DESIGN AND IMPLEMENTATION OF A DIGITAL ARCHIVE OF LEARNING OBJECTS FOR REMOTE ACCESS

Abdullah Al Mahbub, Md. Saifuddin Khalid, Khaled Mahmud Sujan, Mohammed Anwer
School of Engineering & Computer Science
Independent University, Bangladesh
amahbub@iub.edu.bd, khalid@iub.edu.bd, sujan@iub.edu.bd, manwer@iub.edu.bd

Abstract
Digital libraries are fast expanding into the role of independent educational entities that aspire not only to complement traditional classroom teaching, but also allow remote access to Learning Objects. Multifaceted roles of Learning Objects can be realized only if the course content and the related content management system are versatile enough to be captured into any individual’s learning needs. This paper presents the design & implementation of such a Digital Archive, which will facilitate the use/re-use, store & bookmark of Learning Objects focusing to learning needs.

Keywords: Digital Library, E-Learning, Distance Learning

1. INTRODUCTION
Advances in internet technologies have changed many activities common in life. Learning is one such activity that is currently been revolutionized by access and use of this medium. By almost any measure, this form of online, distance and continuous learning, is fast gaining popularity among learners of all ages and interests. Even the parishioners of knowledge – the educators, have found this learning medium to be a convenient and viable means of pedagogy. It is the convergence of the web and learning on all levels, whether it is elementary school, college or business [1]. An institution that allows pedagogy through electronic means via a “large” collection of information is called a Digital Library (DL). Digital libraries are fast expanding into the role of independent educational entities that aspire not only to complement traditional classroom teaching, but also allow open electronic learning for education on demand. Inherently this transforms learning management system from “teacher-centric” to “learner-centric” pedagogical endowment. Using digital libraries, education can be attained in multiple ways. Some students may prefer to go through a fixed set of digitized learning material for self-paced learning. Others may prefer real-time interactive education with the educator and the learner brought together in the virtual medium. In the last century, radio, television, and satellite broadcasting technologies have provided new medium for delivery of distance learning. Since computers offer so much flexibility and variety, so does e learning [2].

The research effort presented in this paper aims to address this issue of standardization of development of learning material and other relevant issues. More specifically it will target the following:
Develop a cohesive and comprehensive “learning object” model, alternately called the “data module” or “learning module”. This refers to the chunks of resources that have educational value. Important issues that need solution here are how to meta-tag (identify and link) and “package” these chunks; how to ensure that the “packaging” is scalable yet malleable to be easily handled across various parts of the digital library; and how to make use/reuse of these chunks efficient while maintain copyright considerations.

2. DIGITAL LIBRARY – THE CONCEPT
The basic functionalities provided by any DL are somewhat parallel to that of traditional libraries, but it also encompasses the role of educational institutions. Books or resources, as in a traditional library, are collected and upon screening and cataloging, are finally made available to its patrons. However the similarities end there. The architecture of digital libraries to support such activities is quite different from those of traditional libraries. The brick-and-mortar buildings are replaced with databases, e-learning tools, content servers, web-portals, etc. The books, in this case are the digital resources and constructs. The common denominator among them is their ability to provide knowledge or education.

2.1. RESOURCE – BOOKS
Since the DL resides in the virtual medium, the learning materials are maintained in digital format. In almost all the field of arts, sciences, engineering, technology and others, the knowledge providing content come in varying sizes and formats and sometimes in mixed formats. Moreover the material can be presented as text, graphics, animated graphics, audio, video, or a combination of these [3]. Although this is basically a source of information, patrons of the library should be able to easily browse, collect, reference and reuse (with permission) these data modules.
2.2. THE CATALOGS, CUE CARDS AND LIBRARY CARDS

In a traditional library, the cue cards hold information about a book in order for easy discovery and locating. Moreover the cue cards in themselves are arranged in some intuitive order for convenience of browsing and subsequent retrieval. Its electronic counterpart – the metadata, performs a similar function in digital libraries. *Metadata* is defined as data about data. It holds information that will assist a user in quickly locating the learning modules of interest. A limited list of elements of interest to any audience is given in Table 1.

