Comparative Analysis of Web 3.0 Search Engines: A Survey Report

R.Aravindhan, Assistant Professor
Dept. of Computer Science and Engineering
Sri Eshwar College of Engineering
Coimbatore,Tamilnadu, India
contact2aravind@gmail.com

Dr.R.Shanmugalakshmi, Associate Professor
Dept. of Computer Science and Engineering / IT
Goverment College of Technology
Coimbatore,Tamilnadu, India
drshanmuhslskdhmi@yahoo.co.in

Abstract —**The amount of information accumulated in the internet through myriad number of databases is massive. The information is searched in the internet with specialized tools known as search engine. A search engine can be viewed as a simple software program that searches for the needed information based on the keywords which the users have typed in. Usually these search and retrieval are based on syntactic analysis(Web 2.0) of keyword, but if these searches and retrieval are based on content analysis, a more meaningful result will be returned to the users. The Semantic Web (Web 3.0) is an extension of the current web in which keyword is given a well-defined meaning. Apart from that another advantage of Semantic Web is its user interface. A large number of Semantic Web search engines have emerged recently which are based on different design principles and provides different levels of support for users and/or applications. However, in spite of the benefits provided by these widely available intuition based search engine, users are reluctant in shifting towards this these sophisticated search engines. In this paper, a survey is done about the semantic search engines to reveal the promising features of the semantic search engines (SSE) and the reason behind the reluctance of the users in adopting these sophisticated search engines.**

Index Terms— Web 3.0, Semantic Search Engines, Semantic Web, Information Retrieval, Web Crawling.

I INTRODUCTION TO SEARCH ENGINE

A. [1] Internet search engines are unique and important and the web users are supported by particular type of web sites to search the important facts which has been spread a little over many number of websites. The great plans and design relating to each and every search engines may very yet they all do three main works: i) They search through the internet and obtain information based on keywords. ii) By retaining the index of keywords they locate their origin. iii)They permit uses to search for words or phrases in that index.

The workflow of Search Engine is depicted in Fig.1.they consists of the following steps a. Creating an index b. Receiving a Query (set of search terms and commands) c. Looking in the index file for any matches d. Gathering the matching page entries and ranking them according to the relevance e. Formatting the results f. Returning the result page in HTML format to the user's web browser. Search engine may also be noticed as a organized set of programs that has

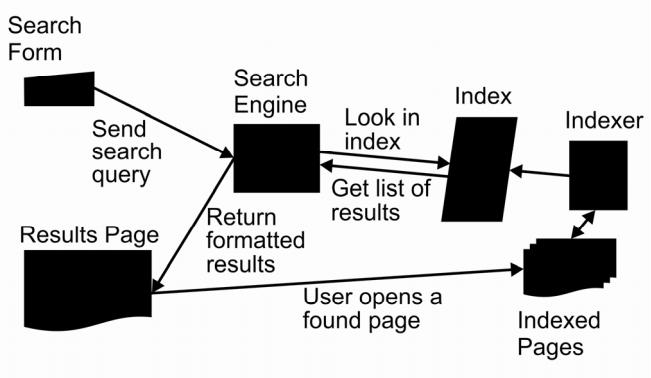

Figure 1. Workflow of Search Engine

(a)A spider which is known as web crawler Figure-1 or web robot goes to every page on website in a thorough by its own in an automated way.

(b)The pages read by the spider is being created as an index known as catalog by a program.

(c)A search request received by a program compares the word in catalog and gives back the results to user's search engine.

[2]The web content usually lacks a proper structure with regard to the information being represented in the web and hence the Traditional search engine has its own limitations and they are 1. poor interconnection of the intended search information and the retrieved information. 2. Unable to ensure trust at all levels as it deals with enormous number of web users and web content. 3. Lack of capability to understand the provided information. 4. lack of Automatic information transfer.

II INTROUDUCTION WEB 3.0

Web 1.0 was mainly intended for searching, whereas Web 2.0 was mainly used for interactive based social activities and Web 3.0 is based on service. [3] Web 3.0 (Semantic Web) has been evolving with an intention of connecting the information searched and the information retrieved. It enables data to be linked from a source to any other source and to be understood by computers so that they can perform increasingly sophisticated tasks on our behalf.

