

THE STUDY OF INTEGRATION OF MULTI-SOURCES HETEROGENEOUS DATA BASED ON THE ONTOLOGY

LUO Jing^{a,*}, DANG An-rong^a, MAO Qi-zhi^a

^aSchool of Architecture, Tsinghua University, Beijing, 100084

Commission WGs WG II/4

Keywords: Ontology; Digital Urban Planning; Data Integration; Multi-sources and Heterogeneous Data;

ABSTRACT:

The problem of integration of multi-sources and heterogeneous data has been the hot spot in the domain of spatial information technology. This paper analyzed the advantages and feasibility of using ontology technology under the ground of analyzing all relative integration technology, and discussed method of using ontology to solve semantic problem during the integration of planning data. In the course of study, this paper attempted to adopt the method of multi-layer ontology sharing mechanism to implement data sharing, and put forward approaches of constructing ontology data integration platform and then use this platform to realize data integration. All of work carried through useful discussion of theory and practice for realization of digital urban planning.

1. INTRODUCTION

The problem of integration of multi-sources and heterogeneous data has been the hot spot in the domain of study of spatial information technology. Different Unit or department has different understanding and definition to the same geographical phenomena, and adopt different data model, GIS software and spatial data format according their each demand; more and more multi-sources and heterogeneous data had been produced. The result of this is to make data exchange more difficult. In order to adequately use existing data fruits and realize the uniform management in the process of data integration, the need of solving the integration problem of these multi-sources and heterogeneous data was becoming more urgent. Even if adopting data exchange criterion, there still exist many problems. The essence of ontology is the sharing and reusing for domain knowledge. The domain ontology having been standardized and formalized can provide good tool for high-level inter-operation between information systems. Therefore it can provide knowledge map described by corresponding semantic description related to meta-data. Also it is the base for realization of virtual reality and computer animation.

The heterogeneous feature has four kinds: system, syntax, structure and semantic. The study of data integration technology started in the middle of 1970s, and it has existed for more than twenty years as so far. It appeared as multi-database integration at the beginning, and now it developed the current heterogeneous data integration. The region and role of data integration all have extended. But up to now, three of these have solved by multi-technology such as CORBA, DCOM except semantic.

1.1 The basic meanings of ontology

Ontology firstly came for one concept of philosophy. From the aspects of philosophy, ontology is an explanation or description of an external system. Ontology cares the pure essence of external fact. In the domain of computer science, Gruber is the first person who has put forward the concept of ontology: "Ontology is the specific and canonical explanation" [1, 3]. Studer has studied various ontology definitions, and has considered ontology is "specific, formal and canonical explanation for sharing conceptual model" [2]. This definition included four kinds of meanings: conceptualized, specific, formalization and sharing. Conceptualized is a summarized model through abstract some corresponding conceptions of phenomena in the real world; specific is to accurately define conceptions and their relationship; formalization is the accurate math description, that it is to say, it can be processed by computer; but sharing reflect the common reorganization of users to the knowledge. Ontology is a canonical and clear description about one domain. It includes class or conception, and every attribute of conception describes various features, attributes and restriction condition of attributes related to conceptions.

1.2 The role of ontology in data integration

Ontology is an old proposition in philosophy, can explain some questions such as "what's existence", "have any characters can explain the existence", "how these characters explain existence". In other words, ontology is a kind of theory about existence and essence. So, using ontology can radically solve the problem of heterogeneous semantic and realize the data integration in semantic level. And ontology has following three important roles in data integration [4]:

(1) Mutual semantic description (Conceptual Definition)

* rudy_luo@163.com Phone: +86-10-62783328, Fax: +86-10-62783328.

* This study was supported by National Natural Science Foundation of China (50678088), National Project of Scientific and Technical Supporting Programs Funded by Ministry of Science & Technology of China (NO.2006BAJ14B08), and Beijing Natural Science Foundation (4063038).

Conceptual definition has two meanings: 1) through the definition of basic conception and relation can express the inner complex conception of ontology and relation; 2) can take ontology as common conception model, and define conception and relation of each data sources under this and take this as the base of data integration.