<table>
<thead>
<tr>
<th>Table 1: Metadata Elements of Interest for Cataloging</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category</strong></td>
</tr>
<tr>
<td>Information about the Contributor</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Basic Metadata</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The first row of metadata pertains to “ownership” or “copyright” aspect of the information. This is generic information like the author(s)’s or publisher’s or contributor’s name, affiliations and contact information. Besides providing inherent copyright value, this kind of data can also be used to identify the content. More common identifiers of information are the title, description, classification and related items as listed in the second row of the table.

2.3. THE LIBRARY

Digital libraries similar to traditional libraries, undertake the tasks of collection, screening, packaging and archiving of educational content for easy development, discovery, delivery, use and reuse. Acquiring contents are done through active harvesting develops collections. Since, as mentioned before, the learning modules are available in heterogeneous formats, they have to be packaged before submission. There are many content packaging methodologies available for bundling resource for e-learning. The packaging has to bind the resources to the metadata and also perform validity and integrity check of the resource files.

3. TECHNOLOGY REQUIREMENTS OF DIGITAL LIBRARY

Currently many technologies attempt to address one or the other need of a digital library administrator, technologies can be broadly classified under the following two categories [4]:

- Technologies for developing learning programs (the “books” and “cue cards”)
- Technologies for handling and delivering learning programs (the “library” architecture)

The technologies for developing the learning program are “many-to-one” type – implying that the user of the technology should have the ability to form a single coherent collection out of many different chunks of resources having heterogeneous educational value. They encompass the hardware and software needed to create the learning materials, prepare them for “publication,” and administer the use of the materials once they are published.

The technologies for handling and delivering the learning programs are “one-to-many” type – implying that the users of these technologies should have the ability to dissipate knowledge possessing modules in a manner that enable anyone requesting them to gain pedagogy via the material.

4. THE LEARNING OBJECT (LO)

4.1. DEFINITION

A learning object is defined as a structured electronic resource that encapsulates high quality information in a manner that facilitates pedagogy. It has a stated objective and audience. It has ownership and associated intellectual property rights.

The definition highlights two aspects namely, “learning” and “object” with the underlying theme being “ownership” and “quality”. It is immutable in the sense that it may not be changed or altered by the user. It must function independently (self-contained and standalone). Also note a LO may have intelligence and additionally, have other LOs embedded in it.

4.2. LEARNING

The “learning” aspect of the LO definition is very important. Learning refers to education. Education is a process. The main aim of DL is to allow continuous education. Digital libraries are more than a simple content repository or a hosting platform, promoting continuous education and life-long learning. Learning from a LO can be envisioned in two ways. The first approach is for the LO to be an independent entity i.e. as an individual topic. In this case, it is important that it be coherent, self-contained and rich in its pedagogical value. This kind of learning suits best to learners who do not have time to follow extensive course layouts or who are seeking a re-fresher to their topic of interest. Another way to gain learning is via
following an ad-hoc path from one learning object to another object. In this case, the learning objects are not only self-contained for that subject but also assists in step-by-step pedagogy from a starting topic to an end topic of desired pedagogical goal.

4.3. OBJECT
The “object” aspect of the LO definition relates to digital/electronic format of the resources. Simply put it the actual physical resource (the files) that forms the learning object. Since digital construct can be of any size and format, so are the objects within the LO. The digital constructs may encapsulate both information and behavior. Since it’s the objects in L0s that have knowledge giving information, the act of accessing, transmitting and distributing them should be as simple for handling across the different spectrum of the modern learning profile. Digital libraries have to assist in streamlining the processing of these objects. Most of these activities occur via the Internet. Additionally, it must be capable of capturing high quality and quantitative metadata about the objects. Further the tool must package the object such that it facilitates fast and efficient delivery. Currently many approaches are at works for information delivery [5].

