Metadata management system: design and implementation

Shien-Chiang Yu
Kun-Yung Lu and
Ruey-Shun Chen

The authors

Shien-Chiang Yu and **Kun-Yung Lu** are both PhD Students and **Ruey-Shun Chen** is Associate Professor, all at the Institute of Information Management, National Chiao-Tung University, Taiwan, ROC.

Keywords

Digital mapping, Organizational processes, Information management, Systems design, Object-oriented computing, Data processing

Abstract

Commonly, an organizational information system may have various data types and directory formats. It usually employs different metadata formats to represent the documents. Although the metadata system can cover the different formats of documents, there still exist the integration problems in various metadata systems. This may lower the performance of data processing and impede information sharing. Research focuses on the model of using multi-XML schema to construct an XML system framework. Through a complete hierarchical tree structure definition of inner elements, the proposed model can settle the weakness of traditional object-oriented languages in information sharing; it can also eliminate the constraints of storage and management among heterogeneous metadata while processing different metadata information.

Electronic access

The Emerald Research Register for this journal is available at
http://www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
http://www.emeraldinsight.com/0264-0473.htm

The Electronic Library
Volume 21 · Number 2 · 2003 · pp. 154-164
© MCB UP Limited · ISSN 0264-0473
DOI 10.1108/02640470310470525

1. Introduction

Owing to the past decades of fast changing information technologies, many varied structures and formats of documents in electronic type are employed to manage the organizational information resources in different areas such as geography, museums, technology, literature, music, etc. This significantly enhances the performance of the operation and management. However, on the other hand, different formats may crucially affect the system integration and information sharing among these organizations. The current information structure has been servicing the needs of printing and audio/video media, but its scope will be much broader because of the evolutional changes of application and development of electronic publication, user interface, and information media. Under this circumstance, it is necessary to develop a new information operation model which shall effectively reduce the cost of system development and to facilitate automatic data transactions.

Nowadays, in addition to the libraries, other domains are also concerned with the research and development of new information organizations. Metadata, a fundamental role of the digital content, has now become an important part of the global information construction in planning, processing, restoring and managing. Several well-known metadata sets are listed below (Vellucci, 2000):

- Computer Interchange of Museum Information (CIMI).
- Federal Geographic Data Committee (FGDC) Content Standards for Digital Geo-spatial Metadata.
- Dublin Core (DC) Metadata Element Set.
- EDUCOM Instructional Management Systems.
- Encoded Archival Description (EAD).
- Government Information Locator Service (GILS).
- IAFA Templates (IAFA/WHOIS++).
- USMARC Formats.
- Resource Description Framework (RDF).

This work was partially supported by the National Science Council of Taiwan under grant NSC 89-2750-P-002-012. For further readings on this project in Chinese, see the National Science Council Web site at http://www.nsc.gov.tw

- Text Encoding Initiative (TEI) Header.
- Visual Resources Association (VRA) Core Data.

By using these metadata sets, heterogeneous documents can be easily integrated.

Commonly, an organizational information system may have various data types and directory formats. It usually employs different metadata formats to represent the documents. Although the metadata system can cover the different format of documents, it still has integration problems in various metadata systems. For instance, this is where there is great difference between the systems of the traditional libraries and digital libraries.

To achieve the target of data exchange among different metadata systems, except the specific format transforming, this paper proposes a multi-XML schema following the XML standards in data exporting and importing to design the metadata system. Through a complete hierarchical tree structure definition of inner element using multi-XML schema, the proposed model can settle the weakness of information sharing of traditional object-oriented languages. In addition, the traditional object component technique has difficulties in presenting the relationships between class and entity. The proposed model can eliminate the constraints of storage and management among heterogeneous metadata, analyze and construct a system while processing information of various metadata, integrate the system framework of retrieval, and interact with the authority control.

2. Relevant literature

In this section, we briefly describe some well-known software techniques and fundamental metadata systems. These techniques are employed as the basic working conceptions of the proposed model.

2.1 Metadata

Metadata is data about data. This definition primary descended from the "Metadata workshop" conference held by the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA) in March 1995 (Weibel *et al.*, 1995). Typically, metadata supports a number of functions: location, discovery, documentation, evaluation, selection and others (Dempsey and Heery, 1997). There are more than 20 different types of international standard (or about to be) metadata existing among the domains for different requirements (El-Sherbini, 2001).

It is an essential goal to upgrade the accessibility of data retrieving and to obtain information among different databases, systems, and Web sites. Using an identical rich-format metadata throughout all platforms and Web sites will efficiently speed up the searching. However, the condition is that the descriptions of metadata must be the same. If the descriptions are not the same, there must be a cross-mapping for different metadata. To sum up: compatibility is necessary; searching the cross-operation between different metadata is expected (Gilliland-Swetland, 1998).

Recently, some projects have been proposed regarding the cross-operation between metadata. The American Congress Library initiated a program, which links up three structures, USMARC, Dublin Core and GILS. Cross-operation is a key requirement to build up a consistent resource repository. It relies on minimizing the dissimilarities among metadata formats. A further ideal is to establish a higher level of super-metadata for all metadata interoperability. In this way, it facilitates the success in integration, and each metadata keeps its own character (Chilvers and Feather, 1998).

