Chapter
12

Semantic Web: Basics & Application: An Overview

Abhijit Bandyopadhyay

Teacher-in-Charge, Raniganj Institute of Computer and Information Science (RICIS), Raniganj, West Bangal, India

Corresponding author: abhijit.ricis@gmail.com

ABSTRACT

Exponential rise of information ramification in the World Wide Web in the recent time calls for proper validation of the concept of Semantic Web. Keyword matching has become a very significant factor in semantic web domain. Traditional algorithms prove to be insufficient and there is a growing need for standard statistical techniques to be properly customized on a case-to-case basis. Semantic Web is used to make the machines to understand the data. The concept of Semantic Web that is an extended version of the existing Web. This research chapter has pointed out certain application areas of Semantic Web, data integration, resource discovery and classification, cataloguing, intelligent software agents, content rating and logical collections of pages for describing intellectual property rights. Relevance of web pages based on the keyword-based web search has been a major area under consideration in this research chapter. Future scope of research has been found to be extremely promising and this chapter aptly addresses that issue giving an intuitive direction for the researchers interested in semantic web domain.

Keywords: Semantic Web, RDF, OWL, Artificial Intelligence, SPARQL, Knowledge Management

INTRODUCTION

Semantic Web is used to make the machines to understand the data. It also makes the computers understand the data so that they can perform efficiently on the data on our behalf [3]. The concept of Semantic Web that is an extended version of the existing Web. It is helpful in standardization of expressing the relationships between computers and information. The primary technical standards for Semantic Web are RDF (*Resource Description Framework*), OWL (*Ontology Web Language*) and SPARQL (*SPARQL Protocol and RDF Query Language*)[8].

In this chapter under review sincere attempt has been to make to define each of them as follows:

- RDF (Resource Description Framework): The data modelling language for the Semantic Web. Semantic Web information is stored and represented in RDF[3]. *RDFS;* It provides data-modelling vocabulary for RDF data[8].
- SPARQL (SPARQL Protocol and RDF Query Language): The query language of the Semantic Web. It is specifically designed to query data across various systems[3].
- OWL (Web Ontology Language): Schema language, or knowledge representation (KR) language, of the Semantic Web. OWL enables to define concepts in composable manner so that these concepts can be reused as much and as often as possible. *Composability* means that each concept is carefully defined so that it can be selected and assembled in combinations with other concepts as needed for many different applications and purposes[3].

The chapter also identified some other standards for Semantic Web like:

- *Uniform Resource Identifier:* Any kind of object or concept is identified by *URI*[19]. Uniform Resource Identifier (URI) is used for resource definition[9].

- *Extensible Mark-up Language (XML)*: It is a simple, flexible text format derived from Standard Generalized Mark-up Language and designed to overcome the challenges of large-scale electronic publishing[2]. XML is used for syntactic representation[9].

It is true that one way to differentiate a Semantic Web application from any other application is the usage of those three technologies. It can be mentioned that the Semantic Web has been called many things, such as Web 3.0 or the Linked Data Web. As some of these names carry great significance, even with regard to the technology stack, the conceptualization of Sematic Web with all its variations and validations is found to have immense importance for the growth of this technology domain.

Review of Literature on Recent Trends in Semantic Web Domain

Latest trends on semantic web can be discussed chronologically. In 2013, roughly semantic web chapters focused on ontology are about 59 percent, search engine and text retrieval are about 12 percent, and rest of the chapters are classified as multimedia retrieval, evaluation of search engine and concepts. Ontology is one of the main topics of semantic and many researches continue to work on in this topic. The search engine is still one of the topics that draw attention by the researchers. Studies on semantic web started with text retrieval and future studies seem to concentrate on multimedia retrieval. For both text and multimedia, the evaluation is key topic to assess the results[11]. In search engines, the authors explained 'Answer Engines' as the Knowledge Graph (or any entity graph) is based on internal, verified and validated structured data that facilitate the existence of answer engines; 'Machine Readable' as The Semantic Web, in the form of structured markup embedded in HTML pages, provides machine-readable information to search engine

