A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines

Junaid Rashid
COMSATS Institute of Information Technology /Department of Computer Science, Wah Cantt, 47040, Pakistan
E-mail: junaidrashid062@gmail.com

Muhammad Wasif Nisar
COMSATS Institute of Information Technology /Department of Computer Science, Wah Cantt, 47040, Pakistan
E-mail: wasifnisar@gmail.com

Abstract—Semantic search engines (SSE) are more efficient than other web engines because in this era of busy life everyone wants an exact answer to his question which only semantic engines can provide. The immense increase in the volume of data, traditional search engines has increased the number of answers to satisfy the user. This creates the problem to search for the desired answer. To solve this problem, the trend of developing semantic search engines is increasing day by day. Semantic search engines work to extract the best answer of user queries which exactly fits with it. Traditional search engines are keyword based which means that they do not know the meaning of the words which we type in our queries. Due to this reason, the semantic search engines super pass the conventional search engines because they give us meaningful and well-defined information. In this paper, we will discuss the background of Semantic searching, about semantic search engines; the technology used for the semantic search engines and some of the existing semantic search engines on various factors are compared.

Index Terms—Semantic, Semantic Searching, Semantic Search Engine, Hakia, RDF.

I. INTRODUCTION

The Semantic Web is an extension of the current Web. This makes the importance of information are described in detail in reference to the well-defined vocabularies that understood by both people and Computer. On the Semantic Web information describes in a new W3C standard called Resource Description Framework (RDF). Semantic Web Search is a search engine for the Semantic Web. Current Web pages can be used by people and computers to locate exactly and collect information published on the Semantic Web. Ontology is one of the most important concepts in the semantic web infrastructure used; RDF (S) (Resource Description Setting / Schema) and OWL (Web Ontology Languages) are two W3C recommended data Representation models. The Semantic Web is supported efficient discovery, automation, integration and reuse of data and provides support for Interoperability problem that cannot be solved with current technologies web.

In the present era, it takes too much time to search for the exact data from different links so; new generations prefer semantic search engines because efficient searching is needed to get a well-defined and quality result. Web search engines give irrelevant information sometimes, but semantic search engines help a lot to find the exact information about relevant data [1].

The main goal of a semantic search engine is, to give such results which can be easily understood by human beings [2].

The development of traditional web resulted from the development of semantic search engines [3]. There are several important components used in search engines. One of them is an ontology which is very important part of search engine [4].

The conventional search engines retrieve data and arrange it in the result list using algorithms like page rank [5]. Each semantic search engine works differently from another semantic search engine. Semantic search engines main approach is to develop machine-readable data [6]. Semantic search engines cannot be used for navigational searches, whereas in searches user tries to get no of documents which will give information about the query enter by user. The main goal of this paper is, to overview the semantic searching and compares the different semantic search engines. As the world progresses the need for information increases day by day. The demand for such need increases as the computer or IT world makes the development. So to fulfill and cope up with the scientific world “search engines” were developed. A question arises in once mind that despite a lot of information on the internet how we can collect it for our personal use? So the tool for it is “search engine”. It collects our demanded data for our queries. The depending upon the work, search, accuracy, efficiency and presentation there are so many search engines which are different from each other. The paper deals with the background of Semantic Searching and some semantic search engines. The paper also explain the some issues in
the semantic search engines, technologies used for the semantic searching, analysis of current search engines and comparison of semantic search engines. The different semantic search engines are compared on the different factors.

Different search engines have different indexes due to this they show different results. Initially, the information was provided by the search engines based on keyword technique. But in this modern era of technology, some queries are answered by semantic search engines. The start when the web was being developed and information was being distributed two important issues were arising which was:

i. How can search engines connect the query typed with the document having information related to that?

ii. How can a search engine recognize what type of query and its result which the users are asking for?

The ontology and other techniques for semantic search engines [7] [8] can solve these problems giving the efficient and required result.

In this paper Section II discusses about semantic searching and Section III describe the approaches of semantic search engines. The Section IV describes the some semantic search engines and Section V explains the some issues in semantic search engines. The Section VI describes the technologies used for semantic search engines and Section VII explains the analysis of current semantic search engine. The Section VIII explains the comparisons of semantic search engines. In the Section IX results and discussion are explain. The last section X gives the overall conclusion.

