2

Semantic Web Technologies for Service Description and Knowledge Processing

Akihiro Fujii
Information and Communications Research Unit

1 Introduction

Web service refers to technology that combines multiple data sources accumulated on the web to provide users with desired information services. The Semantic Web, technology that deals with "semantics," represents in the narrow sense the standard defined by W3C regarding the means of publishing documents for information sharing, and in the broad sense, technology that addresses ontology and other semantic aspects of published documents. These technologies could help build advanced information processing environments in which, for example, knowledge can be shared through the Internet. Current trends in web technologies are thus contributing to building environments for more sophisticated knowledge processing and more diverse service applications.

Technologies in this area have been rapidly standardized recently and are becoming more important in industry. Among a range of structured technologies called web services, this report focuses on the Semantic Web and describes its trends. In some research fields in science and technology, there has been a growing need for knowledge processing, which is a goal of the Semantic Web and is a front-line constituent technology in this technological field. This technology is becoming increasingly significant, notably in bioinformatics, as a tool to derive future directions in research from a massive database of accumulated research results.

Chapter 2 describes trends in information distribution technology focusing on knowledge processing. In other words, it explains how recent advances in web technologies, which have enabled efficient information distribution, have been opening new possibilities in the sophistication of services in the business sector and the utilization of information in the academic research sphere.

Chapter 3 gives an overview, among other web service technologies, of the "ontology" -based Semantic Web technology intended for knowledge processing. This technology is currently being internationally standardized and is actively used for scientific and technological studies. Chapter 4, focusing on the research results presented at major international conferences in recent years on the Semantic Web, outlines research trends as well as how research activities should be promoted in this area. Concluding Remarks are given in Chapter 5.

2 Trends in web service technology

This chapter first outlines how web service technology has evolved, then provides specific examples that imply the ideal forms pursued by current Internet-based services. It then examines the benefits of web services.

2-1 The trend of providing services on the web

Technologies that handle the information available on the Internet have evolved through several steps. The first step was the widespread use of e-mail in the early 1990s and the emergence of the World Wide Web (WWW), through which electronic documents formatted in HTML (HyperText Markup Language) can be published. One characteristic of HTML documents is the ability to express hyperlinks, which act as entrances to other documents. However, this capability alone is insufficient for providing links between published documents.

To complement this, XML (eXtensible Markup Language) was born, allowing electronic information produced by any entity to use a common description format. In XML, the data items that represent the product name and the price in a document, for example, can be defined and identified so that different companies involved in merchandise distribution can share forms and other related documents electronically. Moreover, XML contributes to the structured collection and use of information. Structurally collected information can be easily modified for other purposes. These technological advances have made the collected information, or databases, applicable for diverse uses.

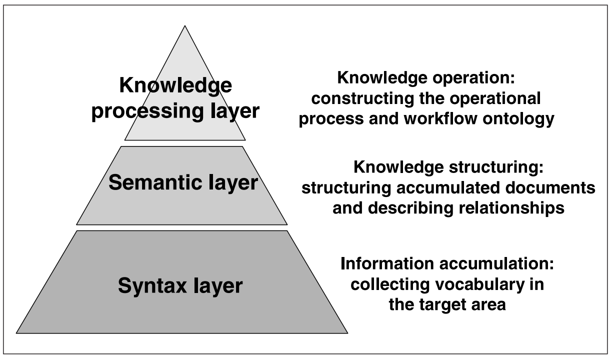
Along with development in sharing and structuring information, environments for document processing and information interoperation have been developed. The programming language called Java, for example, allows information accumulated in XML to be processed to display pages with dynamic movement and enables information to be input from remote places. In other words, these developments have enabled the ability to manipulate documents in various predefined formats in web pages and to provide access to large-scale databases of documented information.

2-2 Providing ideal services

These technologies help construct an integrated environment for operating electronically stored information and documents. In particular, the electronic commerce sector has actively introduced electronic information sharing to manage distribution processes. The services offered in this environment are essentially the content of information provided. This raises the question of what would be the ideal forms of service, or information, that the Internet could provide. This question is explored below, although it may seem to be a hypothetical issue.

