TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 08/18 pp196-205
Volume 14, Number 2, April 2009

Semantic Extraction for Multi-Enterprise Business Collaboration*

SUN Hongjun (孙红俊), FAN Yushun (范玉顺)**

Department of Automation, Tsinghua University, Beijing 100084, China

Abstract: Semantic extraction is essential for semantic interoperability in multi-enterprise business collaboration environments. Although many studies on semantic extraction have been carried out, few have focused on how to precisely and effectively extract semantics from multiple heterogeneous data schemas. This paper presents a semi-automatic semantic extraction method based on a neutral representation format (NRF) for acquiring semantics from heterogeneous data schemas. As a unified syntax-independent model, NRF removes all the contingencies of heterogeneous data schemas from the original data environment. Conceptual extraction and keyword extraction are used to acquire the semantics from the NRF. Conceptual extraction entails constructing a conceptual model, while keyword extraction seeks to obtain the metadata. An industrial case is given to validate the approach. This method has good extensibility and flexibility. The results show that the method provides simple, accurate, and effective semantic interoperability in multi-enterprise business collaboration environments.

Key words: semantic interoperability; semantic extraction; neutral representation format; business collaboration

Introduction

With the global economic integration and rapid development of IT, multi-enterprise business collaboration is becoming the mainstream of current enterprise cooperation[1]. Companies have recognized that they need effective frameworks to share semantics between partners and to use them within interoperating internal systems. Semantic interoperability for the sharing of digitalized knowledge has become essential for modern distributed applications[2]. To enable enterprises to fully participate in collaboration, more research is needed in semantic interoperability field of business information in multi-enterprise business collaboration environments[3].

Received: 2008-01-08; revised: 2008-10-13

* Supported by the National Natural Science Foundation of China (No. 60674080), the Europe Union Project of Software for Ambient Semantic Interoperable Services (FP6-2005-IST-5-034980), and the National High-Tech Research and Development (863) Program of China (Nos. 2006AA04Z166 and 2007AA04Z150)

** To whom correspondence should be addressed.
E-mail: fanyus@tsinghua.edu.cn; Tel: 86-10-62789650

In collaboration environments, companies sharing information resources distribute not only data, but also the schemata of data among multiple previously unknown users. Therefore, the exchange format needs to be understood by a broad community of users. Recently, web ontology language (OWL) has become the determinant standard of the international research community in this area[4]. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL description language (DL), and OWL Full. OWL DL is used to represent data and schema items, since it provides a good balance between knowledge representation and data quantity.

The transformation process from the data schema to the OWL DL ontologies is called semantic extraction. Semantic extraction refers to all the methodologies concerned with the extraction of semantics from a schema without using a pre-existent reference or domain ontology. Thus semantic extraction is the key to semantic interoperability in multi-enterprise business collaboration environments.

Most existing research in semantic extraction is focused on how to directly extract ontology from a specific schemata[5-7]. For example, Li and Du[5] gave an approach for automatically formulating OWL ontology from data in relational databases using a set of learning rules. Shen et al.[6] described groups of semantic mapping rules for extracting a global OWL ontology from relational database. They have also been proposed for XML schema[7]. But there is little work on extracting ontology from other data schema formats. Recently, approaches have been presented to extract ontologies for reverse engineering relational databases[8-11]. Lammari et al.[8] used reverse engineering techniques to identify the inherant links embedded in a relational database. Davies[9] migrated data-intensive web pages into the ontology-based semantic web. Benslimane and Malks[10] acquired OWL ontology corresponding to the content of relational database. The tool RDB2ONT[11] generates and publishes OWL ontologies dynamically from the metadata and structural constraints of relational database systems.

However, existing semantic extraction methods mainly directly extract semantics from specific data schema. The current semantic interoperability environment handles multiple heterogeneous data schemas, such as relational database (RDB) and XML. The ontology construction for each data schema is time-consuming and labor-intensive even with an ontology editor. Furthermore, each company can automatically decide to join or leave an information sharing environment at any time. Thus, the collaboration process within and across enterprises should be as fast as possible in distributed, heterogeneous computing environments. The semantic extraction efficiency needs to be improved to reduce time and avoid failure.

This paper describes a new semantic extraction method based on the neutral representation format (NRF). The goal is to automatically construct the OWL DL ontology. The semantic extraction process is divided into two steps. The first step integrates all the heterogeneous data schema into a common neutral representation format (NRF) through simple mapping rules. In the second step, the NRF is mapped to the OWL DL ontology using conceptual extraction and keyword extraction.

1 Semantic Extraction Approach

1.1 Preliminary definitions

The key definitions related to semantic interoperability in multi-enterprise business collaboration environments are presented for reference.

(1) **Semantic entity (SE)** An SE is an instance of some concept recognized by means of semantic extraction, which is associated with additional information (e.g., metadata, keywords, owner information).

