Mda-Based Modeling for Multi-Layered Subsurface Spatial Ontology

Sanghoon, Lee

Ubiquitous Land Implementation Research Department,
Korea Institute of Construction Technology,
411-712, Goyang-shi, Republic of Korea
E-mail: Sanghoon@kict.re.kr

Abstract - **Today, many heterogeneous data are being distributed via the Web. The heterogeneous data that different suppliers produce make it difficult for users to find and share the data they need. A spatial analysis of subsurface spatial data, that which is part of the National Spatial Data Infrastructure, needs various spatial data such as topographical maps, geological maps, and underground facility maps. The existing methods, however, which use a standard format or a spatial data-warehouse, cannot consider semantic heterogeneity. In this study, the layered ontology model, which consists of a generic concept, a measurement unit, a spatial model, and a subsurface spatial data, was developed. Besides, the current ontology building method designed by experts is an expensive and time-consuming process. The MDA-based meta-model of ontology was developed for easy understanding and flexibility of environmental change. The semantic quality of the developed ontology model was validated with the reasoner and semantic query for retrieving spatial distribution in a high liquefaction potential. Improved semantic sharing and strengthened capacities to develop a GIS experts system using the knowledge representation ability of ontology are expected.**

Keywords: ontology modeling, model-driven architecture, subsurface spatial data, GIS, semantic sharing.

I. INTRODUCTION

With the rapid development of the Internet, many heterogeneous data, including those in documents, images, videos, and maps, are being distributed via the Web. The heterogeneous data that different suppliers produce make it difficult for users to find and share the data they need. Especially in the GIS system, reuse and sharing are very difficult because the central government, municipalities, and entities collect spatial data according to their unique purposes and store large volumes of data in different systems and formats (Devogele, Parent, and Spaccapietra, 1998).

The existing methods for sharing heterogeneous data in the GIS domain are standards development and data warehouse construction. The methods involved in standards development include the development of interchange formats (e.g. SAIF and SDTS), the standardization of ISO/TC211, and the development of GML, WMS, and WFS by OGC(Open Geospatial Consortium). Examples of the spatial data warehouse construction include Korea's National Spatial Data System, NSDI (National Spatial Data Infrastructure, US), and CGDI (Canadian Geospatial Data Infrastructure). The spatial data warehouse classifies spatial data and offers relevant knowledge using metadata, which are defined as "the data of data".

The heterogeneity of data consists of three factors: syntax, structure, and semantics (Stuckenschmidt and Visser, 2000). The aforementioned methods that involve standards development and data warehouse construction can solve the problems of syntactic and structural heterogeneity, but cannot search for the semantic or innate meanings in compliance with the user's demand. In this study, the ontological methodology was used to overcome the semantic heterogeneity in the subsurface spatial data system, which is one of Korea's national spatial information systems. Ontology is a method of sharing semantics, which can extend the range of system applications by supporting free communication between users, applications, and computers, as well as information sharing. It enables the spatial information and GIS system that were developed for specific purposes to be applied to diverse scenarios.

There have been many studies on ontology development for sharing semantics in the GIS domain, but their applications have been limited. This was because it is difficult to develop an ontology that satisfies all users, since GIS is a system that integrates diverse domains. Therefore, a domain ontology was developed that suits subsurface spatial data while sharing the existing GIS ontology as its top-level concept.

In addition, the existing ontology development methodology depends on the manual work of experts (Nicola, Missikof, and Navigli, 2009). It requires much time and cost, and the heuristic methodology that depends on domain experts' experience and intuition may lack subjectivity. In this study, ontology was developed by applying an MDA (Model-driven Architecture) metamodel and using the UML (Unified Modeling Language) model, which is recognized as the standard methodology for software development. With these methods, an ontology that anyone can easily reuse can be designed. Besides, a highly reliable ontology can be reproduced in a short time and at a low cost, because MDA can be adjusted to changes in the environment, including to technology, situations, and infrastructure, simply by changing the metamodel.

II. LITERATURE REVIEW

A. Ontology and Development Methodology

Ontology originated from a domain of philosophy that studies the existence of materials and the relationships between them. It is applied to information systems. There are many definitions of ontology, one of which is that it is “a formal, explicit specification of a shared conceptualization of a domain of interest” (Gruber, 1992). Ontology is similar to the existing data modeling method in that it involves the conceptualization of the subject domain. Besides, ontology can express knowledge of a domain by describing terms and concepts (or meanings) in the domain.

