	Computer Science, Theory & Methods	0.0350
137	Automation & Control Systems	0.0350
138	Physics, Applied	0.0350

TABLE D.286. List of categories on the right wing (red) with projection onto the first principal component. Categories are sorted by absolute values of their component score in descending order.

Right (red) Wing		
No.	Category	PC1
1	Reproductive Biology	0.0380
2	Materials Science, Multidisciplinary	0.0400
3	Neuroimaging	0.0410
4	Telecommunications	0.0420

No.	Category	PC1
5	Nutrition & Dietetics	0.0430
6	Integrative & Complementary Medicine	0.0430
7	Medicine, Research & Experimental	0.0450
8	Radiology, Nuclear Medicine & Medical Imaging	0.0490
9	Sport Sciences	0.0510
10	Immunology	0.0510
11	Health Policy & Services	0.0520
12	Pathology	0.0530
13	Tropical Medicine	0.0530
14	Substance Abuse	0.0540
15	Engineering, Electrical & Electronic	0.0540
16	Public, Environmental & Occupational Health	0.0590
17	Rehabilitation	0.0610
18	Endocrinology & Metabolism	0.0630
19	Nursing	0.0650
20	Medical Laboratory Technology	0.0680
21	Dentistry, Oral Surgery & Medicine	0.0740
22	Infectious Diseases	0.0760
23	Gerontology	0.0770
24	Psychiatry	0.0820
25	Hematology	0.0820
26	Dermatology	0.0830
27	Health Care Sciences & Services	0.0890
28	Geriatrics & Gerontology	0.0940
29	Primary Health Care	0.1160
30	Obstetrics & Gynecology	0.1180
31	Ophthalmology	0.1270
32	Pediatrics	0.1270
33	Anesthesiology	0.1320
34	Oncology	0.1330
35	Otorhinolaryngology	0.1380
36	Allergy	0.1400
37	Peripheral Vascular Disease	0.1400
38	Transplantation	0.1460
39	Clinical Neurology	0.1470
40	Emergency Medicine	0.1470
41	Orthopedics	0.1530
42	Urology & Nephrology	0.1550

No.	Category	PC1
43	Medicine, General & Internal	0.1560
44	Rheumatology	0.1560
45	Respiratory System	0.1600
46	Gastroenterology & Hepatology	0.1650
47	Critical Care Medicine	0.1740
48	Cardiac & Cardiovascular Systems	0.2160
49	Surgery	0.2200

D.5. List of Original Attributes (Categories) in Positive, Negative and Zero Groups Identified by Vector Approximation on the Principal Components

Table D.287. List of original attributes (categories) in positive group identified by vector approximation on the 2^{nd} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC2)		
No.	Attribute	Component Coefficient
1	Cultural Studies	0.2044
2	Humanities, Multidisciplinary	0.1905
3	Asian Studies	0.1887
4	History	0.1877
5	Area Studies	0.1831
6	Literature	0.1765
7	History Of Social Sciences	0.1736
8	Sociology	0.1704
9	Social Issues	0.1688
10	Literature, Romance	0.1666
11	International Relations	0.1541
12	Political Science	0.1452
13	Medieval & Renaissance Studies	0.1451
14	Literary Theory & Criticism	0.1437
15	Ethnic Studies	0.1436
16	History & Philosophy Of Science	0.1376
17	Film, Radio, Television	0.1367
18	Communication	0.1311
19	Literature, African, Australian, Canadian	0.1304

No.	Attribute	Component Coefficient
20	Literature, British Isles	0.1286
21	Folklore	0.1222
22	Classics	0.1200
23	Literature, American	0.1199
24	Geography	0.1196
25	Literature, German, Dutch, Scandinavian	0.1192
26	Planning & Development	0.1180
27	Art	0.1165
28	Anthropology	0.1162
29	Social Sciences, Interdisciplinary	0.1150
30	Religion	0.1118
31	Public Administration	0.1090
32	Philosophy	0.1058
33	Theater	0.1000
34	Ethics	0.0982
35	Literary Reviews	0.0924
36	Law	0.0905
37	Literature, Slavic	0.0884
38	Poetry	0.0856
39	Industrial Relations & Labor	0.0797
40	Women's Studies	0.0762
41	Environmental Studies	0.0755
42	Social Work	0.0733
43	Demography	0.0722
44	Urban Studies	0.0684
45	Medical Ethics	0.0675
46	Social Sciences, Biomedical	0.0661

TABLE D.288. List of original attributes (categories) in zero group identified by vector approximation on the 2^{nd} principal component. Categories with positive component coefficients are sorted by values of their component coefficients in descending order. Categories with negative component coefficients are sorted by absolute values of their component coefficients in descending order. That is, categories in two directions are sorted from the greatest contribution to lowest contribution to the component.

Zero (PC2)		
No.	**Attribute**	**Component Coefficient**
1	Language & Linguistics	0.0603
2	Business	0.0544
3	Information Science & Library Science	0.0521
4	Management	0.0475
5	Linguistics	0.0462
6	Music	0.0457
7	Economics	0.0433
8	Psychology, Applied	0.0430
9	Hospitality, Leisure, Sport & Tourism	0.0429
10	Psychology, Social	0.0420
11	Psychology, Multidisciplinary	0.0418
12	Education & Educational Research	0.0418
13	Criminology & Penology	0.0412
14	Architecture	0.0411
15	Psychology, Psychoanalysis	0.0363
16	Archaeology	0.0362
17	Dance	0.0349
18	Family Studies	0.0279
19	Psychology, Educational	0.0255
20	Business, Finance	0.0236
21	Agricultural Economics & Policy	0.0226
22	Education, Scientific Disciplines	0.0175
23	Education, Special	0.0087
24	Psychology, Mathematical	0.0083
25	Health Policy & Services	0.0038
26	Social Sciences, Mathematical Methods	0.0014
27	Chemistry, Analytical	0.0012
28	Psychology, Experimental	0.0010
29	Transportation	0.0003

No.	Attribute	Component Coefficient
30	Geography, Physical	-0.0179
31	Metallurgy & Metallurgical Engineering	-0.0178
32	Materials Science, Textiles	-0.0177
33	Physics, Condensed Matter	-0.0175
34	Cell Biology	-0.0171
35	Agriculture, Dairy & Animal Science	-0.0171
36	Andrology	-0.0170
37	Ornithology	-0.0169
38	Physics, Atomic, Molecular & Chemical	-0.0169
39	Food Science & Technology	-0.0168
40	Engineering, Petroleum	-0.0165
41	Psychology, Biological	-0.0164
42	Geology	-0.0163
43	Mineralogy	-0.0163
44	Meteorology & Atmospheric Sciences	-0.0162
45	Microbiology	-0.0161
46	Virology	-0.0161
47	Physics, Nuclear	-0.0161
48	Optics	-0.0156
49	Forestry	-0.0156
50	Health Care Sciences & Services	-0.0156
51	Physics, Particles & Fields	-0.0153
52	Nanoscience & Nanotechnology	-0.0151
53	Gerontology	-0.0149
54	Materials Science, Composites	-0.0148
55	Biotechnology & Applied Microbiology	-0.0144
56	Materials Science, Paper & Wood	-0.0143
57	Soil Science	-0.0141
58	Nuclear Science & Technology	-0.0133
59	Thermodynamics	-0.0133
60	Polymer Science	-0.0132
61	Substance Abuse	-0.0130
62	Paleontology	-0.0125
63	Chemistry, Inorganic & Nuclear	-0.0124
64	Psychology	-0.0123
65	Materials Science, Ceramics	-0.0123
66	Statistics & Probability	-0.0122

No.	Attribute	Component Coefficient
67	Chemistry, Medicinal	-0.0122
68	Mathematical & Computational Biology	-0.0117
69	Crystallography	-0.0114
70	Biophysics	-0.0109
71	Materials Science, Coatings & Films	-0.0109
72	Mathematics	-0.0107
73	Chemistry, Organic	-0.0103
74	Mycology	-0.0101
75	Cell & Tissue Engineering	-0.0099
76	Astronomy & Astrophysics	-0.0095
77	Acoustics	-0.0093
78	Public, Environmental & Occupational Health	-0.0091
79	Electrochemistry	-0.0087
80	Logic	-0.0087
81	Engineering, Biomedical	-0.0066
82	Materials Science, Biomaterials	-0.0066
83	Psychology, Clinical	-0.0064
84	Spectroscopy	-0.0047
85	Medicine, Legal	-0.0041
86	Biochemical Research Methods	-0.0028
87	Nursing	-0.0025
88	Audiology & Speech-Language Pathology	-0.0025
89	Microscopy	-0.0019
90	Ergonomics	-0.0014
91	Psychology, Developmental	-0.0011

TABLE D.289. List of original attributes (categories) in negative group identified by vector approximation o the 2^{nd} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC2)		
No.	Attribute	Component Coefficient
1	Engineering, Multidisciplinary	-0.0759
2	Clinical Neurology	-0.0686
3	Medicine, General & Internal	-0.0677

No.	Attribute	Component Coefficient
4	Engineering, Electrical & Electronic	-0.0655
5	Computer Science, Theory & Methods	-0.0621
6	Gastroenterology & Hepatology	-0.0612
7	Surgery	-0.0605
8	Medicine, Research & Experimental	-0.0579
9	Critical Care Medicine	-0.0574
10	Computer Science, Interdisciplinary Applications	-0.0573
11	Peripheral Vascular Disease	-0.0573
12	Respiratory System	-0.0570
13	Engineering, Industrial	-0.0570
14	Automation & Control Systems	-0.0556
15	Computer Science, Information Systems	-0.0555
16	Urology & Nephrology	-0.0554
17	Computer Science, Hardware & Architecture	-0.0550
18	Engineering, Manufacturing	-0.0546
19	Computer Science, Artificial Intelligence	-0.0540
20	Engineering, Mechanical	-0.0532
21	Mathematics, Interdisciplinary Applications	-0.0524
22	Computer Science, Software Engineering	-0.0524
23	Cardiac & Cardiovascular Systems	-0.0516
24	Computer Science, Cybernetics	-0.0512
25	Anesthesiology	-0.0510
26	Hematology	-0.0506
27	Otorhinolaryngology	-0.0503
28	Pediatrics	-0.0502
29	Mechanics	-0.0497
30	Telecommunications	-0.0493
31	Orthopedics	-0.0486
32	Emergency Medicine	-0.0477
33	Rheumatology	-0.0464
34	Endocrinology & Metabolism	-0.0458
35	Engineering, Civil	-0.0449
36	Operations Research & Management Science	-0.0447
37	Dermatology	-0.0429
38	Medical Laboratory Technology	-0.0426
39	Oncology	-0.0412
40	Infectious Diseases	-0.0403

No.	Attribute	Component Coefficient
41	Immunology	-0.0392
42	Obstetrics & Gynecology	-0.0391
43	Biology	-0.0386
44	Pathology	-0.0365
45	Geriatrics & Gerontology	-0.0360
46	Transportation Science & Technology	-0.0352
47	Energy & Fuels	-0.0348
48	Materials Science, Multidisciplinary	-0.0334
49	Allergy	-0.0332
50	Pharmacology & Pharmacy	-0.0328
51	Engineering, Aerospace	-0.0328
52	Transplantation	-0.0326
53	Materials Science, Characterization & Testing	-0.0323
54	Sport Sciences	-0.0317
55	Tropical Medicine	-0.0317
56	Construction & Building Technology	-0.0317
57	Physics, Multidisciplinary	-0.0316
58	Multidisciplinary Sciences	-0.0316
59	Veterinary Sciences	-0.0314
60	Instruments & Instrumentation	-0.0314
61	Mathematics, Applied	-0.0313
62	Physics, Mathematical	-0.0305
63	Dentistry, Oral Surgery & Medicine	-0.0302
64	Nutrition & Dietetics	-0.0302
65	Physiology	-0.0300
66	Engineering, Marine	-0.0299
67	Neurosciences	-0.0292
68	Robotics	-0.0286
69	Geosciences, Multidisciplinary	-0.0282
70	Rehabilitation	-0.0276
71	Engineering, Ocean	-0.0275
72	Psychiatry	-0.0273
73	Marine & Freshwater Biology	-0.0272
74	Reproductive Biology	-0.0270
75	Engineering, Environmental	-0.0267
76	Primary Health Care	-0.0266
77	Ecology	-0.0259

No.	Attribute	Component Coefficient
78	Engineering, Geological	-0.0258
79	Agriculture, Multidisciplinary	-0.0250
80	Chemistry, Multidisciplinary	-0.0248
81	Ophthalmology	-0.0247
82	Imaging Science & Photographic Technology	-0.0245
83	Plant Sciences	-0.0242
84	Engineering, Chemical	-0.0242
85	Parasitology	-0.0241
86	Physics, Applied	-0.0240
87	Radiology, Nuclear Medicine & Medical Imaging	-0.0239
88	Chemistry, Applied	-0.0237
89	Limnology	-0.0235
90	Zoology	-0.0230
91	Agricultural Engineering	-0.0228
92	Integrative & Complementary Medicine	-0.0227
93	Oceanography	-0.0227
94	Chemistry, Physical	-0.0225
95	Environmental Sciences	-0.0225
96	Anatomy & Morphology	-0.0224
97	Entomology	-0.0223
98	Agronomy	-0.0222
99	Neuroimaging	-0.0220
100	Green & Sustainable Science & Technology	-0.0217
101	Genetics & Heredity	-0.0212
102	Evolutionary Biology	-0.0208
103	Remote Sensing	-0.0207
104	Mining & Mineral Processing	-0.0207
105	Physics, Fluids & Plasmas	-0.0202
106	Water Resources	-0.0200
107	Biochemistry & Molecular Biology	-0.0200
108	Behavioral Sciences	-0.0199
109	Geochemistry & Geophysics	-0.0197
110	Biodiversity Conservation	-0.0196
111	Medical Informatics	-0.0196
112	Horticulture	-0.0194
113	Fisheries	-0.0192
114	Toxicology	-0.0192

