# Approximate Query Reformulation for Multiple Ontologies in the Semantic Web

*Jun-ichi Akahani*†*, Kaoru Hiramatsu, and Tetsuji Satoh*

**Abstract**

The Semantic Web has been attracting a great deal of interest as the next-generation Web. It enables intelligent processing of Web information based on semantic structures called "ontologies." The decentralized nature of the Web requires integration of multiple ontologies. Ontology integration requires approximation mechanisms, since there is often no perfectly corresponding ontology. However, most previous research efforts on ontology integration have not provided clear semantics for approximation. Therefore, we propose an approximate query reformulation framework with clear semantics for integrating multiple ontologies. In this article, we provide a brief overview of the Semantic Web and present an approximate query reformulation framework.

## 1. Introduction

Information spaces, the Web on the Internet in particular, continue to grow explosively in a borderless way. In the real world, on the other hand, ongoing globalization has been diversifying our lifestyles. Providing information for a user's daily life in this environment requires semantic interoperability among heterogeneous information services.

To achieve semantic interoperability, semantic structures called "ontologies" are widely used in the fields of artificial intelligence, databases, and the Semantic Web [1]. For example, the Semantic Web enables users to make queries about restaurants, *e. g.*, location and hours of operation, based on an ontology for restaurants. However, the decentralized nature of the Web makes it difficult to construct or standardize a single ontology. Thus, integration of multiple ontologies is one of the key technologies that need to be developed for the Semantic Web.

When integrating ontologies, exact correspondences between them are seldom found. For example, there may be no corresponding class for Cajun restaurants in a Japanese ontology for restaurants. In such cases, one could use an approximation mechanism to replace "Cajun" with "American restaurant" in the Japanese ontology. However, most previous research efforts on ontology integration have not provided clear semantics for approximation.

Our group previously proposed an approximate query reformulation framework with clear semantics to meet the challenge of integrating multiple ontologies [2]. In this framework, a query represented in one ontology is reformulated approximately to a query represented in another ontology. This reformulation is based on an ontology-mapping specification. To achieve closer approximation, this framework provides specialization and generalization operators with clear semantics for approximate query reformulation.

In this article, we first provide a brief overview of the Semantic Web and then describe our approximate query reformulation framework.

## 2. The Semantic Web

### 2.1 Overview

Current Web pages are prepared primarily for use by people. Consequently, it is difficult for computers to process Web information. Consider, for example, the case of searching for a restaurant named "London." Current Web search engines cannot distinguish whether the "London" on each Web page it finds with that word means a location, a person's name, or a restaurant name, so they list numerous irrelevant

† NTT Communication Science Laboratories
Soraku-gun, 619-0237 Japan
E-mail: akahani@cslab.kecl.ntt.co.jp

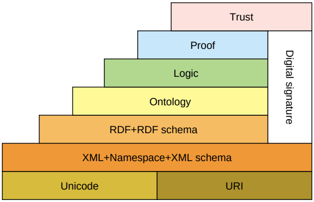


Fig. 1. Semantic Web layers.
http://www.w3c.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

search results. If each Web page had information indicating the meaning of "London", then the search engine could produce much better search results. To define such metadata, we need semantic structures, such as the knowledge that a restaurant has a name, a location, a phone number, and so on. In the Semantic Web, each Web document is annotated with metadata based on semantic structures called "ontologies," and users can make queries based on these ontologies.

Tim Berners-Lee, the creator of the Web, proposed the Semantic Web in the late 1990s. The goal of the Semantic Web is to realize the Web of Trust, shown in Fig. 1. The layers below the XML layer essentially correspond to the current (or near future) Web. The RDF (resource description framework) layer provides a description framework for metadata. Ontologies are described in the ontology layer. The logic and proof layers provide proofs, *i.e.*, how information is derived from resources such as Web documents. Together with the proof of derived information and digital signatures, the Web of Trust is achieved.

The W3C (WWW Consortium) has been standardizing description frameworks for metadata and ontologies [3]. The first version of RDF was standardized as a description framework for metadata in 1999. The W3C is currently standardizing OWL (Ontology Web Language) as a description framework for ontologies.

### 2.2 Ontology, metadata, and contents

The relationship among ontology, metadata, and contents is shown in Fig. 2. The ontology layer provides a framework for describing the semantic structure. Ontologies are described using classes and

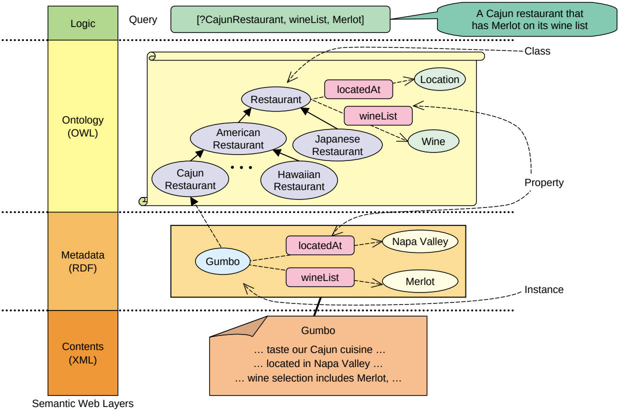


Fig. 2. Relationship among ontology, metadata and contents.

