International Journal of Research and Review
www.ijrrjournal.com E-ISSN: 2349-9788; P-ISSN: 2454-2237

Review Paper

An Overview of Semantic Search Engines

Subham Roy, Akshay Modak, Debabrata Barik, Surajit Goon

Department of Computer Application, Eminent College of Management & Technology, Barasat, Kolkata, West Bengal 700126, India

Corresponding Author: Debabrata Barik

ABSTRACT

Semantic Search is a search technique that improves searching precision by understanding the purpose of the search and the contextual significance of words as they appear in the searchable data space, whether on the web to generate more relevant result. We highlight here about Semantic Search, Semantic Web and discuss about different type of Semantic search engine and differences between keyword base search and Semantic Search and the advantage of Semantic Search. We also give a brief overview of the history of semantic search and its feature scope in the world.

Keywords: Semantic Search, Semantic Wed, Semantic search engine

INTRODUCTION

The word "Semantic" refers to the meaning or summary of something. "Semantics", applied to search, essentially refers to the study of words and its logic. It is a technique of data searching, but it is not only a search query to find keywords, but also to determine the intent and contextual meaning of the search words that a person uses. Therefore, it provides more significant search results by assessing and understanding the search phrase and discovering the most appropriate results in a website, database or any other data store. It operates on linguistic semantics principles. [1] Unlike typical search algorithms, semantic searching is based on the searched phrase's context, substance, intent and concept of the searched phrase. As part of the search, semantic search also organizes place, term synonyms, present trends, word variants and other natural language components. Semantic search ideas are obtained from different search algorithms and methodologies, including keyword-to-concept mapping, graph patterns, and blurred logic.

Web search is a major Web technology that was initially based on a combination of textual keyword search and document ranking depending on the Web's connection structure. That's why it has many constraints, and there are plenty of study operations toward smarter web searching, called Semantic Web search, which is presently one of the hottest study subjects in both Semantic Web and Web search. [2]

Search engine has become a primary need to explore the internet. Without Search Engine, there are no uses of information in website, blog, etc., because without search engine, it is almost impossible to look for one by one website just for search information in internet. Semantics is the study with meaning. It focuses on the meaningful relationship, such as words, sentences, signs, and symbols. The term was coined by the inventor of the World Wide Web, Tim Berners-Lee, who defined it as "a data web that can be processed by machines directly and indirectly". [3]

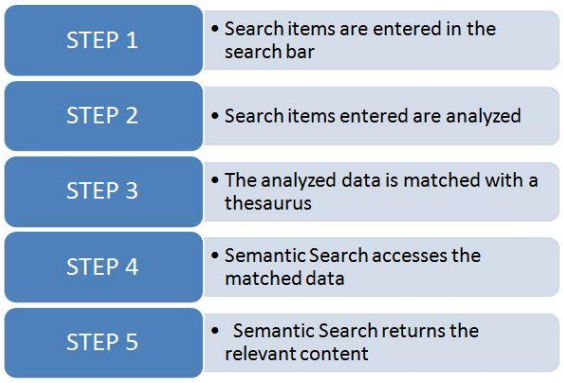


Figure 1: Sematic Search Process

The searching process of Semantic Search is divided into five processes as shown in Figure 1. In the first step the user entered the searching element in the search bar. In the next step the search engine analyzed the search items. In the third step the analyzed data is matched with a thesaurus. Then semantic search accesses the matched data and then semantic search engine returns the relevant content.

HISTORY OF SEMANTIC SEARCH

Semantic is the study and mining of language. It was first used by French philologist Michel Bréal in 1983. It is used to define how the words can be different meanings for distinct individuals because of their experiential backgrounds. For example, French, Dutch, or Hindi, any language can be a natural language, it can be an artificial language, such as a programming language for PCs. In 1967, Robert-Floyd wrote a paper describing the use of language Semantics in computers and received credit for starting the programming language Semantics field, and his work included the analysis and design of algorithms used to locate the most efficient paths in a network, quantile calculation, programming language parsing, and information sorting. [4,5] The featured article 'The Semantic Web' was released by *James Hendler, Ora Lassila*, and *Tim Berners-Lee* in May, 2001.Their paper described a new way to use and to search the Internet, an added dimension with full of new possibilities. Although the text of an HTML web page can be read by a human, but a computer or search engine cannot (unless tags are intentionally placed). This is because HTML is intended to store visual information and is not written in a language of programming. [6]

COMPARISON OVER KEYWORD BASED SEARCH

Similar search queries are typed in both search engines such as *Google* as a keyword search engine and *Hakia* as a Semantic Search engine to compare keyword and semantic search engine. But fundamental search results are provided by the Google search engine. It does not give us a precise and meaningful result. It is the drawback of Keyword Search Engine. In contrast, using Semantic based search engine, most appropriate and precise query outcomes are obtained. That means traditional search engines generate results of a specified query within a specified framework, but Semantic search engines operate on a Semantic based strategy that is helpful for precise and meaningful query data. [7]

In Keyword Search, the obtained data depends on keywords and page ranking algorithms that can yield spam outcomes without using any methodology and does not emphasis on stop words like is, or, and, how because it does not give exact results what user is looking for. It shows all web pages which may or may not fulfill the request of the user and it is a challenging job to select appropriate page from many pages. Keyword Search Engine does not emphasize any words or sentences that are helpful in responding to precise outcomes. It uses HTML, XML language for creation of metadata.

