Преглед НЦД 14 (2009), 43–52

Teo Eterović,
Nedim Šrndić

INTRODUCING THE UNIFIED E-BOOK FORMAT AND A HYBRID LIBRARY 2.0 APPLICATION MODEL BASED ON IT

Abstract: We introduce Unified e-Book Format (UeBF) and an example Hybrid Library 2.0 application model based on it. The UeBF is a file format for storing electronic books and related metadata we have designed and developed. It differs from other electronic book file formats in its ability to store an electronic book in multiple file formats and related metadata using multiple metadata standards, all in a single file. The format is open and extensible, and the contents can be accessed using a ZIP file archiver. Using the said file format, we have succeeded in designing an application model that combines the Hybrid Library and Library 2.0 concepts. We named it Hybrid Library 2.0. It leverages the advantages of a Hybrid Library using the UeBF, like the ability to provide users with paperback and electronic books, and offers Web 2.0 presentation possibilities, such as user interactivity, social aspects and a multimedia experience.

Keywords: unified electronic book format, hybrid library 2.0

1. Introduction

Libraries play an important role in the process of education, and have the function of cultural heritage archiving. Users find the traditional libraries outdated, while the modern web-based libraries like Google Books and Amazon become more popular. In order to avoid this, it's necessary for the libraries to evolve and enhance their services in order to fulfill the requests of modern users that are used to the comfort of computer-based information access [1].

Digitizing new titles and making them available electronically, a recent trend among publishers, reflects a general enthusiasm for the possibilities electronic books offer: they are dynamic, interactive, easily updateable and, with the growth of Internet usage, available anytime anywhere [2].

However, digital and hybrid libraries use multiple, incompatible electronic book file formats [3], [4]. One example is Project Gutenberg Literary Archive Foundation, one of the leading free electronic book archives. At Project Gutenberg, they use 15 different file formats to store electronic media, 11 of those are used to store different types of textual and similar content [5].

The most widely adopted electronic book formats, such as the Portable Document Format (PDF), have a limited support for storing library catalog records and metadata. Some of the e-book file formats are closed and cannot be edited. These facts point to other problems digital and hybrid libraries are facing today: it's difficult, sometimes impossible to exchange electronic books and/or metadata among libraries; the application model used to store and present the electronic books and metadata is complicated.

2. Hybrid Library 2.0

A hybrid library is a library that contains printed and electronic resources and integrates traditional and electronic services. In order to introduce social interaction to hybrid libraries we have combined the hybrid library with the library 2.0 concept [1]. We named this new concept Hybrid Library 2.0.

The Hybrid Library 2.0 represents a link between the availability and the global advantages of the hybrid library and the library 2.0 concept of social interaction, collaboration, openness, knowledge sharing, giving priority to users and building a community, a sense of belonging.

The Hybrid Library 2.0 ideal application model contains two main components:

- ***Hybrid Library Core***, which is used to store and process information and provide an interface to the upper layers of the system. It is built from three basic components: the digital library, the traditional library and the web service with the role of an interface.
- ***User Interface***, the presentation layer, which prepares and formats the information into a form suitable for the end user. It also provides an interface to third party libraries and services. This layer is made of two components: The web application for user services based on the library 2.0 concept, and the desktop application used to manage users and books. It also manages the indexing engine with repositories using Hybrid Library Core Interface and the Management Module.

The ideal Hybrid Library 2.0 application model is shown in Figure 1.

