
Digitization of Dissertations and Other Collections of Department Library (DLIS-A.M.U) using Greenstone

Jafar Iqbal

P M Naushad Ali

Abstract

A digital library is a highly organized collection of electronic resources. The digital libraries provide the required documents directly to the users on the screen enabling the user to interact with digital information. The digital libraries enable persons from different sections of the society in accessing different kind of digital information and knowledge resources for different purposes. This paper presents the efforts made by the authors in digitization of database of bibliographic information and abstracts of the dissertations submitted in the department of Library & Information Science and bound volumes of the journals for the perusal of teachers, research scholars and the students.

Keywords: Digitization, Greenstone

1. Introduction

The concept of libraries has been changed with the rapid advancement in information and communication technologies (ICTs). The advent of digital libraries has made an impact on the role of the library and information professionals as well as end users of information. Digital libraries are revolution the information dissemination with respect to access, speed and availability.

A digital library is a highly organized collection of electronic resources. A digital library is a library in which a significant proportion of the resources are available in machine-readable format. The digital libraries provide the required documents directly to the users on the screen enabling the user to interact with digital information. The digital libraries enable persons from different sections of the society in accessing different kind of digital information and knowledge resources for different purposes.

2. Department Library, Department of Library & Information Science

By 2004, the Seminar library of Department of Library and Information Science, Aligarh Muslim University, Aligarh has been developed into a fully automated library with issue and return of the books through Barcode, OPAC, Access to e-journals and CD workstations. The library of the Department consists of a rich collection of about 5000 books and 350 bound journals. The library currently subscribes to 10 journals from abroad and equal number of journals from India. The Seminar library of Department of Library and Information Science has a collection about 538 dissertations submitted from 1972. The Post. Graduate dissertation is an important document submitted to the universities in partial fulfillment of the requirements for the award of the degree of 'Master of Library and Information Science'. Journals are the medium for publishing original contribution in the different field. Journals and dissertations are valuable primary sources, which report original work

in the specific fields. For the project Databases of contents of the two bound volumes of Library Herald published in the years 2003 and 2005 (Vol. 41 & Vol. 43) and ILA Bulletin published in the years 1999-2000 and 2002 (Vol. 35-38). As far as a database of Dissertation is concerned, MLISc Dissertations submitted from 1972- onwards have been included. However full text of the dissertations has been given only those dissertations submitted from 2005 onwards.

3. Greenstone Digital Library Software

Greenstone is a suite of software for building and distributing digital library collections. It is not a digital library but a tool for building digital libraries. It provides a new way of organizing information and publishing it on the Internet in the form of a fully searchable, metadata-driven digital library. It has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium. It is open source, multilingual software, issued under the terms of the GNU General Public License.

3.1. Features of Greenstone Digital Library Software

The features of greenstone digital library software are as follows:

Greenstone is:

- ◆ Widely accessible: Collections are accessed through a standard web browser.
- ◆ Multi-platform: Collections can be served on Windows and Unix, with an external Web server or (for Windows) a built-in one.
- ◆ Metadata-driven: Browsing (and, if desired, searching) indexes are built from metadata. Metadata may be associated with each document or with individual sections within documents. It must be provided explicitly (often in an accompanying XML or spreadsheet file) or derivable automatically from the source documents.
- ◆ Extensible: Plugins can be written to accommodate new document types. Classifiers can be written to create new kinds of browsing indexes based on metadata.
- ◆ Multi-language: Unicode is used throughout and is converted on the fly to an encoding supported by the user's Web browser. Separate indexes can be built for different languages: a plugin allows automatic language identification for multilingual collections.
- ◆ International: The interface is available in multiple languages: new ones are easy to add.
- ◆ Large-scale: Collections containing millions of documents, and up to several gigabytes, have been built. Full-text searching is fast. Compression is used to reduce the size of the indexes and text.
- ◆ Z39.50 compatible: The Z39.50 protocol is supported for accessing external servers and (under development) for presenting Greenstone collections to external clients.

