Next Generation Semantic Web and Its Application

Soumyarashmi Panigrahi[1] , Sitanath Biswas[2]
[1]Computer Science &Engg.,ITER, Siksha 'O' Anusandhan University,
Bhubaneswar, Orissa,751030, India

[2]Information Technology, ITER, Siksha 'O' Anusandhan University,
Bhubaneswar, Orissa,751030, India

Abstract
Currently, computers are changing from single, isolated devices into entry points to a worldwide network of information exchange and business transactions called the World Wide Web (WWW). However, the success of the WWW has made it increasingly difficult to find, access, present and maintain the information required by a wide variety of users. In response to this problem, many new research initiatives and commercial enterprises have been set up to enrich the available information with machine-process able semantics. This Semantic Web will provide intelligent access to heterogeneous, distributed information, enabling software products (agents) to mediate between user needs and the information sources available. In this paper we describe some areas for application of this new technology. We focus on on-going work in the fields of knowledge management and electronic commerce. We also take a perspective on the semantic web-enabled web services which will help to bring the semantic web to its full potential.

Keywords: *World Wide Web (WWW), Semantic Web, Web Ontology Language (OWL),RDF(Resource Description Framework), XML(Extensible Markup Language).*

1. Introduction

The World Wide Web (WWW) has drastically changed the availability of electronically accessible information. The WWW currently contains some 3 billion static documents, which are accessed by over 300 million users internationally. However, this enormous amount of data has made it increasingly difficult to find, access, present and maintain the information required by a wide variety of users. This is because information content is presented primarily in natural language. Thus, a wide gap has emerged between the information available for tools aimed at addressing the problems above and the information maintained in human-readable form. In response to this problem, many new research initiatives and commercial enterprises have been set up to enrich available information with machine-process able semantics. Such support is essential for "bringing the web to its full potential". Tim Berners-Lee, Director of the World Wide Web Consortium, referred to the future of the current WWW as the "semantic web" - an extended web of machine-readable information and automated services that extends far beyond current capabilities ([Berners-Lee at 2001], . The explicit representation of the semantics underlying data, programs, pages, and other web resources, will enable a knowledge-based web that provides a qualitatively new level of service. Automated services will improve in their capacity to assist uses in achieving their goals by "understanding" more of the content on the web and thus providing more accurate filtering, categorization and searching of information source. The vision of the semantic web is to enable machines to interpret and process information in the World Wide Web.

The aim is to support humans in carrying out their various tasks with the World Wide Web. Several technologies have been developed for shaping, constructing and developing the semantic web. Many of the so far developed semantic web technologies provide us with tools for describing and annotating resources on the Web in standardized ways, e.g. with the Resource Description Framework (RDF) and its binding to XML (eXtensible Mark-up Language).Semantic Web will enable automatic collection and correlation of various parts of information about an object, available at various different web resources. Semantic Web will save our valuable time we spend on navigating from one web resource to another in order to obtain meaningful information on a particular object.

2. Next Generation Semantic Web Application

Our research on next-generation Semantic Web applications originates from our observation and anticipation that intelligent application development will increasingly change owing to the availability of the Semantic Web's large scale; distributed body of knowledge dynamically exploiting this knowledge introduces new possibilities and challenges requiring novel infrastructures to support the implementation of next generation Semantic Web applications. Next-generation Semantic Web applications must address significant problems associated with the Semantic Web's scale and heterogeneity as well as with the widely varying quality of the information it contains.

3. Features and Requirements

Next-generation Semantic Web applications achieve their tasks by automatically retrieving and exploiting knowledge from the Semantic Web as a whole. Unlike early Semantic Web applications, which gathered and engineered knowledge at design time, these new applications explore the Web to discover ontologies relevant to the task at hand. Because dynamic knowledge reuse replaces the traditional knowledge-acquisition task, we can potentially reduce the application development cost. In addition, because such applications can use any semantic information available online, they're not necessarily bound to a particular domain. Any application that wishes to explore large-scale semantics must perform the following tasks:

Find relevant source:

The ability to dynamically locate sources with relevant semantic information is a prerequisite for applications that aim to leverage online knowledge. This feature is important because developers might not be able to judge a particular resource's relevance to the target problem at design time.

