



Emerging Role of Ontology in Semantic Web: Developmental Prospective

Rakesh Kumar Singh

Scientist-D (Information Technology)

G.B. Pant Institute of Himalayan Environment & Development
Almora, Uttarakhand, India

Ranjan Singh

MCA Student

IGNOU
New Delhi, India

Abstract— This paper is focused on the ontology aspects in computer science and its role in semantic web. The ontology has been widely used in recent years in the field of Artificial Intelligence, computer and information science especially in domains such as, cooperative information systems, intelligent information integration, information retrieval and extraction, knowledge representation, and database management systems. The ontology evolution process enable the customisation of the ontology-evolution process to the current need and ensures the consistency of the underlying ontology and all dependent artefacts and also offers advice to the user for continual ontology reengineering. The implementation of the second generation of the WWW, the so-called Semantic Web, has made the idea of the large-scaled ontology-based applications in a business context real. Ontologies have recently become a key technology for semantics-driven modelling. The explicit representation of the semantics of data through ontologies enables applications to provide a qualitatively new level of services, such as verification, justification, gap analysis, etc.

Keywords— Ontology, Artificial Intelligence, Information Retrieval, Ontology Evolution, Semantic web, etc.

I. INTRODUCTION

The word ontology is borrowed from philosophy, where an ontology is a systematic account of existence. In the context of computer science, the ontology definition is “An ontology is a formal, explicit specification of a shared conceptualisation of a domain of interest.” Ontology is used in the fields of computer science as artificial intelligence, software engineering, semantic web, language processing. In the field of computer science, ontology is the foundation of describing a domain of interest; it consists in a collection of terms organized in a hierarchical structure that shape the reality. The following are the components of ontology: 1.concepts, terms; 2.relations between concepts, terms; 3. Properties, attributes of the concepts; 4. Rules, axioms, predicates, constraints. Ontology is used to make assumptions about the meaning of a specific concept. Machine processibility is the basis for the next generation of the WWW, the so-called Semantic Web, which is based on using ontologies for enhancing with formal semantics.

The purpose of an ontology is not to model the whole world, but rather a part of it so called domain. A domain is just a specific subject area or area of knowledge, like medicine, tool manufacturing, real estate, automobile repair, financial management, etc. Therefore, in order to define a domain, it is important to know what an ontology is for. Ontologies serve as a means for establishing a conceptually concise basis for communicating knowledge for many purposes. Ontologies are now ubiquitous in many enterprise-wide information- systems: they are used in e-commerce, knowledge management and in various application fields such as bioinformatics and medicine.

II. TOOLS AND TYPES OF ONTOLOGY

Ontological tools usually provide a graphical user interface for building ontologies, which allows the ontologist to create ontologies without using directly a specific ontology specification language like OntoEdit, OilEd, WebODE, Ontolingua, Ontosaurus, LinkFactory. A great range of languages have been used for implementing ontologies during the last decade like Ontolingua, LOOM, OCML, FLogic, CARIN, OKBC, Telos, Cycl. Many of these languages had been already used for representing knowledge inside knowledge-based applications, other ones were adapted from existing knowledge representation languages, and there is also a group of languages that were specifically created for representing ontologies. Recently, Web-based ontology specification languages have been developed in the context of the World Wide Web: RDF, RDF Schema, SHOE, XOL, OML, OIL, DAML+OIL and OWL. The types of ontologies are as follows:

- **Domain Ontology:** A domain ontology represents concepts which belong to part of the world. Particular meanings of terms applied to that domain are provided by domain ontology. For example the word card has many different meanings. Ontology about the domain of poker would model the "playing card" meaning of the word, while an ontology about the domain of computer hardware would model the "punched card" and "video card" meanings.
- **Upper Ontology:** An upper ontology is a model of the common objects that are generally applicable across a wide range of domain ontologies. It usually employs a core glossary that contains the terms and associated object descriptions as they are used in various relevant domain sets. There are several standardized upper ontologies

available for use, including BFO, Dublin Core, GFO, OpenCyc/ResearchCyc, SUMO, the Unified Foundational Ontology (UFO), and DOLCE.

