

# A Review on Semantic Organization of Data

Muneeb Niaz<sup>1</sup>, Asma Sajid<sup>2</sup> and Jafar Raza Alam<sup>3</sup>

<sup>1</sup>Department of Computer science, Govt. Collage University,  
Faisalabad, Punjab, Pakistan

Department of Computer science, Govt. Collage University,  
Faisalabad, Punjab, Pakistan

<sup>3</sup> Department of Computer science, Govt. Collage University,  
Faisalabad, Punjab, Pakistan

## Abstract

The Information on web resources is increasing day by day because of increment of user's interaction and contribution in web resources. It is becoming a challenge to collect relevant and useful information from this huge amount of data because most the data is not interrelated or organize. Semantic web allows collecting meaningful information from bulk amount of data, there for semantic web helped a lot to optimize the searching technique and facilitate the use to get useful information. The paper discussed some research performed on semantic web and some tools and platform developed by using semantic web approaches in past few years.

**Keywords:** *Semantic web, RDF, Web Scrapping,*

## 1. Introduction

The concept of semantic web was proposed by Tim Berners Lee the inventor of World Wide Web. The question is why semantic web is needed? The basic difference between semantic web and current searching techniques is the machine readability. Semantic web provides us an easy way to find, combine, reuse and share the data based on the information read by machine. Human can read different types of data in different formats but machines are unable to do so. Machines can collect data from different sources but they are unable to process further if the data is in different formats. Semantic web is designed in machine readable format, these types of website can get more accurate and processed data from different types of sources.

It is analyzed that a very limited amount of data on internet is in properly organized and in proper format and the data on internet is increasing day by day because every single user is not contributing in information. A huge struggle and effort is needed to collect meaningful, relevant and useful data. There are many search engines that collect information from different types of sources but their searches are based on relevancy only while semantic web understands need of search and collect meaningful

data from data sources that are very helpful for not only user's perspective but also as organizational perspective.

As the usage of semantic web is increasing the research is also opening new angles of semantic webs. Different tools, frameworks and algorithms are proposed to facilitate the research on internet. The paper is going to discuss some researchers held in semantic web as well as some tools, framework and algorithm developed on the bases of semantic web in last few years. The paper is going to discuss their idea, concepts about semantic web their contribution in related area also the difference between these semantic webs based researches and currently used tools and techniques will be discussed. This research will help to understand the different concepts and contribution of semantic web in World Wide Web.

## 2. Previous work in Semantic Web

A lot of work is being done in semantic web but there is a still a lot of potential research for semantic web. In current section some pervious work is going to be discussed.

### 2.1 Dependable Semantic Web

Thuraisingham (2002) published an article in The National Science Foundation entitled "Dependable Semantic Web." Described that semantic web is basically a technology that helps to understand the web pages. Data from the web pages is categorized according to the quality of that material. For The Categorization of the Data, Semantic web is depended upon some Constraints. In the paper some concepts and ideas are propose for these dependable Semantic Web.

The Different aspects of Dependability of Semantic web are discussed thoroughly in the paper. Thuraisingham discussed the concept of agents involved in dependability of semantic web. These agents are responsible to increase or decrease the quality of data. Both the consumer and the

provider of the web pages are involved in these quality attributes. The quality of Data is named as information quality of the web pages.

## 2.2 A Mental Cognitive Model

Kwan (2002) published an article in Database and Expert Systems Applications named “A Mental Cognitive Model of Web Semantic for E-Customer Profile”. It is discussed that the new era of computer is mainly based upon E-Commerce as the main source of revenue generation is the Customer from e-commerce. So it is a need to solid infrastructure that could access and manage the whole data. The main focus of E-commerce is always the traffic. So the traffic landed to any e-business website must be sufficient for reasonable revenue generation. It is also discussed that a successful e-business based on the relationship and interaction between e-business and customer. To achieve the purpose (increase the interaction between provider and customer) a Model is proposed on the bases of Mental Cognitive. The MC Model investigates and navigates on e-customer by identifying the behavior of customer on e-business website. The MC model uses the click and tick behavior of customer to support e-business.

## 2.3 Building Secure Survivable Semantic Web

Thuraisingham (2002) published an article in IEEE International Conference on Tools with Artificial Intelligence named “Building Secure Survivable Semantic Web”. The paper describes some concept of security and reliability of web semantic based on some ideas of dependable semantic web which was proposed by the same author. The dependability issues are discussed in detail. It is an important issue that semantically organized websites must be secure to exchange the data from authentic sources. The data exchanged by the sources must be high in quality as well as it must survive errors and failures, the data quality aspects are discussed in detail. The paper focused on the dependability of data collection sources and their security issues thoroughly. The paper also discussed the previous work performed for semantic web survivability as well as their security.

