Research on Metadata-driven Enterprise Data Modeling System

Hong Miao[1, 2] Jingsheng Sun[1] Shilun Ge[2] Nianxin Wang[2]

[1]School of Automation, Nanjing University of Science and Technology, 210094, Nanjing, China
[2]School of Economics & Management, JiangSu University of Science & Technology, 212003, Nanjing, China

Abstract

Since the existing mass data resource in enterprises has never generated any reuse-value, this paper proposes a framework for enterprise data modeling, verification and management driven by metadata. It starts from the building-up of metadata base, and then checks and verifies the metadata at different levels. It also put forward a metadata management strategy followed by MOF standard. With such foundation, the model elements and fragments can be drawn out from the metadata items to mapping between the metadata and data model. Such a framework can contribute a lot to develop enterprise data standard and construct a unified data modeling platform.

Keywords: data item, metadata, model verification, enterprise data model, data standard

1. Introduction

Enterprise information systems use the data to describe enterprises themselves and their products, record enterprise activities, reflect enterprise states and support decision-making [1]. In the dynamical and changeable environment, data and data model are increasingly important in enterprise information systems. In recent decades, IT application has permeated greatly in domestic enterprises, and then mass data has been accumulated. Because there lacks a unified data standard and data models representing enterprise essences, the mass data, including structure data about enterprise features and process data about enterprise business, could not turn into valuable resources. Then it becomes very difficult for communication, information analysis and mining among enterprises, not even to update legacy systems in enterprise. So the research topic about enterprise data model, with the goal of reflecting enterprise business principles and facts, keeping consistency with data standard, adapting requirements changes and qualifying software reuse, is promptly critical.

2. Literature Review

2.1. Enterprise data and data model

Enterprise data and data model are always the focus in the domain of information systems. According to the evolution of topics, the research is mainly on the following three aspects.

(1) On the methods, technologies and tools of enterprise data models. Since 1970s, the use of rational model made the rational database dominant [2]. ER model has been widely applied till now [3]. Data flow diagram is always the main method to do structure analysis and design [4]. IDEF1x method is popular for its semantic advantage in concept modeling [5]. In

recent years, UML has been widespread accepted as standard. With the development of software engineering theories, the data modeling methods, technologies and tools will get more mature.

(2) On the instances or demos of data modeling methods. For the methods, technologies and tools are more mature, the study focus turns to the application that is how to build data models in a specific domain. In the process of data being expressed as data models, the more work is people's thinking about information systems understanding and analysis. During such process, the research is in three stages. The focus of the first stage is on business analysis, which mainly centers on the relationships between data and functions, such as the U-C matrix [6], and Zachman framework [7]. The focus of the second one is on strategic data planning. Researchers are gradually realized the commonness of business, the stability of data description and the science of system development. So the work on data should be strategic. James Martine firstly put forward the idea that data is stable while processes are changeable, and advocated the theory of overall data-planning [8]. Sheer built the data structure for typical functions in general enterprises, which was developed as the famous methodology of ARIS [9]. Meanwhile the theory of information resources management was popular [10]. The focus of the third one is on data integration. When enterprises have some IT input and infrastructure, the heterogeneous systems and data sources do exist. The data model of each application system was designed just for function without planning. And for the domain specialties, such data models can not be accumulated and transformed into sustained and reusable resources. So the problems of data sharing and interaction among these systems are increasingly outstanding. Then a series of methods, ideas, description languages and architec-ture technologies are appearing for such problems, for example, agile data modeling, Ontology, XML, data warehouse and semantic web [11].

(3) On the data model standard. The wide application of enterprise data model motivated enterprise data standard being the research focus. Users and researchers are conscious of that the operation and maintenance cost is much higher and the information systems are much more inefficient, the reason for which is the worse data model quality and there are few relative standards. The literature of [12] put forward a series of principles about how to build high quality data model in practice. The literature of [13] pointed out that the management on enterprise data was not only doing strategic planning but also to develop data items standard for database design, implementation and maintenance. In 2003 there was a try of releasing a standard on ERP in China [14], which gave out a series of data items catalog according to an ERP package. Then some special data models and standards are released, such as the Core Product Model released by USA NISR, the Open Assembly Model, Shop Data Model. And international organizations for standardization developed Step standard. ISA and ANSI together developed the ISA95 for integration of management and control. The faculty of Product Development Engineering department in Beihang University put forward a Unified Product Information Model. However the above data models and standards are almost on the product manufacturing, while at the enterprise level there has no the models or standards about products, features, behavior and performance of a type of enterprises.

