Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 38 (2012) 2164 – 2171

Procedia Engineering

www.elsevier.com/locate/procedia

Conference Title

A Machine Learning Technique for Semantic Search Engine

G.Nagarajan[a] and K.K.Thyagharajan[b],a*

[a]*Research Scholar,Sathyabama University,Chennai,India*
Proffessor,Depart. Of Information & Technology,RMK College of Information and Technology,Chennai,India

Abstract

In this world of information technology revolution we human wants a way of automation in every field, but the intelligent technique behind the way how we visualize an information is not equivalent to how an computer system visualize .In web information content is simply a resource with some specific hyper-link path. In this paper we propose an idea of giving semantic to a web page so a system can understand the semantic behind the web page which automatically increases the efficiency of information search. So we represent a way of converting an ordinary Syntactic page into a Semantic web page with corresponding Ontology which would pave the way of advancement in Semantic Web Learning technology.

Keywords: Machine Learning, Semantic Web,NLP, Enity Relationship,RDFS & OWL

1. Introduction

Information is the main source of intelligent. In today's revolutionary era the amount of information available on web is enormous. Just by typing any keyword on any search engine will provide you with millions of information, but the amount of relevant information would be very few and the user again has to search manually on all those million result . So, this kind of search engine is not user friendly . The primary approach of this paper is to design a Intelligent search engine which has to provide only the needed relevant information regarding the given query.

Semantic search engine is the only key answer for this kind of search. As said in [1] here both the machine and the user tries to search some information on web. There are many research papers

* Corresponding author. Tel.: +91 9444173937; fax: +0-000-000-0000 .
E-mail address: nagarajanme@yahoo.co.in

doi:10.1016/j.proeng.2012.06.260

regarding the design of a new semantic search engine. In [1] even listed top five to ten Semantic Search Engine. The main drawback of Semantic Search Engine is that the available Semantic Web page on web is very few. As the concept of Semantic Web had started on around 2000, we have very few Semantic Web page . As the creation and design of Syntactic Web page is ease of work also we have lot of in-built software for it still people are interested in creating simple Syntactic web page with general XHTML,XML,PHP,ect. coding instead of construct ontology for it.

The main focus of this paper is to design an Web Intelligent Framework for converting a Syntactic web page to Semantic Web page. Thus if a framework is available we can convert all those web pages and make it available for Semantic Search Engine thus we can provide a way for efficient search engine where both the User and Machine search an information from Web. Fig.1 shows the key technology of semantic web technology.

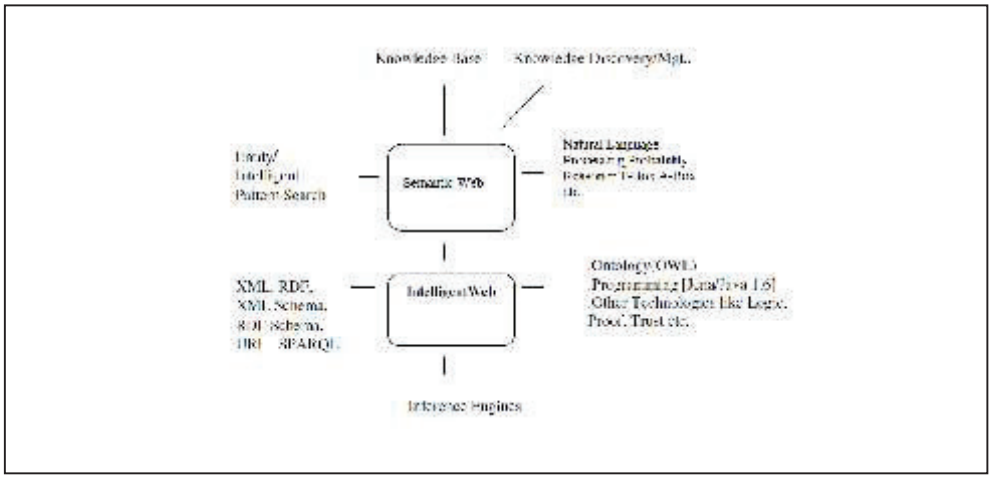

Fig 1. Key Technology of Semantic Web

This paper is design in such a way that first we explained the concept of semantic search engine and there categories , then we start with providing the design of Our Proposed Semantic Search Engine Framework, then we concluded our work with future scope.

