Glimpse of Semantic Web and its Application Across Industries

Ravi Shankar Jha ***Amartya Roy***

Abstract

The most used tools to access data existing in web are the search engines. Traditional search engines are inefficient in the sense that it produces results that are mostly left for humans to interpret and consume and mostly not specific to the user's search context. Semantic web creates machine/application readable data. Semantic web extends the traditional search engines and helps them produce context specific and meaningful search results by reading and assimilating the data before producing the search results. Also, Semantic web has been helping in externalizing data for sharing and reusability across users and applications. This paper explains some of the concepts of semantic web, and touches upon some of its industry applications with focus on knowledge/information management and sharing.

Keywords: Semantic Web, Knowledge Management, Information Management, Library Management, Ontologies, RDF, Search, Instance Data

Introduction

How do we find information – from the internet – we head straight to a search engine and type our question and from there starts the real search – sorting through thousands of pages listing the words we typed to pinpoint the exact thing we needed. Indeed search engines index much of the Web's content, but they have little ability to identify pages that a user truly wants or needs. Use of self-descriptions and similar techniques during web development and content creation can make context-understanding possible and can selectively single out what users want.

Tim Berners-Lee, inventor of World Wide Web, proposed the idea of Semantic Web. He proposed that the Web can be made more intelligent and intuitive about how to address a user's needs by creating an interlinked web – the Semantic Web.

The Wikipedia defines Semantic web as an extension of the Web through standards by the World Wide Web Consortium (W3C). Semantic Web process data in effective & accurate ways, and provides interoperability [1]. Here data is given well-defined meaning enabling humans and computers to work in tandem.

Semantic Web may very well be the next major thing in linked information. It enables pieces of data to be interlinked in a form that would be read and understood by computers enabling them to perform activities by inferring on behalf of the user.

10th Convention PLANNER-2016
NEHU, Shillong, Meghalaya, 09-11 November, 2016

Why Semantic Web?

To explain the importance of semantic web, we need to go back to the evolution of internet (Web 1.0) dating to PCs and Desktop age, all the information was stored in these in the form of documents. When Web 1.0 provided the ability to reference the documents via hyperlinks, Web 2.0 (e.g. Facebook, Wikipedia) let the users read & update over the web. However, it took a physical person/users to identify relation between various information over the web. Here comes the advantage of Semantic Web, which builds intelligence for computers to understand and interpret various information over the web. The Fig on the right depicts the evolution of Sematic web (Web 3.0)

- **URI – Uniform Resource Identifier**

A URI is nothing but provides unique name for items present in a statement across the entire internet. Therefore, each element/component of a statement comprises of "Subject, Predicate and Object" – holds a URI to affirm its identity throughout the era of WWW. Notably, it eliminates the possibility of naming conflicts and ensure that two items are the same or not and can also provide path of additional information.

- **Resource Description Framework**

A language which utilizes three URIs (a triple) is called Resource Description Framework (RDF). XML serialization of RDF is developed by World Wide Consortium (W3C) [3]. RDF XML is considered to be the standard interchange format for RDF on the Semantic Web, however, it is not the only format.

