In: Fensel, D, Hendler, J. Lieberman, H., Wahlster, W. (eds.) (2003): Spinning the Semantic Web.
Bringing the World Wide Web to Its Full Potential. Cambridge: MIT Press 2003, pp. 1 - 25.

SPINNING THE SEMANTIC WEB

INTRODUCTION

Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster

The World Wide Web (WWW) has drastically changed the availability of electronically accessible information. Currently there are around three billion static documents in the WWW that are used by more than 200 million users internationally, and this number is growing astronomically. In 1990, the WWW began with a small number of documents as an in-house solution for around a thousand users at CERN. By 2002, W3C (http://www.w3c.org) expects around a billion Web users and an even higher number of available documents. This success and exponential growth makes it increasingly difficult, however, to find, to access, to present, and to maintain information of use to a wide variety of users. Currently, pages on the Web must use representation means rooted in format languages such as HTML or SGML and employ protocols that allow browsers to present information to human readers. The information content, however, is mainly presented via natural language. Thus, there is a wide gap between the information available for tools to use in creating and maintaining Web pages and the information kept in human readable form on those pages, a gap that causes serious problems in accessing and processing the available information:

▪ *Searching for information*. Already, finding the right piece of information on the Web is often a nightmare. In searching the Web for specific information, one gets lost in huge amounts of irrelevant material and may often miss the relevant matter. Searches are imprecise, often returning pointers to many thousands of pages (and this situation worsens as the Web grows). In addition, a user must read through the retrieved documents to extract the desired information - so even once a truly relevant Web page is found, the search may be difficult or the information obscured. Thus, the same piece of knowledge must often be presented in different contexts on the same Web page and adapted to different users' needs and queries. However, the Web lacks automated translation tools to allow this information to be transformed automatically among different representation formats and contexts.

▪ *Presenting information*. A related problem is that the maintenance of Web sources has become very difficult. Keeping redundant information consistent and keeping information correct is hardly supported by current Web tools, and thus the burden on a Webmaster to maintain consistency is often overwhelming. This leads to a plethora of sites with inconsistent and/or contradictory information.

▪ *Electronic commerce*. Automatization of electronic commerce is seriously hampered by the way information is currently presented. Shopping agents use wrappers and heuristics to extract product information from weakly structured textual information. However, the development and maintenance costs involved are high and the services provided are limited. Business-to-business marketplaces offer new possibilities for electronic commerce; however, they are hampered by the large and increasing mapping costs required to integrate heterogeneous product descriptions.

There is an emerging awareness that providing solutions to these problems, requires that there be a machine-understandable semantics for some or all of the information presented in the WWW. Achieving such a Semantic Web (Berners-Lee 1999) requires

▪ developing languages for expressing machine-understandable metainformation for documents and developing terminologies (i.e., namespaces, or ontologies) using these languages and making them available on the Web,

▪ developing tools and new architectures that use such languages and terminologies to provide support in finding, accessing, presenting, and maintaining information sources.

▪ realizing applications that provide a new level of service to the human users of the semantic Web.

Developing such languages, ontologies, and tools is a wide-ranging problem that touches on the research areas of a broad variety of research communities. Therefore the development of this book brought together colleagues from these different research communities, including those in the areas of databases, intelligent information integration, knowledge representation, knowledge engineering, information agents, knowledge management, information retrieval, natural-language processing, metadata, and Web standards, as well as others. The book is based on a seminar held in Dagstuhl, Germany, in March 2000. The contents of the book are organized as follows. First, a number of arising Web standards are discussed that should improve the representation of machine-processible semantics of information. Second, ontologies are introduced for representation of semantics (in the sense of formal and real-world semantics) in these formalisms. Third, as these semantic annotations allow automatization in information access and task achievement, we discuss intelligent information access based on them. Finally, a number of applications of these new techniques are presented.

The purpose of this chapter is to provide an overall motivation for the book's subject. First, in section 1.1, we discuss in further depth the need for a Semantic Web, mainly as motivated by the shortcomings of the current state of the WWW. We show which kind of new services the Semantic Web will enable, and in section 1.2 we explain how they can be developed.

1.1 Why Is There a Need for the Semantic Web and What Will It Provide?

The Web has brought exciting new possibilities for information access and electronic commerce. It is the Web's simplicity that has fueled its quick uptake and exponential growth, but this same simplicity also seriously hampers its further growth. Here we discuss these bottlenecks with respect to knowledge management and electronic commerce (see Fensel 2001 for more further details).

1.1.1 Knowledge Management

Knowledge management is concerned with acquiring, maintaining, and accessing the knowledge of an organization. It aims to exploit an organization's intellectual assets for greater productivity, new value, and increased competitiveness. Because of globalization and the universal availability of the Internet, many organizations are increasingly geographically dispersed and organized around virtual teams. With the large number of documents made available online by organizations, several document management systems have entered the market. However, these systems have severe weaknesses:

▪ *Searching information*. Existing keyword-based search retrieves irrelevant information that uses the keyword in a context other than the one in which the searcher is interested or may miss relevant information that employs words other than the keyword in discussing the desired content.

▪ *Extracting information*. Human browsing and reading is currently required to extract relevant information from information sources, as automatic agents lack the commonsense knowledge required to extract such information from textual representations and fail to integrate information spread over different sources.

▪ *Maintenance*. Maintaining weakly structured text sources is a difficult and time-consuming activity when such sources become large. Keeping such collections consistent, correct, and up to date requires a mechanized representation of semantics and constraints that help to detect anomalies.

▪ *Automatic document generation*. Adaptive Web sites that enable a dynamic reconfiguration of information according to user profiles or other relevant aspects would be very useful. The generation of semistructured information presentations from semistructured data would require a machine-accessible representation of the semantics of these information sources, and such a representation currently does not exist.

Semantic Web technology will enable structural and semantic definitions of documents providing completely new possibilities: intelligent search instead of keyword matching, query answering instead of information retrieval, document exchange among departments via ontology mappings, and definition of customized views on documents.

