Semantic Web Search Engines : A Comparative Survey

Anita Kumari, Dr. Jawahar Thakur

Department of Computer Science, Himachal Pradesh University, Shimla, Himachal Pradesh, India

ABSTRACT

Search engines play important role in the success of the Web. Search engine helps the users to find the relevant information on the internet. Due to many problems in traditional search engines has led to the development of semantic web. Semantic web technologies are playing a crucial role in enhancing traditional search, as it work to create machines readable data and focus on metadata. However, it will not replace traditional search engines. In the environment of semantic web, search engine should be more useful and efficient for searching the relevant web information. It is a way to increase the accuracy of information retrieval system. This is possible because semantic web uses software agents; these agents collect the information, perform relevant transactions and interact with physical devices. This paper includes the survey on the prevalent Semantic Search Engines based on their advantages, working and disadvantages and presents a comparative study based on techniques, type of results, crawling, and indexing.

Keywords : Semantic Web, Semantic Search Engines, Intelligent Search

I. INTRODUCTION

The web is a medium for accessing a great variety of information stored in different parts of the world. Information is mostly in the form of unstructured data. As the data on the web grows at explosive rates, it has lead to several problems such as increased difficulty of finding relevant information, extracting useful knowledge.

Conventional web search engines like Google and yahoo return thousands of results for users quires because they do not understand the exact meaning of quires. They uses keyword based search approach, thus new technology "Semantic Web" comes into existence, which works on the semantics (meaning) of queries. Semantic web offers a smarter web services which synchronies and arranges all the data over the web in a disciplined manner.

So moving from current web to semantic web particularly depends on enhancement of knowledge and overcome the current web issues. Semantic search plays an important role because it intelligently understands the content and give smart and relevant results according to the quires. Traditional search engines get the answers syntactically accurate but larger in amount. However, semantic engines provide accurate information in terms of well-defined expressions. The main aim of semantic web is to lift the capacity of users and software agents to find documents, information and knowledge from the web and make the content to machine and human readable. This is possible by adding the semantic layer to the current web.

The organization of this document is as follows. In Section 2 (**Literature Survey),** literature survey is explained with comparative study. In Section 3 (**Keyword Based Search**), keyword based search engines and the limitations are discussed. Semantic web, approaches to semantic web and few search engines (Hakia, Falcons, Swoogle, DuckDuckGO and Knigne) and comparison on these search engines are discussed in Section 4(**Semantic Web Search**) based on few parameters. Section 5 (**Conclusion)** concludes the paper by providing a summary of all reported work in semantic web search.

II. LITERATURE SURVEY

There is significant literature work that has been done on semantic web area. Various authors and researchers have contributed very well in this area.

1) **Vidhi Shah1, Akshat Shah [5],** concluded that searching the internet today is a challenge and semantic search engine enhances the traditional web search engines. They discuss various semantic search engines and present a comparison based on the approaches.

2) **G.Anuradha, G.Sudeepthi et al [6],** presents a brief survey of the existing literature regarding intelligent semantic web search engine technologies. Also, review their characteristics respectively. They concluded that existing techniques have limitations particularly in terms of time response, accuracy of results, importance of the results and relevancy of results. An efficient semantic web search engines should meet out these challenges efficiently and compatibility with global standards of web technology.

3) **Maliha Majid Qureshi, Bibi Assma [7],** presents the semantic search engines and broad survey to analyze the semantic search engines has done. Comparative analysis of the semantic search engines has done based on Factors cited in the

pyramid. The research provides deep Understanding of five main semantic search engines based on comparative analysis that may help for future work for semantic web in general and for semantic search engines in particular.

4) **Nikhil CHITRE [8],** analyses the semantic search engines generation into the intelligent web. Their focus is on finding the difference between the traditional and semantic search engines. Explains the various techniques and advantages of few important semantic web search engines that has been developed so far.

