*International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)*
*Volume-2 Number-4 Issue-7 December-2012*

# Semantic Web Mining: Benefits, Challenges and Opportunities

**Syeda Farha Shazmeen[1], Etyala Ramyasree[2]**
Department of Information Technology, Balaji Institute of Technology and Science, Warangal, A.P, India[1,2]

## Abstract

***Semantic Web Mining aims at combining the two areas Semantic Web and Web Mining by using semantics to improve mining and using mining to create semantics. Web Mining aims at discovering insights about the meaning of Web resources and their usage In Semantic Web, the semantics information is presented by the relation with others and is recorded by RDF. RDF which is semantic web technology that can be utilized to build efficient and scalable systems for Cloud. The Semantic Web enriches the World Wide Web by machine process able information which supports the user in his tasks, and also helps the users to get the exact search result .In this paper; we discuss the interplay of the Semantic Web with Web Mining, list out the benefits. Challenges, opportunities of the Semantic web are discussed.***

## Keywords

***Semantic web, Web Mining, RDF (Resource Description Framework), Ontology.***

## 1. Introduction

As web is moving towards Web 2.0 (Semantic Web), it is shifting towards representing things as per their meaning (semantic representation). At the same time semantic web [1] is also an emerging area to augment human reasoning. Resource Description Framework (RDF) which is semantic web technology that can be utilized to build efficient and scalable systems for Cloud .Semantic web provides a framework for control of cloud usage. The unified RDF can replace the disparate cloud database models [6]. By using ontologies, semantic web aims at collecting structured information from web pages and redirect the information to client side to cater needs and preferences of the individual users [2]. Semantic web also reduces cost and complexity of cloud computing by the use of rules laid down in the issue of security, one of the major roadblocks in the success of cloud computing, is resolvable by a wide range of security mechanisms that the semantic web provides.

Web mining is now a reality and the challenge is to carry out semantic web mining. It is about machine understandable web pages to make the web more intelligent and able to provide services to the user. This means information on the web has to be mined so that the machine can understand the content [3].

The following steps show the direction where the Semantic Web is heading:

1. Providing a common syntax for machine understandable statements.
2. Establishing common vocabularies.
3. Agreeing on a logical language.
4. Using the language for exchanging proofs.

**1(a) Benefits of Semantic Web**

- Programs and sites can easily exchange information
- Search Engines can display more relevant information in results
- Data Mashers can combine data from different datasets to find new and outstanding things. Researchers can take large amount of data and try to make sense of it.
- Semantic web helps machines understand what the information on the web page is and the relationship between pieces of information.

**1(b) RDF (Resource Description Framework)**

RDF is used for organizing information. RDFa solves data linking problem**.**RDF consider everything as Resource (Named things). This is uniquely identified by the URI's[4,5]. Because we are using http URI's there resources don't need to be in same database, they can be fetched from web, the resources can be distributed across the web in federated Graph(dataset in different places over web) and when all resources are described this way the web becomes one giant database. **SPARQL** is a specific Query Language for RDF. RDF data is suitable for this analysis. Useful graph mining algorithms can be implemented in SPARQL and OWL.

The Resource can be a person, the properties of a person like name, email-id, DOB. And also relate the

resource to the other resource like the company employee the person how wrote this document. RDF type gives more information about the resources. A Curie is called to shorten the URI's like http://www.lincpersonal-design. The curie can be Lin: me.

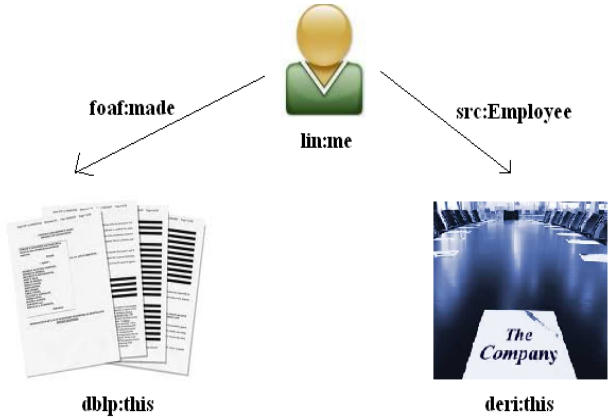


**Fig (1): describes how a person is related to other resources**

The below code shows that how RDF resources and its properties can be linked.

```
<p vocal="http://schema.org/" resource="#Lin" type of="Person">
  My name is
  <span property="name">linc George</span>
  and you can give me a ring via
  <span property="telephone"> 1-800-555-0199 </span>.
  <img property="image" src="http://.lincgeorge.org/images/linc.png" />
  My favorite sports is the <span property="ov: preferredsport">Cricket</span>.
</p>
```

If we assume that the markup above can be found at http://example.org/people, then the identifier for the thing is the address, plus the value in the resource attribute. Therefore, the identifier for the thing on the page would be: http://example.org/people#linc. This identifier is also useful if you want to talk about that same thing on another Web page. By identifying all things on the Web using a unique Uniform Resource Locator (URL), we can start building a Web of things. Companies building software for the Web can use this Web of things to answer complex questions like: "What is Linc George's phone number, what does he look like and what is his favorite sport?"