crawlers on specific topics/subjects; 'Enhanced SERP Displays & Lift' as Search Engines exploit semantic technology to create a better user experience and the carousel and knowledge graph results are also examples of enhanced displays which create a better user experience; 'Validation Of Web Pages' as extracting pertinent structured data, search engines can verify that our pages are, in fact, about the topic we are describing on those pages and this verification adds trust to our webpages; 'Social Network Adoption' as Search engines like Google can leverage semantic search to drive adoption of their Google Plus social network for businesses; 'Google+ Authorship Rich Snippet' as Google+ social network adoption is further driven for personal use by individuals due to the authorship rich snippet; 'Internal Structured Data' as Internal structured data (verified and validated via trusted sources) can be leveraged for many things, such as prediction or recommendation, by remembering the law of all data when utilizing it for results: GIGO (Garbage In, Garbage Out); 'The Future Of Search' as Internal axiomatic facts can allow computers (and search engines) to derive new information and can be leveraged along with viable reasoning mechanisms and at a minimum, search engines like Google are already deriving basic associations by traversing a Semantic network or Knowledge Graph; 'Schema.org Ontology' as search engines leverage some thing like a Schema.org ontology, those concepts are essentially language independent; and at last, they explained 'Understanding User Intent' as Search engines use Semantic Search to better understand user intent which is possibly one of the most important aspects to understand.

Conceptualizing Semantic Web

An analysis to the process of research in Semantic Web by deliberate vivisection of the concept and application of Semantic Web reveals that the Semantic Web is a Web of data which are very different from bank statement data as the latter is controlled by applications, and each application

keeps it to itself[3]. Hence, the vision of the Semantic Web is to extend principles of the Web from documents to data using the general Web architecture using, e.g., URI-s, which means creation of a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, to be processed automatically by tools as well as manually, including revealing possible new relationships among pieces of data[3]. It can be rightly pointed out that certain application areas of Semantic Web, like, data integration, resource discovery and classification, cataloging, intelligent software agents, content rating and logical collections of pages for describing intellectual property rights demand special attention[3]. While explaining rules on the Semantic Web, the logic programming defined by RDF to discover new, implicit relationships is of utmost importance[3]. The future use of Web can be seen in the following schematic figure.

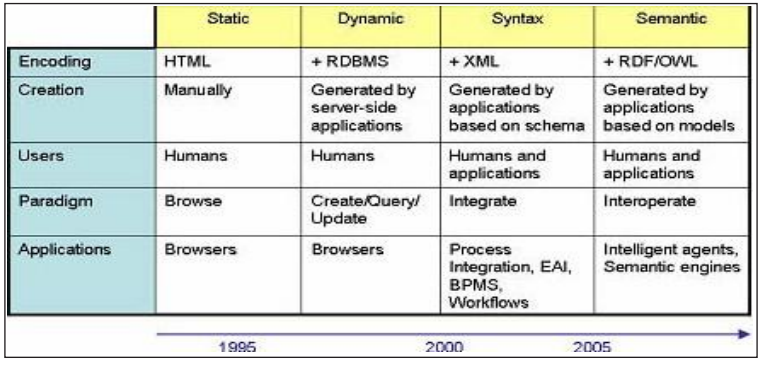

Fig. 1: Use of Web today and in Future[9].

But it is also to be remembered that Semantic Web has come a long way since 2005, as progress seen in the above figure, just to mention some present-day applications of Semantic Web includes

1. Predict certain events in order to better manage operations.

2. Easy and timely access to data to provide the right information at right time, and
3. Combine heterogeneous data to support more effective decision-making processes.

More recent works on Semantic Web reveals the Semantic Web Architecture which clearly explains this modern concept.

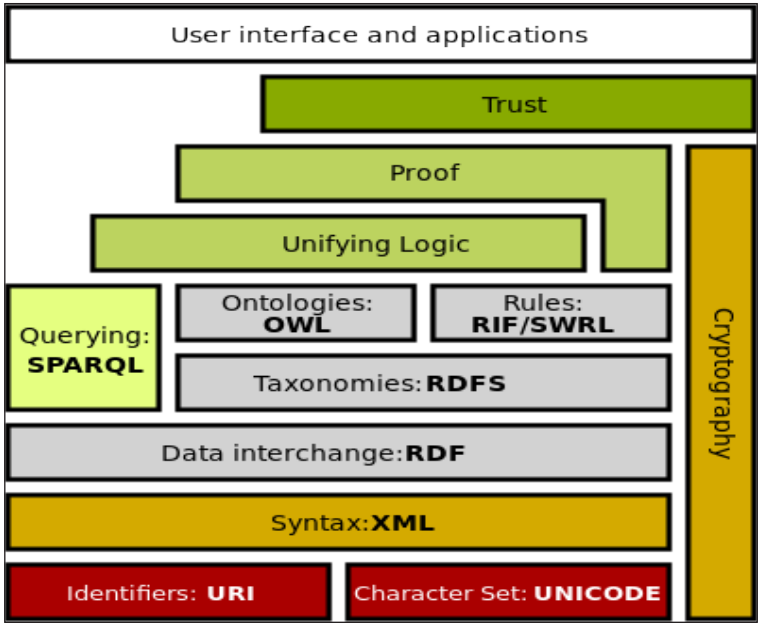

Fig. 2: Semantic Web Layered Architecture[10]