2. Semantic web Search Engine

Information sharing is the main objective of World Wide Web, so it is obvious that the amount of information in internet would be tremendous and we are in need if search engine to retrieve the needed information from web. The main drawback in today's search engine is that it shows trillions of link for any search query. This can be overcome by this Semantic Web technology. Search is one of several applications for the Semantic Web. Where both man and machine can search a content as the content of the web page is framed in such a way that a machine can also understand.

We can categories the types of Semantic Search Engine is different way one of the way suggest by [..] are that it can be mainly categories as Ontology Search Engine and Semantic Search Engine. Ontology is one of the main component of Semantic Web. The reason for the slow revolution of Semantic technology

is also the Ontology as only if we provide a well-formed ontology for each domain the machine can understand the concept behind them.

So, Ontology Search engine provide a way of how to search this already existing ontology for re-using purpose. In Semantic Web the Ontology of each domain is specified by Web Ontology Language famously called as OWL. Thus for this kind of categories we can build a search engine using the meta-data information of OWL or we can crawler the OWL files of each web page. But this kind of search engine is useful only if we have numerous amount of semantic web page which is not available.

The another category of search engine type is Semantic Search Engine where there are three type of engine like Context based search Engine, Evolutionary Search Engine and Semantic Associate Discovery Engine. In Context based search engine, is build on top of traditional search engine

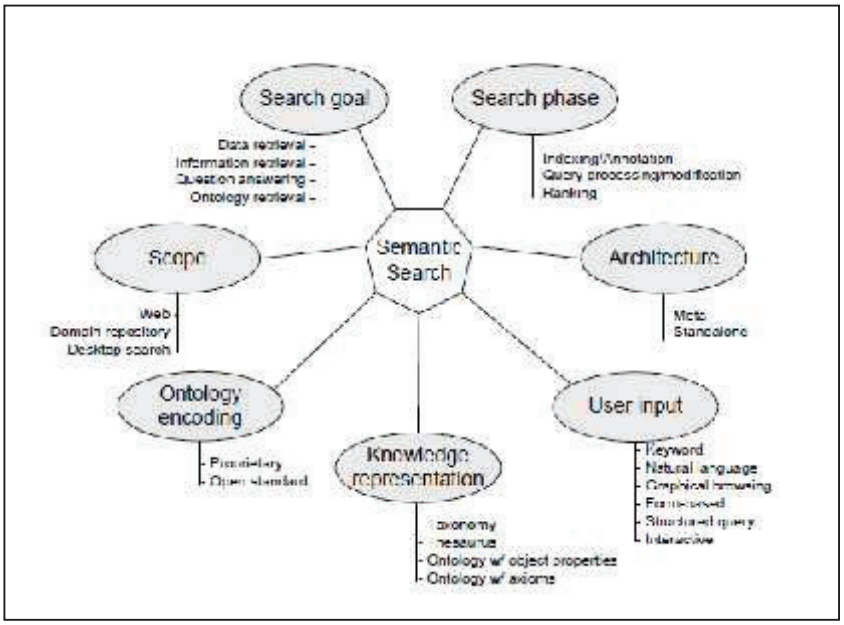

Fig 2. Concept of Semantic Search

As discussed There are many approaches to semantic search, e.g. some rely on semantic annotations while others enhance clustering of retrieved documents., we can also classified semantic search based on an analysis of reviewed literature and related classification schemes which is summarised in Figure 2. As can be seen from the figure, search applications can be categorised along seven dimensions. However, to be classified as a semantic search application, w.r.t. the definition previously presented, the system must utilise some form of structural knowledge that is used to improve the retrieval effectiveness (i.e. relevance and/or user experience). In the following, we elaborate on each of the different aspects of semantic search systems.

A categorization scheme for different types of search engines on the semantic web is summarized in Table I[25].

3. Proposed Semantic Web Search Engine Framework

Figure 3 shows an overall framework for this kind of engines. It should be emphasized that very limited number of engines have all of the functionalities specified in the figure.