The establishment of the ontology can be classified into the bottom-up and top-down methods. The bottom-up method is generally used. In this method, concepts are extracted from the glossary or from an interview with an expert to make the ontology. This method can supply detailed concepts, but the integrity and consistency of the ontology may be defective and its subjectivity may be insufficient if diverse experts cannot participate in the work (Teller, 2007). In the top-down method, the ontology is directly extracted from established databases or models. Its advantages are that an ontology with integrity can easily be created and new users can easily use it. Relevant systems or models must first be established, though.

In this study, the top-down method, wherein the ontology was derived from the subsurface spatial data system of the National Geotechnical Information Database Project, was used. Thus, the development of a heuristic ontology that could minimize the time and cost of repeated manual work and integrity checks became possible.

B. MDA-base Ontology Modeling

MDA, which complies with UML 2.0, is a development methodology that focuses more on model creation than on codes. It has high interoperability, reusability, and portability because it structurally separates the implementation and the specifications of the software. Of the MDA-based methodologies, the UML model offers a visual method in the object-oriented development procedure, including requirement analysis, system design, and implementation, along with the methodology for sharing the design results. The systems to which UML, which is recognized as the industry standard, is applied are considered reliable.

As shown in Figure 1, MDA has a meta-modeling structure with four layers, and consists of MOF (meta-object facility), UML, and XMI (XML metadata interchange). MDA is independent of the platform, as it conducts mutual conversion by mapping the modeling language, the implementation language, and the metamodel. It can effectively comply with large-volume applications and technical changes.

UML is an efficient quantitative concept modeling method and can be used as an ontology modeling language (Cranefield and Purvis, 1999). In a previous study, the UML model was converted into the OWL model using the XMI of ODM (the ontology definition model) (Duric, Gasevic, and Devdzic, 2003), and the UML model was converted into the DAML ontology model (Baclawski, et al., 2001). Using these methods, users can directly express the ontology because they use standard UML graphic tools, and the conversion is automatically conducted using the XMI exchange format and the interface that is defined in MOF. In addition, the changes in the specific object do not influence other objects, due to the modular characteristic of UML, which is an object-oriented model.

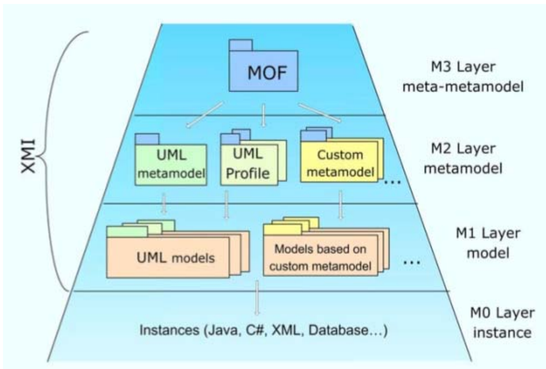

Figure 1. Structure of MDA (Miller and Mukerji, 2003).

C. GIS Ontology Development

The necessity of geospatial information system (GIS) ontology has been suggested for information exchange between heterogeneous GIS systems. Unlike in the other information system ontology, the spatial information must be differentiated in the GIS ontology because of its phase relationship and part-whole relationship.

The existing GIS ontology development efforts have had limited success, though. This is because it is very difficult to develop an ontology that satisfies all users, as GIS is a system that integrates diverse domains (Torres, et al., 2005). The GIS ontology that is being developed describes spatial information so that it can be shared in as many systems as possible. Therefore, it is desirable that it will be used as the top-level ontology. In a specific system, such as that in this study, an appropriate domain and task ontology must be developed.

D. Subsurface Spatial Data System in Korea

Korea's MLTM (the Ministry of Land, Transport, and Maritime Affairs) constructed a database for over 100,000 boreholes nationwide under its National Geotechnical Information Database Project and made it mandatory to register geotechnical survey data to expand the database (MLTM, 2008). The subsurface spatial data, including technical features such as the shapes and locations of layers and the strength required for the subsurface structure design, is inputted in the GIS-DB along with the digital map.

