Framework and Modeling Method for Heterogeneous Systems Information Integration Base on Semantic Gateway

Xianwang Li*, Yuchuan Song*, Ping Yan, and Xuehai Chen
State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing, China
*Corresponding author, Email: lxw023@163.com, syc@cqu.edu.cn, yp@cqu.edu.cn, cxh@cqu.edu.cn

***Abstract*—To realize the integration of heterogeneous systems with greater flexibility and more convenience, a framework for heterogeneous systems information integration based on semantic gateway is constructed firstly, and then the integration process and several relevant technologies are discussed. On the basis of above researches, a modeling method for heterogeneous systems information integration is put forward. According to the integration requirements and traditional information models, the method is able to rapidly generate the information integration models, which provide supports to the implementation and operation of semantic gateway. Finally the research results have been verified in a manufacturing enterprise.**

***Index Terms*—Modeling Method; Information Integration; Semantic Gateway**

I. INTRODUCTION

The application of information systems in manufacturing enterprises have greatly improved the work efficiency and the market competitiveness. However, these systems are mostly dispersed and heterogeneous, and difficult to achieve information sharing and integrating between each other, which bring great difficulties to the collaboration between different departments within the enterprise, as well as the cooperation between enterprises. Therefore, there exist urgent requirements for the integration of heterogeneous systems. On the other hand, to adapt to the environment change more agilely, the information systems of manufacturing enterprises are required to have the ability of rapidly reconfiguration and integration on demand.

With the development of interoperation technologies, such as the appearance of COM, Web Service, MOM, the middleware-based integration technology was put forward and became dominant gradually. In this case, if the amount of systems involved is n, it just needs n times operations because the integration of these systems is conducted not between each other but with the middleware separately. In this way, the integrating is greatly simplified. At the same time, the middleware has the ability to transform data format and convert protocols. As a result, the problem of Heterogeneity is solved and the flexibility is improved.

The middleware-based integration is the research focus. Many researchers pay great attention to this area and lots of solutions were proposed separately. For instance, ref [3] discussed the data integration in banking statement management system. Ref [4] proposed a collaborative and integrated platform to support distributed manufacturing system. Ref [5] introduced a kind of heterogeneous data integration middleware.

The technologies mentioned above have solved the problems of interoperation and greatly improved the flexibility. But most of them lack effective practice. On the other hand, the research mainly concerned technology, ignoring information content.

(3) Integration in the information content layer

In this aspect, how to describe, analyze and map the relevant information is discussed. The research mainly consists of the following three aspects: a) information format; b) information semantics; c) information modeling.

The purpose of information format research is to make Information exchange standards, in order to make the information be identified and analyzed by different systems. STEP, EDI and XML are typical ones [6-8]. These technologies unify message interchange format, that is, has clarified fields constructing information and the data format of these fields. But it does not explicit the semantic information of fields. Therefore, it still needs to write special programs to analyze the interchange message. Because of this, the study of semantic information shows up. It indicates that semantic expression of information is cleared and standard by semantic annotation, ontology modeling and ontology-mapping technology [9-11], so that information can be recognized. There are not yet widely accepted standards now, so that semantic integration lacks effective applications.

Expression of information format and information semantics, as well as the construction of the systems, needs the support of certain models. So information modeling research is very important. Existing information modeling research mainly two aspects: a) traditional information modeling; b) Semantic Information Modeling. Traditional information modeling research mainly focuses on how to build information system

doi:10.4304/jnw.9.4.979-985

models and support design and development of information systems with these models.

UML and IDEF are typical ones [12, 13]. Semantic information modeling research mainly focus on how to build a semantic model of information, usually using ontology modeling technology, formal modeling technology and so on [14, 15].

Currently, traditional information modeling and semantic information modeling are studied separately in most cases, and there are few researches based on the combination of them. Furthermore, because the systems are built separately, information models are constructed respectively in the case of traditional information modeling, without considering information systems integration, namely these models do not contain any integration elements.

In summary, the middleware-based integration technology and semantic information integration technology are currently research hotspots, and they can solve the problems of information integration of heterogeneous systems. But due to the high technical requirements, there is a lack of effective implementation.

The researches on information integration generally consist of the following three aspects:

(1) Integration in the database layer

The information of enterprise information systems is mainly stored in databases, so information integration can be regarded as the integration of databases.

