Chapter 1

THE SEMANTIC WEB AND ITS APPLICATIONS

Jorge Cardoso[1] and Amit Sheth[2]
[1]Department of Mathematics and Engineering,University of Madeira, 9000-390, Funchal, Portugal – jcardoso@uma.pt

[2]Large Scale Distributed Information Systems (LSDIS) Lab, Department of Computer Science, University of Georgia, GA, USA. – amit@cs.uga.edu

1. INTRODUCTION

Currently, the World Wide Web is primarily composed of documents written in HTML (Hyper Text Markup Language), a language that is useful for publishing information. HTML is a set of "markup" symbols contained in a Web page intended for display on a Web browser. During the first decade of its existence, most of the information on the Web is designed only for human consumption. Humans can read Web pages and understand them, but their inherent meaning is not shown in a way that allows their interpretation by computers

The information on the Web can be defined in a way that it can be used by computers not only for display purposes, but also for interoperability and integration between systems and applications. One way to enable machine-to-machine exchange and automated processing is to provide the information in such a way that computers can understand it. This is precisely the objective of the semantic Web – to make possible the processing of Web information by computers. "The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

(Berners-Lee, Hendler et al. 2001). The next generation of the Web will combine existing Web technologies with knowledge representation formalisms (Grau 2004)

The Semantic Web was made through incremental changes, by bringing machine-readable descriptions to the data and documents already on the Web. As already stated, the Web was originally a vast set of static Web pages linked together. Currently the Web is in evolution, as illustrated in Figure 1-1, and different approaches are being sought to come up with the solutions to add semantics to Web resources. On the left side of Figure 1-1, a graph representation of the syntactic Web is given. Resources are linked together forming the Web. There is no distinction between resources or the links that connect resources. To give meaning to resources and links, new standards and languages are being investigated and developed. The rules and descriptive information made available by these languages allow to characterize individually and precisely the type of resources in the Web and the relationships between resources, as illustrated in the right side of Figure 1-1.

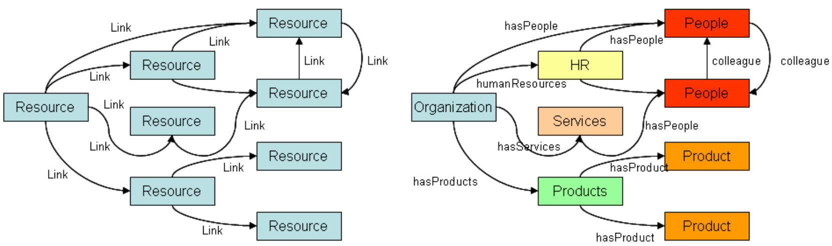

Figure 1-1. Evolution of the Web

Due to the widespread importance of integration and interoperability for intra- and inter-business processes, the research community has tackled this problem and developed semantic standards such as the Resource Description Framework (RDF) (RDF 2002) and the Web Ontology Language (OWL) (OWL 2004). RDF and OWL standards enable the Web to be a global infrastructure for sharing both documents and data, which make searching and reusing information easier and more reliable as well. RDF is the W3C standard for creating descriptions of information, describing their semantics and reasoning (Lassila and Swick 1999), especially information available on the World Wide Web. What XML is for syntax, RDF is for semantics. Both share a unified model and together provide a framework for developing Web applications that deal with data and semantics (Patel-Schneider and Siméon 2002). Relationships are at the heart of semantics (Sheth, Arpinar et al. 2002). Perhaps the most important characteristic of RDF is that it elevates relationships to first class object, providing the first representational basis for giving semantic description. RDF evolved from MCF designed by Guha, which was motivated for representing metadata. Hence RDF is also well suited for representing metadata for Web resources. OWL provides a language for defining structured Web-based ontologies which allows a richer integration and interoperability of data among communities and domains.

According to TopQuadrant (TopQuadrant 2005), a consulting firm that specializes in Semantic Web technologies, the market for semantic technologies will grow at an annual growth rate of between 60% and 70% until 2010. It will grow from its current size of US$2 billion to US$63 billion. According to William Ruh of CISCO, before the end of 2004, RDF was applied under the covers of well over 100 identified products and over 25 information service providers. Existing well known applications that add Semantic Web capabilities include Adobe's Extensible Metadata Platform, RDF of annotation of most of the product data that Amazon receives or digital media content a top mobile carrier receives, and well known infrastructure support include Creative Commons DF based annotations of license information and Oracle's support for RDF data.

Semantic software is being experimentally used by banks to help them to comply with the U.S. government's Patriot Act (the Patriot Act requires banks to track and account for the customers with whom they do transactions), by European police force to follow crime patterns, and by telephone service providers to create applications that provides information about pay-per-view movies (Lee 2005; Sheth 2005). In addition to investment banks, the Metropolitan Life Insurance Company, the U.S. Department of Defense and the Tennessee Valley Authority have also used Semantic software to integrate enterprise data to comply with federal regulations.

2. SEMIOTICS – SYNTAX, SEMANTICS, AND PRAGMATICS

Semiotics is the general science of signs – such as icons, images, objects, tokens, and symbols – and how their meaning is transmitted and understood. A sign is generally defined as something that stands for something else.

The human language is a particular case of semiotics. A language is a system of conventional spoken or written symbols by means of which people communicate. Formal languages, such as logic, are also based on symbols and, therefore, are also studied by semiotics. Compared to the human language, formal languages have a precise construction rules for the syntax and semantics of programs. Semiotics is composed of three fundamental components: syntax, semantics, and pragmatics (Peirce 1960). These components are illustrated in Figure 1-2.

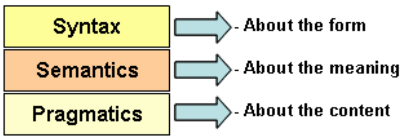

Figure 1-2. Semiotics and its components

Syntax. It deals with the formal or structural relations between signs (or tokens) and the production of new ones. For example, grammatical syntax is the study of which sequences of symbols are well formed according to the recursive rules of grammar. The set of allowed reserved words, their parameters, and the correct word order in an expression is called the syntax of a language. In computer science, if a program is syntactically correct according to its rules of syntax, then the compiler will validate the syntax and will not generate error messages. This, however, does not ensure that the program is semantically correct (i.e., return results as expected).

For example, when XML is used to achieve interoperability and integration of information systems, the data exchanged between systems must follow a precise syntax. If the rules of the syntax are not followed, a syntactical error occurs. For example, using a tag spelled <cust> instead of <customer>, omitting a closing tag, or not following the syntax of a XML Schema (XMLSchema 2004) will generate a syntactical error. It should be noticed, that syntax does not include the study of things such as “truth” and “meaning.”

Semantics. It is the study of relations between the system of signs (such as words, phrases, and sentences) and their meanings. As it can be seen by this definition, the objective of semantics is totally different from the objective of syntax. The former concerns to what something means while the latter pertains to the formal structure/patterns in which something is expressed. Semantics are distinct from the concept of ontology (ontologies will be discusses later in this chapter). While the former is about the use of a word, the latter is related to the nature of the entity or domain referenced by the word. One important and interesting question in semantics research is if the meaning is established by looking at the neighborhood in the ontology that the word is part of or if the meaning is already contained in the word itself. Second important and interesting question is the formal representation language to capture the semantics such that it is machine processable with consistent interpretation. Third important question is the expressiveness of this representation language that balances computability versus capturing the true richness of the real world that is being modeled. Correspondingly, the following three forms of semantics have been defined in (Sheth, Ramakrishnan et al. 2005):

- **Implicit** semantics. "This type of semantics refers to the kind that is implicit in data and that is not represented explicitly in any machine processable syntax."
- **Formal** semantics. "Semantics that are represented in some well-formed syntactic form (governed by syntax rules) is referred to as formal semantics."
- **Powerful** (soft) semantics. "Usually, efforts related to formal semantics have involved limiting expressiveness to allow for acceptable computational characteristics. Since most KR mechanisms and the Relational Data Model are based on set theory, the ability to represent and utilize knowledge that is imprecise, uncertain, partially true, and approximate is lacking, at least in the base/standard models. Representing and utilizing these types of more powerful knowledge is, in our opinion, critical to the success of the Semantic Web. Soft computing has explored these types of powerful semantics. We deem these powerful (soft)

semantics as distinguished, albeit not distinct from or orthogonal to formal and implicit semantics."

