International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 5(2), September - October 2015 ISSN 2278-6856

Ontology Based Search Engine

Vasantha Kavitha[1], Dr. M. Hanumanthappa[2], Dr. B R Prakash[3]

[1]Assistant Professor, Maharani Lakshmi Ammanni College for women, Bangalore.

[2]Professor, Department of Computer Science & Applications, Bangalore University, Bangalore.

[3]Assistant Professor, Dept. of MCA, Sri Siddhartha Institute of Technology, Tumkur.

Abstract

An ontology based search engine helps in identifying the most efficient and useful result for the input query. The result produced by the ontology based search engines are purely based on the literal meaning of the word in the given sentence. It does not take the keyword in the given sentence; instead it takes the meaning of the query submitted. The presence of huge amount of resources on the Web thus poses a serious problem of accurate search. This is mainly because today's Web is a human-readable Web where information cannot be easily processed by machine. Highly sophisticated, efficient keyword based search engines that have evolved today have not been able to bridge this gap. There are many kind of techniques followed in implementing the ontology based search engines. Here, in this paper we identify the some of the techniques to be used in developing the search engine. All of the techniques are different from one other and that the efficiency is also different. These techniques form a special pattern of accuracy and they are disused in the paper. The difference in the working of the keyword based search engines and the ontology based search engines are shown with examples. Also the ontology based search engine that is build up using the fuzzy logic ontology is considered here. An ontology based search engine that is developed in many steps with the help of multi crawlers is also taken into consideration.

Key words: search engines, Ontology, Information retrieval, fuzzy logic, crawler, Semantic .

I. INTRODUCTION

The current existing web mainly concentrates on the human and the document available on web is also human reliable one. Nowadays the web is not only used by humans but also the software agents. This reality case brought the usage of the semantic ontology based search on web. Most of the traditional web users are not sure about their query for which they need the search engine to provide the results. Hence the normal keyword based search will not be in a position to provide the accurate search results to the user. In this situation we need a semantically proven search engine. The figure I represents the general framework of the semantic web. Here levels of the query that must it pass through is clearly port rated. When a user is not sure about the query he will provide only relative words together and in that case the semantic based search engine will compare the words and users the relationship between those words to provide the result.

A **search engine** is a document retrieval system designed to help find information stored in a computer system, such as on the World Wide Web, inside a corporate or proprietary network, or in a personal computer. The search engine allows one to ask for content meeting specific criteria (typically those containing a given word or phrase) and retrieves a list of items that match those criteria.

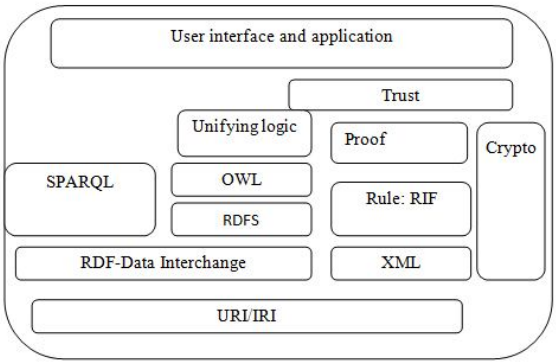


Figure 1 Semantic Web Framework

In this case the result provided will be efficient and more meaningful. Always the users will expect the desired results to appear as the first result and the semantic based search engine will provide the same means rather than using the keyword based or context based search engines. Recently many search engines that are semantically proven are developed using ontology languages like RDF, OWL, HTML. The paper compares the performance of the search engines that are developed using these languages

2.SEMANTIC WEB SEARCH ENGINE

2.1 The working of a regular search engine

For most internet users, a search engine is the starting point of finding desired information in the Web. The most common form of text search used by the majority of popular search engines on the Web is **keyword based search** that is**,** they do their text query and retrieval using keywords. The working of any regular search engine may be summarized as follows:

- Search engine searches its enormous database for the keyword - entered by the user (after pressing the search button.)
- Every engine has its own collection system to fill its database.