(2) Query Model

If data sources can be defined by ontology, the main significance consists in: it can make ontology as agency or deputy, so that large of multi-sources and heterogeneous data is transparent to user. That is to say, user needn't know the structure of data sources, only want to submit a query aimed to ontology; and then, the system can automatically rewrite this query according to semantic definition. In this way, user only put forward the wanted data, needn't point how to discover these data.

(3) Reasoning Base

If data sources can be defined by ontology, the other meaning consists in: we can discover some potential conception and relation from data sources as ontology can construct on the logic base. If users query one conception of ontology, the correlative answers can be found in its child-conception or father-conception. From a specified query conception, we can find its offspring conception and ancestor conception as the potential support answer according to the structure of ontology. Through the inheritance relation, we can confirm the element and attribute of these conceptions. Therefore, the reasoning base mainly embodied in the data integration of multi-source heterogeneous data and largely improves the efficiency of data query.

2. THE ADVANTAGES ANALYSIS FOR DATA INTEGRATION BASED ON ONTOLOGY

The study of data integration started from metaphase of 1980, having existed more than twenty years. From the beginning of multi-database integration to now heterogeneous data integration, both extent and role of data integration all expanded. In the course of development of data integration, there have existed many problems. All basically depend on method of virtual view and data warehouse to solve these problems of data integration. The heterogeneous features of heterogeneous data have been divided into four kinds: system, syntax, structure and semantic. Up to now, the first three kinds have been solved by using various middleware such as CORBA, DCOM and so on.

The early data integration system mostly adopted relation model or object model as common data modal before the appearance of XML. Ontology has close relation with object-oriented model. Ontology attempted to promote human-machine cooperative through describing domain sharing conception model. Most ontology language all used method of object-oriented to describe domain sharing conception model. They considered that there existed class, and class has its instance, and they used attributes to link them in their domain. Therefore, from this meaning, object-oriented is a method of analysis and description of domain. Ontology and object-oriented is different in essence.

Usually, ontology model and object-oriented model have been viewed as different model, and adopted different language to depict them. OWL is a kind of language used to depict ontology, and UML is a kind of language to depict object-oriented

modeling. They have common grounds; also have differentia [5]. Their differentia consisted in following aspects:

(1) The difference in modeling goal: Object-oriented model is abstract, used to eliminate and predigest unnecessary conception and relation, only select necessary knowledge to solve certain fixed problems. Ontology model consist in knowledge expression. It often needs to cover all knowledge in this domain, and emphasize on integrality of knowledge. As the development of knowledge along with time, ontology modeling is a course of ceaseless to be perfected. This feature need development tool of ontology to be ceaseless updated and inter-quoted.

(2) The difference in purpose: Object-oriented model is used in the time of software system development. It can be used to establish accurate and material instances. Ontology model attend to problems in conception level. Though it can explain some fact, but it do not to guide element establishment of material software development. The instance of ontology maybe just a viewpoint, and there maybe existed various viewpoints on one instance in this ontology. Even if various viewpoints, they are not self-contradiction.

The common grounds of ontology model and object-oriented model consist in: both of them are to express conceptions can be processed in the real world; and they all stands on the ground of class and relation, for example sub-class and aggregation relation. Whether ontology or object-oriented, all is to get the reusable model of one domain.

The comparability of ontology model and object-oriented model in one domain is the base of effective development ontology. But the differentia of them result that UML is not unsuitable for ontology modeling, for example, attribute is the first grade model element in ontology, but attribute and relation all not being the first grade in UML. Even the UML2.0 criterion to be issued by OMG in the future can't solve all problems needed by ontology modeling too. To solve these problems, it needs to expand UML so as to realize the object-oriented ontology modeling [1].

The appearance of XML makes different sources and heterogeneous data can be easily combined. It makes query various incompatible data sources to be possible, and provides chances to solve the integration of heterogeneous data sources. Using XML although can effectively solve the problems of data exchange and sharing between different systems and different data sources. Through analyzing those heterogeneous data integration system used XML technology, all of these systems can effectively solve syntax heterogeneous except semantic heterogeneous between data sources.