2.2 XML schema

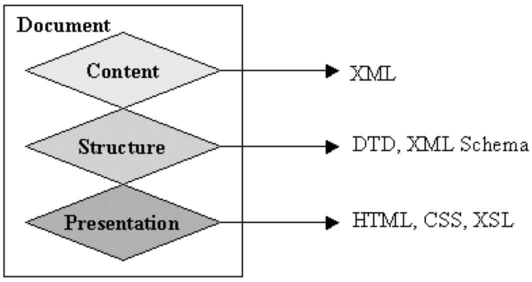
XML is a simplified sub-class of SGML. The SGML experts and W3C announced version 1.0 in February 1998. Unlike the fixed tag sets of HTML, XML allows users to define the tags when necessary. Consequently, the tags clearly explain the meanings of the data content. Using related tags to identify certain data items or data groups makes data in XML documents easily extracted for further process (W3C, 1998).

Based on the language-neutral and platform-neutral of XML, automation systems will be readily modified for data exchanging. Some successful e-commerce examples show that XML has good capability to integrate different types of documents and to exchange data among machines. There are two forms of XML: one is the Well-formed XML, and the other the Valid XML. The existence of document type definition (DTD) makes the difference between these two forms. DTD is a set of rules primarily to define and regulate the structure of an XML document. With DTD, applications and parsers can verify the validity of XML documents and authoring tools can generate XML documents. But, DTD is the sub-class of SGML and is mostly transplanted from the type definitions of SGML. Thus, the syntax of DTD is hard to learn, not to mention the insufficiency of metadata definitions. For instance, DTD handles only text format data, not including the declarations of other formats; DTD provides only the declaration of the default value for attribute field, not the element field; DTD cannot treat an XML document as an object type for redirection (Cagle, 2000).

Owing to the above constraints, the W3C working group studied many proposals from industry for what is to be the replacement of DTD for XML. The first draft of *XML Schema Requirements* was published on 15 February 1993 (St Laurent, 1999). W3C referred the following proposals that have been submitted to the schema working group. They are document content description (DCD), document definition markup language (DDML), resource description framework (RDF), schema for object-oriented XML (SOX) and XML data. W3C integrated the merits and characters from these proposals, and issued the publication of the W3C XML schema specification as a W3C recommendation on 2 May 2001 (W3C, 2001) (Figure 1).

XML schema is an XML-based schema (or metadata) description language that actually provides two pieces of critical data: a definition of the acceptable structure of the elements that make up a valid type of XML document and a representation of the data type used by the document. XML schema is not only an attempt to simplify existing schemas, it is an effort to create a language capable of defining the set of constraints of any possible data resource. Based on the comparison between DTD and XML schema, XML schema provides more advantages (Ioannides, 2000) including:

Figure 1 Document uses different standards in content, structure and presentation



- More expressions in XML schema than in DTD. It is not only usable by a wide variety of applications that employ XML, but simple enough to implement with the modest design and runtime resources.
- Explicitly placed limits on the number of elements contained within a structure, as well as declaring whether a given XML node's content is closed (containing only the specifically declared sub-elements) or open (containing any type of sub-elements).
- Coordinate with relevant W3C specs (XML information set, links, namespaces, pointers, style and syntax, as well as DOM, HTML, and RDF schema). Through the use of namespace, developers recognize the need for universal names for document constructs, and are able to embed many XML schemas into one document.
- Ability to define a wider variety of data types.
- With the ability to define an element reusing information already contained in another element, XML schema can inherit or reuse the information from the other element. Object-oriented characteristics are also definable in the XML schema, such as archetype definition, inheritance and encapsulation.

3. Concept and principle

In this paper, we proposed a metadata system based on XML schema to control and manage the different types of metadata. The main design points are focused on structure, authority control, depth, scope, and database technical. The design concept and principles are described as the following.

3.1 Structure

Metadata is employed to describe the resources. In cross-referring most of the currently used or recognized metadata definitions, it is found that they contain almost the same kind of structure. The only differences among metadata are the complexity and the design point of view. The major execution capabilities of handled metadata include the functionality of the parser for well-formed XML and the authoring tool for editing and validating the XML schema structure.

3.2 Authority control

Traditionally, the purpose of authority control has been to bring consistency to library catalogues. The authority control process is directed at the access points contained in catalogue records, i.e. names, titles, and subjects. It ensures that these access points are unique and consistent in content and form, and provides a network of linkages for variant and related headings in the catalogue. It supports the finding task by ensuring that each entity has a unique name; that only one name is used for each entity; that variant name forms are represented and linked in some way; and that related names, titles, and subjects are collocated in the catalogue (Vellucci, 2000, p. 34).

3.3 Depth

Metadata describes the attributes and characteristics of the resource. It differs little from the purposes of the library category. Fields are based on hierarchical structure and subdivided downwards. It builds up the subordinate relationships between elements and sub-elements. Each element's character relies on the attribute value. Basically, there are two sorts of fields: fixed-length fields and variable fields. Fixed-length fields are general types (e.g. ID number, ISBN, ISSN) and character-indication types (e.g. MARC 21's 008 "GENERAL INFORMATION" tag. Each character position contains its special meaning and range). Sub-field of a field which, be it fixed-length fields or variable fields, shall contain both fixed-length and variable fields. According to the reason above, the process ability of the system must cover the various situations.

3.4 Scope

Metadata not only describes the information resource, more importantly it marks the relationship between objects. Metadata emphasizes the description of the data instance itself and also explains the relationship among different metadata. The cascading connections must be able to span from a simple object to time, space, people, and event. Thus the system must consider the scope of different applications, i.e. various types and formats of metadata; various different types of users; various types of resources; and various data suppliers.

3.5 Database technical

XML and XML schema are of hierarchic structure. Relational database system is based on the relationship between tables. Thus, the database structure of this system building the relational tables shall meet the following three criteria:

(1) Relational database tables with the ability to store hierarchic XML and XML schema.
(2) Index tables for different access points while restoring and managing metadata.
(3) Tables for authority control with different accessing points.

