A DIGITAL LIBRARY USING DESCRIPTIVE METADATA

Su Su Htwe
Computer University, Mandalay
zinzinsoewin@gmail.com

ABSTRACT

Digital libraries are emerging technologies for document management starting, from source through storage of documents and use of the management standard such as metadata for efficient search and retrieval to delivery. Digital libraries have to use all modern network and servers technologies in order to supply services of a high quality. Digital libraries are being created today for diverse communities and in different fields e.g. education, science, culture, development, healthy, government and so on. This paper intends to implement the discovery and identification of the digital document files at which the desired information about education is included. There are two main parts in this system: it is to build a digital library supporting Computer Science and Technology field areas and distribute the digital document files in a LAN.

1. INTRODUCTION

Digital Libraries are being created today for diverse communities and in different fields e.g. education, science, culture, development, health, governance and so on. With the availability of several free digital Library software packages at the recent time, the creation and sharing of information through the digital library collections has become an attractive and feasible proposition for library and information professionals around the world. Library automation has helped to provide easy access to collections through the use of computerized library catalogue.

Digital libraries differ significantly from the traditional libraries because they allow users to gain an access to and work with the electronic versions of full text documents and their associated images. Many digital libraries also provide an access to other multi-media content like audio and video [2]. Additionally, distributed system is a tool for digital libraries. Information accessible also provides digital library based on distributed system. Accordingly, the searching mechanism is controlled and accessed by metadata creation.

Thus, metadata is one of the most important aspects of data storage. It is data about data stored in the database and its users. Descriptive metadata contains information that gives users an easy-to-understand perspective of the information stored in the database [1]. This system is intended to develop a digital library system using descriptive metadata. Descriptive metadata describes a resource for purposes such as discover and identification. Related works are described in section 2. System Architecture and Metadata for Digital Library are explained in section 3. Functional Components of Digital Library are expressed in section 4. Conclusion is described in section 5.

2. RELATED WORK

The University of California at Berkeley has been in the forefront of digital library innovation for many years. Projects begun at Berkeley included the creation of specifications for encoding electronic finding aids that are used to access special collections and archives. This project has since evolved into the Metadata Encoding and Transmission Standard (METS) [4].

Sirsi, vendor of the "Unicorn" integrated library system, has a digital library offering called "Hyperion™ Digital Media Archive." The Hyperion Digital Media Achieve is a tool for building, storing, and maintaing collections of digitally captured material [5]. VTLS, developer of the "Virtua" integrated library system, has a slightly different approach, offering "Visual MIS"(Multimedia and Imaging Solutions), professional services for digitizing and making

available a variety of materials such as documents, maps, glass plates, slides, photographs, catalog cards, manuscripts and microfilm [6].

Luna Imaging's "Insight" high resolution digital imaging software is used to digitize image collections, license digital images from existing image collections, share content with other institutions. The Andrew W. Mellon Foundation is adopting Luna's Insigh software as the platform to distribute ArtSTOR, a new independent, not-for profit organization that will make a large body of digital resources for the study of art, architecture, and other fields in the humanities available for subscription [7].

MuseGlobal, Inc. specializes in information connectivity solutions. Its flagship product, MuseSearch™, is a broadcast search technology that allows unlimited numbers and types of information sources to be searched simultaneously with a single user query. It also incorporates linking features to take users from record results to category indices, content, and references [8].

3. SYSTEM ARCHITECTURE AND METADATA FOR DIGITAL LIBRARY

3.1. Digital library

This system organizes and focuses the collections of digital objects with the methods of access and retrieval. This system retain the several qualities of traditional libraries such as a defined community of users, focused collections, long-term available, the possibility of selecting, organizing, preserving and sharing resources.

The digital libraries are sometimes perceived as institutions, though this is not as dominant as the previous definition. The following definition given by the Digital Library Federation (DFL) brings out the essence of this perception. "Digital Libraries are organization that provide the resources, including the specialized staff to select, structure, offer intellectual access to interpret, distribute, preserve the integrity of and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities".

3.2. Metadata

The metadata refers to information about information or, equivalently, data about data. The term has come to mean structured information that feeds into automated processes, and this is currently the most useful way to think about metadata [3].

An important reason for creating descriptive metadata is to facilitate discovery of relevant information. In addition to resource discovery, metadata can help organize electronic resources, facilitate interoperability and legacy resource integration, and provide digital identification, and support archiving and preservation [1].

In this system, an important reason for creating descriptive metadata is to facilitate discovery of relevant information. Metadata creation is for searching first character of authors or titles and the quantities of that character. In addition to resource discovery, metadata can help organize electronic resources, facilitate interoperability and legacy resource integration, and provide digital identification, and support archiving and preservation. When administrator creates metadata, the system automatically updates.

