

Research on Heterogeneous Data Sharing Based on Ontology

Zhao Guozeng

Luoyang Institute of Science and Technology Computer
and Information engineering department
Luoyang, China, 471023

Ding Guoqiang

Luoyang Institute of Science and Technology Computer
and Information engineering department
Luoyang, China, 471023

Abstract—The traditional method of heterogeneous data sharing can not describe the formal semantics of data and finds it difficult to solve the semantic heterogeneity in the heterogeneous information sharing. Based on the XML-based heterogeneous data integration technology, the heterogeneous data sharing method is proposed based on ontology by introducing metadata and ontology. By metadata's specifying the distributed, heterogeneous data and by describing semantic information of data source, having "ontology" as a common semantic model, semantic match is established through ontology mapping between heterogeneous data sources and semantic difference institutions are shielded, so that semantic heterogeneity problem of the heterogeneous data sources can be effectively solved. Experiments show that the method can solve the problem of semantic heterogeneity of heterogeneous data.

Keywords—*Semantic heterogeneity, Metadata, Ontology, Ontology mapping*

I. INTRODUCTION

With the rapid development of computer network technology, the database technology and wide application of MIS, a series of "Information Islands" came into being due to different running environments, different storage and representation of data. In order to make better use of these distributed, heterogeneous information resources, integration and sharing of heterogeneous data sources in the network becomes an effective way of eliminating the "Information Island".

Heterogeneous data sharing is the premise of ensuring the integrity of data to provide users with a unified access interface. The different sources and different data formats in the logical or physical integration, packaging, handling, shielding the underlying data source differences[1]. XML has a mature technology standard and improve the query mechanism, which fully meet the Internet and distributed heterogeneous environments, it can realize the data that the separation of content and data, enabling heterogeneous data sources through a unified data model the exchange of information, to some extent to achieve the data sharing, but the semantics of heterogeneous information sources is difficult to be effectively addressed. Artificial Intelligence, Ontology, Semantic Web technology for heterogeneous data integration provides new methods and ideas[2]. At present, ontology has been described as a data source and the sharing tools applied to heterogeneous data sharing to solve the semantic heterogeneity problem[3].

Experiments show that the heterogeneous data sharing method based on Ontology and metadata can solve the problem of semantic heterogeneity.

II. RESEARCH ON RELATED TECHNOLOGIES

A. Metadata standards

Metadata is used to describe information resources, interpretation, localization, is a structured information and easy to extract and use[4]. Its purpose is to let the machine can not only read but to understand in order to solve the mapping problem of information and integration issues. In the heterogeneous information resources, "data" is transmitted from the local all terminals, to undergo an "information 1 - Metadata - Information 2" data conversion, this process, its plays a key role. This system, the metadata of information sources and to achieve a clear description of information users, by the description of information sources and information users to convert data between formats, heterogeneous data transfer between data sources.

Solve the resource metadata is the primary means of various forms. There are several different international standards. Resource sharing in modern, generally required to follow some big standards, according to characteristics of the actual needs and resources to build their own resource metadata standards.

Dublin Core metadata standard, referred to as DC, can be used to describe most of the resources, has become an information resource for describing any metadata standard.

B. Ontology

Ontology has a similar philosophy and metadata definition: "the existence of the presence" in the knowledge sharing process, "ontology is a shared conceptual model of clear formal specification." [5] Since the Semantic Web have been proposed, the ontology became interconnected content network resources to solve an important way, the semantic representation in the sharing of resources and reasoning play an important role. Currently, the establishment of various public shared ontology knowledge base using ontology reasoning for knowledge reasoning research is under way.

Heterogeneous data sharing, the ontology as a middleman and it shields the underlying structure of heterogeneous data sources. Users to perform queries, you can not know the structure of various data sources, they do not know how to query data, only need to know what data the system provides a query for the body on it. System can be defined semantics and ontology mapping will automatically query is decomposed into a query for each

data source. Logical reasoning ability through the ontology, in a distributed environment of heterogeneous data sharing, the information can greatly improve the precision and recall.

