ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

# EVOLUTION OF SEARCH ENGINES AND IMPROVED TECHNIQUES OF SEARCH OPTIMIZATION

T.Sarada Kiranmayee, Ritwika Mallick[y], Anhad Saran[z] and Nilesh Kumar[x]
Assistant Professor
Department of Computer Science and Engineering, SRM University
Chennai,Tamil Nadu,India
Email: tskiranmayee@gmail.com, [y]ritwikamalick@gmail.com, [z]anhadsaran@yahoo.com, [x]niel123.kumar@gmail.com

**Abstract—The amount of information accumulated in the internet is massive. The information is searched in the internet using a search engine. The current search engines searches for the needed information based on the keywords which the users have typed. These search and retrieval are based on syntactic analysis of keyword instead of contextual analysis. The Semantic Web is an extension of the current web in which Information retrieval is based on the contextual analysis of user search query which is a more meaningful search. In this paper we discuss the evolution of search engines and present a review on various approaches and techniques used in different search engines, their limitations and future of search engines.**

**Keywords—Web Search, Search Engine Optimization, Keyword Search, Semantic Analysis, Contextual Search.**

## I. INTRODUCTION

A search engine is a program that searches the web for sites based on your keyword search terms. The search engine takes the keyword given by you and returns (SERP) i.e. search engine results pages, with a list of sites it deems relevant or connected to your searched keyword. The goal for many sites is to appear in the first SERP for the most popular keywords related to their business. A site's keyword ranking is very important because the higher a site ranks in the SERP, the more people will see it. Search Engine Optimization or more precisely we can say SEO, is the method used to increase the likelihood of obtaining a first page ranking through techniques such as building links, SEO title tags, keyword search, meta description, and content optimization. The very first tool used for searching on the Internet was called "Archie". (The name stands for "archives" without the "v", not the kid from the comics). It was created by Alan Emtage, a student at McGill University Montreal, in 1990. The program downloaded the directory listings of all the files located on public anonymous FTP (File Transfer Protocol) sites. In 1993, MIT student Matthew Gray created what is considered the first robot, called World Wide Web Wanderer. According to The Web Robots FAQ, "A robot is a program that automatically traverses the Web's hypertext structure by retrieving a document, and recursively retrieving all documents that are referenced. Web robots are sometimes referred to as web wanderers, web crawlers, or spiders. A robot simply visits sites by requesting documents from them. "In 1993, Martijn Koster created Archie-Like Indexing of the Web or more precisely ALIWEB. What it does that it allows the users to submit their own pages to be indexed. Eventually, as it seemed that the Web might be profitable, investors started to get involved and search engines became big business. Yahoo was created by Jerry Yang and David Filo in 1994. It started out as a listing of their favourite web sites. What made it different was that each entry, in addition to the URL, also had a description of the page. Within a year the two received funding and Yahoo, the corporation, was created. Later in 1994, WebCrawler was introduced. It was the first full-text search engine on the Internet; the entire text of each page was indexed for the first time. Lycos introduced relevance retrieval, prefix matching, and word proximity in 1994. It was a large search engine, indexing over 60 million documents in 1996. Larry Page and Sergey Brin, as a part of research project, invented Google at Stanford University. It uses inbound links to rank sites. In 1998 MSN Search and the Open Directory were also started. The Open Directory, according to its Web site, "is the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a vast, global community of volunteer editors." It seeks to become the "definitive catalog of the Web." The entire directory is maintained by human input.

### A. HOW GOOGLE SEARCH WORKS

The three key processes in delivering search results to you are: Crawling: Crawling is the process by which Googlebot discovers new and updated pages to be added to the Google index.We use a huge set of computers to fetch (or "crawl") billions of pages on the web. The program that does the fetching is called Googlebot (also known as a robot, bot, or spider). Googlebot uses an algorithmic process: computer programs determine which sites to crawl, how often, and how

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
ernational Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
l. 4, Special Issue 19, April 2017

many pages to fetch from each site. Google's crawl process begins with a list of web page URLs, generated from previous crawl processes, and augmented with Sitemap data provided by webmasters. As Googlebot visits each of these websites it detects links on each page and adds them to its list of pages to crawl. Indexing: Googlebot processes each of the pages it crawls in order to compile a massive index of all the words it sees and their location on each page. In addition, we process information included in key content tags and attributes, such as Title tags and ALT attributes. Googlebot can process many, but not all, content types. For example, we cannot process the content of some rich media files or dynamic pages. Serving results: When a user enters a query, our machines search the index for matching pages and return the results we believe are the most relevant to the user. Relevancy is determined by over 200 factors, one of which is the PageRank for a given page .In simple terms, each link to a page on your site from another site adds to your site's PageRank. Not all links are equal: Google works hard to improve the user experience by identifying spam linksGoogle's Did you mean and Google Autocomplete features are designed to help users save time by displaying related terms, common misspellings, and popular queries. Like our google.com search results, the keywords used by these features are automatically generated by our web crawlers and search algorithms.