Consider a simple sentence for example

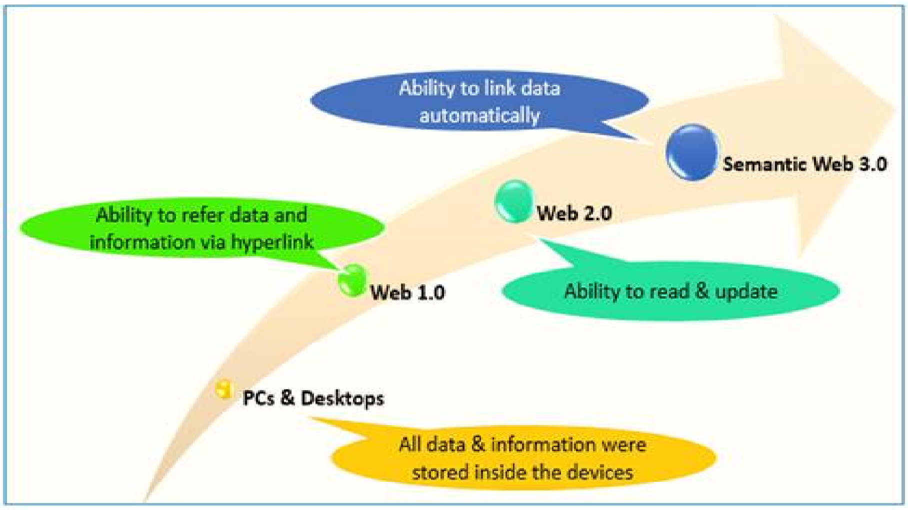


Figure1: Evolution of Semantic Web

Building blocks of Semantic Web

There are four major building blocks that essentially make up what we call Semantic Web [2]. Below diagram depicts various components of Semantic web along with brief description for each. They are -

John Smith is the creator of the resource http://www.johnsmithorg.come/Home

The above statement is represented in RDF/XML as:

<rdf:RDF>

<rdf:Description about=" http://www.johnsmithorg.come/Home">

<s:Creator>John Smith</s:Creator>

</rdf:Description>

</rdf:RDF>

Using RDF the information will maps directly and unambiguously to a knowledge model, which is decentralized, and for which many generic parsers already available. In other words it is adopting RDF for representing the data, identifies those bits of data that are semantics of the application, from the ones that are not.

❖ **WEB Ontology**

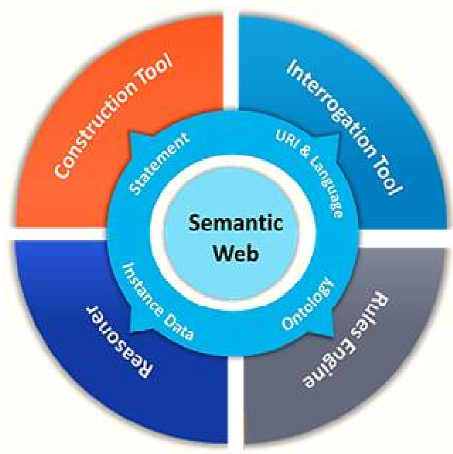
Ontology primarily comprised of statements that define concepts, relationship (Parent-Child or transitivity) and constraints. In other words, it could be same as object oriented class diagram to capture the information of given application and build model domain knowledge

For the Semantic Web to become expressive enough to address wide range of situations, there is a necessity of a powerful logical language to make inferences. Inference is nothing but deriving new information from already known data. For example, querying is a form of inference, which searches for a particular information from a mass of data on web.

Web Ontology Language (OWL) extends the RDFS (Resource Description Framework schema) vocabulary with additional resources that could help to achieve more expressive ontologies and intelligence for the Web.

❖ **Instance Data**

Instance data is the statements containing information about specific instances rather than generic concepts at large. For example, *Activity* is an instance whereas *Transaction* is a class/concept. We can perhaps draw a parallel to object oriented concepts of class and object. Also, Instance data forms the bulk of the semantic web.



Statement	• Foundation of semantic web primarily composed of subject, predicate and Object. • Defines information structure, specific instances/objects and limits on the structure
URI and Language	• Uniform Resource Identified provides a unique name for items contained in a statement by offering separate identity for Subject, Predicate and Object element • Statement expressed in accordance with Semantic web language consist of set of key words which offers instruction to various semantic tools • Provide degree of complexity and semantic expressiveness
Ontology	• Comprise of statements that define concepts, relationship and constraints and build information domain model by leveraging existing domain models.
Instance Data	• Contain information related to specific instances rather than generic concept analogous to class - objects/instances model. Also, it forms the bulk of semantic web used across the intranet
Construction Tools	• Facilitate the construction and integration of a semantic web through the creation or import of statements for the ontology and instances
Interrogation tools	• Enable navigation through the Semantic Web through web services or other mode to achieve requested response . • Few Interrogation methods such as simple graph navigation to search to a full query language.
Reasoner	• Integral in adding inference to Semantic Web by creating logical additions which offers classification and realizations • Classification and realization help to populate class/concept structure, establish relationship (using various properties) to augment the inference ability of Semantic web
Rules Engine	• Act as powerful dimension to knowledge constructs by merging ontologies and other larger logical tasks and provide knowledge base consume by both applications and human being

