Implementation of Intelligent Semantic Web Search Engine

Dinesh Jagtap[*1], Nilesh Argade[1], Shivaji Date[1], Sainath Hole[1], Mahendra Salunke[2]

[1]Department of Computer Engineering,
[2]Associate Professor, Department of Computer Engineering

Sinhgad Institute of Technology and Science,

Pune, Maharashtra, India 411041.

Abstract—**Search engines are design for to search particular information for a large database that is from World Wide Web. There are lots of search engines available. Google, yahoo, Bing are the search engines which are most widely used search engines in today. The main objective of any search engines is to provide particular or required information with minimum time. The semantics web search engines are the next version of traditional search engines. The main problem of traditional search engines is that information retrieval from the database is difficult or takes long time. Hence efficiency of search engines is reduced. To overcome this intelligent semantic search engines are introduced. The main target of semantic search engines is to give the required information within small time with high accuracy. Many search engines will provide result from blogs or various websites. The user can not have a trust on the results because the information on blogs or websites is does not necessarily true. For this purpose we use xml meta-tags and its features .The xml page will contain built in and user defined tags. The metadata info of the pages expected from this XML into resource description framework (RDF).**

Keywords—Intelligent Search, RDF, Semantic Web, XML.

I. INTRODUCTION

The semantic web is next version of traditional web which consisting of well-defined database that understood by users. The semantic web is described by using W3C standard called resource description framework (RDF).

Ontology is one of the most important terms used in semantic web .the ontologies can be represented using rdf(s) and owl of w3c recommended data representation models. Some basic features of semantic web are efficient information retrieval, automation, integration & reusability of information.

The traditional web search engine will not cover the point of trustfulness or reliability. For example: for a particular user search a query likes 'which is the best engineering college in Pune?' the search engine will give thousands of results to user but it's very hard for user to find which information is reliable.

In this paper we propose web based search engine which is called as intelligent semantic web search engines. Here we use xml meta- tags and its features. The xml page will include built in and user defines tags. The metadata info is generated from this xml into RDF .The RDF graph are populated by inputting through x forms.

II. BACKGROUND

The idea of search engine and info retrieval from search engine is not a new concept. The interesting thing about traditional search engine is that different search engine will provide different result for the same query. While information was available in web, we have some fields of problem in search engines. Information retrieval by searching information on the web is not a fresh idea but has different challenges when it is compared to general information retrieval. Different search engines return different search results due to the variation in indexing and search process. Google, Yahoo, and Bing have been out there which handles the queries after processing the keywords. They only search information given on the web page, recently, some research group's start delivering results from their semantics based search engines, and however most of them are in their initial stages. Till none of the search engines come to close indexing the entire web content, much less the entire Internet.

1. Many times this happened that the particular result is available on the web but due to not availability of intelligent retrieval system.

2. The another main program with search engine is result that contain information will scattered in different pages, so there is need of hyperlinking of these pages.
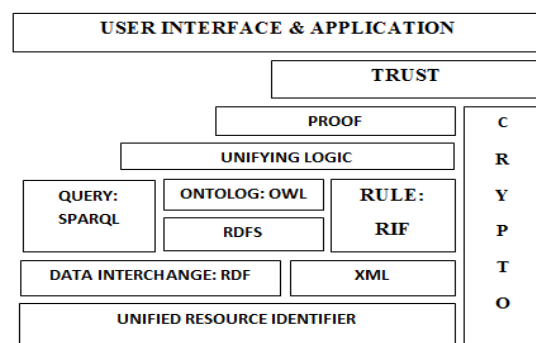

Fig. 1. Semantic Web framework.

Semantic is the process of communicating enough meaning to result in an action. A sequence of symbols can be used to communicate meaning, and this communication can then affect behaviour. Semantics has been driving the next generation of the Web as the Semantic Web, where the focus is on the role of semantics for automated

approaches to exploiting Web resources. 'Semantic' also indicates that the meaning of data on the web can be discovered not just by people, but also by computers. Then the Semantic Web was created to extend the web and make data easy to reuse everywhere.

III.   SEMANTIC WEB SEARCH ENGINE

Currently many of semantic search engines are developed and implemented in different working environments, and these mechanisms can be put into use to realize present search engines.

Alcides Calsavara and Glauco Schmidt propose and define a novel kind of service for the semantic search engine. A semantic search engine stores semantic information about Web resources and is able to solve complex queries, considering as well the context where the Web resource is targeted, and how a semantic search engine may be employed in order to permit clients obtain information about commercial products and services, as well as about sellers and service providers which can be hierarchically organized. Semantic search engines may

Sara Cohen Jonathan Mamou et al presented a semantic search engine for XML (XSEarch).It has a simple query language, suitable for a naïve user. It returns semantically related document fragments that satisfy the user's query. Query answers are ranked using extended information-retrieval techniques and are generated in an order similar to the ranking. Advanced Indexing techniques were developed to facilitate efficient implementation of XSEarch. The performance of the different techniques as well as the recall and the precision were measured experimentally. These experiments indicate that XSEarch is efficient, scalable and ranks quality results highly.