Suppose that you are going to take several days off to travel. You are visiting a travel agency so

Figure 1 : Conceptual model of knowledge processing in the Semantic Web



Source: Prepared by STFC

that its agent can help you plan a vacation. The essential information in planning a trip is where and when to go. Next, limiting factors such as budget need to be considered. In addition, you may have preferences regarding airline company, seat type, and so forth. To have your travel agent create your travel itinerary and make the necessary arrangements, which are services you expect to receive, you have to supply such information to the agent.

The most important element of a good travel agent is the ability to communicate actively with the customer. Naturally, the agent and the customer need to share language, vocabulary, and terms. This is a tacit assumption in human-human communication. The next important element is the ability to make a proposal based on the knowledge of the customer's preferences. This hopefully involves not only a one-way supply of information, such as giving recommendations, but services that meet the known customer needs, such as choosing services that accept discount coupons. Ideally, the agent remembers the personal preferences of the customer so that they are reflected in future travel planning. The agent should also be able to perform the necessary tasks autonomously even when the customer is not present, which means saving the customer waiting if it takes time to complete hotel and ticket booking arrangements.

In short, a good agent collects key information, chooses options according the customer’s preferences, and conducts booking and other services with the consent of the customer. In conducting these tasks, the agent needs to share vocabulary with the provider of the information and the customer, and make decisions based on the acquired information.

The amount of information available on the Internet or the web has been growing explosively, and it will reach 100 times today's size in the next few years, according to a projection. There is no doubt that the number of "hits" for searches will become so massive that narrowing down search results will require enormous effort. If this becomes a reality, conventional search engines that depend on statistical operations to rank search information will be no longer able to return satisfactory results. This will raise the ironic problem of inability to find information even if there is plenty of useful information on the web.

The goal of web services is to assimilate different description formats of information sources and to provide a description format for each demanded service that is adapted to it. This allows the selection of services that meet customer needs from a massive information space. If this technology matures, the following web services will become available.

(i) An environment in which people who know little about information technology can easily input and output information
(ii) An Internet world that is accessible as a huge distributed knowledge database
(iii) Creation of new services by integrating data assets that have originally been constructed for different purposes

Web services with such characteristics are expected to contribute, among others, to narrowing the digital divide, addressing aging societies, supporting lifelong study, and controlling diverse information terminals and robots, consequently facilitating aids linked to people's daily lives.

2-3 An overview of Semantic Web technologies

Figure 1 shows a conceptual model of the technologies that help provide these advanced services. The structure can be divided into three tiers: from bottom to top, (i) the syntax layer, (ii) the semantic layer, and (iii) the knowledge processing layer. There are two points of note when hierarchically expressing a software structure. First, the top layer uses the function of the layer below it to implement its own expected function. Second, functions are defined independently among different layers. In other words, the function of the databases constructed using XML and other technologies is used to describe the semantic world, and this description then serves as the base of knowledge processing.

Knowledge processing-based services are conducted by building vocabulary structures called ontologies at the top knowledge processing layer and by describing services that define the relationship among the ontologies. In fact, this hierarchy can be divided into more detailed layers whose specifications have been set up as a standard technology. Chapter 3 further explains this detailed hierarchy known as the "layer cake."

2-4 Possibilities of web services in hot science and technology fields

Some researchers indicate that "service"-related science is essential today, particularly from the viewpoint of innovation policy. Focusing on service is especially important as a direction of development in information technology[1].

Let us take the foresight study conducted by the National Institute of Science and Technology Policy as an example. From all areas of scientific and technological research, it has selected 130 "hot science and technology areas," assumed to be noteworthy in the next 10 years. A closer examination of these areas reveals that, for many of them, the presence of web services and other large-scale information search tools are effective or even indispensable for technological achievements intended in the area.

This applies to the research area of "support for human intelligence," which was selected in the study from the information and communications field, and to "tailor-made medicine" and "biometric personal authentication" in the life science field. These technological areas would benefit from web services.

Similar emphasis on service in the evolution of information technology can be observed in the review results presented by experts in a recent European study on science and technology foresight[10]. This suggests that web services constitute a technological field in which discussion should not be limited to technological advances within the information technology sphere, but should concretely address the possibility of the fusion of web services with other fields and sectors.