[978-1-4799-2352-6/14/$31.00 ©2014 IEEE]

Web 3.0 is interpreted as a "Web of Things." Each thing has: (a) A name and address on the web. (b) A human-readable page that has been created and (c) machine-readable data that may describe about the thing, what it wants and what it offers. [4]Web 3.0 is also called the "Service Web." A web service is programmed to read machine-readable data about things and uses data-linking to relate a machine's needs to relevant things.

A Comparison of web 2.0 and web 3.0

TABLE 1

	Web 2.0	**Web 3.0**
Permission	Read / Write	Portable Web
User	Group/ Communities	Individuals
Data	Sharing the data	Consolidated Dynamic content
Technology	AJEX	RDF
Example	Google, Wikipedia	Dbpedia, igoogle

An important goal of the semantic web is to make the actual meaning of the intended search explicit ,thereby enabling effective access to information contained in a heterogeneous information environment. Semantic search plays an important role in realizing this goal, as it takes effort and ensures to return relevant answers to user's queries by taking advantage of the availability of explicit semantics of information.

The contemporary Web search engine will not be able to exactly interpret your intended search. It merely searches and retrieves for Web pages which contains the keywords that has been typed by the user. The main drawback of the search engine is that can't ensure, if the Web pages returned is actually what the user is expecting. For example, if an user enters the keyword "price of apple" with an intention of knowing the price of apple fruit, it word instead return results that contains the price of apple phone.

Keywords and contest of your search can be found by a web3.0browser it might return related results and give other contest relevant to you keywords for instance if "tropical holiday destinations under $3,000" is the keyword typed in the web3.0 search engine a number of fun activities or huge restaurants related to the keywords. For every query submitted the web 3.0 search engine make a massive search of the entire web content to return a most useful and relevant information to the user .

III ANNOTATION METHODS

One of the growing problems that the researchers are encountering in SW is annotation which is a mandatory requirement for SW search engines. In order to make the current web pages adopt SW search annotating it by appropriate meta-data is a pre condition. In general generating meta-data for structured data is simpler [24].

Annotations can be classified based on following aspects:

A. Type of meta-data: According to [25] meta-data can be divided into two types i.e Structural and Semantic. In structural meta-data, non contextual information is expressed (e.g. language and format) whereas in the semantic meta-data, contextual information is expressed and it is usually stored as RDF triples.

B. Generation approach: generation approach can be either simple that is to generate meta-data without considering the overall theme of the page and by only using structural information of a page or it can use an ontology in the generation process. The main advantage of this method is that it is based on the contextual information.

C. Source of generation: Usually the source of meta-data generation is a page itself but sometimes it will be beneficial to use other complementary sources. For example [1] and [5] uses information from available networks for accumulating more information for a page. For example if a movie is searched it is possible to extract additional information like director, genre, etc.

IV. HARDLY USED WEB 3.0 SEARCH ENGINES

Hakia

Hakia's: A search including the meanings of words and sentences are made up over three developing methods of scientific knowledge[5].

I. OntoSem(sense repository)

II. QDEX(Query indexing method)

III. Semantic Rank algorithm.

OntoSem is a database of language related ideas in which every words are drawn or traced in to different "sense" they say.

QDEX expands as query detection and extracting inverted index is used to store the web content by more over all the engines. Thus hakia's is replaced by QDEX. It obtains all possible problems relevant to the content of the page at last, semantic rank algorithm used to rank contents on its own based on sentence analysis.

Figure 2. hakia Search Engine

The following are the features of the hakia search engine[6]

- It concentrates on good quality and accurate results
- The information related to date can be obtained
- Without the permission of uses no data is saved in their system.
- Hakia delights the customer by reaching the targeted information in a minimum time.
- Hakia tries to give more accurate and relevant results on a searcher's first attempt at a query
- Based on the query, Hakia returns categorised results.
- Hakia understands polysemy and synonymy and provides retrieves relevant result.

Factbites:

[7]Factbites is another Web search engine that retrieves meaningful results relevant to the topic of the search query. Instead of listing the websites by matching the keyword of the search query, a more precise and accurate results relevant to the meaning of the keyword is retrieved. Factbites searches for matches on the based on the topic rather than your keyword. The Factbites tries to provide the user more genuine and meaningful content.

