

ONTOLOGY-BASED SEMANTIC DESCRIPTION MODEL FOR DISCOVERY AND RETRIEVAL OF GEO-SPATIAL INFORMATION

Qin Zhan^a, Xia Zhang^b, Deren Li^c

^aSchool of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
- qin.zhan@gmail.com

^bPrinting and Packaging School, Wuhan University, 129 Luoyu Road, Wuhan 430079, China - zxx@whu.edu.cn

^cState Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China – dli@wtusm.edu.cn

KEY WORDS: GIS, Integration, Knowledge Base, Retrieval, Representation, Spatial Infrastructures, Model, Method

ABSTRACT:

Finding and accessing suitable geographic information to satisfy various applications in the open and distributed environments of Spatial Data Infrastructures (SDIs) is a crucial task. However, because of the semantic heterogeneity in conventional exploitations and descriptions of geographic information, it is difficult to find suitable geographic information which exactly meets the requirements of application. To solve the problems caused by semantic heterogeneity, this paper presents an ontology-based semantic description model which explicitly represents geographic information semantics in abstract and concrete level. It is an integrated model consisting of three parts: Data Profile, Data Content, and Data Binding. Data Profile tells users what the data are about in abstract level. Data Content describes what the data contain in concrete level. Data Binding tells users where and how to access the data. Moreover, this paper puts forward an ontology based approach to enhance the efficiency of discovery and retrieval of geographic information.

1. INTRODUCTION

Geographic information (GI) is more and more important in various application domains such as planning and decision-making etc. Finding and accessing suitable geographic information to satisfy these applications in the open and distributed environments of Spatial Data Infrastructures (SDIs) is a crucial task. However, it is difficult to find suitable geographic information which exactly meets the requirements of application because of the semantic heterogeneity in conventional exploitations and descriptions of geographic information. Much research work has been done for discovery and retrieval of geographic information by OpenGIS Consortium (OGC). The specifications provided by OGC enable the syntactic interoperability and cataloguing of geographic information. Though the catalogs support discovering, organization, and access of geographic information, they do not yet provide methods to solve problems of semantic heterogeneity (Bernard et al., 2004; Klien et al., 2004). Problems of semantic heterogeneity are caused by synonyms and homonyms in metadata and user's query (Klien et al., 2004). This is because various metadata specifications and various vocabularies are used in metadata and user's queries. In various specifications and vocabularies, different terms may refer to similar concepts, and the same terms may refer to different concepts. So overcoming problems of semantic heterogeneity is the key to enhance efficiency of discovering and retrieval of geographic information.

One possible approach to overcome the problems of semantic heterogeneity is the explication of knowledge by means of ontology, which can be used for the identification and association of semantically corresponding concepts because ontology can explicitly and formally represent concepts and relationships between concepts and can support semantic reasoning according to axioms in it. Ontology has been

developed in the context of Artificial Intelligent (AI) to facilitate knowledge sharing and reuse. It covers many fields such as knowledge engineering, information integration, information retrieval and so on. The reason for ontology being so popular is due to what it promises: a shared and common understanding of some domain that can be communicated between people and application systems. Many definitions of ontology have been proposed. A popular definition of them is: an ontology is a formal, explicit specification of a shared conceptualization (Gruber, 1995). The meanings of the definition is explained (Studer et al., 1998): A "conceptualization" refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon; "Explicit" means that the type of concepts used, and the constraints on their use are explicitly defined; "Formal" refers to the fact that the ontology should be machine readable; "Shared" reflects the notion that an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group.

In this paper, an ontology-based semantic description model is put forward to explicitly represent geographic information semantics in abstract level and concrete level by introducing ontologies. It is an integrated model consisting of three parts: Data Profile, Data Content, and Data Binding. According to the model, descriptions of geographic information can be readable and understandable for computers. Moreover, based on the proposed model, an ontology-based approach is discussed to enhance efficiency of discovering and retrieval of geographic information.

This paper is organized as follows. In section 2, we make a survey on the related works. Then, the ontology-based semantic description model is presented in section 3. In the subsequent section, the ontology-based approach for

discovering and retrieval is made out. Finally, we make conclusions and discuss the future work need to do.

2. RELATED WORKS

Ontology based research for discovering and retrieval of geographic information has been booming. There are many research works done to make efforts to it.

Klien et al (2004) presented an architecture for ontology-based discovery and retrieval of geographic information that can contribute to solving existing problems of semantic heterogeneity. So far they have defined components such as Enhanced Cascading Catalog Service and the Reasoner component and plan to develop a query scenario in which the user is able to formulate a question using terms from the familiar shared vocabularies.