- Indexing system is used to organize the database - permits faster searching
- Returns a list of hit -includes relevant (as well as irrelevant) pages
- This keyword based search technique gives rise to several problems listed as follows:
- The Web is growing much faster than any present-technology search engine can possibly index. In 2006, some users found major search-engines became slower to index new Web-pages.
- Keyword searches have a tough time distinguishing between words that are spelled the same way, but mean something different. This often results in hits that are completely irrelevant to the query.
- Some search engines also have trouble with stemming, i.e., if the word "big," is entered, should it return a hit on the word, "bigger?" What about singular and plural words? What about verb tenses that differ from the word someone entered by only an "s," or an "ed"?
- Search engines also cannot return hits on keywords that mean the same, but are not actually entered in the query. A query on heart disease would not return a document that used the word "cardiac" instead of "heart."
- Users are returned thousands to millions of Web pages in return of their queries, of which majority prove to be irrelevant to the query submitted and is impossible for any user to go through.

In view of the above mentioned problems, come up the concept of semantic Web and semantic Web search engines.

2.2 Semantic Web and Semantic Search Engine

"The Semantic Web is the representation of data on the World Wide Web. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming." – W3C Semantic Web. The Semantic Web is a framework that allows publishing, sharing, and reusing data and knowledge on the Web and across applications, enterprises, and community boundaries [4]. Currently, the Semantic Web, consisting of Semantic Web documents typically encoded in the languages RDF and OWL, is essentially a Web universe parallel to the Web of HTML documents [5]. Knowledge encoded in Semantic Web languages such as RDF differs from both the largely unstructured free text found on most Web pages and the highly structured information found in databases. Such semi-structured information requires using a combination of techniques for effective indexing and retrieval. RDF and the Web Ontology Language (OWL) which are ontology based procedures or representing knowledge on the Web, introduce aspects beyond those used in ordinary XML, allowing users to define terms
(for example, classes and properties), express relationships among them, and assert constraints and axioms that hold for well-formed data. An application of the emerging Semantic Web is a Semantic Web search engine which searches the Semantic Web documents against a user query for accurate results. Our work uses RDF encoded Semantic Web documents which are searched in response to a user query for exact results.

3. ONTOLOGY

An ontology is an explicit specification of some topic. For our purposes, it is a formal and declarative representation which includes the vocabulary (or names) for referring to the terms in that subject area and the logical statements that describe what the terms are, how they are related to each other, and how they can or cannot be related to each other. Therefore, Ontology provides a vocabulary for representing and communicating knowledge about some topic and a set of relationships that hold among the terms in that vocabulary [2] [3].

Why develop an Ontology?

- To enable a machine to use the knowledge in some application.
- To enable multiple machines to share their knowledge.
- To help yourself understand some area of knowledge better.
- To help other people understand some area of knowledge.
- To help people reach a consensus in their understanding of some area of knowledge.

In our project we used Resource Description Framework or RDF to represent knowledge. For example, if we need to describe a subject in terms of its classes and their relationships using RDF, we are creating an Ontology.
As our project deals with the crops domain, the designed ontology is shown in Figure 1. In Figure 2, the general information ontology is depicted. A relational diagram is shown in Figure 3 to depict some classes, instances, and relations among them in the crops domain.

4. FUZZY ONTOLOGY BASED SEARCH ENGINE

The paper [3] describes about providing an efficient method using query refinement process. Once if the user articulates an exact word for search, the search engine will pull the correct result to the top of the page. The query refinement process is implemented in the PASS (Personalized Abstract Search Service) system. All the time, the user cannot provide the correct word for search.
Hence to overcome this problem, in this paper, a method of Fuzzy ontology has been implemented. This method processing takes place in such a manner that it compares

the search query based on keyword based retrieval and also the results provided by the PASS method for that particular user query. The final outcome will be purely based on the comparison of those two results. The PASS system will provide the abstract of the paper when the user clicks on the link and also will provide the list of related papers if they are available. To provide all the features mentioned above, the PASS method is implemented in two dimensions. One is using the structure of the domain and other is using the knowledge of the user. For this process WordNet dictionary is used. Document clustering is the next function done here. And this is handled by using the scatter gather algorithm. Also here, the cosine technology has been used for constructing document similarity networks. In the paper they have mainly concentrated on the construction of fuzzy ontology and query refinement process. The fuzzy ontology uses set of terms with broader and narrower meaning. The broader terms are the inverse of the narrower terms. This method of construction is mainly carried out using the relation between the broader and narrower terms in the query given by the user. The literal motive is to bring out the relationship of the terms.