But in Semantic Search, the retrieved information is independent of keywords and page rank algorithms that generate exact results rather than any insignificant results. It utilizes ontology to establish keyword relationships and takes on stop words and punctuation marks because each and every tiny character is taken into consideration as it impacts search outcomes.

[7] It shows only those results that will answer our query so it does not highlight any words or phrases which are useful in answering getting accurate results. It uses Semantic Web languages like XML, OIL, DAML+OIL, DAML-S, OWL, RDF, WSDL, URI, UDDI for creation of metadata. [8] Metadata means data about any other data in a compiled manner. The user gets the information about any specific information sequentially or separately. It's not smart process. The primary objective of meta data is to search, but it can also be used for another objective.

Search result of Keyword Search Engine:

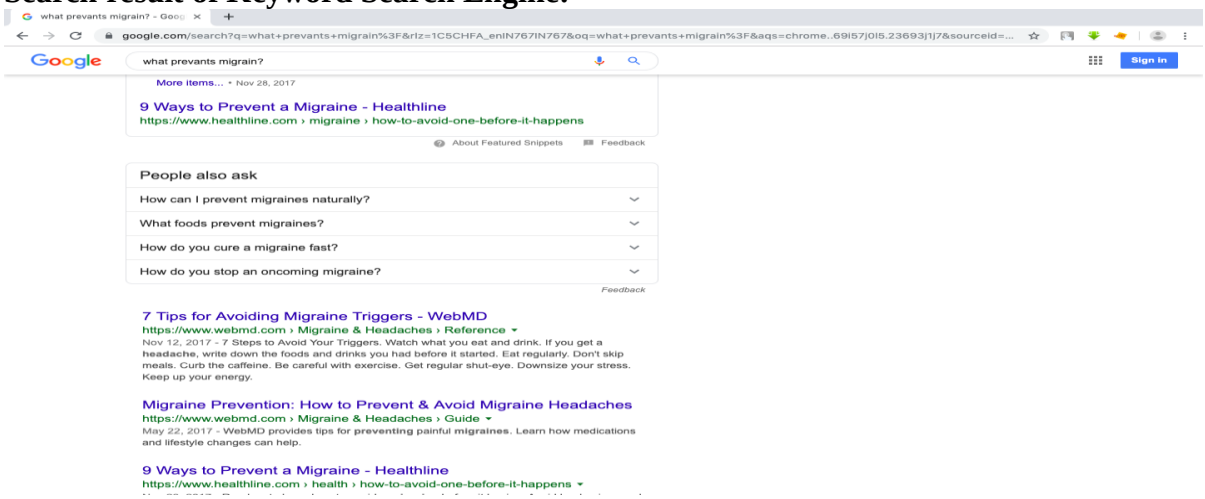


Figure 2: Keyword Search Engine

Search result of Semantic Search Engine:

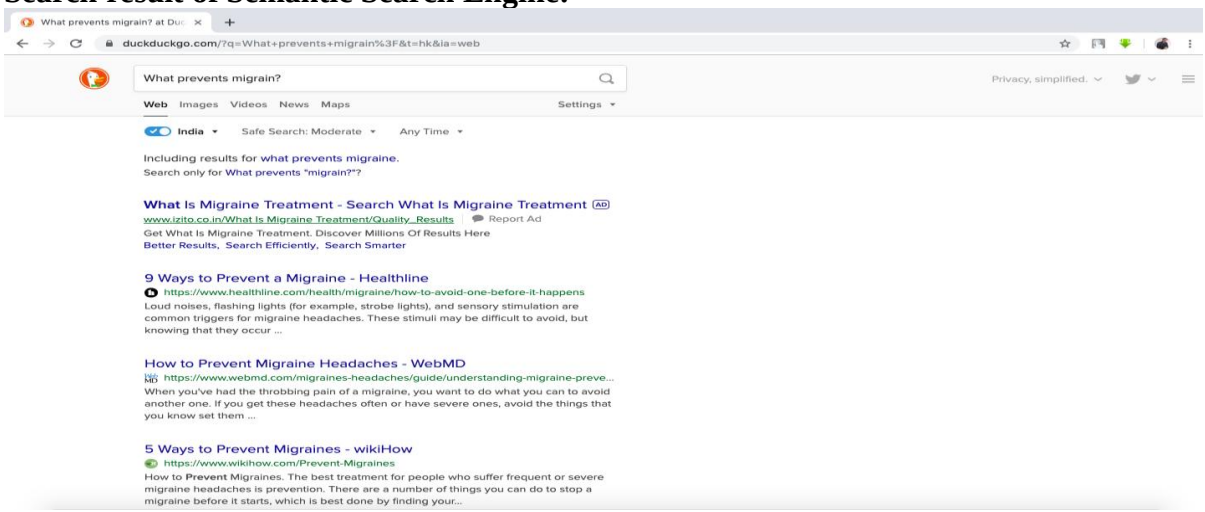
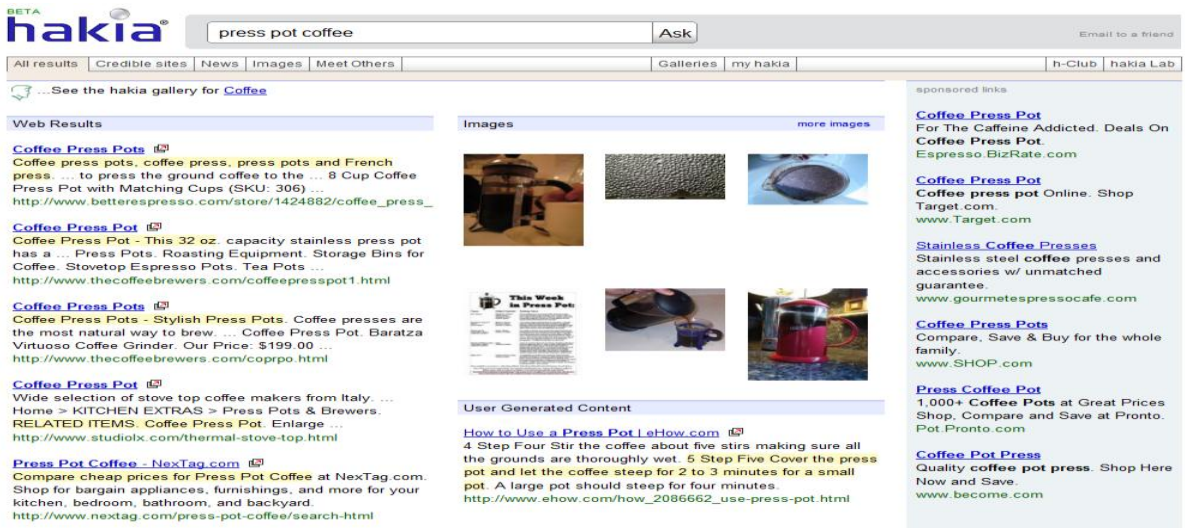


