MDR-based Framework for Sharing Metadata in Ubiquitous Computing Environment

O-Hoon Choi[1], Jung-Eun Lim[1], and Doo-Kwon Baik[1]

[1] Department of Computer Science and Engineering, KOREA University,
SungBok Gu, Anam Dong, 135-701, Korea
{pens,jelim,baik}@software.korea.ac.kr

Abstract. : To resolve the syntax, structure and semantic heterogeneity for sharing information resources, the representative technologies are XML and Metadata. XML is used to represent the syntax and structure of information resources but the various XML schema definitions that have been developed by independent organizations without any standards or guidelines, make it difficult to share the semantic meaning of XML encoded information resources. In this paper, we propose a mechanism, named MSDL that represents the exact meaning of XML tags by describing the structural and semantic differences with standard metadata in metadata registries. MSDL overcomes the limitations of other approaches with respect to exactness, flexibility and standardization, and provides an environment for business partners using different metadata to share their XML encoded information resources.

1 Introduction

XML is a standardized text format designed specifically for transmitting structured data to web applications. With XML, application designers can create sets of data element tags and structures that define and describe the information contained in a document, database, object, catalog or general application, all in the name of facilitating data interchange [1]. XML documents, the unit of data exchange, are based on schemas (XML schema definitions). XML schema is a specification defining the structure of tags and plays a role of metadata on XML documents. Metadata is a data about data. And it is essential to information sharing. It describes attribute, property, value type and context on data.

One important point is that XML documents based on same schema definition have same structure and tags. Therefore, it is possible to share data between applications if they use standardized schema definition. Such standardized schema definitions include CML DTD [2] used in chemistry domain, Math DTD [3] in mathematics, ONIX [4] in digital library domain, and so on. So, common schema definition makes it possible to share tag elements and attributes defined in schema definition, and to understand semantics of the XML documents.

However, in many cases, each system (organization) has developed their own schema definition and use XML documents based on the schema. There is no interoperability with other systems without manual converting processes.
For example, if we represent “author” of a book named “Database Design”, Dublin Core [5] which is a metadata standard for describing general properties of information resources, uses a term “creator”, but ONIX [4] which is a metadata standard for digital contents, uses a term “author”
In this paper, we first categorize heterogeneity between XML documents – syntactic, structural and semantic heterogeneity, and second design a framework (FSMI, Framework for Sharable Metadata Interoperability) to increase the interoperability of XML encoded information resources using different metadata sets. In this framework we propose a mechanism, named MSDL (Metadata Semantic Description Language) that represents the exact meaning of XML tags by describing the structural and semantic differences with standard metadata in metadata registries. With MSDL, users can describe the syntactic, structural and semantic difference of their XML schema to other’s schema, and based on this description, can convert other system’s XML document to a document with their own schema.

2. Related Works

In this section, we introduce some other approaches that enable the interoperability of XML documents having different schema definitions. The most straightforward approach for XML document interoperability is 1:1 mapping. In 1:1 mapping, each system tries to map their XML schema to other’s schema. Microsoft’s BizTalk [6] is the representative software. Users of BizTalk create a mapping table for each other application, and interchange their XML documents. After receiving, it converts the documents according to receiver’s understandable format.
Maintaining integrated global schema is another method. EAI solutions analyze each application’s data and processes, create global mapping tables and rules, and enable each system to exchange XML encoded data. Also, many organizations develop standard metadata (XML schema) and member of the domain use it as a data interchange format. The representative examples are ebXML [7] and Rosettanet [8] in electric commerce domain, and ONIX in digital contents (e-book) domain. This approach is very simple and powerful when all members use it as their data standard.
These three approaches have advantages and disadvantages in exactness, flexibility, initial cost and maintenance cost. In the last section, we will compare our proposed approach to them.

3. Metadata Model

We use data element model of ISO 11179 as a standard metadata model for comparison. Data element is a basic unit of identification, description and value representation unit of a data.

Definition 1 (Data Element) *Data Element is a basic unit for definition, identification and representation of data. It consists of data element concept and permissible value domain.*

DE = {(dec, r)| dec∈DEC, r∈R}
DEC = {(oc, p)| oc∈OC, p∈P}

The combination of an object class and a property is a data element concept (DEC). A DEC is a concept that can be represented in the form of a data element, described independently of any particular representation. In the examples above, annual household income actually names a DEC, which has two possible representations associated with it. Therefore, a data element can also be seen to be composed of two parts: a data element concept and a representation. Data element is composed of three parts as Figure 1 [9].

