제21회 한국정보처리학회 춘계학술발표대회 논문집 제11권 제1호 (2004. 5)

메타데이터 상호운용을 위한 프레임워크

최오훈
고려대학교 컴퓨터학과
e-mail : pens@software.korea.ac.kr

A study on Framework for Sharable Metadata Interoperability

*O-Hoon Choi
*Dept. of Computer Science, Korea University

Abstract

It is necessary to resolve the syntax, structure and semantic heterogeneity for sharing information resources. And the representative technologies are XML and Metadata. XML has been used to represent the syntax and structure, and metadata has been used to represent the semantic meaning of information resources. However, various metadata sets in one or more domains that have been developed by each independent organizations without any standards or guidelines, make it difficult to share their information resource. In this paper, we propose an interoperability framework (FSMI, Framework for Sharable Metadata Interoperability) on MDR (Metadata Registry) to increase the interoperability of XML encoded information resources between systems using different metadata sets

1. Introduction

In these days, to share information in distributed environments, we will consider some matters. Using standard description language as WSDL or meta-level format as metadata is necessary to share different format information. And total framework to manage all things is necessary. In this paper, we propose a framework for sharable metadata. FSMI consists of three interfaces – XML [3][12] service on metadata registry, MSDL (Metadata Semantic Description Language)[13], and DTC (Document Translation Component). XML service generates and services standard set of metadata in metadata registry, MSDL [13] describes the difference between local schema of XML documents and standardized metadata in metadata registry, and DTC translate an XML document to target domain's XML document by referencing MSDL. FSMI overcomes the limitations of approaches using static terminology sets like ontology, wordnet, metanet, and provides an environment for business partners using different metadata to share their XML encoded information resources.

2. Related Works

Although metadata set and standard for describing information resource will increase an interoperability of information in its domains, yet we need to additional process for sharing and exchanging information because of discordance among metadata, variety. These additional processes are schema mapping, integrated schema and XML schema [3][12].

2.1. Schema mapping method

The schema mapping method is one to one (1:1) matching for every data elements. First, the system that wants to share and exchange its metadata will match one to one to other system's metadata. And then matched systems can share and exchange their information. A representative example is BizTalk. It is an industry initiative headed by Microsoft to promote Extensible Markup Language (XML) as the common data exchange language for e-commerce and application integration on the Internet.

The schema mapping method is correct and suitable in changing rule. However, mapping cost will increase

exponentially as participating partners, because of matching their metadata one to one. And Updating, deleting and inserting a new metadata, it must maintain the related systems. So this method has a weak point in cost management.

2.2 Integrated schema method

The integrated schema method consists of two steps. In the first step, after analyzing systems of working independently, it will construct an integrated schema and manage it. This method is used on construction EAI (Enterprise Application Integration) system. EAI is a business computing term for plans, methods, and tools aimed at modernizing, consolidating, and coordinating the computer applications in an enterprise. Typically, an enterprise has existing "legacy" applications and databases and wants to continue to use them while adding or migrating to a new set of applications that exploit the Internet, e-commerce, extranets, and other new technologies. EAI may involve developing a new total view of an enterprise's business and its applications, seeing how existing applications fit into the new view, and then devising ways to efficiently reuse what already exists while adding new applications and data.

2.3 XML schema method

Because of simplified and abundant power of expression, a lots of application system domains will use XML for specifying their data and metadata, exchange them. A representative example is ebXML[11] in e-commerce, Rosettanet and ONIX in e-book and digital content. It is ideal system to exchange data with using a standard XML schema. There is no conflict about meaning, structure, and grammar in this system. However, it is impossible to make a standard system to be satisfied all systems. Also it in necessary to make additional method to solve a conflict about non-standard schema