Figure 2: Building blocks of Semantic Web

Unleash the power of Ontology

Ontology is one of the fundamental building blocks of Semantic Web. It is widely considered as the backbone to support various types of information management including retrieval, storage, and sharing on web. The core objective of Ontology is to capture (business logic plus data) and model the domain knowledge in order to provide a common understanding of the domain by building unified domain knowledge. In other words, an ontology defines the terms and concepts (meaning) used to describe and denote an area of knowledge. For example, in Insurance business some application call out Policy record as "Policy" entity however other application calls it as "Contract". So from end user's perspective, it appears to be two different entities however both have by and large the same meaning (concept). In such situations, Ontology will come to rescue and define a common understanding (for the insurance domain) which will facilitate exchange of information, interoperability among applications and many more. Also, Ontology will help to decouple knowledge model (*marking it as application agnostic*) from the application to unleash the potential of sharing data, reuse between user and applications. In this design consideration, all the core business logics and data will reside in centralized knowledge model. All interacting applications could leverage API economy to create interface to interact with own application specific knowledge models using various mapping capabilities provided by Ontology Web Language

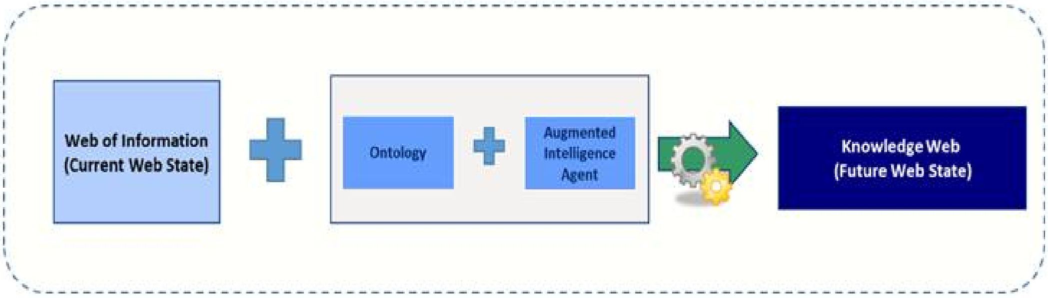


Figure 3: Transition from Current Web to Future Web [4]

Essential Steps to Build Ontology

- Define the scope of Ontology, especially with respect to domain
- Establish methods to collate the details including the scope of re-use in terms of leveraging existing Ontologies
- Identify the important/crucial terms associated with the domain
- Define class and associated hierarchy or build concept and associated relationship
- Define property of classes and constraints
- Define the mapping of classes
- Organize the classes in a taxonomic (sub-super class) hierarchy
- Define the integration mechanism
- Review, Validate and Instantiate the class/ concept

Leading Ontologies Tools/Editors

Quite a few Tools/Editors that make building Ontology easy are now available in the market.

Here we have described four leading ones [5].

