ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 10, October 2014

A Review on Ontology Based Search Engine

R.Aravindhan[1], Mano Chitra.M[2]

Assistant Professor, Computer Science, Sri Eshwar College of Engineering, Coimbatore, India[1]

Student, Computer Science, Sri Eshwar College of Engineering, Coimbatore, India[2]

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. 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 cconsideration.

Keywords: Resource Description Framework, Web Ontology Language, Fuzzy Ontology, Support Vector Machine, Multi-Crawler.

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.

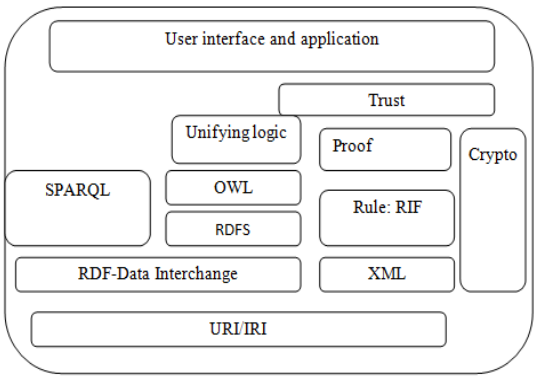

Fig 1: Semantic Web Framework

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

II. KEYWORD AND SEMANTIC BASED SEARCH ENGINES

The paper [1] analyse the comparison between the semantic based search engine and the keyword based search engines. Often the keyword based search engine do not provide the accurate result as it cannot find the meaning of the user query to the expression used in the web pages. For this performance comparison in this paper they have taken two keyword based search engines such as Google and Yahoo and also three semantic based search engines such as DuckDukGo, Hakia and Bing. Although the conventional search engines are getting modernized, they now even suffer from a defect of not able to find the exact meaning of the user query. A survey about these search engines tells that about 25% of them do not return the correct URLs on the first set of results and therefore there efficiency is not up to the extent. Some of the problems with traditional search engines are due to polysemy words, problems with synonymy, and problems with traditional information retrieval technology and finally the problem with low precision and low recall problem.

The precision is the ratio between the number of relevant retrievals to the number of total retrievals.

$$\text{Precision} = \frac{\text{Number of relevant}}{\text{Number of total retrieved}}$$

The recall value is the ratio between the number of relevant retrievals to the number of possible amount of relevant results.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 10, October 2014

$$\text{Recall} = \frac{\text{Number of relevant retrievals}}{\text{Number of possible relevant results}} \quad (3)$$

In the other case, the semantic search engines are just the extension of the normal web were the processing is done through machines. Mostly the information in the web are in the form of HTML, OWL (Web Ontology Language) or in RDF (Resource Description Framework) format. The documents in theses formats are often known as semantic web documents.

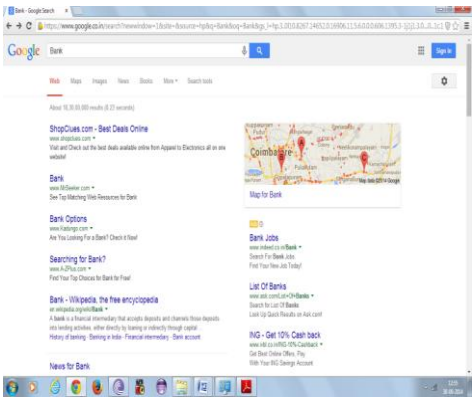

Fig 2: Result Of Google (Query-Bank)

A semantic search engines is capable of storing semantic information about the web resources and also consists of more than one relation between the terms.To make a comparison between these two kinds of search engine, in this paper they have performed a search for the word"Bank" in both the search engines. The result given by both are extremely different. In case of keyword based search engine, the result contained only about the money banks (i.e.) the result was based on frequent visiting of the users and not based on the pure keyword. But in case of semantic based search engine, the result produced is for all kinds of bank that are used worldwide. Hence, this result is based on meaning of the query and id linked with all the facts for which we might use that word.

TABLE II:

THE NATURAL QUERIES

Question no	Google	Yahoo	Bing	Hakia	DuckDuckGo
1	15	16	19	15	16
2	12	16	16	13	16
3	16	13	11	8	16
4	18	13	13	7	11
5	7	10	12	7	6

To compare the efficiency of the above done search, the semantic search engine has provided the best result while the keyword based search engine does not provide them. For the next level of comparison, in this paper they have taken five different queries as the features and all of them are given to all the kind of search engines. The resultant performance is calculated based on the number of relevant results given by each of the search engines. The queries and the performance comparison of each search engine are shown in the below table.

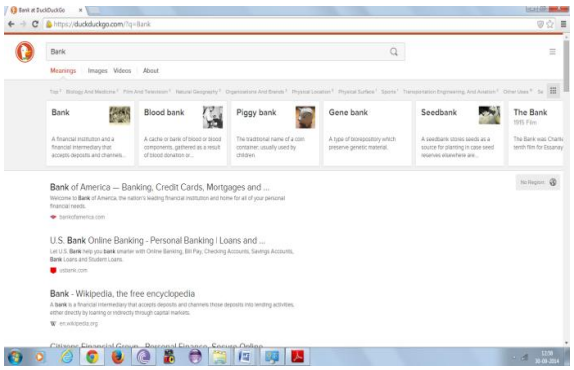

Fig 3: Result Of Duckduckgo (Query-Bank)

The experimental result produced in this paper tells that the Bing, the semantic based search engine has produced an efficient outcome when compared to all other search engines.