1.1.2 Web Commerce

Electronic commerce (B2C) is an important and growing business area for two reasons. First, it is extending existing business models. It reduces costs and extends existing distribution channels and may even introduce new distribution possibilities. One example of such a business field extension is online stores. Second, it enables completely new business models or gives them a much greater importance than they had before. What has up to now been a peripheral aspect of a business field may suddenly receive its own important revenue flow. Examples of new business fields generated by electronic commerce are shopping agents, online marketplaces, and auction houses, which make comparison shopping or meditation of shopping processes into a business with its own significant revenue flow. The advantages of online stores and the success many have experienced has led to a large increase in the

number of such shopping pages. The task for the new Web customer is now to find a shop that sells the product he is looking for, get it in the desired quality and quantity and at the desired time, and pay as little as possible for it. Achieving these goals via browsing requires significant time and even with a sizeable time investment, a customer will cover only a small share of the actual Web offerings. Very early on in B2C development, shopbots were developed that visit several stores, extract product information, and present to the customer an instant market overview. Their functionality is provided via wrappers written for each online store. Such wrappers use a keyword search, together with assumptions on regularities in the presentation format of stores' Web sites and text extraction heuristics, to find information about the requested product and return it to the customer. However, this technology has two severe limitations:

▪ *Effort*. Writing a wrapper for each online store is a time-consuming activity, and changes in the layout of stores result in high levels of required maintenance to keep the wrappers up to date.

▪ *Quality*. The product information extracted by shopbots using such technology is limited (mostly price information), error prone, and incomplete. For example, a wrapper may extract the direct price of a product but miss indirect costs such as shipping.

These problems are caused by the fact that most product information on Web sites is provided in natural language, and automatic text recognition is still a research area with significant unsolved problems. What is required is machine-processible semantics for the information provided. The situation will drastically change when standard representation formalisms for the structure and semantics of data are available. Software agents can then be built that can "understand" the product information the Web sites provide. Meta-online stores can then be constructed with little effort, and this technique will also enable complete market transparency in various dimensions of diverse product properties. The low-level programming of wrappers based on text extraction and format heuristics will be replaced by semantic mappings that translate different formats used to represent products and can be used to navigate and search automatically for the required information.

1.1.3 Electronic Business

Electronic commerce in the business-to-business field (B2B) is not a new phenomenon. Initiatives to support electronic data exchange in business processes among different companies existed already even as long ago as the 1960s. To exchange information about business transactions, sender and receiver have to agree on a common standard (a protocol for transmitting the content and a language for describing the content). A number of standards arose for this purpose; one is the United Nations initiative Electronic Data Interchange for Administration, Commerce, and Transport (EDIFACT). In general, the automatization of business transactions has not lived up to the expectations of its propagandists. This can be explained by the serious shortcomings of an existing approach like EDIFACT: it is a rather procedural and cumbersome standard, making the programming of business transactions expensive, error prone, and hard to maintain. It assumes that business data are exchanged via special networks (extranets), which are not integrated with other document exchange processes, that is, EDIFACT is an isolated standard. Using the infrastructure of the Internet for business exchange significantly improved this situation. Standard browsers can be used to render specifications for business transactions, and these

transactions can be transparently integrated into other document exchange processes in intranet and Internet environments. However, data exchange is currently hampered by the fact that HTML does not provide a means for presenting rich syntax and semantics of data. XML, which is designed to close this gap in current Internet technology, is already changing the situation. B2B communication and data exchange can then be modeled with the same means that are available for other data exchange processes, transaction specifications can easily be rendered by standard browsers, and maintenance will be cheap. XML provides a standard serialized syntax for defining the structure and semantics of data. Therefore, it provides means to represent the semantics of information as part of defining its structure. However, XML does not provide standard data structures and terminologies to describe business processes and exchanged products. Therefore, new Semantic Web technology will have to play important roles in XML-enabled electronic commerce:

▪ First, languages with a defined data model and rich modeling primitives will have to be defined that provide support in defining, mapping, and exchanging product data.

▪ Second, standard ontologies will have to be developed covering various business areas. Examples are Common Business Library (CBL), Commerce XML (cXML), ecl@ss, Open Applications Group Integration Specification (OAGIS), RosettaNet, and UN/SPSC. However, these "ontologies" are quite specific and provide only partial coverage of the domains, with quite limited semantics.

▪ Third, efficient translation services will be required in areas for which standard ontologies do not exist[1] or in which a particular client wants to use his own terminology and needs his terminology translated into the standard.

This translation service will have to cover structural and semantical as well as language differences.

Such support will significantly extend the degree to which data exchange is automated and will create complete new business models in the participating market segments.

The Semantic Web deals with important application areas such as knowledge management and electronic commerce (both B2C and B2B). It may help to overcome many of the current bottlenecks in these areas. The next section will explain how it can help do this.

1.2 How the Semantic Web Will Be Possible

In the preceding section we described new services provided by the Semantic Web. In this section we will discuss how such a new level of functionality can be achieved. First, we describe new languages that allow semantics to be added to the Web. Second, we describe important tools for adding semantics to the Web, and finally, we illustrate by some applications the potential utility of the Semantic Web.

1.2.1 Languages

Languages for the Semantic Web must include two aspects. First, they need to provide formal syntax and formal semantics to enable automated processing of their content. Second,

they need to provide standardized vocabulary referring to real-world semantics enabling automatic and human agents to share information and knowledge. The latter is provided by ontologies.

1.2.1.1 Formal Languages

Originally, the Web grew mainly around HTML, which provide a standard for structuring documents that was translated by browsers in a canonical way to render documents. On the one hand, as noted above, it was the simplicity of HTML that enabled the fast growth of the WWW. On the other hand, HTML's simplicity has seriously hampered more advanced Web application in many domains and for many tasks. This was the reason for defining another language, XML (see figure 1.1), which allows arbitrary domain- and task-specific extensions to be defined (as the figure shows, even HTML got redefined as an XML application, XHTML). Therefore, it is just a logical consequence to define the semantic Web as an XML application. The first step in this direction is taken by RDF, which defines a syntactical convention and a simple data model for representing machine-processible semantics of data. A second step is taken by the RDF Schema (RDFS), which defines basic ontological modeling primitives on top of RDF. A full-blown ontology modeling language as extension of RDFS is defined by the Ontology Inference Layer (OIL) and DARPA Agent Markup Language-Ontology (DAML-ONT), which conclude our discussion on Semantic Web languages.

