

# **A Comparative Study of Keyword and Semantic based Search Engine**

Ankita Malve<sup>1</sup>, Prof. P. M. Chawan<sup>2</sup>

M. Tech Student, Department of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India<sup>1</sup>

Associate Professor, Department of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India<sup>2</sup>

**ABSTRACT:** Keyword based Search engines are unable to give relevant search results as they do not know the exact meaning of the keywords used. This paper compares both keyword and semantic search engines. Semantic Web is advanced version of the current web. It represents information meaningfully for machines and humans. It is based on an Ontology, which is considered as the main component of the Semantic Web. The current Web is transformed from being machine-readable to machine-understandable. Ontology is a key technique with which to annotate semantics and provide a common, comprehensible foundation for resources on the Semantic Web. This paper provides the basic difference between the keyword and semantic based search engine.

**KEYWORDS:** Information retrieval, Ontology, Semantic Web, OWL, URI, Unicode.

## **I. INTRODUCTION**

The main purpose of search engine is to allow user to search and retrieve web documents with queries to get information which they want.[1] The volume of search ratio of the most popular search engines such as Google, Yahoo, Bing are 84.16%, 7.61%, 4.40% respectively [2].The estimated size of the World Wide Web is at least 11.5 billion pages, [3] but a much larger Web, estimated at over 3 trillion pages exists in the databases whose contents are not indexed in the search engine. Google and other popular search engines are also a target for search engine optimizer so there may also be many results returned that only serves as advertisements. Sometimes web pages have many keywords which are designed to attract users to that web page.

Google uses page ranking algorithm to predict relevancy where as semantic search uses semantics or the science of meaning in language, to produce highly relevant search results. The main goal of search engine is to provide the information queried by a user rather than have to sort through a list of non related keyword retrieved results. The main difference between keyword and semantic based search engine is that time consumption. User takes time to sort the retrieved result in keyword engine where as in case of semantic based search engine there is no sorting of search results. User can get the relevant results in short time and which is very useful and most important characteristic of semantic search system.

## **II. RELATED WORK**

### **A. Keyword Search Engine**

Conventional Search Engines are very helpful in finding information on the internet and getting results within some time, but they suffer from the fact that they do not know the meaning of the terms and expression used in the web pages and the relationship between them. Surveys indicate that users who wants to search o the web do not find accurate results in the first set of URLs returned, because of increasing size of links on the web pages [6].Sometimes one word has several meaning and several words have same meaning, in that case if a user wants to search a particular word then it may produce confusion and user will not get what he wanted to search.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

Disadvantages of Keyword Search engine:

- Conventional Information Retrieval (IR) technology is based on the occurrence of words in documents. Therefore it's difficult to get a relevant result.
- Precision and Recall value is low. Precision is the ratio of the documents retrieved that are relevant to the user's information need. And Recall is the ratio of the documents relevant to the query that are easily retrieved. Precision is equal to number of Total Relevant results is divided by the number of total retrieval. And Recall is the number of retrieved relevant results is divided by number of possible relevant results.
- Polysemy words: It means one word having several meanings. E.g. word "FAST". It has several meanings like quick, starves you and fixed. Therefore, it may produce confusion while searching a query.
- Synonymy words: several words having same meaning. E.g. clothes and apparel which have same meanings. It can also produce the confusion while searching a query having same meaning.

## B. Semantic based Search Engine

A semantic search engine intelligently understands the context of what is being searched and give smart and relevant result according to the queries asked. The traditional search engines uses page ranking algorithm to give ranking to the particular link so that search results will be relevant. On the other hand, a semantic search engine uses ontology so that meaningful and accurate results should be retrieved in less time. It provides the guaranty that more relevant results will be retrieved depending upon the meaning and relations of the words not a specific keyword.

Ontology helps to provide related association between the contents. The Semantic Web is an enhanced version of the Web where information is represented in a machine process able way [8]. While the information on the Web is mostly represented as HTML documents, RDF (Resource Description Framework) [9] and OWL (Web Ontology Language) [10] are used for Semantic Web documents. The Semantic Web will contain not just a single relation between resources, but several kinds of relations between different types of resources. Semantic search engine like Hakia[11], Swoogle[12], DuckDuckGo[13] etc. are different from conventional search engine is that the semantic search engines are meaning based.