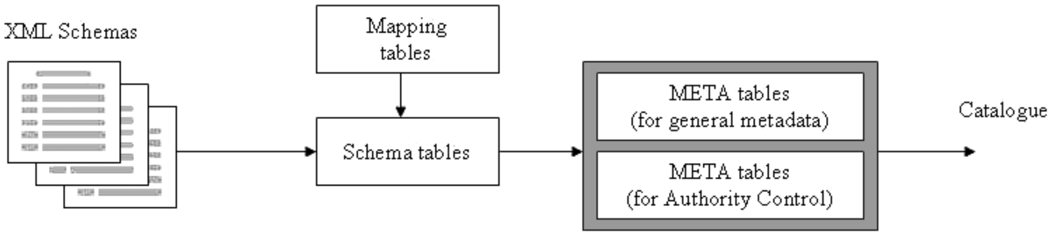
The methodology of database design is using linked list structure (Kruse *et al.*, 1997). The idea of a linked list is, for every metadata in the list, to put a pointer into the structure giving the location of the next structure in the list. By this method, metadata can keep in columns of tables of database to fulfill hierarchical structure, and even one field of metadata can link to another metadata.

4. Integrating mechanism

To meet the above requirements, we construct a multi-XML schema imitating the hierarchic structure of XML as the schema resource of the data structure. It aims at the compatibility of various metadata schemas by the XML framework. To integrate different metadata systems, a database, containing the tables of schema, mapping, META, and authority control, is employed. Figure 2 describes the relationships and working principles of these tables:

- *Schema tables.* Define the metadata structure of the system (for actual consideration, this system can handle XML schema and DTD).
- *Mapping tables.* Integrate the XML schema/DTD and define the access point of each element.
- *META tables (parameter files)*. Combine the mapping table and XML schema to

Figure 2 The referencing relationships of tables in importing XML schemas


XML Schemas
Mapping tables
Schema tables
META tables (for general metadata)
META tables (for Authority Control)
Catalogue

generate catalogue parameter files for the system operation.

- *Authority control tables.*

5. Information organization

In the design phase, the information organization is an important initial step. The method of information arrangement determines the system database structure. In this paper, the system organization based on the principle of unified model language (UML) (Booch *et al.*, 1999) is divided into five groups: pre-work, system, catalogue, indexing, and retrieval. Figure 3 describes the working process of a system activity according to the above classifications. Designers can allocate these groups to relative objects to follow up the analysis.

To cope with the user activity, an UML-type of user-case diagram can be used to describe the outer demand and major system functionality. Figure 4 shows the action of the entire system.

5.1 Module architecture

According to the analysis of information organization mentioned in Figure 3, this paper employs the following modules to construct the proposed model. All these modules are based on XML in designing the structure of processing documents and the type definition (XML schema/DTD).

- *Schema construct module* – supplying XML schema input and transforming into the function of system schema structure.
- *Catalogue module* – including the record maintenance functions of insert/update/delete for authority control and metadata.
- *Metadata import/export module* – using XML as the basic format for import/export, declaring well-formed or valid XML.
- *Enquiry module* – including Web and open public access catalogue (OPAC) interface for retrieving.

5.2 Enquiry function

To meet the increasing requirements of the integrated retrieval of distributed information systems, the retrieval function of the system