2.2. Model verification

From the above study about enterprise data and data model, most of them lacked of the feedback mechanism to verify

them. So such data models were not widespread accepted and popularized. Recently, verification on data models is being concerned. In the domain of information systems, the model checking and verification are always the study area, but are being developed differently in varied directions [15]. Model checking in safety software is always the emphasis and its automated tools have been developed [16-17]. In software requirement engineering area, analyzers recognize the point of capturing accurate requirements and take work on not only the test but also put more attention on the checking and verification on requirement model [18]. While in enterprise modeling area, there are less attention and less work on such model checking. The traditional way to control the modeling quality is always the domain experts' review, and some researchers have begun the study on enterprise model verification process and framework, such as TOVE, which put forward a method of asking competency question to test ontological model [22]. The literature of [15, 20] analyzed the close relationship between modeling activities and checking or verification activity, and designed a formal method based on the enterprise modeling langrage UEML. Meanwhile its important contribution is the list of study issues about checking or verifying models in different granularity and views, and shows the frontier in this academic research area. The literature of [21] gave a method of business modeling and automated verification on business model in transaction systems. While the checking and verification on data models, the most is the general verification according to the normalization theory of relational database, and the review experience of domain experts has not been accumulated and abstracted to form a data modeling process with feedback. Now the necessity of verification and evaluation on concept data models has been obvious

[22-23]. The literature of [24] addressed seven basic verification problems in workflow data modeling. The literature of [25] built the TOVE ISO9000 micro theory on enterprise quality management for two purposes, one is to share the concept data of quality domain, the other is to supply the standard for verifying other quality ontology models. When managing data model in the late life cycle, current normalization work always centers on keeping the static documents of confirmed data models and daily maintaining of physical data model implemented in operation. And few studies are on the existence, evolution and dynamic management of enterprise data and data models [26]. Then as known many data models do not have the strong and sustained reference capability.

2.3. Metadata theory

The concept of metadata is another profile for studying on the verification and management of enterprise data models, and even the standardization of enterprise data. Metadata is the data about data, and people have got the consensus of its importance to present the enterprise data flow, information flow and workflow, and also to support the scalability and reusability of information systems [27, 13].

Nowadays, the metadata is always the topic of data warehouse area for the purpose of business intelligence. In this paper our work, referring to the concept of metadata and its standard MOF, is for the whole life of enterprise IS instead of the only stage of BI.

In summary, checking, verification and management on data model are becoming more important, especially in the domain data modeling practices. To capture a set of data model, which is verified to be rational, effective, complete and exact to express business logic, is significant for accumulating, reusing, and standardizing the current data models.

3. Metadata-driven Enterprise Data Modeling System

With the practical experiences of IT infrastructure building in the large-scale, single-piece and small-batch okp enterprises in these years, we bring forward such a data modeling system driven by metadata.

3.1. Enterprise metadata base

The metadata base is the basis and platform of all study work. The key steps to build metadata base are as followed.

(1) To abstract data items. We take the structure information of enterprises operation data, collected from real applications, as source data, and then abstract data items or items group from the attributes set.

(2) To describe data items. Through the analysis on the enterprise functions, processes and activities at general, industrial and application levels, taking the data items and items group as objects, we should answer a series of questions. What are their general features and domain features? What constraints are set on them in different business activities? What changes will happen and when happens? What are the changing tracks? What influences do the changes bring out to what other data items? What rules will be concluded after such investigation? Finding the answers to these questions is to describe the data items.

(3) To model the metadata. After getting the whole description of data items, we do the modeling work following the metadata standard of MOF. The work is to find a reasonable structure and mode to store the captured commonsense knowledge and rule knowledge.

3.2. Verification and management on enterprise metadata

- Verification on metadata

The metadata in metadata base is used to describe enterprise data. Then its accuracy, rationality, validity and integrity become the key about whether the data is resource. So the verification work is taken in three aspects.

(1) Static checking. For semantic level, according to normalization theory of relational database, we check the general features and domain features. For example the name, format, default values, synonymy semantics and so on.

(2) Dynamical business checking. At the industry level, there are many fact rules and tacit rules, which can be set as theorems and axioms to infer, deduce and prove more facts and rules. Using the formal method of predicate logic, obeying the principles of non-contradiction, integrity and independence, we draw out the theorem formulas and axioms about many aspects, such as temporal relationship, modal relationship, constraints, metadata calls, referential integrity and scalability, to check the dynamical intension of enterprise metadata.

(3) Experiment of simulation checking. Taking the existing and widely accepted enterprise referential data models and other data models successfully applied in such enterprises for reference, we develop the query mechanism and scalability mechanism to investigate whether the metadata in such instances of data models is covered by our metadata base. If not, we need to find the reason, which is naming problem or none-existence or others. It can validate the integrity and effectiveness of metadata base, meanwhile can promote the accumulation of domain knowledge.

