Semantic Search and Social-Semantic Search as Cooperative Approach

[1]Ms. Priya B. Pandharbale, [2] Mr. Rahul S.Kadam,[3]Mr. Jaydeep B. Patil
[1]Assistant Professor, BVCOEK, Shivaji University, Kolhapur, Maharashtra, India,
priyasathe123@gmail.com
[2]Assistant Professor, D.Y.PatilCollege of Engineering,pune,SPPU,Pune,Maharashtra,India,
kadam.rahul49@gmail.com
[3]Assistant Professor, AISSMS's Institute of Information Technology, pune, SPPU, Pune,Maharashtra, India,
er.jaydeep7576@gmail.com

Abstract: Social and semantic web can be combined for searching web resources. A semantic search engine can find accurate results and annotate web resources using this cooperative approach.As the volume of information is growing, the syntactically correct outputs given by traditional search engines for the user queries have enlarged directly. In order to find exact answers for user queries many more Semantic Search Engines (SSE) are developed now a day. The Semantic Search Engines use a wide range of methods for matching the semantics behind user queries and the indexed collection of resources. The survey shows the semantic search engines domain, and presents a miscellaneous of perspectives about the different classification of approaches. A comparative scheme is presented here and the prevalent research directions in SSE with the advancements in it are identified for the efficient searching techniques.

Keywords*-Social web, Semantic Web,Semantic Search Engines, Ontologies, Information Retrieval.*

*****

I. Introduction

The search results provided by the traditional search engines are irrelevant and hence can frustrate the users. These answers are not precise enough for some users demanding a more refined list of results according to the semantics of their queries. This open problem has motivated a new era of search systems that have received the name of Semantic Search Engines.The goal of this work is to study research directions in semantic search engines. An overview of current approaches to semantic search can be seen here. One of the reasons for which Social Web or Web 2.0 became so popular is that it is focused on contents, relations and knowledge and not precisely on technology. Web 2.0 technologies augment the Web power.

II. Related Work

1. Semantic search:

Semantic search technology enables accurate retrieval of information via concept/meaning match. It is very effective, and perhaps the only method, in application to credible and dynamic content. Because most of the credible and dynamic content are statistically flat (infertile) for popularity algorithms to work effectively beyond common queries.

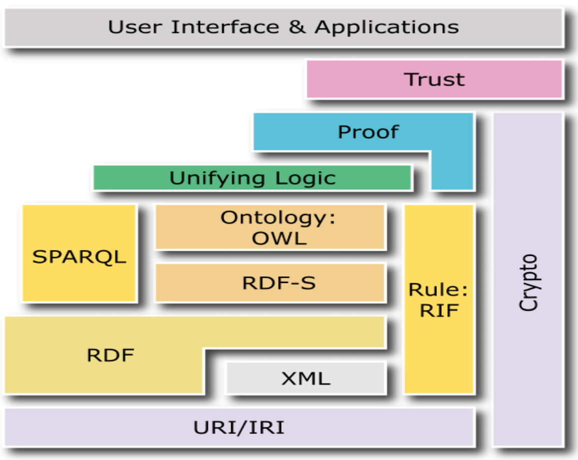

Figure 1: A more recent version of the Semantic Web stack.

A Semantic Search Engine (SSE) can be understood as a semantic Web application that can answer questions based on the meaning of users query specification, resources in the repositories and in many cases it is based on predefined domain semantics or a knowledge model. SSE can return relevant results on your topics that do not necessarily mention the word you searched for explicitly.

1.1 Semantic Web technologies

The Semantic Web technologies have been represented in a stack often called the "Semantic Web cake" or "Semantic Web stack" as shown in Figure 1. The stack shows the layers of technologies required to realize the full Semantic Web vision.

The bottom layers of the stack have been fully realized (at least upto RDF + rdfschema).The Ontology vocabulary layer has been partly realized and is actively being developed. The upper layers are still not quite mature at the web scale though these have been deployed within local or enterprise levels. However, the Semantic Web stack is itself evolving frequently along with new technologies, research and practical challenges coming to the scene. Existing technologies. The basic Semantic Web technologies and frameworks are quite well-established by now. Just as web documents are identified and interlinked by URLs, data resources are identified and interlinked by URIs (Uniform Resource Identifiers) in the Semantic Web. RDF (Resource Description Format)8 has become the standard language used to describe data for the Semantic Web. With the RDF model, all information is represented as (subject, predicate, object) triples, also known as RDF triples.

OWL (Web Ontology Language) has become the standard for representing Semantic Web ontologies. Similarly, SPARQL12 has become the standard for querying in the Semantic Web. Many ontologies have been created for different information domains. Semantic Web technologies are successfully being used in many industrial applications.