3 Structure of Semantic Web technologies

The primary technological schemes for implementing web services are the Semantic Web and UDDI, the latter being explained later. This chapter first addresses the Semantic Web by describing ontology as its key concept, the language known as RDF, which is fundamental to knowledge processing, and the model of service provision. How the Semantic Web relates to UDDI is discussed at the end of the chapter.

3-1 Ontology-based linkage

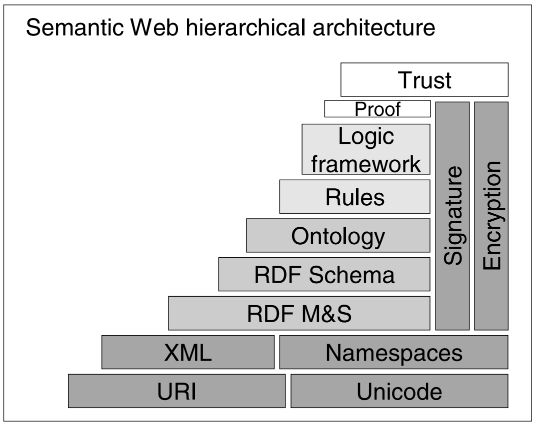
To allow researchers to perform as advanced information searches through a practical-sized database as that described in the example of the travel agent, the construction of ontology is the primary issue.

“Ontology” is a term used in the field of artificial intelligence to represent the structure of vocabulary. For instance, when someone runs a search for the term “Newtonian mechanics,” the search engine is most likely to simply compare text data for matching. In contrast, ontology-based advanced searches refer, for example, to exploiting the relationship of “Newtonian mechanics” with “physics” or even with “relativity theory,” which, like Newtonian mechanics, belongs to physics but is a different piece of knowledge in the area. Describing these relationships requires the accumulation of information according to a specific set of rules known as a logical expression. In summary, ontology-based knowledge processing means constructing systems that can answer diverse queries, using the accumulation of structures.

3-2 The layer cake

The previous section explained that structures are used in the Semantic Web to accumulate and apply information. The diagram in Figure

Figure 2 : Layer cake



Source: Prepared by STFC based on the reference[11]

2, known as the layer cake, defines the process of this structuring. The layers of this hierarchy have been specified one by one, starting from the bottom, and the specification process has so far been completed up to the fifth layer, the Ontology layer. This technological model has been discussed internationally by the World Wide Web Consortium (W3C), a standardization organization[11].

The information available on the web is usually in either text or graphic form and is categorized as “syntax”-level information. This type of information is expressed in syntax. HTML and XML are widely used syntaxes.

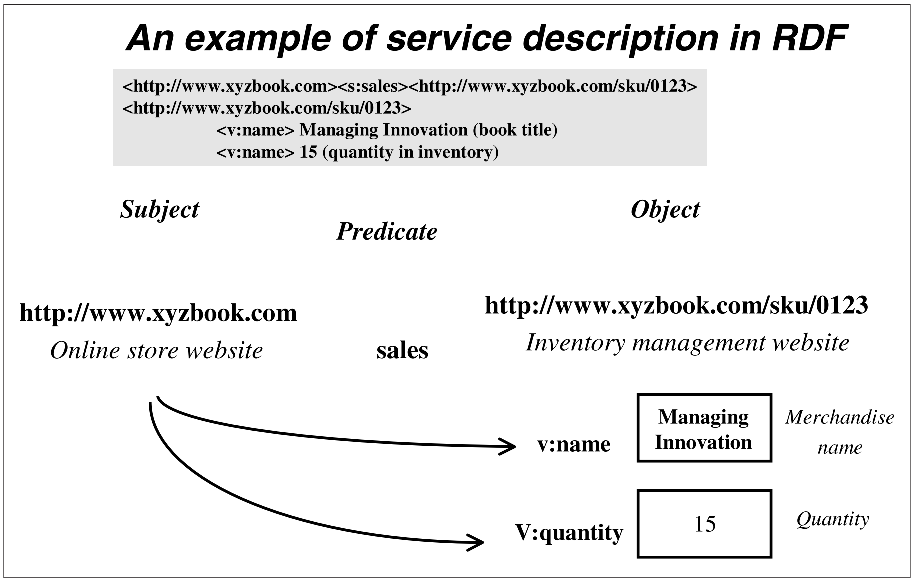
To deal with semantics, some layers are assigned to RDF, a notion explained below. These layers describe a knowledge structure built according to a specific set of rules.