Figure 3: Semantic Search Engine

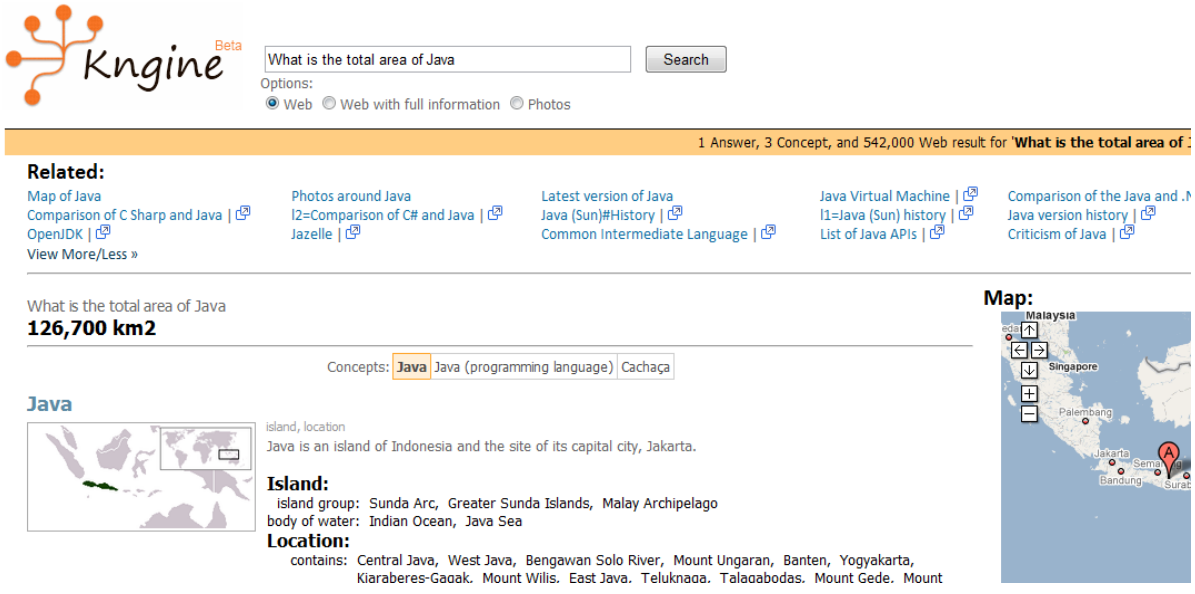
DIFFERENT TYPES OF SEMANTIC SEARCH

Hakia: Hakia is a semantic search engine that delivers appropriate outcomes depend on matching idea instead of matching keywords or prevalence ranking. The engine prompts not only keywords to enter- but a query, a phrasing, or a sentence. They cater search results based on meaning and not on the popularity of search terms. A very essential capability of hakia engine is that it projects results using equivalent terms. The search items are divided into Web, News, Blogs, video and can be arranged by relevance or date. Hakia's semantic search is consist of three technologies. OntoSem (sense repository) is a lexical database where words are classified into the diverse "senses" they impart. QDEX (Query indexing technique) evokes all queries

concerning the content. Semantic Rank algorithm independently orders content. It gives almost exact results collected from reliable sites in less time compare to Keyword Search Engine. [9]

