Semantic Web Search

Research Paper

Submitted to

Mr Santosh Khanal

Deptartment of Computer Science and Engineering
Kathmandu University
Dhulikhel, 5 Kavre

Submitted by

Milan Thapa (52)
Sem 7

Contents

1 Introduction

1.1 History before Semantic Web

From the invention of the internet and web, the inforamtions and data are being accumulated, refined and distributed all over the world. Up to par the web have been thorugh its different versions. **Web 1.0** was all about the *information web*, where people just collected the data over the network. It gets advanced to **Web 2.0** as a *social web* when people started getting connectd through social networking like facebook, linkdien. Meanwhile the data accumulation rate over the web gets 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 provides a phrase which is intended to denote an object of which the user is trying to gather information about. But the poor search engines can't exactly resolve the phrase user is looking for, it is because the engines really don't have any idea(clue) of what they are dealing with. They were just programmed without any intillegence to analyse the meaning of a simple phrase. So with no altenatives, the user tries to locate a number of documents or pages which together will give him/her the information s/he is trying to find. *i.e. every time a user needs a information about something, s/he becomes an explore, exploring the web, filtering the contents, merging all together and finally drawing some desired meaning out of it.*

So that is how **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 web. It can altenatively visualised as a representation that both the machines(computers) and human can interact with ease. 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.

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 work with. Semantic search seeks to improve search accuracy by understanding searcher intent and the contextual meaning of terms as they appear in the searchable dataspace, whether on the Web or within a closed system, to generate more relevant results. Author Seth Grimes lists "11 approaches that join semantics to search", and Hildebrand et al. provide an overview that lists semantic search systems and identifies other uses of semantics in the search process. 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.

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. For example, when searching for news stories about phd students, with traditional searching technologies, we often could only get news entries in which the term "phd students" appears. Those entries which mention the names of students but do not use the term "phd 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 the example of phd 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.

A number of semantic search tools have been recently developed [5, 4, 7, 2,9, 6]. Our overview of the state-of-art semantic search tools reveals that while these tools do enhance 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 engine we present here, SemSearch, provides several means to address this issue.

–SemSearch tackles the problem of knowledge overhead by supporting a Google- like query interface. As will be described in Section 4, the proposed query interface provides a simple but powerful way of specifying queries. SemSearch addresses the problem of existing semantic-based keyword search engines by supporting complex queries. It provides comprehensive means to make sense of user queries and to translate them into formal queries. SemSearch takes the focus of user queries into consideration when generating 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. Thus, SemSearch makes it possible for ordinary end users to harvest the benefits of semantic search and other semantic web technologies without having to know the underlying semantic data or to learn a SQL-like query language. A prototype of the search engine has been implemented and applied in the semantic web portal of our lab1 . An initial evaluation shows promising results. The rest of the paper is organized as follows. We begin in Section 2 by investigating how current semantic search tools approach the issue of end user support. We then present an overview of SemSearch in Section 3. Thereafter, we explain the Google-like query interface in Section 4. We describe the major steps of the semantic search process in sections 5 and 6. In Section 7, we describe the implementation of SemSearch and the experimental evaluation. Finally, in Section 8, we conclude our paper with a discussion of our contributions and future work.

1.3 Why Semantic Web Search ?

The effective exploration of the web content involves with the numerious challanges:

Useability The user on a web searching for something does't care how the structured query is specified. The end user often does not know the query language and the underlying data scheme.

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.

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

2 Available Semantic search portals

Some available cool semantic search portals are tabulated below:

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

3 Vision for Future

We have given a brief overview of approaches to se- mantic search on the Web (also called Semantic Web search), which is currently one of the hottest research topics in both the Semantic Web and the Web search community. In semantic search on the Web, the cur- rent strong research activities of the former to realize search on the Semantic Web are merged with the cur- rent strong research activities of the latter to add se- mantics to Web queries and content when performing Web search. It is through this integration that the rea- soning capabilities envisioned in Semantic Web tech- nologies are coming to Web search and the Web. As we have seen, the formulation of queries and their results in semantic search on the Web is ultimately directed by a third area, namely, the one of question answering sys- tems, which is based on natural language processing. Although many approaches and systems to seman- tic search on the Web already exist, the research in this area is still at the very beginning, and many open re- search problems still persist.

Some of the most pressing research issues are maybe

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

2. how to automatically add semantic annotations to Web content, or alternatively how to automatically extract knowledge from Web content. Another central research issue in semantic search on the Web is

3. how to create and maintain the underlying ontologies.

This may be done either (a) manu- ally by experts, e.g., in a Wikipedia like manner, where different communities may define their own ontolo- gies, or (b) 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 (c) semi-automatically by a combination of (a) and (b). Clearly, the larger the degree of automa- tion, the larger is also the potential size of ontologies that can be handled and the smaller are the costs and efforts for generating and maintaining them. So, for the very large scale of the Web, a very high degree of au- tomation is desirable. A closely related important re- search challenge is (iv) the evolution and updating of and mapping between the ontologies that are underly- ing semantic search on the Web, where it is similarly desirable to have a very high degree of automation. A further important issue is (v) how to consider im- plicit and explicit contextual information to adapt the search results to the needs of the users. For example, the needs and motivations of users may be defined in terms of ontology-based strict and/or soft (weighted) constraints and (conditional) preferences (e.g., similar to [24]), which may then implicitly be expanded into the semantic search query and/or used in the computa- tion of the ranking on objects and search results. Performing Web search in the form of returning sim- ple answers to simple questions in natural language is still science fiction, let alone performing Web search in the form of query answering relative to some concrete domain or even general query answering. However, with the current activities towards semantic search on the Web, we are moving one step closer to making such