Boras, Damir; **Seljan, Sanja**; Lazić, Nikolaj. *Ontologies as the base of Semantic web* // Proceedings of the 28th International Convention MIPRO 2005: Intelligent Systems / Budin, Leo ; Ribarić, Slobodan (ur.). Rijeka : Mipro, 2005. pp.150-154

Ontologies as the base of Semantic web

Damir Boras, Sanja Seljan, Nikolaj Lazić
University of Zagreb, Faculty of Philosophy,
Department of information science,
I.Lucica 3, 10000 Zagreb, Croatia
Tel: +385 1 6120111
e-mail: dboras@ffzg.hr, sseljan@ffzg.hr, nlazic@ffzg.hr

Abstract - World Wide Web is entering its second faze called Semantic web. Today we are talking about new kind of web whose base are ontologies. Broadly speaking, every organized set of subjects can be an ontology, like catalogues and indexes in field of Information retrieval, ER models in Databases, dictionaries and thesauruses in Computational linguistics, object oriented definitions of classes in OO-Programming, etc. Purpose of this paper is to give overview of ontologies, ontological tools and languages, as well as projects dealing with ontologies.

I. INTRODUCTION

When WWW was conceived, it was envisioned not only for the purpose of human-to-human and human-to-machine communication, but also for machine-to-machine communication. With explicitly given information and with the aim of easier machine-to-machine data exchange, semantic Web has become vision of the future Web. Semantic Web is basically an incredible amount of information linked and organized in the efficient way, as much as possible.

Semantic data is embedded into Web documents and describes the properties of the document itself or put on the Web to be reused, repurposed (Palmer, 2001). In case of documents without metadata embedded finding data on the Web, retrieval, indexing and text summarizing is something that cannot be done with machine alone.

The main problem in each of these tasks is lack of machine understanding of semantics. The computer does not understand the meaning of Web pages because of the natural complexity of human languages, because of non-standardized page layout and data presentation and because of different types of data (graphics, multimedia, text). The semantic Web, thought up by Tim Berners-Lee (Barnes-Lee, 2001) (inventor of the URIs, WWW, HTTP) needs work on standardization, improvement and consensus. Although the implementation on the large scale is quite difficult because the global system for publishing data on the Web still does not exist, and if it would exist it couldn't be enforced.

In order to help machines to help people in finding relevant documents one should describe data on the Web in order to ease finding it.

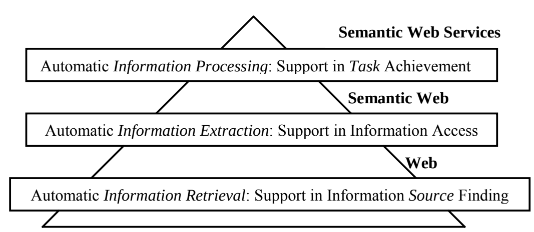
In order to make data more organized and accessible, considerable efforts should be undertaken to express meaning in the machine-processable format:

- creation of common vocabularies (ontologies)
- use of unambiguous names (URIs)
- use of common data model for expressing information (RDF) and especially meta-data.

The Semantic Web is generally built on particular syntax called RDF – Resource Description Framework, defined by World Wide Web Consortium (W3C, 1999) which integrates applications using the XML for syntax for defining customized tagging and the Uniform Resource Identifier – URI for naming, i.e. for data representation.

The main aim of semantic Web is to become flexible enough to represent data, the same as HTML is flexible enough for textual information. The semantic Web, being enough simple and powerful, would provide universal roads for data interchange between diverse systems. This interchange would be enabled by the Semantic Web Logic Language. The first level for the semantic Web is an ontology language which could formally describe the meaning of terminology used in Web documents.

According to Fensel, there are three layer of service that are today available for automatic information retrieval (i.e. finding information).