- **Protégé 3.4** [6] is a knowledge based ontology editor with a well-defined GUI. It enables the construction of domain ontologies, provides better flexibility for meta-modeling and also provides custom data entry forms for data input. It is usually targeted at conceptual modeling and knowledge engineering without having to think about or know the syntax of the output language.
- **IsaViz** [7]: Enables browsing and authoring of RDF models as graphs in a visual environment. This tool is offered by W3C Consortium. IsaViz was developed by Emmanuel Pietriga. IsaViz exports RDF/XML, N-Triples, Portable Network Graphics (PNG) and Scalable Vector Graphics (SVG) and imports RDF/XML and N-Triples.
- **Apollo** [8] is a user-friendly knowledge modeling application based around primitives, such as classes, instances, functions, relations etc. Based upon the internal model of the OKBC protocol an internal model is built as a frame system.
- **SWOOP** [9] is a Web Ontology Language browser and editor. SWOOP offers various presentation syntax views for OWL validation.

It provides a multiple ontology environment and has inbuilt support for reasoning with facilities for easy comparison, merger and editing of ontologies. Description Logic-based definitions, properties and instances of ontologies can be easily compared using SWOOP.

Suggested Approach for Semantic Web Application Implementation

In order to implement semantic web application, it would be preferred to classify the overall piece of work/program across multiple phases and under each phase, define set of key activities which need to carry out. In the suggested approach, we classify program into 5 phases commencing with current state assessment followed by envision & roadmap, Design & Build, Integration & Data Migration and Validation & Rollout. Under each phases, broadly outline set of key activities which could be further slice down to a detailed work breakdown structure.

In addition to set of key activities and phases mentioned above, it is necessary to have key enabling functions in place and their continuous involvement, such as well-defined governance structure, Product/Technology Vendor support, blessings from leadership, Infrastructure support and well managed community of practices, to run the program seamlessly from end to end perspective.

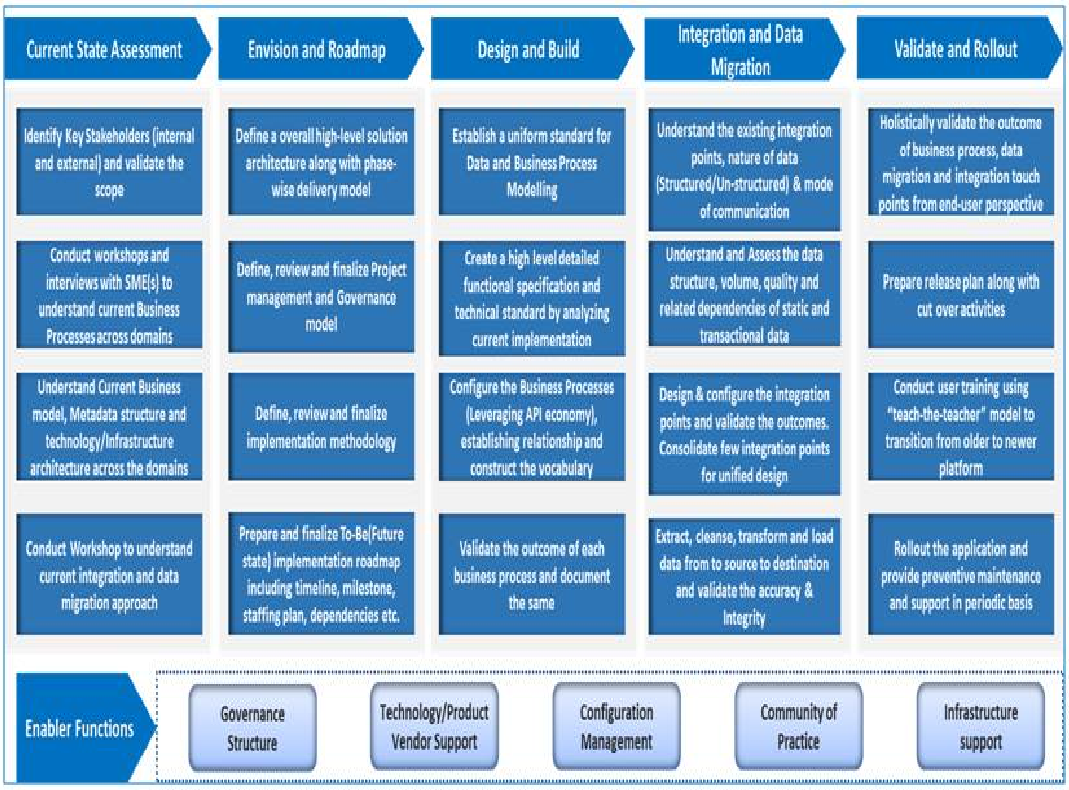


Figure 4: Suggested Implementation approach for Semantic Web Application

Adoption Issues

Here are few inhibitor or adoption issues that may crop up during implementation of Semantic web in terms of people, process and technology