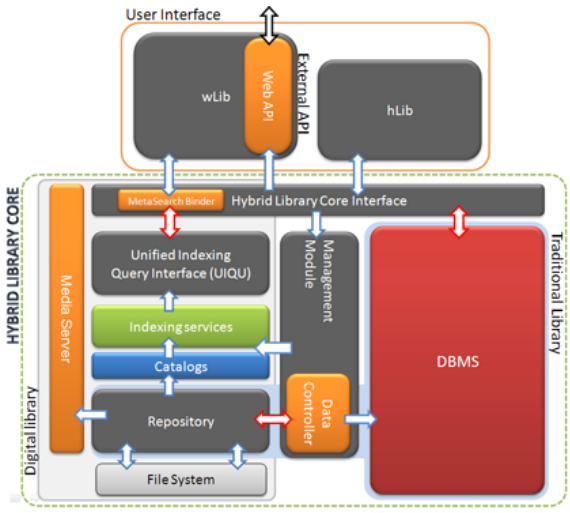

Figure 1: The Hybrid Library 2.0 Application model

The red arrows in Figure 1 represent the link between the traditional and the digital library. The connections are used to bind the two models into a hybrid library. The two top red arrows represent the search result: the full-text search of digital library resources and the book metadata from the traditional library. The search queries are executed over the HLC interface.

The bottom red arrow represents the data flow when adding electronic resources to the digital and traditional library at the same time. The resources are stored in the digital library repository as files and in the database management system of the traditional library as standard library catalog records and metadata.

The metadata may contain the name of the author, ISBN number, publishing date, tags and other information that is required to generate the specified record format (for example the MARC format).

3. Hybrid Library Core

The Hybrid Library Core (Figure 2) represents an entity that offers traditional and digital library services. These are provided over external interfaces to the user interface and to third parties. The Management Module is used to manage external services and the process of adding resources to the hybrid library.

The components of the Hybrid Library Core are the traditional library, the digital library and the interface to the presentation layer.

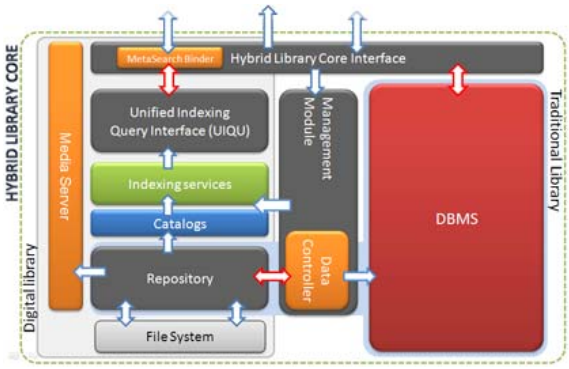

Figure 2: The Hybrid Library Core

3.1 The Traditional Library. The traditional library application model is made of two subcomponents (Figure 3):

- ***DBMS*** – A database that contains the library records and the book metadata. These represent traditional printed resources or digital resources containing format-specific metadata (for example a book, audio, video or other resources). The database also contains the data model of the application services.
- ***Management Module*** - used in the process of adding electronic resources into the repository and metadata into the database at the same time.

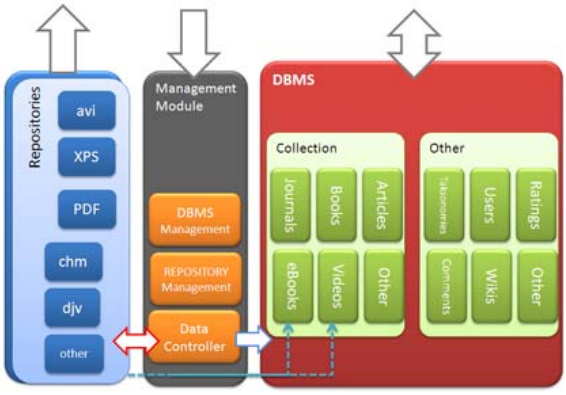

Figure 3: The Traditional Library application model

3.2 The Digital library. The digital library application model (Figure 4) provides full-text search services of electronic resources in the library repository. It also provides the availability to stream digital media live.

The digital library relies on the file system that contains repositories and the Management Module used to insert electronic resources into those repositories. The repositories can be used from third party external services. Our model of the digital library provides two external services:

- ***Media Server*** - used to stream multimedia such as video, sound, images and other resources from the repository.
- ***Indexing Engine*** - used to index text-based electronic resources from the catalog repositories for a fast full-text search. This model is able to use different indexing engines at the same time and bind the query search results into a unified reply.

