# Dependable Semantic Web

**Ms.J.Gayathri, M.Sc., M.Phil., (Ph.D)**
Assistant Professor, Dept of Computer Applications,
KG College of Arts and Science, Coimbatore,

***Abstract-*** *This paper describes some ideas for a reliable semantic web. Semantic web is a technology for perceptive web pages. It is significant that the semantic web is secure. In accumulation, data exchanged by the web has to be of high quality. The processes that the web supports have to congregate certain timing constraints. This paper discusses these aspects, and describes how they endow with a dependable semantic web.*

***Keywords***- Data integrity, Warehouse, Security, Fault tolerance.

## I. INTRODUCTION

Recent developments in information systems technologies have resulted in computerizing numerous applications in various business areas. Data has become a decisive resource in many organizations, and therefore, competent access to data, sharing the data, extracting information from the data, and making use of the information has become an urgent need. As a result, there have been many efforts on not only integrating the various data sources speckled across several sites, but extracting information from these databases in the form of patterns and trends has also become vital. These data sources may be databases managed by database management systems, or they could be data warehoused in a repository from manifold data sources. The advent of the World Wide Web (WWW) in the mid 1990s has resulted in even greater demand for managing data, information and knowledge effectively.

There is currently so much data on the web that managing it with conventional tools is flattering almost impossible. New tools and techniques are desirable to effectively manage this data. Therefore, to endow with interoperability as well as warehousing between the multiple data sources and systems, and to extract information from the databases and warehouses on the web, various tools are being urbanized.

One of the recent developments with the web is the semantic web by Tim Berners Lee, Jim Hendler and others [BERN01]. The semantic web is essentially about machine comprehensible web pages. In the article on semantic web by Berners Lee et al., the semantic web is described to be a web that can comprehend and interpret web pages and manage activities for people.

These activities could be maintaining appointments, giving advice, and in essence making the life of the human as easy as possible. If the semantic web is to be effective, then we need to guarantee that the data and information on the web is timely, accurate, and precise. Note that with dire data one cannot make good decisions. Therefore, we need to extend ways to incorporate quality parameters into the technologies for the semantic web. These technologies include XML, (eXtensibe Markup Language), RDF (Resource Description Framework), and agents.

There is petite work reported on data quality, security and integrity for the semantic web. We need to start investigating the issues while we conduct research on the semantic web. If data quality, integrity and security are added as an addendum, then it will be very difficult to build practical systems. This paper will review the developments with the semantic web and then discuss ways of making the semantic web trustworthy. These include integrating data quality, security, integrity and real-time processing for the semantic web. Because of the web we now have access to all sorts of data.This may compromise quality as data may originate from unfrosted sources and be passed from one to another. The paper also discusses research guidelines.

## II. SEMANTIC WEB DEPENDABILITY ASPECTS

While the semantic web as a concept is still sprouting, there have been many developments in this area. These include RDF, Ontologies, Agents, and Databases. One could envisage a issue and subscribe model for the web where producers publish the services while consumers subscribe for the services. Agents operate on behalf of users. There are various types of agents including brokers. These brokers confer the best deals for their customers. These services could be managing schedules and appointments as well as giving advice and essentially managing all of the activities for a customer.

Consider a hypothetical example of John who is a physician. The web will arouse him up depending on the day

of the year. Then he will be conversant of the day's schedule and appointments. He could get information about the optimal routes to get to his destinations.

He could also be clued-up of where his personal accessories are. Then as the day progresses the web will manage the dynamic situations such as accidents, unexpected traffic, and other unforeseen events. Then the web also gives him information about his patients such as patient history and then when his work is over the web makes measures for John to meet his wife Mary for dinner and go to the theater. One could say that the web has completely taken over John's life, and yes the web would be involved in many aspects of his life. It would certainly make things a lot easier for John and in the end it is up to John to follow the advice and directions. In the end John will decide how he should proceed for the day.

Next let us examine constancy aspects for the semantic web. By Dependability we mean security, fault tolerance, integrity, data quality, and real-time processing. It will be difficult to guarantee that all the constraints are met. The dispute is to develop quality of service constraints for the semantic web. For example, the agents that interrelate with each other need to be secure and meet the timing constraints. We need to make certain that they are fault tolerant. The data being exchanges such as XML documents need to be of high quality. Various access control policies have to be enforced for XML documents. Trust issues are also critical. For example, to what extent do you trust your source? The remaining sections will focus on two aspects, data quality for the semantic web and real-time services for the semantic web.

## III. DATA QUALITY FOR THE SEMANTIC WEB

Data quality (sometimes referred to as "information quality" or IQ) has received increased consideration by industry and academia in recent years. For example, MIT Sloan School runs a Total Data Quality Management program [TQDM], with annual workshops. Early work in the area focused on incorporating data quality attributes for data models such as the entity relationship model. Data quality attributes would include information such as timeliness, accuracy and precision.

We anticipate that the semantic web will require techniques to manage the quality of data on the web. Much of the interest in data quality is due to attempts to integrate data from previously-unconnected systems, often in a data warehouse. The semantic web can be described as a virtual integration of information and services on the web. The semantic web enables relationships between data from previously-unconnected sources, and proliferation of data from one organization to another. Therefore, users of the semantic web will need some way to determine the quality of data used. In core, it is necessary to understand the quality of information if one is to achieve semantic understanding of the information.

