International Journal of Information Movement Vol.3 Issue VI (October 2018)

IFSIJ Impact Factor : 1.575 Website: www.ijim.in ISSN: 2456-0553 (online) Pages 20-24

A NOVEL ONTOLOGICAL DOMAIN MODEL FOR INFORMATION RETRIEVAL WITHIN DIGITAL LIBRARY

B.Samatha
Librarian,
New Science PG College,
Warangal, Telangana.
Email- samathart@gmail.com

Abstract: A digital library is a type of information retrieval (IR) system. The existing information retrieval models generally have problems on keyword-searching. A proposed model to solve the problem by using concept-based approach (ontology) and metadata case base. This model consists of identifying domain concepts in user's query and applying expansion to them. This Model aims at contributing to an improved relevance of results retrieved from digital libraries by proposing a conceptual query expansion for intelligent concept-based retrieval. We need to import the concept of ontology, making use of its advantage of abundant semantics and standard concept. Domain specific ontology can be used to improve information retrieval from traditional level based on keyword to the lay based on knowledge (or concept) and change the process of retrieval from traditional keyword matching to semantics matching. One approach is query expansion techniques using domain ontology and the other would be introducing a case based similarity measure for metadata information retrieval using Case Based Reasoning (CBR) approach. Results show improvements over classic method, query expansion using general purpose ontology and a number of other approaches.

Keywords : Digital library, Domain ontology, Information access, Intelligent information retrieval, Query expansion

1.0 Introduction:

A digital library (DL) is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers [1]. The digital content may be stored locally, or accessed remotely via computer networks. Many digital libraries have evolved from traditional libraries and concentrated on making their information sources available to a wider audience. Today, many companies maintain their own digital libraries, and research and development for digital libraries now includes processing, dissemination, storage, search and analysis of all types of digital information. In contrast to physical libraries, digital libraries enable concurrent access at any time without physical boundaries. As such, digital libraries can be regarded as indispensable tools for today's knowledge workers. Digital libraries have always been an appealing playground for innovative computer science solutions. So they became a prominent research area.

In this paper, we focus on digital library within the efficient information retrieval using domain ontology as a controlled vocabulary to expand the input query string. Nowadays, user faces problems of management and sharing of huge amount of documents saved in the DLs. The work proposes methodology and technological framework allowing the user to be provided with a set of relevant documents based on semantic retrieval. Typically, information is retrieved by matching terms in documents with those of a query. The traditional solution employs keyword based search.

To help end users efficiently retrieve documents relevant to their information needs, this system provides concept-based query expansion and traditional statistical information retrieval algorithms that has given such good results in the IR field. To guarantee delivery of minimal irrelevant information (high precision) while insuring relevant information is not overlooked (high recall), the process of intelligent retrieval system based on the ontology is particularly presented. An increasing number of recent information retrieval systems make use of ontologies to help the users clarify their information needs and come up with semantic representations of documents.

An ontology is a collection of concepts and their interrelationships, which provide an abstract view of an application domain. With regard to converting words to meaning the key issue is to identify appropriate concepts that both describe and identify documents, as well as language employed in user requests. The use of ontology to overcome the limitations of keyword-based search has been put forward as one of the motivations of the Semantic Web since its emergence in the late 90's. While there have been contributions in this direction in the last few years, most achievements so far either make partial use of the full expressive power of an ontology-based knowledge representation, or are based on boolean retrieval models, and therefore lack an appropriate ranking model needed for scaling up to massive information sources.

2.0 Review of The Intelligent Information Retrieval System

Aim at the problem of poor retrieval quality in digital library, the advantage and correlative application of the ontology in digital library's semantics retrieval fields was introduced. And contributing to an improved relevance of results retrieved from digital libraries by proposing a conceptual framework for semantic retrieval. Semantics retrieval technology would improve retrieval quality extremely, and would be the preferred method to solving the lack of semantic relation in traditional retrieval technology. The work proposes methodology and technological framework allowing the user to be provided with a set of relevant documents based on semantic retrieval and case-based metadata. The user is able to enter natural language queries which, in turn, are analyzed. The conceptual representation of the query is matched against the database of conceptual representations to select the closest match. It allows the user to start the search with a relevant document or a natural language or Boolean query. It allows the user to browse related documents once a relevant document is found.

