

Information Retrieval on Web: Ontology Based Vs Traditional Search Engines



Disha Grover, Barjesh Kochar

Abstract: Information Retrieval has become the buzzword in the today's era of advanced computing. The tremendous amount of information is available over the Internet in the form of documents which can either be structured or unstructured. It is really difficult to retrieve relevant information from such large pool. The traditional search engines based on keyword search are unable to give the desired relevant results as they search the web on the basis of the keywords present in the query fired. On contrary the ontology based semantic search engines provide relevant and quick results to the user as the information stored in the semantic web is more meaningful. The paper gives the comparative study of the ontology based search engines with those which are keyword based. Few of both types have been taken and same queries are run on each one of them to analyze the results to compare the precision of the results provided by them by classifying the results as relevant or non-relevant.

Keywords : Information Retrieval, Ontology, Semantic Web

I. INTRODUCTION

The basic function of a search engine is to allow user to retrieve documents from the web on the basis of the queries fired by them. The most widely used search engines include-Google, Yahoo, and Bing etc. The approximate size of the web is 4.54 billion pages and around 1.8 billion websites are available over the internet. The information is so huge that it becomes very difficult for the user to retrieve relevant information. The primary reason being, the pages uploaded over the internet contains hundreds of keyword to attract the user on the search engine and they may contain an advertisement or a video which do not satisfies the user. Hence, there's the need of search engines which are based on semantic web, an extension of the current web. The base of the semantic web is the concept of ontology. The ontology [1] works on the science of meaning so as to produce relevant results. The major limitation of the traditional search engine lies in their inability to understand the meaning of the keywords and expressions used by the user in the search query which may be due to the words having same meanings or the words having several meanings (polysemy). The

semantics search engine tries to make logic or sense of search results based on its context and is able to detect the concepts which structure the texts [2]. For instance, if you search for "elections" a semantic search engine might retrieve the resulting documents that contain the words "vote", "campaign" and "ballot", even if the word "elections" is not there in the source document.

An important property of this search is no ambiguity, in both, the queries and the content on the web. The search engine with the help of natural language processing techniques will understand if user is searching for a big cat or car when user searches for "Jaguar".

The semantic web is based upon Ontology. Ontology can also be defined as hierarchical illustration of classes and sub classes along with corresponding properties and instances. It has led to understand the concepts of the given domain, deriving relationships amongst them and showing them in a machine interpretable language [3]. Ontology in philosophical science means what is the object structure, events, properties in all areas of reality [4]. Role of Ontology in Semantic Web is as below [5], [9]:

- Ontology is considered to be the base of semantic web which plays key role in representing purposeful information from large volume of unstructured data on web
- It eliminates word sense disambiguation (WSD)
- It reuses and analyzes domain knowledge
- It helps to achieve interoperability amongst semantic web

II. LITERATURE REVIEW

The father of the web Tim Berners-Lee coined the different versions of web, as explained below:

A. Web 1.0

This was the crawling phase of the web. It was mostly Read-only and HTML based, the authors were limited and the information was static in nature. The interaction between the sites and the web users was also limited. The users did not ave the option of giving feedbacks or posting any reviews.

B. Web 2.0

This phase was the writable phase which included interactive data. It facilitated the interaction among the sites and the users. It encouraged information sharing. It is a communicative web and is based on XML.

Manuscript published on November 30, 2019.

* Correspondence Author

Disha Grover*, IT Department, Jagan Institute Of Management Studies, Delhi, India. Email:disha.grover@jimsindia.org

Dr Barjesh Kochar, IT Department, Vivekananda Institute of Professional Studies, Delhi, India.Email:bkkochar@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

C. Web 3.0

It is the executable phase of WWW, the applications are dynamic. It includes machine to machine interaction. The machines are capable of interpreting information like humans, intelligently. This is a semantic web, an extension of the current web, which is based upon individuals and the content is made available as per the users need.