- **Hybrid Ontology:** The Gellish ontology is an example of a combination of an upper and a domain ontology.

III. ONTOLOGY EVOLUTION AND SIGNIFICANCE

Ontology evolution can be defined as "Ontology Evolution is the timely adaptation of an ontology to the arisen changes and the consistent propagation of these changes to dependent artefacts". Since a change in the ontology can cause inconsistencies in other parts of the ontology, as well as in the dependent artefacts, the ontology evolution has to be considered as a process. The following characterization will be used by adapting the terminology from the database community:

- **Ontology Management** is the whole set of methods and techniques that is necessary to efficiently use multiple variants of ontologies from possibly different sources for different tasks. Therefore, an ontology management system should be a framework for creating, modifying, versioning, querying, and storing ontologies. It should allow an application to work with an ontology without worrying about how the ontology is stored and accessed, how queries are processed, etc.
- **Ontology Modification** is accommodated when an ontology management system allows changes to the ontology that is in the use, without considering the consistency.
- **Ontology Evolution** is accommodated when an ontology management system facilitates the modification of an ontology by preserving its consistency.
- **Ontology Versioning** is accommodated when an ontology system management allows handling of ontology changes by creating and managing different versions of it.

IV. TECHNICAL ISSUES IN UNDERSTANDING ONTOLOGY EVOLUTION

Ontology Management is the whole set of methods and techniques that is necessary to efficiently use multiple variants of ontologies from possibly different sources for different tasks. Therefore Ontology evolution is not a trivial process, due to the variety of sources and consequences of changes. Building a ontology system has proven to be a difficult task, since there is almost a complete lack of suitable methodology, techniques and tools. There are three challenges for the efficient realisation of the ontology evolution:

- **Complexity:** An ontology model is rich and, therefore, an ontology has an interwoven structure. Each change leads to a change specific workaround. Even when the effects of a change are minor, the cumulative effect of all changes realizing a user's request can be enormous.
- **Dependencies:** Ontologies often reuse and extend other ontologies. Changes in an ontology may affect the ontologies that are based on it. Therefore, changes between dependent ontologies are interrelated, and the immediate synchronisation between dependent ontologies is required. Obviously, the complexity of the ontology evolution increases with the number of dependent ontologies being evolved.
- **Physical Distribution:** Ontology development is a decentralized and collaborative process. Therefore, the physical distribution of the dependent ontologies has to be taken into account. The ontology evolution requires tracking the changes applied to an ontology and broadcasting the group of changes when an explicit request arises.

V. ONTOLOGY BASED SEMANTIC WEB

According to Berners-Lee, the Semantic Web is "A web of data that can be processed directly and indirectly by machines." Semantic Web is a network that takes the apparently infinite amount of data on the World Wide Web, but also connects this information with data in relational databases and other non-compatible archives, for example, the EDI system. Given that relational databases house most of the data from the company, the capability of the semantic Web technology to access and process, along with other data from websites, databases, XML documents, and other systems increases the amount of useful Data exponentially. Moreover relational databases already contain a large amount of semantic information. Databases are organized in tables and columns on the basis of relations between the tasks at home, and these relationships show the meaning (semantics) of data.

The Semantic Web is the next generation of the WWW, which is based on using ontologies for enhancing content with formal semantics. It is worth noticing that the real power of the Semantic Web is realised when many systems that collect Web content from diverse sources, integrate and process the information as well as exchange the results with other human or machine agents are created. Thereby, the effectiveness of the Semantic Web will increase drastically as more machine-readable Web contents and automated services become available. This level of inter-agent communication will require the exchange of "proofs". Furthermore, the Semantic Web will also be the basis for the Web of Trust, which will provide mechanisms to handle authentication, permission, and validation of attribution in a Web where, by design, anyone can contribute content, links, and services. Two important technologies for developing the Semantic Web are already in place: the eXtensible Markup Language (XML) and the Resource Description Framework (RDF). The six layers of the semantic web are presented in Fig.1.

