NSF Annual Report:
Collaborative Research: Curriculum Development: Digital Libraries

Note: Virginia Tech (VT) grant: NSF IIS-0535057 and
University of North Carolina at Chapel Hill (UNC-CH) grant: NSF IIS-0535060

PI (VT): Edward A. Fox;
PI (UNC-CH): Barbara M. Wildemuth
Co-PI (UNC-CH): Jeffrey P. Pomerantz

Date: Nov. 7, 2006

1. Participants

- Project Team at VT:
  - PI: Edward A. Fox (fox@vt.edu)
  - GRA: Seungwon Yang (seungwon@vt.edu)
- Project Team at UNC-CH:
  - PI: Barbara Wildemuth (wildem@ils.unc.edu)
  - Co-PI: Jeffrey P. Pomerantz (pomerantz@unc.edu)
  - GRA: Sanghee Oh (shoh@email.unc.edu)

2. Activities and Findings

2.1 Describe the major research and education activities of the project.

The first-year activities of this project have focused on three goals: to further define a framework for a curriculum in digital libraries (DL), to develop some of the modules that would be included in such a curriculum, and, in order to support these first two activities, to establish and interact with a project advisory board.

2.1.1. A digital library curriculum framework

Our original proposal included a draft framework, organizing the components of a curriculum for digital library education that would be suitable both for computer science programs and for library and information science programs. It included 18 modules: 9 covering core topics and 9 covering related topics.

Two approaches were used to validate this preliminary framework and develop it further. The first was a theoretical approach, using the 5S framework (streams, structures, spaces, scenarios, and societies) to ensure that the framework would
include all the necessary components of a DL. The second was an empirical approach. We analyzed current literature related to digital libraries (189 ACM DL conference papers, 1996-2000; 354 JCDL papers, 2001-2005; and 521 articles published in D-Lib Magazine, July 1995-February 2006). Each of these papers was classified, using the preliminary framework. This manually-classified data may be used to develop a supervised-learning classification system for future classification tasks. Based on this theoretical and empirical work, the framework has now been modified to include 10 core topics and 36 related topics. Our current module set can be accessed at [http://curric.dlib.vt.edu/wiki/index.php/Module_Diagram](http://curric.dlib.vt.edu/wiki/index.php/Module_Diagram). During this period, as we were continuing to refine the framework, it was presented at the 9th International Symposium on Electronic Theses and Dissertations (ETD 2006), by Edward A. Fox.

2.1.2. Development of curriculum modules

Each of the core and related topics in the current framework is a candidate for development as a curriculum module. To support the development of these modules, an analysis of current syllabi was undertaken, with particular attention to the assigned course readings. We considered the DL course syllabi from ALA-accredited programs in Library and Information Science (LIS). Forty courses were identified; of these, 20 syllabi were collected. From these course reading lists, 1,697 titles were classified into the 9 DL core topics included in our preliminary framework. From this data, we could find most frequently assigned books, journal articles, journals, and most frequently assigned authors. The complete data set can be accessed at [http://curric.dlib.vt.edu/wiki/index.php/Link](http://curric.dlib.vt.edu/wiki/index.php/Link). The results from the analysis will be published in *D-Lib Magazine* in November, 2006. A parallel analysis of computer science (CS) course syllabi is currently ongoing.

During the analysis of the LIS course reading lists, we found words and concepts in the reading lists that could not be properly classified into the 9 core topics. Examples of those include ‘project management’, ‘DL evaluation’, and ‘cost/economic issues.’ Therefore, we extended the 9 DL educational modules to 10 modules ensuring coverage of those concepts.

