Information Management for Multimedia Earthquake Science Data

1. Research Team

Project Leader: Prof. Dennis McLeod, *Computer Science*

Other Faculty: Prof. Cyrus Shahabi, *Computer Science*

Graduate Students: Anne Yun-An Chen, Seokkyung Chung, Vesile Evrim, Shan Gao, Vijayakumar Gopalakrishnan, Jong-eun Jun, Leila Kaghazian, Hyun Woong Shin, Bomi Song, Sang-Soo Sung, Seongwook Youn

Industrial Partner(s): JPL

2. Statement of Project Goals

The primary goal of this project is to design, develop, and test methodologies for exploiting semantic aspects of the multimedia earthquake science data. The nature of data in the earthquake science domain is full of variety. That is, the interpretations of data differ from resource to resource and scientist to scientist. Therefore, our system must provide integration portability to manage the interoperability for heterogeneous data. Toward this end, we have developed an information system based on ontologies (collections of key concepts and terms along with their inter-relationships) to provide interoperability for heterogeneous data. The main capabilities of the system include modeling the meaning of the multimedia earthquake science data with mining the semantics, and providing an access to heterogeneous data sources with utilization of web services.

3. Project Role in Support of IMSC Strategic Plan

The representation of meta-data (information describing multimedia information units) is crucial for IMSC three Vision Projects, as well as several of the efforts in the application research projects. For example, with regard to the education project, we require the ability to represent (dynamic) the content of educational experience, and locate relevant information in response to students' inquiries. Thus, the understanding of the concepts and relationships must be involved, which can be supported by ontologies.

4. Discussion of Methodology Used

To support seismic hazard analysis, it is necessary to develop an information system, which manages a variety of types of earthquake science information (i.e., characterizes dynamic earthquake faults). The most critical issue in this project is how to represent and extract the semantic meaning from information contents. Toward this end, we choose an approach based on a domain-dependent ontology to express semantic knowledge. In particular, we seek dynamic ontology management since ontology must evolve with time as the concepts in that domain change.

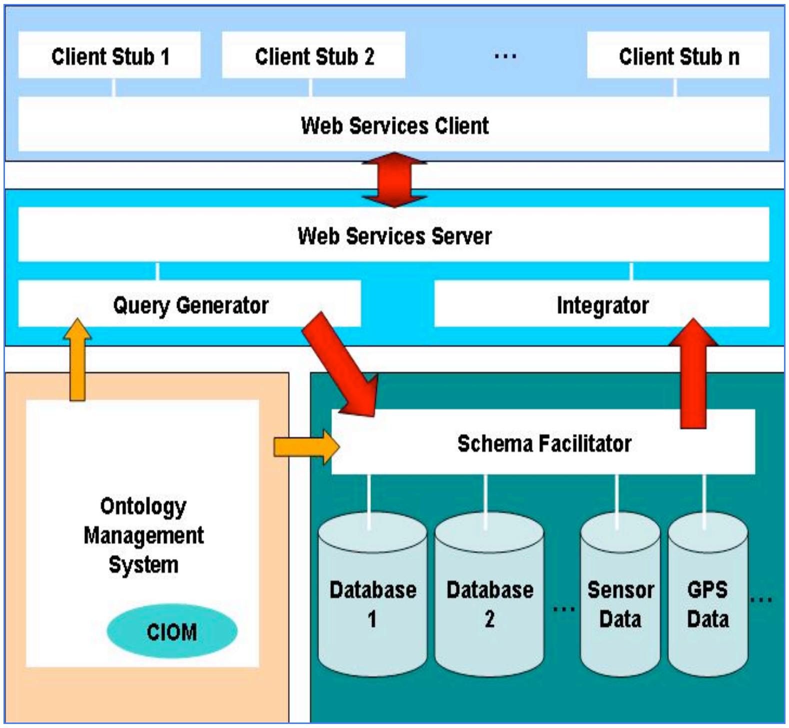

Figure 1. Framework of information management for multimedia earthquake science data

5. Short Description of Achievements in Previous Years

In the past, we focused on defining, designing, developing, deploying, and testing a data semantics based system in order to provide interoperability for heterogeneous data in the earthquake science domain [1]. In particular, we emphasize on the database management aspects of the work. These include modeling the meaning of the data, and providing Web service based access to heterogeneous data sources. The research outcomes will be used in simulations, model developments, and data mining.

