Brian Coté

ILS655 Assignment-Digital Library Project Review Paper 02(02)

02-20-2010

PROJECT BACKGROUND

http://nzdl.sadl.uleth.ca/cgi-bin/library

The New Zealand Digital Library Project was created by a team from the Department of Computer Science at the University of Waikato, New Zealand. Its "…aim is to develop the underlying technology for digital libraries and make it available publicly so that others can use it to create their own collections. (New, n.d.)" The project has built a wide variety of software applications to assist digital library construction, with their most well-known creation being Greenstone Digital Library Software. This is "…a suite of software which has the ability to serve digital library collections and build new collections. It provides a new way of organizing information and publishing it on the Internet or on CD-ROM. (About, n.d.)" In addition their own "…web site provides several document collections, including historical documents, humanitarian and development information, computer science technical reports and bibliographies, literary works, and magazines. (New, n.d.)"

The purpose of their program "…is to explore the potential of internet-based digital libraries. [Their] vision is to develop systems that automatically impose structure on anarchic, uncatalogued, distributed repositories of information, thereby providing information consumers with effective tools to locate what they need and to peruse it conveniently and comfortably. (New, n.d.)"

The New Zealand Digital Library has partnered with two organizations to help spread their software throughout the world: Human Info NGO, which "…is a registered charity responsible for the provision of universal low-cost information access through co-operation between UN Agencies, universities and NGOs (New, n.d.)" and the United

Nations Educational, Scientific and Cultural Organization, which disseminates "…educational, scientific and cultural information throughout the world…. (New, n.d.)"

ORGANIZATION

All of the resources on the Library's main web site are organized with Greenstone software. The categories include the Humanitarian and UN Collections, which covers such topics as the Medical and Health Library, the East African Development Library, and the Technical and Vocational Training Library; the Greenstone Document Collection; demonstration collections such as Project Gutenberg, the Arabic Collection, and Kiwi Aircraft Images; and finally, Documented Sample Collections for Greenstone development. These collections, setting aside the Humanitarian and UN Collections, were mainly devised as proofs of concept for Greenstone software.

The collections themselves house text, image, audio, and video files, in a variety of languages: Arabic, Chinese, French, Maori, and Spanish, and English. (About)"

Greenstone software supports robust methods for metadata. It comes preloaded with four specifications: Dublin Core (qualified and unqualified), RFC 1807 (used mainly for technical reports) (RFC, 2011), NZGLS (New Zealand Government Locator Service), and AGLS (Australian Locator Service). Plug-ins are available to expand the software's capabilities. These include XML, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI, DSpace, and METS (Factsheet, 2007). Metadata can be imported automatically or entered manually (Greenstone, 2007). For example, for a text based collection the plug-ins associated with text documents automatically extract Title and Source/Filename metadata from each document (MSWord, n.d.), whereas for an image collection the plug-

in automatically extracts the name of the file containing the image, the width of the image in pixels, the height of the image in pixels, the name of the gif file containing the thumbnail of image, the width of the thumbnail image in pixels, the height of the thumbnail image in pixels, the full pathname specification of the thumbnail image, and the pathname of the image directory in the collection's associated directory. However, the set represented on the New Zealand Digital Library homepage uses only a minimum of metadata tags.

SERVICE FEATURES

The majority of collections on the project's web site are open to the public. There are, however, a few links that would not allow access. For example, the "How to Build a Digital Library" link produces a 500 Service error. While use of the Youth Oral History Collection is limited to patrons of the Hamilton Public Library in Hamilton, New Zealand.

Searching is conducted through Greenstone's search engine, and browsing is typically limited to a small number of categories, typically one to three. For example, the Youth Oral History collection allows browsing by title and by creator. Whereas The Complete Works of Shakespeare Collection allows browsing by title and by phrase. Clicking on the Phrase button opens up a Java applet that allows users to enter in their desired phrase. Finally the Kiwi Aircraft Images collection allows browsing by Title, Subject, and Phrase. The last one being phrases that appear in the accompanying documentation to the images.