(2) **Terminological semantic entity (SE_T)** An entity recognized in a semantic interoperability environment by semantic extraction, which corresponds to an ontology class. An SE_T corresponds to an OWL DL class.

(3) **NRF** A representational (non semantic) encoding of new elements entering the semantic interoperability environment as representive items in an external schema.

1.2 Semantic interoperability

Figure 1 illustrates how semantic interoperability obtains a link specification reflecting the semantic correspondences between two business documents from different organizations in a multi-enterprise business collaboration environment. In this scenario, the business document from one organization uses a relational schema, whereas the other organization describes their business document on an XML schema.

The whole semantic interoperability process between the two schemata can be achieved by the following four steps as follows. The same approach can be extended to multi-enterprise semantic interoperability.

(1) **Format transformation** The process transforms the two schemata to the NRF. Both schemata are expressed in the structural ontology format (RDB_OWL or XSD_OWL). Then a logical data model (LDM) defined in NRF is used to abstract the different data models from the various schemata.

(2) **Semantic extraction** In the narrow sense, semantic extraction is the process used to obtain semantic information from the NRF. It includes conceptual extraction and keyword extraction. Conceptual extraction entails extracting from the NRF the typical constructs of a conceptual model. The keyword extraction method is used to obtain metadata.

(3) **Semantic annotation** The main aim of semantic annotation is to provide a common framework for the integration of semantic information from the two schemata based on a common ontology. An automatic disambiguation technique is used to further refine the semantic information, which is the process of

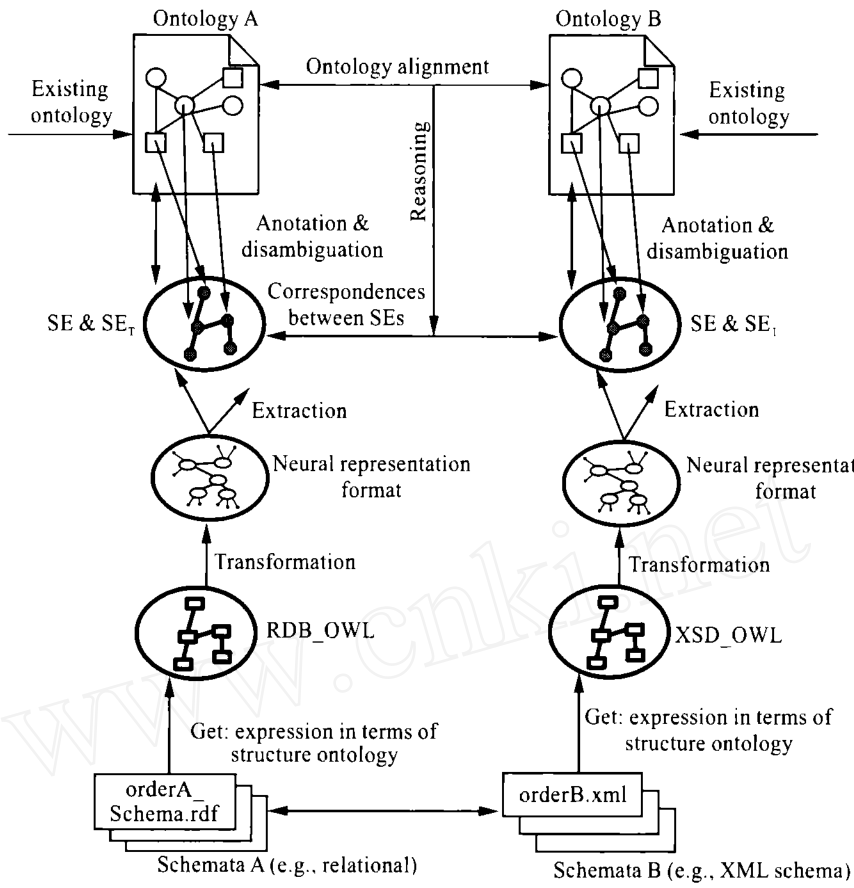


Fig. 1 Semantic interoperability for multi-enterprise collaboration

automatic lexical annotation.

(4) **Semantic mapping and ontology alignment** Once available, the SEs of one organization can be mapped to the SEs of the other organization. The two organizations are interested in the semantic correspondences between their SEs and could establish corresponding SE links to express them. Apart from direct SE mapping, indirect methods are also used to derive the SE correspondences through reasoning. In addition, possible discrepancies between the ontologies have to be reconciled, which is referred to as ontology alignment.

1.3 Semantic extraction method

Semantic extraction obtains semantic representations of a given schema with a high degree of automation to facilitate multi-enterprise business collaboration. The extraction process can be divided into the heterogeneous schemata format transformation step and the semi-automatic semantic extraction step.

The transformation step takes a syntax-specific data model and translates it into the NRF as a normalization step, to remove all the contingencies in the original environment. The next step extracts the semantic information from the neutral representation format of the external schema.