The Hybrid Library 2.0 Unified Index Query Interface (UIQI) provides the feature to query more than one indexing engine in a unified way and represents an interface to the Hybrid Library Core Interface (HLCI).

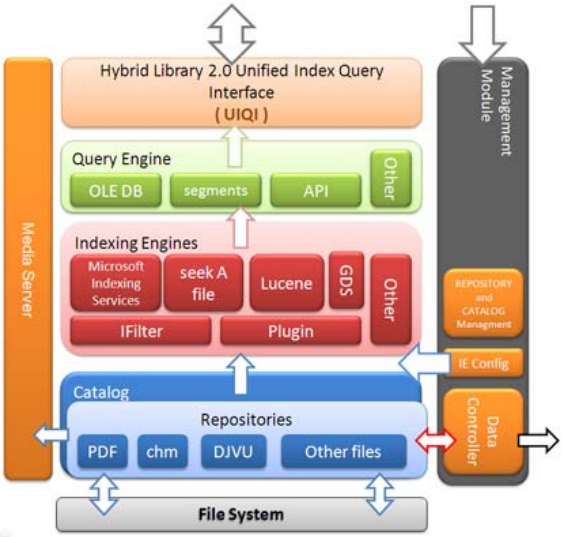

Figure 4: The Digital Library application model

4. User Interface

The user application/interface is placed on the top of the Hybrid Library Core Interface layer. This application model is made of two user applications (Figure 5):

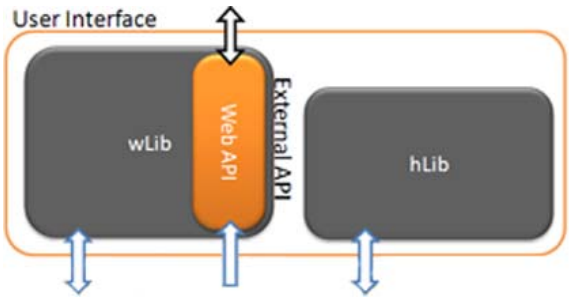

Figure 5: The User Application/Interface

- ***wLib*** - a web application based on the Web 2.0/Library 2.0 principles. The web application and the Hybrid Library Core together represent the Hybrid Library 2.0 concept.
- ***hLib*** - a desktop application used for user, metadata, module data and hybrid library repository management. A user interface of an example hLib application we developed is shown in Figure 6.

4.1 wLib User Application. The wLib user application is a web application based on the Library 2.0 principles supported with the Hybrid Library Core web service. The model of such an application is shown in Figure 7.

Figure 6: A user interface of an example hLib application

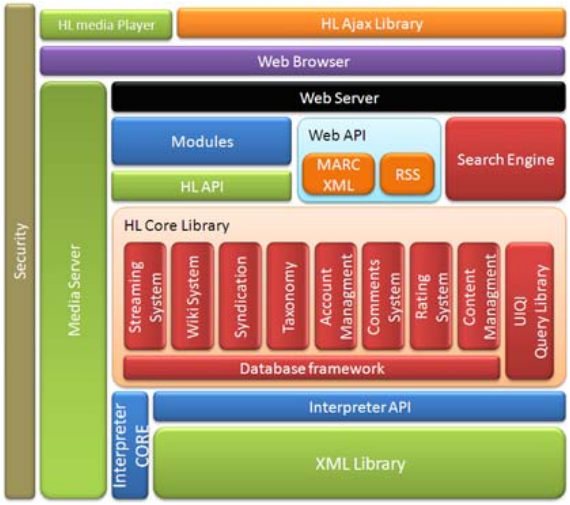

Figure 7: A wLib User Application layered model

The bottom two layers of the presented model represent the interpreter used to create dynamic websites and provide a connection to the HLCI web service.

Using the interpreter we are able to build the HL Core (Software) Library, a framework for future development of modules based on Library 2.0 principles. This framework provides an interface for communication with HLC UIQI. That enables us to provide search features to the end user over an unified interface or to bind services over the Web API (RSS, SRU/SWR).