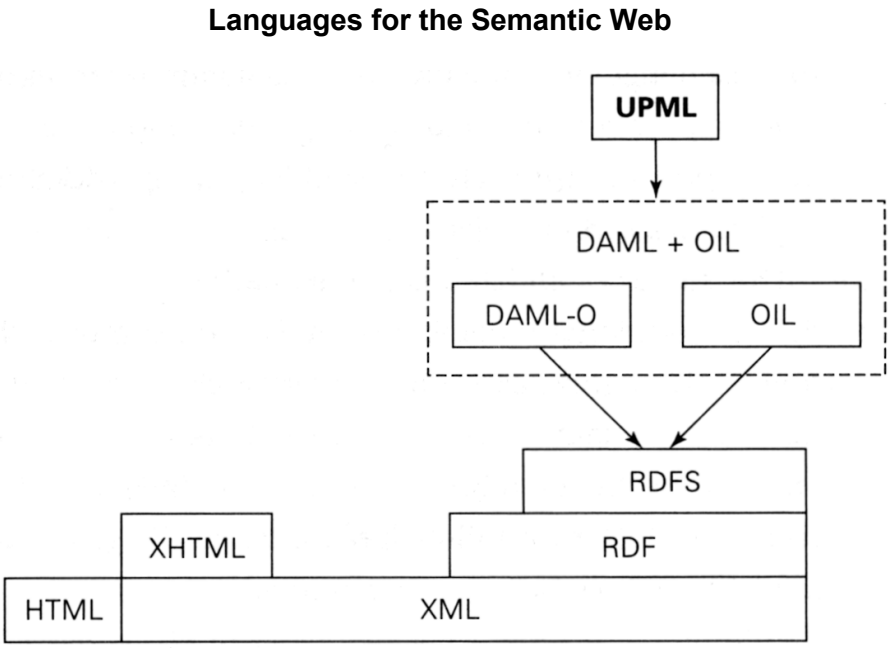

Figure 1.1
Layer language model for the WWW.

RDF is a standard for Web metadata developed by W3C (Lassila 1998). Expanding from the traditional notion of document metadata (such as something like library catalog information), RDF is suitable for describing any Web resources, and as such provides interoperability among applications that exchange machine-understandable information on the Web. RDF is

an XML application and adds a simple data model on top of XML. This data model provides three elements: objects, properties, and values of properties applied to a certain object.

The RDFS candidate recommendation (see Brickley and Guha 2000) defines additional modeling primitives on top of RDF. It allows the definition of classes (i.e., concepts), inheritance hierarchies for classes and properties, and domain and range restrictions for properties. OIL (http://www.ontoknowledge.org/oil) (see Fensel et al. 2001) takes RDFS as a starting point and extends it to a full-fledged ontology language. An ontology language must fulfill three important requirements:

▪ It must be highly intuitive to the human user. Given the current success of the frame-based and object-oriented modeling paradigm, it should have a framelike look and feel.

▪ It must have a well-defined formal semantics with established reasoning properties in terms of completeness, correctness, and efficiency.[2]

▪ It must have a proper link with existing Web languages like XML and RDF, ensuring interoperability.

In this respect, many of the existing ontology languages like CycL (Lenat and Guha 1990), the Knowledge Interchange Format (KIF) (Genesereth 1991), Ontolingua (Farquhar, Fikes, and Rice 1997), and Simple HTML Ontology Extensions (SHOE) (Luke, Spector, and Rager 1996) fail to satisfy these requirements. However, OIL fulfills all three criteria mentioned above. OIL unifies three important aspects provided by different communities: epistemologically rich modeling primitives as provided by the frame community, formal semantics and efficient reasoning support as provided by description logics, and a standard proposal for syntactical exchange notations as provided by the Web community.

Another candidate for such a Web-based ontology modeling language is DAML-ONT (http://www.daml.org) funded by the U.S. Defense Advanced Research Projects Agency (DARPA). However, this language is still in an early stage of development and lacks a formal definition of its semantics.

1.2.1.2 Ontologies

Ontologies were developed in artificial intelligence to facilitate knowledge sharing and reuse. Since the beginning of the 1990s, ontologies have become a popular topic for investigation in artificial intelligence research communities, including knowledge engineering, natural-language processing, and knowledge representation. More recently, the notion of ontology has also become widespread in fields such as intelligent information integration, cooperative information systems, information retrieval, electronic commerce, and knowledge management. The reason ontologies are becoming so popular has to do in large part with what they promise: a shared and common understanding of some domain that can be communicated among people and application systems. Because ontologies aim at consensual domain knowledge, their development is often a cooperative process involving different people, possibly at different locations. People who agree to accept an ontology are said to "commit" themselves to that ontology.

Many definitions of ontologies have been offered in the last decade, but the one that, in our opinion, best characterizes the essence of an ontology is based on the related definitions by Gruber (1993): An ontology is a formal, explicit specification of a shared conceptualization. A "conceptualization" refers to an abstract model of some phenomenon in

the world that identifies the relevant concepts of that phenomenon. "Explicit" means that the type of concepts used and the constraints on their use are explicitly defined. "Formal" refers to the fact that the ontology should be machine understandable. Different degrees of formality are possible. Large ontologies like WordNet (http://www.cogsci.princeton.edu/~wn) provide a thesaurus for over 100,000 terms explained in natural language. On the other end of the spectrum is CYC (http://www.cyc.com), which provides formal axiomating theories for many aspects of commonsense knowledge. "Shared" reflects the notion that an ontology captures consensual knowledge, that is, it is not restricted to some individual but accepted by a group.

1.2.2 Tools

Effective and efficient work with the semantic Web must be supported by advanced tools enabling the full power of this technology. In particular, it requires the following elements:

▪ Formal languages to express and represent ontologies (We already discussed some of these in the last section)

▪ Editors and semiautomatic construction to build new ontologies

▪ Reusing and merging ontologies (ontology environments that help to create new ontologies by reusing existing ones)

▪ Reasoning services (instance and schema inferences that enable advanced query answering service, support ontology creation, and help map between different terminologies)

▪ Annotation tools to link unstructured and semistructured information sources with metadata

▪ Tools for information access and navigation that enable intelligent information access for human users

▪ Translation and integration services between different ontologies that enable multistandard data interchange and multiple view definitions (especially for B2B electronic commerce).

In the following sections, we will briefly describe examples of these technologies.

1.2.2.1 Editors and Semiautomatic Construction

Ontology editors help human knowledge engineers build ontologies. They support the definition of concept hierarchies, the definition attributes for concepts, and the definition of axioms and constraints. They enable the inspection, browsing, codifying, and modification of ontologies and in this way support the ontology development and maintenance task. To be useful in this context, they must provide graphical interfaces and must conform to existing standards in Web-based software development. One example of an ontology editor that fulfills all of these criteria is Protégé (Grosso et al. 1999), developed at Stanford University, which allows domain experts to build knowledge-based systems by creating and modifying reusable ontologies and problem-solving methods. Protégé generates domain-specific

knowledge acquisition tools and applications from ontologies. It has been used in more than 30 countries. It is an ontology editor that can be used to define classes and class hierarchies, slots and slot value restrictions, relationships between classes, and properties of these relationships (see figure 1.2). Protégé's instances tab is a knowledge acquisition tool that can be used to acquire instances of the classes defined in the ontology.

