BUILDING A SEMANTIC WEB FOR ACADEMIC NETWORKS: A CONCEPTUAL ARCHITECTURE

K Srinivas S I Ahson T A V Murthy

Abstract

In this paper an effort has been made to propose a conceptual architecture for building an academic semantic web information retrieval for academic network. It incorporated semantic web, web services, and multi-agent technologies for effective information retrieval in an academic network.

Keywords : Academic Networks

1. Introduction

The Semantic Web is a scheme that was first introduced by Tim Berners-Lee to extend the current web from documents linked to each other in to a Web that recognizes the meaning of information in these documents. “The Semantic Web is not a separate Web but an extension of the current one, in which information is given well defined meaning, enabling computers and people to work in better cooperation” (Berners-Lee et al 2001). In the words of World Wide Web Consortium (W3C) -The goal of the Semantic Web initiative is as broad as that of the Web; to create a universal medium for the exchange of data. Initiation of weaving the Semantic Web into the structure of the existing Web is already in rapid pace. In the near future, the developments in the Semantic Web will usher significantly, as machines become much efficient to process and interpret the data.

The Internet and the World Wide Web have brought a revolution to Information Technology and a paradigm shift in daily lives of most of the people. The essential property of the World Wide Web is its universality. The power of a hypertext link is quantitatively so large it enables to link anything and everything. Web technology now has wide usage but it may not discriminate between the inarticulate text and structured text, between commercial and academic information, or among, languages, media and cultures etc (Semantic web 2001). Information varies along many axes. One of these is the difference between information produced primarily for human consumption and other for the systems. The Web has become a medium of documents for consumption rather than for data and information for the systems to be automatically processed. The objective of the semantic web is not just to connect systems, but to make the collated data and information within the systems interoperable and accessible.

Information interoperability is one of the benefits of semantic technologies. Information on the Web is becoming increasingly fragmented and varied in terms of appropriateness, timeliness, and trustworthiness. Such a tool augments the key role of a search engine.

2. The Semantic Web Overview

The Semantic Web is a scheme that was introduced by Tim Berner-Lee (Boag et al 2002) to extend the current Web from documents linked to each other, in to a Web that recognizes the meaning of information in these documents. Tim Berners – Lee published

a description of the Semantic Web Architecture in a conference talk he presented at the XML-2000. The description has garnered widespread interest within the Semantic Web community and has been cited in many occasions. Here the emphasis is on the protocols and languages that will be used as foundations to technical components.

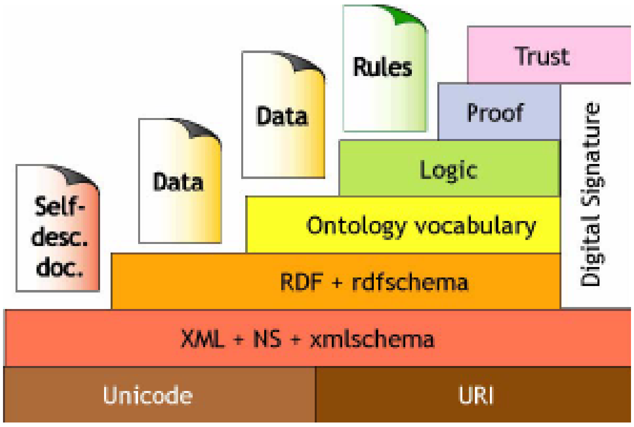


Figure 1: Semantic Web building Blocks (From Berners-Lee, XML 2000 Conference)

The bottom of the diagram shows standards that are well defined and widely accepted. The diagram provides a blueprint for a set of protocols and languages that will provide professionals with expansive capabilities for bringing about truly adaptive computing.

3. Knowledge Representation

For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. Artificial-intelligence researchers have studied such systems much before the Web was developed. Knowledge representation (Shiyong Lu et al 2002) as this technology is often called is comparable to that of hypertext before the advent of the Web: Newell 1982, has given a documentary evidence about the concept. It contains the seeds of important applications, and also possess potential to be operational when linked into a single global system.