Table 1: Semantic Web Search Engine summarization

Semantic Web Search Engines				
Ontology Search Engines		Semantic Search Engines		
Ontology Meta Search Engines	Crawler Based Ontology Search Engines	Context Based Search Engines	Evolutionary Search Engines	Semantic Association Discovery Engines
Here we want to find SWDs specially ontologies. There is two approach in using usual search engines, search only by the name of files and use some options like filetype (rdf,owl,rss,...) or search by labels by converting both documents and queries to intermediate forms that is not ignorable for ordinary search engines. In this type of search engine having a good display module for browsing and navigating the founded ontologies is critical point	Applications of these search engines are like to the previous category. But here we use a specific crawler for finding SWDs on the web, index them and extract some metadata about them. By using these engines we can search be special class or property and even for sample data (ABox). Graph structure of the SWDs on the web can be explored by use of these search engines. Also here visualizing the results is important. Preparing a standard test collection for these engines is a challenging problem.	The final purpose of these engines is enhancing performance of traditional search engines (especially Precision and Recall). It's possible through understanding the content of documents and queries. One of the most important part of this type is annotator which responsible for generating metadata for crawled pages. We need to generate some metadata for user's query in order to detect its context. Here usually after traditional retrieval we combine matching RDF graphs to obtain better quality of results. these engine are the most practical ones, in fact they are the next generation of current search engines. We can evaluate them using traditional performance measures and test collections	This search engine is an answer to a very famous well known problem: automatically gathering information on the one specific topic. The main distinguished behavior of this engine is using external metadata. They usually use an ordinary search engine and display augmented information near the original results. We think that in a large-scale mode like (i.e. in whole web) they will be very similar to a multi context based search engines	There is a specific application of the semantic web for the search capabilities. The goal is finding various semantic relations between input terms (usually two) and then rank the results based on semantic distances metrics. They work better in the context of Knowledge Bases. An upper ontology like WordNet or OpenCyc can be used for evaluating this kind of search engines
Swangler[46] OntoSearch[27]	Swoogle [27,28,29] Ontokhoj[30]	OWLIR[26], QuizRDF[31], ItWorks[32], Corese[33], Intotos[34], SHOE[35] Dose[36], SERSE[37], ALVIS[38], OntoWeb[39],Score[40]	W3C Semantic Search:[11],ABC[41]	SemDis[42]

1) Crawling the semantic web: There is not much difference between these crawlers and ordinary web crawlers and in fact many of the implemented systems uses an existing web crawler as underlying system. For example in [1] haircut is used as underlying system and also [15] uses one that understands special semantic tags. One of the important features of theses crawlers should be the exploration of ontologies that are referred from existing web pages.

*2) Metadata generation: T*here are different ways for metadata generation. For example [1] and [5] use external metadata. [1] Uses AeroText to extract names and expressions and then generates metadata in RDF format. One of the important problems in this regards.

