Heterogeneous Metadata Management and Manipulation using an XML-based Framework

Shihn-Yuarn Chen
Dept. of Computer and Information Science, National Chiao-Tung University, 1001 Ta Hsueh Rd., Hsinchu City, Taiwan 300
sychen@cis.nctu.edu.tw

Hao-Ren Ke
University Library, National Chiao-Tung University, 1001 Ta Hsueh Rd., Hsinchu City, Taiwan 300
claven@lib.nctu.edu.tw

Wei-Pang Yang
Dept. of Computer and Information Science, National Chiao-Tung University, 1001 Ta Hsueh Road, Hsinchu City, Taiwan 300

Dept. of Information Management, National Dong-Hwa University, 1, Sec. 2, Da Hsueh Rd., Shou-Feng, Hualien,Taiwan
wpyang@cis.nctu.edu.tw

Abstract*-Heterogeneous metadata is used in different application systems of digital libraries, online shops, or enterprises. The schema of metadata may change frequently; therefore, issues about metadata management and manipulation may arise. Traditionally, administrators have to update the RDBMS database schema and the application programs while the schema of metadata changes. In this paper, XML-based Metadata Management and Manipulation Framework (XMF for short) is proposed to facilitate these tasks. In XMF, metadata of objects, the information about their relationships, and the keyword information are stored in XML-format. By using XMF, administrators can perform retrieve and maintain metadata, keyword and relationship search, and handle the changes of the schema of metadata easily. To certify the feasibility of XMF, an application system based on XMF is shown.*

Keywords: Metadata Management, XML, Keyword Search, Relationship Information

1. Introduction

There are many application systems used in a digital library, an online shop, or an enterprise. Various kinds of metadata are used to describe digital objects, books, or a customer, and are manipulated in these application systems.

Traditionally, a RDBMS is usually employed to store metadata and administrators have to take care of the changes of metadata schema. If schema change of metadata occurs (e.g. add the reprint information of a book, enlarge the data field length of the address of a company), administrators have to update the RDBMS table schema to meet the change. If there is a new type of metadata (e.g. an online shop starts to sell DVDs), administrators have to create the schema of this new type of metadata. These tasks are important for an application system, but boring for administrators.

To reduce the load about these boring tasks, we survey some research and propose an XML-based framework, the XML-based Metadata Management and Manipulation Framework (XMF for short). In XMF, metadata of an object (a book, a transaction, a customer, etc.) is stored as an XML document. Besides, information about relations between objects is also kept in XML documents.

XMF provides rich methods for retrieving and updating metadata and its structure. Taking into account of practicability, XMF also supplies the keyword search mechanism. User interface representation becomes easier in XMF, because style sheets can format an XML document easily.

In Section 2, some related research would be introduced briefly. The main architecture and design of XMF are shown in Section 3 and 4. In Section 5, an application system based on XMF is introduced to show that XMF is effective. Finally, conclusions and future work are in Section 6.

2. Related Work

Maintaining various types of metadata is always an issue for digital content repositories, such as a digital museum, an online book shop, even an enterprise. Ruey-Shun Chen et al.'s propose a metadata management system to solve this issue [1, 2]. In their system, metadata is stored in XML format.

Besides, the functions of schema construction, cataloguing, and import/export are supplied.

Their approach can well handle the management of the metadata of one object. Even so, one shortcoming of their approach is that relationships between objects are not considered. If their approach is just used in a simple application system, the information on relationships is not necessary; however, the information on relationships between objects becomes very important in a larger scale application system, such as an online book shop.

Morpho [3] is another XML-based framework that focuses on ecological data. In Morpho, Ecological Metadata Language (EML for short) is exploited to model ecological data.

Besides, NeoCore Inc. proposes an XML-based information management system [4]. This system is designed to manage and manipulate the metadata about the research of bioinformatics that is a field developing fast and impossible for administrators to predefine all metadata fields in a RDBMS. As an XML-based framework, it is easy to extend the metadata fields while necessary.

Although these approaches can handle metadata well and provide a new way to store metadata - store them in XML format, they are still short of the relationship management and the support of keyword search mechanism. XMF extends these approaches to solve these issues.

3. XMF Architecture

Research mentioned in previous section presents some solutions for metadata management and manipulation. However, those approaches are still not enough. XMF is proposed in this paper to solve the issues about the relationship information and keyword information in metadata management and manipulation, and Figure 1 shows the architecture of XMF.