The syllabus analysis provides a strong starting point for the development of specific curriculum modules – the next step of our work. In order to develop a workable template for module development, two draft modules (‘Multimedia, Hypertext and Information Access’ and ‘Information needs, Relevance, Evaluation/Effectiveness’) were developed and discussed by the project team and some members of the project advisory board in June, 2006 (please see [http://curric.dlib.vt.edu/wiki/index.php/Modules](http://curric.dlib.vt.edu/wiki/index.php/Modules) for more information). Based on this discussion, we developed a template that will be used for the development of all the modules.
Four modules have now been completed, and will be provided to the advisory board members for discussion at (or shortly after) the meeting of the American Society for Information Science & Technology (ASIST) in November 2006. These four modules are:

- Intellectual Property Rights in a Digital Library (Meredith Weiss and Barbara Wildemuth);
- Information Needs and Assessment of Relevance (Barbara Wildemuth);
- Search Strategies (Sanghee Oh); and
- Reference Services (Jeff Pomerantz).

2.1.3. Advisory board activities

Prior to submitting our proposal, we had recruited an advisory board composed of VA Tech faculty (10), UNC-CH faculty (13), and additional external scholars (12, now expanded to 13). Please see http://curric.dlib.vt.edu/wiki/index.php/Advisory_Board for a list of members.

While we have not yet convened the entire advisory board in one location, we have held several meetings that included 6-12 of the members at each event. These meetings were held:

- At Virginia Tech, Blacksburg, VA, May 1, 2006;
- In conjunction with JCDL, Chapel Hill, NC, June 14, 2006 (including several additional scholars not yet named to our advisory board, but who are willing to participate in the project); and
- In conjunction with ASIST, Austin, TX, November 6, 2006 (planned; expect 7-9 members to attend).


2.2 Describe the major findings resulting from these activities.

Two primary analyses have been conducted this year: 1) a longitudinal analysis of publications on digital libraries, and 2) an analysis of digital library course syllabi in library and information science. The findings from the first analysis were reported at the 2006 Joint Conference on Digital Libraries; the findings from the second are included in a D-Lib Magazine article, currently in press. The primary findings from each are summarized here.

The analysis of publications indicated that the most common topics at DL conferences over the past ten years have been ‘services’ and ‘architecture.’ For D-Lib Magazine articles, ‘legal issues’ as well as ‘services’ and ‘architecture’ were the frequent topics. Based on this analysis, those aspects of our module framework have been expanded.
The analysis of course syllabi found that 29 (of our sample of 55) library and information science programs offer courses in digital libraries. The most frequently-assigned books, articles, and authors were identified. Readings in the areas of ‘project management’ and ‘architecture’ were more frequently assigned than readings in other areas. Based on these findings, we added a module to our framework, specifically dealing with project management.

2.2.1 What research and teaching skills and experience has the project helped provide to those who worked on the project?

The project team consists of two full professors (one at each institution), one assistant professor (at UNC), and two research assistants (one at each institution). Each of the team members has increased their research skills and experience through this project, as well as gaining insights into curriculum development in the area of digital libraries.

In particular, the research assistants have been involved in two areas of the project. First, they have been directly involved in the collection and analysis of data concerning publications about digital libraries (from ACM DL and JCDL conferences and published in D-Lib Magazine). They more fully appreciate the need to gather data systematically and verify its accuracy before drawing conclusions. In addition, they have gained significant experience in reasoning from the raw data to our conclusions about which modules would be most appropriate for inclusion in a digital library curriculum.

Second, they have been involved in the development of content for the curriculum modules. This experience allows them to participate in one aspect of teaching, i.e., course development. For each module, Mr. Yang or Ms. Oh must thoroughly investigate the topic (including reviewing existing syllabi that include the topic) and make judgments about specific areas that should be covered in the module and how those areas should be covered.

2.3 Describe the opportunities for training and development provided by your project.

- For training
  Once all modules are developed, we will create one-semester and two-semester courses from the modules. It will be a flexible design so that the instructors can take out some modules or add other modules to fit their needs. Several people, including some of our advisory board members and other international collaborators (e.g., Dr. Chakraborty, Bose Institute, India, and Dr. Torres, UNICAMP, Brazil), have already agreed to use these modules in their classes. It will be a good training opportunity for the participating instructors and a good learning opportunity for students. Evaluation and update of the courses will follow their initial implementation in the field.
• For development
Along with the module content preparation, we will consider using an authoring/presentation tool which can effectively deliver the module contents to students. If we cannot find tools that satisfy our purpose, we would like to develop one. It will be an open source tool.