Bhagwat and Polyzotis propose a Semantic-based file system search engine- Eureka, which uses an inference model to build the links between files and a File Rank metric to rank the files according to their semantic importance. Eureka has two main parts:

a) Crawler which extracts file from file system and generates two kinds of indices: keywords' indices that record the keywords from crawled files, and rank index that records the File Rank metrics of the files

b) When search terms are entered, the query engine will match the search terms with keywords' indices, and determine the matched file sets and their ranking order by an information retrieval based metrics and File Rank metrics.

Wang et al. project a semantic search methodology to retrieve information from normal tables, which has three main steps: identifying semantic relationships between table cells; converting tables into data in the form of database; retrieving objective data by query languages. The research objective defined by the authors is how to use a given table and a given domain knowledge to convert a table into a database table with semantics. The authors' approach is to denote the layout by layout syntax grammar and match these.

IV.   CURRENT WORKS AND LIMITATION

In today's web "World Wide Web" is a world wide database which causes the lacks of existing of semantic structure. The traditional web search engine returns ambiguous or partially ambiguous results. We used the semantic search engine to overcome these problems.

The semantic search engine is available today is "Hakia." Hakia calls itself a "meaning based search engine." They are providing results based on query matching rather than by popularity. Semantic search uses the technologies semantic web and search engine to improve the search results obtained by current search engine and evolves to next generation of search engine built on semantic web.

In general processes of semantic search engine are:-
A. The user question is interpreted, obtaining the relevant concept from the sentence.
B. That set of concept is used to build a query that is launched against the ontology.
C. The results are presented to the user.
.
The overall structure of this search engine is complex. It gives the many advanced options like reefing, sorting and saving the search. The search results are very easy to navigate.

Hakia is widely used semantic search engine that work like Wikipedia. Hakia calls itself a meaning based search engine. The main goal of these search engines is that they provide search results based on meaning match rather than by popularity of search query. The current news, blogs are processed by hakia's proprietary are semantic technology called QDEXing. It will process any query by its semantic rank technology Sense Bot represent a new type of search engine that prepares a text summary in response to users search query. Sense Bot extracts most relevant results using semantic web technologies from the web. It then summarizes results together for the user as per topic. It uses different text mining algorithm to parse web pages which leads to identification of key semantic concept.

By going through the literature analysis of some of the important semantic web search engines, it is concluded that each search engine has some relative strengths. A summary is given in the below which summarizes the techniques, advantages of some of the important semantic web search engines that are developed so far. XSEarch involves a simple query language, suitable for a naïve user. It returns semantically related document fragments that satisfy the user's query. Query answers ranked using extended information-retrieval techniques and are generated in an order similar to the ranking. Advanced indexing techniques were developed to facilitate efficient implementation of XSEarch[8].The performance of the different techniques as well. XCDSearch is a context-driven search engine. It uses Keyword-based queries as well as loosely structured queries, using a stack-based sort-merge algorithm. It employs Object-Oriented techniques for answering queries. The keyword query is answered by returning a sub graph that satisfies the search keywords[5] It builds the relations between data elements based solely on their labels and proximity to one another, while overlooking the contexts of

the elements, which may lead to erroneous results. It employs stack-based sort-merge algorithm employing context driven search techniques for determining the relationships between the different unified entities. Swoogle is a crawler-based indexing and retrieval system for the Semantic Web documents. It analyzes the documents it discovered to compute useful metadata properties and relationships between them. The documents are also indexed by using an information retrieval system which can use either character N-Gram or Uri's as terms to find documents matching a user's query or to compute the similarity among a set of documents. One of the interesting properties computed for each Semantic Web document is its rank -a measure of the document's importance on the Semantic Web.

V. PROPOSED SYSTEM

The problem mention in previous section related to semantic web search can be solved by maintaining metadata repository for the pages that contain domain knowledge from trusted sources. In this work our search engine first searches for web pages and the gets the result by searching for the metadata. The metadata recording could either be manual or automated. The manual system requires input of information from the administrator of website.

A. Proposed approach for semantic web search engine

The interoperability issue can be resolved by using W3C tools. For representing domain knowledge, W3C proposes ontology's in OWL, while metadata can be represented in graphs as RDF triples. The major part in our semantic web documents is the instances of ontology. These instances are represented as metadata that contains information about the large web pages. We used W3C based tools to ensure semantic interoperability.
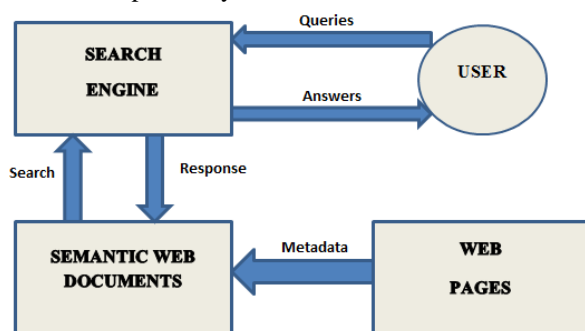

Fig. 2. Design Architecture of Intelligent Semantic Web Search Engine Framework

B. Issues in Search Engine

1) High Recall but Low precision: We not able to say in all cases the semantic technology show its performance due to low precision in mapping concepts (pages) and unwanted recall is high. For example the semantic search engine Ding is using the Google results and form a metadata for the top results obtained from Google in several test cases the semantic produce low precision but high in recall. Hence more optimized mapping required to improve precision.