Greenstone software is in a constant state of improvement. Recently the project has been working on Greenstone 3. Their website states, "Greenstone 3 is a complete redesign and reimplementation which retains all the advantages of Greenstone 2…for example, it is multilingual, multiplatform, and highly configurable. It incorporates all the features of the existing system, and is backwards compatible…. Written in Java, it is structured as a network of independent modules that communicate using XML: thus it runs in a distributed fashion and can be spread across different servers as necessary. (Factsheet, n.d.)"

TECHNOLOGIES

The Greenstone interface is very user-friendly. The first page of a collection typically instructs the user in the access points available to them. The Music Video collection offers users the options of Title, Artist, and 'Title and Artist' as access points. As well, users can search for Some or All of the words they enter into the search box, or they can browse by Artist or Title. There is also a Help menu that provides users with instructions on various concerns. (Music, n.d.)

Greenstone software is able to run "…on Windows, Unix and Mac OS X. The distribution includes ready-to-use binaries for all versions of Windows, and for Linux and Mac OS X. It also includes complete source code for the system, which can be compiled using Microsoft C++ or gcc. Greenstone works with associated software that is also freely available: the Apache Webserver and PERL. The user interface uses a Web browser: typically Netscape Navigator or Internet Explorer. Greenstone is specifically designed to be highly extensible and customizable. New document and metadata formats

are accommodated by writing ‘plugins’ (in Perl). Analogously, new metadata browsing structures can be implemented by writing ‘classifiers.’ The user interface look-and-feel can be altered using ‘macros’ written in a simple macro language. A Corba protocol allows agents (e.g. in Java) to use all the facilities associated with document collections. Finally, the source code, in C++ and Perl, is available and accessible for modification. (About, n.d.)”

Greenstone also allows various types of interoperability. “It incorporates a server that can serve any collection over the Open Archives Protocol for Metadata Harvesting (OAI-PMH), and Greenstone can harvest documents over OAI-PMH and include them in a collection. Any collection can be exported to METS…and Greenstone can ingest documents in METS form. Any collection can be exported to DSpace ready for DSpace's batch import program, and any DSpace collection can be imported into Greenstone. (Factsheet, 2007)”

EVALUATION

The method of digital library evaluation developed by Sudatta Chowdhury will be used in this paper. It focuses on the following usability factors: 1) interface features, 2) search process, 3) query formulation, 4) search options for text, multimedia, etc., 5) search operator, 6) results manipulation, and 7) help options (Chowdhury, 2006).

As stated in the Technologies section, the standard interface features offered by the project are displayed on the opening page of most collections. Here limited options are offered. However, by clicking on the Preferences link at the top of the page, the user is brought to a new page that enables them to configure their search experience. This is a

somewhat inconvenient approach and somewhat misleading. The Preferences page would be more appropriately titled the Advanced Search Options page. Once on the Preferences page, however, some very useful options are available. The user can select Interface Language (over fifty languages are available), Encoding type for foreign alphabets, query mode (which allows users to toggle between simple and Boolean searches), Word Endings (whether the engine matches partial or entire words), and Returns (how many returns are given and how many are displayed per page). The interface itself is well designed. The color scheme (light green and white) is pleasing. As well the balance between and the placement of elements on the page facilitate searching and browsing.

Options for selecting databases are limited. On the main page of any site there will be a collection of either thumbnail images displaying the name of the collection or text links. Generally cross-searching is not available or practical. There is, however, a supercollection facility, that enables more than one collection to be linked together. However, when this option is used, browsing is eliminated. Greenstone advises using the supercollection facility on very large collections only. This is a significant disadvantage for the software. Greenstone maintains that collections of the size appropriate to the supercollection facility would be too cumbersome to allow adequate browsing. (Bibliography, n.d.)

There are no options for searching specifically for multimedia or text. If a collection contains more than one type of media, it must be searched as a whole. The user cannot specify that they only want to search for video files, for instance.