On the top of these is the Ontology layer. In this layer, the target knowledge structures are formed as a vocabulary collected in RDF and as a set of rules to apply it. Genome and other application areas mentioned below are examples of ontologies constructed in this way and systems of using such ontologies.

3-3 Resource Definition Framework (RDF)

Let me briefly give an example of “knowledge description” here. Figure 3 shows a triple syntax called an RDF graph. RDF, an abbreviation for “Resource Definition Framework,” is a concept adopted in defining knowledge structure. Knowledge fragments are expressed in a syntax consisting of three elements: the subject, the predicate, and the object. There are several

Figure 3 : Ontology construction logic



Source: Prepared by STFC

notation methods for this structure.

The example in the above chart refers to an online bookstore. The subject is a collection of books representing the merchandise displayed at the store, for instance. The predicate defines the rule: "sales" based on the "inventory" of the displayed merchandise. Hence, this diagram shows that the web page introduced by the predicate contains descriptions that serve as the object-the properties of the data on individual books-, such as the book title, "Managing Innovation,", and the number of copies in the inventory, "15 copies." The example assumes that the retailer and wholesaler maintain separate databases for merchandise control. The online store in this example uses Semantic Web technology to implement the service of "inventory inquiry."

3-4 Service provision models

Providing services through the web is not a two-fold process simply consisting of the information provider and its user. For example, there may be a business sector in which the provider of service-related information such as inventories of goods is not the business entity engaged in sales. In this case, for consumers to receive more advanced services, this information needs to be shared under specific rules. While inventory information may not be applicable for purposes other than sales, data on academic research, whose applications are not always specified in advance, could be used for a broader range of services.

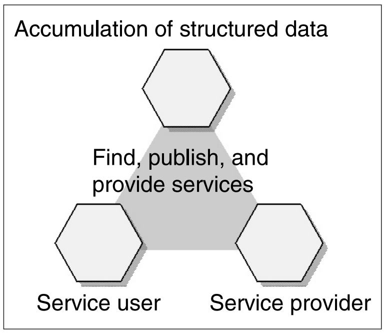
In short, in web services, the three parties - the data provider, the service provider, and the service user - can be independent entities. Therefore, service brokerage for selling certain services is feasible as a business. Figure 4 depicts the conceptual relationship among these entities.

Currently, technologists who promote electronic commerce assume that diverse services in business can be efficiently provided in such a structure. Similarly, in the academic research domain, it would be possible for computers to assist in deriving new academic findings from massive information databases constructed for different purposes in the same field.

3-5 Related technological schemes

For the Semantic Web, W3C has been setting international standards[11] and has published recommendations together with specifications for peripheral technologies. There is another standard developed by the OASIS Consortium, an organization established in 2000. It is called UDDI (Universal Description, Discovery and Integration)[9] and coexists with the Semantic Web standard scheme. The UDDI standard consists of UDDI Data Structure Reference, which is the main part and defines the structure of

Figure 4 : Conceptual model of service brokerage



Source: Prepared by STFC

the data to be stored in the database called the Business Registry, and UDDI Programmer's API, which is a programming interface for developing applications for using UDDI data. It is in this interface section that programmers develop software related to the content of the services intended. UDDI seeks to provide advanced web-based services in the electronic commerce area in particular. Note that under the UDDI standard, no ontology level is specified. This suggests that UDDI depends on descriptions in natural language concerning the semantics of services.

As with Semantic Web technologies, UDDI technologies assume that providing data is independent of the services based on that data, and that constructing services is feasible as a business. This perception is common across all web service technologies.

4 Research trends in the Semantic Web

This chapter describes research activities on Semantic Web technologies and on the application of such technologies to other fields of study, and examines research promotion possibilities in this area.

4-1 Theoretical research on the Semantic Web

In the following section, trends in theoretical research activities are briefly reviewed in connection with major topics. Next, as examples of attempts to extend these activities to applied research, some applied projects in the area of bioinformatics are examined. Ongoing research projects supported by public funds in Japan are then listed with short descriptions.

