

A Novel Architecture of Ontology based Semantic Search Engine

Paras Nath Gupta¹, Pawan Singh², Pankaj P Singh³, Punit Kr Singh⁴, Deepak Sinha⁵

¹A.M. College, Maghad University, Gaya

^{2,3,4}Department of Computer Science, IIMT Engineering College, Meerut

⁵Department of Computer Application, IIMT Management College, Meerut

ABSTRACT

The world wide web is a rapidly going and changing information source. Its growth and change rate make the task of finding relevant information harder. With the dynamic nature of WWW, for a given query the set of relevant web pages is also dynamic, it leads to problem of scalability the assumption of accurate sufficient static image of the web is reduced with its change. Most of the search engines failed to user satisfaction for relevant, complete and updated information. On the part of search desirable to generate the searching technique to get the improvement in the regency and coverage of search engines. In this paper architecture of a search engine is proposed which may lead the user relevant web pages. This architecture uses ontology, semantic based web so as to help user to draw relevant information through search engines.

Keywords: search engine, semantic web, ontology, crawler, indexer, query processor.

1. INTRODUCTION

Finding the relevant information from the WWW has become herculean task as data kept is spread out over millions of machine and pages over the web net. Search engine are tunnels to excess the information required from internet. Search engine consist of four basic units crawler, indexer, ranker & query procedure. The early search engines depend only on matching the words of user query to the words web pages. Due to the growing size of web increase in the numbers and types of user semantic clues and ontology may be utilised to get the relevant data. In the proposed architecture ontology and semantics are used in crawling, indexing and query interface to improve the quality and relevancy in searching through search engine.

2. THE SEMANTIC WEB

Semantic Web came up in 1998 by Tim Berners-Lee which was published in the Roadmap to the Semantic Web on the homepage of the WWW Consortium. As per Tim Berners-Lee the Semantic Web is: "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" [1]. The Semantic Web is considered as a future generation of the Web in which data could be processed by machine and automated agents could recover, extract, and merge data from the Web [2]. In the Semantic Web the semantics of information and services on the Web is defined for the Web to comprehend and satisfy the requests of people and machines to use its content. One of the basic of the Semantic Web is the idea of having explicit semantic information on the Web pages which can be used by intelligent agents to solve complex

problems of information collection and Query-reply. The final objective of the Semantic Web is to be able to keep the accountability of Web content and capability to semantically analyze. It needs a group of structures to model the knowledge, and a link between the knowledge and contents. In this way it depends on two basic components, ontology and semantic annotations. Ontology is used to interpret the textual content of a resource regardless of its format. There have been many fundamental approximations in Semantic Web in which it is supposed that resources have been semantically annotated.

3. ONTOLOGY

Ontology term was initially used preferably by AI researchers and now it is one of the bases of the Semantic Web. It is impossible to envision the Semantic Web with no ontology because Semantic Web is the prime research project concerning ontology. The term Ontology is lent from philosophy. There are different definitions for the concept of ontology applied to information systems, each emphasize a specific aspect. Gruber (1993) [3] defines an ontology as a formal specification of a conceptualization or, in other words, a declarative representation of knowledge relevant to a particular domain. As per Uschold and Gruninger (1996) [4] ontology is as a shared understanding of some domain of interest. Ontology provides "well-defined meaning" to the information enclosed in the Web also the benefit that different parties over the internet now have "shared" definitions about certain key concepts. The most important characteristic of ontology for the present research effort is their role as a structured form of knowledge representation. Ontology is used for the reason of interoperability among systems

based on different schemas and comprehensively describing knowledge about a domain in a structured and sharable way, ideally in a readable format that is processable by a computer.

4. BASIC SEARCH ENGINE ARCHITECTURE

Search engine is used to find the desired information from the internet. It required going through number of phases with in the search engine to find the relevant information. In basic search engine user have to give his query to the search engine interface and in response a list of ranked web pages is returned. The architecture of the basic search engine is as shown in Figure 1.