The Integration technologies of databases mainly include data federation and data warehouse [1]. In the case of data warehouse, data warehouse is created and used for storing the data, which are extracted from one or more original databases, and processed according to the global schema. In the case of data federation, a virtual view is used to collect, but not really store, integration information from original databases.

The main disadvantage of data federation and data warehouse is that they are tightly coupled to the original databases. As a result, the flexibility is insufficient. In addition, they put more emphasis on data integration, regardless of interoperation among databases.

(2) Integration in the application layer

In this aspect, interoperation among systems is mainly focused on.

In the early time, interoperation was realized through p2p-based secondary development according to the integration requirements. E.g. Ref [2] applied the interoperation technology to the integration between PDM and ERP. The advantage is that this technology is easy to realize, so it has been widely used. However, because the p2p-based secondary development is necessary, the integration efforts increase rapidly as the systems involved increase. For example, if the amount of systems involved is n, secondary development must be conducted c_n^2 times. The flexibility and extendibility is poor.

Therefore, the author's research team proposes a new integration technology, namely semantic gateway [16, 17] technology on the basis of the above study. It is intended to provide an effective solution to heterogeneous system integration with more convenience and greater flexibility.

II. Heterogeneous Systems Information Integration Based on Semantic Gateway

A. Framework for Heterogeneous Systems Information Integration Based on Semantic Gateway

The heterogeneous systems information integration framework based on semantic gateway is shown in figure 1. The integration framework consists of information systems, databases, database triggers, database adapters, semantic gateway server, and semantic gateway modeling tools. The database triggers designed as special database triggers for semantic gateway construction, are used to capture data changes; the database adapters are employed for adaption to different types of database system, such as Oracle, SQL Server, MySQL, etc.; the semantic gateway server, with the sematic gateway engine as the key part, is the core of the whole framework; the semantic gateway engine is constructed on basis of message-oriented middleware. Different from conventional one, this message-oriented middleware can conduct two-level semantic parsing and mapping, thus equipped with the ability of semantic understanding; the semantic gateway modeling tools are used to build integrated models, including semantic description model, semantic mapping model, semantic publishing modes and semantic subscription model, etc.

B. The Integration Process and Relevant Technologies

As enterprise information systems, in most cases, are based on databases, the information will be eventually stored in the databases. And various operations in information systems will finally mapping for four types of operations (insert, delete, update, select), and data changes in database follow. Therefore, the integration of information systems can be regarded as data synchronization among databases. From this perspective, the semantic gateway can also been seen as a data synchronization tool of databases. At the same time, different information systems and corresponding databases are mostly heterogeneous. Therefore, semantic gateway is also a tool to eliminate heterogeneity and help heterogeneous systems understand each other.

(1) Publish/Subscribe Process

The publish/subscribe techniques are able to realize decoupling between sender and acceptor. In this way, the sender is only responsible for sending messages to the semantic gateway, without caring about who will receive the message.

The detailed process of publish/subscribe is as follows.

Firstly, according to the integration requirements, the sender establishes database triggers, and database triggers will capture data alteration and send them through database adaptor to semantic gateway for further processing. Then, the acceptor as integration demands initiates subscription to semantic gateway. According to subscription items, semantic gateway engine distributes messages as required, and eventually write into the database of acceptor through database adaptor.