Use of metadata standards to achieve a unified presentation layer of data resources and mutual understanding, the use of ontology technology resources, interconnection semantics, in the research literature [6] has a number of related reports. The basic idea of Heterogeneous resources sharing, as show in Figure.1.

III. HETEROGENEOUS DATA SHARING METHOD BASED ON ONTOLOGY

A. Model of heterogeneous data sharing system

According to the correlation of heterogeneous data sharing and data query technology model, this paper based on Mediator / Wrapper methods, XML technology and the actual demand, a ontology-based heterogeneous data sharing system. The method introduced in the system ontology, semantic ontology to describe the advantages of heterogeneous data sharing to solve the semantic problem.

The model,as show in Figure.2. Model is divided into the physical layer, middle layer and application layer. By a direct interaction with the underlying physical layer is responsible for. Middle layer is responsible for receiving user's request, the core of the system. Interact directly with the user application layer, is responsible to provide users with data query interface. User submits a query is based on the overall global ontology model, users do not care about the underlying data source, distribution and heterogeneity of data sources, query processing done by the middle layer processing.

B. Create ontology

In the ontology, knowledge is the use of classes, instances, relations, functions, and axioms to the standardization of the five primitives, therefore, ontology can be expressed by a group of five dollars. $O=(C, I, R, F, A)$. C, I, R, F, A denote concepts, instances, relations, functions, and axioms. Classes in the ontology are often classified organization to the concept of form, as the five basic primitives to describe the concept [7].

The concept of the body also has a different relationship between the basic relations of specific forms shown in Table I.

The system uses hybrid ontology approach; need to build the global ontology and local ontology. Global ontology and local ontology, mapping relations between them using ontology language (OWL) is described.

In the relational data source, may have different data models, the same data model semantic differences may exist for this situation, through the relational database model E / R model analysis to determine the metadata, then the metadata Information ontology. The relationship between the relational database name, field, and create relationships between entities in the ontology classes, properties and roles. At the same time the process of creating a body to record the correspondence as a local ontology and the mapping information between data sources.

Assuming the data source DataSource1 and DataSource2, courses and instructors have indicated the information; its structure is as follows:

DataSource1: Course(ID, name, time, lecture)

Teacher(TeacherID, name, sex, PresentPosition)

DataSource2: kc(number, kcmc, xs, skz)

Js(bh, xm, xb, zc)

DataSource1 and DataSource2 of the analysis were established corresponding to the local ontology. The local ontology of the DataSource1, as show in Figure.3.

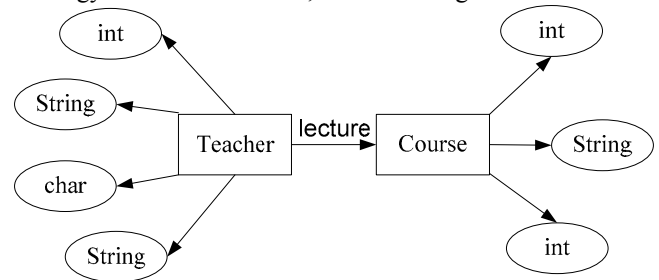


Figure.3 Datasource1 ontology

Tables in the relational model can be represented entities and linkages. Through the relational database model E / R model analysis to determine the metadata, the metadata information to create ontology. The relationship between the relational database name, field, and create relationships between entities in the ontology classes, properties and roles.

Through the analysis of heterogeneous data sources, found here, "Course" and "kc", "Teacher" and "js" to convey the same concept in fact. Similarly, "Course" of the "ID", "name", "time", "teacher" and "kc" the "number", "kcmc", "xs", "skz" has the same meaning. DataSource1 and DataSource2 been integrated global ontology and language to describe the use of owl.

C. Ontology mapping

Ontology mapping is two or more different ontology as input, and then follow the semantic relationship of these ontology elements (concepts, attributes, relationships) to establish the appropriate semantic relations, by mapping to achieve a common understanding of the same things.