- Management on metadata

It is to choose a management framework and strategy of enterprise metadata. Firstly, we set the management principles of high integration, scalability, multi-dimension, and robustness, and adopt the strategy of MOF standard, including data

layer, interface layer and application layer. It used meta-meta model to unify the description of all kinds of meta-model, based on which we set the unified access interface to metadata and classify the metadata dynamically to achieve the multi-dimensional management.

3.3. Building-up, Verification and Management of Enterprise Data Model Based on Metadata

When the metadata model and metadata base are ready, the key point is to resolve the problem of building-up, verifying and managing data model.

(1) Draw out the abstract levels between metadata and data models.

We choose the relational model as the form of enterprise data model, including entity, attributes and relationship. Based on enterprise metadata base, we are inspired by the software component method and put forward the concept of data model components, including two levels.

The first is model elements, which are composed by the metadata from enterprise metadata base and are a series of instances from requirements corresponding to the elements of relational model, such as table, field, and cardinality and so on.

The second is model fragments, which are composed of model elements and can express the requirement scenario of some concrete activity, process and function.

So the transformation between enterprise metadata and data model becomes the mapping and matching among metadata-model element-model fragment.

(2) Enterprise data model verification based on metadata.

The model elements should be verified according to relational normalization theory and integrity reference rules. The business process models in model fragments are transformed into its limited state space from UML. Combined with the requirement specification of a domain,

the rules in business processes are transformed into linear temporal logic. With the model checking technology to verify the business process automatically. If failed, the reason can be explained through the counter-example.

(3) Enterprise data model management based on metadata.

On the ground of enterprise metadata base, according to the domain requirements, we can construct the model elements, fragments and data models. Following the composition relationship and inheritance relationship among model elements, fragments and data models, we set up the BOM of data model components to serve for accumulating and managing the data models orderly. Through the study on the modeling behavior of users with such platform, we can quantify the reusable value of each data model component to support reference in the later domain data modeling work.

4. Conclusion

As we emphasized in the paper, enterprise data model is to support application, while metadata centers much more on reusable and sharable resources. Through the breakdown from enterprise data to metadata and the mapping between metadata with enterprise data model, we can achieve the building-up, verification and management of enterprise data model driven by metadata, which is shown as Fig.1.

Since the enterprise information systems have been developed to today, the knowledge and intelligence is being the trend. Then the higher requirements are put on enterprise models. This modeling system captures the essence of data resources of enterprise data modeling and brings some contributions to promoting standard data environment and sustained development of enterprise information systems.
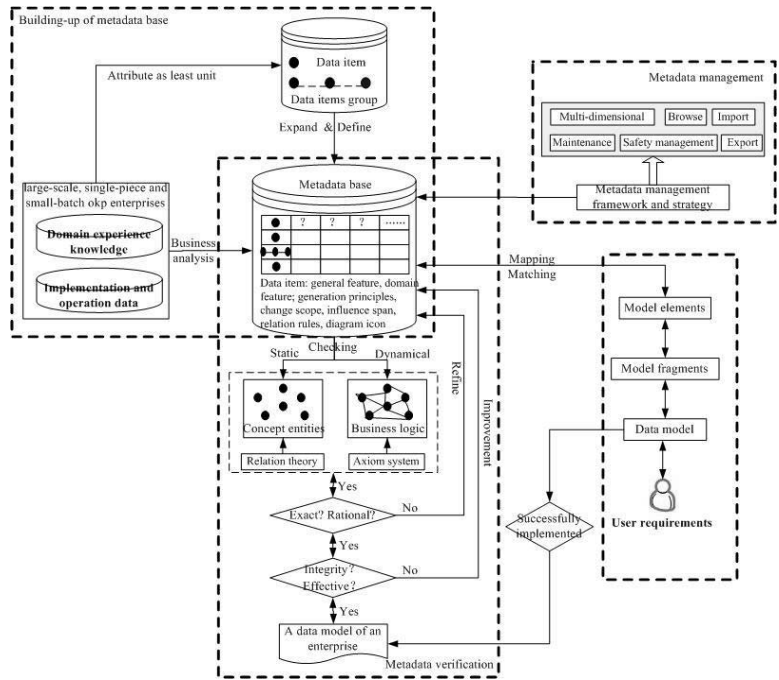
Fig. 1: Metadata-driven enterprise data modeling system.

References

[1] GE Shi-lun, "Information model of large-scale, single-piece and small-batch okp enterprises" (in Chinese), Beijing, Science Press, 2005

[2] Codd EF. "A Relational Model of data for large shared databanks" . *Commun ACM*, pp. 377–387, 1970.13(6).

