International Journal of Recent Trends in Engineering & Research
ISSN(ONLINE) : 2455-1457
IMPACT FACTOR : 3.344

Enhancing Google Search Engine using Semantic Data Search

Karthika K[1],Bhuvaneshwari B[2],Jaya Priya J[3] and Saravana Kumar K[4]
[1,2,3,4]*Information Technology,Kathir College of Engineering*

Abstract-The purpose of this project involves with search engine and search engine optimization methods. Here we are proudly introducing a new technique called as ontology search logs, which will be used for customized search logs according to the user define input. This application will be executed in any of the search engine. Here we are implementing this process in an offline system with some sample data sets. In this, we propose a new system called as Semantic Search log Social Personalized Search which would be able to give results for search query that relates to a particular user's surroundings, his area of interests, his likes and dislikes, the data the he/she might have found to be useful for him while searching.
Keywords - Information retrieval, Intelligent Search, Search Engine, Semantic web.

I. INTRODUCTION

The internet contains waste amount of information that the search engines are able to provide search results that are based on page ranks[1]. But the search results are not related to one particular user's environment. In this project, we introduce a new system called as Semantic Search log Social Personalized Intelligent Search[2]. It would be able to give results for search query that relates to a particular user's environment, his area of interests, his likes and disfavour, the data the he/she might have found to be useful for him while searching. Social networks are the domain in which we could obtain such user oriented information, which we use could use for providing personalized search results. We use governed learning technique to learn about the user based upon his interactions inside the system. This process can be able to create applicable for each and every registered users in this application. User can give their basic data in their profile and get benefits from their each and every search.

When the user getting register with the system it makes an ontological profile, according to the profile created by the user and when he/she obtaining login into the social network and communicates with it the system updates his/her ontological profile based upon the interaction. The search provision can be determining out in their home page after they get login. When the user searches a keyword using the search engine inner part of the social network, it refers to the ontological profile of the user and displays the customized Search results. The system should be able to intelligently identify whether a search result has been helpful to him or not and save it for his future reference when he searches for the same or similar keyword next time.

II. RELATED WORK

Information retrieval by thoroughly scrutinizing information on the web is not a fresh idea but has differentchallenges when it is compared to general information retrieval. Different search engines return different thoroughly scrutinize results due to the variation in indexing and search process. Google have been out there which handles the queries after processing the keywords. They only thoroughly scrutinizing information given on the web page, recently, some research group's begin delivering results from their semantics based search engines, and however most of them are in their initial steps. Till none of the search engines come to close indexing the entire web content, much less the entire Internet. When the information was distributed in web, we have two types of research problems in search engine i.e.

- How can a search engine map a query to documents where information is obtained but does not retrieve in intelligent and meaning full information?
- The query results build by search engines are distributed across different documents that may be connected with hyperlink. How search engine can identify efficiently such a distributed results?

Semantic web, can solve the first problem in web with semantic annotations to make intelligent and meaningful information by using query interface mechanism and ontology's. Other one can be answered by the graph-based query models.