- Layer-1: The XML layer represents the *structure* of data.
- Layer-2: The RDF + RDF Schema layer represents the *meaning* of data.

- Layer-3: The Ontology Vocabulary layer represents the *formal common agreement* about meaning of data.
- Layer-4: The Logic layer enables intelligent *reasoning* with meaningful data.
- Layer-5: The Proof layer supports the exchange of "*proofs*" in an inter-agent communication enabling the common understanding of how the desired information is derived.
- Layer-6: The Trust layer ranges from digital signatures and security to social network analysis.

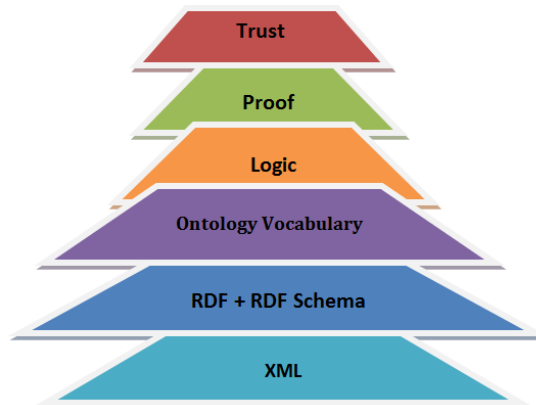


Fig.1: Layers of Semantic Web.

XML allows users to add arbitrary structure to their documents, but says nothing about what the structures mean. The meaning of XML-documents is intuitively clear to humans since the "semantic" mark-up and tags are domain-terms. However, computers do not have intuition. Tag-names per se do not provide semantics. The Resource Description Framework (RDF) provides a means for adding semantics to a document. RDF is an infrastructure that enables encoding, exchange and reuse of structured metadata. RDF, in combination with RDFS, offers modelling primitives that can be extended according to the needs at hand. Basic class hierarchies and relations between classes and objects are expressible in RDFS. Some parts of the RDF and RDFS vocabularies are not assigned any formal meaning, and in some cases, notably the reification and container vocabularies, it assigns less meaning than one might expect.

Ontologies are well suited for describing heterogeneous, distributed and semi-structured information sources (e.g. XML documents) that can be found on the web or in the intranets. The ontologies, which already exist on the Semantic Web, range from simple taxonomies (such as the Yahoo hierarchy), to metadata schemes, to logical theories.

VI. CATEGORY OF ONTOLOGY LANGUAGES

Based on their formal semantics ontology language can be split into two categories which are as follows:

(i). **Frame-based Ontology Languages:** Their central modelling primitives are classes (known as frames) with properties (known as slots). A frame provides a context for modelling a class, which is generally defined as a subclass of one or more other classes, with slot-value pairs being used to specify additional constraints on instances of the new class. For example, the KAON (Karlsruhe Ontology and Semantic Web framework) ontology language, incorporates the essential modelling primitives of frame-based systems, being based on the notion of a concept and the definition of its superclasses and slots. It also treats slots as first class objects that can have their own properties (e.g. domain and range) and can be arranged in a hierarchy. As per KAON ontology language, all information is organised in so-called OI-models (ontology-instance models), containing both ontology entities (concepts and properties) as well as their instances. This allows grouping concepts with their well-known instances into self-contained units. An OI -model may include another OI-model, thus making all definitions from the included OI-model automatically available. The mathematical definition of OI-model is given below.

Definition-1: An OI-model (ontology-instance model) is a tuple $OIM := (E, INC)$, where:
 - E is the set of entities of the OI-model.
 - INC is the set of included OI-models.

An OI-model represents a self-contained unit of structured information that may be reused. It consists of entities (the set E in previous definition) and may include other OI-models (represented through the set INC). Different OI-models may talk about the same entity, so the set of entities of these models need not to be disjoint. Note that the set of ontology entities E contains the ontology entities and an instance pool associated with it.

