2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing

Ontology Based Heterogeneous Data Integration Framework Facing Mobile Environment

Chuanwei Xu[1,2], Shuo Yang[1], Jiting Huang[1], Changqing Ji[1,3 *]
1. College of Physical Science and Technology, Dalian University, Dalian 116622,China
2. College of Navigation, Dalian Maritime University, Dalian 116026, China
3. College of Information Science and Technology, Dalian Maritime University, Dalian, 116026, China
E-mail: jcqgood@gmail.com

***Abstract*—In the context of cloud computing, Mobile Data Integration Platform is put forward, on this basis, we first study on SORE (Simplied Object Repository) system of heterogeneous data integration based on CORBA. Furthermore, by virtue of the design of Heterogeneous Data Source Wrapper, this paper introduces an ontology based distributed information integration framework(OIIF), it implements uniform semantic query on heterogeneous data and enhances the depth and scope of available information. OIIF is innovation, its study has practical significance.**

Keywords- Data Integration; Mobile Cloud Computing; Ontology

I. INTRODUCTION

Cloud computing is "ubiquitous computing" a specific application [1], the transition from the mainframe era to the personal computer makes the computer into the homes of ordinary people, and the emergence of the internet has dramatically changed the calculation mode , but did not change the pattern of computer. Today, mobile devices are becoming smaller and smaller and we are seeing that there is either a physical or economic limit to the amount of storage and processing power that can fit into these devices. It seems that the original model of client-mainframe computing may be a good answer for this situation. However, we can now utilize existing wireless networks to connect mobile devices to servers in massive datacenters, rather than hardwiring all clients to a server. Radio-Frequency Identification (RFID) technology possesses a promising prospect in many practical applications [2]. This idea of connecting to unseen data may be where the term “cloud” came from, since it seems that the extra power is coming out of nowhere. In essence, cloud computing contains three main aspects: firstly, it focuses on how people view computing devices, how these devices can be integrated with the environment to complete the tasks; secondly, it focuses on how to design and layout calculation application to accomplish these tasks; thirdly, it concerns for the environment, as well as how to use the new technologies and features to improve access capacity under environmental services. People obtain these services through the traditional computers, mobile devices, and even the embedded chips in the network (these devices with computing power as chip technology advances have become smaller and smaller trends). With the support of the network, cloud computing environment is that of some combination of mobile, fixed and embedded chips, and that of mutual collaboration to accomplish computing tasks, these three parts constitute a people-centered, task-oriented computing environment.

Cloud Computing is becoming one of the next emerging IT industry technologies, with the rapid development of Internet and information technology. So a great number of distributed heterogeneous data sources appear. It is quite necessary to query and manage these distributed heterogeneous data sources uniformly. Structural heterogeneity means different data models and semantic heterogeneity means terms to be exchanged have different meanings. Semantic conflicts will appear when two information systems use different interpretations for the same information. Goh identifies three main causes for semantic heterogeneity: confounding conflicts, scaling conflicts and naming conflicts [3]. A solution to the problems of semantic heterogeneity should equip heterogeneous and autonomous software systems with the ability to share and exchange information in a semantically consistent way [4]. The development of ontology and its application techniques make the concept based information integration on heterogeneous data sources and implementation of semantic interoperability between different information systems possible. The contributions of the paper are reflected in the introduction of SORE system and the proposal of OIIF. The paper experimentally demonstrates it deponding on tourist information services data sets.

II. HETEROGENEOUS DATA INTEGRATION UNDER MOBILE CLOUD COMPUTING

A. *Characteristic analysis of mobile environment*

In mobile environment, we do not have enough resource like PC: the capacity of PDA is so small that it cannot save too much data. We must treat the system design and choices of ways with caution for the limitation in the mobile environment. There are some challenges in this situation:

- Resource limitation. Either the mobile equipment itself or its CPU disposal rate, or wireless network bandwidth, is much less than normal PC and wire network. So it is hoped that the smart entity is able to be conscious of resource.
- Communication cost. The wireless network is charged by flux and it is too expensive. After all, the wireless network bandwidth is limited resource. As a result, it

* Changqing Ji is corresponding author.