The research project SPIRIT (Spatially-Aware Information Retrieval on the Internet) developed tools and techniques to support spatial search on the internet based on ontology (Jones et al., 2002). Geographical ontologies are constructed to assist spatial search (Fu, 2005a). An ontology-based spatial query expansion method is developed that supports retrieval of information relevant to space by trying to derive its geographical query footprint (Fu, 2005b).

Hartwig H. Hochmair (2005) proposed a conceptual framework to overcome problems of semantic heterogeneity in keyword based retrieval of geographic information. In the architecture, the server-sided knowledge base including domain ontology and rules for query expansion is used to expand the keyword-based searches.

Wiegand and Garcia (2007) developed a task-based ontology approach to automate geospatial data retrieval. In the approach, ontologies of task, data source, metadata and place, along with relationships between them are developed. With the ontology, reasoning can be done to infer various types of information including which data sources meet specific criteria for use in particular tasks.

In this paper, we focus on uniform semantic descriptions of geographic information and approaches for discovery and retrieval based on semantic descriptions.

3. ONTOLOGY-BASED SEMANTIC DESCRIPTION MODEL

In this section, we set forth an ontology-based semantic description model that explicitly represents semantics of geographic information in uniform machine readable and understandable format. Thus, problems of semantic heterogeneity in description of geographic information are solved.

3.1 Ontologies

Ontologies are the key to semantic description of geographic information. In this paper, several ontologies are constructed to explicitly model knowledge for geographic information such as metadata ontology, fundamental geographic information ontology, spatial relation ontology, geometric ontology. They

all belong to geographic information domain ontology and model non-spatial semantics and spatial semantics of geographic information.

Metadata ontology explicitly represents the knowledge of geographic information metadata. It not only formalizes existing metadata specifications such as ISO-19115 (ISO TC/211, 2003) and CSDGM (FGDC, 1998) in Ontology Web Language (OWL) (W3C, 2004), but also builds mappings between them. Therefore, it makes various metadata according to different specifications be able to interoperation in semantic level.

Fundamental geographic information ontology is created to describe non-spatial semantics of geographic features. It defines fundamental geographic feature concepts and relationships of them.

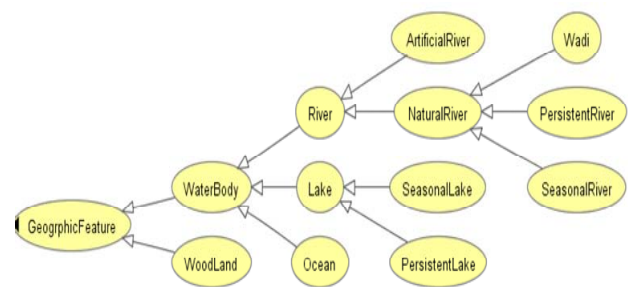


Figure 1. A light ontology for example

Spatial relation ontology and geometric ontology are built to describe spatial semantics of geographic features. Spatial relation ontology defines topological relations, direction relations and distance relations. Geometric ontology defines the concepts and relations of geometric point, line, polygon, surface, etc.

These ontologies are built with a “down-up” abstracting method. First, key concepts in geographic information science domain are abstracted. Then, concepts relationships, geometric relationships, spatial relationships, location relationships are modeled and established. Concepts relationships include “subclass of”, “part-whole”, “member of”, “instance of” and “dependency”, “reference”. Geometric relationships define the composite relationship among point, line, polygon and surface. Spatial relationships define topological relations, direction relations and distance relations. Finally, OWL DL (W3C, 2004) is adopted to describe these ontologies and test their consistency by semantic reasoning engine RACER (Racer Systems GmbH & Co. KG, 2005).

3.2 Ontology-based semantic description model

Metadata is an important part in SDI to describe geographic data for data discovering and sharing. It facilitates data sharing among Geographic Information Communities (GIC) in distributed environments. However, metadata is not enough to solve problems of semantic heterogeneity because different GICs publish metadata according various metadata specifications and various vocabularies. For solving problems of semantic heterogeneity, we proposed an ontology-based semantic description model to represent semantics of geographic data.