5. SCHOOL OF ENGINEERING & COMPUTER SCIENCE DIGITAL LIBRARY: A LEARNING OBJECT MODEL
A learning object is defined as a structured electronic resource that has two components: metadata and resources [5]. In order to facilitate a goal-oriented pedagogy while at the same time preserve copyright, it is important to collect a vast range of the metadata. The quality of the metadata facilitates a natural discovery for the LO. From a technical perspective, the LO is collection of raw resources that are “structured”. Hence, resource validation is an equally important part besides metadata collection. Finally, both these component have to be bundled for transmission and distribution within the digital library. School of Engineering & Computer Science (SECS) Digital Library is such an effort to fulfill the requirement of a unique learning object model. The successful implementation of this learning object model gave us the opportunity to disclose its architectural and technical aspect through this paper.

5.1. THE METADATA COLLECTION
Any Digital Library has to collect, process, store and convey the information about the metadata. An interface must have well defined fields to enter the data. Fields like author’s name, email address etc. will harvest the metadata relating to the author. Further fields will be available to capture the characteristics and technical properties of the LO. These are the educational metadata. Here there are two options. One let the content creator decide the values of the field based on his/her best experience or prior knowledge. Two, the tool contains pre-defined values for the areas so that the content creator can select the best fit for their content [5]. As high quality of metadata is required (later for quality control and high-yield discovery), the mixed approach to allowing creator to fill-in some metadata areas, while selecting others is judged to be most optimum approach to metadata collection. The collected metadata can be stored and transmitted by many means and methods. We have two powerful, new tools at our disposal: the PHP scripting language, and the MySQL database engine. In SECS Digital Library we have used these two tools successfully.

5.2. CONTENT STORING & CATEGORIZATION
As we have discussed in section 2.2, the cue cards hold information about a book in order for easy discovery and locating of a book in a traditional library. But in our case where the information should be kept in electronic form and the book [LO] itself is digital we have to implement some algorithm to store the both in a database in an effective way. Since SECS Digital Library is a database driven system, we have stored the category of the LO as well as the LO in the database. But the categorization should be made in such a order so that any new entry in the existing category in any level is possible as well as there should be the flexibility to add new category in any level of categorization.

Figure 1: Second level categorization

Figure 2: Third level categorization

Here figure # 1 shows the implementation of second level categorization while figure # 2 shows the third level of categorization, which is implemented in SECS Digital Library. In this aspect we have
implemented the concept of R-tree. An R-tree is a height-balanced tree similar to other tree structure with index records in its leaf nodes containing pointers to data objects. Nodes correspond to database, if the index is database resident and the structure is designed so that a special search requires visiting only a small number of nodes. The index is completely dynamic: inserts and deletes can be intermixed with searches and no periodic reorganization is required [6]. Since Digital Library should be capable of storing various types of LO, SECS Digital Library stores the content in such a database, which is a collection of tuples representing special object and each tuple has a unique identifier to retrieve it. Leaf nodes in an R-tree contains index record entries of the form

(I, tuple - identifier)

Where tuple - identifier refers to a tuple in the database and I is a n-dimensional rectangle which is the boundary box of the object indexed [6]

\[ I = (I_0, I_1, \ldots, I_{n-1}) \]

Here \( n \) is the number of dimensions and \( I_i \) is a closed bounded interval \([a,b]\) describing the extent of the object along dimension \( i \). Alternatively \( I_i \) may have one or both endpoint equal to infinity, indicating that the object extends outward indefinitely. Non-leaf nodes contain entries of the form

(I, child-pointer)

Where child pointer is the address of a lower node in the R-tree and \( I \) covers all rectangles in the lower node’s entries [6]. Figure # 3 shows the structure of R-tree.

The backend consists of a number of tables comprises a database which resembles data tuples at the bottom level. To make sure the integrity of the metadata (content categorization data & information about the creator) we have used several keys which is actually index (discussed in the previous section). It has been implemented successfully because MySQL is a relational database and it provides a great way to store and access complex information. The middleware is basically the PHP 4.0 scripting language. PHP belongs to a class of languages known as middleware. This language works closely with the Web server to interpret the requests made from the World Wide Web, process these requests, interact with other programs on the server to fulfill the requests, and then indicate to the Web server exactly what to serve to the client's browser. The middleware is where the user will be doing the vast majority of his work.

