A Model of Extracting ITS Ontology from Relational Database

Ke ZHENG[12a], Dunyao ZHU[1b], Lixin YAN[1c] and Hao CAI[1d]

[1] Intelligent Transport Systems Research Center, Wuhan University of Technology, Wuhan, China
[2] College of Computer and Information Engineering, Nanyang Institute of Technology, Nanyang, China
[a]Doctoral Candidate, E-Mail: zanghshun@163.com
[b] Professor, E-Mail: dunyaoz@kotei.com.cn
[c] Postgraduate Candidate, E-Mail: lizhenyun69@163.com
[d] Doctoral Candidate, E-Mail: 369875181@qq.com

ABSTRACT

In order to realize information sharing, the Intelligent Transport Systems (ITS) should describe information at the semantic level. Ontology is a modeling tool used for domain knowledge sharing and reusing. But most information of ITS is based on the relational database. This paper presented a new modal of ITS domain ontology based on the Chinese ITS's characteristics and the exiting research results, established rules of extracting ITS ontology from a relational database, and proposed an automatic ontology extraction algorithm. Experiments showed that the rules and algorithms can quickly extract ITS ontology from a relational database.

KEYWORDS: ITS, Semantic, Ontology, Relational Database, Automatic Extraction

INTRODUCTION

With the constant improvement of Chinese transportation infrastructure, a real-time, high-efficiency, and accurate new intelligence transportation system (ITS) has become the research hotspot (Editorial Department of China Highway and Transport, 2012). ITS is a comprehensive traffic information management system that uses advanced information technology, data communication transmission technology, electronic sensor technology, control technology, and computer technology to complete information collection, processing, and publishing. With traffic information, it is difficult to achieve information sharing and interoperability between heterogeneous systems traffic information consists of large amounts of heterogeneous and multi-source data. This issue has become the key problem in intelligent transportation systems. To solve this problem, we need a clear semantic description of data so that the systems can achieve common knowledge in intelligent transportation.

As a modeling tool that can describe the conceptual structure of information systems at semantic and knowledge levels (Gruber, 1993), ontology is the sharing and reusing of domain knowledge. Standardization and formalization of domain ontology provide a sound solution for interoperability between disparate systems. Many scholars in the field of intelligent transportation, both at home and abroad, have begun to study ontology modeling and ontology mapping theory to solve the problem of ITS semantic integration and resource sharing. For example， Mnasser, Maha and Kathia (2010) used protégé tools to develop traveler information service ontology. Cao (2010) researched the relationship between the ontology and the data model of multi-strategy mapping method for the city ITS heterogeneous data integration.

Currently, most data of information systems are stored in relational databases and ITS is no exception. The knowledge in the database is usually highly domain correlated, so it can be faster to extract domain-specific ontology from a relational database. This paper will study how to extract domain ontology of ITS from the existing large number of relational databases.

THE MODELING METHOD OF ITS DOMAIN ONTOLOGY

Ontology that provides a clear conceptual model and inference rules can realize the integration of information at the semantic level. Domain ontology describes the domain knowledge by giving specific formal definitions of concepts and relationships, activities, characteristics, and law in a specific domain. Establishing the ITS domain ontology is a tool to achieve the object of sharing and reusing the data resource.

Domain ontology

The standard definition of ontology was put forward by Gruber in 1993: ontology is an explicit specification of a conceptualization (Gruber, 1993). Ontology describes shared and a machine-readable knowledge. A complete ontology should have five basic elements: concept, relation, function, axiom, and instance. Relation represents the interaction between concepts. There are four basic relations: part of， kind of，instance of and attribute of. The actual application of ontology does not have to follow in strict accordance with the above five primitive elements. And to meet the application needs, we can refer to the actual application circumstances defined relations, which are not confined to the above listed four basic relations.