A semantic search engine stores semantic information about Web resources and is able to solve complicated queries. Semantic search integrates the technologies of Semantic Web and search engine to improve the search results gained by current search engines and evolves to next generation of search engines built on Semantic Web. The inventor of WWW, Tim Berners-Lee states that semantic web is not a separate Web but it's an advanced of current web in which information is well structured and well-defined.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

The layers of semantic web architecture represented [14] in Fig.1 are briefly described below:

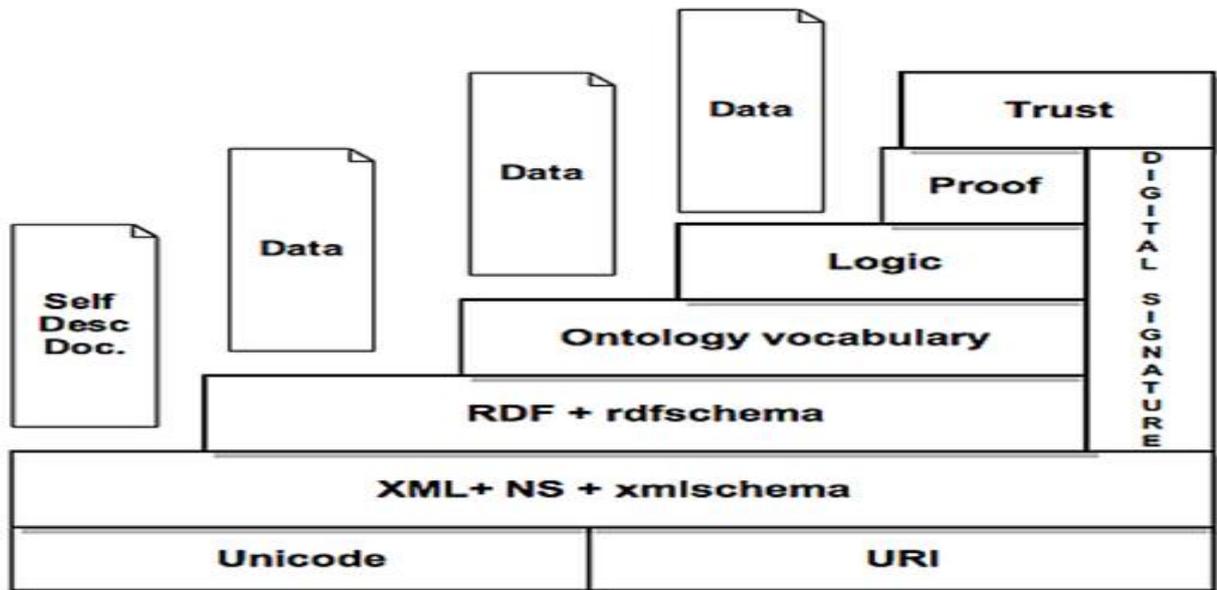


Fig.1 Semantic Web Architecture

- **URI and Unicode:** URI and Unicode are used for identification and location of resources. The URI is used to give a unique name to each resource. Unicode is the standard for computer character representation.
- **Extensible Markup Language (XML):** A markup language, which means that it is machine-readable and has its own format. It is widely known in the WWW community because it has a flexible text format and was designed to describe data and to meet the challenges of large-scale e-business and electronic publishing; it plays an important role in exchanging different types of data on the Web. In fact, it is the basis of a rapidly growing number of software development activities. Each document starts with a namespace declaration using XML Namespace.
- **The Resource Description Framework (RDF) :** The first layer of the Semantic Web. RDF is a framework for using and representing metadata and describing the semantics of information about resources on the Web in a machine-accessible way. URI are used for identification of resources, and a graph model is used to identify and provision of relations between the resources. Simple modeling languages such as RDF Schema are used for description of classes of resources. , it also provides a simple reasoning framework for inferring types of resources.
- **Ontology Vocabulary:** A language which provides a common vocabulary and grammar for published data as well as a semantic description of the data used to preserve the ontologies and to keep them ready for inference. Ontology means describing the semantics of the data, providing a uniform way to enable communication by which different parties can understand each other.
- **Logic and Proof:** In the Semantic Web, the building of systems follows a logic which considers the structure of ontology. A reasoner could be used to check and resolve consistency problems and the redundancy of the concept translation. A reasoning system is used to make new inferences.
- **Trust:** Final layer of the Semantic Web. This component concerns the trustworthiness of the information Web.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

## III. COMPARISON

For comparing keyword and semantic based search engine, similar search query is typed in both the search engines such as Google as a keyword search engine and DuckDuckGo as a semantic search engine. Here, Polysemy words “Fast” is used for search query.