No.	Attribute	Component Coefficient
115	Developmental Biology	-0.0182

TABLE D.290. List of original attributes (categories) in positive group identified by vector approximation o the 3^{rd} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC3)		
No.	Attribute	Component Coefficient
1	Multidisciplinary Sciences	0.1711
2	Biochemistry & Molecular Biology	0.1628
3	Cell Biology	0.1570
4	Biology	0.1392
5	Computer Science, Information Systems	0.1324
6	Computer Science, Theory & Methods	0.1320
7	Computer Science, Interdisciplinary Applications	0.1303
8	Computer Science, Artificial Intelligence	0.1296
9	Biotechnology & Applied Microbiology	0.1262
10	Biophysics	0.1261
11	Developmental Biology	0.1226
12	Computer Science, Software Engineering	0.1199
13	Computer Science, Cybernetics	0.1183
14	Automation & Control Systems	0.1140
15	Computer Science, Hardware & Architecture	0.1122
16	Engineering, Industrial	0.1109
17	Physiology	0.1104
18	Operations Research & Management Science	0.1079
19	Engineering, Electrical & Electronic	0.1044
20	Engineering, Multidisciplinary	0.0991
21	Telecommunications	0.0950
22	Genetics & Heredity	0.0941
23	Microbiology	0.0938
24	Toxicology	0.0914
25	Cell & Tissue Engineering	0.0880
26	Pharmacology & Pharmacy	0.0880
27	Medicine, Research & Experimental	0.0777

No.	Attribute	Component Coefficient
28	Mathematics, Interdisciplinary Applications	0.0746
29	Mathematical & Computational Biology	0.0726
30	Immunology	0.0723
31	Transportation Science & Technology	0.0718
32	Plant Sciences	0.0652
33	Robotics	0.0647
34	Virology	0.0634
35	Chemistry, Medicinal	0.0630
36	Mycology	0.0614
37	Evolutionary Biology	0.0601
38	Imaging Science & Photographic Technology	0.0596
39	Ecology	0.0526
40	Zoology	0.0521
41	Biochemical Research Methods	0.0506

TABLE D.291. List of original attributes (categories) in zero group identified by vector approximation o the 3^{rd} principal component. Categories with positive component coefficients are sorted by values of their component coefficients in descending order. Categories with negative component coefficients are sorted by absolute values of their component coefficients in descending order. That is, categories in two directions are sorted from the greatest contribution to lowest contribution to the component.

Zero (PC3)		
No.	Attribute	Component Coefficient
1	Biodiversity Conservation	0.0485
2	Engineering, Manufacturing	0.0469
3	Social Sciences, Mathematical Methods	0.0467
4	Parasitology	0.0435
5	Anatomy & Morphology	0.0430
6	Veterinary Sciences	0.0428
7	Marine & Freshwater Biology	0.0418
8	Engineering, Aerospace	0.0412
9	Engineering, Civil	0.0406
10	Pathology	0.0399
11	Engineering, Marine	0.0373
12	Acoustics	0.0369

No.	Attribute	Component Coefficient
13	Entomology	0.0364
14	Ornithology	0.0350
15	Engineering, Mechanical	0.0342
16	Food Science & Technology	0.0339
17	Mathematics, Applied	0.0337
18	Engineering, Ocean	0.0329
19	Fisheries	0.0325
20	Remote Sensing	0.0305
21	Agronomy	0.0303
22	Horticulture	0.0295
23	Logic	0.0293
24	Materials Science, Biomaterials	0.0292
25	Transportation	0.0290
26	Agriculture, Dairy & Animal Science	0.0283
27	Literature, Romance	0.0279
28	Agriculture, Multidisciplinary	0.0276
29	Statistics & Probability	0.0273
30	Literary Theory & Criticism	0.0273
31	Reproductive Biology	0.0267
32	Literature, British Isles	0.0259
33	Engineering, Biomedical	0.0258
34	Medieval & Renaissance Studies	0.0255
35	Architecture	0.0248
36	Economics	0.0244
37	Forestry	0.0242
38	Paleontology	0.0236
39	Literature	0.0232
40	Construction & Building Technology	0.0227
41	Literary Reviews	0.0209
42	Classics	0.0209
43	Mathematics	0.0203
44	Geography, Physical	0.0202
45	Humanities, Multidisciplinary	0.0198
46	Literature, American	0.0198
47	Integrative & Complementary Medicine	0.0196
48	Literature, Slavic	0.0191
49	Asian Studies	0.0188

No.	Attribute	Component Coefficient
50	Neurosciences	0.0184
51	History	0.0181
52	Oncology	0.0181
53	Literature, German, Dutch, Scandinavian	0.0177
54	Art	0.0176
55	Business, Finance	0.0175
56	Information Science & Library Science	0.0173
57	Andrology	0.0173
58	History Of Social Sciences	0.0167
59	Philosophy	0.0167
60	Engineering, Geological	0.0159
61	Poetry	0.0149
62	Mechanics	0.0145
63	Folklore	0.0142
64	Soil Science	0.0141
65	Environmental Sciences	0.0140
66	Literature, African, Australian, Canadian	0.0137
67	History & Philosophy Of Science	0.0136
68	Limnology	0.0135
69	Archaeology	0.0131
70	Language & Linguistics	0.0131
71	Management	0.0125
72	Religion	0.0117
73	Theater	0.0116
74	Hematology	0.0099
75	Agricultural Economics & Policy	0.0099
76	Meteorology & Atmospheric Sciences	0.0096
77	Oceanography	0.0083
78	Cultural Studies	0.0082
79	Linguistics	0.0081
80	Planning & Development	0.0078
81	Environmental Studies	0.0076
82	Area Studies	0.0072
83	Instruments & Instrumentation	0.0066
84	Behavioral Sciences	0.0066
85	Business	0.0066
86	Geosciences, Multidisciplinary	0.0064

No.	Attribute	Component Coefficient
87	Film, Radio, Television	0.0063
88	Anthropology	0.0061
89	Geology	0.0053
90	International Relations	0.0053
91	Geography	0.0047
92	Endocrinology & Metabolism	0.0032
93	Urban Studies	0.0032
94	Law	0.0028
95	Dance	0.0022
96	Music	0.0022
97	Political Science	0.0019
98	Agricultural Engineering	0.0009
99	Geochemistry & Geophysics	0.0006
100	Medicine, Legal	0.0002
101	Water Resources	0.0002
102	Physics, Atomic, Molecular & Chemical	-0.0439
103	Medical Ethics	-0.0419
104	Medical Informatics	-0.0412
105	Demography	-0.0353
106	Materials Science, Ceramics	-0.0344
107	Electrochemistry	-0.0339
108	Women's Studies	-0.0339
109	Radiology, Nuclear Medicine & Medical Imaging	-0.0336
110	Polymer Science	-0.0329
111	Transplantation	-0.0307
112	Psychology, Experimental	-0.0304
113	Metallurgy & Metallurgical Engineering	-0.0300
114	Materials Science, Textiles	-0.0294
115	Crystallography	-0.0293
116	Social Issues	-0.0276
117	Sociology	-0.0273
118	Physics, Fluids & Plasmas	-0.0268
119	Criminology & Penology	-0.0262
120	Nutrition & Dietetics	-0.0258
121	Spectroscopy	-0.0246
122	Optics	-0.0242
123	Education & Educational Research	-0.0239

No.	Attribute	Component Coefficient
124	Infectious Diseases	-0.0219
125	Energy & Fuels	-0.0218
126	Audiology & Speech-Language Pathology	-0.0216
127	Materials Science, Composites	-0.0202
128	Education, Scientific Disciplines	-0.0197
129	Thermodynamics	-0.0195
130	Neuroimaging	-0.0191
131	Physics, Nuclear	-0.0166
132	Microscopy	-0.0163
133	Chemistry, Inorganic & Nuclear	-0.0158
134	Materials Science, Characterization & Testing	-0.0145
135	Engineering, Environmental	-0.0136
136	Industrial Relations & Labor	-0.0136
137	Materials Science, Paper & Wood	-0.0135
138	Tropical Medicine	-0.0134
139	Hospitality, Leisure, Sport & Tourism	-0.0116
140	Physics, Mathematical	-0.0116
141	Physics, Particles & Fields	-0.0115
142	Green & Sustainable Science & Technology	-0.0113
143	Ethics	-0.0112
144	Ethnic Studies	-0.0110
145	Mineralogy	-0.0105
146	Chemistry, Applied	-0.0104
147	Mining & Mineral Processing	-0.0102
148	Engineering, Petroleum	-0.0092
149	Nuclear Science & Technology	-0.0085
150	Psychology, Mathematical	-0.0074
151	Ergonomics	-0.0073
152	Communication	-0.0071
153	Psychology, Psychoanalysis	-0.0070
154	Psychology, Biological	-0.0064
155	Chemistry, Organic	-0.0046
156	Chemistry, Analytical	-0.0032
157	Public Administration	-0.0020
158	Astronomy & Astrophysics	-0.0019
159	Social Sciences, Interdisciplinary	-0.0017

Table D.292. List of original attributes (categories) in negative group identified by vector approximation o the 3^{rd} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC3)		
No.	**Attribute**	**Component Coefficient**
1	Medicine, General & Internal	-0.1713
2	Health Care Sciences & Services	-0.1592
3	Primary Health Care	-0.1449
4	Public, Environmental & Occupational Health	-0.1448
5	Health Policy & Services	-0.1363
6	Critical Care Medicine	-0.1320
7	Clinical Neurology	-0.1303
8	Rehabilitation	-0.1201
9	Gerontology	-0.1180
10	Emergency Medicine	-0.1167
11	Geriatrics & Gerontology	-0.1161
12	Pediatrics	-0.1160
13	Psychiatry	-0.1143
14	Surgery	-0.1109
15	Otorhinolaryngology	-0.1095
16	Social Sciences, Biomedical	-0.1085
17	Anesthesiology	-0.1066
18	Respiratory System	-0.1051
19	Psychology, Clinical	-0.1026
20	Nursing	-0.1003
21	Cardiac & Cardiovascular Systems	-0.0977
22	Urology & Nephrology	-0.0928
23	Materials Science, Multidisciplinary	-0.0923
24	Rheumatology	-0.0885
25	Orthopedics	-0.0883
26	Gastroenterology & Hepatology	-0.0880
27	Peripheral Vascular Disease	-0.0865
28	Physics, Applied	-0.0844
29	Obstetrics & Gynecology	-0.0822
30	Social Work	-0.0822
31	Nanoscience & Nanotechnology	-0.0817

No.	Attribute	Component Coefficient
32	Psychology, Multidisciplinary	-0.0805
33	Psychology, Developmental	-0.0801
34	Substance Abuse	-0.0768
35	Family Studies	-0.0766
36	Psychology	-0.0763
37	Physics, Condensed Matter	-0.0713
38	Chemistry, Physical	-0.0708
39	Medical Laboratory Technology	-0.0677
40	Education, Special	-0.0629
41	Allergy	-0.0629
42	Psychology, Social	-0.0580
43	Chemistry, Multidisciplinary	-0.0559
44	Physics, Multidisciplinary	-0.0513
45	Ophthalmology	-0.0508
46	Sport Sciences	-0.0506
47	Dentistry, Oral Surgery & Medicine	-0.0495
48	Materials Science, Coatings & Films	-0.0489
49	Engineering, Chemical	-0.0485
50	Dermatology	-0.0482
51	Psychology, Educational	-0.0472
52	Psychology, Applied	-0.0464

Table D.293. List of original attributes (categories) in positive group identified by vector approximation o the 4^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC4)		
No.	Attribute	Component Coefficient
1	Psychology, Multidisciplinary	0.1975
2	Psychology, Applied	0.1690
3	Psychology, Social	0.1638
4	Social Work	0.1593
5	Psychology	0.1373
6	Family Studies	0.1363
7	Psychology, Clinical	0.1331

No.	Attribute	Component Coefficient
8	Social Sciences, Biomedical	0.1265
9	Management	0.1221
10	Psychology, Developmental	0.1195
11	Psychology, Educational	0.1174
12	Business	0.1161
13	Psychology, Experimental	0.1103
14	Social Sciences, Interdisciplinary	0.1088
15	Education, Special	0.1052
16	Behavioral Sciences	0.0976
17	Information Science & Library Science	0.0974
18	Psychology, Biological	0.0963
19	Planning & Development	0.0947
20	Industrial Relations & Labor	0.0946
21	Demography	0.0924
22	Environmental Studies	0.0923
23	Education & Educational Research	0.0910
24	Social Issues	0.0861
25	Economics	0.0855
26	Sociology	0.0827
27	Health Policy & Services	0.0786
28	Criminology & Penology	0.0782
29	Ergonomics	0.0760
30	Education, Scientific Disciplines	0.0757
31	Public Administration	0.0742
32	Hospitality, Leisure, Sport & Tourism	0.0742
33	Business, Finance	0.0700
34	Psychology, Mathematical	0.0637
35	Public, Environmental & Occupational Health	0.0611
36	Agricultural Economics & Policy	0.0600
37	Biophysics	0.0559
38	Urban Studies	0.0558
39	Gerontology	0.0552
40	Psychiatry	0.0550
41	Biology	0.0543
42	Biochemistry & Molecular Biology	0.0524
43	Women's Studies	0.0510
44	Geography	0.0491

Table D.294. List of original attributes (categories) in zero group identified by vector approximation o the 4^{th} principal component. Categories with positive component coefficients are sorted by values of their component coefficients in descending order. Categories with negative component coefficients are sorted by absolute values of their component coefficients in descending order. That is, categories in two directions are sorted from the greatest contribution to lowest contribution to the component.