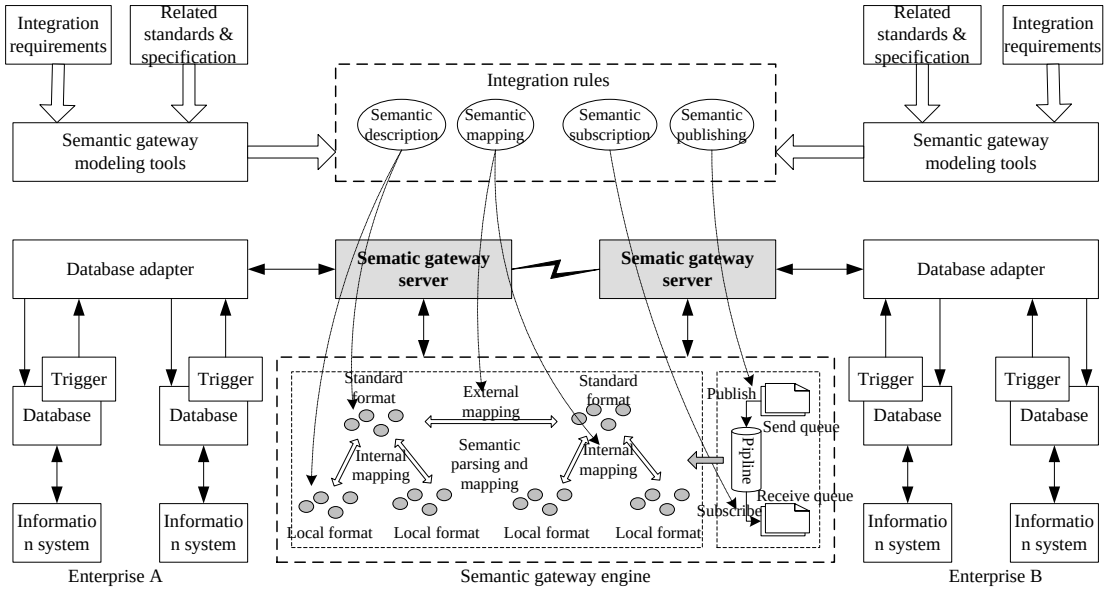


Figure 1. Framework for heterogeneous systems information integration based on semantic gateway

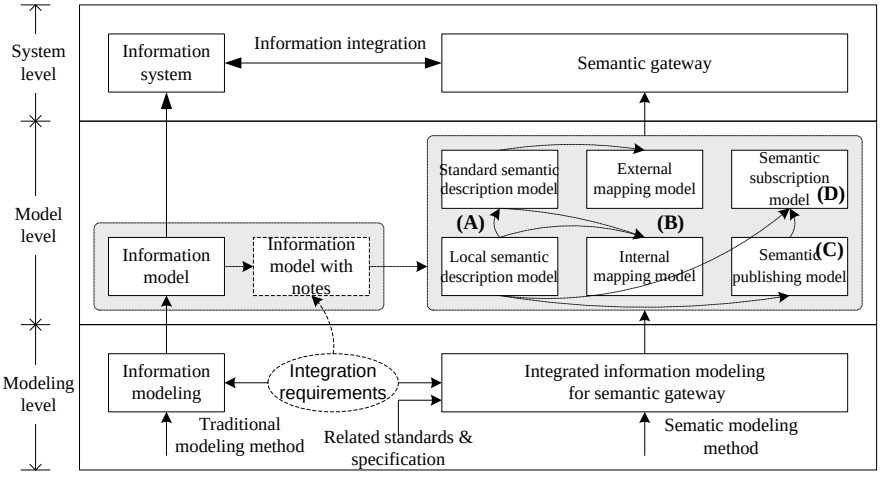


Figure 2. The modeling framework

(2) Two-Level Semantic Parsing and Mapping Process

Firstly, information in each information system has its own internal format called local semantic description. Secondly, every enterprise may set up a set of consolidated message interchange format called standard semantic description. And this is two-level semantic description. The mapping between local semantic description and standard semantic description is called internal mapping, and the mapping among multiple standards is called external mapping.

Internal information integration in enterprise should go through internal mapping twice. Firstly, the local semantic of system A maps the standard semantic of enterprise, and then standard semantic of enterprise maps the local semantic of system B.

Information integration among enterprises should go through internal mapping twice and external mapping once. Firstly, the local semantic of system A maps the standard semantic of enterprise A. Secondly, the standard semantic of enterprise A maps the standard semantic of enterprise B. Finally, the standard semantic of enterprise B maps the local semantic of system B.

III. Modeling Method for Heterogeneous Systems Information Integration Base on Semantic Gateway

A. Modeling Framework for Heterogeneous Systems Information Integration Based on Semantic Gateway

The framework of the heterogeneous system integration modeling method based on semantic gateway is shown in figure 2.

The main content of this framework can be summarized in three levels and two main lines. Three levels refer to modeling method layer, model layer and system layer. The model is produced by modeling method, and the model serves for system.

