Techniques of Ontology and its Usage in Indian Languages - A Review

Rajveer Kaur
Research Scholar
Baddi University of Emerging Sciences and
Technology, Baddi, India

Saurabh Sharma
Assistant Professor
Baddi University of Emerging Sciences and
Technology, Baddi, India

ABSTRACT
Ontology is presently an emerging research topic in the field of artificial intelligence, semantic web, and natural language processing, software engineering, and information architecture etc. Manual Ontology building is essentially a time consuming and tedious task. From the last few decades, different ontology building approaches are being used to build ontology either semi-automatically or automatically. This paper review has been carried out/done on the usage of different ontology building approaches in natural language processing and their usage on Indian languages. It is hereby concluded from the review that in Indian languages usage of ontology is very less because of the challenges to build ontology for minority languages.

Keywords
Ontology Engineering, Multilinguistic Ontology, Information Retrieval, Ontology Approaches.

1. INTRODUCTION
 "Ontology" has originated from Philosophy branch but in the last decades, Ontology in information systems has become common in numerous other fields like Natural Language Processing, Database integration, Internet technologies, Artificial Intelligence, Multiagent Systems etc.

In recent years, to exploit the information for an economical and helpful management of knowledge, the employment or use of ontology in the area of Artificial Intelligence and Natural Language Processing has become a necessity. Informally Ontology describes the terms, concepts, classes and relationship between them. Formally Ontology is defined as: Gruber originally outlined ontology [1, 2] as an "explicit specification of a conceptualization". Borst modified Gruber ontology definition as a "formal specification of a shared conceptualization". This definition need that conceptualization ought to express a shared view instead of individual view and ought to be in formal (i.e. machine readable) format. Studer et al. combines these two ontology definitions and give new definition of ontology [3] as: "formal, explicit specification of a shared conceptualization".

Conceptualization [4] is a simple and abstract view of the world that represents the knowledge for any motive. Explicit implies that the variety of concepts used and therefore the constraints on their utilization are explicitly outlined. Formal make reference to the reality that ontology ought to be machine understandable and readable. Shared reveals the insight that an ontology captures accordant knowledge i.e. it's not individual personal information, however accepted by a group.

In alternative words, abstract illustration of the entities, events and their relationships is known as ontology. Abstraction and composition (part-of) are two main relationships used in ontology. The representational primitives of ontology include the classes, their properties, and relationships. They are employed to represent specific application domain knowledge. The structure of ontology defines the way of gathering and constructing components together to model a correct ontology.

2. LITERATURE REVIEW
2.1 Usage of Ontology in Various Domains
There are two models to represent the knowledge and information in computer systems. These models are Database models i.e. Relational databases and Ontology. For long, relational databases have been used to store and query information but Ontologies have appeared as an alternative to databases in applications that require a more 'improvised' relevance.

C. M. Cruz et al. [5] presents a review of relationship and differences between relational databases and ontologies and mechanisms used for communications. The identical reality is represented by Ontologies and databases, however relying on the matter to be solved or the application to be developed one or another technology presents benefits. Ontology provides a restriction-free framework even in the Web to represent a machine readable reality. This framework supposes an open world where information may be reused, explicitly outlined, and shared. Furthermore, information could even be swap and utilized to create queries and deductions. Thus, investigation of the resemblance and variation between databases and ontologies is developed and therefore the following results classified: ontologies are text and execution independent and run on a higher abstraction level. On the other hand, Data models are located at the lower abstraction level. Knowledge modeled by ontologies put together the specification of structure with real information i.e. instances. A database creates a transparent distinction between data and its schema. The drawbacks of the data model chosen initiate the quantity of loss of semantics within the knowledge representation. The identical drawback happens with ontologies: the semantics of ontologies depends on the language of representation. Therefore, the choice to decide one or other technology totally depends on needs of user's. If the knowledge to be represented must be shared on the web, ontology ought to apparently give a decent solution. Ontologies give a wonderful way to reality representation; however database is actually the better methodology for kept up such knowledge when this is of significant size.

N. M. Meenachi [6] presents a survey of ontology usage in various domains such as Aviation, Medicine, Agriculture, Nuclear weapons, Education, Computer science etc. This survey relies on the obtainable literature; solely on the basis of outlines of existing ontologies built in numerous domains. The findings show that ontology has been created in various domains like Agriculture, Biology, Chemistry, Computer Science, Education, Linguistics, Library, Human resources, Military, News, etc. which indicate that in medicine and education domain, a lot of in-depth work has been done and a

number of ontologies are developed like for medicine domain, Gene ontology, MGED ontology, cardiovascular medicine ontology etc.

For the education domain, efforts are towards developing ontologies that aid e-learning techniques along with formal teaching like Web Based Educational Systems (WBES), Topic Maps For E-Learning (TM4L) etc. Ontologies related to hardware and software product management systems, audio-video indexing and retrieval systems, software engineering life cycle ontology, neural networks and image processing have been developed in the Computer science domain. Enhanced ontology development would aid in the evolution of linguistics web that will result in complete sharing of information in a particular domain. It is however important to understand here that the ontology development is a continuous method and complete success can only be achieved by participation of domain specialists and users. It is prominent from the survey that very limited work has been done in other domains like Aviation, Human resources, nuclear weapons, news and transport.

M. Hazman et al. [7] presents a survey where ontology from different information sources i.e. semi-structured and unstructured data is constructed. Learning from unstructured data includes three approaches i.e. Natural Language Processing approach, Statistical approach and integrated approach. The first approach attempts to integrate NLP with statistical approach and gets concepts and their relations. The system uses the noun frequency count and noun phrases in documents that are retrieved from the web. The system uses shallow parser to extract noun phrases. Sanchez and Moreno build ontology from web documents using this approach. The second approach describes a system that is based on natural language processing. To find the relation between syntactic entities in natural language, this system uses dependency grammar and parsers. Sabou et al. and Cimiano et al. use this approach to construct ontology. Integrated approach includes different methods from various disciplines like information retrieval, lexical database, and machine learning methods. Ontology learning from semi-structured data includes two approaches namely traditional data mining approach and web content mining approach. Karoui et al. discovers ontology from web pages using traditional data mining approach. Bennacer and karoui uses this approach to retrieve conceptual knowledge from web pages. Davulcu et al. use web content mining approach to develop OntoMiner. OntoMiner is an automatic method for bootstrapping and populating specific domain ontologies.

N. A. Astrakhantsev and D. Yu. Turdakov [8] provides a survey of methods for Automatic Construction and Enrichment of Informal Ontologies. Informal ontologies are those ontologies that are built and maintained by user communities such as Internet Encyclopedias (e.g. Wikipedia). On the other hand, Formal ontologies are those ontologies which include classes, functions, axioms, concepts and relationship between concepts. Informal ontologies usage is based on two methods. The first method is to develop a formal ontology that relies on information from an informal ontology. The second method makes use of informal ontology as linked data where links shows relationship between concepts. During survey it is noted that Wikipedia (English version) has more than 4 million articles and more than 28 million pages. There is no generally accepted method of ontology building because there is no commonly accepted notion of ontology. L. Drumond and R. Girardi [9] proposed a separation called "ontology learning layer cake". In this the

list of separated subtasks is as follows: Terminology extraction for a specific domain, Recognition of synonymous terms or forms of the same word, Formation of concepts, Hierarchical organization of concepts, Extraction of relations, properties, or attributes of concepts together with the corresponding application domain, Identification of axioms. Terminology Extraction is the main sub task of Ontology construction. There are two methods of terminology extraction namely Linguistics method and Statistical methods. Linguistic methods include stop word removal, duplicate terms removal etc. The Statistical methods are based on the information about the distribution of words in large data set. The authors consider the ontology enrichment as an optimization problem by introducing a special function of similarity between concepts. Research in the field of enrichment of existing ontologies by terms and concepts of specific domains is much younger than in the field of the ontology construction "from scratch." It is important to note here that most systems for ontology enrichment are semiautomatic. Survey shows that in the case of specific ontologies, it is often impossible to improve accuracy at the expense of completeness because of insufficient amount of available text data.