Select appropriate knowledge:

Applications must select the appropriate knowledge from the set of previously located semantic documents on the basis of application- dependent criteria, such as data quality and adequacy to the task at hand.

Exploit heterogeneous knowledge source:

at runtime. When reusing online semantic information, the application can't make assumptions about the ontological nature of the target elements. Hence, the process must be generic enough to use any online semantic resource. As with the two previous tasks, the application must carry out this activity.

Combine ontologies and resources:

Developers can't expect one unique knowledge source to provide all the required elements for a given application. Therefore, a typical next-generation Semantic Web application must select and integrate partial knowledge fragments from different sources and jointly exploit them.

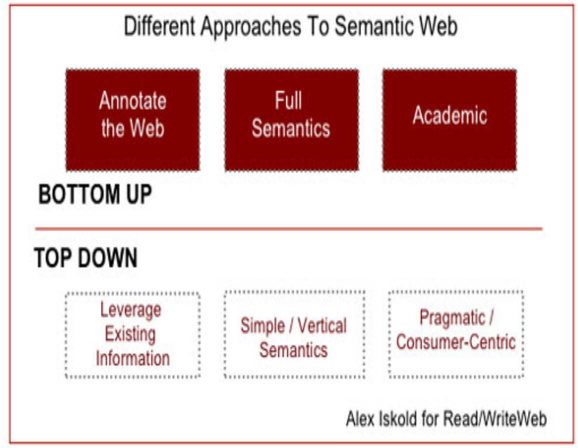

Fig.1 Approaches to Semantic Web

4. Components of theSemantic Web

4.1 RDF(Resource Description Framework)

RDF is a language for stating meta-data about resources. RDF can be used to describe any resource that can be identified by a URI which typically means web-resources, but can also be used for any other resources where natural identifier are available, for instance, ISBN numbers, ISO country codes, or national insurance numbers. RDF documents are built up as graphs, where each node and edge is labeled with a URI. The base RDF unit for representing information is a (Subject, Predicate, Object) triple, where the subject and object are referred to as resources. A collection of these triples form a RDF graph where the subject and objects are nodes and the predicate the edge between them. As RDF is built upon other W3C technologies both the resources and properties are identified by Uniform Resource Indicators (URIs). RDF has some syntactic approach for different collections of objects called containers. There are three types of containers: bag, sequence, and alternative. A bag is unordered, a sequence is ordered, and an alternative is a set of choices. Using a special about each attribute, a document can make statements that apply to every element in a collection.

RDF is a framework for describing web resources (identified by Uniform Resource Identifier or URIs) such as homepage, title, author, content and copyright information of a web page etc. RDF is a data model for objects("resources") and relations between them. RDF provides a simple semantics for this data model, and these data models can be represented in XML syntax. RDF is designed to be read and understood by computers but it is not meant for displaying to human being. RDF describes resources with properties and property values. A resource is anything that can have a URI, such ashttp://www.kolkataabcd.in/NLE.A property is also a resource that has a name, such as "author" or "homepage."A property value can also be another resource. RDF is a basic ontology language and a graphical language used for representing information about resources on the web[5]. Resources are described in terms of properties and property values using RDF statements. Statements are represented as triples, consisting of a subject, predicate and object [S, P, O]. The subject of one statement may be the object of another statement and that is how resources can be merged together. A set of linked statements (triples) forms an RDF Graph.

4.2 RDF Schema(RDFS)

RDFS is an extension to RDF that provides the framework to describe application- specific classes and properties and thus allows resources to be defined as instances of classes, and subclasses of classes. RDFS allows properties to be defined as sub properties of other properties and it enriches the descriptions of what we already have.RDF Schemas are used to declare vocabularies. RDF schema defines the terms that will be used in RDF statements and gives specific meanings to them[6]. It provides a basic type system for use in RDF models. It defines resources and properties such as Class and SubClassOf that are used in specifying application-specific schemas. An RDF Schema provides information about the interpretation of the statements given in an RDF data model.