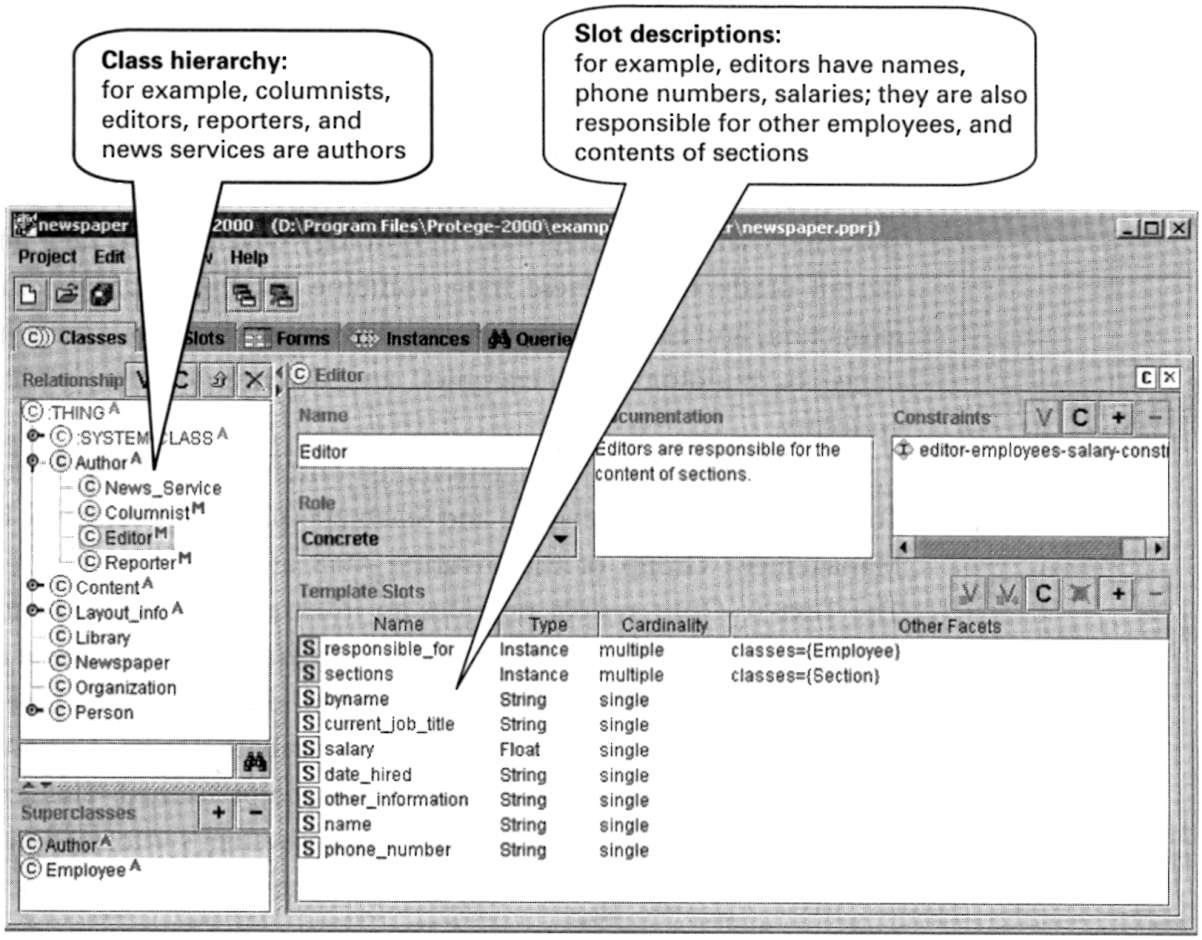

Figure 1.2
Protégé editor

Manually building ontologies is a time-consuming task. It is very difficult and cumbersome to manually derive ontologies from data. This appears to be true regardless of the type of data under consideration. Natural-language texts exhibit morphological, syntactic, semantic, pragmatic, and conceptual constraints that interact to convey a particular meaning to the reader. Thus, such texts transport information to the reader, and the reader embeds this information into his background knowledge. Through the understanding of the text, data are associated with conceptual structures and new conceptual structures are learned from the interacting constraints given through language. Tools that learn ontologies from natural language exploit the interacting constraints on the various language levels (from morphology to pragmatics and background knowledge) in order to discover new concepts and stipulate relationships among concepts. Therefore, in addition to editor support, such semiautomated tools in ontology development help improve the overall productivity. These tools combine machine learning, information extraction, and linguistic techniques. Their main tasks are extracting relevant concepts, building is-a hierarchies, and determining relationships among concepts.

An example of such a semiautomated ontology development tool is Text-To-Onto (figure 1.3) (Mädche and Staab 2000), developed by the Knowledge Management Group of the Institute AIFB at the University of Karlsruhe. The Text-To-Onto system provides an integrated environment for the task of learning ontologies from text. The system's text management module enables the selection of a relevant corpus of domain texts. These texts may be both natural-language texts and HTML-formatted texts. A meaningful text analysis requires that textual preprocessing be performed. The text management module serves as an interface with the system's information extraction server. If a domain lexicon already exists, the information extraction server performs domain-specific parsing. The results of the parsing process are stored in XML or feature value structures.