Ontology has been taken as a tool used to depict conception system of each domain. Through defining ontology, we can explicitly and formally express tolerant and connotative conceptions system of one domain. It provides a possible way of semantic integration and inter-operation for digital urban planning. Ontology system is made up of conception and relationship of conception. The semantic of conception is depicted by sets of attributes of conception class; and relationship of conception is depicted by hierarchical structure of conception class. This hierarchy model assured conception in high level has more pure or general than that in low level. Conceptions in high level depicted the basic features of object, while bottom conceptions depicted rather material object that aim at different applications. Through this manner, it can

explicitly and formally express connotation of conception class and their relationships.

Therefore, using ontology technology to carry through the study of data integration can radically solve the problem of semantic heterogeneous, and can realize the information integration from the semantic level. This method can greatly improve the efficiency of data integration in a way, and enlarge the arrangement and extent of data sharing. From this viewpoint, it has advantages that other technologies don't have.

3. THE CONSTRUCTION OF PLANNING ONTOLOGY OBJECT-ORIENTED

The planning ontology can integrate each feature in one conception through ontology. It is also the study base of handling of spatio-temporal knowledge and helps to the study of different methods aimed to common conception, attribute and relation. So, developing the study of multi-sources heterogeneous data integration through constructing planning ontology is very useful for the expression and handling of spatio-temporal knowledge of planning data, and accumulates valuable experience for the integration of planning data and the application of spatio-temporal reasoning modal in digital city planning at the same time.

From the viewpoint of application, planning ontology is mainly made up of conceptions, relationship of conceptions and depicting language can be processed by computer. The aim of constructing planning ontology is to form the common understanding and cognition to structure of planning information, and then to analyze knowledge of urban planning domain. All of this work provided base for planning semantic web. Therefore, theory of ontology is a methodology about using computer language to standardize expression of planning domain knowledge, organize planning knowledge and carry through planning knowledge service. The theory of planning includes special conception and instances of planning; it is a formally depiction for application knowledge, also definition for various conceptions. In the theory of planning or system of planning knowledge organization, it expresses planning conceptions and their relationship, and can be processed by computer.

In the background of digital city, taking digital urban planning as main study object, using various information technology such as RS, GIS, GPS, VR, web communication, database management, adopting ontology technology to carry through data integration and using material ontology describing language to depict planning data, it can solve the problem of semantic heterogeneous in the course of data integration. There often have three periods in the course of constructing ontology: The first period: Construct sharing glossary (global ontology). This period includes three main steps, that they are analyzing data content needed by every period, looking Primitive Meta language, and defining global ontology;

The second period: Construct local ontology. This period includes two main steps that they are analyzing planning data and defining local ontology;

The third period: Define mapping. This period mainly define mapping relation between global ontology and local ontology so as to solve the problem of semantic heterogeneous.

3.1 The construction of planning ontology

3.1.1 The analysis of preconditions

It needs to analyze the capability of constructed ontology before constructing global ontology. In digital urban planning, the capability of ontology often needs to solve following problems: 1) The kinds of digital urban planning; 2) Which kind of planning data needed to be attended when we are carrying through digital urban planning; 3) What features do these planning data have, and what are their advantages and disadvantages? 4) If only selecting one kind of data, whether it can satisfy the demand of digital urban planning? 5) Which kind of data is frequently used by digital urban planning? 6) In different period, which kinds of data do planning need? 7) Whether each kind of data all need to renew along with time? 8) How long do these time interval need? According to these problems, we know that planning ontology at least includes planning type and data type.

3.1.2 The information description of planning ontology

In digital urban planning, global ontology needs to depict data type and planning type. Using Protege2000, under the root conception "Thing", we can construct the second root conception "Data Type" and "Planning Type". Global ontology is used by various local ontologies which correspond to data source. The form and amount of the bottom heterogeneous data source increasing with time. Therefore, this global ontology also can be called domain ontology.