4.3 XML (eXtensible Markup Language)

This language is extended version of hyper text markup language (html).a html document can contain a only a static page. We can't keep any dynamic data on it & now a days webs are too much flexible.Html is a presentation oriented rather Xml is a syntax oriented language. The Xml is a simple, flexible text format derived from SGML (ISO 8879),the w3c created, developed and continues to maintain the XML specification. The W3C is also the primary center for developing other cross-industry specifications that are based on XML. Some of these are done by within the XML ACTIVITY such as XML QUERY and XML SCHEMA, and some are being done in other W3C areas. The XML activity tries to keep a balance between maintaining stability and backwards compatibility, making improvements that help to encourage interoperability, and bringing new communities into the world of XML. Recently, the W3C has released an alternative to DTDs (Document Type Definition) called XML Schema. XML Schemas provide greater flexibility in the definition of an XML application, even allowing the definition of complex data types. Furthermore, XML Schemas use the same syntactic style as other XML documents. However, XML Schema only gives XML an advanced grammar specification and data typing capability, and still suffers from the same semantic drawbacks as DTDs.

The eXtensible Markup Language (XML) (W3C 2000) is rapidly becoming the premier method for exchanging information across the Internet. The Document Type Definition (DTD) language, which has traditionally been the most common method for describing the structure of XML instance documents, lacks enough expressive power to properly describe highly structured data. XML Schema (W3C 2001), on the other hand, provides a much richer set of structures, types and constraints for describing data and is therefore expected to soon become the most common method for defining and validating highly structured XML documents. Information in schema documents is often used by XML-aware editing systems so that they can offer users the most likely elements to occur at any given location in a document. Checking a document against a Schema is known as validating against that schema; for a DTD, this is just validating, but for any other type of schema the type is mentioned, such as XSD (XML Schema Definitions) Validation or Relax-NG validation. Validating against a schema is an important component of quality assurance. The Service Modeling Language (SML) provides a framework for relating multiple XSD documents to one or more documents in a single validation episode. Since XSD supports associating data types with element and attribute content, it is also used for data binding, that is, for software components that read and write XML representations of computer programming-language objects.

4.4 OIL (Ontology Interference Layer)

OIL was developed by Dieter Fensel, Frank Van Harmelen (Vrije Universities, Amsterdam) and Ian Horrocks (University of Manchester) as part of the IST Onto Knowledge project. Much of the work in OIL was subsequently incorporated into DAML+OIL and the Web Ontology Language (OWL).OIL is a language for describing ontologies on the Web. OIL's semantics are based on description logics, but its syntax is layered on RDF. One of the design goals for OIL was to maximize integration with RDF applications. OIL starts with the basic primitives of RDF, classes and properties. There are two basic types of classes: primitive classes and defined classes. Primitive classes are essentially ordinary RDFS classes, while defined classes provide necessary and sufficient conditions for membership. Defined classes require the use of class expressions, which are Boolean combinations of classes and slot constraints. The standard Boolean operations are provided by oil: AND, oil: OR, and oil: NOT. The advantages of OIL are tied to its description logic basis. If two ontologies used the same set of base terms in their definitions, then it is possible to automatically compute a subsumption hierarchy for the combination of the ontologies. Additionally, the rich modeling constructs allow consistency to be checked, which eases the construction of high-quality ontologies. However, it is possible for logical inconsistencies to arise due to instances, which will be distributed across the Semantic Web and thus harder to control. Three roots of OIL are as follows:-

Description Logics (DL):
DLs describe knowledge in terms of concepts and role restrictions that are used to automatically derive classification taxonomies. The main effort of the research in knowledge representation is in providing theories and systems for expressing structured knowledge and for accessing and reasoning with it in a principled way. DLs (cf. [Brachman&Schmolze, 1985], [Baaderet al., 1991]), also known as terminological logics, form an important and powerful class of logic based knowledge representation languages.

