John Kalkbrenner
ILS655 Assignment - Digital Library Project Review Paper 02 (Virginia Tech University Digital Library and Archive)
February 20, 2011

One: Project Background

I choose to look at the Digital Library and Archive at Virginia Tech University.   The

digital library at Virginia Tech University (http://scholar.lib.vt.edu/index.html) was

originally a Gopher server started in 1989. Gopher was a precursor to the World Wide

Web and was text based rather than having a graphical user interface. The Digital

Library and Archive at Virginia Tech University is one of the oldest digital libraries in

higher education and originally ran on a NeXT 3.3 computer called Borg. Interestingly,

the NeXT Machine was a venture started by Steve Jobs when he left Apple computer in

the 1980s.   The venture failed and Jobs eventually returned to Apple. The digital library

at Virginia Tech University was migrated to a Web server in February 1993 very shortly

after the advent of the World Wide Web. The collection grew quickly and by June of

1995, there was over 4 gigabytes of scholarly electronic publications available through

the Project's servers.

Virginia Tech Digital Library and Archive evolved from the Scholarly Communications

Project that began in 1989. That project focused on working with the university

community to host its unique digital resources. The Virginia Tech Library also began

regional collaborations; in 1992, they hosted the first online news reports from the Roanoke Times & World News.

The Virginia Tech ImageBase hosts historical photographs as well as images that are more contemporary. For example, the International Archive of Women in Architecture features images of architecture designed by women spanning more than a century. In addition to having online resources, like the image database, Virginia Tech Digital Library and Archive created and maintains a Survey Database where data is collected, tabulated, and displayed. Virginia Tech Digital Library and Archive hosts a growing number of faculty archives and Virginia Tech publications.

Virginia Tech Digital Library and Archive hosts documentation resulting from the tragic events of April 16, 2007, when a student shot and killed several students, faculty and staff members. Government reports, local projects, and images from the out-pouring of sympathy are available in this digital collection. After the shootings, Virginia Tech received thousands of cards and letters of support, banners, art, and other unique items from around the world. Many of these are available in a searchable database, the Condolence Archives of the University.

Three grants from the National Science Foundation have been awarded to Virginia Tech in recent years. All relate to digital libraries and scholarly communication. Starting in September 2008, with further funding in August 2009 the National Science Foundation awarded over $425,000 in grant monies for a project called Ensemble.

The Ensemble project, when it is completed, will add a computing pathway to the existing set of NSF Science Technology Engineering and Mathematics (STEM) Digital Libraries. Ensemble will reach across the full range of audiences for computing education from K-12 to graduate and professional education. Ensemble creates a distributed portal providing access to the broad range of existing educational resources while preserving the collections and their associated curation processes. Ensemble will bring together many digital libraries to make searching multiple libraries easier.

This continuing grant is overseen by Program Manager, Herbert H. Richtol, of the National Science Foundation's Division of Undergraduate Education. Investigators under this NSF grant are Edward Fox (Principal Investigator) and Steven Edwards (Co-Principal Investigator) both of Virginia Tech University.

In 2009, Virginia Tech was awarded another $170,000 from the NSF for the K-Gray Engineering Pathway, which provides resources and services to users and collection

developers from a broad range of engineering constituencies. During the last several years, Engineering Pathway substantially grew its user base and number of catalog records in both content and functionality. The Pathways project is very similar to the Ensemble project in that it joins a number of digital libraries. However, the digital libraries are cutting-edge engineering research libraries pertaining to biotechnology, energy, sustainability, security, and nanotechnology.

In 2010, Virginia Tech was awarded another National Science foundation grant of $150,000.   Under principal investigator Clifford Schaffer, the project aims to establish a new model for online educational communities. It will focus on digital community model, where members' discussions, reviews, and ratings of content items available alone with information found by searching a collection. This includes traditional distribution mechanisms such as online forums, email, and online blogs and newer social networking sites including Facebook.