In [2], it described geographical information retrieval with ontology of place that may be used to derive semantic distance measures for use in geographically-referenced information retrieval. The proposed ontology was characterised by a mix of qualitative and quantitative spatial data including topological relations and sparse coordinate data representing the spatial footprints of places. Places were classified according to their geographical categories and were linked to instances of non-geographical phenomena classified by conceptual hierarchies. An hierarchical distance measure is combined with Euclidean distance between place centroids to create a hybrid spatial distance measure.

In the approach [3], a query enrichment approach that used contextually enriched ontologies was proposed to bring the queries closer to the user's preferences and the characteristics of the document collection. The idea is to associate every concept (classes and instances) of the ontology with a feature vector (fv) to tailor these concepts to the specific document collection and terminology used. The structure of the ontology was taken into account during the construction of the feature vectors. The ontology and its associated feature vectors were later used for post-processing of the results provided by the search engine.

In [4], it reported on how ontologies developed in the EU Semantic Web project SPIRIT were used to support retrieval of documents that were considered to be spatially relevant to users' queries. The query expansion techniques presented in this paper were based on both a domain and a geographical ontology. An overview of ongoing research was presented in [5] based on the use of a concept network as the knowledge base for inducing a query expansion based on the concepts deduced from the original query terms. In this system, the quality of this conceptual query expansion depended on the quality of the concept network. Query terms were matched to those contained in the concept network, from which concepts were deduced and additional query terms were selected.

Although most concept-based IR systems used the WordNet as controlled vocabulary to expand query [4], [5] and [6], our proposed approach combined the advantages of concept-based approach with the benefits of statistical approaches based on IR techniques in this paper. Domain specific ontology is used as controlled vocabulary for query expansion. And the basic assumption is that a user composing a search query simultaneously is describing a problem he or she seeks to solve. The case based reasoning component handles an information retrieval request as a description of a problem being part of a case. A good solution for such a case would be a good search result, i.e. a set of links to relevant information with respect to the search query. For this model, a case base has to be created to represent document information (metadata). The system can prove how this approach enables various benefits for intelligent query processing and expansion. The system architecture is shown in figure 1.

The retrieval method which is used by traditional DL based on keyword, it is too unilaterally concerning research of arithmetic to ignore consequence of semantics and mining of semantics of keyword itself. Under the bag of words model, if a relevant document does not contain the terms that are in the query, then that document will not be retrieved. Query expansion is the process of augmenting the user's query with additional terms in order to improve results in computer science.