II. SEMANTIC SEARCHING

Semantic searching gives the exact meaningful and useful information for user queries. Semantic searching gives related information very fast. It does not match the word, but matches the query with data and give the reference links. It extracts relevant data from the query related to documents. Goha et al have divided the searches into the navigational. In navigation type to get the content of documents the users use the search engine as a source. In order to improve the result, the semantic searching integrates the techniques of SSE [9]. The searching is for knowledge and many documents are checked and analyzed by the user for semantic search. The semantic web is a field where it provides the user to facility of publishing reusing data and exchanging it all over the world for the sake of application and other work. It is a W3C approach which is a larger structure. In semantic searching, there is a facility that one can refine the search data to make the gathered information more accurate and reliable [10, 11]. It keeps in view the queries which asked and answer it quite exactly. To differentiate between the web search engine and semantic search engine duckduckgo is the excellent search engine to be used [12]. It shows all answers which can be related to query with short descriptions. By having a glance that user can select his desired information about the word “apple”. The result shows that the word apple can represent fruit, bark, Microsoft Corporation. All of these answers can be possible and have equal chances.

III. APPROACHES OF SEMANTIC SEARCH ENGINE

The semantic search engine is more accurate and efficient search engines as they provide the data related and nearly too exact. Google’s popularity is accepted, but all the other web search engine’s deficiencies are fulfilled by semantic search engines. They are helpful in producing machine readable data. Semantic search engines goal is to develop the program that can simplify and maintain the knowledge-based services [13]. Such purpose languages like RDF [14] OIL, DAML+OIL and OWL [15] are developed. It’s gives relevant information in short time and makes the searching easier. The searching of data in semantic search engine differs according to users for the searching at a low level that is no need of analyzing and checking it.

A. Approaches of Semantic Search Engines

Four approaches of SSE are listed below. Semantic search engine can use any of the following approaches.

- First approach: A Contextual approach to disambiguate and to make the queries single meaning [16].
- Second approach: The focus on reasoning. It can gather new reasoning from the provided facts [17].
- Third approach: Understanding of language. In this type the search engines effort to approach the goal of identifying, the aim of people of getting information. Powerset is the best example [18].
- Fourth approach: The representation of knowledge semantic search engines uses ontology. The system uses the typed query by using of ontology so that the search can be focused as much as possible [18].

Semantic search engines can mix more than one approach to fulfilling different functions. There is room for a variety of search engine which means each size does not fit all semantic search engines.

IV. SOME SEMANTIC SEARCH ENGINES

Search engines play an important role in the access to the web. Search engines helps the internet user in finding their required results easily. These search engines search different web pages related to the search. However to overcome this problem in search engines to retrieve meaningful information accurately and efficiently, semantic Web technologies are playing a major role. [19]. Some semantic search engines are as follow in figure 1.

Hakia: Hakia is also called the general purpose search
A semantic search engine which results based on the meaning of content rather than the page rank [20, 21]. Page rank is an algorithm which is mostly used by conventional search engines like Google [22, 23]. Hakia used the terms QDEX (Query indexing techniques) and CMR (Compressed meaning Representation).

DuckDuckGo: This semantic search engine has many features, which offers us numerous details to use it. When a user searches for a term it gives him a chance to select what was he looking for, with clear results. It provides us disambiguation and meaningful results. That’s why it is better than Google. First, it gives all possible results and then comes to the description in one sentence. [24].Duckduckgo used the terms NLP and XML.

Sensebot: In this semantic search engines uses text mining for identity the key semantic concepts. It performs multi-documents summarization through which it gives us an accurate search summary of the query. It gives an overview of topic related to user query and this way it saves the time in searching.

It is launched as interesting interface consists of many engines which can be questioned such as Google, Yahoo and Bing. Moreover, it offers many languages [25]. It uses text mining to identify semantic Concepts [26].

Cognition: A semantic search engine which uses “Natural Language Processing” [27].It is a linguistic search engine which means it gives us the facility to learn about different aspects of many languages [28].Cognition used the terms ontology.