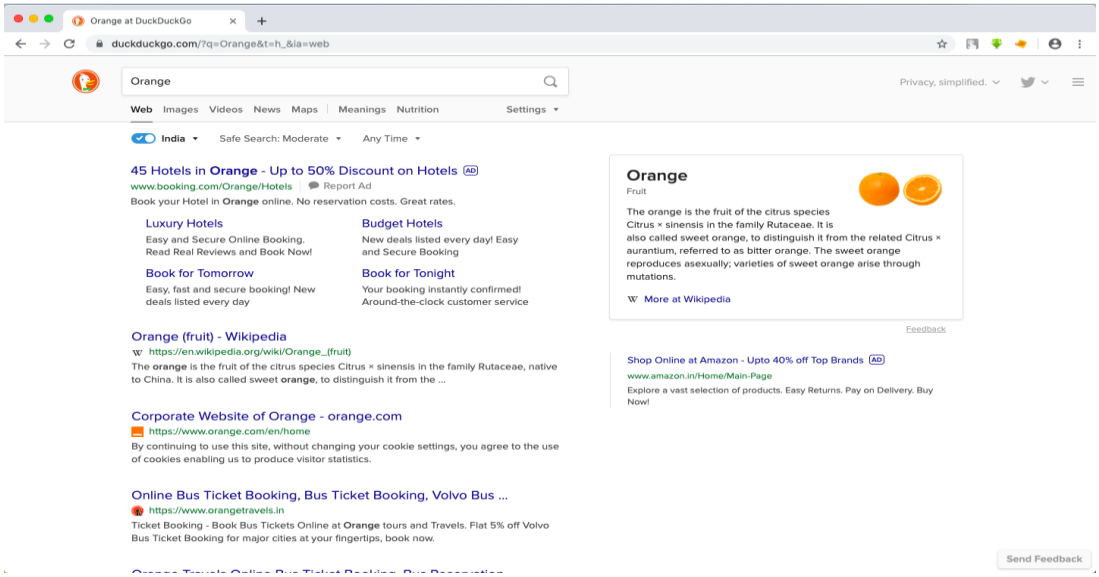


Kngine: Kngine is a smart engine that answers de facto questions commonly posed and carries out actions. Kngine recognizes what users are truly looking for and delivers significant outcomes to them. You can speak or type your question for Kngine. It extracts the factual data from each phrase to build / update our knowledge graph. Kngine results are based either on web results or on image result. In order to achieve state-of - the-art accuracy, Kngine also uses the power of deep learning, big data and unsupervised learning. It learns and improves all the time. Currently, it includes over eight million Concepts where the power of the site resides. [10,11]

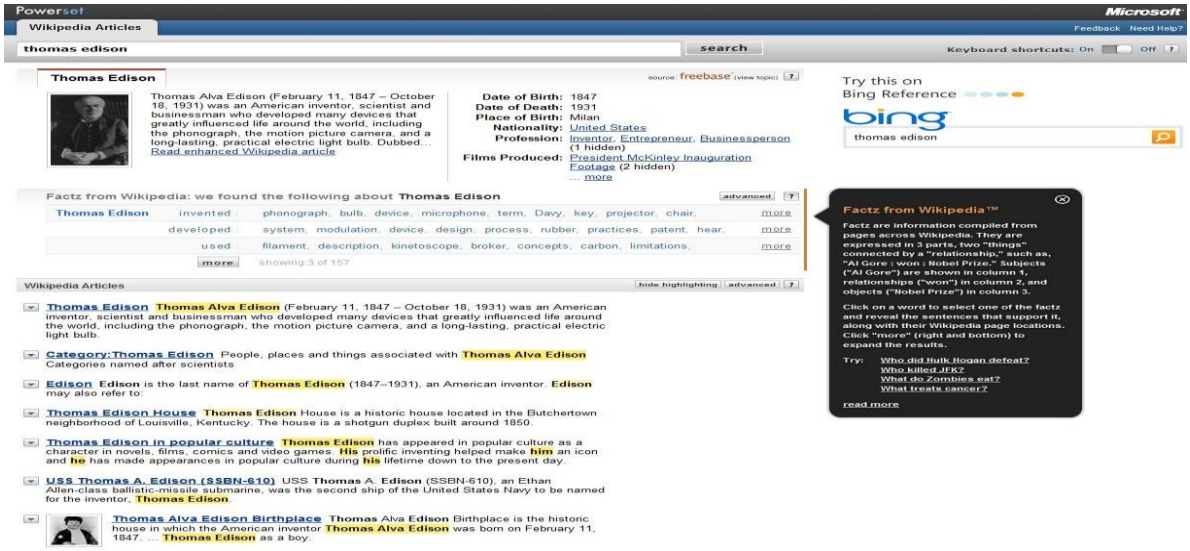


Kosmix: Kosmix is acts as guide to Web. It allows users to browse the Web topic wise and to present a dashboard of appropriate Videos, Images, News, Opinions, Blogs, Forums, Twitter, Amazon, Facebook and references to related topics. The categorization engine of Kosmix arranges the Internet into pages of magazine-style based on the topic. It also uses Deep Web. [12,13]