Traditional knowledge-representation systems typically have been centralized, requiring everyone to share exactly the same definition of common concepts. The vision of Semantic Web is to extend the current web by enriching the information transmitted and accessed over the internet with well-defined meaning, thus enabling computers to do more of the work in assembling and processing data in order to turn it in to highly relevant information and knowledge. The steps taken by W3C are targeted toward filling the gap in data association and collective understanding.

Semantic Web researchers, in contrast, accept that paradoxes and unanswerable questions, that must be paid to achieve versatility. The language for the rules as expressive as needed to allow the Web to reason as widely as desired. But the expressive power of the system made vast amounts of information available, and search engines now produce remarkably complete indices to a lot of the material. The challenge of the Semantic Web, therefore, is to provide a language that expresses both data and rules for reasoning (Berners-Lee, 2001. The figure 2 shows a conceptual framework for the Semantic Web illustrating the way semantic technologies can be added to extend the capabilities of the current web.

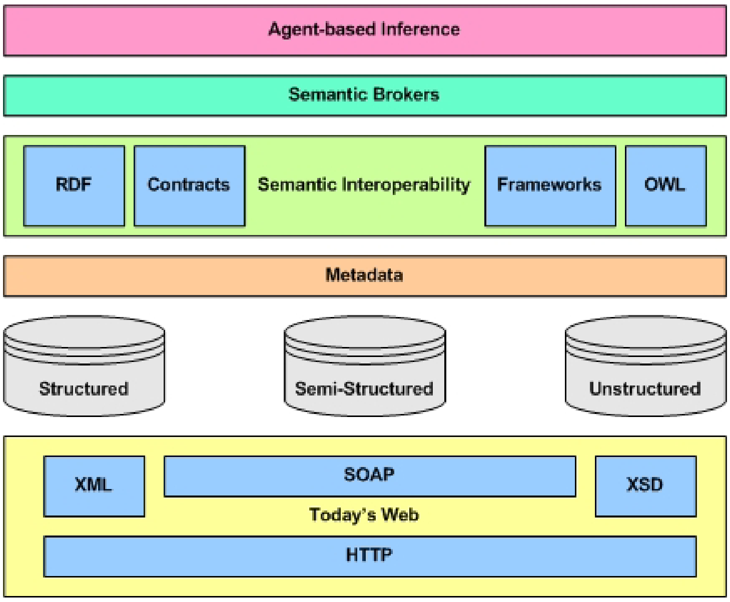


Figure 2: Semantic Web Conceptual Frame work

4. Academic Networks

The scholarly information arena offers researchers an ever-increasing array of resources. Researchers are likely to find material relevant to their subject in a variety of web-based resources: their own library's catalogue; catalogues outside their own library, such as a national or union catalogue or a catalogue or another institution that specializes in similar subjects. The current process of information access for the sake of seeking information is cumbersome and requires some knowledge of the various resources, their access mechanism, the query interface they provide, and the type of results they return. It also requires a manual comparison between the results returned from several resources and does not enable the user to move from one resource to another for further discovery and navigation.

With ever-increasing information overload, web information retrieval systems are facing new challenges for helping people not only locating relevant information precisely but also accessing and aggregating a variety of information from different resources automatically.

5. Approach

A conceptual architecture of building an academic semantic web information retrieval system is based on the following ideas:

The semantic information retrieval system concerns three main kinds of users: The "scholar" who is searching for the information, the "provider" (Academic Libraries) who is holding the information and the "interface" who is enabling the communication between the "scholar" and the "provider".

The integrated or unified management of academic contents and the related services need to be carried out through different levels including the description of capabilities and requirements, querying, discovering, selection and aggregation.

Proving a gateway to all the information needs of the scholar is also another priority. The scholar is only interested in certain parts of the resources available in the network; the personalization functionality and the integration of different applications are very much required.

6. A conceptual Architecture

A conceptual architecture for academic semantic web information retrieval system is illustrated in figure 3.