XMF contains three main layers: storage layer, constraint layer, and metadata manipulation and presentation layer. In the storage layer, **XML Document Repository** (XDR for short) takes responsibility of XML document storage and access. In the early development stage, XMF uses RDBMS to handle access tasks for simplifying the issues of storage, access, and efficiency.

The constraint layer handles the tasks of metadata retrieval according to the given constraints, such as a keyword, or a unique id of an object. Currently, this layer contains three modules that can process these constraints: **Relation Constraint Module** (RCM for short), **Filter Constraint Module** (FCM for short) and **Metadata Object Module** (MOM for short). More modules may be added in the future, when necessary.

To find out related information about an object is a frequent action in an application system. While this action is taken, the identifier of the object is used as a relationship constraint. In XMF, relationships between objects are maintained, and this relationship constraint can be well handled by RCM. RCM can retrieve the related objects according to the relationship constraint easily, and these would be returned as a metadata object collection. For example, if a user wants to review his/her historical shopping logs, RCM would retrieve the historical shopping logs from XDR according to the relationship constraint, the identity of this user. And these logs would be returned as a collection of metadata objects.

A filter constraint is usually a keyword given by a user, and FCM can handle this constraint and facilitate the keyword search mechanism. In XMF, any field in the metadata of all object types can be configured as a keyword field. If a field is configured as a keyword field, the content of this field would be collected automatically for keyword search during the cataloging process. The detail about automatic keyword collection and the management of keyword fields would be described in Subsection 4.2. While receiving a filter constraint, a keyword from the user, FCM retrieves the metadata of objects which match the keyword.

MOM, the primary component of XMF, handles the metadata of a digital object and takes responsibility of XML-format metadata retrieval, cataloguing, and update. Furthermore, it cooperates with RCM and FCM to handle the keyword information and the relationship information that are assigned by one user or retrieved from XDR. It also performs simple metadata presentation and supplies the XML-format metadata to the **Presentation Transformation** (PT for short) component for other format presentations.

The PT component can transform the presentation from XML format into others, such as HTML. The transformation can be achieved by CSS, XSLT, and programs.

4. XMF-based Application System Design

Metadata management and manipulation is achieved by some operations, such as cataloging, update, and retrieval. Besides, XMF also provides the mechanisms of keyword field management and automatic keyword collection to meet the requirement of keyword search in general application systems. In the following subsections, these operations of XMF are described.

4.1. Metadata Catalogue

Figure 2 is the process of metadata cataloguing in XMF. A cataloguer can input the metadata of an object through the user interface. The metadata would be stored as a metadata object, which also records the relationship information that the cataloguer assigns, and the keyword information in the metadata. The keyword fields of one kind of objects are predefined and are able to be changed dynamically.

Then, MOM deals with the metadata object, stores the metadata and relationship information into the XDR, and stores the keyword information into the proper keyword information structure in the XDR. Then the cataloguing task is completed.

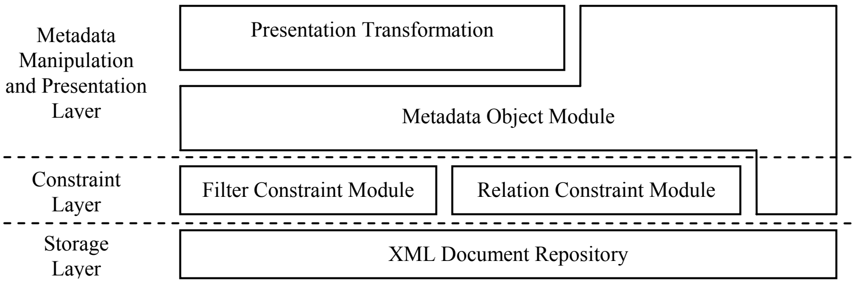


Figure 1. The architecture of XML Metadata Framework (XMF).