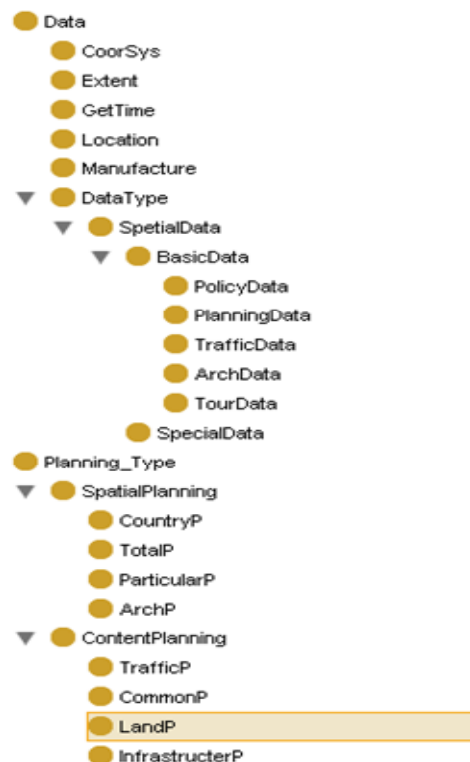


Figure 1 Ontology classes of digital urban planning

The domain ontology refers to that it is a ontology that it can provide activities, main theory and basic principle: it can be reused in one domain, and can provide conception definition and relationship between conceptions for one domain. Conceptions in one domain usually have these features:

connotation defined accurately and extension can be confirmed. Further more, domain ontology and application ontology is not divided strictly. Application ontology in one domain maybe is domain ontology in another domain. They supplement each other.

Before constructing global ontology, it needs to analyze conceptions and their relations of this domain. This is the basis of whole ontology modeling. In digital urban planning, starting with the basic conceptions of data, we confirmed following basic conceptions: spatial data, attribute data. Under this basis, the lower level conceptions are: data number, data acquisition time, user name and department etc.

3.1.3 Confirm the attribute sets of ontology conceptions

Domain experts confirm the attributes sets of ontology to every conception. This operation had to be done from the upper root to the bottom. That it is to say, firstly confirm attributes of the highest level, and then is the rest [6]. The aim of it is to make use of the features of the highest level that has little sets of attribute and highly abstracted. When defining sub-class conceptions, firstly they inherit some sharing attributes from their father class by succession mechanism, and then add special attributes belong to sub-classes. In digital urban planning, the sets of attributes of data ontology confirmed are: extent, location, coordination, time, data provider. These attributes described the common attribute of data, and it can be used to integrate all kinds of data type.

3.1.4 Adopt OWL to describe ontology

Using OWL to describe ontology is to use that well-defined meta-ontology of OWL to formally express conceptions and their relations. In this process, the most things are to define classes, sub-classed, attributes and their features. In the course of ontology construction, it automatically created OWL files because of using Protégé tool.

3.1.5 Create unit

Ontology construction usually started with defining classes. In digital urban planning, firstly listed general conceptions of this domain as root class, then express each level class illustrated in Fig. 1. After this, we created relevant units according to these constructed classes. In digital urban planning, aiming to grid data class, we created two units: RS and DWG; while to vector data, we created some units like GIS.

3.1.6 The features of planning ontology

The planning ontology has following four features:

- (1) Spatio-temporal feature of planning information. The planning has strong feature of space and periodicity in itself. So it needs to clearly express the geographical information relative to city planning;
- (2) Diversity of planning information. Planning involves broad extent such as spatial information, time information, geographical information of planning and industry information (policy, rule of law etc.);
- (3) Diversity of expression. There exists different semantic to the same conception because of the different intent;
- (4) Diversity of data source.

