Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
IJMTER
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

Semantic Web-An Extensive Literature Review

Rashmi Bakshi[1], Abhishek Vijhani[2]
[1] *Assistant Professor, VSIT,VIPS*
[2] *MCA Student, JIMS*

Abstract - The purpose of this paper is to explore the field of semantic web by conducting an extensive literature review of the research done in the past. This paper attempts to understand architecture of the semantic web as well as its scope in other areas of research such as data mining, information retrieval, language processing and bio-informatics. A tabular format is used to list the research work done. Categorisation is done based on the key points discussed in the research paper, problems faced during the research and an outcome of the research article. This research paper divides the field of semantic web into sub fields-semantic architecture and security web, semantic mining and semantic context aware web

Keywords: Semantic web, Context aware, linked data, RDF, Ontology

I. INTRODUCTION

Semantic web was introduced by Tim Berner lee in 2000.Semantic web is an extension of the current web in which information is given well defined meaning, enabling computers and people to work in cooperation efficiently. It requires strong language and set of rules that can express the data and link the data, reason for the data in a most efficient way. It was discovered because web 2.0 was not machine understandable. World Wide Web is universal that means “anything can link to anything”. There must not be any discrimination among commercial and academic information or among cultures, languages and media and so on. Web has developed for people but not for automatic processing by machines. It was not context aware or adaptive. There was no formal semantics of the data. Contents were machine readable but not machine understandable. Web 2.0 is like a book with multiple hyperlinked documents. Index of the keywords is present but the contexts in which those keywords used are missing. To check which one is relevant we have to read the corresponding pages. In Semantic Web this limitation is eliminated via ontologies where data is given with well-defined meanings, which can be understandable by machines. Traditional search engines are unable to reach and give accurate precise results due to their lack of context aware. Semantic web on the other hand is adaptive and context aware. [16]

Semantic web technologies are entering in the industrial mainstream. Schema.org and the Face book Open Graph Protocol are bringing metadata to bear on the Web large-scale. IBM's Watson and Apple's Siri incorporate Semantic Technologies. Google is improving its search engine and is going more semantic in implementing their knowledge graph. These are just few of the examples. The main idea is that Semantic web is becoming popular and attracting industry as well as academicians.

II. ARCHITECTURE OF SEMANTIC WEB

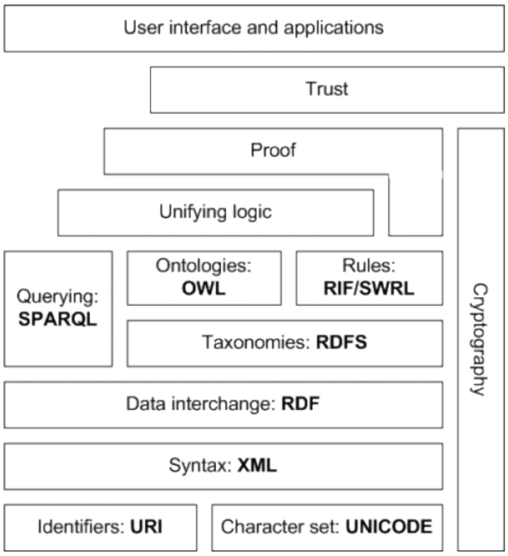

Figure2.Architecture of Semantic Web; Source-W3C

Semantic Web is an effort to make web intelligent by modifying the structure of data on the web so that it can be easily linked, processed and delivered to the users. Most of the data on the web is unstructured (text/images/HTML) or semi structured (XML). Semantic Web is an initiative to convert this data into "web of data" build upon RDF data model. It is built upon RDF schema which is based on the Web Ontology Language (OWL) where all RDF classes and properties are stored that describe entities and their relationships. It is a fairly new standard for data exchange on web. It can be referred as database of web. It uses URIs to link data i.e. to describe relationships between entities.

URI: A universal resource identifier is a formatted string that is used to identify an abstract or physical resource.RDF URIs identifies the real objects. They are different from HTTP URI that identifies HTML documents. It is just a way to distinguish data at human end (HTML) from data at machines.

Unicode: It provides a unique number for every character, independent of the underlying platform, program, or language. Previously ASCII and EBCDIC were used before the development of Unicode. The diverse encoding made the manipulation of data complex. A computer was asked to be compatible with different encodings. There could be risk of conflicts between different encoding schemes. Unicode solved this problem by being universal.

