METADATA VISUALIZATION OF SCHOLARLY SEARCH RESULTS

TARANEH KHAZAEI

Metadata Visualization of Scholarly Search Results

by

© *Taraneh Khazaei*

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Sciences

Department of Computer Science

Memorial University of Newfoundland

October 2012

St. John's Newfoundland

Abstract

Studies of online search behaviour have found that searchers often face difficulties formulating queries and exploring the search results sets. These shortcomings may be especially problematic in digital libraries since library searchers employ a wide variety of information-seeking methods (with varying degrees of support), and the corpus to be searched is often more complex than simple textual information. To address these problems, an interactive Web-based library search interface is presented, which has been designed to support strategic retrieval behaviour of library searchers. This system takes advantage of the rich metadata associated with academic documents and employs information visualization techniques to provide searchers with additional information-seeking tools. These tools are designed to facilitate visual and interactive query refinement, search results exploration, and citation navigation. User evaluations illustrate the potential benefits of the design choices in comparison to a list-based digital library search interface.

Acknowledgements

Firstly, I would like to sincerely thank my supervisor, Dr. Orland Hoeber, for the support and guidance he offered me throughout my master's program. This research would not have been possible without his insightful advice and patient encouragement.

I would also like to acknowledge the financial, academic, and technical support of the Department of Computer Science and Memorial University. This research is financially supported by my supervisor's NSERC Discovery Grant and the scholarship that I have received from the School of Graduate Studies of Memorial University.

Finally, I would like to thank my best friend, my husband, who gave me unconditional support and encouragement throughout this process. My love and thanks also go to my parents and my brother for always being supportive of my academic pursuits regardless whether abroad or at home.

Part of this research was presented at the 2012 International Conference on Knowledge Management and Knowledge Technologies [71].

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

The ongoing development of automated search technology, along with the rapid growth of available information, has made search a fundamental part of peoples' lives [51]. Search systems are now capable of providing direct access to large information spaces, and are continuing to evolve and grow at a rapid pace. As a result, searching for information is now an integral task undertaken by people daily and is regarded as the second most frequently used online application [43]. As such, research on information retrieval systems has been an active area of study for years.

In the last few decades, improving retrieval and ranking algorithms has been a primary focus of information retrieval research while less attention has been paid to the human aspect of search systems [51]. Nevertheless, research on search behavior [104, 64, 101, 63] has shown that the current interfaces are not well matched with retrieval behavior, providing limited support for searchers during their search pro-

cess [49]. This lack of support can be especially problematic in digital libraries, where the existence of diverse metadata elements (e.g., title, authors, venues, keywords, and abstracts) has led to a wide variety of information-seeking methods in library searches. Meanwhile, digital libraries are becoming the main repository of mankind's knowledge, and the primary means of accessing them are search systems [12]. It has been reported that almost 70% of faculty and students use Web-based search engines daily to access North American research libraries, while about 90% of college students start their information seeking process with search engines [26]. Thus, in order to take advantage of the wealth of information existing in digital libraries, further research to improve digital library search interfaces is required.

In studies on search behaviour, it has become evident that queries crafted by searchers are often poorly formulated and do not reflect their information needs properly and accurately. This may be due to searchers' tendency to formulate short queries [67, 83, 104], their incomplete knowledge about their information needs [38], or their inability to express their information needs due to a lack of terminology [38]. Searchers' inabilities to specify accurate queries has long since been documented as one of the main issues in traditional libraries, with reference librarians reporting that few people know how to ask reference questions [69]. In traditional libraries, the active engagement of reference librarians in the search process may enable searchers to subsequently express their information needs properly. However, in the context of digital libraries, little assistance is provided for searchers to craft and reformulate their queries.

Similar problems exist with the interfaces for presenting search results, where the simple list-based format is commonplace. Such a representation requires searchers to

extensively utilize their cognitive abilities to evaluate and compare result items by reading document surrogates (i.e., titles, snippets or abstracts, and URLs) one-by-one. In addition, since the list-based representation provides only a few rudimentary interaction mechanisms, there is little support for navigation and exploration within retrieved documents. This lack of support is even more problematic for large and complex information structures such as content-rich metadata-enhanced digital libraries, where more advanced exploration features are needed to support different search activities.

Although an effective ranking method can help searchers for targeted queries, there is still a cognitive burden for exploratory searches. Exploratory search tasks are often motivated by a complex information need, a poor understanding of terminology and the information space [117], or a desire to learn [84]. Such conditions are common starting points for library searchers, resulting in their desire to initiate a search process.

Given that current search interfaces lack the ability to support exploratory search tasks, the design and study of useful interfaces which aid searchers in their search activities is a vital concern in the digital library context. Hence, with the aim of promoting searchers' information-seeking behaviour in an academic digital library, this thesis outlines a new approach called Bow Tie Academic Search. The main purpose of this approach is to support the key search activities of search results exploration and query reformulation. Bow Tie Academic Search moves beyond the traditional information retrieval interfaces, and instead provides a "metadata-enhanced visual interface" that aims to incorporate the human element into library search systems [99]. The main goal is to allow searchers to take an active role within the search process,

which is particularly beneficial for exploratory search tasks [49].

1.2 Approach

In this research, a structured approach has been employed to alleviate the problems associated with the current list-based search interfaces. As the first step, the existing literature on online search behaviour as well as library search strategies has been reviewed and surveyed to identify the difficulties searchers may encounter during their library search process. Next, the previous approaches that provide visual and interactive support for information retrieval tasks has been surveyed. Then, the knowledge gained from prior literature has been used to design a library search interface intended to promote searchers' information-seeking behaviour by utilizing effective visualization and interaction techniques.

Finally, an empirical evaluation was conducted with users in order to explore the effectiveness of the specific design choices in comparison to the current search interfaces. Moreover, this user study was designed to validate the hypotheses made regarding the use of visualization and interaction in search interfaces to support library retrieval tasks. Overall, a design-oriented research methodology [36] was employed in which the main contribution is the knowledge gained form studying and evaluating a designed artifact. This methodology is in contrast to research-oriented process in which the product development is the focus of study.

The resulting interface, called Bow Tie Academic Search, has been designed and developed to facilitate exploratory searches by providing interactive query refinement and interactive search results exploration support. Considering the significant role of

metadata elements for retrieval purposes [99] and the great potential of information visualization to improve user experience [110], Bow Tie Academic Search makes use of the rich metadata associated with academic documents and employs information visualization techniques to support searchers during their search process.

This system provides additional interface extensions to the conventional list-based format. Visual representations of citation metadata are added to the search results list, allowing searchers to visually scan, evaluate, and compare search result items in order to find the potential documents based on their citation characteristics. In addition, when a particular document is selected, searchers are provided with an interactive exploration tool to navigate within the citations of the document. The interactive query refinement is supported by providing a visual and interactive overview of the keywords associated with the top search results. These representations provide concise and compact visual encodings of the metadata, and were designed to be lightweight representations in order to prevent searchers from getting overwhelmed with visual complexity.

Incorporating such metadata visualizations into the library search interface supports searchers in their retrieval tasks by enhancing their abilities to perceive, interpret, and understand features and relationships among the search results and their associated metadata. In addition, by providing interaction methods, an effective combination of searcher control and system retrieval power can be achieved. This is an example of next-generation information retrieval systems supporting the search process beyond the simple query box and search results list [49, 51].

1.3 Research Questions

The main objective of this research is to address the fundamental issues associated with the library search interfaces that follow the traditional paradigm of query box and search results list through the use of visualization and interaction techniques. Since this approach moves beyond the traditional search paradigm, it leads to some fundamental research questions related to the value of the design and prototype implementation of Bow Tie Academic Search:

How can metadata visualizations and metadata-based interactions be designed to support interactive search results exploration and interactive query refinement in digital library search interfaces?

The value of information visualization to improve different aspects of digital libraries is well-recognized [123, 11, 8], with information access and retrieval being one of the potential directions. Even though a variety of interactive and visual tools have been designed and proposed to support information retrieval activities, not all of them have shown to be effective [43]. These differences can be attributed to the specific design choices used in these approaches, indicating the challenges and complexities associated with the design of useful and effective visual representations for search interfaces. Therefore, it is necessary to investigate the best ways to design interactive and visual tools that can effectively support search activities of search results exploration and query refinement. This research question will be answered via the design of Bow Tie Academic Search (Chapters 3 and 4).

What is the impact of Bow Tie Academic Search on efficiency, effectiveness, and subjective impressions of the search process?

Bow Tie Academic Search includes some interactive and visual tools that allow searchers to take an active role in their information seeking process. As such, it can be classified as an interactive information retrieval system. In human-computer interaction studies, a key research purpose is to investigate the usability of interactive systems [61], which is defined as "the capability to be used by humans easily and effectively" [98]. Therefore, one of the research questions of this thesis is to explore the impact of Bow Tie Academic Search on the usability measures of efficiency and effectiveness. In this research, "retrieval efficiency" refers to the time it takes for searchers to complete their retrieval tasks, and "retrieval effectiveness" refers to the quality of the search outcome.

Sometimes quantitative measures may not reflect the user experience, particularly when the search tasks are vague or ambiguous [50]. Since the approach for representation and exploration of search results is particularly targeted at exploratory tasks, subjective measures can also be helpful in validating the potential benefits that the system may provide. This research question will be answered by the user study (Chapter 5).

What are the searchers' perceptions of the usefulness and ease of use of each specific component of Bow Tie Academic Search?

Beside the possible impacts of the entire system on the retrieval process, it is critical to explore the potential benefits and drawbacks of each of the specific components of the proposed interface. The Technology Acceptance Model (TAM) [28] has been tested in many empirical studies and is considered a robust and reliable instrument to measure user's acceptance of an information technology [47]. As such, TAM measures are used to assess the searchers' perception of ease of use and usefulness of each of the

components of Bow Tie Academic Search. This research question will be answered via the user evaluation as well (Chapter 5).

1.4 Organization of the Thesis

The remainder of this thesis is organized as follows: An overview of the related work is provided in the next Chapter. Chapter 3 explains the approach used to support search results visualization and exploration, and Chapter 4 outlines the techniques used to support visual and interactive query refinement. The design and results of the user evaluation are provided in Chapter 5. The thesis concludes in Chapter 6 with a summary of the research contributions, along with an overview of future research activities.

Chapter 2

Related Work

2.1 Information Visualization

Information visualization is defined as the use of computer supported, interactive, visual representations of abstract data to reinforce human cognition [17]. The main purpose of information visualization is to transform abstract information into a visual representation that takes advantage of rapid processing capabilities of the human visual perception. Considering the abilities of human visual system in rapid interpretation of specific visual features, information visualization can be seen as an effective and a useful tool. As Ware [116] stated: "Combining a computer-based information system with flexible human cognitive capabilities, such as pattern finding, and using the visualization as an interface between the two is far more powerful than an unaided human cognitive process.", information visualization can be seen as a visual tool that aims at supporting the cognitive system of the user.

The human visual processing system operates so fast that it is for all intents

and purposes parallel processing of information. Information visualization systems use some basic visual features to represent multiple dimensions or attributes of the data. Examples of such features include colour, size, shape, and spatial proximity. These visual attributes can be used to devise an information intensive visualization; however, it should be noted that not every possible permutation of these features can be easily and separately decoded by the human visual processing system [115]. For instance, since colour and spatial location can be easily separated by the visual system, they can be utilized simultaneously to represent different attributes of the data, but the use of colour and shape for the same purpose is not as easily decoded [115].

There are a number of well-known theories and design principles that can be used to design effective visual representations of information. Based on human perceptual capabilities, Cleveland and McGill [25] have provided a ranking of the most effective visual features to represent quantitative information. Mackinlay [80] has extended this ranking to include more data types (i.e., ordinal and nominal). For instance, according to this ranking, position and length are the most effective visual features to encode quantitative data, while position and colour hue are the most effective ones to represent nominal information. When designing visual representations of information, this ranking can be used to assess the relative effectiveness and accuracy of alternative design choices.

In information visualization, the use of colour is extremely useful as it supports visual distinction of objects based on their colour differences [115]. Drawn from physiological features of the human vision, the *opponent process theory of colour* provides a solid foundation for effective use of colour to represent data attributes [115].

According to this theory, there are six elementary colours, which are perceptually opponent pairs along three axes. These colour pairs are black-white, red-green, and yellow-blue [115]. As such, these six distinct colours are the most effective choices when encoding nominal information with colour hue. In addition to this theory, suggestions by Tufte [109] can guide the colour selection process as he recommends the use of soft colours instead of using strong and bright ones.

The *pre-attentive processing principle* introduces a specific set of visual features that can be identified even after a brief exposure by human visual perception [115]. These features include visual form (e.g., line length, line width, and size), colour (hue and intensity), motion (flashing and direction of motion) [115], and spatial position (e.g., 2D position and stereoscopic depth) [115]. These visual attributes can be used to encode the most important aspects of information, facilitating their easy and instantaneous identification and distinction from the surrounding area.

The other fundamental aspect of computer-mediated visualizations, in addition to envisioning information, is interaction. Interaction is highly intertwined with visual representation as an interaction with a system may trigger a change in representation [122]. Interaction techniques can enhance users' cognitive abilities by allowing them to manipulate and control information being visually represented. Visualizations without interaction would become a static image which can only address a very limited number of tasks [31], and its usefulness may be adversely affected as the underlying collection becomes large and dense [122].

Yi et al. [122] categorized the interaction methods in information visualization into seven categories based on users' intention of performing a method. The first method is called *select* by which the user is able to mark a set of data items to keep track of

them in different stages of interaction. *Explore* method allows users to navigate and examine different subsets of large information spaces. *Reconfigure* provides users with the ability of viewing information from different perspectives by re-arranging the data points. *Encode* technique allows users to change the way that information is encoded such as transforming a pie chart to a histogram. *Abstract/Elaborate* enables users to specify the level of detail to be represented in the visualization. *Filter* method allows users to narrow the represented dataset down to a subset based on a criterion. Although filtering is categorized as a different technique, it can be seen as an exploration method since iterative filtering allows the investigation and exploration of different subsets of a dataset. Finally, *connect* shows relationships among data items or shows the hidden items related to a specified one. An example of this category is a brushing and linking technique using which allows identification of a selected data item in different views of a dataset.

In order to design any visual interface, special care should be taken to select and use both effective visual features and useful interaction techniques as two core components of information visualization. Even with the guidance of the aforementioned theories and principles, designing useful interactive and visual tools is a challenging task that requires careful consideration and investigation of the potential effects of choosing a design alternative on the effectiveness of the other design choices. Due to the complex interaction among design choices, in addition to design development, conducting thorough evaluations through user studies is now an expectation in the information visualization field.

2.2 Citation Visualization

In library collections, academic documents are often linked together as scientific literature refer to or cite one another. The information associated with such relationships has always been considered valuable as it can be used to assess the importance of authors, documents, and topics; and to interpret how these elements relate to each other over the time. Although displaying citation information is a challenging task, a number of approaches have been proposed to visually represent documents' citations, their relations, and their characteristics.

The most common way of representing citation information is the use of node-link diagrams, which have been used for the purpose of citation analysis for years [39]. Despite efforts to create more innovative views of node-link diagrams to represent citation information [102, 7], there are inherent issues associated with such views for large and dense datasets, and it has been shown that users are not particularly comfortable with them [113]. In addition, citation graphs fail to provide other critical metadata, such as documents' titles and authors within the graph representation, and they normally require user actions to provide such information on demand. As such, users' initial evaluations and comparisons can only be based on a small set of metadata, which may adversely affect information seeking efficiency and effectiveness.

Therefore, some researchers have moved beyond node-link diagrams, and employed other visualization methods to represent citation metadata. For example, Butterfly [81], which is a 3D search interface, uses a butterfly layout, where a document metadata is shown at its head in a textual format (see Figure 2.1). One wing of the butterfly includes a list of backward citations (i.e., documents referenced by the orig-

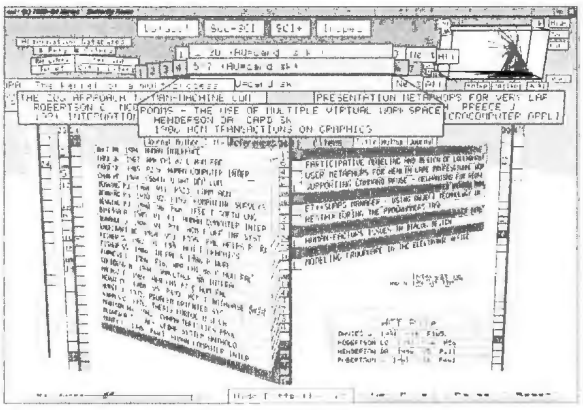


Figure 2.1: A screenshot from Butterfly [81].

inal one), while the other wing lists document's forward citations (i.e., documents that cited the original one). Colour is used to encode the source database as well as the number of forward citations for each document, and to show if a document was already visited. In addition to the general problems users may face in understanding 3D representations [37], Butterfly has scalability issues as each wing can only represent 22 citations.

In PaperLens visualization [77], rather than explicitly showing the relations among documents, simple views of different pieces of information, such as "year by year top 10 cited documents" and "popularity of topic by year", are provided (see Figure 2.2). These views are tightly coupled with brushing and linking interaction methods, providing searchers with a powerful tool to recognize influential documents and to understand trends and topics in a field. However, searchers are not able to understand the citation characteristics of individual documents that are not in the top 10 cited documents, or to navigate within the citations of the documents.

Another attempt is the CiteWiz [35] interface, which consists of three different

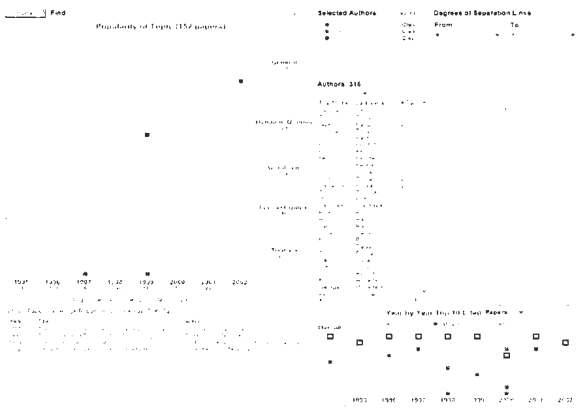


Figure 2.2: A screenshot from PaperLens [77].

views including timeline displays that show the general chronology and importance of documents and authors in a citation network, and a node-link diagram of keyword and authorship metadata to let searchers gain insight into this metadata. In addition, the growing polygons technique [34] is adopted and enhanced to represent the citation information of a particular subset of documents. These visualizations are augmented with some interaction techniques to support navigation among the citation network and details-on-demand of the entire citation chain for a document of interest. Even though preliminary user studies of CiteWiz provided positive results, the citation visualization component becomes difficult to understand as the selected document set grows.

Even though these approaches replace common node-link diagrams with novel visualization methods to represent citation metadata, there are still complexity and scalability issues that need to be addressed. Considering the rapid growth of scientific literature, and the value of citation metadata for evaluation, comparison, and navigation purposes, compact and scalable visualizations are required to support document

discovery and navigation in digital library collections.

2.3 Search Results Visualization

In recent years, many visual interfaces have been designed and developed to provide a better representation of search results and to support exploration and navigation within the search results set. These interfaces employ a number of different approaches when visualizing the search results, including augmentation of the list with visual representations that encode query-document relationships, spatialization of documents in a 2D or 3D interface, representation of document collections (flat or hierarchical), and visualization of auxiliary information derived from the search results set. Each of these approaches will be explained in more detail in the sections that follow.

2.3.1 List Augmentation

In some studies on search results visualization, researchers have proposed adding interface extensions to the conventional list-based interface in order to address its shortcomings to some degree. In most of these approaches, the search results list is augmented by adding small visual representations alongside each document, with each representation visualizing the relation of query terms with that document.

One of the early attempts at list augmentation is the TileBars interface [45] which simultaneously and compactly shows relative document length, query term frequency, and explicit query term distribution in a full text information access environment (see Figure 2.3). In the TileBars visualization, a rectangular icon is shown beside each search result item. The rectangle length represents the relative length of the

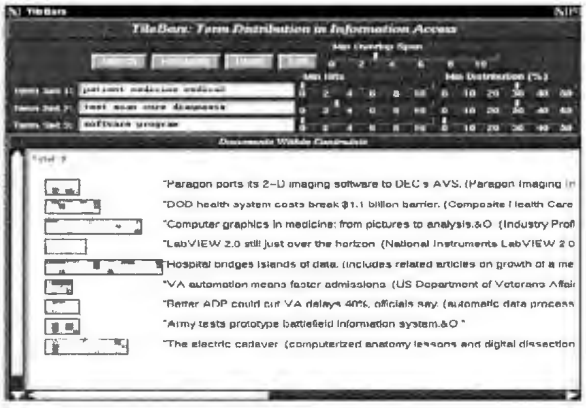


Figure 2.3: A screenshot from TileBars [45].

document. It is further subdivided into columns representing document segments (using paragraphs, section breaks, or units chosen by TextTilling algorithm [42]); each row is also considered as a representative of each query term. Then the query terms frequency within each segment is represented using grey scale encoding.

In the work by Heimon and Jhaveri [46], a small document shaped icon is presented on the left side of the result item. This icon contains four equal sized rows representing document sections (independent of document size). The number of occurrences of the all query terms within a 20-word window for each section is depicted using the same method as in TileBars [45]. Both of these approaches require access to the full textual contents of the document, which is inefficient when the underlying search engine is inaccessible.

Another work intended to support interactive exploration of search results is HotMap [54, 58], which visualizes the frequency of each query term in each document surrogate by a colour-coded square located alongside the corresponding result item (see Figure 2.4). In addition, each query term is shown vertically at the top

Figure 2.4: A screenshot from HotMap [54, 58].

of the corresponding column of squares enabling the searcher to easily understand which square is related to which query term. Furthermore, searchers are able to perform a nested sorting on the search results by clicking on the query term labels. It also offers a zoomed-out representation of search results allowing searchers to see and compare how their query terms are being used across the set of documents. In similar approaches, the relations between query terms and documents are shown using a colour-coded pie chart [2] in which segment size indicates the relative frequency of a query term, and a bar chart [112] in which overall and single keyword relevance is shown using the length of bars.

WaveLens [92] also intends to let searchers take an active role during their search process by allowing them to dynamically zoom into the document surrogate of interest through focus+contex representation. Focus+context visualization allows searchers to see the result item of interest in full detail, while at the same time the surrounding items are shown in less detail.

Another variation on the idea of enriching results lists by displaying query-document

relationships is to show miniaturized and small versions of the visual appearance of the document, known as thumbnails. Thumbnails typically include highlighted colour coded query terms [91]. While this approach provides little support for the manipulation and exploration of search results, it has shown to be beneficial when searchers want to re-find a document from a previous session [119].

These approaches take advantage of the list simplicity, consistency, and scalability, and add extra visual representations in order to better support search activities. For search interfaces, which are being used by millions of people daily, a drastic change may cause problems of adoption [43]. One of the notable advantages of list augmentation methods is that they avoid this issue by keeping the list as the main part of the search interface. However, in these approaches, designing information-bearing visual representations is a challenging task due to the space limitations.

2.3.2 Document Spatialization

Many researchers have proposed spatialization of documents or document surrogates to a 2D or 3D visual overview, wherein spatial proximity indicates documents' similarity [43]. The main differences between these approaches is how they specify and calculate document similarity and how they organize and represent the documents in 2D and 3D spaces. In most of these interfaces, documents are represented as small glyphs mapped to a specific point in the spatially-oriented interface based on two or three of their attributes. In addition, more attributes of the documents can be depicted through visual features of the glyph itself such as its colour, shape, and/or size.

