An Application Programming Interface Based Architectural Design for Information Retrieval in Semantic Organization

Usman Shafique virk[1], Zeshan Akhter[2]

Department of Computer Science,
University of Agriculture,
Faisalabad, Punjab, Pakistan

usmanvirk59@yahoo.com, zeshanakhter786@gmail.com

ABSTRACT: The World Wide Web (WWW) allows users to share the information or data from the large database repositories all over the world. In World Wide Web the data is increasing on internet day by day but small amount of data on internet provides us semantic access. It is a great need to have a framework to organize data structurally and in a fast way. The World Wide Web is used as the source of information. The purpose of this research is to make conceptual semantic scraping framework for fast and structured extraction of reusable information from one specific domain. There are many convenient applications that are playing important roles to facilitate users in a specific domain but are not fulfill the desires due to exponential growth of data over internet. It is difficult to retrieve information from unstructured data because of messy Hyper Text Markup Language (HTML) pages, cookies check, and data privacy and exponential growth and enormous size of Web sources. In this study, an API based framework for data handling is introduced that extracts the data from targeted sources, organize it semantically and provide it in a structured or tabular format. It makes the process of generating the query related to his/her topic very reliable and convenient.

KEYWORDS: API, Information Retrieval, Semantic Organization

1 INTRODUCTION

Computer is one of the greatest inventions made by the human. With the passage of time modern communication have given rise through internet. Internet is a huge source of information. Is the era of science computer devices are one of the greatest inventions ever made by the human. Modern communications have also going to raise using internet with the passage of time. WWW (World Wide Web) over a domain is great source of knowledge. On the internet huge amount of data as well as information is placed over internet in many formats and forms. Unfortunately, this large amount of data major part is unstructured. Data over the WWW is increasing with the passage of time.

The Semantic web idea was proposed by inventor Tim burners-Lee, Tim Burner is also called the inventor of the World Wide Web (WWW) and different sites like URL etc. At the present time the semantic web is well know and popular and very familiar in peoples of the researcher. In 2013, semantic web are more than four million Web spaces contained a Semantic Web markup. The utilization of semantic web is expanding step by step, not just vast and major sites like Google are utilization semantic web way to deal with collaborate a huge number of sites yet numerous little sites are additionally utilizing semantic web to deal with gather information inside they could call their own destinations. While gathering information from sites on little scale or vast scale, scrapping is a best standout method amongst the most helpful procedure to gather information. Resource Description Framework and Ontology Web Language are two buildings blocks on the Semantic Web; Different algorithms' techniques are applied by this scrapper to gather unarranged and unstructured needed information form a large scale. At whatever point we can get to information on the web as per our needs not very many sites are accessible on the web are utilizing semantic web innovations. As not very many information are organized and arranged semantically on web for getting to execute different undertakings for information resembling discovering, again reuse or brushing to achieve a few different valuable data so there are many number of issues are interfacing while we discovering related information on websites, an excess of unimportant information is additionally consolidated likewise with our obliged information.

Villamor designed a scraping model on basis of three different level scraping service, syntactic scraping and semantic scraping model. Scraping service presented skills to collect the information from websites at high level by providing an interface to intelligent agents and generic applications.

SEMANTIC WEB ARCHITECTURE

The fundamental structural planning of semantic web contains Identifiers (Uniform Resource Identifiers) and

character code as Unicode. Over this layer is the Syntax layer, characterizing the syntactical relationship and the base here is XML. Over this layer is the Data Interchange layer with RDF characterizing. Above it the query handling part is taken care of by SPARQL and the scientific categorizations is determined by RDFS. The Ontology is represented by OWL and principles by RIF/SWRL. Above it is the unifying and the evidence layer. All the previously stated layers were scrambled utilizing Cryptology over these is the trust layer.