There is increasing demand for the use of geotechnical data for sharing with other GIS data including topographical maps, geological maps, and land use maps, or for its use in the design of foundations and subsurface structures in construction projects. To obtain the subsurface spatial data that a user requires, however, the analysis of GIS experts and geotechnical engineers, as well as understanding of the DB architecture and system, is needed. The additional data analysis by experts is needed because users cannot easily understand the semantics of the data. Accordingly, the development of the subsurface spatial data system ontology helps users search and share the data, as they can understand the semantics without the need for additional data analysis.

III. DESIGN OF THE SUBSURFACE SPATIAL ONTOLOGY LAYER

A. Configuration of the Ontology Layer

Ontologies have different levels of dependence according to specific tasks or views, and can be divided into the top-level, domain, task, and application ontologies (Guarino, 1997). In this study, the ontology was divided into layers for the mapping of the subsurface spatial data ontology and for the semantic sharing application. Especially, the ontology of the general concept, term dictionary, measurement unit, and spatial model, which are difficult to include in the subsurface spatial data domain, was organized in layers, as shown in Figure 2.

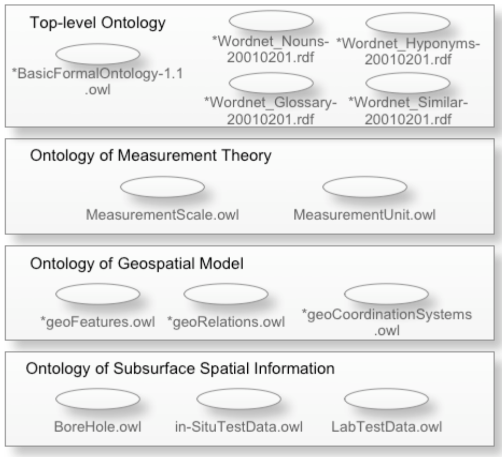

Figure 2. The layered ontology model.

Top-level ontology

In the top-level ontology, the concept and relationships of the ontology are classified based on human recognition. The basic classification is conducted via philosophy, epistemology, and psychology (Mika, et al, 2004). It describes the general concepts that can be identically treated in many domains, and aims to produce the widest range of semantic sharing. The top-level ontology that was developed includes Cyc, BFO (basic formal ontology), DOLCE, GFO, IDEAS, Wordnet, and SUMO. BFO, which can express geological phenomena such as earthquakes and landslides, and Wordnet, which can be used as a natural language dictionary, were employed in this study.

BFO consists of two sub-ontologies at different levels (SNAP and SPAN). SNAP describes the present objects in one, two, and three dimensions to understand them, and SPAN describes the processes that are generated at given time intervals. BFO can express stationary or dynamic, and spatial or time-based, objects at the same time by defining the mutual relationships between two sub-ontologies. Wordnet was introduced as a semantic network based on epistemology, and extended its scope to the dictionary. As a top-level ontology that includes most general concepts, it is widely used to study natural languages with diverse semantic relationships, including "isKindOf (hypernym/hyponym)" and "isPartOf (holonym/meronym)".

Measurement ontology

Subsurface data describes survey and test results, including boreholes. These data are measured according to their name, order, interval, and ratio (Chrisman, 1995). Nominal and ordinal units such as soil classification and USCS(Unified Soil Classification System) codes cannot be mathematically calculated due to their qualitative properties, but the interval and the ratio, such as the depth and the N value, are quantitative properties that can be calculated. The ratio can be divided or multiplied, but the interval can only be added or subtracted. The quantitative properties can be quantitatively described using measurement units (e.g., g/cm^2), and the measurement units that were required for the subsurface spatial information model were expressed in “MeasurementUnit.owl”.

Thus, measurement units describe the semantics of quality for measured values. BFO, which was selected as the top-level ontology, does not quantitatively and qualitatively classify the measurement units. Therefore, “MeasurementScale.owl” was written as the extension of the concept of “quality” in BFO by adding the measurable quantity and the measurable quality.

Spatial ontology

To organize the spatial ontology, the phase relationship, distance, direction, and whole-part relationship, as well as the object definition, must be described. First, OGM (the open geodata model), which is the spatial reference system, and GML were applied to the object definition. GML expresses the spatial object using XML and includes the object’s geometry, properties, and reference system. The spatial relationship offers details of the objects. Well-known relationships include the feature type (e.g., face, line, point, location, and dimension) and the phase relationship (e.g., disjoint, equal, contains, overlaps, and touches). In this study, the spatial model ontology was based on GeoOntologies (Geoontologies, 2004), which was made with the GML-based OWL ontology. GeoOntologies consists of spatial descriptors (“geoCoordinateSystems.owl”) such as points and multi-polygons, spatial features (geoFeatures.owl”) such as cities and buildings, and relationships between spatial descriptors (“geoRealtions.owl”).