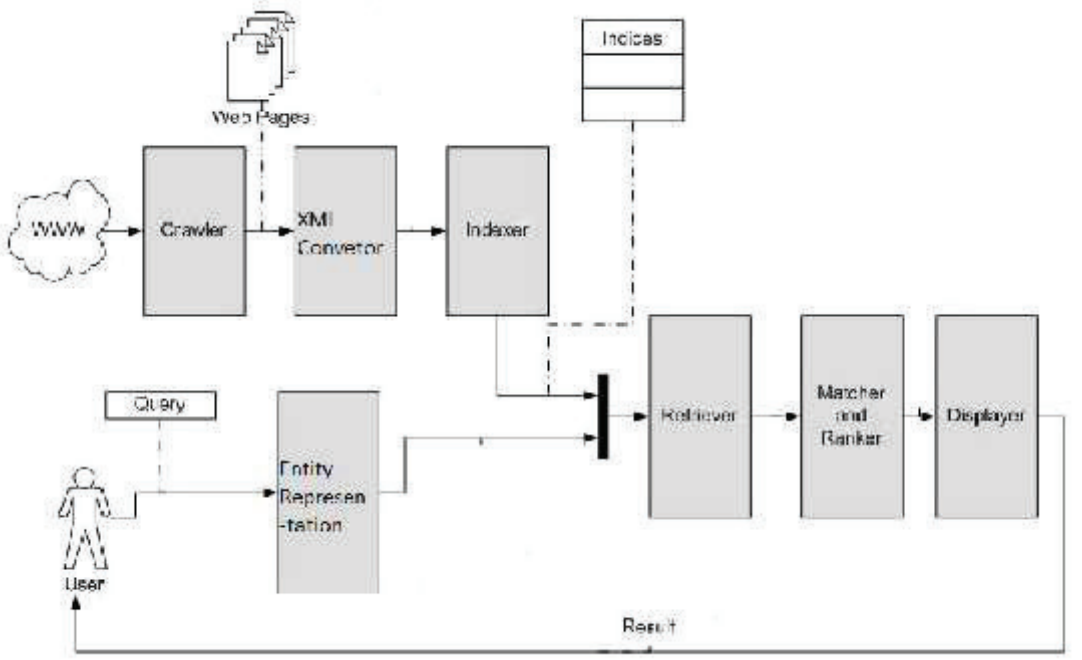

Fig. 3.Proposed Framework

is semantic normalization [25] meaning to generate metadata for different resources in same form. For example [12] is a non-standard example in which metadata is represented in ad hoc model. In semantic portal [23] producers should generate annotations and there is a summarization and collection of metadata in the central server. As explained before, metadata generation is simpler and more accurate when the theme of a page is known. For example in [15] using a tool named Knowledge Annotator terms of ontology is used to describe information in a given page. Also [18] proposed a method for generating and managing metadata according to already defined ontologies. But if ontology for a page is not known in advance, it is possible to use clustering techniques like what explained in [25] to find an appropriate ontology. Knowledge Parser [24] is a kind of complete system using important techniques from different areas like NLP, Text Engineering, Document Structure Processing, and Layout Processing.

3) Indexing: Most of the engines does not provide any special functionality regarding indexing. OWLIR [1] uses Swangling explained earlier. [6] Introduces Ontological Indexing in which indexing is done based on a reference ontology. Also in [18] possibility of dividing documents to smaller parts is used to improve indexing performance. Also in p2p architecture of [22] for each of concepts in the reference ontology there exist an agent that maintains information corresponding to it.

4) Accepting user's requests: There are two different approaches: term-based and form-based. In term-based approach used in [5], [23], and [24], it is tried to find the search context from entered keywords. In the form-based approach used in [1], [15], [23], and [24], user interface is generated according to the ontology selected by user.

5) Generating meta data for user requests: This operation is very similar to generating metadata for documents. For example in [18] the same Semantic Mapper is used for generating metadata both for documents and user requests. Often Wordnet is used to expand user requests. For example in [20] for termed entered by a user, using Wordnet, synonyms is found and used to expand the query.

6) Retrieval and ranking model: Usually an ordinary VSM model [30] is used and then based on RDF graph matching results are pruned. In [9] from the equivalence of RDF graphs and Conceptual Graphs (CG), already existing operations on CGs is used to match user request and documents. Semantic Distance concept is often used to estimate similarity of concepts in a matching process. In [21] this measure is defined for different elements in graphical representations. It is also possible to use graph similarity for ranking results. However, in [7] a fuzzy approach is used for this purpose.

7) Display of results: A major difference of semantic search engines and ordinary ones is the display of results. One of the primary tasks is to filter the results (for example for eliminating repetitions). In [6] in addition to normal display of results, a number of classes is displayed and when a user selects one, only those results having instances of the classes is shown. In [23] display is a kind of hierarchy in which top concepts of ontology is shown and by selecting one, detail of it according to the ontology is displayed.

4. Conclusion

If we have knowledge about what we are searching for, we can easy retrieve the desire information. The main drawback in information retrieval procedure in web technology is that the technology doesn't know the semantic and syntax of what the user searching for. This gives birth to the Semantic Web Technology. In this paper we have proposed an idea of retrieving the web information by providing semantic to it. Here we designed a Semantic retrieving system using the web Intelligent. In our future work we will integrate the Web Intelligent with Knowledge Intelligent.

References

[1] J. Mayfield, T. Finin, and B. County, "Information retrieval on the semantic web: Integrating inference and retrieval," in *SIGIR Workshop on the Semantic Web*, Toronto, Canada, 2004.

[2] T. Finin, J. Mayfield, C. Fink, A. Joshi, and R. S. Cost, "Information retrieval and the semantic web," in *Proceedings of the 38th International Conference on System Sciences*, Hawaii, United States of America, 2005.

[3] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. C. Doshi, and J. Sachs, "Swoogle: A search and metadata engine for the semantic web," in *Proceedings of the Thirteenth ACM Conference on Information and Knowledge management*, 2004.