Logic layer makes a characterized organization and standard regulations for the internet searchers which are dependable to create the final data in semantic web. This layer which is situated up to the Ontology layer is utilized to clarify the reasonable words in machine level. Regulations verification implies that the machines must lay on them recovering the data in large amount measure of the data. These regulations are truth be told scripts and the software programs. One of the objectives of Semantic Web is to achieve trust. As in Semantic Web the pursuit operation happens utilizing the web crawlers, the clients must have the feeling of dependability about the security, improvement and modification and organized, recovering the data. This layer is vital in internet shopping from security and trust perspective. URI layer shows the text and the sending method of them to the web environment. Also, it is possible to use this layer and make all web environment resources accessible and explainable. The HTML pages are the documents which are implemented in this layer. URI is used to identify the concepts in Semantic Web. In fact it is a kind of URI to identify the resources in web. Unicode is used to support the multilayered language. XML is a standard language for arranging the data using tagging and like HTML; it is script language to some extant goes through the way. But it differs in this way that the information on it is saved in syntactic form and is easily accessible or to be linked. This specification makes different elements of the web pages attributed different and favored specifications. RDF (Resource Description Framework) is a XML (Extensible Markup language) based language which is created to explain the concepts and create the documents in SW. And using a set of mathematic and semantic relations, it is possible to form the relation between the data in a logic way which is accessible directly. RDF documents are able to explain the words in a way that it is understandable to the machine. Now, the different languages are used for presenting the contents of the web pages like HTML and XML.

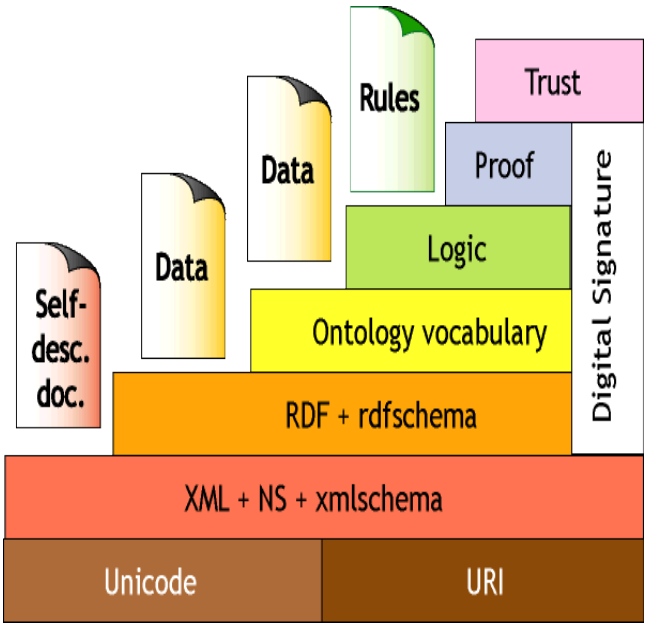


Fig. 1. Layered Approach to Semantic Web

2 MATERIAL AND METHODS

There is a huge amount of information are placed on internet in many different forms and formats. Majority of information is unorganized and unstructured it is almost impossible to organized information in proper format. Only the addition is not only from professionals but an average internet user is also playing his/her role to increase data. Because of these kinds of involvements or non-professionals it is difficult task or collects the right, relevant and useful information from that huge amount of data. So semantic web is helpful to collect relevant amount of data and resolve these kinds of problems.

From the previous couple of years site is changing to semantically sorted out networks to encourage their clients. There are numerous web scrapper in view of semantic web as it encourages to get focused on and significant information from huge measure of information. Be that as it may, there is not any scrapper which can gather information from more than one entrance of research websites. As the quantity of documents, paper, and publication and research articles from over the world is expanding, numerous portal of research are create to collaborate among these publication there is an in number requirement for a system to create connection among these research gateways.

3 GOALS OF THE RESEARCH

There is a lot of need to gather information which is more relevant and precise, efficient, well organized amount of data collected from many research journal and portals. To solve such kind of problems we need to develop a framework which is capable to understand the functionality of different research portals to gather information and manage it semantically.

The goal of this exploration research work is to build up an extended new approach that make the strategy simple to plan general data retrieval framework for semantic organization to meet the requirement of user.

4 NEW API-BASED FRAMEWORK

A Framework for Semantic Organization of information retrieval that is base on Application Programming Interface (API) focuses to facilitate users using different devices with structured and tabular data that is portable. Information technology is facilitating many different fields but also increase of data is creating many ambiguities. It is very hard to collect relevant and authentic data from huge amount of data available on internet. There are many convenient applications that playing their roles to facilitate users in a specific domain, the research is going to open a new field to facilitate users in semantic organization.