XML: XML (eXtensible markup language) with XML namespace and XML schema definitions makes sure that there is a common syntax used in the semantic Web.

RDf: On the top of XML, is the resource description framework for representing information about resources in a graph form. It is usually a triplet having subject-predicate-object expressions where subject represents the resource (thing), predicate means property or an attribute of the resource and object means the ultimate value.

RDF Schema (RDFS) defines the vocabulary of RDF model. It provides a mechanism to describe domain-specific properties and classes of resources to which those properties can be applied, using a set of basic modelling primitives (class, subclass-of, property, subproperty-of, domain, range, type). However, RDFS is rather simple and it still does not provide exact semantics of a domain.

Ontology: It comprises a set of knowledge terms, including the vocabulary, the semantic interconnections, simple rules of inference and logic for some particular topic. Ontologies applied to the Web are creating the semantic Web.

Logic, Proof and Trust: The logic layer is used to enhance the ontology language further and to allow the writing of application-specific declarative knowledge. The proof layer involves the actual deductive process as well as the representation of proofs in Web languages and proof validation. Finally, the Trust layer will emerge through the use of digital signatures and other kinds of knowledge, based on recommendations by trusted agents or on rating and certification agencies and consumer bodies.

III. SCOPE OF SEMANTIC WEB

Semantic web is slowly gaining power and collaborating with other areas of research like bioinformatics, eCommerce, eGovernment and social web. It's most significant use is seen in the field of Bioinformatics A fast developing trend in biomedical network analysis is about combining multiple biomedical associated data, which can be highly heterogeneous into coherent bio-molecular interaction networks to enable integrated network analysis. This is possible due to progress in semantic web.
Applications like genomic ontologies, semantic web services, automated catalogue alignment, ontology matching, blogs and social networks are constantly increasing, driven by companies like Google, Amazon, YouTube, Face book and LinkedIn. The need for combining information in a meaningful way creates the potential and demand for research in Semantic web.

IV. LITERATURE REVIEW

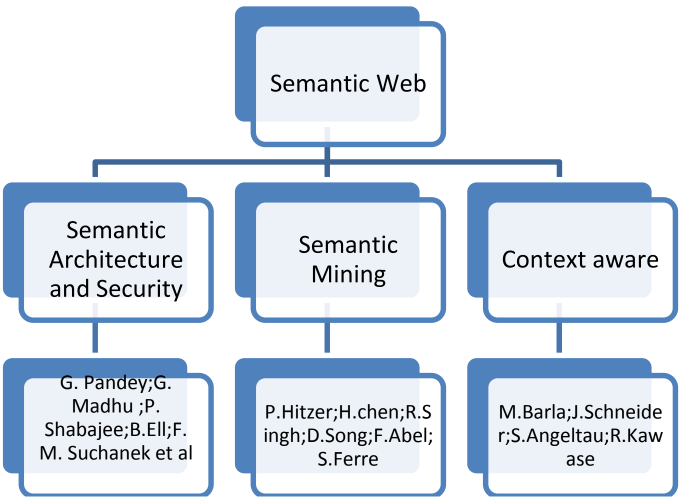

Figure 4.1 Classification of Semantic web

Research papers for Semantic web are classified into 3 categories.
Category I- comprises of those research papers that studies architecture of semantic web and security related research in semantic web.
Category II- comprises of those research papers where semantic ontologies are improved or modified to boost the overall system performance. It consists of information retrieval algorithms to increase the efficiency of semantic web search engines.

Category III- comprises of those research papers where semantic web is utilised as context aware, adaptive portals to develop usable and user friendly interfaces.
Table 4.1, discusses relevant research papers that contributed in the field of semantic web.