[4] T. Finin, L. Ding, R. Pan, A. Joshi, P. Kolari, A. Java, and Y. Peng, "Swoogle: Searching for knowledge on the semantic web," in *Proceedings of the AAAI 05*, 2005.

[5] R. Guha, R. McCool, and E. Miller, "Semantic search," in *Proc. of the 12th international conference on World Wide Web*, New Orleans, 2003, pp. 700–709.

[6] J. Davies, R. Weeks, and U. Krohn, "Quizrdf: Search technology for the semantic web," in *WWW2002 Workshop on RDF and Semantic Web Applications*, 2002.

[7] T. Priebe, C. Schlaeger, and G. Pernul, "A search engine for RDF metadata," in *Proc. of the DEXA 2004 Workshop on Web Semantics*, 2004.

[8] Y. Zhang, W. Vasconcelos, and D. Sleeman, "OntoSearch: An ontology search engine," in The Twenty-fourth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cambridge, 2004.

[9] O. Corby, R. Dieng-Kuntz, and C. Faron-Zucker, "Querying the semantic web with the corese search engine," in *Proc. 15th ECAI/PAIS*, Valencia, Spain, 2004.

[10] C. Halaschek, B. Aleman-Meza, I. Arpinar, and A. Sheth, "Discovering and ranking semantic associations over a large RDF metabase," in *30th International Conference on Very Large Data Bases(VLDB)*, Toronto, Canada, 2004.

[11] H. Yu, T. Mine, and M. Amamiya, “An architecture for personal semantic web information retrieval system,” in *14th international conference on World Wide Web table of contents*, Chiba, Japan, 2005.

[12] [12] B. Sigrist and P. Schubert, “From full text search to semantic web: The Infofox project,” in *Proceedings of the Tenth Research Symposium on Emerging Electronic Markets*, 2003, pp. 11–22.

[13] L. Bangyong, T. Jie, and L. Juanzi, “Association search in semantic web: Search+ Inference,” in *International World Wide Web Conference*, 2005.

[14] B. Aleman-Meza, C. Halaschek-Wiener, I. B. Arpinar, and A. Sheth, “Context-aware semantic association ranking,” in *First International Workshop on Semantic Web and Databases*, Berlin, Germany, 2003, pp. 33–50.

[15] J. Heflin and J. Hendler, “Searching the web with SHOE,” in *AAAI-2000 Workshop on AI for Web Search*, 2000.

[16] C. Rocha, D. Schwabe, and M. de Aragao, “A hybrid approach for searching in the semantic web,” in *Proceedings of the 13th international conference on World Wide Web*, New York, NY, USA, 2004, pp. 374 – 383.

[17] W. Buntine, K. Valtonen, and M. P. Taylor, “The ALVIS document model for a semantic search engine,” in *2nd Annual European Semantic Web Conference*, Heraklion, Crete, 2005.

[18] D. Bonino, F. Corno, and L. Farinetti, “DOSE: a distributed open semantic elaboration platform,” in *ICTAI 2003, The 15th IEEE International Conference on Tools with Artificial Intelligence*, Sacramento, California, 2003.

[19] K. van der Sluijs, “Search the semantic web,” Master’s thesis, Department of Mathematics and Computer Science, Technical University of Eindhoven, 2004.

[20] J. Robin and F. Ramalho, “Can ontologies improve web search engine effectiveness before the advent of the semantic web?” in *SBBD 2003*, Manaus, Brazil, 2003, pp. 157–169.

[21] H. Zhu, J. Zhong, J. Li, and Y. Yu, “An approach for semantic search by matching RDF graphs,” in *In Proceedings of the Special Track on Semantic Web at the 15th International FLAIRS Conference (sponsored by AAAI)*, Florida, USA, 2002.

[22] V. Tamma, I. Blacoe, B. Smith, and M. Wooldridge, “SERSE: searching for semantic web content,” in *In Proceedings of the 16th European Conference on Artificial Intelligence, ECAI 2004*, Valencia, Spain, 2004.

[23] P. Spyns, D. Oberle, R. Volz, J. Zheng, M. Jarrar, Y. Sure, R. Studer, and R. Meersman, “OntoWeb - A semantic web community portal,” in *Proceedings of the 4th International Conference on Practical Aspects of Knowledge Management*, 2002, pp. 189 – 200.