Two - Organization of Resources, Scope, Collections/Databases

Virginia Tech Digital Library and Archive has developed a range of online collections since its origins as the Scholarly Communications Project in 1989. Some were

developed for the university at large, such as Virginia Tech ImageBase (digital image database) and ETDs (electronic theses and dissertations). Others grew out of continuing collaborations with various colleges of the university. For example, the Virginia Tech Library worked with the College of Architecture to develop the International Archive of Women in Architecture, and with the College of Culinary and Hospitality to create the Peacock-Harper Culinary History Collection.

Several collections have resulted from the Virginia Tech Library working directly with Virginia Tech faculty, for example, the publication of several online academic journals, varying form the Journal of Technology Education to Nantahala: A Review of Writing and Photography in Appalachia. Virginia Tech Digital Library and Archive hosts projects in the Faculty Archives, which are directly related to the academic, research, and teaching or to the extra-curricular interests of the Virginia Tech faculty. These archives include both scanned and so-called "born-digital" projects, which never existed in non-digital format.

Virginia Tech Digital Library and Archive collaborates with other universities as well to expand the use of technology for digital libraries. Its participation in the Meta Archive Cooperative, for example, has contributed to the development of a distributed digital preservation network. One outcome is the preservation of more than 70 unique library

collections securely stored at six geographically dispersed sites. The grants mentioned above, funded by the National Science Foundation, will further the sharing of this digital libraries information with others around the nation.

Three - Service Features

Searching the Virginia Tech Digital library is very easy. Users can search the entire library with a very simple Google search box located on the front page. An advanced search option is also available. While the advanced option is very similar to the familiar Google advanced search, it has additional features. For example, a user can specify that the words being searched for be located only in the URL, the text of the item, or both. The search screen will be familiar to anyone who is familiar with Google; and since that is one of the leading search engines, training users about how to use the system is simplified. See Figure 1, below which shows the search screen.

While the search function is easy to use, after a little experience with the Virginia Tech Digital library I realized that my searches were retrieving more information than I could really use. I was retrieving hundreds of items and I had to sift through then one by one. A way to narrow my searches would be to choose specific collections within the digital

library. However, the content of each collection is not very well documented. I visited

the "ask a librarian" webpage hoping to access a reference librarian, perhaps by a chat

mechanism.   Instead, I found a list of frequently asked questions and a web form that I

could complete to send a question to a reference librarian. I competed the form and

submitted it asking for information about early computers. To date, I have not heard

back from a librarian. Working with such a large and diverse digital library, I see that

reference librarians are really required to use this resource effectively. The number of

collections and volume of information is tremendous.   The names of collections helps a

patron look in the right collection, but some guidance from a librarian would be very

welcome.

Figure 1.   Virginia Tech Digital Library and Archive search screen.
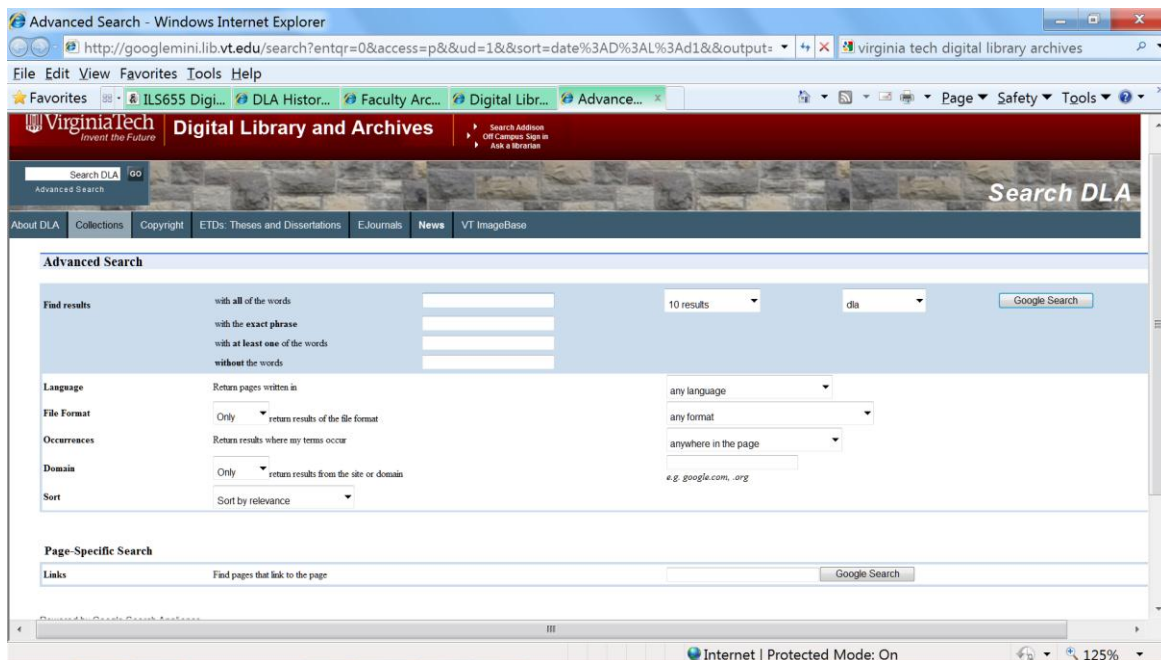
Alternatively, users can look through a list of collections within the Virginia Tech Digital Library and Archive. The Collections include culinary history, eJournals, faculty, manuscripts, rare books, university libraries publications, university history and the Virginia Tech ImageBase as well as a collection of theses and dissertations.