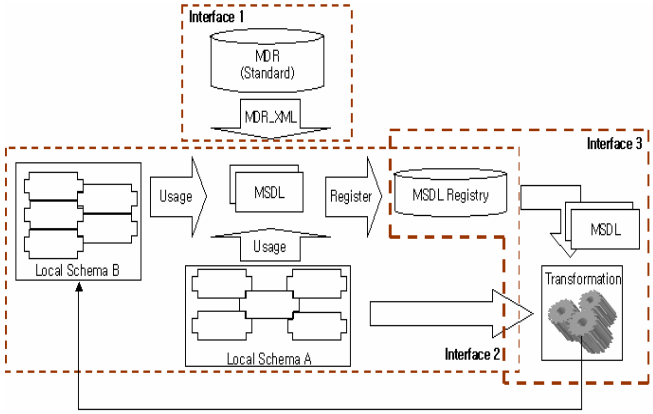
3. FSMI

3.1. Metadata Registry

In this paper, we use MDR (Metadata Registry) for registering, searching, updating and deleting metadata. MDR is useful to manage a standard metadata and to share it. MDR is described in ISO/IEC 11179. A data element is defined as electronic or written representation of the properties of natural-world object classes and is the smallest unit of data that is shared and held in common. It is composed of three parts: Object class, Property and Representation. Object classes are the things about which we wish to collect and store data. Properties are what humans use to distinguish or describe objects classes. And representation is a set of valid value for a data element.

In a data model, an attribute is a characteristic of an entity (entity type, object class, etc) that the enterprise chooses to record as data. Data elements names created from the data model are typically composites of the entity name and the entity attributes names. In an object model, class or object names are used in combination with the class or object attributes to produce data element name. Object models differ from data models in that they may contain additional information about the object or class, such as behavior or operations. A data model or object model, an attribute is thus usually equivalent to a data element. The data element attributes shall be registered and controlled in a standard way in order to achieve consistency in the exchange of information on data elements among data element dictionaries and to enable the comparison of the data elements used in different data management environments. [1] [2][5]

3.2. Modeling for FSMI



<Figure 1> Modeling of FSMI

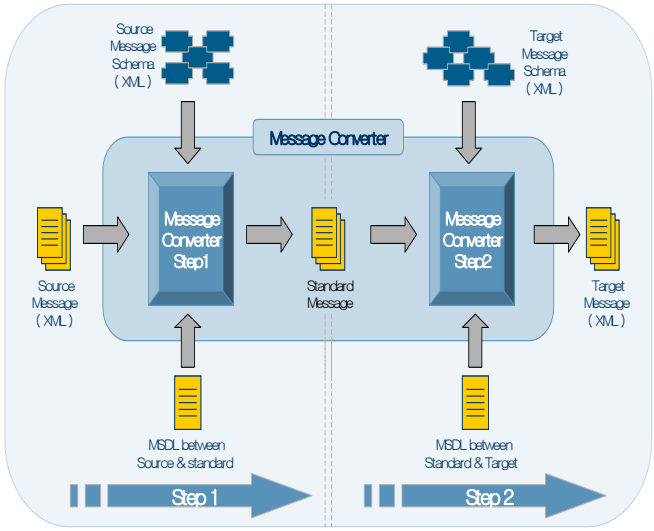
FSMI has two kinds of XML documents which are MDR_XML[13], MSDL[13]. MDR_XML reflect standard data elements of MDR at special time. MDSL is a metadata semantic description language that describes meaning, structure, representation and difference on real system. And it defines a proper communication method, data processing protocol, and service rule among real systems [4].
FSMI consists of three kinds of interface. An interface 1 is a metadata management system based on ISO/IEC 11179, and MDR_XML that is created from MDR. MDR_XML is shortly a set of standard data element in MDR. An interface 2 is modeling about difference between MDR_XML and legacy system then it describes and stores the difference with MSDL. The last one is an interface 3. An interface 3 translates documents among real systems with using MSDL.With interface 3, we will share documents that are represented domain-specified data elements. <Table.1> shows definition and rule of FSMI. <Figure 1> shows an interoperability of FSMI.