Pragmatics. It is the study of natural language understanding, and specifically the study of how context influences the interpretation of meaning. Pragmatics is interested predominantly in utterances, made up of sentences, and usually in the context of conversations (Wikipedia 2005). The context may include any social, environmental, and psychological factors. It includes the study or relations among signs, their meanings, and users of the signs, and the repercussions of sign interpretations for the interpreters in the environment. While semantics deals with the meaning of signs, pragmatics deals with the origin, uses, and effects of signs within the content, context, or behavior in which they occur.

3. SEMANTIC HETEROGENEITY ON THE WEB

Problems that might arise due to heterogeneity of the data in the Web are already well known within the distributed database systems community (e. g. (Kim and Seo 1991), (Kashyap and Sheth 1996)). Heterogeneity occurs when there is a disagreement about the meaning, interpretation, or intended use of the same or related data. As with distributed database systems, four types of information heterogeneity (Sheth 1998; Ouskel and Sheth 1999) may arise in the Web: system heterogeneity, syntactic heterogeneity, structural or schematic heterogeneity, and semantic heterogeneity.

- **System heterogeneity**: Applications and data may reside in different hardware platforms and operating systems.
- **Syntactic heterogeneity**: Information sources may use different representations and encodings for data. Syntactic interoperability can be achieved when compatible forms of encoding and access protocols are used to allow information systems to communicate.
- **Structural heterogeneity**: Different information systems store their data in different document layouts and formats, data models, data structures and schemas.
- **Semantic heterogeneity**: The meaning of the data can be expressed in different ways leading to heterogeneity. Semantic

heterogeneity considers the content of an information item and its intended meaning.

Approaches to the problems of semantic heterogeneity should equip heterogeneous, autonomous, and distributed software systems with the ability to share and exchange information in a semantically consistent way (Sheth 1999). In the representation languages to support the Semantic Web approach, as recommended by the W3C, XML supports ability to deal with syntactic heterogeneity; XML, XPath, and XQuery provide ability to transcend certain structural heterogeneity, while RDF and OWL (or other ontology representation languages) provide a key approach to deal with semantic heterogeneity.

One solution is for developers to write code which translates between the terminologies of pairs of systems. When the requirement is for a small number of systems to interoperate, this may be a useful solution. However, this solution does not scale as the development costs increase as more systems are added and the degree of semantic heterogeneity increases. Assuming the development of bidirectional translators, i.e. translators that enable the interoperation of system A to system B and from system B to system A, to allow the interoperability of 'n' systems we need (n-1)+(n-2)+…+1 translators. Figure 1-3 shows the translators required to integrate 6 systems.

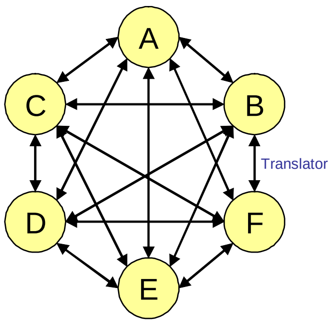

Figure 1-3. Using translators to resolve semantic heterogeneity

A more suitable solution to the problem of semantic heterogeneity is to rely on the technological foundations of the semantic Web. More precisely, to semantically define the meaning of the terminology of each distributed system data using the concepts present in a shared ontology to make clear the relationships and differences between concepts.

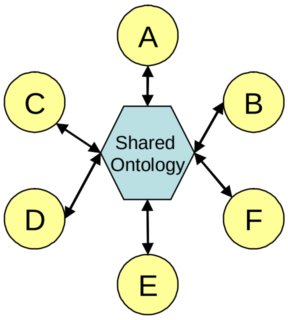

Figure 1-4. Using a shared ontology to resolve semantic heterogeneity

Figure 1-4 shows a possible architecture that achieves interoperability using the semantic Web and ontologies. This solution only requires the development of 'n' links to interconnect systems.

4. METADATA

Metadata can be defined as "data about data." The goal of incorporating metadata into data sources is to enable the end-user to find items and contextually relevant information. Data sources are generally heterogeneous and can be unstructured, semi-structured, and structured. In the semantic Web, a data source is typically a document, such as a Web page, containing textual content or data. Of course, other types of resources may also include metadata information, such as records from a digital library.

Metadata can exist in several levels. These "levels of metadata" are not mutually exclusive; on the contrary, the accumulative combination of each type of metadata provides a multi-faceted representation of the data including information about its syntax, structure, and semantic context (Fisher and Sheth 2004).

The process of attaching semantic metadata to a document or any piece of content is called semantic. Metadata extraction is the process of identifying metadata for that document or content. This process could be manual, semiautomatic (e.g., (Handschuh, Staab et al. 2002)) or fully automatically (e.g., Semantic Enhancement Engine (Hammond, Sheth et al. 2002) or SemTag (Dill, Eiron et al. 2003)). Semantic applications are created by exploiting metadata and ontologies with associated knowledgebase (Sheth 2004). In essence, in the semantic Web, documents are marked up with semantic metadata which is machine-understandable about the human-readable content of documents. Other approaches, which are less expressive, consist on using purely syntactic or structural metadata.

4.1 Syntactic Metadata

The simplest form of metadata is syntactic metadata. It describes non-contextual information about content and provides very general information, such as the document's size, location, or date of creation. Syntactic metadata attaches labels or tags to data. The following example shows syntactic metadata describing a document:

```
<name> = "report.pdf"
<creation> = "30-09-2005"
<modified> = "15-10-2005"
<size> = 2048
```

Most documents have some degree of syntactic metadata. E-mail headers provide author, recipient, date, and subject information. While these headers provide very little or no contextual understanding of what the document says or implies (assuming value of author is treated as a string or ordered sets of words, rather than its full semantics involving modeling of author as a person authoring a document, etc.), this information is useful for certain applications. For example, a mail client may constantly monitor incoming e-mail to find documents, related to a particular subject, the user is interested in.

4.2 Structural Metadata

Structural metadata provides information regarding the structure of content. It describes how items are put together or arranged. The amount and type of such metadata will vary widely with the type of document. For example, an HTML document may have a set of predefined tags, but these exist primarily for rendering purposes. Therefore, they are not very helpful in providing contextual information for content. Nevertheless, positional or structural placement of information within a document can be used to further embellish metadata (e.g., terms or concepts appear in a title may be give higher weight to that appearing in the body). On the other hand, XML gives the ability to enclose content within more meaningful tags. This is clearly more useful in determining context and relevance when compared to the limitations of syntactic metadata for providing information about the document itself.

For example, a DTD or XSD outlines the structural metadata of a particular document. It lists the elements, attributes, and entities in a document and it defines the relationships between the different elements and attributes. A DTD declares a set of XML element names and how they can be used in a document. The following lines, extracted from a DTD, describe a set of valid XML documents:

```
<!ELEMENT contacts (contact*)>
<!ELEMENT contact (name, birthdate)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT birthdate (#PCDATA)>
```

Structural metadata tell us how data are grouped and put in ordered arrangements with other data. This DTD sample indicates that a “contacts” element contains one or more “contact” elements. A “contact” element contains the elements “name” and “birthdate”, and the “name” and “birthdate” elements contain data.