Four - Technologies

The Virginia Tech Digital Library and Archive has over six terabytes of data in stored on twenty-three servers. That data is managed using dynamic load-adjusting servers, which include six real hardware servers, three virtual servers, and ten Web servers.

In addition, there are three physical hardware servers providing MetaArchive preservation cache services: two are production systems and one is the MetaArchive Test Network server. An additional Sun server hosts ContentDM local services. The scripts that manage the ImageBase are written in PHP, and read and write information to a MySQL database.

According to the Virginia Tech web site, the change in servers reflects a major reorganization that took place in the wake of the May 2009 systemic failure of the dynamic load-adjusting servers. The major focus of change has been to remove some

of the unnecessary complexity of the original dynamic load-adjusting cluster setup and to streamline operations onto fewer machines.

The digital collections are developed using ContentDM from OCLC. Web pages have embedded code from Google Analytics to facilitate tracking usage both in terms of quantity of usage and type of patron.

Five – Comments and Suggestions

What impressed me most about the Virginia Tech Digital Library and Archive is the amount of information available there. While many libraries are just now wading into the world of digital libraries, Virginia Tech has been doing it for 22 years. The volume of data and the complex computing infrastructure are extensive and impressive

The more recent projects, funded by the National Science Foundation, are related to how to join the Virginia Tech digital library with other digital libraries around the nation. I became aware of how challenging it can be to navigate a large digital library.   In my first review of a digital library, I looked at the Villanova University Digital Library. That collection has less than ten thousand records.   It was relatively easy to find items.   The categories/subjects seemed very clear and discrete.   The Virginia Tech Digital Library

and Archive has over a million records and many of those items fit in with several

collections.

Clearly, the profession of reference librarian is not endangered. On the contrary, if the

Virginia Tech Digital Library and Archive is any indication of what large digital libraries

will be like, they are hard to navigate and the assistance of a reference librarian is more

important than ever. Virginia Tech has built a wonderful digital collection. However, with

size comes complexity, and navigating that huge wonderful complex collection is a

challenge if you are trying to do it without assistance.

References:

Gopher protocol article, http://en.wikipedia.org/wiki/Internet_Gopher, accessed
2/17/2011

M. Goncalves, B. Moreira, E. Fox and L. Watson, (2004) "What is a Good Digital
Library?" – A Quality Model for Digital Libraries, Department of Computer Science,
Federal University of Minas Gerais, Brazil
http://what.csc.villanova.edu/~cassel/9010DL/F07/qualityIPMWversion2.pdf

NeXt computer article, http://en.wikipedia.org/wiki/NeXT, accessed 2/18/2011

NSF web site about the Digital Libraries and the grant to Virginia Tech University (2008)
http://www.nsf.gov/awardsearch/ accessed 2/18/2011

Virginia Tech University Digital Library and Archive,   http://scholar.lib.vt.edu/index.html

accessed 2/14-20/2011