High quality data is critical for e-commerce transactions. These transactions may involve large sums of money, and if the data is of poor quality, the consequences may be devastating. Many of the businesses that have studied data quality (within their internal systems) have found that quality problems can be directly traced to momentous loss of revenue. Some commercial tools now address data quality issues, but many of these are exclusively tuned to common problems with customer names, addresses, and phone numbers. This tuning is not directly pertinent to e-commerce and the semantic web, where contact with customers is largely electronic (although, one could imagine similar tools for email addresses).

Some more flexible tools are budding that use data mining techniques to identify statistically anomalous data. For example, such a tool might notice that all but a few entries in a given database field are five digit numbers, without knowing that the field represents US postal codes. One impending limitation of these tools is that they identify statistically rare occurrences, and thus may miss quality problems that affect large fractions of the data. A more complete data quality management approach would also need a way to propose possible actions to correct the problems identified.

Most proposed data quality attributes are aggregate or average measures of specific quality attributes. These metrics are needed so that users of the semantic web can decide how to use particular sources (i.e., sites). For example, RDF can be used to make assertions about the content provided by a source. We have also proposed per item quality attributes, which would allow assertions about the quality of individual data objects (given some form of identifying those objects) [Rosenthal 2001]. For example, one may assert that the stock recommendations in one thread of messages are suspect, even though other threads provided by the same source are reliable.

The semantic web will also involve data quality attributes defined from both provider's and consumer's viewpoints. We have discussed attributes for precision, timeliness, and source of information, which replicate the producer's viewpoint. However, consumers often need to know the value of the information to a decision or business process that uses it. Quality attributes like value are prejudiced, and it is not generally necessary for others to understand their semantics. However, in the semantic web, the

decisions and business processes of one user will often be used to create information products used by others. The semantic web enables information supply chains (including long ones) by essential semantics for the products at each link in a chain. Like other kinds of supply chains, an information supply chain using the semantic web will only be as dependable as its weakest link. So, the web will need to proliferate quality attributes from producers to consumers, and users will need ways to define or derive quality attributes for products from those of the information they use to produce those products.

## IV. REAL-TIME PROCESSING AND THE SEMANTIC WEB

Web services are very popular. Web services include e-business, services for ensuring that the semantic web operates appropriately as well as services for locating. Web services are very popular. Web services include e-business, services for ensuring that the semantic web operates suitably as well as services for locating resources on the web. Web services have to be secure as well as meet timing constraints for many applications. Consider the case of financial services on the web. We need to get stock quotes in a timely manner. If it is too late, the result is of no use to us resources on the web.
Often real-time processing means that we need end to end quality of service. That is, the clients, the services, the middleware, the agents and all the processes have to convene timing constraints. There has been lot of work on real-time agents, real-time middleware, and real-time serves and databases. The challenges are to incorporate all of these technologies to ensure end to end quality of services. Furthermore, these agents and servers also have to be sheltered. This means that we have to develop flexible policies to meet timing constraints as well as the security constraints.

## V. DIRECTIONS

This paper has discussed the need for a dependable semantic web. We first endow with an overview of the semantic web and then discuss dependability issues. These include integrating security, real-time processing, fault tolerance and data quality. Then we discussed data quality for the semantic web as well as real-time processing aspects.

## VI. CONCLUSION

This paper provides only preface directions for dependable semantic web. The ideas reachable in this paper have to be developed further before dependable semantic web becomes a reality. We need to establish soon before it is too late to incorporate features like security and data quality into the semantic web.

## VII. SCOPE FOR FUTURE WORKS

There is scope for lot of work some developers oppose on whether the Semantic Web should rely more heavily on rules or on ontologies. Critics also say that the project is enormously impractical. First, people don't actually think in terms of the graphs that RDF uses. Second, it seems unlikely that businesses and existing sites will actually devote the time and resources it would take to add all the necessary metadata. In the future, off-the-shelf software might include options for toting up metadata when creating new documents, but that tool still might not make the project feasible on a larger scale when combined with pertinent research in related areas such as trust, archiving and digital libraries.

## REFERENCES

[1] [BERN01] Berners Lee T. et al, The Semantic Web, Scientific American, May 2001.
[2] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” Foundations and Trends in Information Retrieval, vol. 2, nos. 1/2, pp. 1-135, 2008.
[3] [Rosenthal 2001] Rosenthal, A.,etal,Data Quality in the Small:Providing Consumer Information, Information Quality Workshop (IQ 2001), November 2001.
[4] B. Pang and L. Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, 2004
[5] Bing Liu and Lei Zhang, “A Survey of Opinion Mining and Sentiment Analysis”, SpringerLink, 2012 .
[6] JesusSerrano-Guerrero, Jose A.Olivas, Francisco P.Romero, EnriqueHerrera-Viedma, “Sentiment analysis: A review and comparative analysis of web services”, Information Science, August 2015.
[7] KumarRavi, VadlamaniRavi, “A survey on opinion mining and sentiment analysis: Tasks, approaches and applications”, Science Direct, November 2015.
[8] Sarala.R, “Augmenting Data warehouses with Big Data – A Survey”, International Journal of Multidisciplinary Educational Research (IJMER), Volume 6, Issue:10(3),Oct,2017.
[9] Aswanandhini.S, “A Customizable Multi-Agent System for Distributed Data Mining”, National Conference on Innovative Trends in Information Technology, 2014.