S. Ramakrishnan et al. [10] presents a study on contributions of recently implemented ontology visualization tools in the development of cognitive support features. The domain taken to build ontology is HIV-AIDS domain. This ontology includes causes, symptoms of HIV and information of HIV-AIDS patients living in Tamilnadu area. The HIV-AIDS ontology data set includes 15 classes having 2,573 instances. First 13 classes are used to build is-a taxonomy tree and remaining 2 classes are used to employed multiple inheritance. There are several ontology visualization tools available that enhance user's cognitive support such as SOVA, Snow OWL, and OWL diff, OntoGraf, OWL2Query, and OWL viz, Top-Braid composer, Neon-toolkit, Swoop, DL Query etc. The experiment was conducted by three visualization tools (i.e. DL Query, OntoGraf and OWL2Query) to evaluate user's satisfaction on cognitive support features. Before evaluation, some preliminary tests were conducted to decide that which ontology visualization method should be used in the experiment. Features that were introduced to the users are layout editor, property editor, toolbar, variable editor, prefix editor, query graph, SPARQL query view, edge editor SPARQL-DL preview, result panel, property editor and Abox mode editor, Rbox mode editor and Tbox mode editor. The experiment built five queries to extract information from HIV-AIDS ontology. The result shows that users can solve each query faster when using OntoGraf as compare to two other tools such as OWL 2 Query and DL query.

When OntoGraf tool was used with comparison of OWL2 Query take 59.6% less time to solve the queries and had a 25.4% lower TLX. When OntoGraf tool was used with comparison of DL Query approximately 33.2% less time is spent by the users and had a 16.4% lower TLX. These comparisons conclude that OntoGraf is much better than OWL2Query and DL Query for HIV-AIDS ontology use case. The results of survey provide understanding of new features that enhance the cognitive support.

H. Alani et al. [11] proposed Automatic knowledge extraction from web documents that is based on ontology. The proposed system incorporates Artequakt project which automatically retrieve information from web related to artists, then create a knowledge base, and utilized this knowledge base to create

personal biographies of artists. The knowledge-extraction tool was linked with ontology by Artequakt project to attain knowledge support and way of extracting the information. The extraction tool seeks on-line documents and extracts the information that is equivalent to the specified classification structure. The extracted machine- readable knowledge is kept up automatically in a knowledge base (KB). Lexicon-based term expansion technique provides extended terminology of ontology. This terminology is utilized to upgrade the knowledge extraction. Artequakt joins together experience and proficiency from three different projects: First project is European Artiste project that construct art images distributed database. Second project is The Equator IRC that uses descriptive methods to model and present information. And last project is The AKT IRC project that analyze all facet of the knowledge life cycle.

Artists and paintings domain ontology is constructed during first phase of Artequakt. Many information extraction tools and methods are applied that model the ontology automatically with information extraction from web that is based on ontology's structure and lexicons of WordNet. The system keeps information in knowledge bases and finds duplications in stored information. Narrative construction tools were developed during second phase. Through an ontology server, these tools are used to query the knowledge bases, search and extract related text, and create a particular artist biography. In first experiment, Artequakt begins to search and retrieve information of 10 artists from the web. The knowledge extraction tool analyzes 100 web pages and has set to choose the 10 relevant pages for each artist. The extraction procedure recognized over 800 passages and thousands of entities and their relations. Most of the identified entities are duplicate entities. Entities and relations associate persons with different type of information like artists professional relations to other artists, family members, their home location and study information, and personal details like date of birth and date of death. The result shows that 280 unique person entities and 600 unique relations were remained after knowledge extraction and consolidation process.

H. Kong [12] proposed a design of domain specific automatic ontology building. There's the problem that the domain ontology is usually made freshly, although the same domain ontology was existed somewhere on the web. This problem is overcome by the proposed method. This method identifies the semantic similarity and relationship between two ontologies and then builds the ontology automatically by using the Ontomerge algorithm. Ontomerge algorithm integrates two or more existing ontologies and constructs the new ontology. The process of ontology building includes three steps. First is user access the system and enter the domain on which ontology will be build. System retrieved the domain from WordNet and automatically builds the frame ontology. Second step is: user enters the domain knowledge and this knowledge is added to the frame ontology by the system. User can edit the domain knowledge throughout the ontology building process. In Third step, OWL language is used by system to build the domain ontology. The wine ontology is compared through proposed system based on proposed method. The experimental result shows that wine ontology based on proposed method has less abundance content as compare to W3C's wine ontology. The efficiency of the system improves, if richer domain knowledge data provides by the domain expert.

Z. Sui et al. [13] proposed a method of building medical domain system in Chinese that is based on CBD (Concept

based description model). This method gives the answers of two questions. First question is: Is there any knowledge description model that is computer understandable existed on web? To answer this question, Concept based description model is used that represent the rich data in ontology that is computer understandable. Second question is: how to reduce domain expert work using NLP techniques? Two NLP techniques are used to give an answer to this question. First technique is Automatic term extraction and second one is integrated approach of relations extraction. When using first approach, the performance of proposed method based on CBD model is compare with C-value method. The result shows that proposed approach has high precision (47%), recall (44.2%) and f-measure (45.6) values. So it is concluded that proposed approach enhance the C-value approach performance. When second approach is used, www extraction method uses as baseline one and DIPRE (Dual interactive pattern expansion) method use as baseline two. Ten random concepts of medical domain ontology are taken to evaluate the integrated approach. The results shows that proposed method enhance the F-score by 2.32% from the baseline one and precision by 12.38% from the baseline two. So it is concluded from the results that proposed method is better method for building domain ontologies.

C. Hu et al. [14] proposed a technique of building domain ontology automatically from relational databases. Material science domain data is taken as relational databases to build domain ontology. Three mapping rules are employed to extract the information from schema of relational databases to components of ontology and initial ontology is built. The instances of ontology are generated from database using extraction of metadata and migration of data. An experimental prototype is implemented through this proposed method. The result of experiment shows that proposed method is effective in ontology building from relational databases.

Q. Guo and M. Zhang [15] present a question answering system that is based on semantic web and ontology. In proposed system, NL parser is utilized to implement the human-machine interface of Chinese natural language and ontology is utilized to represent the domain knowledge. The system consists of three models: Ontology and Semantic Web based question's semantic comprehension model, question similarity match model based on FAQ, and automatic answer fetching model based on document warehouse. The Question answering system is evaluated to check the feasibility, effectiveness and extensibility of the system. The testing set includes 100 sample questions from a large number of questions asked by students. These questions are protected with the covering scale i.e. no answer situation. The evaluation shows that it is feasible to build the semantic QA system based on semantic web and ontology. Users can get the direct answers which show system works perfectly. In the situation of no answer, the proposed system takes large amount of proportion that shows the good extensibility of the system. The answer can be easily and quickly send into the knowledge ontology without any conflict with the ontology semantic relations. The results show that recall of the question answering system is 91.95% when using ontology and semantic web.

D. Huang and J. Gao [16] introduce a method of building domain ontology based on SECI model. First, authors analyze and identify the issues that come during creation of domain ontology and try to overcome these issues by integrating previous ontology building method with SECI model. SECI model is a spiral model that consists of four processes such as

Socialization, Externalization, Combination, and Internalization. Ontology of East China Sea fishery domain is built by using EHSO (ECSI Hierarchy-Spiral-Ontology) method. This method includes three layers namely Analysis layer, knowledge accumulate layer, and development and application layer. SECI model happens in and between every layer of the EHSO technique. Analysis layer analyze the domain knowledge that may be used for constructing domain ontology. Knowledge accumulate layer gather the required knowledge of specific domain. Application and development layer is employed for ontology building and evolution. This technique of building ontology is concentrated on runtime conversion of implicit knowledge to explicit knowledge, and on raise of knowledge transfer. Therefore this technique builds the ontology more perfectly.