Currently, we have developed the middleware for two groups of experts in order for them to manage the raw data, retrieve the data, and obtain the wrapped data for the further usage, for example, in simulation programs. A fault database has been established to handle fault parameters. An initial domain ontology is developed by both the computer scientists and earthquake science experts.

5a. Detail of Accomplishments During the Past Year

Our current status is as follows. We have developed an initial domain ontology and three different databases. One of the databases contains processed data of California faults and the sources are journal articles in the field of paleoseismology. The other contains data of California layers. Another contains the California Geological Survey data. We have also developed a simple XML-based distributed web service system for these three databases. At present, we employ a SOAP server on Tomcat. Several clients for our Web services have been developed using a WSDL interface. These clients are used to generate SOAP requests. Our users now can use the client stub to request the information and extract the results for literature references, data integration, and even graphical simulations with virtual reality tools.

The key issue in this project is how to define, develop, and test domain ontologies for earthquake science. Toward this end, we have developed an ontology management system, which is referred to as CIOM-plus. CIOM-plus is based on Classified Interrelated Object Model (CIOM) that is a subset of Semantic Data Model (SDM) [5]. CIOM-plus is equipped with the following five capabilities:

- Facilitate the domain ontology creation and update
- Associate the ontology/metadata with observational and hypothetical data
- Learn new concepts, relationships, and patterns among the metadata and data
- Support user (scientist) data and meta-data discovery/search
- Provide the base for the semantic wrapping of information sources

There are several advantages of using CIOM for the ontology representation. Since CIOM is an object-based model, it is easy to express concepts and their interrelationships in the ontology. That is, concepts in the ontology are mapped into classes and instances in CIOM; relations in the ontology are mapped into attributes, inheritances, memberships, and other descriptions of relations in CIOM. Since CIOM is a subset of a database model, storing and organizing ontologies become a simple task with database techniques.

CIOM-plus is in the middle layer, which implements all creation and maintenance operations. It retrieves different ontology schema (e.g., DB, XML, DAML/OIL), generates the class inheritance hierarchy, maintains instances and their attributes, and provides Java APIs to the Web-based user interface. The prototype is built using Java for its platform-independent feature.