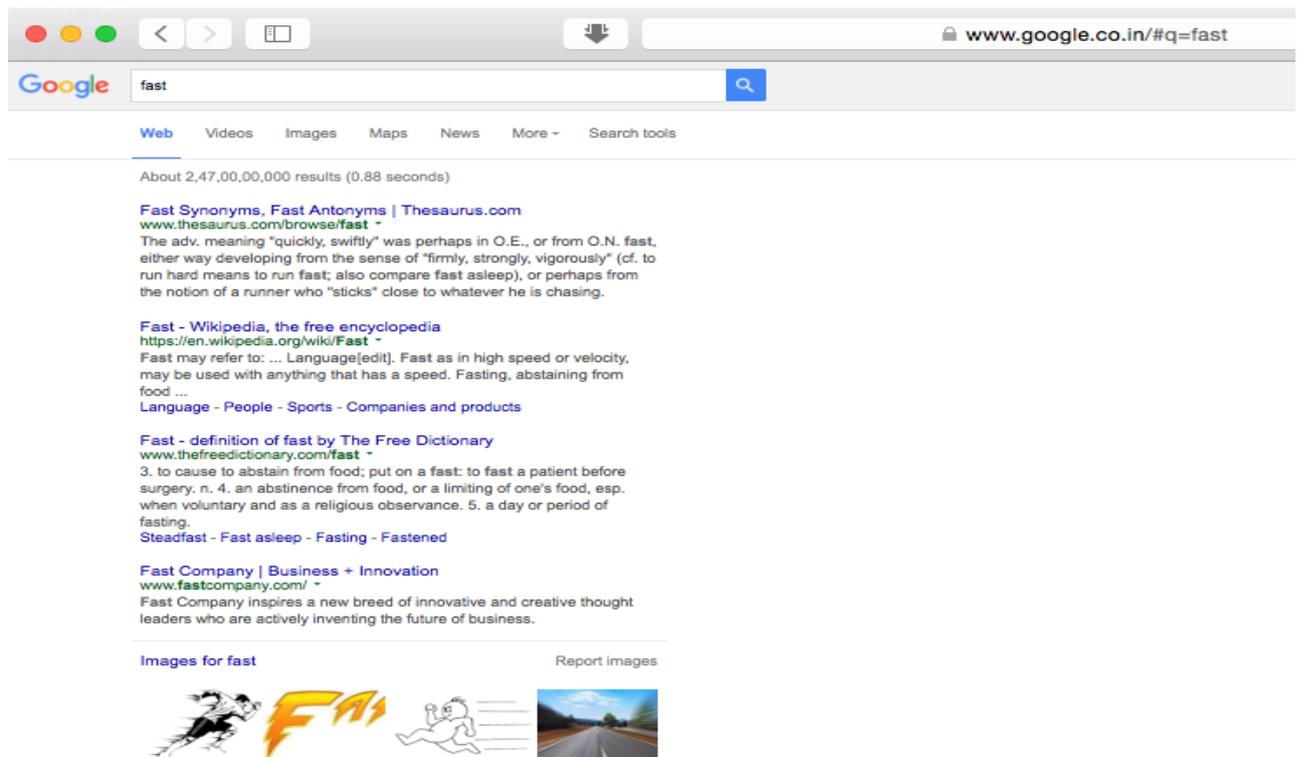


Fig.2 Google search engine

In above figure, Google search engine provide a basic search result such as FAST synonyms, Wikipedia and one of company whose name is Fast. It doesn't provide us a relevant and accurate result. There are two meanings of Fast but here in Goggle search only one meaning is displayed which is quickly, other meaning is not mentioned which is one of drawback of keyword search engine.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