3.2 Ontology Database of digital urban planning

3.2.1 Structure of ontology database

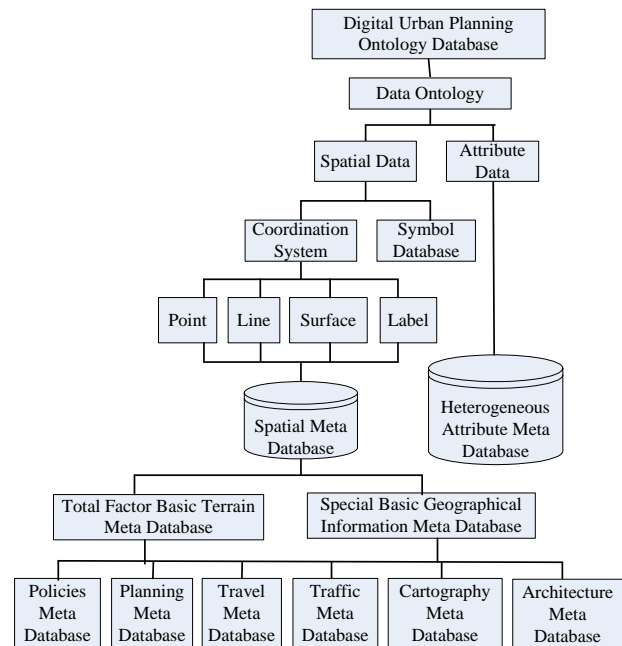


Figure 2 Ontology Database of Digital Urban Planning

Figure 2 illustrated part of the structure of digital urban planning ontology database. Meta-data is a kind of data regarding data, and is descriptive information about data and information resources. Spatial meta-data is descriptive information about geographical spatial data and relative information resources [7]. Data object in digital urban planning often are spatial data and attribute data. Spatial data is abstracted from spatial meta-database including points, lines, surfaces and spatial labels. Attribute data come from different units because different units formed many heterogeneous attribute data while described the same object. However, Transformation between different data format usually existed information loss in a certain degree. To avoid this problem, we adopted seamless integration of multi-sources data that can provide uniform database input interface for attribute data relative to spatial data so as to construct application ontology in the condition of keeping the format of multi-sources data.

3.2.2 Multi-level ontology sharing realization of digital urban planning

In the course of multi-level ontology sharing realization, the upper level is taken as ontology that describes the next level ontology. The top conception described those conceptions of the most abstract and extensive in this domain, that it is domain ontology introduced before; the bottom is those application ontology oriented to material application. During material application, ontology in different levels can carry through description of semantic coordination by sharing upper ontology so as to realize consistent between them. Multi-level ontology sharing mechanism is defined: it creates new application ontology level through the rule of date sharing and data reusing under the semantic coordination of domain ontology. The core is: divided geographical entities in the same domain, same time

and same space into different levels so as to realize sharing and reusing of huge urban information.

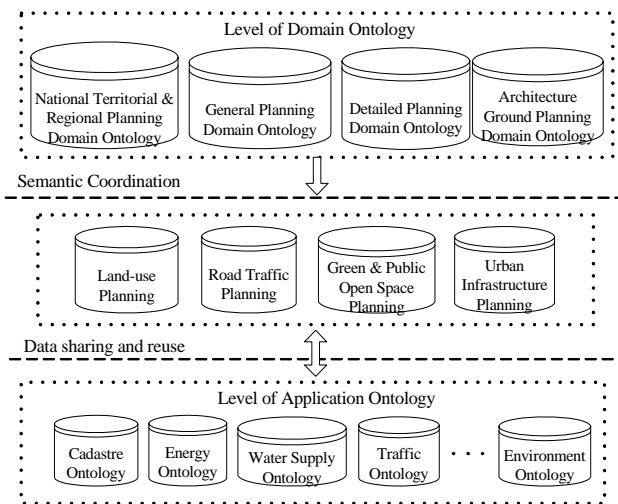


Figure 3 Realizing Mechanism of Multi-levels Ontology Sharing

Taking public areas planning as example, we need to confirm the land status whether belongs to public land, traffic status is convenient, water supply system and energy supply system is perfect and so on when we carried through detailed planning. Thus, we need to share cadastre data, traffic data, water supply data and energy data. But traditional database can implement this aim because single knowledge, so we need a semantic coordination mechanism to implement above data sharing. While multi-level ontology sharing mechanism just is the tool of realizing this aim. Fig.3 illustrated this multi-level ontology sharing mechanism implemented in the course of construction of planning ontology.

3.3 Data Input

Multi-sources heterogeneous urban planning data input is the base of implementation this system, also is the difficulty. It included basic spatial data input and multi-sources heterogeneous attribute data input. Basic spatial data mostly was the format of DWG drawn by local city planning academy. These data were drew under CAD environment that mostly serving cartography demand, only having spatial information not having attribute information. Therefore, these data can't satisfy the demand by GIS. CAD emphasizes on graphical information transmission, while GIS emphasizes on spatial data process and analysis. Before spatial data input, it is the spatial graph transmission usually implemented by different GIS platform.

3.4 System structure

Based on the above work, we constructed database application system illustrated in Fig 4. Users can query through semantic query interface, while administrator can continually modify and consummate digital urban planning ontology database through ontology development interface.