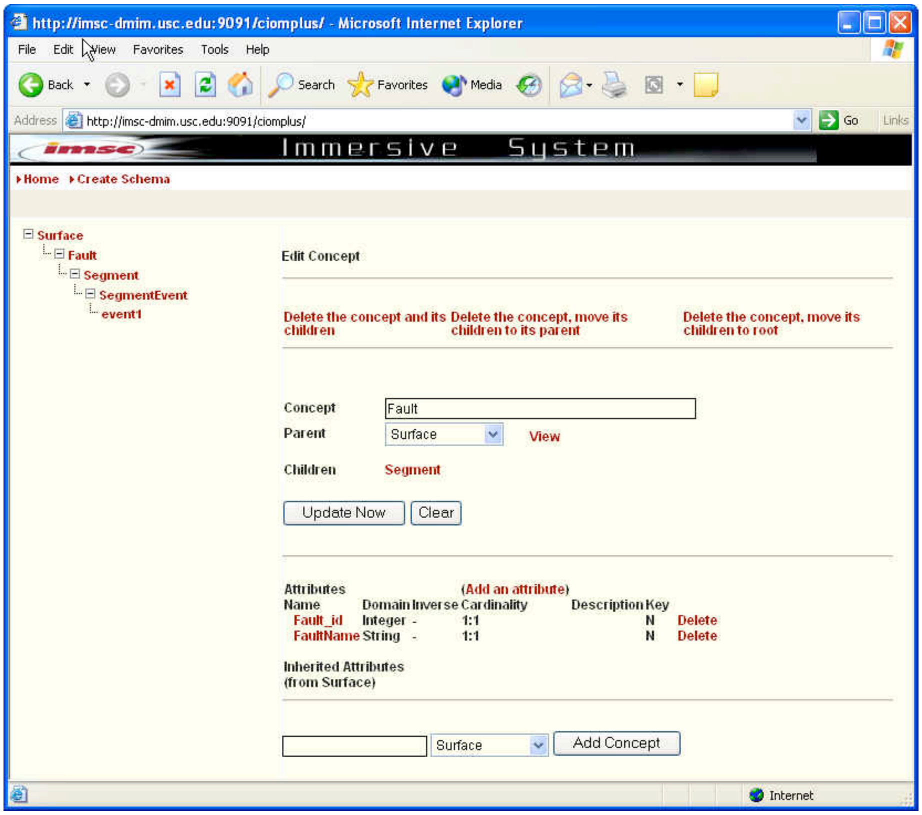

Figure 2. Sample screen-shot for CIOM-plus

6. Other Relevant Work Being Conducted and How this Project is Different

Recently, XML is proposed as a standard for representing and exchanging data (and meta-data). However, XML can only reflect syntactic property of data, and it has a limitation in representing semantic aspect of information. Therefore, we choose XML Schemas to format data for web services. XML Schema extends the properties of XML and includes more semantic definitions. XML Schema is machine-accessible and web-compatible. Furthermore, the Semantic Web research efforts, including current technologies such as RDF, DAML/OIL [3.4.6.7.14], are closely related with our research. In addition, our system has strong connections with the following specific areas: "metadata services", "federated database system", "data assimilation", "data mining", and "Web services". We will complement our work by exploring and implementing existing models in these fields.

With regard to uniqueness, CIOM-plus is unique ontology management system in that it exploits idea of CIOM, which is developed by Semantic Information Research Group at the University of Southern California. See http://imsc-dmim.usc.edu:9091/ciomplus/ for details.

7. Plan for the Next Year

Leading our plans for next year is a careful analysis of the use of our research outcomes in three Vision Projects: education, communication, and entertainment. Our initial main focus will be how to efficiently apply our ontology management system to Communication Vision Project. In addition, we plan to study the applicability of ontology framework to behavioral immersidata for Education Vision Project.

There are a number of ontology management systems that have been developed for the Semantic Web. Currently, we are in the process of evaluating several key systems such as Protégé [13]. Our goal is to highlight the strengths and weaknesses for each of them, and develop our own ontology management system (like CIOM-plus) to address the key limitations of other systems. The developed one is expected to have a graphical user interface for creating and editing ontologies, and read and write files in diverse ontology languages.

8. Expected Milestones and Deliverables

Spread across the four-year scope of this project, we will provide several deliverables including ontology management system, and dynamic ontology learning module. To accomplish the deliverables, each milestone represents the culmination of research and the realization of the project requirements.

Analysis of existing ontology management systems (Year 2):
There are a number of ontology management systems that have been developed for the Semantic Web [13.15.16]. For example, Protégé is one of the widely used systems in Semantic Web applications. However, there are some limitations in existing systems. For example, Protege does not allow more than one relation type and Web support. Thus, to account for necessary requirements and functionalities for our system, careful analysis and evaluation on existing ontology management systems must be conducted. Toward this end, we plan to investigate the diverse ontology systems, including Protégé, KAON, and OntoEdit. This experience will help us to develop a relevant ontology management system.

Migration to Oracle (Year 2):
To support the earthquake fault databases, we utilized MySQL, which is a state-of-the-art commercially-available general-purpose database management system. However, MySQL has a limitation in flexibility to add information dynamically. Thus, we plan to port our fault data to Oracle DBMS.

Development of dynamic ontology management systems (Year 2-3):
Although ontology-authoring tools have been developed in the past decades, manually constructing ontologies whenever new domains are encountered is an error-prone and time-consuming process. Thus, integration of knowledge acquisition with data mining, which is referred to as ontology learning, is necessary [12]. In our previous research efforts, we successfully applied topic mining to extract ontological information (e.g., metadata, topics, and concepts) from news streams [2], and exploited ontologies for intelligent news information selection [8.9.10.11]. Our current ontology management system involves manual ontologies editing. In our future work, coupling with topic mining, dynamic ontology learning module will be developed and connected to ontology management systems to learn new concepts as the new data is added to the system. As a result, we can reduce ontology maintenance efforts of knowledge engineers.

Provision of interoperability for heterogeneous data (Year 3-4):
To provide interoperability, we implement the query generator, schema facilitator, and the integrator to wrap the data to the format compatible with the web service architecture. When the server receives the request from any user stub, the query generator generates queries to heterogeneous data sources with the facilitation of the domain ontology. The schema facilitator incorporates the domain ontology in order to tune the individual database schemas or metadata. The query results are sent to the integrator for final wrapping with XML Schema statements, and the server is ready to response to the client. Users are able to access heterogeneous data with uncomplicated methods and the light-weighted protocol provided by the web services technology.

