IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-4 E-ISSN: 2347-2693

Ontology based Domain Specific Web Search Engine

Daksh Agrawal[1], Hirali Sanghani[2], Sonali Jadhav-3 and Supriya Shinde[4*]

[1,2,3,4*]*Department of Computer Science, Mumbai University, India*

www.ijcseonline.org

Received: Mar/30/2015 Revised: Apr/12//2015 Accepted: Apr/18/2015 Published: Apr/30/ 2015

***Abstract*—** Most of the existing search engines retrieve web pages by means of finding exact keywords. Traditional keyword-based search engines suffer several problems. First, synonyms and terms similar to keywords are not taken into consideration to search web pages. Users may need to input several similar keywords individually to complete a search [1]. Second, traditional search engines treat all the keywords as the same importance and cannot differentiate the importance of one keyword from that of another. Third, traditional search engines lack an applicable classification mechanism to reduce the search space and improve the search results. In this system, we develop a Semantic Search Engine. First, a fuzzy ontology is constructed by using fuzzy logic to capture the similarities of terms in the ontology, which offering appropriate semantic distances between terms to accomplish the semantic search of keywords. Second, users can check or uncheck the pages results based on their needs to show or hide it next time they search it. The totally satisfactory degree of keyword scam be aggregated based on their degrees of importance and degrees of satisfaction [2] [3]. Third, the domain classification of web pages offers users to select the appropriate domain for searching web pages, which excludes web pages in the inappropriate domains to reduce the search space and to improve the search results.

***Keywords*—** Information retrieval, Clustering, Semantic Web, Fuzzy ontology

1.0 INTRODUCTION

This project falls in the area of information retrieval and semantic web, and aims to improve the evaluation of web search tools. The huge number of information on the web as well as the growth of new inexperienced users creates new challenges for information retrieval; certainly the current search engines (such as Google, Bing and Yahoo) offer an efficient way to browse the web content. However, these types of tools do not take into account the semantics driven by the query terms and document words. This project proposes a new semantic based approach for the evaluation of information retrieval systems; the goal is to increase the selectivity of search tools and to improve how these tools are evaluated.

1.0.1 Purpose of the project

Nowadays the volume of the information on the Web is increasing dramatically. Facilitating users to get useful information has become more and more important to information retrieval systems. While information retrieval technologies have been improved to some extent, users are not satisfied with the low precision and recall. With the emergence of the Semantic Web, this situation can be remarkably improved if machines could "understand" the content of web pages. The existing information retrieval technologies can be classified mainly into three classes.

- The traditional information retrieval technologies are almost based merely on the occurrence of words in documents. It is only limited to string matching. However, these technologies are of no use when a search is based on the meaning of words, rather than on words themselves.
- Search engines are limited to string matching and link analysis. The most widely used algorithms are the PageRank algorithm and the HITS algorithm. The PageRank algorithm is based on the number of other pages pointing to the Web page and the value of the pages pointing to it. Search engines like Google combine information retrieval techniques with PageRank. In contrast to the PageRank algorithm, the HITS algorithm employs a query dependent ranking technique. In addition to this, the HITS algorithm produces the authority and the hub score.
- The widespread availability of machine understandable information on the Semantic Web offers some opportunities to improve traditional search. If machines could "understand" the content of web pages, searches with high precision and recall would be possible.

1.0.2 Scope of the Project

The user will have the option to search for the information as per the user needs. The results will be shown not only based on the keywords but also on the related pages relevant to the user. The users will get only the relevant information for what they are searching.
The activities for searching the information on the web by the user will be recorded for the maintenance and analysis

purpose. Based on this recorded information the user search related pages will be shown based on the time they have stayed on the visited pages previously.

1.0.3 Constraints of the Project

The major difficulty in implementing is the extra time spent to annotate web pages. It is possible that not all of the relevant web pages are annotated and therefore only part of them can be returned. As a result, there is not sufficient page volume.
Moreover, in contrast to a typical web search interface for keyword entering, the semantic search systems are more complicated. There has to be a way to choose detailed information such as ontology, category, and etc.
The search methodology will be only text based. So no images and audio mechanisms will be used in this search engine.

