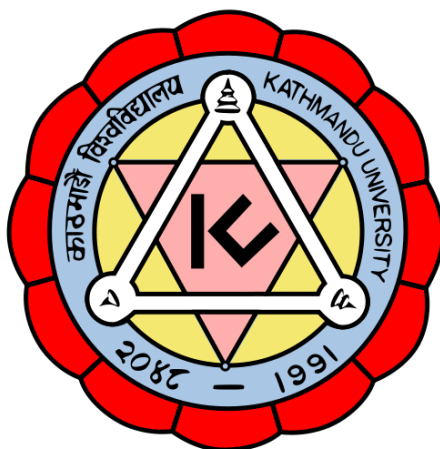


Semantic Web Search

Review Paper

January 15, 2013



Submitted to

Mr. Santosh Khanal

Department of Computer Science and Engineering

Kathmandu University

Dhulikhel, 5 Kavre

Submitted by

Milan Thapa (52)

Semester 7

milans.thapa78@gmail.com

Contents

1	Introduction	1
1.1	Search History before Semantic Web	1
1.2	Semantic Web —an Intelligent Web	2
1.3	Why Semantic Web Search ?	3
2	Semantic Search Procedures	5
2.1	Available Semantic search portals	5
2.2	SemSearch-an overview[1]	5
2.3	Implementation of Semantic Web	6
3	Intended Benefits of Semantic Search	7
4	Implementation Problems and Future Possibilites	9
5	Conclusion	10

1 Introduction

1.1 Search History before Semantic Web

From the invention of the internet and web, the informations and data are being accumulated, refined and distributed all over the world. Up to par, the web have been through its different versions. **Web 1.0** was all about the *information web*, where people just collected the data over the network. It got advanced to **Web 2.0** as a *social web* when people started getting connectd through social networking like hi5,linkdien,facebook. Meanwhile the data accumulation rate over the web got its exponential hit and the process is still being growing progressively. The search engines during the days of web 2.0 were just programmed to serve the user's query with some algorithms of page ranking *like Goolge PageRank in old days to predicate relevant result, before its knowledge graph*. With those chaos of the information all over the web, a user using some searching techniques like: the user provided a phrase which is intended to denote an object of which the user was trying to gather information about. But the poor search engines couldn't exactly resolve the phrase user was actually looking for, it is because the engines really didn't have any idea(clue) of what informations and data they were dealing with. They were just programmed, without any intillegence to analyse the meaning of a simple phrase. So with no altenatives, the user tried to locate a number of documents or pages which together would give him/her the information s/he was really after.

i.e. every time a user needed a information about something, s/he became an explore, exploring the web, filtering the contents, merging all together and finally drawing some desired meaning out of those.— number of procedures to do simple search.

So that is why **semantic web** emerged as an extension of the current web by standards and technologies that help machines to understand the information on the Web, to support richer discovery, data integration, navigation, and automation of tasks.

1.2 Semantic Web —an Intelligent Web

Semantic web —an intillegent web is a web designed to solve all the chaos of informations and data spreading all over the current internet. It's an evolving collection of knowledge, built to allow anyone on the Internet to add what they know and find answers to what they want to know[8] Information on the Semantic web, rather than being in natural language text, is maintained in a structured form which is fairly easy for both computers and people to interact and work with. The mechanism of linking the search engine to the semantic web give rise to the new terminology called semantic web search or so called semantic search. Its a search engine designed to improve search accuracy by understanding searchers intent and the contextual meaning of terms as they appear in the searchable dataspace (on the web), to generate more relevant results. As an explanatory example

When searching for "laptop", then one is looking for laptops or synonyms / related concepts (such as "notebook"), but also for special kinds of laptops that are not synonyms / related concepts, such as e.g. IBM/Lenovo ThinkPads. That is where web needs an intelligence and reply the desired results.ie semantic search.

One important goal of the semantic web is to make the meaning of information explicit through semantic mark-up, thus enabling more effective access to knowledge contained in heterogeneous information environments, such as the web. Semantic search plays an important role in realizing this goal, as it promises to produce precise answers to users queries by taking advantage of the availability of explicit semantics of information. Another example could be like this: when searching for news stories about "UNG students", with traditional searching technologies, we often could only get news entries in which the term "UNG students" appears. Those entries which mention the names of students but do not use the term "UNG students" directly will be missed out. Such news entries however are often the ones that the user is really interested in. In the context of the semantic web, where the meaning of web content is made explicit, the semantic meaning of the keyword (which is a general concept in this example of UNG students) can be figured out. Furthermore, the underlying semantic relations of metadata can be exploited to support the retrieving of information which is closely related to the keyword. Thereby, the search performance can be significantly improved by expanding the query with instances and relations.