Metadata creation in this system is as follows: Metadata for title is shown in Table 1 and for author in Table 2. In Table 1, the first character for each title and number of books for that title are stored into two columns. In table 2, the first character for each author and number of books for that author are stored into two columns, respectively.

Table 1 . Metadata for Title

Sr.no	Staring with Character	Number of Books
1	A	2
2	B	4
3	C	6
4	D	5
5	E	6
6	F	3
7	G	6
...	......	

Table 2 . Metadata for Author

Sr.no	Starting with Character	Number of Books
1	A	2
2	B	3
3	C	6
4	D	4
5	E	1
6	F	9
7	G	7
...	...	...

If the user search by title, the system sort title name start with character and then also search by author name. Metadata schemes (also called schema) are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. The definition or meaning of the elements themselves is known as the semantics of the scheme. The values given to metadata elements are the content. Metadata schemes generally specify names of elements and their semantics. Optionally, they may specify content rules for how content must be formulated (for example, capitalization rules), and allowable content values (for example, terms must be used from a specified controlled vocabulary).

3.3. Client/Server Technology

Client/server technology is the computer architecture used in this system. It is divided functions into client (requestor) and server (provider) subsystems, with standard communication methods to facilitate the sharing of information between them.

In this system, there are two main processes such as client site and server site. Client user can search the books through metadata for book title or book author. The administrator can create and modify metadata and books to the system. The books are created in server site related to Computer Science and Technology field area such as Multiagent, the Design and Analysis of Computer Algorithms, Programming Languages, Software Engineering, Data Mining, Web-based Mining,

Networking. In this system, number of books is organized as hundred lengths in server database.
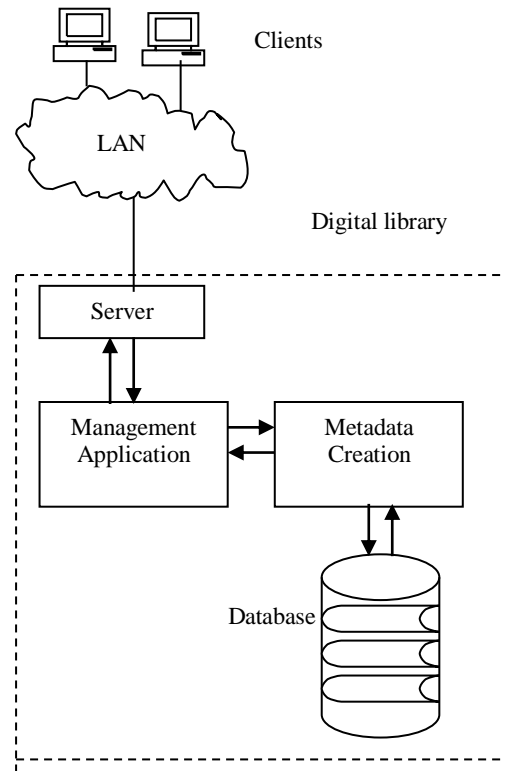

Figure 1. System Architecture for digital library

3.4. Motivation of the system

This digital library is being created for Computer Science and Technology area using descriptive metadata and client/server technology. This system retains the several qualities of traditional libraries such as a defined community of users, focused collections, long-term available, the possibility of selecting, organizing, preserving and sharing resources. Metadata is one of the most important aspects of data storage. Metadata is one of the most important aspects of data storage. Metadata contains information that gives users an easy-to-understand perspective of the information stored in the database. Client/server technology is used in this system as standard communication method to

facilitate the sharing of information between them.

4. FUNCTIONAL COMPONENTS OF DIGITAL LIBRARY

In this section, digital library is explained in 4.1, implemented procedure is described in 4.2 and implementation interfaces are in 4.3.

4.1. Digital library

In this system, four components are shared as:
Selection and Acquisition: the typical process converted in this component includes the selection of documents to be added, the subscription of database and the digitization and conversion of documents to an appropriate digital form.
Organization: the key processes involve in this component is the assignment of the metadata to each document being added to the collection.
Indexing and Storage: this component carries out the indexing and storage of documents and metadata for efficient search and retrival.
Search and Retrival: this is the digital library interface used by the end users to browse, search retrieve and view the contents of the digital library.