The semantics are extracted using keyword extraction and conceptual extraction. Conceptual extraction expresses relations among the resulting entities that are needed to relate classes in the OWL.

Reverse engineering is defined as the process of analyzing a system to identify all the system's components and the relationships between them[12]. This is used for conceptual extraction, and to establish the corresponding SE links in a semi-automatic extraction process. A set of rules is used for the mapping from the NRF to the conceptual schema. The keyword extraction is based on the RELEVANT method[13] which calculates a set of relevant values according to the designer selections given as an input list of attribute values. The relevant values are keywords that provide a synthesized view of the values of a node directly extracted from the data.

2 NRF Acquisition

2.1 Schema representation format based on OWL

Multi-enterprise business collaboration environments have a variety of heterogeneous data schemas. An OWL ontology is defined for each schema format which abstractly describes the format[3,14]. A schema in a given format can then be represented as instances of the related schema format OWL ontology. Thus, the representation does not extract any semantic from the schema, but represents a basic schema in the OWL format.

The schema representation format based on the OWL format is defined as a schema S of the given model M. The representation format based on OWL of schema S as an instance of the related schema format OWL ontology is called S_OWL. For example, the relational database schema can be mapped to RDB_OWL.

2.2 NRF

NRF is used to describe incoming elements into the semantic interoperability environment from an external schema. The NRF representation is an instance of an OWL ontology which contains generic concepts, i.e. classes and properties, to abstract the various input schemata from the different data models to obtain a unified representation. This ontology is called LDM_OWL, where LDM means logical data model.

2.2.1 LDM_OWL

The abstract of the syntactical aspects of a specific data model can be used to map data models to a unified representation, represented by the LDM. The LDM corresponds to a graph with directed labeled edges, which allows the representation of classes/concepts, relations, and attributes. Concepts are organized in the familiar "is-a" hierarchy. Relations are subject to constraints such as the specification of domain and range, plus cardinality constraints. Thus, from a representation point of view, the LDM model is comparable to a conceptual model[14].

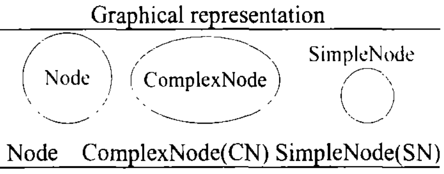
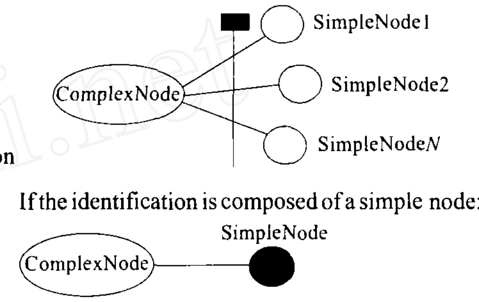
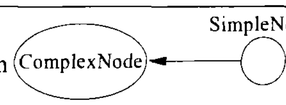
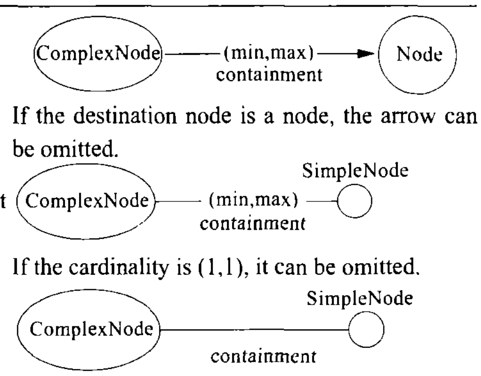
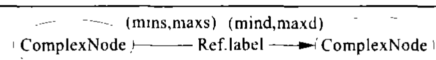
Here, the LDM is represented as an instance of OWL ontology called LDM_OWL. The approach by An et al.[15] achieves the transformation.

2.2.2 NRF definition

Given a model M, a schema S of the model M, and the structural ontology of M, the representation of the schema S as an instance of the LDM_OWL is called the neutral representation format of schema S.

NRF is a directed and labeled graph with several types of nodes and edges. A basic graphical representation is given for each element of the LDM model in Table 1.

Table 1 Graphical representation of the LDM model[16]

Concept	Graphical representation
Node (N)	
Identification	
Qualification	
Containment	
References (R)	

2.3 Acquisition process

The NRF acquisition has a "get" step and a "transformation" step. NRF is the ultimate result of the acquisition of information describing the external environment.

2.3.1 "Get" process

In the "Get" process, all heterogeneous schemas must be represented as instances of the related schema format OWL ontology. The input is a schema and the output is the encoding of this schema in OWL. Saw[7] described the representation approaches used here in detail for mapping each schema X to instances in the

OWL ontology (X_OWL) by analyzing their structures.