Two main lines refer to information modeling line and semantic gateway information integration modeling line. Information modeling line means the conventional information system modeling process, using conventional

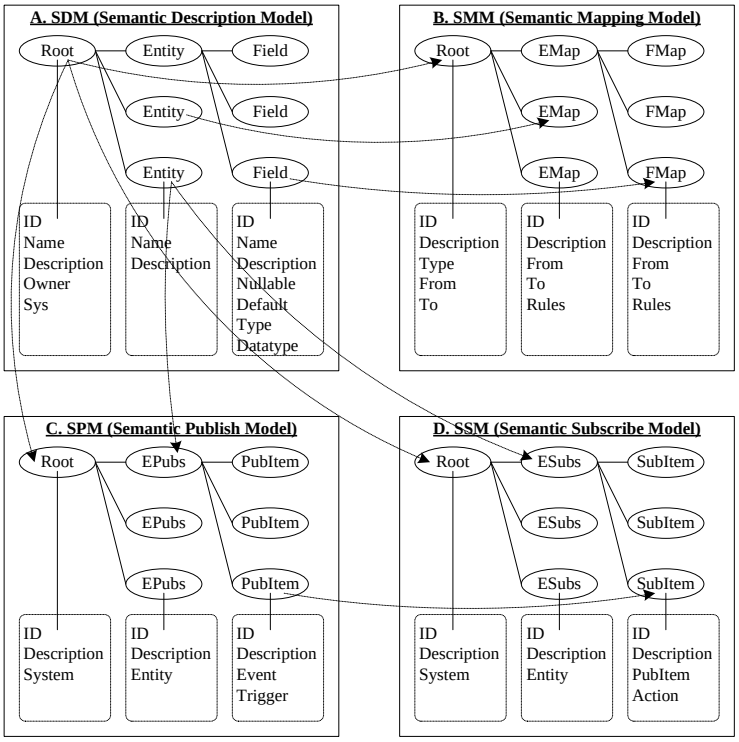


Figure 3. The information integration models

modeling method, such as UML, IDEF1x, etc., to construct information models of information systems. It will serves for the construction and development of information systems. Semantic gateway information integration modeling line is used to express the process of the heterogeneous systems integrated modeling method based on semantic gateway, using semantic modeling method and leading in relevant standards, to construct semantic gateway information integration models. It eventually supports the configuration and operation of semantic gateway. There is an obvious difference between semantic modeling method and conventional modeling method. The generated models of semantic modeling method should be analyzed by semantic gateway, so every model element and symbol needs specific semantic. Semantic gateway information integration model includes local semantic description model, standard semantic description model, internal semantic mapping model, external semantic mapping model, semantic publishing model and semantic subscription model.

Integration requirements are the link between information modeling and semantic gateway information integration modeling. Conventional information modeling has no regard for integration requirement. In order to support semantic gateway integration, the conventional information modeling method should be extended to add integration requirements description. Just adding the mark "☑" and corresponding annotations to model elements required to participate in integration, the prototype of the semantic gateway information integration models will appear without interruptions to the normal information modeling process.

B. The Models and Modeling Process

Semantic gateway information integration model includes local semantic description model, standard semantic description model, the internal semantic mapping model, the external semantic mapping model, semantic publishing model and semantic subscription model. And these six models can be divided into semantic description model, semantic mapping model, semantic distribution model and semantic subscription model. As shown in Figure 3.

(1) Semantic Description Model (SMM)

Semantic Description Model is defined as follows.

$$SMM = \{ID, Name, Description, Type, Owner, (Entity, Entity, Entity...)\} \quad (1)$$

ID is the unique number, *Name* is the designation, and *Description* is the additional information of the model. *Type* is the type of the model, valued as $\{\text{Local} \mid \text{Standard}\}$. *Local* stands for local semantic description model and *Standard* stands for standard semantic description model. *Owner* is used to represent which system or enterprise does the model belongs to.

Entity is similar to the table model of database. Each *SMM* model contains more than one *Entity* model.

$$Entity = \{ID, Name, Description, (Field, Field, Field\ldots)\} \quad (2)$$

ID is the unique number, *Name* is the designation and. *Description* is the additional information of *Entity* model.