## II. ONGOING PROJECTS OF SEMANTIC SEARCH

In keyword based search, that search engines use keywords to match results to queries.The more exactly the query phrase matches a phrase repeatedly used on your web page the more likely your web page is to be delivered to a searcher as a relevant result. That search engines use keywords to match results to queries. Semantic web is being developed to overcome the following main limitations of the current Web. 1. The web content lacks a proper structure regarding the representation of information. 2. Ambiguity of information resulting poor interconnection of information. 3. Unable to deal with enormous number of users and content ensuring trust at all levels. 4. Incapability of machines to understand the provided information due to lack of a universal format. 5. Automatic information transfer is lacking.

### A. Google Knowledge Graph:-

The Knowledge Graphis a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources. Knowledge Graph display was added to Google's search engine in 2012, starting in the United States, having been announced on May 16, 2012.[1] It provides structured and detailed information about the topic in addition to a list of links to other sites. The goal is that users would be able to use this information to resolve their query without having to navigate to other sites and assemble the information themselves.According to some news websites, the implementation of Google's Knowledge Graph has played a role in the page view decline of various language versions of Wikipedia. As of the end of 2016, knowledge graph holds over 70 billion facts.

### B. Implementation of Semantic Web Mining on E-Learning:-

In recent ten years online web portals for e-learning, e-government and e-commerce became a very popular part of Web. For a better quality of service, web portals were in a need of a semantic structure and an intelligent logic. Nowadays Web Mining and Semantic Web are popular topics in Web. These two disciplines can be used together since they both are addressing each other to fulfill the aim of understanding or examining large amount of data in an automated way and also discovering and obtaining meaningful results. By applying Semantic Web Mining on educational purposes, especially on distance learning and course management systems where both can be used as a support to traditional education and distance learning intentions. The aim of discovering students learning model and personalization of services over current e-learning portals and course management systems are achievable viasemantic tools such as Web Services or Semantic Web Agents. Web Usage Mining tries to find out what users are looking for while they are using Web, and WUM also helps to find the patterns for a particular group of people belonging to a region or depending on their interest. Web Content Mining is a kind of text mining application on Web content. This method could be used for creating metadata for Learning Objects (LO) forbuilding ontology and semantic structure. The word semantic web is a product of Web2.0 (second generation web) which makes the web itself to understand andsatisfy the user requests and web agents or machines to use the content of web.

### C. MICROSOFTS SATORI KNOWLEDGE BASE

Microsofts Satori (named after a Zen Buddhist term for enlightenment) is a graph-based repository that comes out of Microsoft Researchs Trinity graph database and computing platform. It uses the Resource Description Framework and the SPARQL query language, and it was designed to handle billions of RDF triples (or entities). For a sense of scale, the 2010 US Census in RDF form has about one billion triples. Satori needs to be able to handle queries from Bings Microsofts Satori (named after a Zen Buddhist term for enlightenment) is a graph-based repository that comes out of Microsoft Researchs Trinity graph database and computing platform. It uses the Resource Description Framework and the SPARQL query language, and it was designed to handle

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
ernational Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
l. 4, Special Issue 19, April 2017

billions of RDF triples (or entities). For a sense of scale, the 2010 US Census in RDF form has about one billion triples. Microsofts Satori (named after a Zen Buddhist term for enlightenment) is a graph-based repository that comes out of Microsoft Researchs Trinity graph database and computing platform. It uses the Resource Description Framework and the SPARQL query language, and it was designed to handle billions of RDF triples (or entities). For a sense of scale, the 2010 US Census in RDF form has about one billion triples

D. Clerezza: An Apache Project for the Semantic Web

A new project entered the Apache incubator that can remove many of the difficulties associated with developing for the Semantic Web. Clerezza is an OSGi-based applica-tion that helps developers to build applications that integrate perfectly with the Semantic Web. The project also provides a set of bundles in machine-understandable RDF/JSON formats for building RESTful Semantic Web applications and web services. It uses technologies like jQuery, Jena, Apache Felix, Jetty, and Jersey. Clerezza can also be used as a platform, providing compile and runtime requirements for building se-mantic applications. DZone interviewed Reto Bachmann-Gmr, the ”father” of Clerezza who worked on the technology before it was donated by Trialox . The project has two aspects: web application development and RDF storage and manipulation.