It can be pointed out the definition of URI (Universal Resource Identifier) as a formatted string serves as a means of identifying abstract or physical resource[9]. It is a crucial point that Semantic Web needs unique identification to allow provable manipulation with resources in the top layers[10]. It has also been noted that Unicode provides a unique number for every character, independently of the underlying platform, program, or language[9]. Semantic Web should also help to bridge documents in different human languages, so it should

be able to represent them[10]. That XML (eXtensible markup language) with XML name space and XML schema definitions makes sure that there is a common syntax used in the semantic Web[9], proves immensely for proliferation of Semantic Web Domain. Most notably, it has been identified that XML is the Resource Description Framework (RDF), for representing information about resources in a graph form. RDF is based on triples O-A-V that form a graph data with a relation among object (a resource), an attribute (a property), and a value (a resource)[9]. This is extremely noteworthy that great amount of ground work on Semantic Web has been done by different researchers in this field to prove that RDF Schema (RDFS) precisely defines the vocabulary of RDF model[9]. Another term, Ontology can also be mentioned that it comprises a set of knowledge terms, including the vocabulary, the semantic interconnections, simple rules of inference and logic for some particular topic and Ontologies applied to the Web are creating the semantic Web[9]. It can be also explained that SPARQL as a RDF query language can be used to query any RDF-based data. Querying language is necessary to retrieve information for semantic web applications[10]. The RIF can be defined as a rule interchange format that cannot be directly described using description logic used in OWL[10]. It should be noted that Cryptography can be achieved by appropriate digital signature of RDF statements[10].

While mentioning Logic, Proof and Trust, it can be deftly explained that the logic layer is used to enhance the ontology language further and to allow the writing of application-specific declarative knowledge[9] while the proof layer involves the actual deductive process as well as the representation of proofs in Web languages and proof validation[9] and 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[9]. User interface, as can be clarified, is the final layer that will enable humans to use semantic web applications[10].

Applications of Semantic Web

We can aptly delineate the applications of Semantic Web as the following.

1. Knowledge Management: Knowledge is power. It used to be conventional wisdom that the organization with the most information wins. The Semantic Web can bring structure to information chaos[13].

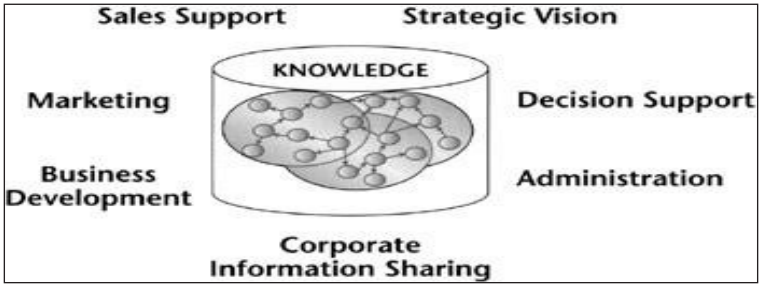

Fig. 3: Uses of the Semantic Web in your enterprise[13]

2. Decision Support: Using Semantic Web technologies will help in creating semantic-enabled decision support systems (DSSs) that focus on software agent analysis and interaction between the end user and computer system for decision making[13].
3. Business Development: Semantic Web. E-commerce industry experts believe that the Semantic Web can be used in matchmaking for e-business, by which businesses are put in contact with potential business partners or customers to create a smarter CRM solution[13].
4. Information Sharing and Knowledge Discovery: Using Semantic Web-enabled Web services can allow us to create a registry corporate knowledge base[13].
5. Administration and Automation: Semantic Web-enabled

Web services have the potential to automate menial and complex tasks in your organization[13].

It must be admitted that insights the world of Semantic Web and its application in web search will go a long way in making this technology impeccable for the users of his domain.