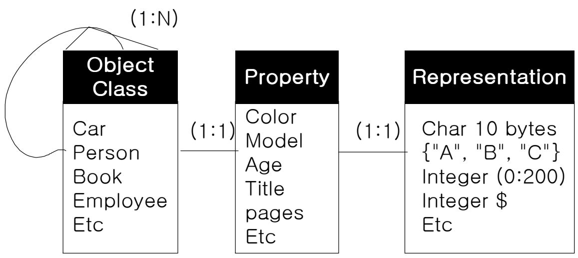

Fig.1. Structure of Data Element

Object class (OC) is a set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behavior follow the same rules. Examples of object classes are cars, persons, books, employees, orders, etc. Property (P) is what humans use to distinguish or describe objects. Examples of properties are color, model, sex, age, income, address, price, etc. The most important aspect of the representation (R) part of a data element is the value domain. A value domain is a set of permissible (or valid) values for a data element. For example, the data element representing annual household income may have the set of non-negative integers (with unit of dollars) as a set of valid values. Figure 2 shows some examples related on Definition 1.

OC={person, car, employee, book, product, ...}
P={name, age, title, cost, ...}
R={char 30byte,{blue, red, green},{0..100},...}
DEC={personname, carname, booktitle, productcost, ...}
DE={(personname, char 30byte), (carname, char 30byte), (productcost,dollar),...}

Fig.2. Examples on Data Element

A metadata registry (MDR) supports registration, authentication and service of such data elements. Each data elements have a unique identifier, all users and programs using the metadata registry identify a data element by the identifier. A data element also contains it definition, name, context, representation format, and so on.

4. MSDL

4.1. Information Heterogeneity

Table 1 shows the type of heterogeneity between standard metadata based on ISO 11179 MDR and XML schema definitions. Semantic heterogeneity means that the two xml tags has same meaning but different names. Structural heterogeneity means that two sets of xml tag have same meaning but different structures including composition, decomposition and rearrangement. Syntactic heterogeneity means the difference of data values in XML tags. Conflicts in this heterogeneity include measurement unit, data type and data code sets.

Table 1. Heterogeneity of XML Documents

	Description and Examples
Semantic Heterogeneity	Different terminologies have same semantics in context. ex) title – book name, price - cost
Structural Heterogeneity	Structurally different with same semantics ex1) name - first name, last name ex2) first name, last name – name ex3) last name - first name, last name
Syntactic Heterogeneity	Semantically, structurally same, but have different data type, measurement unit and code sets ex1) {kr, jp, cn} - {korea, japan, china} ex2) mile - kilometer ex3) integer - float

4.2. MSDL

The key idea of this paper is that users of each system describe the difference between their XML schema and standard metadata registered in metadata registry. Then they interchange their XML documents with other system as resolving the difference of other's document using the descriptions.

We propose MSDL (Metadata Semantic Description Language) as a tool for describing the difference between local XML schema and standard metadata in MDR.

MSDL consists of namespace on the location of MDR and local schema, correspondent information (MAP) and transformation rules for converting local document to/from standard document.

Definition 2 (MSDL) *A language for describing the difference between local schema and standard metadata. It uses XML syntax.*

MSDL= (mdrNameSpace, xmlNameSpace, MAP, CODESET)

Where, mdrNamespace and localNamespace must appear once in a MSDL document, but MAP appears as much as the number of tags to be converted.

Definition 2.1 (MAP) *MAP defines semantic, structural and syntactic conflicts between tags in XML documents and metadata in MDR.*

$$MAP=\{(mdrE, localE, maptype) \mid mdrE \subset DE, localE \subset XS, maptype \in MapT\}$$

Where, mdrE is a subset of standard metadata in MDR, and has one or more elements. localE is a subset of metadata defined in local XML schema, and also has one or more elements.

Definition 2.2 (Mapping Type) *MapT defines structural conflicts and syntactic conflicts between metadata with same semantics.*

$$MapT=\{(sct, vct \mid sct \in SCT, vct \in VCT)$$
$$SCT=\{substitution, composition, decomposition, rearrangement\}$$
$$VCT=\{(CodeSetmdr, CodeSetlocal), (MUnit, cvalue),(TypeCasting, cfunction)\}$$

Where, number(CodeSetmdr) = number(CodeSetlocal),
Munit = measurement unit (m, km, kg, etc.),
Cvalue = constant for measurement unit change,
cfunction = name of system support function for type casting

SCT (Structural Change Type) has one of the four elements – substitution, composition, decomposition, and rearrangement. Where, substitution is for resolving semantic conflict and other three elements are for structural conflict. For example, "book title" and "book name" have same meaning but different name, in this case, "substitution" is applied. On the other hand, a "name" tag may correspond to two tags "first name" and "last name" In this case, "composition" or "decomposition" is applied.