Approaches using spatialization mainly use a 2D scatterplot where documents are plotted based on two of their attributes corresponding to the x and y axis. Early attempts include systems such as xFind [3], Envision [90], and FilmFinder [1]. As these interfaces may overwhelm some searchers, Shneiderman et al. proposed a simplified display called Hieraxes [100] (see Figure 2.5). This display uses categorical and hierarchical axes in which documents are represented in a grid using either a set of colour-coded dots or a bar chart. Also, the mapping of particular attributes to visual representations such as the x-axis, y-axis, icon size, and icon shape, is customizable using drop-down menus.

In Citiviz [68], an interactive animated scatterplot is used with a hyperbolic tree in a single interface, wherein each document is represented in the scatterplot by a tower of colour-coded cylinders (see Figure 2.6). Each level of a tower represents a category to which a document belongs; therefore, the taller a tower is, the more categories the document belongs to. Citiviz uses a city skyline metaphor in which documents metadata can be shown simultaneously; however, there are problems with

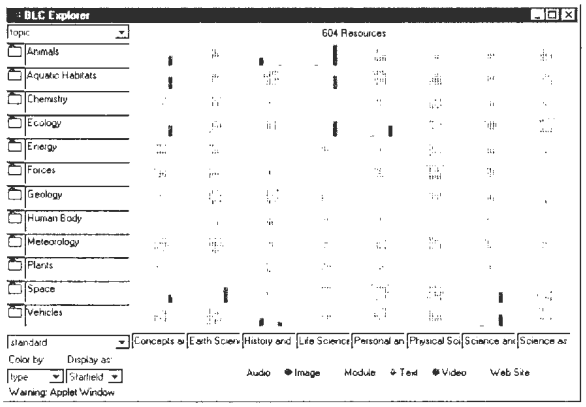


Figure 2.5: A screenshot from Hieraxes [100].

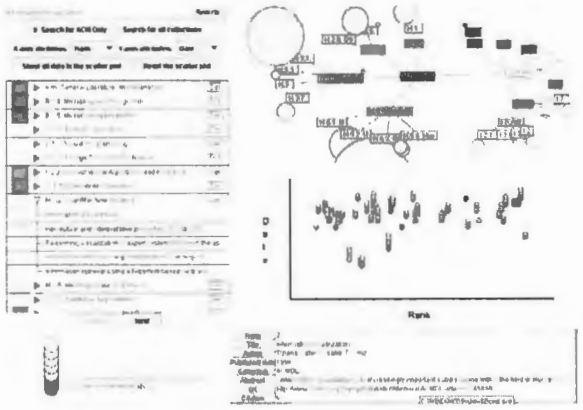


Figure 2.6: A screenshot from Citiviz [68].

occlusion and navigation within the space.

In two other related approaches, documents are organized and mapped around the center of a circle based on their similarity to the query while their proximity to the center indicates their relevance to the query. In DART [23], the circular space contains several concentric circles so that the searcher can easily evaluate and compare the distance of a document from the query in the center. This space is divided into pie shaped sections, each representing a predefined cluster. As such, documents are mapped to the display based on both their similarity to the query, and the cluster to which they belong to.

RankSpiral [105] uses the average rank issued by multiple search engines as an indication of relevance. Then documents are mapped and organized in a spiral in which the most relevant documents are closer to the spiral center. Also, other visual features of the glyph representing each document are utilized to depict which search engine the document is coming from.

Document spatialization approaches have been proposed frequently, and they are

intended to help searchers gain insight into the relationships between documents in the search results set, grasp potentially unexpected patterns in document collections, and find important documents that otherwise might be missed [43]. However, these kinds of graphical overviews of large document spaces have to be proven useful and understandable by searchers. Evaluations conducted so far mostly indicate that searchers face difficulties understanding such spatialized representations [43].

2.3.3 Document Cluster Visualization

Some attempts have been made to organize search results into meaningful groups, and then visually represent these collections in order to help searchers gain an overview of the search results and easily determine their next step in the search process. One way to classify a document space is through clustering. Document clustering refers to the grouping of documents based on some measures of similarity. Some of the clustering algorithms create hierarchical clusters. This section deals with non-hierarchical clustering while the following section deals with hierarchical clustering as well as other approaches for generating hierarchies from search results.

In most of the systems described in the previous section, in addition to document spatialization, the thematic groups or topics are derived from the text based on some measurements and are further displayed graphically by adding visual cues. For example, in Envision [90], similarly sized elliptical icons represent a set of relevant documents while the number of documents is shown inside the ellipse, and labels below these icons indicate the rank of the two most relevant documents in the cluster (see Figure 2.7). In the enhanced version of Envision for digital libraries [114], the

size of the cluster ellipse is related to the number of documents it contains (see Figure 2.8). In addition, document icons are shown in the cluster icon, allowing searchers to select an individual document from within the cluster. Documents are placed in the cluster icon by locating them on concentric ellipses with diameter differences of equal magnitude from outside in.

In xFind [3], the VisIsland interface was developed in addition to the scatterplot and a list-based representation to display thematic clustering of search results. First,

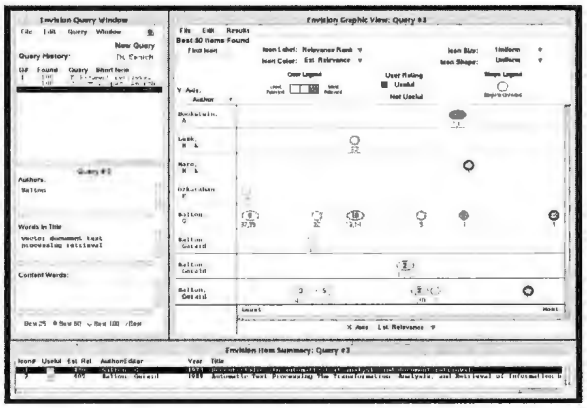


Figure 2.7: A screenshot from Envision [90].

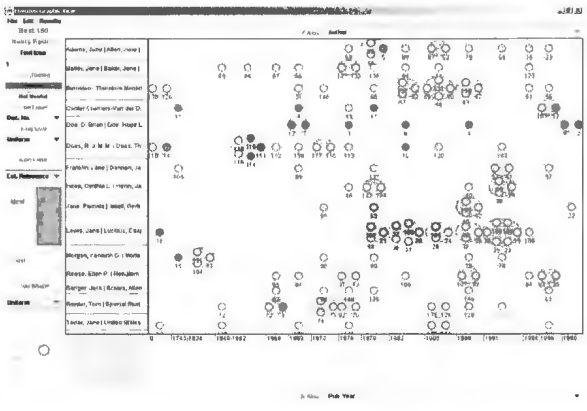


Figure 2.8: A screenshot from the enhanced version of Envision [114].

cluster centroids are randomly mapped to the display, and then documents are located in a ring around the corresponding cluster while documents that are more similar are placed closer to each other. This procedure finally leads to a representation in which related topics are displayed as islands.

Kohonen's self-organizing feature map algorithm [72] is used to organize document collections in number of studies [21, 78]. For example, in [78], self-organizing clusters are shown as adjacent polygons in a 2D map in which their size and shape indicate how frequently documents are assigned to the corresponding cluster. The adjacency of regions reflects semantic relations of clusters within the collection. Moreover, when searchers hover the mouse cursor above any polygon, a pop-up window will be displayed showing the titles of documents closely linked to the corresponding cluster.

Clustering can clarify new and interesting patterns and trends hidden in the document space, and it can be done automatically on any text collection. However, choosing understandable and descriptive labels for clusters is a challenging task, which can be problematic in search results exploration [43]. Moreover, clustering may make it difficult for searchers to compare documents within different clusters, and searchers may neglect very small clusters even though they might contain the most relevant documents.

2.3.4 Document Hierarchy Visualization

Another method to classify document collections is the use of category systems. In category systems, documents are assigned to some organized and meaningful labels

that represent the domain concepts. Category system structures are usually either hierarchical, or faceted (which is discussed further in Section 2.4.1). As mentioned, some clustering algorithms also build a hierarchy of clusters. Hierarchical clustering is often considered as the better quality clustering approach, but it is computationally expensive as the size of the collection increases [106]. Although category systems are only applicable in well-structured collections and their automated methods are about 75% correct on average [97], their superiority to clustering methods has been shown in terms of usability [43].

Traditional methods of presenting hierarchical information, namely listing, outlines, and static tree diagrams are not feasible because the extraction of information from large hierarchies is quite difficult since the navigation is a great burden, and contents of information are often hidden within nodes [65]. In addition, visualization of large hierarchies in a limited-size screen is a serious challenge.

One of the fundamental ways of hierarchy representation is the table-of-contents view used in books and other information systems [43]. Such a tree-based hierarchy outline has been used with hyperlinked Web pages to support both search results organization and navigation [22, 20]. More sophisticated variations of tree diagrams have been also proposed and used in search results visualization. For example, Cat-a-Cone [44] makes use of available subject headings in library systems and represents categories associated with highly ranked documents in the search results as well as their ancestors and siblings using a 3D ConeTree [94] (see Figure 2.9). In Citiviz [68], a hyperbolic tree is used to show the hierarchical structure of documents using the ACM Computing Classification System.

One of the well-known techniques for representing hierarchical information is the

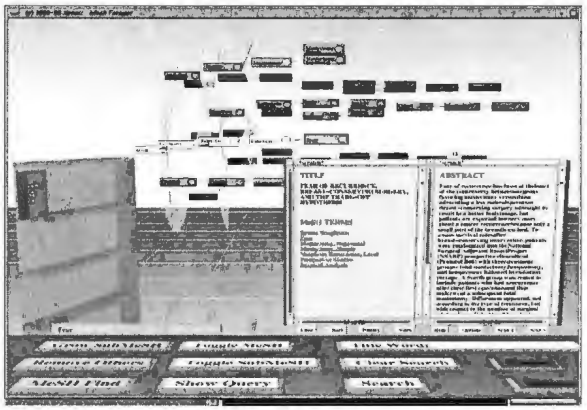


Figure 2.9: A screenshot from Cat-a-Cone [44].

use of a TreeMap [65]. The original algorithm for TreeMap creation is based on dividing a rectangle with vertical partitions to the number of branches from the root, and then performing the same process for emerging rectangles, but with horizontal partitions. This recursive approach continues until the leaves of the tree are reached. Further, the area of each leaf is specified based on the amount of information stored there. In search results visualization literature TreeMap is used and evaluated in a number of studies [41, 13, 24].

For example, in ResultMaps [24], a TreeMap has been used to encode a digital library's full content, rather than encoding relevant items to a query, according to the available hierarchical taxonomy classification (see Figure 2.10). Presenting the entire repository hierarchy let searchers gain knowledge about the whole library content as a side effect, which can be beneficial for their future information seeking tasks. The TreeMap view and search results list are linked via brushing techniques, allowing the searcher to highlight documents in the search results list to see where they are in the digital library hierarchy, or to highlight regions of the digital library hierarchy to

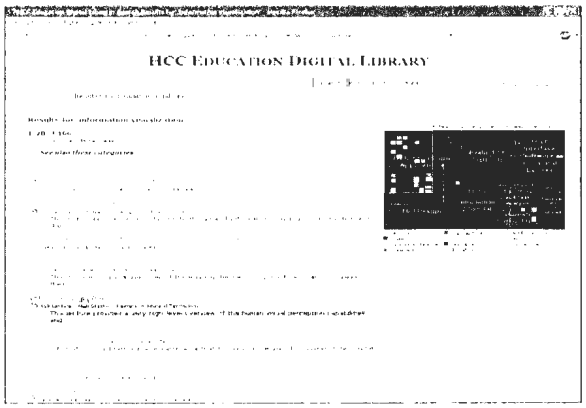


Figure 2.10: A screenshot from ResultMaps [24].

show those documents in the search results that match this information.

As TreeMap is a space-filling approach, it can show large trees containing thousands of branches, and the amount of information on each branch of the tree can be easily visualized, which makes it suitable for library collections. In addition, navigation among branches is fairly easy. However, in TreeMaps, the hierarchical structure is not clear [115], and they have not been proven to be successful for representing textual information [43].

2.3.5 Visual Representation of Auxiliary Information

In another set of approaches to support search results exploration, rather than visualizing documents themselves, a visualization of auxiliary information derived from the search results set is offered to the searcher independent of the results list. Those search systems augmented with such interface extensions often include interaction tools to let searchers dynamically filter or re-sort search result items. Moreover, many of these visual representations of auxiliary information support query formulation and

reformulation, which will be discussed in more detail in the next section.

Tag clouds allow searchers to quickly perceive frequent and potentially valuable terms used in the text. In the information retrieval context, tag clouds have been used on the available text material related to the highly ranked documents, such as document surrogates, in order to provide a visual summary of the search results. For instance, PubCloud [75], supports the visual exploration of search results from the PubMed database of biomedical literature by providing tag cloud representations of documents' abstracts. In addition, selecting a term allows the searcher to navigate to the relevant subject matter that otherwise might be hidden far down in the ranked list. Tag clouds have also been used for explaining the underlying personalization approach [52] in which searchers are able to see the rationale behind the personalized ranking, and interactively re-rank the search results list by clicking on the term found as the relevant one to their information need.

In VisGets [32], in addition to tag clouds as a topical overview of the results set, two other visual representations are designed to show a summary of the amount of information published in different periods, and the geographic location of search result items. These three coordinated views allow searchers to identify the interrelationships between these attributes while performing multidimensional exploration in the underlying collection.

In WordBars [53, 57], a histogram is used to visually depict the most frequent terms found in the highly ranked search result items retrieved by the initial query (see Figure 2.11). This vertically oriented histogram, which is incorporated in the main interface, allows the searcher to grasp an overview of the frequent and potentially prominent topics used in the search results. In addition, if the searcher clicks on the

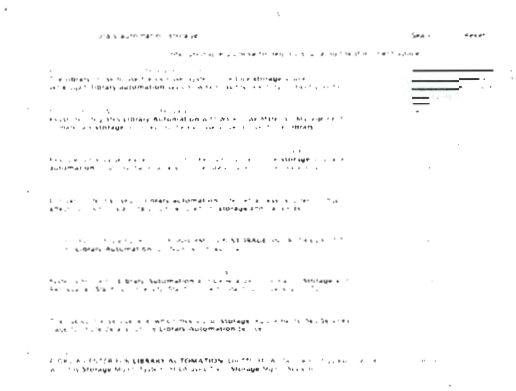


Figure 2.11: A screenshot from WordBars [53, 57].

potential term, the search results set is re-sorted based on the frequency of the selected term within the titles and snippets of the search result items. This feature supports interactive search results exploration process. The interactive exploration capability of this system is empowered in subsequent modifications as the searcher is able to highlight relevant and non-relevant terms in WordBars2 [56], and visually specify a utility function in WordBars3 [48], both within the histogram view.

The visual representations of auxiliary information provide searchers with a summary of search results, allowing them to interpret the overall properties of the search results set. Providing such an overview may allow searchers to quickly evaluate whether the retrieved set fulfill their information needs. Even though these tools usually provide exploration support as well, their exploration feature is of little value for those searches where the initial query is poor and no relevant document is retrieved.

2.4 Query Representation and Visualization

The first step of the information retrieval process is query formulation or query specification, in which searchers convert internalized abstract concepts into language, and further convert the expression of language into a query format in order to express their information needs [43].

Since searchers are often unable to craft and formulate queries that accurately reflect their information needs [43], a number of approaches exist to visually help them reformulate their initial query to provide a better description of the information they are seeking. In general, there are two types of approaches to visually support query reformulation: visualizing faceted navigation and providing additional relevant terms for the purpose of query expansion.

2.4.1 Faceted Navigation Visualization

One of the main research directions for the application of information visualization to queries is on faceted search systems. Faceted search systems use faceted classification in which documents can be assigned to any number of categories (flat or hierarchical) based on the existing metadata within the collection. These categories are referred to as facets. In faceted search, searchers are able to navigate within the underlying collection by applying different orders of the facets. In addition, it allows searchers to elaborate their queries progressively, seeing the effects of each choice in one facet on the available choices in other facets [111]. Since this approach makes use of available metadata to filter the search results, it can be seen as an interactive query construction method.

There are number of studies that provide visual and interactive support for searchers to choose better facets and to easily navigate within the collection. Flamenco [121] makes use of hyperlinked Web pages to facilitate navigation within facets (see Figure 2.12). Every hyperlink is displayed with the number of results that will be shown by choosing that facet. Flamenco aims to keep the interface uncluttered and simple by hiding links leading to zero results and using simple point-and-click interaction techniques.

Relation Browser [124, 16] is another faceted search system that provides extra visual and interactive features to allow better use of facets. This system uses a graphical bar to show the relative frequency of each category in the results set. One of the notable advantages of Relation Browser is that it allows the searcher to quickly explore a document space using dynamic queries that can be issued by hovering over facet items.

The mSpace [79] system uses a restricted version of faceted navigation in order to manage high-dimensional spaces. The searcher is allowed to navigate via only one

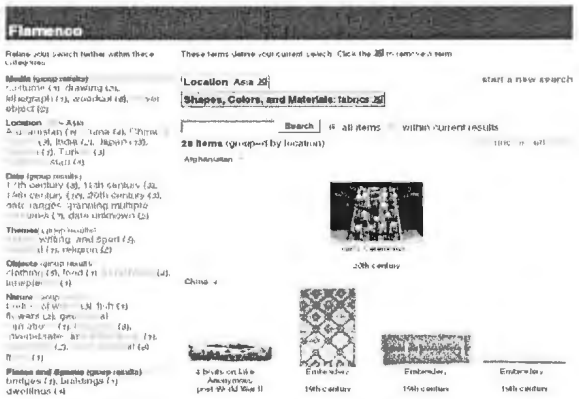


Figure 2.12: A screenshot from Flamenco [121].

category within each facet at a time by using a left-to-right set of columns, where the left-most column is the top level of the hierarchy and the right-most is at the bottom.

VisGets [32] supports searchers to dynamically construct their queries by providing a set of coordinated visual representations designed to represent three major dimensions of information, namely publication date, geographic location, and author-supplied tags. Searchers are able to incrementally specify their queries by manipulating these three aspects visually, and immediately see the visual evidence of the query manipulation in provided views.

Although faceted search systems have shown great potential, and have been used successfully over some document sets, there are some challenges for assigning quality meta-data to every document in large collections and for applying the same meta-data to every result and every query in heterogeneous collections [107].

2.4.2 Candidate Terms Visualization

One of the most important query reformulation techniques is query expansion, which is the process of appending meaningful terms to the initial query in order to obtain useful results. Query expansion can assist searchers in their query reformulation stage by using either automatic or interactive techniques. In interactive query expansion, automatically derived terms are offered as candidates to the searchers, allowing them to choose the terms of interest for the query modification purpose.

Different ways have been proposed to extract the most useful and relevant terms for the purpose of query reformulation. One of the major techniques is relevance feedback. In the relevance feedback process, searchers evaluate retrieved documents, and

mark those considered to be relevant. Then features derived from selected documents are used to modify the initial query. This method places the burden of relevance judgment tasks on the searchers, and might not be desirable from the searchers' point of view.

Potential query terms can be also extracted from knowledge structures. In these approaches, queries are simply expanded by looking up related terms in the appropriate structure. Knowledge structures can be either collection-independent such as manually constructed thesauruses or ontologies (e.g., Wikipedia and WordNet [87]), or collection-dependent which are constructed automatically by discovering words relationships based on corpus analysis. This method is also referred to as global analysis technique. Different ways have been proposed to construct these structures such as stemming, clustering, and term co-occurrence. The effectiveness of collection-dependent knowledge structures highly depends on the corpus structure as it only works if there are sufficient relevant documents to work with, and also that these documents contain a reasonable set of terms that represent the subject area for the queries [10].

In local analysis, instead of the whole corpus, the top-ranked documents retrieved by the original query are involved in the term extraction process. Similar to global analysis, different statistical techniques have been used to choose potential terms from retrieved documents. For example, the frequency of terms used in the initial search results set has been considered as a good criteria to choose and weight those terms as useful candidates for query expansion. Xu and Croft [120] compared local and global analysis using different query sets and corpora, and they found that although global analysis has some advantages; local analysis is generally more effective. However, it

should be noted that their effectiveness depends on the quality of the initial query.

Query logs have been also exploited as a valuable source from which expansion terms can be selected. Many search engines are accumulating a large amount of query logs daily. These logs typically contain searchers' IP address, submitted queries, timestamps, and the URL of the clicked documents. If searchers click on a result item, it is likely that the document is relevant to the query at least to some degree. Thus, search logs can indicate documents' relevancy implicitly. As such, some probabilistic correlations between query terms and document terms can be established and further used for selecting expansion terms from documents for similar queries [27, 88].

In interactive query expansion, in addition to choosing the most effective and efficient approach to extract candidate terms, it is necessary to explore different methods to effectively present these terms to the searcher. However, in current search systems, extracted candidate terms are often presented in a simple list in which searchers are just able to select a term of interest in order to add it to the initial query.

There are number of studies focusing on interactive query expansion, and exploring visual and interactive ways to present the candidate terms to the searcher. For instance, Joho et al. [66] used cascading menus to visually represent hierarchically structured expansion terms (see Figure 2.13). These terms are automatically derived based on co-occurrence of terms from a set of documents retrieved from an original query.

In VisiQ [59, 60] a knowledge base containing relationships between terms and concepts has been used to visually represent the relationships between the initial query and suggested terms. The system first searches the knowledge base to derive relevant concepts to query terms. Then additional terms that have been used to

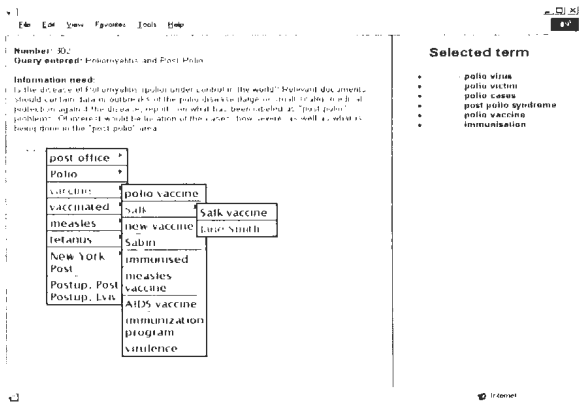


Figure 2.13: A screenshot from the work of Joho et al. [66].

describe these concepts are extracted. VisiQ further generates a graph-based visual representation of these terms and concepts, allowing searchers to quickly understand how their initial query is related to other concepts and terms. Searchers are able to add terms to their query by double-clicking on the node representing that term. By doing so, the visual representation is updated by newly extracted concepts and terms.

WordBars [53, 57], which was also discussed in Section 2.3.5, supports interactive query refinement. Using a histogram, which contains all the unique words found in the top search results, the searcher is able identify the potential terms and interpret their relative frequencies. In addition, the re-sort feature allows searchers to see and evaluate whether the selected term can bridge the gap between the issued query and their information need. Searchers are able to remove and add terms to the query from this histogram view.

It has been argued that interactive query expansion can significantly increase retrieval effectiveness [82] as it supports recognition of relevant terms instead of requiring searchers to remember them. In addition, these approaches incorporate human

control and decision making in the process of query reformulation, which is inherently a human activity [53]. However, by providing a static and simple list of terms, searchers may not be able to make effective decisions [82, 96]. As such, it is necessary to investigate potential visual and interactive methods to let searchers choose effective terms among the provided alternatives. Interactive query expansion methods may not be useful where the searcher is unfamiliar with the domain terminology.