Most existing web development frameworks tend to hide away core concepts of the Web (Rest, Uri, Representations) to emulate a desktop applications environment, due this the application built with them do not benefit from core features of the web: scalability, device independence (and thus acces-sibility), collaboration (with persistent ”deep” Uris). Clerezza by contrast is designed from the ground up to leverage the power of the web-stack, and it does not map these to desktop paradigms or to traditional RDMS database models, as the RDF (linked data) model is much more flexible and seamlessly extends web concepts. [6] discussed about Reconstruction of Objects with VSN. By this object reconstruction with feature distribution scheme, efficient processing has to be done on the images received from nodes to reconstruct the image and respond to user query. Object matching methods form the foundation of many state- of-the-art algorithms. Therefore, this feature distribution scheme can be directly applied to several state-of- the-art matching methods with little or no adaptation.

E. Lexxe Search Engine:-

The word semantic web is a product of Web2.0 (second generation web) which makes the web itself to understand and satisfy the user requests and web agents or machines to use the content of web.

F. SenseBot

The technology powering this engine creates a summary of the top results that are returned for a user query, often negating the need to drill down into the URLs to get the information that one is seeking. Semantic Engines LLC, the company behind the engine provides a variety of products around this technology. There is Link Sensor, a tool that can be used on major blogging platforms (WordPress, Blogger, etc.) for automatically picking up key concepts from the post and linking them to related articles from the same blog or publisher. It is possible to point to other venues as well, e.g. to another blog from the same publisher perhaps with a higher CPM. The tool increases user engagement. The company has also started providing APIs for returning summaries of results for a query from a set of URLs that are also passed in as parameters to the APIs. This is one interesting approach that helps save time when an exact answer is what one is looking for.

## III. ONGOING PROJECT ON CONTEXT BASED SEARCH

Contextual understanding is about giving people the infor-mation they want when they need it the most. Remove the hard work for them and deliver what they will likely want at that exact moment.As an example: I am a frequent user of Uber and GetTaxi, but what if I land in a city where neither of those apps work? I can either call a taxi or find a comparable app, though its unlikely that I am going to take out my phone right then and start searching the app store.

A. Apple Siri Setting Change And More

Rather than tap and swipe a few times to get where you want to, just ask. You can say ”make the screen brighter,” ”cancel all alarms” and ”Do not disturb” when you want to silence everything. If you can’t find an app icon, just say ”open calculator” or ”Launch Kindle.” If your hands are not free, just say ”read my new messages.” Finally, you can also use Google as a verb: say Google the Museum of Natural History in London.

B. Apple Siri Sort through photo mess

Siri has access to your iDevice’s photo album, so you can ask it to search for specific photos with a condition of time, place or album.For instance, you can ask ”Show me photos taken on November 10 this year” or ”show me photos of my trip to Agra.” If you’re an OS X user with ’faces’ set up in iPhoto (and a synced iOS device), you can also ask Siri to search for photos of a particular person.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
ernational Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
l. 4, Special Issue 19, April 2017

C. Apple Siri: Contextual awareness

This ability came in with iOS 9 – let's say you were searching the web for a restaurant. You can ask Siri "When will it open?" and Siri will realise that 'it' means the restaurant you're looking for. Or if you get an email with some action required, ask Siri to "remind me about this." Siri will figure out what app you were in when you asked and what you were looking at when it reminds you later, it'll open everything back to that same content.

D. Microsoft Cortana: Location Based Reminder

Quite like Google, Cortana can set location-based reminders for Windows 10 devices. You just have to speak out the reminder and the name of the location. The next time you visit that place, Cortana will show you the reminder once it recognizes your location.

E. Microsoft Cortana: Person Recognition

Cortana comes with built-in voice recognition. This helps it create a personal voice recognition of your particular voice. This way only voice commands said by you are acted upon faster and other commands are ignored.

F. Google Now: Follow up questions

If you have asked a specific question to Google, you can also ask follow up questions. For instance, if you ask for "Movie review of Inception," you can follow it up with "Who is the lead actor in it?" -Google will automatically understand that you are talking about the same movie and answer accordingly. In the same way, if you ask "What's the weather like in Mumbai?" you can follow that up with "How about Delhi?"