S. Janzen and W. Maass [17] proposed Natural Language Processing system based on ontology for In-store shopping. Within In-store shopping situations, formal ontologies are used to represent the web based product information and dialogue system. A domain specific dialogue system i.e. Conversational Recommendation Agent (CoRA) system is presented. The architecture of CoRA includes four basic components such as middleware layer, mobile client, knowledge base and dialogue web service. The in-store sale is improved by using this dialogue system and product information because customer can access product details easily through this system. To evaluate the CoRA dialogue system, a user study is conducted. This study includes 16 subjects taken from German University. The dialogue system is presented to each and every subject. The study shows that it is usable and useful to use CoRA dialogue concept for acquiring in-store product related information.

J. Mustafa et al. [18] proposed ontology based semantic information retrieval framework that enhances the search results precision. The framework includes various modules such as source model, Crawler, semantic matcher, query ranker and reformulator. Thematic similarity procedure is used for retrieval of information that interprets the concepts and their relationships by matching RDF triples (i.e. subject, predicate, and object) as a substitute of matching keywords. RDF triples is also used to define domain ontology. To store the metadata triples, Oracle network data Model is used. To evaluate the proposed semantic information retrieval system, corpus of 300 manually created documents is taken. Each document holds 30 to 50 RDF triples. Research publication ontology is employed in experiment as a knowledge base of particular domain that includes 451 concepts. A set of 93 queries having different number of triples are accomplished with proposed system. Simple semantic information retrieval system is compared with proposed semantic retrieval system that considers context. The upper bound and lower bound correlation between precision and recall is 0.82 and 0.39 respectively of simple semantic retrieval system and 0.79 and 0.49 respectively of proposed semantic retrieval system. The proposed system shows strong relationship between precision and recall coefficients. Therefore proposed semantic retrieval system based on ontology gives better performance than simple semantic retrieval system.

J. Uma Maheswari and G. R. Karpagam [19] proposed a Conceptual framework for retrieval of information that is based on ontology. The conceptual framework includes various phases such as parsing of query, stemming of words, ranking, matching of different ontologies, weight assignment, and retrieval of information. E-Governance in Academics is a case study that is implemented using this framework. The proposed system gives the better approach for seeking the data on the web. It finds the documents that are relevant to the user query by using techniques called word stemming, ontology matching, weight assignment, rank calculation etc. In the ontology matching stage an improved matching algorithm is used to enhance the document retrieval relevance. The highest score amalgamation is ranked as the best and the equivalent information is retrieved.

S. Chaware and S. Rao [20] proposed an Integrated Approach for Ontology Development methodology. The proposed system is related to Shopping Mall Domain. This approach integrates the existing methodologies in order to remove the limitations and therefore enhance the overall methodology to ontology development. An integrated approach includes different modules to build ontology. These modules are motivating user scenarios module, formulation of formal/informal questions and answers module, extraction of terms and constraints module and build ontology module. Ontology concepts and relations are developed dynamically. This approach uses top-down structure to build ontology and performs faster as compare to previous approaches.

Y. C. Kiong et al. [21] developed an ontology system for Health domain. The Health ontology system provides a method to combine data related to healthcare institutes in the configuration of a shared progressive ontology. Progressive ontology integrates domain relevant ontologies. The development of health ontology system includes three software tools such as Ontology generator, distiller, and accumulator. Ontology generator builds the health ontology from selected information source (i.e. MS Access database). The reverse process of Ontology generator is known as Ontology distiller. Ontology distiller creates the database by retrieving the concepts and their relations from ontology OWL file. The resulted database is either MS Access database or SQL server database. Ontology accumulator creates the rules list for integration of similar ontologies. It includes classes list of cumulative ontology. This Integrated system of Health Ontology make a start for merging existing data with ontologies which will be helpful for constructing semantic agents for healthcare domain.

T. Sun et al. [22] proposed a method for semi-automatic ontology building based on two-layer ontology model. Two layer ontology model (TMOL) includes basic model layer and derivative layer. Basic model layer consists of classes and their attributes, individuals, and base type. Derivative layer (i.e. inheritance layer) derived the ontology from basic model layer and produced hierarchy. The method of building ontology is divided into five phases that are based on TLOM. These five phases are: determine the scope of ontology and consider reusing existing ontology, modularize the requirements of ontology and assign tasks to project members, prepare list of extracted terms and their relations in the field, representation of ontology, and evaluation and evolution of ontology. The proposed method based on two layer ontology model is useful for fast creation of ontology and ontology evolution.

D. tang and Y. Xu [23] proposed framework of ontology for plain text evolution. Plain text evolution is a process of exploring change in the semantics of data. The plain text semantics (i.e. intrinsic and extension semantics) of genome in biology are represented as ontology called DATAGENOME. DATAGENOME ontology includes CONNOTATION ontology and CONTEXT ontology to represent the semantics of plain text. Intrinsic semantics are represented by CONNOTATION ontology and extension

semantics are represented by CONTEXT ontologies. The construction of CONNOTATION ontology relies on plain text chapter structure. A tree is constructed with instances of CONNOTATION ontology according to the relationships between instances. CONTEXT ontology includes eight interlinked facet specifically when, what, where, how, who, which and why that represents the practical environment which results in the knowledge existence, together with the occurrence that affected knowledge, event occurrence time, event location, event reason, input and output etc. Intrinsic and extension semantics correlation are shown by mapping between CONNOTATION and CONTEXT ontology tree. Data genome must grow to indicate the text evolution when semantics of the data are modifying with change of context.

B. Abdelbasset et al. [24] proposed an agent based approach for creating ontology from text. A multi-agent system is implemented by using this approach. The proposed methodology includes various steps such as specify the domain of ontology, construct glossary of terms, construct hierarchies of concepts, construct diagram of binary relation, build a dictionary, and describe the components of ontology. The three levels are used to organize the proposed methodology. These levels are Syntactic level, Semantic level, and Formal level. Syntactic relations between entities are extracted by the syntactic level using natural language techniques. At semantic level, combination of two algorithms (i.e. Formal concept analysis (FCA) and Relational concept analysis (RCA)) are used to move from syntactic level to semantic level. FCA combines the set of objects that exhibit same set of attributes. The set of binary relations between the formal concepts are extracted by RCA; an extension of FCA. At formal level, ontology is build using OWL language. The testing has done on sample text corpus of animal domain. It is noted from the results that proposed ontology building process in not fully automatic because users get involved to assign names to concepts and their relations.

Athira P. M. et al. [25] proposed the architecture of natural language question answering systems based on domain ontology. The proposed framework includes four modules that are best suited for upgrading current QA capabilities with the ability of processing complex queries. The four modules were Question Processing, Document Extraction, Document Processing and Answer Processing. User queries were identified and categorized by the question processing module. Extracted documents were processed by document processing module and responses were generated by answer processing module. To reformulate queries and identifying the relations, domain ontology was used. The goal of the proposed system is to generate small and specific answers to the queries asked in natural language in a particular domain. The experiment demonstrates that proposed framework can filter semantically matching sentences and their relations effectively and efficiently and rank the answer higher in the result list.

Z. Sellami et al. [26] developed an interactive tool DYNAMO-MAS: a Multiagent system for ontology growth from text. This tool is based on an adaptive multi-agent system (AMAS). In the protégé ontology editor, DYNAMO-MAS is performed as a plug-in. DYNAMO-MAS extend TextViz plug-in of protégé. Three corpora are used as knowledge sources from three different domains in this project. First corpora is research in archeology about techniques that is in French language, second corpora is automotive diagnosis (automotive components, symptoms, engine failures, etc.) that is also in French language and third

one is software bug reports that is written in English language. Three different ontologies and associated corpora have been taken to conduct the test using DYNAMO-MAS.