The base of the semantic web is ontology where the data is stored in the form of triples and have relationships among them. Languages such as Resource Description Framework support Semantic Web [6], [7]

Table- I: Differences between WWW and Semantic Web

S. no	World wide web (WWW)	Semantic web (SW)
1	It is human focused web	It's a www extension that will automatically manipulate information without any human involvement.
2	It discovers documents for humans.	It discovers documents for collecting relevant information
3	It only deals resources related to media such as web pages or photos or images.	It deals with resources like pages, images, photos and people.
4	WWW keeps hold of hyperlinks between resources	SW holds different kinds of relations showing association among various resources.
5	It does not use ontology concepts.	SW makes use of ontology which helps users organize information
6	It does not have formal semantics of context. The contents are readable by machine but not understandable	SW uses web ontology language to generate data i.e. It has formal semantics of context
7	Traditional search engines (e.g. Yahoo) can access only few pages of information.	Semantic search engines (e.g. Bing) can access complete information.

III. METHODOLOGY

The search engines based on ontology are compared with the conventional ones. Few from each category are taken and compared on the basis of the relevant results they provide. Ontology based search engines Google, Bing and DuckDuckGo and traditional engines Yahoo and Askme.com are used for comparison [8]. Set of ten user search queries in natural language are taken and executed on each of the selected search engines. The top ten results/URLs retrieved are analyzed manually and are marked relevant or non-relevant to obtain the relevancy metrics.

Formula used: Relevancy Metric = No. of Relevant results retrieved/Total no. of relevant & non relevant results.

Table- II: List of queries for analysis

Query #	Description
QUERY1	Which country won the first cricket world cup?
QUERY2	Who is the prime minister of Pakistan?
QUERY3	How to cook shahi paneer?
QUERY4	What to eat in fever?
QUERY5	When is the best time to introduce food to a baby?
QUERY6	Which laptop is best?
QUERY7	How to drive a car?

Query #	Description
QUERY8	Who is Narendra Modi's wife?
QUERY9	What is magnetic field?
QUERY10	How steam engine works?

IV. OBSERVATION RESULT

In the analysis performed, it was observed that the ontology based search engines i.e Google, Bing and DuckDuckGo gave much precise and accurate results when they were compared with the non-ontological keyword based search engines like Ask.com and Yahoo.

Table-III: Comparative analysis of ontology based and non-ontological search engines

Query#	Ontology Based			Non Ontology based	
	Google	Bing	DuckDuckGo	Ask.com	Yahoo
Query1	0.6	0.9	1.0	0.5	0.3
Query2	0.8	0.6	0.7	0.7	0.3
Query3	1.0	1.0	1.0	0.6	0.8
Query4	0.9	0.7	0.8	0.9	0.5
Query5	1.0	0.8	0.8	0.3	0.3
Query6	0.9	0.7	0.6	0.5	0.5
Query7	0.7	0.7	0.8	0.6	0.4
Query8	0.7	0.9	0.8	0.6	0.5
Query9	0.9	0.8	1.0	0.5	0.6
Query10	1.0	1.0	0.9	0.6	0.6
Average Relevancy Metric	0.85	0.81	0.84	0.58	0.48

The average relevancy metric, which quantifies the relevance of the information retrieved by the ontological model, is found to be highest in Google followed by DuckDuckGo and Bing. The Search engines which do not use ontology and primarily use traditional keyword based information retrieval approach i.e ask.com and yahoo, have comparatively very low average relevancy metric as depicted in the table above.

The average relevancy metric for each engine is graphically plotted below, which explains that Ontology provides highly relevant results. The search engines based on ontology have come out to be more relevant as compared to the ones which are based on traditional approach.



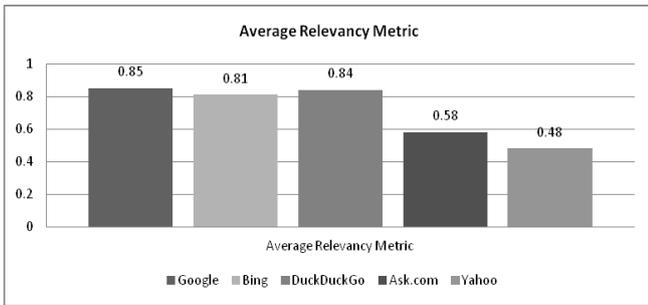


Fig. 1. Average Relevancy Metric

The below graphs shows query wise comparison.