[24] J. Contreras, V. R. Benjamins, M. Blzquez, S. Losada, R. Salla, J. Sevilla, D. Navarro, J. Casillas, A. Momp, D. Patn, L. Rodrigo, P. Tena, and I. Martos, “International Affairs Portal: A semantic web application,” in *ECAI Workshop on Application of Semantic Web Technologies to Web Communities*, 2004.

[25] Kyumars Sheykh Esmaili & Hassan Abolhassani "A Categorization scheme for semantic web search Engines",2005.

[26] T. Finin, J. Mayfield, C. Fink, A. Joshi, and R. S. Cost, “Information retrieval and the semantic web,” in *Proceedings of the 38th International Conference on System Sciences*, Hawaii, United States of America, 2005.

[27] Y. Zhang, W. Vasconcelos, and D. Sleeman, “OntoSearch: An ontology search engine,” in *The Twenty-fourth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, Cambridge, 2004.

[28] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. C. Doshi, and J. Sachs, “Swoogle: A search and metadata engine for the semantic web,” in *Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, 2004.

[29] T. Finin, L. Ding, R. Pan, A. Joshi, P. Kolari, A. Java, and Y. Peng, "Swoogle: Searching for knowledge on the semantic web," in *Proceedings of the AAAI 05*, 2005.
[30] C. Patel, K. Supekar, Y. Lee, and E. Park, "Ontokhoj: A semantic web portal for ontology searching, ranking and classification," in *Proceedings of ACM Fifth International Workshop on Web Information and Data Management (WIDM)*, New Orleans, 2003, pp. 58–61.
[31] J. Davies, R. Weeks, and U. Krohn, "Quizrdf: Search technology for the semantic web," in *WWW2002 Workshop on RDF and Semantic Web Applications*, 2002.
[32] T. Priebe, C. Schlaeger, and G. Pernul, "A search engine for RDF metadata," in *Proc. of the DEXA 2004 Workshop on Web Semantics*, 2004.
[33] O. Corby, R. Dieng-Kuntz, and C. Faron-Zucker, "Querying the semantic web with the corese search engine," in *Proc. 15th ECAI/PAIS*, Valencia, Spain, 2004.
[34] B. Sigrist and P. Schubert, "From full text search to semantic web: The Infofox project," in *Proceedings of the Tenth Research Symposium on Emerging Electronic Markets*, 2003, pp. 11–22.
[35] J. Heflin and J. Hendler, "Searching the web with SHOE," in *AAAI-2000 Workshop on AI for Web Search*, 2000.
[36] D. Bonino, F. Corno, and L. Farinetti, "DOSE: a distributed open semantic elaboration platform," in *ICTAI 2003, The 15th IEEE International Conference on Tools with Artificial Intelligence*, Sacramento, California, 2003.
[37] V. Tamma, I. Blacoe, B. Smith, and M. Wooldridge, "SERSE: searching for semantic web content," in *In Proceedings of the 16th European Conference on Artificial Intelligence, ECAI 2004*, Valencia, Spain, 2004.
[38] W. Buntine, K. Valtonen, and M. P. Taylor, "The ALVIS document model for a semantic search engine," in *2nd Annual European Semantic Web Conference*, Heraklion, Crete, 2005.
[39] P. Spyns, D. Oberle, R. Volz, J. Zheng, M. Jarrar, Y. Sure, R. Studer, and R. Meersman, "OntoWeb - A semantic web community portal," in *Proceedings of the 4th International Conference on Practical Aspects of Knowledge Management*, 2002, pp. 189 – 200.
[40] A. Sheth, C. Bertram, D. Avant, B. Hammond, K. Kochut, and Y. Warke, "Managing semantic content for the web," *IEEE Internet Computing*, vol. 6(4), pp. 80 – 87, Sep 2002.
[41] R. Guha, R. McCool, and E. Miller, "Semantic search," in *Proc. of the 12th international conference on World Wide Web*, New Orleans, 2003, pp. 700–709.
[42] C. Halaschek, B. Aleman-Meza, I. Arpinar, and A. Sheth, "Discovering and ranking semantic associations over a large RDF metabase," in *30th International Conference on Very Large Data Bases(VLDB)*, Toronto, Canada, 2004.