Let *C* = (*a*1, *a*2, ¼ *a*n) be a collection of articles *a*i, where each article *a* = (*t*1, *t*2, ¼ *t*m) is represented by a set of terms *t*j. Let *occur* (*t*j,*a*) denote the occurrence of *t*j in article *a*. The membership degree of *occur* (*t*j,*a*) is defined by m*occur* (*t*j,*a*) = *f*(|*t*j|), which in general is a function of term's frequency of occurrence. In the information retrieval community, the function *f* can be viewed as the normalized within document term weighting method. Let *NT*(*t*i, *t*j) denote that *t*i is *narrower* than *t*j. The Membership degree of NT (*t*i, tj), represented

NT(*t*i, *t*j), is defined by

$$\mu NT(t_i, t_j) = \frac{\sum_{a \in C} \mu\, occur\,(t_i, a) \square \mu occur(t_j, a)}{\sum_{a \in C} \mu\, occur\,(t_i, a)} \quad (1)$$

In (1), Σ denotes a fuzzy conjunction operator. In current implementation, we use a binary function for the *f* function so that Σ *occur*(tj,*a*) = 1 if the occurrence frequency of *t*j Σ 0, or Σ *occur*(*t*j,*a*) = 0 otherwise.

Using the binary function will turn Equation 1 into the same equation regardless the selection of fuzzy conjunction operator. Let *BT*(*t*i, *t*j) denote that *t*i is *broader* than *t*j. Because the notion of broader term is basically the inverse of narrower term notion, the membership value of *BT*(*t*i, *t*j) is derived from the membership value of

$$NT(t_i, t_j)\ \mathrm{BT}(t_i, t_j) = \Sigma\ \mathrm{NT}(t_j, t_i) \quad (2)$$

The fuzzy ontology construction is done in two major steps. They are building fuzzy ontology from fuzzy narrower terms and by fuzzy ontology pruning. In the first step, the membership values of two NT relations are calculated. During this process the redundant terms, meaningless terms and unrelated terms re found and eliminated. In case of membership value being zero indicates that the two terms are unrelated. tn the second step of fuzzy ontology creation, next level of reducing the relations is carried out by making an analysis on the set of relations.

Table1: The Concatenation of query refinement with Fuzzy Ontology

Sl.No	Ontology	Percentage
1	Related terms	37
2	Broader terms	47
3	Narrower terms	16

Finally, the experimental results show that this system is built on the fuzzy ontology and automatic technique for PASS system. The method collaboration is one of the idle results provided in the paper. The efficiency of the system can be improved even by combined use of PASS features.

Finally, the experimental results show that this system is built on the fuzzy ontology and automatic technique for PASS system. The method collaboration is one of the idle results provided in the paper. The efficiency of the system can be improved even by combined use of PASS features.

5. CRAWLING SEARCH

This method follows a crawler based search engine for implementation and this architecture is called the PSSE (Personalized Semantic Search Engine). The system mainly concentrates on minimizing the processing time. For this they have followed web page clustering. Annotation agents and ontology matching are the concepts utilized in this paper. Annotation is the process of just assuming that the derived feature is correct the then continue with the next level of processing.In the architecture the processing phase is split up into two different phases. One is working in online phase and other is working in offline phase. In the offline phase the crawling of web and pre-processing of pages takes place. The first and foremost step in the architecture is the crawling process. In the crawling process, as this approach uses multi crawlers, they traverse the World Wide Web and finds the web resources and finally stores in their database. Here the crawler's job is to find the related links for the user query and provide them.