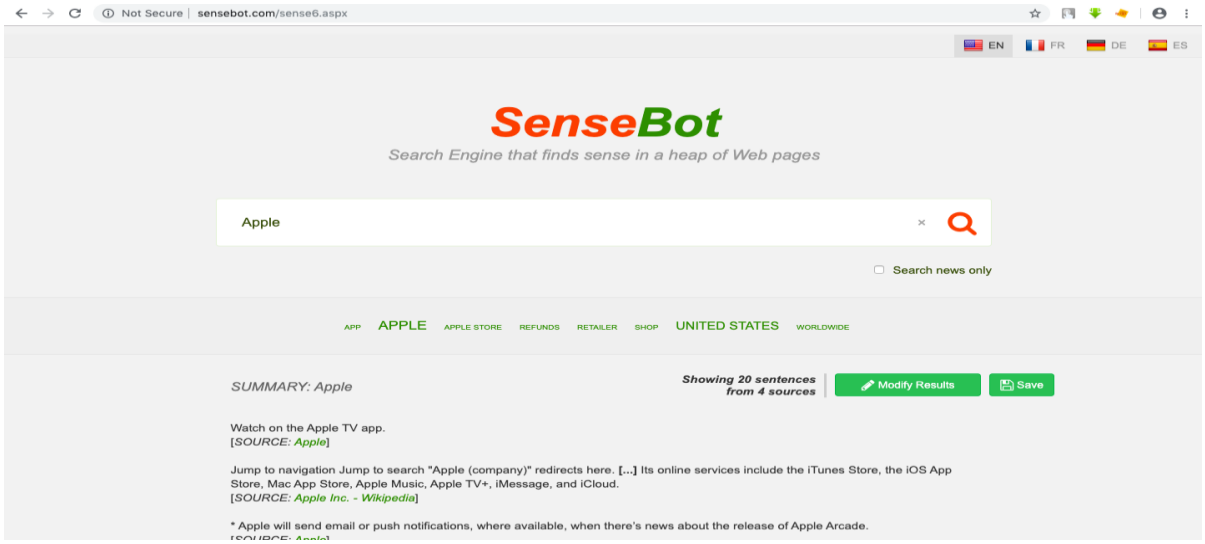
DuckDuckGo: If you use Google, forget it, because unlike Google, DuckDuckGo is a feature-rich Semantic Search Engine. If you search for a term that has more than one meaning, with its disambiguation results, you will be able to select what you originally searching for. For example, if you browse for the term Orange, it will give list of possible meanings including fruit, online reservation system, corporate site, business service, etc. [13]



Powerset: In Sep 2005, the firm introduced with the aim of making search easier and more intuitive. Later, Microsoft acquired it on July 1, 2008. It concentrates on doing just a single thing and doing it really well by using natural language processing to understand the nature of the question and return pages containing the answer. It comprises all the search results from Wikipedia, using semantics Search terms can be articulated as questions, which will be answered, or as simple terms, and results will be accumulated from all the significant pages on Wikipedia from different resources. [14,15]



Sensebot: Sensebot utilizes text mining to analyze web pages and recognize important semantic ideas. It then conducts a multi-document overview of content to generate a consistent overview which, depending on the query provided, gives a summarized precise search outcome. The synopsis gives A nice concept of the subject of the query. [15] The overview can be read and is consistent. It saves time. To get the outcomes, the user does not need to go through many web pages. The search engine itself attempts to comprehend the query idea, in fact what it includes and provides a suitable outcome. [16]



Cognition: Cognition finds meaning formula in the quest. It provides Link results. It is a search engine that provides the ability to learn about various elements of distinct languages and promotes Ontology, Morphology, and Synonyms. The use of this technology could vary from better enterprise-wide search to more appropriate advertising. It offers access to this technology with APIs. [17]

Swoogle: A crawler-based indexing and retrieval Semantic Web. It analyzes the documents it has found in order to calculate beneficial metadata characteristics and their relationships. [18] Identified documents are likewise recorded by a data recovery framework which can utilize either character N-Gram or URIrefs as keywords to search specific documents and to figure out similarities among number of documents. [19] By using text-mining and multi-records summarization, it extracts meaning from web pages. It finds the semantic web's suitable ontology and instance data structure. [20]

Factbites: Factbites searches for material that is factual and accurate. To abstract meaning from web pages, it utilizes text mining and multi-record summarization. [21] The outcome provides us a feeling of carefully and precisely understanding stuff. It is best to filter out those non-relevant sites. [22]

COMPARISON OF DIFFERENT SEMANTIC SEARCH ENGINES

The following Table 3 shows the comparison of different Semantic Search Engine.