2.5 Discussion

In this Chapter, the benefits and advantages of utilizing information visualization to enhance human cognitive abilities were discussed. In addition, primary design principles and guidelines that can assist in effective visualization and interaction design were explained. Moreover, an overview of the previous approaches that provide citation visualization, search results organization and visualization, and query representation, along with their advantages and disadvantages, were provided. The purpose of this Section is to explain and justify the design decisions made in this thesis in light of these studies.

As explained in Section 1.2, Bow Tie Academic Search provides visual representations of available metadata elements in library collections. In order to provide effective metadata visualizations, the guidelines provided by Mackinlay [80] are followed to represent different types of metadata. In addition, the opponent process theory of colour [115] as well as Tufte's suggestions [109] guide the colour encodings in the system. As well, to represent more important metadata elements, the pre-attentive processing theory is used to choose those visual attributes that can be pre-attentively

perceived.

These visual representations are further augmented with interaction methods to let searchers perform more information retrieval activities. Following the classification provided by Yi et al. [122], the abstract/elaborate method is provided to enable searchers to view a more detailed representation of a document. The explore method is used to let searchers navigate within and examine different subsets of the underlying collection. This exploration feature is facilitated by the support provided for iterative filtering of documents. In addition, searchers are provided with the connect method using which enables them to view a different representation of a document. These interaction methods are facilitated by simple and easy to learn low-level interaction mechanisms such as the well-known point-and-click technique [95].

As noted in Section 2.2, a large portion of citation visualization studies provides citation graphs using node-link diagrams. Even though citation relations can convey valuable information to the searcher, there are many issues associated with their node-link representations (as discussed in Section 2.2). To overcome these problems, Bow Tie Academic Search provides an abstract visual representation of the citation graph associated with each document. This visual representation is intended to allow easy and fast comparison and evaluation of documents based on their citation information.

In addition to the abstract visualization of citation metadata, Bow Tie Academic Search provides a detailed representation of this information. This representation is similar to the approach provided by Buttery [81] (see Section 2.2). However, it uses a 2D layout to avoid the problems associated with 3D representations [115]. Moreover, it includes extra visual and interaction methods to facilitate more search activities and to alleviate the scalability problem of Buttery.

Section 2.3 provides an overview of the studies on search results visualization. As discussed in Section 2.3.2, searchers normally face difficulties using spatially-oriented visual representations of documents. In addition, as discussed in Sections 2.3.3 and 2.3.4, there are some drawbacks associated with visualization of document collections in search interfaces. Considering the issues associated with such visual representations of search results and the advantages of the list augmentation (discussed in Section 2.3.1) and visualization of auxiliary information (discussed in Section 2.3.5), these two sets of approaches can be considered as the most effective ones. Hence, in Bow Tie Academic Search, the list-based format is augmented with the abstract encodings of documents' citation information. Moreover, keyword metadata associated with the highly ranked documents is used as auxiliary information and is visually represented independent of the list.

In Section 2.4, different techniques to visually support query formulation and reformulation are reviewed. Since there are some issues in extending faceted search systems to heterogeneous and large information spaces, candidate terms visualization is chosen to support visual query refinement. Among the term extraction techniques (discussed in Section 2.4.2), local analysis has been shown to be more effective as it is query-oriented. Therefore, Bow Tie Academic Search uses the keyword information of the top retrieved documents to provide candidate terms to the searcher. These terms are provided in a visually enhanced list to facilitate an effective decision-making process. In addition, intuitive interaction mechanisms are added to the terms using which searchers are able to modify their queries.

Chapter 3

Interactive Search Results Exploration and Discovery

3.1 Motivation

Similar to Web search interfaces, the main method of presenting library search results is through a list-based format, wherein documents are ordered based on their relevance to the query, and then are textually represented based on a set of their metadata elements in a list. This list-based representation of search results can help searchers where the target is well known but may not be effective for exploratory search tasks due to the little support it provides for comparing documents, finding multiple relevant documents, and exploring within the retrieved set of results [58].

In addition, the representation of search results in a textual format requires searchers to utilize their cognitive resources to read the document surrogates one-by-one and to evaluate them individually. This cognitive load may prevent searchers from

finding the most relevant documents buried deep in the ranked list. Searchers may find this shortcoming more problematic in digital libraries as studies on information-seeking behaviour indicate that library searchers may employ a variety of information-seeking methods to explore document collections [33, 4].

To address these shortcomings, a library search interface, called Bow Tie Academic Search, has been designed and developed. This interface provides an augmented list of search results, wherein the citation metadata of documents are visually provided. The motivation for this work is based on the value of citation metadata for evaluating and comparing library documents. In addition, since browsing within citations is known as one of the most commonly employed information-seeking strategies in library collections [33], searchers are provided with an interactive tool to navigate within backward and forward citations of documents.

In this Chapter, in addition to the data source features, the visual and interactive components of Bow Tie Academic Search that support search results visualization and navigation are described in detail. Where necessary, illustrative examples are provided to depict the potential benefits these tools may provide in the retrieval process. In addition, the implementation details are described. At the end of the chapter, the potential advantages, issues, and limitations of this approach are discussed.

3.2 Approach

3.2.1 Data Source

Bow Tie Academic Search is a meta-search engine that uses Microsoft Academic Search (MAS) API [85] data and functions to provide scholarly search results to the library searcher. Even though numerous digital library sources exist such as DBLP [29] and IEEE Xplore [62], MAS is one the few publicly available digital libraries that provides information on both backward and forward citations. Since MAS is still undergoing development, sometimes the metadata is incomplete. However, considering the wide range of metadata elements it provides as well as its fast improvement in covering more information, MAS can be seen as the best option among the alternatives.

In response to the searchers' queries, MAS provides various metadata associated with different entities such as documents, authors, and venues. Bow Tie Academic Search primarily focuses on the use of the metadata associated with the document entities. Table 3.1 includes the metadata elements that MAS provides to describe documents, along with their descriptions. It also shows which of these metadata elements are included in the design of Bow Tie Academic Search. Note that the descriptions are derived from MAS API User's Manual [86].

Metadata	Description	Usage
ID	A unique identity among all documents.	No
Forward Citation Count	The number of forward citations.	Yes
Backward Citation Count	The number of backward citations.	Yes
Forward Citation List	The list of documents that have cited this document.	Yes
Backward Citation List	The list of documents cited by this document.	Yes
Type	The type of this document; for example, paper, book, poster, etc.	No
Year	The publication year of this document.	Yes
Title	The title of this document.	Yes
Abstract	The abstract of this document. When a full text search term is provided, only a contextually relevant portion of the abstract is returned. In other cases, only the first 60 words of the abstract is provided.	Yes
Author	The authorship of current document.	Yes
Venue	The venue where this document is published.	Yes
Keyword	A list of author-specified keywords of this document.	Yes
URL	A list of URLs that can be used to download this document.	Yes

Table 3.1: Metadata elements that MAS API provides to describe document features.

3.2.2 List Augmentation using Bow Ties

3.2.2.1 Visual Representation of Citation Metadata

Representing citation metadata in library search interfaces can provide valuable support in finding seminal and important published works [76]. In addition, citations can be used to analyze research trends, identify new or active areas of science, and find out where and how often a particular article is cited [76]. However, current systems do not provide an intuitive way for searchers to evaluate and compare documents based on their citation information or navigate and explore through citations of a document.

As discussed in Section 2.2, previous research mostly used textual information or node-link diagrams to represent citation information. In this research, we opted for a simpler approach to enable searchers to compare documents with relative ease and accuracy. In this approach, a visual representation of citation information is located alongside each document as an interface extension to the well-known list-based representation.

In order to provide a compact visualization of citation information, details of the citation graph for each document are abstracted away, generating a representation that takes the form of a bow tie (Figure 3.1). This bow tie representation conveys time in two different levels of detail; therefore, it consists of two different timelines: A high-level timeline which is mapped to the horizontal dimension, and a low-level one which is mapped along the vertical dimension. In this representation, the document node is placed in the middle, containing the year that the document was published. The left side represents information about backward citations, while the right side

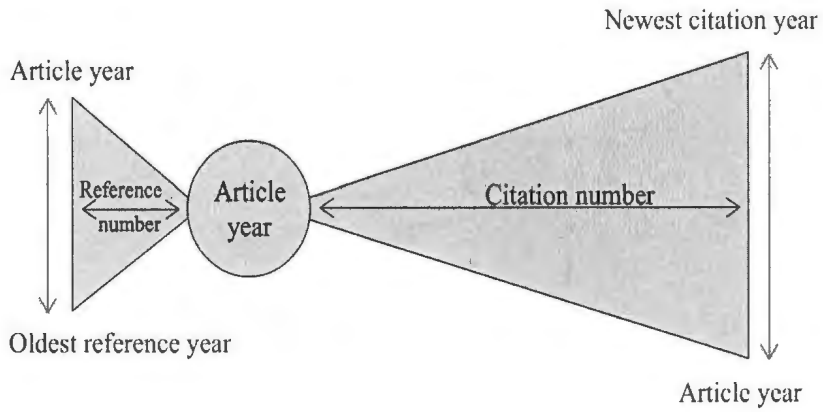


Figure 3.1: Metadata regarding document citation information is mapped to visual features of the bow tie representation.

represents forward citation information. This order of backward citations to the left and forward citations to the right follows the common representation of time flowing from left to the right [108].

The number of backward citations in a document is mapped to the width of the left side, and the right side's width indicates how often the article is cited in other works. Although direct decoding of the widths of the bow tie representation into the number of citations will not be possible, the relative differences can be perceived by ensuring that the center of the document nodes in a collection of bow tie representations are lined up vertically when they are added to the list.

The height of the left side shows the period from the publication year of the oldest backward citation to the year that the document itself was published. This mapping allows searchers to find those documents that provide an extensive coverage of the earlier studies that cover a long period. The right side's height represents the period between the document publication year and its most recent forward citation, letting

searchers distinguish newly cited documents from potentially obsolete ones.

These bow tie representations visually convey citation metadata to the searcher, and can readily be interpreted with little training. Table 3.2 contains different possible forms of bow tie representations as well as their corresponding interpretations. Locating these representations alongside the search results list may provide searchers with the ability to perceive and interpret this metadata, and then to make judgments on the suitability of each document to their given retrieval purpose.

Figure 3.2 provides an example of the bow tie representations within a set of search results. As can be seen, by comparing the widths of bow tie representations, the third item can be easily identified as a seminal work since it is cited many times.

Table 3.2: Based on citation information of a document, different forms of bow tie representations can be generated, with each bow tie conveying specific characteristics of the corresponding document.

Backward Citations		Forward Citations	
	Many backward citations that cover a long period.		Many forward citations that cover a long period.
	Many backward citations that cover a short period.		Many forward citations that cover a short period.
	A few backward citations that cover a long period.		A few forward citations that cover a long period.
	A few backward citations that cover a short period.		A few forward citations that cover a short period.

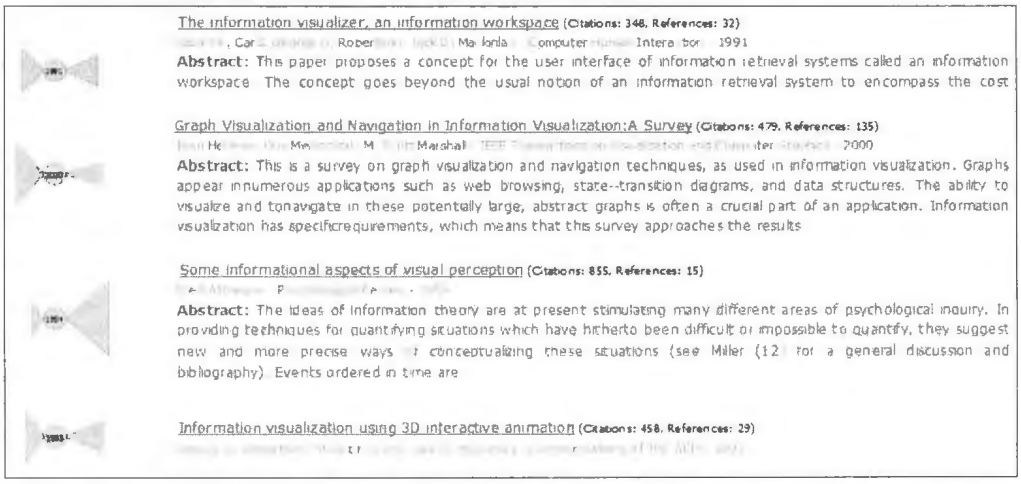


Figure 3.2: Augmenting the search results list with bow tie representations supports visual identification of potential documents by encoding features of the citation lists, along with the year of publication.

In addition, its height indicates that it is also recently cited and so is not obsolete. On the other hand, if someone is looking for a new document that provides a good coverage of the previous work, the second one is potentially a satisfactory one.

3.2.2.2 Document Selection for Detailed Evaluation

The search results interface augmented with bow tie representations assists searchers in comparing documents and finding potentially relevant ones based on their citation information. However, it encodes an abstraction of the backward and forward citations. Therefore, after the recognition of potentially interesting documents in the augmented list of search results, searchers are able to click on the corresponding bow tie representation to perform further detailed evaluation of that document.

When the searcher clicks on a bow tie representation of a document, a detailed bow

tie visualization of the document citation information is shown in a new page. The visual and interactive features of the detailed bow tie representations are explained in the following section.

3.2.3 Document Focus with Detailed Bow Ties

3.2.3.1 Detailed Visualization of Citation Metadata

The detailed representation of the selected document consists of the detailed bow tie representation of citation metadata and a description of the other metadata elements associated with the document in a textual format. Two columns containing the lists of backward and forward citations of the document are also provided (see Figure 3.3).

For designing a detailed visualization of documents, the premise is that the distribution of citations through different years is an important criterion for evaluating individual documents. For example, among two documents that have been cited an

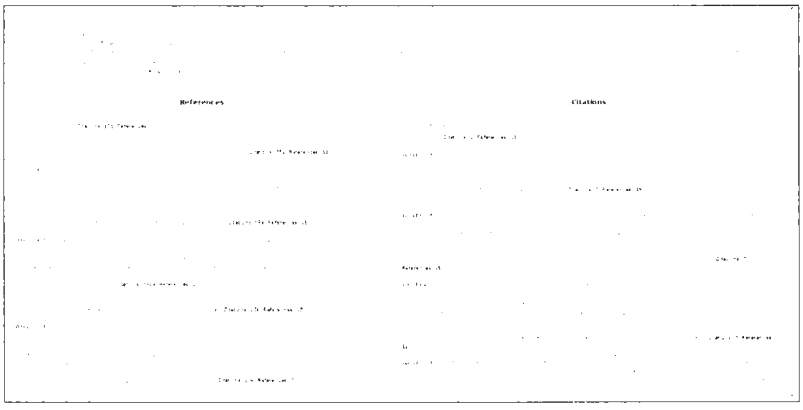


Figure 3.3: Detailed representation of the document containing the detailed bow tie visualization and citation lists, along with the filtering feature, facilitates further evaluation of individual documents and navigation among citations.

equal number of times, the one that has the greater number of recent forward citations may better represent a hot topic. Perceiving this valuable information in textual format is a difficult task that requires careful consideration and evaluation of each document in the citation list. Instead, a visual representation of this information is offered to the searchers, allowing them to evaluate each document in more detail.

In order to visually convey this information to the searcher, the number of backward and forward citations in each year is extracted from the metadata of the document and is counted. The height of the bow tie representation of the selected document, which represents the publication period, is subdivided into bars, each representing one year in that period. The number of citations published in each year is mapped to the length of these bars, providing a detailed bow tie representation (Figure 3.4). In addition, hovering the mouse over a year bar activates a tooltip, which shows the corresponding year and the number of citations published in that year. As the human visual system can assess relative length quickly and precisely [115], searchers are able to understand how citations of a document are distributed in different years.

In current search interfaces, if searchers want to view backward or forward citation list of an article, they need to click on a link which shows a new page containing just the requested list, while the original document is left behind. Providing document information in detail on top of the citation lists enables searchers to navigate through these lists without losing focus on the original document.

3.2.3.2 Interactive Citation Exploration

Two strategies widely used by library searchers are "backward chaining" and "forward chaining" in which searchers often start with one or more relevant research materials, and then search for other works through articles that were referenced by this document, or articles that cite this document [33]. To support this strategic behaviour associated with information seeking, the detailed bow tie representation is further augmented with interaction techniques. Searchers are provided with a filtering function to explore within the documents' citations. This feature provides an easy and effective way for searchers to perform backward and forward navigation among documents.

To activate this function, searchers can select different years from the detailed bow tie representation by single clicking on the bar representing that year. This

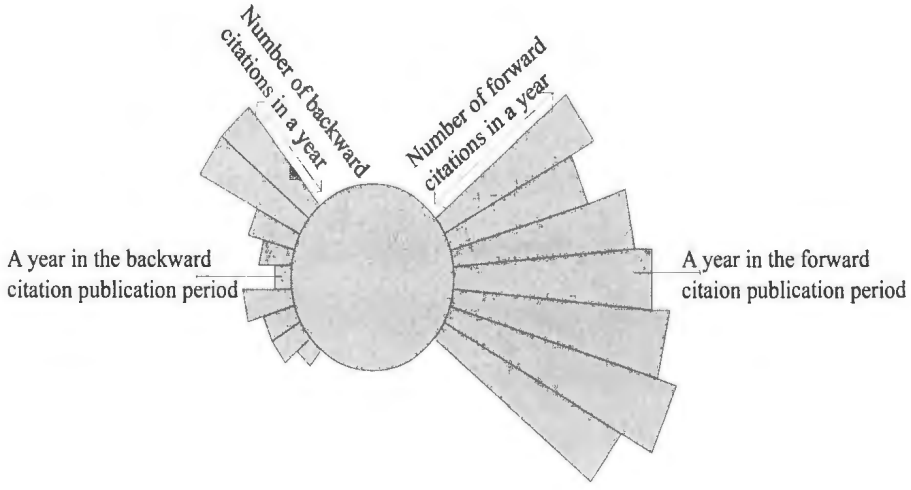


Figure 3.4: The detailed bow tie representation visually convey the number of backward and forward citations in different years.

Figure 3.5: The filtering operation allows searchers to navigate within the backward and forward citation lists.

operation causes the backward or forward citation list to be filtered to only show the corresponding data. When a filter is requested, the colour of the year bar is changed as a visual reminder that the filter is in effect (see Figure 3.5). Re-clicking on the already selected bar deactivates the filter.

Providing this interaction technique promotes searchers' involvement in the information retrieval process and facilitates navigation within citations of documents. It also allows searchers to narrow down the scope of their search to a manageable set of documents, enabling them to focus on documents of interest. This tool can be particularly helpful for the documents that have many backward or forward citations. Although these kinds of documents are often good starting points for finding other relevant ones, it might be beyond the searchers' tolerance to search within their citations without exploration support.

3.3 Example

In order to illustrate the value and benefits that the bow tie representations and the detailed bow tie visualization can provide to a searcher, an example is provided here. This example shows how a searcher can use these tools to evaluate and compare the search results in order to find a potentially relevant document and then explore within this document's backward and forward citation lists. These information retrieval steps are supported by the visualization of the citation metadata associated with each document.

Suppose a searcher starts with an initial query "content based image retrieval" with the goal of finding a set of interesting documents in the field. When the searcher submits the query, the system retrieves the documents, along with their metadata from MAS API, and then provides an augmented list of search results to the searcher. Figure 3.6 shows the interface after submitting the query.

By scanning the augmented list, the searcher is able to visually evaluate and compare documents based on their citation metadata as well as the information provided in the regular list-based representation. The searcher may quickly identify seminal documents that are being cited over a long period as well as important published works that have many backward citations. For instance, as can be seen in Figure 3.6, the searcher can easily realize that the first and the third documents of the page are relatively old documents that are cited many times. In addition, the fourth document's bow tie representation indicates that this document is providing an extensive coverage of the previous work with many backward citations covering a long period.

During the search process, the searcher may decide that a particular document is

worthy for further examination and may want to explore within its citations. Therefore, the searcher may click on the bow tie representation of a document to get the detailed bow tie representation and citation lists. Figure 3.7 shows the system after clicking on the bow tie representation of the second document retrieved by the initial query.

Figure 3.6: The augmented list of search results is generated by the searcher's query.

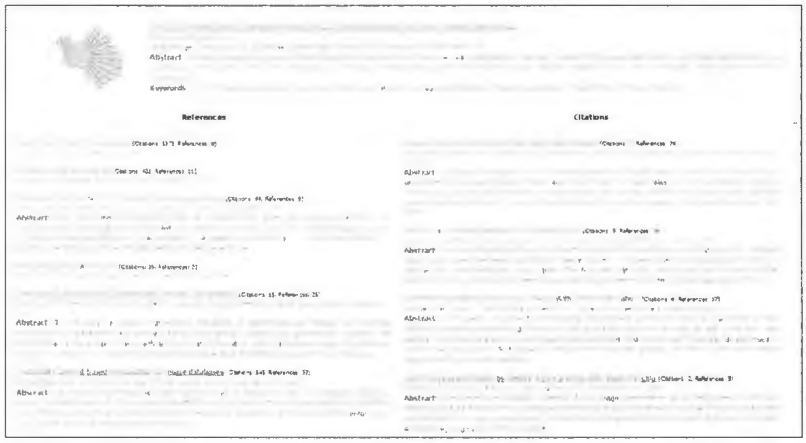


Figure 3.7: The searcher is able to evaluate an individual document and navigate within its citation lists.

Using the detailed bow tie representation, the searcher can observe how the backward citations of the document are distributed in different years and how it is being cited over the years. By examining this information as well as the detailed metadata elements provided in a textual format, the searcher may conclude that the clicked document is relevant and may select it as the first choice.

Moreover, by interpreting the information provided in the detailed bow tie representation, the searcher may conclude that the current document is also suitable for the citation navigation purpose. Therefore, the searcher may decide to explore those citations published in the similar timeframe to the publication year of the document of focus to find some other relevant document. As such, the searcher may filter the backward citations of the document to the most recent ones by selecting the lower bars in the right side of the detailed bow tie, and filter the forward citations to the oldest ones by selecting the upper bars of the left side of the detailed bow tie representation (Figure 3.8).

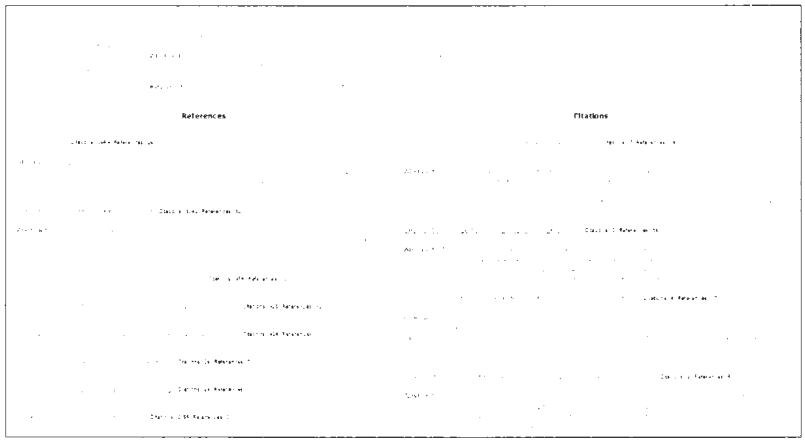


Figure 3.8: The searcher may filter the citation lists to the documents that are published in the similar timeframe to the document of interest.

