© 2016 IJSRSET | Volume 2 | Issue 1 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

# Prototype for Enhancing Search Engine Performance Using Semantic Data Search

**Divya. J, C. Senthil Kumar**

Department of Information Technology, SNS College of Technology, Coimbatore, Tamil Nadu, India

## ABSTRACT

Information's on Internet are vast that are retrieved by the search engines based on page ranks. But the search results are not related to one particular user's environment. Many researches had been possessed to provide better results. In this project, we propose a new system called as Semantic Search log Social Personalized Search which would be able to provide results for search query that relates to a particular user's environment based on the users area of interests, his likes and dislikes etc.., Social networks are such domain in which we could obtain the user oriented information, which can be used for providing personalized search results. Here a supervised learning technique is used to learn about the user, based upon his interactions inside the system. This process can be able to make applicable for each and every registered user in this application. This can be done by proving the user basic information in their profile and get benefits from their each and every search.

When the user gets register with the system, it creates an ontological profile, when the user gets login into the social network and interacts with it the system updates the user ontological profile based upon their interaction. The search provision can be finding out in their home page after they get login. When the user searches a keyword using the search engine inside the social network, it refers to the ontological profile of the user and displays the Personalized Search results. The system should be able to intelligently identify whether a search result has been useful to him or not and save it for his future reference when he searches for the same or similar keyword next time. The main objective of this project involves with search engine and its optimization methods. A new technique called as ontology search logs is introduced, which will be used for customized search logs according to the user's define input based on his/her area of interests, his/her likes and dislikes,. This application will be processed in any type of the search engine.

**Keywords :** Search Engine, Ontology technique, Semantic Search, Google search, web crawler.

## I. INTRODUCTION

The traditional adaptive web focuses on personalized services tailors the specific needs of individual users based on the user context along with device and environment properties and also according to the content recommendation, navigation adaptation and presentation customization that unsuit's for increasing diverse customers[1,2].Figure 1 shows the process a search engine.

The Semantic Web is as process which is an extension of the current web which displays information in a well-defined manner, meaning and allows people to work in co-operation .The study on the categories of things that exist or may exist in some domain constitutes ontology in which a catalog of things that are assumed to exist in a domain of interest from the perspective of a person and their language for the purpose of conversing about domain of interest. It provides advanced machine processing of information which ultimately improves the application interoperability, data integration, sharing and

availability. It represents the predicates, concept and relation types of the language based on the discuss topics in the domain. Ontology drives based on the growing end user expectations and requirements as well as commercial considerations. At present web is not used only by humans, software agents are also becoming users of the web too which led to the development of the semantic web [3, 4]. Information retrieval technology can draw massive benefits from using semantic web vision. Ontologies represents relationships between concepts which in turn improve search results [5, 6, 7] makes ontologies to rank semantic web which is one of the motivations of the Semantic web vision that has been subjected to many researches [8, 9, 10] who founds enhance search process either statically or dynamically [11].

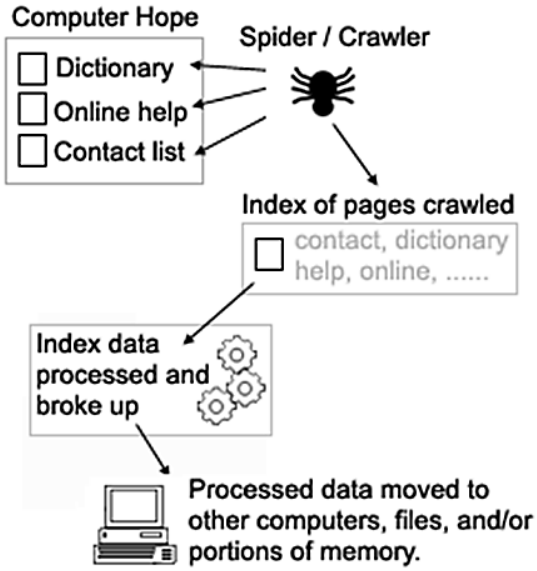



**Figure 1 :** Displays the work of a search engine.

The main objective of this project involves with search engine and its optimization methods. A new technique called as ontology search logs is introduced, which will be used for customized search logs according to the user's define input based on his/her area of interests, his/her likes and dislikes,. This application will be processed in any type of the search engine. The process is implemented in an offline system with some sample data sets.