4.3 Semantic Metadata

Semantic metadata adds relationships, rules, and constraints to syntactic and structural metadata. This metadata describe contextually relevant or domain-specific information about content based on a domain specific metadata model or ontology, providing a context for interpretation. In a sense, they capture a meaning associated with the content. If a formal ontology is used for describing and interpreting this type of metadata, then it lends itself to machine processability and hence higher degrees of automation.

Semantic data provides a means for high-precision searching, and, perhaps most importantly, it enables interoperability among heterogeneous data sources. Semantic metadata is used to give meaning to the elements described by the syntactic and structural metadata. These metadata elements allow applications to “understand” the actual meaning of the data.

By creating a metadata model of data, information, and relationships, we are able to use reasoning capabilities such as inference engines to draw logical conclusions based on the metadata model, or path identification and ranking using graph based processing leading to mining and discovery. For instance, if we know that the ABC Company sends every year a gift to very good customers, and that John is a very good customer, then by inference, we know that the company will ship a gift to John next year. Or if we find a potential customer has a business partner with another person who is on the Bank of England list of people involved in money laundering, the potential customer is a suspect according to the government's anti-money regulations. Figure 1-5 (Sheth 2003) shows the types of metadata we have discussed.

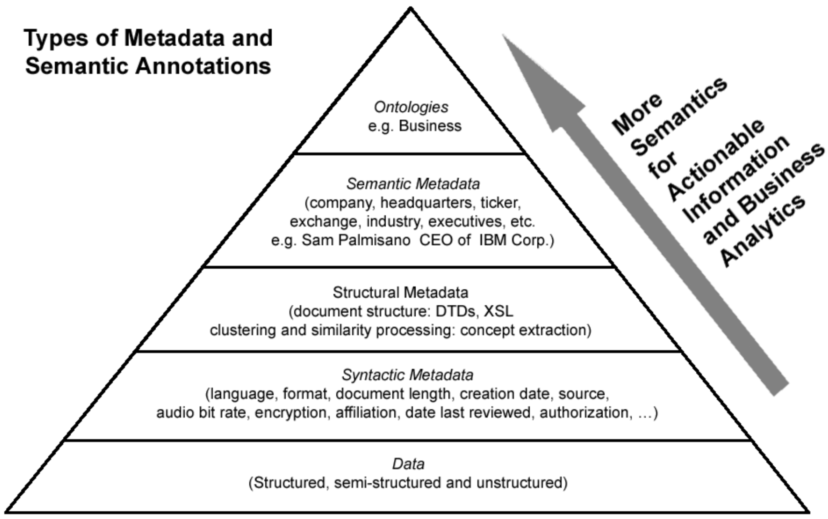

Figure 1-5. Types of metadata

4.4 Creating and Extracting Semantic Metadata

In order to extract optimal value from a document and make it usable, it needs to be effectively tagged by analyzing and extracting relevant information of semantic interest. Many techniques can be used to achieve this based on extracting syntactic and semantic metadata from documents (Sheth 2003). These include:

Semantic lexicons, nomenclatures, reference sets and thesauri: Match words, phrases or parts of speech with a static or periodically maintained dictionary and thesaurus. Semantic lexicon, such as WordNet (Voorhees 1998) which groups English words into sets of synonyms called synsets and records semantic relations between synonym sets, can be used to identify and match terms in different directions, finding words that mean the same or are more general or more specific. WordNet supports various types of relationships such as synonyms, hypernyms, hyponyms, holonym, and meronym which can de effectively used to find relationship between words and extract the meaning of words.

Document analysis: Look for patterns and co-occurrences, and apply predefined rules to find interesting patterns within and across documents. Regular expressions and relationships between words can be used to understand the meaning of documents.

Ontologies: Capturing domain-specific (application or industry) knowledge including entities and relationships, both at a definitional level (e.g., a company has a CEO), and capturing real-world facts or knowledge (e.g., Meg Witman is the CEO of eBay) at an instance or assertional level. If the ontology deployed is "one size fits all" and is not domain-specific, the full potential of this approach cannot be exploited.

The last option, also known as ontology-driven meta data extraction, is the most flexible (assuming the ontology is kept up to date to reflect changes in the real world) and comprehensive (since it allows modeling of fact-based domain-specific relationships between entities that are at the heart of semantic representations).

5. EMPIRICAL CONSIDERATIONS ON THE USE OF SEMANTICS AND ONTOLOGIES

Semantics is arguably the single most important ingredient in propelling the Web to its next phase to provide standards to seamlessly enable interoperability of applications. Semantics is considered to be the best framework to deal with the heterogeneity, massive scale, and dynamic nature of the resources on the Web. Issues pertaining to semantics have been addressed in other fields like linguistics, knowledge representation, and AI. Based on the research on semantics, semantic Web, and real-world applications deployment, we present a set of empirical observations, considerations, and requirements for the construction of future applications, extended from the original set presented in (Sheth and Ramakrishnan 2003):

- It is the "ontological commitment" reflecting agreement among the experts defining the ontology and its uses that is a key basis for semantic integration. A good case in point is the Gene Ontology (GO) which despite its use of a representation with limited expressiveness has been extremely popular among the genomic scientists.
- Ontologies can capture human activities (e.g., modeling domains of travel or financial services) or natural phenomena and science (e.g., protein-protein interactions or glycan structures). Schemas modeling some domain, especially those modeling natural phenomena and science could be quite large and complex. For example, the Gycomics Ontology GlycO (http://lsdis.cs.uga.edu/projects/glycomics/) has over 600 classes, pushes the expressiveness of the OWL language in modeling the constraints, and is eleven levels deep.
- Ontology population which captures real world facts and trusted knowledge of a domain is critical. In the near future, it will not be uncommon to find ontology with millions of facts. Since it is obvious that this is the sort of scale Semantic Web applications are going to be dealing with, means of populating ontologies with instance data need to be automated.
- Semi-formal ontologies, possibly based on limited expressive power focusing on relationships but not constraints, can be very practical and useful. Ontologies represented in more expressive languages such as OWL (compared to RDF/S) have in practice yielded little value in industrial applications so far. One reasons for this could be that it is difficult to capture the knowledge that uses the more expressive constructs of a representation language. At the same time, when modeling more complex domains have required use of more expressive languages and more intensive effort in schema design as well as population.
- Large scale metadata extraction and semantic annotation is possible, as exemplified by Semantic Enhancement Engine of Semagix Freedom (Hammond, Sheth et al. 2002) and SemTag/SemSeeker of IBM WebFountain (Dill, Eiron et al. 2003).

Storage and manipulation of metadata for millions to hundreds of millions of content items requires best applications of known database techniques with challenge of improving upon them for performance and scale in presence of more complex structures.

- Support for heterogeneous data is key – it is too hard to deploy separate products within a single enterprise to deal with structured and unstructured data. New applications involve extensive types of heterogeneity in format, media and access/delivery mechanisms. Database researchers have long studied the issue of integrating heterogeneous data, and many of these come handy.
- A vast majority of the Semantic (Web) applications that have been developed rely on three crucial capabilities: ontology creation, semantic annotation, and querying/reasoning. A good percentage of reasoning used in real world applications is related to path finding and rule processing, rather than academically popular inferencing. All these capabilities must scale to millions of documents and concepts.

6. APPLICATIONS OF SEMANTICS AND ONTOLOGIES

The intention of this section is to present the state of the art of the applications that use semantics and ontologies. We describe various applications ranging from the use of semantic Web services, semantic integration of tourism information sources, and semantic digital libraries to the development of bioinformatics ontologies.