TABLE III:

NUMBER OF RELEVANT RESULTS

Question number	**Query**
1	Which country won the first cricket world cup?
2	Who is the current prime minister in India?
3	Why the colour of the sky is blue?
4	What is the weather now in new Delhi ?
5	Where is java?

Then DuckDuckGo stays second after Bing. This experiment proves that the semantic based search engines produced an efficient output result when compared to the keyword based search engines.

III. PERSONALIZED MOBILE SEARCH ENGINE

The paper [2] offers a different and advanced approach that matches with the development of the amount of web information. In the OBPMSE (Ontology Based Personalized Mobile Search Engine), they have captured each user's interest and personalized search results in unique user profile. Then these concepts are classified into two major categories namely content concepts and location concepts. The system design is composed of client-server model. This model should meet with three requirements such as SVM (Support Vector Machine) training, data transmission between the client servers must be minimized and finally the precise user preference

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 10, October 2014

should be stored in the client side in order to maintain the user privacy. In the content and location search, search is carried out by first asking the user weather they need the web search or places search. If the user's choice is web search, then the general search results for that particular user query is gathered from the normal Google sever to our ranking module. Then later our ranking algorithm is applied on those results and then the sorted results are given to the user's output screen page.If the user has selected the place search, then this module will first takes the user's own lattidude and longitude pair that can be automatically calculated using the GPS methodology. Then based on the user location the results are provided to the output screen page to the user.After the user submits the query, it is taken to the client and the related query containing the user content and location content are given to the server to provide the search results. In the client side, the re –ranking process takes place by following two ways. one is based on highest number of clicks and another one is displaying the results based on same number of click values. In the paper they have made use of UDD algorithm that is mainly used to avoid the duplication in the values of the final list that will given to the user.

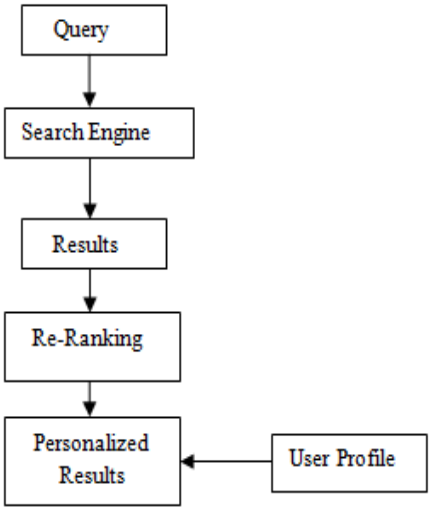

Fig 4: Re Ranking Process

IV. 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 $C = (a1, a2, ¼ an)$ be a collection of articles ai, where each article $a = (t1, t2, ¼ tm)$ is represented by a set of terms tj. Let $occur\ (tj,a)$ denote the occurrence of tj in article a. The membership degree of $occur\ (tj,a)$ is defined by m$occur\ (tj,a) = f(|tj|)$, 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 $NT(ti, tj)$ denote that ti is *narrower* than tj. The Membership degree of NT (ti, tj), represented by NT (ti, tj), is defined by

$$\mu NT(ti, tj) = \sum_{a \in C} \mu occur\ (ti, a) \otimes \mu occur(tj, a) \Big/ \sum_{a \in C} \mu occur\ (ti, a)$$

In (1), $\sum$ denotes a fuzzy conjunction operator. In current implementation, we use a binary function for the f function so that $\sum occur(tj,a) = 1$ if the occurrence frequency of $tj \sum 0$, or $\sum occur(tj,a) = 0$ otherwise. Using the binary function will turn Equation 1 into the same equation regardless the selection of fuzzy conjunction operator. Let $BT(ti, tj)$ denote that ti is *broader* than tj. Because the notion of broader term is basically the inverse of narrower term notion, the membership value of $BT(ti, tj)$ is derived from the membership value of $NT(ti, tj)$ $BT(ti, tj) = \sum NT(tj, ti)$

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.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 10, October 2014

TABLE I:

THE CONCATENATION OF QUERY REFINEMENT WITH FUZZY ONTOLOGY

S.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.