## II. METHODS AND MATERIAL

### A. Literature Survey

Personalized systems are developed to help users to find relevant information based on the key search. The main challenge of effective personalization systems is to accurately identify the user search interests. In most personalized techniques the user's preferences will be in the form of a user profile or personal profile. At presentCross-enterprise collaboration [18,19] is one of challenges on the business-to-business integration research with sematic search. In [12] the user explicitly selects the categories fits best of the user interests and the user profile is then defined by storing the whole path based on each category of interest. When a new query is issued Search results are classified according and then the distance between the hierarchical structure of the user profile and the results' topics are calculated which re-rank results. This method lacks in maintaining the user's changing preferences.

Pedrinaci et al [15] reviews Business process management (BPM) concept and also illustrates SBPM approach which have been developed to enhancing existing BPM solutions in order to achieve more flexible, manageable and dynamic business processes.

Comparison Table with different methods:

| S.NO | TITLE | METHOD | ADVANTAGES | DISADVANTAGES |
|---|---|---|---|---|
| 1 | Semantic business process integration based on ontology Alignment [13] | Supervised Learning Technique | Solving some linguistic mismatch between ontology | Ontology merging and alignments are difficult task |
| 2 | A Conceptual Framework for Dynamic Cross Enterprise Collaboration[14] | SBPM (Semantic Business Process Methodology) and BizKB | It is more flexible, dynamic and manageable business process | BPM suffers from a lack of automation |

| | | | | |
|---|---|---|---|---|
| | approach | | | |
| **3** | An Ontological Approach for Dynamic Cross-Enterprise Collaboration [16] | OHTN (Ontological Hierarchical Task Network) | Framework can quickly access customer need and provide the collaborative models to meet need of customers | Web service discovery has to be improved |
| 4 | Genesis – Dynamic collaborative business process formulation based on business goals and criteria[17]. | BOWL & Genesis with CBP'S | Genesis traverse through the ontology and dynamically produces required output | Users business goals are not fulfilled here |

### B. Problem Definition

The existing system drawbacks are:

- No user defined Login.
- Basic search engine should not produce a customized 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.

### C. Proposed System

According to the proposed system the web crawler will be concentrated more. This proposed system will be furnished in technical as follows:

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

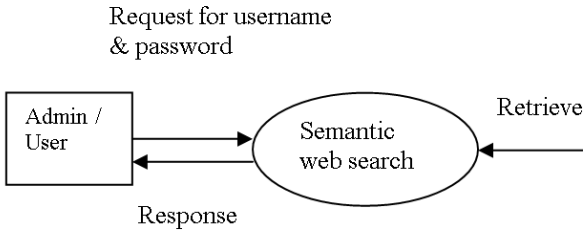
In order to scale to hundreds of millions of web pages, Google has a fast distributed crawling system. A single URL server serves lists of URLs to a number of crawlers. Both the URL server and the crawlers are implemented in Python. Each crawler keeps roughly 300 connections open at once. This is necessary to retrieve web pages at a fast enough pace. At peak speeds, the system can crawl over 100 web pages per second using four crawlers. This amounts to roughly 600K per second of data. A major performance 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 connections can be in a number of different states: looking up DNS, connecting to host, sending request, and receiving response. These factors make the crawler a complex component of the system. It uses asynchronous IO to manage events, and a number of queues to move page fetches from state to state.

But this problem had not come up until we had downloaded tens of millions of pages. Because of the immense variation in web pages and servers, it is virtually impossible to test a crawler without running it on large part of the Internet. Invariably, there are hundreds of obscure problems which may only occur on one page out of the whole web and cause the crawler to crash, or worse, cause unpredictable or incorrect behavior. Systems which access large parts of the Internet need to be designed to be very robust and carefully tested. Since large complex systems such as crawlers will invariably cause problems, there needs to are significant resources devoted to reading the email and solving these problems as they come up.

Advantages of proposed system:

- 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.

## D. Methodology




**User creation:**

This is the initial module of this project. Here the user can create their account in order to register their details into semantic logs. While creating the account user should provide their security information and they can create their own username and password.

A Social network is implemented here for the user sharing purpose in alone with user can create their friends group. This user profile creation will be done automatic admin authentication.

**Ontology Search Engine Updates**

The Ontology Search Engine updates module is involving with admin process where the data center is available with huge number of data sets according to the user define search. An Ontology table structure is created in order to update the data which will be retrieved according to the user define search. A site will be updates in the data set with the prior http link, Data content and hit ratio.