6.1 Semantic Web services

Web services are modular, self-describing, self-contained applications that are accessible over the Internet (Curbera, Nagy et al. 2001). Currently, Web services are described using the Web Services Description Language (Chinnici, Gudgin et al. 2003), which provide operational information. Although the Web Services Description Language (WSDL) does not contain semantic descriptions, it specifies the structure of message components using XML Schema constructs. Semantic Web services are the result of the evolution of the syntactic definition of Web services and the semantic Web as shown in Figure 1-6.

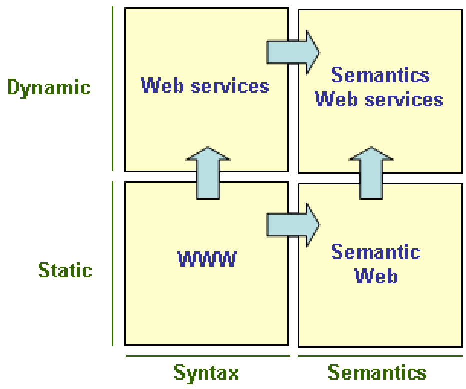

Figure 1-6. The nature of semantic Web services

One solution to create semantic Web services is by mapping concepts in a Web service description to ontological concepts. Using this approach, users can explicitly define the semantics of a Web service for a given domain. Significantly different approaches to specifying semantic Web services are exemplified by four submissions to the World Wide Web consortium (W3C): OWL-S (OWL-S 2004), WSMO (WSMO 2004), FLOWS (SWSF 2005) and WSDL-S (Akkiraju, Farrell et al. 2005). WSDL-S is the most standard compliant and incremental approach that extends WSDL2.0, W3C's recommendation for Web service specification. Figure 1-7 illustrates METEOR-S WSDL-S Annotator tool (Patil, Oundhakar et al. 2004) and the mapping that have been established between WSDL descriptions and ontological concepts.

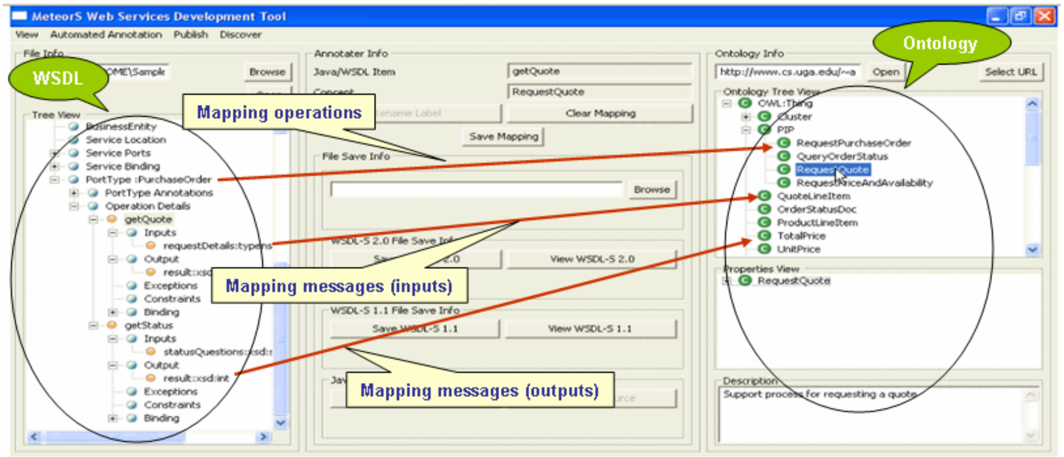

Figure 1-7. Annotating Web services with ontological concepts

Based on the analysis of WSDL descriptions, three types of elements can have their semantics increased by annotated them with ontological concepts: operations, messages, and preconditions and effects. All the elements are explicitly declared in a WSDL description.

Operations. Each WSDL description may have a number of operations with different functionalities. For example, a WSDL description can have operations for both booking and canceling flight tickets. In order to add semantics, the operations must be mapped to ontological concepts to describe their functionality.

Message. Message parts, which are input and output parameters of operations, are defined in WSDL using the XML Schema. Ontologies – which are more expressive than the XML Schema – can be used to annotate WSDL message parts. Using ontologies, not only brings user requirements and service advertisements to a common conceptual space, but also helps to use and apply reasoning mechanisms.

Preconditions and effects. Each WSDL operation may have a number of preconditions and effects. The preconditions are usually logical conditions, which must be evaluated to true in order to execute a specific operation. Effects are changes in the world that occur after the execution of an operation. After annotating services' operations, inputs, and outputs, preconditions and effects can also be annotated.

The semantic annotation of preconditions and effects is important for Web services since it is possible for a number of operations to have the same functionality, as well as, the same inputs and outputs, but different effects.

6.2 Semantic Web service discovery

Given the dynamic nature of e-business environment, the ability to find best matching Web services that can also be easily integrated to create business processes is highly desirable. Discovery is the procedure of finding a set of appropriate Web services, select a specific service that meets user requirements, and bind it to a Web processes (Verma, Sivashanmugam et al. 2004). The search of Web services to model Web process applications differs from the search of tasks to model traditional processes, such as workflows. One of the main differences is in terms of the number of Web services available to the composition process. In the Web, potentially thousands of Web services are available. Therefore, one of the problems that need to be solved is how to efficiently discover Web services (Cardoso and Sheth 2003).

Currently, the industry standards available to register and discover Web services are based on the Universal Description Discovery and Integration specification (UDDI 2002). Unfortunately, discovering Web services using UDDI is relatively inefficient since the specification does not take into account the semantics of Web services, even though it provides an interface for keyword and taxonomy based searching as shown in Figure 1-8.

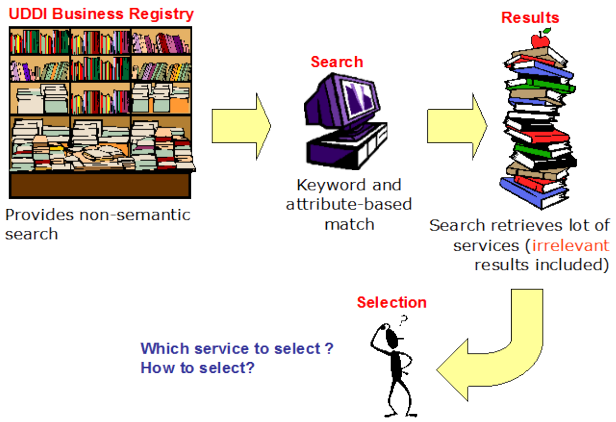

Figure 1-8. State of the art in discovery (Cardoso, Bussler et al. 2005)

The key to the discovery of Web services is having semantics in the description of services itself (Sheth and Meersman 2002) and then use semantic matching algorithms (e.g. (Smeaton and Quigley 1996; Klein and Bernstein 2001; Rodríguez and Egenhofer 2002; Cardoso and Sheth 2003), to find Web services. An approach for semantic Web service discovery is the ability to construct queries using concepts defined in a specific ontological domain. By having both the description and query explicitly declare their semantics, the results of discovery will be more relevant than keyword or attribute-based matching.

The semantic discovery of Web services has specific requirements and challenges compared to previous work on information retrieval systems and information integration systems. Several issues that need to be considered include:

- Precision of the discovery process. The search has to be based, not only on syntactic information, but also on data, functional, and non-functional/QoS semantics.
- Enable the automatic determination of the degree of integration of the discovered Web services and the Web process host.
- The integration and interoperation of Web services differs from previous work on schema integration due to the polarity of the schema that must be integrated (Cardoso and Sheth 2003).

Adding semantic annotations to WSDL specifications and UDDI registries allows improving the discovery of Web services. The general algorithm for semantic Web service discovery requires the users to enter Web service requirements as templates constructed using ontological concepts. There phases of the algorithm can be identified. In the first phase, the algorithm matches Web services based on the functionality (the functionality is specified using ontological concepts that map to WSDL operations) they provide. In the second phase, the result set from the first phase is ranked on the basis of semantic similarity (Cardoso and Sheth 2003) between the input and output concepts of the selected operations and the input and output concepts of the initial template, respectively. The optional third phase involves ranking the services based on the semantic similarity between the precondition and effect concepts of the selected operations and preconditions and effect concepts of the template.