2.3.2 "Transformation" process

The transformation process targets various source schemata/models including RDB and XML-schemata with the target schema/model being the LDM_OWL. This process is characterized by simple 1-1 correspondences between the entities of an input schema and those of the output schema. The transformations are specified as a set of rules for mapping rdb:Table to ldm:ComplexNode; for mapping rdb:Column to ldm: SimpleNode; for mapping rdb:hasColumn to ldm: Containment; for mapping rdb:KeyConstraint to ldm: Identification; and for mapping rdb:ForeingKey to ldm: references[16]

3 Extraction

The keyword and conceptual extraction are used to obtain the semantics from the NRF. The keyword extraction result is SEs, while the conceptual extraction result is SE_Ts.

3.1 Conceptual extraction

Conceptual extraction entails extracting the typical constructs of a conceptual model from a schema (NRF), such as relationships and constraints. This is a semi-automatic semantic extraction process using reverse engineering techniques with the OWL DL ontology as the target, the NRF as the source.

3.1.1 NRF and OWL DL relationships

The NRF is the input to the resource of reverse engineering conceptual extraction, which results in a common representation of multiple data formats within the multi-enterprise collaboration environment. The result describes the ontology used to reflect the structure of various input schemata as well as the unified logical data model. The NRF does not need any semantic.

The output, also called the reverse engineering target, is an OWL DL ontology constituted by a set of concepts, a set of classes and properties, which is called a "conceptual schema".

3.1.2 Mapping rules

The correspondence between the NRF and the OWL DL is defined by a set of translation rules organized into four groups[6,17-19]:

(1) Rule for constructing classes

All ComplexNodes which satisfy the following rules are Classes, and all OWL DL Classes come from ComplexNodes.

Rule 1 For ldm:ComplexNode $cn_1, cn_2, \ldots, cn_i$, suppose that:

(a) idf_1 = ldm:Identification(cn_1),

(b) idf_2 = ldm:Identification(cn_2), …,

(c) idf_i = ldm:Identification(cn_i).

If $cn_1(idf_1)=cn_2(idf_2)=\cdots=cn_i(idf_i)$, then $cn_1, cn_2, \ldots, cn_i$ may be mapped to an ontological class c_i.

Rule 2 If Rule 1 cannot be satisfied, for ldm:ComplexNode cn_i, suppose that: idf_i = |ldm:Identification (cn_i)| =1, then, cn_i can be mapped to an ontology class c_i.

Rule 3 If Rule 1 cannot be satisfied, for ldm:ComplexNode cn_i, suppose that |ldm:Identification (cn_i)| > 1, and there exists idf_i, $idf_i \in$ ldm:Identification (cn_i), and $idf_i \notin$ ldm:Reference(cn_i). Then cn_i can be mapped to an ontology class c_i.

(2) Rule for constructing properties

Those ComplexNodes which indicate the associations between two other ComplexNodes can be mapped into an ontological ObjectProperty.

Rule 4 For ldm:ComplexNode cn_i, if |ldm:Indentification(cn_i)| $\geqslant 1$, then p_i ($p_i \in$ ldm:SimpleCode(cn_i)) is mapped to the property of class c_i (c_i is the corresponding class of cn_i).

Rule 5 If Rule 4 is satisfied, for ldm:ComplexNode cn_i, cn_j, if the following conditions are satisfied.

(a) |ldm:Reference(cn_i)| $\geqslant 1$,

(b) $cn_i(rfn_i) \subseteq cn_j(rfn_j)$,

($rfn_i \in$ ldm:Reference(cn_i), and $rfn_j \in$ ldm:Reference(cn_j)).

The Reference(s) can be mapped to the ObjectProperty op_i of class c_i, with the domain of op_i as c_i and the range of op_i as c_j (c_i and c_j are the corresponding classes of cn_i and cn_j).

Rule 6 If Rule 4 is satisfied, for ldm:ComplexNode cn_i, if

(a) p=ldm:SimpleCode(cn_i)−ldm:Indentification(cn_i)−ldm:Reference(cn_i)),

(b) $|p| \geqslant 1$.

$p_i (p_i \in p)$ can be mapped to DataTypeProperty dp_i of class c_i, with the domain of dp_i as c_i.

Rule 7 For LDM:ComplexNode cn_i, cn_j, if the following two conditions are satisfied:

(a) |ldm:Identification(cn_i)|>1,

(b) ldm:Reference(cn_i) $\subset$ ldm:Identification(cn_j),

where ldm:Reference(cn_i) refers to cn_j.

Then, two ontological object properties "has-part" and "is-part-of" can be created, where the domain and range of "is-part-of" are c_i and c_j, and the domain and range of "has-part" are c_j and c_i. "has-part" and "is-part-of" are two inverse properties (the classes c_i and c_j are correspond to cn_i and cn_j).

Rule 8 For LDM:ComplexNode cn_i, cn_j, and cn_k, if the conditions are satisfied:

(a) ldm:Identification(cn_i) $\cup$ ldm:Identification(cn_j)= ldm:Reference(cn_k),

(b) ldm:Identification(cn_i) $\cap$ ldm:Identification(cn_j)= $\varnothing$.