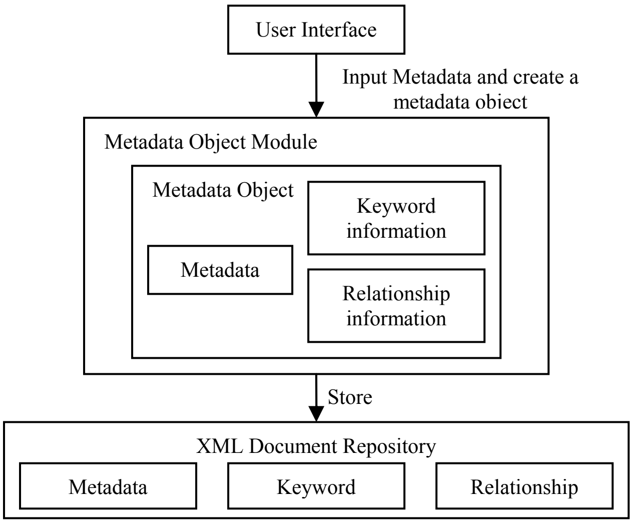


Figure 2. The process of catalogue in XMF.

4.2. Automatic Keyword Collection

To achieve automatic keyword collection, XMF has to perform three processes.

1. Maintain the **Keyword Information Structure Collection** (KSC for short);
2. collect the keyword information while cataloging;
3. store the keyword information in the proper keyword information structure.

The last two processes are easy and mentioned in Subsection 4.1, so they would not be discussed in this subsection.

However, the first process is the most important to achieve automatic keyword collection. Without this, XMF cannot recognize the keyword type while receiving a filter constraint, and cannot make out which object a keyword belongs to, etc. The KSC is formulated as follows:

$$KSC = \{ KS_{(1,T1)}, KS_{(2,T1)}, \ldots, KS_{(i,T1)}, KS_{(1,T2)}, KS_{(2,T2)}, \ldots, KS_{(j,Tk)}, \} \qquad (1)$$

$$KS_{(j,Tk)} = \{K_{(Fj,\,O1)}, K_{(Fj,\,O2)}, K_{(Fj,\,O3)}, \ldots, K_{(Fj,\,Om)}\}, \qquad (2)$$

where $\{O_1, O_2, O_3, \ldots, O_m\}$ are type of T_k

The KSC is composed of a number of Keyword Information Structure (KS for short), as shown in equation (1). Each KS is used to store the keyword information of a metadata field of a type of objects, as shown in equation (2); in other words, each KS(j,Tk) is composed of keywords which belong to the metadata field, Fj, of the same type of objects, O1 to Ok .

If a user gives a keyword of a metadata field of a type of objects as a filter constraint to perform a search action, XMF searches the keyword in the proper KS(j,Tk) and returns the metadata of objects which matches the keyword.

Besides, because the metadata field of a type of objects is not required while searching, a keyword search for all fields of all types of objects is permissible. XMF would search the whole KSC while receiving one search of this kind.

After constructing the KSC and the proper KS for a metadata field of a type of objects, XMF is able to collect the keyword information automatically while cataloging or updating, and is able to provide keyword search mechanism.

4.3. Metadata Retrieval

The retrieval operation is used frequently in an application system. XMF encompasses three types of metadata retrieval. The processes are shown in Figure 3.

First of all, for the most part, XMF retrieves the metadata of an object via the MOM by the unique-id of the object. Retrieving the metadata of a book is an example, and the processes are step 1 and 2 in Figure 3.

Second, a user may want to search for objects in the repository, and he/she can use a keyword of a metadata field as input of FCM. This component would retrieve the metadata of objects which match the keyword of the metadata field, and return the results as a metadata objects collection. The processes are step 3 to 6 in Figure 3.

In addition, RCM is also useful, and its mechanism is similar to FCM except that the input of RCM is the unique-id of an object. For example, a user may want to browse other works of an artist while browsing one of them in an online shop. He/She can click the hyper-link of the artist and use the unique-id of the artist as input of RCM, and this component would retrieve and return the related works of this artist. The processes are step 7 to 10 in Figure 3.

4.4. Metadata Update

There are two types of metadata update: metadata field value update and metadata field schema update. Metadata field value update means that the data of a field is modified, for example, the edition number of a book changes while a new edition is published. Metadata field schema update means that the schema of metadata of one type of objects changes, for example, the warranty field of a product is replaced by two fields, international warranty and local warranty.

If the value of the metadata of one object needs modification, XMF retrieves the metadata, keyword information and relationship information into one metadata object. Then, the metadata object shows the information in the user interface. After the modification, the metadata object updates the value of metadata in XDR. The keyword information and relationship information are updated in the same manner if necessary.

The task of dealing metadata field schema update is more complex. The keyword information may need reconstruction while the keyword field of metadata is changed. For example, if the warranty field is a keyword field of one product and this field is replaced by two keyword fields, local warranty and international warranty, XMF has to drop the whole KS of warranty, and create two KSes for local warranty and international warranty, as shown in Figure 4.