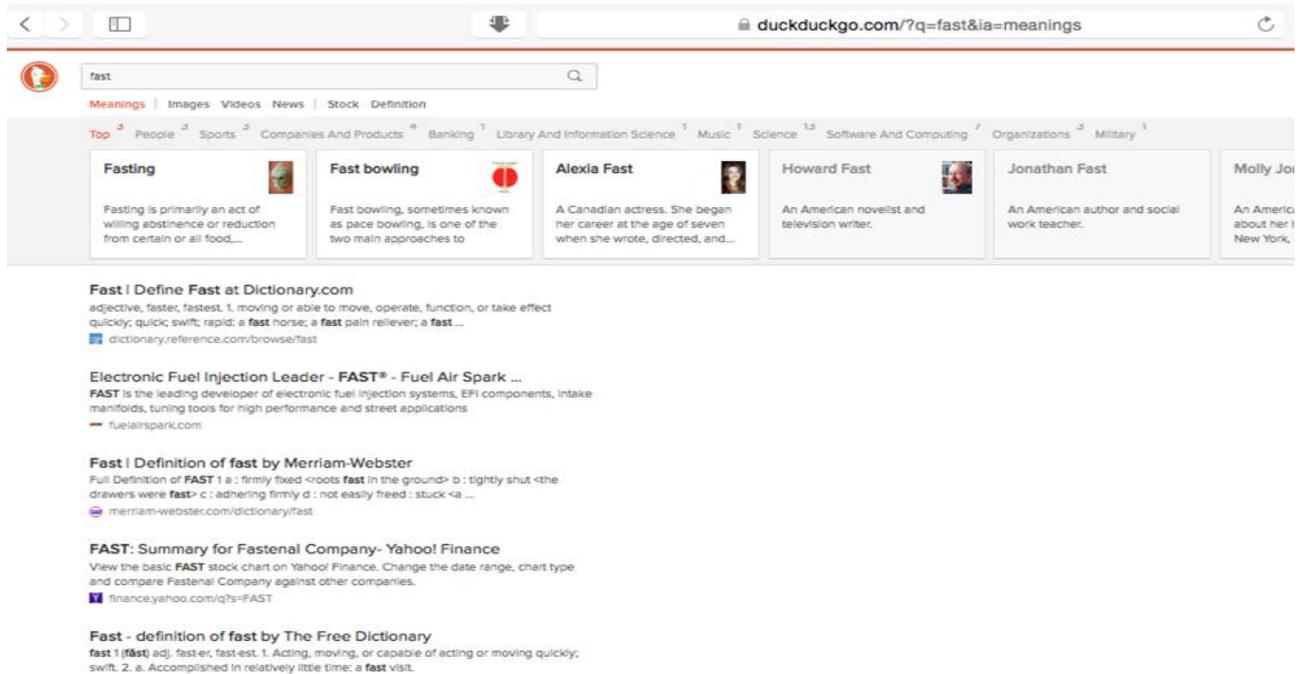


Fig.3 DuckDuckGo search Engine

In above figure, using semantic based search engine, most relevant and accurate results of given query is retrieved. Hence, from the figures 2 and 3 it is clear that Google is not able to handle Polysemy words while DuckDuckGo which is a Semantic Search Engine provide results in all possible meaning to the word “Fast”

TABLE I  
COMPARISON BETWEEN KEYWORD AND SEMANTIC SEARCH ENGINE

Keyword Search Engine	Semantic Web Search Engine
1. It is a traditional search engines that produce results of given query within the given context.	1. It works on Semantic based approach which is useful for having accurate and relevant information about the given query.
2. The information which is retrieved is dependent on keywords and page ranking algorithms that can produce spam results.	2. The information retrieved is independent of keywords and page rank algorithms that produce exact results rather than any irrelevant results.
3. It does not focus on stop words like is, or, and, how because it does not give accurate results what user is searching to get information.	3. It focuses on stop words and punctuation marks because it takes into account each and every small character as it affects search results.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2015

4. 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.	4. It will show only those results that will answer our query.
5. It does not highlight any words or phrases which are useful in answering getting accurate results.	5. It highlights the sentences or words that give answer to query asked by the user.
6. It makes use of keywords to expand query instead of using any methodology.	6. It uses ontology to get relations between the keywords.
7. It uses HTML, XML language for creation of metadata.	7. It uses Semantic Web languages like OWL, RDF for creation of metadata.

Above table shows the basic difference between keyword and semantic based search engine which is useful for studying significance of semantic based search engine over keyword search engine. By comparing both the search engine it is clear that semantic search engines have more advantages than keyword search engine.

### IV. CONCLUSION

This paper concludes that Semantic base search engines have more advantages over keyword search engine in terms of accuracy of getting results. Search process in the semantic search engine is based on the semantics of the query. It provides an assurance to the user to get more accurate and relevant results based on the meaning of the word that is being searched by the user, instead of page rank algorithms and keywords. User can retrieve relevant results using semantic based search engine. This paper also lists the basic differences between Keyword search engine and Semantic based search engine and provides a clear idea about significance of semantic search engines over keyword search engine.

### REFERENCES

- [1] Jagendra Singh, Dr. Aditi Sharan, "A Comparative Study between Keyword and Semantic Based Search Engines", <https://www.rgpv.ac.in/iccbdt/Papers/CL-112.pdf>
- [2] Website: [www.netmarketshare.com/search-engine-market-share.aspx?qprid=4](http://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4) accessed on 3rd November, 2012.
- [3] A. Gulli and A. Signorini, "The Index able Web is more than 11.5 billion pages," 2013.
- [4] Efrati and Amir, "Google Gives Search a Refresh," The Wall Street Journal, 2012.
- [5] Guha, R. McCool and Miller, "Semantic Search," 2011.
- [6] W. Roush, "Search beyond Google," Technology Review, 2004.
- [7] G. Antoniou and V. Harmelen, "A Semantic Web Primer," MIT Press Cambridge, Massachusetts, 2004.
- [8] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web," Scientific American, 2001.
- [9] Website: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210>, 2004.
- [10] Website: <http://www.w3.org/TR/owl-features>, 2009
- [11] Website: <http://www.hakia.com/>
- [12] Website: <http://swoogle.umbc.edu/>
- [13] Website: <http://duckduckgo.com/>
- [14] Mohammad Mustafa Taye, "Understanding Semantic Web and Ontologies: Theory and Applications", [arxiv.org/pdf/1006.4567](http://arxiv.org/pdf/1006.4567)