4. Objectives of the Project

Major objectives of the project are:

- ◆ Digitization of the MLISc dissertations submitted in the seminar library;
- ◆ Developing a database of collections in the library using Greenstone software.

- ◆ To promote fast and efficient delivery of information economically to all of the users of the library

5. Scope of the Project

The main motive of the present project is to provide digital database of bibliographic information and abstracts of the dissertations submitted in the department of Library & Information Science and bound volumes of the journals for the perusal of the teachers, research scholars and the students.

Preservation of the documents is a very important aspect because it is often very difficult to replace the original document once deteriorated. After getting digitized the original document, it can be easily kept away from the users since they may use the digital copy. Since the dissertations and bound journals are not available in soft form, the present project has a great importance for this purpose and when the digital database has a link to the Internet it can be access anyplace by many users at one time.

Abstract is a concise form of the document and it provides central idea of a document. The bibliographic information and abstracts give outline of the documents. Thus user can easily understand that a particular document is relevant or not for his/her needs.

Digital documents have several advantages over conventional documents. If it is available online, it can be access at anyplace and anytime by the user and it can be accessed by a number of users simultaneously. Users can conduct searches by title, author, content or full text searches or combination of these, i.e. multiseaches using Boolean logic.

6. Methodology

For this project, the dissertations submitted from 1972 onwards and two bound journals such as Library Herald (V-41 and V-43) and ILA Bulletin (V-38 and V-35) have been selected for the digitization. Then, bibliographic information along with abstracts has been prepared for the dissertations and bound journals except dissertations submitted in 2005.

After that metadata for the dissertations and bound journals have been created and then bibliographic information and abstracts of the dissertations and bound journals have been uploaded on the greenstone. Full text of the dissertations submitted from 2005 onwards also been given on the greenstone.

7. Digitization of Dissertations and Other Collections of Department Library Using Greenstone

7.1. Selection of the problem

The problem for the present project is entitled "Digitization of Dissertations and Other Collections of Department Library Using Greenstone".

7.2. Selection of the dissertations and journals

For this project the dissertations from 1972 onwards have been selected for the digitization. Due to the limitation of the time among bound volumes of journals, only two bound journals such as Library Herald (V-43 and V-41) and ILA Bulletin (V-38 and V-35) have been selected for the digitization.

7.3. Bibliographic information and abstracts of the dissertations

In the third step, the bibliographic information along with abstracts of the dissertations have been prepared.

7.4. Bibliographic information and abstracts of the bound journals

In the fourth step, the bibliographic information along with abstracts of the bound journals have been prepared.

7.5. Creation of Metadata

Metadata is another issue central to the development of digital libraries. Metadata is the data to describe the content and attributes of any particular item in a digital library. Metadata is important in digital libraries because it is the key to resource discovery and use of any document.

Assignment of metadata or in simple terms cataloguing of the digital collection is one of the most important steps in developing a digital library. The information seeking behaviour of different users may be different. Therefore it is necessary that the metadata be assigned keeping all the possibilities in mind.

There are three categories of metadata that are frequently used to describe objects in a digital library. These are:

- 1. Descriptive** - Information describing the intellectual content of the object, such as MARC cataloguing records, finding aids or similar schemes. It is typically used for bibliographic purposes and for search and retrieval.
- 2. Structural** - Information that ties each object to others to make up logical units (e.g., information that relates individual images of pages from a book to the others that make up the book).
- 3. Administrative** - Information used to manage the object or control access to it. This may include information on how it was scanned, its storage format, copyright and licensing information, and information necessary for the long-term preservation of the digital objects

In this step, metadata for dissertations and bound journals have been prepared. In Greenstone, There are already a number of metadata sets are available such as Dublin Core. But they do not fulfill the requirements because some elements are not available in the existing metadata sets.

So, two different metadata sets have been created, one for the dissertations and other for the bound volumes of the journals.