Figure 3 System activity diagram

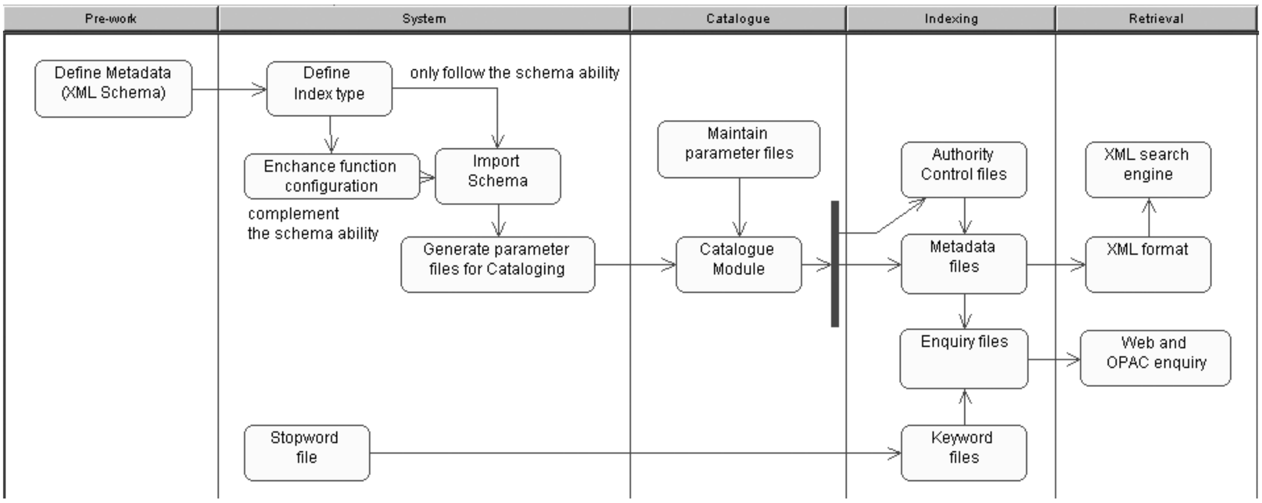
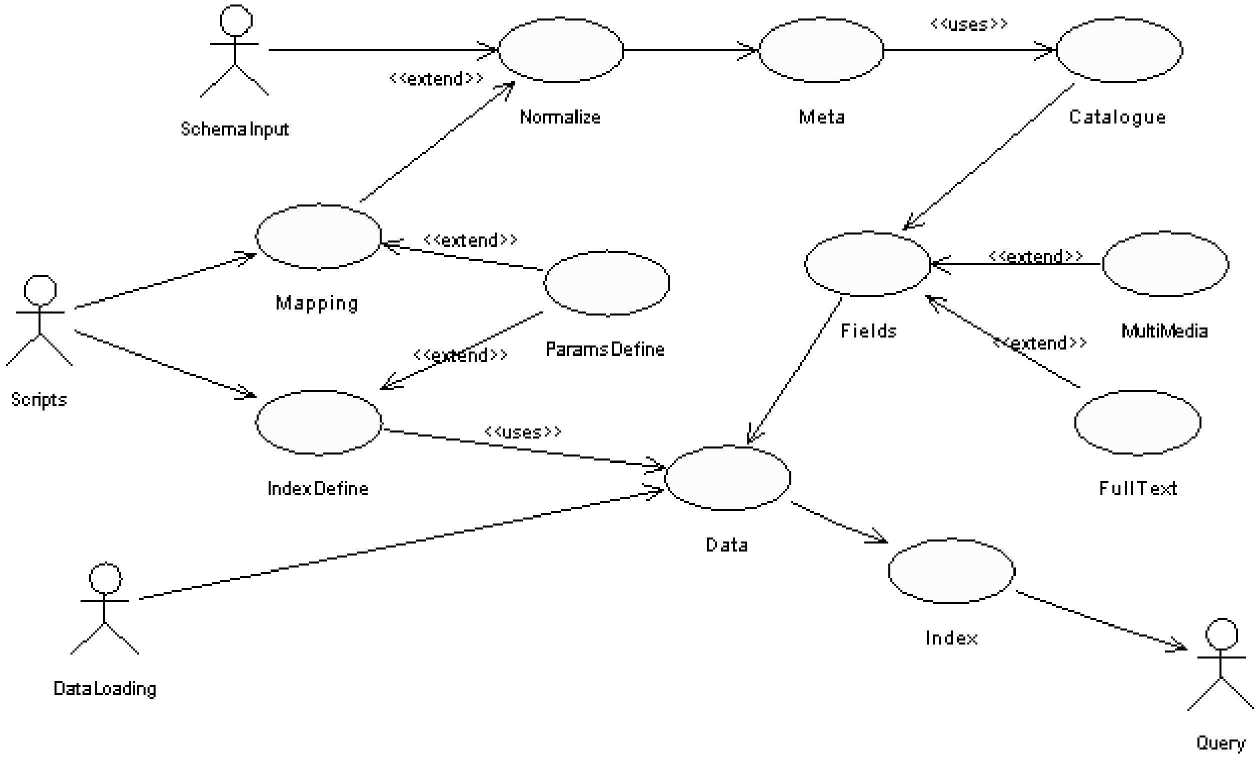


Figure 4 User-case diagram showing the activities of entire system



should offer integrated, faster and exact searching for metadata documents. It should also offer deeper and broader information for presenting the value of metadata. Though access methods and query methods may be different and the data files may have various structures, the basic principle of retrieval is similar, particular in database retrieval, through the standard interface of structured query language (SQL). However, while the system is processing metadata, the differences in searching between English and CJK (Chinese, Japanese, and Korean) is one of the major concerns:

- Besides Unicode characters, an English character is processed by only one byte; CJK characters are processed by at least double-bytes.
- The length of English vocabulary is not fixed, but CJK vocabulary is.
- English vocabularies often use space as a delimiter, but CJK vocabulary does not.
- English vocabularies are arranged in alphabetical order. CJK uses stroke-count, strokes, and radicals for arrangement.

Based on the above differences of word processing, English and CJK vocabularies in metadata document content must be separated and indexed. The main design rules are the following:

- vocabulary retrieval is for English, phrase retrieval is for CJK;
- fault-tolerance retrieval is necessary for English, e.g. sounded search;
- English vocabulary has the characteristic of singular, plural, third person, and tense;
- English vocabulary is normal, while CJK phrase needs to be accessed by truncated retrieval;
- space phrase and wrap lines phrase retrievals are necessary for CJK;
- phrase segmentation retrieval is necessary for CJK.

5.3 System functions

Based on the demands of all modules in this system, the completed main functions are listed below:

- Meta schema for loading XML schema to establish database.
- Accessing points and authority control.
- Management and description of digitized data.
- General enquiry.
- Enquiry and reference for authority control items.
- User priority management and control.
- Multilingual interfaces.

6. Implementation

The proposed system is composed of the modules of schema constructor, catalogue, metadata import/export, and enquiry. The main functions of these modules are described below.

6.1 Schema construct module

The schema constructor is employed to provide the function for importing XML schema and establishing the system schema. It contains three main execution parts:

(1) *Storing mechanism.* A database is employed to store the information of imported XML schema. It is a nested structure similar to the tables of object-oriented database (see Figure 5). The XML schema file should be decomposed into tables before importing into the system. This mechanism simplifies the complexity of front-end software development and rear-end database accessibility by using the relational database tables as physical structure.

(2) *Mapping mechanism.* The XML schema declaration does not include the field definitions of access point, extra function, input length, the item of authority control and default value. The system operators must do some exceptional operations for the imported XML schema such as transformation, integration and use partial. To meet these requirements, this system supplies a set of mapping tables as the intermediary files for XML schema definition transfer.