Frame-based systems:
The central modeling primitives of predicate logic are predicates. Frame based and object-oriented approaches take a different point of view. Their central modeling primitives are classes (i.e., frames) with certain properties called attributes. These attributes do not have a global scope but are only applicable to the classes they are defined for (they are typed) and the ”same” attribute (i.e., the same attribute name) may be associated with different value restrictions when defined for different classes.

DARPA Agent Markup Language(DAML):
It is a agent markup language developed by the DARPA for the semantic web. The DAML program has generated the DAML+OIL markup language. The submission of the DAML+OIL language to the World Wide Web consortium captures the work done by DAML contractors and the EU/U.S. Joint Committee on Markup Languages. This submission was the starting point for the language to be developed by W3C’s web ontology working group, WebOnt.

Ontology Layer technology:
The ontology has been developing a layered architecture, which is often represented using a diagram as follows and 1st proposed by Tim Berners-Lee, developer of the web.

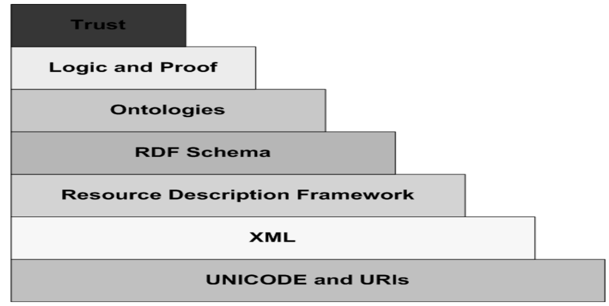

Fig.2 Different layers of ontology

These layers are briefly described as follow:-
Unicode and URI: Unicode, the standard for computer character representation, and URIs, the standard for identifying and locating resources (such as pages on the Web), provide a baseline for representing characters used in most of the languages in the world, and for identifying resources.

XML: XML and its related standards, such as Namespaces, and Schemas, form a common means for structuring data on the Web but without communicating the meaning of the data. These are well established within the Web already.

Resource Description Framework: RDF is the first layer of the Semantic Web proper. RDF is a simple metadata representation framework, using URIs to identify Web-based resources and a graph model for describing relationships between resources. Several syntactic representations are available, including a standard XML format.

RDF Schema: A simple type modeling language for describing classes of resources and properties between them in the basic RDF model. It provides a simple reasoning framework for inferring types of resources.

Ontologies: A richer language for providing more complex constraints on the types of resources and their properties.

Logic and Proof: An (automatic) reasoning system provided on top of the ontology structure to make new inferences. Thus, using such a system, a software agent can make deductions as to whether a particular resource satisfies its requirements (and vice versa).

Trust: The final layer of the stack addresses issues of trust that the Semantic Web can support. This component has not progressed far beyond a vision of allowing people to ask questions of the trustworthiness of the information on the Web, in order to provide an assurance of its quality.

OWL (Web Ontology Language)

The Web Ontology Language (OWL) was developed to be a more formal and more powerful ontology language than RDFS. Whereas RDFS only really allows describing shared vocabularies or schemas, i.e. only specifying the symbols used, OWL goes further towards specifying the actual semantics of the classes using the Description Logics formalism (Baader et al.,2003).OWL provides concepts such as cardinality constraints, restriction based sub-classing, functional properties, and class/individual equality. Three flavors of owl exist: owl Lite, owl Dl and owl Full. Each flavor increases the expressiveness and also encompasses the previous levels, so every legal owl lite ontology is also owl Dl ontology, and every legal owl Dl ontology is also a legal owl Full ontology.owl makes some efforts to be compatible with RDF and RDFS, and it is possible to express all owl facts as triples, OWL and RDF are not fully inter-operable OWL-S supplies a core set of ontology concepts for describing the properties and capabilities of Web services in unambiguous, computer-interpretable form. OWL-S markup of Web services is designed to facilitate the automation of Web service tasks including automated Web service discovery, execution, interoperation, composition and execution monitoring.