B. Structure of the Subsurface Spatial Data

The most frequently used subsurface spatial data in construction and environmental surveys is the result of drilling test, followed by the indoor test and the in-situ test. The geological survey, physical survey, and field repair test are also important (PEER, 2004).

The subsurface spatial data in this study consisted of the hole-related data in Figure 3. The site indicates the drilling locations on the aerial photographs or topographic maps of the area for the hole or sample collection. The hole is the result of the boring for the subsurface space survey, and the layer describes the textures and physical properties of materials at specific depth intervals in the soil. The component is a physical form that is observed at a specific depth or scope in the soil. It expresses physical, chemical, biological, and mineral characteristics and geological behavior with the lapse of time, and exists in a layer or includes several layers. The core is an sample with a specific scope, which is extracted from holes, and the specimen is a sample that is separated from the core for the purposes of description and indoor testing of the subject point. The water content, grading, and liquid/plastic limit are measured from the specimens as basic soil properties. Besides, the subsurface spatial data is expressed via the in-situ test and the indoor test.

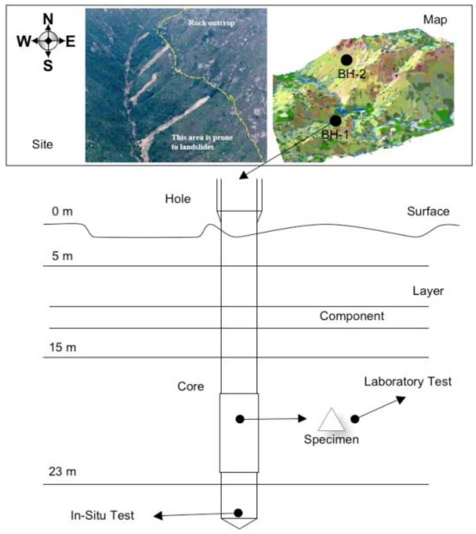

Figure 3. Composition of the hole-related data.

IV. MDA-BASED OWL MODEL CONSTRUCTION

To construct the MDA-based ontology, the class diagram of the UML must be expanded to the metamodel, the UML profile. In this study, a stereotype was used to expand the meaning of the UML and to create new components. Via the stereotype, the UML meta-class was defined as a virtual subclass, and additional semantics, including the class, association, and dependence, were assigned. As the shown in Figure 4, the UML profile model was converted into XMI documents, which were in turn converted into the OWL model using XSLT (eXtensible Stylesheet Language Transformations) as an XML parser. Poseidon for UML v4.0, which supports UML XMI 1.4, was used as the UML modeling tool. Xalan Java v2.7.1 was used as an XSLT processor to convert the UML model into an XMI document. In addition, Protege v3.4 was used to edit the implemented ontology model.

The package of the class diagram was displayed as the stereotype “<<ontology>>”, and each class was displayed as “<<OntClass>>”. Besides, classes in the ontology, including Enumeration, Union, Intersection, Complement, Restriction, and AllDifferent, were defined as stereotypes. Simple data-type class attributes were displayed as “<<DatatypeProperty>>”, and the attributes that were related to complex data types or relationships were displayed as “<<ObjectProperty>>”. According to such rules, a UML profile model was constructed using Poseidon for UML, as shown in Figure 7. The UML model was converted using the Xalan XSLT processor, and an OWL-DL model was finally developed with the previously designed ontology layer. Figure 8 shows the class of the OWL model that was constructed using Protege-OntoViz.