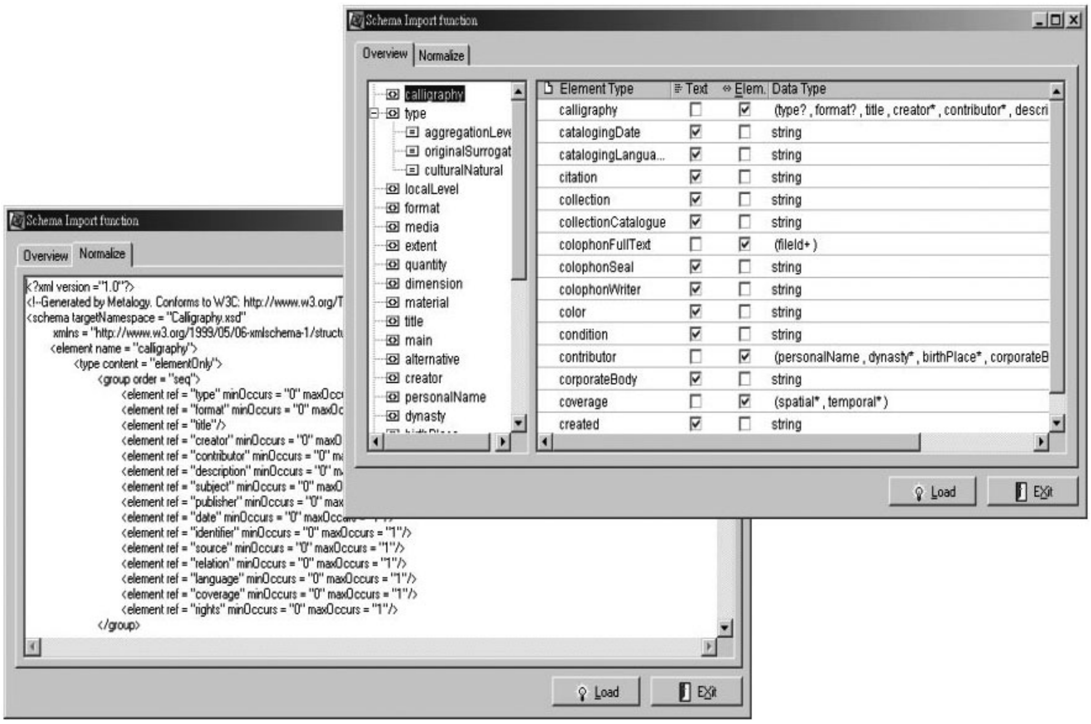
(3) *Verifying mechanism.* Usually, mapping tables can be generated automatically while importing XML schema. However, the system operator needs to manually examine the data format, extra function, input length, item of authority control, and index of every element. Occasionally, it must be edited before importing XML schema. Figure 6 shows an example of verifying the imported XML schema.

Based on the results of comparing the original definitions of XML schema and the mapping table, the system produces the internal meta-structure. This is the source for data type editing.

6.2 Catalogue module

The catalogue module provides the function of cataloguing for different metadata and authority control. Based on the activity diagram depicted in Figure 7, the flow procedure of activity within the overall catalogue function is shown – the Figure also represents the interior process behavior and the flow of control among objects. This

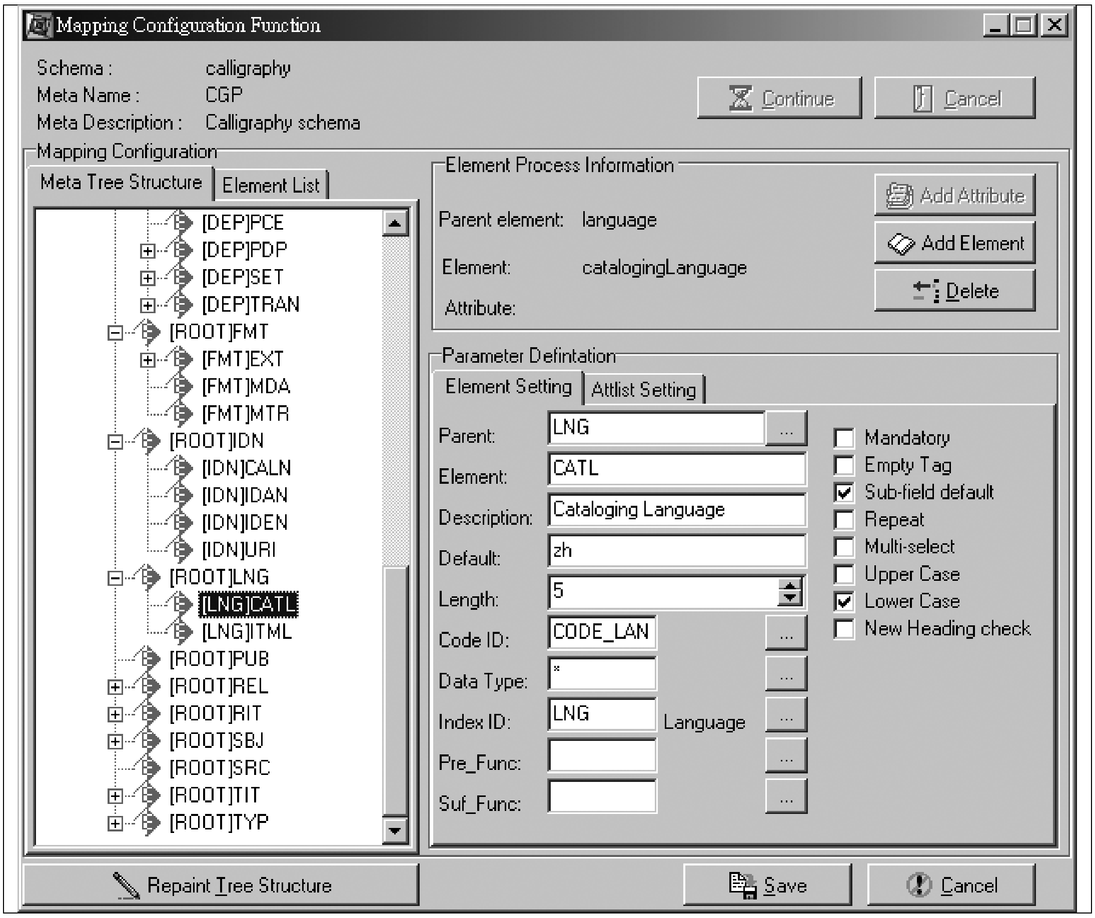
Figure 5 Metadata schema used for importing XML schema



```
k?xml version ="1.0"?>
<!--Generated by Metalogy. Conforms to W3C: http://www.w3.org/T
<schema targetNamespace = "Calligraphy.xsd"
      xmlns = "http://www.w3.org/1999/05/06-xmlschema-1/structu
   <element name = "calligraphy">
      <type content = "elementOnly">
         <group order = "seq">
            <element ref = "type" minOccurs = "0" maxOcc
            <element ref = "format" minOccurs = "0" maxOc
            <element ref = "title"/>
            <element ref = "creator" minOccurs = "0" maxO
            <element ref = "contributor" minOccurs = "0" ma
            <element ref = "description" minOccurs = "0" m
            <element ref = "subject" minOccurs = "0" maxO
            <element ref = "publisher" minOccurs = "0" max
            <element ref = "date" minOccurs = "0" maxOcc
            <element ref = "identifier" minOccurs = "0" maxOccurs = "1"/>
            <element ref = "source" minOccurs = "0" maxOccurs = "1"/>
            <element ref = "relation" minOccurs = "0" maxOccurs = "1"/>
            <element ref = "language" minOccurs = "0" maxOccurs = "1"/>
            <element ref = "coverage" minOccurs = "0" maxOccurs = "1"/>
            <element ref = "rights" minOccurs = "0" maxOccurs = "1"/>
         </group>
```