## 2.4 Scalable Template-based Query Containment Checking for Web Semantic Cache

Amiri (2003) published an article in International Conference of Data Engineering entitled “Scalable template-based query containment checking for web semantic cache”. Concept of semantic cache was proposed

for client-server database system but in recent years it is used in dynamic web systems. These caches are helpful to fast query execution and optimization. Different algorithms are proposed for different kind of dynamic systems. But the problem is that the containments of these queries have limitations. Solution is proposed for these limitations in the article. Efficient algorithms are proposed for containment of predicated of similar queries. If the queries are arranged by the similarities of predicates it will improve the efficiency of search queries, I technique to do so is proposed in research paper. The basic purpose of these techniques and algorithms is to reduce the cost and workload.

## 2.5 Semantic Web Challenges and New Requirements

Toja (2005) published an article in International Workshop on Database and Expert Systems Applications named “Semantic Web Challenges and New Requirements”. The paper discussed the history of Semantic web and the participation of Tim Berners-Lee in Semantic web Area. Article also covers the initial role and goals of semantic web as well as limitations occur in beginning of semantic web. The future of semantic web and the extensions of semantic web with other areas are also discussed like some fields of data engineering.

## 2.6 Integrating Web Services and Web Contents

Amaiya (2005) published an article in International Conference on Web Services entitled “An Architecture for Personal Semantic Web Information Retrieval System – Integrating Web Services and Web Contents”. The paper discussed that the main advantage and purpose of semantic web is to process the information automatically, most of the researches in semantic web area are performed for collecting information from different web services but author focused the concept on whole life cycle and proposed architecture for this purpose. The architecture is basically a conceptual architecture and based on multi agent technologies. The architecture is hybrid architecture of semi automatic and automatic processing of web resources through which restricted and prohibited content and information can be access and processed. According to the author it is a need for next generation of semantic web to access any kind of data targeted by the researcher. The processing between two different sources of data is on the bases of similar ontology which is definitely a limitation in this architecture and need more attention on mapping the ontology to solve this problem.

## 2.7 Agent Based Semantic Web

Wang (2005) published an article in ITHET 6<sup>th</sup> Annual international Conference entitled “An Agent Based Semantic Web Department Content Management System”. Paper described the semantic web and its benefits in context of Metadata, also discussed the intelligence of semantic web to collect information from various sources. The paper proposes an agent based system that is architecture for content management. The system targets higher education institute’s departments to manage data. The system collects all information about the entities of the department; the data includes their information like name, address and age as well as their academic’s details. To collect and manage all this information RDF is used. The paper also discussed the scope and potential of semantic web in educational institutes. The system is implemented in department related to computer science of Acadia University.

## 2.8 Semantic Web Shifting from Knowledge Push to Knowledge Pull

Naeve (2005) has published an article entitled “The human semantic web shifting from knowledge push to knowledge pull”. Described that semantic web is a coordination and a conceptual interference between human and machine. The semantic web has three levels which define it properly. These levels are collaboration, isolation and coexistence. The highest goal of semantic web is collaboration between human and web that can be achieved by conceptual collaboration. This conceptual collaboration creates a bridge between different ontologies. While semantic isolation is described as a concept to know the exact location of the data being searched. If a query is generated and the data is not relevant to the query or the keyword generated by the human this stage can be called as coexistence and it occurs when data is searched without knowing the exact location of data.

## 2.9 A Semantic Web Editor

Pereire (2006) published an article in International Conference on Internet and Web Applications and Services entitled “SWedt: A Semantic Web Editor Integration Ontologies and Semantic Annotation with Resource Description Framework”. The article describes the increment of tools related to semantic web since last three years but these semantic tools has very strong limitations. The common limitation of these tools is that these are designed to work with a very few or limited layers of architecture related to semantic web. So the

market strongly needs to develop a tool which can integrate different layers to utilize maximum number of layers in minimum time frame. Architecture is proposed in the research paper for that specific limitation named “SWedt”. The tool focus on fully standardized layers and their functionalities as well as just standardized layers to integrate and it allows the user to create semantic web documents easily.