Information for semantic web uses conceptual representation of content beyond plane keywords such as knowledge bases. This architecture handles the concept representation of content. Quires extractions match the semantic relevant results. The plane key words entered by the user is expended at converted into the Symantec query by the different ways such as by matching, by using the links by using the thesaurus. When the semantic query matches the semantically indexed web content should match with two parts keywords and concepts of the semantic query. The www (World Wide Web) allow user to share the information and data from the large data repositories all over the world the user will also generate the query related to his/his topic. The framework extract the data from targeted resources, organize it symmetrically and provides to the user in tabular form in structured format such as XML and JSON form.

The basic architecture of Framework is as followed, The SUD (system under development) will generate a query to web services server, and the web services server will generate the query on both research portals (IEEE Explore and Microsoft Academics) and collect the relevant data from these two portals, after that these results will provided to SUD by web services server. The working of framework in IEEE Explore and springer will be discussed further in detail.

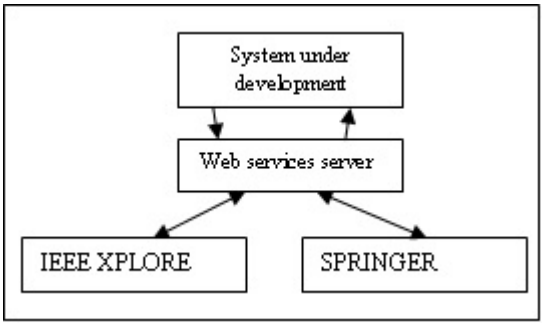


Figure 4.1: Main architecture framework

ARCHITECTURE OF SPRINGER

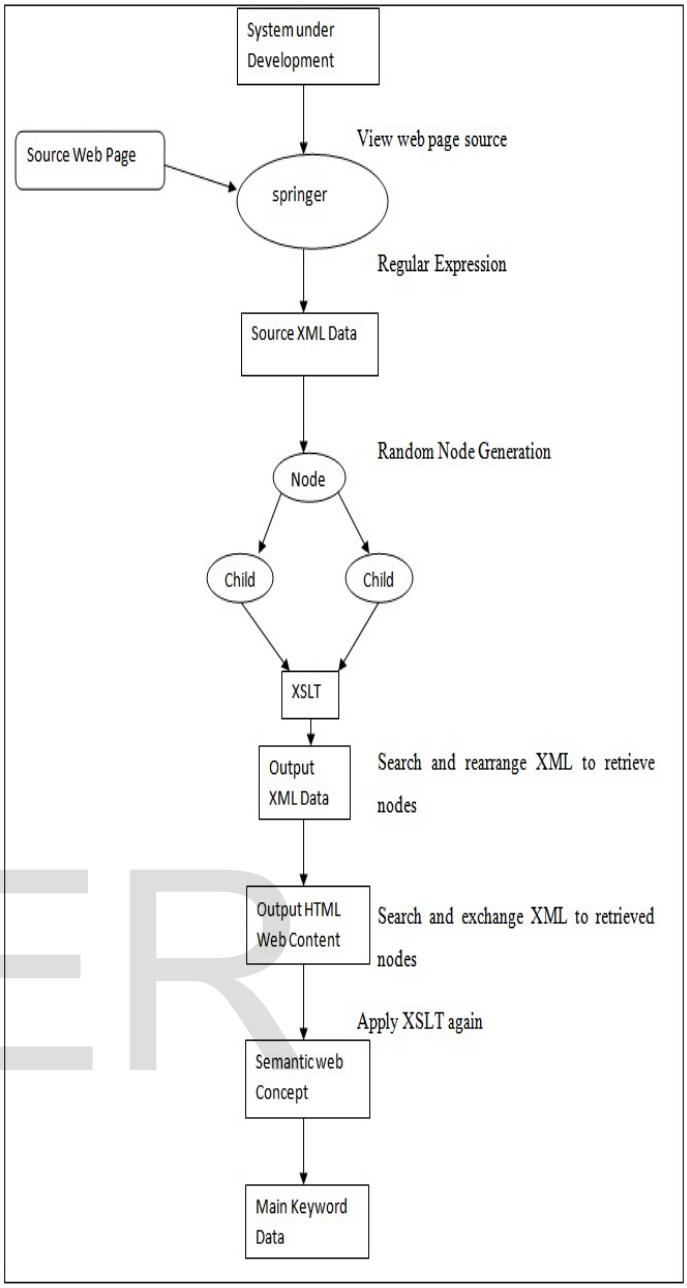


Figure 4.2 Architecture of Springer

ARCHITECTURE OF IEEE EXPLORER

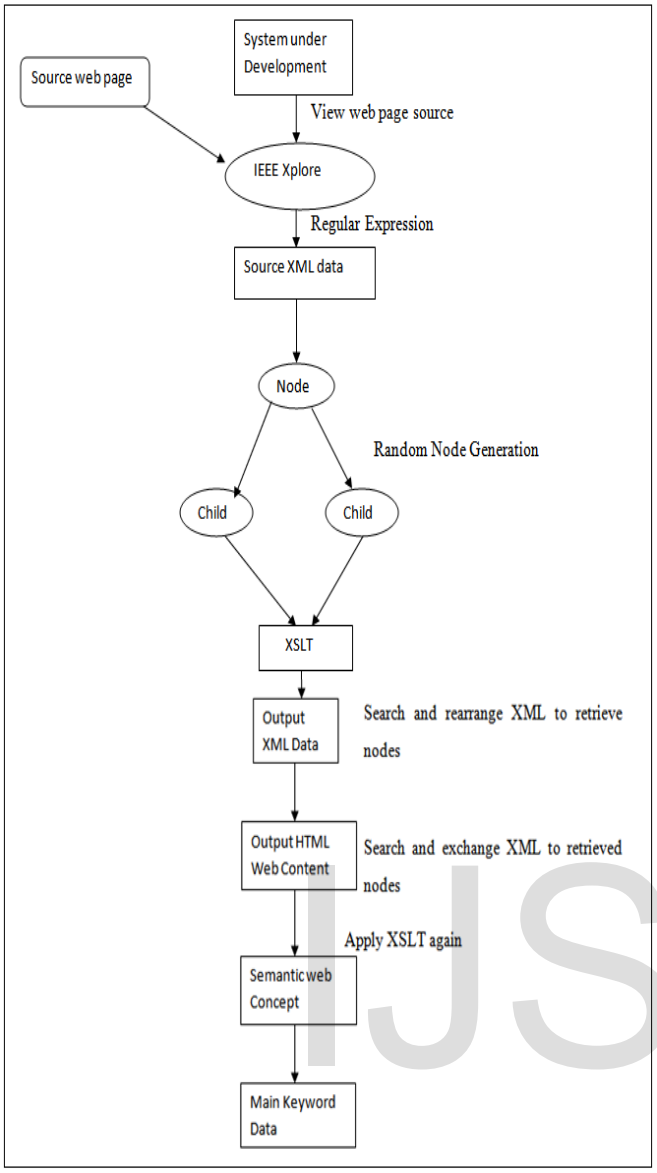
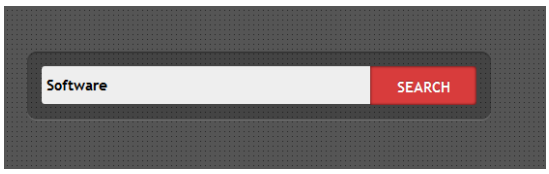


Figure 4.3 Architecture of IEEE

5 RESULTS

Since we alter the query and different results are produced here some queries are generated and results are retrieved are given below. The result shown by the query "Reverse Engineering" is given below.



The result shown by the query "Software Quality Assurance" is given below.

