RESEARCH ARTICLE OPEN ACCESS

An Ontology Based Approach For Semantically Enhanced Information Retrieval

Archana V. Bhoyar[1], Mrs. Mona Mulchandani[2]
[1, 2](Jhulelal Institute Of Technology, Lonara, Nagpur, Maharashtra, India)

Abstract:

Today's content explanation and query processing techniques for IR are depend on keywords. Therefore, it has limited capabilities to understand and exploit conceptualizations involved in user needs and content meanings. To minimize the drawback of keyword-based models, search the content by meaning rather than literal string, the idea of conceptual search has been the focus of a wide body of research in the IR field. In this paper, we have done the survey on semantic searching techniques and find different works that have been done in semantic search by different researchers. In the survey, following two fields of research are noticed, first is graph ontology based IR model and second is vector space IR model. A input text can represented as Graph, in which, vertex may be feature term and edge relation can be a relation between the feature terms. The semantic graph based method is a proper way of representing text document. It improved results of analysis and searching over traditional models.

Keywords **- Semantic Web, Information Retrieval, semantic search,Graph Analysis.**

I. INTRODUCTION

As the online information on the web is increasing, searching and managing the large amount of data have become increasingly challenging. Search engines, such as Google, Yahoo is continuously introducing new implementations to improve the users' search experience, which includes the different information such as news, blogs, forums or books; the introduction of metadata by publishers to enhance the visualization of results; or the use of personal and contextual information, such as social networks, location, etc., to particularize results according to users' tastes, interests and situations.

Even though there is enhancements in the search engine technology, the data description and query processing methods, Information Retrieval technology are basically based on keywords, and hence it provide limited capabilities to capture the conceptualizations involved in user needs and content meanings. To overcome the drawback of keyword-based models, we introduce the semantic search, it search the content by meanings rather than literal strings, has been the focus of awide body of research in the IR and the Semantic Web communities.

In Today's environment text is the most common form for storing the information. The document representation is an important step in the process of searching. Hence, the challenging task is the proper representation of the textual information which will capable of representing the semantic information of the text. Existing models like the vector space model consider numerical feature vectors in a Euclidean space. Thus, proper method for semantic search is necessary.

Here we discuss some techniques of semantic search and information retrieval.

II. LITERATURE REVIEW

A. Vector Space Model

Vector space model [1] is an algebraic model. It uses Euclidian space and considers mathematical attributes vectors. The concept of vector space model is simple, but it has the following drawbacks:

The meaning of a words and design is not understandable.

Each word is not dependent on other, and does not represent sequence of text or other dependency.

If there are two records having same meaning, but words include in them are different, then it is difficult to compute similarity between them.

All words are divided in unit, passage, sentences and articles for defining the content of record. Therefore the relation between various contents of a record, their position and numbering are valuable to know the record in specific way. The solution to this problem is graph based representation model [2]. Graph based representation is nothing but mathematical representation and can construct relationship and arrange information in a systematic manner. Graph based representation of a word record is effective because it can helpful and powerful in most of operations in word such as topology, relation, ontology, statistics, etc.

B. An Ontology Based Concept

An Ontology based approach depends on principles from semantically enhanced information retrieval [2], wherever a general framework is developed to influence ontology's within the frame of a standard vector space IR model. During this work, they address the additional challenges concerned in creating the approach possible on massive and heterogeneous data repositories, as needed to focus on sensible and realistic settings like the web.

The core linguistics search model relies on associate degree adaptation of the classic keyword-based IR model. It contains four main methods of IR system: indexing, querying, searching and ranking. However, opposite to the existing keyword-based IR models, in this method, the query is written in query language which is based on ontology (SPARQL), and also the external resources which are external and used for process of indexing and query processing have an ontology with its KB. The indexing method is similar to a semantic annotation method. In the ontology-based IR model, rather than making inverted index wherever the keywords are related to the documents, the inverted index includes semantic entities related to the documents wherever they seem. The relation between a semantic entity and a document is called annotation.

Drawback of semantic retrieval on the Web

Heterogeneity: This ontology gives reasonably good information of knowledge areas such as geographical locations, organizations, etc. Nevertheless, there is a large number of contents available on the Web which has a unlimited number of domains.

Scalability: Scalability is the major issue in ontology-based technologies. Scaling the model of the Web environment implies, not only to accomplish all the increasing semantic metadata which is used for providing a good knowledge of contents, but also to arrange large amounts of data which is in the form of unstructured content.

Usability: Usability is the important requirement to extend the ontology-based retrieval model in the Web environment to provide users with an easy way to use query user interface. This means that users do not require previous knowledge of ontology-based query languages, or they don't need to formulate their queries.