7.5.1. Metadata Sets

Two different metadata sets have been created, DLIS Dissertation and DLIS Bound Journal.

DLIS Dissertation

In this metadata set, various elements have been created which are as follows:

1. Title: The name given to the dissertation by the author.
2. Supervisor: The person under whom supervision dissertation work had been done.
3. Author: The person responsible for the dissertation work.
4. Accession: A unique no. given to the dissertation by the seminar library.
5. Year: The year the dissertation was made available in its present form.
6. Location: The place where dissertation is available.
7. Study: The category of the study of dissertation, such as bibliometric study, bibliography, or survey.

DLIS Bound Journal

In this metadata set, various elements have been created which are as follows:

1. Title: The name given to the article by the author.
2. Author: The person responsible for the article.
3. Journal: The name of the journal.
4. Publisher: The publisher of the journal.
5. ISSN: ISSN no. of the journal.
6. Year: The year the journal was made available in its present form.
7. Volume: The volume no. of the journal.
8. Issue: The issue no. of the journal.
9. Place: The place from where journal is to be published.
10. Keywords: The keywords of the article.
11. Accession: A unique no. given to the journal by seminar library.

7.6. Development of the collection

In this step the bibliographic information along with abstracts of dissertations and bound journals have been uploaded on the Greenstone. The simplest way to build new collections is to use greenstone's "librarian" interface (GLI). This allows user to collect sets of documents, import or assign metadata, and build them into a greenstone collection.

7.6.1. The Librarian Interface

It supports five basic activities, which can be interleaved but are nominally undertaken in this order.

- ◆ Copy documents from the computer's file space, including existing collections, into the new collection. Any existing metadata remains "attached" to these documents. Documents may also be gathered from the web through a built-in mirroring facility.
- ◆ Enrich the documents by adding further metadata to individual documents or groups of documents.
- ◆ Design the collection by determining its appearance and the access facilities that it will support.
- ◆ Build the collection using Greenstone.
- ◆ Preview the newly created collection, which will have been installed on user Greenstone home page as one of the regular collections.

To convey the operation of Greenstone's librarian interface, let's work through a simple example. Figures 1 to 10 are screen snapshots at various points during the interaction. This example uses documents in the DLIS Bound Journals.

7.6.2 Getting started

Launch the librarian interface under windows by selecting Greenstone digital library from the programs section of the start menu and choosing librarian interface. To begin a user must either open an existing collection or start a new one. (Figure 1) shows the user in the process of starting a new collection. User has selected new from the file menu and begun to file out general information about the collection—its title, the E-mail address of the person responsible for it, and a brief description of the content—in the popup window. The collection title is a short phrase used throughout the digital library to identify the collection's content. When user types the title, the system assigns a unique mnemonic identifier, the collection "name", for internal use. The E-mail address specifies the first point of contact for any problems encountered with the collection.

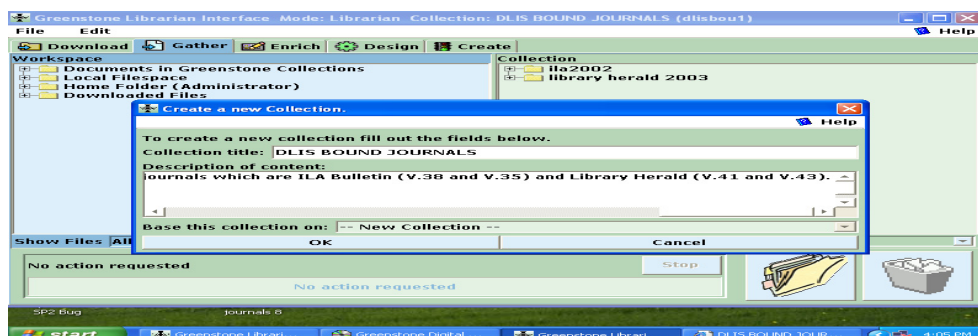


Figure 1. Starting a new collection

The brief description is a statement describing the principles that govern what is included in the collection. It appears under the heading About this collection on the collector's initial page.