(1) Semantic integration

In the Semantic Web, more than one ontology can sometimes be created in the same field. This occurs, for example, when document databases developed over many years in different research institutes are to be integrated for interoperation. In this case, technologies for integrating multiple semantics, such as a technology for correlating different ontologies, will play an important role. This underlines the significance of research in these areas as correlation between ontologies and correlation between a global ontology and a local ontology. Studies in these areas are considered to be extensions of data integration research in the database sector of artificial intelligence.

(2) Description logic reasoning

To describe ontologies, logical expressions need to be configured. This process uses a syntax called predicate logic. Ontologies are written in OWL (Web Ontology Writing Language), whose standardization has been conducted by W3C. Describing a knowledge structure in predicate logic corresponds to constructing a set of elements that meet a certain condition, such as "If A, then B." The resulting set is the database fundamental to knowledge processing in the Semantic Web.

Knowledge processing based on predicate logic takes the form of generating answers from a collection of fragments of knowledge, such as "If A is true, then B is satisfied" and "If B, then C," to queries such as "Is Z true, if A holds?" This process is referred to as the reasoning mechanism. An important research direction in the Semantic Web is constructing ontology databases in the target knowledge areas, followed by effectively implementing the reasoning mechanism using OWL in each area.

(3) Finding, linking, and implementing services

The process of finding the necessary service in a network and receiving it corresponds to a user inputting keywords to search through a network for the desired service and executing the program. This can be regarded as an extension of typing keywords in an online search engine and finding the desired website. In this case, the requested service is expressed as a data sequence in certain structure. From this perspective, researchers are studying techniques to determine whether the description of a demanded service matches the description by the service provider as well as the efficiency of such techniques.

Some research projects are seeking to apply query schemes to practical applications and to benchmark such schemes. Others are attempting to dynamically combine services by using reasoning engines. This includes studies on algorisms to identify relations between metadata attached to services to describe the content of service. Some others adopt service condition comparisons and syntax comparisons in these studies.

4-2 Applied research project examples

(1) eScience

A group of researchers including Hendler, at the University of Maryland, U.S., is proposing a new direction in research by naming research that aims to solve problems in diverse scientific fields on grid computer systems “eScience.” To avoid confusion, note that “eScience” addressed here is different from the project carried out in the U.K. on science and the computer[7].

Grid computing is the notion of connecting numerous general-purpose CPUs to perform computations in parallel. To predict the characteristics of a crystal, for example, a researcher needs to examine its chemical structure using existing databases. Grid computing is ideal for large-scale database searches and for relatively simple reasoning tasks based on the search results. This makes grid computing a promising candidate for the Semantic Web processing engine.

Meanwhile, databases of technical knowledge, although accumulated over time, have not necessarily been constructed in a unified manner. They vary in their data format and operational procedure. In other words, data size, computational complexity of service implementation, compatibility with other operations, and so forth differ from one database to another. This implies the need for interoperability. More specifically, this means constructing a platform through which services are provided to write the results of numeric simulations and other processing on the Semantic Web for common use.

The eScience project members have been trying to build a common information base by applying Semantic Web technologies to several existing databases on science and technology. To extract the knowledge that researchers want from fragments of existing information, it is necessary to define uniform semantics across various databases. The scientists involved in the project assume that this will enable services to support scientific and technological research efforts across a wide range of disciplines.

(2) Application examples in bioinformatics

Representative examples of application of the Semantic Web to bioinformatics are three Genome Ontology (GO) projects: myGrid, MBOY-Service, and Semantic-MBOY. These are systems to derive network structures of diverse reaction pathways (e.g. metabolic pathways) in the living body from data relationships.

It is assumed that in the living body there are reaction pathways for many different substances that are networked under complex control. In fact, researchers have so far identified around 17,000 rules associated with reaction. These reaction pathways are used to transfer signals. GO systems serve as tools to effectively retrieve pathways from the literature. They help find rules regarding the base pairs and proteins relevant to the target effect from millions of chemical reactions.

Common goals pursued in these research projects include automatic information service, providing structure messages and middleware for implementing this service, desirable forms of displaying information service results and of interfacing, measures to meet the different needs of individual users, and constructing complex ontologies.