Zero (PC4)		
No.	**Attribute**	**Component Coefficient**
1	Nursing	0.04648
2	Biotechnology & Applied Microbiology	0.04589
3	Multidisciplinary Sciences	0.04389
4	Substance Abuse	0.04214
5	Materials Science, Coatings & Films	0.04184
6	Medical Ethics	0.04176
7	Genetics & Heredity	0.04107
8	Nanoscience & Nanotechnology	0.04048
9	Cell Biology	0.04041
10	Evolutionary Biology	0.04001
11	Developmental Biology	0.03994
12	Ecology	0.03957
13	Ethics	0.03898
14	Chemistry, Physical	0.03862
15	Neurosciences	0.03823
16	Biodiversity Conservation	0.03741
17	Transportation	0.03695
18	Materials Science, Biomaterials	0.03552
19	Physics, Condensed Matter	0.03445
20	Rehabilitation	0.03412
21	Social Sciences, Mathematical Methods	0.03378
22	Materials Science, Ceramics	0.03213
23	Microbiology	0.03182
24	Biochemical Research Methods	0.03135
25	Communication	0.03085
26	Ethnic Studies	0.03068
27	Microscopy	0.02999
28	Spectroscopy	0.02837
29	Zoology	0.02801

No.	Attribute	Component Coefficient
30	Polymer Science	0.02792
31	Mycology	0.02721
32	Plant Sciences	0.02681
33	Ornithology	0.02608
34	Materials Science, Multidisciplinary	0.02593
35	Physics, Applied	0.02525
36	Electrochemistry	0.02495
37	Chemistry, Multidisciplinary	0.02231
38	Health Care Sciences & Services	0.02183
39	Chemistry, Applied	0.02157
40	Audiology & Speech-Language Pathology	0.02000
41	Metallurgy & Metallurgical Engineering	0.01990
42	Chemistry, Inorganic & Nuclear	0.01859
43	Chemistry, Analytical	0.01793
44	Marine & Freshwater Biology	0.01708
45	Mathematical & Computational Biology	0.01674
46	Chemistry, Medicinal	0.01671
47	Entomology	0.01642
48	Cell & Tissue Engineering	0.01630
49	Anatomy & Morphology	0.01606
50	Forestry	0.01589
51	Materials Science, Composites	0.01556
52	Engineering, Chemical	0.01534
53	Crystallography	0.01524
54	Physics, Atomic, Molecular & Chemical	0.01482
55	Virology	0.01396
56	Political Science	0.01342
57	International Relations	0.01296
58	Neuroimaging	0.01286
59	Physiology	0.01233
60	Agronomy	0.01204
61	Geriatrics & Gerontology	0.01113
62	Law	0.01097
63	Horticulture	0.01089
64	Toxicology	0.01060
65	Nuclear Science & Technology	0.00955
66	Acoustics	0.00939

No.	Attribute	Component Coefficient
67	Materials Science, Textiles	0.00926
68	Medical Informatics	0.00833
69	Optics	0.00825
70	Fisheries	0.00817
71	Astronomy & Astrophysics	0.00816
72	Agriculture, Multidisciplinary	0.00790
73	Food Science & Technology	0.00790
74	Environmental Sciences	0.00778
75	Soil Science	0.00776
76	Limnology	0.00704
77	Physics, Nuclear	0.00627
78	Materials Science, Paper & Wood	0.00604
79	Medicine, Legal	0.00584
80	Mineralogy	0.00500
81	Green & Sustainable Science & Technology	0.00449
82	Parasitology	0.00438
83	Engineering, Biomedical	0.00425
84	Operations Research & Management Science	0.00420
85	Materials Science, Characterization & Testing	0.00384
86	Anthropology	0.00288
87	Geography, Physical	0.00116
88	Oceanography	0.00114
89	Thermodynamics	0.00112
90	Physics, Particles & Fields	0.00096
91	Integrative & Complementary Medicine	-0.04716
92	Pediatrics	-0.04666
93	Endocrinology & Metabolism	-0.04477
94	Engineering, Multidisciplinary	-0.04188
95	History Of Social Sciences	-0.04154
96	Infectious Diseases	-0.03875
97	Mathematics, Interdisciplinary Applications	-0.03662
98	Engineering, Electrical & Electronic	-0.03188
99	Pathology	-0.03050
100	Tropical Medicine	-0.02927
101	Automation & Control Systems	-0.02893
102	Archaeology	-0.02788
103	Language & Linguistics	-0.02619

No.	Attribute	Component Coefficient
104	Physics, Mathematical	-0.02588
105	Computer Science, Interdisciplinary Applications	-0.02418
106	Engineering, Mechanical	-0.02278
107	Architecture	-0.02272
108	Mechanics	-0.02194
109	Computer Science, Theory & Methods	-0.02168
110	Mathematics, Applied	-0.01957
111	Pharmacology & Pharmacy	-0.01910
112	Computer Science, Hardware & Architecture	-0.01910
113	Engineering, Aerospace	-0.01871
114	Area Studies	-0.01860
115	Computer Science, Software Engineering	-0.01823
116	Engineering, Manufacturing	-0.01806
117	Engineering, Civil	-0.01778
118	Computer Science, Artificial Intelligence	-0.01776
119	Telecommunications	-0.01753
120	Immunology	-0.01737
121	Physics, Multidisciplinary	-0.01712
122	Music	-0.01672
123	Dance	-0.01654
124	Logic	-0.01641
125	Reproductive Biology	-0.01305
126	Primary Health Care	-0.01291
127	Mathematics	-0.01283
128	Statistics & Probability	-0.01270
129	Psychology, Psychoanalysis	-0.01236
130	Instruments & Instrumentation	-0.01165
131	Construction & Building Technology	-0.01116
132	Sport Sciences	-0.01064
133	Energy & Fuels	-0.01060
134	Computer Science, Information Systems	-0.01037
135	Geosciences, Multidisciplinary	-0.01004
136	Engineering, Ocean	-0.01001
137	Engineering, Marine	-0.00997
138	Veterinary Sciences	-0.00907
139	Nutrition & Dietetics	-0.00826
140	Robotics	-0.00819

No.	Attribute	Component Coefficient
141	Andrology	-0.00798
142	Geology	-0.00756
143	Mining & Mineral Processing	-0.00711
144	Remote Sensing	-0.00661
145	Engineering, Industrial	-0.00660
146	Physics, Fluids & Plasmas	-0.00621
147	Imaging Science & Photographic Technology	-0.00612
148	Geochemistry & Geophysics	-0.00589
149	Transportation Science & Technology	-0.00467
150	Engineering, Geological	-0.00304
151	Agriculture, Dairy & Animal Science	-0.00266
152	Linguistics	-0.00255
153	Computer Science, Cybernetics	-0.00201
154	Meteorology & Atmospheric Sciences	-0.00191
155	Chemistry, Organic	-0.00190
156	Paleontology	-0.00164
157	Engineering, Environmental	-0.00151
158	Engineering, Petroleum	-0.00147
159	Water Resources	-0.00112
160	Agricultural Engineering	-0.00001

TABLE D.295. List of original attributes (categories) in negative group identified by vector approximation o the 4^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC4)		
No.	Attribute	Component Coefficient
1	Literature	-0.1482
2	Surgery	-0.1389
3	Literature, Romance	-0.1364
4	Literature, British Isles	-0.1318
5	Literary Theory & Criticism	-0.1295
6	Gastroenterology & Hepatology	-0.1279
7	Medieval & Renaissance Studies	-0.1242
8	Critical Care Medicine	-0.1241

No.	Attribute	Component Coefficient
9	Respiratory System	-0.1227
10	Literature, American	-0.1203
11	Medicine, General & Internal	-0.1171
12	Cardiac & Cardiovascular Systems	-0.1162
13	Peripheral Vascular Disease	-0.1155
14	Urology & Nephrology	-0.1152
15	Classics	-0.1145
16	Otorhinolaryngology	-0.1128
17	Literature, German, Dutch, Scandinavian	-0.1118
18	Anesthesiology	-0.1118
19	Medical Laboratory Technology	-0.1088
20	Literature, African, Australian, Canadian	-0.1088
21	Clinical Neurology	-0.1087
22	Poetry	-0.1041
23	Asian Studies	-0.1015
24	Rheumatology	-0.0990
25	Orthopedics	-0.0990
26	Emergency Medicine	-0.0988
27	Literature, Slavic	-0.0953
28	Literary Reviews	-0.0949
29	Humanities, Multidisciplinary	-0.0907
30	Art	-0.0857
31	History	-0.0835
32	Dermatology	-0.0807
33	Folklore	-0.0704
34	Cultural Studies	-0.0703
35	Hematology	-0.0687
36	Theater	-0.0670
37	Dentistry, Oral Surgery & Medicine	-0.0656
38	Allergy	-0.0646
39	Transplantation	-0.0628
40	Film, Radio, Television	-0.0614
41	Obstetrics & Gynecology	-0.0606
42	Oncology	-0.0573
43	Ophthalmology	-0.0564
44	Medicine, Research & Experimental	-0.0563
45	Philosophy	-0.0560

No.	Attribute	Component Coefficient
46	Religion	-0.0559
47	Radiology, Nuclear Medicine & Medical Imaging	-0.0507
48	History & Philosophy Of Science	-0.0495

Table D.296. List of original attributes (categories) in positive group identified by vector approximation o the 5^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Positive (PC5)		
No.	Attribute	Component Coefficient
1	Ecology	0.2341
2	Biodiversity Conservation	0.2128
3	Environmental Sciences	0.2122
4	Marine & Freshwater Biology	0.2103
5	Geosciences, Multidisciplinary	0.2101
6	Limnology	0.2056
7	Geography, Physical	0.1976
8	Oceanography	0.1783
9	Water Resources	0.1724
10	Zoology	0.1493
11	Paleontology	0.1397
12	Forestry	0.1384
13	Agronomy	0.1314
14	Evolutionary Biology	0.1306
15	Geology	0.1298
16	Environmental Studies	0.1263
17	Agriculture, Multidisciplinary	0.1263
18	Soil Science	0.1261
19	Meteorology & Atmospheric Sciences	0.1226
20	Plant Sciences	0.1203
21	Ornithology	0.1189
22	Engineering, Environmental	0.1185
23	Geochemistry & Geophysics	0.1127
24	Fisheries	0.1028
25	Entomology	0.1019

No.	Attribute	Component Coefficient
26	Horticulture	0.0994
27	Geography	0.0923
28	Planning & Development	0.0888
29	Agricultural Economics & Policy	0.0876
30	Green & Sustainable Science & Technology	0.0835
31	Biology	0.0835
32	Agricultural Engineering	0.0830
33	Urban Studies	0.0820
34	Mycology	0.0784
35	Economics	0.0761
36	Engineering, Geological	0.0698
37	Public Administration	0.0661

TABLE D.297. List of original attributes (categories) in zero group identified by vector approximation o the 5^{th} principal component. Categories with positive component coefficients are sorted by values of their component coefficients in descending order. Categories with negative component coefficients are sorted by absolute values of their component coefficients in descending order. That is, categories in two directions are sorted from the greatest contribution to lowest contribution to the component.

Zero (PC5)		
No.	**Attribute**	**Component Coefficient**
1	International Relations	0.0635
2	Mineralogy	0.0610
3	Business, Finance	0.0595
4	Political Science	0.0518
5	Engineering, Ocean	0.0501
6	Area Studies	0.0501
7	Mining & Mineral Processing	0.0475
8	Business	0.0442
9	Social Sciences, Mathematical Methods	0.0441
10	Engineering, Petroleum	0.0433
11	Archaeology	0.0395
12	Remote Sensing	0.0393
13	Surgery	0.0389
14	Engineering, Chemical	0.0385

No.	Attribute	Component Coefficient
15	Anthropology	0.0379
16	Radiology, Nuclear Medicine & Medical Imaging	0.0369
17	Parasitology	0.0357
18	Management	0.0351
19	Engineering, Marine	0.0307
20	History Of Social Sciences	0.0301
21	Genetics & Heredity	0.0294
22	Medical Laboratory Technology	0.0290
23	Gastroenterology & Hepatology	0.0286
24	Industrial Relations & Labor	0.0258
25	Tropical Medicine	0.0257
26	Physics, Multidisciplinary	0.0251
27	Physics, Applied	0.0245
28	Engineering, Biomedical	0.0243
29	Emergency Medicine	0.0236
30	Materials Science, Multidisciplinary	0.0232
31	Demography	0.0227
32	Cardiac & Cardiovascular Systems	0.0226
33	Engineering, Civil	0.0215
34	Respiratory System	0.0210
35	Anesthesiology	0.0209
36	Materials Science, Paper & Wood	0.0209
37	Microbiology	0.0198
38	Critical Care Medicine	0.0194
39	Orthopedics	0.0193
40	Food Science & Technology	0.0183
41	Physics, Condensed Matter	0.0182
42	Agriculture, Dairy & Animal Science	0.0179
43	History	0.0172
44	Materials Science, Characterization & Testing	0.0172
45	Energy & Fuels	0.0170
46	Law	0.0170
47	Nanoscience & Nanotechnology	0.0169
48	Urology & Nephrology	0.0168
49	Dermatology	0.0154
50	Rheumatology	0.0153
51	Ethnic Studies	0.0152

No.	Attribute	Component Coefficient
52	Otorhinolaryngology	0.0150
53	Materials Science, Composites	0.0147
54	Sociology	0.0146
55	Transportation	0.0143
56	Materials Science, Coatings & Films	0.0140
57	Architecture	0.0139
58	Physics, Fluids & Plasmas	0.0137
59	Biotechnology & Applied Microbiology	0.0136
60	Social Issues	0.0128
61	Oncology	0.0127
62	Metallurgy & Metallurgical Engineering	0.0127
63	Clinical Neurology	0.0122
64	Materials Science, Textiles	0.0122
65	Chemistry, Physical	0.0120
66	Medicine, General & Internal	0.0113
67	Astronomy & Astrophysics	0.0106
68	Microscopy	0.0102
69	Optics	0.0100
70	Physics, Atomic, Molecular & Chemical	0.0098
71	Construction & Building Technology	0.0096
72	Transplantation	0.0096
73	Peripheral Vascular Disease	0.0085
74	Electrochemistry	0.0085
75	Nuclear Science & Technology	0.0084
76	Physics, Particles & Fields	0.0083
77	Cultural Studies	0.0079
78	Hospitality, Leisure, Sport & Tourism	0.0077
79	Thermodynamics	0.0072
80	Hematology	0.0071
81	Infectious Diseases	0.0071
82	Physics, Nuclear	0.0069
83	Physics, Mathematical	0.0058
84	Dentistry, Oral Surgery & Medicine	0.0052
85	Veterinary Sciences	0.0049
86	Asian Studies	0.0047
87	Materials Science, Ceramics	0.0045
88	Pathology	0.0045