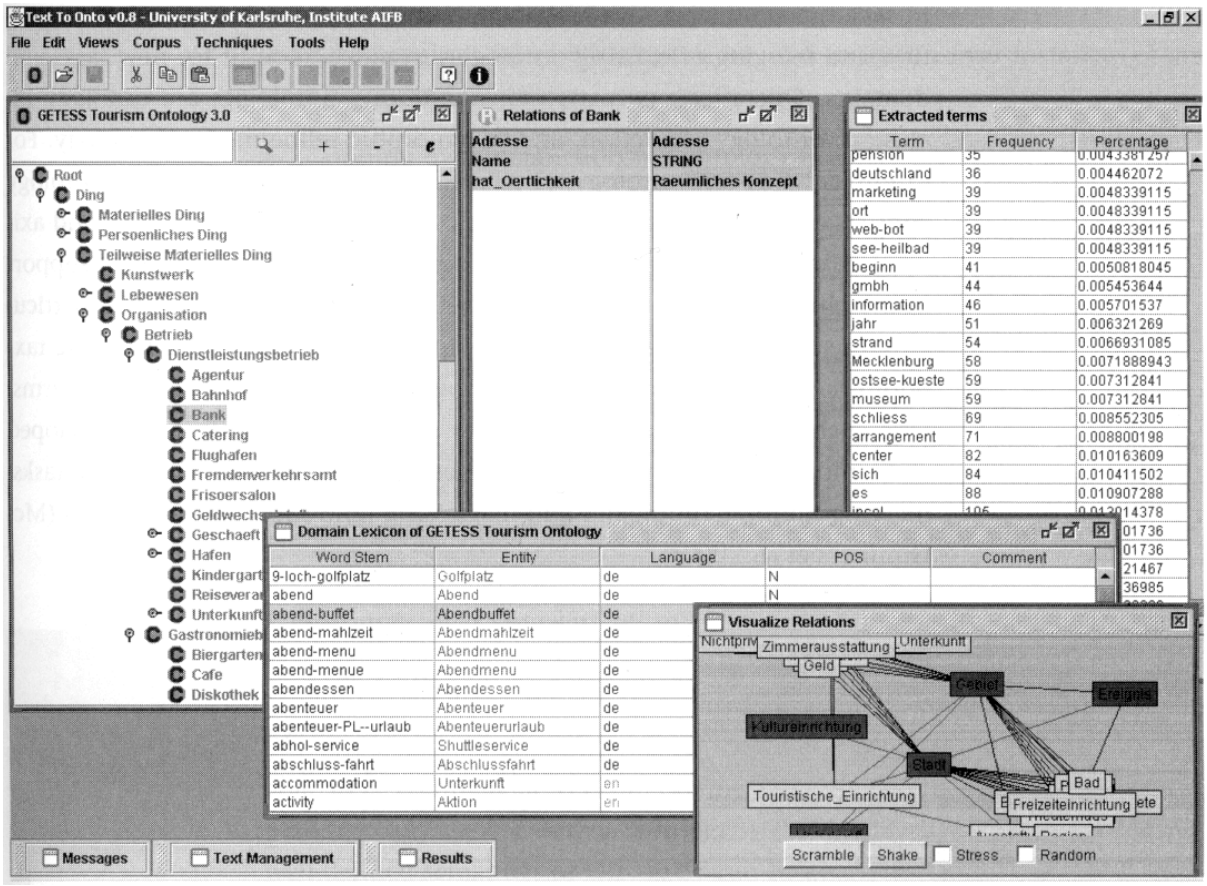

Figure 1.3
Text-To-Onto.

Text-To-Onto's management module offers all existing learning components to the user. Typically these components are parameterizable. Existing knowledge structures (for example, a taxonomy of concepts) are incorporated as background knowledge. The system's learning component discovers, on the basis of the domain texts it processes, new knowledge structures, which are then captured in the ontology modeling module to expand the existing ontology.

1.2.2.2 Ontology Environments

Assuming that the world is full of well-designed modular ontologies, constructing a new ontology is a matter of assembling existing ones. Instead of building ontologies from scratch,

one wants to reuse existing ontologies to save time and labor. Tools that support this approach must allow adaptation and merging of existing ontologies to make them fit for new tasks and domains. Operations necessary for combining ontologies are ontology inclusion, ontology restriction, and polymorphic refinement of ontology. For example, when one ontology is included in another, the composed ontology consists of the union of the two ontologies (their classes, relations, and axioms). The knowledge engineer needs a number of different kinds of support in merging multiple ontologies together and diagnosing ontologies, particularly in such tasks as using ontologies in differing formats, reorganizing taxonomies, resolving name conflicts, browsing ontologies, and editing terms. One such ontology environment tool is Chimaera (figure 1.4), developed at Stanford University, which provides support for two important tasks: merging multiple ontologies and diagnosing (and evolving) ontologies (McGuinness et al. 2000).

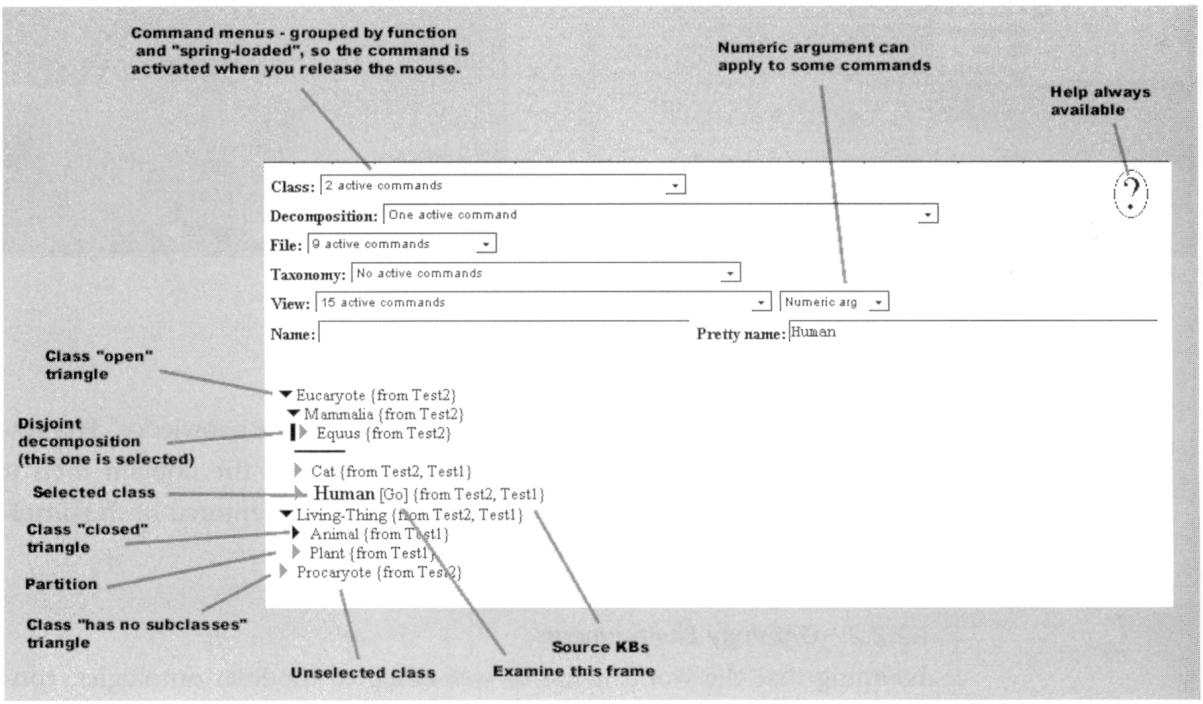

Figure 1.4
Chimaera.

1.2.2.3 Reasoning Services

Inference engines for ontologies can be used to reason over instances of an ontology or over ontology schemes.