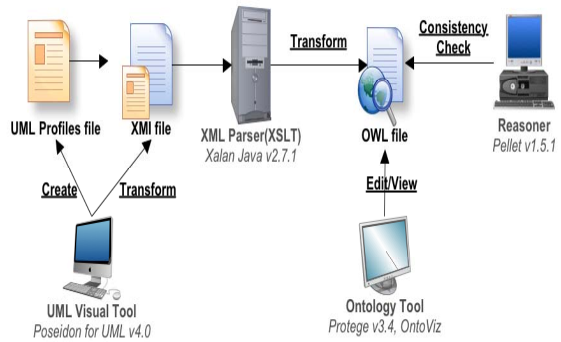

Figure 4. Composition of the hole-related data

V. VALIDATION AND DISCUSSION

A. Ontology Validation via the Reasoner

The quality of the ontology, which has multi-dimensional aspects, can be divided into syntactic, semantic, pragmatic, and social qualities (Burton-Jones, et al, 2005). The syntactic quality is evaluated using the formal syntax; and the semantic quality, using the existence of conflicting concepts. The pragmatic quality is evaluated based on the usefulness of the ontology for users; and the social quality, based on the universality of the connection of the ontology to other ontologies.

In this study, the semantic quality was first considered. The syntactic validity was already verified during the conversion of the UML profile into the OWL model using XMI. The social and pragmatic qualities were achieved through the universality of the top-level ontology in the ontology layer. To evaluate the semantic quality, integrity validation was conducted via the reasoner. In this study, Pellet v1.5.1, an OWL reasoner provided by Mindswap, was used to validate that the concept of the subsurface spatial ontology model and the class hierarchy both have integrity without any error, as shown in Figure 5 (Consistent: Yes). After ontology consistency was checked, individuals of ontology model were defined by the existing subsurface database.

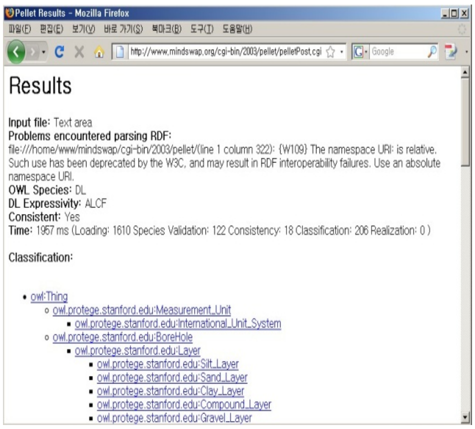

Figure 5. The ontology verification using Pellet.

B. Ontology Validation via Semantic Query

It is neccssary for semantic query to integrate with various spatial data sources for corresponding with user's semantic. For example, there is sample query , "Which area within 1km of Meyoung-Gi Boulevard is to be expected in the high liquefaction potential ?". This query includes various concepts to optimize spatial decision making in subsurface data. These concepts have the location of "Meyoung-Gi Boulevard" indicated by user, the definition of "within" intended by user, the earhquake engineering definition of "liquefaction potential", and so on. This query was sucessfully conducted using SPARQL query engine. Figure 6 depicts UML sequence diagram for semantic query procedure. From step 1 to step 13, user can obtain semi-results to interpret first query through developed ontology model and GIS database, and choose second query suitable to user's semantics. From step 14 to step 25, user can discover final results.

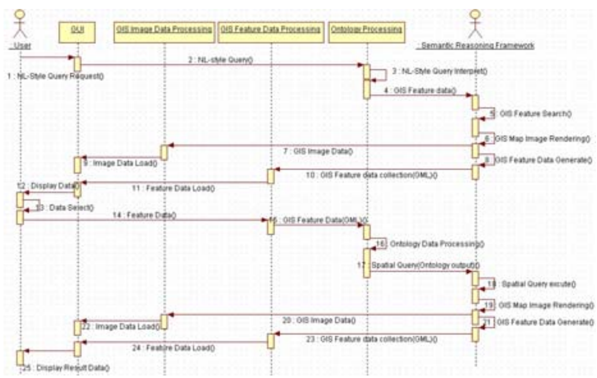

Figure 6. Sequence diagram for semantic query procedure.