- Artal ontology is software bug reports ontology in English language, which includes corpus of 287 documents having 887 terms and 582 concepts.

- Arkeotek ontology is archeology traditional techniques ontology in French language that include corpus of 299 documents having 733 terms and 380 concepts.

- Actia ontology is automotive diagnosis ontology in French language that include corpus of 710 documents having 579 terms and 330 concepts.

Quality and performance Evaluation are two scenarios to evaluate DYNAMO-MAS. When 21 new documents were added to the Artal ontology data set, the DYNAMO-MAS attained *67%* and 56% relevant term proposals and concept proposals respectively. DYNAMO-MAS have ability to recommended terms and concepts that have not been manually identified and analyzed (i.e. 12 terms and 9 concepts). When 12 new documents are added to the Arkeotek ontology dataset, DYNAMO-MAS attained *68.75%* and *59.26%* of relevant term proposals and concept proposals respectively. DYNAMO-MAS recommended 18 terms and 14 concepts that are not manually identified. When 50 new documents are added to the Actia ontology dataset, DYNAMO-MAS attained only *16.98 %* of relevant term proposals and *22.22%* of relevant concept proposals. DYNAMO-MAS recommended 5 terms and 5 concepts that are not manually identified. It is noted from the results that poorer results are obtained with the Actia ontology dataset because the new documents added to the corpora contain very little new knowledge. To sum up, DYNAMO-MAS runs ontology evolution in different domains and it supports two languages (English and French).

Q. Yu et al. [27] proposed a method for automatic evaluation of domain-specific ontology that is based upon ontology co-relation network. The evaluation of ontology has done by mapping ontologies and then ranks the ontologies. Graph based ranking algorithm (i.e. Google's PageRank) is used to rank the ontology. The data set includes 150 ontologies and their metadata, count of mapping to one another. Ontologies that have self-mapping are removed from the data set and 145 ontologies are remained in data set. To calculate MGP for 145 ontologies, Matlab 7.0 is used. MGP gives the rank score list of all ontologies. It is noted from the list that SNOMED clinical terms ontology has highest mapping count (i.e. 620720). If mapping count of ontology is highest then ontology gets highest rank score. Therefore SNOMED clinical terms ontology has highest rank score (i.e. 11,094). Therefore proposed method tells that which ontology is the best ontology.

S. Chimalakonda and K. V. Nori [28] present an educational modeling framework called IDont for Adult literacy instructional design that is based on ontology. The architecture of the IDont is based upon existing ontologies. Existing ontologies are extended for instructional design of adult literacy. IDont framework includes several modular ontologies such as Context ontology, Goals ontology, Process ontology, Content ontology, Evaluation ontology, Environment ontology, and Domain ontology. The main aim of this proposed framework is to use ontology to present separation of concerns approach to consistently extract totally

different aspects of ID. Research work is ongoing to coming up with a technology platform which will support the IDont framework.

J. Monti et al. [29] proposed an approach based on ontology for cross-lingual information retrieval. The methodology is based upon the Lexicon-grammar (LG). LG is a technique to formalize natural language, automatic text analysis and parsing. Two lexical resources are built for processing such as Electronic dictionary and Grammar that is helpful to build information retrieval systems. The goal of proposed framework is to map data and metadata that utilized the semantic information and morph syntactic stored in the electronic dictionaries. This proposed ontology based cross language information retrieval system improves precision and recall of Information extraction and information retrieval systems. During first phase, translation and transformation rules are applied. The information is extracted from user query and match with existing ontological domain concepts. The feasibility test is conducted to check that proposed cross language information retrieval is feasible or not. Archaeological domain is used to test the feasibility of proposed approach. Translation experiment has been carried out from Italian to English language. The result shows that proposed architecture ensures coverage of big data and maintains semantic relations which unite the different languages.

I. Mathur et al. [30] presents the development of ontology for communicable disease domain. The methodology to build ontology includes various phases such as Specification, Conceptualization, Creation of Instances, and Ontology Visualization. A reasoner based on DL query is employed that check the classes, its properties and relationships to verify the constructed ontology. For using description based queries, FaCT++ reasoner in protégé is configured. When class name of ontology is given to the reasoner, it extracts the query in terms of classes, super classes, its properties, relations, domain, and range. The result of the reasoner shows that ontology is constructed perfectly without any defect.

G. Gottlob et al. [31] developed a Query answering in Datalog+/- ontologies under probabilistic uncertainty that is based on Markov logic networks. Datalog+/- is a part of ontology languages that is helpful for representing and reasoning over light-weight ontologies. Datalog+/- plays main role in context of question answering and information extraction for the semantic web. Apart from dealing with probabilistic uncertainty, they also proposed ways to handle uncertainty arising from inconsistency in probabilistic Datalog+/− ontologies. Many algorithms were also given, namely a basic approach, an anytime one, and one based on heuristics, which are sure to return sound results. The main focus is on scalable algorithms for answering threshold queries to probabilistic guarded Datalog+/− ontologies. Result shows that answering threshold queries to probabilistic guarded Datalog+/− ontologies may be done in polynomial time.

T. Lukasiewicz et al. [32] proposed Preference-Based Query Answering in Probabilistic Datalog+/− Ontologies. The proposed system is an extension of the Datalog+/− part of ontology languages that uses two models: one representing user preferences and one representing the (probabilistic) uncertainty with which inferences are made. It is assumed that preference is given to the more probable answers. Four Specific operators are proposed to find preferences i.e. ask how to rank answers to a user's queries. The semantics and Computational properties of operators were analyzed. Authors

developed four specific preference combination algorithms, i.e. egalitarian algorithms (ComPrefsGen and Com-PrefsRank) and user-biased algorithms (ComPrefsPT and ComPrefsSort). These algorithms fulfill the prerequisite of a preference combination operator and semantic properties.

A basic algorithm was also developed for answering k-rank queries based on the iteration of the well-known skyline answers to a question, along with proofs of correctness and running time. It shows that, under certain conditions, k-rank queries can be answered in polynomial time in the data complexity. A real word application with preference based query answering is used to evaluate preferences based probabilistic Datalog+/-. Internet Movie Database (IMDB) is used as real world application. To test the algorithms performance, experiments were conducted with two parameters such as SPO size and the number of answers (*k*). SPO (Strict partial order) is represented as a directed graph. The size of SPO is given by the number of nodes and edges of graph. When numbers of nodes were varies from 1,000 to 1,300 with seed density of 0.01% and k fixed at 200, ComPrefsGen algorithm is performed better as compare to other algorithms. On the other hand, ComPrefsSort gives worst performance if the number of nodes increases, as it include sorting each iterated skyline, that was quite large for these relatively low-density graphs. When number of nodes in SPO was fixed at 2,000, edge density varies from 0.2% to 0.4% and k fixed at 200, in this case, ComPrefsSort performance is same as ComPrefsRank and Com- PrefsPT, because in this case the number of nodes to sort is fixed. When k varies, number of nodes was fixed at 2,000 and density at 20%, in this case ComPrefs- Rank and ComPrefsSort are influenced by the density of edges in the graph.

N. B. Mustapha et al. [33] proposed query based technique of contextual ontology module learning by using web snippets. The objective of this approach is to create automatic and domain independent light-weight ontology modules to information retrieval applications that do not use predefined ontologies. English language is used as target language in proposed system. The input sources include disambiguated queries (using user feedback) and web snippets provided by available search engines. The proposed system consists of five main stages. These stages are Query analysis and topic signature construction, Web snippets extraction, Generation of dependencies' list, weighted concept map construction, Concept map enrichment with taxonomic and non-taxonomic relationships, and Ontology module formalization using Typed Attributed Graph (TAG). To identify question concepts and to import related concept and relations, a top-level ontology (i.e. Mesh, Sensus) is used. Top-level ontologies describe general conceptualization which is equivalent over all knowledge domains. Ontology module is built on the obtained concept map. Typed attributed graphs are used to represent the ontology module. Two case studies were considered to check the proposed approach feasibility. These case studies are the context of a domain-independent search in web e.g. Google, Yahoo, and Bing search and the context of a search associated with particular domain e.g. PubMed (Medical digital library ontology).