▪ Reasoning over instances of an ontology involves deriving a certain value for an attribute applied to an object. Inference services of this type can be used to answer queries about the explicit and implicit knowledge specified by an ontology. The powerful support it provides in formulating rules and constraints and in answering queries over schema information is far beyond that available in existing database technology. These inference services are the equivalent of SQL query engines for databases, however, they provide stronger support (for example, recursive rules) than such query engines. An example of a system for reasoning

over instances of an ontology is Ontobroker (Fensel, Angele, et al. 2000), available commercially through the company Ontoprise (http://www.ontoprise.de).

▪ Reasoning over concepts of an ontology automatically derives the right position for a new concept in a given concept hierarchy. One system with such a capacity, FaCT (Fast Classification of Terminologies) (Horrocks and Patel-Schneider 1999), developed at the University of Manchester and available in a commercial version can be used to derive concept hierarchies automatically. It is a description logic (DL) classifier that makes use of the well-defined semantics of OIL. FaCT can be accessed via a Corba interface. It has been developed at the University of Manchester and currently an internet start up may go for implementing a commercial version. It is one of the most, if not *the* most, efficient reasoner for the kinds of tasks it handles.

Both types of reasoners help to build ontologies and to use them for advanced information access and navigation, as we discuss below.

1.2.2.4 Annotation Tools

Ontologies can be used to describe a large number of instances. Annotation tools help the knowledge engineer to establish such links via:

▪ linking an ontology with a database schema or deriving a database schema from an ontology (in cases of structured data)
▪ deriving an XML DTD, an XML schema, and an RDF schema from an ontology (in cases of semistructured data)
▪ manually or semiautomatically adding ontological annotation to unstructured data

More details can be found in Erdmann and Studer 2001 and Klein et al. 2000.

1.2.2.5 Tools for Information Access and Navigation

The Web is currently navigated at a very low level: clicking on links and using keyword searches is the main (if not the only) navigation technique. It is comparable programming with assembler and go-to instructions instead of higher-level programming languages. This low-level interface may significantly hamper the growth of the Web in the future for a number of reasons:

▪ Keyword-based search retrieves irrelevant information that uses a particular word in a different meaning from the one intended, and it may miss relevant links in which different words than the keyword are used to describe the content for which the user is searching. Navigation is supported only by predefined links; current navigation technology does not support clustering and linking of pages based on semantic similarity.

▪ Query responses require human browsing and reading to extract the relevant information from the information sources returned. This burdens Web users with an additional loss of time and seriously limits information retrieval by automatic agents, which lack all commonsense knowledge required to extract such information from textual representations.

▪ Keyword-based document retrieval fails to integrate information spread over different sources.

▪ Current retrieval services can retrieve only information that is directly represented on the WWW. No further inference service is provided for deriving implicit information that must be derived from the explicit text.

Ontologies help to overcome these bottlenecks in information access. They support information retrieval based on the actual content of a page. They help the user navigate the information space based on semantic, rather than lexical, concepts. They enable advanced query answering and information extraction services, integrating heterogeneous and distributed information sources enriched by inferred background knowledge. This provides two main improvements over current methods:

▪ Semantic information visualization, which groups information not on location but on contents, providing semantic-based navigation support. Examples are the hyperbolic browsing interface of Ontoprise (see figure 1.5) and the page content visualization tool of Aidministrator (http://www.aidministrator.nl) (see figure 1.6).

▪ Direct query answering services based on semistructured information sources.

1.2.2.6 Translation and Integration Services

Around 80% of the Web's electronic business will be in the B2B area, in which all experts expect exponential growth. Many studies estimate that around 10,000 B2B marketplaces will be set up during the next few years. However, there is one serious obstacle to the projected growth: the heterogeneity of product descriptions on Web sites and the exponentially increasing effort that must be devoted to mapping these heterogeneous descriptions as the number of Web sites increases. Therefore, effective and efficient content management of heterogeneous product catalogues is the critical point for B2B success. Traditional B2B did not change the business model of the companies involved: it only helped reduce the transaction costs associated with the existing model. It required one mapping from one supplier to one customer or *N* mappings from one supplier to *N* customers. The new business model of B2B marketplaces, in contrast, changes the business model, bringing electronic commerce to its full economical potential: individual product search, corporative product search, market transparency, easy access, and negotiation.[3]

An Internet-based marketplace can help significantly to bring the sides of a business interaction together. It will provide instant market overview and offers comparison shopping. Such a marketplace will significantly change the business model of this market segment where it operates. Basically, it will replace or at least compete with traditional mediation agents, like wholesale traders. However, the number of required mappings will explode in comparison to that required in traditional B2B.