## 2.10 Relation-Based Search Engine in Semantic Web

Li (2007) published an article in IEEE Transactions on Knowledge and Data Engineering entitled “A Relation-Based Search Engine in Semantic Web”. The article declared the semantic web as a breakthrough in data processing and named as “big bang”. The most important and useful tool for collecting data from different sources of data is search engines. The machines now these days in semantic web cares only about the location and the type of data, there for a lot of loopholes still exist in search engines and the most popular and advance search engines are still unable to satisfy user. The article proposed a search engine based on relations and implemented in virtual environment. In that virtual environment of semantic web tested the architecture and major algorithm related to relation-based search engine. As the search engine is based on relations, these relations and concepts are defined in RDF. The search engine retrieves data by following these concepts and relations.

## 2.11 A Flexible Ontology Reasoning Architecture

Pan (2007) published an article in IEEE Transactions on Knowledge and Data Engineering entitled “A Flexible Ontology Reasoning Architecture for the Semantic Web”. The article defines that any system based on knowledge can use the power of Semantic and enhances its capacity to retrieve useful as well as meaning full data from different sources. The power of semantic to retrieve data is based on ontologies. Ontologies are helpful for collection of data defined by users and it supports customized data, its predicated and its data type. The paper proposed a reasoning architecture which is flexible as perspective of unfolding the prototype and web ontology languages. These two features of architecture provides the facility to the user to define their own predicates and data types as well as user can change its data type or add new one and the concept will remain same. It will be helpful to customize the research and user can get data according to their desires and requirements.

## 2.12 A Reasonable Semantic Web

Hitzler (2010) Published an article in journal of semantic web entitled “A reasonable semantic web” described semantic and shared information. Semantic web is described as a most advance technique to capture shared information. When an agent like a web server, web service, human dialog or a database completes a message, the agent have many unspoken queries in mind generally. So the message generated by the agent contains more meaning. A reasonable semantic web must be able to define the meaning in unique way.

## 2.13 Three Level Scraping Model

Villamor (2011) designed a scraping model on bases of three levels syntactic scraping, semantic scraping model and scraping service. Scraping service provides facility to collect the data from websites at high level by providing an interface to intelligent agents and generic applications. A semantic RDF model is defined in semantic scraping model. It provides a declarative approach in scraping process by dividing HTML into fragments. While the syntactic scraping provides implementation of the defined model for specific technology using the fragments.

## 2.14 Intelligent Semantic Web Search Engines

Madhu (2011) published a paper in International journal of Web & Semantic Technology. The title of the article is “Intelligent Semantic Web Search Engines: A Brief Survey”. The article described the working of search engines like bing, yahoo and google. The main aim of these search engines is to extract meaningful data. Semantic web is playing a very important role to achieve this aim however the search on semantic search engines is in beginning stage. In these search engines the information is described using Resource Description Framework (RDF) a new W3C standard. Also the second most important part of a semantic web is ontology. Web ontology language (OWL) and Resource Descriptive Framework used to represent ontology. These two models are recommended by W3C.

## 2.15 Evaluation Factors of Semantic Web Application

Yu (2012) published an article entitled “Balance: A Key Factor For The Evaluation Of Semantic Web Application”. Paper describes that there is a key role of balance not only in science but also in our daily life. As the development of semantic web increasing day by day,

there is a strong need to evaluate the quality of semantic web. Currently available tools to evaluate semantic web are also based on some specific applications or same technology of semantic web which is not enough to evaluate the general features of a semantic web. The research is based of the evaluation of semantic web in general perspective. The characteristics of a semantic web are gathered. The critical points are collected from these characteristics to evaluate the balance of the semantic web. After summarization of these critical points the key factors of balance evaluation are presented. The research is helpful to open a new angle to analyze the semantic web.

## 2.16 Intelligence in Information Retrieval

Aroma (2013) Published an article entitled “A Semantic Web: Intelligence in Information Retrieval”. Article describes that the web resources are increasing day by day and this increment of resources creates a strong need for an advance research system for retrieval of information from these sources. The interaction between user and web resources is also increasing day by day, almost every user contribute to increase the information of web resources. Normally this huge information has no connections with other sources and also not organizes so well so that proper and meaningful information can be collected from these bulky web resources. There is a basic difference between searching schemas that is currently in use and semantic web search schema. The current web schemas collect information based on relevancy only while semantic web search schema understand the need of search and collect suitable results. For this purpose semantic web utilize the support of machine readability. The paper provides an overview of mapping in semantic web and also proposes a mechanism to find out an efficient semantic web mapping algorithm.

## 3. Semantic Web in 10 years

Grobelink published an article entitled” Semantic Web in 10 years: Semantics with a purpose” described that semantic web evolution starts from the very basic concept KRR (knowledge representation and reasoning) and now it is moving towards very specific targets. The main aim of semantic web is to be able to make systems and organization, these systems and organizations would be to work together in almost all areas businesses and research. According to the article in these days semantic web is using lighter semantic approaches and will be able to use advance semantic approaches in 10 years like it would be able to handle big data and heavily involve in area of machine translation and text reading.