2.0 THE PROBLEM AREA

The exponential growth rate of Web-data leads to many complications in retrieval of relevant information. In addition to this, the navigation of many links in an attempt to find desired information cause wastage of user time and makes the user annoyed. Therefore, Web-searchers are introduced domain specific Web-search engines. But most of the cases search engines produced their main search results by appending title tag, URL and meta-tag information or first few words from the Web-page content. Now, few commonly used products such as book, mobile handset, medicine, jewellery, etc., where most of the cases Web searchers main intention to know some basic information about each domains. [8]

In general, our main goal is to identify the basic information from the search result for a Web-searcher. The example illustrates the problem in the search result, which can overcome by using our search model.
Example 1. John is looking for a mobile handset. He prefers a Samsung I5500 Galaxy handset, but preferably costing not more than $200. He planned to search with search string “Samsung I5500 Galaxy”. Unfortunately he has received lots of search results. He needs to open each and every search result links to get the basic information about the handset, which is not only time consuming, but also it deals with Web-page download cost.

3.0 OVERALL SYSTEM DESCRIPTION

3.0.1 Existing System

Google runs on a unique combination of advanced hardware and software. The speed you experience can be attributed in part to the efficiency of our search algorithm and partly to the thousands of low cost PC's we've networked together to create a superfast search engine.

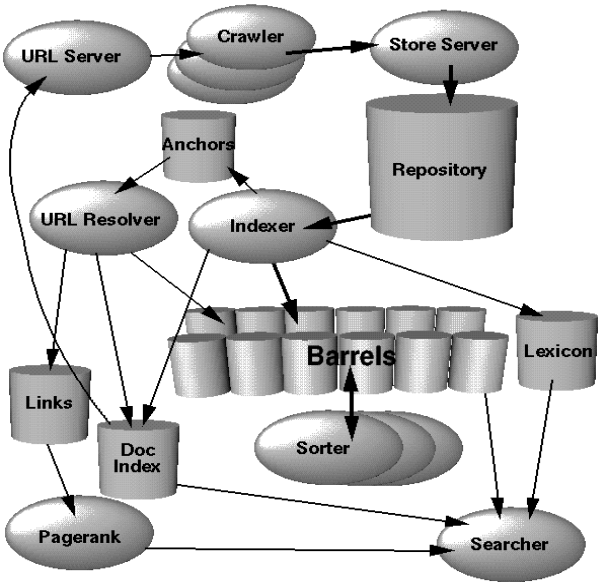

Fig. 3.0.1 Google Search Approach

Swoogle is a crawler based indexing and retrieval system for the semantic web. It extracts Meta data for each discovered documents and computes relation between documents

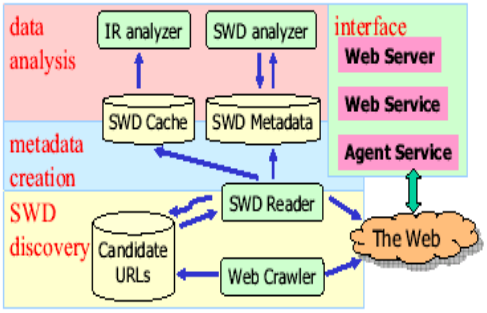

Fig. 3.0.2: Swoogle Search Approach

The functional process of this system is as follows:
The SWD discovery component discovers potential SWDs throughout the Web and keeps up-to-date information about SWDs.
The metadata creation component caches a snapshot of a SWD and generates objective metadata about SWDs at both the syntax level and the semantic level.

The data analysis component uses the cached SWDs and the created metadata to derive analytical reports, such as classification of SWOs and SWDBs, rank of SWDs, and the IR index of SWDs.