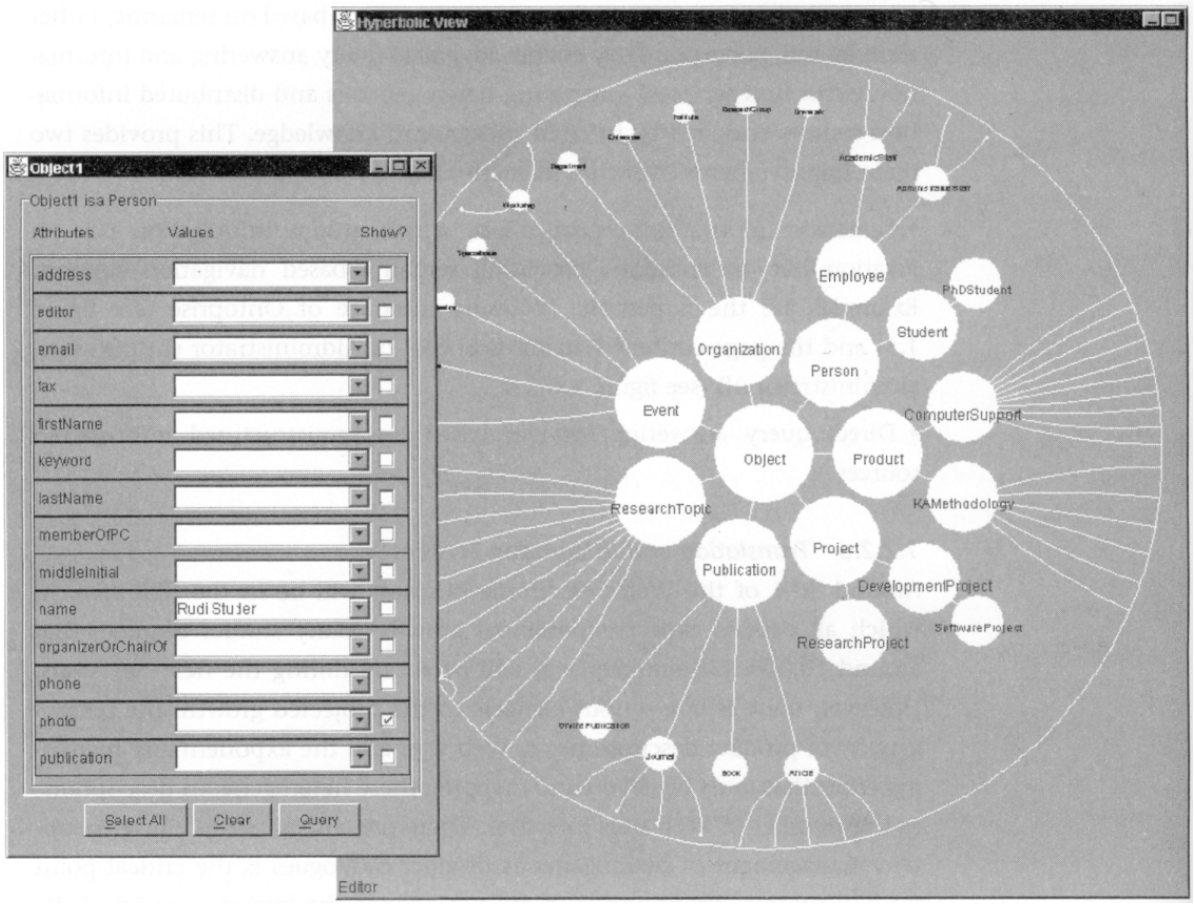

Figure 1.5
Hyperbolic browsing interface.

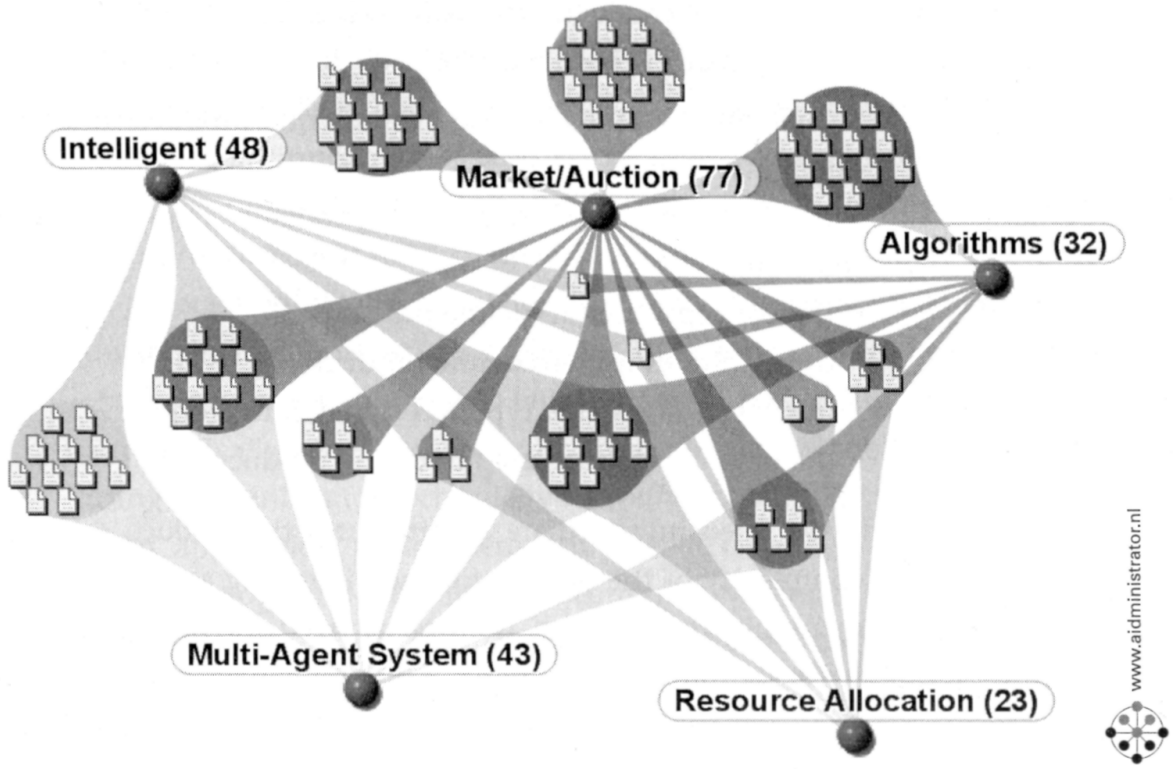

Figure 1.6
Automatically generated semantic structure maps.

In an Internet-based marketplace, M companies will exchange business transactions electronically with N companies in a fragmented market. In consequence one will need $M * N$ mappings. These mappings will arise at two levels:

▪ Different representations of product catalogs must be merged, as different vendors may use different representation of their catalog data. For example:

- A product catalog in EXPRESS must be merged with a product catalog in XML.
- A product catalogue in XML with DTD1 must be merged with a product catalogue in XML with DTD2.
- Different vocabularies used to describe products must be merged. Differences may appear in
- the languages used to describe products (English, Spanish, French, German, etc.)
- the concepts used to define products
- the attributes used to define products
- the values and value types used to define products
- the overall structure used to define products.

We need intermediate architectures that reduce drastically the inherent complexity of the process for each mapping and that reduce the number of mappings itself. Given the urgent need for flexible tools for mapping between ontologies, not many actual tools have been developed. A promising approach based on a metalevel architecture is described in Bowers and Delcambre 2000.

1.2.3 Applications

At the beginning of the chapter we sketched three application areas for Semantic Web technologies: knowledge management, B2C Web commerce, and B2B electronic business. This section provide some prototypical examples for such applications. It is not meant as a representative survey of the field, which would require much more space and would be a chapter (if not a book) all its own.

On-To-Knowledge[4] (Fensel, van Harmelen, et al. 2000) builds an environment for knowledge management in large intranets and Web sites. Unstructured and semistructured data are automatically annotated, and agent-based user interface techniques and visualization tools help the user navigate and query the information space. On-To-Knowledge continues a line of research that was initiated with SHOE (Luke, Spector, and Rager, 1996) and Ontobroker (Fensel et al. 1998): using ontologies to model and annotate the semantics of information resources in a machine-processible manner. The developers of On-To-Knowledge are carrying out three industrial case studies-with SwissLife (http://www.swisslife.ch), British Telecom (http://www.bt.com/innovations), and Enersearch[5] - to evaluate the tool environment for ontology-based knowledge management. In this context, CognIT (http://www.cognit.no) extended its information extraction tool

Corporum to generate ontologies from semistructured or unstructured natural-language documents. Important concepts and their relationships are extracted from these documents and used to build up initial ontologies. Figure 1.6 shows an automatically generated semantic structure map of the EnerSearch Web site using Aidministrator technology (http://www.aidministrator.nl).