[8]Factbites makes the web search efiicient by making a clean search i.e it gives the user a meaningful, relevant responses. It does not do just a mere keyword matching, it ensures that the page returned has exactly the information you are looking for rather than focusing on the number of keywords that has been matched. Since it analyses the meaning of the page that has been retrieved it filters the spam sites before retiring the list of websites

Figure 3. facbites Search Engine

The following are the Features of Factbites

- It provides additional factual data.
- It has the ability to Filters out spam websites in the the search results.
- It searches based on the topic rather than the keywords.

KNGINE

[9]Knowledge Engine abbreviated as kNGINE, is a web 3.0 semantic search engine which focuses on returning what the user actually expected. It analyzing the entered keywords to form a relationships between them and then the meaning is interpreted based on the relationship and then the results are returned. This interpretation is made by a hybrid implementation of Knowledge-Based approach and statistical approach.

It has two main components

A. The Extraction Engine

Kngine's Extraction Engine collect and organizes widely available unstructured data by crawling the web. By the use of natural language processing, machine language and data mining algorithms the some important and useful abstract principle, information and relations among each keyword is learnt. Knowledge graph is made up based on this engine.

B. The Question Answering Engine

The Question Answering Engine understand the question and move across the knowledge graph to solve the answer with the help of natural language processing and machine learning. Other than this it also searches clues unproved ideas and obtainable answers in unstructured data all these searches are done statistically

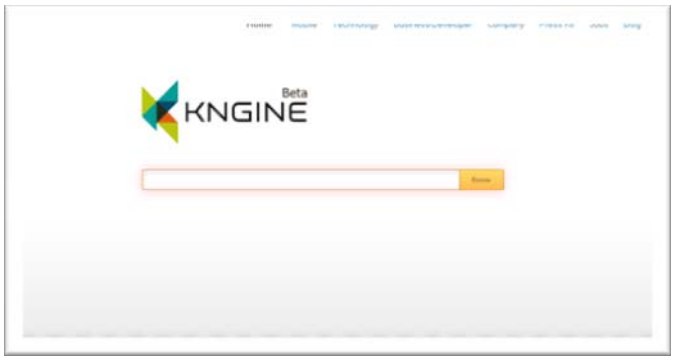

Figure 4. KENGINE Search Engine

Features if Kngine

- It can also provide search results in the form of images and pictures
- Kngine has the ability to allow the users to search in parallel manner.
- It is multi-lingual

SENSEBOT

[12]SenseBot, a search engine that has a unique feature of summarizing the top results for a query . It uses Natural Language processing combined with Multi Document summarization techniques to present a summary which appears as the main link for the result.

[13] Usually a user has to visit various links returned by the normal search engine to acquire the required information, but in

the case of SenseBot , a summary of what the user is actually searching is presented at the top of the search results.

Figure 5 SnesBot Search Engine

SWOOGLE

[14]Swoogle is a specialized content based search engine which analyses, discovers, and indexes knowledge in the web. Swoogle uses web crawlers to discover RDF documents and then extracts metadata related to each discovered document Then the relationship between these documents is computed. Swoogle is also referred as a Prototype Semantic Web search engine that has three main activities.

A. Finding appropriate ontologies

Finding appropriate ontologies is a key factor in this algorithm because, if appropriate ontology was not returned then it leads to the creation of new ontology. then a ontology rank algorithm is used to rank the returned ontologies.

B. Finding instance data

Swoogle enables querying Semantic Web Data based on the classes and properties defined by them and as a result unstructured data is collected appropriately.

C. Characterizing the Semantic Web:

The semantic web is characterized by collecting metadata especially inter-document relations.