[3] Chen PP. "The entity–relationship model: toward a unified view of data".

IBM Systems Journal, pp. 31-53, 1982. vol.21.No.1.

[7] Zachman, J.A. "A Framework for Information Systems Architecture". *IBM Systems Journal*, pp.276-292, 1987.vol.26.No.3.

ACM Trans Database System, pp. 19–36, 1976.

[4] Gane C, Sarson T. "Structured systems analysis: tools and techniques". Prentice, Englewood Cliffs. 1979.

[5] FIPS Publication 184 released of IDEF1X by the Computer Systems Laboratory of the National Institute of Standards and Technology (NIST). 1993.12.31.

[6] Zachman, J.A. "Business Systems Planning and Business Information Control Study: A Comparison".

[8] James martine. "Strategic Data Planning Methodologies", 1982.

[9] Scheer, August-Wilhelm. "ARIS – Business Process Modeling", Third Edition, Springer-Verlag 2000. Berlin: Springer-Verlag, 1992.

[10]   Eileen M. Trauth, "The Evolution of Information Resource Management", *Information & Management*, pp.257-268, 1989.

[11]   http://en.wikipedia.org/wiki/Data_modeling.

[12]   Matthew West and Julian Fowler. "Developing High Quality Data Models". *The European Process Industries STEP Technical Liaison Executive (EPISTLE)*, 1999.

[13]   Nikki Swartz. "Data Management Problems Widespread". *The Information Management Journal*, pp.28-30, 2007.vol.9.No.10.

[14]   Deng chao. "The Application Guideline to Enterprise Resource Planning" (in Chinese). Beijing, Publishing House of Electronics Industry, 2003.9.

[15]   V.Chapurlat, C.Braesch. "Verification, validation, qualification and certification of enterprise models: Statements and opportunities". *Computers in Industry*, pp.711-721, 2008(59).

[16]   Clarke E, Grumberg O, Peled D. Model checking. MIT Press, 2000

[17]   Moebius, N.; Haneberg, D.; Reif, W.; Schellhorn, G. "A Modeling Framework for the Development of Provably Secure E-Commerce Applications". *Software Engineering Advances, 2007. ICSEA 2007*. International Conference on 25-31 Aug. 2007 Page(s):8 - 8

[18]   Banerjee, A.; Ray, S.; Dasgupta, P.; Chakrabarti, P.P.; Ramesh, S.; Vignesh, P.; Ganesan, V., "A dynamic assertion-based verification platform for UML statecharts over rhapsody". TENCON 2008. *IEEE Region 10 Conference*, pp.1-6, 2008.

[19]   Mark S. Fox, "Michael Gruninger. Enterprise Modeling". *AI Magazine*, pp. 109-121, 1998.19(3).

[20]   V.Chapurlat, B.Kamsu-Foguem, F. Prunet, "A formal verification framework and associated tools for Enterprise Modeling: Application to UEML". *Computers in Industry*, pp.153-166, 2006(57).

[21]   Y. Liu, S. Muller, K. Xu. "A static compliance-checking framework for business process models". *IBM Systems Journal*, pp.335-361, 2007.5.Vol.46, No.2.

[22]   Graeme Shanks, Elizabeth Tansley, and Ron Weber. "Using Ontology to Validate Conceptual Models ". *Communication of The ACM*, pp.85-89, 2003.Vol.46.No.10.

[23]   Chery L. Dunn, Gregory J. Gerard, Severin V. Grabski. "Critical evaluation of conceptual data models ". *International Journal of Accounting Information System*, pp.83-106, 2005.6.

[24]   Shazia Sadiq, Maria Orlowska, Wasim Sadiq, Cameron Foulger. "Data Flow and Validation in Workflow Modelling". ADC'2004 Dunedin, New Zealand, *Conferences in Research and Practice in Information Technology*, 2004.Vol.27.

[25]   Henry M. Kim, Mark S. Fox, Arijit Sengupta. "How to build enterprise data models to achieve compliance to standards or regulatory requirements (and share data) ". *Journal of the Association for Information Systems*, pp.105-108, 2007.Vol.8.Issue 2.

[26]   Xi Jianqing, Guo Yubin, Tang Deyou. "Data Gene : Model of Managing the Evolution of Data" (in chinese). *Computer Science*, pp.12-16, 2007.vol.34.No.1.

[27]   Pat Franks, Ph.D, CRM, and Nancy Kunde, CRM, CA. "Why Metadata Matters". *The Information Management Journal*, pp.55-61, 2006.Vol.9.No.10.