(3) Research projects in Japan

Listed below are examples of Japanese projects seeking to construct ontologies in science and technology selected from the studies presented

Table 1 : Major Japanese research projects on the Semantic Web

Research content	Affiliation of researchers
Structuring the knowledge on nanomaterials technology	Osaka Univ. Institute of Scientific and Industrial Research
Support for developing and operating space systems	Japan Aerospace Exploration Agency, Osaka Univ.
Knowledge structuring in biology	National Institute of Advanced Industrial Science and Technology, Computational Biology Research Center
Drug function	RIKEN, Genomic Sciences Center

Source: Prepared by STFC

at the Japanese Society for Artificial Intelligence. Two of them are briefly described here.

The Institute of Scientific and Industrial Research (ISIR) of Osaka University has been working on constructing and using ontologies to structure the knowledge on nanomaterials technology. Nanotechnology research is related to a wide range of existing research areas. Aware of this, ISIR intends to develop a “conceptual” interface through which common notions across multiple fields are provided. Its current project specifically addresses knowledge structuring by analyzing patent information.

The Genomic Sciences Center, on the other hand, has been developing a knowledge base of reaction pathways that drugs have at the molecular level. The first step is to describe interactions between substances in RDF using the relationship between the three elements of “drug,” “biomolecule,” and “output from interactions.” On the database that has accumulated their relations, drug interaction ontology is implemented. The basic rules regarding the combination of interactions are then defined to enable reasoning on the database. The result will be a tool that runs effective queries on drug-interaction relationships. With this tool, researchers will be able to infer reaction pathways from retrieved interactions and thereby discover new reaction pathways.

(4) The significance of Semantic Web research in bioinformatics

One study published in the domain of bioinformatics mentions the following as general findings from the research activities of authors for the application of the Semantic Web.

First, constructing ontologies regarding knowledge structures intrinsic to a field is extremely difficult. If database assets already exist, this difficulty increases because of the need to ensure compatibility with their data format. The authors point out that these factors make research in bioinformatics take a significant amount of time and human resources. They also note that an approach in which the users of the Semantic Web describe the services they want using a syntax of the Semantic Web so that they can find the appropriate services through the notation is a promising model applicable to many other fields. Bioinformatics is a representative field where pioneering studies are conducted on the Semantic Web. Adopting search tools that involve knowledge processing based on a large-scale database is likely to spread to other scientific disciplines in the future. Therefore, research techniques developed in bioinformatics could have a ripple effect on research on specialized databases in many fields.

4-3 How research should be promoted

Last November, the 3rd International Semantic Web Conference, 2004, or ISWC2004, the major international forum on research in Semantic Web technologies, was held in Hiroshima, Japan. Of the total participants of 450, over 300 came from overseas, creating an exceptionally cosmopolitan mood for an international conference in Japan. Among the papers presented, whose quality was generally high, that delivered by a researcher who appeared to be in his twenties was particularly impressive.

The keynote speech was made by Edward Feigenbaum, a professor at Stanford University, who is known as the father of artificial intelligence and the 1994 winner of the Turing Award. In his speech, he clearly expressed how Semantic Web research should be promoted. His strong message about this challenge was "Give me something that works!" As demonstrated in constructing and operating/managing ontologies, Semantic Web research applies the outcomes of advanced studies on artificial intelligence. Through this message, Feigenbaum probably intended to stress the importance of practical application in this research area, where language models and other theoretical studies are often the focus of attention. Not only did the professor underline the significance of real-world applications, he also added his hopes for the "spiral development of theory and practice."

If approaches to research promotion were split between the discipline type and the mission type, the former would refer to efforts voluntarily made within the framework of conventional activities in scientific societies, and the latter would indicate approaches that cross the existing boundaries of specialties that are conducted under top management control as projects progress. From this point of view, the spiral development mentioned by Feigenbaum requires mission-type projects to be set up.

One example of experiments aimed at demonstrating the effectiveness of the Semantic Web is the Semantic Web Challenge, an international contest of application systems[8]. This event, presented by a group of researchers and funded by several scientific societies, intends to better illustrate Semantic Web technologies to society and stimulate current research activities to higher goals.