The semantic description model is built in abstract level and concrete level, which is showed in Figure 2 in Unified Modeling Language (UML). It is an integrated model consisting of three parts: Data Profile, Data Content and Data Binding. **Data Profile** tells users what the data are about by describing the profile of the geographic information, such as the dataset name, data format, acquisition methods, involved processing, spatial extent, spatial resolution, provider, theme and so on. Data Profile belongs to abstract level description. It is built by introducing metadata ontology. **Data Content** tells users what the data contain by describing the content semantics

of geographic information data in concrete level. Conventionally, geometries of spatial data are explicitly represented. But spatial relationships are not explicitly or formally represented. In this paper, geometric ontology and spatial relation ontology are introduced into Data Content to explicitly describe spatial semantics of geographic data. In addition, fundamental geographic information ontology is introduced to describe non-spatial data semantics. **Data Binding** tells users where and how to access the data by providing URL, contact information and procedure to access data.

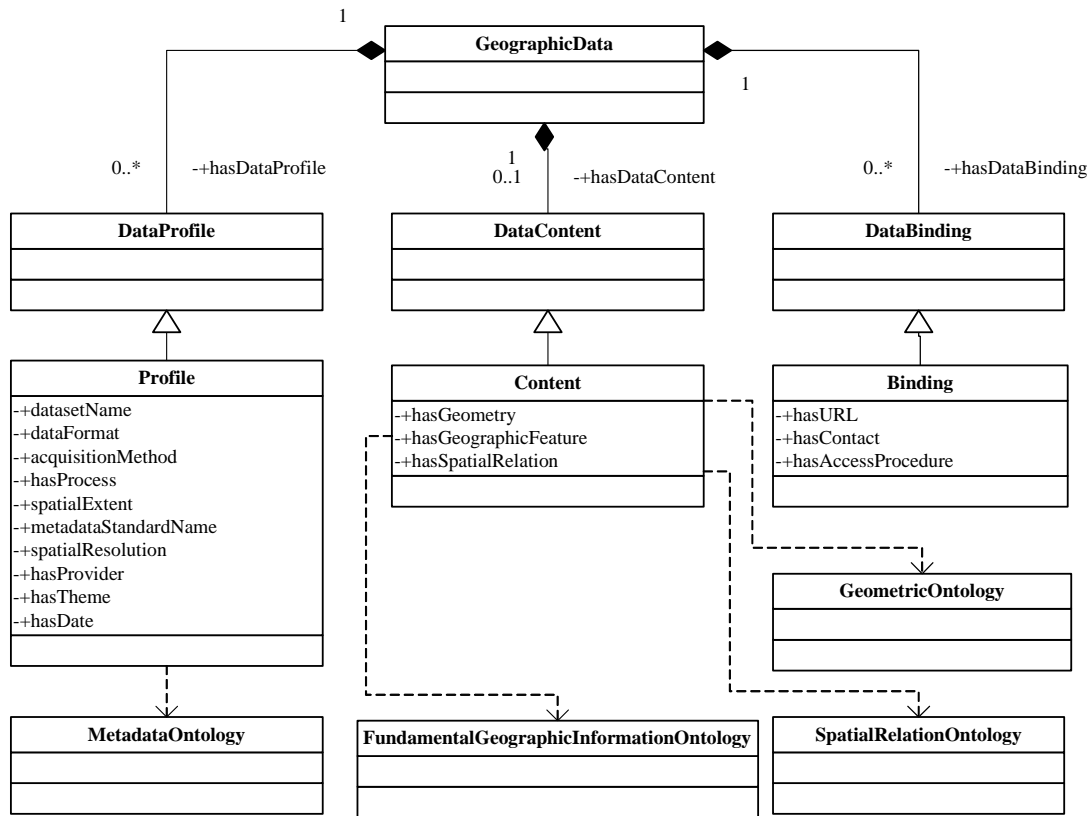


Figure 2. Ontology-based semantic description model

4. DISCOVERY AND RETRIEVAL BASED ON ONTOLOGIES

Conventionally, discovery and retrieval for geographic information is carried through based on keywords. However, keywords are not sufficient to find exactly suitable geographic information because lack of semantics and inference mechanism usually causes search results often to be too narrow or too large (Hartwig H. Hochmair, 2005). The emergence of ontology provides possibility to enhance discovery and retrieval. In this section, we discuss an ontology-based approach for discovery and retrieval of geographic information.

4.1 Conceptual framework

The key to improve discovery and retrieval is to solve problems of semantic heterogeneity between user's search and description of geographic information in SDI. Thus, user ontology is needed to cooperate with ontologied introduced into the semantic description model proposed above. User

ontology models concepts and relationships between concepts in user's view during search.