The results shown that precision average is below 50%, it implies that half of the new concepts searched by the proposed system appeared in MeSH. Thus, it can be argued that PubMed is able to detect automatically new domain concepts from the Web, so as to help the users search in digital libraries. The analysis of question answering system

has discovered that the accuracy of the results was considerably improved by using modular ontologies.

S. K. Dwivedi and A. Kumar [34] present the university ontology construction for aSPOCMS (An Agent-based Semantic Web for Paperless Office Content Management System). To develop the ontology, generalized structure and workflow processes of 10 Indian universities, 30 schools and 150 departments have been taken. This ontology is the combination of four ontologies: Top level ontology, task ontology, Domain Ontology, and application ontology. The aSPOCMS use Owl, RDF, and RDFS for acknowledgement of metadata and reasoning rule. The aSPOCMS system architecture includes four modules namely access control, reasoners, communicator, and knowledge manager. DL query is employed to extract information from constructed university ontology. Some queries are entered to test the ontology of university by employing inference engine known as RacerPro. The result of proposed method shows that system is working perfectly.

J. N. K. Liu et al. [35] proposed graph model of domain ontology and its usage in classification of Chinese text. The proposed system includes an ontology leaning approach to build semiautomatic Graphical Ontology Structure known as Ontology Graph. The ontology graph is a representation of ontology and knowledge conceptualization model. After generation of Domain Ontology Graphs (DOG), Document Ontology Graphs (DocOG) are generated and Ontology Graph based Text Classification has performed. Main Objective of Document Ontology Graph is to carry out Ontology Text Classification by matching the single Document Ontology (DocOG) with Domain Ontology.

The corpus includes 2,814 Chinese documents of ten different domains having 965 Chinese characters in each document. Out of these documents corpus (D1) 70% documents are for learning process and 30% documents are for testing process. Different domains are Education, Arts and Entertainment, Politics, Environment, Economics, Military, Traffic, Health and Medical, Computer and Information Technology, and Sports. To model the knowledge, thirteen different kinds of Domain Ontology Graph are constructed from each domain having number of terms 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200 and 300 respectively.

Innovative Ontology Extraction System called KnowledgeSeeker has been developed which automatically generate the domain ontology with Domain Ontology Graph. Knowledge seeker executes four modules: Ontology graph modeling, Ontology learning, Ontology generation and Ontology querying. The ontology graph modeling includes several conceptual units. These Conceptual Units are Terms, Concepts, Concept Cluster, and Ontology graph. These conceptual units are necessary for extraction in Ontology learning process. The five sub processes of learning process are Term Extraction, Term-to-Class relationship mapping, Term-to-Term relationship mapping, clustering of concepts and generation of ontology graph. Term Extraction includes elimination of special symbols, duplicate terms, and stop words etc. Chinese Dictionary and Initial Word list are also used to extract terms from input text. Chi-Square base statistical learning method is applied to find Term-to-Class relationship mapping and Term-to-Term relationship mapping. The unclassified Document corpus (D2) of 57,208 documents having average of 2,349 Chinese characters in each document is used for Term-to-Term relationship mapping and clustering of concepts. Two experiments were conducted; first is performance testing of ontology graph-

based approach and second is evaluating the optimal parameters (i.e., k and h) of domain specific ontology graph for best categorization result. To conduct first experiment three different approaches are used. These approaches are Traditional tf-idf approach, term dependency approach, and ontology graph based approach. The second experiment use equivalent techniques that are used in first experiment. It varies the number of terms utilized in each approach and the parameter h for ontology graph based approach. The results shown that ontology graph-based text categorization approach achieved a higher accuracy (i.e. 89.2%) at term size 80 in text classification as compare to other two approaches. It is concluded from the experiment that ontology graph based approach is useful and efficient for text classification systems implementation.

A. Rainer and W. Matthias [36] proposed an approach for social media analysis that is based on Ontology. The goal of this approach is to improve the data quality of ontology for social data analysis by automated knowledge extraction from previous enterprise application systems. Social data is extracted from social media channels for business applications. To analyze the qualitative data/information from social media, ontology building approaches are linked with text mining tools. Ontology is used as dictionary which includes keywords and concepts relevant to social media for social media analysis. The proposed framework of social media analysis based on ontology includes four modules. These modules are Business databases, Ontology engineering, Text mining framework, and social media. Business database module includes information about company and their products and processes. This information is utilized to build ontology. Ontology engineering process automatically builds the ontology by using existing business databases. Database interfaces (i.e. JDBC or ODBC) are used to retrieve the data from business databases. This module is also used to map database structures into ontology by applying algorithms i.e. tables are mapped into classes, attributes are mapped into properties etc. The resulting ontology is use as dictionary for text mining tools. Text mining framework analyzes the unstructured text (e.g. text files, books etc.) from social media channels (e.g. Consumer forums and social networking sites). The sources of information that are untapped are represented mostly by Social media module. Social data gets accessible by utilizing already existing interfaces from different platforms. These platforms are Twitter or Facebook application programming interface (API). The mapping algorithms (such as DB2OWL, R2O, RDB2ONT, and RONTO) were analyzed based upon four aspects: Interoperability, Data source integration, ontology creation, and Individualization of the mapping result. It is shown from the result that DB2OWL automatically map data (i.e. unstructured and structured) and structure of relational databases to ontologies.

J. A. Bateman et al. [37] present a space linguistic ontology for natural language processing. Ontological organization is developed by covering language involved with space, space actions and spatial relationships. The common behavior and characteristics of ontology are modeled to point out a way to extend linguistic ontology particularly with space. The linguistic ontology is domain independent; it is an extension of original GUM ontology. It is midway between linguistic expressions and their interpretation. Three aspects are included to model the spatial language: first one is address the linguistic phenomena of spatial language use, Second is address the systematization of spatial language interpretations and last one is address the computational representation of processing schemes for natural language that include space.

Gum ontology includes total 270 concepts and 110 relations, out of which 98 concepts (36%) and 28 relations (25%) are spatial related. The DL expressivity of GUM ontology is little affected by spatial extension. The design of the system includes semantic representations that are produces by entering voice or written input data to module of analysis. The domain knowledge is represented by problem solving component using ontologies and Pellet and Racer description logic reasoners are used to perform the task. Seven spatial language corpuses that consists English and German natural languages are employed to build the linguistic ontology. Three spatial language corpora were chosen to analyze English: the HCRC Map Task, the Trains 93 Dialogues, and the IBL (Instruction Based Learning) Corpus. Four spatial corpora has chosen for analyze German language: Aibo2, the Bielefeld SFB Corpus, Stuga (unpublished route description corpus) and Rolland. These corpora are of spatial domain that includes task-oriented dialogues; they consist of large number of terms related to space together with isolate expressions, descriptions of spatial activity, and descriptions of route. Linguistic ontology of space is evaluated to scope visualization, uniformity, ability to create predictions, and suitability. It is shown from the result that the linguistic ontology gives the term semantics, however once this is attached to contextualized interpretations of space additional variety of information could naturally be involved.

2.2 Use of Ontologies in Indian Languages

In this section, we present the literature review of usage of ontologies in Indian languages.