The interface component focuses on providing data services

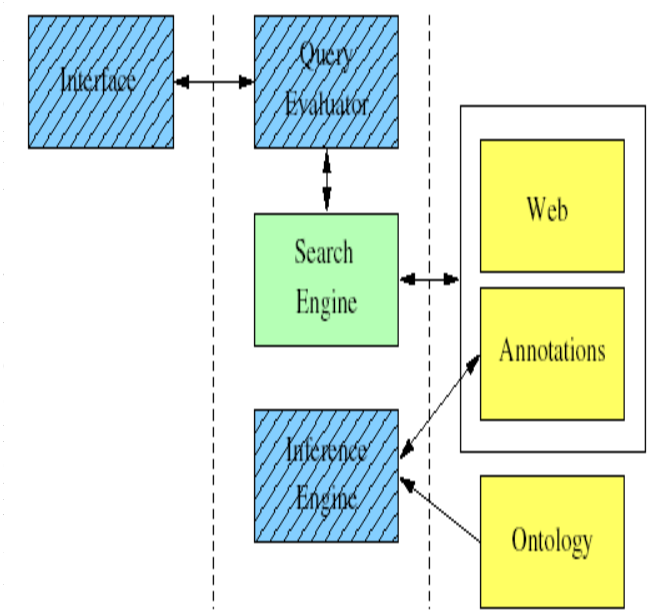

Fig. 3.0.2: Proposed System

Interface:
This is the user interface through which the user will enter keywords or phrases to search.
Query Evaluator:
This block filters out the keywords from the user entered phrases and generates the synonyms to it.

Search Engine:
It takes the keywords from the query evaluator and checks it in the web document for the relevant pages which is returned to the inference system.

Inference System:
Using background ontology Inference system adds all properties that can be deduced / induced from the ontology and returned to the web documents for other relevant pages.

Time Rank Mechanism:
We try to implement a Time Rank Mechanism for ranking the pages which user searches. This is a simple mechanism which ranks the pages based on the amount of time user has stayed on it previously. Higher the time, higher would be the rank of the page.