**Profile Search Engine Optimization**

The Profile Search Engine Optimization is an advanced search engine technique, which will be get optimize according to the user internal search. Basically a search engine will search according to the Meta data as well as according to the hit ratio. Here the database is designed according to the user define search logs to that retervial will not be more complex.

**Semantic Web search**

This is the core module of this project where user can interact with the semantic web search , Here user will be get login with their corresponding user name and password. Initially the search will be done according to the profile information of the user, a ontology interface process is implemented here in order to get the relevance feedback information from the user define output.

**Search log contents**

Here initially the user search content information will be displayed according to the current given output. But if in case of user searching ay of their related links the searched links will be displayed at the top most searches from their next search.

Comparison Table with Google and Semantic Search

| **GOOGLE** | **SEMANTIC SEARCH** |
|---|---|
| 1. It is one of traditional search engines that produce results of given query within the given context. | 1. It is one of semantic search engines that work on Semantic based approach which is useful for having deep information about the given query |
| 2. The information retrieved by Google is dependent on keywords, phrases or predefined algorithms that lead to spam results. | 2. The information retrieved by semantic is independent of these keywords and predefined algorithms that produce exact results instead of spam oriented results |
| 3. It uses HTML, XML language for creation of metadata. | 3. It uses Semantic Web languages like OWL, RDF for creation of metadata |
| 4. It does not focus on stop words like **is, or, and**, **how** because it does not produce exact results what user is looking for. Omitting these words will not affect results. | 4. It focuses on stop words and punctuation marks because it takes into account each and every small character as it affects search results. |
| 5. Google does not give suggestions until we press Search button. | 5. Semantic presents some suggestions before pressing Search button. |
| 6. It displays all web pages that may or may not satisfy user‘s query and to select relevant page from many pages is difficult task. | 6. It will show only those results that will answer our query. |
| 7. It makes use of keywords to expand query instead of using any methodology. | 7. It uses Fuzzy logic to expand query. Fuzzy logic is a problem solving methodology that uses original queries to produce accurate results. |
| 8. It does not highlight any words or phrases | 8. It highlights the sentences or words that give |

| | |
|---|---|
| that are most useful in answering query. | answer to user query. |
| 9. It does not offer any option to see users who are searching for same information. | 9. It uses an app **—Meet Others‖** that allows various users to come together and discuss important issues. |
| 10. It uses Page Rank algorithm that is used to rank search results in order of popularity of page. | 10. It uses Semantic Rank algorithm that is used to rank search results in order of content of page in response to given query. Better the content, better the rank of page. |

## III. RESULTS AND DISCUSSION

**Figure 2 :** User Registration checking username availability

The index page is displayed from which the search engine is created. Figure 2 shows user registration checking the username availability. Here the user must provide their details and get registered. User name availability will be checked once the user has been registered. After the details has been provided by the users the registration process will be completed, and the users profile will be successfully created.


**Figure 3:** Login Validation Checking

Figure 3 shows the validation checking or authentication for login. The user can now login by providing user name and password. The login validation checking will be done in this stage, if the user name and password is valid, the user home page will be displayed to the user.

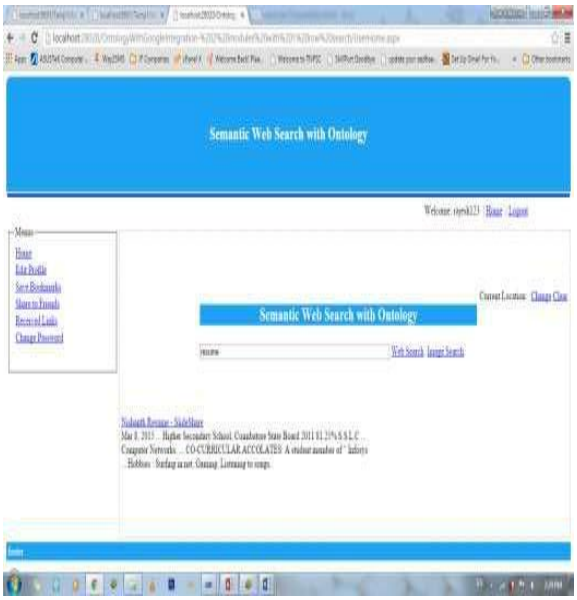



**Figure 4:** Profile Location based search

This form searches the information based on the location provided by the user. If the user needs to change the location he/she as already provided, he/she can update their current location with the location change dialog box.


**Figure 5:** Location Changed

Now the location has been changed according to the user provide location. Keywords can be given to get the particular information in the search bar. 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.