Table 4.1, Literature Review of previous work

S No.	Author's Name & Year of Publication	Title of the Paper	Key points discussed in the research paper	Problems addressed in the research paper	Conclusion
1	G. Pandey 2012 [6]	The semantic web –an introduction and issues	Description of the semantic web stack and working of its layers	Strong ontology is required to increase interoperability. Semantic web is loosely coupled, in a way it is an open system so standardization is required. Proof and trust is needed. Need for multi linguism and visualisation	Semantic web is better than traditional web in respect to structure of the data, linking of the data, simplification of the information and easy to search and access data.
2	G. Madhu et al. 2011 [5]	Intelligent Semantic Web Search Engines: A Brief Survey	It discusses the role of search engines in semantic web through preliminary survey	Problems faced by semantic web search engines are low precision and high recall. Inaccurate queries Identity intention of the user. Individual user patterns can be extrapolated to global users.	Brief survey of existing literature regarding intelligent semantic search technologies is done based on designer's and user's perception, low precision, high recall and lack of experiment tests.
3	P. Hitzler et al. 2010 [10]	A Reasonable Semantic Web	This paper argues that semantic web is shared inference. Machine learning methods along with deductive ones are needed to be explored in terms of precision and recall value rather than just soundness and completeness of the system	Inference systems cannot scale up to the size of web. Input data from the open semantic web cannot be clean enough for classical reasoning systems to draw useful inferences from them.	Incomplete and inconsistent systems are useful as long as they cater and provide usability to subset of a data if it can be qualitatively accessed.
4	P. Shabajee	Informed	This paper	It is difficult to	Policy aware web

	2006 [11]	Consent on the Semantic Web – Issues for Interaction and Interface Designers	discusses the differences between traditional and semantic web based on the privacy and ethical issues	secure the data when it is integrated from multiple sources. Informed consent is one way of ensuring the trust of the user	including interaction design is required to ensure the security.
5	H. Chen et al. 2009 [7]	Semantic web for integrated network analysis in biomedicine	Semantic web technology is explored to represent, integrate and analyse the knowledge in various biomedical networks. Semantic graph mining framework is introduced for network data analysis.	Semantic Web brings up additional complexities in decentralized mining, automated data process flow, heterogeneous graph fusion and data provenance management. It is difficult to achieve semantic consistency, graph transformation, scalability and performance	RDF, RDFS and OWL not only capture the topology of complex biomedical networks, but also preserve the semantics of these networks by annotating nodes and edges with the vocabulary defined in corresponding Semantic Web ontologies.
6	M. Barla et al. 2007 [9]	Adaptive portal framework for Semantic Web applications	A framework is designed for the creation of adaptive portal solutions for the semantic web		Framework is used for Reusability, component-based design, personalization and interoperability taking advantage of ontologies, adaptive navigation and presentation.
7	R.Singh et al. 2013 [13]	Semantic web and application in ERP	Semantic web mining has been proposed for an ERP application based on educational domain	No problems mentioned	Precision and recall values are calculated and compared with traditional queries and semantic web is used as an improved method for web retrieval
8	J. Schneider et al. 2013 [8]	A Review of Argumentation for the Social Semantic Web	It investigates ontologies and tools which may be useful for argumentation on the Social Semantic Web	Great research is needed in creating ontologies, easing human annotation, improving techniques in detecting and mining argumentation, marking the	Ontologies are studied to represent informal social web arguments. Context is investigated and users are motivated to reduce human effort.

				context and history of dialogues	
9	S. Angeletou et al. 2011 [14]	Modelling and Analysis of User Behaviour in Online Communities	Method to infer user roles in online communities: Ontology to model behavioural features and support community role inference: Analysis of community health through role composition	Dynamic nature of communities results into many unclassified users online. Churn-loss of community members is an identified risk. Different behavioural patterns are possible for analysing different communities	This paper presented ontology to capture the behavioural characteristics of users as numeric attributes and explained how semantic rules can be employed to infer the role that a given user has.
10	D. Song et al. 2011 [2]	Automatically Generating Data Linkages Using a Domain-Independent Candidate Selection Approach	Candidate selection algorithm is designed for rdf data i.e. finding non-overlapping blocks of instances such that all instances in a block will be compared to each other or locating similar instance pairs.	Limitation of this research is that the algorithm targets the datasets that are composed of strings and same measure is adopted for numerical value. Fuzzy matching is not implemented and therefore exact match is difficult to retrieve from incorrect datasets having errors and misspellings	This paper presented an index based domain-independent candidate selection algorithm for scalable detecting *owl:sameAs* links. Predicates are identified for candidate selection through unsupervised learning. By indexing the instances on the learned predicates, object values, the algorithm is able to efficiently look up similar instances.
11	B. Ell et al. 2011 [1]	Labels in the Web of Data	Human readable labels are produced for the web of data.	Quality and usefulness of labels need to be improved depending upon the application that will use the knowledge	This research paper defined metrics to access the completeness, efficient accessibility, un ambiguity and multi linguality.
12	F. Abel et al. 2011 [3]	Leveraging the Semantics of Tweets for Adaptive Faceted Search on Twitter	An adaptive personalised faceted search framework for twitter messages is proposed	No problems faced during the research were mentioned in the research paper	A framework is developed that added semantics to twitter messages extracting entities and enriching them with external resources in order to create facets (e.g. persons, locations,

