www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3 Issue 11 November, 2014 Page No. 9129-9132

A Study on Semantic Web Languages and Technologies

Anthony Narzary, Gypsy Nandi
Department of Computer Science & Engineering and IT
School of Technology
Assam Don Bosco University
Guwahati-781017, India
anthony.narzary@gmail.com

Department of Computer Science & Engineering and IT
School of Technology
Assam Don Bosco University
Guwahati-781017, India
gypsy.nandi@dbuniversity.ac.in

***Abstract* - The World Wide Web (WWW) has changed the way people communicate with each other, how information is spread and retrieved. The word semantic web includes techniques that promise to dramatically improve the current WWW and its use. The Semantic web area has seen rapid development in current era with the improvement of technologies every now and then. The Semantic web is propagated by the World Wide Web consortium (W3C), an international standardization body for the web. The main purpose of the Semantic web is to help users locate better, organize and access the information on the web. The language used should be a natural language so that users find it comfortable to deal with any web pages or Applications. This paper gives us the overview of web languages and technologies used now-a-days in present era.**

Keywords—XML, Semantic, DAML, RDF, UDDI ,W3C

I. INTRODUCTION

The word Semantic Web is initiative by Tim Berners-Lee, the very person who invented the WWW in the late 1980, a vision where the meaning of information played a far more important role than it does in today's Web. The development of the Semantic Web has a lot of industry momentum, and governments are investing heavily. The U.S. government has established the DARPA Agent Markup Language (DAML) Project, and the Semantic Web is among the key action lines of the European Union's Sixth Framework Programmed [1]. Most information is currently available in a weakly structured form, for example, text, audio, and video. From the knowledge management perspective, the current technology suffers from limitations in the following areas:

- Searching for information for eg. In Companies usually depend on keyword-based search engines, the limitations of which we have outlined.
- For extracting information, human time and effort are required to browse the retrieved documents for relevant information. Current intelligent agents are unable to carry out this task in a satisfactory fashion.
- For maintaining information currently there are problem such as inconsistencies in terminology and failure to remove outdated information.
- For uncovering information new knowledge implicitly existing in corporate databases is extracted using data mining.
- For viewing information we are often desirable to restrict access to certain information to certain groups of employees. "Views," which hide certain information, are known from the area of databases but are hard to realize over an intranet (or the Web).

The aim of the Semantic web is to allow much more advanced knowledge management systems:

- Such that Knowledge will be organized in conceptual spaces according to its meaning.
- The automated tools which will support maintenance by checking for inconsistencies and extracting new knowledge with the help of semantic web.
- The Keyword-based search will be replaced by query answering; requested knowledge will be retrieved, extracted, and presented in a human friendly way.
- Also Query answering over several documents will be supported in semantic web.

- In semantic web defining who may view certain parts of information (even parts of documents) will be possible by the semantic web.

II. Semantic Web Technologies

We can reasonably claim that the challenges to extend the current human-readable web by encoding some of the semantics of resources in a machine-process able form, for Semantic web are engineering and technology adoption rather than a scientific one: partial solutions to all important parts of the problem exist. At present, the greatest needs are
in the areas of integration, standardization, development of tools, and adoption by users. But, of course, further technological progress will lead to a more advanced Semantic Web than can, in principle, be achieved today [1].

Technologies Used in Semantic Web:
Two important technologies for developing the Semantic Web already in used are:

1. eXtensible Markup Language (XML).
2. Resource Description Framework (RDF).

eXtensible Markup Language (XML):
XML lets us create our own tags—hidden labels such as or that annotate Web pages or sections of text on a page. Scripts, or programs, can make use of these tags in sophisticated ways, but the script writer has to know what the page writer uses each tag for. XML allows users to add arbitrary structure to their documents but says nothing about what the structures be the use of mean. The Semantic Web will enable machines to understand semantic documents and data, not human speech and writings. Meaning is expressed by RDF, which encodes it in sets of triples, each triple being rather like the subject, verb and object of an elementary sentence. These triples can be written using XML tags. In RDF, a document makes assertions that particular things (people, Web pages or whatever) have properties (such as "is a sister of," "is the author of") with certain values (another person, another Web page) [3].