DISCUSSION

While existence of efficient algorithm for Web Search and Sufficiency of standard statistical techniques are under systematic consideration, it has been found to address the issue of insufficient algorithm and standard statistical techniques. It has been tried to relate Semantic Web and Artificial Intelligence and establish by referring to Takahira Yamaguchi's Group (doing Research on Semantic Web) that in semantic web, Software canbe made to understand the meaning of web-pages by adding headings called "metadata" to web-pages. Here, the emphasis should be pursued not in the Internet field, but independently of the internet, in the field of artificial intelligence. Underdevelopment is the system where the answer can be calculated using data and Ontologies to give reply automatically. In addition, the combination semantic web and robots is the order of the day. Hence, we put effort to relate Semantic Web and Social Network and opine that the Semantic Web would function with highly formal ontologies with minimal ambiguity and thus a minimal need for human interpretation, we now encounter the limitations of increasing the formality of knowledge in the global, dynamic environment of the Web.[14]. But we must admit that a design of efficient algorithm has neither been displayed, nor the insufficiency of statistical techniques have been eradicated by suggesting any superior and effective method. Hence,we suggest incorporation of Singhad Technical Education Society's researcher Grishma Y Bobhate and Prof Usha A Jogalekar's study in this area, namely "An Efficient Algorithm to Reduce the Semantic Gap between

Image Contents and Tags" where it has been suggested that the following algorithm may the Semantic Gap[8] and an efficient algorithm can also be suggested as below:

1. Give a query and obtain n feature vectors for all nodes in the database.
2. Perform a first t-steps random walk and get a new vector value.
3. Rank the relevant list of the nodes in the highest order.
4. Make a new feature vector value according to the entries with highest rank=1 otherwise 0.
5. Again perform a new t-steps random walk and get the retrieve results.

This sort of improved algorithm has been suggested by this work authors to make the work more suggestive for future researchers and improve credibility.

LIMITATIONS

The present research works are having certain limitations as under:

1. It comprises of those research chapters that studies architecture of semantic web and security related research in semantic web, which may not always be the sole case.
2. It comprises of those research chapters where semantic ontologies are improved or modified to boost the overall system performance while it does not consist information retrieval algorithms to increase the efficiency of semantic web search-engines.
3. It comprises of those research chapters where semantic web is utilized as context aware, adaptive portals to develop usable and user-friendly interfaces but a more detailed work needs to be done for developing the course of action of AI enabled interface.

FUTURE SCOPE

It has been tried to visualize the scope of future research in a myriad of ways. It has been opined that Semantic web is slowly gaining power and collaborating with other areas of research like bioinformatics, e-Commerce, e-Government and social web, and the most significant use being 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. 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[15]. While getting engaged in the field of Bioinformatics we completely agree with the view of the application of Semantic Webin Bioinformatics and e-healthcare. The reference to Semantic and Web Intersection (The Pioneers Approach), where the pioneers envisioned the use of Semantic Web where, inferencing is a major aspect, hence, inferencing and reasoning would essentially lead to knowledge management. Not only that, the it can also be found to be extremely creditable on the part of the distinguished researchers engaged in this field to augment the Semantic Web applications on a regular basis. Here we can refer James Handler and Sir Tim Berners-Lee's argument that once we have enough semantic data everyone would want to become a Semantic Web user including governments and the open data project which more and more governments are joining is a clear indication towards this phenomenon. Even James Handler talks about the application of Semantic Web and how pervasive it has become. He provides enough examples to prove that knowingly or unknowingly everyone uses the fruits of Semantic Web. Facebook's open graph

protocol, Oracle's Semantic Web extensions, Google's search result etc. are all ways in which everyone is already a user of Semantic Web. Hence, overall the review of literature, recent trends in Semantic Web and scope of further research can aptly be explained by continuous research and investigation which demands special credit and merit for review and investigation including the present one, putting systematic and continuous effort in this domain.

CONCLUSION

The rationale behind the present review chapter has been on three counts, namely, whether the proper conceptualization of Semantic Web and its main goal and its primary usage have been explained so far, whether the existence of efficient algorithm for web search and sufficiency of standard statistical techniques have been suggested in the domain and lastly whether the review of literature on Semantic Web, recent trends in Semantic Web Domain and scope of Semantic Web have been aptly delineated of late. The review chapter found that different authors have done commendable work in almost all three counts, though there is always room for improvement and suggestions have also been given towards that. The rate at which data gets generated is increasing day by day with semantics in it and increasing amount of semantically annotated information on the Web, a lot more structured data is becoming available. Hence this includes information from scientists and governments publishing data on the Web and the ever-increasing amount of information available about each of us, individually and as societies-in the form of our social interactions, location and health data, activities, and interests. Working with this data, and understanding its diverse and often contradicting nature, to provide really meaningful services and to improve the quality of our lives, is something that researchers in both industry and academia are beginning to tackle. Statistical and machine-learning methods become

more powerful and computational resources continue to improve. So, some of the semantic knowledge that researchers had to construct manually they can now learn automatically, tremendously increasing the scale of the use of semantics in understanding and processing Web data. Similarly, our very understanding of the nature of the semantics that intelligent systems produce and leverage is changing, and with it, our vision for the future of the Semantic Web. It is sincerely expected that future researchers and academicians will take up this issue more and more at their individual capacity and a better world will arise for usage of Semantic Web. A technically progressive and ethically sound society can always reap the benefits of advances of Semantic Web.