The media player is placed on the top layer of the application and is used to reproduce the multimedia resources streamed by the Media Server.

An important aspect of this application is the security. The application is directly exposed to potential attackers so the security layer extends through the whole layered model.

5. Unified E-Book Format

As a possible answer to the presented problems, we have designed and implemented a new electronic book file format. It is called the Unified e-Book Format, abbr. UeBF. We have set the following design goals for the format:

- treating a book and corresponding metadata as a single piece of information;
- storing the same library book as an electronic book in different file formats;
- storing metadata in various formats;
- using a standard file format to encapsulate the electronic book and the corresponding metadata in a single file;
- expandability and customizability of the format for different usage scenarios and different users;
- ability to add new electronic book and metadata formats without changing the file structure;
- use of standard, open formats.

In order to achieve these goals, we have created a file format consisting of a single ZIP archive file with a predefined directory tree inside. The tree consists of:

- zero or more instances of an electronic book in some of the file formats used for electronic books (for example PDF, DjVu, CHM, TXT etc.). The files can be differentiated from each other by the file name extension. The format does not make any limitations on the file name of the electronic book;
- a directory named "meta" that is used to store metadata describing the book. This directory is mandatory, and contains the following:
  - an image file named "cover". This file contains the image of the book's cover. The file can be of any graphical file format (for example JPG, PNG etc.). The file name extension corresponds to the file format used;
  - zero or more standard library metadata files. UeBF defines the following standard names and formats, that are being used in [3], [4] and [5], description and usage is shown in [6] and [7]:
    - Dublin Core - dublin.dc
    - MARC - marc.mrc
    - MARC XML - marc.xml
    - UNIMARK – unimarc.xml;
- a directory named „custom". Users should store their specific information inside custom-made subdirectories of this directory. It is recommended to use this directory to store user-specific data, such as images or metadata in a format not defined by this standard, information on the physical location of the book inside the library building etc.

The recommended file name extension for UeBF files is „.ebook.zip".

We see that the Unified e-Book Format treats the electronic book and the corresponding metadata as a single piece of information. We find that behavior more natural than to separate the information into a file containing the contents of the electronic book, and another file containing the metadata describing the book.

The recommended file format enables libraries using different standards to store and exchange books and metadata. The process of adding new books into the catalog is simplified. All information about the book and the book itself can be kept in a single file.

Another advantage accomplished using standard and widespread technologies is the ability to create, share and view UeBF files without having to use special software. The only necessary software is a ZIP file archiver, that can be obtained free of charge on most modern operating systems. To view the contents of the e-book or metadata files, one must use the corresponding software applications.

Figure 8 shows a layout of an example UeBF file.

6. Format Conversion Process

For the UeBF to become more widely adopted, it is necessary to achieve compatibility with other existing formats. Figure 9 shows the process of creating a UeBF file.