No.	Attribute	Component Coefficient
89	Social Sciences, Interdisciplinary	0.0041
90	Ophthalmology	0.0036
91	Polymer Science	0.0031
92	Chemistry, Applied	0.0018
93	Allergy	0.0014
94	Virology	0.0007
95	Literature, Slavic	-0.0280
96	Mathematical & Computational Biology	-0.0271
97	Psychology, Mathematical	-0.0270
98	Chemistry, Medicinal	-0.0259
99	Nutrition & Dietetics	-0.0243
100	Literature, African, Australian, Canadian	-0.0243
101	Literature, Romance	-0.0239
102	Classics	-0.0222
103	Nursing	-0.0221
104	Imaging Science & Photographic Technology	-0.0208
105	Pediatrics	-0.0208
106	Criminology & Penology	-0.0207
107	Information Science & Library Science	-0.0199
108	Instruments & Instrumentation	-0.0199
109	Mathematics, Applied	-0.0195
110	Engineering, Aerospace	-0.0188
111	Medical Informatics	-0.0182
112	Integrative & Complementary Medicine	-0.0177
113	Immunology	-0.0176
114	Developmental Biology	-0.0175
115	Women's Studies	-0.0175
116	Engineering, Mechanical	-0.0172
117	Chemistry, Organic	-0.0172
118	Health Policy & Services	-0.0170
119	Engineering, Manufacturing	-0.0170
120	Toxicology	-0.0166
121	Medieval & Renaissance Studies	-0.0158
122	Reproductive Biology	-0.0157
123	Logic	-0.0153
124	Medicine, Research & Experimental	-0.0144
125	Acoustics	-0.0142

No.	Attribute	Component Coefficient
126	Cell & Tissue Engineering	-0.0141
127	Andrology	-0.0140
128	Theater	-0.0136
129	Philosophy	-0.0131
130	Music	-0.0121
131	Multidisciplinary Sciences	-0.0118
132	Biochemical Research Methods	-0.0117
133	Chemistry, Inorganic & Nuclear	-0.0116
134	Psychology, Psychoanalysis	-0.0112
135	History & Philosophy Of Science	-0.0111
136	Statistics & Probability	-0.0107
137	Chemistry, Multidisciplinary	-0.0106
138	Mechanics	-0.0102
139	Art	-0.0100
140	Humanities, Multidisciplinary	-0.0099
141	Religion	-0.0090
142	Obstetrics & Gynecology	-0.0090
143	Dance	-0.0089
144	Medicine, Legal	-0.0086
145	Medical Ethics	-0.0079
146	Neuroimaging	-0.0078
147	Anatomy & Morphology	-0.0067
148	Spectroscopy	-0.0054
149	Ethics	-0.0054
150	Mathematics	-0.0053
151	Folklore	-0.0049
152	Health Care Sciences & Services	-0.0048
153	Materials Science, Biomaterials	-0.0040
154	Chemistry, Analytical	-0.0037
155	Crystallography	-0.0032
156	Film, Radio, Television	-0.0031
157	Primary Health Care	-0.0020
158	Communication	-0.0002

TABLE D.298. List of original attributes (categories) in negative group identified by vector approximation o the 5^{th} principal component. Categories are sorted by values of their component coefficients in descending order, that is, from the greatest contribution to lowest contribution to the component.

Negative (PC5)		
No.	**Attribute**	**Component Coefficient**
1	Computer Science, Interdisciplinary Applications	-0.1071
2	Computer Science, Artificial Intelligence	-0.1066
3	Computer Science, Information Systems	-0.1022
4	Computer Science, Theory & Methods	-0.1016
5	Computer Science, Cybernetics	-0.0991
6	Psychology	-0.0978
7	Computer Science, Software Engineering	-0.0934
8	Computer Science, Hardware & Architecture	-0.0922
9	Automation & Control Systems	-0.0921
10	Psychology, Clinical	-0.0865
11	Psychology, Multidisciplinary	-0.0830
12	Engineering, Electrical & Electronic	-0.0817
13	Psychology, Developmental	-0.0813
14	Telecommunications	-0.0807
15	Psychology, Experimental	-0.0757
16	Engineering, Multidisciplinary	-0.0717
17	Psychology, Educational	-0.0651
18	Education, Special	-0.0650
19	Psychiatry	-0.0617
20	Behavioral Sciences	-0.0608
21	Family Studies	-0.0602
22	Psychology, Social	-0.0597
23	Psychology, Biological	-0.0590
24	Gerontology	-0.0549
25	Robotics	-0.0536
26	Physiology	-0.0532
27	Geriatrics & Gerontology	-0.0524
28	Mathematics, Interdisciplinary Applications	-0.0515
29	Rehabilitation	-0.0511
30	Operations Research & Management Science	-0.0500
31	Substance Abuse	-0.0480

No.	Attribute	Component Coefficient
32	Neurosciences	-0.0445
33	Literary Theory & Criticism	-0.0419
34	Public, Environmental & Occupational Health	-0.0417
35	Education, Scientific Disciplines	-0.0396
36	Audiology & Speech-Language Pathology	-0.0395
37	Linguistics	-0.0393
38	Literature, British Isles	-0.0379
39	Education & Educational Research	-0.0378
40	Engineering, Industrial	-0.0377
41	Transportation Science & Technology	-0.0375
42	Language & Linguistics	-0.0375
43	Social Work	-0.0366
44	Literature, American	-0.0363
45	Ergonomics	-0.0354
46	Social Sciences, Biomedical	-0.0349
47	Psychology, Applied	-0.0346
48	Literature	-0.0341
49	Biophysics	-0.0339
50	Pharmacology & Pharmacy	-0.0328
51	Biochemistry & Molecular Biology	-0.0328
52	Poetry	-0.0326
53	Cell Biology	-0.0325
54	Literary Reviews	-0.0322
55	Endocrinology & Metabolism	-0.0319
56	Literature, German, Dutch, Scandinavian	-0.0301
57	Sport Sciences	-0.0292

APPENDIX E

Appendix of Chapter 6: Descriptive Statistics for Citation Counts

E.1. List of Categories with the Descriptive Statistics Obtained for the Number of Citation for Each of Categories. Categories are Sorted by the Average Citation Count Assigned to the Category.

TABLE E.1. List of categories with the descriptive statistics obtained for the number of citation

Set	N	Max	Min	μ	Q_1	Q_2	Q_3	σ	SE
Cell Biology	23,108	978	0	20.67	6	12	24	32.37	0.213
Chemistry, Multidisciplinary	55,907	2,210	0	18.90	3	9	21	37.29	0.158
Multidisciplinary Sciences	53,140	3,004	0	18.32	4	9	18	44.42	0.193
Chemistry, Physical	58,065	2,288	0	18.11	5	10	20	32.86	0.136
Nanoscience & Nanotechnology	35,050	2,464	0	18.11	2	8	20	40.29	0.215
Critical Care Medicine	3,982	320	0	17.53	5	11	21	22.63	0.359
Allergy	1,765	488	0	17.41	4	10	20	26.99	0.642
Neuroimaging	2,702	527	0	17.24	6	12	22	23.59	0.454
Cell & Tissue Engineering	2,455	261	0	16.81	4	10	20	24.64	0.497
Medicine, General & Internal	16,179	3,908	0	16.01	2	4	10	68.91	0.542
Gastroenterology & Hepatology	10,943	667	0	15.50	4	9	18	24.99	0.239
Oncology	34,339	8,234	0	15.24	4	9	17	52.34	0.282
Hematology	9,096	399	0	15.17	4	9	18	22.19	0.233
Endocrinology & Metabolism	14,622	1,388	0	14.95	5	10	18	22.18	0.183
Genetics & Heredity	19,512	6,756	0	14.94	4	8	17	59.04	0.423
Materials Science, Biomaterials	8,040	280	0	14.74	4	10	19	17.46	0.195
Electrochemistry	15,663	290	0	14.56	4	10	19	16.46	0.132
Biochemistry & Molecular Biology	47,490	1,829	0	14.56	4	9	16	30.65	0.141
Immunology	18,270	515	0	14.52	4	9	17	21.29	0.157
Peripheral Vascular Disease	8,700	1,104	0	14.19	4	9	16	24.90	0.267
Rheumatology	3,942	481	0	14.16	4	9	17	19.92	0.317
Neurosciences	32,972	808	0	14.15	5	9	17	18.72	0.103
Green & Sustainable Science & Technology	6,412	304	0	14.10	3	9	19	17.12	0.214
Respiratory System	7,666	800	0	14.08	4	9	17	21.08	0.241
Virology	6,270	680	0	14.06	5	9	17	17.98	0.227
Cardiac & Cardiovascular Systems	16,369	813	0	14.02	3	8	16	22.63	0.177
Evolutionary Biology	5,742	1,085	0	13.96	5	9	17	21.70	0.286
Engineering, Environmental	14,614	367	0	13.87	2	8	19	18.81	0.156
Physics, Condensed Matter	27,316	2,288	0	13.83	2	6	14	35.28	0.213
Microbiology	17,252	680	0	13.73	4	9	17	18.98	0.145
Biochemical Research Methods	15,050	4,744	0	13.56	3	8	15	65.25	0.532
Agricultural Engineering	3,727	496	0	13.50	3	9	19	16.76	0.274

Set	N	Max	Min	μ	Q_1	Q_2	Q_3	σ	SE
Astronomy & Astrophysics	22,825	1,243	0	13.24	2	7	16	24.20	0.160
Nutrition & Dietetics	9,416	244	0	13.11	4	9	17	14.13	0.146
Engineering, Chemical	29,171	1,105	0	13.06	3	7	16	22.10	0.129
Developmental Biology	3,593	217	0	13.04	4	8	16	15.59	0.260
Infectious Diseases	12,521	515	0	12.74	4	8	15	18.28	0.163
Biotechnology & Applied Microbiology	26,286	4,744	0	12.70	3	7	14	49.30	0.304
Ecology	16,760	392	0	12.55	4	8	16	15.76	0.122
Meteorology & Atmospheric Sciences	12,318	1,659	0	12.42	3	8	15	24.22	0.218
Clinical Neurology	22,127	1,104	0	12.32	4	8	15	18.68	0.126
Environmental Sciences	42,082	1,105	0	12.28	3	7	15	21.18	0.103
Chemistry, Analytical	21,490	290	0	12.14	4	8	15	14.68	0.100
Geriatrics & Gerontology	4,742	513	0	12.13	4	8	15	15.78	0.229
Psychiatry	16,055	317	0	12.11	3	8	15	15.51	0.122
Psychology, Developmental	4,390	306	0	11.83	4	8	15	13.60	0.205
Anesthesiology	2,943	326	0	11.82	4	8	15	14.57	0.269
Energy & Fuels	44,202	1,105	0	11.81	0	5	16	22.14	0.105
Parasitology	5,683	680	0	11.79	4	8	14	17.54	0.233
Urology & Nephrology	8,264	994	0	11.75	3	7	14	19.56	0.215
Physics, Particles & Fields	13,203	1,508	0	11.72	2	6	14	22.81	0.199
Chemistry, Organic	17,941	210	0	11.61	3	8	15	13.11	0.098
Chemistry, Applied	14,058	174	0	11.50	3	8	15	12.27	0.104
Physics, Atomic, Molecular & Chemical	17,010	1,143	0	11.46	3	7	14	21.09	0.162
Environmental Studies	7,811	775	0	11.30	2	6	13	21.16	0.239
Medicine, Research & Experimental	19,744	978	0	11.24	2	6	13	21.63	0.154
Toxicology	9,613	338	0	11.17	4	8	14	12.80	0.131
Polymer Science	18,017	358	0	11.15	3	7	15	14.02	0.104
Materials Science, Multidisciplinary	112,912	2,464	0	10.99	0	4	12	27.89	0.083
Biophysics	12,630	488	0	10.97	3	7	14	13.87	0.123
Geochemistry & Geophysics	10,023	213	0	10.92	3	7	14	13.30	0.133
Psychology, Experimental	6,784	572	0	10.75	3	7	14	14.55	0.177
Psychology, Clinical	6,860	282	0	10.57	3	7	14	13.03	0.157
Psychology	6,989	300	0	10.56	3	7	14	14.75	0.176
Transplantation	4,105	559	0	10.47	3	6	13	17.34	0.271
Physiology	9,009	231	0	10.47	4	8	13	11.12	0.117
Pharmacology & Pharmacy	30,713	323	0	10.47	3	7	14	11.92	0.068
Gerontology	2,531	513	0	10.42	3	7	13	17.23	0.342
Reproductive Biology	3,984	454	0	10.40	4	7	13	13.95	0.221
Behavioral Sciences	5,922	186	0	10.37	4	8	14	9.92	0.129
Biology	9,917	517	0	10.36	2	6	13	16.08	0.162
Biodiversity Conservation	4,705	352	0	10.31	2	6	13	14.79	0.216
Chemistry, Medicinal	12,456	530	0	10.31	4	7	13	12.84	0.115
Substance Abuse	3,433	226	0	10.29	3	7	13	12.62	0.215
Geography	3,908	775	0	10.27	2	6	13	18.95	0.303
Psychology, Mathematical	538	572	0	10.25	2	5	10	31.19	1.345
Materials Science, Composites	4,277	357	0	10.11	2	5	13	15.48	0.237
Geography, Physical	6,806	378	0	10.09	2	6	13	16.09	0.195
Food Science & Technology	20,414	499	0	9.86	3	7	13	11.06	0.077
Materials Science, Coatings & Films	7,226	129	0	9.78	3	7	13	10.76	0.127
Orthopedics	10,538	234	0	9.76	3	7	13	11.35	0.111
Pathology	7,217	373	0	9.76	3	6	12	13.98	0.165