## IV. CONCLUSION

An effective and flexible distributed scheme is proposed with explicit dynamic data support that is used to ensure the correctness of Semantic web search engine optimization methods. We rely on erasure correcting code in the web distribution preparation which provide redundancies and guarantee the data dependability. This process drastically reduces the communication and storage overhead when compared to the traditional replication-based file distribution techniques. By utilizing the ontology techniques with distributed verification of erasure-coded data, the scheme achieves i) storage correctness insurance and ii) data error localization. Whenever data corruption has been detected while storing correctness verification, the proposed scheme can almost guarantee the simultaneous localization of data errors, i.e., the identification of the misbehaving server.

It is concluded that the application works well and satisfy the owner and customers. The application is tested and errors are properly debugged. When the site is simultaneously accessed from more than one system then it is monitored and tested. The application works according to the restrictions provided in the respective browsers. The speed of the transactions is increased efficiently.

## V. SCOPE OF FUTURE WORK

The system can be further enhanced by adding new features and facilities. Now the system is platform dependent and it can be made as platform independent software.

In added with the future system will be developed in cloud architecture with the 3 components called as Software as a service (SaaS), Platform as a services (PaaS), Infrastructure as a services (Iaas). In case of implementing in a Saas, this will be featured in an enhanced content manager like ebay, amazon and most of the commercial web searched.

As implemented in a platform as a services it can be implements as a browser based application such as implementation will be possible in linux, mac, Solaris and etc. User define customized process also be implanted in case of the updating of data in any web server.

## VI. REFERENCES


[1] Brusilovsky, P.: Adaptive Hypermedia. User Model. UserAdapt. Interact., Vol. 11, 2001, No. 1–2, pp. 87–110.

[2] Brusilovsky, P.—Kobsa, A.—Nejdl, W. Eds.: The Adaptive Web, Methods and Strategies of Web Personalization. LNCS 4321, Springer 2007.

[3] 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.

[4] T. Burners-Lee, J. Hendler, and O. Lassila, “The semantic web,” Scientific American, vol. 284(5), May 2001

[5] 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)

[6] Ramprakash, S. K. Malik, N. Prakash, S. Rizvi , "A Comparative Study of Different Types of Search Engines in Context of Semantic Web”, National Conference on Advancements in Information & Communication Technology (NCAICT) , on March 15-16, 2008.

[7] W.A. PINHEIRO, A. Maria , C. Moura. "Semantic Search in Portals using Ontologies". I Workshop de Web Semântica (WWS2004), Brasília, 22 de outubro de 2004.

[8] C.Patel, K.Supekar, YLee, and E.Park, ”Ontokhoj: A semantic web portal for ontology searching ,ranking and classification ,”in Proc.of ACM 5th International Workshop on Web Information and Data Management (WIDM),New Orleans, pp.58-61,2003.

[9] E. Thomas, J. Z. Pan, D. H. Sleeman. "ONTOSEARCH2: Searching Ontologies

Semantically". In Proc. of OWL Experiences and Directions Workshop, 2007.

[10] H. Alani, N. Noy, N. Shah, N. Shadbolt, M. Musen "Searching Ontologies Based on Content: Experiments in the Biomedical Domain In: 4th International Conference on Knowledge Capture.ACM Press; p.55–62, 2007.

[11] D. Taibi, M. Gentile, L. Seta. "A semantic search engine for learning resources". Third International Conference on Multimedia and Information & Communication Technologies in Education.2005.

[12] Chirita, P. Nejdl, W. Paiu, R. and Kohlschutter, C. (2005) Using ODP Metadata to Personalize Search. Proc. SIGIR '05, Salvador, Brazil, 15-19 August, pp. 178-185, ACM Press, New York, NY, USA.

[13] Jason J. Jung," Semantic business process integration based on ontology alignment" Sciencedirect, Volume 36, Issue 8, October 2009, Pages 11013–11020.

[14] N.T. Nguyen, R. Kowalczyk and S.M. Chen," BizKB: A Conceptual Framework for Dynamic Cross-Enterprise Collaboration", Springer ICCCI 2009,LNASI 5796,pp. 401-4012,2009.

[15] Pedrinaci, C. ; Knowledge Media Inst., Open Univ., Milton Keynes ; Domingue, J. ; Brelage, C. ; van Lessen, T. "Semantic Business Process Management: Scaling Up the Management of Business Processes ",Semantic Computing, 2008 IEEE International Conference on Aug. 2008, pp:546 – 553