Figure 6. Swoogle Search Engine

The main features of the Swoogle are as follows[15]

- It efficiently queries a comprehensive database of SWD metadata
- It enables "distributed" knowledge sharing
- It makes knowledge visible and easy to access
- It Finds appropriate ontologies
- It Finds instance data
- It Characterizes the Semantic Web

LEXXE[16]

Lexxe is a Natural Language Processing based semantic search engine. It is best alternate to traditional Specific information search an General information search

Figure 7. Lexxe Search Engine

Lexxe uses a database based on content and knowledge as well as linguistic algorithms to predict the real topics associated with each web page. It also predicts the informativety (i.e the degree of how much the information is relevant to a particular topic) It will only return the documents that are most relevant in content to the specific query. popularity also plays a small part in ranking the relevance systems. It give the user a pleasant experience by providing a serious-cutting edge research.

V COMPARISON OF VARIOUS WEB 3.0 SEARCH ENGINE METHODOLOGIES

[10]There are various factors that characterize these search engines and they are as follows:

A. Architecture

There are two different architectures and they are

Stand-alone search engine.

The crawler browses each document and stores the meta data associated with each document in an index .

Meta search engine.
It is not a stand-alone search engine and it does not form an index. It distributes queries to other subordinate search-engines and then combines the results.

B. Coupling
There are two approaches:

Tight coupling
In tightly coupled approaches the meta data associated with each page refer explicitly to specific ontology. Sometimes, each documents associated with a page are considered as individuals in an ontology. However, it incurs a potentially high semantic document annotation cost. These are suitable not only to special purpose information systems but also to more general application areas.

Loose coupling
In the case of loose coupling most of the documents are not associated to any specific ontology. Therefore In the loosely coupled case, it is very difficult to choose an appropriate ontology for a given domain. However, this is feasible if only a small fraction of documents is semantically annotated Loosely coupled systems can be implemented as meta search engines easily.

C. User Context
User context is a factor that directly contributes to the usefulness of the documents which has been returned to the user. [11]In order to provide a better user context the following strategies are implemented.

Learning
User context is usually extracted from the user interaction dynamically. Based on the user's query and query-refinement history the system makes a prediction about desired results.

Hard-coded
In the hard-coded approach, queries are categorized in a question-categories that which specify what the user actually needs. The system provides a fixed number of question-categories that are exploited during query evaluation. Typical categories can define the kind of information need such as "location of …", or "general resources for …" or the context of the information need such as 'Jazz'. In the latter case, if the user searches for 'Miles' the system will return no documents about the Miles College or Air Miles but documents on Miles Davis, only.

Query modification
query needs to be modified in order to return the relevant results and it can be done in the following ways

Manually:The easiest and simplest way to make query modification is to allow the user to modify the query manually. When the user enters a query, the system returns not only documents, but also an appropriate part of an ontology. The user navigates the ontology and reformulates his query, i.e., adds or removes query terms.

Query rewriting: Query rewriting is driven by the idea that a query can be optimized by the system. We observe three different ways, augmentation, trimming and term substitution.

Graph-based: The third technique to optimize user queries requires tight coupling between the document base and the ontology. It perceives both, ontological concepts and documents as the nodes of a graph. Query terms are used to find relevant nodes in the graph.

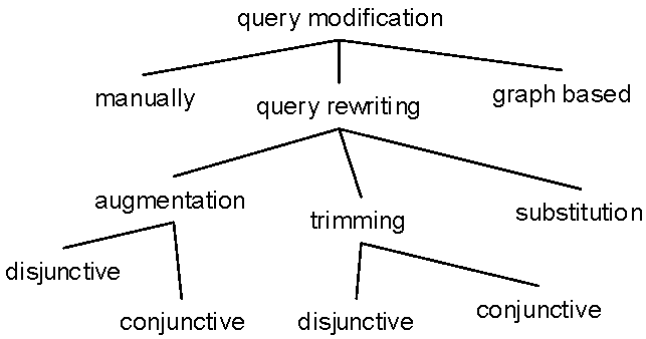

Figure 8. Structure of Query Modification

D. Ontology structure
Ontology-based semantic search engines rely on certain ontology structures. Ontologies are usually built from concepts, properties, constraints and possibly axioms. We observe that semantic search exploits properties only and distinguish the following cases:

- Anonymous properties: In the case of anonymous properties, the system disregards the name and the semantics of the property. The interrelation between two concepts indicates that they share the same context, only.
- Standard properties: We only found a small set of 'common sense' properties in our survey. The propert ies are synonym_of, hypernym_of, meronym_of, instance_of and negation_of. The homonym_of property does not have to be modelled explicitly since it is equivalent to term equality. The usage of standard properties enhances semantic search capabilities. However, it also introduces dependencies on ontological structures. For an overview of how standard properties can be exploited for semantic search see, e.g., Bates (1990).
- Domain specific properties: Besides standard properties, a system can exploit domain specific properties, as e.g., 'camera type' in a photograph retrieval system.