Set	N	Max	Min	μ	Q_1	Q_2	Q_3	σ	SE
Ophthalmology	7,830	615	0	9.75	3	6	12	15.19	0.172
Sport Sciences	8,368	184	0	9.69	3	7	13	11.48	0.126
Radiology, Nuclear Medicine & Medical Imaging	21,014	527	0	9.69	2	6	12	14.58	0.101
Psychology, Biological	1,527	294	0	9.67	3	7	12	13.22	0.338
Mathematical & Computational Biology	8,015	4,744	0	9.67	1	4	9	76.31	0.852
Mycology	1,829	232	0	9.58	3	6	12	13.51	0.316
Plant Sciences	21,321	449	0	9.54	2	6	12	12.81	0.088
Tropical Medicine	3,696	318	0	9.36	3	6	12	12.96	0.213
Surgery	30,805	559	0	9.35	2	6	12	12.95	0.074
Limnology	1,941	217	0	9.35	3	6	12	12.16	0.276
Thermodynamics	13,852	238	0	9.33	1	5	13	12.56	0.107
Obstetrics & Gynecology	9,883	454	0	9.32	3	6	12	12.25	0.123
Psychology, Applied	3,523	404	0	9.27	2	6	12	15.31	0.258
Geosciences, Multidisciplinary	24,644	383	0	9.26	2	6	12	13.90	0.089
Mineralogy	2,550	139	0	9.24	3	6	12	10.52	0.208
Physics, Multidisciplinary	22,930	1,178	0	9.21	1	3	9	23.95	0.158
Health Care Sciences & Services	7,999	288	0	9.21	3	6	12	12.78	0.143
Chemistry, Inorganic & Nuclear	12,591	173	0	9.19	2	6	12	11.15	0.099
Psychology, Social	3,548	193	0	9.15	3	6	11	11.46	0.192
Psychology, Educational	2,112	151	0	9.09	2	6	12	12.17	0.265
Soil Science	4,800	149	0	9.04	2	5	12	11.85	0.171
Psychology, Multidisciplinary	8,332	717	0	8.84	1	5	10	17.32	0.190
Public, Environmental & Occupational Health	25,493	655	0	8.77	2	5	11	13.79	0.086
Water Resources	13,997	367	0	8.76	2	5	11	11.81	0.100
Physics, Applied	78,796	2,288	0	8.73	0	3	9	24.73	0.088
Dermatology	5,793	249	0	8.58	2	5	11	11.86	0.156
Marine & Freshwater Biology	10,124	157	0	8.53	3	6	11	9.66	0.096
Instruments & Instrumentation	17,090	419	0	8.48	1	4	10	14.16	0.108
Pediatrics	13,364	309	0	8.43	2	5	10	12.32	0.107
Health Policy & Services	5,318	159	0	8.42	2	6	11	10.38	0.142
Social Sciences, Biomedical	3,003	132	0	8.40	3	6	11	9.50	0.173
Urban Studies	2,309	334	0	8.11	1	4	10	14.43	0.300
Medical Laboratory Technology	2,598	337	0	8.05	2	5	10	11.81	0.232
Ergonomics	1,431	109	0	7.94	3	6	10	8.76	0.232
Metallurgy & Metallurgical Engineering	16,898	209	0	7.92	1	4	10	11.09	0.085
Integrative & Complementary Medicine	3,453	88	0	7.91	2	6	11	7.80	0.133
Geology	2,153	108	0	7.86	2	5	10	9.18	0.198
Emergency Medicine	2,627	143	0	7.82	2	5	10	10.49	0.205
Engineering, Biomedical	17,786	368	0	7.82	0	3	10	13.45	0.101
Dentistry, Oral Surgery & Medicine	8,502	445	0	7.80	2	5	10	10.12	0.110
Spectroscopy	7,388	173	0	7.75	2	5	10	9.13	0.106
Forestry	4,472	148	0	7.72	2	5	10	8.95	0.134
Crystallography	6,932	743	0	7.71	2	5	10	13.28	0.160
Materials Science, Ceramics	6,222	257	0	7.69	1	5	11	9.85	0.125
Statistics & Probability	9,532	4,744	0	7.65	1	3	7	68.73	0.704
Physics, Fluids & Plasmas	9,704	229	0	7.64	2	5	10	10.26	0.104
Fisheries	4,702	152	0	7.61	2	5	10	9.42	0.137
Demography	948	197	0	7.46	2	5	10	10.29	0.334

Set	N	Max	Min	μ	Q_1	Q_2	Q_3	σ	SE
Andrology	391	36	0	7.35	3	6	10	6.00	0.303
Oceanography	7,417	200	0	7.33	1	5	10	9.74	0.113
Primary Health Care	1,269	253	0	7.29	2	5	10	10.76	0.302
Audiology & Speech-Language Pathology	2,052	95	0	7.23	3	5	9	7.64	0.169
Medicine, Legal	1,711	139	0	7.21	2	5	9	9.62	0.233
Business	9,394	404	0	7.19	0	2	9	14.92	0.154
Agriculture, Multidisciplinary	6,406	203	0	7.16	1	4	10	9.49	0.119
Paleontology	2,503	92	0	7.09	2	5	9	7.31	0.146
Family Studies	2,229	80	0	7.05	2	5	9	7.59	0.161
Political Science	5,106	293	0	7.04	1	4	9	10.89	0.152
Operations Research & Management Science	11,879	271	0	6.99	0	3	9	12.09	0.111
Agronomy	8,651	148	0	6.98	1	4	9	9.42	0.101
Otorhinolaryngology	4,797	432	0	6.96	2	5	9	9.65	0.139
Management	14,339	304	0	6.92	0	2	8	13.30	0.111
Imaging Science & Photographic Technology	9,353	737	0	6.91	0	2	7	17.96	0.186
Anthropology	3,149	84	0	6.86	1	4	9	8.92	0.159
Planning & Development	4,115	147	0	6.81	0	3	9	11.38	0.177
Transportation	4,035	113	0	6.80	1	4	9	8.93	0.141
Hospitality, Leisure, Sport & Tourism	2,998	144	0	6.80	0	4	9	10.55	0.193
Physics, Mathematical	10,426	412	0	6.79	1	4	8	11.69	0.114
Rehabilitation	7,791	182	0	6.78	2	5	9	7.84	0.089
Information Science & Library Science	4,565	162	0	6.62	0	3	8	11.35	0.168
Agricultural Economics & Policy	880	143	0	6.62	1	3	8	11.32	0.382
Criminology & Penology	2,015	91	0	6.60	2	4	9	8.41	0.187
Physics, Nuclear	7,876	324	0	6.58	1	3	8	12.36	0.139
Communication	3,200	169	0	6.55	1	3	8	10.80	0.191
Sociology	4,725	192	0	6.54	1	4	8	9.17	0.133
Medical Informatics	3,991	368	0	6.54	0	3	8	12.07	0.191
Materials Science, Paper & Wood	1,963	358	0	6.51	2	4	8	10.92	0.247
Materials Science, Textiles	2,548	358	0	6.19	1	3	8	10.78	0.214
Economics	22,338	217	0	6.11	1	3	7	10.78	0.072
Social Issues	1,296	72	0	6.04	2	4	8	7.09	0.197
Engineering, Geological	4,573	222	0	6.00	0	2	8	10.39	0.154
Zoology	11,218	234	0	6.00	2	4	8	7.50	0.071
Anatomy & Morphology	1,889	62	0	5.97	2	4	8	6.95	0.160
Agriculture, Dairy & Animal Science	6,163	289	0	5.96	1	4	8	7.58	0.097
Education, Special	1,666	56	0	5.95	1	4	8	6.71	0.164
Remote Sensing	11,388	469	0	5.94	0	1	6	13.97	0.131
Mathematics, Interdisciplinary Applications	10,072	572	0	5.89	1	2	7	13.36	0.133
International Relations	2,941	125	0	5.88	1	3	7	8.70	0.160
Entomology	5,704	116	0	5.88	2	4	8	7.38	0.098
Engineering, Civil	22,127	222	0	5.83	0	2	8	9.93	0.067
Social Work	2,114	78	0	5.82	2	4	8	6.69	0.145
Nursing	6,637	101	0	5.72	2	4	8	6.42	0.079
Computer Science, Interdisciplinary Applications	29,153	4,744	0	5.67	0	1	6	40.84	0.239
Construction & Building Technology	12,078	167	0	5.64	0	1	7	9.91	0.090

Set	N	Max	Min	μ	Q_1	Q_2	Q_3	σ	SE
Acoustics	6,935	236	0	5.60	0	2	7	10.04	0.121
Industrial Relations & Labor	879	57	0	5.59	1	3	7	7.48	0.252
Women's Studies	1,341	81	0	5.50	1	3	7	7.26	0.198
Ethics	1,928	61	0	5.46	1	3	7	7.49	0.170
Optics	47,737	1,277	0	5.44	0	2	6	14.16	0.065
Veterinary Sciences	11,502	289	0	5.42	1	3	7	7.17	0.067
Archaeology	2,118	89	0	5.41	1	3	7	7.45	0.162
Engineering, Industrial	10,362	532	0	5.36	0	1	6	11.99	0.118
Public Administration	2,204	81	0	5.34	0	2	7	8.54	0.182
Mechanics	33,545	229	0	5.33	0	1	7	9.77	0.053
Computer Science, Artificial Intelligence	41,210	2,100	0	5.31	0	1	5	20.59	0.101
Microscopy	1,319	129	0	5.16	1	3	7	7.98	0.220
Ethnic Studies	675	42	0	5.14	1	3	7	5.76	0.222
Medical Ethics	674	57	0	5.08	1	3	7	6.05	0.233
Automation & Control Systems	29,427	2,100	0	5.02	0	1	4	18.32	0.107
Ornithology	1,008	70	0	4.92	1	4	7	5.54	0.175
Computer Science, Cybernetics	3,652	205	0	4.88	0	1	4.25	12.22	0.202
Mining & Mineral Processing	2,687	81	0	4.70	0	2	6	7.39	0.143
Telecommunications	40,550	2,028	0	4.50	0	1	3	21.47	0.107
Engineering, Manufacturing	13,102	236	0	4.48	0	1	6	8.88	0.078
Business, Finance	7,214	162	0	4.48	0	1	5	9.24	0.109
Transportation Science & Technology	8,411	201	0	4.39	0	1	5	9.97	0.109
Mathematics, Applied	27,982	253	0	4.33	0	2	5	8.03	0.048
Engineering, Multidisciplinary	21,144	696	0	4.26	0	1	4	10.58	0.073
Social Sciences, Mathematical Methods	3,496	225	0	4.21	0	1	5	9.46	0.160
Horticulture	5,338	105	0	4.19	0	2	5	6.75	0.092
Law	3,574	77	0	4.03	0	2	5	6.55	0.110
Computer Science, Software Engineering	17,103	696	0	4.00	0	1	4	10.52	0.080
Nuclear Science & Technology	11,359	229	0	4.00	0	2	5	6.72	0.063
Linguistics	5,921	112	0	3.96	0	2	5	6.47	0.084
Engineering, Ocean	2,352	200	0	3.83	0	0	5	8.62	0.178
Materials Science, Characterization & Testing	3,878	246	0	3.82	0	1	5	7.44	0.119
Engineering, Electrical & Electronic	174,272	2,028	0	3.80	0	0	3	13.96	0.033
Engineering, Aerospace	4,435	270	0	3.74	0	1	5	7.92	0.119
Education, Scientific Disciplines	6,308	238	0	3.72	0	1	5	8.72	0.110
Computer Science, Information Systems	45,865	715	0	3.63	0	1	3	12.79	0.060
Cultural Studies	945	190	0	3.55	0	1	4	9.13	0.297
Engineering, Petroleum	1,930	120	0	3.53	0	1	4	6.63	0.151
Robotics	8,491	175	0	3.38	0	1	4	7.58	0.082
History & Philosophy Of Science	2,199	104	0	3.35	0	2	4	5.96	0.127
Computer Science, Hardware & Architecture	18,489	532	0	3.34	0	1	3	10.96	0.081
Mathematics	25,450	253	0	3.19	0	2	4	5.54	0.035
Education & Educational Research	20,087	149	0	3.14	0	1	4	6.85	0.048
Area Studies	2,046	44	0	2.92	0	1	4	4.44	0.098
Engineering, Mechanical	50,972	262	0	2.90	0	0	2	7.27	0.032
Social Sciences, Interdisciplinary	11,035	92	0	2.73	0	0	3	5.62	0.053
Film, Radio, Television	398	31	0	2.59	0	1	3	4.58	0.229

Set	N	Max	Min	μ	Q_1	Q_2	Q_3	σ	SE
History Of Social Sciences	879	97	0	2.49	0	1	3	4.24	0.143
Computer Science, Theory & Methods	55,591	737	0	2.45	0	1	2	8.79	0.037
Art	725	89	0	2.33	0	1	3	5.55	0.206
Logic	1,786	38	0	2.33	0	1	3	3.28	0.078
Engineering, Marine	2,110	101	0	2.27	0	0	2	5.12	0.112
Philosophy	3,657	61	0	2.22	0	1	3	4.06	0.067
Language & Linguistics	5,174	112	0	2.20	0	1	3	4.25	0.059
Psychology, Psychoanalysis	345	40	0	2.17	0	1	3	4.04	0.218
Music	888	28	0	2.06	0	1	3	3.10	0.104
Religion	2,335	81	0	1.70	0	1	2	3.52	0.073
History	3,487	97	0	1.53	0	1	2	2.84	0.048
Asian Studies	877	32	0	1.10	0	0	1	1.99	0.067
Literature, African, Australian, Canadian	59	6	0	1.07	0	1	1.5	1.26	0.164
Humanities, Multidisciplinary	2,559	53	0	1.02	0	0	1	2.83	0.056
Architecture	1,376	145	0	1.00	0	0	1	4.45	0.120
Literature	1,608	18	0	0.81	0	0	1	1.77	0.044
Theater	300	9	0	0.77	0	0	1	1.19	0.069
Medieval & Renaissance Studies	485	11	0	0.68	0	0	1	1.16	0.053
Literature, American	75	5	0	0.68	0	0	1	1.02	0.117
Classics	325	9	0	0.66	0	0	1	1.15	0.064
Folklore	134	7	0	0.65	0	0	1	1.26	0.109
Dance	74	8	0	0.55	0	0	0	1.30	0.152
Literature, Romance	269	6	0	0.49	0	0	1	0.92	0.056
Literature, British Isles	220	4	0	0.43	0	0	1	0.74	0.050
Literary Theory & Criticism	498	45	0	0.39	0	0	0	2.34	0.105
Poetry	42	3	0	0.38	0	0	1	0.70	0.108
Literature, German, Dutch, Scandinavian	128	4	0	0.36	0	0	1	0.70	0.061
Literature, Slavic	35	3	0	0.29	0	0	0	0.67	0.113
Literary Reviews	35	2	0	0.17	0	0	0	0.51	0.087

Bibliography

[1] Wittgenstein, L. (2009) *Philosophical Investigations.* Fourth Edition, Revised, P. Hacker, J. Schulte, Eds., John Wiley & Sons.