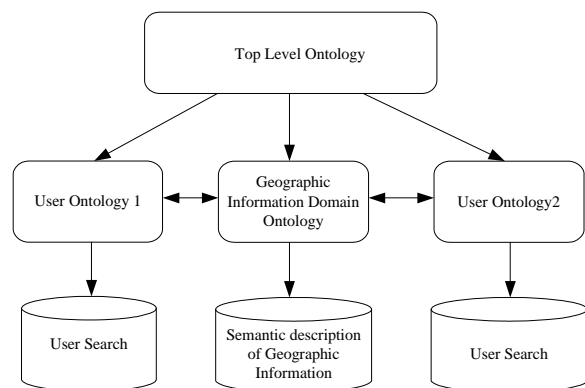


Figure 3. The hybrid ontology approach (Visser, U. & H. Stuckenschmidt, 2002) (modified)

The hybrid ontology approach (Visser, U. & H. Stuckenschmidt, 2002) is used, as shown in Figure 3, in our conceptual framework to make user ontology and domain ontologies used in description of geographic information. Top Level Ontology is a general ontology shared by user and description provider. It's a medium component to communicate between user ontology and domain ontologies in the proposed model above.

4.2 Ontology-based method

Based on the above hybrid ontology approach, a method is developed for discovery and retrieval for geographic information, as showed in Figure 4.

There are three main steps in the method. First step, user's search terms are mapped into concepts in domain ontology based on the hybrid ontology approach. Second step, the concepts are expanded based on the hierarchy of concepts in domain ontology. If the concepts have subclasses, then the subclasses are added to expand the query. Otherwise, the superclass is checked. If there is superclass, then the superclass is added to expand the query. Third step, based on the query expansion, suitable descriptions of geographic information are searched and return to users. If the results are suitable, search is finished. Otherwise, repetition from the second step is needed.