In general, ontology that is geared to the needs of specific fields describes the concept of domain model. An important application of ontology is establishing domain knowledge modeling. Ontology can build a set of shared terms and information structure for a specific field so that we can implement information integration and the sharing of multiple source, which can become homogeneous through ontology.

OWL (Web Ontology Language) is a W3C recommended language used to describe the ontology relations, which enables the computer to automatically understand ontology-related ideas. OWL supports descriptions of the axioms, attribute feature constraints, property type restrictions, cardinality restrictions, logical combination of the class, etc. In OWL, ontology is composed of categories, properties, instances, and the relationship between instances. This paper will formally define ontology by

$$O ::= \{C, P, D, I, A\}$$

where O is the domain ontology, C is the set of classes identifier, P is the set of property identifier which can be classified into DP (data type property) and OP (object type property) by the attribute values, D is the set of data type identifier that has predefined in OWL, I is the set of instance identifier, A is the set of axiom identifier which can be classified into three types: A_{class} (class axiom), $A_{property}$ (property axiom)， and $A_{instance}$ (instance axiom).

This paper will describe the domain ontology that is extracted from the ITS relational database with OWL.

The Establishing Model of ITS Domain Ontology

Building a domain ontology method can vary due to different areas and different purposes. We should completely and deeply understand a specific domain and clear purpose of constructing ontology when we build ontology.

In China, the ITS characteristic mainly reflects in the high distribution state of the traffic information system. The main purpose of building ITS domain ontology is to achieve the knowledge reusing and sharing between people, departments, and software systems in different backgrounds, languages, tools, and techniques by providing the shared vocabulary (Cao, 2010). Therefore, building ITS domain ontology must first describe existing domain resources at the semantic level.

The existing domain resources of ITS include a relational database that is used in all information systems of ITS. “Domain knowledge resources” are the specification standards or experience and knowledge stored in the mind of domain experts, and “domain entity” is the department and user who design ITS.

The relational database is a tool used to store and manage knowledge of a particular field, and its maturity and stability has been widely proven. The knowledge in the database is usually has a high domain correlation, so it can be faster to extract domain-specific ontology from a relational database. Most data of ITS information system are stored in relational database, so we should make full use of the existing ITS relational database. But because of the restriction of data volume, completeness, and semantic information in relational database, the ontology extracting from the relational database is not complete and lacks equivalent concept and concept hierarchy. As a consequence, the authors would improve ontology that is extracted from relational database semantically by using the Word Net and integrating the existing ontology, so that eventually the authors could give a more complete and accurate available ontology. Figure 1 shows the building model of ITS ontology. This paper mainly studies how to extract domain ontology from its existing database.

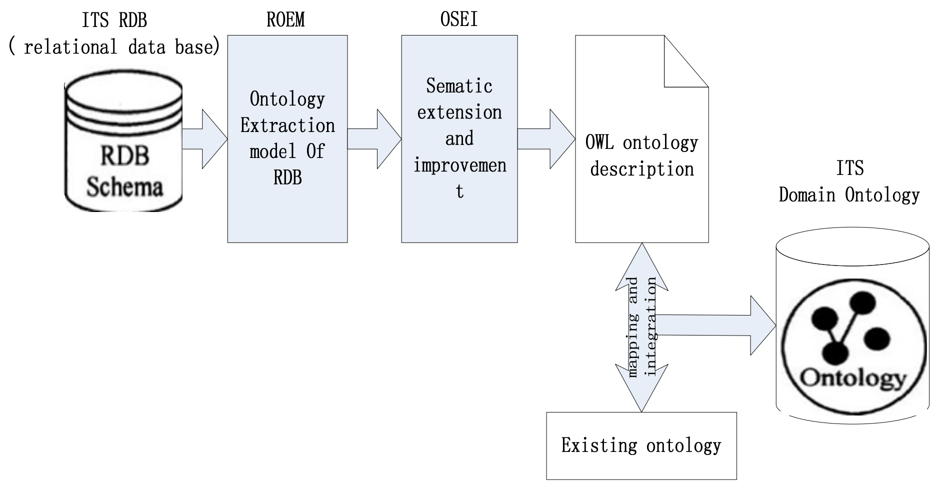

Figure 1. Building model of ITS ontology.