properties. For example, “Restaurant” and “Location” are classes, and “locatedAt” and “wineList” are properties. Classes are characterized by their properties. For example, “Restaurant” has “locatedAt” and “wineList” properties, and the value of the “locatedAt” property is in the “Location” class. OWL provides several built-in properties, such as “subClass-Of,” for describing class hierarchies. For example, Fig. 2 shows that “American Restaurant” is a subclass of “Restaurant.” (We denote the “subClass-Of” property as a solid arrow for easy visualization.) There is also a built-in “subPropertyOf” property for describing the property hierarchy, which will be discussed later.

Contents are annotated with metadata based on ontologies. For example, the Web page for “Gumbo” has the metadata shown in Fig. 2. The metadata indicates that “Gumbo” is an instance of “Cajun Restaurant” and gives its location and wine list. (The notions of class and instance are similar to those in object-oriented programming, but properties are not inherited as subclasses in the Semantic Web. We assume the “locatedAt” and “wineList” properties are also defined for the “Cajun Restaurant” class.)

The Semantic Web allows users to make queries based on ontologies. For example, one can make a query about a Cajun restaurant that has Merlot on its wine list, as shown in Fig. 2. (Note that there is no standard query language as of April 2003. The query syntax shown in the figures is not a standard one.) A Semantic Web query engine answers “Gumbo” to this query by using the ontology and metadata.

## 3. Need for integration of multiple ontologies

As shown in the previous section, ontologies play the central role in the Semantic Web. However, the decentralized nature of the Web makes it difficult to construct or standardize a single ontology. Regional information is one reason for this because ontologies vary with the region due to the cultural differences among regions. For example, the Yahoo! Japan site has different information categorization (*i.e.*, ontology) from the Yahoo! site in the U.S.

Furthermore, it is desirable for a user to obtain information based on his or her own ontology. For example, users of the Yahoo! Japan site may want to access U.S. regional information but with a categorization that is familiar to them. It is also desirable for a user to publish information based on his or her own ontology.

Ontologies also vary over time. For example, ontologies for regional information reflect continuing changes in industrial structures, culture, and so on. From the viewpoint of scalability, it is impossible to maintain global consistency with a single huge ontology. Furthermore, different people may update or customize an ontology independently. We thus have to handle different versions of ontologies.

Therefore, integration of multiple ontologies is one of the key technologies that need to be developed for the Semantic Web. Ontology integration requires approximation mechanisms, since ontologies generally do not perfectly correspond with each other.

## 4. Approximate query reformulation

### 4.1 Overview

To integrate multiple ontologies, our group proposed an approximate query reformulation framework with clear semantics [2]. In this framework, a query represented in ontology A (say, in the United States) is reformulated approximately to a query represented in ontology B (say, in Japan) based on an ontology-mapping specification, as shown in Fig. 3. A Semantic Web query engine processes the reformulated query for metadata based on ontology B.

In our framework, ontology-mapping specifications are described as an ontology. For example, the ontology mapping in Fig. 3 specifies that the “American Restaurant” class in ontology A is equivalent to the “Beikoku RyouriTen” class in ontology B and that the “wineList” property in ontology A is a subproperty of the “menu,” “adultMenu,” and “drinkMenu” properties in ontology B.

Our framework provides reformulation operators for reformulating queries. There may be many possible reformulated queries, but we prefer close approximation. Therefore, we specifiy two kinds of reformulation: specialization and generalization.

With the specialization operators, the reformulated query is maximally covered by the original query; *i.e.*, all the answers of the reformulated query are precise, but some precise answers may be missed. With the generalization operators, the reformulated query minimally covers the original query; *i.e.*, the answers of the reformulated query contain all of the precise answers, but there may also be imprecise answers.

### 4.2 Example

In the example shown in Fig. 3, ontology B has no class corresponding to “Cajun Restaurant”; however, there is a “Beikoku RyouriTen” class, which is equivalent to “American Restaurant,” a super-class of

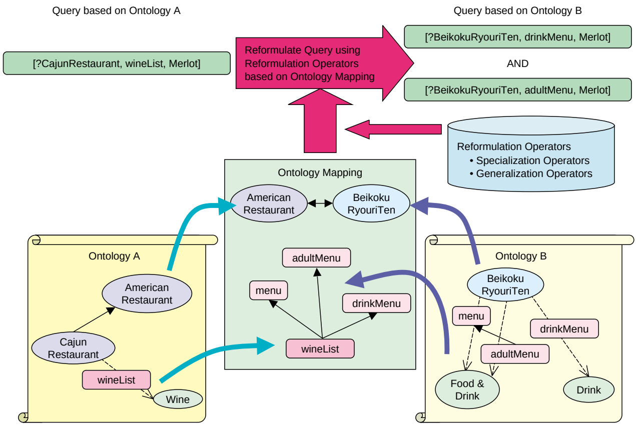


Fig. 3. Approximate Query Reformulation.