Performing these information retrieval steps with the assistance of bow tie representations and the detailed bow tie visualization has guided the searcher in comparing and evaluating the search results, inspecting a particular search result in detail, and performing further exploration of the citation information. Performing similar activities with current search interfaces would require significant cognitive effort as searchers have to perform them manually due to the limited support provided.

3.4 Implementation Details

3.4.1 Platform and Web Technologies

Bow Tie Academic Search is a Web-based application developed using JavaServer Pages (JSP). The system is built on Java Development Kit 1.7, and it is developed using Eclipse as the integrated development environment and Tomcat as the Web application server.

The visualization and interaction features of Bow Tie Academic Search are developed by means of JavaScript and CSS. In particular, the methods provided by D3 [14] and jQuery libraries are used to dynamically access and alter Document Object Model (DOM) elements. These visual representations are rendered using Scalable Vector Graphics (SVG) as it creates resolution independent graphical representations.

3.4.2 System Architecture

The system architecture of Bow Tie Academic Search consists of five primary components. The *meta-search component* retrieves documents, along with their metadata

elements, from MAS API. The *backward citation analysis component* performs calculation and analysis on the backward citations retrieved by the meta-search component. The *forward citation analysis component* analyzes the retrieved sets of forward citations. The *data integration component* combines the information provided by these three components. Finally, the *data visualization component* generates visual representations of these metadata elements. This system architecture is illustrated in Figure 3.9.

Once a query is submitted by the searcher, the meta-search component retrieves the first 10 documents (documents of the first page) as well as their metadata elements from MAS API. For each of these documents, this component also sends two other calls containing document's ID to the API in order to retrieve its backward and forward citation lists. Then the meta-search component passes backward citation sets and documents' year information to the backward citation analysis component. It also sends forward citation sets as well as publication years to the forward citation analysis component. In addition, it directly passes documents' titles, authors, years, URLs, abstracts, and citation counts to the data integration component.

In order to find the oldest backward citation of each document, the backward citation analysis component searches into the corresponding list of backward citations. Then it calculates the period between the document publication year and its oldest backward citation. When the backward citation sets of all documents are processed, this component passes a list of the calculated citation metadata to the data integration component. Meanwhile, the forward citation analysis component finds the most recent forward citation for each document, calculates the period between the publication year of this citation and the document publication year, and then passes a list of the

calculated metadata to the data integration component. When year metadata is not available in the underlying collection, MAS provides zero as publication years. It is the responsibility of these two components to handle such anomalies in data.

The data integration component combines the metadata elements directly received from the meta-search component with the calculated metadata elements received from the citation analysis components, providing a complete record for each docu-

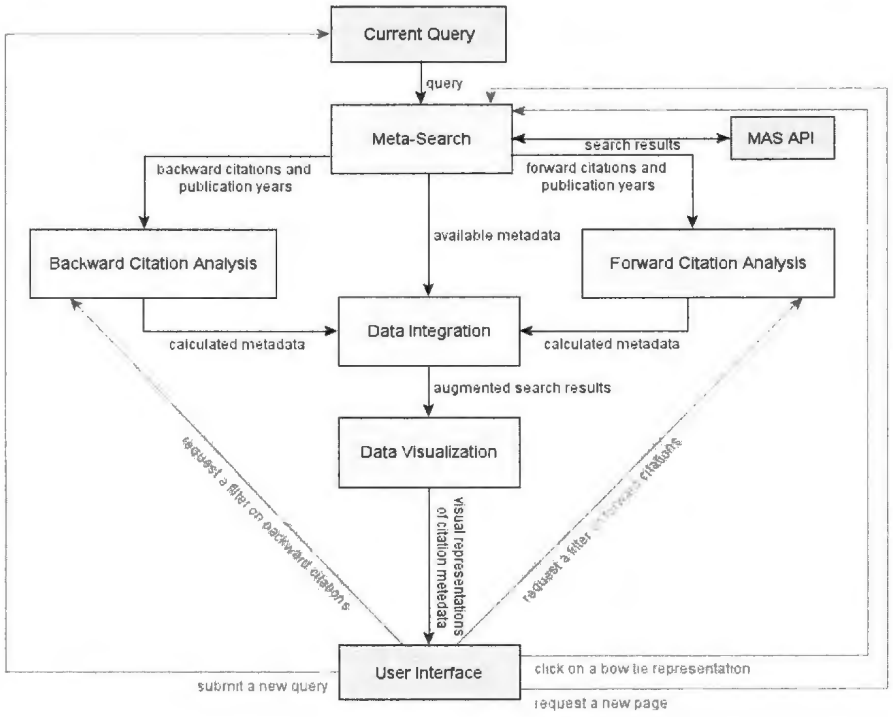


Figure 3.9: The system architecture of Bow Tie Academic Search includes five primary components: meta-search component, backward citation analysis component, forward citation analysis component, data integration component, and data visualization component. Data exchanges among these components are shown using black arrows, while red arrows show interactions with the system.

ment. Then these records are passed to the data visualization component, which uses the aforementioned Web technologies to generate visual representations of citation information and to create the augmented list of search results.

Using the augmented list, searchers may request to view another page of search results. Once a page button is clicked, the current query, along with the requested page number, is sent to the meta-search component. The meta-search component then uses this information to retrieve the new set of documents from MAS API.

In addition, when a bow tie representation is clicked by the searcher, using the ID of the corresponding document, the meta-search component retrieves the document's metadata elements in more detail as well as its backward and forward citation lists. At this point, in addition to the publication period, the citation analysis components count the number of citations in different years. Then the data integration and data visualization components perform the same processes explained above to provide the detailed representation of the document. Moreover, when a filter is requested, the corresponding citation analysis component receives the selected year and filters the citation set.

3.5 Discussion

In this Chapter, an approach for incorporating search activities of library searchers to digital library search interfaces was described. In particular, this approach supports two fundamental activities of citation-based evaluation and comparison of documents as well as citation navigation and manipulation by providing visual representations of citation information.

As noted in Section 2.2, one of the main challenges in creating visualizations of citation metadata is to overcome the scalability and complexity issues associated with the current approaches. Bow tie representations compactly and simultaneously convey backward and forward citation metadata as well as document year information to the searcher and can represent documents with a large number of citations.

The design of these bow tie representations was guided by the information visualization principles and theories. Mackinlay [80] ranking was used to map quantitative information to visual representations. According to his ranking, position is the best visual form to represent quantitative data. However, using position to simultaneously encode two attributes, results in a 2D spatialization of documents, which requires a relatively large space. In addition, the spatialization of documents cannot be easily interpreted by the searcher (as discussed in Section 2.3.2). As such, length, as the second visual form in the ranking, has been chosen to represent quantitate data including citation counts and publication periods. According to the pre-attentive processing theory, length can also be pre-attentively perceived.

Due to space limitations in the bow tie representation, the exact values of citation counts and publication periods could not be precisely mapped to length. Instead, binning is used to encode these two values. However, since the citation counts may cover a wider range of values compared to the publication periods, they are grouped into larger bins. This difference in binning scales is not expected to be problematic as they encode different pieces of information. As such, even though searchers may not be able to quantify exact values, they are still able to interpret the relative differences of these two features for different documents, which may lead them to an efficient and effective preliminary analysis of documents. Even so, there may be issues with the

scalability of the approach in extreme situations.

Since bow tie representations simultaneously convey a collection of metadata elements, the use of other visual features such as colour to encode further information can be overwhelming for the searcher. Therefore, colour is not used as a visual encoding mechanism, but is used to only label bow tie representations. Using strong and bright colours for visual representations has been shown to produce unpleasant effects [109]. Instead, these bow ties representations are labeled with a relatively soft colour, while having enough contrast to be easily distinguished from the background.

Given the advantages of the list augmentation approaches to visualize the search results (as discussed in Section 2.3.1), these bow tie representations are added to the list-based format. The integration of these two forms of representations allows searchers to visually compare and evaluate documents based on their citation information while they can still benefit from the list advantages of clarity and consistency in further evaluation of individual documents. Furthermore, searchers may not feel comfortable using search interfaces that are drastically different from what they normally use [43]. This approach avoids this problem by keeping the list as the central component of the interface.

The initial visualization of citation information is an abstract representation of this metadata element. In order to facilitate detailed evaluation of documents, the abstract/elaborate interaction method is provided, letting searchers request and view a more detailed representation of individual documents. This interaction method can be employed via the simple mechanism of single clicking on the bow tie representation of a potentially interesting document.

The detailed bow tie representations provide searchers with the opportunity to

evaluate backward and forward citation sets and to use this information to evaluate the document of focus. The structural similarity between the detailed bow tie visualization and the bow tie representations added to the list is meant to enhance the learnability of the system. Moreover, the detailed bow tie representation uses a space filling approach by which the available space is optimally used. Considering the need for compact representations in search interfaces (see Section 2.3), such a representation can be a better choice in comparison to the alternatives such as a bar chart.

In addition to revealing citation patterns over the time, the detailed bow tie representation is augmented with the explore interaction method, supporting flexible navigation within citations. This exploration support is provided by the metadata-based filter method that can be iteratively applied by the searcher. Metadata-based filtering has been shown to be an effective way to support exploratory searches [118]. The year metadata has been chosen for classification and filtering purposes as it can be an important criteria in judging the suitability of documents. Moreover, unlike some other metadata elements, the publication year is consistently available in all of the metadata standards.

The low-level interaction method used to activate and deactivate filters has been chosen to be the intuitive and easy-to-perform point-and-click technique. However, in space-filling radial layouts, as the number of bars grows, their size decreases. This limitation may make it hard for searchers to click on year bars and to apply filters when the citations are distributed over many different years.

In addition, when a filter is requested, the colour of the selected year bar is changed to let searchers easily perceive which filters are in effect. As the colour of the detailed

bow tie is blue, yellow has been chosen to label the selected year. Following the opponent process theory of colour as well as the pre-attentive processing principle, this colour difference between the non-selected and selected years can be pre-attentively processed [115], facilitating easy and fast identification of the selected filters.

Following these design theories and principles, many other visual and interactive representations can be designed to represent citation metadata. In order to validate the specific design choices made in Bow Tie Academic Search, a user study evaluation was conducted, which will be explained in detail in Chapter 5.

Chapter 4

Interactive Query Refinement

4.1 Motivation

One of the most important tasks in traditional libraries is the reference interview, the goal of which is to clarify what searchers really need via query negotiation [30]. The establishment of the reference interview in the early stages of librarianship clearly indicates that the need to support searchers in the query refinement process has been identified for years. In digital libraries, however, in order to formulate a better query, searchers are required to solely rely on the knowledge they might gain from the documents retrieved by the initial query as well as the features of the search interface. In current library search interfaces, little support is provided for searchers to construct better queries as they mostly provide a text box for manual query specification. This lack of support may lead to the use of underspecified and inaccurate queries, reducing the effectiveness of information access and retrieval.

Suggesting candidate terms may support searchers in this endeavor, resulting in

reformulation of more effective queries that better represent searchers' information needs. Term suggestion approaches have been studied frequently [19], with the focus being on the methods to generate and rank the candidate terms. Even though an important aspect of these approaches is the incorporation of human decision-making and control in the query reformulation process, less effort has been made to explore different methods to better present these terms to the searchers. Given that searchers are not necessarily able to choose effective terms when they are simply presented in a list [82, 96], sophisticated tools are required to enhance searchers' abilities to make better decisions when reformulating their queries [53, 57, 59, 60].

With the aim of promoting human decision-making in the query reformulation process, Bow Tie Academic Search provides a visual and interactive tool, allowing searchers to evaluate and compare the suggested terms and to refine their initial queries. This tool, which also uses MAS API data, consists of a histogram of the most frequently used keywords in the top search results and a compact visual encoding that represents document similarities based on the co-use of keywords.

The use of a histogram to support query refinement process was inspired by WordBars [53, 57]; however, here, the histogram is enhanced with visual encodings of additional information. Using this enhanced histogram, searchers are able to explore the search results and refine their queries by adding, removing, or replacing the terms with the initial query. Once the searcher submits the refined query, an updated histogram will be generated, facilitating iterative query reformulation, which is a common activity in exploratory searches [118].

This Chapter provides a detailed explanation of the visual and interactive features of the enhanced histogram. An example is provided to illustrate how it can

support searchers in exploring the search results set and in formulating better queries. Moreover, an updated version of the system architecture is described. Finally, the advantages, issues, and limitations of this approach are discussed.

4.2 Approach

4.2.1 Visual Representation of Keyword Metadata

In academic libraries, some terms are often assigned to documents in order to represent the core concepts of the documents. Representing such keyword information associated with retrieved documents can support identification of important aspects of the search results set. As such, a compact visualization of keyword metadata is offered to the searchers. This visual representation allows searchers to recognize the most frequent keywords used in the top search results, to perceive their relative frequencies, and to understand keyword-document relationships.

As the documents are retrieved from the MAS API, the system counts the keywords associated with the top 200 documents. The keyword frequencies are visually represented using a vertically oriented histogram located at the right side of the search results list, where the size of each bar represents the frequency of the associated keyword (Figure 4.1). Due to the space limitations, only the first 25 keywords are displayed in the histogram. This method is similar to the approach provided in WordBars [53, 57] in the Web search context.

By browsing this information, searchers are able to understand the general topics of the retrieved documents. This information along with the keyword frequencies

can also be used to evaluate whether the retrieved set is capable of satisfying their information needs. For exploratory search activities, the histogram can be used to identify other potentially interesting topics relevant to the given query for the subsequent searches.

To represent interrelations of the documents, the histogram is enhanced with a visual depiction of which documents in the current page of the search results set use each keyword in the histogram. A relation grid [6] is provided, where each row represents a keyword in the histogram and each column represents a document currently being shown to the searcher. Since there are ten documents shown per page of search results, and 25 keywords extracted from the top search results, this grid is of size 10×25 . When a specific keyword is associated with a document of the page, the corresponding cell in the relation grid is highlighted with a colour and tagged with a number, which is the document rank in the current page.

This relation grid can assist searchers in their retrieval process in different ways. For example, searchers are able to perceive which documents share the same keyword by scanning across the grid for a particular keyword. Scanning down each column reveals the keywords used by each of the search results on the current page. Furthermore, searchers are able to see which documents are using the keywords that are interesting to them. Alternatively, after identification of an interesting keyword, the searcher can scan across to find a document, scan down to find other keywords used by this document, and then scan across again to find other documents that use these keywords. Moreover, it can be used to check if a specific document is among the most frequently used ones. Since the most common keywords may be the most relevant ones to the searchers' information needs, providing this representation may

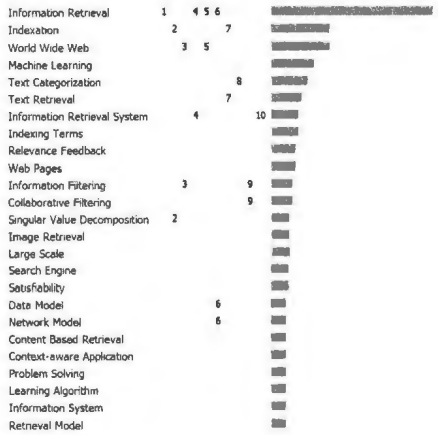


Figure 4.1: The enhanced histogram of the most frequent keywords allows searchers to identify the most frequent keywords in the search results, their relative frequencies, and document interrelations in terms of their keywords.

allow searchers to find potentially relevant documents from the search results list.

4.2.1.1 Interactive Search Results Exploration

After identifying the potentially relevant documents by using the relation grid, searchers are able to select the corresponding cell to perform further evaluation of the document. This selection causes the search results list to be scrolled to target the corresponding document. The document is also highlighted by a yellow colour as it can be pre-attentively distinguished form the background [115]. The use of such a colour allows searchers to readily identify the selected document in the list of search results.

The coordination between the relation grid and the list of search results supports searchers in an interactive exploration of search results, allowing them to both gain an insight into the features of the documents currently being shown as well as investigate the documents in more detail. Moreover, the interactivity of this approach allows

the searcher to jump from one document to another in order to further examine its relevance.

4.2.1.2 Interactive Query Refinement

Further interaction techniques are provided to support interactive query refinement by allowing searchers to iteratively convert their initial queries to a well-defined one. When searchers start with a vague and underspecified query, it may be difficult for them to determine alternative ways to improve the query. Here, the histogram assists searchers in identifying useful alternative terms.

Providing candidate keywords to the searchers supports recognition of relevant terms instead of requiring searchers to remember them [53, 57]. Providing the keyword frequencies offers additional support in the term selection process. In addition, using the relation grid allows searchers to quickly jump to the documents that are using a particular keyword, and then to evaluate its relevance to their information need. Assuming that searchers start with an initial query that is at least somewhat relevant to their information needs, the enhanced histogram may allow them to find more accurate descriptions of what they are seeking.

Using the checkbox provided at the left side of each keyword, the searcher is able to add or remove terms from the current query. The keywords that consist of multiple terms are added in quotes, allowing searchers to focus on a more specific set of search results. In addition, the searcher is able to change the focus of the search by single clicking on any keyword, which replaces the current query with the selected keyword. After refining the query, searchers can click on the search button to indicate the end of the query refinement process and to retrieve a new set of search results.

4.3 Example

In this section, an example is provided to illustrate the potential benefits of the enhanced histogram in the exploratory search processes. This example shows how searchers can benefit from the relation grid and the histogram in exploring the search results set and in refining their queries. These search activities are supported by visual representations of keyword metadata associated with the top ranked documents.

Consider the same search scenario explained in Section 3.3 in which the initial query was "content based image retrieval". When the searcher submits the query, in addition to the augmented list of search results, the histogram of the most frequent keywords used in the top retrieved documents, along with the relation grid, is generated and presented to the searcher (Figure 4.2).

Besides using bow tie representations to discover relevant documents, the searcher is able to use the enhanced histogram in multiple ways to discover potentially relevant

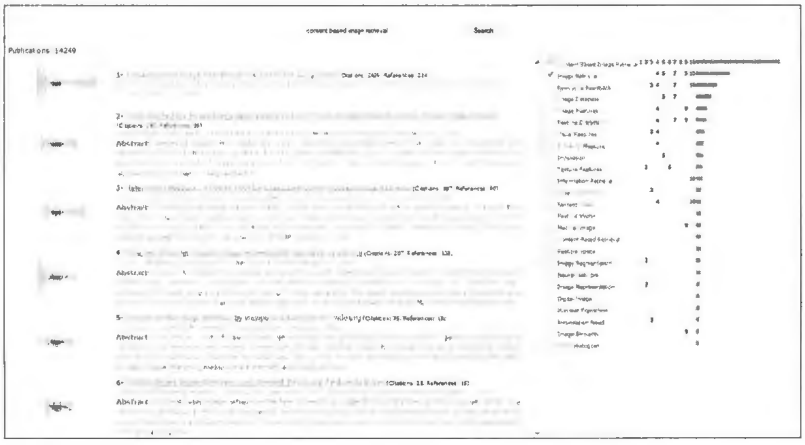


Figure 4.2: The augmented list of search results and the enhanced histogram are generated by the searcher's initial query.

documents based on their keyword metadata. The searcher may decide to first look down the document columns in the relation grid to find the documents that have most of their keywords among the most frequent ones. By employing this method, the searcher can find potentially relevant documents that may not be identified by considering their citation information.

Then the searcher may look for the terms of interest in the histogram and may want to inspect the documents using these terms. For example, the searcher may recognize "relevance feedback" as a relevant term and may click on the seventh document cell to further evaluate it (Figure 4.3). At this point, the searcher may find this document relevant and may wish to check for the other documents using the same keyword as this document.

After reviewing the first set of results, the searcher may wish to refine the query and explore the underlying collection. Considering keyword frequencies as well as the relevancy of the documents using "relevance feedback" as their keyword, the searcher

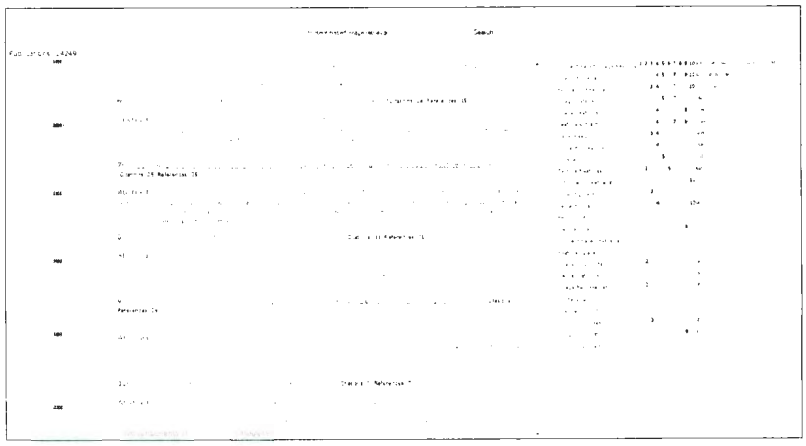


Figure 4.3: The searcher can use the enhanced histogram to explore the search results and to inspect an individual document in more detail.

may decide to choose this keyword to refine the initial query. The searcher may want to focus on a more specific set of search results by checking this keyword, which adds it to the initial query. Alternatively, the searcher may wish to change the focus of the search from image search to relevance feedback studies by clicking on this term, which replaces it with the initial query. After refining the query, the searcher may decide to click the search button to submit the new query and review the new set of search results. Figure 4.4 shows the interface after submitting the expanded query.

4.4 System Architecture

In order to incorporate the enhanced histogram into Bow Tie Academic Search, the *keyword analysis component* is added to the system architecture explained in Section 3.4.2. The updated system architecture is illustrated in Figure 4.5.

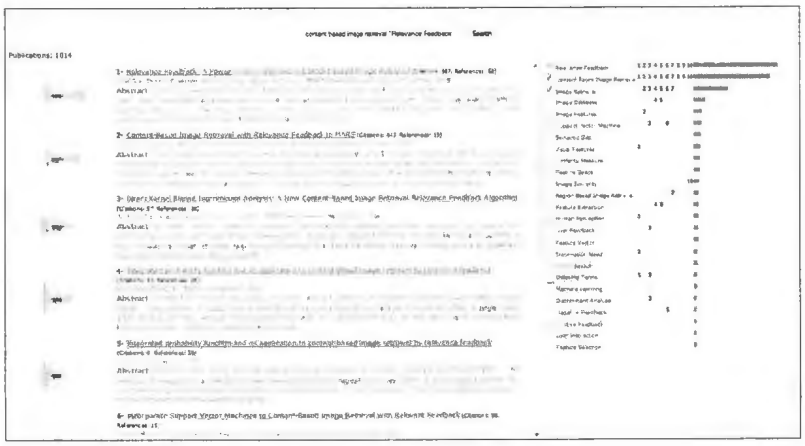


Figure 4.4: After the refinement of the query by the searcher, the visual representations of keyword and citation metadata are updated based on the characteristics of the new set of search results.

After the retrieval of search results by the meta-search component, in addition to the operations described in Section 3.4.2, it generates two lists: the top keywords list that contains the keywords associated with the top 200 documents, and the keyword-document list that contains the documents of the current page as well as their corresponding keywords. These two lists are then passed to the keyword analysis component.