5) **Yi Jin Zhuying Lin, Hongwei Lin [9],** states that search engines plays important role in success of web; search engine helps any internet user to rapidly find relevant information. Architecture of semantic search engine is explained and how elements of semantic search engine are used in performing the fundamental task of information retrieval is explained. An improved algorithm based on TFIDF algorithm is also proposed.

6) **Kyumars Sheykh Esmaili, Hassan Abolhassani [10],** presents a categorization scheme for web search engines. For each category, components are explained according to a proposed general architecture and various approaches are used in these components also represented. It should note that some of these engines are constructed through a portal and therefore they have not individual names.

7) **Ranjana Jain, Neelam Duhan el at [11],** states that web 2.0 search engines are unable to present direct answers against the users fired query. Web 3.0 is semantic web using RDF format organization information in more structure form, which helps the semantic search engines to present results in a more direct way. Also include these semantic search engines, review their working, and compare them based on few factors.

8) **Jyoti Chaurasia, Jagdish Raikwal[12],** states that semantic web technology plays an crucial role in enhancing traditional web search. They proposed

a semantic search engine based on ontology. An algorithm is also introduced to overcome the various challenges in the fields of unstructured data using semantic web technology and provide the complete description.

9) **Ankita Malve , P.M. Chawan [13],** concluded that semantic based search engine have more advantages over keyword based search engine in terms of accuracy of getting results. Search process in semantic search engine is based on semantics of queries. It provides assurance to user to get accurate and relevant results based on meaning of word. They also include the basic

difference between the keyword based and semantic based search engines.

10) **G.Madhu1 and Dr. A. Govardhan [14],** presents a brief survey of the existing literature regarding intelligent semantic search technologies. 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.

TABLE I

COMPARISONS BASED ON LITERATURE SURVEY

Reference no.	Semantic Search Engines	Language	Objects	Conclusion
[6]	Falcons, Kosmix, Sunbot, Hakia, Swoogle, XCDSearch, XSearch, Kngine, Powerset, DuckDuckGO.	RDF, OWL, OIL, DAML+OIL,ODEX.	Differentiate between semantic search engines and presents their advantages and disadvantages.	Combining semantic search with current search engines to make content understandable by human as well as machine.
[7]	Hakia, Sensebot, Powerset, Cognition, Lexxe.	QDEX, NLP.	Research is carried out to explore the different dimensions of semantic search engines.	This paper presents pyramid based comparison. Values of pyramid includes: search environment, intrinsic problem, query type, and iterative and exploratory.
[8]	DuckDuckGO, Hakia, Kngine.	RDF, DAML+OIL, OIL, OWL, QDEXing.	Analyse semantic search engines generation and the part of search technologies and intelligent web.	The semantic search engine improves the limitation of traditional search engine and provide 100% relevant result.
[9]	Architecture of semantic search engine and improved algorithm based on TFIDF are presented.	OWL, XML, TFIDF algorithm.	How fundament task of retrieval is performed.	Shows that how the fundamental elements of semantic search engines perform the information retrieval.
[10]	Content meta semantic engine, crawler based ontology search engine, evolutionary search	RDF,OWL.	Differentiate between ontology semantic search engines and	Provide a categorization scheme for different type of search engines on semantic web.

	engines, semantic association discovery engine.		semantic search eninge.	
[11]	Architecture of search engines, meta-search engines, semantic search engines (Swoogle, Falcons, Hakia, SWSE).	RDF, OWL, CSV, XML.	Why we need web search engines.	Comparative study of Google, Falcons, Swoogle, Hakia, SWSE.
[13]	Google, DuckDuckGO	OWL, RDF, XML, HTML.	Comparison between keyword based search engines and semantic search engine.	Search process in semantic search engine is based on semantic of query and provide assurance to get 100% accurate data.
[12]	Google, a new algorithm proposed in addition with Google.	RDF, OWL, DAML+OIL, OIL, QDEXing.	Propose a semantic search engine based on domain ontology.	Proposed an algorithm with ontology's to give better search experience.