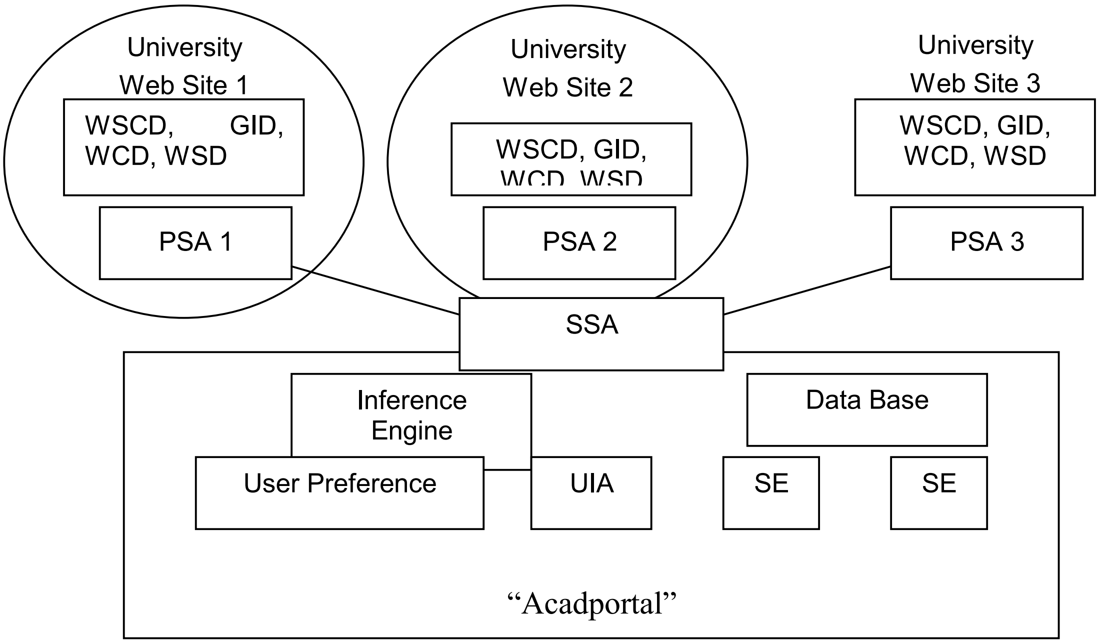


Figure 3 : A Conceptual Architecture

WSCD	:	Web Site Capability Description
GID	:	General Information Description
WCD	:	Web Content Capability Description
WSD	:	Web Service Description
PSA	:	Provider (Library) Search Agent
SSA	:	Scholar Search Agent
UIA	:	User Interface Agent

Each academic libraries (providers) describes their capabilities in what can be called as WSCD (web site capacity description) and is assigned a PSA (Provider search agent). Each Scholar describes the individual's requirements including preferences. It is assigned a Scholar search agent (SSA) and also has a user interface agent (UIA) that provides an intelligent unified interface to the user. The SSA and PSA will function as mediators between a scholar and provider by communicating with each other to fulfill the searching and accessing task.

7. “Acadportal” Functionality

“Acadportal” plays a very important role in the information retrieval architecture. It links the user to all the information from the different universities. It resites on the scholar’s own desktop or local server and is designed to satisfy scholar’s personal information requirements and to be mastered freely by the user. The information can be shared by other scholars with proper authority.

“Acadportal” is composed of three main functional componets: Core component, Scholar component and Provider component.

The core component provides basic support for semantic technologies and information management. It consists of “knowledge ware house (KW)”, knowledge management (KM)” and “interface engine (IE)”.

For scholar the portal brings together a variety of necessary information form different resources automatically or semi automatically for the user. As a provider, the contents and services of “acadportal” can be consumed by humans aswell as machines. The humans can be scholars or other permitted persons and the machine can be local or remote. The interface for browsing, searching and facilitating web contents are services need to be provided. The functionality of “Acadportal” is illustrated in Figure 4.