Name	Features	Search Methodology	Result Summary	Type of Result	Search Results	Result Explanation	Multilingual	Advantages
Hakia	Excellent summaries, related searches, algorithm of semantic rank, CMR	Pure content analysis.	The content of the documents is important	It leads to Link & Free Text	The outcomes of the search are divided into Web, News, Blog, Videos	Yes	Yes	It easily collects information appropriate to our query from various reliable sites. It identifies Focus and Saves time information
Kngine	It answers de facto questions commonly posed and carries out actions, it learns and improves all the time	It uses the power of deep learning, big data and unsupervised learning	It extracts the factual data from each phrase to build / update our knowledge graph	It also results in a summary form	web results or image result	Yes	Yes	You can speak or type your question for Kngine
Kosmix	It has features such as Digg, Buzz, Flikr, Fark, and Youtube on its first page. Article of the function of the image	Classification of contents	In Search Query, it gives significance.	It provides the results of the search	Amazon, Facebook, Video, Websites, News, Blog, Images, Forums	No	Yes	It organizes its user's internet outcomes
DuckDuckGo	Emphasizes privacy, it does not monitor private data of the user, it results in the use of other sources or search engines including its own web crawler	It is a Meta search engine capable of collecting data from other search engines	Gives user query text / content-related results	It results in a summary form	Images, local search, self-suggestions, news, weather, cooking recipes etc.	Yes	Yes	It accumulates results from various sources like Yahoo and Wikipedia, etc.
Powerset	It helps to search results in more easy and intuitive way	It uses natural language processing to understand the nature of the question and return pages containing the answer	Multiple web pages generate a text overview	It provides Link results	results are gathered from all the significant pages on Wikipedia from different resources	Yes	Yes	It concentrates on doing just a single thing and doing it really well by accumulating search results from various sources.
Sensebot	It provides overview in the result of search query, text mining, multi-document summary rather than links to other website pages. It discovers top search results and then summarizes this information	To abstract meaning from webpages, it utilizes text mining and multi-record summarization	Multiple web pages generate a text overview	It results in a summary form	Summary of all web pages on our subject	Yes	Yes	It provides us a summary of the outcomes rather than all connections associated with our question
Cognition	It is a linguistic search engine that uses Ontology, Synonyms and Morphology	Processing of natural languages	Retrieves the search meaning formula.	It provides Link results.	It allows 4 domains for searching: Law, Medicine, Wiki, Bible	Yes	Yes	It is a search engine that allows us to learn about various elements of distinct languages and promotes ontology, morphology and synonyms
Swoogle	It offers various search services using the REST interface	Indexes records that use RDF	Gives outcomes from the semantic	It provides O WL, RDF re sults	Online ontology, Documents, Terms,	No	No	It discovers the suitable ontology and instance of the semantic

			web		Published web-based data			web's data structure
Factbites	Search for questions and answers, search for dictionaries, authentic and relevant search / content	Searches for authentic and informative material	To abstract meaning from web pages, it utilizes text mining and multi-record summarization	It provides results that are summarized, oriented and appealing	Coherent phrases and connections	Yes	No	The outcome provides us a feeling of carefully and precisely understanding stuff. It is best to filter out those non-relevant sites

STATISTICAL ANALYSIS

A Semantic Search Engine, namely, DuckDuckGo and a Keyword based Search Engine, namely, Google were selected to compare the search results. Subsequently, five topics which consist of one or two terms and five natural language queries were randomly chosen as shown in Table 1. Then, the first ten results of searched queries were assessed and the relevant results out of ten were listed individually for each search engine as shown in Table 2.

RESULT

In our study, it was seen that relevant document searched by DuckDuckGo is more (76 out of 100) than Google. Figure 4 depict the mean precision ratio of search engine for first 10 results and Figure 5 represent no of related document searched based on given topics.

Query no	Query
Q1	Deep Web
Q2	Atom
Q3	Windows
Q4	Orange
Q5	Firewall
Q6	What is the weather of Kalimpong
Q7	Which country won first hockey World Cup
Q8	What is the GDP of India
Q9	Why the plastic should be banned
Q10	What is total no of tree in the world

Table 1: Query List

Query no	Google	DuckDuckGo
Q1	7	9
Q2	6	4
Q3	7	10
Q4	0	2
Q5	8	9
Q6	7	6
Q7	8	8
Q8	8	9
Q9	10	10
Q10	7	9
Total	69	76
Average	**69%**	**76%**

Table 2: No of related document searched

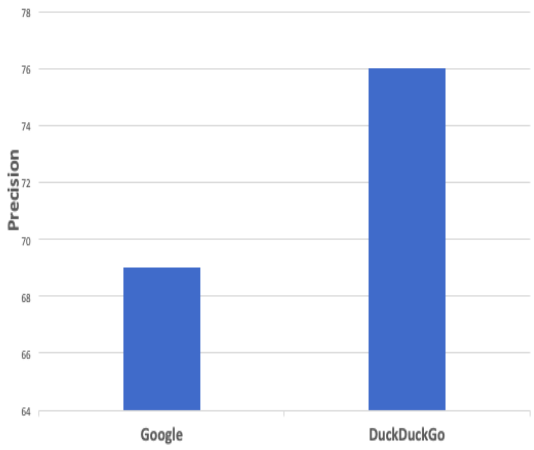


Figure 4: mean precision ratio of search engine

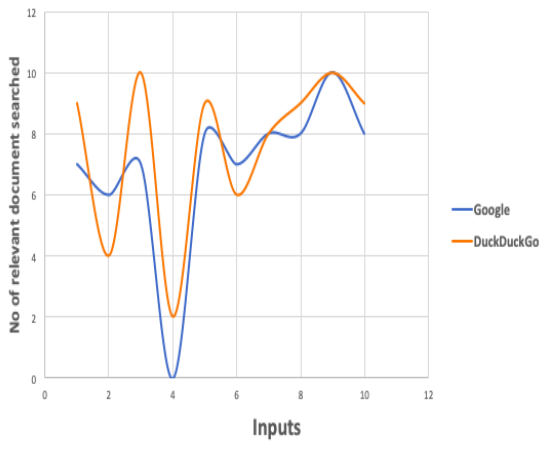


Figure 5: no of related document searched for 10 topics

MAJOR ADVANTAGES OF SEMANTIC SEARCH

i. The main advantage originates in the extra semantic relationships that search engines deliver better results during query routing. These relationships offer a more dynamic, communicative and dialog-based result pages or SERPs. Semantic search technologies enable individuals to monitor information by idea rather than by a defined match of keyword or key phrase. This implies that people can differentiate more easily what's going to be on a page while choosing which one to click to. Search engines have everything to do with what the user wants. The semantic strategy is more effective in ensuring that people do not become dissatisfied after turning up somewhere SERPs recommend. [23]