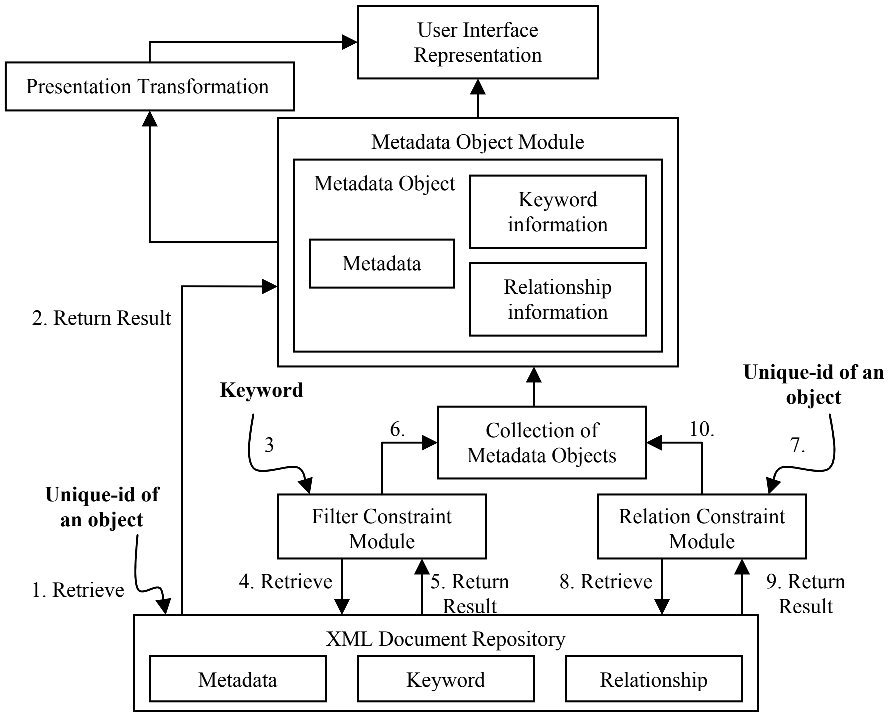


Figure 3. The process of metadata retrieval in XMF.

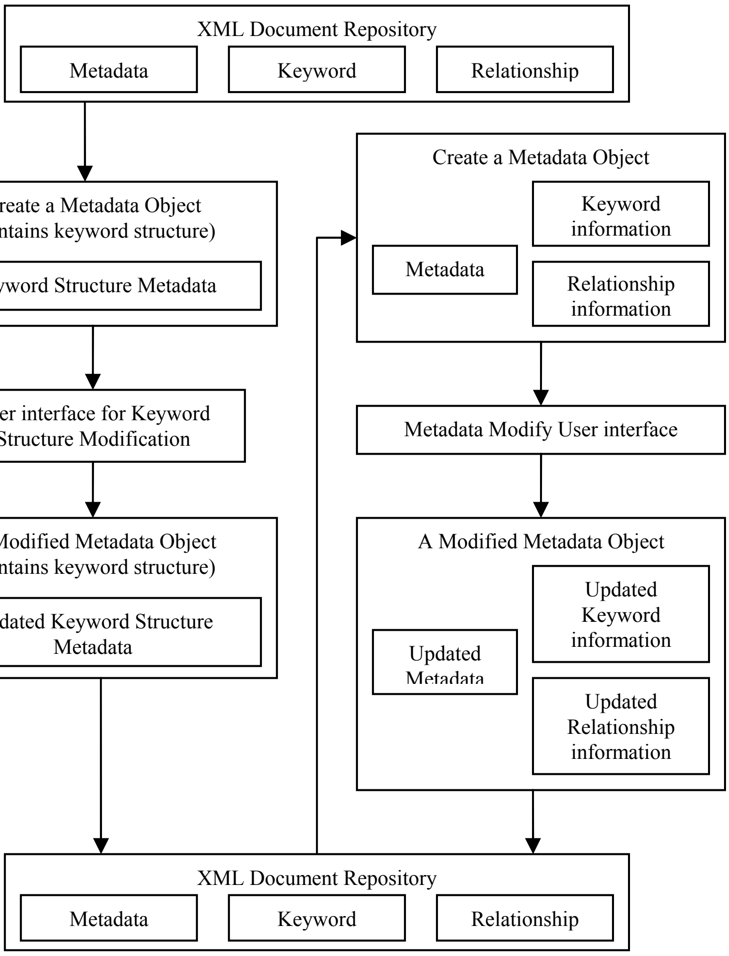
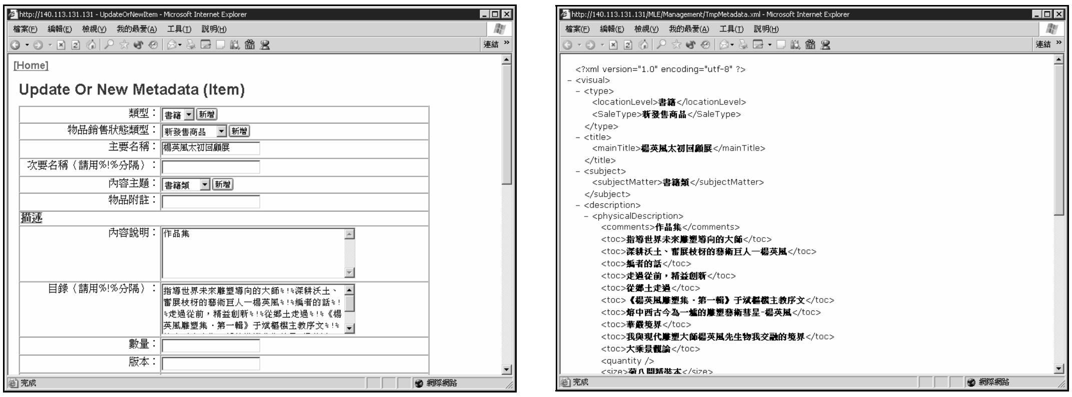