III. KEYWORD BASED SEARCH

Search Engines are used to find the information on the internet and getting results in less time but it lacks the existence of the semantic structure so it becomes difficult for the machine to understand the meaning of the terms and expressions provided by the users. As the results, search engines returns the ambiguous data sets. The survey indicate that user do not find the accurate results in the first set of URLs returned because of the increased size of the links on the web pages. Because sometime one word has several meanings and several words have same meaning, in this case user will not get the result what he wanted to search [13]. Semantic web is developed to overcome the following limitations of the current web search:-

- These engines are based on the occurrence of words in the documents, so it is difficult to get the relevant results.
- Lack of proper **semantic structure** according to the representation of information.
-

- Low value of **Recall** and current **Precision** parameters.
- More **network congestion** because multiple crawlers try to crawl the same web pages form web servers.
- **Polysemy words:-** engines do not differentiate between the same spelled words having different meaning hence give irrelevant pages.
- **Stemming:** - also these search engines don not differentiate between the singular and plural words. Hence may produce the confusion while searching a query.
- **Synonymy:** - it is the problem of understanding the different spelled words with the same meaning and it produce the confusion while searching.
- Lack of semantic understanding [6][13][12].

IV. SEMANTIC WEB SEARCH

A. Semantic Web

It is an extension to WWW and international certified standard by W3C, which aims in handling

mostly unstructured data. It describes the relationship between things and properties of things. The main idea of semantic web is to add the metadata into the current web pages to improve the description of online resources.

Tim Berners-Lee originally expressed the vision of the semantic web as follows —I have a dream for the Web [in which computers] become capable of analysing all the data on the Web – the content, links, and transactions between people and computers. A Semantic Web, which should make this possible, has yet to emerge, but when it does, machines talking to machines will handle the day-to-day mechanisms of trade, bureaucracy and our daily lives. The intelligent agent's people have touted for ages will finally materialize [1]. It presents the easier way to share, reuse and discover the online resources [2]. Also represent the information in the way that human and machines can understand using semantics [2].
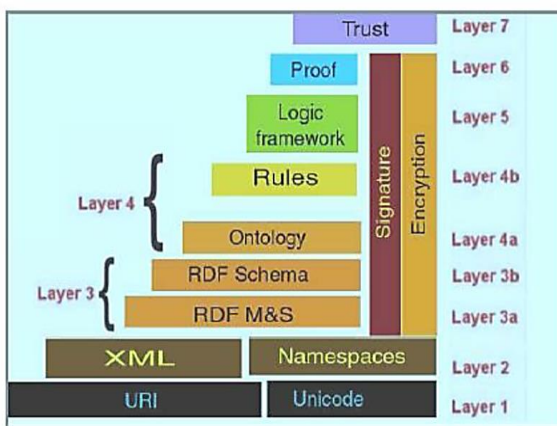

Figure 1. Architecture of Semantic Web

B. Approaches to Semantic Web

There is not only one approach in the semantic search. Every approach has a lot to contribute. These main four approaches are used; different search engine may use one or more approaches'.

1) The **First approach** uses contextual analysis to disambiguate the search queries.
2) The **Second approach** focus on reasoning.

3) The **Third approach** focus on natural language understanding.
4) The **Fourth approach** makes use of ontology to represent the knowledge about a Particular domain and expand quires [6][8].

C. Semantic Based Search Engines

The semantic search describes that search engine will give most accurate search results possibly by understanding searcher intent, query content and the relation between words. Alternatively, we can say that semantic search engines will understand the natural language as the way human can understand and it improves the search accuracy of query related data and delivers exact and smart results. Semantic search engines uses ontology to get the relevant and accurate results in less time. Ontology provides the related association between the contents. RDF (Resources Description Framework) and OWL (Web Ontology Language) are used for semantic web documents. Rather than using ranking algorithms (PageRank to predict relevancy), Semantics are used. It provides the guaranty that most relevant results will retrieve depending upon the meaning and relations of the words not a specific keyword. The goal of semantic search is to use the meaning (semantic) to improve the search experience of the users.