At this point, the user decides whether to base the new collection on the same structure as an existing collection, or to build an entirely new kind of collection. In (Figure 1) user has chosen to base it on the DLIS Bound Journals (Figure 2) subset collection.

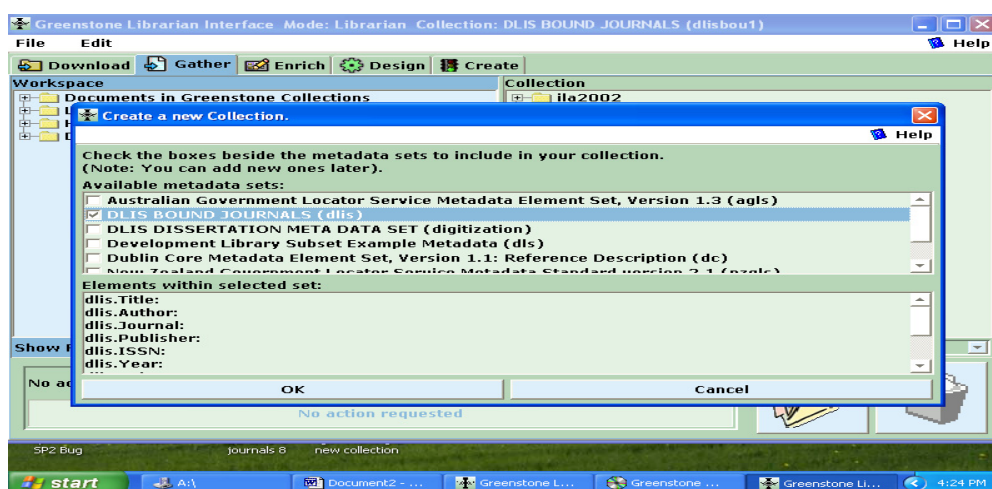


Figure 2. Selection of metadata for new collection

This implies that the "DLIS Bound Journals" metadata set (Figure 2) which is used in this collection will be sued for the new collection.

If, instead, the user had chosen "new collection" at this point, user would be asked to select what metadata sets be used in the new collection. Three standard sets are pre-supplied: Dublin core, the DLS metadata set, and a set that comprises metadata elements extracted automatically by Greenstone from the document in the collection. The user can also create new metadata sets using "Greenstone Editor for Metadata Set" menu.

7.6.3. Assembling the source material

After clicking the OK button on the "new collection" popup, the remaining part of the interface, which was grayed out before, becomes active. The gather panel, selected by the eponymous tab of (Figure 1), is displayed initially. This allows the user to explore the local file space and existing collections, gathering up select documents for the new collection. The panel is divided into two sections, the left for browsing existing structures and the right for the documents in the collection.

Care is taken to deal appropriately with name clashes when files of the same name in different parts of the computer's directory structure are copied into the some folder of the collection.