S. Saraswathi, S. M. Asma et al. [38] proposes information retrieval system for English and Tamil language i.e. bilingual system on the festival domain. System development process includes four modules: User Interface, Keyword Extraction, Information Retrieval and Information Extraction. To build the ontology of festival domain, Indian language Tamil is used by system apart from English. Domain ontological tree is constructed to answer retrieval in the same language as of input question entered by user. Naïve algorithm was used for information retrieval. To determine the keywords from the given question, a Part-of-speech (POS) tagger is employed. Total 200 documents were collected for both English and Tamil language. For English language, Documents related to Christmas, New Year, Easter and Good Friday were collected and For Tamil language, documents related to Diwali, Pongal, Navarathiri, Kaarthigai Deepam and Vinayagar chathurthi were collected. The System was tested for information retrieved from the Ontology that is based upon using the keywords alone and with the use of concept words. It was noted from the results that the relevance of documents was improved by 40% for English and 60% for Tamil language.

S. Ningombam et al. [39] describes procedure of creating Manipuri-English dictionary by ontology implementation that is machine readable dictionary. Words can be stored for ontology by using OWL and Protégé 3.2 beta software. Manipuri-English lexicon is implemented to build the ontology. The taxonomic structure of the ontology includes two components namely Part-Of-Speech (POS) and Word structure. All Manipuri vocabularies are represented through POS. POS has two main classes such as noun and verb. Word structure also has two classes such as simple word and compound word. Development of ontology based Manipuri-English lexicon includes two stages: Look up and data entry. Look up module provides the user interface and helpful for search information about vocabularies of Manipuri language. It extracts the lexical information for the given Manipuri word

such as synonyms, antonyms, pronunciations and structure of word. Data entry module is accessible only to Administrator. Vocabularies are inserted in ontology by using data entry module. The interface of the system was developed by using Java and Java expert system shell.

S. Chaware and S. Rao [40] proposed an approach of semantic matching based on ontology for Hindi and Marathi language inference system. Knowledge is represented by using ontology. The proposed approach design includes seven modules such as user interface in local language, query formation module, inference module, translation module, ontology module, synsets module, and output display module. Grocery shop domain is used to build ontology. User enters the input query through local language user interface either in Hindi or in Marathi language. To retrieve English ontology terms, bi-lingual dictionary is used to translate keywords of local language into English. Each keyword is searched in database; if match found then that keyword is interpreted as ontology term. In this way, all terms are found in database and ontology is constructed using these terms and their relationships found in database. Now ontology has been enlarged into inference system. The terms of ontology fed to the inference module to model information as a solution to the inference question. The information retrieval from ontology and results of inference system has been shown to user by using display module. For Hindi and Marathi user, the precision and recall values are different for single query. For single query, the precision is 20% for Hindi user and 18% for Marathi user and Recall is 65% for Hindi user and 90% for Marathi users. When numbers of queries are more than one then precision is 45% and recall is 95%, which is same for both Hindi and Marathi users. It is shown from the results that system based on proposed approach will give inference accurately for any type of query.

A. K. Durga et al. [41] presents a technique of ontology based classification for Telgu documents. This model relies on classic Vector- space model and analyzes the terms on the sentences and documents level. The Ontology based classification is performed in three steps: Ontology Creation, Calculating Relevance Score, and Text Classification. Telgu text documents are tokenized and words are sent to the morphological analyzer to find root words. Then domain keywords are identified and classified using text classifier. The model effectively differentiates non-important terms and terms which hold the concepts that represent the sentence meaning. The testing has been done on 400 Telgu documents of two domains namely Rajakeeyalu (Politics), Aatalu (Sports). Out of these 400 documents, 80% data is Training data and 20% data is testing data. The constraints of keyword based search are overcomes by ontology based text categorization. This technique proved that efficiency of Text categorization of term is better when ontology model is used for Telgu documents instead of Conventional model.

K. R. Ananthapadmanaban et al. [42] designed subtle User profile ontology for Tourism domain in Tamilnadu. By identifying interest of user by using user profile ontology, process of finding for the perfect tourism package improves. This ontology extracts the user personal data by requesting to user to register and fill personal details in few forms. Personal detail includes number of persons in travel, no. of persons related to each other and their age group is assessed by asking no. of kids, adults, teenagers and persons aged more than sixty traveling with him/her. The area of interest of users in touristy like historical places, Beaches, Hill Stations, Zoological Parks, Botanical Gardens etc., is additionally collected. User'

leisure activities like Swimming, Cycling, Golf etc., is additionally known through an interface. The user is additionally requested to provide trip date based on the data regarding persons with the user, their area of interest perfect vacation destination to the user will be known. Inferences will be created based on the user profile and touristy ontology. The results are generated with higher level of accuracy and suitable tourist location of user interest in Tamilnadu area can be generated.

B. Bhatt et al. [43] presents a k-partite graph learning algorithmic program for Indian languages that extract ontology from unstructured text. By using English and Hindi text corpora, ontology of Health domain is built. The English corpus includes 15589 sentences and 16498 terms while a Hindi corpus includes 16002 sentences and 14794 terms. Based on information content of the terms, algorithm divides the initial set of terms into completely different partitions and then builds ontology by detecting kind of relation between terms in numerous partitions (i.e. Top, Mid, Bottom, Top-mid, Mid-bottom, and Top-bottom). Two experiments were conducted to evaluate the performance of the proposed system. First experiment was Layer wise evidence detection and second was Comparison with Gold standard. First experiment shows that the relationship between concepts is identified more often in top partition using WordNet but proofs from lexico-syntactic patterns are more frequent in bottom partition. This is accordant with given hypothesis that top level concepts are general concepts and definitions of these concepts may be found in WordNet. In second experiment, the resulting ontology is compared with manually constructed ontology to evaluate the quality. The lexical precision and recall is calculated for both cases: with partition and without partition. The results shown that precision is higher for partition algorithm that means proposed approach improves the precision while not sacrificing F-score. The proposed approach not simply reduces the quantity of computation needed for ontology construction however additionally provides a further level of term filtering. It additionally avoids false association between the terms.

S. Saraswathi et al. [44] proposed Semi-Automatic ontology primarily based Bilingual Information retrieval System in Pilgrimage domain in South India. For the input query and document retrieval, Tamil and English language has been used. Modules of proposed system are Building of Semiautomatic Ontological tree, Design and Implementation of Semiautomatic Ontological tree. Semi-automatic ontology tree was built using XML and has four levels. These levels include State, District, Pilgrimage place and their attributes. Designing User Interface in Java programming language is the first phase of proposed system implementation. User can enter the query through user interface. The sub modules of User interface Design include Language Selection, Input Query, Keyword Extraction, Keyword Extraction and Translation, and Information Retrieval. The user can enter the query in either Tamil or English and the resultant document will be displayed in input query language. When user has entered the query, the proposed system searches the keywords and tries to retrieve the relevant concept words for the given keywords from the existing Ontology. If ontology contains relevant information then concept words would be utilized to retrieve the information from the Ontology. If information for the pilgrimage domain is not present in ontology then information will be extracted from search engines (e.g. Google, yahoo etc.) and updated in the Ontology. Two parameters (i.e. precision and recall) have been taken to evaluate the performance of the system. Different types of input queries

were entered by user and their resultant documents are analyzed using manually created Ontology and Semi-Automatic Ontology in the Pilgrimage domain. Five types of queries were examined namely. Type A query includes details about the pilgrimage place in a particular district. Type B includes details about the deity worshipped in any pilgrimage place. Type C includes questions related to the location of a pilgrimage. Type D includes pilgrimage place where a particular deity is worshipped, and Type E query includes Bus/Train/Air facility available to reach a pilgrimage place.

Input to the implemented system includes more than 20 questions for each type and the obtained results were analyzed. The results are shown that for the Type A, B, C and D queries, the Precision is high when using semi-automatic ontology instead of manually created Ontology. The results of the proposed system are same as that of manually constructed Ontology system for the Type E queries. The Recall is high in proposed system for type A, B, D and E queries and for Type C queries, Recall factor is same for both systems. This is because the location of a particular pilgrimage place does not change. It is concluded that the overall performance of the system is better when Semi-Automatic Ontology is used in Bilingual Information Retrieval system.