1.2 Different Schemes for Comparison and Classification of SSE

Extensive classification of approaches:

- Related searches/queries: The SSE recommends searches that are in some "sense" similar to the user search.
- Reference results: SSE is responding with resources that define the search terms, via a dictionary look-up, or elaborately, pulling Wikipedia pages.
- Semantically annotated results: SSE returns pages or documents with high-lighting of text features, especially named or pattern-defined entities.
- Full-text similarity search: SSE use a block of text ranging submitted from a phrase to a full document, rather than a few keywords.
- Search on semantic/syntactic annotations. Users define the semantic of search by means of indicate the syntactic role the term play.
- Concept search: The SSE identifies specific concept to seek the original and their equivalent concepts semantically.
- Ontology-based search: SSE can understand hierarchical relationships of entities and concepts as in taxonomy, and more complex inter-entity relations.
- Semantic Web Search: SSE capture data relationships and make the resulting "Web of data" query able.
- Faceted search: It provides a means of exploring results according to a set of predefined, high-level categories called facets.
- Clustered search: It is like faceted search, but without the predefined categories. Meaning is inferred from topics extracted from the search results.
- Natural language search: The SSE understands the semantic behind the questions, and present answers in natural language.

A summary about the classifications is presented in Table 1. The goal is identify the main active areas in SSE domain [1].

2. Analysis of current SSE

According to the researchers to improve traditional engines, including features like: user feedback; results explanation and compressive presentation of results; and more dialogue with the users about possible problem with their request, e.g ambiguity advertisement is needed.

To study the interoperability has become important, that is, to analyze the kind of interoperability present in SSE,whether the SSE is a machine or informatic agent query able. The SSE exploration is summarized in Table 1[3], which show the following 8 parameters:

- **Main approach(es)**: This field identifies the type of approach used by each SSE.
- **Features**: It is a description about the main SSE qualities.
- **Type of Result**: It specifies the query result: summary, link, free text or other.
- **User feedback**: This is useful when there are multiple controlled terms that match with the free text input semantically. There are two ways. The first one is called "pre-query disambiguation", allow us to select the intended term be-fore it is processed by the search algorithm. The second way is called "post-query disambiguation"; feedback is taking into account on the results.
- **Multilingual**: Multiple language support.
- **Interoperability**: It evaluates if the SSE is able to exchange machine understable content by mean of a standard protocol.
- **Result explanation**: Here the SSE argue the query answer, justifying by means a graph, conceptual structure or other.
- **Ambiguity alarm**: In many cases, there are results that match with the query. SSE must advert to user about the different senses that satisfy the query.

Additional two features are presented available in online version, as following:

- **Geospatial component**: It allows evaluate as if the SSE takes into account additional richness aspects, such as geospatial location information when is required to complement or clarify the semantic or to confirm the result sense. i.e. Washington state instead of Washington president (see RDF online).
- **Availability**: It examines if the Web application is available now (see RDF).

Web applications and publications describe their approaches from a very abstract view-point. Classifying the SSE according to their external description and comparing with similar semantic search engines is being done without having deep knowledge about them. In table, the symbol "-" represents unknown information. The main parameter of comparison in the table is the second column "Main approach(es)". It allows us to identify the research areas with the more intense activity in the semantic search.

ISSN: 2321-8169
110 – 114

Table 1.Comparison of semantic search Engines

Engine	Main Approach(es)	Features	Type of Result	Ma	Interoperability	RE b	AA c
SenseBot	Concept Search	Text mining	Summary	Yes	SOAP, REST	No	No
BotPowerseet	Natural Language Process-ing (NLP)	Free text input, disambiguate.	Summary	Yes	-	Yes	Yes
DeepDyve	Semantic/Syntactic An-not., Reference results	Analysis across large amounts of data	Summary	Yes	-	No	No
Cognition	NLP	Business, APIs	Link	Yes	API	Yes	Yes
Hakia	Related searches, NLP	Excellent re-sumes	Link & Free text	Yes	Yes	Yes	No
TrueKnowl-edge	Ontology-based search, Semantically annot. results	Questions – answering	Summary and clas-sification	No	Direct Answer API, Query API	Yes	Yes
Open Mind Common Sense	NLP, concepts search	Learn general knowledge	Free text	No	-	No	No
Swoogle	Semantic Web search	Semantic Web documents.	OWL, RDF	No	REST web service	No	No
TrueVert	Concept search, NLP and Clustered results	model of word relations in con-text	Free text	Yes	-	No	No
Wolfram Alpha	Reference results, Ontol-ogy-based search, Clus-tered search	Web, parallel computing, mathematical, grid knowledge	Taxonomy, graph	Yes	REST API	Yes	No
Duck Duck Go	Clustered search, NLP	Zero-click Info above links, Dis-ambiguation	Summary	Yes	XML-based API	-	Yes