That is cn_k is related with cn_i and cn_j, ldm:Identification(cn_k) and ldm:Identification(cn_j) can be mapped to the ObjectProperties op_i and op_j with the domain of op_i as c_i and the range of op_i as c_j. The domain of op_j is c_j, and the range of op_j is c_i. op_i and op_j are the inverse of each other. (c_i and c_j are the corresponding classes of cn_i and cn_j, respectively).

(3) Rule for constructing hierarchies

In the NRF, a node can exist as a generalisation for one or more other nodes. The generalisation element is synonymous with the OWL:subClassOf construction.

Rule 9 For ldm:ComplexNode cn_i and cn_j, if the conditions are satisfied as follows:

(a) idf_i= ldm:Identification (cn_i),

(b) idf_j= ldm:Identification (cn_j),

(c) $cn_i(idf_i) \subseteq cn_j(idf_j)$.

The class/property corresponding to cn_i is a subclass/subproperty of the class/property corresponding to cn_j.

(4) Rule for constructing restrictions

In the OWL DL, the owl:allValueFrom constraint belongs to value constraints with cardinality constraints mainly including minCardinality and maxCardinality[4]. Furthermore, an OWL property can be globally declared as a functionalProperty or inverseFunctional.

Rule 10 For ldm:ComplexNode cn_i and cn_j, if the conditions are satisfied:

(a) |ldm:Indentification(cn_i)| $\geqslant$ 1,

(b) |ldm:Reference(cn_i)| $\geqslant$ 1,

(c) $cn_i(rfn_i) \subseteq cn_j(rfn_j)$.

($rfn_i \in$ ldm:Reference(cn_i), $rfn_j \in$ ldm:Reference(cn_j)),

Then the ObjectProperty op_i of class c_i has a restriction allValueFrom, which refers to the corresponding inclusion dependency class c_j (c_i and c_j are the corresponding classes of cn_i and cn_j)

Rule 11 For ldm:ComplexNode cn_i, if the following conditions are satisfied:

(a) A=ldm:Identification(cn_i) $\cup$ ldm:Reference(cn_i),

(b) $A \neq \varnothing$.

The restrictions minCardinality and maxCardinality of each property $p_i (p_i \in A)$ are both set equal to 1, or restrictions cardinality of each is set equal to 1.

Rule 12 For ldm:ComplexNode cn_i, if the following three conditions are satisfied:

(a) np = ldm:SimpleCode(cn_i) – ldm:Indentification (cn_i)– ldm:Reference(cn_i)),

(b) $|np| \geqslant 1$,

(c) $np_i \in np$.

Then p_i is declared UNIQUE and the restriction maxCardinality of Property p_i (p_i is the corresponding property of np_i) is set equal to 1.

Rule 13 For each ldm:ComplexNode cn_i, if the three conditions are satisfied:

(a) np = ldm:SimpleCode(cn_i) – ldm:Indentification (cn_i)– ldm:Reference(cn_i)),

(b) $|np| \geqslant 1$,

(c) $np_i (np_i \in np) = \varnothing$.

The restriction minCardinality of Property p_i is set equal to 0, if $np_i \neq \varnothing$, the restriction minCardinality of Property p_i is set equal to 1.

Thus, the OWL DL ontology can be automatically constructed using Rules 1-13.

3.2 Keyword extraction

The approaches given by Chiang et al.[12] and

Beneventano[16] were used for the keywords extraction. In this process, a synthesized view of the values of a SimpleNode is directly extracted from the data. These keywords are called "relevant values" since they represent the domain of a SimpleNode with a synthetic, though extensional description using a reduced set of values. Since the output of this method is values associated with an attribute which corresponds to a SimpleNode in the LDM model to represent these values, the data property in Table 2 is used in the OWL ontology.

Table 2 SimpleNode property: hasRelevantValue

Rdf:ID	Rdfs:domain	Rdfs:range	Rdfs:comment
HasRelevantValue	Ldm:SimpleNode	Xsd_owl:string	A SimpleNode has a set of relevant values

4 Case Study

A case is given to illustrate implementation of this proposed approach in a multi-enterprise business collaboration environment.

4.1 Multi-enterprise business collaboration scenario

The machine embroidery industry is a service chain of multiple enterprises cooperating, which includes collaborative design, manufacture, and supply. Effective management of the supply chain requires complete integration of the companies. Collaboration on the extended supply chain connects material flows, information flows, and fund flows. Figure 2 describes the scenario for the multi-enterprise business collaboration in the machine embroidery industry.