Mapping to establish the specific methods can be divided into the following steps:

Step1: concept mapping. Concept mapping can be regarded as a binary relation; mapping rules can be formally described as a collection of M1:

$$M_1 = \{ \langle c, g \rangle \mid c \in C_L \wedge g \in C_G \} \quad (1)$$

Step2: Attribute mapping. Similarly, the attribute mapping rules set M2, can be formalized as:

$$M_2 = \{ \langle c, g \rangle \mid c \in A_L \wedge g \in A_G \} \quad (2)$$

Step3: Relationship mapping. The set of mapping rules between defined M3. The formal description is:

$$M_3 = \{ \langle r, r' \rangle \mid r \in R_L \wedge r' \in R_G \} \quad (3)$$

Where, RG' represents RG relation derived using the path of its collection. Between local ontology for each "r", assuming "r" is connected to the concept of "c1" and "c2" of the relationship, the mapping has been established by M1 can be found in "c1" and "c2" corresponds to the global ontology The concept of "g1" and "gn". Let "r"= R_1, R_2, \dots, R_n " is the relationship between the path of "g1" to "gn", the establishment of mapping "r \rightarrow r'", or "r= R_1, R_2, \dots, R_n " added to the collection in the M3.

Step4: By combining M1, M2, M3, get a collection of ontology mapping rules, and to build the global ontology and local ontology mapping table.

In the owl (Ontology file), by adding "equivalentClass" syntax, the concept of the global ontology and local ontology in the concept of association, that the mapping between classes; by adding "equivalentProperty" syntax, the properties of the global ontology Attributes associated with the local ontology, mapping between that property.

In the mapping algorithm, based on the concept of ontology, properties, instances of similarity, to determine the match between the ontology elements. Similarity calculation can be expressed as:

$$\text{Sim}(e_{o1,i}, e_{o2,j}) > s, (s \in [0, 1]) \quad (4)$$

Among them, the formula "s" that set the threshold, if they meet the conditions on the type, indicating that the element "e_{o1,i}" and "e_{o2,j}" similar to the mapping relationship can be established.

D. Query processing

Heterogeneous query processing is the main research data sharing, the goal is to distributed heterogeneous data sources from multiple query data, and the integration of the results of these queries, and then the final result will be presented to the user. User query against the global ontology concepts and properties in the query, the query Q can be defined as a triple $\langle S, F, W \rangle$, where "S" to "select" clause, "F" for the "from" Clause, "W" on behalf "where" clause. Specific steps are as follows:

Step1: Query interface to receive and resolve user queries, generates a Onto-QL (global query).

Step2: A global query is decomposed into sub-queries each local ontology.

Step3: Convert the query results.

Step4: Semantic analysis, aggregation, sort, and remove the redundancy, back to the user.

For example, the details of the curriculum name "artificial intelligence" and the instructor name, course number, teaching hours, then submitted to the global ontology query:

Select c, d, e

from Teacher=a, a.lecture =b, a.name=c, b.kcID=d, b.time=e

where b.name='artificial intelligence'

According to mapping the content to be decomposed into sub-query:

DataSource1: select name, id, time from Course, eacher

Where name='artificial intelligence' and course.lecture=Teacher.TeacherID

DataSource2: select xm, number, xs from kc, Js

Where kcmc='artificial intelligence' and kc.skz=Js.bh

After several tests show that this design method can solve the heterogeneous database semantic heterogeneity.

IV. CONCLUSION

In this paper, ontology technology and meta-data of distributed, heterogeneous data source specification to solve the heterogeneous information sharing between. An example test demonstrate that the proposed ontology-based metadata and heterogeneous data sharing method. At present, the address data sharing in heterogeneous semantic heterogeneity needs to be done a lot of exploration and research. With the Semantic Web technology continues to mature, "body" will solve the semantic heterogeneity heterogeneous data integration problems are more and more applications.