Resource Description Framework (RDF):
Meaning is expressed by RDF, which put codes in sets of triples, each triple being rather like the subject, verb and object of an elementary sentence. These triples can be written using XML tags. In RDF, a document makes assertions that particular things (people, Web pages or whatever) have properties (such as "is a brother of," "is the writer of") with certain values (another person, another Web page). This structure turns out to be a natural way to describe the vast majority of the data processed by machines.
The triples of RDF form webs of information about related things. Because RDF uses URIs to encode this information in a document, the URIs ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web, imagine that we have access to a variety of databases with information about people, including their addresses. If we want to find people living in a specific zip code, we need to know which fields in each database represent names and which represent zip codes [3].

III. Languages use in Semantic web

Ontology Languages - Ontologies play a key role in the semantic Web by providing vocabularies that applications can use to understand shared information. DAML+OIL an ontology language designed specifically for use in the semantic Web. It was produced by merging two ontology languages, OIL and DAML. OIL integrates features from frame-based systems and description logics (DLs), and has an RDF-based syntax. DAML is more tightly integrated with RDF, enriching it with a larger set of ontological primitives. Because DAML+OIL are based on description logic, a DL reasoned can be used to compare (semantically) descriptions written in DAML+OIL. This provides a powerful framework for defining and comparing e-commerce service descriptions [4].

Service Description Language - Choosing the appropriate service ontology is an important part of the matchmaking prototype.

WSDL - WSDL (Web Services Description Language) is an XML format for describing network services in abstract terms derived from the specific data formats and protocols used for implementation. As communication protocols and message formats are standardized in the Web community, it becomes possible and important to describe communications in a structured way. WSDL addresses this need by defining an XML grammar for describing network services as collections of communication endpoints capable of exchanging messages. WSDL service definitions provide documentation for distributed systems and serve as a recipe for automating the details involved in application communications. However, WSDL does not support semantic description of services. For example, it does not support the definition of logical constraints between its input and output parameters, although it has the concept of input and output types as defined by XSD [4].

UDDI - UDDI (Universal Description, Discovery, and Integration) is another upcoming XML-based standard for Web service description. It gives a business to describe its business and services, discover other businesses that offer desired services, and integrate with these other businesses by providing a registry of businesses and Web services. UDDI describes businesses by their physical attributes, such as name, address, and the services they provide. UDDI descriptions are augmented by a set of attributes, called tModels, which describe additional features, such as the
classification of services within taxonomies like NAICS (North American Industry Classification System). Since UDDI does not represent service capabilities, however, the tModels it uses only provide a tagging mechanism, and the search performed is only done by string matching on some fields they

have defined. Thus, it is of no use for locating services based on a semantic specification of their functionality [4].

DAML-S - DAML-S permits Web service providers with a core set of markup language constructs for explaining the properties and capabilities of their Web services in unambiguous, computer-interpretable form. DAML-S markup of Web services is intended to ease the automation of Web service tasks, including automated Web service discovery, execution, interoperation, composition, and execution monitoring. In DAML-S, service descriptions are structured into three essential types of knowledge: a ServiceProfile, a ServiceModel and a ServiceGrounding. The ServiceProfile is typically required in a matchmaking process because it provides the information needed for an agent to discover a service that meets its requirements [4].

IV. LIMITS OF TODAY'S WEB

With the current state of the Web, there are only two real methods of gaining broader information about documents. The first is to use a directory or portal site, and thus rely on human editors to scour the Web and appropriately categorize pages and their associated links.

Such portals are the heroes of today's Web. After all, the most effective information management tool on Earth is still the human librarian, and probably will be for years to come. The problem is that directories take tremendous effort to maintain. Finding new links, updating old ones, and maintaining the database technology add to a portal's administrative burden and operating costs.

Search engines are the alternative. Good search engines pay special attention to metadata in the pages that they spider and add to their index databases. In the simplest case, this metadata might take the form of content in <meta> tags. More advanced search engines, like Google, rely on more subtle information. For instance, Google's widely touted algorithm evaluates not only the occurrence of keywords on a page, but also the number of outside links to the page itself, as a measure of its importance or popularity.

Search engines take less human effort on the content management end, but they require a frightfully large resource investment. It's also very difficult to produce valuable indices efficiently. It's no secret that some of the most advanced search engines are so primitive that queries often turn up an unmanageable number of poorly differentiated hits.

The Web needs to support something in between portals and search engines. Of course, until there's a server as sophisticated as HAL 9000 (but, hopefully, not as neurotic), we probably won't be able to completely replace the human portal editor with a computer program. But if we could provide standardized means for Web publishers to catalog and classify their own content, then we could develop more effective agents that work on this substrate of better-organized information.