Nidhi and V. Gupta [45] planned and enforced two new algorithms (Ontology based classification and Hybrid Approach) for classification of Punjabi documents as previously no alternative Punjabi document classifier is obtainable in the world. Text classification is enforced in three phases: Pre-processing phase, Feature Extraction phase and Processing Phase. Pre-processing phase includes stop words removal, stemming, punctuations marks removal, and special symbol removal. Feature extraction phase includes statistical approach and linguistics approach. Statistical approach is used to extract the relevant features, e.g. words less than threshold value 2 are not considered as features. For linguistic approach, gazetteer list was prepared for Punjabi language. Processing phase includes text classification techniques to classify Punjabi text documents such as Centroid based classification, Bayesian classification and new Hybrid approach which is combination of Naïve Bayes and Ontology based classification. The classification corpus includes 180 Punjabi text documents and 45 files are used as training data that contains 3313 words. The corpus contains only Sports related documents that are taken from the Punjabi News Websites. Results conclude that Ontology based classification (85%) and Hybrid classification (85%) provides higher results as compared to Naïve Bayes classification (64%) and Centroid based classification (71%) algorithms for Punjabi documents.

D. Mavaluru [46] proposed Hybrid model for Telgu English cross language information retrieval system, that allows users to merges the retrieval model and the translation model. This approach includes four modules: query preprocessing, information retrieval from internet, Re-ranking and content presentation. The user entered the query in local language (i.e. Tamil). Query processor finds the query in ontology and identified meaningful terms that are related to user query. If terms not found in ontology then terms are searched on web and sends related terms to the user. Then re-ranking algorithm is used to re-rank the related terms results. Query is translated to the target language (i.e. English) by using bi-lingual ontology. The system was evaluated over two versions. First one is using the Ontology and second is using the Regular Lexicon. Java Programming Language is employed to implement the general system. Testing has done to check system accuracy. The proposed system results are compared

with Google results. Testing has been done on around 300 queries corresponding to various scenarios. It is concludes from the results that Ontology and rule based cross language information retrieval (CLIR) performs much better than regular CLIR model.

S. M. Chawre [47] proposed an ontology approach for cross-language information retrieval. The proposed approach create ontology from relational databases with query formation rules and translation/transliteration rules, which may be used for information retrieval from one language to a different i.e. Hindi/Marathi to English. Different modules of proposed system are User Interface module, Parsing module, Question/Answer module, Stemmer module, Translator/Transliteration module, Query module, and Database module. The ontology is build dynamically as per user's requirements, which can provide overall knowledge domain to the user. Ontology of Grocery shop domain is built. The entered local language keywords are stemmed to eliminate stop words before searching in dictionary. The translation/transliteration module is used to translate the entered local language keyword into English. For keyword translation, Hindi to English and Marathi to English machine-readable dictionaries were used. If the keyword is found in dictionary, then it acquire the terms required to create ontology from the database. In some cases, transliteration will be needed to translate local language string into English and vise-versa. The result shows the whole, simple and easy approach of creating ontology from databases.

P. Talita et al. [48] presents the challenges that are arising during construction of ontology for Minority languages. Several minority languages are morphological rich languages. Due to lack of sufficient resources, ontology building for minority language is a challenging task. Iban (i.e. Indigenous language) is considered as main language to build ontology for agriculture domain and to search the problems in building ontology. This ontology will be helpful for developing WordNet database of Iban language. The challenges in creating domain ontology for indigenous languages are: insufficient language semantics, Ambiguity in the related terms, different dialects of language, time bound and cost to build ontology. It is noted that apart from these challenges, there is one another problem to create ontology that is lack of expertise and knowledge.

3. CONCLUSION

In this paper, we present a review on usage of techniques of ontology and their usage in Indian languages. It is concluded from the review that very little work has been done for developing ontology in Indian languages. The reason for this can be attributed to the fact that the, number of challenges for the construction of ontology for minority languages are, many and varied. Another reason is the lack of knowledge about different Indian languages. There is lot of work related to ontology that will be done in future for different Indian minority languages as the literature or data sets, in these languages will grow and become standardised as well.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] Gruber, T. R. "Toward principles for the design of ontologies used for knowledge sharing," In International Journal Human-Computer Studies 43, Substantial revision of paper presented at the International Workshop on Formal Ontology, Padova, Italy, pp. 907-928, March, 1993.

[2] Chandrasekaran, B., John, R., and Banjamins, V. R. "What Are Ontologies And Why We Need Them," *IEEE Intelligent Systems*, pp. 20-26, Feb. 1999.

[3] Staab, S., and Studer, R. *Handbook on Ontologies,* International handbooks on Information Systems, 2nd ed., London, New York: Springer Dordrecht Heidelberg, 2009.

[4] Bermeja, J. "A Simplified Guide to Create an Ontology," *The Autonomous Systems Laboratory, Madrid, Universidad Politecnica de Madris*, 2007.

[5] Cruz, C. M., Blanco, I. J., and Vila, M. A. "Ontologies versus relational databases: are they so different? A comparison," *Artificial Intelligence Review, Springer, Netherlands*, vol. 38, pp. 271-290, Dec. 2012.

[6] Meenachi, N. M., and Baba, M. S. "A survey on usage of ontology in different domains," *International Journal of Applied Information Systems, IJAIS,* vol. 4, pp. 46-55, Sep. 2012.

[7] Hazman, M., Beltagy, S. R. E., and Rafea, A. "A survey on ontology learning approaches," *International Journal of Computer Applications, IJCA,* vol. 22, pp. 36-43, May 2011.

[8] Astrakhantsev, N. A., and Turdakov, D. Y. "Automatic construction and enrichment of Informal Ontologies," *Programming and computer software,* vol. 39, pp. 34-42, 2013.

[9] Drumond, L., and Girardi, R. "A Survey of Ontology Learning Procedures," In *Proceedings of the 3rd Workshop on Ontologies and Their Applications*, 2008, pp. 1-12.

[10] Ramakrishnan, S., and Vijayan, A. "A study on development of cognitive support features in recent ontology visualization tools," *Artificial Intelligence Review, Springer, Netherlands,* vol. 41, pp. 595-623, April 2014.

[11] Alani, H. et al. "Automatic ontology based knowledge extraction from web documents," *IEEE Intelligent Systems,* vol. 18, pp. 14-21, Feb. 2003.

[12] Kong, H., Hwang, M., and Kim, P. "Design of the automatic ontology building system about the specific domain knowledge," In *Proceedings of 8th International Conference on Advanced Communication Technology, ICACT, IEEE ,* 2006, pp. 1405- 1408.

[13] Sui, Z., Zhao, J., Kang, W., and Zhao, Q. "The building of a CBD-based domain ontology in Chinese," In *Proceedings of IEEE/WIC/ACM International conference on Web Intelligence and Intelligent Agent Technology, WI-IAT, 2008,* p. 303-306.

[14] Hu, C., Li, H., Zhang, X., and Hao, C. "Research and implementation of domain-specific ontology building from relational database," In *Proceedings of third China Grid Annual Conference, IEEE,* 2008, p. 289-293.

[15] Guo, Q., and Zhang, M., *Question answering system based on ontology and semantic web,* Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2008, vol. 5009.

[16] Huang, D., and Gao, J. "Technology of building domain ontology based on SECI," In *Proceedings of World Congress on Software Engineering, WCSE, IEEE,* 2009, pp. 519-523.

[17] Janzen, S., and Maass, W. "Ontology-based Natural Language Processing for In-store shopping situations," In *Proceedings of IEEE International Conference on Semantic Computing, ICSC,* 2009, pp. 361-366.

[18] Mustafa, J., Khan, S., and Latif, K. "Ontology based semantic information retrieval," In *Proceedings of 4th International IEEE Conference on Intelligent Systems,* 2008, pp. 22-14 – 22-19.

[19] Maheswari, J. U., and Karpagam, G. R. "A conceptual framework for ontology based information retrieval," *International Journal of Engineering Science and Technology,* vol. 2, pp. 5679-5688, 2010.