An application of the Semantic Web technology in the B2C area has been developed by Semantic Edge (http://www.semanticedge.com) that offers front-end voice-based and natural-language access to distributed and heterogeneous product information. The technology will enable the human user, instead of manually browsing large volumes of product information, to ask simple questions like "Where can I get a cheap color printer for my Mac?"

Finally, the B2B area may become the most important application area of Semantic Web technology in terms of the market volume. Companies like VerticalNet (http://www.verticalnet.com) which builds many vertical marketplaces, or ContentEurope (http://www.contenteurope.com) which provides content management solutions for B2B electronic commerce, all face the same problem: integrating heterogeneous and distributed product information. Naturally such companies make use of ontology-based integration techniques to reduce the level of effort required to provide integrated solutions for B2B marketplaces.

Notes

1. Given the current situation, there will be many "standards" requiring interchange.

2. Note that we are speaking here about the Semantic Web.

3. Fixed prices turned up at the beginning of the 20th century, lowering transaction costs. However, negotiations and auctions (like those available on some Web sites) help allocate resources more optimally. Still, the effort required for negotiation may outweigh the advantages in resource allocation and lead to unreasonably high demands on time (and transaction costs). Automated negotiation agents and auction houses reduce these high transaction costs and allow optimized resource allocation.

4. On-To-Knowledge is a European IST project (http://www.ontoknowledge.org).

5. See further http://www.enersearch.se. Enersearch research affiliates and shareholders are spread over many countries: its shareholding companies include IBM (United States), Sydkraft (Sweden), ABB (Sweden/Switzerland), PreussenElektra (Germany), Iberdrola (Spain), ECN (Netherlands), and Electricidade do Portugal.

References

Bowers, S., and L. Delcambre. 2000. Representing and Transforming Model-Based Information. In Electronic Proceedings of the ECDL 2000 Workshop on the Semantic Web at the Fourth European Conference on Research and Advanced Technology for Digital Libraries (ECDL-2000), Lisbon, Portugal, September 21, 2000. Available from http://www.ics.forth.gr/proj/isst/SemWeb/program.html.

Berners-Lee, T. 1999. *Weaving the Web*. London: Orion Business.

Brickley, D., and R. Guha. 2000. Resource Description Framework (RDF) Schema Specification 1.0 (candidate recommendation). World Wide Web Consortium. Available from http://www.w3.org/TR/2000/CR-rdf-schema-20000327.

Erdmann, M., and R. Studer. 2001. How to Structure and Access XML Documents with Ontologies. *Data and Knowledge Engineering* 36:317-335.

Farquhar, A., R. Fikes, and J. Rice. 1997. The Ontolingua Server: A Tool for Collaborative Ontology Construction. *International Journal of Human-Computer Studies* 46:707-728.

Fensel, D. 2001. *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Berlin: Springer-Verlag.

Fensel, D., J. Angele, S. Decker, M. Erdmann, H.-P. Schnurr, R. Studer, and, A. Witt. 2000. Lessons Learned from Applying AI to the Web. *Journal of Cooperative Information Systems* 9(4):361-382.

Fensel, D., S. Decker, M. Erdmann, and R. Studer. 1998. Ontobroker: The Very High Idea. In *Proceedings of the 11th International Flairs Conference (FLAIRS-98), Sanibel Island, Florida, USA, May*, ed. D. J. Cook (pp. 131-135). Menlo Park, CA: AAAI.

Fensel, D., I. Horrocks, F. Van Harmelen, D. McGuiness, and P. Patel-Schneider. 2001. OIL: Ontology Infrastructure to Enable the Semantic Web. *IEEE Intelligent Systems* (March/ April):38-45.

Fensel, D., F. van Harmelen, M. Klein, H. Akkermans, J. Broekstra, C. Fluit, J. Van der Meer, H.-P. Schnurr, R. Studer, J. Hughes, U. Krohn, J. Davies, R. Engels, B. Bremdal, F. Ygge, U. Reimer, and I. Horrocks. 2000. On-To-Knowledge: Ontology-based Tools for Knowledge Management. In *Proceedings of the eBusiness and eWork (EMMSEC-2000) Conference, Madrid, Spain, October*. Available from http://www.ebew.net/.

Genesereth, M. R. 1991. Knowledge Interchange Format. In *Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning (KR-91)*, ed. J. Allen et al. San Francisco: Morgan Kaufman.

Grosso, W. E., H. Eriksson, R. W. Fergerson, J. H. Gennari, S. W. Tu, and M. A. Musen. Knowledge Modeling at the Millennium (The Design and Evolution of Protege-2000). In *Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management (KAW-1999)*, Banff, Alberta, Canada, October 16-21, 1999. Available from http://smi-web.stanford.edu/pubs/SMI_Abstracts/SMI-1999-0801.html.

Gruber, T. R. 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 5:199-220.

Horrocks, I, and P. F. Patel-Schneider. 1999. Optimizing Description Logic Subsumption. *Journal of Logic and Computation* 9(3):267-293.

Klein, M., D. Fensel, F. van Harmelen, and I. Horrocks. 2000. The Relation between Ontologies and Schema-Languages: Translating OIL-Specifications to XML-Schema In *Proceedings of the Workshop on Applications of Ontologies and Problem-Solving Methods, 14th European Conference on Artificial Intelligence ECAI-2000, Berlin, Germany, August 20-25, 2000*, ed. V. R. Benjamins et al. Available from http://www.cs.vu.nl/~mcaklein/papers/.

Lassila, O. 1998. Web Metadata: A Matter of Semantics. *IEEE Internet Computing*, 2(4):30-37.

Lenat, D. B., and R. V. Guha. 1990. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Reading, MA.: Addison-Wesley.

Luke, S., L. Spector, and D. Rager. 1996. Ontology-Based Knowledge Discovery on the World Wide Web. In *Working Notes of the Workshop on Internet-Based Information Systems at the 13th National Conference on Artificial Intelligence (AAAI96)*. Available from http://www.csl.sony.co.jp/person/amf/iis96.html.

Mädche, A., and S. Staab. 2000. Mining Ontologies from Text. In *Knowledge Acquisition, Modeling, and Management: Proceedings of the European Knowledge Acquisition Conference (EKAW-2000)*, ed. R. Dieng et al. Lecture Notes in Artificial Intelligence (LNAI). Berlin: Springer-Verlag.

McGuinness, D. L., R. Fikes, J. Rice, and S. Wilder. 2000. An Environment for Merging and Testing Large Ontologies. In *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR-2000), Breckenridge, Colorado, April 12-15* (pp. 483-493). San Francisco: Morgan Kaufmann.