ISSN:0975-9646

Comparative Study of Web Search Methods Using Ontology

Anushree Mehta
Department of Computer Engineering
Dwarkadas J. Sanghvi College of Engineering
Mumbai University, Mumbai

Shreyansh Zatakia
Department of Computer Engineering
Dwarkadas J. Sanghvi College of Engineering
Mumbai University, Mumbai

Prof. Khushali Deulkar
Department of Computer Engineering
Dwarkadas J. Sanghvi College of Engineering
Mumbai University, Mumbai

***Abstract* – An enormous amount of information is present on the Internet. If any user wants to access some data, they would tend to enter their query into a search engine. This is known as Web Searching. Web Searching has increasingly become a complicated task over the years. To add to this, it often provides inaccurate results. Thus there is a dire need for a more accurate and fast search result providing system. This problem may be solved by the Semantic Web. One of the pillars of the Semantic Web is Ontology. Usage of ontologies during web search has provided some promising results. This paper discusses some of the proposed methods of incorporating ontologies in web search. After a brief description is provided for each of the methods, they are compared on the basis of the ontology technologies supported by them and the methods utilized for implementation.**

Keywords – Ontology, Web Search, Semantic Web

I. INTRODUCTION

The Internet provides a platform for people to publish various kinds of information at any point in time. This convenience led to a large amount of electronic information to be present on the Internet today. Navigating through millions of pages of information is in itself a challenging task and often leads to inefficient results. Today, web search results are obtained by crawling through web pages from one hyperlink to another. The pages obtained in this manner are then searched for clues linking them to the user query by using query understanding techniques, synonyms, search tools, etc. The shortlisted pages are then displayed in a certain order, which is obtained as a result of a ranking algorithm. The inefficiency and inaccuracy of results obtained using this method can be avoided by either improving search tools or by utilizing the Semantic Web.

One of the main shortcomings of the World Wide Web (WWW) is that queries can often be misinterpreted due to lack of semantic meaning associated with it. This drawback is overcome by the Semantic Web using ontology. Ontologies are explicit formal definitions and explanations of the terms in a domain and relations among them as described by Natalya Noy and Deborah McGuinness in [1]. They contain definitions and semantic meanings of terms which help derive the intended meaning of the user query correspondingly leading to accurate results. Ontologies can be incorporated into Semantic Web searches in numerous ways. These ways generally employ different kinds of ontologies at different stages of web search. In this paper, we discuss and compare some of these methodologies.

The rest is structured as follows. Section II discusses different methods of web searching using ontologies. Section III contains a comparison table and description of parameters while in Section IV we arrive to a conclusion based on the comparison.

II. METHODS OF WEB SEARCHING USING ONTOLOGY

A. DLOSSS

The aim of web searching is to find the most relevant information first. This is facilitated in DLOSSS by converting user query into query ontology in accordance with the domain ontology and allowing the user to choose some or all of the possible meanings of this query ontology. Performing this step allows for the results to be refined as per the semantic context of the user query. A set of relevant web pages is obtained based on the meaning(s) chosen by the user, which is then sent for semantic analysis. [2] Semantic analysis is done with the help of morphological, semantic, and syntactic analysis agents. Before semantic analysis, terms, concepts, taxonomic relations are extracted from the previously obtained collection of web pages, to create a corpus. This corpus is analysed under a given criterion and the ontology is enriched with novel information from this corpus. Through this, ontology is constantly updated and refined to obtain better results the next time around. [2]

B. MIRO

MIRO is an indexing and searching system that is based on automatically evolving ontologies. It also offers a guided search tool. The search process in this system starts with a query syntax check. Next lexical treatment is performed on the query to normalize it and to remove stop words. The concepts that are requested in the query are then identified and searched for in the ontology. Two lists are obtained as output of this step. The first list includes the recognized concepts, i.e. present in the ontology and the second list contains the unrecognized concepts which are used to update the ontology. If the user provides a concept synonym then the concept corresponding to that synonym

is used during searching. Finally, the user query is converted into an SQL query based on the recognized concepts, This SQL query is used to retrieve results from the index. Here, ontologies are used to solve the problems caused due to the noise and silence of web pages. [3]

C. Context Synonymy

In this paper, the authors propose a structural approach for unstructured knowledge over the Internet.[4] They look at two major problems in information retrieval namely Polysemy (word having multiple meanings) and Synonymy (multiple words with same meaning) and propose a system which resolves both these problems. In this system, as the user query is entered, keywords are extracted from it, and passed to their context synonym ontology which determines the various contexts available for the keywords. The user selects the desired context if multiple contexts are available. Next, the respective ontology is retrieved and the corresponding URL list is retrieved from web repository. Ranking algorithms are employed to sort the obtained pages as per relevance ratio and results are displayed to the user. The authors observed that more focussed results are obtained as a result of this method.