TABLE I. FIELD TYPE

Type Name	Sample	Additional Semantic Description
Primary	Customer.ID	{Rules}, Rules represents the encoding rules of primary key
Foreign	Order.Customer.ID	{Entity, Field}, Entity represents the associated entity, Field represents the primary key of the associated entity
Calc	Total-price = Unit-price * Amount	{Fields, Formula}, Fields represent one or more fields which are reliant, Formula represents computational formula.
Enum	Sex	{(Item, Item, Item …) | Item= {Value, Description}}, Value is the value of enumeration term, Description is the semantic description of enumeration term.
Normal	Customer.Name	

TABLE II. DATA TYPE OF FIELD

Type Name	Sample	Additional Semantic Description
Decimal	16.7	{Length, Precision}, Length represents data bits, Precision represents decimal digits.
Currency	$99.0	{Type}, Type represents kind of currencies.
Date	8:23:15	{Local, Format}, Local represents time zone, Format represents date format.
String	"hello"	

Field is similar to the field model of database. Each *Entity* contains more than one *Field* model.

$$Field = \{ID, Name, Description, Nullable, Default, Type, Datatype\} \quad (3)$$

ID is the unique number, *Name* is the designation and *Description* is the additional information of the model. *Nullable* represents whether the field value can be empty or not. Its value is $\{Yes \mid No\}$. *Default* is the default value of field. *Type* is the type of the model, as shown in Table 1, and *Datatype* is the data type of field, as shown in Table 2.

Through the analysis of the semantic description model and comparing it with the information model, it can be concluded that most model elements of semantic description model are similar, or even identical to the model elements of information model, such as Entiy.Name, Field.Name and Field.Nullable, ect. So these model elements can be got directly from information model, but there are also some model elements not existing in information model, such as ID, Field.Type, which should be added and perfected.

The steps of establishing semantic description model are as follows:

Step 1: The main model elements of local semantic description models are formed through information models with integration notes.

Step 2: The complete local semantic description models are formed through supplements to the above models.

Step 3: The standard semantic description models are formed through two ways. One is extracting model elements from local semantic description models and conducting further integration (bottom-up molding); the other is building the models from scratch according to the relevant standards (top-down modeling).

(2) Semantic Mapping Model (SMM)

Semantic mapping model is defined as follows.

$$SMM = \{ID, Description, Type, From, To, (EMap, EMap, EMap...)\} \quad (4)$$

ID is the unique number and *Description* is the additional information of the model. *Type* is the type of the model, valued as "internal" or "external" which respectively indicates the internal semantic mapping and the external semantic mapping. *From* and *To* represent both sides of the semantic mapping.

EMap is entity mapping model. Each *SMM* contains more than one *EMap*.

$$EMap = \{ID, Description, From, To, (FMap, FMap, FMap...)\} \quad (5)$$

ID is the unique number and *Description* is the additional information of the model. *From* and *To* represent both sides of semantic mapping.

FMap is field mapping model. One *EMap* contains more than one *FMap*.

$$FMap = \{ID, Description, From, To\} \quad (6)$$

ID is the unique number and *Description* is the additional information of the model. *From* and *To* represent both sides of semantic mapping.

Semantic mapping model and semantic description model share analogous hierarchical structure. And every semantic mapping model depends on two semantic description models.

The steps to construct semantic mapping model are as follows:

Step 1: According to the integration requirements, the relationships between information system and semantic gateway and the relationships between semantic gateway and other semantic gateways should be definite. Thus a top-level model of semantic mapping model is formed, namely the Root node.

Step 2: The mid-level model of semantic mapping model, namely the EMap nodes are formed through identifying the relationships between entities from different semantic description models.

Step 3: The bottom-level model of semantic mapping model, namely the FMap nodes are formed through

identifying the relationships between fields from different semantic description models.

(3) Semantic Publishing Model (SPM)

Semantic publishing model is defined as follows.

$$SPM = \{ID, Description, System, (EPubs, EPubs, EPubs...)\} \quad (7)$$

ID is the unique number and *Description* is the additional information of the model. System indicates which information *System* represents which system does conduct information publishing.

EPubs represents the publishing rules based on entity. One *SPM* contains more than one *EPubs* .

$$EPubs = \{ID, Description, Entity, (PubItem, PubItem, PubItem...)\} \quad (8)$$

ID is the unique number and *Description* is the additional information of the model. *Entity* describes on the basis of which entity the *EPubs* are established.

PubItem is one of the publishing rules. One *EPubs* contains more than one *PubItem* .

$$PubItem = \{ID, Description, Event, Trigger\} \quad (9)$$

ID is the unique number and *Description* is the additional information of the model. *Event* represents the causes of data changes, including INSERT, UPDATE and DELETE. *Trigger* represents the events triggering condition.