In order to access structured data, a variety of search engine has been introduced which works on semantics(meaning) of data and provide the accurate results in less time as compare to the traditional search engine. The broad categories of semantic search engine according to their workings are [10]

1) Crawler Base Search Engines,
2) Human powered directories,
3) Hybrid Search Engines,
4) Meta Search Engines.

Some of the important semantic search engines are:

1) **SWOOGLE: -** It is a crawler-based indexing and retrieval system for semantic web documents (SWD). It uses the crawler to discover the RDF and HTML documents. It extracts metadata and computes relations between documents. Swoogle is also a content-based search engine that used to analyse, discover, and index the knowledge from the web. It is based on the page rank algorithm and OWL language [11].

2) **HAKIA:-** It is a meaning based search engine that works on matching the meaning rather than keyword match. The searching capacity of this engine is not limited to single keyword; you can search sentences, a phrases or questions. This engine gets the results using equivalent terms. Hakia engine essentially builds around mainly three evolving technologies i.e., Sense Repository, Query Indexing Technique and Semantic Ranking algorithm [8][6][11].

3) **FALCONS: -** It is a keyword based ontology search engines. It is develop around the concept sharing and ontology reusing. It display the required information on snippets, therefore user does not need to explore the pages. This engine provides the best user interface. This engine integrates these two terms i.e., Concept-Level-Search and Ontology- Level-Search by recommending ontology and allows filtering the concepts using ontology's. Through this simple process user can easily find the ontology's that satisfy their need within a few seconds [11].

4) **KNGINE: -** It is intelligent search engine and the result is divided into either web results or image results. They are preceded by the information about the "concepts". For instance, attempting to find a town can be preceded by using statistics about the metropolis, nearby points of interest, activities, climate and hotels. It provides the high degree of related data to the user's queries. It analyze the main keywords from the users queries and form the relations between them and return results. It is a question-answer engine that understands the questions and tries to answer with the help of natural language processing and machine learning [8][6].

5) **DUCKDUCKGO: -"The search engine that does not track you"** it is a feature rich search engine that gives many reasons to put Google away. It focuses on maintaining user privacy and therefore does not store any personalized search results. This engine guaranty that they does not store users IP address, does not users login information. This engine only uses cookies to maintain the privacy of the users. It is a semantic search technology based search engine that presents relevant match or concept match for example if a user searches for a keyword that has more than one connotation, it gives you the options to choose what you were originally searching for along with the disambiguate outcome. This engine has the most important property that it provides users privacy which Google does not give [8][6][10].

TABLE II

COMPARISONS BETWEEN DIFFERENT SEMANTIC SEARCH ENGINES

	SWOOGLE	FALCONS	HAKIA	KNGINE	DUCKDUCKGO
APPROACH	Content based approach.	Keyword Based approach.	Content match, NLP	knowledge based , Statistical	cluster search, NLP.
TECHNIQUE	1.Rational Surfer	Object	OntoSem(Sense	knowledge-	Cluster search,

	Model, 2.It uses crawler based indexing semantic search technique that search ontology and instance data.	Ranked based on combination of their relevance to the query and their popularity. Uses cosine similarity measures between the query and virtual documents of objects.	Repository), QDEX(Query Indexing Method), Semantic Algorithm.	Based-Approach, Statistical Approach.	NLP
TYPE OF RESULTS	Entity(OWL,RDF)	Entity(gives the result in snippet itself)	Documents (Links and free Text)	Web Results or Image Results	Summaries(provide Privacy)
INDEXING TECHNIQUE	Swangling technique.	Inverted Indexing	Query detection and extraction		Provide Instant Answers
IT CRAWEL	RDF Data	RDF Data	HTML Web Pages	Web pages	RDF Data
FEATURES	**1.**Find exact ontology's, **2.**Finds exact instance data structure of the semantic web.	**1.**It integrate concepts level search and ontology level search, **2.**User can find results quickly.	**1.**Easily identify information from credible sites, **2.**Save time, **3.**Excellent response time.	**1.**Display results in the form of image, **2.**It is multi-lingual, 3.It allows the user to search in parallel manner.	**1.**Zero click information, **2.**Provide privacy, **3.**Does not record users information, **4.**Produces the result based on many source and also from its own crawler.