In 2003, the first year of the contest, among the 10 projects that participated, the winner was CS AKTive Space, an application developed by the University of Southampton, U.K. It was a system designed for searching a database of British computer scientists. For the 2004 contest, 18 teams have submitted applications as of December. There has been only one attempt from Japan, "Semblog," which was submitted in 2003 by the National Institute of Informatics. More active participation of Japanese universities in such events is required.

Professor Feigenbaum also said that to develop research in the Semantic Web area, where connections with application fields are essential, drawing the "path" of research is important in formulating research plans. To explain the path, he cited robot soccer tournaments as an example. He pointed out that although this idea may have caused laughter among researchers at first, continued events year after year have produced a steady stream of research outcomes in the area of distributed artificial intellisgence. The professor compared this with an opposite example in which research on certain pattern recognition has led to an impasse. He stated that given today's improved processing capacity, pursuing recognition accuracy alone at the level of a fraction of a percent is meaningless as research.

Since the 1980's boom in artificial intelligence research, when large projects were launched and implemented, activity in Japanese academic communities in knowledge processing is assumed to remain relatively high. The problem, however, is that many research projects have followed the discipline model. For future promotion in the Semantic Web area, interdisciplinary research projects should be managed with clearly defined missions, as Feigenbaum's message suggests. In other words, it is essential to set up, beyond the spiral path, environments and goals that can bring meaningful achievements in both theory and application.

5 Conclusion

This article has focused on Semantic Web technologies because I consider that their inherent orientation toward service is important not only for future directions in information technology advancement but also as a research topic in many interdisciplinary fields of science and technology.

Currently, active discussions are held to implement Japan's Third Science and Technology Basic Plan. Information and communications technology is considered to continue to be a key field in the nation's science and technology policy. In interdisciplinary areas between this field and others, in particular, there are many technological development issues that have been recognized as critical.

To promote research in these interdisciplinary areas, mission-oriented management that emphasizes the "path" is essential, and this is the very type of management approach that Semantic Web research needs.

References

[1] National Innovation Initiative, "Innovate America: Thriving in a World of Challenges and Change," Dec. 2004

[2] Junichi Akahani, Noriaki Izumi, Takahiro Kawamura, and Hideaki Takeda, "Semantic Web Research Trends Observed in ISWC2003," material for a study group of the Japanese Society for Artificial Intelligence, SIG-SW&ONT-A302-10 (in Japanese)

[3] Interoperability Technology Association for Information Processing, Japan, "Research Report on Semantic Web Technologies," March 2003. (in Japanese)

[4] Tatsuya Hagino, "What is the Semantic Web?" Joho Shori, Vol.43, No. 7, July 2002 (in Japanese)

[5] Masaki Kiyono, "Web Ontology Language for the Semantic Web," same journal issue as above (in Japanese)

[6] Sheila A. McIlraith, Dimitris Plexousakis, and Frank van Harmelen (Eds.), "The Semantic Web-ISWC2004," Lecture Notes in Computer Science, Springer, Nov. 2004

[7] D. De Roure, Y. Gil, and J. Hendler, "E-Science," IEEE Intelligent Systems, 19 (1): pp. 24-25, 2004

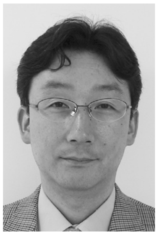
[8] Website of the Semantic Web Challenge: http://challenge.semanticweb.org/

[9] Website of the OASIS Consortium on UDDI standardization: http://www.uddi.org/

[10] European foresight research project on information and communications technology: http://fistera.jrc.es/

[11] Website for the standardization of the Semantic Web: http://www.w3.org/2001/sw/

[12] Tim Berners-Lee and James Hendler, "Publishing on the semantic web," Nature, 410, 1023-1024 (26 Apr., 2001)



Akihiro FUJII, Ph. D.
Information and Communications Research Unit, Science and Technology Foresight Center
Dr. Eng: After engaging in research on distributed computing and communications protocols, he has conducted a project to construct an electronic commerce system. His current area of interest is the potential impact of innovations in information and communications technology on business administration and national policies.

(Original Japanese version: published in April 2005)