The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies. The languages are characterized by formal semantics and RDF/XML-based serializations for the Semantic Web. OWL is endorsed by the World Wide Web Consortium and has attracted academic, medical and commercial interest. Web Ontology Language (OWL) is a language for defining and instantiating web ontologies (a W3C Recommendation). OWL ontology includes description of classes, properties and their instances. OWL is used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. Such representation of terms and their interrelationships is called ontology. OWL has facilities for expressing meaning and semantics and the ability to represent machine interpretable content on the Web.OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans. This is used for knowledge representation and also is useful to derive logical consequences from OWL formal semantics.

OWL provides three increasingly expressive sublanguages designed for use by specific communities of implementers and users. These are as follows:-

OWL Lite:

It supports those users primarily needing a classification hierarchy and simple constraints. For example, while it supports cardinality constraints, it only permits cardinality values of 0 or 1. It should be simpler to provide tool support for OWL Lite than its more expressive relatives, and OWL Lite provides a quick migration path forthesauri and other taxonomies. Owl Lite also has a lower formal complexity than OWL DL.

OWL DL:

It supports those users who want the maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time). OWL DL includes all OWL language constructs, but they can be used only under certain restrictions (for example, while a class may be a subclass of many classes, a class cannot be an instance of another class).OWL DL is so named due to its correspondence with *description logics*, a field of research that has studied the logics that form the formal foundation of OWL.

OWL Full:

It is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees. For example, in OWL Full a class can be treated simultaneously as a collection of individuals and as an individual in its own right. OWL Full allows an ontology to augment the meaning of the pre-defined (RDF or OWL) vocabulary. It is unlikely that any reasoning software will be able to support complete reasoning for every feature of OWL Full.

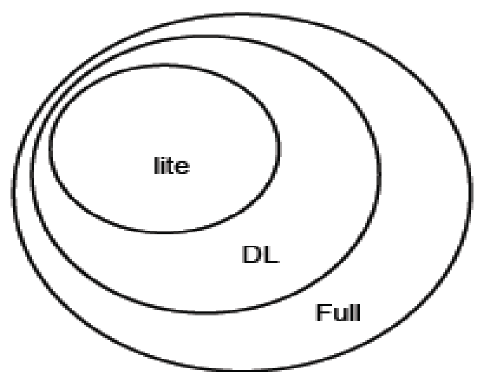

Fig.3 Diagrammatic Representation of OWL Sublanguages

Each of these sublanguages is an extension of its simpler predecessor, both in what can be legally expressed and in what can be validly concluded. The following set of relations hold.

Every legal OWL Lite ontology is a legal OWL DL ontology.

Every legal OWL DL ontology is a legal OWL Full ontology.

Every valid OWL Lite conclusion is a valid OWL DL conclusion.

Every valid OWL DL conclusion is a valid OWL Full conclusion.

Ontology developers adopting OWL should consider which sublanguage best suits their needs. The choice between OWL Lite and OWL DL depends on the extent to which users require the more-expressive constructs provided by OWL DL. The choice between OWL DL and OWL Full mainly depends on the extent to which users require the meta-modeling facilities of RDF Schema (e.g. defining classes of classes, or attaching properties to classes). When using OWL Full as compared to OWL DL, reasoning support is less predictable since complete OWL Full implementations do not currently exist. OWL Full can be viewed as an extension of RDF, while OWL Lite and OWL DL can be viewed as extensions of a restricted view of RDF. Every OWL (Lite, DL, Full) document is an RDF document, and every RDF document is an OWL Full document, but only some RDF documents will be a legal OWL Lite or OWL DL document. When the expressiveness of OWL DL or OWL Lite is deemed appropriate, some precautions have to be taken to ensure that the original RDF document complies with the additional constraints imposed by OWL DL and OWL Lite.

5. Future of the Semantic Web

Semantic Web has gone from being fairly small scale and mainly of interest to researchers within knowledge-representation to being an important force in the development of the future Web in general, with large industrial players now taking an active part. Consider for instance the Data Portability effort, which includes Google, Face book, Flickr, LinkedIn and many other large web-companies. The increased interest and activity surrounding the Semantic Web is undoubtedly joined with the recent evolution of the normal Web, the Web2.0 movement (O`Reilly, 2005) has already gone far in moving the focus on the web from web-pages to data.