ACKNOWLEDGEMENTS

This work is a part of CISMPH-1204**,** Ph.D.-CIS Program of Raiganj University, WestBengal, India, and I am thankful to Dr. A. Sarkar, Department of CIS, Raiganj University, West Bengal, India. Special thanks to Dr. P. K. Paul, Department of CIS, Raiganj University, West Bengal, India, for his encouragement and support to publish this work in a publication format.

REFERENCES

1. Berners-Lee, Tim., S.T. 2006. Linked Data. *The World Wide Web Consortium*.W3C.

2. Bobhate, Grishma Y. and Jogalekar Usha, A. 2013. An Efficient Algorithm to Reduce the Semantic Gap between Image Contents and Tags. *International Journal of Computer Applications*, **72**: 38-40.

3. Cambridge Semantics. 2022. Introduction to the Semantic Web. [online] Available at: <https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web> [Accessed 9 March 2022].

4. Cordoso, Jorge. 2007. The Syntax and the Semantic Web. Semantic Web Services: Theory, Tools and Applications. *Idea Group*, pp. 1-23.

5. Cordoso, J. 2007. The Semantic Web Vision: Where Are We? *IEEE Intelligent Systems*, **22**: 84-88.

6. Hendler, J. 201.. "Why the Semantic Web will Never Work". [Performance]. Keynote, European Semantic Web Conference, Heraklion, June, 2011.

7. Hendler, J. 2014 "Semantic Web: The Insider Story", [Online]. Available: <http://www.slideshare.net/jahendler/semantic-web-the-inside-story.> [Accessed 9 March 2022].

8. Kariera Future Processing. 2022. "Introduction to Semantic Web"[online] Available at: <http://www.future-processing.pl/blog/introduction-to-semantic-web/> [Accessed 9 March 2022].

9. Bernstein, M. (2009). "MIT CSAIL Research," [Online]. Available at: < http://haystack.csail.mit.edu/blog/2009/10/25/tales-of-a-semantic-web skeptic/>. [Accessed 9 March 2022].

10. Mironov, V., Seethappan, N., Blonde, W., Antezana, E., Lindi, B. and Kuiper, M. Benchmarking triple stores with biological data. Semantic Web Applications and Tools for Life Sciences (SWAT4LS), Berlin, Germany.

11. Menemencioglu, O. and Orak, I.M. 2014. A Review on Semantic Web and Recent Trends in Its Applications, pp. 297-303.

12. Paul, P.K., Govindarajan, S., Chaterjee, D. and Bhatnagar, R. 2012. Web Technology: From Web 0.5 to Web 4.0 Emphasizing Semantic Web in The Sci-Tech. *Journal of Science & Technology,* **1**: 162-167.

13. Paul, P.K. 2013. Web 2.0: An Account of Its Existing, Emerging and Possible Dimension- the Modern way of Communication. International Journal of Computer Sciences and Engineering Systems, ISSN-0973-4406,71-76.

14. Paul, P.K. 2013. Web 2.0 and Information Centre and Services. *Abhinav National Journal of Science and Technology,* **2**: 28-33.

15. Paul, P.K. 2013. Sports 2.0: The Emerging Dimension of Web 2.0: The Sophisticated Electronic Information and Communication. *International Scientific Journal of Sport Sciences,* **2**: 69-73.

16. Paul, P.K. and Aithal, P.S. 2019. Web Security: An Overview and Current Trend, *World Academics Journal of Engineering Sciences,* **6**: 52-54.

17. Popitsch, Niko and Haushofer, B. 2011. DS Notify. A Solution for Event Detection and Link Maintenance in Dynamic Datasets. Web Semantics. *Science,* pp. 266-83.

18. Salih, I. and Talal, S. 2007. A Literature Review on Semantic Web. Understanding the Poineers' Perspective. *Computer Science & Information Technology,* pp. 15-28.

19. Berners-Lee, S.T. 2009."The next Web," TED Talks. [online] Available at: <https://www.ted.com/talks/tim_berners_lee_on_the_next_web?language=en>[Accessed 9 March 2022].

20. Semantic Web Technology. 2015 "Layered Architecture". [online] Available at: <http://www.srmuniv.ac.in>[Accessed 9 March 2022].

21. W3.org. 2022. "W3C Semantic Web FAQ". [online] Available at: <http://www.w3.org/RDF/FAQ> [Accessed 9 March 2022].