The keyword analysis component first analyzes the top keywords list, counts the occurrences of each keyword, and sorts the list based on keyword frequencies in as-

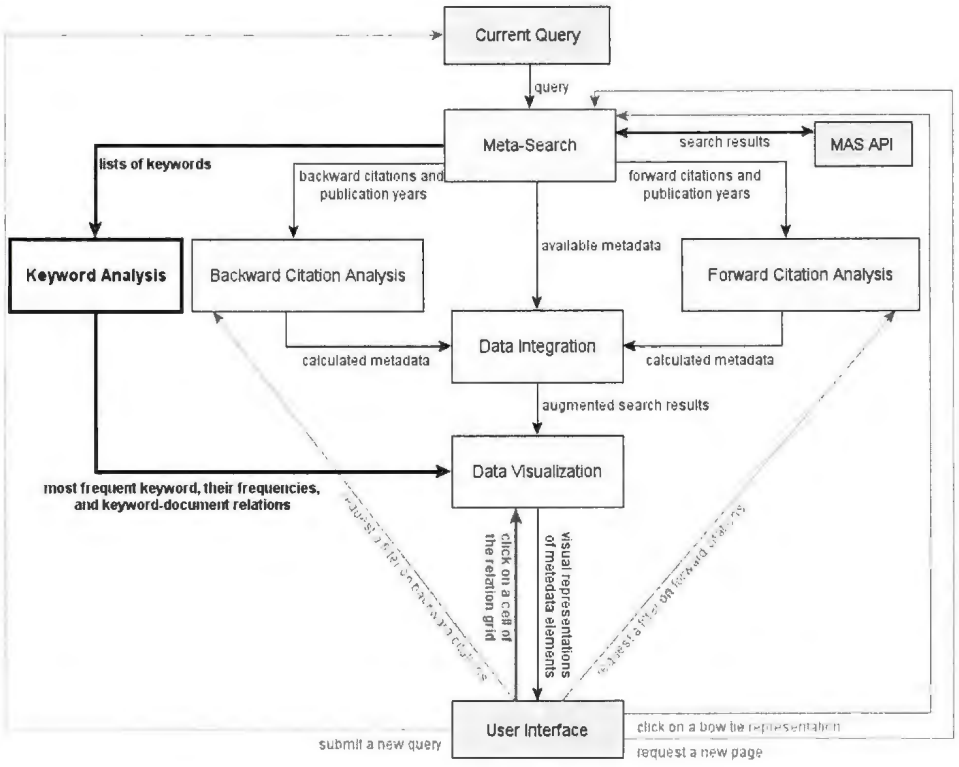


Figure 4.5: The keyword analysis component is incorporated in the Bow Tie Academic Search architecture.

cending order. It then updates the list to only include the first 25 keywords and their corresponding frequencies. This list is then compared with the document-keyword list to find whether a frequently used keyword is associated with one or multiple documents of the current page. After adding this information to the top keywords list, this updated list is sent to the data visualization component by which the enhanced histogram is generated. In addition, when a cell is clicked in the relation grid, the data visualization component uses this list to find the corresponding document, scrolls the search results list to target this document, and highlight its textual representation in the augmented list of search results.

4.5 Discussion

Considering the capabilities of interactive query refinement methods to increase retrieval effectiveness (discussed in Section 2.4.2), the enhanced histogram has been designed and developed to support searchers in their query refinement process. Even though the primary purpose of designing this tool was to provide query support, it also supports search results visualization and visual exploration via visualization of auxiliary information.

As discussed in Section 2.3.5, these approaches provide a summary of search results and allow searchers to perceive the general characteristics and attributes of the retrieved set of documents. The enhanced histogram compactly depicts what keywords are commonly used in the top ranked document, their frequencies, and which documents of the page are using these keywords.

Similar to the design of citation visualizations, the enhanced histogram has been

designed based on information visualization theories and guidelines. The mapping of keyword frequencies, as quantitative information, to bar lengths, allows searchers to accurately and pre-attentively understand and perceive relative keyword frequencies in the top and potentially prominent retrieved documents [115].

The use of the relation grid instead of a node-link diagram provides searchers with a compact visual representation of keyword-document relations. In addition, the relation grid has been shown to outperform node-link diagrams for the purpose of understanding relations between nodes [40]. The incorporation of the relation grid into the histogram provides a unified representation of keyword information, allowing searchers to perceive and understand both keyword frequencies and keyword-document relationships.

Since the histogram and the relation grid represent two different pieces of information, they are labeled with different colours. Guided by the opponent process theory of colour [115], green is used to highlight cells as the human visual system can easily separate it from the blue colour used to show the histogram bars. Moreover, the highlighted cells are tagged with a number, which is the document rank in that page, to facilitate quick identification of the corresponding documents.

Search results exploration support is provided by the coordination between the augmented list of search results and the relation grid, allowing searchers to focus on particular elements of their information needs. This coordination is provided by the connect interaction method [122], enabling searchers to view a detailed representation of a potentially interesting document identified in the enhanced histogram. Even though this exploration support may not be useful for very poor queries, using the information provided in the enhanced histogram, searchers might be able to evaluate

whether the retrieved set contains what is being sought.

One of the well-known usability principles is *recognition rather than recall*, which aims at minimizing users' memory load by facilitating recognition of information rather than requiring users to recall it [89]. Providing a list of frequently used keywords, promotes this principle and supports searchers in recognition of potentially relevant topics without the need to remember the terms and topics that are relevant to their information needs. Moreover, the intuitive interaction mechanisms added to these terms let searchers "see-and-point" instead of "remember-and-type" when refining their queries [89].

Even though providing a list of potentially relevant keywords allows recognition of relevant topics, the overall effectiveness of the query refinement process is impacted by the searchers' ability to choose appropriate terms. The extra visual information of keyword frequencies and keyword-document relationships, along with the knowledge that may be gained through the exploration activities, can help searchers make informed decisions when choosing terms for the query modification purpose.

However, as these terms are extracted using a local analysis approach (see Section 2.4.2), their quality and usefulness depends on the quality of the initial query and the retrieved documents. Therefore, the query refinement process may not be effective when a very poor initial query is provided. However, the lack of relevant terms in the histogram may inform searchers that the initial query cannot reflect what they are looking for and needs to be manually refined.

Chapter 5

User Study

5.1 Purpose

User studies are now an integral part of any user-oriented research field such as human computer interaction or information visualization [18]. User studies let researchers verify their design choices, confirm or reject hypotheses, and make comparisons between different systems [50]. In the last few decades, an increasing number of interactive information retrieval systems have been developed that enhance the human element of search systems by promoting searchers' control in the search process. Therefore, the incorporation of searchers into the evaluation of search systems has become an important part of the interactive information retrieval research [70, 15].

Designing and conducting user-centered evaluations of information retrieval systems is a difficult task due to the complex cognitive steps that normally take place during the retrieval process. Especially for exploratory search systems that promote high levels of interaction, searchers may perform a series of unobservable cognitive

activities that may affect their subsequent interactions and behaviour [70]. The inclusion of all of these parameters within the design of a user study may not be feasible. Therefore, in most studies, researchers make simplifying assumptions about searchers, their information needs, and the complex concept of relevance [70].

Despite all these challenges and complexities, it is necessary to understand the behaviours, experiences, and preferences of searchers engaged in search processes. It is also necessary to explore the retrieval effectiveness as well as subjective impressions of the specific features of information retrieval systems. As such, to obtain empirical evidence regarding the potential benefits of Bow Tie Academic Search in comparison to a baseline system, a user study was conducted in a controlled laboratory setting.

The Web-based interface to MAS provides document snippets when presenting search results in the list, while MAS API provides document abstracts. Due to this difference in document representation, the default MAS interface could not be used as a baseline system. Instead, a list-based interface was developed using MAS API data and functions. This interface consists of a query box, a list of document surrogates (including their title, number of forward citations, authors, and an abstract), a link to the actual document, and a link to a list of forward citations. These features resemble a typical library search interface, making this interface a reliable baseline system. Figure 5.1 shows a screenshot of this system after submitting a query.

In order to compare and evaluate documents based on their backward citation information or to navigate within backward citations, searchers are required to click on the documents' URL, view documents, and then use the backward citation list provided in the actual document. To compare and evaluate documents based on their forward citation counts, searchers can use the search results list to read forward

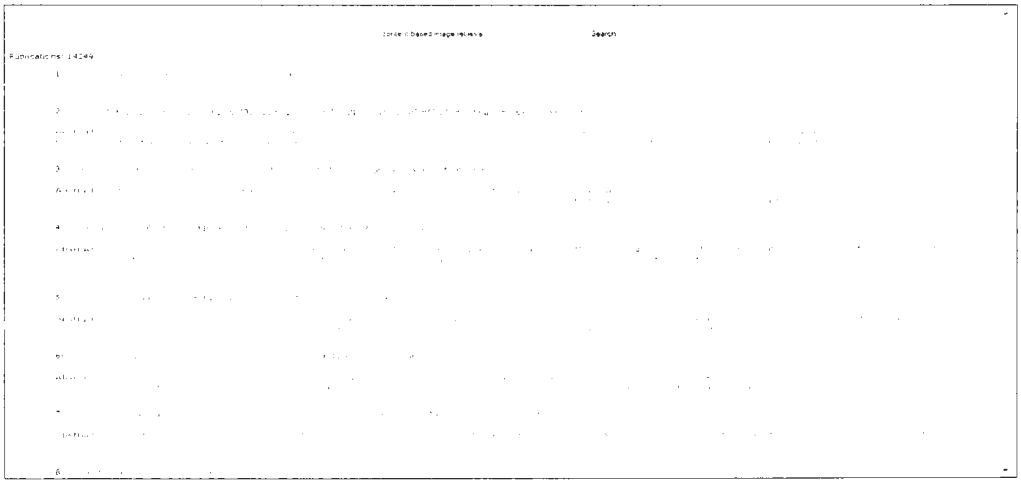


Figure 5.1: Following the general features of a typical library search interface, the list-based interface is developed as a baseline system.

citation counts provided beside document titles in a textual format. However, in order to understand other features of forward citation sets or to explore within forward citations, they are required to click on the hyperlinks associated with forward citation counts, which loads a new page containing a list-based representation of forward citations. Query refinement can only be performed manually by the searcher.

5.2 Hypotheses

Based on the observations and knowledge about the two library search interfaces under study here (i.e., the list-based interface and Bow Tie Academic Search), two sets of hypotheses were formulated. The first set predicts the impacts of Bow Tie Academic Search on efficiency, effectiveness, and subjective measures on exploratory search processes (named as H1), while the second set predicts participants' subjective opinions about each of the components of the system (named as H2).

The first set of hypotheses are as follows:

H1.1: *Participants will find relevant documents faster with the list-based interface than with Bow Tie Academic Search.*

The list-based interface uses a static and simple list, and provides little interaction and exploration support compared to Bow Tie Academic Search. As such, we expect that the participants will immediately start working on the task rather than interacting with the interface and exploring the search results. On the other hand, as Bow Tie Academic Search provides analysis and exploration support, it is expected that participants will spend more time performing their information retrieval tasks.

H1.2: *Participants will be able to find documents that are more relevant using Bow Tie Academic Search than using the list-based interface.*

As Bow Tie Academic Search provides participants with additional information about documents, participants will be able to judge a result's relevancy more effectively by considering multiple features of the documents (e.g., citation information, keyword frequencies, and keyword-document relations). Furthermore, using visual features to represent metadata supports easier comparison and evaluation of documents, resulting in a better assessment of search results before completing the search tasks, and the filtering function may enable participants to find the relevant citations buried deep in the citation lists that otherwise might remain unnoticed. As such, it is expected that participants will complete their search tasks with a better set of selected documents using Bow Tie Academic Search than using the list-based interface.

H1.3: *Participants will be able to reformulate their queries more effectively using Bow Tie Academic Search than using the list-based interface.*

To reformulate queries, the list-based interface provides a simple query box that requires participants to articulate their information needs based solely on what they have been able to learn while doing the search. Bow Tie Academic Search, however, provides query reformulation support via the enhanced histogram, allowing participants to recognize potentially relevant keywords and inspect documents that use these keywords when deciding whether to choose them to refine their queries. Consequently, it is expected that Bow Tie Academic Search may positively affect participants' ability to refine their queries in comparison to the list-based interface.

H1.4: *Participants will report a higher level of confidence in finding relevant documents using Bow Tie Academic Search than using the list-based interface.*

We expect that visualizations of metadata elements, along with exploration and navigation support provided by Bow Tie Academic Search, will allow searchers to better evaluate and compare search results during their search process. Thus, it is expected that participants' feelings of confidence in finding relevant documents will be more positive for Bow Tie Academic Search than for the list-based interface.

H1.5: *Participants will report that the list-based interface is easier to use than Bow Tie Academic Search.*

The list-based interface provides a simple representation of search results and supports a few basic interaction methods. Therefore, the expectation is that participants will report more positive impressions of the ease of use of the list-based interface in comparison to Bow Tie Academic Search, which is a relatively complex interface.

H1.6: *Participants will report that Bow Tie Academic Search is more useful for finding academic documents than the list-based interface.*

It is expected that participants will be able to find documents that are more rele-

vant using Bow Tie Academic Search than using the list-based interface (hypothesis H1.2). In addition, the expectation is that participants will be able to construct better queries using Bow Tie Academic Search than using the list-based interface (hypothesis H1.3). Therefore, it is expected that they will be able to conduct their information retrieval tasks more effectively with Bow Tie Academic Search, and they will provide more positive feedback regarding the usefulness of this system in comparison to the list-based interface.

H1.7: *Participants will prefer Bow Tie Academic Search over the list-based interface.*

Even though it is expected that participants will be faster using the list-based interface and will find it easier to use, their enhanced ability in finding a good set of documents and in crafting better queries, along with their positive impressions of the usefulness of the system will outweigh these drawbacks. Therefore, it is expected that participants will prefer Bow Tie Academic Search over the list-based interface.

The following are the second set of hypotheses formulated to predict the perceived ease of use and perceived usefulness of the specific features of Bow Tie Academic Search:

H2.1: *Participants will report that bow tie representations are easy to use.*

We expect that the various forms of bow tie representations can readily be interpreted by searchers with little training. In addition, the simple point-and-click interaction method supported by bow tie representations does not require much effort to be learned and used. Therefore, it is expected that participants will have a positive perception of the ease of use of bow tie representations.

H2.2: *Participants will report that bow tie representations are useful for finding*

academic documents.

Bow tie representations convey backward and forward citation metadata as well as document year information to the searcher. Since citation metadata provides valuable information for scholarly searches, we expect that participants will provide positive feedback regarding the usefulness of this component of the system.

H2.3: *Participants will report that the detailed bow tie representation is easy to use.*

It is expected that once participants become familiar with the detailed bow tie visualization and interaction methods, they can interpret the citation information of individual documents and explore within the citations with little effort. Therefore, we expect that participants will find the detailed bow tie representations easy to use.

H2.4: *Participants will report that the detailed bow tie representation is useful for finding academic documents.*

The detailed bow tie representation allows detailed evaluation of an individual document, and it simplifies two common library search strategies of backward and forward chaining. As such, it is expected that participants will have a positive perception of the usefulness of detailed bow tie representations.

H2.5: *Participants will report that the enhanced histogram is easy to use.*

The enhanced histogram uses relatively simple visualization and interaction methods to encode keyword metadata and to support query refinement activity. Moreover, it is coordinated with the list of search results. Therefore, we expect that participants will provide positive feedback regarding its ease of use.

H2.6: *Participants will report that the enhanced histogram is useful for finding academic documents.*

The enhanced histogram allows searchers to recognize the most frequent keywords used in the top search results, to perceive their relative frequencies, to understand keyword-document relationships, and to explore within the search results. In addition, it supports the common search activity of query refinement. Because of all the information-seeking supports it provides to searchers, we expect the participants' perception of its usefulness to be positive.

5.3 Methodology

The user study was designed as a within-subject study to facilitate direct comparisons of the two interfaces (i.e., the list-based interface and Bow Tie Academic Search). Even though search tasks are not treated as independent variables, they function as variables due to their learning effects. Therefore, each participant was asked to perform a different exploratory search task with each interface.

5.3.1 Tasks

In order to successfully evaluate exploratory search interfaces, it is important to specify well-grounded and realistic exploratory search tasks. Exploratory search tasks should indicate ambiguity in information need, suggest knowledge acquisition, comparison, and discovery, and provide a low level of specificity about the information and how to find it [73, 74]. They also need to provide enough imaginative context for participants, allowing them to relate to the situation and apply their knowledge to it [73]. To evaluate Bow Tie Academic Search, as an exploratory search interface, the tasks were designed to have these desirable characteristics of exploratory search

tasks.

In order to ensure that participants have a basic knowledge and experience in scholarly search, a purposeful sampling approach [93] was followed by recruiting participants from the computer science and computer engineering graduate student body. The tasks were designed to ask these graduate students find multiple documents for writing research papers [73, 74]. The following are the two composed exploratory search tasks, with each task consisting of descriptions of the information needs and the corresponding initial queries:

Task 1: Imagine you are taking a class called "Information Visualization." For this class you need to write a research paper about visualization and interaction techniques to represent graph layouts. Use the assigned academic library search interface to find highly relevant research articles as the starting point for your research paper.

initial query: "graph visualization"

Task 2: Imagine you are taking a class called "Information Organization and Retrieval." For this class you need to write a research paper about methods to retrieve multimedia information. Use the assigned academic library search interface to find highly relevant research articles as the starting point for your research paper.

initial query: "multimedia retrieval"

These search tasks are intended to let us understand whether the new features of Bow Tie Academic Search are indeed an improvement over the current search interfaces. As such, the tasks are followed by four sub-tasks specified to ask participants use the metadata elements available within the search interfaces (or to manually find the metadata from within the document if is not provided in the interface):

1. Find three highly relevant documents in the first set of search results retrieved by the given query.

2. Using document #n, perform backward chaining and find three highly relevant documents.

3. Using document #n, perform forward chaining and find three highly relevant documents.

4. Refine the initial query. Then review the new set of search results and judge the quality of the refined query.

The documents that were selected for the citation navigation sub-tasks (the second and the third sub-tasks) were highly relevant documents selected from the first page of search results. In addition, they had a similar number of forward citations and backward citations to further control the potential effects of the task complexity on the results.

5.3.2 Procedure

As the first step of the study, pre-study questionnaires were administered to determine participants' educational background, their prior experience with scholarly search, and their scholarly search behaviour. After the completion of pre-study questionnaires, a brief description of the new features of Bow Tie Academic Search was provided to ensure that each participant had a preliminary understanding of the interface.

Participants were asked to perform a different exploratory search task with each interface. In order to alleviate the biasing effects of the task order as well as the interface order, such as the impact of learning and fatigue, a Graeco-Latin Square design was used to systematically rotate task and interface variables in the study (Table 5.1). To control the potential impacts of the participants' individual differences on the results of the study, they were assigned to different groups in a round-robin fashion.

Participants were then asked to follow the sub-tasks and find highly relevant documents. During each sub-task, participants were asked to verbally indicate the documents they deemed relevant to the investigator, who was responsible for noting them in the questionnaire. To avoid the potential learning effects that may be caused by reading the documents' content, participants were asked to only consider the document surrogates for relevance assessments. However, to perform backward chaining with the list-based interface, they could use the list of backward citations at the end of the actual documents since it is not possible to directly perform backward chaining using most of the current search interfaces.

Participant Groups	Time 1	Time 2
G_1	I_1, T_1	I_2, T_2
G_2	I_1, T_2	I_2, T_1
G_3	I_2, T_1	I_1, T_2
G_4	I_2, T_2	I_1, T_1

Table 5.1: Graeco-Latin Square rotation of search tasks and search interfaces.

During each task, efficiency measurements of the overall time to complete the task and the amount of time to complete each of the sub-tasks were made. At the end of each task, participants were provided with an in-task questionnaire to measure their confidence in finding a relevant set of documents, their topical familiarity with the search task, as well as their perceptions of the quality of the refined query.

Once both tasks were complete, post-study questionnaires designed based on the guidelines of the Technology Acceptance Model (TAM) [28] were given to the participants, collecting their subjective opinions of the ease of use and usefulness of the both interfaces. Moreover, questions were asked regarding the ease of use and usefulness of the specific features of Bow Tie Academic Search. Finally, the participants were asked to indicate their preference for a library search interface, and to provide comments about the features of the interfaces and the tasks. Video and audio recording was used during the study to facilitate further analysis of the participants exhibiting exceptionally successful or unsuccessful performance. This entire procedure took about 60 minutes for each participant.

5.3.3 Analysis

To compare retrieval efficiencies of the two interfaces, the overall time to complete the tasks as well as the time to complete each of the sub-tasks were used. Since these measures are quantitative, they were analyzed via analysis of variance (ANOVA). The ANOVA results of retrieval efficiency are analyzed and reported independently for each task.

Another aspect of usability measured in this study is the retrieval effectiveness.

Since measuring the quality of the outcome is a common method to assess the effectiveness of a system [61], retrieval effectiveness is measured based on the quality of the sub-tasks' outcomes.

For the first three sub-tasks, the participants were asked to select highly relevant documents to a specified information need. The quality of the selected documents are measured based on experts' judgment of the relevancy of the document to the specified topic. As the task topics are highly related to the research areas of the author and her thesis supervisor, the documents selected by the participants were inspected by them as experts to verify their relevance to the information needs.

Evaluating document surrogates independently, each expert assigned relevance scores to the selected documents using the four-point relevance scale [103, 50] (see Table 5.2). In case of any conflict in these scores, documents were re-evaluated and discussed to reach an agreement. Documents that were assigned a relevance score of one or two were considered relevant. This process results in ground truth relevance scores [50]. To analyze the quality of the selected documents ANOVA is used, and the results are independently analyzed and reported for each task.

In the query refinement sub-task, the participants were asked to refine their queries

Score	Description
1	This document is relevant. I would probably click on it.
2	This document is probably relevant. I would likely click on it.
3	This document is probably not relevant. I might click on it.
4	This document is not relevant. I would not click on it.

Table 5.2: The relevance scores used to rate document surrogates.

to target a sub-topic they find interesting. Since these results can only be assessed based on participants' intended purpose [9], the quality of the refined query is measured based on the participants' perception of the quality elicited from the in-task questionnaires. This qualitative data was analyzed using Wilcoxon-Mann-Whitney test.

Other qualitative data collected from in-task and post-study questionnaires (i.e., confidence, perceived ease of use, and perceived usefulness) were analyzed using Wilcoxon-Mann-Whitney tests as well. The preference ranks were analyzed pair-wise using Wilcoxon signed rank. Finally, to analyze the open-ended questions, thematic analysis was performed.

5.4 Results

5.4.1 Participant Demographics

24 participants were recruited from graduate computer science and computer engineering courses to participate in this study. Pre-study questionnaires were designed to elicit some basic measures to describe participants' scholarly search experience, along with their search behaviour. The results from these questionnaires are presented in Table 5.3.

Even though the participants reported a wide range in the numbers of scholarly searches they regularly perform, all of them indicated that they conduct at least few scholarly searches per week. This finding was expected since the participant pool was limited to only include graduate students.

Table 5.3: Features of the participant demographics.