Picture 1. Three layers of service

The remaining task of extracting the information and using the information to solve a given task are left to the human user. Semantic Web projects like Ontobroker and On-To-Knowledge add an additional level of service on top by providing automated information extraction support, helping the user with information access and interpretation.

Finally, Semantic Web services project like IBROW also provide a reasoning service that supports users in task fulfilment.

II. ONTOLOGIES

The backbone of the Semantic Web is ontology. Ontology is a document or file that formally defines the relations among terms, defines the terms used to describe and represent an area of knowledge. Typical ontology has a taxonomy and a set of inference rules. They provide a shared understanding of knowledge domains that can be communicated between people and computers. Originally their origin lies in philosophy but with the development of artificial intelligence and WWW they have become the major topic in modern computer science. Ontologies could be compared to databases because they both are organized in the way to describe and represent certain knowledge areas in order to be usable by humans and by machines. Therefore, they include computer-usable definitions of terms and relations encoding knowledge that can be reused. Ontologies are explicitly defined in the logic-based language, consistent, detailed, with clear distinctions among classes, properties and relations. Some ontologies can also perform automated reasoning and thus serve for semantic retrieval, language understanding, knowledge management etc. They are of great importance for the Semantic Web because represent the semantics of documents in the way that it could be used by Web applications.

An ontology provides metadata information that describes the semantics of the data, and has similar function to a database schema. What distinguishes ontology from database are the following features/properties (Meersman, 2000):

- language for defining ontologies is syntactically and semantically richer
- information described by an ontology consists of a semi-structured natural language texts
- ontology must be a shared and consensual terminology; and
- ontology provides a domain theory (not just the structure of a datacontainer)

Knowledge representation as understood within the field of AI deals with a wide range of knowledge that is computable i.e. expressed by strict rules of logic. There are for essential features of a knowledge representation language which make them logic:

- **vocabulary** - logic must have collection of symbols represented by characters, words, icons, diagrams, or even sound: 1) domain-independent logical symbols like $\vee$ or $\wedge$; 2) domain-dependent constants which identify individuals, properties, or relations in the universe of discourse or in the application domain; 3) the variables whose range of application is governed by quantifiers; 4) punctuation like commas and parentheses that separate or group the other symbols
- **syntax** - a logic must also have grammar rules or formation rules that determine how the symbols are combined to form the grammatical or well formed sentences. The rules could be stated in a conventional linear grammar, a graph grammar or an abstract syntax that is independent of any concrete notation
- **semantics** - to make meaningful statement, the logic must have a theory of reference that determines how the constants and variables are associated with things in the universe of discourse. It must also have a theory of truth that is able to distinguish true statements from false statements
- **inference rules -** to be more than a notation a logic must include rules to determine how one pattern can be inferred from another. If the logic is sound the rules of inference must preserve truth as determined by the semantics (Sowa, 2000)

An ontology is built following a basic logical procedure and this results in a classification structure with clearly defined classes and conceptual relationships that, for instance, can be expressed through formalised structures called 'conceptual graphs' and formatted in a machine processable way (Sowa, 2000).

As ontologies are formal theories about a certain domain of discourse they require formal logical language to express them. They can be deviden in three categories: first-order predicate logic languages, like CycL and KIF, oriented on predicate logic whose modelling primitive are predicates; frame-based and object-oriented approaches, like Ontolongua and FrameLogic, whoce modeling primitive are classes (i.e. frames) with attributes (i.e. properties); description logic that describes knowledge in terms of concepts and role restrictions used to automatically derive classification taxonomies.

All of the ontology languages must fulfil some important requirements. First, they must be highly intuitive to the human user. Given the current success of the frame-based and object-oriented modelling paradigm, they should have a frame-like look and feel. Secondly, the must have a well-defined formal semantics with established reasoning properties in terms of competences, correctness, and efficiency. Of course, what is maybe most important they must ensure interoperability which means that they need to have a proper link with existing Web languages like XML and RDF.