[2] V.A. Lefebvre, V.A. (2010) *Lectures on the Reflexive Games Theory.* Los Angeles, CA, Leaf & Oaks Publisher.

[3] Mel'čuk & I.A, Žolkovskij, A.K. (1970) "Towards a functioning 'Meaning-Text' model of language." *Linguistics.* **8(57)**: 10–47. https://doi.org/10.1515/ling.1970.8.57.10

[4] Mel'čuk, I.A. (2012) *Semantics: From Meaning to Text.* Volume 1 [Studies in Language Companion Series, 129]. Amsterdam, The Netherlands, John Benjamins Publishing Company.

[5] Mel'čuk, I.A. (2013) *Semantics: From Meaning to Text.* Volume 2 [Studies in Language Companion Series, 135]. Amsterdam, The Netherlands, John Benjamins Publishing Company.

[6] Mel'čuk, I.A. (2015) *Semantics: From Meaning to Text.* Volume 3 [Studies in Language Companion Series, 168]. Amsterdam, The Netherlands, John Benjamins Publishing Company.

[7] Shchedrovitsky, G.P. (1974) "Sense and meaning", In: Solntsev, V.M. (ed.) *Problems of Semantics*, Moscow, "Nauka", 76–111.

[8] Thurstone, L.L. (1934) "The vectors of mind." *Psychol. Rev.* **41(1)**: 1–32 . https://doi.org/10.1037/h0075959

[9] McCrae, R.R., & Costa, P.T. (2004) "A contemplated revision of the NEO five-factor inventory." *Personal. Individ. Differ.* **36(3)**: 587–596. https://doi.org/10.1016/s0191-8869(03)00118-1.

[10] Fehrman, E., Egan, V., Gorban, A.N., Levesley, J., Mirkes, E.M., & Muhammad, A.K. (2019) *Personality Traits and Drug Consumption. A Story Told by Data.* Cham, Springer. https://arxiv.org/abs/2001.06520

[11] Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957) *The measurement of meaning* (No. **47**). University of Illinois press.

[12] Osgood, C. E. (1952) "The nature and measurement of meaning." *Psychological bulletin*, **49(3)**: 197–237. https://doi.org/10.1037/h0055737

[13] Suzen, Neslihan (2019): LSC (Leicester Scientific Corpus). figshare. Dataset. https://doi.org/10.25392/leicester.data.9449639.v2

[14] *WoS Subject Categories.* (2019, October). Retrieved from https://images.webofknowledge.com/WOKRS56B5/help/WOS/hp_subject_category_terms_tasca.html

[15] Suzen, Neslihan (2019): LScDC (Leicester Scientific Dictionary-Core). figshare. Dataset. https://doi.org/10.25392/leicester.data.9896579.v3

[16] Suzen, Neslihan (2020): LScDC Word-Category RIG Matrix. University of Leicester. Dataset. https://doi.org/10.25392/leicester.data.12133431.v2

[17] Bornmann, L., & Daniel, H. D. (2008) "What do citation counts measure? A review of studies on citing behavior." *Journal of documentation.*

[18] van Raan, A. F. (2005) "For your citations only? Hot topics in bibliometric analysis." *Measurement: interdisciplinary research and perspectives*, **3(1)**: 50–62.

[19] Smith, A. T., & Eysenck, M. (2002) "The correlation between RAE ratings and citation counts in psychology". University of London, Royal Holloway.

[20] Small, H. (2004) "On the shoulders of Robert Merton: Towards a normative theory of citation." *Scientometrics*, **60(1)**: 71-79.

[21] Woolgar, S. (1991) "Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy." *Science and Public Policy*, **18(5)**: 319-326.

[22] Browne, C., Culligan, B. & Phillips, J. (2013). *The New Academic Word List.* Retrieved from http://www.newgeneralservicelist.org

[23] Süzen, N., Gorban, A. N., Levesley, J., & Mirkes, E. M. (2020) "Automatic short answer grading and feedback using text mining methods." *Procedia Computer Science*, **169**: 726–743.

[24] Mohler, M., Bunescu, R. and Mihalcea, R. (2011) "Learning to grade short answer questions using semantic similarity measures and dependency graph alignments," *in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* **1**: 752–762.

[25] Reynolds, C.R., Livingston, R.B., Willson, V.L. and Willson, V. (2010) *Measurement and assessment in education.* Pearson Education International.

[26] McDaniel, M.A., Anderson, J.L., Derbish, M.H. and Morrisette, N. (2007) "Testing the testing effect in the classroom." *European Journal of Cognitive Psychology* **19 (4–5)**: 494-513.

[27] Leacock, C. and Chodorow, M. (2003) "C-rater: Automated scoring of short-answer questions." *Computers and the Humanities* **37 (4)**: 389–405.

[28] Mitchell, T., Russel, T., Broomhead, P., Aldridge N. (2002) "Towards robust computerized marking of free-text responses," in M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK

[29] Sukkarieh, J.Z., Pulman, S.G. and Raikes, N. (2004) "Auto-marking 2: An update on the UCLES-Oxford University research into using computational linguistics to score short, free text responses." *International Association of Educational Assessment*, Philadelphia.

[30] Basu, S., Jacobs, C. and Vanderwende, L. (2013) "Powergrading: a clustering approach to amplify human effort for short answer grading." *Transactions of the Association for Computational Linguistics* **1**: 391–402.

[31] Page, E.B. (1966) "The imminence of... grading essays by computer." *The Phi Delta Kappan* **47 (5)**: 238–243.

[32] Wresch, W.(1993) "The imminence of grading Essays by computer– 25 Years Later." *Computers and composition* **10** (2): 45–58.

[33] Landauer, T.K.(2003) "Automatic essay assessment." *Assessment in education: Principles, policy and practice* **10 (3)**: 295–308.

[34] Shermis, M.D., Burstein, J., Higgins, D. and Zechner, K. (2010) "Automated essay scoring: Writing assessment and instruction." *International encyclopedia of education* **4 (1)**: 20–26

[35] Burrows, S., Gurevych, I. and Stein, B. (2015) "The eras and trends of automatic short answer grading." *International Journal of Artificial Intelligence in Education* **25 (1)**: 60–117.

[36] Jordan, S. and Mitchell, T. (2009) "e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback." *British Journal of Educational Technology* **40 (2)**: 371–385.

[37] Bachman, L.F., Carr, N., Kamei, G., Kim, M., Pan, M.J., Salvador, C. and Sawaki, Y., (2002) "A reliable approach to automatic assessment of short answer free responses," *in Proceedings of the 19th international conference on Computational linguistics* **2**: 1–4. Association for Computational Linguistics.

[38] Pulman, S.G. and Sukkarieh, J.Z. (2005) "Automatic short answer marking," *in Proceedings of the second workshop on Building Educational Applications Using NLP* (pp. 9–16). Association for Computational Linguistics.

[39] Cutrone, L. and Chang, M. (2011) "Auto-Assessor: Computerized assessment system for marking student's short-answers automatically," *in Technology for Education (T4E)*, 2011 IEEE International Conference on (pp. 81–88). IEEE.

[40] Thomas, P. (2003) "The evaluation of electronic marking of examinations." *ACM SIGCSE Bulletin* **35 (3)**: 50–54.

[41] Mohler, M. and Mihalcea, R. (2009) March. "Text-to-text semantic similarity for automatic short answer grading," *in Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 567–575). Association for Computational Linguistics.

[42] Horbach, A., Palmer, A. and Wolska, M. (2014) "Finding a Tradeoff between Accuracy and Rater's Workload in Grading Clustered Short Answers." *In LREC* (pp. 588–595).

[43] Manning C.D., and Schütze, H. (1999) *Foundations of Statistical Natural Language Processing*, MIT Press.

[44] Provost, F. and Fawcett, T. (2013) *Data Science for Business: What you need to know about data mining and data-analytic thinking.* O'Reilly Media, Inc.

[45] Everitt, B.S., Landau, S., Leese, M., Stahl, D. (2011) *Cluster Analysis.* John Wiley & Sons. Ltd., NY.

[46] Bock, H.H. (2007) "Clustering methods: a history of k-means algorithms." *In Selected contributions in data analysis and classification* (pp. 161–172). Springer, Berlin, Heidelberg.

[47] Cox, D.R. (1957) "Note on grouping." *Journal of the American Statistical Association* **52 (280)**: 543–547.

[48] Fisher, W.D. (1958) "On grouping for maximum homogeneity." *Journal of the American statistical Association* **53 (284)**: 789–798.

[49] Steinbach, M., Karypis, G. and Kumar, V. (200) "A comparison of document clustering techniques," *in KDD workshop on text mining* **400 (1)**: 525–526).

[50] Luo, C., Li, Y. and Chung, S.M. (2009) "Text document clustering based on neighbors." *Data and Knowledge Engineering* **68 (11)**: 1271–1288.

[51] Chiang, M.T. (2009) "Intelligent K-means clustering in L2 and L1 versions: experimentation and application" *(Doctoral dissertation, University of London).*

[52] Kaufman, L. and Rousseeuw, P. (1987) *Clustering by means of medoids.* North-Holland, Amsterdam.

[53] Xu, R. and Wunsch, D. (2005) "Survey of clustering algorithms." *IEEE Transactions on neural networks* **16 (3)**: 645–678.

[54] Park H.S., and Jun, C.H. (2009) "A simple and fast algorithm for K-medoids clustering", *Expert Systems with Applications* **36 (2)**: 3336–3341.

[55] Hamming, R. W. (1950) "Error detecting and error correcting codes." *The Bell System Technical Journal* **29 (2)**: 147-160

[56] Zielesny, A. (2016) *From curve fitting to machine learning: an illustrative guide to scientific data analysis and computational intelligence.* Springer, Switzerland.

[57] Motulsky, H. and Christopoulos, A. (2004) *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting.* Oxford University Press, Oxford.

[58] ResearchGate. (2019, October 14). *Why a commitment to open science makes a great job candidate* [Blog Post]. Retrieved from https://www.researchgate.net/scientific-recruitment/blog/post/why-a-commitment-to-open-science-makes-a-great-job-candidate

[59] Matthews, D. (2016, April 7). *Do academic social networks share academics' interests?*. Times Higher Education. Retrieved from https://www.timeshighereducation.com/features/do-academic-social-networks-share-academics-interests

[60] Sebastiani, F. (2014, April) "Text quantification." *In European Conference on Information Retrieval* (pp. 819-822). Springer, Cham.

[61] Leydesdorff, L., & Welbers, K. (2011) "The semantic mapping of words and co-words in contexts." *Journal of Informetrics*, **5(3)**: 469-475.

[62] Osgood, C. E. (1952) "The nature and measurement of meaning." *Psychological bulletin*, **49(3)**: 197-237.

[63] *Web of Science.* (2019, October). Retrieved from https://clarivate.com/webofsciencegroup/solutions/web-of-science/

[64] Suzen, N. (2019). *LSC (Leicester Scientific Corpus)* [Dataset]. figshare. Retrieved from https://doi.org/10.25392/leicester.data.9449639.v1

[65] *Research Areas in WoS.* (2019, October). Retrieved from https://images.webofknowledge.com/images/help/WOS/hp_research_areas_easca.html

[66] *Times Cited in WoS Core Collection.*. (2019, October). Retrieved from https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Times-Cited-accessibility-and-variation?language=en_US

[67] Suzen, N. (2019). *LScD (Leicester Scientific Dictionary)* [Dataset].figshare. Retrieved from https://doi.org/10.25392/leicester.data.9746900.v2

[68] Suzen, N. (2019). *LScDC (Leicester Scientific Dictionary-Core)* [Dataset]. figshare. Retrieved from https://doi.org/10.25392/leicester.data.9896579.v2

[69] Browne, C., Culligan, B. & Phillips, J. (2013). *The New General Service List.* Retrieved from http://www.newgeneralservicelist.org

[70] Coxhead, A. (2000) "A new academic word list." *TESOL quarterly*, 34(2): 213-238.

[71] Suzen, N. (2019). *LScD-LEICESTER SCIENTIFIC DICTIONARY CREATION*. Available from https://github.com/neslihansuzen

[72] Biber, D. (1993) "Representativeness in corpus design." *Literary and linguistic computing*, 8(4): 243-257.

[73] McEnery, T., Xiao, R., & Tono, Y. (2006) *Corpus-based language studies: An advanced resource book*. Routledge.