Element Type	Text	Elem.	Data Type
calligraphy	□	☑	(type? , format? , title , creator* , contributor* , descri
catalogingDate	☑	□	string
catalogingLangua...	☑	□	string
citation	☑	□	string
collection	☑	□	string
collectionCatalogue	☑	□	string
colophonFullText	□	☑	(fileld+)
colophonSeal	☑	□	string
colophonWriter	☑	□	string
color	☑	□	string
condition	☑	□	string
contributor	□	☑	(personalName , dynasty* , birthPlace* , corporateB
corporateBody	☑	□	string
coverage	□	☑	(spatial* , temporal*)
created	☑	□	string

Figure 6 Mapping table configuration – to supply the extended definitions of each element



Mapping Configuration Function

Schema : calligraphy
Meta Name : CGP
Meta Description : Calligraphy schema

Mapping Configuration

Element Process Information
Parent element: language
Element: catalogingLanguage
Attribute:

Parameter Defintation

Element Setting | Attlist Setting

Parent: LNG
Element: CATL
Description: Cataloging Language
Default: zh
Length: 5
Code ID: CODE_LAN
Data Type: *
Index ID: LNG Language
Pre_Func:
Suf_Func:

□ Mandatory
□ Empty Tag
☑ Sub-field default
□ Repeat
□ Multi-select
□ Upper Case
☑ Lower Case
□ New Heading check

module contains two sub-functions: data input and authority control.

6.3 Data input function

This function provides capability of metadata editing. The editable metadata permitted by the framework developed in this study is decided by the XML schema imported previously. That is how many XML schemas are imported and then how many metadata can be administrated.

As shown in Figure 8, the system generates corresponding META tables (schema parameter tables) while loading an XML schema. Operators can catalogue the metadata by selecting the meta-group to which where the metadata belongs. The system provides the functions of duplicating, deleting, inserting sub-field, code, and connecting to multimedia files to each field of metadata. Operators also can use the enquiry function during cataloguing and, furthermore, copying metadata previously edited. If operators consider the subject heading check is not necessary, or need to simplify the processing of catalogues, operators can use offline any text-editor to edit records, and then import in batch these records by full-text editor.

The major function of the full-text editor is providing an independent editing capability. This system can mark up full-text into metadata, as report, document, article, monograph, and so on. The editor is divided in to two parts, the left side is the selection element list of schema, and the right side is the edit zone of full-text content that can be imported by the operator or pasted from the clipboard. Operators can directly modify the content in the edit zone. The method of mark up is to highlight a section of content, and then drag-drop the highlighted section to the tag of theelement which is in the left side. XML restricts a set of a few characters to special roles in certain contexts, and requires that if these characters are used in any other way that they need be written as an escape sequence. The reserved characters include: &, <, >,", '. When the full-text mark up process is finished, the system will convert the reserved characters into general entities (St Laurent, 1999). For example, if the