XML Document Repository
Metadata
Keyword
Relationship
Create a Metadata Object (contains keyword structure)
Keyword Structure Metadata
User interface for Keyword Structure Modification
A Modified Metadata Object (contains keyword structure)
Updated Keyword Structure Metadata
Create a Metadata Object
Metadata
Keyword information
Relationship information
Metadata Modify User interface
A Modified Metadata Object
Updated Metadata
Updated Keyword information
Updated Relationship information
XML Document Repository
Metadata
Keyword
Relationship

Figure 4. The process of metadata field schema change in XMF.

5. XMF-based System

NCTU e-go, an online shop, is an application of XMF. In the backend system, administrators can catalogue and update goods information, manage the keyword information structure, and deal with orders. Users can browse goods, search for goods by keyword, make an order, etc. Figure 5 and Figure 6 show the cataloguing operation, the metadata of a good in XML format, and the management of the keyword information structure respectively. Figure 7 shows searching results and detail of a good browsing.


http://140.113.131.131 - UpdateOrNewItem - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
連結
[Home]
完成
網際網路

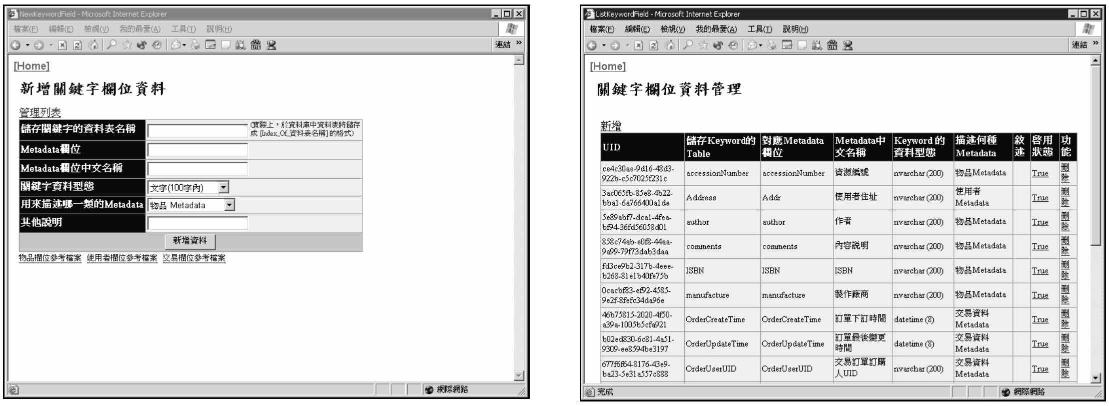
Update Or New Metadata (Item)