6.3 Semantic Integration of Tourism Information Sources

Dynamic packaging technology helps online travel customers to build and book vacations. It can be described as the ability for a customer to put together elements of a (vacation) trip including flights, hotels, car rentals, local tours and tickets to theatre and sporting events. In the offline world, such packages used to be put together by tour operators in brochures. The new technology includes the ability to combine multiple travel components on demand to create a reservation. The package that is created is handled seamlessly as one transaction and requires only one payment from the consumer, hiding the pricing of individual components.

Current dynamic packaging applications are developed using a hard-coded approach to develop the interfaces among various systems to allow the interoperability of decentralized, autonomous, and heterogeneous tourism information systems. However, such an approach for integration does not comply with the highly dynamic and decentralized nature of the tourism industry. Most of the players are small or medium-sized enterprises with information systems with different scopes, technologies, architectures, and structures. This diversity makes the interoperability of information systems and technologies very complex and constitutes a major barrier for emerging e-marketplaces and dynamic packaging applications that particularly affects the smaller players (Fodor and Werthner 2004-5).

Two emerging technologies can enable the deployment of a more integrated solution to implement dynamic application (Cardoso 2005):

Web services and semantics. As opposed to the hard-coded approach, Web services take a loosely coupled software components approach, which can be dynamically located and integrated on the Web. Web services are flexible to easily design processes that model dynamic packaging applications. Semantics are important to dynamic packaging applications because they provide a shared and common understanding of data and services of the tourism information systems to integrate. Semantics can be used to organize and share tourism information, which allow better interoperability and integration of inter- and intra-company travel information systems.

Figure 1-9 illustrates the integration of various tourism information systems to support the concept of dynamic packaging. As it can be seen, new communication links are established among the various participant of the distribution model to integrate tourism products.

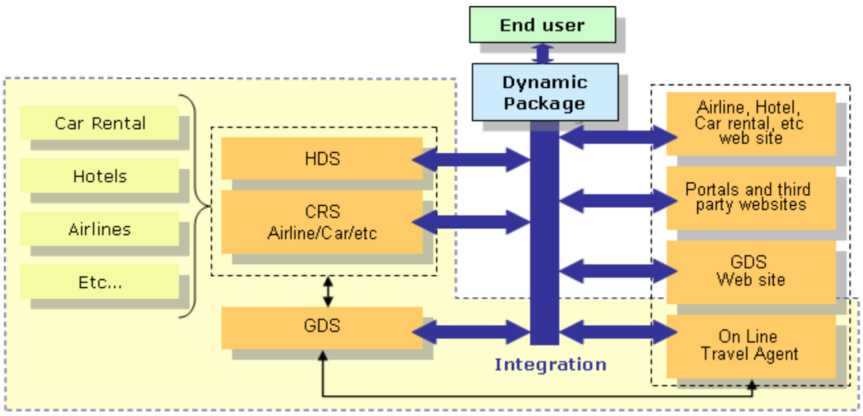

Figure 1-9. Integration of tourism information systems

So far, the travel industry has concentrated its efforts on developing open specification messages, based on XML, to ensure that messages can flow between industry segments as easily as within. For example, the OpenTravel Alliance (OTA 2004) is an organization pioneering the development and use of specifications that support e-business among all segments of the travel industry. It has produced more than 140 XML-based specifications for the travel industry (Cardoso 2004).

The development of open specifications messages based on XML, such as OTA schema, to ensure the interoperability between trading partners and working groups is not sufficiently expressive to guarantee an automatic exchange and processing of information to develop dynamic applications. A more appropriate solution is the development of suitable ontologies for the tourism industry that can serve as a common language for tourism-related terminology and a mechanism for promoting the seamless exchange of information across all travel industry segments. Ontologies are the key elements enabling the shift from a purely syntactic to a semantic interoperability. An ontology can be defined as the explicit, formal descriptions of concepts and their relationships that exist in a certain universe of discourse, together with a shared vocabulary to refer to these concepts. With respect to an ontology a particular user group commits to, the semantics of data provided by the data sources to integrate can be made explicit. Ontologies can be applied to the area of dynamic packaging to explicitly connect data and information from tourism information systems to its definition and context in machine-processable form. Ontologies can be used to bring together heterogeneous Web services, Web processes, applications, data, and components residing in distributed environments. Semantic Web processes, managing dynamic package determine which Web services are used, what combinations of Web services are allowed or required and specific rules determine how the final retail price is computed (Cardoso, Miller et al. 2004).

6.4 Semantic digital libraries

Libraries are a key component of the information infrastructure indispensable for education. They provide an essential resource for students and researchers for reference and for research. Metadata has been used in libraries for centuries. For example, the two most common general classification systems, which use metadata, are the Dewey Decimal Classification (DDC) system and the Library of Congress Classification (LCC) system. In the United States, the DDC is used in 95% of all public and K-12 school libraries, in 25% of college and university libraries, and in 20% of special libraries. The DDC system has 10 major subjects, each with 10 secondary subjects (DDC 2005). The LCC system uses letters instead of numbers to organize materials into 21 general branches of knowledge. The 21 subject categories are further divided into more specific subject areas by adding one or two additional letters and numbers (LCCS 2005).

As traditional libraries are increasingly converting themselves to digital libraries, a new set of requirements has emerged. One important feature for digital libraries is the availability to efficiently browse electronic catalogues browsed. This requires the use of common metadata to describe the records of the catalogue (such as author, title, and publisher) and common controlled vocabularies to allow subject identifiers to be assigned to publications. The use of a common controlled vocabulary, thesauri, and taxonomy (Smrz, Sinopalnikova et al. 2003) allows search engines to ensure that the most relevant items of information are returned. Semantically annotating the contents of a digital library's database goes beyond the use of a controlled vocabulary, thesauri, or taxonomy. It allows retrieving books' records using meaningful information to the existing full text and bibliographic descriptions.

Semantic Web technologies, such as RDF and OWL, can be used as a common interchange format for catalogue metadata and shared vocabulary, which can be used by all libraries and search engines (Shum, Motta et al. 2000) across the Web. This is important since it is not uncommon to find library systems based on various metadata formats and built by different persons for their special purposes. By publishing ontologies, which can then be accessed by all users across the Web, library catalogues can use the same vocabularies for cataloguing, marking up items with the most relevant terms for the domain of interest. RDF and OWL provide a single and consistent encoding so implementers of digital library metadata systems will have their task simplified when interoperating with other digital library systems.

6.5 Semantic Grid

The concept of Grid (Foster and Kesselman 1999) has been proposed as a fundamental computing infrastructure to support the vision of e-Science. The Grid is a service for sharing computer power and data storage capacity over the Internet and goes well beyond simple communication providing functionalities that enable the rapid assembly and disassembly of services into temporary groups.

Recently, the Grid has been evolving towards the Semantic Grid to yield an intelligent platform which allows process automation, knowledge sharing and reuse, and collaboration within a community (Roure, Jennings et al. 2001). The Semantic Grid is about the use of semantic Web technologies in Grid computing; it is an extension of the current Grid. The objective is to describe information, computing resources, and services in standard ways that can be processed by computers. Resources and services are represented using the technologies of the semantic Web, such as RDF. The use of semantics to locate data has important implications for integrating computing resources. It implies a two-step access to resources. In step one, a search of metadata catalogues is used to find the resources containing the data or service required by an application. In the second step, the data or service is accessed or invoked.