Swoogle: Swoogle is crawler–based indexing and recovery framework for the semantic web records. It explores the records. It found to register helpful metadata properties and connections between them. Swoogle used the terms RDF, Topic Exploration and deep web. The deep web is an unfriendly area for web crawlers. It is evolving as a material source [29].

Kosmix: It is the first general purpose and topic exploration having a lot of information and elaborative search engine. Its search results are relevant to user query. It classified its Contents [30]. It uses semantic in an effort to link the records from all over the web.

Factbites: Like other semantic engines it gives its user meaningful and relevant results. It does not stick on some specific keywords [31].

The word factbites means “Where results make sense”. It works like a Filter for those sites which are not relevant to user query [32]. It gives summarized data in sequence. Factbites has attractive interface [33].

V. SOME ISSUE IN SEMANTIC SEARCH ENGINES

Common issues in the semantic search engines are as follow

i. Sometimes results are not correct and repeated again and again [34, 35].
ii. The intention identification is very vital in such search engines [36].
iii. In the start the user could enter into global web i.e. they offered disambiguation option [37].
iv. Some of the intelligent based semantic search engines not showing the importance of precision and recall. Google is not a semantic search engines but its precision is high and recall is low.

v. The user usually domain-specific knowledge. The users may not describe all potential but synonyms and variations in the query actually have a problem that users are not sure how Phrase.

vi. The World Wide Web got trillions distributed information on the subject based search only to extract the subject information in pages and produce a result to the user. But if we go for semantic based Search engines, it is able to make several decisions for individual user keyword. The metadata key-based search respectively for processing Web pages so the individual Crawler is not sufficient to do the job.

VI. TECHNOLOGIES USED FOR SEMANTIC SEARCH ENGINES

Phase diagram: Phase diagram for technology used for semantic search engines shown in figure 2.

RDF is the basis of the semantic search engine. It is standard language of W3C. Semantic search engine consists of information in the term of collected data or data collection in the place of document form. RDF is useful to help for this [38]. It rates the information or collected data in many sources and compares with hyperlinks [39].
**Resource:** The result and things described or explained by RDF is a source. The each resource has a URI and it can be whole web or a segment of any web [40].

**Property:** Any quality describing or explaining a resource is its property. A property itself can also become a resource. It describes or explains a resource [41].

**Statement:** The link between a property and a resource is called a statement and a value. They can also be called named as a predicate, subject and an object.

Ontology increases the information and vocabulary for the explanation of property. It is a study of the relationship between the entities/objects and their properties. It can be used to explain different types of concepts. It is mostly very useful in representing information or knowledge [42]. Its development is very expensive, difficult and it is basic of the semantic search engine. Its creation is time taking as well. The ontology developers are representing the interesting domain which is built on different classes and properties [43].

The several researchers [44-46] give the concept of ontology that it can be used to examine the knowledge in a field by modeling the ideas relevant to many applications. The several approaches are using the ontologies as the semantic basis for the source of theoretical schemas [47-52].

Metadata is knowledge or information of any other knowledge or information. They give compiled information of any other information. The user can get information about any particular information one by one or separately. It is not a wise method. Meta data’s main goal is searching but it can be used for other purposes as well [53].

XML is a famous language and popular technology. It is very efficient for the tags of the semantic web. It is machine readable and flexible. It is useful if anyone want to make its document in the search engine because it does not prohibit it [54].

URI (Uniform Resource Identity) is used to encode identify and make any resource uniform. It contains a lot of data and is used for distributing data. Anyone can be making URI for its own resource [55].

**VII. ANALYSIS OF CURRENT SEMANTIC SEARCH ENGINE**

The method applied to semantic search engines was based on four steps including reviewing the features, projects and publications secondly parameters of each search engines, thirdly contacting with authors for completing the information and the fourth was analyzing with previous work [56-58]. Then all the facts are assembled and published in a group portal IAAA [59].The data is linked for getting feedback of users. Researchers and other web developers are aware of their duty to improve and upgrade search engines in several steps. Many of these are relating to man or human understanding. It is important to study what operation search engines can provide.

The current information retrieval all techniques are not so good that searches the semantic knowledge in the document and gives the specific results [60].The semantic web [61-63] solve the problem in the field of the web with semantic and gives the meaningful results by a user query.

**VIII. COMPARISON OF SEMANTIC SEARCH ENGINES**

In this section, we compare all the semantic search engines which discuss in the paper.