The client layer consists of a common web browser that communicates with middleware’s web server. A general description of the architecture is shown in figure # 4.

**5.3. SYSTEM ARCHITECTURE**

The SECS Digital Library is based on three-tier architecture capable of handling metadata on the Web. Where, both the content and the categorization of the content have been stored in a database, which is one of the three tiers of the system. But any user through web browser has to be given the complete access to those contents. For that we have to design a middle tier that will work as an intermediary between the two. As a result the data recovery becomes more easy and reliable. More precisely PHP 4.0 & MySQL works in such a manner to support this system architecture.

**5.4. CONTENT MANAGEMENT**

Content management is the task of application server and performance operation on the content and the related metadata. One of the main tasks is to assist learners is search and recovery of the desired LO. This process should support the natural discovery process of the learners to zoom-in on the LO via provided search elements. In SECS Digital Library, three types of search are implemented.

- **Simple Keyword Search**
  A keyword search is executed by collecting the search query and parsing it for given words. Stop words are dropped and stemming is done wherever possible. Once this is complete a sequence of queries is executed (either in a sequential manner or simultaneously) using sets of procedures and results are obtained from the database containing metadata stored as reverse indices. The results of the search are sorted in the order of degree of match and precedence, and presented to the patron. The user then has the option of viewing the metadata, previewing or even downloading the learning object.

- **Title / Author Search**
Collecting the content title / author name from the user and parsing it to the database executes a search. This search is also executed in the above manner. But the only difference is that it can find the similar titles also, which facilitates the user to the more relevant contents at the same time.

- **Taxonomical Search**

SECS Digital Library provides its users the opportunity to browse through its collection using a pre-defined taxonomy. A multi-level taxonomy (*discussed earlier as content categorization*) has been designed to sub-classify each fields into several subgroups. This provides a unique opportunity to a user to track his learning resource based on taxonomical classification. Future plans include making available a combination of advanced/keyword search along with browse functionality.

### 5.5. USER MANAGEMENT

Throughout the above discussion, access control has been the central issue in patrons interface with the library. When patrons register with SECS Digital Library, a profile is created that is stored in the user profile repository (Database). This is equivalent to the “library card” in traditional libraries with added security of password protection. Since SECS Digital Library is a stand alone project of Independent University, Bangladesh, it’s user authentication is only given to those who are the registered students of the university and faculty members or part of the management team. An automated “Sing Up” system has been developed to provide the authentication to the feasible users.

### 6. ECONOMIC, SOCIAL AND LEGAL ISSUES OF DIGITAL LIBRARY

Digital libraries are not simply technological constructs; they exist within a rich legal, social, and economic context, and will succeed only to the extent that they meet these broader needs. Rights management, economic models for the use of electronic information, and billing systems to support these economic models will be needed. User privacy needs to be carefully considered. There are complex policy issues related to collection development and management, and preservation and archiving. SECS Digital Library may shed some light on these questions. The update is we are very much aware of these issues and from the very beginning of its development we are maintaining such issues in SECS Digital Library.

### 7. CONCLUSIONS

A digital library can be successful only if the underlying system architecture is robust. While the hardware and software aspects play a major part in implementation, it is the interaction and functionalities that are provided by the entities associated with digital library that will allow efficient pedagogy. The modern profile of the digital library is such that these entities, while acting independently, have a unifying goal to sustain the paths to knowledge from the educator to the learner. It is modular and hence more flexible in meeting the changing landscape of education endowment. Furthermore, digital libraries must themselves be scalable to integrate various libraries representing different disciples together. All this is possible only because of the central and critical role played by metadata. The educational content is processed based on the values of elements in the metadata. Thus, metadata and content should inherently be grouped together by association. The learning object model defines such an association.

### References