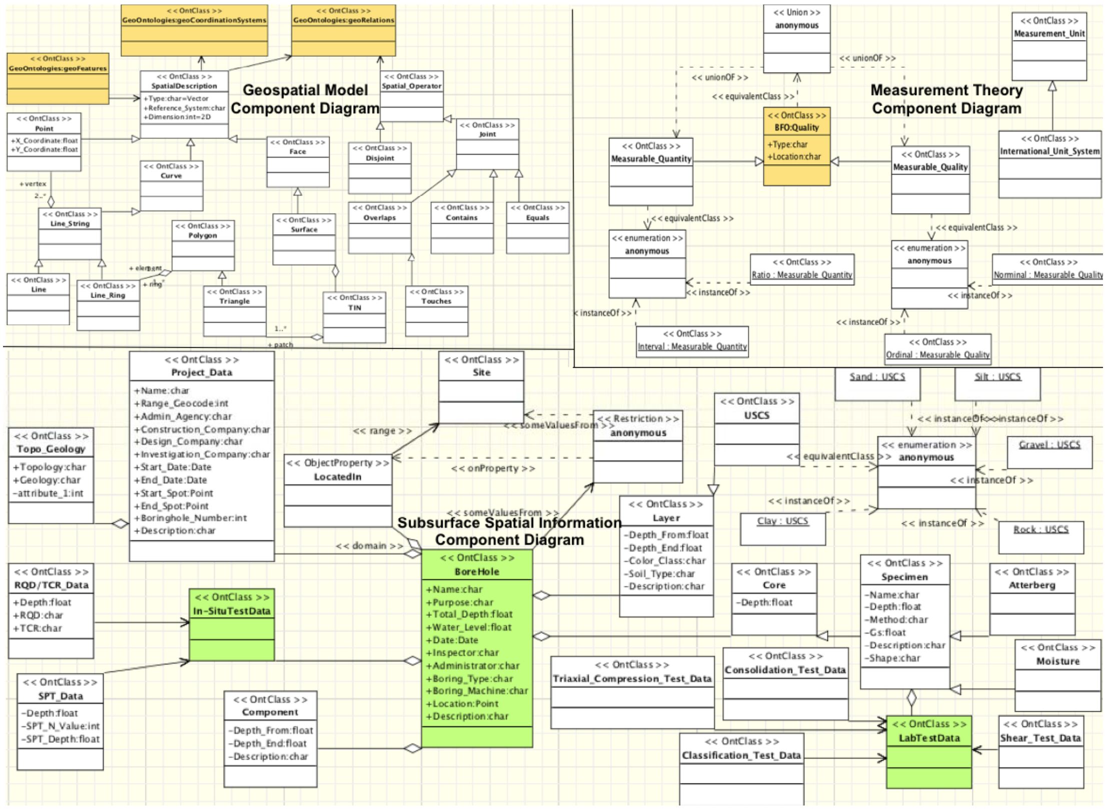

Figure 7. An excerpt of the UML profile model for subsurface spatial layer using Poseidon-UML.

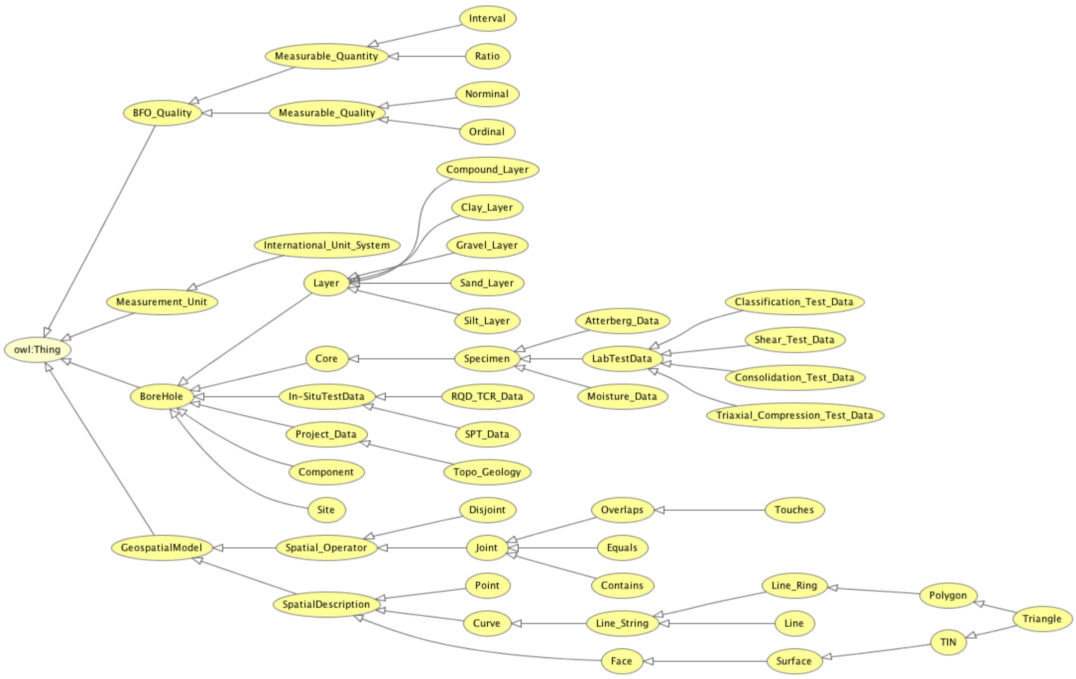

Figure 8. An excerpt of the OWL model(class) for subsurface spatial layer using Protégé-OWLViz.