D. Personal Social Dataset with Ontology-guided Input

Personal Social Dataset (PSD) is a featherweight ontology used for collective filtering of data and Ontology-Guided Input (OGI) facilitates query refining and multifaceted browsing [4]. In this method, the above mentioned concepts are combined to form a system which can improve search efficiency by filtering results without bring hindered by incompatible data formats. Initially, a PSD is created for the user and filtration of search engine results is done according to correlation of results. Then the OGI is used to filter results further and to make up a basis for traversing through the results obtained in the previous step. Indexing and ranking of documents is performed while creating the PSD through a Vector Space Model which represents documents as vectors of keywords. Relevancy ranking is performed on the basis of document similarities and comparison of deviation angles between document vectors and query vector. [5]

E. A Multi-Agent System for personalized Web Search

To give accurate search results to the users, one needs to build user profiles to record their interests. This paper proposes to build a dynamic user profile based on multi-agent approach. Here, the behaviour of the user is tracked and classified into short and long term interests. User interests are represented as ontological conceptions which can be built by indexing web pages visited by a user to a reference ontology. [6] Interest weights are then assigned to the concepts where, uninteresting concepts are assigned 0 and interesting concepts are associated with frequency value and a document that contains all the visited web pages that are identified as interesting to the users. This document then filters webpages and displays only those pages that are interesting to the users. When a user enters a query, it is mapped to the user profile to identify similar concepts that represents the query. It is also passed to any search engine to retrieve initial search results. Next, the extent of synonymy between search results and user interest document is calculated and a value is assigned to each result. As similarity increases the value also increases. Finally, the results are ranked in descending order and presented to the user.

F. Personalized Web Search with Location Preferences

This paper proposes Ontology based Multi-Facet (OMF) personalization. The users' preferred content and location preferences are identified using different methods and techniques. User's clickthrough plays an important role in identifying the preferences. Content ontology and location ontology are then used to store the content and location preferences respectively. Content and location preferences are maintained by using two different adaptation process. Thus, weights are assigned to content and location preference, to integrate the two different processes together. Here, the query is analysed to identify content and location entropies. It is useful in measuring the diversity of content and location information. The clickthrough of the user helps in profiling of the user. It helps us identify the type of the user and with this information we can personalize the web search for the user. Based on this information the search results are ranked so as to personalize it as much as possible. Ranking SVM is used for the re-ranking of the search results. When compared to existing methods OMF provides more accurate results and this is proved by experimental results. [7]

G. Organization oriented Web Search

This is a hybrid web search method that comprises of semantic web and keyword matching approach that is traditionally used. Semantic web conceptualized knowledge into both human and machine understandable languages. This method consists of a knowledge base which is constructed using ontology. The main function of this knowledge base is capturing the semantic of common terminology and, if it is not understandable to human then it should be replaced with something that humans can understand. This method essentially analyses keywords using ontology and understands the underlying relationships between various keywords. If the background knowledge base is too large then entities have to be ranked. This ranking is done on the basis of a heuristic function which uses the concept of weights. Thus the resulting query that is provided as a solution gives better search results than the initial query. [8]

H. Fuzzy-Go

Fuzzy-Go, as the name suggests, uses fuzzy logic theory and semantic web search technique. The synonyms or words similar to keywords given in the web pages are identifiable and searchable in Fuzzy-Go. To accomplish this, a fuzzy ontology is created by using fuzzy logic that

captures similarity of terms. For this purpose, a data mining approach is used which calculates the similarities between different terms. A web crawler is created that classifies web pages and stores it according to their domain. An option is provided to the users that give them the freedom to assign priority or weights to their search keywords based on their needs. Then during the search, the webpages that are not in the domain are excluded from the search so as to reduce the search space. Next, using fuzzy ontology, the keywords are expanded to find the synonyms and similar words. The search results thus obtained are then ranked based on factors such satisfaction of keywords, degree of importance of keywords, relevance of domain etc. As a result, improved search results are obtained. [9]

III. COMPARITIVE STUDY

The above mentioned approaches have been compared in Table I. The parameters for comparison are approach, ontology technology and analysis methods. In approach the different ways of performing web search are looked at. In ontology technology, the language or technology used for implementing the ontology used in method is mentioned. Analysis method talks about the models used for ranking and Indexing of pages.