LIMITATIONS	1.Poor indexing of documents, 2.Poor response time for find queries.	1.This engine does not rank the objects according to query.	1.URL Canonization, 2.Privacy Session IDS, 3.Does not index everything, 4.Need other search engines too, 5.Contains virtual content and dynamic content.	1.Does not provides silent mode option.	1.It provides itself an option of privacy but they are just as opaque as Google is, since they don't open source their code, how do you know they are not tracking you?

V. CONCLUSION

This paper includes the survey on the prevalent Semantic Search Engines based on their advantages, working and disadvantages and presents a comparative study based on techniques, type of results, crawling, and indexing. It concluded that semantic based search has more advantages over the keyword-based search and provide the relevant and accurate results in first URL. Semantic Web organizes the information in more structure form by using RDF or OWL format, which helps the semantic search engines to, presents the results in way that is more accurate.

Future work can be extended to the development of new semantic search engine technologies that overcomes their limitations.

VI. REFERENCES

[1]. Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." Scientific american 284.5 (2001): 34-43.

[2]. Shadbolt, Nigel, Tim Berners-Lee, and Wendy Hall. "The semantic web revisited." IEEE intelligent systems 21.3 (2006): 96-101.

[3]. Berendt, Bettina, Andreas Hotho, and Gerd Stumme. "Towards semantic web mining." International Semantic Web Conference. Springer, Berlin, Heidelberg, 2002.

[4]. Al-Feel, Haytham T., Magdy Koutb, and Hoda Suoror. "semantic web on Scope: A New Architectural Model for the semantic web." (2008).

[5]. Shah, Vidhi, Akshat Shah, and Asst Prof Khushali Deulkar. "Comparative Study Of Semantic Search Engines." International Journal Of Engineering And Computer Science 4.11 (2015).

[6]. Sudeepthi, G., G. Anuradha, and M. Surendra Prasad Babu. "A survey on semantic web search engine." International Journal of Computer Science Issues (IJCSI) 9.2 (2012): 241.

[7]. Qureshi, Maliha Majid, Bibi Asma, and Hikmat Ullah Khan. "Comparative analysis of semantic search engines based on requirement space pyramid." International Journal of Future Computer and Communication 2.6 (2013): 562.

[8]. Chitre, Nikhil. "Semantic Web Search Engine." International Journal 4.7 (2016).

[9]. Jin, Yi, Zhuying Lin, and Hongwei Lin. "The research of search engine based on semantic web." International Symposium on Intelligent Information Technology Application Workshops. IEEE, 2008.

[10]. Esmaili, Kyumars Sheykh, and Hassan Abolhassani. "A categorization scheme for semantic web search engines." Computer Systems and Applications, 2006. IEEE International Conference on.. IEEE, 2006.

[11]. Jain, Ranjna, Neelam Duhan, and A. K. Sharma. "Comparative study on semantic search engines." International Journal of Computer Applications, ISSN (2015): 0975-8887.

[12]. Chaurasia, Jyoti, and Jagdish Raikwal. "Enhanced Semantic Web Retrieval Technique: Domain Ontology." International Journal of Computer Applications 119.16 (2015).

[13]. Malve, Ankita, and P. M. Chawan. "A Comparative Study of Keyword and Semantic based Search Engine." (2015): 11156-11161.

[14]. Madhu, G., Dr A. Govardhan, and Dr TV Rajinikanth. "Intelligent semantic web search engines: a brief survey." arXiv preprint arXiv:1102.0831 (2011).

Cite this article as :