Table.1 FSMI Interface

Name	Description
Interface 1 (MDR Service)	- Define XML style service structure and metadata storage. - MDR Metadata Specification MDR_XML Specification & Service
Interface 2 (MSDL)	- Define description about difference of metadata and local schema. -MSDL Service Specification MSDL Registry Architecture & Service
Interface 3 (XML Document Translation)	- Define XML document translation method - XML document translation process and rule specification

3.3. Translation process of FSMI

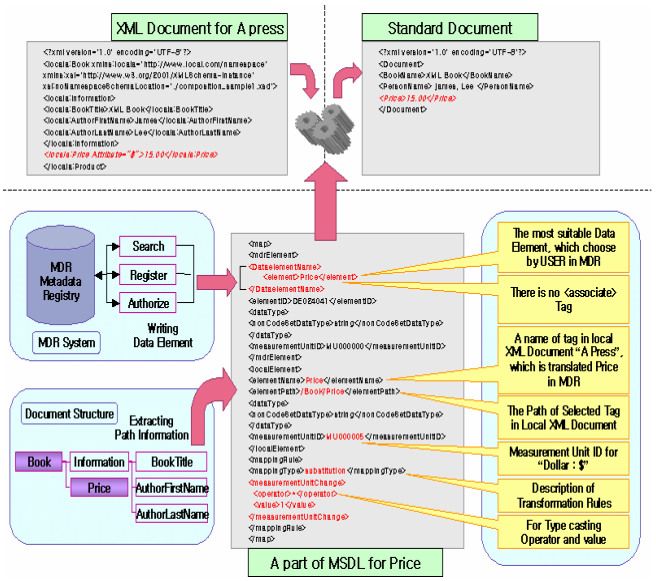
FSMI is a system for translating XML document style source data to different style XML document. This system can create a needed style's document with using MSDL document and XML style's source document. <Figure.2> shows a process of documents translation.



<Figure 2> Process of document translation

A source document's metadata is mapping to a standard metadata in MDR. The mapping rule is described in MDSL. We make a standard style document with using these two documents. The standard style document is a perfect by itself, so any system that want to translate this to their style documents can do without any additional preprocessing.

We will apply two steps mapping rule to solve discordance in document translation processing. In the first step, we translate a source document to a standard document that is structured by standard data elements using source document, its XML schema, and its MSDL. In the second step, we translate a standard document to a target document that is used by other system using target document's XML schema and its MSDL. In each step, we apply translation rules that are described in MSDL. It is need to apply lots of mapping rules to solve discordance of representation and structure. With two steps document translation process, we get a compound mapping rules. <Figure 3> shows a example of translating a source document to standard document using MDSL.



<Figure 3> Translation using MSDL

3. Evaluation

Table.2 Evaluation of FSMI

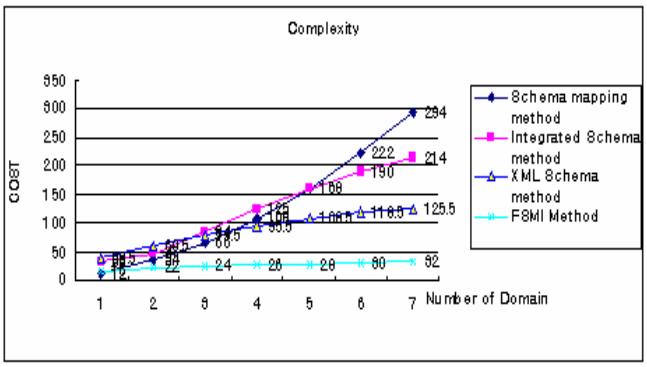
	Schema Mapping method	Integrated Schema Method	XML Schema Method	FSMI Method
Translation Correct	Very High	High	Middle	High
Developing Cost	Low	Middle	Very High	High
Maintenance	Very High	Middle	Low	Low
Extension	Low	Middle	Middle	High
Necessity Of Standard	Not Need	Need (Bottom-up)	Need (Top-down)	Need (Hybrid)
Other's schema overview	Need	Need in Center	Not Need	Not Need
Domain	None	Enterprise Integration	Specific Domain	Inter Domain