Figure 7 The activity diagram of the catalogue module

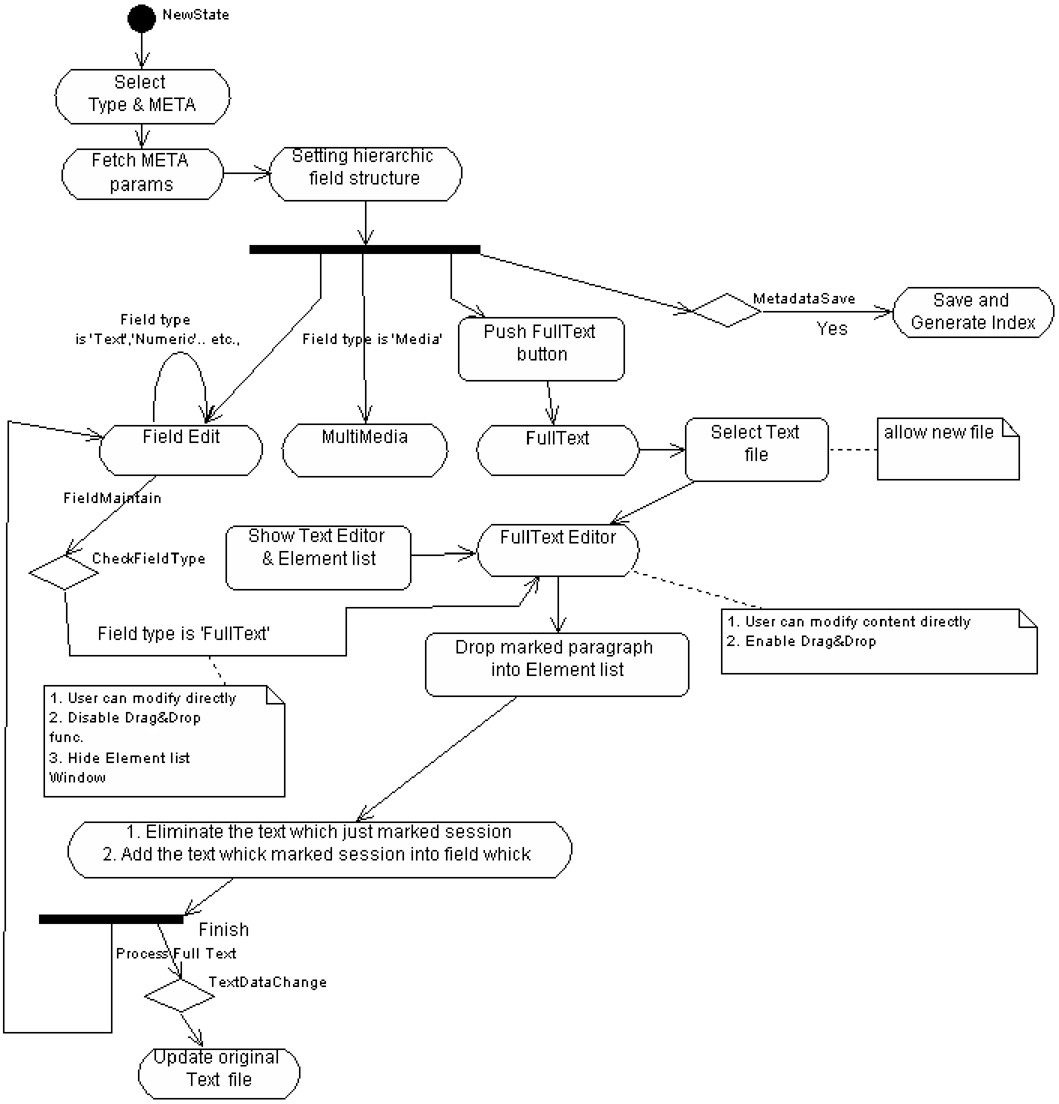
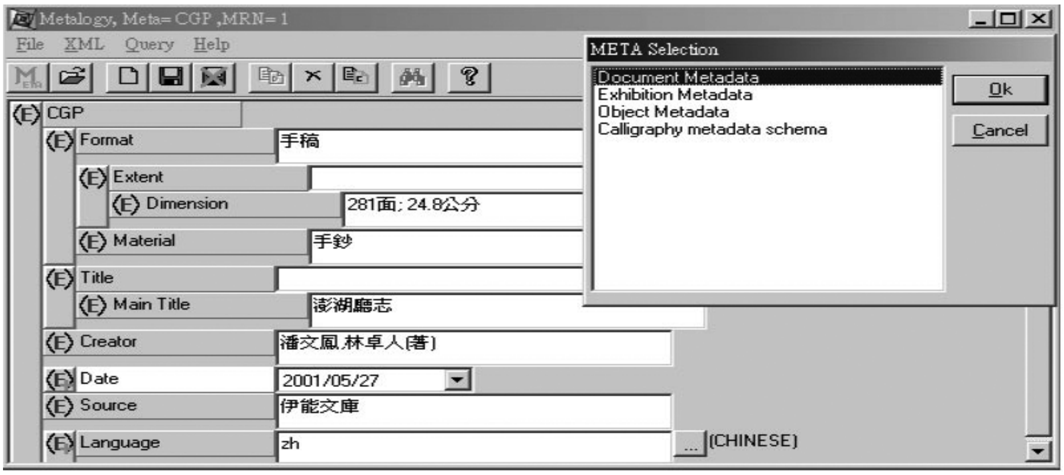

NewState
Select Type & META
Fetch META params
Setting hierarchic field structure
MetadataSave
Yes
Save and Generate Index
Field type is 'Text','Numeric'.. etc.,
Field type is 'Media'
Push FullText button
Field Edit
MultiMedia
FullText
Select Text file
allow new file
FieldMaintain
CheckFieldType
Show Text Editor & Element list
FullText Editor
1. User can modify content directly
2. Enable Drag&Drop
Field type is 'FullText'
Drop marked paragraph into Element list
1. User can modify directly
2. Disable Drag&Drop func.
3. Hide Element list Window
1. Eliminate the text which just marked session
2. Add the text whick marked session into field whick
Finish
Process Full Text
TextDataChange
Update original Text file

Figure 8 Operators can choose the catalogue format of data according to the amount of XML schema



CGP
Format: 手稿
Extent:
Dimension: 281面; 24.8公分
Material: 手鈔
Title:
Main Title: 澎湖廳志
Creator: 潘文鳳,林卓人(著)
Date: 2001/05/27
Source: 伊能文庫
Language: zh (CHINESE)

contents involve the "&" character, the system will convert it into "&amp".