ii. Semantic search takes user intent and user data into account. Google's machine-learning algorithm has trained itself to interpret what you really want based on the aggregate data of millions of searches. Perhaps this is the most important thing about searching semantic. The best results are not those with lots of keywords, an optimized H1, and a well-designed title tag. [24]

iii. A rich semantic network not only recovers query-related documents, but also information that is conceptually similar. Where keyword and statistical technologies for higher precision use synonyms and other techniques and recall research approach results in more false results and errors. The idea behind the technology enabled by semantic overcomes these difficulties because, through the strength of a semantic network, it eliminates the ambiguities that get the real meaning of the terms. [25]

iv. Whenever you're happy with anything, try throwing in some related words, for example-If you're reading a Panda bear article, you're likely to find words like "Bamboo, China, Animal and Mammal" in that particular article. If you can identify the right words and include them in your search engine, transmit the subject of your content and thereby promote visibility. [26]

v. The technique of using a semantic strategy and structured information for content planning uses a data model to characterize the information. *Jarno van Driel* focuses on the importance and authority of using structured information for company intelligence and provides case studies and how-to. Read proof that this approach benefits locations that incorporate semantic advantages. Semantic information produces data recovery a progressively beneficial procedure.

vi. How Internet users find the content they want is dynamically advancing towards a semantic “shape” to reach that data. One way to clarify it is that the natural ordering and logic of web data has a semantic reason and job. Business owners now have more capacity to customize and control the text copy of your digital content.

vii. Organized information may look to the human eye like a lot of mixed words, but for search engines, it explains the words on a page. In the task of sorting out data, it is basic for *GoogleBot* to discover mechanisms to more precisely understand a web page’s content. By becoming more like authentic learning machines, search engines are interpreting content much quicker and more precisely. These new advances offer colossal advantage for search capacities to be robust in terms of scalability, efficiency, and resilience to failure from indiscernible search

queries. So, by using schematics, we make this easier for them. [27]

SOME CONTROVERSIES REGARDING SEMANTIC SEARCH ENGINES

i. The findings are sometimes incorrect and repeated over and over again.
ii. Identification of intention in such search engines is very important
iii. The user could enter the worldwide web at the beginning, i.e. they provided the choice of disambiguation. [28]
iv. Some of the smart semantic search engines that do not demonstrate the significance of accuracy and recall. Google is not a semantic search engine, but elevated accuracy and low recall. [29]
v. Usually the knowledge specific to the domain of the user. Users may not define all possibilities but the query's synonyms and variants have an issue that consumers are not sure how to use Phrase. [30]
vi. Trillions of information on the subject-based search was distributed to the World Wide Web only to extract the subject information in pages and produce an outcome for the user. But if we go for search engines based on semantic, it can create several choices for the keyword of the individual user. The metadata key-based search for the processing of web pages is not adequate for the individual Crawler. [31]

CONCLUSION

This paper concludes an overview of Semantic Search, Semantic Web and different Semantic Search Engines and advantages of Semantic Search. This paper gives a brief overview of comparison of different Semantic Search Engine so that anyone can do a further research on it. We get to know that it is the best and hottest research topic and some of the most promising research issues are how to automatically translate natural language queries into formal ontological queries, and how to automatically add semantic annotations to Web content, or alternatively how to automatically extract knowledge from Web content. Another central research issue in semantic Web search is how to create and maintain the underlying ontologies and a closely related important research challenge is the evolution and updating of and mapping between the ontologies that are underlying Semantic Web Search, where it is similarly desirable to have a very high degree of automation. A further important issue is how to consider implicit and explicit contextual information to adapt the search results to the needs of the users. Performing Web search through returning simple answers to simple questions in natural language is still science fiction, not to mention performing Web search in the form of query answering relative to some concrete domain or even general query answering. A number of future directions need to consider for this work. So, it will be the most important and powerful search technique in the world in future. The improvements suggested to be made in the future indicate the development of such an effective search engine search technology that is able to meet the difficulties effectively and is compatible with worldwide web technology standards.