To link ontology and Web languages several standards have been developed. The most popular ones are:

- XOL - frame-based language with an *XML syntax* for the exchange of ontologies for molecular biology. It includes schema information (metadata), like class definition form object database, as well as, non-schema information (facts) like object definition.

- OIL – unifies three core paradigms: formal semantics and efficient reasoning support (from description logic), epistemologically rich modelling primitives (from frame-based systems) and standard proposal for syntactical exchange notation (from web languages: XML and RDF based syntax).
- DAML+OIL – provides modelling primitives commonly found in frame-based languages (such as an asserted subsumption hierarchy and description or definition o classes through slot fillers) and has a clean defined semantics.
- OWL (Web Ontology Language) (W3C, 2004) – is part of W3C Semantic activity and was derived from DAML+OIL language (based on description logic) and build upon RDFS. W3C is preparing a SQL like language for RDF called SPARQL (W3C, 2004) which will ease searching through ontologies.

When looking at language features both XOL and OIL are XML and Frame-based, while OIL is RDF and description logic based as well. On the other hand DAML+OIL and OWL are based on RDF and description logic. Of all these mentioned, only OWL has the best chances of surviving on the Semantic Web.

Manually building ontologies is difficult and time-consuming task. Natural language text exhibit morphological, syntactic, semantic, pragmatic and conceptual constraints that interact in order to convey a particular meaning to the reader. Ontology editors help human knowledge engineers to build ontologies, they support the definition of concepts hierarchies, the definitions of concept hierarchies, the definition attributes for concepts, and the definition of axioms and constraints. They enable inspection, browsing, codification and modification of ontologies and thus support their development and maintenance. Some of the currently popular editors are:

- **Protégé** - developed by the Knowledge Modelling Group at Stanford Medical Informatics to assist developers in the construction of large electronic knowledge bases. Protégé allows developers to create, browse and edit domain ontologies in a frame-based representation. It allows developers to customize it directly by arranging and configuring the graphical entities in forms that are attached to each class in the ontology for the acquisition of instances. This allows application specialists to enter domain information by filling in the blanks of intuitive forms and by drawing diagrams composed of selectable icons and connectors. Knowledge bases can be stored in several formats, among others RDF. It also has a plugin for exporting in OWL, thus making it able to produce ontologies to be used on the Web.
- **Onto Edit** is an ontology engineering environment developed at the Knowledge Management Group of the University of Karlsruhe. Currently OntoEdit supports representation languages such as F-logic, OIL, and RDFS.
- **WebOnto** is a Java applet coupled with a customized Web server which allows users to browse and edit ontologies over the Web. It was developed at the Knowledge Media Institute of the Open University in Milton Keynes.
- **OilEd** is a simple editor, developed by the University of Manchester, which allows the user to create and edit OIL ontologies. OilEd tends to provide simple and free editor that stimulates and demonstrates the use of DAML+OIL. As it is not intended as a full ontology development environment, it does not support the development of large-scale ontologies, the migration and integration of ontologies, versioning, argumentation and many other activities that are involved in ontology construction.

III. APPLICATION OF ONTOLOGIES

One of possible ontology applications are *ontology based Web portals*. With the main aim to provide different types of information about certain topic and to become starting point, Web portals should be indexed (usually by community members). Tagging the content using simple metatags represents possible way to identify the topic. In order to become more standardized and therefore usable and efficient, it is possible to recommend or to define certain ontology for web indexing, providing therefore terminology for content description.

Ontologies could be also used to provide *semantic annotations* for non-textual objects, i.e. for *multimedia elements* that are usually indexed by caption or metatags. Using certain ontology (media specific and content specific) would help to describe properties of different media in order to improve searching results (e.g. ontology for video description, for description of images of antique furniture, tunes or dance, etc.).

Several interrelated ontologies (of professional terms and of ordinary expressions) could be also used by *large associations and companies*, hotels, agencies, etc. describing their work, documentation, products and case studies in order to index documents and provide better search results.