Individual types of media (text, audio, video, etc) share standard access points. Meaning all text documents will typically be assigned the same types of metadata. However, individual collections can change the number of points depending on their specific needs. Typically these points are explained on the first page of the collection. For example, Project Gutenberg's main page states that users can search for words that appear in the texts and titles, as well as by the creators (Gutenberg, n.d.), whereas in the Kiwi Aircraft Images collection users can search for words in the text, title, and subject, as well as by phrases (Kiwi, n.d.). This is only a sampling of the actual amount of access points that could be searched. As stated above in the Organization of Resources section, the standard metadata produced during the creation of an image collection is much more extensive, covering data such as image size or thumbnail height. On the projects main web page the search process is simplified for users.

Navigation through the return list is a little awkward. For example, in the Youth Oral History collection, when searching for the term "jazz," four documents are returned. Their titles are listed along side a small image of a closed book. By clicking the book, the record (i.e., the book) will open up. This image doesn't tell the user much about the content of the document: whether it contains audiovisual material or only text, for example. At the top of the record's page is a list of the documents contained in the record. By clicking on an image of a page of text the user is brought to a new page in the record. In this case, the page contains an audio clip and a short description of its contents. Back on the record's main page there is another list further down titled Contents. This is a list of links whose titles restate the list of records at the top of the page. Clicking on these links doesn't appear to do anything except change the color of

the links. However, in the right hands searching and browsing can be accomplished very nicely. In contrast, Auburn University Library's Piano Bench Collection (Piano, n.d.) uses more descriptive thumbnails. Instead of the standard 'book' thumbnail, they use a thumbnail image of the actual document from the record. Here the navigation is intuitive and simple.

Printing options are limited to the facilities of the individual browsers. However, in the Piano Bench collection, there is the option of viewing books in high resolution, which can make for very high quality printouts if the user has decent equipment at their disposal.

Finally, as stated earlier, the Help menu is very informative in explaining the operations of the Greenstone interface. It leads users through various concerns, including Search Terms, Query Type, Scope of Queries, Advanced Searching, Language Preferences, and Presentation Preferences.

Overall, the New Zealand Digital Library Project has done a commendable job in putting together free, open source digital library software. Its power lies in its simplicity. However, with such projects as Auburn University Library's Piano Bench Collection, a very advanced presentation can be created. The main points of contention: a slightly misleading Preferences page (i.e., the lack of an Advanced Search option on the first search page), no options for searching for specific types of media, and navigation through the return list are mainly due to the skill of the developer in crafting the presentation of their collections. In well-trained hands the Greenstone software is able to craft a very sophisticated digital library.

References

About the Greenstone Software. (n.d.). Retrieved from http://tinyurl.com/4ppkn9q

Bibliography Supplement. (n.d.). Retrieved from http://www.nzdl.org/cgi-bin/library.cgi?a=p&p=about&c=cltext-e&ct=1&qt=1

Chowdhury, S., Landoni, M., & Gibb, F. (2006). Usability and Impact of Digital Libraries: A Review. *Online Information Review, 30*(6), 656-680.

Factsheet: Greenstone Digital Library Software. (2007) Retrieved from http://www.greenstone.org/factsheet

Greenstone Digital Library Developer's Guide: Chapter 2 Getting the Most out of Your Documents. (2007). Retrieved from http://www.greenstone.org/manuals/gsdl2/en/html/Chapter_getting_the_most_out_of_your_documents.htm

Kiwi Aircraft Images. (n.d.). Retrieved from http://www.nzdl.org/gsdlmod?a=p&p=about&c=aircraft

MSWord and PDF Demonstration. (n.d.). Retrieved from http://www.nzdl.org/gsdlmod?a=p&p=about&c=wrdpdf-e

Music Videos. (n.d.). Retrieved from http://www.nzdl.org/gsdlmod?a=p&p=about&c=musvid

The New Zealand Digital Library. (n.d.). Retrieved from http://nzdl.sadl.uleth.ca/cgi-bin/library

The Piano Bench Collection. (n.d.). Retrieved from http://diglib.auburn.edu/gsdl/cgi-bin/library?site=localhost&a=p&p=about&c=sheetmus&ct=0&l=en&w=utf-8

Project Gutenberg Collection. (n.d.). Retrieved from http://www.nzdl.org/gsdlmod?a=p&p=about&c=gberg

RFC 1807 – A Format for Bibliographic Records. (2011). Retrieved from http://www.faqs.org/rfcs/rfc1807.html