The result of having better standard metadata would be a Web where users and agents could directly tap the latent information in linked and related pages. This would help free us from having to scour for information site by site, and from relying on portals and search engines. It wouldn't be hard to outfit each user with personal portal generators and search agents tailored to their particular interests, needs, and constraints. These agents might even be configured to learn and respond to personal details with the help of artificial intelligence techniques [5].

V. THE SEMANTIC WEB'S CHALLENGES

It's fine to talk about enabling each Web publisher to properly place content in context, but there are several problems to overcome before any such initiative will gain critical mass:

- *The complexity to develop:* Any technology that the average Web developer can't grasp in a day and apply proficiently in a week is doomed. In addition, a successful technology will have to be integrated into current Web development and maintenance tools. Semantics are quite arcane, and it won't be easy for semantic technologies to meet this criteria.
- *Control of abuse:* Practices like meta-tag spamming, and even trademark hijacking, show that any system that lets people set their own context is subject to abuse. Knowing the value of the Burton snowboards brand, another unscrupulous manufacturer might want to tell an agent that it is the Burton Company in hopes of directing some undeserved attention to its site. Semantic Web technologies will need a mostly automated system for establishing trust in the assertions that Web publishers make. This concept is often referred to as the *Web of trust*:
- *Open for implementation:* Because of the diversity in developers and development tools, Semantic Web technology will have to be politically and technically open for implementation and use. If it requires royalty payments to any party, open source advocates and competing web technology vendors will boycott it. If it requires a specific plug-in or module, most developers and users won't even bother installing it [5].

VI. WHY SEMANTIC TECHNOLOGIES?

The word semantic technologies represent a fairly diverse family of technologies that have been in existence for a long time and seek to help derive meaning from information. Examples of semantic technologies include natural language processing (NLP), data mining, artificial intelligence (AI), category tagging, and semantic search. Some examples of existing semantic technologies being used today include:

- Natural-language processing(NLP) technologies attempt to process unstructured text content and extract the names, dates, organizations, events, etc. that are talked about within the text.
- Data mining: Data mining technologies employ pattern-matching algorithms to tease out trends and correlations within large sets of data. Data mining can be used, for example, to identify suspicious and potentially fraudulent

trading behavior in large databases of financial transactions.

- Artificial intelligence or expert systems. AI or expert systems technologies use elaborate reasoning models to answer complex questions automatically. These systems often include machine-learning algorithms that can improve the system's decision-making capabilities over time.
- Classification: Classification technologies use heuristics and rules to tag data with categories to help with searching and with analyzing information.
- Semantic search: Semantic search technologies allow people to locate information by concept instead of by keyword or key phrase. With semantic search, people can easily distinguish between searching for John F. Kennedy, the airport, and John F. Kennedy, the president.

Many other modern technologies can be called semantic technologies. While all of these technologies have an overall goal in common—helping to make sense of large or complex sets of data without being supplied with any preordained knowledge about the data—they do not share much more than that. They are implemented using many different programming languages, produce data (signal) in many different formats, rely on very different underlying formalisms, and rarely work well together without investing a significant amount of effort in integration engineering.

The key ideas of the Semantic Web, namely, common shared meaning (ontology) and machine processable metadata, establish a promising approach for satisfying the e-learning requirements. It can support both semantic querying and the conceptual navigation of learning materials:

- *Learner-driven:* Learning materials, possibly by different authors, can be linked to commonly agreed ontologies. Personalized courses can be designed through semantic querying, and learning materials can be retrieved in the context of actual problems, as decided by the learner.
- *Flexible access:* Knowledge can be accessed in any order the learner wishes, according to her interests and needs. Of course, appropriate semantic annotation will still set constraints in cases where prerequisites are necessary. But overall nonlinear access will be supported.
- *Integration:* The Semantic Web can provide a uniform platform for the business processes of organizations, and learning activities can be integrated in these processes [6].