978-1-4799-5079-9/14 $31.00 © 2014 IEEE
DOI 10.1109/DASC.2014.72

will not be changed a lot after the 3G network is prevalence.

- Visualization and Human-Computer interaction.
- Mobile and collaboration put forward some new questions to data management. Wireless network and mobile equipment need to keep connection between users and data service while the position is changing, which is helpful to guarantee the data interaction coherence. In some applications, it is required that mobile agents possess geography position awareness [5].

B. *Mobile Data Integration Platform*

- User of handhold equipment sends requirement to system. Through Wireless Markup Language (WML) parser, the requirement will be disposed. The detail of WML parser will be introduced in the next section.
- Terminal recognition module judges the terminal type through WML requirement head. In the WML requirement head, the User-Agent (terminal description information provided by user terminal provider) and Accept (MIME document that can be accepted by user terminal) are related to terminal type.
- Call on model data—XML virtual database.
- Underlying data can be integrated to virtual database based on XML through forgoing heterogeneous data integration architecture. Fig. 1 describes Mobile Data Integration Platform.

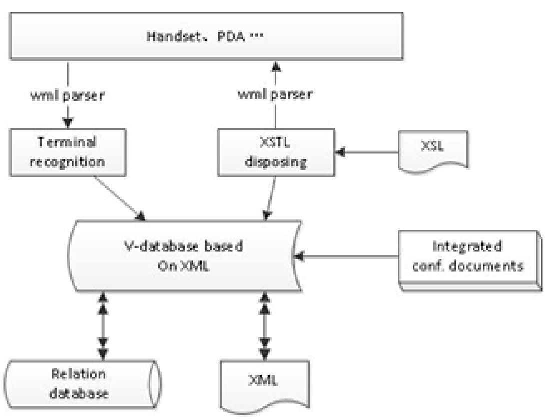

Fig.1. Mobile Data Integration Platform

III. SORE HETEROGENEOUS DATA INTEGRATION SYSTEM

SORE (Simplified Object Repository) is an intelligent heterogeneous data integration system based on CORBA. In the SORE local heterogeneous data source is packed as CORBA object by CORBA package technology and UDM (Uniform Data Model) which is defined by our project group. Packed CORBA object can provide an outward uniform accessing interface and a standard data exchange pattern[6]. At the end of client, the global common model and data operation can only be visible for users. In the inner of integration, actual user data manipulation is decomposed into some sub-manipulations of each certain data source, and then the result of each sub-manipulation is summarized and returned to user by integration system. Fig. 2 describes SORE system integration framework based on soft component and bus of CORBA.

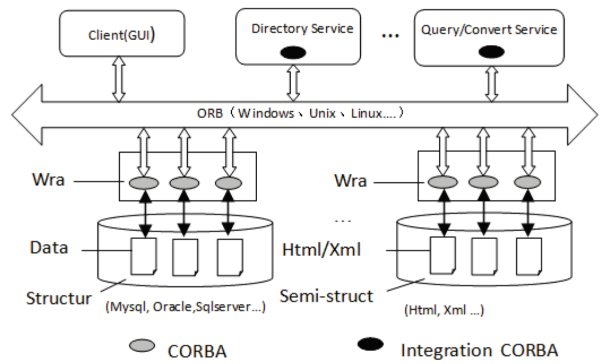

Fig.2. SORE SYSTEM INTEGRATION

SORE is an integration system on structural level and in this system the data model can not make its own meaning explicit. Therefore, it cannot implement the query on related information of data sources on conceptual and semantic level [7]. After the introduction of ontology, an ontology based distributed information integration framework which can implement semantic interoperability between heterogeneous information is presented by rebuilding the original system.

IV. ONTOLOGY BASED DISTRIBUTED INFORMATION INTEGRATION FRAMEWORK (OIIF)

OIIF faces some specific domain applications. Fig. 3 describes the ontology based distributed information integration framework. Reasoner module, Metadata Service module and Query Processor module are all connected with CORBA ORB bus.