According to several mentioned examples, there are several ontology characteristics that are always mentioned:

- stability and consistency providing as much as possible balance between expressivity and scalability
- compatibility with other commonly used Web standards
- supporting internationalization, i.e. development of multilingual ontologies suitable for different cultures

- simplicity for maintaining and enlarging with new terms and definitions
- shared ontologies should be publicly available and easy too use

Globalization and advanced Internet features have influenced the increase creation of globally dispersed companies or organizations, are very much dependant on ontologies to enable knowledge share and reuse. Therefore, ontologies can find their application in the fields of: *knowledge management* (for acquiring, maintaining and accessing organization's knowledge together with enabling collaboration while capturing, presenting and interpreting the knowledge resource of their organizations), *web commerce* (enables individual and product search, market transparency, easy access and adaptability); *electronic business* (improving exchange of electronic data between businesses).

As Web languages, namely XML, do not provide standard data structures and terminologies to describe business process and exchange products, ontologies are a of great importance for electronic commerce (Fensel, 2004):

- *Standard ontologies* have to be developed covering the various business areas. In addition to official standards, on-line marketplaces may generate de facto standards. If they can attract significant shares of the on-line transactions in a given business area they will create a true standard ontology for this area.
- *Ontology-based translation services* between different data structures in areas where standard ontologies do not exist or where a particular client wishes to use his own terminology and needs translation service from his terminology into the standard. This translation service must cover structural and semantical as well as language differences.

The ontology-based trading will significantly extend the degree to which data exchange is automated and will create completely new business models in the participating market segments.

IIIa.ON-TO-KNOWLEDGE

On-to-Knowledge was a project in the 5th European Information Society Technologies Framework program. Its goal was to improve information access in digital networks with efficient and effective knowledge management. Main tasks were acquiring (by applying text mining and extraction techniques to extract semantic information form texts), maintaining (enabling automatic maintenance of knowledge by using RDF and XML to describe syntax and semantics of structured information sources) and accessing (using push service and agent technologies) weakly structured on-line information sources through implementation of ontologies. On-to-Knowledge develops a three-layered architecture for information access. First level or information level processes weakly structured information sources to extract machine-processable meta-information from them. Second level or representation level uses this meta-information to provide automatic access, creation, and maintenance of these information sources. The last level or access level uses agent-based techniques as well as state-of the art querying and visualization techniques that fully employ formal annotation to guide user access of information. (Fensel, 2004).

The results of the project are several tools (i.e. software) which are developed to support users both in accessing information and in the maintenance, conversion and acquisition of information sources. These tools are:

- *QuizRDF* is ontology-based tool for knowledge discovery which combines traditional keyword querying of WWW resources with the ability to browse and query against RDF annotations of those resources
- *OntoShare* enables the storage of best practice information to relevant, and browsing or searching the ontology to retrieve relevant information to the problem
- *Spectacle* organizes the presentation of information using ontologies to generate exploration context related to specific task.
- *OntoEdit* enables modification of ontologies at a conceptual level

IIIb. SEMANTIC NETS

Semantic nets, next to electronic dictionaries and corpora represent basis for the computational analysis of the lexical meaning. In the era of electronic media, of information and communication technologies, electronic accessibility of language resources became necessary. With the main aim of providing lexical distinguishing in the text, semantic nets are applied in different language processing fields, such as document retrieval, computational linguistics, indexing and text summarizing, data extraction, machine translation, semantic tagging, for creation of lexicons, thesauri, etc.

One of the semantic net projects is Word Net, developed at Princeton University. Word Net is an *electronic lexical database*, which is considered to be one of the most important resources available to language processing researchers. It is an English lexical database where four types of words (nouns, verbs, adjectives and adverbs) are considered to have lexical meanings. Words are organized into *synonym sets* (*synsets*), each representing one underlying concept.

