WUJNS
Wuhan University Journal of Natural Sciences

Vol.11 No.3 2006 525-528

Article ID: 1007-1202(2006)03-0525-04

Semantic Query Mechanism on Peer to Peer XML-Based Information Systems

☐ **JIANG Min[1,2], HUANG Yi[1], GU Jinguang[2], HE Guiming[3]†**

1. School of Computer, Wuhan University, Wuhan 430072, Hubei, China;
2. College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China;
3. School of Electrical Engineering, Wuhan University, Wuhan 430072, Hubei, China

Abstract: To enable accessing web information at semantic level, this paper develops a semantic query rewriting mechanism on peer to peer XISs with complex ontology mapping technology. It discusses the patterns of complex ontology mappings at first, and the ontology-based query mechanism in peer to peer environment. The extension of XML query algebra and XML query rewriting mechanism are dissussed in detail.

Key words: ontology; query rewriting; P2P; XML-based information systems

CLC number: TP 391

0 Introduction

Due to its ability to express semi-structured information, XML based information system (XIS) plays a key role in web information community, and XML is rapidly becoming a language of choice to express, store and query information on the web. Problems that might arise due to heterogeneity of the data are already well known within the distributed database systems community: structural heterogeneity and semantic heterogeneity, semantic heterogeneity considers the content of an information item and its intended meaning[1]. How to access distributed information with a consistent semantic environment and how to make the XML query mechanism with semantic enabled are the problems should discussed.

The use of ontologies for the explication of implicit and hidden knowledge is a possible approach to overcome the problem of semantic heterogeneity. Ontologies can be used to describe the semantics of the XIS sources and to make the content explicit. With respect to the data sources, they can be used for the identification and association of semantically corresponding information concepts.

Two major bottlenecks exist in the area semantic query answering on XISs. One is ontology mapping discovery, which means how to find the similarities between two given ontologies, determine which concepts and properties represent similar notions, and so on. Many technologies, such as heuristics-based, machine learning based[2] or Bayesian network based[3] methods, were discussed in recent years, a survey of ontology mapping is discussed in Ref. [4]. The other is how to use ontology mapping technologies to support semantic XML querying, few research has been focus on this topic.

This paper focuses on how to enable semantic level quer-

Received date: 2005-09-01

Foundation item: Supported by the Natural Science Foundation of Hubei Province(2005ABA235) and Key Project of Science Research of Education Agency of Hubei Province (Z200511005)

Biography: JIANG Min(1975-), female, Ph.D. candidate, Lecturer of Wuhan University of Science and Technology, research direction: distributed computing system, computer vision. E-mail: jiangminwust@gmail.com

† To whom correspondence should be addressed. E-mail: gmhe@whu.edu.cn

ying on peer to peer XISs with ontology technologies. It uses ontology mapping technology to get a consistent semantic environment, and extends the XML querying to enable semantic querying on XISs.

1 The Peer to Peer XML-Based Information Systems

In the distributed peer to peer environment, every local site contains the local ontology based structured or semi-structured information source, the information source maybe the relational database, native XML database, web sites, XML based applications or other autonomous systems. From the opinion of web based or semi-structured information processing, all the local information sites can be expressed as the collections of XML instances. An XML based information system (XIS) can be denoted as $S=(K, W)$, where K is a finite set of XML instances, W is the ontology based wrapper which is described in the following subsection in detail. A peer to peer XML based information system can be denoted as $P=(\{S_i\}_{i\in I}, L)$, where I is a set of sites, S_i is an XIS for any $i\in I$, L is a symmetric, binary relation on the set I. Ontologies are used for the explicit description of the information source semantics, a detail introduction about the ontology based XIS has been described in Ref. [5].

2 The Patterns of Ontology Mapping

2.1 The Definition of Ontology Mapping Based on the Semantic Similarity

The patterns of ontology mapping can be categorized into four expressions: direct mapping, subsumption mapping, composition mapping and decomposition mapping[6], a mapping can be defined as:

Definition 1 A Ontology mapping is a structure $M:=(S, D, R, v)$, where S denotes the concepts of source ontology, D denotes the concepts of target ontology, R denotes the relation of the mapping and v denotes the confidence value of the mapping, $0\leqslant v\leqslant 1$.