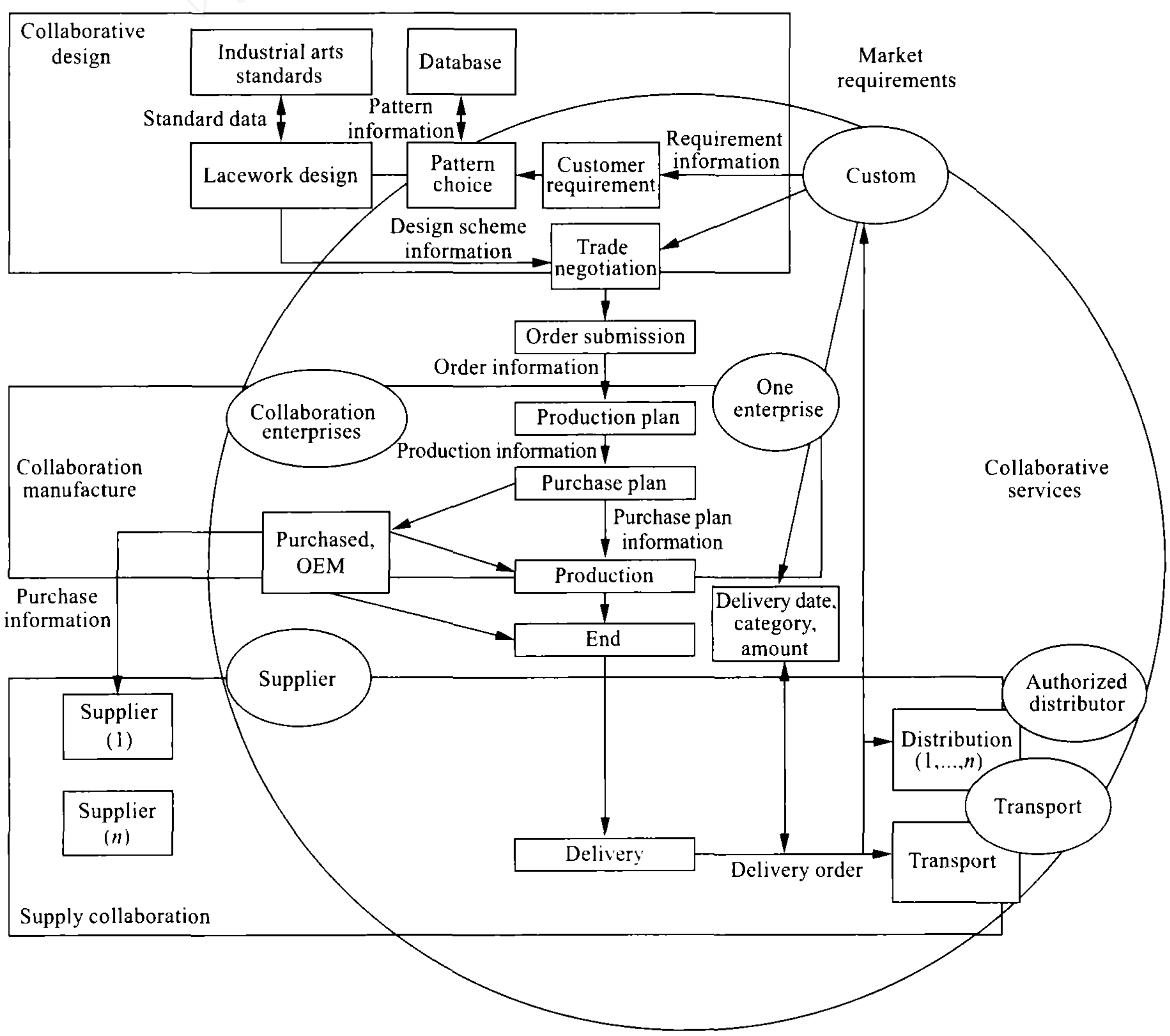


Fig. 2 Multi-enterprise business collaboration scenario

The embroidery industry collaboration has extensive business information communication including order information, production plans, material forms, and purchasing plans. Various companies may have different business information formats, such as database schema, XML schema, and EDI files. This semantic interoperability among these heterogeneous business information system based on semantic extraction is necessary. The semantic extraction of order information shown in Fig. 3 from a database format is given as an example to show the process for this approach.