Scholary Searches	0 times per week: 0% 1-5 times per week: 33% 6-10 times per week: 25% 11-15 times per week: 25% 15+ times per week: 17%
Use of Forward Citation Metadata	Never: 0% Seldom: 8% Sometimes: 42% Often: 37% Always: 13%
Use of Backward Citation Metadata	Never: 4% Seldom: 8% Sometimes: 29% Often: 38% Always: 21%
Use of Keyword Metadata	Never: 8% Seldom: 17% Sometimes: 13% Often: 54% Always: 8%
Use of Backward and Forward Chaining	Never: 0% Seldom: 0% Sometimes: 4% Often: 67% Always: 29%
Likelihood of Refining a Query	Never: 0% Seldom: 0% Sometimes: 17% Often: 33% Always: 50%

Most of the participants reported the use of forward citation, backward citation, and keyword metadata for evaluating and comparing search results, although there was some variability in these results. They also reported a frequent use of backward and forward chaining methods when searching for academic documents. Similarly, query refinement was reported as a common search activity among the participants. Therefore, according to the discussion on information retrieval behaviour in Section 1.1, it can be assumed that they were representative of the real-world academic library

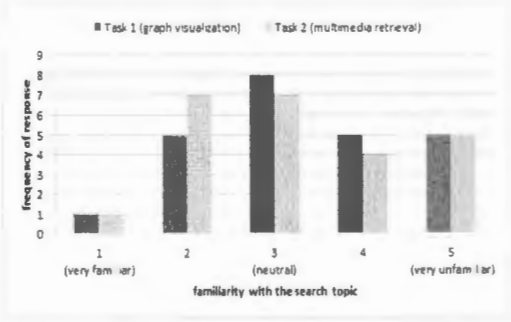


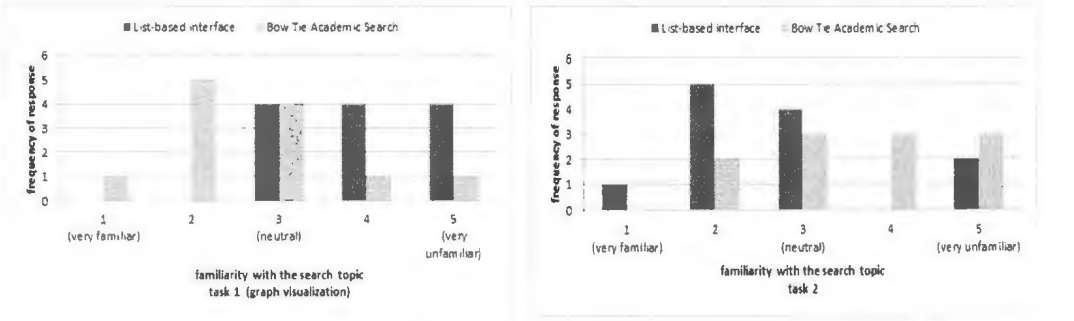
Figure 5.2: Familiarity with the assigned search topic regardless of the assigned interface as reported by the participants.

searchers.

After completing each task, the participants were asked to report their familiarity with the assigned search topic. Figure 5.2 shows the collected responses for this question. These results indicate that the majority of participants had some level of familiarity with the both search topics, while they also had the expected knowledge gap that needs to be resolved through the exploratory search process. To facilitate further analysis of the results, these findings are also reported independently for each interface (see Figure 5.3). As it can be seen, the participants reported a wider range of familiarity levels between the groups.

5.4.2 Retrieval Efficiency

Prior to the user study, it was expected that participants will find relevant documents faster with the list-based interface than with Bow Tie Academic Search (hypothesis H1.1). In order to compare efficiency measures, it is critical to precisely specify the criteria by which the task is considered complete [55]. In this study, the time to complete the task is the total time it takes for a participant to complete all the



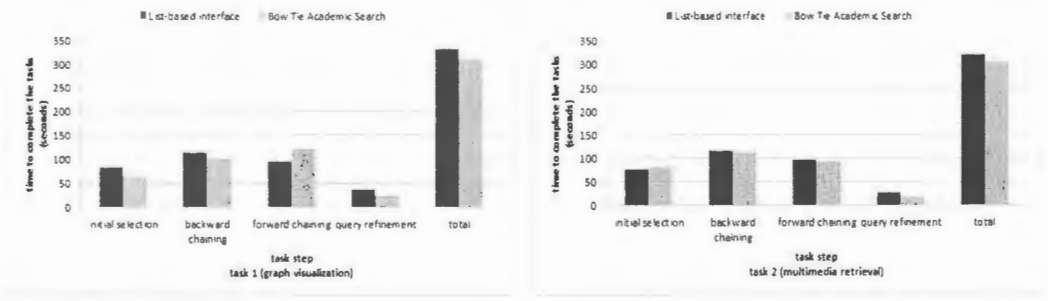
(a) (b)

Figure 5.3: Familiarity with the assigned search topic of Task 1 (a) and Task 2 (b) as reported by the participants.

sub-tasks. As such, each sub-task was assigned a completion criterion. For the first three sub-tasks, upon finding the third highly relevant document, the task was considered complete. For the query refinement sub-task, submitting the refined query was specified as a completion benchmark. Figure 5.4 illustrates the average time the participants took to complete each sub-task as well as the average of the total time needed to complete each task.

Note that one outlier data point was removed for the analysis as the corresponding participant spent about 25 minutes more that the average time of 10 minutes that it took other participants to complete both tasks. By consulting the video recording, it could be concluded that he spent this extra time to evaluate individual documents as he was looking for a more detailed answer than the other participants were.

On average, the participants were able to complete the backward chaining and query refinement sub-tasks faster with Bow Tie Academic Search than the list-based interface. For the initial selection and forward chaining sub-tasks, however, there are mixed results for different search tasks. Overall, these time differences lead to



(a) (b)

Figure 5.4: The average time to complete Task 1 (a) and Task 2 (b), along with the average time to complete their sub-tasks.

a faster completion of both search tasks with Bow Tie Academic Search than the list-based interface. However, none of the time differences is proven to be statistically significant. The results of ANOVA tests are provided in Table 5.4.

These results indicate that the participants were able to complete the tasks with both interfaces in similar times even though they were required to interact with Bow Tie Academic Search interface and perform exploration activities via its features. This finding is promising and is even better than what was expected before the study (hypothesis H1.1). However, since it was the participants' first exposure to the Bow Tie Academic Search tools, it can be argued that as they become more experienced

Table 5.4: Statistical analysis (ANOVA) of the responses for time to completion data.

Task	Initial Selection	Backward Chaining	Forward Chaining	Query Refinement	Total
Task 1	$F(1,21) = 0.44$, $p = 0.51$	$F(1,21) = 0.42$, $p = 0.52$	$F(1,21) = 0.73$, $p = 0.39$	$F(1,21) = 1.69$, $p = 0.20$	$F(1,21) = 0.17$, $p = 0.67$
Task 2	$F(1,21) = 0.05$, $p = 0.81$	$F(1,21) = 0.02$, $p = 0.86$	$F(1,21) = 0.02$, $p = 0.86$	$F(1,21) = 4.00$, $p = 0.05$	$F(1,21) = 0.04$, $p = 0.82$

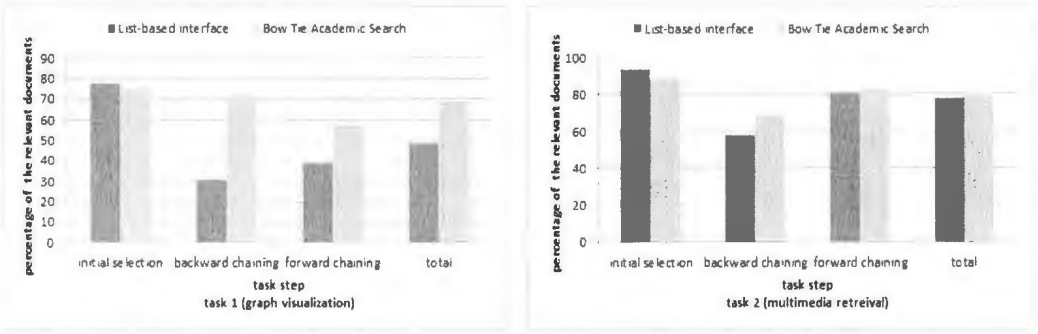
using these features, their efficiency may further improve.

5.4.3 Retrieval Effectiveness

5.4.3.1 Selected Documents Quality

Based on the features of the both interfaces, it was hypothesized that participants will be able to find a better set of documents using Bow Tie Academic Search than using the list-based interface (hypothesis H1.2). The ability of the participants to find a good set of documents was evaluated based on the experts assessments of the documents' relevancies. The percentage of the relevant documents to the total selected documents (known as precision within the information retrieval literature) is provided in Figure 5.5, for different sub-tasks and different interfaces. Except for the initial selection sub-task, the number of relevant documents selected using Bow Tie Academic Search is more than the relevant documents selected using the list-based interface. However, the results of ANOVA tests indicate that these differences are statistically significant only for the backward chaining sub-task of Task 1 and the entire set of documents selected in this task (see Table 5.5).

When performing the initial sub-tasks, most of the participants only considered the first page of the search results (10 documents) using either the list-based interface or Bow Tie Academic Search. Given that they were provided with the same set of documents, the negligible differences in the number of relevant documents selected in this sub-task can be explained. However, these differences are more considerable for the citation navigation sub-tasks and even statistically significant in one case (backward citation sub-task of Task 1). This improvement may be a result of the



(a) (b)

Figure 5.5: The percentage of the relevant documents to the total selected documents in different steps of Task 1 (a) and Task 2 (b).

Table 5.5: Statistical analysis (ANOVA) of the quality of the selected documents.

Task	Initial Selection	Backward Chaining	Forward Chaining	Total
Task 1	$F(1,70) = 0.07$, $p = 0.78$	$\mathbf{F(1,70) = 14.71}$, $\mathbf{p < 0.001}$	$F(1,70) = 2.75$, $p = 0.10$	$\mathbf{F(1,214) = 8.68}$, $\mathbf{p < 0.01}$
Task 2	$F(1,70) = 0.71$, $p = 0.40$	$F(1,70) = 0.94$, $p = 0.33$	$F(1,70) = 0.09$, $p = 0.76$	$F(1,214) = 0.25$, $p = 0.61$

exploration support provided via the filtering operation.

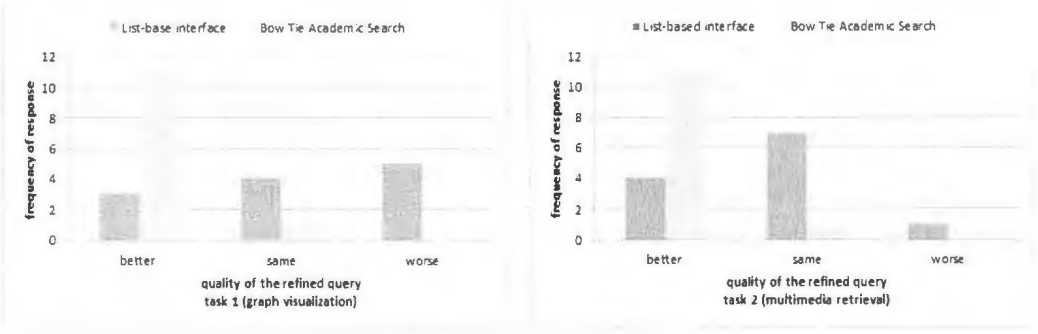
Overall, these mixed results suggest that the tools provided in Bow Tie Academic Search may not necessarily affect the searchers' ability to better assess the relevance of individual documents. As such, for this measure, superiority of any interface over the other could not be confirmed. However, these results may be specific to the given tasks and the features of the retrieved documents that were evaluated.

5.4.3.2 Refined Query Quality

Considering the query refinement support provided by Bow Tie Academic Search, it was expected that participants will be able to reformulate their queries more effec-

tively using this system than using the list-based interface (hypothesis H1.3). After each task was complete, the participants were asked to judge the relative quality of the refined query by indicating whether its quality is better, the same, or worse that the quality of the initial query. These judgments were considered to be accurate. The goal was to evaluate the participants' ability to improve the search process by refining their queries using the two different interfaces. The responses to this question are reported in Figure 5.6. As can be seen, the results were in favour of Bow Tie Academic Search for both tasks. These differences are also shown to be statistically significant via Wilcoxon-Mann-Whitney tests (see Table 5.6).

These results suggest that the participants could reformulate better queries using Bow Tie Academic Search than using the list-based interface, and they confirm the prior expectation. This superiority can be attributed to the support provided by the visual and interactive features of the enhanced histogram in recognition of relevant terms and their meta-attributes; and in modifying a query by adding and removing the terms.



(a) (b)

Figure 5.6: The participants' perceptions of the quality of the refined query in comparison to the initial query for Task 1 (a) and Task 2 (b).

Table 5.6: Statistical analysis (Wilcoxon-Mann-Whitney tests) of the responses for the perception of the quality of the refined query in comparison to the initial one.

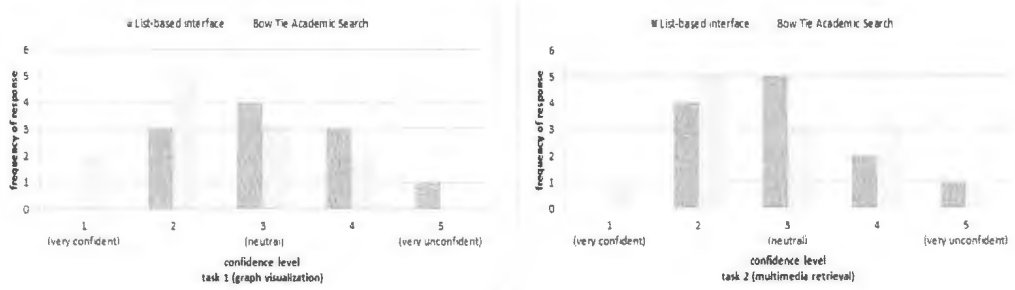
Task	List-based interface vs. Bow Tie Academic Search
Task 1	$\mathbf{Z = -3.29, p < 0.001}$
Task 2	$\mathbf{Z = -2.89, p < 0.001}$

Although the group of the participants that performed Task 2 with Bow Tie Academic Search reported a lower level of familiarity with this task in comparison to those that used the list-based interface (see Figure 5.3b), they could still perform better in the query refinement step. This finding implies that regardless of the searchers' prior knowledge of the search topic, the enhanced histogram can improve searchers' ability to better refine their queries.

5.4.4 Confidence

Due to the extra visual and interactive features of Bow Tie Academic Search, the expectation was that participants will report a higher level of confidence in finding relevant documents using Bow Tie Academic Search than using the list-based interface (hypothesis H1.4). Using in-task questionnaires, the participants reported their confidence level in selecting a highly relevant set of documents (Figure 5.7). For both tasks, the participants expressed a higher degree of confidence using Bow Tie Academic Search than using the list-based interface. However, the results of Wilcoxon-Mann-Whitney tests indicate that these differences are not statistically significant (see Table 5.7).

Even though Bow Tie Academic Search improved this aspect, its superiority over



(a) (b)

Figure 5.7: The participants' confidence degree in selecting a good set of document when performing Task 1 (a) and Task 2(b).

Table 5.7: Statistical analysis (Wilcoxon-Mann-Whitney tests) of the responses for the degree of confidence in selecting a good set of documents.

Task	List-based interface vs. Bow Tie Academic Search
Task 1	$Z = -1.71, p = 0.08$
Task 2	$Z = -0.76, p = 0.44$

the list-based interface could not be confirmed. One may argue that searchers' familiarity with the search topic is a stronger predicator of searchers' confidence in selecting document. As such, the differences for this measure may not be significant for the searchers who have low or high level of familiarity with the topic, but for those who are moderately familiar with the search topic. Further study with different groups of participants is required to verify this argument.

5.4.5 Perceived Ease of Use and Perceived Usefulness

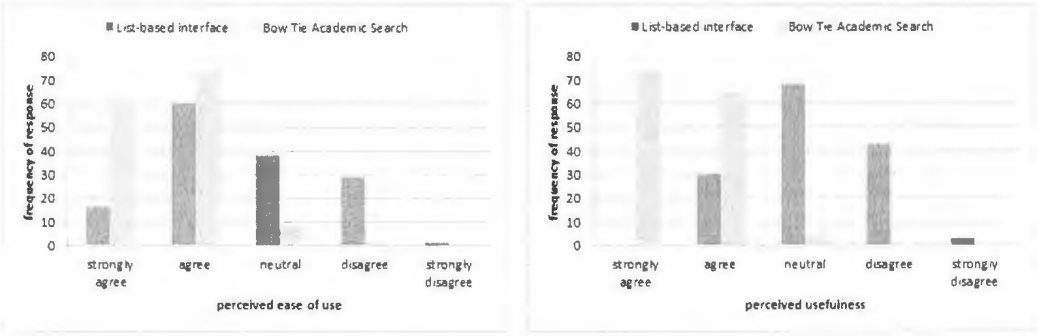
After completing both search tasks, the participants provided their perceptions of the ease of use and usefulness of the list-based interfaces as well as Bow Tie Academic

Search. They also indicated their perceptions of these measures for the specific components of Bow Tie Academic Search, namely bow tie representations, detailed bow tie representations, and the enhanced histogram. This data was collected through post-study questionnaires using the TAM instrument. The TAM instrument uses multiple questions to measure the different aspects of ease of use and usefulness. In this study, the data collected from these questions was aggregated based on the purpose of the questions.

5.4.5.1 List-based interface vs. Bow Tie Academic Search

For the perceived ease of use, it was expected that participants will report that the list-based interface is easier to use than Bow Tie Academic Search (hypothesis H1.5). For the perceived usefulness, however, it was expected that participants will report that Bow Tie Academic Search is more useful for finding academic documents than the list-based interface (hypothesis H1.6). Figure 5.8 shows the frequency of response to the different statements asked about the perceived ease of use and perceived usefulness. For both tasks, the participants reported that Bow Tie Academic Search is easier to use than the list-based interface. As well, they found Bow Tie Academic Search more useful for retrieving academic documents. Wilcoxon-Mann-Whitney tests were performed using a pair-wise grouping of the interfaces. Statistical significance was found for both measures and for both tasks (See Table 5.8).

These results indicate the superiority of Bow Tie Academic Search over the list-based interface for the perceived ease of use measure. This finding is better than what was expected prior to the study. The conventional list-based interface is designed to be easy to use, while the Bow Tie Academic Search is a relatively complex interface.



(a) (b)

Figure 5.8: The frequency of response to the different statements of the perceived ease of use (a) and perceived usefulness (b) for the two interfaces.

Table 5.8: Statistical analysis (Wilcoxon-Mann-Whitney tests) of the responses for the perceived ease of use and perceived usefulness.

Measure	List-based interface vs. Bow Tie Academic Search
Perceived Ease of Use	$\mathbf{Z = -8.45, p < 0.001}$
Perceived Usefulness	$\mathbf{Z = -13.33, p < 0.001}$

However, to perform complex analysis tasks such as forward and backward citation analysis, the visual exploration support provided by Bow Tie Academic Search improved searchers control and ability, resulting in their more positive perceptions of its ease of use over the list-based interface.

As expected, Bow Tie Academic Search could also positively affect the participants' perceptions of the usefulness of the search system. This superiority can be attributed to the participants' enhanced ability in finding relevant documents through backward and forward chaining (see Section 5.4.3.1) as well as constructing more effective queries (see Section 5.4.3.2).

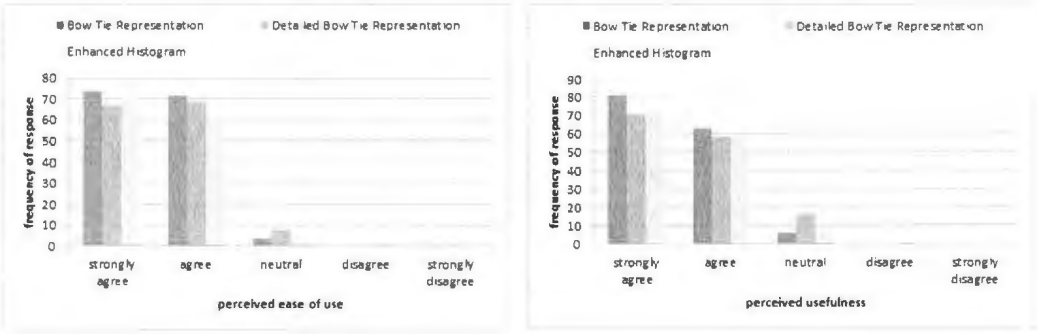
5.4.5.2 Bow Tie Academic Search Components

In the second set of hypotheses formulated to predict the perceived ease of use and perceived usefulness of the individual components of Bow Tie Academic Search, it was hypothesized that participants will find bow tie representations easy to use (hypothesis H2.1). Moreover, it was expected that participants will find bow tie representations useful in the document retrieval process (hypothesis H2.2). Similarly, for the detailed bow tie representation, the expectation was that participants will find it easy to use (hypothesis H2.3) and useful (hypothesis H2.4) when retrieving academic documents. The enhanced histogram was also expected to be perceived as an easy to use (hypothesis H2.5) and useful (hypothesis H2.6) retrieval tool.

The frequency of response to the perceived ease of use and perceived usefulness of the different components of Bow Tie Academic Search is shown in Figure 5.9. As can be seen, the participants either "strongly agreed" or "agreed" with the majority of the statements of the TAM questionnaire. These results are consistent for both measures and for all of the interface features.

This finding confirms the hypotheses formulated about the ease of use and usefulness of bow tie representations (hypotheses H2.1 and H2.2), suggesting that the participants found the encoded citation information useful for academic search. They were also able to easily understand and perceive this information from the visual features of bow tie representations.

The results for ease of use and usefulness of the detailed bow tie representation are also consistent with prior expectations (hypotheses H2.3 and H2.4). This finding indicates that providing detailed information of citation metadata as well as citation



(a) (b)

Figure 5.9: The frequency of response to the different statements of the perceived ease of use (a) and perceived usefulness (b) for individual interface features.

navigation support was considered useful and effective by the participants. As well, the visual attributes to encode this data and the low-level interaction mechanism to facilitated exploration were easy to understand and use.

Moreover, the findings regarding the ease of use and usefulness of the enhanced histogram confirms the corresponding hypotheses (hypotheses H2.5 and H2.6), indicating that the participants had a positive perception about the effectiveness of keyword metadata encoded in the enhanced histogram as well as the search results exploration feature in the query refinement process. In addition, the interactive features to support exploration and query modification were found easy to use by the participants.

In total, these results confirm the choices made regarding the metadata elements to be visually represented and the visual and interactive features to represent these metadata elements. One may relate the few neutral and negative responses to the very little time the participants had to learn about all of the visual and interactive features of these tools.

5.4.6 Preference

It was expected that participants will prefer Bow Tie Academic Search over the list-based interface for retrieving academic documents (hypothesis H1.7). In post-study questionnaires, the participants were asked to indicate their preference for an academic library search interface. 22 out of 24 participants indicated that they prefer Bow Tie Academic Search (92%) over the list-based interface for their scholarly searches. A Wilcoxon signed rank test proved this measure to be statistically significant ($\mathbf{Z = -4.89, p < 0.001}$). This result can clearly confirm the hypothesis formulated regarding the participants' preference of Bow Tie Academic Search over the list-based interface.

5.4.7 Open-ended Questions

At the end of post-study questionnaires, the participants were asked to provide their comments about the features of the interfaces and the tasks. However, all of the comments provided by the participants were about the visual tools of the Bow Tie Academic Search.

In general, the comments were positive about these tools. For bow representations, the general consensus was that it is useful and helpful during the retrieval process. The participants appreciated the unified representation of backward and forward citation lists and the interactive citation exploration feature of the detailed bow tie representation. Regarding the citation navigation support, it was mentioned that "having a quick way to filter based on both citations and references is very helpful." The keyword histogram was also mentioned to be very useful in the query refinement

process by some of the participants. One of these participants particularly liked the ability of the enhanced histogram "to provide proper terminology related to the body of work without the searcher explicitly knowing about it or having to search for it."