C. Semantically Enhanced IR

The aim is to retrieve and find the most important and related information on the World Wide Web. The focus is to defeat the drawbacks of existing keyword based search model .A semantic search removes the drawbacks of overloaded and mismatch data related with keyword based search. A semantic search makes the use of ontology and clustering algorithm [3] to find related and important retrieval of web documents. This method is for improving the information retrieval process using semantic search. Its aim is to get the most useful data from the large amount of data present on the World Wide Web.

With the help of Semantic search we can improve search accuracy by understanding searcher meaning and the contextual meaning of terms. It uses the pre-processing and document clustering for semantic information retrieval.

D. Semantic Graph Mapping

This paper gives the method for summarizing document with the help of creation of semantic graph of the actual document. Next step is to find the substructure of that a graph which can be used to find sentences for a document summary [4]. The first step in this process is a syntactic analysis (deep) of the document, next step is for each sentence, extract logical form triples, subject–predicate–object from text. Then next step is semantic normalization, co-reference resolution and cross-sentence pronoun resolution for refining the set of triples. Last step is to merge them into a semantic graph. This method is applied to documents and corresponding summary extracts of it. They make the use of Support Vector Machine on the logical form triples for the automatic creation of document summaries.

E. Semantic Search Based on Graph Analysis

The Graph structures consist of nodes which indicate feature terms and edges which indicates the relationship between terms. A Relationship may be co-occurrence [6, 7, 8], grammatical [9], semantic [1, 10] or conceptual. When a text document represented as Graph, no. of graph analysis methods can be applied on it to process graph. Graph operation such as Graph union, Graph intersection, topological properties such as degree coefficient, clustering component and vertex ranking, small world property found effective and efficient text document analysis for different applications. In this analysis there is no requirement of detailed semantic knowledge, domain or language specific knowledge.

III. PROPOSED METHODOLOGY

The conclusion from literature survey is the realization of different semantic retrieval model which has deep level of conceptualization and which helps search in large, open and heterogeneous repositories of unstructured data.

To overcome this drawback, the system have proposed method for more research in this direction because it facing the limitations on heterogeneous, massive repository such as web, the proposed system overcome this by constructing a complete semantic retrieval approach.

Building of a complete semantic retrieval approach

The input of the proposed system is, Natural language queries, And the Output may be Specific answers which are in the form of ontology entities and documents which are ranked semantically which overcome the drawback of basic model in which ontology based language is used for input query.

Challenges Faced in the Web environment:

Heterogeneity: The proposed system is able to potentially cover a large amount of domains reusing the ontology's and KBs available online Semantic coverage enhancement would directly result in retrieval performance improvement.

Scalability: The semantic indexing (annotation) process of proposed system can maintain huge amounts of unstructured data and semantic metadata without any predefined restriction.

Usability: Queries are written in NL, so, it is useful for user interface.

For research purpose, proposed system will use graph creation method [5]. As all paper about ontology based system are based on the tree structure. To improve the speed of search system will use the semantic graph mapping technique**.**

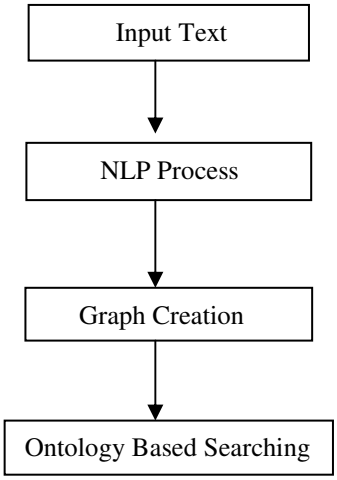

Fig. 1.The block diagram of the proposed system

IV. Result and Discussion

To understand the proposed work, we have prepared a set of natural language queries for the mobile domain. In the first step these queries are given to the natural language processing module where the semantic graphs of queries are created. Natural language processing is useful in the second step which is preprocessing. The preprocessing module has three steps: taking user query as input, tokenization, and POS tagging.

User query as input:

It takes the entered query of a mobile domain of user in which he wants to retrieve information about. Then this user entered query is sent to part of speech tagger for next processing.

Tokenization:

Here, the user query is divided into a no. of tokens with the help of part of speech tagger, in this step words of the query is separated from one another for tagging by part of speech tagger.

Part Of Speech tagging:

Tokenization gives the token for further process in which part of speech tagger tagged the each word with different part of speech in the English language such as a noun, adjective, pronoun, etc. by POS tagger. With the help of this tagging, the system creates the semantic graph of query and removes stop words.