A direct mapping relates ontology concepts in peer to peer environment directly, and the cardinality of direct mapping could be one-to-one. A subsumption mapping is used to denote concept inclusion relation especially in the multiple IS-A inclusion hierarchy. The composition mapping is used to map one concept to combined concepts. For example, the mapping address = contact (country, state, city, street, postcode) is a composition mapping, in which the concept address is mapped to combined concept "contact, country, state, street, and postcode" of local schema elements. The decomposition mapping is used to map a combined concept to one local concept. These four mapping patterns can be described in Ref. [7].

A subsumption ontology mapping is a 6-tuple $S_m:=(D_m, R_m, B_m, \leq_m, I_m, v_m)$, where: D_m is a direct mapping expression, which maps one source ontology concept to (possibly) many target ontology concepts; R_m is the first target concept, which is the most specialized ontology concept; the mapping between the source ontology and R_m is denoted as Root ontology concept mapping; B_m is the last target concept, which is the most generalized ontology concept; the mapping between the source ontology and B_m is denoted as Bottom ontology concept mapping; $\leq_m$ is inclusion relation between target ontology concepts with the expression $C_i \sqsubseteq \forall \text{Subsume}.C_{i+1}$, where the ontology concept C_i is included in C_{i+1} and i is a positive integer; I_m is the inverse mapping with the mapping expression $C_i \equiv \exists \text{Subsume}^{-}.C_{i+1}$, where the inverse of ontology concept C_i is C_{i+1}, and i is a positive integer; v_m is the confidence value of the mapping.

A composition ontology mapping is a 4-tuple $C_m:=(F_m, A_m, B_m, v_m)$ where: F_m is a direct mapping expression, which maps one source ontology concept to the first target ontology concept, which is the first node of chaining target role(s); A_m is chaining of role(s) between target ontology concepts with the expression $C_i \sqsubseteq \forall \text{Associate}_{i+1}.C_{i+1}$, where the target concepts are chained in order; B_m is the last target symbol, which is the node of chaining target role(s); v_m is the confidence value of the mapping.

The decomposition ontology mapping $D_m:=(A_m, B_m, L_m, v_m)$ is the reverse of composition ontology mapping.

2.2 The Properties of Semantic Mapping

This section defines some properties of semantic mapping which are useful in the task of querying. The first property is transitivity, for the mapping $M_{i-1,i}=(C_{i-1}, C_i, R, v_{i-1,i})$, $M_{i,i+1}=(C_i, C_{i+1}, R, v_{i,i+1})$, a new mapping $M_{i-1,i+1}=(C_{i-1}, C_{i+1}, R, v_{i-1,i+1})$ can be created to satisfy the mapping relation R. The second property is symmetric, which means that the mapping $M=(S, D, R, v)$ is equal to the mapping $M'=(D, S, R, v)$. The third property is strong mapping property, it can be described as follows.

Definition 2 A set of mappings M_i $(1 \leqslant i \leqslant n)$ are strong if they can satisfy the following conditions: They share the same mapping relation R, and the mapping relation is transitivity; For $\forall_{i,j,k}, v_i, v_j, v_k$ are the confidence value of mapping M_i, M_j, M_k, then $v_i \leqslant v_j + v_k$.

3 XML Query Rewriting with the Extension of Ontology Mapping

3.1 The Extension of XML Algebra with Semantic Query Enhanced

This paper extends XML algebra such as TAX[8] and OrientX[9] to enable semantic querying on peer to peer XISs, TAX uses Pattern Tree to describe query language and Witness Tree to describe the result instances which satisfy the pattern tree. Pattern tree is redefined for the purpose of optimizing result construction, and is renamed as Source Pattern Tree or Constructor Pattern Tree[9]. The definition of pattern tree with ontology extension can be described as follows:

Definition 3 A Ontology Enhanced Pattern Tree is a 2-tuple $Q := (T, F)$, where $T := (V, E)$ is a tree with node identifier and edge identifier. F is a combination of prediction expressions.