1.3 Why Semantic Web Search ?

The effective exploration of the web content involves with the numerous challenges like:

Useability The user on a web searching for something doesn't care how the structured query is specified. The end user often does not know the query language and the underlying data scheme. For example searching pdf,docs,tex in google one simply couldnot know that use of tag filetype:pdf/doc/tex could be handy.

Scalability As the amount of available data is ever growing, the ability to scale the most desired results is essential.

Imprecise Information Needs The information needs expressed by the user might be imprecise. An effective search solution should be able to consider this aspect to deliver relevant results. For example: a user might search phrases like vdo for video,fbk for facebook,anime for animation,b4 for before, bday for birthday, etc.

Data Change The Web of Data is continuously changing. Thus, efficient mechanisms for index update at the Web scale are needed when data changes.

So understanding the user's intent by analysing keywords he/she is trying to search with the aid of dynamic web resources knowledge scattered over the internet is the first thing what semantic search is designed for. Thus it explored the possibility of search engines with a broad sense that a user happily could trust the search results that matched his/her intention. He/she can always put search phrase in clean and simple search interface with zero knowledge of any syntax and procedures, i.e, in his/her own ways and with his/her own language.

The need of the semantic search or semantic web is fundamental aspect of the current web.It could be generalise with some simple examples:

Example1: From the interview of Sir Tim Berns Lee "*Looks Back The In-Web Addresses Was Unnecessary[2]*" ,he had had and exmaple of going to an event which is self explanatory.

If you are going to a event, you need to first read the text on a web page... you have to do things with it. and then you need to tell your computer seperately when it is and where it is... which is very silly ! But if your computer may be your car knew where you are going, it might be able to take you there...to the right place via the right direction.

Example2: We can take that to another level. In streets of Kathmandu, many tourists buy "I Love Nepal" T-shirts. However, the word "love" is replaced by a heart. We know what that means. But how would a these mechanical servants:like computers?

Example3: Another one would be about a search condition we all been through once, since infomation retrival on the web is such a fascinating science. If we know the name of something, it's typically not too difficult to find it on the Web using a search engine. But when we don't know the name of something and have to use a description, search usually falls over. i.e.

Ask any Nepali about a fruit found in terai and hills, yellow-white colour when completely ripe, contains soft core inside covered with small granules. He'll tell you it's a 'Guava' 'Amba'. But just throwing a few ingredients into a search engine may not get us such a quick or even correct response.

2 Semantic Search Procedures

2.1 Available Semantic search portals

Some available list of semantic search portals are tabulated below:

Search Portals	Description
Bing	particularly its new Knowledge Graph feature first semantic search engine on the Internet(2002)
Google	
GoPubMed	
Hakia	
iGlue	
Kosmix	semantic search engine with realtime annotator plugin/bookmarklet which adds a smart layer to every website
Lexxe	
Swoogle	
Yummly	social media semantic search beta in early 2011
Rendipity	
	food and recipe semantic search
	image semantic search

Describing about what really is happening under the hood of those semantic search portals, we could have an overview of the SemSearch[1], a simple semantic search engine.

2.2 SemSearch-an overview[1]

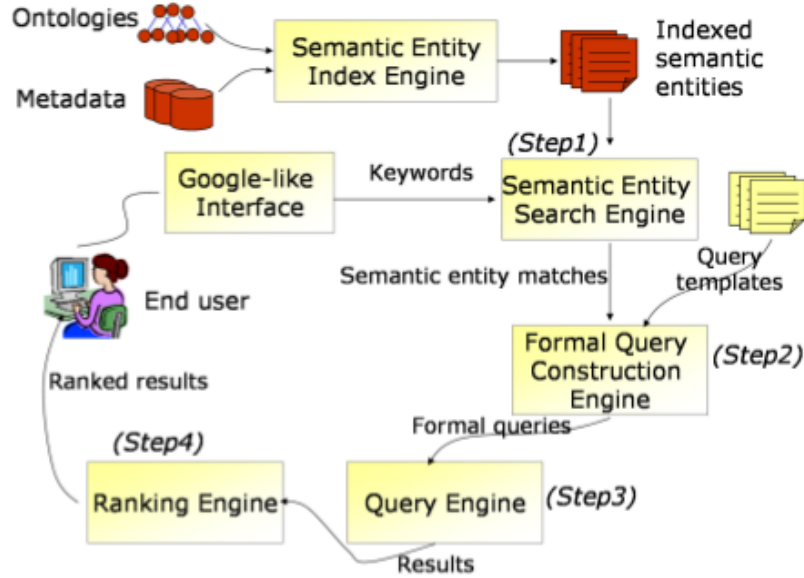
One major goal of SemSearch is to hide the complexity of semantic search from end users and to make it easy to use and effective for naive users. An overall overview of SemSearch could be visualised in the diagram. The major steps involved in SemSearch are:

Step1. Making sense of the user query, which is to find out the semantic meanings of the keywords specified in a user query.

Step2. Translating the user query into formal queries.

Step3. Querying the back-end semantic data repositories using the generated formal queries.

Step4. Ranking the querying results.



SemSearch provides a Google-like query interface allowing users to specify queries without requiring any knowledge about ontologies or specific languages. User queries consist of two or more keywords, whose semantic meaning is taken into account to reformulate the queries themselves according to a formal query language syntax. Keywords are assigned a semantic meaning by matching them against a collection of classes, properties, and instances in semantic data repositories. Since each keyword can match a class, a property, or an instance, several combinations of semantic matchings of the keywords are considered. For instance, it can be the case that every keyword matches a class, or that the first keyword matches a class, while the second matches a property, and so on. All the combinations of matchings are taken into account in the reformulation process, and each combination leads to a distinguished formal query, obtained from a pre-determined set of query templates. After the reformulation, formal queries are exactly evaluated, and this yields results that are semantically related to all the user keywords.

2.3 Implementation of Semantic Web

Up to par, semantic web consists of set of technologies such as semantic graphs, descriptive logic, tagged data, ontologies defining relationships between entities of web, etc. Some of standards being followed in recent time are RDF,

OWL, HTTP, RDFa, etc. OWL and RDF have been growing very popular.

OWL(Web Ontology Language) is intended to be used when the information contained in documents needs to be processed by applications, as opposed to situations where the content only needs to be presented to humans. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms.

RDF (Resource Description Framework) is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed.

Some of the tools used in Semantic Web are Protege, Jena, Sesame, Semantic Media Wiki, etc. A set of popular artifacts implementing Semantic Web are Dbpedia, Cyc and Linkeddata.org.

DBpedia is a project aiming to extract structured content from the information created as part of the Wikipedia project. It allows users to query relationships and properties associated with Wikipedia resources, including links to other related datasets. Cyc is an artificial intelligence project that attempts to assemble a comprehensive ontology and database of everyday common sense knowledge, with the goal of enabling AI applications to perform human-like reasoning. One stated goal is that of providing a completely free and unrestricted semantic vocabulary for use in the Semantic Web.

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods.

3 Intended Benefits of Semantic Search

Using the semantic web search end user could experience new level of satisfaction. Semantic Web as a whole would certainly improve the search, automate topic recognition, better target advertisements, elevate commercial possibilities, save time, save money and so on. There are lots of benefits that everybody on the web could entertain. Some intended benefits that are possible out of this engineering work are :

Intelligent Search: Semantic web is representation of the data enveloped inside the documents, multimedias, maps, memos, etc. So using a search to locate those relationally linked data well structured in some

format, the more meaningful search results could be feed to the end users.

Enhanced Result: With Semantic Web the web could be programmed to serve more precise information with an ability to manipulate the actual data using some natuaral language processing algorithms or graphs or machine learnings technologies and finally serve the results in multiple dimensions like charts,graphs,pictures,audios,videos or any non verbal apporach rather than a single hyperlink to a page containg the information all in simple human cognitive forms.

For example: When somebody search for a piece of information on the web thorough the interface like sound recording mic/picture, he or she could have some intention to grab some practical and informative results out of it. Meanwhile if the semantic search somehow results some non verbal resutlts like pictures or videos or media associated with it, it would mimimise the gap of the digital divide between the novice user of the web and the data out there on the web. For instance, a farmer growing his/her crop at the field enable enough to use a mobile device could know about what disease his crops are infected with using this search extension. Thats an enhancement, what semantic search is envisioned for.