Behind the scenes of the Web2.0 pages there are many technical advances that go well with the Semantic Web vision, data interchange formats such as Micro formats and JavaScript Object Notation (JSON) can easily interoperate with RDF, and the REST architecture encourages building Web applications in a data-oriented manner from the ground up. While the normal Web has been making progress becoming more Semantic Web compatible, there has also been more interest in the web part of the Semantic Web. In the following graph a informal mapping have done which shows the future aspects of semantic web in corresponding fields.

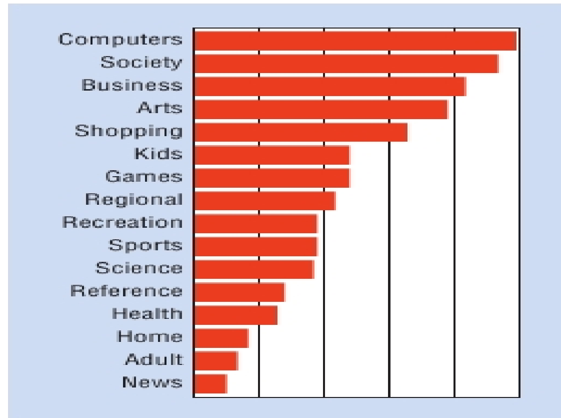

Fig.4 Aspects of Semantic Web

6. E-Commerce & Semantic Web

Semantic web technologies will give digital assistants and agents the ability to search the web for products that correspond best to the specific needs of a certain user. While consumers today have to rely on the limited number of offers available on centralized e-commerce portals when looking for products, future applications will be able to provide users with a search process based on product attributes, which will include all products published in this form on the Internet. Furthermore, in a next step semantic web services will enable digital assistants to handle business processes like selling and buying or even negotiations automatically. The success of semantic web relies heavily on its wide spread adoption by the mainstream web development community. Unfortunately, a successful application of semantic web needs a large amount of semantic data, which is difficult for the small knowledge representation and semantic web community to provide.

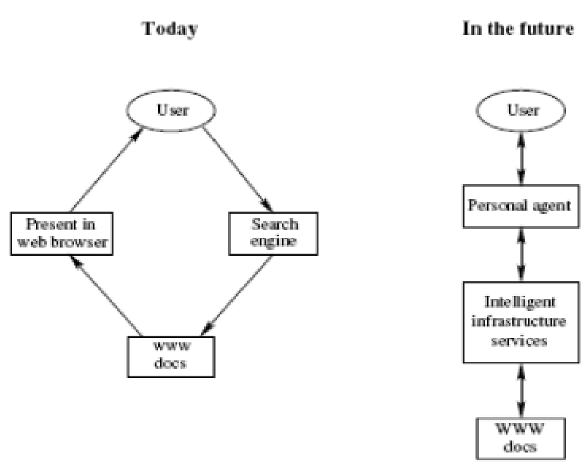

Fig.5 Application of Semantic Web

7. Knowledge Management

Based on the semantic network knowledge representation formalism, which enables packaging information in the form of object-attribute-value statements, so called triplets. By assuming that terms used in these statements are based on the formally specified meaning (for the community of interest), i.e. ontologies, these triplets can be semantically processed by machine agents. Most of the current Semantic Web applications are based on using such atomic statements as pure facts, which we can reason about. So, a machine agent can understand information that a concrete patient, who suffers from disease X, is treated by medicine Y. Moreover, the agent can use this information in the communication with other machine agents (e.g. to make an appointment with the doctor W), making the vision of the Semantic Web real. But useful statements, which can be exchanged between agents, are not always related to concrete individuals - instances (e.g. patient X, disease Y), but also to a group of individuals with some common characteristics (e.g. statements about female patients older than 60 who suffer from disease Y). Moreover, atomic statements could be combined in a more expressive way as simple conjunction, for example in the conditional form (e.g. Precondition: the patient is male and suffers from X; Action: he has to be treated by medicine Y). On the implementation level this form can be represented using the If-Then statements, forming in that way reasoning atoms for inference- and trust- services on the Semantic Web.