2) Identify User Intention: For any semantic search engine it must act upon what the user is intend to search, using its intelligence it produce the result more relevant to user what he thought to search. For example the analysis has been made in research how to match user intention and Semantic search hence it produce most suitable results for user.

3) Inaccurate Quires: The user who search in search engine only haves domain specific knowledge so he give the keyword or sentence. In search space the user will not include any synonyms or potential variation in query, that exactly matches results here the user have a problem but aren't sure how to phrase.

4) Efficiency of Crawler: The World Wide Web got trillions of distributed information for the topic based search engines it only extract the topic in pages and produce a result to user. But once we go for semantic based search engines it is capable of making multiple choices for single user keyword. Second thing is it forms metadata key based search for processing each web documents or pages so single crawler is not sufficient to do the task.

VI. EXPERIMENT AND RESULT

In this paper we implemented proposed system with intelligent semantic search engine

Fig. 3. Homepage of Proposed System [GUI].

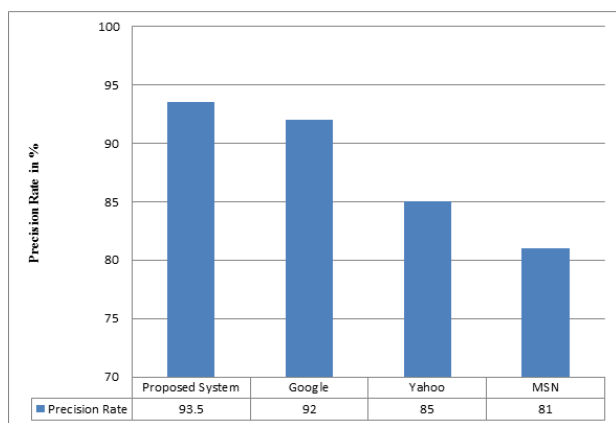below chart shows the results for precision rate

Fig. 4.   Precision Rate Graph.

Precision rate:-precision is fraction of retrieved document that are relevant to searched result.
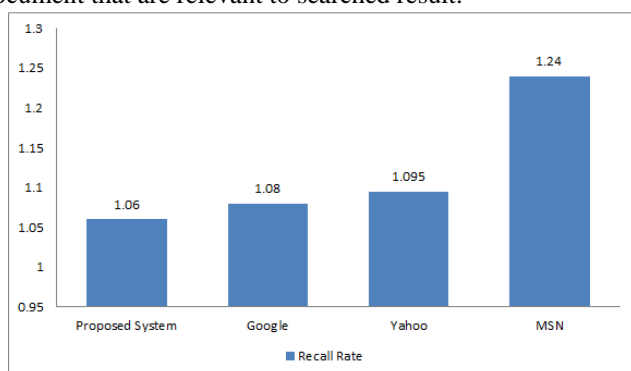

Fig. 5.   Recall Rate Graph.

Recall rate:-Recall in information retrieval is the fraction of the document that is relevant to query that are successfully retrieved.

VII.   CONCLUSION

In this paper, we make a brief survey of some the semantic web search engine that uses various methods to search experience for users. In addition, we discussed intelligent semantic web search engine technique and engine concluded based on perspectives high recall but low precision, identify user intension, inaccurate queries, efficiency of crawler.

For searching the information on web pages using W3C compliant tools, here we use semantic web tool. In the future, our work will be developing an efficient semantic web search technology and that can be also answered the complex query. This paper also gives a brief overview of some of the best semantic search engines that uses various approaches in different ways to yield unique search experience for users. It is concluded that searching the internet today is a challenge and it is estimated that nearly half of the complex questions go unanswered. Semantic search has the power to enhance the traditional web search. Weather a search engine can meet all these criteria continues to remain a question. Future enhancements include developing an efficient semantic web search engine

technology that should meet the challenges efficiently and compatibility with global standards of web technology.

REFERENCES

[1] Berners-Lee, T., Hendler, J. and Lassila, O. "The Semantic Web", Scientific American, May 2001.

[2] Deborah L. McGuinness. "Ontologies Come of Age". In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, 2002.

[3] G.Sudeepthi1, G. Anuradha, Prof. M.Surendra Prasad Babu, "A Survey on Semantic Web Search Engine", International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.

[4] Ramprakash et al "Role of Search Engines in Intelligent Information Retrieval on Web", Proceedings of the 2nd National Conference; INDIACom.

[5] R. Bekkerman and A. McCallum, "Disambiguating Web Appearances of People in a Social Network," Proc. Int "l World Wide Web Conf. (WWW "05), pp. 463-470, 2005.

[6] G. Salton and M. McGill, Introduction to Modern Information Retrieval. McGraw-Hill Inc., 1986.

[7] F. Manola, E. Miller, and B. McBride, RDF primer. *W3C recommendation*, Vol. 10, No., 2004.