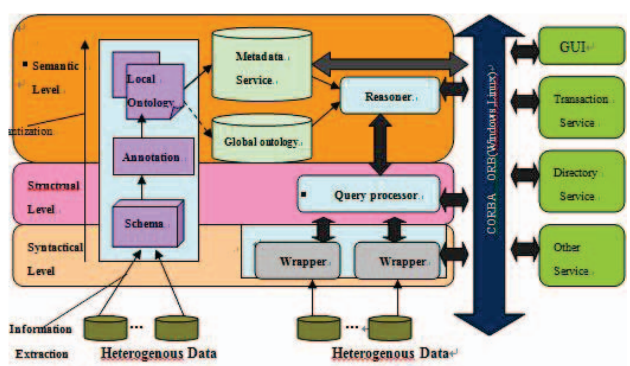

Fig.3. Architecture of OIIF

This framework has three levels: syntactical level, structural level and semantic level. On syntactical level and structural level, information is expressed uniformly and structurally using UMD model [8]. On semantic level, concept information is expressed uniformly using OWL (Ontology Web Language) which is W3C specification. In this framework, hybrid ontologies approach is employed and the semantic information of each data source is described by its own local ontology. In order to make the local ontologies compare with each other, a global shared vocabulary is established. Also it is a global ontology library [9].

Global ontology library: It is a global shared vocabulary of the specific application domain built in advance with the participation of domain experts. The precise definition and detailed declaration of domain ontologies should be supplied. The shared vocabulary includes basic concept terms and subsumption relationships about a domain. As expansion of local ontologies, it deals with the semantic heterogeneity between different local ontologies. At the same time, it satisfies the query need on different local ontologies. There is no period after the "et" in the Latin abbreviation "et al."

Local ontologies: Local ontologies, which are built automatically or semi-automatically based on extracted pattern information, are local views of heterogeneous data sources on conceptual level. They map with corresponding data sources.

Information semantization: The method is to extract pattern information from data sources firstly, then build local ontologies automatically or semi-automatically with the help of domain experts [10]. During the building process, the terminology of global ontology should be referred. So all local ontologies accomplished refer to the same sharing vocabulary which make reasoning between different local ontology workable. The local ontology information are registered as metadata in metadata service module. This metadata information is described in OWL.

Reasoner: Reasoner is the most important part in OIIF. It mainly supplies service for reasoning between terminology and finding implicit knowledge information. A reasoning mechanism based on semantic similarity and semantic correlativity is employed. Semantic similarity is calculated using word distance in linguistics. The bigger the distance between concepts in a concept subsumption relationship is, the lower the semantic similarity is. In the contrary, the smaller the distance is, the higher the semantic similarity is. The correlativity is divided into two types: the explicit correlativity type and implicit correlativity type. The explicit correlativity type includes direct correlativity and indirect correlativity. The implicit correlativity type includes direct inheriting correlativity and direct inheriting correlativity. Indirect correlativity means two concepts correlate with each other through at least two relationships. In the same principle, direct inheriting correlativity means two concepts correlate with each other through at least two inheriting relations. The correlativity between two concepts can be calculated by gave a parameter to each relationship. Synthesizing the similarity and correlativity between concepts, an index can be obtained. By this index, the concepts concerned can be arranged.

Metadata service: The main function of this module is to store the ontology metadata corresponding with data sources [11]. When querying, it supply service for reasoner to load ontology metadata. Each local ontology registers in metadata service module in advance as metadata describing corresponding data source.

Semantic query process: User can choose global query or local query. Choosing concepts from ontology registered in metadata service, if it is a global query, reasoning load corresponding mapping information and corresponding ontology and get concepts according their semantic similarity and semantic correlativity by calculating. Then reasoner sends the results to query processor and the query processor transform concepts to UDM model. In succession, a query sentence rewrote by query processor is sent to wrapper for structural query. The results got at last are concepts and their instance data. If the query is just a local query, the reasoner loads only aiming local ontology. Other processing are alike with global query.

V. A CASE STUDY - DOMAIN ONTOLOGY MODEL FOR INTELLIGENT CITY UNDER MOBILE ENVIRONMENT

An ontology (O) organizes domain knowledge in terms of concepts (O), properties (P) and relations (R) and can be formally defined as follows [12].