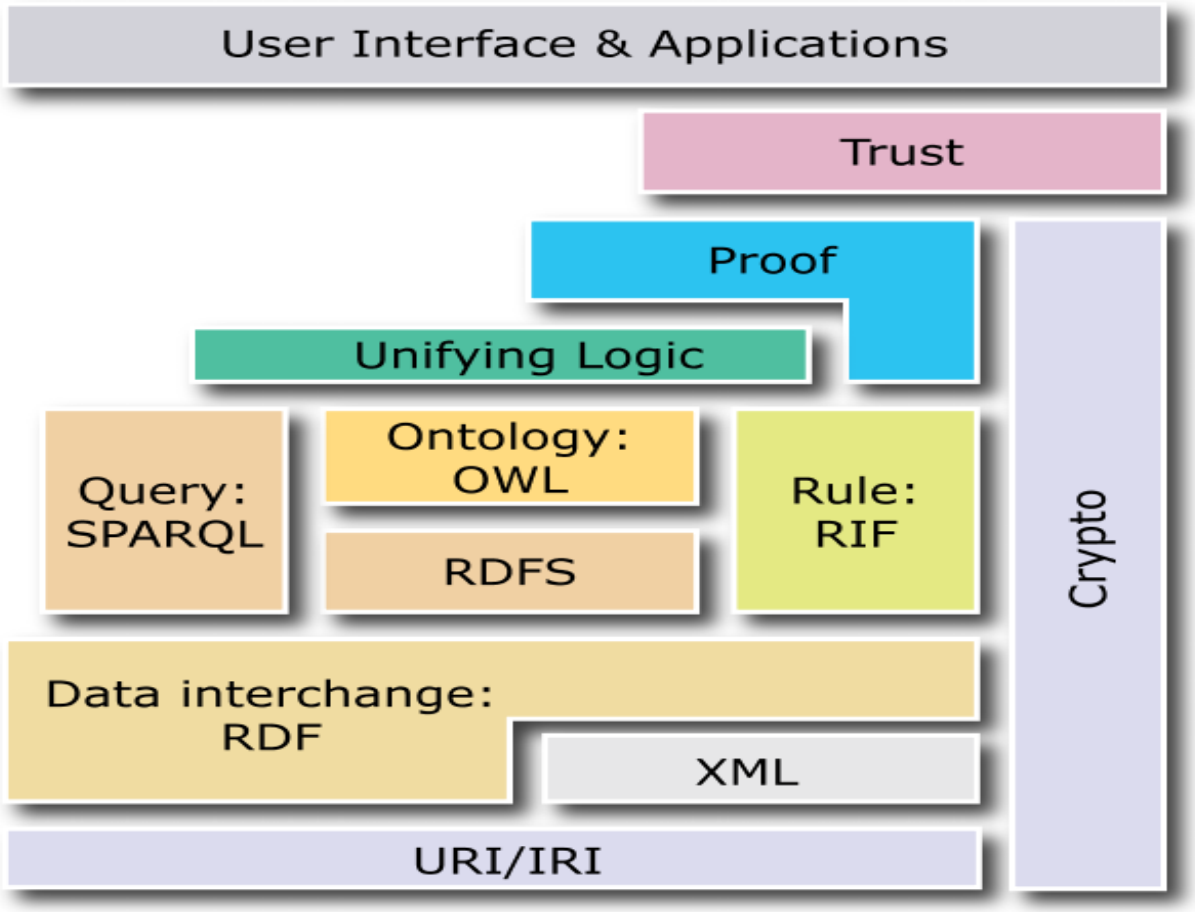

Fig .1. Semantic Web Frame Work

III. EXISTING SYSTEM

In a basic search engine, the web crawling is done by various distributed crawlers. There is a URL server that sends lists of URLs to be conveyed to the crawlers. The web pages that are transported are then sent to the store server. The store server then compresses and reserves the web pages into a repository. Every web page has an associated ID number called a doc ID which is allocated whenever a new URL is parsed out of a web page. The indexing function is performed by the indexer and the sorter. The indexer executes a number of functions. It reads the repository; uncompressed the documents, and parses them. Each document is changed into a set of word occurrences called hits. The hits record the word, position in document, an estimation of font size, and capitalization. The indexer distributes these hits into a set of "barrels", making a partially sorted forward index.

The indexer executes another important function. It parses out all the links in every web page and reserves important information about them in an anchors file. This file contains enough information

to find where each link points from and to, and the text of the connection. The URL resolve reads the anchors file and transforms relative URLs into absolute URLs and in turn into doc IDs. It puts the anchor text into the forward index, connected with the doc ID that the anchor points to. It also makess a database of connections which are pairs of doc IDs. The connections database is used to calculate Page Ranks for all the documents. The sorter extracts the barrels, which are arranged by doc ID, and resorts them by word ID to generate the inverted index. This is done in place so that little non-conformed space is important for this operation. The sorter also produces a list of word IDs and offsets into the inverted index. In this existing system has some drawbacks. That are,

- No user defined Login.
- Basic search engine should not produce a modified search.
- Denial of search attacks will not properly implement so that the hit ratio of the websites are not function correctly.
- Search engine produces more junk or advertisement websites from the top level of the search.
- Updates will be possible only through the admin web server of the search engine.
- No ontology profile is implemented.

IV. PROPOSED SYSTEM

Running a web crawler is a demanding task. There are tricky performance and reliability issues and even more powerfully, there are social issues. Crawling is the most powerful application since it involves interacting with hundreds of thousands of web servers and various name servers which are all past the control of the system.

Fig.2.Request and Response in Semantic search using Ontology

In order to scale to hundreds of millions of web pages, Google has a quick distributed crawling system. A single URL server serves numbers of URLs to a number of crawlers. Two of the URL server and the crawlers are implemented in Python. Each crawler keeps roughly 300 connections open at once. This is powerful to retrieve web pages at a fast enough pace. At peak speeds, the system can crawl 100 above web pages per second using four crawlers.

This amounts to roughly 600K per second of data. A major action stress is DNS lookup. Each crawler maintains its own DNS cache so it does not need to do a DNS lookup before crawling each document. Each of the hundreds of links can be in a number of different states: looking up DNS, linking to host, sending request, and receiving reply. These factors make the crawler a complex component of the system. It uses asynchronous IO to handle events, and a number of queues to move page fetches from state to state.But this issue had not arise until we had downloaded tens of millions of pages. Because of the immense variation in web pages and servers, it is virtually very difficult to test a crawler without running it on large part of the Internet. Invariably, there are hundreds of obscure issues which may only occur on one page out of the whole web and cause the crawler to crash, or worse, cause unpredictable or wrong behaviour. Systems which access large parts of the Internet need to be designed to be very robust and meticulous tested. Since large complex systems such as crawlers will invariably cause problems, there needs to be important resources devoted to

reading the email and solving these problems as they come up. Proposed system have some Advantages. That are

- Most of the complexity will be run through the proposed system, using the web crawler and DNS.
- User Login will be possible in this search engine.
- Customized search is possible according to the ontology web search logs.
- So that user can get their perfect search information according to their profile information.
- No advertisement sites will be pushed up in search.
- Frequently viewing website will have more priority in their every search.

V. CONCLUSION

The design and implementation, as well the analysis, of efficient and effective Web Search Engines (WSEs), are becoming more and more powerful as the size of the Web has continually kept growing. Furthermore, the development of systems for Web Information Retrieval represents a very interesting task whose complexity imposes the knowledge of several concepts coming from many other areas: databases, parallel computing, artificial intelligence, statistics, etc.Thus we have implemented ontology concept successfully as per we committed in the introduction chapter. Here the testing was done on the local host successfully. As per our future enhancement this concept can be implemented in web server or else in any server like cloud, grid and etc. Thus the thesis explains the trends, threads and process of the Semantic web using Ontology, which was implemented in search engine in successfully manner.

REFERENCES

1. D. Zhang, Y. Chee, A. Mondal , A. Tung, and M. Kitsuregawa, “Keyword search in spatial databases: Towards searching by document, ”in Proc. IEEE Int. Conf. Data Eng., 2009, pp. 688–699.
2. D. Felipe, V. Hristidis, and N. Rishe, “Keyword search on spatial databases,” in Proc. IEEE 24th Int. Conf. Data Eng.,2008, pp. 656–665.