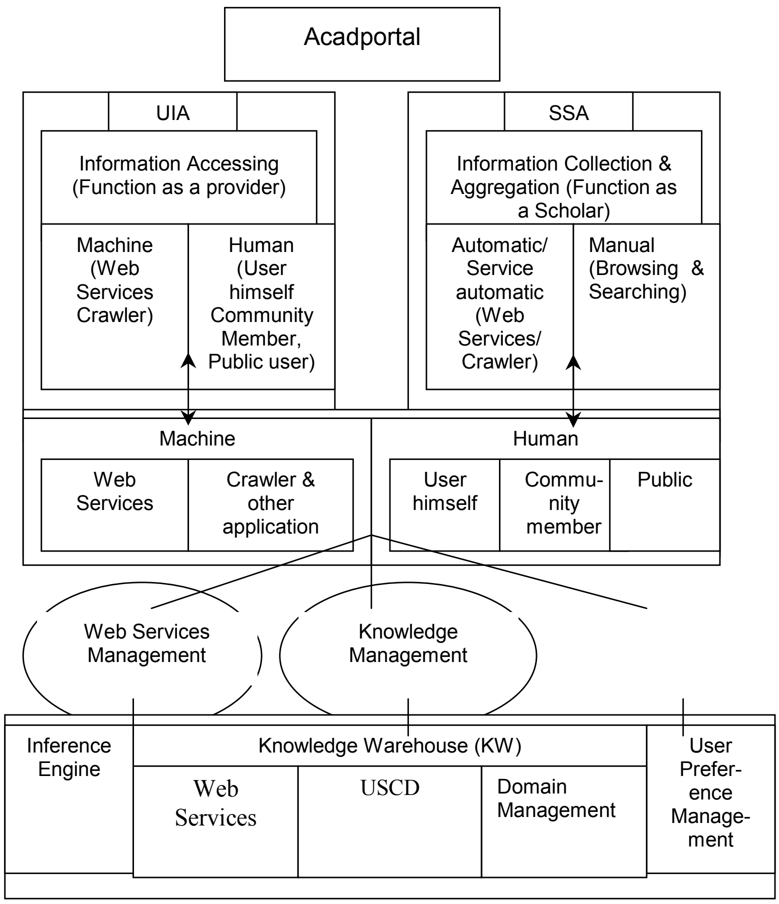


Figure 4 : Structure of “Acadprotal”

8. Communication Interfaces

In order to fulfill the information retrieval task, the interfaces between providers and scholars including query language and protocol for communicating those queries need to be defined. As Semantic Web information is based on RDF to represent data, a standard interface for querying and accessing RDF data is ideal for the interoperability between heterogeneous environments. The W3C RDF data accessing working group (DAWG) has published their working drafts RDF query Language SPARQL and SPARQL protocol that are expected to be standards in this field. The RDF query language SPARQL expresses queries over RDF graphs and SPARQL protocol for RDF defines a protocol for communicating those queries to an RDF data service. The applications can access and combine semantic web information across the academic network by combining SPARQL query language and protocol for RDF. The proposed architecture is designed for any reasonable communication interface, but at present it is confined to use the SPARQL RDF query language and SPARQL protocol as the communication interfaces between the provider and scholar.

10. Process Flow

The process of the proposed information retrieval system emphasizes the following aspects:

When the scholar sends his request for relevant information inside the knowledge warehouse the searching will be carried out first for the authenticity and only when the confirmation comes from the database the searching will start. As one can tend to repeatedly and frequently use a certain amount of information.

11. User requirements

The user requirements are reflected by the preferences, profile and constraints along with a query. A user interface, which enables the input of all these information. The formal scholar query is composed of three types of element fields: user preferences (Ups), content query(CQ) and web service query (SQ) . Even if the user does not explicitly describe their requirements on Web services for each query, searching for Academic web service potentially relevant to him/her will automatically be carried unless he/she explicitly refuses and seaching.

12. Key Capabilities of Semantic Technologies

New innovations and approaches in the form of web semantics are emerging in world wide web, connecting high quantum of information for machine manipulation and processing. These innovations will facilitate information interoperability for relevant information within the World Wide Web. Table 1 details the key capabilities of semantic computing and its impact for stakeholders.