[74] Johansson, S., Leech, G. N., & Goodluck, H. (1978). *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computer*. Department of English, University of Oslo. Retrieved from http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM

[75] Onwuegbuzie, A. J., & Leech, N. L. (2007) "A call for qualitative power analyses." *Quality & Quantity*, 41(1): 105-121.
[76] Salton, G. (1975) *A theory of indexing* (Vol. **18**). SIAM.
[77] Salton, G., & McGill, M. J. (1983) *Introduction to modern information retrieval*. McGraw-Hill.
[78] Salton, G., & Buckley, C. (1988) "Term-weighting approaches in automatic text retrieval." *Information processing & management*, 24(5): 513-523.
[79] Ramos, J. (2003, December) "Using tf-idf to determine word relevance in document queries." *In Proceedings of the first instructional conference on machine learning* (Vol. **242**, pp. 133-142).
[80] Bowker, L. (2003) "3.2 Specialized lexicography and specialized dictionaries." In van Sterkenburg, P. (Ed.)., *A Practical Guide to Lexicography* (pp. 154-164). John Benjamins Publishing.
[81] Francis, W. N., & Kucera, H. (1967) *Computational analysis of present-day American English*. Brown University Press.
[82] Burnard, L., & Aston, G. (1998) *The BNC handbook: exploring the British National Corpus with SARA*. Edinburgh University Press.
[83] Leech, G. (2013) "The Development of ICAME and the Brown Family of Corpora." *Bergen Language and Linguistics Studies*, **3(1)**.
[84] Leech, G., & Rayson, P., Wilson, A. (2014) *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.
[85] Leech, G. N. (1992) "100 million words of English: the British National Corpus (BNC)." *Language Research*, 28(1): 1-13.
[86] *British National Corpus.* (2007). Retrieved from https://web.archive.org/web/20191017093659/http://www.natcorp.ox.ac.uk/corpus/index.xml
[87] *The Oxford English Corpus.* (2019, June 10). Retrieved from https://web.archive.org/web/20190610180141/https://languages.oup.com/our-story/corpus
[88] *The Oxford English Corpus.* (2019). Retrieved from Sketch Engine: https://www.sketchengine.eu/oxford-english-corpus/.
[89] *The Oxford Corpus of Academic English.* (2019). Retrieved from Sketch Engine: https://www.sketchengine.eu/oxford-corpus-of-academic-english/
[90] Rösiger, I. (2016, May) "Scicorp: A corpus of english scientific articles annotated for information status analysis." In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 1743-1749).
[91] D. Lewis. (1997). *Reuters-21578 text categorization test collection, Distribution 1.0 , AT& T Labs-Research* [Dataset]. Retrieved from http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html

[92] Kim, J. D., Ohta, T., Tateisi, Y., & Tsujii, J. I. (2003) "GENIA corpus—a semantically annotated corpus for bio-textmining." *Bioinformatics*, **19** (suppl_1), i180-i182.

[93] Brümmer, M., Dojchinovski, M., & Hellmann, S. (2016, May) "Dbpedia abstracts: A large-scale, open, multilingual NLP training corpus." *In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 3339-3343).

[94] Steels, L. (2017) "Human language is a culturally evolving system." *Psychonomic bulletin & review*, **24(1)**: 190-193.

[95] Satoshi. (2019) *In Oxford English dictionary online.* Oxford University Press. Retrieved from https://www.oed.com/view/Entry/78036595?redirectedFrom=satoshi#eid

[96] Yeesh. (2019) *In Oxford English dictionary online.* Oxford University Press. Retrieved from https://www.oed.com/view/Entry/75048500?redirectedFrom=yeesh#eid

[97] Simit. (2019) *In Oxford English dictionary online.* Oxford University Press. Retrieved from https://www.oed.com/view/Entry/61488496?redirectedFrom=simit#eid

[98] Crystal, D. (1987) "How many words?" *English Today*, **3(4)**: 11-14.

[99] *How many words are there in English?.* (2019, September 18) Merriam-Webster. Retrived from https://web.archive.org/web/20190918100123/https://www.merriam-webster.com/help/faq-how-many-english-words

[100] Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Pinker, S. (2011) "Quantitative analysis of culture using millions of digitized books." *Science*, *331(6014)*: 176-182.

[101] *The Oxford English Dictionary.* (2019, October 18). Retrieved from https://web.archive.org/web/20191018041346/https://public.oed.com/about/

[102] *Oxford English dictionary online.* (2019). Retrieved from https://www.oed.com/

[103] Gove, P. B. (Ed.). (1961) *Webster's third new international dictionary (unabridged).* United States of America: Merriam-Webster Inc.

[104] Ritter, R. M. (Ed.). (2014) *New Oxford dictionary for writers and editors.* Oxford University Press.

[105] Lea, D. (2014) "Making a learner's dictionary of academic English." *In Proceedings of the XVI EURALEX International Congress: The User in Focus. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.*

[106] Stedman, T. L. (Ed.). (2005) *Stedman's Medical Dictionary.* Lippincott Williams & Wilkins. Retrieved from https://stedmansonline.com/public/LearnMore.aspx?resourceID=Medical

[107] Younger, P. (2007) "Stedman's Medical Dictionary", 28th ed. *Reference Reviews*, **21(1)**: 41-42.

[108] *The OEC: Facts about the language.* (2012, January 4) Retrieved from https://web.archive.org/web/20111226085859/http://oxforddictionaries.com/words/the-oec-facts-about-the-language

[109] Word for Office 365. (2019) *Show word Count.* Retrieved from https://support.office.com/en-us/article/show-word-count-3c9e6a11-a04d-43b4-977c-563a0e0d5da3

[110] American Psychological Association. (1983) *Publication manual.* Washington, DC: American Psychological Association.

[111] Gloviczki, P., & Lawrence, P. F. (2017) "Information for Authors." *Journal of Vascular Surgery*, **65(1)**: A16 - A22.

[112] Suzen, N., Mirkes, E. M., & Gorban, A. N. (2019) "LScDC-new large scientific dictionary." *arXiv preprint* arXiv:1912.06858.

[113] Blumberg, R., & Atre, S. (2003) "The problem with unstructured data." *DM Review*, **(13)**: 42-49.

[114] Thomas, A. (2013) "Common Prefixes, Suffixes and Roots." *Center for Development and Learning* [PDF file]. Retrieved from https://www.cdl.org/wp-content/uploads/2013/12/Common-Prefixes-Suffixes-and-Roots-8.5.13.pdf

[115] Ramasubramanian, C. & Ramya,R. (2013) "Effective pre-processing activities in text mining using improved porter's stemming algorithm." *International Journal of Advanced Research in Computer and Communication Engineering*, **2(12)**: 4536-4538.

[116] Bouchet-Valat, M., & Bouchet-Valat, M. M. (2019) *SnowballC: Snowball Stemmers Based on the C 'libstemmer' UTF-8 Library. R package version 0.6.0.* Retrieved from https://CRAN.R-project.org/package=SnowballC

[117] Feinerer, I., Hornik, K., & Meyer, D. (2008) "Text Mining Infrastructure in R." *Journal of Statistical Software*, **25(5)**: 1-54.

[118] Zipf, G. K. (1949) *Human behavior and the principle of least effort.* Oxford, England: Addison-Wesley Press.

[119] Luhn, H. P. (1957) "A statistical approach to mechanized encoding and searching of literary information." *IBM Journal of research and development*, **1(4)**: 309-317.

[120] Salton, G., Wong, A., & Yang, C. S. (1975) "A vector space model for automatic indexing." *Communications of the ACM*, **18(11)**: 613-620.

[121] Torrisi, A., Bevan, R., Atkinson, K., Bollegala, D., & Coenen, F. (2019, June) "Automated Bundle Pagination Using Machine Learning." *In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (pp. 244-248). ACM.

[122] Yang, Y., & Pedersen, J. O. (1997, July) “A comparative study on feature selection in text categorization.” *In Icml*, **97(412-420)**: 35.

[123] Rigouste, L., Cappé, O., & Yvon, F. (2005, May) “Evaluation of a probabilistic method for unsupervised text clustering.” *In Proceedings of the international symposium on applied stochastic models and data analysis (ASMDA), Brest, France.*

[124] Wayne, W. P. (2018) *Subjectivity Detection through Quotation Analysis in Wikipedia.* Master’s thesis, University of Koblenz · Landau, Germany. Retrieved from https://d-nb.info/1162134461/34

[125] Nugumanova, A., Bessmertny, I., Baiburin, Y., & Mansurova, M. (2016, September) “A new operationalization of contrastive term extraction approach based on recognition of both representative and specific terms.” *In International Conference on Knowledge Engineering and the Semantic Web* (pp. 103-118). Springer, Cham.

[126] Hasan, K. S., & Ng, V. (2014, June) “Automatic keyphrase extraction: A survey of the state of the art.” *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume* ***1****: Long Papers)* (pp. 1262-1273).

[127] Weeber, M., Vos, R., & Baayen, R. H. (2000) “Extracting the lowest-frequency words: Pitfalls and possibilities.” *Computational Linguistics*, **26(3)**: 301-317.

[128] Comeau, D. C., & Wilbur, W. J. (2004) “Non-word identification or spell checking without a dictionary.” *Journal of the American Society for Information Science and Technology*, **55(2)**: 169-177.

[129] Pareto, V. (1964) *Cours d’économie politique* (Vol. **1**). Librairie Droz.

[130] Pareto, V. (2014) *Manual of political economy: a critical and variorum edition.* OUP Oxford.

[131] Pareto, V. (1897) “The new theories of economics.” *Journal of Political Economy*, **5(4)**:, 485-502.

[132] Geerolf, F. (2016) A theory of Pareto distributions. UCLA manuscript.

[133] Newman, M. (2000) “Applied mathematics: The power of design.” *Nature*, **405(6785)**: 412-413.

[134] Pickands III, J. (1975) “Statistical inference using extreme order statistics.” *the Annals of Statistics*, **3(1)**: 119-131.

[135] Hardy, M. (2010) “Pareto’s law.” *The Mathematical Intelligencer*, **32(3)**: 38-43.

[136] Kleiber, C., & Kotz, S. (2003) *Statistical size distributions in economics and actuarial sciences* (Vol. **470**). John Wiley & Sons.

[137] West, M. P. (Ed.). (1953) *A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology.* Longmans, Green.

[138] Carroll, J. B. (1971) "Statistical analysis of the corpus." *In The American Heritage Word Frequency Book* (pp. xxi–xl). Boston: Houghton Mifflin.

[139] Carroll, J. B., Davies, P., & Richman, B. (1971) "Guide to the alphabetical list." *In The American heritage word frequency book.* Boston: Houghton Mifflin

[140] Breland, H. M., Jones, R. J., Jenkins, L., Paynter, M., Pollack, J., & Fong, Y. F. (1994) "The college board vocabulary study." *ETS Research Report Series*, **1994(1)**: i-51.

[141] Hedgcock, J. S., & Ferris, D. R. (2018) *Teaching readers of English: Students, texts, and contexts.* Routledge.

[142] Hauke, J., & Kossowski, T. (2011) "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data." *Quaestiones geographicae*, **30(2)**: 87-93.

[143] Dodge, Y. (2008) *The concise encyclopedia of statistics.* Springer Science & Business Media.

[144] Ogden, C. K., & Richards, I. A. (1923) *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism* (Vol. **29**). K. Paul, Trench, Trubner & Company, Limited.

[145] Putnam, H. (1975) "The meaning of 'meaning'." *In Language, Mind, and Knowledge, ed. by K. Gunderson, Minnesota Studies in the Philosophy of Science*, **7**: 131–193.

[146] Carston, R. (2002) "Linguistic meaning, communicated meaning and cognitive pragmatics." *Mind & Language*, **17(1-2)**: 127-148.

[147] Michaelis, L. (2003) "Word meaning, sentence meaning, and syntactic meaning." *Cognitive approaches to lexical semantics*, **23**: 163-209.

[148] Shchedrovitsky, G.P. (1995) *Selected Works*, Moscow, The School of Cultural Politics.

[149] Habermas, J. (1984) *Theory of Communicative Action, V. 1: Reason and the Rationalization of Society.* Boston, MA, Beacon Press.

[150] Habermas, J. (1987) *Theory of Communicative Action, V. 2: Lifeworld and System: A Critique of Functionalist Reason.* Boston, MA, Beacon Press.

[151] Chen, G. (2012) *Fuzzy Logic in Data Modeling: Semantics, Constraints, and Database Design.* Springer, NY.

[152] Cho, K, van Merrienboer, B, Bahdanau, & D, Bengio, Y (2014), On the properties of neural machine translation: Encoder-decoder approaches. in *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8), October 25, 2014, Doha, Qatar*, Association for Computational Linguistics, 103–111. https://arxiv.org/abs/1409.1259

[153] Putnam, H. (1988) *Representation and Reality.* MIT press.

[154] Miller, G. A. (1998) *WordNet: An electronic lexical database.* MIT press.

[155] Samsonovic, A. V., & Ascoli, G. A. (2010) "Principal semantic components of language and the measurement of meaning." *PloS one*, **5(6)**: e10921.

[156] Turney, P. D., & Pantel, P. (2010) "From frequency to meaning: Vector space models of semantics." *Journal of artificial intelligence research*, **37**: 141-188.

[157] Anderson, A. B. (1970) "Structure of semantic space." *Sociological Methodology*, **2**: 308-325.

[158] Bentler, P. M., & LA VOIE, L. (1972) "An extension of semantic space." *Journal of Verbal Learning and Verbal Behavior*, **11(2)**: 174–182.

[159] Lowe, W. (2001) "What is the dimensionality of human semantic space?." *In Connectionist Models of Learning, Development and Evolution* (pp. 303-311). Springer, London.

[160] Harris, Z. S. (1954) "Distributional structure." *Word*, **10(2-3)**: 146-162.

[161] Sahlgren, M. (2006) *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces* (Doctoral dissertation).

[162] Lapesa, G., Evert, S., & Im Walde, S. S. (2014, August) "Contrasting syntagmatic and paradigmatic relations: Insights from distributional semantic models." *In Proceedings of the Third Joint Conference on Lexical and Computational Semantics* (* SEM 2014) (pp. 160-170).

[163] Schutze, H. (1993) "Word space." *In Advances in neural information processing systems* (pp. 895-902).

[164] Firth, J. R. (1957) "A synopsis of linguistic theory: 1930-1955" *Studies in linguistic analysis.*

[165] Dinu, G., & Lapata, M. (2010, October) "Measuring distributional similarity in context." *In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 1162-1172). Association for Computational Linguistics.