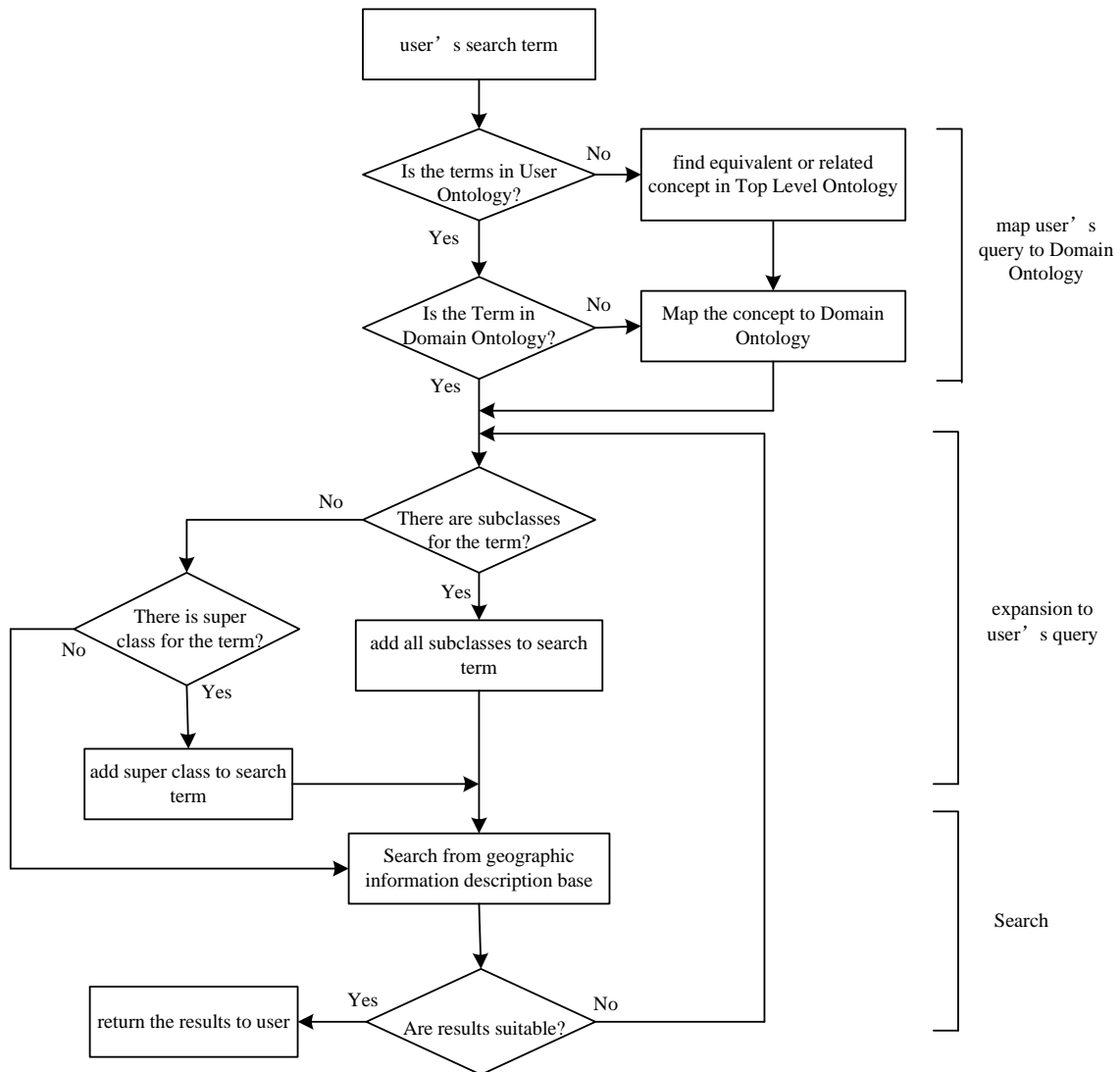


Figure 4. Ontology-based method for discovery and retrieval

5. CONCLUSION

In this paper, we proposed an ontology-based semantic description model for geographic information. The model is able to solve problems of semantic heterogeneity in description by introducing ontologies which enrich semantics of geographic information. Besides, an ontology-based method for discovery and retrieval is discussed to overcome shortage

of approaches based on keywords. The method make user's query more exactly in expression meanings of request. Thus, suitable geographic information are discovered and retrieved.

Of course, there are still some works need to do in future. Sufficient algorithms for building mappings between ontologies still need to present. Besides, implementation for

ontology-based discovery and retrieval should be developed too.

REFERENCE

- Bernard, L., U. Einspanier, S. Haubrock, S. Hubner, E. Klien, W. Kuhn, R. Lessing, M. Lutz, and U. Visser, 2004. Geotechnologien Science Report "Ontology-Based Discovery and Retrieval of Geographic Information in Spatial Data Infrastructures". Report No. 4. http://www.delphi-imm.de/meanings/index_eng.html.
- FGDC, 1998. FGDC-STD-001-1998, Content standard for digital geospatial metadata, Washington, D.C.
- Fu, F. Jones, C.B. Abdelmoty, A.I. 2005a. Building a Geographical Ontology for Intelligent Spatial Search on the Web. In: *Proceedings of IASTED International Conference on Databases and Applications (DBA2005)*.
- Fu, F. Jones, C.B. Abdelmoty, A.I. , 2005b. Ontology-based Spatial Query Expansion in Information Retrieval. *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE*. Springer. pp. 1466-1482.
- Gruber T. R, 1995. Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, Academic Press, Volume43, Issues 5-6, pp. 907-928.
- Hochmair H, 2005. Ontology matching for spatial data retrieval from internet portals. In Rodriguez M A, Cruz I F, Egenhofer M J, and Levashkin S (eds): *Proceedings of the First International Geospatial Semantics Conference*. Berlin, Springer Lecture Notes in Computer Science No 3799: 166-82.
- ISO TC 211, 2003. ISO 19115: Geographic information – Metadata.
- Jones, C.B. et al, 2002. Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT project. In: *SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Tampere, Finland, ACM Press, pp.387 - 388.
- Klien, E., Lutz, M., Einspanier, U., and Hübner, S., 2004. An Architecture for Ontology-Based Discovery and Retrieval of Geographic Information. In: Toppen, F. and Painho, M. (eds.). *Proceedings of the 7th Conference on Geographic Information Science (AGILE 2004)*, Heraklion, Greece, pp. 179-188.
- Racer Systems GmbH & Co. KG, 2005. RacerPro Reference Manual Version 1.9, <http://www.racer-systems.com/products/racerpro/manual.phtml>.
- Studer, R. Benjamins, V.R. and Fensel, D., 1998. Knowledge engineering: principles and methods. *Data Knowledge Engineering*, Elsevier Ltd, Volume 25, Issues 1-2, pp.161-197.
- Visser, U. & H. Stuckenschmidt, 2002. Interoperability in GIS - Enabling Technologies, in: Ruiz, M., M. Gould & J. Ramon (ed.): *5th AGILE Conference on Geographic Information Science*, pp. 291-297.
- Wiegand, N. Garcia, C., 2007. A Task-Based Ontology Approach to Automate Geospatial Data Retrieval. *Transactions in GIS*, Vol. 11, No. 3, pp. 355-376.
- W3C, 2004. OWL Web Ontology Language. <http://www.w3.org/TR/2004/REC-owl-features-20040210>.

ACKNOWLEDGEMENTS

This paper is supported by project of Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing: No.WKL (06) 0301, and National Natural Science Foundation of China (No.40701156).