Word forms in Word Net are represented in their familiar orthography; word meanings are represented by synonym sets (synsets) - lists of synonymous word forms that are interchangeable in some context. Two kinds of relations are recognized: lexical and semantic. While lexical relations hold between word forms; semantic relations hold between word meanings.

Although the synonymy principle is the basic one, nouns are also interrelated with other semantic relations such as antonyms, hyponymy (the semantic relation of being subordinate or belonging to a lower rank or class), meronyms (the semantic relation that holds between a part and the whole), polysemy (the ambiguity of an individual word or phrase that can be used in different contexts to express two or more different meanings), homonymy (sameness of name or designation of things or persons which are different), etc.

Semantic nets are also translated into ontologies and in that way represent linguistic knowledge that could be used to make ontologies for translating terms. Existing ontologies can than be used on another set of terms by just including translation ontology.

For Croatian language there is ongoing work on collecting linguistic data of written language and planning on building something like WordNet (Šojat, K. et al, 2004).

IIIc. ONTOLOGY INTEROPERABILITY

Semantic Web Advanced Development projekt in Europe (named SWAD-Europe Thesaurus activity) aims for creating thesaurus for mappings between different ontologies. This approach can help agent-based technologies in finding data in not so obviously connected documents.

Similar path is taken by SIMILE project (Semantic Interoperability of Metadata and Information in unLike Environments). It consists of several projects spanning from RDF browser, RDF modeller , XML to RDF dataset converter and digital repository manager.

IV. CONCLUSION

The Semantic Web actually represents future vision of the more useful Web: with machine-processable meaning data could be automatically more easily retrieved, shared and combined at the global level. Ontologies have their application in content management, in language processing, in information retrieval, computational linguistics, machine translation, etc. Since Web represents international tool, the semantic Web should enable better search, communication and data exchange between cultures, nations and professionals. Supporting compatibility with commonly used standards, as well as internationalization and integration with other stable and consistent ontologies - this could be one step to enhance Web functionality in order to help people organize and reorganize knowledge.

REFERENCES:

1) Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. Scientific American, May, 2001. (http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2)
2) Brooks, A.B. The Semantic Web, universalist ambition and some lessons from librarianship. Information Research, 7(4), 2002. (http://InformationR.net/ir/7-4/paper136.html)
3) Fensel, D. Ontologies: A silver bullet for knowledge management and electronic commerce. Berlin: Springer Verlag, 2004.
4) Meersman, R.A. The use of lexicons and other computer-linguistic tools in semantics, design and cooperation of database systems. // CODAS Conference Proceedings. Berlin: Springer Verlag, 2000.
5) OWL Web Ontology Language: Use Cases and Requirements. W3C Recommendation, 2004. (http://www.w3.org/TR/webont-req)
6) Palmer, Sean B. The Semantic Web: An Introduction, 2001. (http://infomesh.net/2001/swintro).
7) RDF Primer. W3C, 2004 (http://www.w3.org/TR/rdf-primer/)
8) Semantic Web Development: Technical Proposal. Principal Investigator: Tim Berners-Lee, MIT W3C, 2000. (http://www.w3.org/2000/01/sw/DevelopmentProposal)
9) SIMILE Project. (http://simile.mit.edu)
10) Sowa, J. F. Knowledge representation: logical, philosophical and computational foundations. Pacific Grove, CA: Brooks/Cole, 2000.
11) SPARQL Query Language for RDF. W3C, 2004. (http://www.w3.org/TR/rdf-sparql-query/)
12) SWAD-Europe Thesaurus Activity. (http://www.w3.org/2001/sw/Europe/reports/thes)
13) Šojat, K; Bekavac, B; Tadić, M. Zašto nam treba hrvatski WordNet? Zbornik radova HDPL, Split, 2004.
14) Word Net (http://wordnet.princeton.edu/)