3.0 A Proposed Ontological Module Model Architecture:

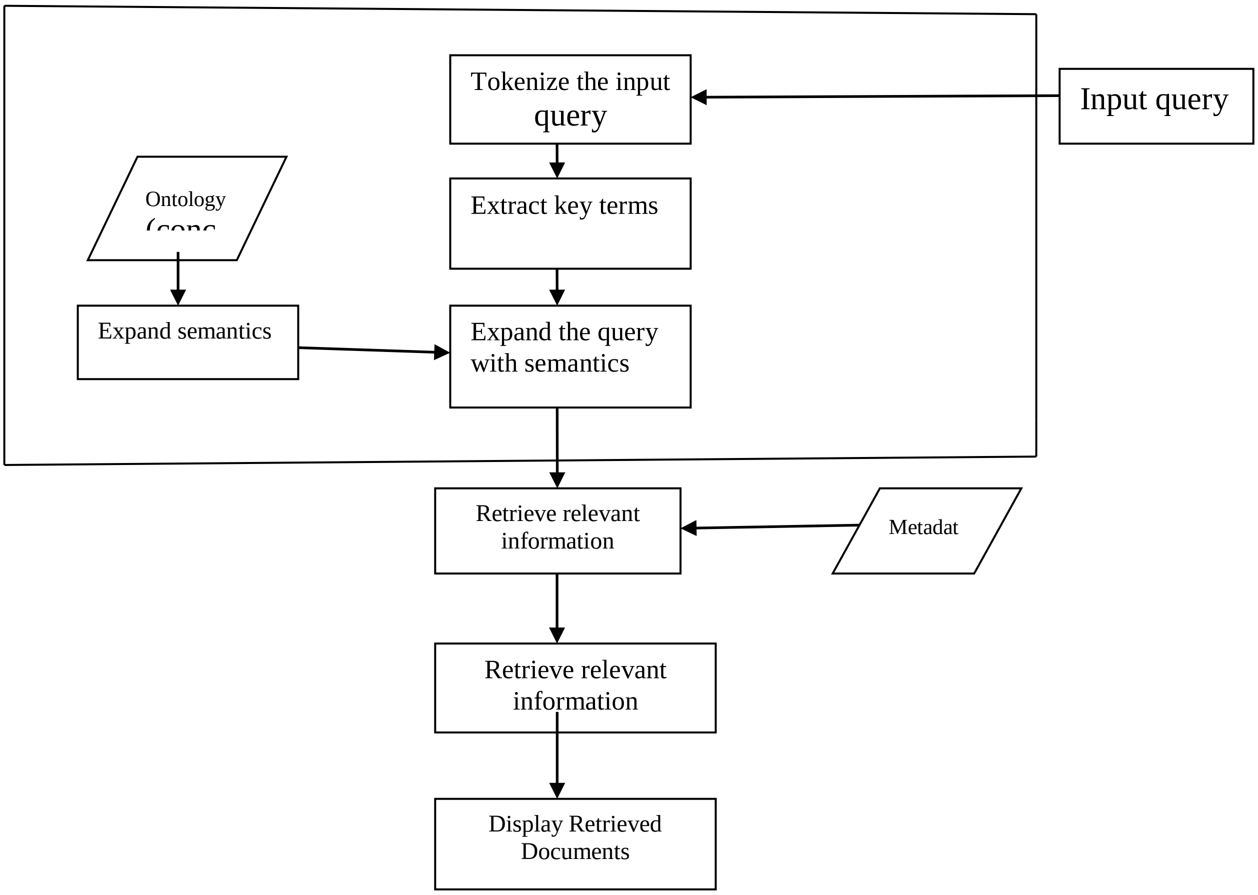
Tokenize the input query
Input query
Ontology (conc
Extract key terms
Expand semantics
Expand the query with semantics
Retrieve relevant information
Metadat
Retrieve relevant information
Display Retrieved Documents

Figure.1 The architecture of the proposed system

The main problem with traditional IR systems is that they typically retrieve information without an explicitly defined domain of interest to their user. Consequently, the system presents a lot of information that is of no relevance to the user. The research presented in this paper examines how ontologies can be efficiently utilized for traditional vector-space IR systems. The ontologies are adapted to the document space within multi-disciplinary domains where different terminology is used. The objective is to enhance the user-experience by improvement of search result quality for large-scale search systems.

During searching and retrieval process, a novel and promising approach is concept-based search [7], [8], [9]. Ontology-based approach to IR is presented. With this approach, the burden of knowing how the documents are written is taken off the user and hence the user can focus on searching on a conceptual level instead. One problem with this approach is to find good concepts. Domain ontology is useful for query expansion by proliferating the input words with the relevant domain concepts. The system is based on a domain concepts representation schema in form ofontology. With the use of ontology, concepts and relations representing concepts about a particular document in domain specific terms are built.

A query expansion method was described in [10] which based on the expansion of geographical terms by means of WordNet synonyms and meronyms. This method was used for the participation to the GeoCLEF 2005 English monolingual task, while using the well-known Lucene search engine for indexing and retrieval. The obtained results show that the proposed method was not suitable for the GeoCLEF track, while WordNet can be used in a more effective way during the indexing phase, by adding synonyms and holonyms to the index terms. There are two key problems in using an ontology-based model: one is the extraction of the semantic concepts from the keywords and the other is the document indexing. With regard to the first problem, the key issue is to identify appropriate concepts that describe and identify documents on the one hand, and on the other, the language employed in user requests. In this it is important to make sure that the irrelevant concepts will not be associated and matched, and that relevant concepts will not be discarded.