- ❖ Dearth of unified framework for data organization
  - ♦ Congregation of heterogeneous systems and associated data format
  - ♦ Data redundancy, each system working in silos leading to inconsistent behavior
  - ♦ Lack of uniform and standard semantic Ontology vocabulary to communicate with other semantic ontology
- ❖ Technical barriers to adoption arise from the complexity of RDF/XML: (Resource description Framework)
  - ♦ Information written in this way is much more verbose than the attribute value pair in XML
  - ♦ The RDF/XML document has complex syntactic rules
  - ♦ Because of complexity, the RDF parsers are 5-20 times slower than XML parsers. Also, RDF/XML parsers are not widely available
- ❖ Classifying information is inherently hard: When the information is coming from so myriad sources including many groups and people, classifying even the objective information becomes hard attributed to (Vastness and Vagueness)
- ❖ Organization Change Management
  - ♦ Lack of awareness about semantic web and people are still running behind manual or document era approach rather than embracing new semantic web based approach
  - ♦ Limited commitment from the organization leadership in offering guidance on talent enablement, logistic and infrastructure planning and clear vision and roadmap to evolve and adapt sematic web technology

Applications of Semantic Web Principles across industries

Semantic Web is used in various industries and they are reaping huge benefits already. Let's look at few of how adoption of Semantic Web has addressed pain-points in various industries.

Case Study 1: Semantic Web in Library and Information Management Systems

Type of Industry	Library and Information Management
Key Challenges	♦ Too much information to sort through ♦ Information distributed in heterogeneous media ♦ No quick and easy way to identify the exact piece of information ♦ Same word may have completely different meanings under different contexts
Semantic Web Adoption	A framework/application that is enabled with Semantic Web features will bring a lot of enhanced capabilities to the end users of the system. The capability of the system that enables it to read, understand and infer from information assets written in natural languages will aid in the correct context specific data representation. When information retrieval systems can understand the meaning of items within given context it will be able to provide greater retrieval precision during end user's search or information-seeking tasks. The primary aim of Semantic Web is to enable information to be machine-interpretable in order to derive meaning.A fully enabled system is reusable and can contribute to the betterment of information and library management. Integrating library collection metadata with web based and publicly available information sources, like Google or Amazon to provide context specific and meaningful results.
Benefits	♦ Standard framework for information management and data description ♦ Help computers/applications understand the contents of information assets written in natural languages (English, Spanish, French, etc.) ♦ Assist the user in better formation of search query ♦ Efficient conceptual search leading to time and effort saving ♦ Students, scholars and researchers benefit immensely by having context specific information ♦ Store and manage domain specific information and knowledge ♦ Help the reuse of the framework by organizations and libraries world wide ♦ Infer new knowledge from already existing data and information assets
References	International Conference on Digital Libraries (ICDL) 2013, Shantanu Ganguly, P. K. Bhattacharya[10]

Reference Architecture for Implementation of Semantic Web in Library and Information management systems.