While the detailed bow tie representation was appreciated by most of the participants, a few mentioned that it was difficult for them to select filters due to the small size of the year bars. These comments suggest the need to provide a method to make this task easier such as binning the years when citations are distributed over many different years.

Some of the participants provided suggestions for improvement in the system. It was noted that adding an undo button that deactivates all the filters at once in the detailed bow tie representation would be useful. The undo button was also asked to be added to the enhanced histogram to enable searchers to use their previously formulated queries. In addition, it was mentioned by a participant that it would be helpful to separately provide a list of the years that are selected as filters in the detailed representation of documents.

5.5 Discussion

In this evaluation, a user study was conducted in a controlled laboratory setting to validate the potential value of Bow Tie Academic Search compared to a baseline system (the list-based interface). Moreover, the specific design decisions incorporated in the Bow Tie Academic Search tools were evaluated. This study was designed to minimize the potential biasing effects by providing two different search tasks and by varying the order in which the participants were exposed to the tasks and to the

interfaces.

By providing an imaginative context and ambiguous information needs as well as asking the participants to find multiple documents in the search tasks, their active engagement and exploration during the retrieval process could be ensured. In addition, to ensure that participants have a basic knowledge to judge the documents' relevancy, the participants were recruited from computer science and computer engineering students to perform tasks that they might be required to do as part of taking a graduate course in these disciplines.

For the comparison between the two interfaces, the usability measures of retrieval effectiveness (quality of the selected documents and the quality of the refined query) and retrieval efficiency as well as the subjective measures of confidence, perceived ease of use and perceived usefulness were made. The analysis of the retrieval efficiency data provided better results than our prior expectation, indicating that searchers may not necessarily perform slower with Bow Tie Academic Search although they have extra features to use. Another interesting finding was about the perceived ease of use measure. Even though it was expected that participants will find the list-based interface easier to use, they reported a higher degree of ease of use for Bow Tie Academic Search.

The results for the quality of the refined query are consistent with what was expected, as the analysis indicates that the participants were able to refine their queries more effectively using Bow Tie Academic Search. Similarly, as expected, the participants reported more positive perceptions of the usefulness of Bow Tie Academic Search for retrieval purposes. Another measure for which the results confirm our expectation is the preference. As an answer to their preference of a library search

interface, 92% of the participants preferred Bow Tie Academic Search.

The analysis of the quality of the selected documents provided mixed results for different tasks. Even though the improvement is consistent for the two tasks, the differences are only statistically significant for the backward chaining sub-task of Task 1 and the entire set of documents selected in Task 1. Although the tasks were designed to be similar in terms of complexity, this difference in statistical significance can be attributed to the potential differences in the tasks and in the retrieved sets of documents. As such, further study with a more comprehensive set of tasks may reveal what features of a search task lead to a better performance with Bow Tie Academic Search.

Following each task, the participants were asked to rate their degree of confidence in selecting a good set of documents. For both of the tasks, the results of the confidence are slightly skewed more positively for Bow Tie Academic Search compared to the list-based interface. However, the differences are not statistically significant and the corresponding hypothesis could not be confirmed.

To evaluate our specific design choices, the perceived ease of use and perceived usefulness for each of the features of Bow Tie Academic Search were measured. As expected, when asked to assess these two measures, the participants reported a high degree of their feelings about the ease of use and usefulness of all of the extra features of the system.

Overall, the results of this user study provide evidence regarding the positive impacts of Bow Tie Academic Search on exploratory search processes, and it supports the design assumptions incorporated in each of the tools of the system. More importantly, it validates the fundamental assumption regarding the value of metadata

visualizations and metadata-based interactions and explorations in library search interfaces. However, given the constraints of time and complexity, only two search tasks were used. In addition, the results could be influenced by the participants' prior topic experience. As a result, it is difficult to draw strong and generalizable conclusions from this study. As such, it can be considered a preliminary study, and a more detailed study with a broader range of library search tasks and different groups of participants with even prior experience in a topic is warranted.

Chapter 6

Conclusions and Future Work

The main objective of this research has been to promote and incorporate the behaviours and exploratory activities of library searchers engaged in research using digital library search interfaces. In order to accomplish this goal, visual representations of widely used metadata elements were designed to provide a novel search interface intended to let searchers intuitively perform common library search strategies. By utilizing citation metadata, an approach was proposed to enable citation-based evaluation and comparison of documents as well as forward and backward citation navigation (Chapter 3). To let searchers subsequently elaborate and refine their queries, keyword metadata was used to provide a visual and interactive query refinement tool (Chapter 4). Finally, a user study was conducted to answer the research questions about the impacts of the proposed tools on the library information retrieval tasks (Chapter 5). The primary contributions of this research as well as the potential future directions are outlined in the reminder of this Chapter.

6.1 Research Contributions

6.1.1 Design of Bow Tie Academic Search

While the Web facilitates fast and direct access to large collections of academic resources, the ways people obtain information from these collections are based on Web search technologies and interfaces. One of the main contributions of this research is the design and development of a novel search interface that can support library exploratory search processes by addressing the fundamental issues of current search interfaces. This support is provided through the visualization of metadata elements to enable fast and easy interpretation of document features as well as the incorporation of metadata-based interactions to enable manipulation and exploration of the retrieved set.

A fundamental research question raised about providing such support was, *how can metadata visualizations and metadata-based interactions be designed to support interactive search results exploration and interactive query refinement in digital library search interfaces?* To answer this research question, library search behaviour studies were reviewed to identify what metadata elements library searchers commonly use and what search activities they normally perform using these metadata elements. Using the knowledge drawn from these studies, citation and keyword metadata was chosen to form the basis of the new interface components. Information visualization and human-computer interaction principles and guidelines were followed in the design of visual and interactive representations of these metadata elements, allowing searchers to iteratively explore the search results set and refine their queries.

Bow tie representations were designed to visually represent citation information

of each document, providing searchers with the ability to make quick judgments about the potential importance and relevance of the documents. The detailed bow tie representations depict the distribution of citations of a document over different years, allowing searchers to visually evaluate an individual document based on the characteristics of its backward and forward citation sets. The filtering operation of the detailed bow tie representation lets searchers navigate within the citation sets of a document, aiding them in employing the common backward and forward chaining strategies.

The enhanced keyword frequency histogram was designed to provide a visual overview of the search results, letting searchers evaluate the overall relevance of the retrieved set, interpret the relevance of each keyword, and perceive document similarities based on the co-use of keywords. Interactive query refinement is supported through the interaction methods of the histogram by which searchers are able to add, remove, or replace the suggested keywords with the initial query.

Even though these tools are designed based on information visualization and human-computer interaction principles and theories, Bow Tie Academic Search is still one of the many possible designs to represent these metadata elements. In order to validate the specific design choices that were made, a user study evaluation was conducted. The main contributions made from the findings of this user study are explained in the next section.

6.1.2 User Study Findings

After the design and development of Bow Tie Academic Search, the second research question emerged as *What is the impact of Bow Tie Academic Search on efficiency, effectiveness and subjective impressions of the search process?* In order to answer this question, a user study was conducted to compare Bow Tie Academic Search with a list-based interface in a controlled laboratory setting.

The results of this study show that although the participants were expected to spend more time using Bow Tie Academic Search, they could complete the assigned tasks in similar times with both interfaces. As discussed in Section 5.4.2, it can be expected that as searchers become experienced with Bow Tie Academic Search, they may be able to perform their tasks even faster. In terms of retrieval effectiveness, the participants were able to find a more relevant set of documents using Bow Tie Academic Search and these differences are statistically significant for the backward citation navigation sub-task of Task 1 and the entire set of documents selected in this task. They also were able to construct more effective queries using Bow Tie Academic Search than using the list-based interface. Moreover, Bow Tie Academic Search could positively affect participants' perceptions of the ease of use and usefulness of a search interface. Furthermore, the participants reported that they prefer Bow Tie Academic Search for library retrieval tasks over the list-based interface.

As a result of this user evaluation, it can be concluded that Bow Tie Academic Search is capable of improving retrieval effectiveness without a negative impact on retrieval efficiency. As well, it can positively affect searchers' subjective opinions of a search interface. Overall, these results support the fundamental hypothesis regarding

the effectiveness of metadata-based visual and interactive tools in exploratory search processes.

Finally, regarding the subjective opinions of the specific components of the system, the research question was, *What are the searchers' perceptions of the usefulness and ease of use of each specific component of Bow Tie Academic Search?* By analyzing the results of the user evaluation, it was concluded that the participants reported a high level of the ease of use and usefulness for all of the components of the system.

This finding confirms our decisions made in regard of the metadata elements to be visually represented as well as the metadata-based explorations to be interactively supported. Moreover, it validates our design choices made in the creation of each of the Bow Tie Academic Search components.

6.2 Future Work

6.2.1 Further Enrichment of the Proposed Tools

The results of the user study have shown that the participants found Bow Tie Academic Search and its individual components easy to use, and they did not encounter any difficulties using them (see Section 5.4.5). This finding indicates that there is an opportunity to further enhance the Bow Tie Academic Search tools with extra visual and interactive features in order to support more retrieval activities.

For instance, bow tie representations can be augmented with extra visual cues, such as small dots in bow tie sides, to show if a document's backward or forward citation set includes another document retrieved by the given query. Such a visual

depiction can represent citation connections among the entire retrieved set. Another way to enhance bow tie representations is to use colour coding (similar to the approaches presented in TileBars [45] and HotMap [54, 58]) to visually represent other features of the citation sets such as the average year or the average rank of the documents included in the set. Similarly, for the detailed bow tie representation, the overall features of the citations published in different years can be encoded through the colour of the year bars, which may support searchers in fast and effective selection of filters. The enhanced histogram can also be augmented by using additional interaction mechanisms to let searchers filter or re-sort search results based on their keyword information, which can allow searchers to evaluate their decisions prior to applying it.

6.2.2 Exploring Visualization Techniques to Represent Other Metadata Elements

Even though the focus of Bow Tie Academic Search is currently on citation and keyword metadata, library search behaviour includes a wide variety of information seeking methods using various metadata elements. There is value in supporting more retrieval techniques by providing visualization of other commonly used metadata elements.

For example, "journal run", which is also a very common strategy [5], involves identification of a source and then selection of the relevant documents from the selected source. Visual representations of the relevant and core venues along with useful metadata elements, such as their citation count, can be designed and incorporated

to the search interface to support this strategic behaviour. Another example of a frequently used method is "author searching" [5], which can be supported by proving visual representations of author relations in terms of citation links or co-authorship. Metadata elements associated with authors such as their h-index or research interest can also be visualized to support evaluation and comparison of authors, which may lead to a more effective evaluation and comparison of documents.

6.2.3 Further Evaluations

In this research, Bow Tie Academic Search was evaluated in a controlled laboratory setting. Laboratory studies allow researchers to control the potential impacts of the unwanted variables; hence, they can facilitate a more accurate comparison [18]. However, their ability to capture subjective measures is limited to a small set of participants and a small number of pre-specified tasks. As such, additional evaluation methods are needed to gain an insight into the potential effects of the system on a more comprehensive set of measures such as learning, cognitive load, engagement, and enjoyment, which are the recommended metrics to assess the performance of exploratory search systems [118]. In addition, further studies against a wider range of approaches may be of value to better assess the effects of the system.

In this context, field trials may be helpful as the evaluation is based on the real world tasks. However, for Bow Tie Academic Search, which is particularly designed to support exploratory tasks, longitudinal evaluations may be more effective as they measure change over an extended period of time [118, 70]. An appealing alternative is to follow the structured and systematic approach of the stepped evaluation model

[55], which includes multiple evaluation methods as well as multiple refinements of the system based on the outcome of each evaluation.

Appendix A

User Study Documentation

This appendix includes the formal approval received from the Interdisciplinary Committee on Ethics in Human Research (ICEHR) for the user evaluation and the evaluation documents.

ICEHR Number:	**20130340-SC**
Approval Period:	**July 20, 2012 – July 31, 2013**
Funding Source:	Supervisor's NSERC funding
Responsible Faculty:	Dr. Orland Hoeber Department of Computer Science
Title of Project:	*A comparison of list-based and visual search interfaces in digital libraries*

July 20, 2012

Mrs. Taraneh Khazaei
Department of Computer Science
Memorial University of Newfoundland

Dear Mrs. Khazaei:

Thank you for your submission to the Interdisciplinary Committee on Ethics in Human Research (ICEHR) seeking ethical clearance for the above-named research project.

The Committee has reviewed the proposal and appreciates the care and diligence with which you have prepared your application. We agree that the proposed project is consistent with the guidelines of the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans* (TCPS2). *Full ethics clearance* is granted for one year from the date of this letter.

However, the Committee requires that you make the following clarifications:

a. In the consent form, on data storage specify "a minimum of five years, as per Memorial University policy on Integrity in Scholarly Research."
b. In the recruitment poster: [i] add the ICEHR approval statement; [2] relocate the compensation information so that it is not at the top of the document; [3] correct typing and spelling errors.

Please forward the revised documents to Theresa Heath at [illegible]

If you intend to make changes during the course of the project which may give rise to ethical concerns, please forward a description of these changes to Theresa Heath at [illegible] for the Committee's consideration.

The TCPS2 requires that you submit an annual status report on your project to ICEHR, should the research carry on beyond July 31, 2013. Also, to comply with the TCPS2, please notify us upon completion of your project.

We wish you success with your research.

Yours sincerely,

Michael Shute, Th.D.
Chair, Interdisciplinary Committee on
Ethics in Human Research

MS/th

copy: Supervisor – Dr. Orland Hoeber Department of Computer Science, Faculty of Science

Director, Office of Research Services

Informed Consent Form

Title: *A comparison of list-based and visual search interfaces in digital libraries*

Researchers: *Taraneh Khazaei, Dept. of Computer Science, Memorial University of Newfoundland;* [redacted]

Dr. Orland Hoeber, Dept. of Computer Science, Memorial University of Newfoundland; hoeber@mun.ca

You are invited to take part in a research project entitled "*A comparison of list-based and visual search interfaces in digital libraries*".

This form is part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. It also describes your right to withdraw from the study at any time. In order to decide whether you wish to participate in this research study, you should understand enough about its risks and benefits to be able to make an informed decision. This is the informed consent process. Take time to read this carefully and to understand the information given to you. Please contact the researcher, *Taraneh Khazaei*, if you have any questions about the study or for more information not included here before you consent.

It is entirely up to you to decide whether to take part in this research. If you choose not to take part in this research or if you decide to withdraw from the research once it has started, there will be no negative consequences for you, now or in the future.

My name is Taraneh Khazaei and I am a graduate student in Computer Science. As part of my Master's thesis, I am conducting research under the supervision of Dr. Hoeber.

The user study will help in analyzing the value of using visualization techniques in library search interfaces. The primary objective of the study is to determine the benefits of the specific design choices used in our proposed search interface in comparison to the traditional list-based search interfaces.

In this study, you will be asked to complete two information retrieval tasks using two different search interfaces. For each task, you will be given a specified information need and an initial query to start the search process. Each task consists of four information-seeking sub-tasks in which you will be asked to identify highly relevant documents for the assigned information need in the first three sub-tasks, and then refine the query and indicate its quality in the final sub-task. You will be asked to perform each task with only one of the interfaces.

Three different sets of questionnaires will be administrated in this study. A pre-study questionnaire will be used to ask about your prior experience with scholarly search and your educational background. In-study questionnaires, which will be administrated following each

task, will be used to assess your familiarity with the topic of the search task and to evaluate your confidence in finding and selecting documents. A post-study questionnaire will be used to capture your feelings and experiences with using the interfaces.

In a nutshell, you will be required to perform two retrieval tasks, and answer pre-task, in-task, and post-task questionnaires. These activities will require approximately 60 minutes of your time, and will be conducted in Earth Sciences Building (ER-2003) at Memorial University.

You can withdraw your participation in this study at any time, and your decision to participate in this study, and your subsequent involvement in it, will have absolutely no bearing on any other dealings you have with Mrs. Khazaei or Dr. Hoeber.

You will be compensated $10 for participation in this study, regardless of your performance or ability to complete the tasks. Participation in this could be a great experience for you in terms of studying and observing research methods in practice. This might help you in improving your own research (at present or in the future).

Knowledge of your identity is not required. You will not be asked to write your name or any identifying information on the research questionnaires. The video and audio recordings will be used for analysis purposes and the comments you make relevant to the assigned tasks or to the use of the interfaces will be transcribed and identified only using a participant ID. The original raw data will only be accessed by the investigators. All research materials will be held confidential by the Principal Investigator and kept in a secure on-campus location and on password-protected computers for a minimum of five years, as per Memorial University policy on Integrity in Scholarly Research. After this period, the data will be destroyed.

Your activities during the study will be video and audio recorded (over your shoulder, focusing on the computer screen, keyboard, and mouse) to support post-hoc analysis of the results.

The results of the user study will be used for analysis and discussion in Mrs. Khazaei's M.Sc. thesis as well as in conference presentations (2012 International Conference on Knowledge Management and Knowledge Technologies) and journal publications (Information Processing & Management). You are welcome to study the results after they have been published. You can obtain copies of the results in this study, upon completion, by contacting Dr. Hoeber or myself, in care of the Department of Computer Science, Memorial University.

You are welcome to ask questions at any time during your participation in this research. If you would like more information about this study after its completion, please contact: *Taraneh Khazaei at taraneh.khazaei@mun.ca.*

The proposal for this research has been reviewed by the Interdisciplinary Committee on Ethics in Human Research and found to be in compliance with Memorial University's ethics policy. If you have ethical concerns about the research (such as the way you have been treated or your rights as

a participant), you may contact the Chairperson of the ICEHR at icehr@mun.ca or by telephone at 709-864-2861.

Consent:
Your signature on this form means that:

- You have read the information about the research.
- You have been able to ask questions about this study.
- You are satisfied with the answers to all your questions.
- You understand what the study is about and what you will be doing.
- You understand that you are free to withdraw from the study at any time, without having to give a reason, and that doing so will not affect you now or in the future.
- You understand that should you choose to withdraw from the study, any data collected from you up to this point will be deleted/destroyed.

If you sign this form, you do not give up your legal rights and do not release the researchers from their professional responsibilities.

Your signature:
I have read and understood what this study is about and appreciate the risks and benefits. I have had adequate time to think about this and had the opportunity to ask questions and my questions have been answered.

☐ I agree to participate in the research project understanding the risks and contributions of my participation, that my participation is voluntary, and that I may end my participation at any time.

A copy of this Informed Consent Form has been given to me for my records.

Participant Name (please print legibly): _____

Signature : _____

Date : _____

Researcher's Signature:
I have explained this study to the best of my ability. I invited questions and gave answers. I believe that the participant fully understands what is involved in being in the study, any potential risks of the study and that he or she has freely chosen to be in the study.

_____ _____
Signature of Principal Investigator Date

Investigators:
Mrs. Taraneh Khazaei
M.Sc. Student
Department of Computer

Dr. Orland Hoeber
Assistant Professor
Department of Computer

A Comparison of List-based and Visual Search Interfaces in Digital Libraries
PRE-STUDY QUESTIONNAIRE

Please answer the following questions in regards to your background. Circle the answer the best describes you or your opinion.

1. What is your area of research?

   _____

2. How many times a week do you search for academic documents on the Web (e.g., Google Scholar, ACM digital library, IEEE Xplore, etc.)?

0 1-5 6-10 11-15 15+

3. How often do you use citation information (such as the number of times that a document is cited, and/or the year of the publication of the citing documents) to evaluate and compare search results?

Never Seldom Sometimes Often Always

4. How often do you use reference information (such as the number of references of a document, and/or the year of the publication of the references) to evaluate and compare search results?

Never Seldom Sometimes Often Always

5. How often do you use keyword information of documents to evaluate and compare search results?

Never Seldom Sometimes Often Always

6. How often do you browse within the citations or references of a document to find other potential documents?

Never Seldom Sometimes Often Always

7. When the search results do not fulfill your information need, how often do you modify the original query?

Never Seldom Sometimes Often Always

A Comparison of List-based and Visual Search Interfaces in Digital Libraries
TASK SCENARIO

Imagine you are taking a class called "Information Visualization". For this class you need to write a research paper about visualization and interaction techniques to represent graph layout. Use the assigned academic library search interface to find highly relevant research articles as the starting points for your research paper.

Initial query: *graph visualization*

Please perform the following sub-tasks to complete the task:

1. Find three highly relevant documents in the first set of search results retrieved by the given query.

   When you find a relevant document, please indicate its number to the corresponding investigator who will note it on the form below.

   _____
   _____
   _____

2. Using document #2, perform backward chaining and find three highly relevant documents.

   When you find a relevant document, please indicate its title to the corresponding investigator who will note it on the form below.

   _____
   _____
   _____

3. Using document #2, perform forward chaining and find three highly relevant documents.

   When you find a relevant document, please indicate its title to the corresponding investigator who will note it on the form below.

   _____
   _____
   _____

4. Refine the initial query. Then review the new set of results, and judge the quality of the refined query in comparison to the initial query.

A Comparison of List-based and Visual Search Interfaces in Digital Libraries
TASK SCENARIO

Imagine you are taking a class called "Information Organization and Retrieval". For this class you need to write a research paper about methods to retrieve multimedia information. Use the assigned academic library search interface to find highly relevant research articles as the starting points for your research paper.

Initial query: *multimedia retrieval*

Please perform the following sub-tasks to complete the task:

1. Find three highly relevant documents in the first set of search results retrieved by the given query.

   When you find a relevant document, please indicate its number to the corresponding investigator who will note it on the form below.

   _____
   _____
   _____

2. Using document #3, perform backward chaining and find three highly relevant documents.

   When you find a relevant document, please indicate its title to the corresponding investigator who will note it on the form below.

   _____
   _____
   _____

3. Using document #3, perform forward chaining and find three highly relevant documents.

   When you find o relevant document, please indicate its title to the corresponding investigator who will note it on the form below.

   _____
   _____
   _____

4. Refine your initial query. Then review the new set of results, and judge the quality of the refined query in comparison to the initial query.

A Comparison of List-based and Visual Search Interfaces in Digital Libraries
IN-TASK QUESTIONNAIRE

1. How did you find the quality of the refined query in comparison to the initial query?

Better Same Worse

2. How familiar are you with the topic of the search?

1	2	3	4	5
(Very familiar)		(Neutral)		(Very unfamiliar)

3. How confident are you in selecting documents for this task?

1	2	3	4	5
(Very confident)		(Neutral)		(Very unconfident)

A Comparison of List-based and Visual Search Interfaces in Digital Libraries
POST-TASK QUESTIONNAIRE

The following set of questions relate to your experiences using **Microsoft Academic Search (MAS)** as your interface for finding academic documents.
INSTRUCTIONS: Please rate how strongly you agree or disagree with each of the following statements by circling the appropriate number.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Learning to operate **MAS** was easy for me.
1	2	3	4	5	I found it easy to get **MAS** to do what I wanted it to do.
1	2	3	4	5	My interaction with **MAS** was clear and understandable.
1	2	3	4	5	I found **MAS** to be flexible to interact with.
1	2	3	4	5	It was easy for me to become skillful at using **MAS**.
1	2	3	4	5	I found **MAS** easy to use.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Using **MAS** for retrieving academic documents enabled me to accomplish this task more quickly.
1	2	3	4	5	Using **MAS** improved my retrieval performance.
1	2	3	4	5	Using **MAS** increased my retrieval productivity.
1	2	3	4	5	Using **MAS** enhanced my effectiveness when retrieving academic documents.
1	2	3	4	5	Using **MAS** made it easier to retrieve academic documents.
1	2	3	4	5	I found **MAS** useful for retrieving academic documents.