EXTRACTING ITS ONTOLOGY FROM RELATIONAL DATABASE

Relational Schema

A relational database mainly uses relational schema to describe the objective world (Chang et al., 2004). The relational model is a data model that expresses entity type and the relationship between them by using two-dimensional table structure. Relational schema can be formally expressed as

$$R = \{P, D, Dom, F\}$$

where R is the name of relationship, P is a set of relationship properties identifiers, D is definitional domain of properties, Dom is range of properties, and F is a set of relationship of properties data. Relational schema can be defined due to its structure as the following.

Definition 1: General Relational Schema

A relationship schema that does not include a foreign key is called the general relational schema. It is taken for R_B. That can be formally expressed as

$$\forall R_i \in R, FKey(R_i) = \phi \Rightarrow R_i \in R_B$$

where $\mathrm{FKey}(R_i)$ is the foreign key of R_i.

Definition 2: Many-to-Many Relational Schema

A relationship schema whose primary key is composed of two foreign keys is called the many-to-many relational schema. It is taken for R_{MM}. That can be formally expressed as

$$\forall R_i, R_j, R_t \in R, PKey(R_t) = FKey(R_t), |FKey(R_t)| = 2,$$

$$FKey(R_{t_1}) = PKey(R_i), FKey(R_{t_1}) = PKey(R_i) \Rightarrow R_t = R_i \leftrightarrow R_j \in R_{MM}$$

where $\mathrm{PKey}(R_i)$ is the primary key of the R_i. If the schema has no properties except composite keys, then it is taken for R_{MF}. Otherwise, it is taken for R_{MN}.

Definition 3: One-to-Many Relational Schema

If a foregin relationship schema is another schema's primary key, then this schema is called the one-to-many relational schema. It is taken for R_{OM}. That can be formally expressed as

$$\forall R_i, R_j \in R, FKey(R_i) = \mathrm{PKey}(\mathrm{R}_j) \Rightarrow R_i \leftrightarrow R_j \in R_{OM}$$

If the number of properties values is one, then R_{OM} is called one-to-one relational schema and is taken for R_{OO}.

Extracting Model of ITS Ontology from RDB

Most data of the ITS information system use the RDB (relational database) to store and manage knowledge. The knowledge in the database usually has a high domain correlation, so it can be faster to extract domain-specific ontology from a relational database. According to the relationship of a given RDB model, using certain transformation rules and the algorithm to generate a set of corresponding OWL DL ontology concepts with a set of attribute constraints, mapping set and instances is called extracting ontology from RDB. The procedure is divided into two steps. First is converting the relational schema to ontology classes, properties, and axiom. Second is converting tuple to ontology instances. Figure 2 shows the model of extracting ITS ontology from the RDB.

Two conversion processes must comply with a rules set that has been defined in advance. According to the relational schema structure, the transformation rules are divided into three categories: concept extraction, attribute extraction, and axiom extraction. This paper, according to the RDB of ITS, defines the transformation rules.