The prediction expression F supports the following atomic condition or selection condition[10]. Atomic condition has the form of $X o Y$, where: $o \in \{=, \neq, <, \leqslant, >, \geqslant, \sim$, instance of, isa, is part of, before, below, above$\}$; X and Y are conditional terms, which are attributes, types, type values $v:\tau$ and $v \in \mathrm{dom}(\tau)$, ontology concepts and so on; $\sim$ stands for the estimation of semantic similarity.

The selection condition is: ① Atom conditions are selection conditions; ② If c_1 and c_2 are selection conditions, then $c_1 \wedge c_2$, $c_1 \vee c_2$ and $\neg c_1$ are both selection conditions; ③ No others selection conditions forms.

3.2 XML Query Rewriting

In order to simplify the discussion, this paper just pays attention to the rewriting mechanism of the selection operation. A selection operation can be expressed as $\sigma_{P_i,P_o,P_e}(X) = \{x \mid x \prec X, P_o(x), P_e(x)\}$, P_i is input pattern tree, P_o is output pattern tree, P_e is predication list. Briefly, it can be expressed as $\sigma(X, Y)$, $\{X(P_i(P_o, Y(P_e\}$, operator $\bowtie$ and $\bigcup$ represent Union and Join operation respectively.

Firstly, for rewriting pattern tree (which is the X element of expression $\sigma(X, Y)$), there maybe several cases as follows:

① X is one of the elements of input pattern tree or output pattern tree, and it is also a concept in VMT table of local ontology based wrapper. X_i $(1 \leqslant i \leqslant n)$ are the concepts for different local ontologies. X and X_i are combined into one concept with strong direct mappings, which means that X and X_i can match each other, and then rewrites X as $X \bigcup \bigcup\limits_{1 \leqslant i \leqslant n} X$;

② The concept of X is generated by the subsumption mapping or composition mapping of X_i $(1 \leqslant i \leqslant n)$, then rewrites X as $\bigcup\limits_{1 \leqslant i \leqslant n} X_i$;

The responding selection rewriting can be expressed as:

$$\sigma(X_1 \bigcup X_2, Y) = \sigma(X_1, Y) \bigcup \sigma(X_2, Y) \qquad (1)$$

And then, for rewriting the predication expressions (which is the Y element of the expression $\sigma(X, Y)$), there are also several cases, which can be described as follows:

① If there are lots of concept Y_i $(1 \leqslant i \leqslant n)$ combined in the concept Y in VMT table, rewrites Y as $Y \bigcup \bigcup\limits_{1 \leqslant i \leqslant n} Y_i$;

② If the concept Y is generated by the subsumption mapping of Y_i $(1 \leqslant i \leqslant n)$, rewrites Y as $\bigcup\limits_{1 \leqslant i \leqslant n} Y_i$;

③ If the concept Y is generated by the composition mapping of Y_i $(1 \leqslant i \leqslant n)$, supposes the composition condition is F, rewrites Y as $(Y_1 + Y_2 + \cdots + Y_n) \bigcap F$.

Accordingly, the corresponding selection rewriting can be described as the following expression:

$$\sigma(X, Y_1 \bigcup Y_2) = \sigma(X, Y_1) \bigcup \sigma(X, Y_2) \qquad (2)$$

$$\sigma(X, (Y_1 + Y_2) \bigcap F) = \sigma(X, Y_1 \wedge F) \bowtie \sigma(X, Y_2 \wedge F) \qquad (3)$$

It is worth to point out that rewriting process maybe a recursion for the transitivity property of semantic mapping. The process of rewriting pattern trees and predication expressions can be described as Algorithm 1 and 2.

Algorithm 1: SELRewriteX(X)

Input: X is the pattern tree of selection query $\sigma(X, Y)$

```
for each $x \in X$ switch Mappings of $X$ node do
    case funsion-node:
    $x \leftarrow x \bigcup \bigcup_{1 \leqslant i \leqslant n} x_i$
    for each $x_i$ SELRewriteX($x_i$)
    case subsumption or composition:
    $x \leftarrow \bigcup_{1 \leqslant i \leqslant n} x_i$
```

```
    for each $x_i$ SELRewriteX($x_i$)
end case
end for
```