In this chapter, to evaluate our supposed FSMI system, we compared FSMI with representative XML document translation rules. These are schema mapping method used in Biztalk, integrated schema method use in EAI, XML schema method used in ebXML.[6][7]

<Table.2> shows the result of functional comparison between FSMI and other method. Schema mapping method manually mapped data element by one to one, so this method has very high degree of correctness. XML schema method could not use undefined metadata, so this method has a loss of data element meaning. In the other side, integrated schema

method an FSMI method managed the needed metadata in center storage system. So it does not appear loss about absence of needed metadata.

<Figure.4> shows cost evaluation graph. The cost evaluation consists of maintenance cost, center complexity, user complexity and construction cost. It shows FSMI is suitable for maintain and construct than other methods. And FSMI based on MDR is a hybrid structure system. It is easy to apply other domains system and framework.[8][9][10]



<Figure 4> Cost Graph of Complexity

4.Conclusion

This paper suggested framework for shareable metadata interoperability (FSMI) to increase information resource interoperability among the systems using different metadata. Using MDR (Metadata Registry) serves standard metadata concept, and MSDL (Metadata Semantic Description Language) describes difference of standard metadata and legacy system's metadata, FSMI can share each other systems' information resource represented with XML and translation component. FSMI do not use statically defined meaning-structure set. Because it dynamically describes relationship of meaning and structure between metadata registry and local XML document using standard model, it can describe a detail element in XML document. So it can reduce loss of means, structure and grammar during XML document translation.

The suggested framework easily manages domain-related standards, and solves the collision of structure and semantics, which may arise during message exchange between different systems, by standardizing the specification of related data elements. Also, it can provide an integrated-environment e-business frame to existing system, without considering the implementation of new system or system extension.

References

[1] "Information Technology - Metadata Registries", ISO/IEC 11179 Part 1 - 6, ISO, 2003

[2] Rachel Heery, Harry Wagner,"A Metadata Registry for the Semantic Web", D-Lib Magazine 8(5), 2002.

[3] D. Connolly, J. Bosak, "Extensible Markup Language (XML)", http://www.w3c.org/XML, 1997.

[4] C. Paice, "A Thesaural Model of Information Retrieval", Information Processing and Management, 27(5):433-447, 1991.

[5] IEEE Learning Technology Standards Committee's Learning Object Meta-data Working Group, Approved Working Draft WD5 Learning Object Meta-data Scheme, http://ltsc.ieee.org/wg12/

[6] C. Batini, M. Lenzerini and S.B. Navathe, "A comparative analysis of methodologies for database schema integration", ACM Computing Surveys, 18(4):323-364, December 1986.

[7] Bayardo, R. et al. "InfoSleuth: Agent-based Semantic Integration of Information in Open and Dynamic Environments". Proceedings of ACM SIGMOD Conference on Management of Data, Tucson, Arizona, pp. 195-206, May 1997

[8] N. Guarino, C. Masolo and G. Vetere, Ontoseek: Content-based Access to the Web, IEEE Intelligent Systems, Vol. 14, No. 3, pp. 70-80, May/June 1999

[9] B. Amann and I. Fundulaki. Integrating Ontologies and Thesauri to Build RDF Schemas. In ECDL-99: Research and Advanced Technologies for Digital Libraries, Lecture Notes in Computer Science, pages 234--253, Paris, France, September 1999. Springer-Verlag.

[10] Zhan Cui, Dean Jones and Paul O'brien,"Semantic B2B Integration: Issues in Ontology-based Approaches", SIGMOD Record, Vol. 31, No. 1, March 2002.

[11] www.ebxml.org

[12] D. Connolly and J. Bosak, "Extensible Markup Language (XML)," 1997, http:// www.w3c.org/XML/

[13] O-Hoon Choi, Jungil-Kim, Doo-kwon Baik, "A MDR based Hybrid e-Business Framework", pp 67~69, 2002, KISS