6.6 Semantic Enterprise Information Integration

The challenges for today's enterprise information integration systems are well understood. In order to manage and use information effectively within the enterprise, three barriers that increase the complexity of managing information have to be overcome: the diverse formats of content, the disparate nature of content, and the need to derive "intelligence" from this content.

Current software tools that look at structuring content by leveraging syntactic search and even syntactic metadata are not sufficient to handle these problems. What is needed is actionable information from disparate sources that reveals non-obvious insights and allows timely decisions to be made. The new concept known as semantic metadata is paving the way to finally realize the full value of information. By annotating or enhancing documents with semantic metadata, software programs can automatically understand the full context and meaning of each document and can make correct decisions about who can use the documents and how these documents should be used.

Semantic is a key enabler for deriving business value via enterprise information integration and can enable the next generation of information integration and analysis software in the following areas (Sheth 2003):

- Extract, organize, and standardize information from many disparate and heterogeneous content sources (including structured, semi-structured, and unstructured sources) and formats (database tables, XML feeds, PDF files, streaming media, and internal documents)
- For a domain of choice, identify interesting and relevant knowledge (entities such as people's names, places, organizations, etc., and relationships between them) from heterogeneous sources and formats.

- Analyze and correlate extracted information to discover previously unknown or non-obvious relationships between documents and/or entities based on semantics (not syntax) that can help in making business decisions.
- Enable high levels of automation in the processes of extraction, normalization, and maintenance of knowledge and content for improved efficiencies of scale.
- Make efficient use of the extracted knowledge and content by providing tools that enable fast and high-quality (contextual) querying, browsing, and analysis of relevant and actionable information.

6.7 Semantic Web Search

Swoogle[1] is a crawler-based indexing and retrieval system for the semantic Web built on top of the Google API. It was developed in the context of a research project of the ebiquity research group at the Computer Science and Electrical Engineering Department of the University of Maryland.

In contrast to Google (Google 2005), Swoogle discovers, analyzes, and indexes Semantic Web Documents (SWD) written in RDF and OWL, rather than plain HTML documents. Documents are indexed using metadata about classes, properties, and individuals, as well as the relationships among them. Unlike traditional search engines, Swoogle aims to take advantage of the semantic metadata available in semantic Web documents. Metadata is extracted for each discovered document and relations (e.g. similarities) among documents are computed. Swoogle also defines an ontology ranking property for SWD which is similar to the pageRank (Brin and Page 1998) approach from Google and uses this information to sort search results. Swoogle provides query interfaces and services to Web users. It supports software agents, programs via service interfaces, and researchers working in the semantic Web area via the Web interface.

Swoogle's database does not stores all of the content of the SWD discovered. It only stores metadata about the documents, the terms, and the individuals they define and use. Currently, the database has information on more that 275 thousand semantic Web documents which contain more than 40 million triples and define more than 90 thousand classes, 50 thousand properties, and 6 million individuals.

[1] http://swoogle.umbc.edu/

A much earlier and commercial effort in building semantic search was Taalee's MediaAnywhere A/V search engine (Townley 2000; Sheth 2001). In this system, ontology driven metadata extraction automatically extracted and refreshed semantic metadata associated with audio/video content rich Web sites. It used ontologies in areas such as Sports, Entertainment, Business and News. Ontology-driven forms based querying supported specification of semantic queries.

6.8 Semantic Web and AI

The merit of the semantic Web is that its concepts and vision are pragmatically oriented. This is a contrast to the speculative aims of Artificial Intelligence (AI). A sharp distinction between semantic Web and AI can be made between the relevance and understanding of data and programs. AI is concerned with highly complex programs being able to understand data, e.g. texts and common sense. The semantic Web is more concerned in making its data "smart" and giving them some machine-readable semantics. While, AI tends to replace human intelligence, semantic Web asks for human intelligence.

Inference mechanisms that can deal with the massive number of assertions that would be encountered by semantic Web applications are required. The claimed power behind many of the proposed applications of semantic Web technology is the ability to infer knowledge that is not explicitly expressed. Needless to say, this feature has attracted the attention from the AI community since they have been dealing with issues relating to inference mechanisms in the past. Inference mechanisms are applicable only in the context of formal ontologies. The idea is to use rules and facts to assert new facts that were not previously known. One of the most common knowledge representation languages has been Description Logic (Nardi and Brachman 2002) on which DAML, one of the earliest semantic Web languages is based.

6.9 Semantic Web and databases

Although an ontology schema may resemble at a representational level a database schema, and instances may reflect database tuples, the fundamental difference is that ontology is supposed to capture some aspect of real-world or domain semantics, as well as represent ontological commitment forming the basis of semantic normalization. Nevertheless, many researchers in the database community continue to express significant reservations toward the semantic Web. The following list shows some examples of remarks about semantic Web technology (Sheth and Ramakrishnan 2003).

"As a constituent technology, ontology work of this sort is defensible. As the basis for programmatic research and implementation, it is a speculative and immature technology of uncertain promise."

"Users will be able to use programs that can understand semantics of the data to help them answer complex questions … This sort of hyperbole is characteristic of much of the genre of semantic web conjectures, papers, and proposals thus far. It is reminiscent of the AI hype of a decade ago and practical systems based on these ideas are no more in evidence now than they were then."

"Such research is fashionable at the moment, due in part to support from defense agencies, in part because the Web offers the first distributed environment that makes even the dream seem tractable."

"It (proposed research in Semantic Web) pre-supposes the availability of semantic information extracted from the base documents -an unsolved problem of many years, …"

"Google has shown that huge improvements in search technology can be made without understanding semantics. Perhaps after a certain point, semantics are needed for further improvements, but a better argument is needed."

These reservations likely stem from a variety of reasons. First, this may be a product of the goals of the semantic Web as depicted in (Berners-Lee, Hendler et al. 2001). Specifically, database researchers may have reservations stemming from the overwhelming role of description logic in the W3C's Semantic Web Activity and related standards. The vision of the semantic Web proposed in several articles may seem, to many readers, like a proposed solution to the long standing AI problems. Lastly, one of the major reservations is related to the concern about the scalability of the three core capabilities for the semantic Web to be successful, namely the scalability of the (a) creation and maintenance of large ontologies, (b) semantic annotation, and (c) inference mechanisms or other computing approaches involving large, realistic ontologies, metadata, and heterogeneous data sets.

6.10 Bioinformatics Ontologies

The integration of information sources in the life sciences is one of the most challenging goals of bioinformatics (Kumar and Smith 2004). In this area, the Gene Ontology (GO) is one of the most significant accomplishments. The objective of GO is to supply a mechanism to guarantee the consistent descriptions of gene products in different databases. GO is rapidly acquiring the status of a *de facto* standard in the field of gene and gene product annotations (Kumar and Smith 2004). The GO effort includes the development of controlled vocabularies that describe gene products, establishing associations between the ontologies, the genes, and the gene products in the databases, and develop tools to create, maintain, and use ontologies (see http://www.geneontology.org/). GO has over 17,000 terms and it is organized in three hierarchies for molecular functions, cellular components, and biological processes (Bodenreider, Aubry et al. 2005).

Another well-known life science ontology is the Microarray Gene Expression Data (MGED) ontology. MGED provides standard terms in the form of an ontology organized into classes with properties for the annotation of microarray experiments (MGED 2005). These terms provide an unambiguous description of how experiments were performed and enable structured queries of elements of the experiments. The comparison between different experiments is only feasible if there is a standardization in the terminology for describing experimental setup, mathematical post-processing of raw measurements, genes, tissues, and samples. The adoption of common standards by the research community for describing data makes it possible to develop systems for the management, storage, transfer, mining, and sharing of microarray data (Stoeckert, Causton et al. 2002).

If data from every microarray experiment carried out by different research group were stored with the same structure, in the same type of database, the manipulation of data would be relatively easy. Unfortunately, in practice, different research group have very different requirements and, therefore, applications need mappings and translations between the different existing formats (Stoeckert, Causton et al. 2002).