In the case of engineering judgment, definition of "high liquefaction (if silt or silt-sand, N values <= 20 and under the ground water level) (Seed, et al., 2003)" described in subsurface ontology model. Here, N value is quantitative data that are represented as numerical values in SPT (Standard Penetration Test). The purpose of SPT is to measure the relative density of sand or gravel. It is used to estimate the strength parameter as a good guide for the determination of subsurface conditions, especially liquefaction. These restriction properties in liquefaction condition are described in subsurface ontology, as shown in Figure 9.

bore:LiquefactionGround =
(Sand and SPT_N some integer(<="20"integer])
and DEPTH some integer(<=GWL integer])
or (Silt and SPT_N some integer(<="20"integer])
and DEPTH some integer(<=GWL integer])
or (LooseSand or VeryLooseSand)
bore:LiquefactionGround = bore:hasUSCS and
bore:hasSPT

Figure 9. An excerpt of restriction tag of "LiquefactionGround" in the subsurface ontology.

C. Ontology Application for Semantic Sharing

The semantics of spatial information can be divided into concepts and relationships. The semantic concept involves identical symbols for different concepts. In this study, the semantic concept was examined via the general concept and the terminology dictionary of the top-level ontology, the measurement unit of the measured ontology, and the subsurface spatial data ontology.

To overcome this semantic heterogeneity in the existing information sharing system, experts' analysis was essential. The ontology layer model that was developed in this study facilitates GIS sharing, considering semantics, as shown in Figure 10, as the basis for making automatic sharing available through reasoning. Besides, general users easily make and update ontology for the specific domains without direct help of the diverse experts, via the developed ontology development methodology (e.g. multi-layered ontology model, UML profile, and xslt).

The ontology enables free communication between users and computers, and is used to express or store relevant knowledge with which to expand the scope of the system application. The results of this study can be used to establish a GIS expert system for the application of the GIS system, which has so far been applied only to specific tasks, to diverse scenarios.

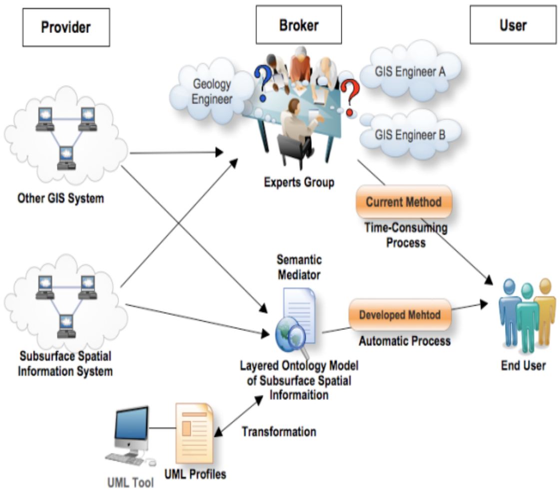

Figure 10. The semantic sharing application using the ontology.

VI. CONCLUSION AND FUTURE STUDIES

The present GIS system is basically heterogeneous, as it manages entities to produce spatial information according to diverse purposes. The syntactic and structural heterogeneity of the information can be ensured through format conversion or standardization, but it is difficult to address the semantic heterogeneity.

In this study, the ontological methodology was used to share the semantics in the subsurface spatial data. The developed ontology model consists of the top-level, measurement, space, and subsurface spatial data ontologies in the form of layers. Thus, a basis was established on which the subsurface spatial data system can share data by considering semantics. In addition, a standard software development methodology, MDA, was used, instead of the heuristic ontology development methodology that depends on domain experts' experience and intuition. MDA provides an ontology development environment using the metamodel, which anyone can easily use. UML has intuitive and diverse visualization tools, and easily reproduces the ontology merely by changing the metamodel in the changing environment. Therefore, the subsurface spatial ontology can easily be modified or supplemented by merely modifying the UML profile model.

