IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-6, Issue-4 E-ISSN: 2347-2693

# Web Searching an Art

**Chinu[1*], Ramil Gupta[2], Ekta[3]**

[1]Dept. CSE, Baba Farid College of Engineering and Technology, Mrsptu, Bathinda, India
[2]Dept. CSE, Baba Farid College of Engineering and Technology, Mrsptu, Bathinda, India
[3]Dept. CSE, Baba Farid College of Engineering and Technology, Mrsptu, Bathinda, India

**Corresponding Author: verma.chinu92gmail.com*

**Available online at: www.ijcseonline.org**

Received: 20/Mar/2018, Revised: 28/Mar/2018, Accepted: 19/Apr/2018, Published: 30/Apr/2018

***Abstract*** In today's world searching information over the internet is one of the difficult task. Specialized tools called search engines plays an important role to retrieve information from world wide web. There are many of search engines available today but retrieving meaningful information is difficult. However to overcome this problem in search engines to retrieve meaningful information intelligently, semantic web technologies are playing a major role. In this research work we present a study on the search engine generations and the role of search engines in intelligent web search technologies.

***Keywords***— Search engine, intelligent search, and semantic web.

## I. INTRODUCTION

The term WWW refers to the World Wide Web or simply the Web consisting of all the public Web sites connected to the Internet worldwide, including the client devices (such as computers and cell phones) that access Web content. The World Wide Web is the cosmos of network-accessible information, an personification of human knowledge [2]. The WWW is one of many applications of the Internet and computer networks [1].The World Web is based on these technologies HTML,HTTP,Web servers and Web browsers The World Wide Web is a system of interlinked hypertext documents accessed via the Internet. With the help of web browser, one can view web pages that may contain text, images, videos, and other multimedia, and navigate between them via hyperlinks [3]. In this research work we present review on the search engine generations and the role of search engines in intelligent web search technologies.
In this research work, Section I contains the introduction of web search engines, Section II contain the evolution of various search engines Section III Search engine algorithms Section IV contain the description about Intelligent web search engine. Section V concludes research work with future directions.

## II. THE EVOLUTION OF WEB SEARCH ENGINES

Web Search is the process of researching the Internet in order to obtain significant and relevant information based on the query or task at hand. [4]Web search is the act of looking for webpages.[5]]A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages (SERPs).[6].When looking for something using a search engine, it is a good idea to use words like AND, OR, and NOT to specify the search. Using these boolean operators, we can usually get a list of more relevant sites. [7] Web search engines work by sending out a *spider* to fetch as many documents as possible. Another program, called an *indexer*, then reads these documents and creates an index based on the words contained in each document. Each search engine uses a proprietary algorithm to create its indices such that, ideally, only meaningful results are returned for each *query*. [8]

In view of the taxonomy discussed so far we identify three stages in the evolution of web search engines:
• ***First Generation*** -- uses mostly on-page data (text and formatting) and is very close to classical IR. Supports mostly informational queries. This was state-of-the art around 1995-1997 and was exemplified by AltaVista, Excite, WebCrawler, etc.
• ***Second Generation*** -- use off-page, web-specific data such as link analysis, anchor-text, and click-through data. This generation supports both informational and navigational queries and started in 1998-1999. Google was the first engine to use link analysis as a primary ranking factor and DirectHit concentrated on click-through data. By now, all major engines use all these types of data. Link analysis and anchortext seems crucial for navigational queries.

 

• ***Third Generation*** -- emerging now, attempts to blend data from multiple sources in order to try to answer ‘‘the need behind the query’’. For instance on a query like San Francisco the engine might present direct links to a hotel reservation page for San Francisco, a map server, a weather server, etc. Thus third generation engines go beyond the limitation of a fixed corpus, via semantic analysis, context determination, dynamic data base selection, etc. The aim is to support informational, navigational, and transactional queries. This is a rapidly changing landscape [16].

## III SEARCH ENGINE RANKING ALGORITHMS

Each search engine has its own method for calculating relevance, usually based on an analysis of the content of the destination webpage, including:

- meta title (visible at the top of the web browser window);
- metadata;
- Number of incoming links, (commonly referred to as the page’s ‘popularity’). Popularity-based ranking assumes that the more incoming links a webpage has, the more likely it is to be a subject ‘authority’;
- incoming link text: a search engine may make assumptions about the content of a website based on how other people have described it through the text they have used to link to a site;
- use of appropriate semantic markup, for example, use of heading elements; and
- Page text.