TABLE II.

Search Engine / Parameters	Hakia	Factbites	Kngine	Sense bot	Swoogle	Lexxe
Architecture	Meta	Meta	Meta	Meta	Meta	Meta
Coupling	Tight coupling	Tight coupling	Tight coupling	Loose coupling	Loose coupling	Loose coupling
User Context	Hard-coded	Learning	Hard-coded	Learning	Learning	Hard-coded
Query modification	Manually	Manually	Query rewriting	Manually	Query rewriting	Manually
Ontology structure	Anonymous	Standard	Standard	Standard	Standard	Anonymous

VI CONCLUSION

This paper gives a brief overview of some of the best semantic search engines that uses various approaches in different ways to yield unique search experience for users. We summarized our discussions this in Table II. It is concluded that searching the internet today is a challenge and it is estimated that nearly half of the complex questions go unanswered Semantic search has the power to enhance the traditional web search. Whether a search engine can meet all these criteria continues to remain a question. Future enhancements include developing an efficient semantic web search engine technology that should meet the challenges efficiently and compatibility with global standards of web technology.

ACKNOWLEDGMENT

The authors like to thank Dr.S.Annadurai and Dr.S. Gunasekeran.

REFERENCES

[1] Frank McCown, ”Introduction to Web Search Engines” http://www.harding.edu/fmccown/search_engines/intro.html.

[2] "Introduction to Search Engines" http://www.yutrend.com/topics/basics/index.php?clanak=seaeng&rubrika=beginners.

[3] Kendra Gaines "A Brief Introduction to Web 3.0" http://www.1stwebdesigner.com/design/web-3-introduction.

[4] Introduction to the Semantic Web "http://www.cambridgesemantics.com/semantic-university/introduction-to-the-semantic-web"

[5] Hakia’s Semantic Search "The Answer to Poor Keyword Based Relevancy" http://www.searchenginejournal.com/hakias-semantic-search-the-answer-to-poor-keyword-based-relevancy/5246.

[6] Jennifer Zaino "Hakia Semantic Search Engine Powers New Stock Intelligence Service" http://semanticweb.com/hakia-semantic-search-engine-powers-new-stock-intelligence-service

[7] Rapid Intelligence "Introduction to factbits" http://www.factbites.com/

[8] John Goldsmith, "Factbites: New Type of Search Engine" http://www.detools.ca/?p=14

[9] Why,What,How K Engine "We're Kngineers" http://www.kngine.com/company.html.

[10] G.Madhu, Dr.A.Govardhan, Dr.T.V.Rajinikanth "Intelligent Semantic Web Search Engines: A Brief Survey" International journal of Web & Semantic Technology (IJWesT) Vol.2, No.1, January 2011

[11] G.Sudeepthi,G.Anuradha,Prof.M.Surendra Prasad Babu, "A Survey on Semantic Web Search Engine",IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012

[12] "The Search Engine that finds sense in a heap of Web pages "http://www.sensebot.net/about.htm

[13] Arun Radhakrishnan,"Summarization, the Answer to Web Search : Interview with Dmitri Soubbotin of SenseBot" http://www.searchenginejournal.com/summarization-the-answer-to-web-search-interview-with-dmitri-soubbotin-of-sensebot/6094.

[14] Arun Radhakrishnan, "Semantic Search Swoogle : An Engine for the Semantic Web" Search Engine Journal August 12, 2007.

[15] Li Ding,Tim Finin. Anupam Joshi,Yun Peng, R.Scott Cost,Joel Sachs,Rong Pan,Pavan Reddivari,Vishal Doshi"Swoogle: A Semantic Web Search and Metadata Engine".

[16] Arun Radhakrishnan, "Semantic Search Lexxe : Search Engine that Answers Exact Queries",Search Engine Journal July 20,2007.