Hakia gives the pure analysis of results and gives those documents which are relevant to our content. It gives result in Link & Free Text. Its search results are divided into web, News, Blog, Videos. It gathers data relevant to our query from different credible sites. It identifies data with focus and saves time.

DuckDuckGo is a Meta search engine that can gather information from other search engines. It gives result related to the text of user query. It gives result in the form of a summary. The results given by it are gathered from different sources including Yahoo and Wikipedia etc.

Sensebot uses text mining and multi-records summarization to abstract sense from Web pages. Sensebot generates a text summary of multiple Web pages. It gives us a summary of results instead of giving all links related to the query.

Cognition retrieves formula for meaning in the search. It gives result in LinkedIn. It is a search engine that gives the facility to learn about different aspects of different languages and supports Ontology, Morphology, and Synonyms.

Swoogle uses text mining and multi-records summarization to abstract sense from Web pages. It finds appropriate ontology’s and instance data structure of the semantic web.

Kosmix provides meaning in the search query. It gives the organize search results. It organizes web results for its user.

Factbites searches for authoritative and informative content. It uses text mining and multi-records summarization to abstract sense from Web pages. Its result gives us a sense of knowing the things closely and accurately. It is best for filtering out those sites which are not relevant.

**A. Some Term used**

Some terms which are used in table 1 are discussed
A Study on Semantic Searching, Semantic Search Engines and Technologies
Used for Semantic Search Engines

below.

**Multilingual:** multi-language Support.

**Result Explanation:** In this we identifies that whether the semantic search engine argues the query answer, explaining by means of a graph, conceptual structure or other.

Features: It is an explanation of semantic search engine qualities.

Type of Result: It shows the peculiarity of the query results these are of different types Link and Free Text etc. Comparison of different semantic search engines shows in table 1.

Table 1. Comparison of different semantic search engines

<table>
<thead>
<tr>
<th>Name</th>
<th>Features</th>
<th>Search Methodology</th>
<th>Result Summary</th>
<th>Type of Result</th>
<th>Search Results</th>
<th>Result Explanation</th>
<th>Multilingual</th>
<th>Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hakia</td>
<td>Excellent Resumes, Related Searches, Semantic Rank Algorithm, CMR.</td>
<td>Pure Analysis of contents.</td>
<td>Given Document are relevant the content</td>
<td>It gives result in Link &amp; Free Text</td>
<td>Its search results are Divided into web, News, Blog, Videos.</td>
<td>Yes</td>
<td>Yes</td>
<td>It gathers data relevant to our query from different credible Sites with ease. It identifies data with Focus and Saves time.</td>
</tr>
<tr>
<td>DuckDuckGo</td>
<td>Emphasizes Privacy, It does not track user personal information, It gives result using other sources or search engines including its own web crawler.</td>
<td>It is a Meta search engine that can gather information from other search engines.</td>
<td>Gives result related to text/conten of user query.</td>
<td>It gives result in form of summa ry</td>
<td>Images, local search, auto-suggest, weather, recipes and more.</td>
<td>Yes</td>
<td>Yes</td>
<td>Results given by it are gathered from different sources including Yahoo and Wikipedia etc.</td>
</tr>
<tr>
<td>Sensebot</td>
<td>It delivers summary instead of links to other website pages in the outcome of search query, Text mining, Multi-documents summarization, It finds top result related to search query and then makes summary of this data.</td>
<td>It uses text mining and multi-records summarization to abstract sense from Web pages.</td>
<td>Generates a text summary of multiple Web pages.</td>
<td>It also gives result in form of summa ry</td>
<td>Summary of all web pages related to our topic .</td>
<td>Yes</td>
<td>Yes</td>
<td>It gives us summary of results instead of giving all links related to our query</td>
</tr>
<tr>
<td>Swoogle</td>
<td>It is a linguistic search engine that supports: 1-Morphology 2-Ontology 3-Synonym</td>
<td>Natural language processing</td>
<td>Retrieves formula for meaning in the search.</td>
<td>It gives result in Link.</td>
<td>It permits searching across four domains 1-Law 2-Medicine 3-Wiki 4-Bible</td>
<td>Yes</td>
<td>Yes</td>
<td>It is a search engine that gives us facility to learn about different aspects of different languages and supports Ontology, Morphology and Synonyms</td>
</tr>
<tr>
<td></td>
<td>It provides different search services using REST interface</td>
<td>Indexes documents using RDF</td>
<td>Gives Semantic web results.</td>
<td>It gives result in OWL, RDF.</td>
<td>1-Web ontology’s 2-Documents 3-Terms 4-Data published on the web.</td>
<td>No</td>
<td>No</td>
<td>It finds Appropriate ontology’s and appropriate instance Data structure of the semantic web.</td>
</tr>
</tbody>
</table>
IX. RESULTS AND DISCUSSION