## 4. Figures

Graphical representation of semantic web concepts.

The figure given below explains the relationship among its tools , techniques and concepts.

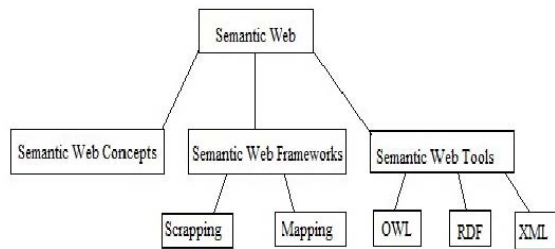


Fig. 1 Semantic web overview.

## 5. Conclusion

Paper is based on the previous researches on semantic web technologies; it discussed some basic concepts and advantages of semantic web. Previous researches using different concepts and approaches are also discussed in paper. Paper focuses on usage of semantic web to facilitate the end user like semantic web in e-commerce. It discussed some general concepts like scraping and role of semantic web in search engines as well as some different concepts like reliability and security. Paper also discussed some implemented frameworks and tools using semantic web approaches to collect and organize data.

## References

[1] B.Thuraisingham, E.Hughes, and D.Allen, "Dependable semantic web In Object-Oriented Real-Time Dependable Systems", IEEE 2002, pp. 0305-0305.  
 [2] I. S.Kwan, "A mental cognitive model of Web semantic for e-customer profile: In Database and Expert Systems Applications", IEEE, 13<sup>th</sup> International Workshop, 2002, pp. 116-120.  
 [3] B.Thuraisingham, "Building Secure survivable semantic webs Tools with Artificial Intelligence", IEEE, International Conference ICTAI 2002.  
 [4] K.Amiri, S.Park, R.Tewari & S.Padmanabhan, " Scalable template-based query containment checking for web semantic caches: In Data Engineering", IEEE, 19th International Conference 2003, pp. 493-504.

[5] A.M.Tjoa, A.Andjomshoaa, F Shayeganfar and R.Wagner, "Semantic Web challenges and new requirements: In Database and Expert Systems Applications, IEEE, Sixteenth International Workshop 2005, pp. 1160-1163.  
 [6] Y. Haibo, M.Tsunenori and A.Makoto , "An architecture for personal semantic Web information retrieval system ntegrating Web services and Web contents", IEEE, 2005.  
 [7] Y.A.Wang and S.Elhadi. "An agent-based semantic Web department content management system: Information Technology Based Higher Education and Training",IEEE, 6th International Conference, 2005.  
 [8] A.Naeva "The human Semantic Web shifting from knowledge push to knowledge pull", International Journal on Semantic Web and Information Systems (IJSWIS), 1(3), 1-30, 2005  
 [9] R.G.Pereira, and M.F.Mario, "SWedt: A semantic web editor integrating ontologies and semantic annotations with resource description framework, IEEE, International Conference on Internet and Web Applications and Services/Advanced International, 2006.  
 [10] L. Yufei, Y. Wang, and X. Huang, "A relation-based search engine in semantic web: Knowledge and Data Engineering", IEEE ,2007.  
 [11] P. Jeff z. "A flexible ontology reasoning architecture for the semantic web: Knowledge and Data Engineering", IEEE, 2007.  
 [12] P.Hitzler, and V.F.Harmelen, "A reasonable semantic web: Semantic Web", IEEE, 2010  
 [13] V.Fernández and I.José, "A Semantic Scraping Model for Web Resources-Appling Linked Data to Web Page Screen Scraping." (2011): 451-456.  
 [14] G.Madhu, D.A.Govardhan, and D.T.Rajinikanth, "Intelligent Semantic Web Search Engines: A Brief Survey", IEEE, 2011.  
 [15] N.Gherabi, K.Addakiri, and M.Bahaj, "Discovering new technique for mapping relational database based on semantic web technology, IEEE, Communications and Information Technology (ICCIT), 2012, pp. 27-32  
 [16] H.Yu,T.Mine, and M.Amamiya, "Balance: A Key Factor for the Evaluation of Semantic Web Application", IEEE, Advanced Applied Informatics (IIAIAI), 2012, pp. 203-208.  
 [17] R. J.Aroma,and M.Kurian,"A Semantic web: Intelligence in the information retrieval", IEEE, Emerging Trends in Computing, Communication and Nanotechnology (CE-CCN), 2013, pp. 203-206.