Definition-2: An ontology structure of an OI-model is an 11-tuple:
 $O(OIM) := (C, P, S, T, INV, H_C, H_P, domain, range, mincard, maxcard)$, where:

- $C \subseteq E$ is a set of concepts.
- $P \subseteq E$ is a set of properties.
- $R \subseteq P$ is a set of relational properties, i.e. relations.
- $A = P \setminus R$ is a set of attribute properties, i.e. attributes.

- $S \subseteq R$ is a subset of symmetric properties.
- $T \subseteq R$ is a subset of transitive properties.
- $INV \subseteq R \times R$ is a symmetric relation that relates inverse properties:
if $(p_1, p_2) \in INV$, then p_1 is an inverse property of p_2
- $HC \subseteq C \times C$ is an acyclic relation called concept hierarchy:
If $(c_1, c_2) \in HC$ then c_1 is subconcept (or child) of c_2 , c_2 is superconcept (or parent) of c_1 , HC^* is the reflexive, antisymmetric and transitive closure of HC
- $HP \subseteq P \times P$ is an acyclic relation called property hierarchy:
If $(p_1, p_2) \in HP$ then p_1 is subproperty (or *child property) of p_2 , p_2 is superproperty (or parent property) of p_1 , HP is the reflexive, ntisymmetric and transitive closure of HP
- Function $domain: P \rightarrow 2^C$ gives the set of domain concepts for some property $p \in P$
- Function $range: R \rightarrow 2^C$ gives the set of range concepts for some property $p \in R$
- Function $mincard: CXP \rightarrow N_0$ gives the minimum cardinality for each concept-property pair
- Function $maxcard: CXP \rightarrow (N_0 \cup \{\infty\})$ gives the maximum cardinality for each concept-property pair.

The ontology models the university domain is shown in Fig.3 as example. It contains the set of concepts such as "Professor", "Student", "Project", and a set of properties between them (e.g. "hasFirstName", "includes", etc.). Note that concepts are interpreted as sets of elements whereas properties establish the relations between these elements. Each property may have domain concepts as well as range concepts. For example, the domain concept for the property "includes" is the concept "Project" whereas the range concept is the concept "Person". Property domain and property range both constrain the types of instances to which the properties may be applied. Note that the properties are first class citizens. This means that a property can exist without being attached to any concept. Few properties may be marked as symmetric and/or transitive, and it is possible to say that two properties are inverse of each other. For each concept-property pair, it is possible to specify the minimum (mincard) and maximum (maxcard) cardinalities, defining how many times a property may be specified for instances of that concept. Concepts and properties can be arranged in a hierarchy, as specified by Hc and $*Hp$ r*espectively. This relation relates directly connected concepts (properties) whereas Hc (Hp) is the transitive closure of Hc (Hp). In the example shown in Fig.2, the concept "PhD Student" is a direct child of the concept "Student", and an indirect child of the concept "Person". Therefore, $(\text{"PhD Student"}, \text{"Student"}) \in HC$ and $(\text{"PhD Student"}, \text{"Person"}) \in HC^*$.

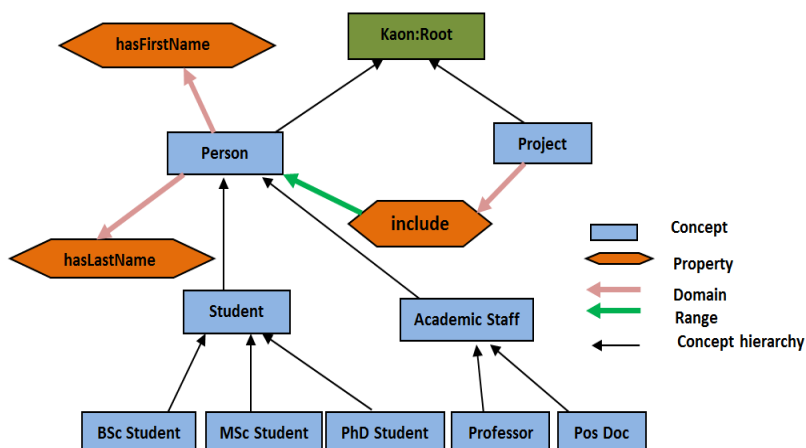


Fig.2: Ontology Model for University Domain.