Definition (Ontology) - An Ontology is a triplet of the form $O = (C, P, R)$, where:

- C is a set of concepts defined for the domain. A concept is often defined as a class in an ontology.
- P is a set of concept properties. A property $p \in P$ is defined as an instance of a ternary relation of the form $p(c, v, f)$, where $c \in C$ is an ontology concept, v is a property value associated with c and f. f defines restriction facets on v.
- $R = \{ r \mid r \subseteq C \times C \}$ is a set of binary semantic relations defined between concepts in O. A set of basic relations is defined as {synonym of, kind of (is a), part of (whole-part), instance of, property of}.

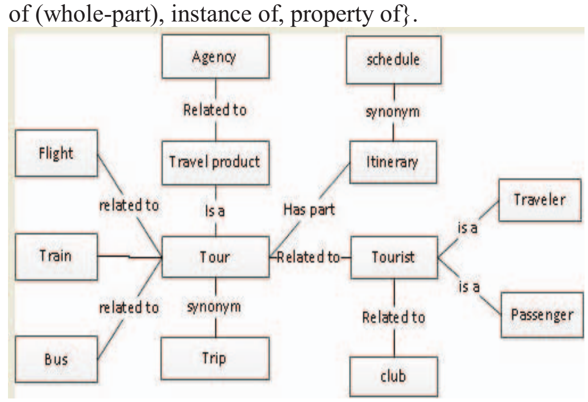

Fig.4. Tourist Information Services Domain Ontology

The individual information services are the vital part of the traveler information services, which assemble correlative information from multi-data-sources (weather bureau, news, agency etc.) for satisfying the personal needs of travelers. The typical application is the travel information afforded for the tourist, in which the partial travel domain ontology is shown in Fig. 4. The main concepts include tourist, agency, trip, itinerary etc. The concepts can be related in various ways. The four most commonly used are synonym (equivalent), is-a (generalization), whole-part (has part) and related-to.

In OWL the domain ontology is represented as follows:

<owl:Class rdf:ID= “Tourist”/>
<owl:Class rdf:ID= “Traveler”>
<rdfs:SubClassOf rdf:resource= “#Tourist”/>
</owl:Class>
<owl:Class rdf:ID= “Passenger”>
<rdfs:SubClassOf rdf:resource= “#Tourist”/>
</owl:Class>
<owl:Class rdf:ID= “Trip”>
<owl:Class rdf:ID= “Tour”>
<owl:equivalentClass rdf: resource= “#Trip”>
</owl: class>

VI. SEMANTIC INFORMATION RETRIEVAL

Ontology, with the excellent concept hierarchy and appropriately supporting for logic reasoning, is used widely in information retrieval, especially in the knowledge retrieval. For the emphasizing of matching based on knowledge and semantic, semantic retrieval has good performance in recall and precision [13].

A. Information retrieval based on "synonymy" relation

"Differing terms for the same concepts (synonyms)" is a common phenomenon[14]. The traditional retrieval technology omits the same semantic information, but if we use the relation of synonyms in ontology, recall ratio can be improved. For example, we search tour’s information, using the query statement as follows:

SELECT tour FROM Web data source

Using the synonymy relation, the expansion of the statement is as follows:

SELECT tour FROM Web data source OR SELECT trip FROM Web data source

B. Information retrieval based on "kind of" relation

Inheritance relation (i.e. kind of) that is the most important semantic relations means the inclusion relation between concepts. All the sub-classes should be searched out, when we search a father class. For example, when we search information about “Tourist”, the query statement is:

SELECT Tourist FROM Web data source

Using the inheritance relation, the expansion of the statement using is as follows:

SELECT TouristFROM Web data sourceOR

SELECT Traveler FROM Web data source OR

SELECT Passenger FROM Web data source

C. Information retrieval based on whole-part relation

Whole-part relation is one of the most semantic relations, too. All the parts of information should be searched out when we search information about whole entity. For example, when we search information about “Tour”, the query statement is:

SELECT Tour FROM Web data source

Using the whole-part relation, the expansion of the statement is as follows:

SELECT tour FROM Web data source OR

SELECT Itinerary FROM Web data source

VII. CONCLUSIONS

This paper introduces an ontology based distributed information integration framework (OIIF). Ontology based information integration framework resolves the problem of semantic interoperability between heterogeneous data sources on semantic level. It implements uniform semantic query on heterogeneous distributed information and enhances the depth and scope of information query greatly. Mobile informationization is the trend of enterprise constructing informationization. But some difficulties exist all the same such as how to build or obtain the global ontology about some domain, how to implements automatic information extraction and ontology building completely without manual work, etc. In the future, deep research work on the problems above will be done.