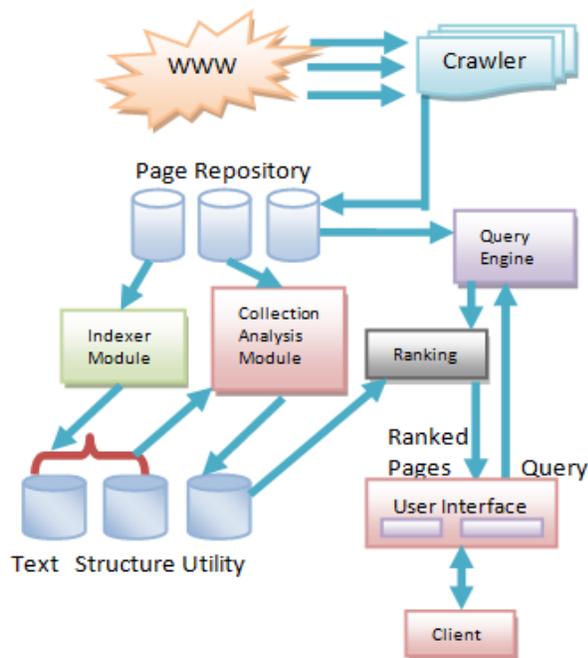


Figure 1: Basic Search Engine Architecture

Conceptually Search Engine retrieves a URL according to query of a user can be achieved in the given steps:

- Client gives the query (keyword) to the Query Engine.
- But in the background, crawler works. Crawler fetches the data from the WWW and store in the database.
- It is store up to end mark, it finds out all in links and out links.
- Then index module indexed the data, text store in the database and structure in the structure database.
- Collection analysis module collect the data from the database and indexed data, then analysis the utility data.

f. Then utility data ranked according to highest page rank, means that sites open so many times. Its priority is higher.

g. And gives the reply to the client.

5. A NOVEL ARCHITECTURE OF ONTOLOGY BASED SEMANTIC SEARCH ENGINE

The world wide web is a rapidly going and changing information source. Its growth and change rate make the task of finding relevant information harder. With the dynamic nature of WWW, for a given query the set of relevant web pages web pages is also dynamic, it leads to problem of scalability the assumption of accurate sufficient static image of the web is reduced with its change. Most of the search engines failed to user satisfaction for relevant, complete and updated information. On the part of search desirable to generate the searching technique to get the improvement in the regency and coverage of search engines. In this proposed work on search engine architecture is suggested as shown in Figure 2 which makes use of ontology at different places to improve the searching result it involved to different modules to process with.

5.1 Web Servers

Web pages of different sites are been collectively stored in the web server. When a request from user is generated to access a web page from a dedicated web site server a client side is activated which sends a request signal to the server side agent to get the web page downloaded from the web server user access information is transferred from client side agents to server side agents. This information is stored in the server log tiles in the web server.

5.2 Ontology based Semantic Web Crawler

For the web search engine a crawler is a program used to download and stores web page and access information from web server. It starts by placing URL's in a queue to be retrieved and then kept. The approach is to provide a crawler the collection of semantic base information for better working of crawler [5]. The goal can be achieved by employing the semantic matching process between the semantic descriptions of web pages ontology.

5.3 Page Repository

In the search engine the page repository is used to store the web pages and other data being crawled by the crawlers. This data base is being used by the indexer to categorize the data and used by the query engine to match the required keywords.

5.4 Search Engine Interface

A search engine interface where user interacts with search engine. It is a front end where user uses to give its query in terms of either words or series of words which in turn are processed by it and then transferred to the semantic query interface. The final result of search is also returned to search engine interface in the form of ranked list of web pages.

5.5 Semantic Query Interface

After receiving a query from the users search engine interface the query is transferred to the semantic query interface where key words are used to correct a normal query into semantic query where lexical database of English language is used. After analyzing the link between ideas, finding the synonyms and then expand the key word set so that the searching gives quality results [6].