Table 1: Computing Capabilities Assessment (Adapted by Richard Murphy)

Capability	Purpose	Stakeholders	Impact	Take-away
Semantic Web Services	Provides flexible look and discovery and schematic transformation	System Developers and System Integrators	Reduced friction in web services	More automated and flexible data connections
Information Integration and Interoperability	Reduces integration complexity	Data and Metadata	Reduced cost to integrate heterogeneous data sources	Increased interoperability at improved speed and reduced cost
Intelligent Search	Provides context sensitive search, queries on concepts and specific filtering	Business and Technology Managers, Analysts and individuals	Reduced human filtering of search results, more relevant searches	Increased search accuracy translates into greater productivity

13. Ontology Considerations

The description of Academic Networks capabilities and the management of data in “acadportal” must be based on formally defined vocabularies in order to make them machine understandable and processable. Ontology is used to formally define terms and the relationships between them. A wide and deep ontology for categorization is necessary and narrow and deep ontolgoies are also needed for the user’s specific interests such as research topic etc.

With ontology pages on the Web, solutions to terminology problems begin to emerge. Pointers from the page to ontology can define the meaning of terms or XML codes used on a Web page. This kind of confusion can be resolved if ontologies provide equivalence relations. Ontologies (Obrst Leo et al 2004) can enhance the functioning of the Web in many ways. They can be used in a simple fashion to improve the accuracy of Web searches—the search program can look for only those pages that refer to a precise concept instead of all the ones using ambiguous keywords. More advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules.

14. Semantic Web Agents for Academic Networks

The real power of the Semantic Web will be realized when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programs. The effectiveness of such software agents (Ankolekar et al 2001) will increase exponentially as more machine-readable Web content and automated services (including other agents) become available. The Semantic Web promotes this synergy: even agents that were not expressly designed to work together can transfer data among themselves when the data come with semantics.

When agents are equipped with intelligence and mobility, the conventional client/server-computing paradigm might be replaced by an agent-based distributed computing paradigm, in which agents can migrate from one site to another, carrying their codes, data and intelligence and fulfill their missions autonomously and intelligently.

An important facet of agents’ functioning will be the exchange of proofs written in the Semantic Web’s unifying language. The functionality of Proof and Trust layers is highly dependent on creation of accurate and trustworthy metadata.

Another vital feature will be digital signatures, which are encrypted blocks of data that computers and agents can use to verify that the attached information has been provided by a specific trusted source. Agents should be skeptical of assertions that they read on the Semantic Web until they have checked the sources of information.

15. Integration of Web Services and Web Contents in Academic Networks

Conventional web contents target at human consumption and are published with standard languages such as HTML, which can be accessed through client browser applications. Standard HTTL protocol is used for the communication between a web server and a client. Web services on the other hand target at machine consumption and are applications, which can be realized at heterogeneous systems.

However in the semantic web information is marked up with metadata and can be manipulated by autonomous agents on behalf of their users. So web contents are in the process of becoming data with well-defined meaning that can also be consumed by machines. Since they target the same scholar , web services and web contents have the necessary common ground to be managed together in a Academic web information retrieval system.

16. Conclusion

In this paper an effort has been made to address main aspects of semantic web information retrieval system architecture. The semantic web is not merely the tool for conducting individual tasks that have discussed so far. In addition, if properly designed, the Semantic Web can assist the evolution of human knowledge as a whole. The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together. The “acadportal” aims at constructing a fully personalized user’s local web portal, which is adapted to user preferences and satisfies all the requirements of a user’s web usage. Currently the assumption is that all the portals, users and agents in a community agree on a common ontology that involved and use it to represent the semantics of web portal capabilities and the web services

Semantic Web communities will appear and grow first then the interaction and interoperation among different communities will finally interweave them. The main idea of a portal for academic network aims at constructing a fully personalized web portal, which is adapted to user preferences and satisfies the requirements of a user data usage.