Each of these aspects of the webpage is scored and then weighted. For example, a search engine may assign a greater weighting to meta title text than other aspects of the webpage. In this case, a webpage that includes the query in its meta title text may then be ranked higher than a webpage where the meta title does not include the query [10].

*Different Types of Search Engines*

**Crawler-Based Search Engines:**

Crawler-based search engines use automated software programs to survey and categories web pages. The programs used by the search engines to access the web pages are called ‘spiders’, ‘crawlers’, ‘robots’ or ‘bots’. A spider will find a web page, download it and analyse the information presented on the web page. This is a seamless process. The web page will then be added to the search engine’s database. Then when a user performs a search, the search engine will check its database of web pages for the key words the user searched on to present a list of link results. The results (list of suggested links to go to), are listed on pages by order of which is ‘closest’ (as defined by the ‘bots’), to what the user wants to find online. Crawler-based search engines are constantly searching the Internet for new web pages and updating their database of information with these new or altered pages.Examples of crawler-based search engines are:Google (www.google.com), Ask Jeeves (www.ask.com) [11]

**Human-Powered Directories**

A ‘directory’ uses human editors who decide what category the site belongs to. They place websites within specific categories in the ‘directories’ database. The human editors comprehensively check the website and rank it, based on the information they find, using a pre-defined set of rules. There are two major directories at the time of writing:

Yahoo Directory (www.yahoo.com), Open Directory (www.dmoz.org)[11].

**Hybrid Search Engines**

Hybrid search engines use a combination of both crawler-based results and directory results. More and more search engines these days are moving to a hybrid-based model. Examples of hybrid search engines are:

Yahoo (www.yahoo.com), Google (www.google.com) [11]

**Meta Search Engines**

Meta search engines take the results from all the other search engines results, and combine them into one large listing. Examples of Meta search engines include:

Metacrawler (www.metacrawler.com), Dogpile (www.dogpile.com)

**Specialty Search Engines**

Specialty search engines have been developed to cater for the demands of niche areas. There are many specialty search engines, including:

Shopping - Froogle (www.froogle.com)
- Yahoo Shopping (www.shopping.yahoo.com)
- BizRate (www.bizrate.com)
- PriceGrabber (www.pricegrabber.com)
- PriceSpy (www.pricespy.co.nz)

Local Search
- NZPages (www.nzpages.co.nz)
- SearchNZ (www.searchnz.co.nz)
- NZS (www.nzs.com)

Domain Name Search
- iServe (www.iserve.co.nz)
- Freeparking (www.freeparking.co.nz)

Freeware & Shareware Software Search
- Tucows (www.tucows.com)
- CNET Download.com (www.download.com) [11][12]

## IV. INTELLIGENT WEB SEARCH ENGINE

The Semantic Web is an extension of the current Web [13] that allows the meaning of information to be precisely described in terms of well-defined vocabularies that are understood by people and computers. On the Semantic Web, information is described using a new W3C standard called the Resource Description Framework (RDF). Semantic Web Search is a search engine for the Semantic Web. Current Web sites can be used by both people and computers to precisely locate and gather information published on the Semantic Web. Ontology [14] is one of the most important concepts used in the semantic web infrastructure, and

 

RDF(S) (Resource Description Framework/Schema) and OWL (Web Ontology Languages) are two W3C recommended data representation models which are used to represent ontologies. The Semantic Web will support more efficient discovery, automation, integration and reuse of data and provide support for interoperability problem which cannot be resolved with current web technologies. Currently research on semantic web search engines are in the beginning stage, as the traditional search engines such as *Google, Yahoo, and Bing (MSN)* and so forth still dominate the present markets of search engines.

Most of the search engines search for keywords to answer the queries from users. The search engines usually search web pages for the required information. However they filter the pages from searching unnecessary pages by using advanced algorithms. These search engines can answer topic wise queries efficiently and effectively by developing state-of art algorithms.

However they are vulnerable in answering intelligent queries from the user due to the dependence of their results on information available in web pages. The main focus of these search engines is solving these queries with close to accurate results in small time using much researched algorithms. However, it shows that such search engines are vulnerable in answering intelligent queries using this approach. They either show inaccurate results with this approach or show accurate but (could be) unreliable results. With the keywords based searches they usually provide results from blogs (if available) or other discussion boards. The user cannot have a satisfaction with these results due to lack of trusts on blogs etc. To overcome this problem in search engines to retrieve relevant and meaningful information intelligently, semantic web technology deals with a great role. Intelligent semantic technology gives the nearer to desired results by search engines to the user [15].