(ii). **Description Logic based Ontology Languages:** Description logic enables reasoning with concept descriptions and the automatic derivation of classification taxonomies. There are now efficient implementations of description logic reasoners able to perform these tasks. The Ontology Web Language (OWL) inherits from description logic both their formal semantics and efficient reasoning support. Ontology language gives greater machine interpretability of Web content to support for XML, RDF, and RDF Schema. They do this by providing additional vocabulary along with formal semantics.

OWL is possible to implement a semantic description of a specific domain, indicating the concepts and relationships between concepts. There are three particular sublanguages: OWL Lite, OWL DL and OWL Full. We can define "ontology" in connection with the Semantic Web as a system that formally defined hierarchies and relations between different resources. Semantic Web ontology is taxonomy and a set of rules for the inference that the machines can make logical conclusions. Taxonomy, in this context is the system of classification, groups of resources into classes and sub- classes based on their relationship and common property. Since the taxonomy to express the hierarchical relationships between resources, OWL can be used to assign characteristics of classes of resources and allow their sub-classes that inherit the same characteristics.

VII. SEMANTIC OF ONTOLOGY CHANGES

The purpose of the ontology evolution is to ensure that the application of ontology changes should result in an ontology conforming to the set of ontology consistency constraints. An ontology change preserves the ontology consistency if and only if it preserves all constraints of the underlying ontology model. However, applying an ontology change alone will not always leave an ontology in a consistent state. As shown in Fig.3, making the concept "PhD Student" a parent concept of the concept "Person" causes the inconsistency since the invariant IC_2 - *Concept Hierarchy Invariant* related to the cycle in a concept hierarchy would be violated. Namely, the concept "Person" would be in the same time the parent concept of the concepts "PhD Student" (through the concept "Student") as well as its child concept.

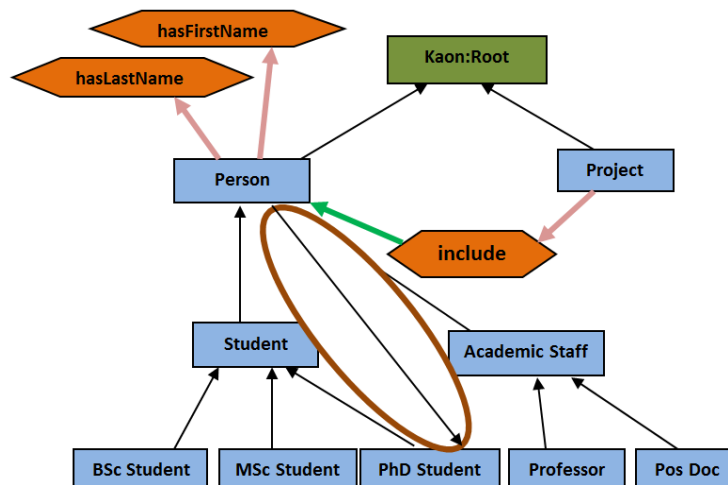


Fig.3: Inconsistent ontology due to a cycle in the concept hierarchy.

Above example shows that the ontology consistency has to be examined. Since verification in general concerns the correctness, the ontology verification is checking of the correctness of an ontology with the respect to the ontology consistency definition. There are two approaches to confirm ontology consistency: The first approach first executes a change, and then validates the updated ontology to check whether it satisfies the consistency constraints. The second approach defines a respective set of preconditions for each change.