17. References

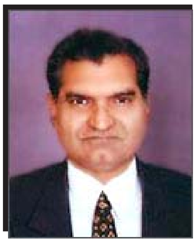
1. Ankolekar, A., et al., (2001), “DAML-S: Semantic markup for web Services.”, : The First Semantic Web Working Symposium, PP. 411-430, Heidelberg: Springer-Verlag Heidelberg.
2. Berners-Lee, T., J. Hendler and O. Lassila : Semantic Web Scientific American, May 2000.
3. Berners-Lee. T. (2001) Services and Semantics : Web Architecture <http://www.w3.org/2001/04/30-tbl>.
4. Berners-Lee,Tim. “Semantic Web: Where to Direct Our Energy?” Invited talk at the International Semantic Web Conference (ISWC), October 2003.
5. Berners-Lee, T. (2001), Weaving the Web: The original Design and Ultimate Destiny of the World Wide Web, Harper Collins Publishers inc., New York, NY.
6. Bedford, Denise. “Charter Statement of Taxonomy and Semantics Special Interest Group”. 2004

7. Boag .S, Chamberlin. D, Fernandez. F.M, Florescu. D, Robie. J, Simeon.J and Stefanescu. M Xquery 1.0: An XML query Language. <http://www.w3.org/TR/xquery 30 April 2002, W3C Working craft..
8. Erdmann, M., Studer, R. (2001). "How to structure and access XML documents with ontologies." Data and Knowledge Engineering, 36(3), 317-335.
9. E. Prud' hommeeaux A. Seaborne, SPARQL, Query Language for RDF, April 19,2005. http://www.w3c.org/TR/rdf-sparql-query/.
10. Fromm, Kenneth, and Jeffrey T.Pollock. "Semantic Computing's Building Blocks" Enterprise Architect, April 15,2004 (Vol.2, No.8).
11. H.Yu, T. Mine, and M. Amamiya. Towards Automatic Discovery of Web Portals – Semantic Description of Web Portal Capabilities-, In Semantic Web Services and Web Process Composition: First International Workshop, SWSWPC 2004, LNCS 3387/2005, pages 124-136,2005
12. K.G. Clark. SPARQL Protocol RDF, January 14,2005. http://www.w3c.org/TR/rdf-sparql-protocol/.
13. Massimo Paolucci, Katial Sycara, Takahiro Kawamura. Delivering Semantic Web Services. In Proceedings of Twelves World Wide Web Conference, WWW2003, pages 111-118, May 2003.
14. Newell, The Knowledge Level, Artificial Intelligence 18(1) (1982) 87 – 127
15. Obrst, Leo J; Howard Liu and Robert Wray. "Ontologies for Corporate Web Applications: AI Magazine (Fall 2003); PP. 49-62.
16. Obrst, Leo. "Ontologies for Semantically Interoperable Systems." MITRE, Centre for Innovative Computing & Informatics. Presentation to the KM.gov Semantics Interoperability Community of Practice. April 14,2004.
17. R. Guha, R. MacCool, and E. Miller. Semantic Search. In Proceedings of WWW2003, pages 700-709,2003
18. Shiyong Lu, Ming Dong , Farshad Fotouhi "The Semantic Web: Opportunities and Challenges for Next-Generation Web applications" July 2002.
19. W3C. "Resource Description Framework." World Wide Web Consortium. August 2004.
20. W3C. "Web Ontology Language (OWL). "World Wide Web Consortium. August 2004.

About Authors

Sh. K Srinivas is a Systems Analyst at National Institute of Educational Planning and Administration, New Delhi.

Prof. S I Ahson is a Professor & Head, Department of Computer Science in Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi.



Dr. T.A.V. Murthy is the Director of INFLIBNET and President of SIS. He holds B Sc, M L I Sc, M S L S (USA) and Ph.D. He carries with him a rich experience and expertise of having worked in managerial level at a number of libraries in many prestigious institutions in India including National Library, IGNCA, IARI, University of Hyderabad, ASC, CIEFL etc. and Catholic University and Case Western Reserve University in USA. His highly noticeable contributions include KALANIDHI at IGNCA, Digital Laboratory at CIEFL etc. He has been associated with number of universities in the country and has guided number of PhDs and actively associated with the national and international professional associations, expert committees and has published good number of research papers. He visited several countries and organized several national and international conferences.
Email : tav@inflibnet.ac.in