Automated Question Answering: This can be best explained though Knowledge Graph. The Knowledge Graph enables the user to search for things, people or places that Google knows about - landmarks, celebrities, cities, sports teams, buildings, geographical features, movies, celestial objects, works of art and more and instantly get information thats relevant to the query.

Facilitate controlled vocabulary development: Web-based solution for simplified development and management of interconnected controlled vocabularies can also facilitated. It would support business stakeholders who need to collaborate on defining and linking enterprise vocabularies, taxonomies, thesauri and ontologies used for information integration, customization and search.

4 Implementation Problems and Future Possibilities

A number of semantic search tools have been recently developed. Present semantic search tools are enhancing the performance of traditional search technologies, they are however not suitable for naive users, i.e. ordinary end users who are not necessarily familiar with domain specific semantic data, ontologies, or SQL-like query languages. The semantic search engines should be implemented in such way that it could address these issues:

- the search engines should tackle the problem of knowledge overhead by supporting a Google like query interface such that the proposed query interface provides a simple but powerful way of specifying queries and also help avoid the undesired typing mistakes.
- it should address the problem of existing semantic-based keyword search engines by supporting complex queries. It should provide comprehensive means to make sense of user queries and to translate them into formal queries, thus being able to produce precise results that on the one hand satisfy user queries and on the other hand are self-explanatory and understandable by end users. For example support for two or more search phrases separated by space or some characters: *poverty: Nepal NorthAmerica SouthKorea*
- it should be properly designed and tested so that it serves promising results that is so quick, clean and clear.

We have given a brief overview of approaches to semantic search on the Web, which is currently one of the hottest research topics in both the Semantic Web and the Web search community. With the integration of semantic value to the existing web, the reasoning capabilities envisioned in Semantic Web technologies are coming to Web search and the Web. Although many approaches and systems to semantic search on the Web already exist, the research in this area is still at the very beginning, and many open research problems still persist. Some of the most pressing research issues[3] are maybe:

1. how to automatically translate natural language queries into formal ontological queries,

2. how to normalise the data to be secure or, alternatively how to maintain appropriate privacy?
3. how to manage e-commerce complexity, the marketers could have had practise about advantage over the individuals data.
4. how to automatically add semantic annotations to Web content, or alternatively how to automatically extract knowledge from Web content.
5. how to consider contextual information to adapt the search results to the needs of the users.
6. how to create and maintain the underlying ontologies.

and some of them could be answered as:

- manually by experts, e.g., in a Wikipedia like manner, where different communities may define their own ontologies.
- automatically, e.g., by extraction from the Web, eventually coming along with existing pieces of ontological knowledge and annotations (e.g., from existing ontologies or ontology fragments, and/or from existing annotations of Web pages in microformats or RDFa), or
- semi-automatically by a combination of both above.

5 Conclusion

As a summary it could be stated that with Semantic Web and Semantic Web Search integration with web 2.0, the data on the web would be in more manageable form i.e data would be more organised, easy to retrieve in meaningful manners. Even with its primitive exploration, Web is getting smarter, more people are using it, large sets of data are being classified to look patterns, people are using predictive modeling, huge commercial usage is elevating and so on. With this semantic empowerment of the web, there is a clean structuring of the data and user looking behind it. Also these mechanical servants (electronic devices) are being more useable to our day to day activities. But with this semantic evolution of web there comes some very serious implications like privacy implications, and other issues such as misuse of the data for commercial benefits or other advertisements.

Up to now, the semantic web search is at its developing stage. Performing Web search in the form of returning simple answers to simple questions in natural language is still science fiction. However, with the current activities towards semantic search on the Web, we are moving one step closer to making such desired environment.

References

- [1] Y. Lei, V. S. Uren, and E. Motta. *SemSearch: A search engine for the Semantic Web*. In Proc. EKAW-2006, LNCS 4248, pp.238–245. Springer, 2006.
- [2] Interview with Sir Tim Berns Lee, *Looks Back The In Web Addresses Was Unnecessary*
- [3] Bettina Fazzinga and Thomas Lukasiewicz. *Semantic search on the Web*.IOS Press,2010.
- [4] Goole. www.google.com/insidesearch.
- [5] Linked Data. <http://linkeddata.org/>.
- [6] Semanticweb. <http://semanticweb.org/wiki/Cyc>.
- [7] Wikipedia. <http://en.wikipedia.org/wiki/DBpedia>.
- [8] Sandro Hawke. *How the Semantic Web Works*. Overview.html,v 1.15 2002/04/22 19:11:29 sandro Exp.