9. Member Company Benefits

We have a new direct connection with earthquake science research at JPL, which has strong requirements for the kinds of ontology and user modeling techniques we are developing. We now have substantial support from NASA on an application study directly connecting to the research (see http://www-aig.jpl.nasa.gov/public/dus/quakesim/index.html for the description of QuakeSim project).

10. References

[1] A.Y. Chen, A. Donnellan, D. McLeod, G. Fox, J. Parker, J. Rundle, L. Grant, M. Pierce, M. Gould, S. Chung, and S. Gao. Interoperability and semantics for heterogeneous earthquake science data. In *Proceedings of the International Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*, 2003.

[2] S. Chung, and D. McLeod. Dynamic topic mining from news stream data. In *Proceedings of International Conference on Ontologies, Databases, and Applications of Semantics*, 2003.

[3] S. Decker, D. Brickley, J. Saarela, and J. Angele. A query and inference service for RDF. In *Proceedings of the Query Language Workshop*, 1998.

[4] D. Fensel, F. van Harmelen, I. Horrocks, D. McGuiness, and P. Patel-Schneider. OIL*:* An ontology infrastructure for the Semantic Web, *IEEE Intelligent Systems*, 16(2): 38-45, 2001.

[5] M. Hammer, and D. McLeod. Database description with SDM: A semantic database model. *ACM Transactions on Database Systems*, 6(3):351-386, 1981.

[6] J. Heflin and J. Hendler. A portrait of the Semantic Web in action, intelligent systems. *IEEE Expert*, 16(2): 54 -59, 2001.

[7] Ian Horrocks, and Peter F. Patel-Schneider. Reviewing the design of DAML+OIL: an ontology language for the Semantic Web. In *Proceedings of the 18th National Conference on Artificial Intelligence*, 2002.

[8] L. Khan, and D. McLeod. Audio structuring and personalized retrieval using ontologies. In *Proceedings of IEEE Advances in Digital Libraries, 2000*.

[9] L. Khan, and D. McLeod. Effective retrieval of audio information from annotated text using ontologies In *Proceedings of ACM SIGKDD Workshop on Multimedia Data Mining*, 2000.

[10] L. Khan, and D. McLeod. Disambiguation of annotated text of audio using onologies. In *Proceeding of ACM SIGKDD Workshop on Text Mining*, 2000.

[11] L. Khan, D. McLeod, and E. Hovy. Retrieval effectiveness of an ontology-based model for information selection. *The VLDB Journal*, 13(1):71-85, 2004.
[12] A. Maedche, and S. Staab. Ontology learning for the Semantic Web. *IEEE Intelligent Systems*, 16(2), 2001.
[13] N.F. Noy, M. Sintek, S. Decker, M. Crubézy, R.W. Fergerson, and M.A. Musen. Creating Semantic Web contents with Protégé-2000. IEEE Intelligent Systems, 6(12):60-71.
[14] Peter F. Patel-Schneider, and Dieter Fensel. Layering the Semantic Web: problems and directions. In *Proceedings of the 1st International Semantic Web Conference,* 2002.
[15] Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, and D. Wenke. OntoEdit: Collaborative ontology development for the Semantic Web. In *Proceedings of International Semantic Web Conference*, 2002.
[16] R. Volz, D. Oberle, S. Staab, and B. Motik. KAON server - A Semantic Web management system. In *Proceedings of WWW*, 2003.

Acknowledgement This work was supported, in part, by NASA's Computational Technologies Project. Portions of this work were carried out by the Jet Propulsion Laboratory, California Institute of Technology under contract with NASA.

Resources and Collaborators: This research has been conducted by the Semantic Information Representation Group at the University of Southern California (see http://imsc-dmim.usc.edu for details). However, it is not an isolated effort at all. We have several key collaborators, mainly earthquake science specialists from various important perspectives. These include Dr. Andrea Donellan, Mr. Jay Parker, Robert Granat and their research group at JPL; Professor John Rundle and his research laboratory and group at University of California, Davis; Professor Lisa Grant and her research group at University of California, Irvine; Professor Geoffrey Fox, Dr. Marlon Pierce and their research group from Indiana University; Professor Terry Tullis and his research group at Brown University; and Mr. Philip Maechling and his group from the Southern California Earthquake Center. More information on our current collaboration can be found at http://www-aig.jpl.nasa.gov/public/dus/quakesim/people.html.