TABLE I
Comparative Study of Methods of Web Searching using Ontology

Method	Approach	Ontology Technology	Analysis Model
DLOSSS [2]	Semantic analysis combined with Ontology evolution	WordNet	Ranking: Morphological, Semantic, Syntactic analysis Indexing: Unclear
MIRO [3]	Concept based	OWL (supports all via plug-ins)	Indexation: Concept Similarity Ranking: Unclear
Context Synonymy[4]	Context based	Unclear	Ranking: Relevance ratio
Personal Social Dataset and Ontology-guided Input [5]	Keyword and Vector based	RDFS	Indexing and Ranking: Relevance based derived from deviation angle of vector document
Multi-agent system for personalized web search [6]	Multi-agent approach for creating a dynamic user profile	Unclear	Ranking: ranking of results based on user profile document Indexing:Unclear
Personalized Web Search with Location Preferences [7]	Ontology based multi-facet personalization	Unclear	Ranking:Based on training of SVM (Support Vector Machine) Indexing:hybrid index to handle both content and location aware queries
Organization Oriented Web Search [8]	Ontology Based Analysis of keywords	RDF based Ontology	Ranking: heuristic function based on weights Indexing:Unclear
Fuzzy-Go [9]	Fuzzy Logic Theory	Suggested Upper Merged Ontology (SUMO)	Ranking: Based on parameters like keyword satsifaction, keyword importance, doamin relevance etc. Indexing:Unclear

IV. CONCLUSION

Various methods of web searching using ontology have been developed in the recent decade. This paper looks at some of these methods and their attributes. Here, it can be noticed that MIRO supports multiple technologies of ontologies which is very advantageous. On the other hand, some methods like Personalized Web Search with location preferences and Personal Social Dataset and Ontology-guided Input, employ both indexing and ranking instead of only one thus improving their result accuracy. In this manner all the methods have been compared and a conclusion has been reached that depending on the criterion for web search these methods would perform better than the others.

REFERENCES

[1] Natalya F. Noy, Deborah L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, 2000.

[2] Yi Liu, Yu Wang, Deli Yang, "DLOSSS: An Improved Ontology-based Semantic Web Search System" in *Fifth International Conference on Intelligence Control an Information Processing*, 2014.

[3] Fraihat Salam, "New Semantic Indexing and Search System based on Ontology" in *Fourth International Conference on Emerging Intelligent Data and Web Technologies*, 2013.

[4] Eakansh Manglik, Priyanka Sharma, Dr. Paramjeet Rawat, Nidhi Tyagi, "Ontology based Context Synonymy Web Searching" in *International Conference on Information Systems and Computer Networks*, 2013.

[5] Juan Miguel Gomez, Ricardo Colomo Palacios, Giner Alor-Hernandez, Ruben Posada-Gomez, Angel Garcia Crespo, "Search in the Eye of the Beholder: Using the Personal Social Dataset Ontology-guided Input to Improve Web Search Efficiency", in *Fifth Latin American Web Congress*, 2007.

[6] Ahmad Hawalah, Maria Fasli, "A multi-agent System Using Ontological User Profiles for Dynamic User Modelling" in *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 2011.

[7] Kenneth Wai-Ting Leung, Dik Lun Lee, Wang-Chien Lee, "Personalized Web Search with Location Preferences" in *ICDE Conference* 2010.

[8] Yuzhen Dong, Jianqiang Li, "Organization Oriented Web Search Management" in *6th International Conference on Information Management, Innovation Management and Industrial Engineering*, 2013.

[9] Lien-Fu Lai, Chao-Chin Wu, Pei-Ying Lin, Liang-Tsung Huang, "Developing a Fuzzy Search Engine Based on Fuzzy Ontology and Semantic Search" *in IEEE International Conference on Fuzzy Systems* June 27-30, 2011.