POST-TASK QUESTIONNAIRE (CONTINUE)

The following set of questions relate to your experiences using **Bow Tie Academic Search (BAS)** as your interface for finding academic documents.
INSTRUCTIONS: Please rate how strongly you agree or disagree with each of the following statements by circling the appropriate number.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Learning to operate **BAS** was easy for me.
1	2	3	4	5	I found it easy to get **BAS** to do what I wanted it to do.
1	2	3	4	5	My interaction with **BAS** was clear and understandable.
1	2	3	4	5	I found **BAS** to be flexible to interact with.
1	2	3	4	5	It was easy for me to become skillful at using **BAS**.
1	2	3	4	5	I found **BAS** easy to use.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Using **BAS** for retrieving academic documents enabled me to accomplish this task more quickly.
1	2	3	4	5	Using **BAS** improved my retrieval performance.
1	2	3	4	5	Using **BAS** increased my retrieval productivity.
1	2	3	4	5	Using **BAS** enhanced my effectiveness when retrieving academic documents.
1	2	3	4	5	Using **BAS** made it easier to retrieve academic documents.
1	2	3	4	5	I found **BAS** useful for retrieving academic documents.

POST-TASK QUESTIONNAIRE (CONTINUE)

The following set of questions relate to your experiences using **Bow Tie Representations** as one of the components of Bow Tie Academic Search interface for finding academic documents.

INSTRUCTIONS: Please rate how strongly you agree or disagree with each of the following statements by circling the appropriate number.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Learning to operate the **Bow Tie Representations** was easy for me.
1	2	3	4	5	I found it easy to get the **Bow Tie Representations** to do what I wanted it to do.
1	2	3	4	5	My interaction with the **Bow Tie Representations** was clear and understandable.
1	2	3	4	5	I found the **Bow Tie Representations** to be flexible to interact with.
1	2	3	4	5	It was easy for me to become skillful at using the **Bow Tie Representations**
1	2	3	4	5	I found the **Bow Tie Representations** easy to use.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Using the **Bow Tie Representations** for retrieving academic documents enabled me to accomplish this task more quickly.
1	2	3	4	5	Using the **Bow Tie Representations** improved my retrieval performance.
1	2	3	4	5	Using the **Bow Tie Representations** increased my retrieval productivity.
1	2	3	4	5	Using the **Bow Tie Representations** enhanced my effectiveness when retrieving academic documents.
1	2	3	4	5	Using the **Bow Tie Representations** made it easier to retrieve academic documents.
1	2	3	4	5	I found the **Bow Tie Representations** useful for retrieving academic documents.

POST-TASK QUESTIONNAIRE (CONTINUE)

The following set of questions relate to your experiences using **Detailed Bow Tie Representation** as a component of Bow Tie Academic Search interface for finding academic documents.

INSTRUCTIONS: Please rate how strongly you agree or disagree with each of the following statements by circling the appropriate number.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Learning to operate the **Detailed Bow Tie Representation** was easy for me.
1	2	3	4	5	I found it easy to get the **Detailed Bow Tie Representation** to do what I wanted it to do.
1	2	3	4	5	My interaction with the **Detailed Bow Tie Representation** was clear and understandable.
1	2	3	4	5	I found the **Detailed Bow Tie Representation** to be flexible to interact with.
1	2	3	4	5	It was easy for me to become skillful at using the **Detailed Bow Tie Representation**.
1	2	3	4	5	I found the **Detailed Bow Tie Representation** easy to use.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Using the **Detailed Bow Tie Representation** for retrieving academic documents enabled me to accomplish this task more quickly.
1	2	3	4	5	Using the **Detailed Bow Tie Representation** improved my retrieval performance.
1	2	3	4	5	Using the **Detailed Bow Tie Representation** increased my retrieval productivity.
1	2	3	4	5	Using the **Detailed Bow Tie Representation** enhanced my effectiveness when retrieving academic documents.
1	2	3	4	5	Using the **Detailed Bow Tie Representation** made it easier to retrieve academic documents.
1	2	3	4	5	I found the **Detailed Bow Tie Representation** useful for retrieving academic documents.

POST-TASK QUESTIONNAIRE (CONTINUE)

The following set of questions relate to your experiences using **Enhanced Histogram** as a component of Bow Tie Academic Search interface for finding academic documents.

INSTRUCTIONS: Please rate how strongly you agree or disagree with each of the following statements by circling the appropriate number.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Learning to operate the **Enhanced Histogram** was easy for me.
1	2	3	4	5	I found it easy to get the **Enhanced Histogram** to do what I wanted it to do.
1	2	3	4	5	My interaction with the **Enhanced Histogram** was clear and understandable.
1	2	3	4	5	I found the **Enhanced Histogram** to be flexible to interact with.
1	2	3	4	5	It was easy for me to become skillful at using the **Enhanced Histogram**.
1	2	3	4	5	I found the **Enhanced Histogram** easy to use.

strongly disagree	disagreed	neutral	agree	strongly agree	
1	2	3	4	5	Using the **Enhanced Histogram** for retrieving academic documents enabled me to accomplish this task more quickly.
1	2	3	4	5	Using the **Enhanced Histogram** improved my retrieval performance.
1	2	3	4	5	Using the **Enhanced Histogram** increased my retrieval productivity.
1	2	3	4	5	Using the **Enhanced Histogram** enhanced my effectiveness when retrieving academic documents.
1	2	3	4	5	Using the **Enhanced Histogram** made it easier to retrieve academic documents.
1	2	3	4	5	I found the **Enhanced Histogram** useful for retrieving academic documents.

POST-TASK QUESTIONNAIRE (CONTINUE)

Please rank your library search interface preference (order from 1 to 2):
_____ Microsoft Academic Search
_____ Bow Tie Academic Search

Please make any other comments about the library search interfaces and/or the search tasks.

Bibliography

[1] C. Ahlberg and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 313–317, 1994.

[2] T. J. Anderson, A. Hussam, B. Plummer, and N. Jacobs. Pie charts for visualizing query term frequency in search results. In *Proceedings of the International Conference on Asian Digital Libraries*, pages 440–451, 2002.

[3] K. Andrews, V. Sabol, W. Lackner, C. Gütl, and J. Moser. Search result visualization with xFIND. In *Proceedings of the Workshop on User Interfaces to Data Intensive Systems*, pages 50–60, 2001.

[4] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.

[5] M. J. Bates. Where should the person stop and the information search interface start? *Information Processing and Management*, 26(5):575–591, 1990.

[6] R. Becker, S. Eick, and A. Wilks. Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics*, 1(1):16 –28, 1995.

[7] P. Bergstrom and D. C. Atkinson. Augmenting the exploration of digital libraries with Web-based visualizations. In *Proceedings of the International Conference on Digital Information Management*, pages 1–7, 2009.

[8] E. Bertini, T. Catarci, L. Di Bello, and S. Kimani. Visualization in digital libraries. In M. Hemmje, C. Niedere, and T. Risse, editors, *From Integrated Publication and Information Systems to Information and Knowledge Environments*, pages 183–196. Springer-Verlag, Berlin, Heidelberg, 2005.

[9] N. Bevan. Measuring usability as quality of use. *Software Quality Journal*, 4:115–130, 1995.

[10] J. Bhogal, A. Macfarlane, and P. Smith. A review of ontology based query expansion. *Information Processing and Management*, 43(4):866–886, 2007.

[11] K. Börner and C. Chen. Visual interfaces to digital libraries. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 425–425, 2002.

[12] K. Börner and C. Chen. Visual interfaces to digital libraries: Motivation, utilization, and socio-technical challenges. In *Visual Interfaces to Digital Libraries*, pages 1–12, 2002.

[13] K. Börner, Y. Feng, and T. McMahon. Collaborative visual interfaces to digital libraries. *In Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 279–280, 2002.

[14] M. Bostock, V. Ogievetsky, and J. Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301 –2309, 2011.

[15] J. Callan, J. Allan, C. L. A. Clarke, S. Dumais, D. A. Evans, M. Sanderson, and C. Zhai. Meeting of the MINDS: An information retrieval research agenda. *SIGIR Forum*, 41(2):25–34, 2007.

[16] R. G. Capra and G. Marchionini. The relation browser tool for faceted exploratory search. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, page 420, 2008.

[17] S. Card, J. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, San Francisco, CA, 1999.

[18] S. Carpendale. Evaluating information visualizations. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, pages 19–45. Springer, Berlin, Heidelberg, 2008.

[19] C. Carpineto and G. Romano. A survey of automatic query expansion in information retrieval. *ACM Computing Surveys*, 44(1):1–50, 2012.

[20] Carrot Search. `http://search.carrot2.org`, July 2012.

[21] H. Chen, A. Houston, R. Sewell, and B. Schatz. Internet browsing and searching: User evaluations of category map and concept space techniques. *Journal of the American Society for Information Science*, 49(7):582–603, 1998.

[22] M. Chen, M. Hearst, J. Hong, and J. Lin. Cha-Cha: A system for organizing Intranet search results. In *Proceedings of the Conference on USENIX Symposium on Internet Technologies and Systems*, pages 11-14, 1999.

[23] E.-I. Cho and S. H. Myaeng. Visualization of retrieval results using DART. In *Proceedings of RIAO (Recherche d'Information Assiste par Ordinateur-Computer Assisted Information Retreival)*, pages 1434–1439, 2000.

[24] E. Clarkson, K. Desai, and J. Foley. ResultMaps: Visualization for search interfaces. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1057–1064, 2009.

[25] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[26] C. Cook. *LibQUAL+® 2006 Survey.* Association of Research Libraries, 2006.

[27] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma. Query expansion by mining user logs. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):829–839, 2003.

[28] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319-340, 1989.

[29] DBLP. http://www.informatik.uni-trier.de/~ley/db/, August 2012.

[30] B. Dervin and P. Dewdney. Neutral questioning: A new approach to the reference interview. *Research Quarterly*, 25(4):506–513, 1986.

[31] A. Dix and G. Ellis. Starting simple: Adding value to static visualisation through simple interaction. In *Proceedings of the Conference on Advanced Visual Interfaces*, pages 124–134, 1998.

[32] M. Dörk, S. Carpendale, C. Collins, and C. Williamson. VisGets: Coordinated visualizations for Web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1205–12, 2008.

[33] D. Ellis. A behavioural model for information retrieval system design. *Journal of Information Science*, 15(4-5):237–247, 1989.

[34] N. Elmqvist and P. Tsigas. Causality visualization using animated growing polygons. In *IEEE Symposium on Information Visualization*, pages 189 –196, 2003.

[35] N. Elmqvist and P. Tsigas. CiteWiz: A tool for the visualization of scientific citation networks. *Information Visualization*, 6(3):215–232, 2007.

[36] D. Fallman. Design-oriented human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 225–232, 2003.

[37] S. Few. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, Oakland, CA, 2009.

[38] G. Furnas, T. Landauer, L. Gomez, and S. Dumais. The vocabulary problem in human- system communication. *Communications of the ACM*, 30(11):964–971, 1987.

[39] E. Garfield. *Citation Indexing-Its Theory and Application in Science, Technology, and Humanities.* John Wiley, New York, NY, 1979.

[40] M. Ghoniem, J.-D. Fekete, and P. Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization*, pages 17–24, 2004.

[41] L. E. Good, A. C. Popat, W. C. Janssen, and E. A. Bier. A fluid interface for personal digital libraries. In *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries*, pages 162–173, 2005.

[42] M. Hearst. Multi-paragraph segmentation of expository text. In *Proceedings of the Meeting of the Association for Computational Linguistics*, pages 9–16, 1994.

[43] M. Hearst. *Search User Interfaces.* Cambridge University Press, Cambridge, 2009.

[44] M. Hearst and C. Karadi. Cat-a-Cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 246–255, 1997.

[45] M. A. Hearst. TileBars: Visualization of term distribution information in full text information access. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 59–66. 1995.

[46] T. Heimonen and N. Jhaveri. Visualizing query occurrence in search result lists. In *Proceedings of the Conference on Information Visualization*, pages 877–882, 2005.

[47] A. R. Hendrickson, P. D. Massey, and T. P. Cronan. On the test-retest reliability of perceived usefulness and perceived ease of use scales. *MIS Quarterly*, 17(2):227 230, 1993.

[48] O. Hoeber. Exploring Web search results by visually specifying utility functions. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 650–654, 2007.

[49] O. Hoeber. Web information retrieval support systems: The future of Web search. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence - Workshops (International Workshop on Web Information Retrieval Support Systems)*, pages 29–32, 2008.

[50] O. Hoeber. User evaluation methods for visual Web search interfaces. In *Proceedings of the International Conference of Information Visualization*, pages 139–145, 2009.

[51] O. Hoeber. Human-centered Web search. In C. Jouis, I. Biskri, J.-G. Ganascia, and M. Roux, editors, *Next Generation Search Engines: Advanced Models for Information Retrieval*, pages 217-238. IGI Global, Hershey, PA, 2012.

[52] O. Hoeber and H. Liu. Comparing tag clouds, term histograms, and term lists for enhancing personalized Web search. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence - Workshops (International Workshop on Web Information Retrieval Support Systems)*, pages 309 –313, 2010.

[53] O. Hoeber and X. D. Yang. Interactive Web information retrieval using WordBars. In *Proceedings of the IEEE/WIC/ACM Conference on Web Intelligence*, pages 875–882, 2006.

[54] O. Hoeber and X. D. Yang. The visual exploration of Web search results using HotMap. In *Proceedings of the International Conference on Information Visualization*, pages 157–165, 2006.

[55] O. Hoeber and X. D. Yang. User-oriented evaluation methods for interactive Web search interfaces. In *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops (Intelligent Web Interaction Workshop)*, pages 239–243, 2007.

[56] O. Hoeber and X. D. Yang. Visual support for exploration within Web search results lists. In *Conference Compendium of the IEEE Information Visualization Conference*, pages 80–81, 2007.

[57] O. Hoeber and X. D. Yang. Evaluating WordBars in exploratory Web search scenarios. *Information Processing and Management*, 44(2):485–510, 2008.

[58] O. Hoeber and X. D. Yang. HotMap: Supporting visual explorations of Web search results. *Journal of the American Society for Information Science and Technology*, 60(1):90–110, 2009.

[59] O. Hoeber, X. D. Yang, and Y. Yao. Visualization support for interactive query refinement. In *Proceedings of the IEEE/WIC/ACM Conference on Web Intelligence*, pages 657 – 665, 2005.

[60] O. Hoeber, X. D. Yang, and Y. Yao. VisiQ: Supporting visual and interactive query refinement. *Web Intelligence and Agent Systems: An International Journal*, 5(3):311–329, 2007.

[61] K. Hornbæk. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2):79–102, 2006.

[62] IEEE. IEEE Xplore. `http://ieeexplore.ieee.org`, August 2012.

[63] B. J. Jansen and U. Pooch. A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology*, 52(3):235–246, 2001.

[64] B. J. Jansen and A. Spink. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management*, 42(1):248–263, 2006.

[65] B. Johnson and B. Shneiderman. Tree-Maps: A space-filling approach to the visualization of hierarchical information structures. In *Proceedings of the Conference on Visualization*, pages 284–291, 1991.

[66] H. Joho, C. Coverson, M. Sanderson, and M. Beaulieu. Hierarchical presentation of expansion terms. In *Proceedings of the ACM Symposium on Applied Computing*, pages 645 649, 2002.

[67] S. Jones, S. J. Cunningham, R. McNab, and S. Boddie. A transaction log analysis of a digital library. *International Journal on Digital Libraries*, 3(2):152–169, 2000.

[68] N. Kampanya, R. Shen, S. Kim, C. North, and E. A. Fox. Citivz: A visual user interface to the CITIDEL system. In *Proceedings of the European Conference on Digital Libraries*, pages 122–123, 2004.

[69] W. Katz. *Introduction to Reference Work: Reference Services and Reference Processes.* McGraw-Hill, New York, NY, 1978.

[70] D. Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1-2):1–224, 2009.

[71] T. Khazaei and O. Hoeber. Metadata visualization of scholarly search results: Supporting exploration and discovery. In *Proceedings of the International Conference on Knowledge Management and Knowledge Technologies*, pages 21:1–21:8, 2012.

[72] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[73] B. Kules and R. Capra. Designing exploratory search tasks for user studies of information seeking support systems. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 419–420, 2009.

[74] B. Kules, R. Capra, M. Banta, and T. Sierra. What do exploratory searchers look at in a faceted search interface? In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital libraries*, pages 313–322, 2009.

[75] B. Y.-L. Kuo, T. Hentrich, B. M. Good, and M. D. Wilkinson. Tag clouds for summarizing Web search results. In *Proceedings of the International Conference on World Wide Web*, pages 1203–1204, 2007.

[76] S. Lawrence, C. Lee Giles, and K. Bollacker. Digital libraries and autonomous citation indexing. *Journal of Computer*, 32(6):67–71, 1999.

[77] B. Lee, M. Czerwinski, G. Robertson, and B. B. Bederson. Understanding research trends in conferences using PaperLens. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, pages 1969–1972, 2005.

[78] X. Lin, D. Soergel, and G. Marchionini. A self-organizing semantic map for information retrieval. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 262–269, 1991.

[79] m. c. schraefel, M. Wilson, A. Russell, and D. Smith. mSpace: Improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, 49(4):47–49, 2006.

[80] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, 1986.

[81] J. D. Mackinlay, R. Rao, and S. K. Card. An organic user interface for searching citation links. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 67–73, 1995.

[82] M. Magennis and C. J. van Rijsbergen. The potential and actual effectiveness of interactive query expansion. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 324–332, 1997.

[83] M. Mahoui and S. J. Cunningham. Search behavior in a research-oriented digital library. In *Proceedings of the Conference on Research and Advanced Technology for Digital Libraries*, pages 13–24, 2001.

[84] G. Marchionini. Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[85] Microsoft. Microsoft Academic Search. `http://academic.research.microsoft.com/About/Help.htm`, June 2012.

[86] Microsoft. Microsoft Academic Search API User Manual. `http://academic.research.microsoft.com/about/MicrosoftAcademicSearchAPIUserManual.pdf`, June 2012.

[87] G. A. Miller. WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

[88] P. Ngok and Z. Gong. Log mining to support Web query expansions. In *Proceedings of the International Conference of Information and Automations*, pages 375–379, 2009.

[89] J. Nielsen. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 152–158, 1994.

[90] L. T. Nowell, R. K. France, D. Hix, L. S. Heath, and E. A. Fox. Visualizing search results: Some alternatives to query-document similarity. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 67–75, 1996.

[91] W. Ogden, M. Davis, and S. Rice. Document thumbnail visualization for rapid relevance judgments: When do they pay off. In *Proceedings of the Text Retrieval Conference*, pages 1–7, 1999.

[92] T. Paek, S. Dumais, and R. Logan. WaveLens: A new view onto Internet search results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 727–734, 2004.

[93] M. Patton. *Qualitative Research & Evaluation Methods*. SAGE Publications, Newbury Park, CA, 2001.

[94] G. Robertson, J. Mackinlay, and S. Card. Cone Trees: Animated 3D visualizations of hierarchical information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 189–194, 1991.

[95] Y. Rogers, H. Sharp, and J. Preece. *Interaction Design: Beyond Human-Computer Interaction.* Wiley, New York, NY, 2011.

[96] I. Ruthven. Re-examining the potential effectiveness of interactive query expansion. In *Proceedings of the ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 213–220, 2003.

[97] F. Sebastiani. Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47, 2002.

[98] B. Shackel. Usability - context, framework, definition, design and evaluation. In B. Shackel and S. J. Richardson, editors, *Human Factors for Informatics Usability*, pages 21–37. Cambridge University Press, Cambridge, 1991.

[99] A. Shiri. Metadata-enhanced visual interfaces to digital libraries. *Journal of Information Science*, 34(6):763–775, 2008.

[100] B. Shneiderman, D. Feldman, A. Rose, and X. F. Grau. Visualizing digital library search results with categorical and hierarchical axes. In *Proceedings of the ACM Conference on Digital Libraries*, pages 57–66, 2000.

[101] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz. Analysis of a very large Web search engine query log. *SIGIR Forum*, 33(1):6–12, 1999.

[102] H. Small. Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9):799–813, 1999.

[103] A. Spink. A user-centered approach to evaluating human interaction with Web search engines: An exploratory study. *Information Processing and Management*, 38(3):401 – 426, 2002.

[104] A. Spink, D. Wolfram, B. J. Jansen, and T. Saracevic. Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3):226–234, 2001.

[105] A. Spoerri. RankSpiral: Toward enhancing search results visualizations. In *Proccedings of the IEEE Symposium on Information Visualization*, pages 18–19, 2004.

[106] M. Steinbach, G. Karypis, and V. Kumar. A comparison of document clustering techniques. In *the proceedings of the KDD Workshop on Text Mining*, pages 109–111, 2000.

[107] J. Teevan, S. Dumais, and Z. Gutt. Challenges for supporting faceted search in large, heterogeneous corpora like the Web. In *Proceedings of the Human-Computer Information Retrieval Workshop*, pages 6–8, 2008.

[108] E. Tufte. *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, CT, 1983.

[109] E. Tufte. *Envisioning Information.* Graphics Press, Cheshire, CT, 1990.

[110] E. Tufte. *Visual Explanations: Images and Quantities, Evidence and Narrative.* Graphics Press, Cheshire, CT, 1997.

[111] D. Tunkelang. *Faceted Search.* Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publisher, San Rafael, CA, 2009.

[112] A. Veerasamy and N. Belkin. Evaluation of a tool for visualization of information retrieval results. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 85–92, 1996.

[113] F. Vigas and J. Donath. Social network visualization: Can we go beyond the graph. In *Proceedings of the Conference on Computer Supported Cooperative Work [Workshop on Social Networks]*, pages 6–10, 2004.

[114] J. Wang, A. Agrawal, A. Bazaza, S. Angle, E. A. Fox, and C. North. Enhancing the ENVISION interface for digital libraries. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 275–276, 2002.

[115] C. Ware. *Information Visualization: Perception for Design.* Morgan Kaufmann, San Francisco, CA, second edition, 2004.

[116] C. Ware. Visual queries: The foundation of visual thinking. In S.-O. Tergan and T. Keller, editors, *Knowledge and Information Visualization*, volume 3426, pages 121–144. Springer-Verlag, Berlin, Heidelberg, 2005.

[117] R. W. White, B. Kules, S. Drucker, and m. c. schraefel. Supporting exploratory search: Introduction to special section. *Communications of the ACM*, 49(4):36–39, 2006.

[118] R. W. White and R. A. Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009.

[119] A. Woodruff, A. Faulring, R. Rosenholtz, J. Morrsion, and P. Pirolli. Using thumbnails to search the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 198–205, 2001.

[120] J. Xu and W. B. Croft. Query expansion using local and global document analysis. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 4–11, 1996.

[121] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 401–408, 2003.

[122] J. S. Yi, Y. A. Kang, J. Stasko, and J. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–31, 2007.

[123] P. Zaphiris, K. Gill, T. H. Y. Ma, S. Wilson, and H. Petrie. Exploring the use of information visualization for digital libraries. *New Review of Information Networking*, 10(1):51–69, 2004.

[124] J. Zhang, T. Shearer, G. Marchionini, M. Efron, and J. Elsas. Relation Browser++: An information exploration and searching tool. In *Proceedings of the National Conference on Digital Government Research*, pages 109:1–109:2, 2004.