2.4 Outreach Activities

Our outreach activities during the first year have included both face-to-face activities and online activities. We have met several times with our advisory board members, and have made multiple conference presentations about the project. In addition, we have established a project homepage and wiki (please see section 3.2).

PI Fox has given two tutorials to help disseminate our results and approach, as follows:
• Introduction to (Teaching / Learning about) Digital Libraries. Half-day tutorial, ECDL 2006, Sept. 17, 2006, Alicante, Spain
• Introduction to (Teaching / Learning about) Digital Libraries. Full-day tutorial with M.A. Goncalves, JCDL 2006, June 11-15, 2006, Chapel Hill, NC

3. Publications and Products

3.1 What have you published as a result of this work?

3.2 What Web site or other Internet site have you created?

- Project homepage: [http://curric.dlib.vt.edu](http://curric.dlib.vt.edu)
  It contains links to the NSF proposal, abstract, FAQ, new and interviews, and project member contact information.
- Project wiki: [http://curric.dlib.vt.edu/wiki](http://curric.dlib.vt.edu/wiki)
  It includes our publications, presentation slides, current module set diagram, module template, draft modules, and resources such as DL syllabi reading classification data and educational resources taxonomy.
  It shows discussion summary, project meeting results, pictures of visits, etc.

3.3 What other specific products (databases, physical collections, educational aids, software, instruments, or the like) have you developed?

The project activities have resulted in the development of four course modules so far (listed in section 2.1.2). In addition, the raw data from the analysis of course syllabi will be posted on the project website and made available to other scholars in this area.

4. Contributions

4.1 Contributions within Discipline

This project was interdisciplinary from its inception. While its products (the curriculum framework and modules) will contribute to digital library education in both computer science and information and library science, it is our goal that they will bring these two disciplines closer together in their work on digital libraries. Specifically, the high quality DL curriculum (the final product of this project) will produce DL researchers, designers, and administrators who understand all the necessary aspects of DL system and services as well as his or her DL specialty area. These researchers, designers, and administrators will come from both of the participating disciplines.

DLs incorporate various technologies such as database systems, information retrieval systems, user interface designed by human-computer interaction knowledge, network systems, multimedia systems, recommender systems, etc. Our DL curriculum project will provide students in both computer science and information and library science with opportunities to understand how each of the above technologies work and how they can interact together efficiently. Doing research to create better DLs will require and lead to the advancement of technologies used in DLs. Those advanced technologies will be used to develop more efficient and effective DLs. Therefore, DLs and the related technologies help each other and in the center of this interaction, there are DL researchers who are educated by our curriculum.
4.2 Contributions to other disciplines of science or engineering

DL experts produced by our curriculum will be working in other disciplines of science and engineering, developing high-quality DLs. Valuable knowledge in those disciplines will be stored, organized and served efficiently by the DLs. The access to the information will become easier. Users will find high-quality information. Our contributions should lead to the overall improvement of research efficiency and the advancement of any discipline that relies on digital libraries.

4.3 Contributions to the development of human resources;

Courses created by this project will provide an organized learning experience to graduate-level students and current DL designers and administrators, increasing the possibility of improving their job performance and growth.

4.4 Contributions to other aspects of public welfare beyond science and engineering, such as commercial technology, the economy, cost-efficient environmental protection, or solutions to social problems.

DLs can be developed for many purposes, each benefiting the public welfare. For example, a DL that provides information about cost-efficient environmental protection will help people involved in this area. They might be able to find the most recent methods of cost-efficient protection. They can also upload and share data, case studies, and best and worst practices in the field. Use of a well-designed and well-managed digital library can lead to improved methods and reduced cost for environmental protection. Our curriculum project will contribute to the public welfare by producing DL experts who can create and administer such DLs.