## 2. Semantic Web Architecture

The architecture of semantic web is illustrated in the figure below. A program collects the Web pages on the Internet with its semantic mark (RDF label) and corresponding ontology, which is described in an OWL Document in the Internet.

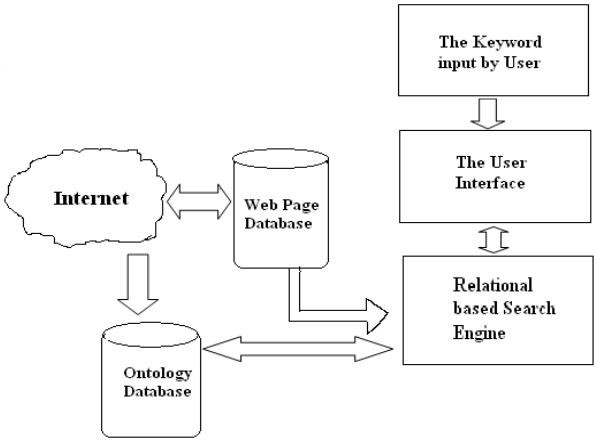


**Fig (2): Shows Semantic Web System Architecture**

When the users search for any keyword in the web, for the efficient and exact search result the Relational Based Search Engine is used. The collected Web pages are transported to a Web page database to be stored for the use of future retrieving URLs and corresponding Web pages. OWL maps the ontology into a relational database. The reason for mapping the ontology to relational database is relational database is mature and widely used [9].

The idea is to improvise the results of web mining by taking advantage of the new semantic structures on the Web; and also, making use of web mining, for building up the semantic web[7,8] by extracting similar meanings, useful patterns, structures, and semantic relations from existing web resources.

Figure 3 shows solution architecture for semantic web mining for knowledge Extraction. The architecture is Primarily divided into logical modules, namely knowledge extraction block and domain specific knowledge base. The User Query Interface is connected to Query Engine which maps all the Queries of the users to the different resources and tries to find the rules.

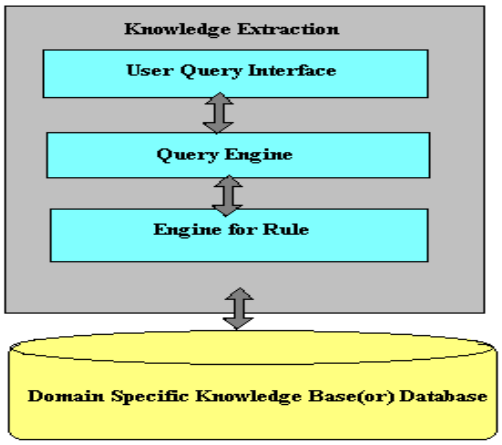


**Fig (3): Shows the Knowledge Extraction in Semantic Web Mining**

## 3. Opportunities in Semantic web

The Semantic web provides different searching opportunities for the different Search Engines [11]. For Instance, we have considered the Hospital Search Engine where the Patient wants to search for dental treatment process. The Search Engine offering drill down to the deepest level of search, from location of the Hospitals to the details of all the Doctors. Figure 3) shows the possible browsing experience. As the Patient searches for the list of Hospitals, he gets to choose between different locations. After selecting the Hospital location, the entire list of different Hospitals in the area is displayed, he can select and view the information related to his interest. By using the Relational Graph between all the resources the searching can be made easy and efficient.

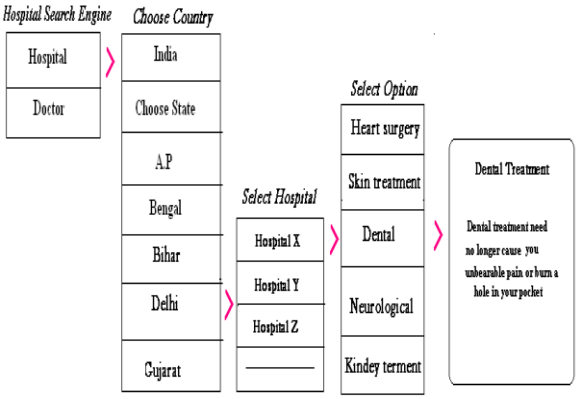


**Fig (4): Shows the Hospital Search Engine**

## 4. Challenges in Semantic web

"Web" has critically changed the perspective of how the organizations extract information from the available data in today's world of dynamic business. Therefore, the most important differentiator between a successful and an unsuccessful business is how an organization manages its data. The critical aspect in today's business scenario is how data is converted into Information and subsequently how information is converted into knowledge. It is very crucial to extract knowledge from un-structured data which is available in various formats and generated by heterogeneous sources across a big organization [10]. Most of the business information exists as unstructured data – commonly appearing in e-mails, blogs, discussion forums, wikis, official memos, news, user groups, chatting scripts on social networking sites, project reports, business proposals, public surveys, research and white papers.