Title	Abstract
Software Quality Engineering: Making It Happen	Software quality engineering calls for a formal management of quality throughout the full lifecycle of software or a system. Several quality models were developed in the course of past three decades, some of them recognized mostly by the scientific community, others also gaining recognition within the industry. This chapter presents the most widely known models of McCall, Boehm, Dromey, ISO/IEC 9126, and ISO/IEC 25010. In software quality engineering, measurement is a pivotal concept. The evaluation of software product quality is important to both the acquisition and development of software. Quality design depends on requirements and their completeness, feasibility, and quality. The chapter analyzes conflicts that can appear within the basic change control process of a software project. The chapter presents four diagrams that help the potential user of Software Quality Implementation Model (SQIM) to map this model to a development process of the most popular choice.
IEEE Approved Draft Standard for Software Quality Assurance Processes	This standard establishes requirements for initiating, planning, controlling, and executing the Software Quality Assurance processes of a software development or maintenance project. This standard is harmonized with the software life cycle process of ISO/IEC 12207:2008 and the information content requirements of ISO/IEC/IEEE 15289:2011.
IEEE Draft Standard for Software Quality Assurance Processes	This standard establishes requirements for initiating, planning, controlling, and executing the Software Quality Assurance processes of a software development or maintenance project. This standard is harmonized with the software life cycle process of ISO/IEC 12207:2008 and the information content requirements of ISO/IEC/IEEE 15289:2011.
IEEE Standard for Software Quality Assurance Processes	Requirements for initiating, planning, controlling, and executing the Software Quality Assurance processes of a software development or maintenance project are established in this standard. This standard is harmonized with the software life cycle process of ISO/IEC/IEEE 12207:2008 and the information content requirements of ISO/IEC/IEEE 15289:2011.
IEEE Trial-Use Standard--Adoption of ISO/IEC TR 15026-1:2010 Systems and Software Engineering--Systems and Software Assurance--Part 1: Concepts and Vocabulary	This trial-use standard adopts ISO/IEC TR 15026-1:2010, which defines terms and establishes an extensive and organized set of concepts and their relationships for software and systems assurance, thereby establishing a basis for shared understanding of the concepts and principles central to ISO/IEC 15026 across its user communities. It provides information to users of the subsequent parts of ISO/IEC 15026, including the use of each part and the combined use of multiple parts. Coverage of assurance for a service being operated and managed on an ongoing basis is not covered in ISO/IEC 15026.
IEEE Draft Trial-Use Standard for Adoption of ISO/IEC TR 15026-1:2010 -- Systems and Software Engineering -- Systems and Software Assurance -- Part 1: Concepts and Vocabulary	This Technical Report defines terms and establishes an extensive and organized set of concepts and their relationships thereby establishing a basis for shared understanding of the concepts and principles central to ISO/IEC 15026 across its user communities. It provides information to users of the other parts of this International Standard including the use of each part and the combined use of multiple parts. Coverage of assurance for a service being operated and managed on an ongoing basis is not covered in this International Standard.
Supply Chain Quality Integration: Antecedents and Consequences	This study extends quality management from an individual company perspective to a supply chain perspective. We propose a concept of supply chain quality integration (SCQI) that consists of internal, supplier, and customer integration for quality improvement, and develop a model that specifies the relationships among competitive hostility, the organization-wide approach to quality, three types of SCQI, and quality-related performance. We test the model using data collected from 291 high-performance manufacturing plants from ten countries. The results indicate that competitive hostility has a positive effect on the organization-wide approach to quality, and that both have positive effects on SCQI. In addition, internal quality integration significantly enhances external quality integration with both suppliers and customers. Further, internal quality integration significantly improves all quality-related performance (i.e., product quality, cost, delivery, and flexibility), and both supplier and customer quality integration significantly improve cost performance. Whereas customer quality integration significantly improves delivery performance and supplier quality integration significantly improves quality performance, only internal quality integration can improve flexibility performance. The findings reveal how different types of SCQI are related to quality-related performance and highlight internal quality integration as a core strategic resource for quality improvement. As such, they provide important managerial insights for supply chain quality managers to improve quality-related performance.
Towards a new telecommunications industry quality standard	The quality technologies developed by Bell Communications Research (Bellcore) and adopted by its owner/clients are described. At the core of the philosophy is the concept of supplier accountability for product and service quality. A brief history of quality in the former Bell System is presented. A supplier career path for today's telecommunications suppliers is shown, and the various characteristics that accompany each phase of the buyer/supplier relationship are described. Programs that enable buyers and suppliers to achieve a long-term cooperative relationship are presented chronologically according to the phases of the product life cycle. The customer/supplier quality process, which is the newest and most comprehensive of the quality technologies available in the telecommunications field, is summarized

On the basis of above figure that is showing the results generated from our system, it is shown that this Knowledge Base system is domain specific and retrieved useful and relevant information from specific domains. This is the actual purpose of web semantics based of any (KBS) Knowledge Base System. Using our proposed IRSE (Information Retrieval Search Engine), one can search relevant information from different Ontology's in case of concept existence in KB (Knowledge Base). Our proposed system is capable to answer quickly, however it also depends upon the internet speed at the time of processing. In this search, we put a query "Computer Science" for retrieval of information and on the basis of API (Application Programming Interface), it search this query from KB (Knowledge Base) where KB consists of two search engines. One of them is IEEE and second one is Springer. IEEE is returning the value in the form of XML.