The multimedia catalogue offers options of batch-import or single record import with brief description. Multimedia linking can be done before catalogue or be edited afterwards. The major purpose of batch-import is saving operation time. Multimedia have various kinds of format, such as image, voice, picture, animation, video, etc. These media may have huge sizes that do not suit database to storage directly. Operators may decide whether the storage is a database or other repository depending on the size of the media. If the storage is another repository, the system will build a virtual pointer to link these media objects belonging to the metadata.

6.4 Authority control function

In our research, while the system is loading an XML schema, if the operator decides the XML schema is imported for authority control, the system will generate the mapping META table of authority control. When operators catalogue the metadata, the system will check the parameters of authority control (such as year, name, or location fields) to automatically proceed with the field linking. Operators can, if necessary, add the same entity but different subject heading of the authority control records – the system will then generate the relative authority control records. The system also executes the functions below, according to the META table imported to the database:

- Importing the authority control records built by other systems.
- Using XML as the syntax of authority control import and export.
- Adding and modifying metadata records will also change the linked authority control records.
- Display the same or similar authority control records for operators to choose.
- Authority control records can be used for the expanded index for enquiry.

6.5 Enquiry module

With this module, users can search the catalogued data in the status of maintenance or export. The enquiry module can display the results with the formats pre-set by parameters, such as main display, brief display, and detailed display. Users can make a general or authority-data enquiry on the metadata of one single meta-group or the metadata of all meta-groups. For the related characteristics of metadata fields, the system can integrate more index items into one access point (or called "search item," "search field name"), and combine similar field attributes, to obtain better search results. In this way, it offers more flexibility to user enquiries. For example, users can set the index items for title, sub-title, and other titles, but choose the "title" as the real access point to make enquiry on those three indexes at one step.

7. Discussion and conclusion

This paper has not set out to design a complete system for a particular region, but only constructed a metadata system using the XML framework compatible with various metadata schemas. Employing the XML schema to define the system schema structure allows a system to exist with multiple XML schemas and thus meet the demand of concurrent processing of multiple different types of metadata in import and export. Users can easily retrieve various types of metadata through the integration of multi-access points and the consistent enquiry function of the system. Importantly, the metadata can be accessed by other systems. Based on this framework, libraries, museums, culture centers, educational institutions, or enterprises can depend on their demand to easily integrate user interfaces or other applications in order to accomplish a complete system.

In summary, the proposed system presents plentiful abilities including the following:

- the compatible and integrated ability in imported heterogeneous XML schema;
- the system allows various types of XML schema to co-exist at the same time;
- users can retrieve various formats of data simultaneously;
- data import/export complies with XML schema formats;
- the processing contents include the data of fields, multimedia structure and full text;
- system management includes the function of access control, transaction log, etc.

In reality, metadata is a useful and popular media for managing and storing the knowledge in heterogeneous formats. With the compatibility of various metadata, the proposed system is a useful tool in creating a knowledge system by integrating the related knowledge. This will efficiently enhance the performance of data processing and sharing.

References

Booch, G., Rumbaugh, J. and Jacobson, I. (1999), *The Unified Modeling Language User Guide*, Addison-Wesley Inc., Reading, MA, pp. 265-77.

Cagle, K. (2000), *XML Developer's Handbook*, SYBEX Inc., San Francisco, CA, p. 272.

Chilvers, A. and Feather, J. (1998), "The management of digital data: a metadata approach", *The Electronic Library*, Vol. 16 No. 6, pp. 365-71.

Dempsey, L. and Heery, R. (1997), "Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats", available at: www.ukoln.ac.uk/Metadata/desire/overview/rev_01.htm

El-Sherbini, M. (2001), "Metadata and the future of cataloging", *The Electronic Library*, Vol. 50 No. 1, pp. 16-27.

Gilliland-Swetland, A.J. (1998), "Setting the stage: defining metadata", in Baca, M. (Ed.), *Introduction to Metadata: Pathways to Digital Information*, Getty Research Institute, Los Angeles, CA, pp. 1-8.

Ioannides, D. (2000), "XML schema languages: beyond DTD", *Library Hi Tech*, Vol. 18, pp. 9-14.

Kruse, R., Tondo, C.L. and Leung, B. (1997), *Data Structures & Program Design in C*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, pp. 154-61.

St Laurent, S. (1999), *XML: The Primer*, M&T Books, Foster City, CA, pp. 103-4.

Vellucci, S.L. (2000), "Metadata and authority control", *Library Resources & Technical Services*, Vol. 44, pp. 33-43.

Weibel, S., Godby, J. and Miller E. (1995), "OCLC/NCSA metadata workshop report", available at: www.oclc.org/oclc/research/conferences/Metadata/dublin_core_report.html

W3C (1998), "Extensible markup language (XML) 1.0: W3C recommendation", available at: www.w3.org/TR/1998/REC-xml-19980210

W3C (2001), "XML schema, W3C recommendation", available at: www.w3.org/XML/Schema

Further reading

Loagze, C. and Lynch, C.A. (1996), "The Warwick framework: a container architecture for aggregation sets of metadata", available at: www.ifla.org/documents/libraries/cataloging/metadata/tr96/593.pdf