Marketing material, official and business presentations and most of the web pages on WWW [13 & 14]. Usually humans can understand such text straightaway. However, with enormous quantity of such data content being available nowadays, both online and inside the enterprise, it becomes critical to mine such text using computers as it becomes very difficult and complex for a human being to mine huge data manually. So, Unstructured data is also important and taken in to consideration in Semantic web.

## 5. Conclusion

As we know Semantic Web Mining aims at combining the two areas Semantic Web and Web Mining [12], the RDF is a simple Meta model for defining and exchanging information on the semantic web. We have also discussed in the paper regarding the Architecture of the semantic web and also illustrated semantic web mining for a Hospital domain. A user-oriented semantic search engine is the need of today. These fields if explored in a right manner will provide unlimited opportunities to extract knowledge from the data available across the globe.

## References

[1] The semantic web: Roles of XML and RDF, Stefan Decker and Sergey Melnik, Frank Van Harmelen, Dieter Fensel, And Michel Klein Jeen

Broekstra Michael Erdmann Ian Horrocks, IEEE Internet Computing, October 2000, vol. 15, nr. 3, pgs. 63--74.

[2] K. Chang, B. He, Z. Zhang (2004, December). "Mining semantics for large scale integration on the web: evidences, insights, and challenges". ACM SIGKDD Explorations Newsletter, Volume 6, Issue 2. pp. 67-74.

[3] Gómez-Pérez, A. and O. Corcho, 2004. Ontology languages for the semantic web. IEEE Intelligent Systems, 17: 54-60.

[4] L. Ding, T. Finin, A. Joshi, Y. Peng, R. Pan, and P. Reddivari. Search on the semantic web. IEEE Computer, 10, 2005.

[5] L. Ding, R. Pan, T. Finin, A. Joshi, Y. Peng, and P. Kolari. Finding and ranking knowledge on the semantic web. In Proceedings of the 4th International Semantic Web Conference, pages 156–170, 2005.

[6] Hsinchun Chen, Xin Li, Michael Chau, Yi-Jen Ho and Chunju Tseng, "Using Open Web APIs in Teaching Web Mining," IEEE Transactions on Education, vol. 52 no. 4, pp. 482–490, November 2009.

[7] Yi Zhao, Xia Wang and Wolfgang A.Halang, "Security in Semantic Interoperation", IGI Global, 2009.

[8] R. Cooley. Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data. PhD thesis, University of Minnesota, Faculty of the Graduate School, 2000.

[9] Han and Kamber. Data Mining. Concepts and Techniques. Morgan Kaufmann, San Francisco, LA, 2001.

[10] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific Am., vol. 284, no. 5, pp. 34-43, 2001.

[11] Resource Description Framework (RDF) Schema Specification. (2000) In W3CRecommendation.

[12] G. Stummea, A. Hotho, B. Berendt. (2006). "Semantic Web Mining State of the art and future directions". Journal of Web Semantic. Web Semantics: Science, Services and Agents on the World Wide Web 4. pp. 124–143.

[13] Carroll, J.J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, and K.: Jena: implementing the semantic web recommendations. In: Feldman, S.I., Uretsky, M., Najork, M., Wills, C.E. (eds.) WWW (Alternate Track Papers & Posters). pp.74 {83. ACM (2004).

[14] Huang, H., Liu, C.: Query evaluation on probabilistic RDF databases. In: Vossen,G., Long, D.D.E., Yu, J.X. (eds.) WISE. Lecture Notes in Computer Science, vol. 5802, pp. 307{320. Springer (2009).

**Syeda Farha Shazmeen** received her Master's in Software Engineering from JNTU HYD in the year 2010. Now working as Asst.Professor in the Department of Information Technology, Balaji Institute of Technology and Science, Warangal, India. Published 3 papers in IEEE International Conference. Her research interest includes Data Mining, Databases, Cloud Computing, Semantic Web and Machine Learning.

**Etyala Ramyasree** received bachelor degree from Sri Kavitha Engineering College in the year 2011. Now working as Asst. Professor in the Department of Information Technology, Balaji Institute of Technology and Science, Warangal, India. Her research interest includes Cloud Computing and Semantic Web.