Each word of the query is tagged by its corresponding part of speech with the help of part of speech tagger. After that it extracts the entity as well as attributes from the query for creation of a semantic graph. It considers noun as an entity and adjective as an attribute to create a semantic graph from the query. A semantic graph of user entered queries is designed in different levels such as by considering domain at first level, entities at the second level and attribute at the third level of the graph.

The output

Using the above index, the score of each term is calculated using TF-IDF [15]. The product of this TF-IDF score and assigned weights is semantic TF-IDF score. Using five documents, table-2 shows the expected retrieval ranking of documents and actual retrieval result using both scores for user query2.

TABLE I I
DOCUMENT RETRIEVAL

Documents	Expected ranking	TF-IDF score - ranking	Semantic TF-IDF score - ranking
PAPER-1	1	0.63 - 2	0.96 - 1
PAPER-2	2	0.90 - 1	0.95 - 2
PAPER-3	3	0.51 - 3	0.91 - 3
PAPER-4	4	0.38 - 5	0.89 - 4
PAPER-5	5	0.45 - 4	0.87 - 5

By comparing TF-IDF score and semantic TF-IDF score we can conclude that the accuracy of semantic TF-IDF score is more than normal TF-IDF score.

V. Conclusion

The semantic search is the best alternative to the existing information retrieval system. Ontology based approach is the best way for creating semantic information retrieval. This paper is about ontology based system which is based on the tree structure. Semantic graph mapping technique is use to improve the speed of search in the given system. The graph based analysis does not require detailed linguistic knowledge, domain or language specific collection. It is highly portable to other domains and languages. Processing of the information in various fields like document clustering, classification, prepositional phrase attachment, etc. are provided by graph based representation of text elements.

REFERENCES

1. A.Tulika Narang, B.Prof. R.R. Tewari, Towards Semantically Enhanced Information Retrieval, International Journal of Latest Trends in Engineering and Technology (IJLTET), 2012.
2. Jure Leskovec1, Marko Grobelnik2 and Natasa Milic-Frayling3, Learning Semantic Graph Mapping for Document Summarization, International Journal of Web & Semantic Technology (IJWesT),2012.
3. S. S. Sonawane, Dr. P. A. ,Graph based Representation and Analysis of Text Document: A Survey of Techniques ,International Journal of Computer Applications, 2014 ,Volume 96.
4. Wei WeiJin and Rohini Srihari, Graph-based text representation and knowledge discovery. In proceedings of the SAC conference,2007, pp 807-811.
5. Bordag, S., Heyer, G., Quasthoff, U. Small worlds of concepts and other principles of semantic search . In T. Bhme, G. Heyer, H. Unger (Eds.), IICS, 2003, lecture notes in computer science Vol. 2877, pp. 10-19.
6. Francois Francois Rousseau, Michalis Vazigiannis, Graph-of-word and TW-IDF: New Approach to Ad Hoc IR. Proceedings of the 22nd ACM international conference on Conference on information and knowledge management 2013, pp. 59-68.
7. Lakshmi Ramachandran, Edward F. Gehringer, Determining Degree of Relevance of Reviews Using a Graph-Based Text Representation. Proceedings of the 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, 2011, pp.442-445.
8. Steyvers, M., Tenenbaum, J. B. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Cognitive Science, 2005, Pp.41-78.
9. Motter, A.E et al Topology of the conceptual network of language. Phy. Rev. E.Stat. Nonlin. Soft Matter Phys., 65, 2002.
10. Dr.Avinash J. Agrawal,Dr.O. G. Kakde, Semantic Analysis of Natural Language Queries Using Domain Ontology for Information Access from Database, I.J. Intelligent Systems and Applications, 2013, 12, 81-9
11. Motter, A.E et al Topology of the conceptual network of language. Phy. Rev. E.Stat. Nonlin. Soft Matter Phys., 65, 2002.
12. X. Ning, H. Jin, and H. Wu, "RSS: A framework enabling ranked search on the semantic web", *Information Processing & Management,* 2007, in a press.
13. Salton, "Introduction to Modern Information Retrieval", McGraw-Hill, NewYork, NY, USA, 1986.
14. Mingyong and Liunand Jiangang Yang, "An improvement of TFIDF weighting in text categorization", International Conference on Computer Technology and Science(ICCTS 2012)
15. Juan Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries", Rutgers University, 23515 BPO Way, Piscataway, NJ, 08855
16. Salton, G. & Buckley, C. "Term-weighting approach in automatic text retrieval", In *Information Processing & Management* (1988), 24(5): 513-523.