For example, when synchronizing the "Order" information between ERP and MES, it only needs to construct the PubItem model for the UPDATE event in ERP and the triggering condition which is the order status changes from "releasing" to "released".

The steps of establishing semantic publishing model are as follows:

Step 1: Identify the system with information to be integrated and its local semantic description model, and form top-level model of semantic publishing model, namely the Root node.

Step 2: Analyze which entity can provide information needed by integration and further find out the business operations associated with these entities. Then the operations conforming to integration rules are located, along with the corresponding database events and triggering conditions, to form EPubs models and PubItem models. And complete semantic publishing model will eventually be constructed

(4) Semantic Subscription Model (SSM)

Semantic Subscription Model is defined as follows.

$$SSM = \{ID, Description, System, (ESubs, ESubs, ESubs...)\} \quad (10)$$

ID is the unique number and *Description* is the additional information of the model. *System* indicates which information system is targeted to conduct information subscription.

ESubs represents the subscription rules based on entity. One *SSM* contains more than one *ESubs* .

$$ESubs = \{ID, Description, Entity, (SubItem, SubItem, SubItem...)\} \quad (11)$$

ID is the unique number and *Description* is the additional information of the model. *Entity* describes on the basis of which entity the *ESubs* are established.

SubItem is one of the subscription rules. One *ESubs* contains more than one *SubItem* .

$$SubItem = \{ID, Description, PubItem, Action\} \quad (12)$$

ID is the unique number and *Description* is the additional information of the model. *PubItem* represents that on the basis of which publishing rule the subscription information is distributed. *Action* is the operation after receiving subscription.

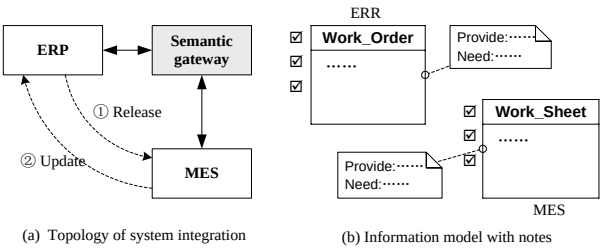
The steps to establish semantic subscription model are as follows:

Step1: Identify the information systems requiring information integrated from other ones and locate the corresponding local semantic description models to form the top-level model of semantic subscription model, namely Root node.

Step2: Identify the entity data in need of synchronization and integration with others and further analyze the integration items. Establish the relationships between subscription items and distribution to eventually obtain complete semantic subscription model.

IV. APPLICATION VERIFICATION

The research results of this paper have been verified in one manufacture enterprise which produces the core components of marine diesel engine. The core information systems in this enterprise are ERP and MES. And they need to interchange the "Worksheet" information, as follows: after ERP system finishes the worksheets assignation, then these assigned worksheets will appear in MES system; in turn, any updated worksheet information in MES system should be synchronized with ERP system in real time. As shown in figure 4(a).



(a) Topology of system integration (b) Information model with notes

Figure 4. The worksheet's information integration of ERP/MES

The process of establishing the information integration model based on semantic gateway:

(1) Mark the ERP/MES information models separately and then form information models with the integration

notes according to integration requirements. As shown in figure 4(b).

(2) Establish the worksheet's local semantic description model of ERP and MES respectively based on the information models with integration notes.

(3) Construct the worksheet's standard semantic description models combined with the worksheet's local semantic models of ERP/MES according to relevant standards.

(4) Establish the mapping from the worksheet's local semantic description models of ERP/MES to the worksheet's standard semantic description models separately.

(5) Establish the worksheet's semantic publishing models of ERP system according to the integration requirements.

(6) Establish the worksheet's semantic publishing models and the worksheet's semantic subscription models of MES according to the integration requirements.

Through the above steps, semantic gateway information integration models are established. The models can transform into semantic gateway setting, with the ability to support the rapid implementation of semantic gateway.

V. CONCLUSIONS

In this paper, a framework for heterogeneous systems information integration based on semantic gateway has been proposed. With the support of the framework, information integration of heterogeneous systems can be realized with more convenience and greater flexibility.