a Multilingual, b Result explanation, c Ambiguity alarm

3. Social-semantic search

Social and Semantic Web are two approaches complementary and each must draw from the other's strengths. In this regard, the ontology metadata provides the benefit of enabling a semantic search engine to find accurate results and to apply reasoning procedures on the metadata. Respect to the social dimension, Wu et al. in states that social annotations remove the high barrier to entry because web users can annotate web resources easily and freely; it directly reflects the dynamics of the vocabularies of the users and thus evolves with the users. This cooperative approach is called, social-semantic web. To rniai et al. it will let 'creating, managing and sharing information through combining the technologies and approaches from Web 2.0 and the Semantic Web'. Merging the best of both approaches can play a crucial role in the OER search: semantic enrichment of tags or content created by users through social tools, social annotations for recommend social systems and folksonomies to populate ontology.

4.1 Structured Data Production in the Social Semantic Web
In spite of the challenges, the combination of social and Semantic Web technologies is definitely promising and a lot of work has been done and are being done in this area. The combination is significant for the production of structured contents required for the practical realization of the Semantic Web. Hence, creation of structured contents in the social Semantic Web is a major focus. The sources of structured data may be different. For e.g., data may come from the users, existing web pages, users desktop, unstructured text, databases, etc. In some systems users actively contribute structured data. In other approaches, users continue to use the existing systems and semantic contents are derived from these indirectly without involving the users. Some systems only produce structured instance data while some produce concepts and ontologies too. Users may participate independently or collaboratively for content creation.

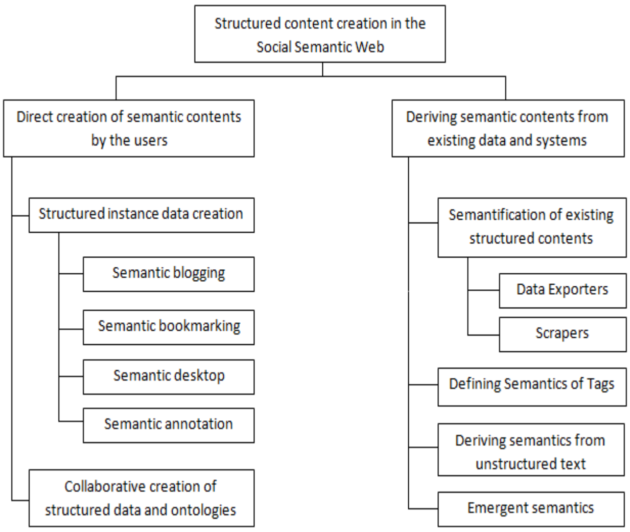

Figure 2: Structured content creation in the social Semantic Web

3.1.1 Direct creation of semantic contents by the users

In this broad category, the users explicitly create the semantically structured contents. This may further be classified into two groups based on the type of content created.

3.1.1.1 Structured instance data creation

In this category, the users directly contribute structured instance data only based on some existing ontology or concept schema. Usually the users contribute data independently without any collaborative effort with others in the community. However, the entire community benefits from the collection of individual personal contributions. There are several works in this category semantic blogging is such a work. By it blogs have made publishing information on the web very easy. Blogs serve as dynamic media showing the latest posted information. Blogs can effectively capture informal knowledge from several users and cater to the entire community. Conventional database driven information systems are rigid and do not covers all types of information that people may want to share within an organization or community. Informal snippets in blogs can cover a wide variety of information. However, traditionally blog entries do not have much structure and organization and cannot be processed effectively. Semantic blogging is a technology that builds upon blogging and enriches blog items with metadata. Semantic blogging exploits the easy publishing paradigm of blogs and enhance them with semantic structure. It combines desirable features of both blogging and the Semantic Web. Blogging provides an easy platform for online publishing along with mechanisms likes RSS, comments and trackbacks. The semantic web can provide well-defined structure to information based on ontologies so that it can be processed by machines. This also enables interoperability between different systems and facilitates information exchange. Pieces of structured data in semantic blogs can be interlinked with semantic relations. This enables meaningful navigation and organization of related contents in blogs. Semantic blogging can extend blogging for decentralized informal knowledge management. Some works done in semantic blogging are as follows.

The Semantic Blogging Demonstrator is a semantic blog for the bibliographic domain. Blog entries contain bibliographic items as metadata.