4.2. Implemented Procedure

In this system, there are two main processes such as server site and client site. The administrator can create/modify metadata, author name, publishers and books information. Administrator can also view the metadata to the system as shown in Figure 2.
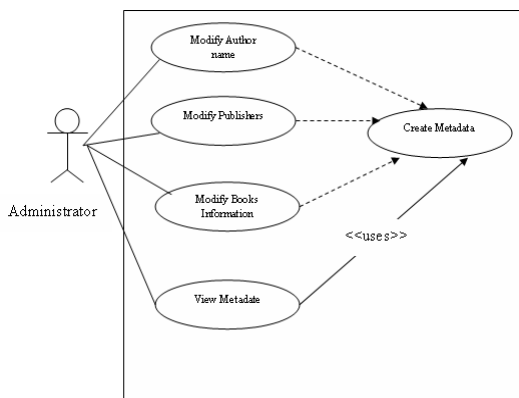

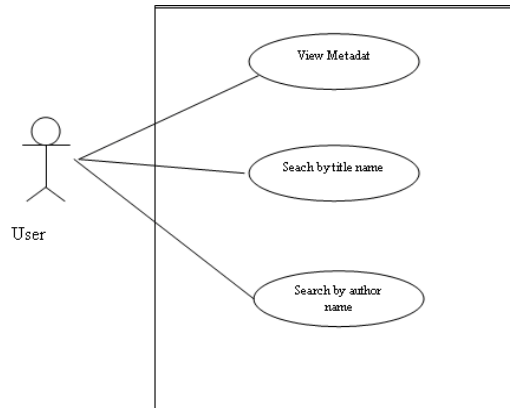**Figure 2.** Use Case Diagram (Sever)

Figure 3. Use Case Diagram (Client)

Client user can search the books by author name and book title as shown in Figure 3. Client can also view the metadata.

4.3. Implementation Interfaces

This system contains client and server potion. In client potion, user can view the metadata as shown in Figure 4. And user can also view all books list into the system's database. That database are stored in sever potion.

Figure 4. Metadata Form

User can retrieve books by title and author name as shown in Figure 5 User can view all information of books in the system.
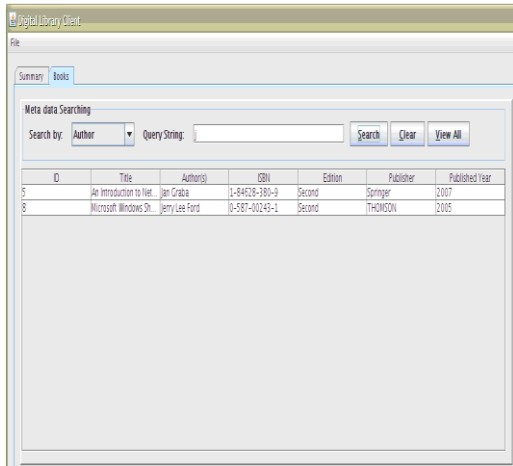

Figure 5. Searching Book by Author Name

Full-text search engines and metadata repositories have so far investigated very different approaches to search, mainly due to their separate and different storage systems for information and data. Metadata is a set of information that describes the characteristics of a data set.

When user search character which starts with letter j for title, user can view all titles starting with j. And then when user chooses required title and retrieve full text information of book as shown in Figure 6.

Figure 6. Full Text Information of Book

Full text information of books is stored in server database. Information of books (metadata) is created in server site.
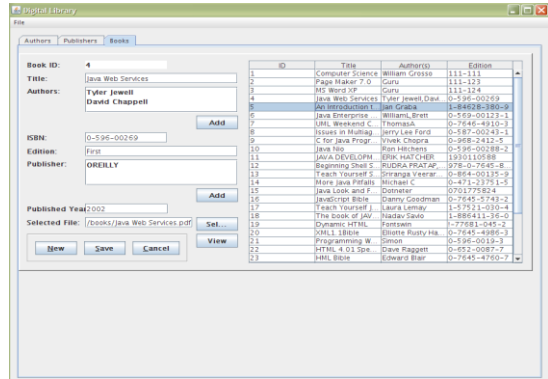

Figure 7. Entry Process for Information of Books (Server)

Administrator creates and modifies author name, publisher name and information of book in server site. Full text information of book is stored in server database. Administrator also view author name publisher name and information of books are stored. And then, Administrator search and retrieve required full text information of books. This process is shown in Figure 7.

5. CONCLUSION

This paper intends to implement the discovery and identification of the digital document files at which the desired information about education is included. There are two main parts in this system: it is to build a digital library supporting Computer Science and Technology field areas and distribute the digital document files in a LAN.

Metadata reduces time to search the information of book. This system cannot be implied with text similarity. And then, users search and retrieve the information through metadata appropriately. This digital library system is to provide only for Computer Science and Technology field area.

REFERENCES

[1] A Practical Approach ," *Alberta, Canada,*

Metadata Resources Guide", ISBN 0-7785-3109-0.

[2] Ibrahim Usman Alhaji, "*Digitization of Library Resources and the Formation of Digital Libraries*".
[3] Peter Horsman, Archiefschool, Amsterdam, "*Understanding Metadata*".

[4] http://www.university of california.com

[5] http://www.sirsi.com

[6] http://www.vtls.com

[7] http://www.luna-imaging.com

[8] http://www.museglobal.com