VCT (Value Change Type) has one of the three elements – code set conflict, measurement unit conflict and data type conflict. In code set conflict, we consider only the case of two code sets having same number of elements. A MSDL document must have one or more mapping type to describe the difference between standard metadata and local schema. Figure 3 shows an example of MDSL document mapping local element "Book price" to standard metadata element "Price" In this example, two elements have semantic conflict and syntactic (measurement unit) conflict.

```
<map>
 <mdrElement>
  <DataelementName>
    <element>Price</element>
```

```
  </DataelementName>
  <elementID>DE024041</elementID>
  <dataType>
   <nonCodeSetDataType>string</nonCodeSetDataType>
  </dataType>
  <measurementUnitID>MU000000</measurementUnitID>
 </mdrElement>
 <localElement>
  <elementName>Price</elementName>
  <elementPath>/BooK/Price</elementPath>
  <dataType>
   <nonCodeSetDataType>string</nonCodeSetDataType>
  </dataType>
  <measurementUnitID>MU000001</measurementUnitID>
 </localElement>
 <mappingRule>
   <mappingType>substitution</mappingType>
 </mappingRule>
</map>
```

Fig.3. An Example of MSDL Document

4.3. XML Conversion Process using MSDL

MSDL just specifies information about conflicts of local schema and standard metadata. Thus, it is necessary to develop a system which transform imported XML document to self-understandable one using MSDL. We call this system as FSMI (Framework for Sharable Metadata Interoperability). FSMI is a system for transforming XML encoded source data to fit other system's XML document using each system's predefined MDSL documents. Figure 4 shows the transforming process.

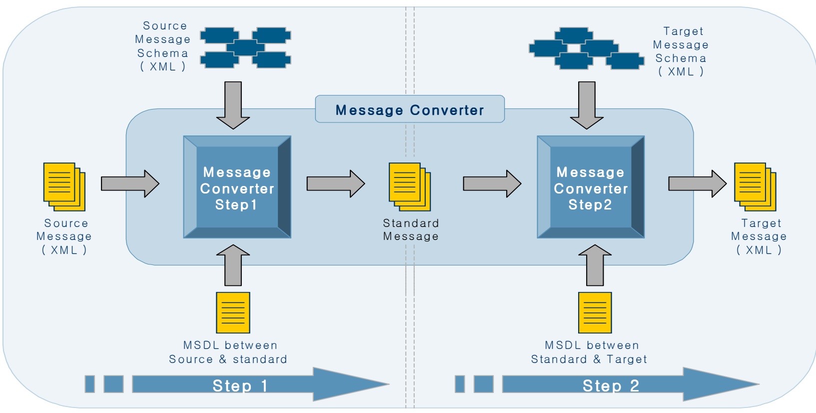

Fig.4. XML document transform process

XML document converting process consists of two steps. The first step is to generate a standard document with standard data element in MDR from source document, local

XML schema and MSDL of the schema. Then, in the second step, standard document is converted to target document using target document's XML schema and MSDL.
In the step 1 and step 2, FSMI applies mapping rules in MSDL to resolve structural and syntactical heterogeneities (conflicts). Generally, these two types of heterogeneity co-exist in a document, it is necessary to apply mapping rules with combined manner.
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 it's 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 it's 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 5 shows an example of translating a source document to standard document using MSDL.

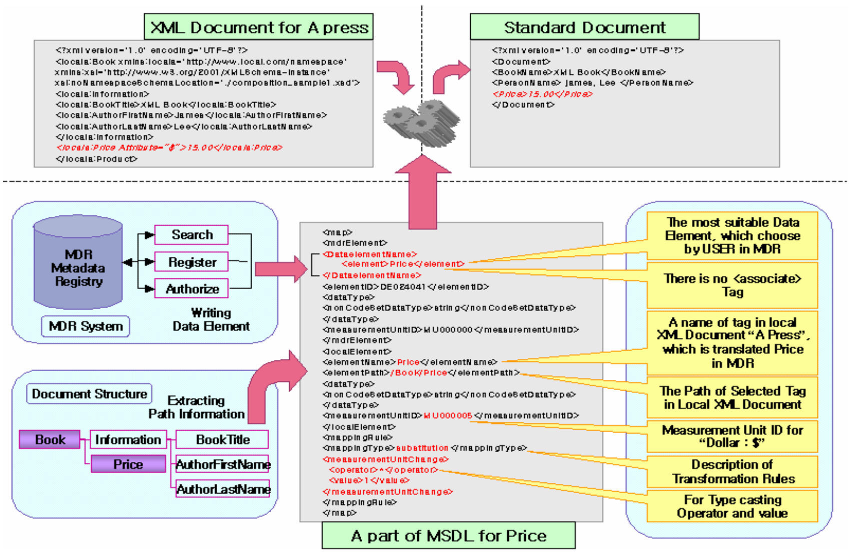

Fig 5. Translation using MSDL

5. Comparison

1:1 mapping approach is relatively exact because experts of each system directly map to other system's schema. But, the mapping cost increases exponentially as the number of related systems. Especially, in case of schema change, the maintenance cost will be high. Global schema can cover large parts of each system's characteristics of schema through analysis process, and the maintaining cost is very low in case of schema change because systems only have to change related parts of global schema. But the initial building cost of global schema may be high.