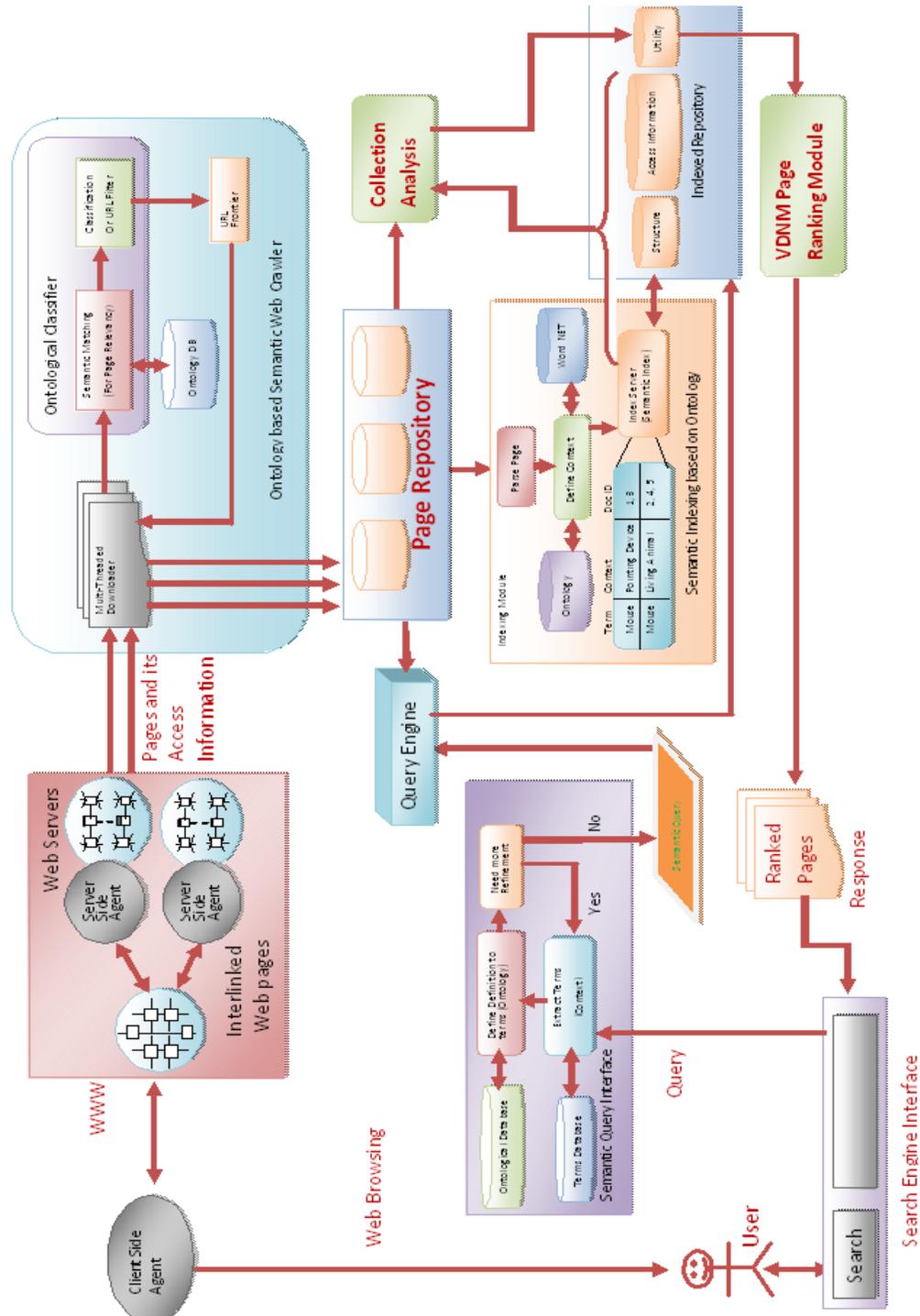


Figure 2: A Novel Architecture of Ontology-based Semantic Search Engine

5.6 Query Engine

The output of semantic query interface is semantic query which in turn submitted to the query engine which used to find the required URL either from the page repository or from the indexed repository. The numbers of possible URLs are selected based on the semantic query through query engine.