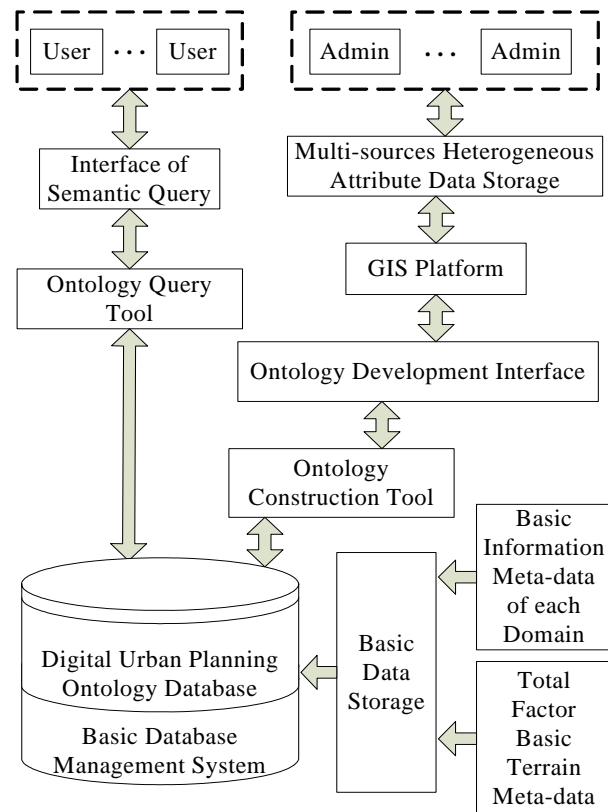


Figure 4 Figure of Ontology Application System Structure

4 CONCLUSIONS

Digital city planning regards spatial planning as its emphases not involving non material spatial planning, and contains a large extent including city, block and plot; also involve national planning, regional planning and so on. All of these features of it want to accurately understand the real world; and ontology is an abstract subject that mainly concerns the meaning of conception. They all concern explaining and describing to the real world in systematization way. Therefore, it is significant that applying conceptual structure constructed by ontology to construct the comparatively physical structure, and carry through digital city planning under this.

REFERENCES

- Chen Kai, He Keqing, Li Bing, Liang Peng. Research on Ontology Modeling Using Object-Oriented Technology, Computer Engineering and Applications, 2005-02-012
- WU Hao, XING Gui-fen. Research on technology of information integration based on ontology, Computer Applications, 2005-02-01
- Gruber T R. A Translation Approach to Portable Ontology Specifications [S], Knowledge Acquisition, 1993
- H.Wache, T.Vogele, U.Visser, H.Stuckenschmidt, G.Schuster, H.Neumann, and S.Hübner. "Ontology-based Integration of Information-A Survey Existing Approaches," In: Proceedings of IJCAI_01 Workshop: Ontologies and Information Sharing, Seattle, WA, 2001, Vol.pp.108-117.

Holger Knublauch, Thomas Rose. Round-Trip Engineering of Ontologies for Knowledge-Based Systems [C], In: Proceeding of the 12th International Conference on Software Engineering and Knowledge Engineering, 2000: pp. 239-247

WANG Xiao Lin, LUO Ying Wei, CONG Sheng Ri etc. A STUDY OF SPATIAL METADATA AND ITS APPLICATIONS, Journal of Computer Research and Development, 2001-03-014

An Yang, Zhao Bo. Spatial information integration based on ontology, Journal of Huazhong University of Science and Technology (Nature Science Edition), 2006-S1-023

WANG Zhenxing, DAN Tangren, YE Changeqing etc. Research on Ontology Integration Framework, Computer Engineering, 2007-02-002

ZHANG Lei, XIA Shi-xiong, NIU Qiang. Ontology-based approach to heterogeneous database integration, Computer Engineering and Design, 2007-14-007

Ontology Development 101: A Guide to Creating Your First Ontology

<http://www.w3.org/RDF/>

ACKNOWLEDGEMENTS

This study was supported by National Natural Science Foundation of China (50678088), National Project of Scientific and Technical Supporting Programs Funded by Ministry of Science & Technology of China (NO.2006BAJ14B08), and Beijing Natural Science Foundation (4063038).