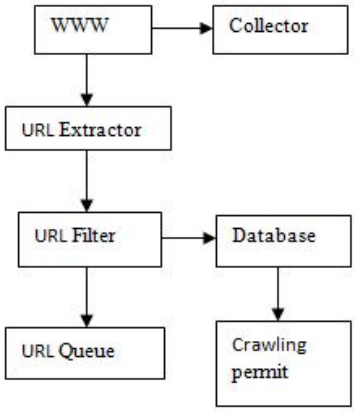
In the pre-processing stage the time consumption will be less because the indexer will generate the graph for all crawled pages. The graph will be acting as a special cluster that holds similar data within each cluster. The resultant cluster will be processed using link analysis technique. This process is carried out for the authorization

of each web documents. This can be done by using the below given formula,

$$PR(A)=(1-d)+d \sum_{i=1} \frac{PR(Ti)}{C(Ti)} \quad (3)$$

Also the annotation process can be done after the measurement is carried out. The weight assignment for each annotation can be done by using the calculation by finding the relevancy of the document. This feature can be concadinatedly. The calculation can be performed by using the cosine function that is mentioned below.

$$Wij=tfij * \log 2 (N/n) \quad (4)$$



COMPARISON BETWEEN THE METHODS:

Table 2: COMPARISON OF METHODS

Title	Approach	Method
Fuzzy ontology	Keyword based	Fuzzy technique
Keyword And Semantic Search Engines	Keyword and semantic based	Precision ratio
Personalized Mobile search engine	Keyword based	Client-server Architecture
PSSE	Crawler based	Muliti crawler

Summary:

We have identify some of the techniques to be used in developing the search engine. As these techniques are different from one other and that the efficiency is also different and also, these techniques form a special pattern of accuracy as they are discussed. The normal search engine is not satisfied by browsers. So we have compared with Semantic web engine, Ontology search engine with fuzzy Ontology, depending on the requirement. The paper describes the comparison and analysis between various methods involved in developing ontology based search engines. It makes clear that the usage of ontology based search engine will provide accurate results depending on the literal meaning of the query and the semantic search engines will produce results based on the query logic. Future work can be implemented by combining the multi crawlers and fuzzy logic to form a new approach for an efficient ontology based search engine.

REFERENCES

[1]. Jagendra Singh, Dr. Aditi Sharan,"A Comparative Study between Keyword and Semantic Based Search Engines" International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV School of Computer & Systems Sciences Jawaharlal Nehru, University New Delhi, India

[2]. Mrs. Rashmi A. Jolhe, Dr. Sudhir D. Sawarkar," An Ontology Based Personalised Mobile Search Engine" Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4,Issue 2(Version 1), February 2014, pp.69-74 Department of Information Technology, Datta Meghe College of Engineering, Airoli, NaviMumbai

[3]. Dwi H. Widyantoro , John Yen, " A Fuzzy Ontology-based Abstract Search Engine and Its User Studies" Department of Computer Sciences Texas A&M University College Station, TX 77843-3112, USA School of Information Sciences and TechnologyPennsylvania State University University Park, PA 16801-3857, USA, A. Gulli and A. Signorini, "The Index able Web is more than11.5 ,billion pages," 2013.

[4]. Efrati and Amir, "Google Gives Search a Refresh," The Wall Street, Journal, 2012.

[5]. Guha, R. McCool and Miller, "Semantic Search," 2011.

[6]. W. Roush, "Search beyond Google," Technology Review, 2004.

[7]. G. Antoniou and V. Harmelen, "A Semantic Web Primer," IT , Press Cambridge, Massachusetts, 2004.

[8]. Natalya F. Noy and Deborah L. McGuinness "A Guide to Creating Your First Ontology"; Stanford U. Report.

[9]. T. R. Gruber. "A translation approach to portable ontologies. Knowledge Acquisition,"5(2):199-220, 1993.

[10]. T. Berners-Lee et al 2001. "The Semantic Web. Scientific American". May 2001.

AUTHOR

Vasantha Kavitha, MCA in 2005, Mphil(Cs) in 2007, MBA in 20011 , MBADS in 2012. I earlier worked for SJCIT, currently working in Maharani Lakshmi Ammanni College.

Dr. M Hanumanthappa currently working as Professor and Chairman, in the departmetn of Computer Sciecne and Applications, Bangalore University, Bangalore. He is Having more than 20 years of expereince in Teachin and Research. He is publised More than 100 Papers in Conferences and Journals

Dr. B R Prakash currently working as a Assistant Professor in the Department of MCA, Sri Siddhartha Institute of Technology, Tumkur. He completed Ph.D from Bangalore University, Bangalore. He is published more than 35 papers in Reputed Conference and Journals.