VIII. CURRENT STATUS OF SEMANTIC WEB

Current status of the Semantic Web will be described in three domain like research, education and industry as follows:

Research Achievements in Semantic Web:

- W3C have standards for XML, RDF, OWL and active working groups for these and semantic web services e.g. SPARQL and SWRL
- Ontology alignment system and framework
- Ontology alignment API's
- Editor of the development of ontology, alignment and querying
- Regular lectures and workshops with participants from more different nationalities and geographical location

Education Achievement in Semantic Web:

- Virtual institution for semantic web education content curriculum and events organized in a well manner and delivered in the semantic web infrastructure
- Dedicated curriculum for master program and PhD studies
- Annual Summer schools with increasing participation
- Active research teams in the world's best universities

Industry Achievement in Semantic Web:

- Participation of business and economic sector in research development activities of Semantic Web. The business and economic sectors includes, Health, Telecom, Automotive, Energy, Food, Media, Transport, Space, Publishing, Banking, Manufacturing, and Technology
- Oracle introduced the industry's first RDF management system, oracle 10.2 provides the facility to upload the RDF
- Adobe, Google, yahoo all these are organization are using intelligent web
- Jena API of HP, With HP investing more in semantic computing

IX. TECHNOLOGIES ASSOCIATED WITH SEMANTIC WEB

There are lot of technologies available to create semantics on the web. Some of these core technologies are briefly discuss as follows:

- **Unified Modeling language (UML):** UML provides a collection of models and graphs to describe the structural and behavioural semantics of any complex information system. Class and object diagrams are used to specify the semantic structure of the system. Object diagrams are the instance of class diagrams. Use case model and scenarios are used to collect the user requirement and functionality of the system. Scenario is the instance of use case. Activity diagrams specify workflows. Interactions diagrams define that how group of object work together in some behaviour. State diagrams define the dynamic behaviour of an object in system. Physical diagrams define the implementation structure of the system.
- **XML and XML Schema:** These are the tools used to go beyond the fixed-oriented structure of the HTML page that provides vocabulary. With XML, it is possible to describe the structure of data and documents under a personal or Community defined vocabulary. These vocabularies can be a kind of semantics and support for an open exchange of data within communities and the tools to understand the vocabulary.
- **RDF and RDF Schema:** RDF is based on existing XML and URI (Uniform Resource Identifier) technology, a URI to identify all the resources, and the use of URI to make statements about resources. RDF statements are often called triple, which consists of a subject, predicate and object. It does not structure the syntax of the data, but defines semantic meaning for data on the web. Multiple semantic perspectives of the same data are possible. The technology is based on lower level technologies: URI's to identify web resources and Namespaces to identify different vocabularies. RDFS used to create the vocabulary that describes groups of RDF-related resources and the relationship between these resources. An RDF vocabulary permitted defines properties that can be allocated to the RDF resources within a given domain. RDFS types of resources can be created that share common characteristics.
- **Topic Maps:** Topic maps are a form of semantic web technology (in the broadest sense) and some work on interoperability between the W3C RDF / OWL / SPARQL family of standards for the Semantic Web. Topic maps define arbitrarily complex semantic knowledge structures and allow the exchange of information needed to build common and maintaining the index of knowledge. They provide a more comprehensive approach to RDF, in principle, because they are not limited to use on the Web.

X. CONCLUSION

Ontology and Semantic Web leaves its impact on information technology and different research areas such as Knowledge Engineering / Management, Software Agents and Web Services. An important objective of the Semantic Web is to hand over most of the information to software agents that users are doing by themselves now a day on the web. Although the Semantic Web has achieved many milestones, but still it faces many challenges in the form of a vision to reality. Ontologies are becoming popular largely because of what they promise: a shared and common understanding that reaches across people and application systems with the use of semantic web.