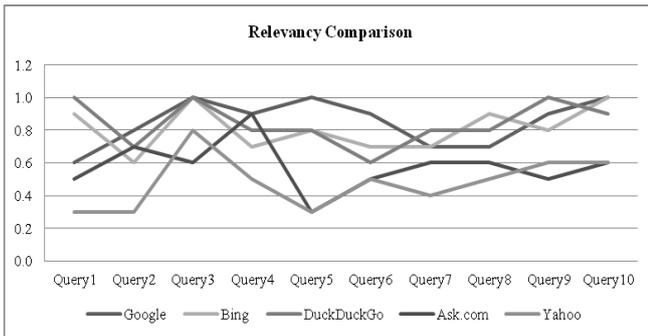


Fig. 2. Relevancy Comparison

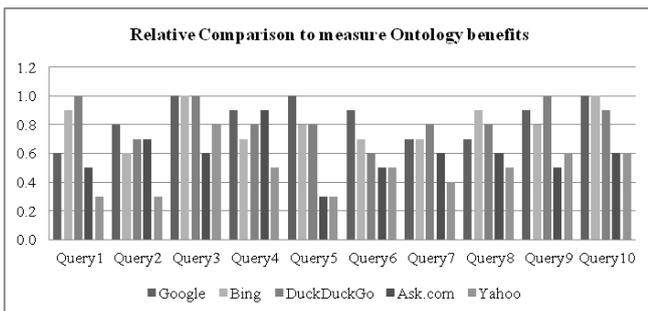


Fig. 3. Comparative analysis to show ontology better

V. LIMITATIONS

The survey was carried over on the limited data set that is only few queries and 5 search engines; this can be extended further on the extended data set with large number of queries from extended variety of users from different domains on large number of search engines.

VI. CONCLUSION

Based on the above analysis for the information retrieval, it can be said that the Ontology based semantic search engines give better performance with respect to retrieval of relevant results as compared to the traditional keyword based search engines. The search engines - Google and Bing have produced highly relevant results in contrast with the other search engines i.e. ask.com and yahoo.

REFERENCES

1. A. S. Amjad Farooq, "Ontology Development Methodology for Semantic Web Systems," *Pakistan Journal of Life and Social Sciences*, vol. 6, no. 1, pp. 50-58, 2008.
2. N. D. A. S. Ranjna Jain, "Comparative Study on Semantic Search Engines," *International Journal of Computer Applications*, vol. 131, no. 14, 2015.

3. J. F. Sowa, "Ontology, metadata, and semiotics Lecture Notes in AI #1867, Springer-Verlag, Berlin,," in *Lecture Notes in AI #1867, Springer-Verlag, Berlin*, 2000.
4. B. K. Y. S. Disha Grover, "Introduction of Ontology based Query Retrieval Mechanism in Cloud Computing Architecture," *International Journal of Computer Applications*, vol. 125, no. 10, pp. 37-42, 2015.
5. G. S. N. N. D. V. J. Usha Yadav, "Ontology Engineering and Development Aspects: A Survey," *International Journal of Education and Management Engineering(IJEME)*, vol. 3, pp. 9-19, 2016.
6. V. J. Gagandeep Singh, "Information Retrieval (IR) through Semantic Web : An Overview," in *Confluence 2012 – The Next Generation Information Technology Summit*, 2012.
7. S. S. K. B. Prabhakaran, "Study on Ontology used in Cloud Computing," *International Journal of Societal Applications of Computer Science*, vol. 2, no. 12, pp. 491-495, 2013.
8. A. Radhakrishnan, "8 Semantic Search Engines That Will Change the World of Search," 2009.
9. V. J. M. S. Gagandeep Singh, "Ontology Development using Hozo and Semantic Analysis for Information Retrieval in Semantic Web," in *IEEE Second International Conference on Image Information Processing (ICIIP-2013)*, 2013.

AUTHORS PROFILE



Disha Grover is associated with Jagan Institute of Management Studies, Rohini affiliated to Guru Gobind Singh Indraprastha University - Dwarka, from past 10 years. Currently, she is working as an Assistant Professor in the field of Information Technology. She is pursuing her PhD. from Jagannath University, Jaipur in Cloud Computing. She has multiple research papers published in various national and international journals to her credit. She also has keen interest in the field of Database and Robotics.



Dr Barjesh Kochar is a PhD from Maharishi Dayanand University, Rohtak. He is currently working as a Professor at Vivekananda Institute of Professional Studies, Pitampura affiliated to Guru Gobind Singh Indraprastha University – Dwarka. Having worked with various reputed academic institutes across India, Dr Kochar has over 25 years of work experience including industrial and academia. He has more than 50 national and international publications to his credit. He has keen interest in the field of Distributed Systems.