The results describe that semantic search engines are more efficient than others keyword-based search engines. The semantic search engines give the relevant results according to user queries. The semantic search enhances the power of searching. The semantic searching gives the relevant information very fast. The semantic search engines like duckduckgo remove the problem of polysemy words.

The description of semantic search engines as it provides a clear explanation of the semantic technology. A description of the components and the processes of semantic search engines gave a more comprehensive step-by-step method on the execution of searching. The analysis of different SSEs showed a layout of all the search engines. The comparison of different semantic search engines may include the following terms (features, search methodology, result in summary, type of results, search results, results summary, results in explanation, multilingual, advantages).

Although SSE provides the relevant results according to a user query, but these search engines provide a lesser number of results for a user query like when working on Hakia its observe that, its gives only a few results considering the semantic context of the given query. The result pages of these SSE’s are different from one another. The semantic search engines provide their own result pages according to specific features owned by search engines.

The depending on the context of giving queries some semantic search engines retrieve relevant documents directly, some of them makes a summary of relevant documents, and some use semantic keys to retrieve the results.

The duckduckgo semantic search engine provides results in categories form and this way the user can choose relevant terms. DuckDuckGo would remove the problem of Polysemy. Hakia topped in terms of Natural Language query processing and provide correct answers than other all semantic search engines. The after studying all existing information about semantic search engines the result describes that to develop a semantic search engine that allows the user to play a portion of the answers, before and after the query, which is pre-query visibility of ambiguity exists, and feedback to improve future answers.

X. CONCLUSION AND FUTURE WORK

In short, this brief survey of existing literature regarding semantic searching, semantic search engines and their technology yields unique search experiences for users. So, it is concluded that these semantic search engines are best because of their richness and efficiency. These semantic search engines provide a wide variety of related results regarding the requirements and thus, these are convenient to use and time-saving. Even if a search engine meets all the domains of requirements even then continues to remain a question. The paper can be a vital reference for further study of the use of semantic search engines. The comparison may include the following terms (features, search methodology, result in summary, type of results, search results, results summary, results in explanation, multilingual, advantages) in future: speed, the relevance of results, and ranking of results.

The enhancements proposed to be made in future, suggest developing such an efficient semantic search engine technology that is capable of meeting the challenges efficiently and has a compatibility with global standards of web technology.

REFERENCES

sciences ravensburg-weingarten, Germany.


Authors’ Profiles

Junaid Rashid has received his Master of Science in Computer Science degree from Department of Computer Science, COMSATS Institute of Information Technology, Wah, Pakistan in 2016. He received his BS (CS) degree in 2014 from COMSATS Institute of Information Technology, Wah, Pakistan.

He has been interested in research domains like Semantic Web, Semantic Search engine, Software Quality Assurance, Software Configuration management, Model versioning, Model Diff, Model Merge and Model Driven Engineering etc.

Email:junaidrashid062@gmail.com, Contact#:+92-343-5079668

Muhammad Wasif Nisar received his PhD degree candidate in computer science from Institute of Software, GUCAS China in 2008. He received his BSc degree in 1998 and MSc degree computer science in 2000 from University of Peshawar, Pakistan. His research interest includes software estimation, software process Improvement, distributed systems, semantic search engines, databases and CMMI-based project management.

Email:wasfinasar@gmail.com

How to cite this paper: Junaid Rashid, Muhammad Wasif Nisar, "A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines", International Journal of Information Technology and Computer Science (IJITCS), Vol.8, No.10, pp.82-89, 2016. DOI: 10.5815/ijitcs.2016.10.10