V. PERSONALIZED SEMANTIC SEARCH ENGINE

The paper [4] implements a method using the multi crawlers to collect the information from both the semantic based search engine as well as the traditional search engine. 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,

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

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.

$$W_{ij}=tf_{ij} * \log_2 (N/n)$$

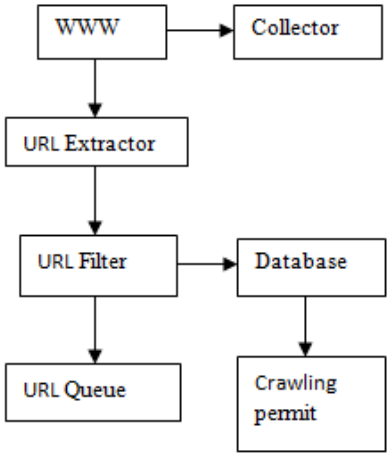

Fig 5: Crawling

In the online phase the actual searching process is carried out. Once the user query is received from the user, in the online phase the system checks for the query, process the query and the ranks them accordingly. As the final stage, the online system will provide the result to the user in this stage. This process will be carried out in majorly two steps. The first stage is searching stage and the next stage is the ranking stage. This technique enables conceptualizing user search. In the paper in the first phase, the main motive is to automate the whole process.

This can be done by making the entire process work without any human intervention. This is one of the main motives of the many researches carried out recently. To sustain this capability, here they correlate the ontology to provide a document annotation. Ranking is the next phase in this proposed method. The system ranks the user query by considering three major factors. They are mentioned as follows. One is the page authorization, second is content relevancy and finally the last facto is the user interest.

VI. INTELLIGENT SEMANTIC WEB SEARCH ENGINES

The paper[5] is different from all other search engines, as it works and produces the result by majorly concentrating on the metadata for the given user query rather than concentrating on the keywords applicable to the submitted query .In Semantic Web information it describes using a new W3C standard called the Resource Description Framework (RDF). The semantic web based search engine is called Intelligent Semantic Web Search Engines. We use the power of XML meta-tags deployed on the web page to search the queried information. The XML page will be consisted of built-in and user defined tags. The metadata information of each pages are extracted from this XML into RDF format. The RDF graphs are populated by inputting through XForms. The main method used in the system is graph identification and the RDF triples. The work has been done in this system to identify the low precision and high precision values for the output results that has been produced. This precision values are more

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 10, October 2014

important in delivering the results to the user query. The results for the query submitted by the user are provided based on the precision ratio of the query. The precision ratio value can be calculated by considering the match that is found by the search engine. The paper uses the RDF format to extract the data from the index of the particular search engine. The system is said to be intelligent as the efficiency of the outcome. The query is submitted to the search engine in the form of tags and while processing the query, each of its named tags are considered. While processing , the lower precision values are not taken into account. And always the query result that issues an efficient and higher precision values are taken for the next level of processing. Thus , this sequence of approach for providing the resultant page is the strength of the implementation.

COMPARISON BETWEEN THE METHODS:

TABLE III:

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
Intelligent semantic web search engine	Metadata based	Graphs, RDF and meta data collection

CONCLUSION

The paper describes the comparison and analysis between various methods involved in developing ontology based search engines. It also illustrates that there are many techniques or patterns that can be followed for getting an efficient result. The kind of comparison that has been done actually reflects that the efficiency of search engine differs from each method. From the above comparison, 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

[4] A. M. Riad,Hamdy K. Elminir, Mohamed Abu ElSoud, Sahar. F. Sabbeh," PSSE: An Architecture For A Personalized Semantic Search Engine" doi:10.4156/aiss.vol2.issue1.9Alzarka Higher institute for administration & computer sciences m_sabbeh@yahoo.com,

[5] Ritu Khatri, Kanwalvir Singh Dhindsa, Vishal Khatri," Investigation and Analysis of New Approach of Intelligent Semantic Web Search Engines" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-1, April 2012

[6] Billsus, D. & Pazzani, M. (1998). Learning Collaborative Information Filters. In Proc. of the Intl. Conference on Machine Learning. Morgan

K. Publishers, Madison, WI.

[7] Fellbaum, C. (1998). WordNet, an Electronic Lexical Database, MIT Press.

[8] Cutting, D.R., Krager, D.R., Pederson, J.O. and Tukey, J.W. (1992). Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections. In Proceedings of the 15th ACM SIGIR Conf. onResearch and Development in Information Retrieval, 318-329.

[9] Salton, G., and McGill, M.J. (1983). Introduction to Modern Information Retrieval. New York, McGraw-Hill.

[10] A. Gulli and A. Signorini, "The Index able Web is more than11.5 billion pages," 2013.

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

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

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

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

[15] N. R. Shadbolt, W. Hall, and T. Berners-Lee, "The semantic web: Revisited," IEEE-Intelligent Systems, vol. 21, issue 3, pp. 96–101, May 2006. [2] T. Burners-Lee, J. Hendler, and O. Lassila, "The semantic web," Scientific American, vol. 284(5), May 2001.

[16] R. Guha, R. McCool, and E. Miller. "Semantic Search". In Proceedings of the 12th international conference on World Wide Web, pages 700–709, 2003.

[17] T.Finin, J.Mayfield, C.Fink, A.Joshi, and R.S. Cost," Information retrieval and the semantic web," in Proceeding of the 38th International Conference on System Sciences, Hawaii, USA, 2005.

[18] E. Mäkelä. "Survey of semantic search research". In: Proceedings of the Seminar on Knowledge Management on the Semantic Web, Department of Computer Science, University of Helsinki (2005)