Table 2. Comparison to Other Approaches.

	1:1 Mapping	Global Schema	Standard Schema	MSDL
Conversion accuracy	Very High	High	Middle	High
Initial cost	Low	Middle	Very high	High
Maintenance cost	Very High	Middle	Low	Low
Extensibility	Low	Middle	Middle	High
Global schema	Not exist	Exist (Bottom-up)	Exist (Top-down)	Exist (Hybrid)
Application Domain	Not limited	Enterprise Integration	Specific Domain	Inter Domain

If all systems use standardized XML schema, then there are no syntactic, structural and semantic conflicts. It is very ideal status on data sharing and integration. But It is impossible to make a standard system (schema) that support all the case of system, and necessary to consider another processes for data that standard does not cover.
Also, a domain may have more than one metadata standards. For example, in digital content domain have metadata standard including ONIX, MARC, DC, and MPEG7 and in electric commerce domain, they have standard including cXML, ebXML, Rosettanet, eCo, and so on. Table 2 shows the characteristics of each approaches including MSDL. With MSDL, conversion accuracy is relatively high, because all the needed part of schema for sharing can be covered in MDR using registering process. Figure 6 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. [10,11,12]

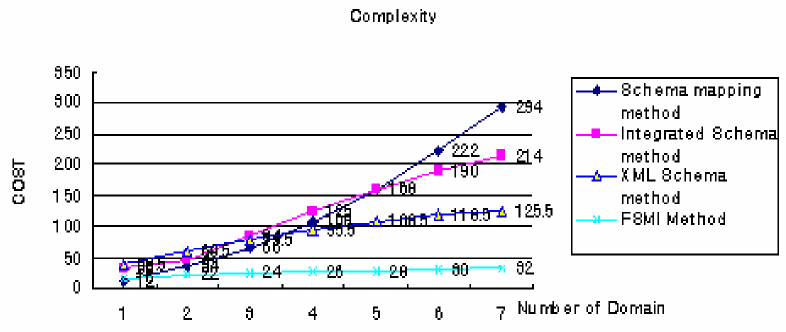

Fig 6. Cost Graph of Complexity

Most of all, maintenance cost in case of changing schema, is very low and extensibility in case of adding new metadata is relatively high. However, the building cost of

MDR is very high and time consumed, and there may be some data losses in two steps transforming process.

6. Conclusion and Future Works

XML is a de facto standard for representing information resources in distributed environment. But it is necessary to share metadata of XML document defining syntax, structure and semantic of the document.

In this paper, we propose MSDL as a tool for describing difference between standard metadata and local XML schemas to support interoperability of XML documents. With MSDL, users can describe the syntactic, structural and semantic difference of their XML schema to other's schema, and based on this description, can convert other system's XML document to a document with their own schema.

Users, who wish to exchange their document, just describe the difference of their XML schema with standard metadata in MDR, and make a MSDL document on the schema, and then other system can convert the document to understandable ones.

MSDL overcomes limitations of 1:1 mapping, global schema and standard schema approaches, providing low maintenance cost on schema changes, high extensibility in increase of systems, and domain independent document sharing.

In the future, we have plans to develop the converting system using MSDL, and extend it to support legacy solutions including EAI, B2Bi, Web Service, and so on.

References

1. Rik Drummond, Kay Spearman, "XML Set to Change the Face of E-Commerce", Network Computing, V.9, N.8, pp.140-144, May 1998.
2. "Chemical Markup Language", Version 1.0, Jan.1997. available at http://www.venus.co.uk/omf/cml/.
3. "Mathematical Markup Language (MathML) 1.0 Specification", W3C, July 1998. available at http://www.w3c.org /TR/1998/REC-MathML-19980407
4. ONline Information eXchange, http://editeur.org/
5. http://dublincore.org/
6. http://www.microsoft.com/biztalk/
7. http://www.ebxml.org
8. http://www.rosettanet.org/RosettaNet/
9. "Information Technology - Metadata Registries", ISO/IEC 11179 Part 1 - 6, ISO, 2003.
10. 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
11. 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.
12. Zhan Cui, Dean Jones and Paul O'brien,"Semantic B2B Integration: Issues in Ontology-based Approaches", SIGMOD Record, Vol. 31, No. 1, March 2002.