REFERENCES

1. Guha, R., McCool, R., & Miller, E. (2003, May). Semantic search. In *Proceedings of the 12th international conference on World Wide Web* (pp. 700-709). ACM.
2. Fazzinga, B., & Lukasiewicz, T. (2010). Semantic search on the Web. *Semantic Web*, *1*(1, 2), 89-96.
3. Jayavel, S., Anouncia, M., & Kapoor, A. (2013). Semantic Search Engine. *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, *1*(2), 19-21.
4. Berners-Lee, T. (1998, September). Semantic Web Road map. Retrieved from https://www.w3.org/DesignIssues/Semantic.html

5. Manjula, D., & Geetha, T. V. (2004). Semantic search engine. *Journal of Information & Knowledge Management*, *3*(01), 107-117.
6. Foote, K. D. (2016, May 9). A Brief History of Semantics. Retrieved from https://www.dataversity.net/brief-history-semantics/
7. Malve, A., & Chawan, P. (2015). A Comparative Study of Keyword and Semantic based Search Engine. *International Journal of Innovative Research in Science, Engineering and Technology*, *4*(11), 11156-11161.
8. Arroyo, S., Ding, Y., Lara, R., Stollberg, M., & Fensel, D. (2004, March). Semantic web languages. strengths and weakness. In *International Conference in Applied computing (IADIS04)* (pp. 23-26).
9. Sudeepthi, G., Anuradha, G., & Babu, M. S. P. (2012). A survey on semantic web search engine. *International Journal of Computer Science Issues (IJCSI)*, *9*(2), 241.
10. Kngine. (n.d.). Retrieved from https://www.crunchbase.com/organization/kngine
11. Khan, J. A., Sangroha, D., Ahmad, M., & Rahman, M. T. (2014, November). A performance evaluation of semantic based search engines and keyword based search engines. In *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)* (pp. 168-173). IEEE.
12. Kosmix. (n.d.). Retrieved from https://www.crunchbase.com/organization/kosmix#section-overview
13. Rashid, J., & Nisar, M. W. (2016). A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines. *International Journal of Information Technology and Computer Science (IJITCS)*, *8*(10), 82-89.
14. Powerset. (n.d.). Retrieved from https://www.crunchbase.com/organization/powerset
15. Sudeepthi, G., Anuradha, G., & Babu, M. S. P. (2012). A survey on semantic web search engine. *International Journal of Computer Science Issues (IJCSI)*, *9*(2), 241.
16. Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., & Warke, Y. (2002). Managing semantic content for the Web. *IEEE Internet computing*, *6*(4), 80-87.
17. Kassim, J. M., & Rahmany, M. (2009, August). Introduction to semantic search engine. In *2009 International Conference on Electrical Engineering and Informatics* (Vol. 2, pp. 380-386). IEEE.
18. Renteria-Agualimpia, W., López-Pellicer, F. J., Muro-Medrano, P. R., Nogueras-Iso, J., & Zarazaga-Soria, F. J. (2010). Exploring the advances in semantic search engines. In *Distributed Computing and Artificial Intelligence* (pp. 613-620). Springer, Berlin, Heidelberg.
19. Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., ... & Sachs, J. (2004, November). Swoogle: a search and metadata engine for the semantic web. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management* (pp. 652-659). ACM.
20. Rashid, J., & Nisar, M. W. (2016). A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines. *International Journal of Information Technology and Computer Science (IJITCS)*, *8*(10), 82-89.
21. FactBites : The Results in Meaningful Sentences. (2007, August 27). Retrieved from https://www.searchenginejournal.com/factbites-the-results-in-meaningful-sentences/5553/
22. Renteria-Agualimpia, W., López-Pellicer, F. J., Muro-Medrano, P. R., Nogueras-Iso, J., & Zarazaga-Soria, F. J. (2010). Exploring the advances in semantic search engines. In *Distributed Computing and Artificial Intelligence* (Vol. 79, pp. 613-620). Springer, Berlin, Heidelberg.
23. Jeannie Hill. (2016, April 16). 6 Benefits That Semantic Search Brings to Digital Marketing. Retrieved from https://www.hillwebcreations.com/benefits-semantic-search-brings-digital-marketing/
24. Everything You Need To Know About Semantic Search And What It Means for Your Website. (2017, July 15). Retrieved from https://www.crazyegg.com/blog/everything-about-semantic-search/
25. Advantages of Semantic Search | Omneek. (n.d.). Retrieved from https://www.omneek.com/en/resources/advantages-of-semantic-search/
26. Semantic Search: What Is It And How Can You Benefit From It? (2018, January 24).

Retrieved from https://www.imarkinfotech.com/semantic-search-can-benefit/
27. Taibi, D., Gentile, M., & Seta, L. (2005). A semantic search engine for learning resources. In *Third International Conference on Multimedia and Information & Communication Technologies in Education*.
28. Madhu, G., Govardhan, D. A., & Rajinikanth, D. T. (2011). Intelligent semantic web search engines: a brief survey. *arXiv preprint arXiv:1102.0831*.
29. Ding, D., Yang, J., Li, Q., Wang, L., & Wenyin, L. (2004, May). Towards a flash search engine based on expressive semantics. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters* (pp. 472-473). ACM.
30. Lee, C. H., & Liu, A. (2005, July). Toward intention aware semantic web service systems. In *2005 IEEE International Conference on Services Computing (SCC'05) Vol-1* (Vol. 1, pp. 69-76). IEEE.
31. Jenice Aroma, R., & Kurian, M. (2012). A Survey on need for Semantic web. *International Journal of Scientific Research and Publications*.

How to cite this article: Roy S, Modak A, Barik D et.al. An overview of semantic search engines. International Journal of Research and Review. 2019; 6(10):73-85.

******