Software programs utilizing the MGED ontology generate forms for annotation, populate databases directly, or generate files in an established format. The ontology can be used by researchers to annotate their experiments as well as by software developers to implement practical applications.

7. CONCLUSIONS

Since its creation, the World Wide Web has allowed computers only to understand Web page layout for display purposes, without having access to their intended meaning. Now the Web has advanced to a lot more than a medium to publish data and documents; a Web resource can be a component of what is called deep web (such as a queryable database) or a service that wraps an application. The semantic Web aims to enrich this Web with a layer of machine-understandable metadata to enable the machine processing of information and services. The semantic Web is not a separate Web but an extension of the current one, in which information and services are given well-defined meaning, thereby better enabling computers and people to work in cooperation. To make possible the creation of the semantic Web the W3C (World Wide Web Consortium) has been actively working on the definition of open standards, such as the RDF (Resource Description Framework) and OWL (Web Ontology Language), and encourage their use by both the industry and academia. These standards are also important for e-commerce and e-science, involving sharing of services and the integration for intra- and inter-business processes that have become widespread due to the development of business-to-business and business-to-customer infrastructures.

To fully appreciate the objective of semantics and the semantic Web, it is essential to comprehend what is the place and role of semantics in science in general and computer science in particular. The heterogeneity of the data occurs when there are differences in syntax, representation (e.g. format or structure), and semantics of data. Dealing with heterogeneity has continued to be a key challenge since the time it has been possible to exchange and share data between computers and applications. Given the ease of publication and sharing of data and services on the Web, and the scale involved, the problem has assumed greater importance on the Web. From the various types of heterogeneity, the semantic heterogeneity is a particularly vexing problem. It arises due to a disagreement about the meaning, interpretation, or intended use of the same or related data. One approach to the problems of semantic heterogeneity is to rely on the technological foundations of the semantic Web. More precisely, to define the meaning of the terminology of data using the concepts present in an ontology to make clear the relationships and differences between concepts.

The theories, methodologies, algorithms, and technologies associated with semantic Web make this approach to application and data integration a strong candidate to solve many problems that current systems face. Currently, Web services, tourism information systems, digital libraries, and bioinformatics are some of the leading areas that are studying the potential brought by semantics and ontologies to solve the integration and interoperability problems they have been confronted for many years. For example, semantic Web services are the result of the evolution of the syntactic definition of Web services and the semantic Web. The idea behind Web services is to map concepts in a Web service description to ontological concepts. Using this approach, users can explicitly define the semantics of a Web service for a given domain. Afterwards, using the semantics added to Web services we are able to construct queries using concepts defined in an ontological domain to enable the discovery of service obtaining search results that are more relevant than keyword or attribute-based matching algorithms. Even more significant advantages can be realized when developing mappings for exchanging messages between services participating in a process.

8. QUESTIONS FOR DISCUSSION

Beginner:

1. Why is the search provided by Google, Yahoo! and MSN not semantic?
2. Why and how can metadata help in dealing with unstructured, semi-structured, and structured data?

Intermediate:

1. Why almost all of the semantic metadata efforts involve textual data? Does it make sense to have an ontology of icons or symbols?

2. What would it take to represent concepts found in the natural world, such as compounds and molecules?
3. Distinguish between database schemas and ontologies in terms of conceptual models or representation languages, intentions or uses, and development methodologies.
4. List various techniques used for metadata extraction from different computer science areas.
5. What are the differences in metadata for Web resources that are data versus services?
6. How would Amazon benefit from the use of a product ontology?

Advance:
1. Discuss how would you define the quality of an ontology.
2. Distinguish between ontologies (representation, extraction/population, etc.) when modeling human activities (e.g., travel, financial services, sports, entertainment) versus natural phenomena and sciences (e.g., earthquakes, complex carbohydrates, protein-protein interactions, cancer research).

Practical Exercises:
1. Identify unstructured, semi-structured and structured documents on the same subject matter, such as a new on a football game (although actual content may be different). Develop a small ontology related to this subject matter. Annotate each of these documents.
2. Obtain at least one RDF(S) and one OWL ontology and load it using an ontology editor (e.g., Protégé).
3. Look up a tool or service on the Web for annotating Web pages and Web services.
4. Take a Web page on a news site. Design a small ontology related to the subject matter or domain of that page. Write syntactic, structural, and semantic metadata of that page.

9. SUGGESTED ADDITIONAL READING

- Antoniou, G. and van Harmelen, F. *A semantic Web primer*. Cambridge, MA: MIT Press, 2004. 238 pp.: This book is a good introduction to Semantic Web languages.
- Pollock, J. and Hodgson, R. *Adaptive Information: Improving Business Through Semantic Interoperability, Grid Computing, and*

Enterprise Integration, Wiley-Interscience, September 2004: Practitioners should find this book to be quite valuable companion.

- Gómez-Pérez, A., Fernandez-Lopez, M., and Corcho, O. *Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web* (Advanced Information and Knowledge Processing), Springer-Verlag, October 2003, 420 pp.: The book presents the practical aspects of selecting and applying methodologies, languages, and tools for building ontologies and describes the most outstanding ontologies that are currently available.

10. REFERENCES

Akkiraju, R., J. Farrell, et al. (2005). Web Service Semantics - WSDL-S, http://lsdis.cs.uga.edu/projects/meteor-s/wsdl-s/.

Berners-Lee, T., J. Hendler, et al. (2001). The Semantic Web. Scientific American. **May 2001**.

Bodenreider, O., M. Aubry, et al. (2005). Non-Lexical Approaches to Identifying Associative Relations in the Gene Ontology. Pacific Symposium on Biocomputing, Hawaii, USA, World Scientific.

Brin, S. and L. Page (1998). The anatomy of a large-scale hypertextual Web search engine. Seventh World Wide Web Conference, Brisbane, Australia.

Cardoso, J. (2004). Issues of Dynamic Travel Packaging using Web Process Technology. International Conference e-Commerce 2004, Lisbon, Portugal.

Cardoso, J. (2005). E-Tourism: Creating Dynamic Packages using Semantic Web Processes. W3C Workshop on Frameworks for Semantics in Web Services, Innsbruck, Austria.

Cardoso, J., C. Bussler, et al. (2005). Tutorial: Lifecycle of Semantic Web Processes. The 17th Conference on Advanced Information Systems Engineering (CAiSE'05), Porto, Portugal.

Cardoso, J., J. Miller, et al. (2004). "Modeling Quality of Service for workflows and web service processes." Web Semantics: Science, Services and Agents on the World Wide Web Journal **1**(3): 281-308.

Cardoso, J. and A. Sheth (2003). "Semantic e-Workflow Composition." Journal of Intelligent Information Systems (JIIS). **21**(3): 191-225.

Chinnici, R., M. Gudgin, et al. (2003). Web Services Description Language (WSDL) Version 1.2, W3C Working Draft 24, http://www.w3.org/TR/2003/WD-wsdl12-20030124/.

Curbera, F., W. Nagy, et al. (2001). Web Services: Why and How. Workshop on Object-Oriented Web Services - OOPSLA 2001, Tampa, Florida, USA.

DDC (2005). Dewey Decimal Classification, OCLC Online Computer Library Center, http://www.oclc.org/dewey/.

Dill, S., N. Eiron, et al. (2003). SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation. 12th international conference on World Wide Web, Budapest, Hungary, ACM Press, New York, NY, USA.

Fisher, M. and A. Sheth (2004). Semantic Enterprise Content Management. Practical Handbook of Internet Computing. C. Press.

Fodor, O. and H. Werthner (2004-5). "Harmonise: A Step Toward an Interoperable E-Tourism Marketplace." International Journal of Electronic Commerce **9**(2): 11-39.