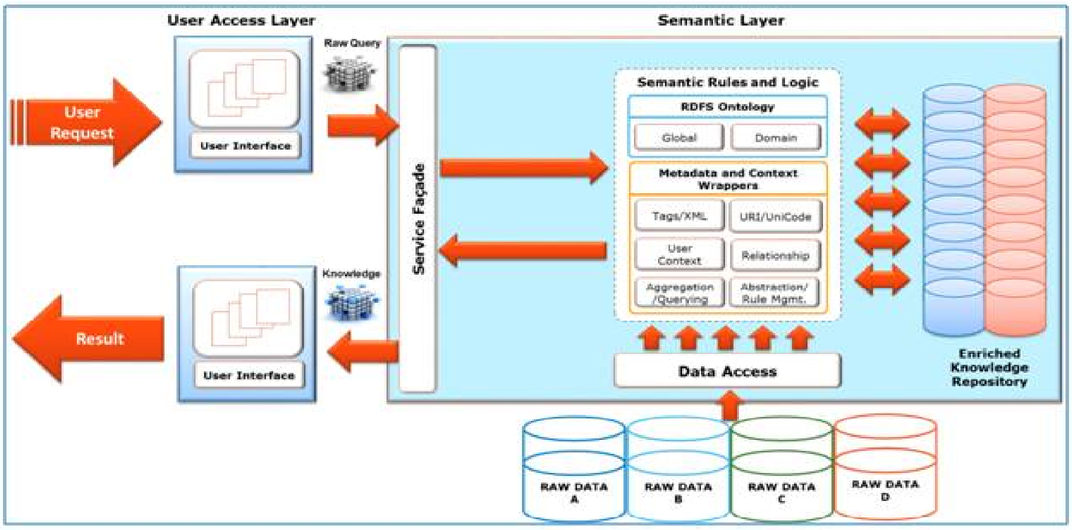


Figure 5: Reference Architecture for Semantic Web Application

Case Study 2: Semantic Web in Healthcare

Type of Industry	**Healthcare**
Key Challenges	Biomedical data are obtained from various sources in a hospital. There are a lot of reference points For example, claim data, provider data, policy data, medical codes etc. These biomedical data are not always stored in a centralized database in a hospital. There could be many Health Information Systems (HIS) like Radiological Information System (RIS), Laboratory Information Management System (LIMS) and these HISs are not interconnected. The medical data won't be structured or in a unified form to be process across various medical systems.Non uniform data makes data retrieval tedious which are sitting in multiple HISs. This also makes its difficult in biomedical research to improve healthcare services
Semantic Web Adoption	Semantic Web creates a framework for the biomedical data to be in a unified & standard format, which can be used by more than one Health Information systems. One such structured data in claims are the International Classification of diseases (ICD -10). ICD-10 is a structured format, which helps claim assessors to use the uniform disease codes across multiple hospitals. To address the inconsistency in process data in the radiology, Radiological Society of North America (RSNA) department developed

	RadLex with the goal unifying the variety of terminologies that radiologists use into one unified lexicon to serve all their needs.
Benefits	♦ Creates a standard and unified framework for data organization ♦ Enable easy transfer of information across different formats ♦ Monitoring financial activity by legal institutions to regulatory standards made easier ♦ Improved risk management due to transparency of instruments
References	http://semanticday.makolab.pl/sites/default/files/Semantic%20 Web%20for%20Banking%20and%20Insurance%20Industries_0.pdf [12]

Best Practices

- ❖ Build enriched ontology / vocabulary repositories of educational material along with demo applications from end user perspective
- ❖ Establish standard design constructs for building ontologies including naming conventions, mapping, integration, link to related course material etc.
- ❖ Build a standard reference architecture for Semantic web application spanning across its major components - data interface, search engine, persistence layer, remote data sources, integration service and few more
- ❖ Leverage tools (in-house or open source) not only to ascertain metrics about the completeness and reusability of the source metadata, but also generate mapping quality metrics to get feedback for improving the defined mapping
- ❖ Provide adequate design consideration especially on 4 key aspects – Scalability, Interoperability (Flexibility), Re-usability and User experience
- ❖ Using a web-based graphical user interface to define mappings from the provider's source metadata format to a common/unified format used in the service. The creation of the mappings is an iterative/continuous process that comprises of validation and refining of mapping as it evolve over a period of time
- ❖ Formulate a working group or community of practice on different areas such as data interface, metadata mapping and relationship building etc.
- ❖ Develop a framework to evaluate (in terms of cost-benefit analysis) list of tools including open source applications available internally or at marketplace for ontologies editor
- ❖ Build a well-defined governance structure to control and manage changes in the semantic web structure