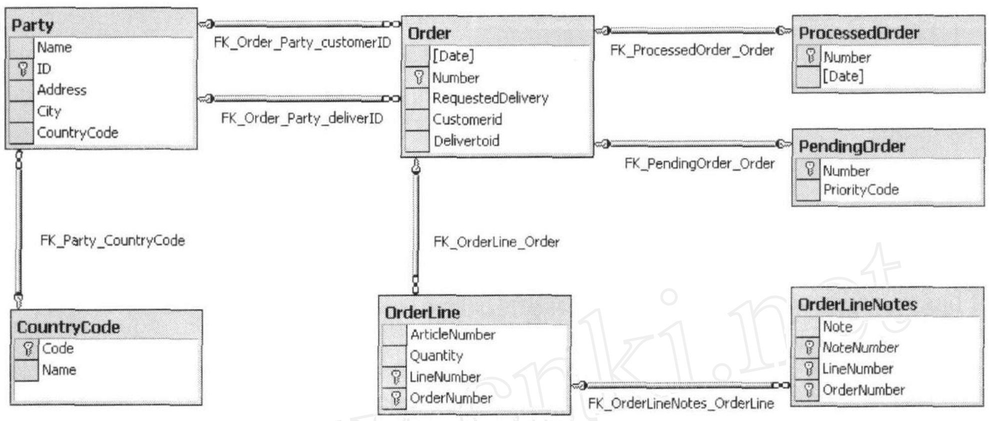


Fig. 3 Example of relational schema for ordering information

4.2 Format transformation

In Fig. 3, all the attributes except for the "Requested-Delivery" and "DelivertoID" attributes of the table Order do not allow NULL values. This relational schemata was coded into RDB_OWL as described previously[18] using the NRF transformation rules. Figure 4 shows a graphical representation of the NRF for the ordering schema in the LDM_OWL format.

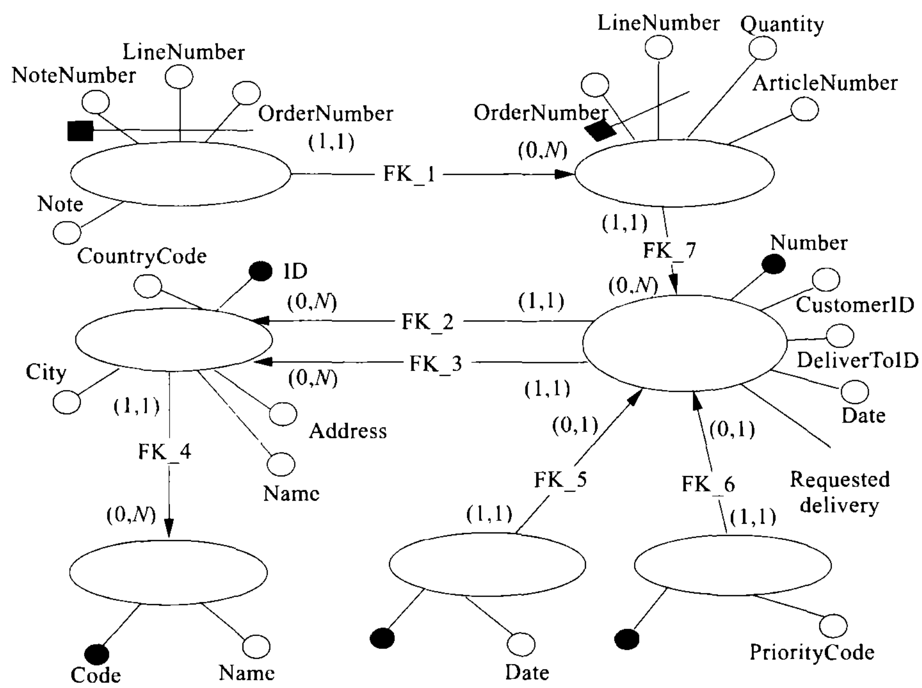


Fig. 4 Graphical representation of the NRF for the embroidery ordering schema

4.3 Semantic extraction

The conceptual extraction extracts the ontology structure from the order information, while the keyword extraction obtains the metadata.

4.3.1 Conceptual extraction

Figure 4 is used as the input to the reverse engineering conceptual extraction process to illustrate the ontology construction. The extraction process creates a conceptual schemata of the enbroidery order schema from the NRF represented in the OWL_DL ontology using Rules 1-13.

4.3.2 Keyword extraction

The SimpleNode "ArticleCategory" belonging to the ComplexNode "OrderLine" is used here as an example to show the keyword extraction. The extracted values of "ArticleCategory" include "Foodstuff", "Food", "Beverage", "Food", "Cocktail", "Beverage", "Cooking", "Beverage", and "Foodstuff".

The keyword extraction finds the relevant values of "ArticleCategory" by identifying the instances of the SimpleNode that correspond to an attribute. The "ArticleCategory" attribute returns the values "Food" and "Beverage" with the following clusters:

Food = (Foodstuff, Cooking, Food),

Beverage = (Beverage, Cocktail).

4.4 Discussion

The same process is used to map the other business information used by the machine embroidery companies to the OWL ontology. Manual refinement and lexical annotation is used to further refine the extracted semantic information. Then, the mapping relations are built among these information results to relate business information semantics in this multi-enterprise business collaboration environment.

An NRF transformation and semantic extraction tools are needed to effectively implement this method in the machine embroidery industry. These are the key functional components in the semantic interoperability system framework for the extraction of heterogeneous business information to the OWL ontology with a high degree of automation in a collaborative environment.

5 Conclusions and Future Work

This paper presents a semi-automatic semantic extraction method to acquire semantics from multiple heterogeneous data schemata. The extraction process includes: heterogeneous schemata format unification and the acquisition of semantic information from the NRFs with the result given in the OWL DL ontology. A case study illustrates how to implement the business information semantic extraction in a multi-enterprise business collaboration environment.