5.7 Semantic Indexing Based on Ontology

It is an indexing module used to get the web pages from the page repository and parse them so as to speed up the searching. It work in two manner understanding the various context fetched from web pages using WordNet[7] data of English vocabulary [8]. Finalizing the context the term is matched against ontology of related class. Next is to link ideas with the terms. The context is stored in the database of indexer. It helps to find the quality results for search engines.

5.8 Index Repository & Collection Analysis

Index repository is the database of the indexing module where all proceeded data in the form of web pages, web graph structure and excess information is stored for the use of collection analysis which generated the utility data from the indexed data and transferred the utility data to the page ranking module.

5.9 Page Ranking Module

Page ranking module give the priority order [9] to the different pages on the bases of different utility data(index server data, structure data and excess information)and factor of web net. The final response to the user query is generated in form of ranked web pages.

6. COMPARISON OF BASIC SEARCH ENGINE WITH ONTOLOGY BASED SEMANTIC SEARCH ENGINE

- a. Basic Search Engine takes less time than Ontology based Semantic Search Engine to provide a required result to user.
- b. Searched Web pages through Ontology based Semantic Search Engine are highly relevance than Basic Search Engine.
- c. Basic Search Engine crawls the pages of all domains for the submitted keywords in user query without considering the user's priority but in semantic web crawler it considers not only words but also logical meaning so as to get the users desired information.
- d. In Basic Search Engine the indexing module is blind but in Ontology based Semantic Search Engine the indexing module is quiet intelligent which index the information on the basis of ontological definition to fasten the overall searching.
- e. In Ontology based Semantic Search Engine the user query is converted into semantic query to provide the better understanding of user's requirement, but in the Basic Search Engine it is simply set of keywords rather than meaningful words which is in Ontology based Semantic Search Engine.
- f. The Ontology based Semantic Search Engine reduces the effort of navigation of user to find the desired information.
- g. In Ontology based Semantic Search Engine the cost of accountancy is high than Basic Search Engine as to store meaning required more space.
- h. The cost of implementation of Ontology based Semantic Search Engine is high due to its complex nature of calculation.

7. CONCLUSION AND FUTURE SCOPE

This paper proposes a new way of carrying out semantic search based on ontology by making use of existing resources, but searching time may be large because semantic search engines transform simple query into the semantic query which pass through the different phases, crawler gets the pages based on ontological information and indexing takes the semantic view based on ontology. This method can provide better, accurate and most relevant result with compare to traditional search engine results. The future work includes the implementation of this architecture to generate a faster and relevant web page or information provider search engine.

REFERENCES

- [1] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. Scientific American, 284(5): 34-43, 2001.
- [2] Tim Berners-Lee, James Hendler, and OraLassila. The semantic web. Scientific American, 284(5):34-43, 2001.
- [3] Thomas R. Gruber, A translation approach to portable ontology specifications, KnowledgeAcquisition 5 (1993), no. 2, 199-220.

- [4] M. Ushold and M. Gruninger, Ontologies: Principles, methods and applications, The Knowledge Engineering Review, 1996.
- [5] Ram Kumar Rana, NidhiTyagi, "A Novel Architecture of Ontology-based Semantic Web Crawler", International Journal of Computer Applications (0975 – 8887) Volume 44– No18, April 2012.
- [6] Sajendra Kumar, Ram Kumar Rana, Pawan Singh, "A Semantic Query Transformation Approach Based on Ontology for Search Engine", International Journal on Computer Science and Engineering (IJCE), May 2012. (688-693).
- [7] KarinK. Breitman, Marco Antonio Casanova, Walt Truskowski, Semantic Web: Concepts, Technologies and Applications, Springer VerlagBerlin Heidelberg London2007.
- [8] Sajendra Kumar, Ram Kumar Rana, Pawan Singh, "OntologyBased Semantic Indexing Approach for Information RetrievalSystem", International Journal of Computer Applications, 49(12):14-18, July 2012.
- [9] L. Page, S. Brin, R. Motwani, T. Winograd. "The PageRank Citation Ranking: Bringing Order to the Web", Stanford Digital Library Technologies Project.