Semantic Web and Way forward

There has been a significant increase in the adoption of Semantic Web technologies across the industries over the recent years. Majority of the adoption has resulted from an emerging need to integrate, establish relationship/linkage and analyze data across multiple databases, applications and social communities. It has been fascinating to witness the breadth of solutions that have been implemented to meet these needs.

With the increasing speed of digitalization across all domains - education, healthcare, manufacturing and others, Semantic web will certainly augment the power of web in terms of carrying/executing user tasks, making meaningful information through reasoning and inferences to ease decision making process. Also, sematic web will pave the path for transformation from the age of information to the age of knowledge by collating the meaning of vast amount of data, compiling them, identifying the hidden relationship, establish linking of data and share the compiled canned knowledge to drive the decision.

Success for the semantic web will depend upon the slope of the adoption curve because people need to understand associated technology to realize its full potential. The current pulse on ground is that concept of Semantic web found itself not only difficult for people but also for organization. In order to achieve the full blown success of Semantic Web it is equally essential to educate people by working together to accept its standards (GRDDL, RDFa, OWL, SPARQL, and others) and its naming and tagging ontologies. The real power of the semantic web will be realized when people create many programs that collect available data on Web created through diverse sources, process the information and exchange the results with other programs

In coming times, there will certainly be an increase in adoption of Semantic Web technologies within both industry and academic settings including Library Science. It is most likely that many implementations will focus on light-weight approaches (in terms of web agent or artificial intelligence agent consuming information on web and establishing knowledge using machine readable format) that enable the linking of data across silos, while a few large-scale efforts will certainly depend upon the use of heavy-weight ontologies as a top-down approach to data integration and sharing. Additionally, we are also expecting a continued convergence of the Semantic Web, Big Data and social networking, thereby enabling a more collaborative approach to future implementations.

A semantically intelligent integrated system will provide functioning of any information management system like a library with much more efficiency. Thanks to Semantic web libraries as we know today are set to be transformed into information hubs that will cater to the world hungry for access to meaningful knowledge assets. Libraries and information repositories will collaborate to create information super-highways that will provide seamless access to the users into the ocean of connected knowledge; removing the barriers of physical and logical isolation. This will accelerate the process of research and development and will be a key contributor to the nation's development at a time when information is the key driving force in the world. The filed is open and we are limited only by our own reservations and reluctance to adapt and adapt fast.

References

1. https://en.wikipedia.org/wiki/Semantic_Web
2. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. Scientific American, May 2001

3. https://en.wikipedia.org/wiki/Resource_Description_Framework

4. International journal of Web & Semantic Technology (IJWesT) Vol.1, Num.3, July 2010 12

5. A Comparative Study Ontology Building Tools for Semantic Web Applications http://airccse.org/journal/ijwest/papers/0710ijwest01.pdf

6. http://protege.stanford.edu

7. http://www.w3.org/2001/11/IsaViz/Overview.html

8. http://apollo.open.ac.uk/index.html

9. Aditya Kalyanpur, Bijan Parsia, Evren Sirin, Bernardo Cuenca Grau, James A. Hendler: "Swoop: A Web Ontology Editing Browser". J. Web Sem. 4(2): 144-153 (2006).

10. International Conference on Digital Libraries (ICDL) 2013, Shantanu Ganguly, P. K. Bhattacharya

11. Utilizing Semantic Web Technologies in Healthcare

12. Semantic Web For Banking and Insurance Industries

About Authors

Mr. Ravi Shankar Jha, Lead Consultant, Infosys Limited.

Mr. Amartya Roy, Project Manager, Infosys Limited.