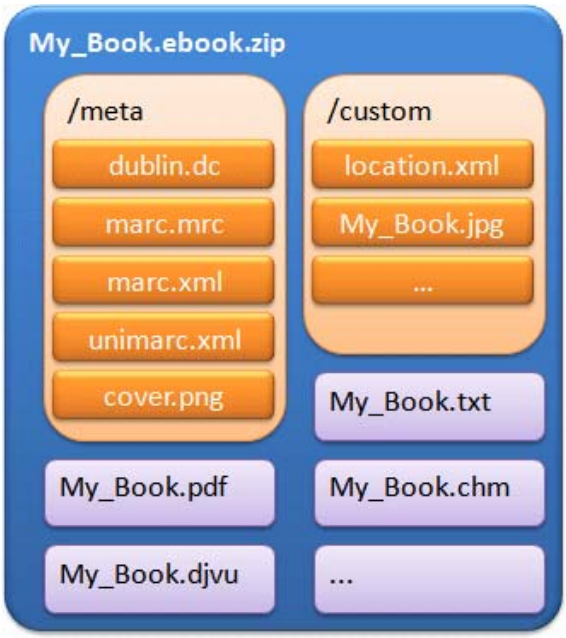

Figure 8: Layout of an example UeBF file

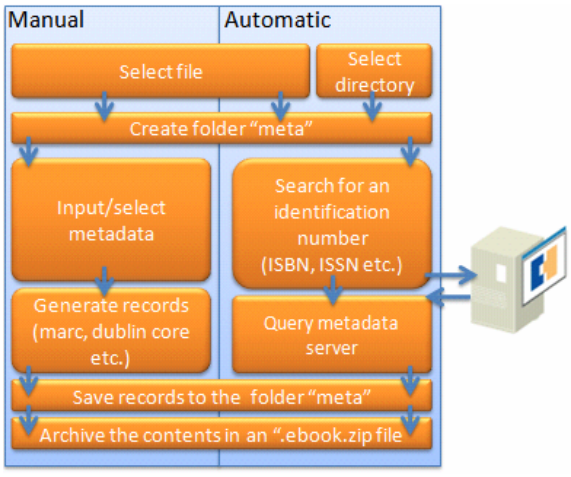

Figure 9: The process of converting existing e-book format files into UeBF

In general, there are two ways to convert an existing e-book format file into UeBF:

- ***Manual*** - First, one creates a root directory and a directory named "meta" inside. All instances of the selected electronic book in different formats are renamed to have the same, arbitrary file name, differing only in the file name extension, and then copied to the root folder. Then one needs to obtain or create standard catalog records, metadata and/or a picture of the book's cover and save them to the "meta" directory. The final step is to create a ZIP archive file with an arbitrary file name ending with the ".ebook.zip" file name extension. This file should contain the contents of the root folder.

- ***Automatic*** – The automatic conversion is made using a software tool. We have developed one for testing purposes. The process begins with selecting a file or directory that should be converted into UeBF. If a directory is selected, all the electronic books in that directory are batch processed. The application first searches for the electronic book's identification number (ISBN, ISSN, ASIN or other) and then communicates to a server that provides metadata search services. Given the identification number, the server returns the corresponding metadata, if existent. The rest of the process is similar to the manual conversion process, only it is being made by a software tool.

7. Hibrid Library 2.0 Application Model

We have designed and developed a hybrid library application model based on the UeBF. We have defined the simplicity of design and use of standard technologies as the main design goals for this model. Figure 10 shows an example hybrid library application model that doesn't use UeBF.

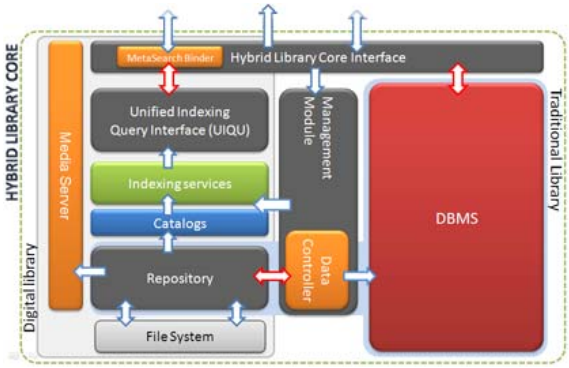

Figure 10: An example hybrid library application model that doesn't use UeBF

The illustrated application model uses a database management system (DBMS) to store metadata and file system paths pointing to individual electronic book files, whereas the files themselves are kept on a file system repository. This is not practical, as it introduces room for errors when manipulating those data. These errors can be made when creating, updating, retrieving information on or deleting an electronic book from the library. That is the reason it is necessary to use a Data Controller and a Hybrid Library Core Interface (HLCI). The Data Controller is used to separate the book and the metadata on insertion, whereas the HLCI combines the information from the book and metadata in a single piece of information on selection.

Because the UeBF is designed to treat a book and corresponding metadata as a single piece of information, there is no need to use a Data Controller and a HLCI, so the application model can be simplified.

We can also eliminate the DBMS from the application model, because the metadata is kept inside UeBF files. In this way we can fully utilize the advantages that the indexing services/engines provide. Storing all information about electronic books inside files makes backup a lot easier.