The semantic extraction process closely follows the standards of the OWL ontology and each data schema. The completeness and correctness of the semantic extraction is ensured. The processes transform multiple data structures into the NRF, which avoids the complex direct mapping from any one-data schema to the OWL ontology. The semantic extraction efficiency is greatly improved, and the interoperability is improved.

The conceptual extraction process may include some constraints implicit, or even miss some constraints, so suitable mapping rules are difficult for these constraints. Therefore, future work should concentrate on further refinements of the ontological structure.

References

[1] Fan Y S, Sun H J. Architecture and key technologies for SOA-based business collaborative management system. In: Proceedings of the 4th International Conference on Digital Enterprise Technology. Bath, UK: CIRP, 2007: 44-51.

[2] Silva C F D, Ghafour S A, Hoffmann P, Ghodous P. Semantic interoperability of heterogeneous semantic resource. *Electronic Notes in Theoretical Computer Science*, 2006, **150**(2): 1-85.

[3] Beneventano D, Montanari D. Mapping of heterogeneous schemata, business structures, and terminologies. In: The 8th International Conference on Database and Expert Systems Applications. Regensburg, Germany, 2007: 412-418.

[4] Wang H H, Li Y F, Sun J, Zhang H Y, Pan J. Verifying feature models using OWL. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2006, **5**(2): 117-129.

[5] Li M, Du X Y. Learning ontology from relational database. In: Proceedings of the Fourth International Conference on Machine Learning and Cybernetics. Guangzhou, China, 2005, **6**: 3410-3415.

[6] Shen G H, Huang Z Q, Zhu X D, Zhao X F. Research on the rules of mapping from relational model to OWL. In: OWLED 2006. Athens, Georgia, USA, 2006.

[7] Saw N T H. OWL-based approach for semantic interoperating and accessing heterogeneous XML sources. In: The 2nd Information and Communication Technologies. 2006: 522-527.

[8] Lammari N, Comyn-Wattiau I, Akoka J. Extracting generalization hierarchies from relational databases: A reverse engineering approach. *Data & Knowledge Engineering*, 2007, **63**(2): 568-589.

[9] Davies J. Reverse engineering of relational databases to ontologies. In: ESWS. New York: Springer-Verlag, 2004: 327-341.

[10] Benslimane S M, Malks M. Acquiring OWL ontologies from data-intensive Web sites. In: Proceedings of the 6th International Conference on Web Engineering. Palo Alto, California, USA, 2006, **263**: 361-368.

[11] Quang T, Barker K, Alhajj R. RDB2ONT: A tool for generating OWL ontologies from relational database systems. In: Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services. 2006: 170-178.

[12] Chiang R H L, Barron T M, Storeyv C. Reverse engineering of relational databases: Extraction of an EER model from a relational database. *Data & Knowledge Engineering*, 1994, **12**(2): 107-142.

[13] Bergamaschi S, Guerra F, Sartori C. Relevant values: New metadata to provide insight on attribute values at schema level. In: International Conference on Enterprise Information Systems. Funchal, Madeira-Portugal, 2007: 12-16.

[14] Kupfer A, Eckstein S, Neumann K, et al. Handling changes of database schemas and corresponding ontologies. In: International Conference on Conceptual Modeling. Heidelberg: Springer-Berlin, 2006, **4321**: 227-236.

[15] An Y, Borgida A, Mylopoulos J. Discovering the semantics of relational tables through mappings. *Journal on Data Semantics*, 2006, **4244**(7): 11-32.

[16] Ceravolo P, Damiani E, Viviani M. Bottom-up extraction and trust-based refinement of ontology metadata. *IEEE Trans. Knowledge and Data Engineering*, 2007, **19**(2): 149-163.

[17] Beneventano D. D232 semantic languages v0.4.1_2nd_RELEASE_CANDIDATE sc1.doc. http://www.stasis-project.net/intranet/folder.cfm?f=351&show=d1204, 2007.

[18] Maedche A, Staab S. Ontology learning for the Semantic Web. *IEEE Intelligent Systems*, 2001, **16**(2): 72-79.

[19] Astrova I, Stantic B. An HTML forms driven approach to reverse engineering of relational databases to ontologies. In: Proceeding of the 23rd IASTED International Conference on Databases and Applications (DBA). Innsbruck, Austria, 2005: 246-251.

University of Pennsylvania President Visits, Cooperation Agreement Signed

Professor Amy Gutmann, President of the University of Pennsylvania, visited Tsinghua on January 8, 2009. During her visit, a cooperation agreement was signed between University of Pennsylvania and Tsinghua.

Tsinghua University President Gu Binglin had a talk with Professor Gutmann. They discussed cooperation established between the two universities and exchanged ideas on enhancing collaborations in engineering, law, management, faculty exchange, and joint research. Tsinghua Vice President Xie Weihe also attended the discussion.

(From http://news.tsinghua.edu.cn，2009-01-09)