Foster, I. and C. Kesselman (1999). The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann.

Google (2005). Google Search Engine, www.google.com.

Grau, B. C. (2004). A Possible Simplification of the Semantic Web Architecture. WWW 2004, New York, USA.

Hammond, B., A. Sheth, et al. (2002). Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content. Real World Semantic Web Applications. V. Kashyap and L. Shklar, IOS Press**:** 29-49.

Handschuh, S., S. Staab, et al. (2002). S-CREAM - Semi-automatic CREAtion of Metadata. LNCS - Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web. London, UK, Springer-Verlag. **2473:** 358-372.

Kashyap, V. and A. Sheth (1996). Semantic heterogeneity in global information systems: The role of metadata, context and ontologies. Cooperative Information Systems: Current Trends and Applications. M. Papzoglou and G. Schlageter. London, UK, Academic Press**:** 139-178.

Kim, W. and J. Seo (1991). "Classifying schematic and data heterogeinity in multidatabase systems." IEEE Computer **24**(12): 12-18.

Klein, M. and A. Bernstein (2001). Searching for Services on the Semantic Web Using Process Ontologies. International Semantic Web Working Symposium (SWWS), Stanford University, California, USA.

Kumar, A. and B. Smith (2004). On Controlled Vocabularies in Bioinformatics: A Case Study in Gene Ontology. Drug Discovery Today: BIOSILICO. **2:** 246-252.

Lassila, O. and R. Swick (1999). Resource Description Framework (RDF) model and syntax specification., W3C Working Draft WD-rdf-syntax-19981008. http://www.w3.org/TR/WD-rdf-syntax.

LCCS (2005). The Library of Congress, Library of Congress Classification System, http://www.loc.gov/catdir/cpso/lcco/lcco.html.

Lee, Y. L. (2005). Apps Make Semantic Web a Reality (http://68.236.189.240/article/story-20050401-05.html). SDTimes.

MGED (2005). Microarray Gene Expression Data Society, http://www.mged.org/.

Nardi, D. and R. J. Brachman (2002). An Introduction to Description Logics. Description Logic Handbook. F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi and P. F. Patel-Schneider, Cambridge University Press.**:** 5-44.

OTA (2004). OpenTravel Alliance.

Ouskel, A. M. and A. Sheth (1999). "Semantic Interoperability in Global Information Systems. A brief Introduction to the Research Area and the Special Section." SIGMOD Record **28**(1): 5-12.

OWL (2004). OWL Web Ontology Language Reference, W3C Recommendation, World Wide Web Consortium, http://www.w3.org/TR/owl-ref/. **2004**.

OWL-S (2004). OWL-based Web Service Ontology. **2004**.

Patel-Schneider, P. and J. Siméon (2002). The Yin/Yang web: XML syntax and RDF semantics. 11th international conference on World Wide Web, Honolulu, Hawaii, USA.

Patil, A., S. Oundhakar, et al. (2004). MWSAF - METEOR-S Web Service Annotation Framework. 13th Conference on World Wide Web, New York City, USA.

Peirce, C. (1960). Collected Papers of Ch. S. Peirce (1931-1935). Cambridge, Mass, Harvard University Press.

RDF (2002). Resource Description Framework (RDF), http://www.w3.org/RDF/.

Rodríguez, A. and M. Egenhofer (2002). "Determining Semantic Similarity Among Entity Classes from Different Ontologies." IEEE Transactions on Knowledge and Data Engineering (in press).

Roure, D., N. Jennings, et al. (2001). Research Agenda for the Future Semantic Grid: A Future e-Science Infrastructure http://www.semanticgrid.org/v1.9/semgrid.pdf.

Sheth, A. (1998). Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. Interoperating Geographic Information Systems. M. F. Goodchild, M. J. Egenhofer, R. Fegeas and C. A. Kottman, Kluwer, Academic Publishers**:** 5-30.

Sheth, A. (2001). Semantic Web and Information Brokering: Opportunities, Early Commercialization, and Challenges. Workshop on Semantic Web: Models, Architectures and Management, Lisbon, Portugal.

Sheth, A. (2003). Semantic Meta Data For Enterprise Information Integration. DM Review Magazine. **July 2003**.

Sheth, A. (2004). From Semantic Search & Integration to Analytics. Dagstuhl Seminar on Semantic Interoperability and Integration, http://www.dagstuhl.de/04391/Materials.

Sheth, A. (2005). Enterprise Application of Semantic Web: the Sweet Spot of Risk and Compliance. IFIP International Conference on Industrial Applications of Semantic Web (IASW2005), Jyväskylä, Finland, Springer.

Sheth, A., B. Arpinar, et al. (2002). Relationships at the Heart of Semantic Web: Modeling, Discovering, and Exploiting Complex Semantic Relationships. Enhancing the Power of the Internet: Studies in Fuzziness and Soft Computing. M. Nikravesh, B. Azvin, R. Yager and L. A. Zadeh, Springer-Verlag.

Sheth, A. and R. Meersman (2002). "Amicalola Report: Database and Information Systems Research Challenges and Opportunities in Semantic Web and Enterprises." SIGMOD Record **31**(4): pp. 98-106.

Sheth, A. and C. Ramakrishnan (2003). "Semantic (Web) Technology In Action: Ontology Driven Information Systems for Search, Integration and Analysis." IEEE Data Engineering Bulletin, Special issue on Making the Semantic Web Real **26**(4): 40-48.

Sheth, A., C. Ramakrishnan, et al. (2005). "Semantics for the Semantic Web: The Implicit, the Formal and the Powerful." Intl. Journal on Semantic Web and Information Systems **1**(1): 1-18.

Sheth, A. P. (1999). Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. Interoperating Geographic Information Systems. C. A. Kottman, Kluwer Academic Publisher**:** 5-29.

Shum, S. B., E. Motta, et al. (2000). "ScholOnto: an ontology-based digital library server for research documents and discourse." International Journal on Digital Libraries **3**(3): 237-248.

Smeaton, A. and I. Quigley (1996). Experiment on Using Semantic Distance Between Words in Image Caption Retrieval. 19th Intemational Conference on Research and Development in Information Retrifval SIGIR'96, Zurich, Switzerland.

Smrz, P., A. Sinopalnikova, et al. (2003). Thesauri and Ontologies for Digital Libraries. 5th Russian Conference on Digital Libraries (RCDL2003), St.-Petersburg, Russia.

Stoeckert, C. J., H. C. Causton, et al. (2002). "Microarray databases: standards and ontologies." Nature Genetics **32**: 469 - 473.

SWSF (2005). Semantic Web Services Framework (SWSF) Overview, W3C Member Submission 9 September 2005, http://www.w3.org/Submission/SWSF/.

TopQuadrant (2005). TopQuadrant, http://www.topquadrant.com/. **2005**.

Townley, J. (2000). The Streaming Search Engine That Reads Your Mind, http://smw.internet.com/gen/reviews/searchassociation/.

UDDI (2002). Universal Description, Discovery, and Integration.

Verma, K., K. Sivashanmugam, et al. (2004). "METEOR-S WSDI: A Scalable Infrastructure of Registries for Semantic Publication and Discovery of Web Services." Journal of Information Technology and Management (in print).

Voorhees, E. (1998). Using WordNet for Text Retrieval. WordNet: An Electronic Lexical Database. C. Fellbaum. Cambridge, MA., The MIT Press**:** 285-303.

Wikipedia (2005). Wikipedia, the free encyclopedia, http://en.wikipedia.org/. **2005**.

WSMO (2004). Web Services Modeling Ontology (WSMO). **2004**.

XMLSchema (2004). XML Schema Part 2: Datatypes Second Edition, W3C Recommendation 28 October 2004.