VII. USE OF SEMANTIC WEB TECHNOLOGIES IN DAILY LIFE

Semantic Web technologies that are transforming drug discovery and health care are being applied to more general situations. One example is Science Commons, which helps researchers openly post data on the Web. The nonprofit organization provides Semantic Web tools for attaching legally binding copyright and licensing information to those data. This capability allows a scientist, for example, to instruct a software applet to go find information about a particular gene—but only information that comes with a free license. DBpedia is an effort to smartly link information within Wikipedia's seven million articles. This project will allow Web surfers to perform detailed searches of Wikipedia's content that are impossible today, such as, “Find me all the films nominated for a Best Picture Academy Award before 1990 that ran longer than three hours.” As applications develop, they will dovetail with research at the Web consortium and elsewhere aimed at fulfilling the Semantic Web vision. Reaching agreement on standards can be slow, and some skeptics wonder if a big company could overtake this work by promoting a set of proprietary semantic protocols and browsers. But note that numerous companies and universities are involved in the consortium's semantic working groups. They realize that if these groups can devise a few well-designed protocols that support the broadest Semantic Web possible, there will be more room in the future for any company to make money from it. Some observers also worry that people's privacy could become
compromised as more data about them from disparate sources is interlinked. But Semantic Web advocates argue that the protections are the same as those used in the non-linked world. If two databases joined by the Semantic Web have different privacy criteria, then the software will have to honor both sets of rules and create a set that covers both. When SAPPHIRE joins patient databases, it adheres to the privacy requirements of both or it won't proceed; the nurses who had formerly performed the same mergers manually imposed the same practice.

The Semantic Web will probably operate more behind the scenes than the World Wide Web does. We won't see how it helps Eli Lilly create personalized drugs; we'll just buy them. We won't know how Vodafone makes cool ring tones so readily available, but we'll appreciate how easy they are to download. And yet, soon enough the Semantic Web will give more direct power to us, too, allowing us to go on eBay and not just say “find me the Toyota Priuses for sale” but “find me only used, red Priuses for sale for less than $14,000 by people who are within 80 miles of my house and make them an offer.” Grand visions rarely progress exactly as planned, but the Semantic Web is indeed emerging and is making online information more useful than ever [7].

According to Peter Mika in his paper, Flink is a presentation of the professional work and social connectivity of Semantic Web researchers. This community is those researchers who have submitted publications or held an organizing role at any of the past International Semantic Web Conferences (ISWC02, ISWC03, and ISWC04) or the Semantic Web Working Symposium (SWWS01). This means a community of 608 researchers from both academia and industry, covering much of the United States, Europe and to lesser degree Japan and Australia (see Figure 1).

Fig 1: Semantic Web Researchers and their Connections across the Globe [8]

VIII. CONCLUSION

This paper has given an overview of the developing Semantic Web infrastructure, showed how this relates to typical hypermedia research topics and given comprehensive pointers to the relevant literature. Four important areas of research that need to be addressed to allow the Semantic Web to realize its full potential have been described. Originally, hypertext research aimed to bring user interaction with digitally stored information closer to the semantic relations implicit within the information. Much of the more "hypertext-specific" research, however, turned to system and application-oriented topics, possibly through the lack of an available infrastructure to support more explicit semantics. The introduction of the Web, as a highly distributed, but relatively simple, hypermedia system has also influenced the character of hypermedia research. The existence of XML and RDF, along with developments such as RDF Schema and DAML+OIL, provides the impetus for realizing the Semantic Web. During these early stages of its development, we want to ensure that the many hypertext lessons learned in the past will not be lost, and that future research tackles the most urgent issues of the Semantic Web.

References

[1] Grigoris Antoniou & Frank Van Hermeln, "A Semantic web primer" Second Edition 2008.(ebook)

[2] Boris Katz Jimmy Lin," Annotating the Semantic Web Using Natural Language" , MIT Artificial Intelligence Laboratory 200 Technology Square Cambridge, MA 02139, USA.

[3] Tim Berners-Lee, James Hendler and Ora Lassila," The Semantic Web ", May 17, 2001(www.scientificAmerican.com) access on 14.09.2014.

[4] Lei Li and Ian Horrocks ,"A Software Framework for Matchmaking Based on Semantic Web Technology", *International Journal of Electronic Commerce* / summer 2004, Vol. 8, No. 4.

[5] http://www.drdobbs.com/the-languages-of-the-semantic web/184414544(Semantic web Challenges)access on 10.09.2014.

[6] http://www.cambridgesemantics.com/semantic-university/semantic-web-vs-semantic-technologies(Semantic Technology)access on 11.09.2014.

[7] Lee Feigenbaum, Ivan Herman, Tonya Hongsermeier, Eric Neumann and Susie Stephens; "The Semantic Web in Action - Scientific American - December 2007", http://thefigtrees.net/lee/sw/sciam/semantic-web-in-action.

[8] Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks(research paper by Peter Mika,Department of Computer Science ,Vrije Universiteit Amsterdam (VUA) De Boelelaan 1081, 1081HV Amsterdam, The Netherlands).