ACKNOWLEDGMENT

We would like to acknowledge Dr. R.K. Bisht, Associate Professor (Computer Applications), Amrapali Institute, Haldwani, Uttarakhand, India and Dr. P.P. Dhayni, Director, G.B. Pant Institute of Himalayan Environment & Development, Kosi-Katarmal, Almora, Uttarakhand, India for their continuous support and encouragement for writing this paper.

REFERENCES

- [1] An Hai Doan et al, "Learning to Match Ontologies on the Semantic Web", *The VLDB Journal*, Springer-Verlag, vol. 12, pp. 303-319, 2003.
- [2] Wikipedia, "Semantic Web", http://en.wikipedia.org/wiki/Semantic_Web, cited on 12-09-2008.
- [3] Berners-Lee, "What is the Semantic Web," http://www.altova.com/semantic_web.html, cited on 12-10-08.
- [4] Saleem A, "Semantic Web Vision: survey of ontology mapping systems and evaluation of progress," Master Thesis, Dept. Intelligent Software System, Blekinge Institute of Technology, 2006.
- [5] Peter Norvig, "Semantic Web Ontologies: What Works and What Doesn't" <http://jslee0.blogspot.com/2006/05/semantic-web-ontologies-what-works-and.html>. Cited on 02-12-08.
- [6] R. Mizoguchi, Ontology Engineering Environments. In S. Staab, R. Studer (Eds.) Handbook on Ontologies, Berlin: Springer, 275-298, 2004.
- [7] A.J. Duineveld, R. Stoter, M.R. Weiden, B. Kenepa & VR. Benjamins, WonderTools? A comparative study of ontological engineering tools, *International Journal of Human-Computer Studies*, 52(6), 2000, 1111-1133.
- [8] T. Berners-Lee, J. Hendler, & O. Lassila, The Semantic Web, *Scientific American*, 279(5), 2001, 34-43.
- [9] Z. Zdrahal, P. Mulholland, J. Domingue, & M. Hatala, Sharing engineering design knowledge in a distributed environment, *Behaviour and Information Technology*, 19(3), 2000, 189-200.
- [10] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, L.A. Stein, OWL Web Ontology Language Reference. W3C Recommendation, 2004.
- [11] Guarino N. and P. Giaretta (1995). 'Ontologies and knowledge bases: Towards a terminological clarification', In Mars, N., Editor, IOS Press, p. 25-32.
- [12] Buitelaar, P., Cimiano, P., Magnini B., (2005) *Ontology from Text: Methods, Evaluation and Applications* Frontiers in Artificial Intelligence and Applications Series, Vol. 123, IOS Press.
- [13] O. Lassila and R. R. Swick W3C Resource Description Framework (RDF) Model and Syntax Specification <http://www.w3.org/TR/REC-rdf-syntax>

- [14] A. Madche, S. Staab, R. Studer, Y. Sure and R. Volz SEAL — Tying Up Information Integration and Web Site Management by Ontologies IEEE Data Engineering Bulletin Mar 2002. <http://www.research.microsoft.com/research/db/debull>
- [15] Carr, L., Bechhofer, S., Goble, C., & Hall, W. (2001). Conceptual linking: Ontology-based open hypermedia. Paper presented at the Tenth International World Wide Web Conference, 1-5 May, Hong Kong.
- [16] Dolog, P., Henze, N., Nejdl, W., and Sintek, M. (2004b). Personalization in distributed e-learning environments. Paper presented at the International World Wide Web Conference, 17-22 May, New York, USA
- [17] Kampa, S., Miles-Board, T., Carr, L., & Hall, W. (2001). Linking with meaning: Ontological hypertext for scholars. Technical report, University of Southampton, retrieved October 10, 2004 from <http://www.bib.ecs.soton.ac.uk/data/5163/PDF/lwm.pdf>.
- [18] OWL (2003). OWL, retrieved October 10, 2004 from <http://www.w3.org/2001/sw/WebOnt/>.
- [19] Berners-Lee, T., "Semantic Web Roadmap", <http://www.w3.org/DesignIssues/Semantic.html> cited on 10-09-2008.