*4.2 Background*

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. Current web is the biggest global database that lacks the existence of a semantic structure and hence it makes difficult for the machine to understand the information provided by the user. When the information was distributed in web, we have two kinds of problems in search engine i.e.

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

Semantic web can solve the first problem in web with semantic annotations to produce intelligent and meaningful information by using query interface mechanism and ontology's. Other one can be solved by the graph-based query models. The Semantic web would require solving extraordinarily difficult problems in the areas of knowledge representation, natural language understanding. The following figure depicts the semantic web frame work it also referred as the semantic web layercake by W3C [15].

*4.3 Current Web & Limitations*

Present World Wide Web is the longest global database that lacks the existence of a semantic structure and hence it becomes difficult for the machine to understand the information provided

by the user in the form of search strings. As for results, the search engines return the ambiguous

or partially ambiguous result data set. Semantic web is being to be developed to overcome the following problems for current web [15].

• The web content lacks a proper structure regarding the representation of information.

• Ambiguity of information resulting from poor interconnection of information.

• Automatic information transfer is lacking.

• Usability to deal with enormous number of users and content ensuring trust at all levels.

• Incapability of machines to understand the provided information due to lack of a universal format.

Hakia is a general purpose semantic search engine that search structured text like Wikipedia. Hakia calls itself a "meaning-based (semantic) search engine". They're trying to provide search results based on meaning match, rather than by the popularity of search terms.

**Types of Semantic Search Engines**

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 behavior. Semantic web is being developed to overcome the following main limitations of the current Web [15]:

- The web content lacks a proper structure regarding the representation of information.
- Ambiguity of information resulting from poor interconnection of information.
- Automatic information transfer is lacking.
- Unable to deal with enormous number of users and content ensuring trust at all levels.
- Incapability of machines to understand the provided information due to lack of a universal format.

*4.4.1 Semantic Search Engines*

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 seriously contribute to the development of electronic business applications since it is based on strong theory and widely accepted standards.

**Ontology Search Engines**

They have an integrated approach for ontology searching, reuse and update. In its architecture, an ontology registry is designed to store the metadata about ontologies and ontology server stores the ontologies. The ontologies in distributed ontology servers can be created, replicated and evolved. Ontology metadata in ontology registry can be queried and registered when a new ontology is created[15].

## V. Conclusion

In this paper, we make a brief survey of the intelligent semantic search technologies and review their characteristics respectively. In addition, the issues within the reviewed intelligent semantic search methods and engines are concluded based on four perspectives-differentiations between designers and users' perceptions, static knowledge structure, low precision and high recall and lack of experimental tests. In the future, we can focus on the deeper and broader research in the field of intelligent semantic search, with the purpose of concluding the current situation of the field and promote the further development of intelligent semantic search engine technologies.

## REFERENCES

[1]http://compnetworking.about.com/cs/worldwideweb/g/bldef_www.htm
[2]http://searchcrm.techtarget.com/definition/World-Wide-Web
[3] http://en.wikipedia.org/wiki/World_Wide_Web
[4] http://www.freelancer.com/jobs/Web-Search/
[5]http://en.wikibooks.org/wiki/PlanoTse_Handbook_for_Job_Search_Automation/What_is_ web_search%3F
[6] http://en.wikipedia.org/wiki/Web_search_engine
[7] http://www.techterms.com/definition/searchengine
[8] http://www.webopedia.com/TERM/S/search_engine.html
[9]http://www.pcmag.com/encyclopedia/term/54339/web-search-engines
[10] http://www.motive.co.nz/glossary/search.php
[11]http://www.zeald.com/Blog/x_post/types-of-search-engines.html
[12]http://www.yuanlei.com/studies/articles/is567-searchengine/page2.htm
[13] Berners-Lee, T., Hendler, J. and Lassila, O. "The Semantic Web", Scientific American, May 2001.
[14] Deborah L. McGuinness. "Ontologies Come of Age". In Dieter Fensel, J im Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, 2002.
[15] G.Madhu1, Dr.A.Govardhan and Dr.T.V.Rajinikanth, "Intelligent Semantic Web Search Engines: A Brief Survey", International journal of Web & Semantic Technology (IJWesT) Vol.2, No.1, January 2011.
[16] Andrei Broder, "Taxonomy of Web Search"

 