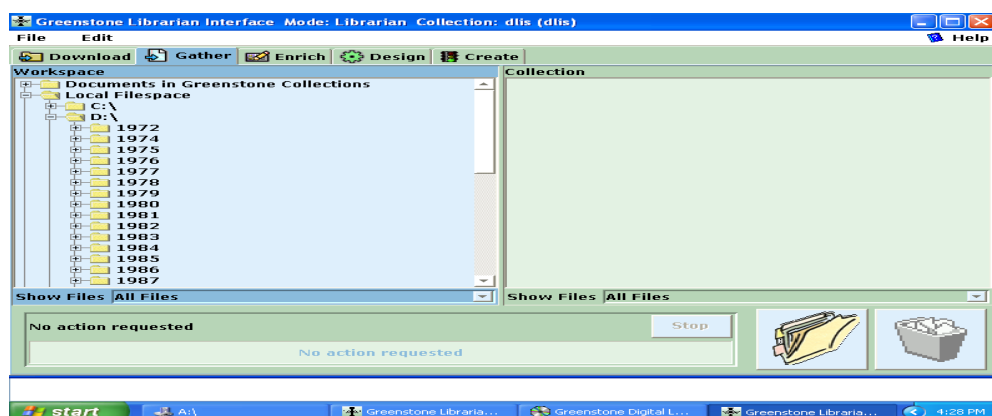


Figure 3. Exploring the local file space

In (Figure 3) the user is using the interactive file tree display to explore the local file system. At this stage, the collection on the right is empty; user populates it by dragging and dropping files of interest from the left to the right panel. Such files are "copied" rather than "move": so as not to disturb the original file system. The usual techniques for multiple selection, dragging and dropping structuring the new collection by creating subdirectories ("folders"), and deleting files from it by moving them to a trashcan, are all available.

When large file sets are selected, dragged, and dropped into the new collection, the copying operation may take some time---particularly if metadata conversion is involved. To involved progress, the interface shows which file is being copied and what percentage of files has been processed.

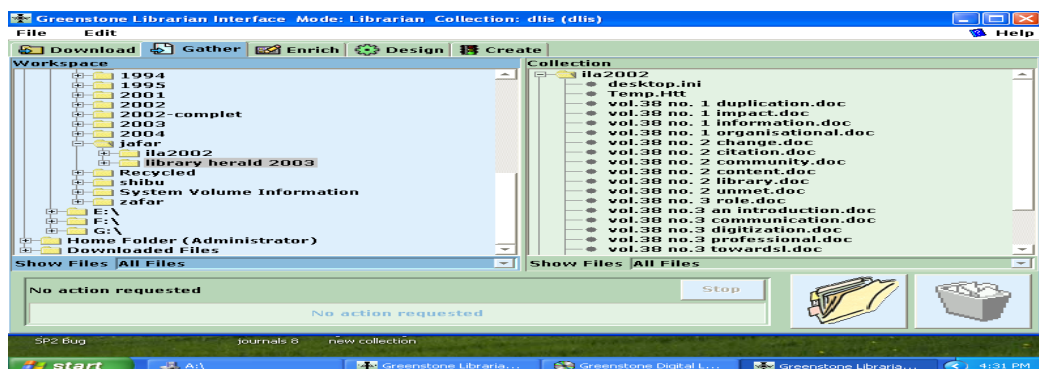


Figure 4. Filtering the file trees

In (Figure 4), only the Word files are being shown.

7.6.4. Enriching the documents

The next phase in collection building is to enrich the documents by adding metadata. The enrich tab brings up a new panel of information (**Figure 5**), which shows the document tree representing the collection on the left and on the right allows metadata to be added to individual documents, or groups of documents.

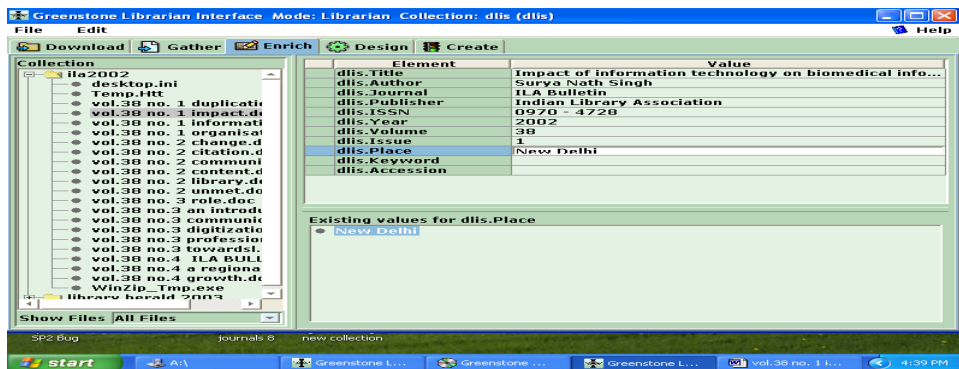


Figure 5. Assigning metadata using Enrich view

Document that are copied during the first step come with any applicable metadata attached. The enrich panel allows metadata values to be assigned to document in the collection.