[166] Lowe, W. (2001) "Towards a theory of semantic space." *In Proceedings of the Annual Meeting of the Cognitive Science Society*, **23(23)**.

[167] Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2013) *Handbook of latent semantic analysis.* Psychology Press.

[168] Landauer, T. K., & Dumais, S. T. (1997) "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge." *Psychological review*, **104(2)**: 211.

[169] Landauer, T. K.,& Dumais, S. (2008) "Latent semantic analysis." *Scholarpedia*, **3(11)**: 4356.

[170] McCarthy, D., Apidianaki, M., & Erk, K. (2016) "Word sense clustering and clusterability." *Computational Linguistics*, **42(2)**: 245-275.

[171] Murphy, G. L. (1997) "Polysemy and the creation of novel word meanings." *In T. B. Ward*, S. M. Smith, & J. Vaid (Eds.), Creative thought: An investigation

of conceptual structures and processes (p. 235–265). American Psychological Association.

[172] Lopukhina, A., Laurinavichyute, A., Lopukhin, K., & Dragoy, O. (2018) "The mental representation of polysemy across word classes." *Frontiers in psychology*, **9**: 192.

[173] Hanks, P. (2000) "Do word meanings exist?." *Computers and the Humanities*, **34(1/2)**: 205-215.

[174] Kintsch, W. (2007) "Meaning in context." *Handbook of latent semantic analysis*, 89-105.

[175] Dumais, S. T. (2004) "Latent semantic analysis." *Annual review of information science and technology*, **38(1)**: 188-230.

[176] Evangelopoulos, N. E. (2013) "Latent semantic analysis." *Wiley Interdisciplinary Reviews: Cognitive Science*, **4(6)**: 683-692.

[177] Erk, K., McCarthy, D., & Gaylord, N. (2013) "Measuring word meaning in context." *Computational Linguistics*, **39(3)**: 511-554.

[178] Pustejovsky, J. (1991) "The generative lexicon." *Computational linguistics,* **17(4)**: 409-441.

[179] Kintsch, W. (2001) "Predication." Cognitive science, **25(2)**: 173-202.

[180] Kintsch, W., & Mangalath, P. (2011) "The construction of meaning." *Topics in Cognitive Science*, **3(2)**: 346-370.

[181] Klein, D. E., & Murphy, G. L. (2001) "The representation of polysemous words." *Journal of Memory and Language*, **45(2)**: 259-282.

[182] Murphy, G. (2004) *The big book of concepts. MIT press.*

[183] Frazier, L., & Rayner, K. (1990) "Taking on semantic commitments: Processing multiple meanings vs. multiple senses." *Journal of memory and language*, **29(2)**: 181-200.

[184] Frisson, S., & Frazier, L. (2005) "Carving up word meaning: Portioning and grinding." *Journal of Memory and Language*, **53(2)**: 277-291.

[185] Pickering, M. J., & Frisson, S. (2001) "Processing ambiguous verbs: Evidence from eye movements." *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **27(2)**: 556.

[186] Frisson, S. (2009) "Semantic underspecification in language processing." *Language and Linguistics Compass*, **3(1)**: 111-127.

[187] Erk, K. (2012) "Vector space models of word meaning and phrase meaning: A survey." *Language and Linguistics Compass*, **6(10)**: 635-653.

[188] Asher, N., & van de Cruys, T. (2018) "Content vs. function words: The view from distributional semantics." *In Proceedings of Sinn und Bedeutung*, 22(1): 1-21.

[189] Asher, N., Van de Cruys, T., & Abrusán, M. (2015) "Lexical semantics today." *The Routledge Handbook of Linguistics, N. York/London: Routledge*, 169-186.

[190] Erk, K., & Padó, S. (2010, July) "Exemplar-based models for word meaning in context." *In Proceedings of the acl 2010 conference short papers* (pp. 92-97).

[191] Erk, K., & Padó, S. (2008, October) "A structured vector space model for word meaning in context." *In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 897-906).

[192] Mitchell, J., & Lapata, M. (2008, June) "Vector-based models of semantic composition." *In proceedings of ACL-08*: HLT (pp. 236-244).

[193] Thater, S., Dinu, G., & Pinkal, M. (2009, August) "Ranking paraphrases in context." *In Proceedings of the 2009 Workshop on Applied Textual Inference (TextInfer)* (pp. 44-47).

[194] Mitchell, J., & Lapata, M. (2010) "Composition in distributional models of semantics." *Cognitive science*, **34(8)**: 1388-1429.

[195] Mitchell, J., & Lapata, M. (2009, August) "Language models based on semantic composition." *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1* (pp. 430-439). Association for Computational Linguistics.

[196] Van de Cruys, T., Poibeau, T., & Korhonen, A. (2011, July) "Latent vector weighting for word meaning in context." *In Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1012-1022). Association for Computational Linguistics.

[197] Van de Cruys, T. (2010) "Mining for meaning." *The Extraction of Lexicosemantic Knowledge from Text.*

[198] Paradis, C. (2012) "Lexical semantics." *The encyclopedia of applied linguistics*, 3356-3357.

[199] Beltagy, I., Roller, S., Cheng, P., Erk, K., & Mooney, R. J. (2015) "Representing meaning with a combination of logical form and vectors." *arXiv preprint* arXiv:1505.06816.

[200] Sahlgren, M. (2005) "An introduction to random indexing." *In Methods and applications of semantic indexing workshop at the 7th international conference on terminology and knowledge engineering.*

[201] Suzen, Neslihan (2019): LScD (Leicester Scientific Dictionary). figshare. Dataset. https://doi.org/10.25392/leicester.data.9746900.v3

[202] Shannon, C. E. (1948) "A mathematical theory of communication." *Bell system technical journal*, **27(3)**: 379-423.

[203] Largeron, C., Moulin, C., & Géry, M. (2011, March) "Entropy based feature selection for text categorization." *In Proceedings of the 2011 ACM symposium on applied computing* (pp. 924-928). ACM.

[204] Suzen, Neslihan (2020): LScDC Word Clouds and Tables to Visually Present the Most Informative Words in Subject Categories. figshare. Figure. https://doi.org/10.25392/leicester.data.12191604.v1

[205] Suzen, N., Mirkes, E. M., & Gorban, A. N. (2020) "Informational Space of Meaning for Scientific Texts." *arXiv preprint* arXiv:2004.13717.

[206] Wold, S., Esbensen, K., & Geladi, P. (1987) "Principal component analysis." *Chemometrics and intelligent laboratory systems*, **2(1-3)**: 37-52.

[207] Jackson, D. A. (1993) "Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches." *Ecology*, **74(8)**: 2204-2214.

[208] King, J. R., & Jackson, D. A. (1999) "Variable selection in large environmental data sets using principal components analysis." *Environmetrics: The official journal of the International Environmetrics Society*, **10(1)**: 67-77.

[209] Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005) "How many principal components? Stopping rules for determining the number of non-trivial axes revisited." *Computational Statistics & Data Analysis*, **49(4)**: 974-997.

[210] Zwick, W. R., & Velicer, W. F. (1986) "Comparison of five rules for determining the number of components to retain." *Psychological bulletin*, **99(3)**: 432.

[211] Fukunaga, K., & Olsen, D. R. (1971) "An algorithm for finding intrinsic dimensionality of data." *IEEE Transactions on Computers*, **100(2)**: 176-183.

[212] Gorban, A. N., Golubkov, A., Grechuk, B., Mirkes, E. M., & Tyukin, I. Y. (2018) "Correction of AI systems by linear discriminants: Probabilistic foundations." *Information Sciences*, **466**: 303-322.

[213] Mirkes, E. M., Allohibi, J., & Gorban, A. N. (2020) "Fractional norms and quasinorms do not help to overcome the curse of dimensionality." *arXiv preprint arXiv:2004.14230*.

[214] Dunteman, G. H. (1989) *Principal components analysis*, (No.**69**). Sage.

[215] Pearson, K. (1901). LIII. "On lines and planes of closest fit to systems of points in space." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* **2(11)**: 559-572.

[216] Guttman, L. (1954) "Some necessary conditions for common-factor analysis." *Psychometrika*, **19(2)**: 149-161.

[217] Kaiser, H. F. (1960) "The application of electronic computers to factor analysis." *Educational and psychological measurement*, **20(1)**: 141-151.

[218] Yeomans, K. A., & Golder, P. A. (1982) "The Guttman-Kaiser criterion as a predictor of the number of common factors." *The Statistician*, 221-229.

[219] Cangelosi, R., & Goriely, A. (2007) "Component retention in principal component analysis with application to cDNA microarray data." *Biology direct*, **2(1)**: 2.

[220] Süzen, N., Gorban, A. N., Levesley, J., & Mirkes, E. M. (2020) "Principal Components of the Meaning." *arXiv preprint* arXiv:2009.08859.

[221] Aksnes, D. W. (2003) "Characteristics of highly cited papers." *Research evaluation*, **12(3)**: 159-170.

[222] Suzen, Neslihan (2020): LScDC Word Clouds and Tables to Visually Present the Most Informative Words in Subject Categories. University of Leicester. Figure. https://doi.org/10.25392/leicester.data.12191604.v1

[223] Salton, G. (1971) *The SMART system. Retrieval Results and Future Plans.*

[224] Robertson, S. (2004) "Understanding inverse document frequency: on theoretical arguments for IDF." *Journal of documentation.*

[225] Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008) "Interpreting tf-idf term weights as making relevance decisions." *ACM Transactions on Information Systems (TOIS)*, **26(3)**: 1-37.

[226] Lund, K., & Burgess, C. (1996) "Producing high-dimensional semantic spaces from lexical co-occurrence." *Behavior research methods, instruments, & computers*, **28(2)**: 203-208.

[227] Berry, M. W. (2004) "Survey of text mining." *Computing Reviews,* **45(9)**: 548.

[228] Lewis, D. D. (1992, February) "Feature selection and feature extraction for text categorization." *In Proceedings of the workshop on Speech and Natural Language* (pp. 212-217). Association for Computational Linguistics.

[229] Salton, G., Buckley, C., & Smith, M. (1990) "On the application of syntactic methodologies in automatic text analysis." *Information Processing & Management*, **26(1)**: 73-92.

[230] Stavrianou, A., Andritsos, P., & Nicoloyannis, N. (2007) "Overview and semantic issues of text mining." *ACM Sigmod Record*, **36(3)**: 23-34.

[231] Salton, G. (1988, June) "Syntactic approaches to automatic book indexing." *In Proceedings of the 26th annual meeting on Association for Computational Linguistics (pp. 204-210).* Association for Computational Linguistics.

[232] Lewis, D. D. (1992, June) "An evaluation of phrasal and clustered representations on a text categorization task" *In Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, (pp. 37-50).

[233] Quan, X., Liu, G., Lu, Z., Ni, X., & Wenyin, L. (2010) "Short text similarity based on probabilistic topics." *Knowledge and information systems*, **25(3)**: 473-491.

[234] AlSumait, L., & Domeniconi, C. (2007, January) "Local semantic kernels for text document clustering." *In SIAM international conference on data mining workshop on text mining.*

[235] Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005) "Text classification using machine learning techniques." *WSEAS transactions on computers*, **4(8)**: 966-974.

[236] Church, K., & Gale, W. (1999) "Inverse document frequency (idf): A measure of deviations from poisson." *In Natural language processing using very large corpora* (pp. 283-295). Springer, Dordrecht.

[237] Xu, R., & Wunsch, D. (2008) *Clustering* (Vol. **10**). John Wiley & Sons.

[238] Aksnes, D. W., & Sivertsen, G. (2004) "The effect of highly cited papers on national citation indicators." *Scientometrics*, **59(2)**: 213-224.

[239] Chapman, W. W., Fizman, M., Chapman, B. E., & Haug, P. J. (2001) "A comparison of classification algorithms to automatically identify chest X-ray reports that support pneumonia." *Journal of biomedical informatics*, **34(1)**: 4-14.

[240] Davis, J., & Goadrich, M. (2006, June) "The relationship between Precision-Recall and ROC curves." *In Proceedings of the 23rd international conference on Machine learning* (pp. 233-240).

[241] Hanley, J. A., & McNeil, B. J. (1982) "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology*, **143(1)**: 29-36.

[242] Jiang, L., Cai, Z., Wang, D., & Jiang, S. (2007, August) "Survey of improving k-nearest-neighbor for classification." *In Fourth international conference on fuzzy systems and knowledge discovery (FSKD 2007)*, **Vol. 1**: 679-683. IEEE.

[243] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003, November) "KNN model-based approach in classification." *In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 986-996). Springer, Berlin, Heidelberg.

[244] Arif, M., Malagore, I. A., & Afsar, F. A. (2012) "Detection and localization of myocardial infarction using k-nearest neighbor classifier." *Journal of medical systems*, **36(1)**: 279-289.

[245] Tan, S. (2005) "Neighbor-weighted k-nearest neighbor for unbalanced text corpus." *Expert Systems with Applications*, **28(4)**: 667-671.

[246] Fisher, R. A. (1936) "The use of multiple measurements in taxonomic problems". *Annals of eugenics*, **7(2)**: 179-188.

[247] Muhammad, A. K. (2018) " Model selection, union and assembling in practical data analysis: methods and case study" *Doctoral dissertation*, University of Leicester.

[248] Vehtari, A., Gelman, A., & Gabry, J. (2017) "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC." *Statistics and computing*, **27(5)**: 1413-1432.

[249] Mirkes, E.M. Supervised PCA (2016). https://github.com/Mirkes/SupervisedPCA/wiki

[250] Mirkes, E.M. Supervised PCA (2016). https://github.com/Mirkes/SupervisedPCA

[251] Times Cited in WoS Core Collection. (27 August 2020). Available: https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Times-Cited-accessibility-and-variation?language=en_US

[252] Patience, G. S., Patience, C. A., Blais, B., & Bertrand, F. (2017) " Citation analysis of scientific categories". *Heliyon*, **3(5)**, e00300.