ACKNOWLEDGMENT

This work was supported in part by the Science Technology Foundation of Dalian under No. 2013J21DW001, National Science Foundation of China under grant nos. of 61370199, Natural Science Foundation of Liaoning Province of China under grant nos. of 201400154 and 201405, Project of College Students' Innovative and Entrepreneurial Training Program under grant nos. of 201411258048, 201411258010, Project of College Students' Innovative and Entrepreneurial Training Program of Dalian University of China under grant nos. of 2013080, 2013081, 2013082 and 2013259, and the Teaching-reform Project of Dalian University of China under Grant nos. of 2013122G1 and 2013123G1.

REFERENCES

[1] I. Foster, Y. Zhao, I. Raicu, and S. Lu, “Cloud Computing and Grid Computing 360-Degree Compared”, IEEE Grid Computing Environments Workshop, 2008, pp. 1-10

[2] Xiulong Liu, Keqiu Li, Senior Member, IEEE, Geyong Min, Kai Lin, Bin Xiao, Yanming Shen, and Wenyu Qu, Efficient Unknown Tag Identification Protocols in Large-Scale RFID Systems, To appear in IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 2013.

[3] Isabel F. Cruz, Huiyong Xiao and Feihong Hsu, “An Ontology-based Framework for XML Semantic Integration,” Proceedings of the International Database Engineering and Applications Symposium (IDEAS’04), 2004, pp.50-58.

[4] H Cui Zhan and Paul O’Brien, “Domain Ontology Management Environment,” Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000, pp.1-9.

[5] Yang Li, Jun Zhai and Yan Chen, “Using Ontology to Achieve the Semantic Integration of the Intelligent Transport System,” Proceedings of 2005 International Conference on Management Science & Engineering (12th) (Volume III), 2005, pp. 2528-2532.

[6] J. Javier Samper, Vicente R. Tomás, J. José Martinez and Leo van den Berg,“An Ontological Infrastructure for Traveller Information Systems,” in Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference, (Toronto, Canada), 2006, pp. 1197-1202.

[7] J. Li, J. Y. Song, H. Zhong, “Ontology-based query division and reformulation for heterogeneous information integration,” Journal of Software, vol. 18, no. 10, pp. 2495í2506, 2007. (in Chinese)

[8] Avancha S, Patel C, Joshi A. Ontology-Driven Adaptive Sensor Networks[C]. 1st Annual International Conference on Mobile and UbiquitousSystems：Networking and Services, Boston, 2004: 194-202

[9] J. Javier Samper, Vicente R. Tomás, J. José Martinez and Leo van den Berg,“An Ontological Infrastructure for Traveller Information Systems,” in Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference, (Toronto, Canada), 2006, pp. 1197-1202

[10] M. Wilson, “Constructing and Managing Appliances for Deployments from Repositories of Reusable Components”, Workshop on Hot Topics in Cloud Computing, 2009

[11] GruberT R:A Translation Approach to Portable Ontology Specifications[M]. Knowledge Acquisition,2009,5:199~220

[12] Yang Li, Jun Zhai and Yan Chen, “Using Ontology to Achieve the Semantic Integration of the Intelligent Transport System,” Proceedings of 2005 International Conference on Management Science & Engineering(12th)(VolumeIII),2005,pp.2528-2532

[13] J. Javier Samper, Vicente R. Tomás, J. José Martinez and Leo van den Berg,“An Ontological Infrastructure for Traveller Information Systems,” in Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference, (Toronto, Canada), 2006, pp. 1197-1202

[14] P. Castells, M. Fernandez, D. Vallet, “An adaptation of the vector-space model for ontology-based information retrieval,” IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 2, pp. 261-272, 2007.