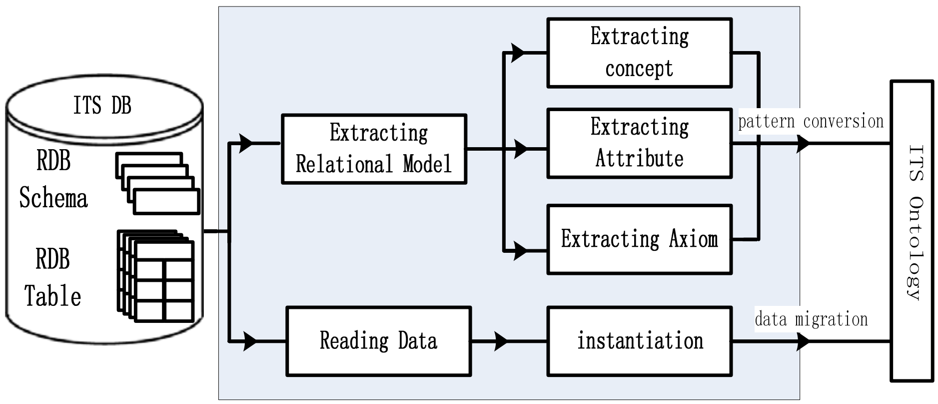

Figure 2. Extracting Model from RDB to Ontology.

The Rules of Extracting Concept

Rule 1: Each relationship except R_{MN} in relational database model must be converted into the concept.

$$\forall R_i \in R, R_i \notin R_{MN} \Rightarrow C_i = \varphi(R_i), \mathrm{C_i} \in \mathrm{C}$$

where $\varphi(\ )$ creates a identifier based on input name and C_i is the identifier of class.

The rules of Extracting Attribute

Attributes are general facts or specific facts of individuals about the class members. In ontology, the attribute can be used to express the relationship of individuals or between individual and numerical. The attribute that associates individual with individual is called OP (Object Properties), and the attribute that associates individual with numerical is called DP (Data Properties).

Rule 2: All the attributes except foreign keys in relational schema must convert to *DP* of concept. Its attribute domain is concept C_i that is generated from the relational schema, and its range is *D* (date type) in OWL, which corresponds to data type in RDB.

$$\forall P_{ij} \in P, P_{ij} \neq FKey(R_i) \Rightarrow DP_{ij} = \varphi(P_{ij}), \mathrm{DP_{ij}} \in \mathrm{DP}$$

Object property of ontology is usually expressed in foreign keys in the relational schema. But the inheritance relationship between classes is also expressed in foreign keys; in this case, the foreign key cannot be converted to object properties. The inheritance relationship can be reflected through the subclasses using the parent primary key as the primary key.

Rule 3: All the foreign keys in relational schema except the inheritance relationship must convert to *OP* of concept. Its attribute domain is concept C_i that is generated from the relational schema, and its range is the concept C_j that is the class generated from the relational schema, which is foreign key references.

$$\forall P_{ij} \in P, P_{ij} = FKey(R_i), FKey(R_i) \neq PKey(R_i) \Rightarrow OP_{ij} = \varphi(P_{ij}), \mathrm{OP_{ij}} \in \mathrm{OP}$$

The Rules of Extracting Axiom

Rule 4: For a relational schema that only has one foreign key and that is the primary key creation subclasses axiom.

$$\forall \mathrm{R_i}, R_j \in R, R_i \neq R_j, R_i \leftrightarrow R_j \in R_{OO} \Rightarrow \varphi(R_i) \subseteq \varphi(R_j)$$

Rule 5: For each relationship creates class axioms.

Rule 6: Converts data integrity constraints of relational database to attribute characteristics or constraints of the ontology concept.

The Rules of Creating Instance

Rule 7: Map the tuples to an instance of ontology.

$$\forall \mathrm{R_i} \in R, t_{ij} \in R_i \Rightarrow createIn(t_{ij})$$

where t_{ij} is a tuple in relational table, and *create In(t_{ij})* is to create instance.

EXPERIMENT

The essence of the intelligent transportation is using computer technology to manage traffic information including road information, traffic flow information, and interactive information. Road information is the core and foundation of the traffic information.

The author used the ITISS (intelligent transportation information service system) that has been put into use to extract intelligent transportation ontology. The ITISS is a BS mode information release system based on Java EE, whose RDBMS is MySQL5.1. Due to limited space, this paper will take the RDB of road network as the experimental object. Figure 3 shows the relation schema of the road network.