Springer is returning response in the form of JSON. After getting both of results, there is a task to manage both above forms to show information in tabular form that is understandable by our Users. We make some algorithms to show only desired information based on research paper Title and Abstract. Both of these values are show in tabular form over web page.

On the basis of above figure that is showing the results generated from our system, it is shown that this Knowledge Base system is domain specific and retrieving useful and relevant information from specific domains. This is the actual purpose of web semantics based of any (KBS) Knowledge Base System. Using our proposed IRSE (Information Retrieval Search Engine), one can search relevant information from different Ontology's in case of concept existence in KB (Knowledge Base). Our proposed system is capable to answer quickly, however it also depends upon the internet speed at the time of processing. In this search, we put a query" Reverse Engineering " for retrieval of information and on the basis of API (Application Programming Interface), it search this query quickly from KB (Knowledge Base) where KB consists of two search engines. One of them is IEEE and second one is Springer. IEEE is returning information in the form of XML. Springer is returning response in the form of JSON. After getting both of results, there is a task to manage both above forms to show information in tabular form that is understandable by our Users.

CONCLUSION

In the era of science and technology semantic web is playing a very important rule in many frameworks to organized data in a meaningful way. User is also capable to communicate with each other data over different sources at the same time using these web semantic approaches. This technology gives an organized data to facilitate their researcher and clients. In this study we designed and implemented semantic web framework in such a way that researcher can find their desired data in a meaningful way and gather research related data from IEEE Explore and Springer. The results found in this study are on the basis of API (Application Programming Interface) this framework shows the working of project in a detail way. Some of other Ontology is also proposed together data in a semantically organized way. The main purpose of this study is to collect data from many web portals in semantically way to facilitate client/ users as well as researcher. In future work we can add many other data resources to achieve this different task. This study is API (Application Programming Interface) has features to provide data for different devices at the same time.

REFERENCES

[1] Kwan, I. S. 2002. A mental cognitive model of Web semantic for e-customer profile. IEEE 13th International Workshop on In Database and Expert Systems Applications, 2002. Proceedings. 116-120.

[2] Berners-Lee, T. J. Hendler, and O. Lassila. 2001. The Semantic Web. *Scientific american,* 284(5):28-37.

[3] Baker, L., T. H. Wagner, S. Singer, and M. K. Bundorf, 2003. Use of the Internet and Email For Health Care Information: Results from a National Survey. *Jama,* 289(18):2400-2406.

[4] Villamor, F., J. Ignacio, J. B. Garcia, C. A. I. Fernandez, and M. G. Ayestaran, 2011. A semantic scraping model for web resources-Applying linked data to web page screen scraping, 451-456.

[5] Horrocks, I. and P. F. Patel-Schneider, 2003. Three theses of representation in the semantic web. In Proceedings of the 12th international conference on World Wide Web, 39-47.

[6] Celik, D., and Elci, A. 2006. IEEE Discovery and scoring of semantic web services based on client requirement through a semantic search agent. *In Computer Software and Applications Conference, 30th Annual International* 2: 273-278.

[7] Namgoong, H., M. Chung, K. Kim, H. Cho and Y. Chung. 2006. Effective Semantic Web Services Discovery Using Usability. *The 8th International Conference in Advanced Communication Technology.*

[8] Bell, D., S. Cesare, M. Lycett, N. Mustafee, S. J. Taylor. 2007. semantic web service Architecture for simulation model reuse. *11th IEEE International Symposium on* Distributed Simulation and Real-Time Applications, 129-136.

[9] Taswel, C. 2008. Doors to the Semantic Web and Grid with a Portal for Biomedical Computing. *IEEE Transactions on Information Technology in Biomedicine,* 12(2): 191-204.

[10] Hitzler, P. and F. V. Harmelen. 2010. A Reasonable Semantic Web. SEMANTIC Web, 1(1): 39-44.

[11] Choi, H. and R. W. Seop. 2012. Distributed Semantic Sensor Web Architecture. *IEEE Region 10* Conference in TENCON on National Research Foundation of Korea funded by the Ministry of Education, Science and Technology, 1-6.

[12] Thangaraj, M. and G. Sujatha, 2014 . An architectural design for effective information retrieval in semantic web. *Expert Systems with Applications* , 41(18) : 8225-8233.