Figure 11 shows the hybrid library application model we designed and implemented. Comparing it to the previous model one can notice that it is basically a subset of it.

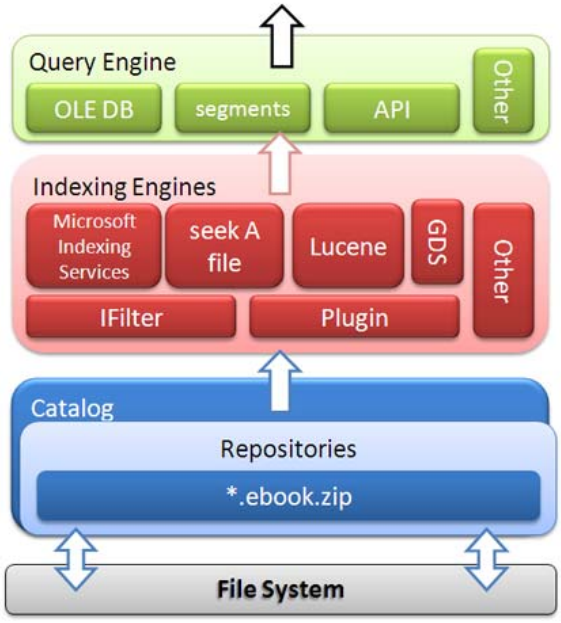

Figure 11: The designed and implemented hybrid library application model.

8. Conclusion

The presented Hybrid Library 2.0 concept is able to successfully solve the problems of previous library models by logically combining them and introducing the Web 2.0/Library 2.0. These principles represent a growing trend in solving some remnants in the modernization of the presented library models.

Considering that many experts agree that the hybrid library as well as the Library 2.0 concept is the future of libraries, we can conclude that the Hybrid Library 2.0 combines two promising concepts and has the potential to succeed as one of the library concepts of the future.

We believe that the introduction of the UeBF format into the libraries would provide a higher level of electronic book and metadata portability among them and also between the library and the library services and users.

The features of the presented simplified application model are multiple. The implementation of the whole model can be done relatively easily. It isn't necessary to develop a new software, and that saves money and time.

The UeBF format and the presented application model have been tested in the laboratory. The preliminary results are promising. We have succeeded in constructing and testing a prototype of the presented format without problems. We have also succeeded to implement the application model fast and easy.

The next step would be to confirm the usability of the model and perform more thorough and long-lasting tests in a real-world environment.

References

[1] "Do libraries matter? The rise of Library 2.0" Ken Chad, Paul Miller, Version 1.0 - November, 2005
[2] Education Guardian weekly. http://education.guardian.co.uk/egweekly/story/0,5500,680788,00.html, 2008.
[3] Northwestern University Library: Inventory of Metadata Standards.
http://staffweb.library.northwestern.edu/dl/metadata/standardsinventory/, 2008.
[4] Library of Congress. http://www.loc.gov/, 2008.

[5] Project Gutenberg Literary Archive. http://www.gutenberg.org/, 2008.
[6] Northwestern University Library: Metadata standards comparison grid in MS Excel format http://staffweb.library.northwestern.edu/dl/metadata/standardsinventory/saimsgrid.xls, 2008.
[7] JISC Collections: Metadata for e-books. http://www.jisc-collections.ac.uk/media/documents/jisc_collections/factfiles-ebooks01v4.pdf, 2008.
[8] Michelle Wake and Linda Lisgarten "User Education in The Hybrid Library – Double Trouble?"
[9] AWDEN, David and ROWLANDS, Ian. Digital Libraries: developing a conceptual framework, The New Review of Information Networking, 5, 1999, 71-89.
[10] Jack M. Maness, University of Colorado - Library 2.0 Theory: Web 2.0 and Its Implications for Libraries
[11] Shanhi, R. (2006). Web 2.0: data, metadata, and interface. June 10, 2006

teo@etf.ba
nedim.sh@gmail.com