In the future, the study will be expanded to OWL-S (web ontology language for services) based on the subsurface spatial ontology, the OWL model, to establish the semantic GIS by which semantic information sharing and spatial queries could be made available in the spatial web service.

REFERENCES

[1] T. Devogele, C. Parent, and S. Spaccapietra. “On Spatial Database Integration.” Intl. Journal of Geographic Information Science, 1998, pp. 335-352.

[2] H. Stuckenschmidt, and U. Visser. “Semantic translation based on approximate re-classification.” in Proc. 7th Intl. Conf. on Principles of Knowledge Representation and Resoning, 2000.

[3] A. D. Nicola, M. Missikof, and R. Navigli. “A Software engineering approach to ontology building.” Information Systems, 2009, pp. 258-275.

[4] T. Gruber. A Translation Approach to Portable Ontology Specification. Knowledge Systems Laboratory, Stanford University, Stanford, CA: Technical Report KSL, 1992, pp. 92-71.

[5] J. Teller. Ontologies for an Improved Communication in Urban Development Projects-Ontologies for Urban Development, Springer, 2007, pp. 1-14.

[6] S. Cranefield and M. Purvis. “UML as an Ontology Modeling Language.” in Proc. of the Workshop on Intelligent Information Integration, 1999.

[7] D. Duric, D. Gasevic, and V. Devdzic. “A MDA-based Approach to the Ontology Definition Metamodel.” In Proc. 6th Int. Conf. on Information Technology, 2003, pp. 193-196.

[8] K. Baclawski et al. “Extending UML to Support Ontology Engineering for the Semantic Web.” in Proc. 4th Intl. Conf. on UML. Toronto, 2001.

[9] J. Miller and J. Mukerji. “MDA Guide Version 1.0.”, OMG, 2003.

[10] M. Torres, et al. “Ontology-Driven Description of Spatial Data for Their Semantic.” in Proc. 5th Intl. Conference GeoS, 2005, pp. 242–249.

[11] MLTM. “the 2008 National Geotechnical Information Database Project.” 2008, pp. 7-10.

[12] N. Guarino. Semantic matching: formal ontological distinctions for information organization, extraction, and integration. Information Extraction, A Multidisciplinary Approach to an Emerging Information Technology. Springer Verlag., 1997, pp. 139–170.

[13] P. Mika, et al. “Foundations for Service Ontologies: Aligning OWL-S to DOLCE.” in Proc. WWW2004, New York, 2004.

[14] N.R. Chrisman. “Beyond Stevens: A Revised Approach to Measurement for Geographic Information.” in Proc. AUTO-CARTO 12 Charlotte 4, 1995, pp. 271–280.

[15] Geoontologies, 2004, Available: http://www.mindswap.org/2004/geo/geoOntologies.shtml [Nov. 1, 2009].

[16] PEER. Archiving and Web Dissemination of Geotechnical Data: Development of a Pilot Geotechnical Virtual Data Center, Consortium of Organizations for Strong-Motion Observation Systems, 2004.

[17] Burton-Jones, et al. “A semiotic metrics suite for assessing the quality of ontologies.” Data Knowledge Engineering, 2005.

[18] Seed R.B., et al. “Recent advances in soil liquefaction engineering : a unified and consistent framework.” in Proc. 26th Annual ASCE Los Angeles Geotechnical Spring Seminar, 2003, pp. 2-7.

ACKNOWLEDGEMENT

This research was supported by the grant (09CCTI-A052531-02-000000) from the Ministry of Land, Transport and Maritime of Korean government through the Core Research Institute at Seoul National University for Core Engineering Technology Development of Super Long Span Bridge R&D Centre.

AUTHOR BIOGRAPHY

SANGHOON, LEE is Team-leader of Ubiquitous Land Implementation Research Department at Korea Institute of Construction Technology sponsored by government. He obtained his M.Sc. and Ph.D. in Geographic Information & Geomatics from the Seoul National University in 2002 and 2010, respectively. He worked as a GIS consultant for supporting government policy making and technical planning in the past 10 years. He also has participated in many national and commercial research projects. His current research interests include spatial decision-making, spatial data mining, ubiquitous-based GIS, subsurface modeling, semantic sharing and reasoning.