8. Conclusion

Semantic Web applications in learning require the authoring of different educational resources, business field and social network that may be in different formats not easy for traditional authorized people. Authoring tools that integrate the different formats of these fields, being easy to use by authorized people have a challenge. Each specific Semantic Web application and its related information should be analyzed to take the best design decisions.

9. References

[1] O.Lassila and R. Swick, "W3C resource description framework (rdf)model and syntax specification", Available at: http://www.w3.org/TR/REC- rdfsyntax/(October 25,2002).

[2] D. Fensel, F. Van Harmelen, I. Horrocks, D. McGuinness, and P. F. Patel- Schneider,(2001), OIL: An ontology infrastructure for the semantic web. IEEE Intelligent Systems, 16(2),38-45, URL download/2001/IEEE-IS01.pdf.

[3] P. Dolog and M.Bielikova,(Sept. 2002),Towards variability modelling for reuse in Hypermedia engineering. In Y. Manolopoulos and P. Navrat, editors, Proc.of 6th East-European Conference on Advances in Databases and Information Systems - ADBIS'2002, Bratislava,Slovakia,388–400.

[4] F. BryandS. Schaffert,(June 2002), A gentle introduction into xcerpt, a rule-based query and transformation language for xml. In International Workshop on Rule Markup Languages for Buisiness Rules on the Semantic Web, Sardinia, Italy.

[5] P.Dolog, R.Gavriloaie, W.Nejdl and J.Brase, (May 2003), Integrating adaptive hypermedia techniques and open rdf-based environments, In International World Wide Web Conference, Budapest, Hungary.

[6] P. Brusilovsky,(1996),Methodsand techniques of adaptive hypermedia. User Modelling and User Adapted Interaction,6(2-3),87–129.

[7] D.M. Pisanelli, A. Gangemi, and G. Steve,(2002) "Ontologies and Information Systems:The Marriage of the Century?", Proceedings of the LYEE Workshop, Paris.

[8] G. Zhao, Y. Gao, and R. Meersman, (2004),Ontology-Based Approach to BusinessModelling. Proceedings of the International Conference of Knowledge Engineering and Decision Support (ICKEDS2004). In press.

[9] Michael K. Bergman (2009) Comprehensive Listing of 250 Semantic Web Tools. Available at http://www.mkbergman.com/?p=291(Accessed on 6/1/2009).

[10] D. Fensel et al. OIL: An Ontology Infrastructure for the Semantic Web.IEEE Intelligent Systems 16, 2 (2001).

[11] Maria Vargas-Vera, Enrico Motta and John Domingue.(2004) "AQUA: An Ontology-Driven Question Answering System" In Maybury, Dr M. T., Eds. Proceedings AAAI Spring Symposium, New Directions in Question Answering, pages pp. 53-57, Stanford University, CA USA.

[12] A. Passant, T. Hastrup, U. Bojars, and J. Breslin.Microblogging(June 2008), "A Semantic Web and distributed approach."In Proceedings of the 4th Workshop on Scripting for the Semantic Web.

Authors Profile

Soumyarashmi Panigrahi is presently working as a Assitant Professor in Computer Science & Engineering Dept., AIET, Bhubaneswar under Biju Pattnaik University of Technology(BPUT) . Also, she is continuing her M.Tech. in CSE Dept. in Institute of Technical Education & Research (ITER), Bhubaneswar under Siksha O Anusandhan University. She is having 2 years of teaching experience and she is having keen interest in the areas such as Semantic Web and Ontology Building.

Mr. Sitanath Biswas is presently workig as a Assitant Professor in Information Technology Dept., Institute of Technical Education & Research(ITER), Bhubaneswar under Siksha O Anusandhan University. He is having 3 years of teaching experience. He got his M.E. Degree form Utkal University,Bhubaneswar.
He published 15 research papers in International and National journals and delivered invited lectures at various conferences, seminars and workshops. He guided a large number of students for their B.Tech. and M.Tech degrees in Computer Science and Engineering and Information Technology. His current research interests are Artificial Intelligence, Semantic Web, Ontology Engineering.