4.0 THE SEARCH MECHANISM:

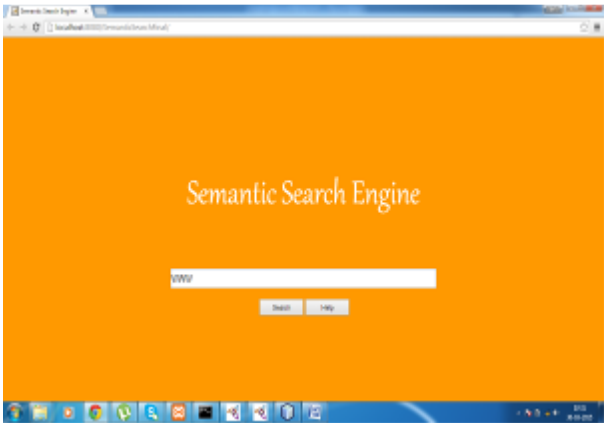

Fig. 4.1

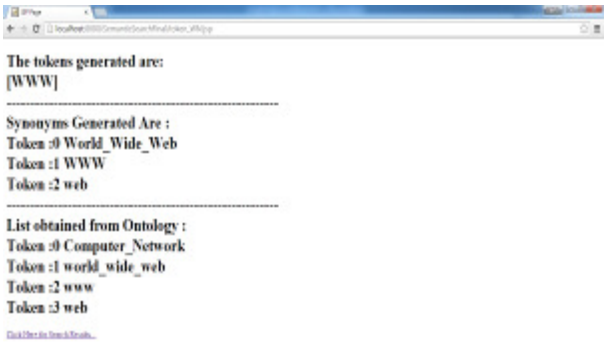

Fig. 4.2

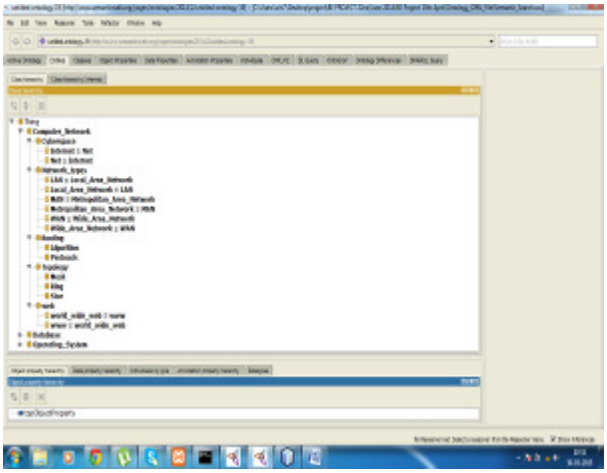

Fig. 4.3

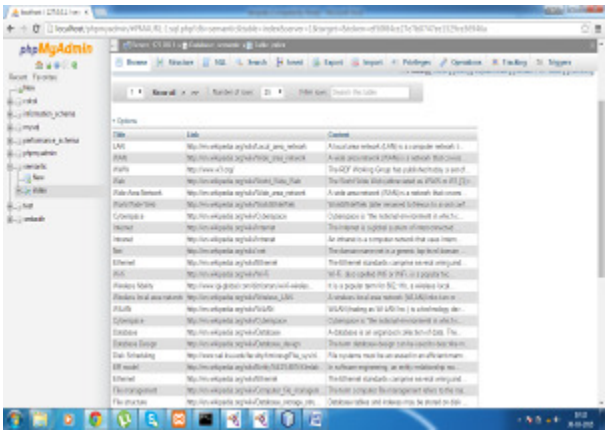

Fig. 4.4

In fig. 4.1 we have shown the start page of the search engine where the user can enter the keywords to be searched.

In fig 4.2 the keywords will be tokenized and passed through the Wordnet dictionary software and the synonyms will be retrieved and final list of keywords will be displayed.

In fig 4.3 the list of keywords will be passed through the ontology and the subclass, equivalence class and the super class will be retrieved and passed to the database created.

In fig 4.4 this is the database which contains all the URL's of the requested keywords by the used which will be retrieved and displayed in the final page.

5.0 ACKNOWLEDGMENTS

The authors wish to acknowledge the Department of Computer Engineering, Rajiv Gandhi Institute of Technology for their support and technical help throughout this work.

6.0 REFERENCES

[1] Lien-Fu Lai, Chao-Chin Wu, Pei-Ying Lin, "Developing a Fuzzy Search Engine Based on Fuzzy Ontology and Semantic Search". Dept. of Computer Science and Information Engineering National Changhua University of Education Changhua, R.O.C.

[2] en.wikipedia.org/wiki/Web_Ontology_Language

[3] en.wikipedia.org/wiki/Semantic_search

[4]gaia.isti.cnr.it/straccia./software/FuzzyOWL/index.html

[5] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum, 1981.

[6] P.T. Chang, K.C. Hung, K.P. Lin, and C.H. Chang, a Comparison of Discrete Algorithms for Fuzzy Weighted Average, IEEE Transactions on Fuzzy Systems, pp.:663-675, Oct. 2006.

[7] K.W. Church and P. Hanks Word Association Norms, Mutual Information and Lexicography, Computational Linguistics 16(1):22-29, Mar. 1990.

[8] D. Dubois and H. Prade. Fuzzy sets and systems: theory and applications. New York, London, 1980.

[9] L.F. Lai, C.C. Wu, M.Y. Shih, L.T. Huang, and W. Chiou. Parallel Processing for Fuzzy Queries in Human Resources Websites. Journal of Internet Technology, 7(11):943-953, Dec. 2010.

[10] Y.C. Lin, L.F. Lai, C.C. Wu, and L.T. Huang. A Self-Adaptation Approach to Fuzzy-Go Search Engine. The 2010 InternationalComputer Symposium (ICS 2010), pp. 1020-1025, Dec. 2010.