4.0 Semantic Analysis Component:

Semantic analysis reasoning is the key of implementation semantics retrieval function. It is just that analyzing the semantics of search terms which is submitted by users. Expanding the classification structure of words semantics then retrieving accordingly data to user interface. We need to identify concepts in information resources (Computer Science documents) and user queries. We need to do conceptual matching between extracted concepts. At this stage it is easy to find exact concept matching but the important part is to match remaining relevant concepts with the help of knowledge repository that is used. The knowledge repository gives information about concepts and their relationships with other concepts. So this stage requires a knowledge repository that does not miss any concepts and any relationships in the application domain
Firstly, it needs to tokenize the user input query. And then the key domain terms form the tokenized words are extracted. And the only domain terms are expanded with the relevant concepts from the ontology. In this case, an important novelty is that prunes irrelevant concepts and allows relevant ones to associate with documents and participate in query generation. These processes are automatically carried out that is without any user intervention or feedback. This mechanism generates queries with appropriate and relevant concepts terms through knowledge encoded in ontology form. In this component, query expansion that is the process of supplementing additional terms or phrases to the original query plays as an important role in order to improve the retrieval performance. There are three different ways of expanding the query: Manual, Interactive and Automatic. Manual and Interactive query expansion requires users involvement. Automatic query expansion is the process of supplementing additional terms or phrases to the original query to improve the retrieval performance without user's intervention. Sometime user may not be able to provide sufficient information for query expansion, therefore query expansion methods are needed which do not require user's involvement.

5.0 Conclusion:

Nowadays, semantics retrieval technology based on ontology has been the popular research direction. It brings hope to solve problems of lack semantics correlativity in traditional retrieval technology. On exploring the idea of using the concepts in ontology to improve search results. In the proposed approach, the query terms are used to match conceptual terms in the ontology. The ontology concepts are adapted to the domain terminology. Our query expansion method was applied to the digital library format and then digital libraries enable concurrent access at any time without physical boundaries. As such, digital libraries can be regarded as indispensable tools for today's knowledge workers. Digital libraries have always been an appealing playground for innovative computer science solutions. So they became a prominent research area.

6.0 References:

[1].Greenstein, Daniel I., Thorin, Suzanne Elizabeth. The Digital Library: A Biography. Digital Library Federation (2002) ISBN 1933645180. Accessed June 25, 2007

[2]. Geographical Information Retrieval with Ontologies of Place

[3]. Tomassen, S.L., Gulla, J.A., Strasunskas, D.: Document Space Adapted Ontology: Application in Query Enrichment. 11th International Conference on Applications of Natural Language to Information Systems. Springer, Klagenfurt, Austria (2006)

[4]. Gaihua Fu , Christopher B. Jones , Alia I. Abdelmoty.: Ontology-based Spatial Query Expansion in Information Retrieval. Lecture Notes in Computer Science, Volume 3761, Meaningful Internet Systems: ODBASE: OTM Confederated International Conferences, Vol.

[5]. F. A. Grootjen , Th. P. Van Der Weide.: Conceptual Query Expansion. Data & Knowledge Engineering, Volume 56, (2004) 174-193

[6]. Davide Buscaldi, Paolo Rosso and Emilio Sanchis Arnal.: Using the WordNet Ontology in the GeoCLEF Geographical Information Retrieval Task. Lecture Notes in Computer Science in Accessing Multilingual Information Repositories.

[7]. Grootjen, F.A., van der Weide, T.P.: Conceptual query expansion. Data & Knowledge Engineering .
[8]. Qiu, Y., Frei, H.-P.:Concept based query expansion. Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval. ACM Press, Pittsburgh, Pennsylvania, USA (1993) 160-169

[9]. Chang, Y., Ounis, I., Kim, M.: Query reformulation using automatically generated query concepts from a document space. Information Processing and Management 42 (2006) 453-468
[10]. Buscaldi D., Rosso P., and Arnal E. S.: "A WordNet-based Query Expansion method for GeographicalInformationRetrieval", 2005.

[11].http://en.wikipedia.org/wiki/Category:Computer_science
12 . E. Hatcher and O. Gospodnetic. Lucene in Action (In Action series). ManningPublications Co., Greenwich, CT, USA, 2004.