"Cajun Restaurant." Therefore, "Beikoku RyouriTen" contains all Cajun restaurants in Japan, as well as other subclasses of "American Restaurant," *e.g.*, Hawaiian restaurants.

Similarly, ontology B has no property corresponding to "wineList", but there are three super-properties ("menu," "adultMenu," and "drinkMenu") of "wineList." Reformulation into only "menu" is over generalized since children's menus could also be included in restaurant menus. As "adultMenu" is defined to be a sub-property of "menu," we can do without "menu." In this manner, we take into consideration only the lower bounds of super-properties (or super-classes). In this case, two properties, "adultMenu" and "drinkMenu," are the lower bounds of the super-properties. Reformulation into only "adultMenu" is also over generalized as it may contain food items for adults.

In our approximate query reformulation framework, generalization operators reformulate a class (or property) by using the lower bounds of super-classes (or super-properties). Consequently, a query about a Cajun restaurant that has Merlot on its wine list is reformulated into a query about a "Beikoku RyouriTen" that has Merlot on its menu for adults and on its menu for drinks.

The theoretical foundation of the Semantic Web is artificial intelligence and database theory. In particular, query processing in the Semantic Web is based on description logic. We previously demonstrated the correctness of our approximate query reformulation framework based on description logic [4].

## 5. Conclusion

Ontologies play the central role in the Semantic Web, and integration of multiple ontologies is one of the key technologies that need to be developed for the Semantic Web. We have developed an approximate query reformulation framework for integrating multiple ontologies. The correctness of the framework was shown theoretically.

The approximate query reformulation framework has been incorporated into the GeoLinkAgent system, in which agents coordinate regional information services. Approximate query reformulation is required for domains that have cross-cultural aspects because ontologies vary from region to region due to cultural differences.

The prototype system is being used as a platform for our Real-World Semantic Web project. The goal of the project is to enrich communication among people in a ubiquitous environment based on Semantic Web technologies. We are developing technologies for understanding conversations based on ontologies, a personal repository for inter-personal communication, and semantic integration of ubiquitous contents, which include data from distributed heterogeneous sensors such as GPS devices attached to users and video cameras located in the environment. We envision these technologies as tools that will enable us to advance our work in much more diverse and effective ways.

## References


[1] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific American, Vol. 284, No. 5, pp. 34-43, 2001.

[2] J. Akahani, K. Hiramatsu, and K. Kogure, "Coordinating Heterogeneous Information Services based on Approximate Ontology Translation," AAMAS-2002 Workshop on Agentcities: Challenges in Open Agent Systems, pp. 10-14, 2002.

[3] W3C Semantic Web Activity, http://www.w3.org/2001/sw/.

[4] J. Akahani, K. Hiramatsu, and T. Satoh, "Approximate Query Reformulation based on Hierarchical Ontology Mapping," International Workshop on Semantic Web Foundations and Application Technologies (SWFAT), pp. 43-46, 2003.



**Jun-ichi Akahani**

Senior Research Scientist, Supervisor, Social Communication Laboratory, NTT Communication Science Laboratories.


He received the B.E. and M.E. degrees in Applied Mathematics and Physics from Kyoto University in 1983 and 1985, respectively. In 1985, he joined the Electrical Communication Laboratories. From 1989 to 1990, he was a Visiting Scholar at the Robotics Laboratory, Stanford University. He has been conducting research in the fields of Artificial Intelligence and the Semantic Web, and serves as a member of academic committees including the Organizing Committee for the International Semantic Web Conference 2004 (ISWC-2004). He is a member of the IEEE, the Information Processing Society of Japan and the Japanese Society of Artificial Intelligence.


**Kaoru Hiramatsu**

Social Communication Laboratory, NTT Communication Science Laboratories.


He received the B.S. degree in Electrical Engineering and M.S. degree in Computer Science from Keio University in 1994 and 1996, respectively, and the Ph. D. in Informatics from Kyoto University in 2002. In 1996, he joined NTT Communication Science Laboratories and has been working on Digital Cities and the Semantic Web. From 2003, he is a Visiting Research Scientist at the Maryland Information and Network Dynamics Laboratory, University of Maryland. He is a member of the ACM, the Information Processing Society of Japan and the Japanese Society of Artificial Intelligence.


**Tetsuji Satoh**

Executive Manager, Social Communication Laboratory, NTT Communication Science Laboratories.


He received the B.E. degree in Electrical Engineering from Yamanashi University in 1980 and the PhD in Electrical Engineering from Osaka University in 1994. In 1980, he joined the Electrical Communication Laboratories, Nippon Telegraph and Telephone Public Corporation (now NTT). He has been conducting research in the fields of Database and the Semantic Web, and serves as a member of academic committees including the Information and System Society Steering Committee of the Institute of Electronics, and Communication Engineers (IEICE). From 2002, he is a Visiting Professor at Osaka University. He is a member of the IEICE and the Information Processing Society of Japan.