					organizations etc.) and facet-values that describe the content of tweets.
13	F. M. Suchanek et al. 2011 [4]	Watermarking for Ontologies	Ownership of ontologies is introduced in this paper	Artificial facts are not considered in this research	Watermarking the ontologies by removing the facts instead of altering it thereby precision value does not lower. It is done to protect the data from theft
14	R. Kawase et al. 2011 [12]	Generating Resource Profiles by Exploiting the Context of Social Annotations	Contextual information related to tag, user and its resource along with tag metadata is used to generate user profile	-	In this paper, authors proposed novel approaches to generate and enrich resource profiles that exploit the multiple types of contextual information available in most social tagging systems.
15	S. Ferre et al. 2011 [15]	Semantic Search: Reconciling Expressive Querying and Exploratory Search	In this paper, the query language, named LISQL, generalizes existing semantic faceted search systems, and covers most features of SPARQL. It introduces Query-based Faceted Search (QFS), the combination of an expressive query language and faceted search, to reconcile the two paradigms.	-	Query based faceted search (QFS) is introduced for RDF datasets. It combines the expressiveness of SPARQL and the benefits of exploratory and faceted search.

V. CONCLUSION

In this research paper, we tried to study about semantic web and classify the past research into 3 categories-semantic architecture and security, semantic mining and semantic context aware. In future, we would like to expand our study and include more referred journals.

REFERENCES

1. B.Ell, D.Vrandecic and E.Simperl. Labels in the Web of Data, KIT, Karlsruhe , *ISWC*-2011
2. D.Song and J.Hefli. Automatically Generating Data Linkages Using a Domain-Independent Candidate Selection Approach.ISWC-2011
3. F.Abel, I.Celik, G.J.Houben,P. Siehndel. Leveraging the Semantics of Tweets for Adaptive Faceted Search on Twitter. ISWC-2011
4. F.M. Suchanek, D.G.Amblard and S.Abiteboul. Watermarking for Ontologies. ISWC-2011

5. G.Madhu, Dr.A.Govardhan and Dr.T.V.Rajinikanth. Intelligent Semantic Web Search Engines: A Brief Survey. *International journal of Web & Semantic Technology*, Vol.2, No.1, 2011
6. G.Pandey.The Semantic Web: An Introduction and Issues**.** *International Journal of Engineering Research and Applications,* Vol 2 Issue 1, 2012
7. H. Chen, L. Ding, Z. Wu, T. Yu, L. Dhanapalan,J.Y. Chen.Semantic web for integrated network analysis in biomedicine. *Briefings in Bioinformatics*, Oxford University Press, 2009
8. J.Schneider, T.Groza and A.Passant. A Review of Argumentation for the Social Semantic Web.*The open university*, IOS Press,2013
9. M.Barla, P.Bartalos, M. Bielikova,R Filkorn and M.Tvarozek. Adaptive portal framework for Semantic Web applications. Adaptation and Evolution in Web Systems Engineering,2007
10. P. Hitzler, F.V. Harmelen.The reasonable web.Semantic web,IOS Press, 2010
11. P. Shabajee. Informed Consent on the Semantic Web - Issues for Interaction and Interface Designers.ISWC-2006
12. R.Kawase, G.Papadakis,F.Abel. Generating Resource Profiles by Exploiting the Context of Social Annotations. ISWC-2011
13. R.Singh, S. Sharma. Semantic web and application in ERP. *International Journal of Emerging Trends and Technology in Computer Science*, Vol 2, Issue 1,2013
14. S. Angeletou, M. Rowe, H. Alani. Modelling and Analysis of User Behaviour in Online Communities. Knowledge Media institute, ISWC-2011
15. S. Ferre and A.Hermann. Semantic Search: Reconciling Expressive Querying and Exploratory Search.IRISA.ISWC-2011
16. T.B.Lee,J.Hendler and O. Lassila. The Semantic Web. Scientific American,2001