REFERENCES

- [1] Amann B, Beeri C, Fundulaki I. Querying XML Sources Using Ontology Based Mediator. Proceedings of Coop IS/ DOA/ODBASE. Springer- Verlag Berlin Heidelberg, 2003, PP: 429- 448
- [2] T.R.Gruber.Towards Principles for the Design of Ontology Used for Knowledge Sharing. Proceeding of the International Workshop on Formal Ontology. Padova, Italy, 2008, PP: 50-66
- [3] LI Xing-yi, GAO Wen-hao, SHI Hua-ji. Heterogeneous Databases Integration Approach Based on Ontology [J].Computer Engineering and Design, 2009, 30(8), PP: 1931-1934
- [4] ZHANG Yu, JIANG Dong-xing, LIU Qi-xin. Metadata-Based Integration Scheme for Heterogeneous Datasets [J]. J Tsinghua Univ (Sci &Tech), 2009, 49(7), PP: 1021-1022
- [5] SONG Wei, ZHANG Ming. Concise Guide to the Semantic Web [M]. Beijing: Higher Education Press, 2004, PP: 56-60
- [6] WANG Li, GAO Zhong-Li. Web Services-based integration of heterogeneous data sources [C]. Proceedings Third Conference on Intelligent Computing in China, Jinan, 2009, PP: 79-80
- [7] Sheth.Changing focus on interoperability in information system: Form system, syntax, structure to semantics [C].Interoperating Geographic Information Systems, Boston: Kluwer Academic Publisher, 1999, PP: 13-14

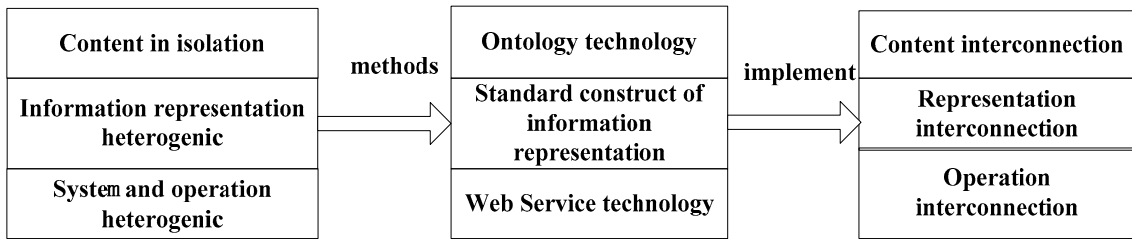


Figure.1 Basic idea model of Heterogeneous resources sharing

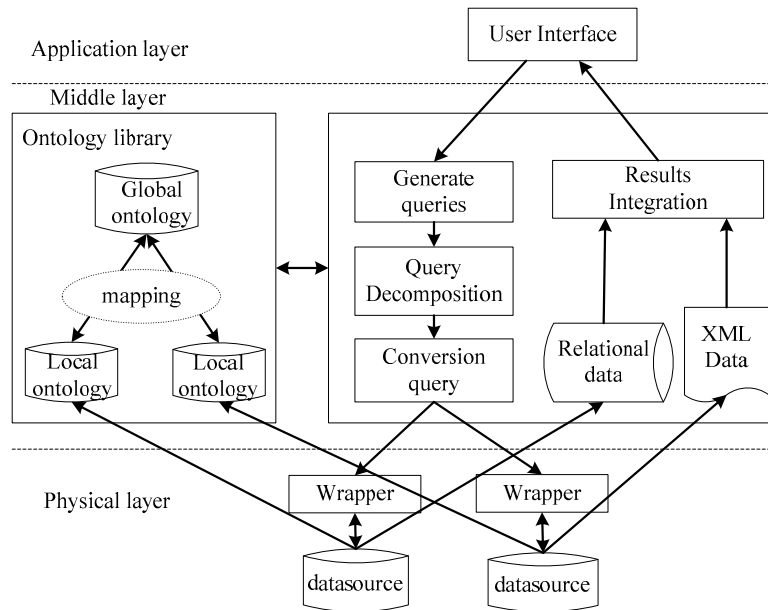


Figure.2 System framework

TABLE I. RELATIONSHIP BETWEEN CONCEPTS

Relation name	Describe the relation	Example
part-of	Part and whole	Tire and Automotive
kind-of	Inheritance	Transport and car
instance-of	Instance and concepts	Santana 2000 and car
attribute-of	Attribute	Cars and car color