類型：	書籍 新增
物品銷售狀態類型：	新發售商品 新增
主要名稱：	楊英風太初回顧展
次要名稱（請用%!%分隔）：	
內容主題：	書籍類 新增
物品附註：	
描述	
內容說明：	作品集
目錄（請用%!%分隔）：	指導世界未來雕塑導向的大師%!%深耕沃土、奮展枝枒的藝術巨人一楊英風%!%編者的話%!%走過從前，精益創新%!%從鄉土走過%!%《楊英風雕塑集・第一輯》于斌樞機主教序文%!%
數量：	
版本：	

http://140.113.131.131/MLE/Management/TmpMetadata.xml - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
連結
完成
網際網路

```
  <?xml version="1.0" encoding="utf-8" ?>
- <visual>
  - <type>
      <locationLevel>書籍</locationLevel>
      <SaleType>新發售商品</SaleType>
    </type>
  - <title>
      <mainTitle>楊英風太初回顧展</mainTitle>
    </title>
  - <subject>
      <subjectMatter>書籍類</subjectMatter>
    </subject>
  - <description>
    - <physicalDescription>
        <comments>作品集</comments>
        <toc>指導世界未來雕塑導向的大師</toc>
        <toc>深耕沃土、奮展枝枒的藝術巨人一楊英風</toc>
        <toc>編者的話</toc>
        <toc>走過從前，精益創新</toc>
        <toc>從鄉土走過</toc>
        <toc>《楊英風雕塑集・第一輯》于斌樞機主教序文</toc>
        <toc>熔中西古今為一爐的雕塑藝術彗星-楊英風</toc>
        <toc>華嚴境界</toc>
        <toc>我與現代雕塑大師楊英風先生物我交融的境界</toc>
        <toc>大乘景觀論</toc>
        <quantity />
        <size>菊八開精裝本</size>
```

Figure 5. Cataloguing operation.


NewKeywordField - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
連結
[Home]
網際網路

新增關鍵字欄位資料

管理列表

儲存關鍵字的資料表名稱		(實際上，於資料庫中資料表將儲存成 [Index_Of_資料表名稱] 的格式)
Metadata欄位		
Metadata欄位中文名稱		
關鍵字資料型態	文字(100字內)	
用來描述哪一類的Metadata	物品 Metadata	
其他說明		

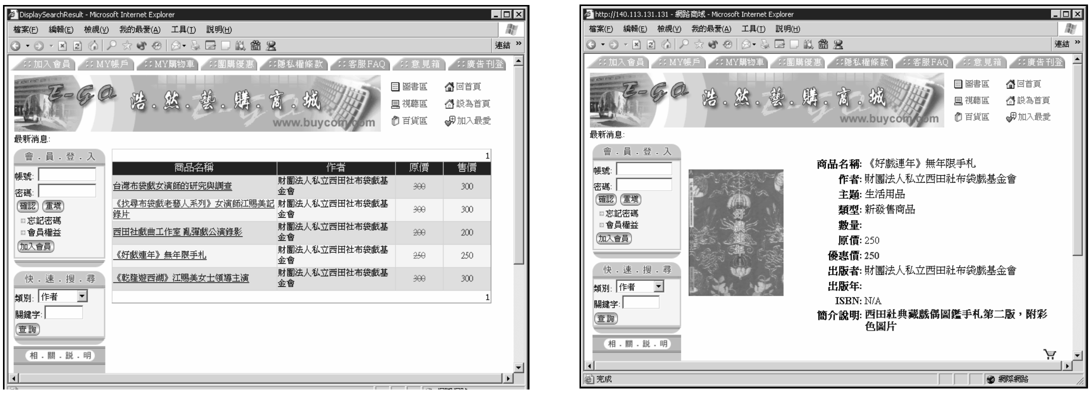
新增資料
物品欄位參考檔案 使用者欄位參考檔案 交易欄位參考檔案
ListKeywordField - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
連結
[Home]
完成
網際網路