Algorithm 2: SELRewriteY(Y)

```
    Input: $Y$ is the predication list of selection query $\sigma(X, Y)$
for each $y \in Y$
switch Mappings of $Y$ node do
    case funsion-node:
    $y \leftarrow y \cup \bigcup_{1 \leqslant i \leqslant n} y_i$
    for each $y_i$ SELRewriteY($y_i$)
    case subsumption:
    $y \leftarrow \bigcup_{1 \leqslant i \leqslant n} y_i$
    for each $y_i$ SELRewriteY($y_i$)
    case composition:
    $y \leftarrow (y_1 + y_2 + \cdots + y_n) \cap F$
    for each $y_i$ SELRewriteY($y_i$)
end case
end for
```

An selection querying $\sigma(X_1 \cup X_2, Y)$ is redundancy if it satisfies

$$\exists (i,j)\{X_i \in P_o \wedge X_j \in P_o \wedge X_i \cap X_j \neq \varnothing\} \tag{4}$$

then the corresponding rewriting of selection can be described as:

$$\sigma(X_1 \cup X_2, Y) = \sigma(X_1, Y) \cup \sigma(X_2 - (X_1 \cap X_2), Y) \tag{5}$$

The advantage of complex ontology mapping with semantic similarity enhanced can be expressed as follows:

① It can match the semantic similar concepts more exactly, especially for the concepts which are part of concept hierarchy; ② It can reduce the semantic inconsistent by solving problem semantic absent; ③ It can reduce the redundancy of the querying by finding more semantic matching in subsumption and composition (decomposition) mappings; ④ The complex mapping mechanism refines the process of querying, and it makes the result more precisely.

4 Conclusion

The paper mainly discusses the extending querying on peer to peer XISs with wrapped ontologies. It discusses the complex ontology mapping patterns with semantic similarity enhanced, such as subsumption mapping, composition mapping and so forth. It also discusses the semantic query mechanism, which primarily extends XML query algebra based on TAX, on the XISs wrapped with local ontologies. Because common XML query languages such as XQuery and XUpdate can be transferred into XML query algebra based on TAX, so the extension is feasible. Complex ontology mapping ensures distributed querying can solve the problem of the inconsistency of semantic and increases the efficiency by refining on the querying and reducing redundancy.

References

[1] Wache H, Vgele T, Visser U, *et al.* Ontology-based Integration of Information a Survey of Existing Approaches[C]// *Proceedings of IJCAI-01 Workshops: Ontologies and Information Sharing*. Seattle, WA: Springer, 2001:108-117.

[2] Doan A, Madhavan J, Domingos P, *et al.* Learning to Map between Ontologies on the Semantic Web[C]// *Proceedings of the Eleventh International Conference on World Wide Web*. USA:ACM Press, 2002: 662-673.

[3] Pan R, Ding Z, Yu Y, *et al.* A Bayesian Network Approach to Ontology Mapping[C]// *Proceedings of the 4th International Semantic Web Conference*. Ireland: Springer, 2005:1-15.

[4] Kalfoglou Y, Schorlemmer M. Ontology Mapping: The State of the Art [J]. *The Knowledge Engineering Review*, 2003, **18**(1):1-31.

[5] Gu Jinguang, Chen Heping, Chen Xinmeng. An Ontology-based Representation Architecture of Unstructured Information [J]. *Wuhan University Journal of Natural Science*, 2004,**9**(5): 595-600.

[6] Kwon J, Jeong D, Lee L S. Intelligent Semantic Concept Mapping for Semantic Query Rewriting/Optimization in Ontology-based Information Integration System [J]. *International Journal of Software Engineering and Knowledge Engineering*, 2004,**14**(5):519-542.

[7] Gu Jinguang. *Research on the Mechanisms of the Semantic based Semi-Structured Data Processing*:[D]. Wuhan: Wuhan University, 2005.

[8] Jagadish H V, Lakshmanan L V S, Srivastava D, *et al.* TAX: A Tree Algebra for XML[J]. *Lecture Notes in Computer Science*, 2001,**2379**:149-164.

[9] Meng X, Luo D, Jiang Y, *et al.* OrientXA: An Effective XQuery Algebra [J]. *Journal of Software*, 2004,**15**(11): 1648-1660 (Ch).

[10] Hung E, Deng Y, Subrahmanian V S. An Extension of TAX with Ontologies and Similarity Queries[C]// *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*. Paris:ACM Press, 2004:719-730.

□