G. Google Now: Location-based reminders

Since 'Ok Google' works in conjunction with Google Now, you can use the service to quickly set location-based reminders. Just include the location or store name in your reminder. Like "Remind me to buy slippers when I get to Select Citywalk." When you visit the mall in question, your reminder will pop-up.

H. Google Now: Smart calling

Almost every smartphone user knows about calling with help of voice assistant. However, if you have added a relationship status of that contact (mom, dad, sister, brother) you can just say "Call Dad" and the voice assistant will automatically find and call the relevant contact. The same trick also works for contacts with nicknames

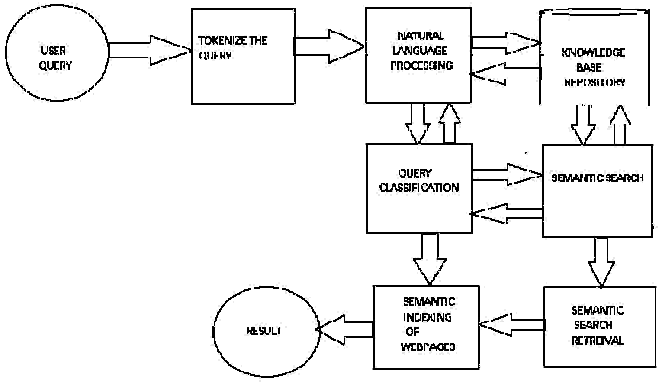

Fig. 1. Basic Architecture

## IV. PROPOSED ARCHITECTURE FOR SEMANTIC SEARCH

Get the user query: Get the user query in the form of keywords (i.e. generally a string)Tokenizer the query: Syntactic analysis of the Query takes place in a tokenizer. All the keywords are separately evaluated.Natural Language Process-ing: It includes processing of the keywords in user query in intelligent systems . It basically includes four tasks-Sentence segmentation, part-of-speech tagging, parsing, analysis. NLP aims to reduce the gap between user intended query and the machine specification.Knowledge Base and repository: It is a part of decision making systems that contains some prede-fined set of rules and other rules imposed due to intelligent learning processes. These rules gathered can be used for future reference.Query Classification: Classification, clustering and aggregation of related searches are an important part of Search optimization and also necessary to relate the semantics (i.e.meaning).Semantic Search: Relating the keywords with their meanings instead of the syntax. A word with a related meaning with another word are clustered together and backed up the knowledge base.Search result retrieval and ranking: List of related search results are obtained after processing. These search results are now filtered and ordered according to their frequency of occurrence and on demand-basis, which is known as search ranking.

## V. CONCLUSION

The future of search engines is about understanding user behavior online and offline.Perhaps the future of optimization has nothing to do with the internet at all, but everything to do with optimizing the users experience. In this paper, we made a brief survey of the existing search engines. We reviewed the ongoing research on context-based and semantic search engines. In addition, the limitations within the reviewed semantic search methods. In the future, our work will focus on the deeper and broader research in the field of intelligent semantic search, with the purpose of introducing personalised

recommendation systems based on semantic and contextual analysis.

REFERENCES

[1] Seth Earley,Cognitive Computing,Analytics,and Personalization,IEEE Computer Society,2015

[2] Amit Seth,Internet Of Things To Smart IOT Through Semantic,Cognitive And perceptual Computing,IEEE Intelligent System,2016.

[3] http://www.wordstream.com/articles/internet-search-engines-history

[4] https://en.wikipedia.org/wiki/KnowledgeGraph

[5] http://www.webreference.com/authoring/searchhistory/index.html

[6] Christo Ananth, M.Priscilla, B.Nandhini, S.Manju, S.Shafiqa Shalaysha, "Reconstruction of Objects with VSN", International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST), Vol. 1, Issue 1, April 2015, pp:17-20

[7] G.Madhu,Dr.A.Govardhan and Dr.T.V.Rajinikanth,Intelligent Semantic Web Search Engines: A Brief Survey,2013

[8] Jyoti chaurasia and Prof. Jagdish Raikwa,Survey on Semantic Web Search Engine: Using Domain Ontology,2008

[9] Ouz Mustapaa, Dilek Karahocaa, Adem Karahocaa, Ahmet Ycela and Huseyin Uzunboylua,Implementation of Semantic Search onE-Learning, Department of Software Enginnering, University of Baheehir, Istanbul-Turkey, 2009

[10] http://www.lexxe.com/

[11] https://en.wikipedia.org/wiki/WebBot

[12] https://en.wikipedia.org/wiki/Lexxe

13]http://searchcontentmanagement.techtarget.com/definition/natural-language-processing-NLP.