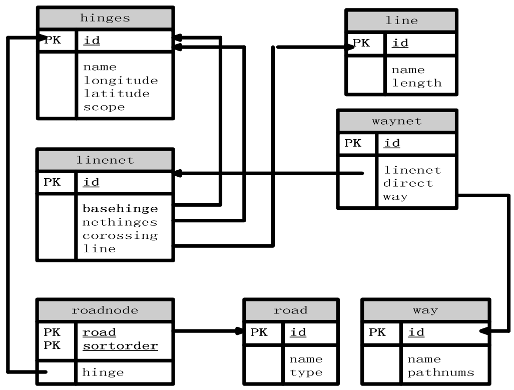

Figure 3. Relation Schema of Road Network.

The author designed the algorithm according to the model and rules aforementioned and implemented the algorithm using Java language. The input of the algorithm is the DDL of relational database, and the output is domain ontology, which is described in OWL DL files. Figure 4 is the OWL DL file fragment that is the output of the algorithm.

```
<owl:Class rdf:ID="Waynet">
<owl:DatatypeProperty rdf:ID="id">
  <rdfs:domain rdf:resource="#Waynet" /> <rdfs:range  rdf:resource="&xsd:int"/>
</owl:DatatypeProperty>
<owl:ObjectProperty rdf:ID="linenet">
  <rdfs:domain rdf:resource="#Waynet"/><rdfs:range rdf:resource="#Linenet"/>
</owl:ObjectProperty>
<owl:DatatypeProperty rdf:ID="direct">
  <rdfs:domain rdf:resource="#Waynet" />  <rdfs:range  rdf:resource="&xsd:boolean"/>
</owl:DatatypeProperty>
<owl:ObjectProperty rdf:ID="way">
  <rdfs:domain rdf:resource="#Waynet"/><rdfs:range rdf:resource="#way"/>
</owl:ObjectProperty>
</owl:Class>
```

Figure 4. Part of the OWL ontology.

The algorithm can easily get OWL DL ontology from a relational database. Table 1 shows the ontology of road network.

Table 1. The Ontology of Road Network Extracting from RDB

Classes	The Number of DP	The number of OP
Hinges	5	0
Line	3	0
Way	3	0
Road	3	0
RoadNode	1	2
LineNet	2	3
WayNet	3	1

CONCLUSION

To implement information sharing at the semantic level, a modern transportation system is important in order to achieve real "smart" information, and ontology is the tool that will help to achieve this goal. Most of the existing information in ITS is stored in RDB, and the knowledge in RDB is generally highly domain correlated, so extracting ontology from RDB is feasible. This paper analyzed the structure of ITS RDB, gave the transformation rules from relational schema to ontology, established the extraction model of intelligent transportation ontology, and verified the correctness of the algorithm through the experiments.

ACKNOWLEDGEMENTS

The research is supported by the Foundation of National Key Technology Program of China (2011AA110402) and Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control and the Tackling-key Project (122102210536) of Henan Department of Science and Technology.

REFERENCES

Cao, Y. (2010). "Application research on semantic integration of city ITS based on ontology theory." Ph.D. thesis, Dalian Maritime University, College of Transportation Management.

Chang, K. C. C., He, B., Li, C., Patel, M., and Zhang, Z. (2004). "Structured databases on the web: Observations and implications" *ACM SIGMOD Record*, 33(3), 61-70.

Editorial Department of China Highway and Transport. (2012). "An academic research summary on China highway and transport: 2012." *China Journal of Highway and Transport*, 25(3), 2−50.

Gruber, T. R. (1993). "A translation approach to portable ontology specifications." *Knowledge acquisition*, 5(2), 199-220.

Houda, M., Khemaja, M., Oliveira, K., and Abed, M. (2010). "A public transportation ontology to support user travel planning." *Research Challenges in Information Science (RCIS), 2010 Fourth International Conference on,* IEEE, 127-136).

Zheng, K. and Liu, K. (2012). "The research of intelligent transportation system based on formal concept ontology models." *JCIT: Journal of Convergence Information Technology*, 7(10), 158-165.