A modeling method for heterogeneous systems information integration based on semantic gateway has been put forward. The models and Modeling process has been discussed in detail. This method is able to rapidly generate the information integration models, which provide supports to the implementation and operation of semantic gateway

The application verification indicates that the research results of this paper have very good feasibility and effectiveness.

ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China under Grant No. 51075414

REFERENCES

[1] Xian-Di Yang, Zhi-Yong Peng, Jun-Qiang Liu, Xu-Hui L. I. An Overview of Information Integration. *Computer Science.* 2006, 33(7) pp. 55-59, 80.

[2] Yu Jun-He, Qi Guo-Ning, Wu Zhao-Tong, Gu Xin-Jian The College Of, Zhejiang University, Hangzhou, et al. Research on Method and Application of Integrating PDM and ERP. *Computer Integrated Manufacturing Systems.* 2001(06) pp. 49-53.

[3] Li Zhi-lin. The Data Integration in Banking Statement Management System Based on Messageing Middleware. *South China University of Technology*, 2012.

[4] Valilai Omid Fatahi, Houshmand Mahmoud. A collaborative and integrated platform to support distributed manufacturing system using a service-oriented approach based on cloud computing paradigm. *Robotics and Computer-Integrated Manufacturing.* 2013, 29(1) pp. 110-127.

[5] He Rongmao, Qin Futong, Hu Ran, Yu Xin. Research and Design on Heterogeneous Data Integration Middleware Based on SOA. *Ship Electronic Engineering.* 2012, 32(1) pp. 77-78, 124

[6] Zhouyang Li, Xitian Tian, Guoding Chen. Study on CAD/CAPP/CNC Integrated System Based on STEP-NC. *China Mechanical Engineering.* 2006, 17(21) pp. 2243-2248.

[7] Wang Ying-Lin, Qi Ke-Tao, Zhang Shen-Sheng. XML/EDI Integration Framework for Agile Supply Chain. *China Mechanical Engineering*. 2003, 14(22) pp. 1922-1925.

[8] Li Qing, Mak Hon Chung, Zhao Jianmin, Zhu Xinzhong. OXML: an Object XML Database Supporting Rich Media Indexing and Retrieval. *Journal of Multimedia*. 2011, 6(2) pp. 115-121.

[9] Yahia Esma, Lezoche Mario, Aubry Alexis, Panetto Herv é Semantics enactment for interoperability assessment in enterprise information systems. *Annual Reviews in Control.* 2012, 36(1) pp. 101-117.

[10] Yong Feng, Ming-Yu Wang. Lightweight data integration method based on semantics. *Computer Engineering and Design*. 2012, 33(1) pp. 402-406.

[11] Buitelaar Paul, Cimiano Philipp, Frank Anette, Hartung Matthias, Racioppa Stefania. Ontology-based information extraction and integration from heterogeneous data sources. *International Journal of Human-Computer Studies.* 2008, 66(11) pp. 759-788.

[12] Yang Tian-Jian, Wang Bo, Guo Wei, Zhao Yan-Bin. Integrated Modeloing Method of CAPP System with IDEFX and UML. *Journal of Tianjin University*. 2004(06) pp. 553-558.

[13] Kim Cheol-Han, Weston R. H., Hodgson A., Lee Kyung-Huy. The complementary use of IDEF and UML modelling approaches. *Computers in Industry.* 2003, 50(1) pp. 35-56.

[14] Guo Gang, Tang Hua-Mao, Luo Yu. Semantic -based Product Functional Formal Modeling. *Computer Integrated Manufacturing Systems*. 2011(6) pp. 1171-1177

[15] Yuan Qiing-Ni, Xie Qing-Sheng, Xu Ming-Heng, Li Shao-Bo. Research on manufacturing resources ontology model based on semantic. *Journal of wuhan university of technology.* 2009(10) pp. 121-125

[16] Song Yu-Chuan, Lei Qi, Liu Fei. A semantic gateway architecture for system integration in networked manufacturing. Laubisrutistr. 24, Stafa-Zuerich, CH-8712, *Switzerland: Trans Tech Publications Ltd*, 2010 pp. 419-420, 453-456.

[17] Qi Lei, Yu-Chuan Song, Xian-Wang L. I. Heterogeneous system integration framework and key technologies with the support of semantic gateway. *Journal of Chongqing University (Natural Science Edition).* 2010, 33(11) pp. 27-32.