[20] Chaware, S., and Rao, S. "Integrated approach to ontology development methodology with case study," International *Journal of Database management Systems, IJDMS,* vol. 2, pp. 13-19, Aug. 2010.

[21] Kiong, Y. C., Palaniappan, S., and Yahaya, N. A. "Health ontology system," In *Proceedings of 7th International Conference on Information technology in Asia, CITA, IEEE,* 2011, pp. 1-4.

[22] Sun, T., Yang, F., Zhang, D., and Yang, L. "Ontology building based on two-layer ontology model," In *Proceedings of International Conference on Industrial Control and Electronics Engineering, ICICEE, IEEE,* 2012, pp. 1492-1494.

[23] Tang, D., and Xu, Y. "An ontology framework for evolution of plain text," In *Proceedings of 3rd International Conference on System Science Engineering Design and Manufacturing Information, IEEE,* 2012, pp. 5-8.

[24] Abdelbasset, B., Okba, Pr. K., and Sofiane, M. "Agent-based approach for building ontology from text," In *Proceedings of IEEE International Conference on Computer Applications Technology*, 2013, pp. 1-6.

[25] Athira, P. M., Sreeja, M., and Reghuraj, P. C. "Architecture of an ontology-based domain specific natural language question answering system," *International Journal of Web & Semantic technology, IJWesT,* vol. 4, pp. 31-39, Oct. 2013.

[26] Sellami, Z., Camps, V., and Gilles, N. A. "DYNMO-MAS: a multi-agent system for ontology evolution from text," *Journal on Data Semantics, Springer-Verlag, Berlin, Heidelberg,* vol. 2, pp. 145-161, June 2013.

[27] Yu, Q., Jiang, C., and Jianxin, W. "Automatic evaluation of domain specific ontology based on ontology co-relation network," In *Proceedings of 5th International Conference on Measuring Technology and Mechatronics Automation, IEEE,* 2012, pp. 317-322.

[28] Chimalakonda, S., and Nori, K. V. "IDont: An ontology based educational modeling framework for instructional design," In *Proceedings of IEEE 13th International Conference on Advanced Learning Technologies,* 2013, pp. 253-255.

[29] Monti, J., Monteleone, M., Buono, M. P. D., and Marano, F. "Natural Language Processing and Big data - An ontology-based approach for cross-lingual information retrieval," In *Proceedings of IEEE International Conference of Social Computing,* 2013, pp. 725-731.

[30] Mathur, I., Darbari, H., and Joshi, N. "Domain ontology development for communicable diseases," *CS & IT-CSCP,* vol. 3, pp. 350-360, 2013.

[31] Gottlob, G., Lukasiewicz, T., Martinez, M. V., and Simari, G. I. "Query answering under probabilistic uncertainty in Datalog+/- ontologies," *Annals of Mathematics and Artificial Intelligence, Springer, Dordrecht,* vol. 69, pp. 37-72, March 2013.

[32] Lukasiewicz, T., Martinez, M. V., Simari, G. I., and Marciuska, O. T. "Preference-based query answering in probabilistic Datalog+/- ontologies," *Journal on Data Semantics, Springer, Berlin, Heidelberg,* vol. 3, May 2014.

[33] Mustapha, N. B., Aufaure, M. A., Zghl, H. B., and Ghezala, H. B. "Query-driven approach of contextual ontology module learning using web snippets," *Journal of Intelligent Information Systems, Springer, New York,* July 2013.

[34] Dwivedi, S. K., and Kumar, A. "Development of University ontology for aSPOCMS," *Journal of Emerging Technologies in Web Intelligence,* vol. 5, pp. 213-221, Aug. 2013.

[35] Liu, J. N. K., He, Y., Lin, E. H. Y., and Wang, X. "Domain ontology graph model and its application in Chinese text classification," *Neural Computing & Applications, Springer, London,* vol. 24 , pp. 779-798, March 2014.

[36] Rainer, A., and Matthias, W. "Towards an ontology-based approach for Social Media analysis," In *Proceedings of Twenty Second European Conference on Information Systems, Tel Aviv,* 2014, pp. 1-10.

[37] Bateman, J. A., Hois, J., Ross, R., and Tenbrink, T. "A linguistic ontology of space for natural language processing," *Artificial Intelligence, Elsevier,* vol. 174, pp. 1027-1071, Sep. 2010.

[38] Saraswathi, S., Siddhiqaa, A. M., Kalaimagal, K., and Kalaiyarasi, M. "BiLingual information retrieval system for English and Tamil," *Journal of Computing,* vol. 2, pp. 85-89, April 2010.

[39] Ningombam, S., Meitei, S. P., and Purkayastha, B. S. "Building Manipuri-English machine readable dictionary by implementing ontology," *International Journal of Engineering Science and Technology, IJEST,* vol. 3, pp. 7682-7689, Oct. 2011.

[40] Chaware, S., and Rao, S. "Ontology supported inference system for Hindi and Punjabi," In *Proceedings of IEEE International Conference on Technology Enhanced Education,* 2012, pp. 1-6.

[41] Durga, A. K., and Govardhan, A. "Ontology based text Categorization- Telgu documents," *International Journal*

of Scientific & Engineering Research, vol. 2, pp. 1-4 , Sep. 2011.

[42] Ananthapadmanaban, K. R, and Srivatsa, S. K. "Personalization of user profile: creating user profile ontology for Tamilnadu Tourism," *International Journal of Computer Applications, IJCA,* vol. 23, pp. 42-47, June 2011.

[43] Bhatt, B., and Bhattacharyya, P. "Domain specific ontology extractor for Indian languages," In *Proceedings of 10th Workshop on Asian Language Resources, COLING, Mumbai,* 2012, pp. 75-84

[44] Saraswathi, S., Jemibha, P., Sugandhi, M., Mathimozhi, M., Sophia, A. L., and Nagarathinam, A. "Semi-automatic ontology based bilingual information retrieval system (Pilgrimage tourism in South India)," *International Journal of Intelligent Systems and Applications, MECS,* vol. 4, pp. 48-55, April 2012.

[45] Nidhi and Gupta, V. "Domain based classification of Punjabi text documents using ontology and hybrid based approach," in *Proc. of 3rd Workshop on South and Southeast Asian Natural Language Processing, SANLP, COLING, Mumbai,* 2012, pp. 109-122.

[46] Mavaluru, D., and Shriram, R. "Telgu English cross language information retrieval: A case study," In *Proceedings of National Conference on Recent Trends in Web Technologies, Rtwt,* 2013, pp. 78-*83.*

[47] Chaware, S. M., and Rao, S. "Ontology approach for cross language information retrieval," *International Journal of Computer Technology and Applications,* vol. 2, pp. 379-384, 2011.

[48] Talita, P., Yeo, A. W., and Kulathuramaiyer, N. "Challenges in building domain ontology for minority languages," in *Proc. of IEEE International Conference on Computer Applications and Industrial Electronics, 2010, pp. 574-578.*

6. AUTHOR'S PROFILE

Rajveer Kaur received her B.Tech degree in Computer Science and Engineering from GGS College of Modern Technology, Kharar, Punjab, India in 2011. Now she is a research scholar (M.Tech) at Computer Science and Engineering Department in Baddi University of Emerging Science and Technology, Baddi, HP, India. Her interests involve natural language processing (NLP), text mining and ontology engineering.

Er Saurabh Sharma is an Assistant Professor in Computer Science and Engineering Department at Baddi University of Emerging Science and Technology, Baddi, HP. He is GATE qualified and has done M.E in Computer Science and Engineering from University Institute of Engineering and Technology, Panjab University, Chandigarh in 2012. He has done B.Tech in Computer Science and Engineering from Swami Devi Dayal Institute of Engineering and Technology, Barwala in 2007. His research area is the field of computational linguistics with specific focus to text mining. His research work has been published in reputed International Journals.