關鍵字欄位資料管理

新增

UID	儲存Keyword的Table	對應Metadata欄位	Metadata中文名稱	Keyword 的資料型態	描述何種Metadata	敘述	啓用狀態	功能
ce4c30ae-9d16-48d3-922b-c5c7025f231c	accessionNumber	accessionNumber	資源編號	nvarchar (200)	物品Metadata		True	刪除
3ac065fb-85e8-4b22-bba1-6a766400a1de	Address	Addr	使用者住址	nvarchar (200)	使用者Metadata		True	刪除
5e89abf7-dca1-4fea-bf94-36fd56058d01	author	author	作者	nvarchar (200)	物品Metadata		True	刪除
858c74ab-e0f8-44aa-9a99-79f73dab3daa	comments	comments	內容說明	nvarchar (200)	物品Metadata		True	刪除
fd3ce9b2-317b-4eee-b268-81e1b40fe75b	ISBN	ISBN	ISBN	nvarchar (200)	物品Metadata		True	刪除
0cacbf83-ef92-4585-9e2f-8fefc34da96e	manufacture	manufacture	製作廠商	nvarchar (200)	物品Metadata		True	刪除
46b75815-2020-4f50-a39a-1005b5cfa921	OrderCreateTime	OrderCreateTime	訂單下訂時間	datetime (8)	交易資料Metadata		True	刪除
b02ed830-6c81-4a51-9309-ee8594be3197	OrderUpdateTime	OrderUpdateTime	訂單最後變更時間	datetime (8)	交易資料Metadata		True	刪除
677f6f64-8176-43e9-ba23-5e31a557c888	OrderUserUID	OrderUserUID	交易訂單訂購人UID	nvarchar (200)	交易資料Metadata		True	刪除

Figure 6. Management of keyword information structure.



商品名稱	作者	原價	售價
台灣布袋戲女演師的研究與調查	財團法人私立西田社布袋戲基金會	~~300~~	300
《找尋布袋戲老藝人系列》女演師江賜美記錄片	財團法人私立西田社布袋戲基金會	~~300~~	300
西田社戲曲工作室 亂彈戲公演錄影	財團法人私立西田社布袋戲基金會	~~200~~	200
《好戲連年》無年限手札	財團法人私立西田社布袋戲基金會	~~250~~	250
《乾隆遊西湖》江賜美女士領導主演	財團法人私立西田社布袋戲基金會	~~300~~	300

http://140.113.131.131 - 網路商城 - Microsoft Internet Explorer
完成
網際網路

商品名稱: 《好戲連年》無年限手札
作者: 財團法人私立西田社布袋戲基金會
主題: 生活用品
類型: 新發售商品
數量:
原價: 250
優惠價: 250
出版者: 財團法人私立西田社布袋戲基金會
出版年:
ISBN: N/A
簡介說明: 西田社典藏戲偶圖鑑手札第二版，附彩色圖片

Figure 7. Keyword searching and goods browsing.

6. Conclusions and Future Work

XMF is a general XML-based framework that supplies a new way for metadata management and manipulation. Relationship management and keyword search mechanism are also provided to meet the demands of practical application systems. Besides, application systems based on XMF can take metadata management and manipulation easy.

In the future, XMF will be used to build up more application systems which store various types of metadata, such as an online comic showroom, and a distant learning system. Besides, the functions of XMF will be extended, and the support of XML Schema and customized representations will be added in. As a result, XMF can import and export XML Schema of an object, and generate proper XML Schema for the metadata of an object. Customized representations enable XMF to be used for building up some personalized information environment of application systems, such as a personalized bookshelf of a digital library, a personalized schedule of an enterprise system, etc.

Acknowledgement

The authors would like to thank the National Science Council, Taiwan, R.O.C., for financially supporting this research under Contract No. NSC 92-2422-H-009-0012.

References

[1] Ruey-Shun Chen, Shien-Chiang Yu, "Developing an XML framework for metadata system," Institute of Information Management, National Chiao-Tung University, 1001 Ta Hsuch Road, Hsin-Chu city, Taiwan, ROC.

[2] Shien-Chiang Yu, Kun-Yung Lu, Ruey-Shun Chen, "Metadata management system: design and implementation," *Electronic Library*, vol. 21, Issue 2, pp.154-164, 2003.

[3] Daniel Higgins, Chad Berkley, Matthew B. Jones, "Managing Heterogeneous Ecological Data Using Morpho," *Proceedings of the 14th International Conference on Scientific and Statistical Database Management*, pp.69-76, July 2002.

[4] Direen, H., Brandin, C., Jones, M., Hedgepeth, C., Shin, D., "Knowledge Management through a Fully Extensible, Schema Independent, XML Database," *Proceeding of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 4, pp.3676-3679, Oct. 2001.