4. Socialsemantic searching technologies

Some Social semantic searching technologies are enlisted below:

- **Social context in online communities:** Here is an approach to extract social context from online social communities, and a prototype that exploits this information in the browsing process. By using the SIOC ontology we can have access to high-quality data with rich structure, which can be directly analyzed for implicit social relations. Relations between people can be derived from their online interactions, such as content that they create or reply to.
- **Blogs:** Easy usable user interfaces to update contents.Easy organization of contents. Easy usage of contents. Easy publishing of comments.Social collaborative (single users but strongly connected)
- **Wikis:** Wiki à invented by Ward Cunningham. Collection of HTML sites for reading and editing. Most famous and biggest Wikiexample is Wikipedia (MediaWiki) can also often used in Intranets. The Problems are solved socially instead of technically. It has Flexible structure. Uses Background algorithms + human intelligence
- **Delicious:** It allows the tagging of bookmarks. Community aspect usage is for suggestion of tags that were used by other users. Availability of tag clouds for bookmarks of the whole community. Possibility to browse related bookmarks based on tags.
- **Semantic Blogging:** Creating blog entries in a structured fashion.Based on the ontologies. This allows Acquiring complementary information from the Web.
- **Semantic Wikis:**A semantic wiki is a wiki that has an underlying model of the knowledge described in its pages. Regular, or syntactic, wikis have structured text and untyped hyperlinks. Semantic wikis, on the other hand, allow the ability to capture or identify information about the data within pages, and the relationships between pages, in ways that can be queried or exported like database data.
- **SoftWiki**: Ontology for Requirements Engineering In order to semantically support the requirements engineering process the SoftWiki Ontology is developed for Requirements Engineering (SWORE) in accordance with standards of the requirements engineering community. Central to the approach are the classes.

ISSN: 2321-8169
110 – 114

- **MyOntology:** A tool which helps specialists and ontology experts to collaborate easily. MyOntology uses the Web 2.0 paradigm gives collaboration of specialists and ontology experts. In first phase (until lightweight ontologies).It provides High usability, Integration and Reusing of web knowledge. (Web 2.0: Folksonomies, Flickr, YouTube, Wikipedia, etc.)

III. Conclusion

The best of both semantic and social approaches can play a crucial role in semantic enrichment of tags or content created by users through social tools, social annotations for recommend social systems and folksonomies to populate ontology.

References

[1] Walter Renteria-Agualimpia, Francisco J. López-Pellicer, Pedro R. Muro-Medrano, Javier Nogueras-Iso, and F.JavierZarazaga-Soria,.International"Advances in Intelligent and Soft-Computing". Springer, 2010, vol. 79, p. 613-620.

[2] Piedra, N., Chicaiza, J., López, J., Tovar, E., and Martínez, O., 'Openeducational practices and resources based on social software, UTPLexperience'. Proc. Int. Conf. Euro American Conference on Telematicsand Information Systems, ACM, 2009, pp. 1-8

[3] Gruber, T., 'Collective Knowledge Systems: Where the Social Web meets the Semantic Web', Journal of Web Semantics, 6 (1), 2008, pp. 4-13.

[4] Dong, H., Hussain, FK., and Chang, E., A survey in semantic search technologies, Second IEEE International Conference on Digital Ecosystems and Technologies, 2008.

[5] Figueira, F., Porto de Albuquerque, J., Resende, A.; Geus, Lício de Geus, P., Olso, G., A visualization interface for interactive search refinement. In: Proc. 3rd Annual Workshop on Human-Computer Interaction and IR, pp. 46-49, Washington DC, 2009.

[6] Hildebrand, M., Ossenbruggen, J., and Van Hardman. L., An analysis of search-based user interaction on the semantic web. Report, CWI, Amsterdam, Holland, 2007.

[7] O'Sullivan, D. Dooley, L.: 'Collaborative innovation for the management of information technology resources', International Journal of Human Capital and Information Technology Professionals, 1, (1), 2010, pp. 16-30

[8] Zhi-Qiang, D. Jing, H. Hong-Xia, Y. Jin-Zhu, H., 'The Research of the Semantic Search Engine based on the Ontology'. Wireless Communications, Networking and Mobile Computing, 2007.

[9] Torniai, C., Jovanovic, J., Gaševic, D., Bateman, S., &Hatala, M., 'Elearningmeets the Social Semantic Web'. Proc. Int. Conf. Eighth IEEEInternational Conference on Advanced Learning Technologies,Washington DC, USA, IEEE Computer Society, 2008, pp. 389-393

[10] Ankolekar, A., Krötzsch, M., Tran, T., and Vrandecic, D., 'The two cultures: Mashing up Web 2.0 and the Semantic Web'. Journal of Web Semantic, Elsiver, 6 (1), 2008, 70-75.

[11] Herzog, C. Luger, M. Herzog, M., 'Combining Social and Semantic Metadata for Search in a Document Repository', Proceedings of the European Semantic Web Conf., Innsbruck-Austria, 2007, pp. 14-21.

[12] Wu, X. Zhang, L. Wu, X., 'Exploring social annotations for the semanticweb', Proceedings of the 15th international conference on WorldWideWeb,2006,pp.417-426.