In (Figure 5) the user has selected the folder vol.38 no.1 impact and assigned "New Delhi" as its Place metadata. The button of updating and removing metadata become active depending on what selections have been made.

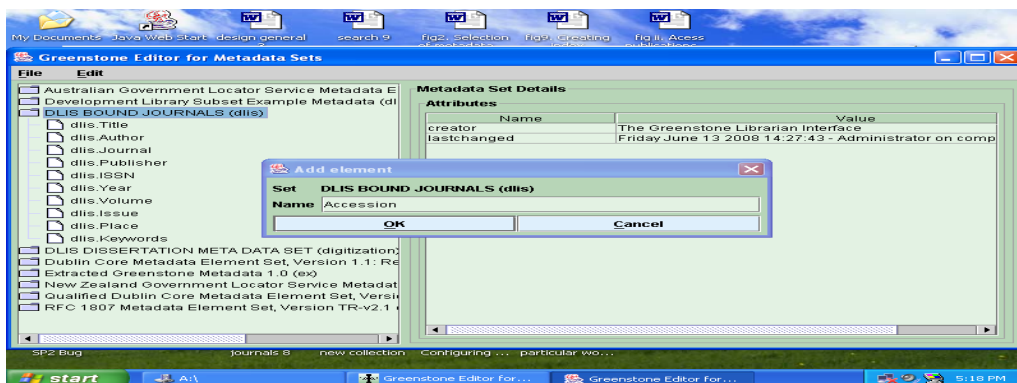


Figure 6. Editing the metadata set

The panel in (Figure 6) allows the user to edit metadata sets. Here, the user is looking at the Accession element of the DLIS Bound Journals set.

7.6.5. Designing the collection

The design panel (Figures 7) allows user to specify the structure, organization, and presentation of the collection being created. The result of this process is recorded in a "collection configuration file," which is Greenstone's way of expressing the facilities that a collection requires. This step involves a series of separate interaction screens, each dealing with one aspect of the collection design.

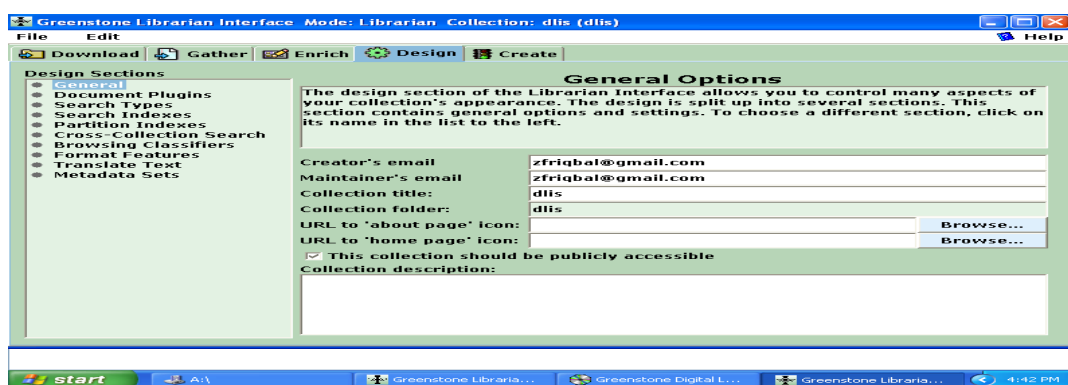


Figure 7. Designing the collection

In (Figure 7) the user has clicked the design tab and is reviewing the general information about the collection, entered when the new collation was created. On the left are listed the various facets that the user can configure: general, document plug-ins, search types, search indexes, partition indexes, cross-metadata sets.



Figure 8. Creating Index for new collection

In (Figure 8) Indexes are assigned to the new collection.

7.6.6. Building the collection

The create panel (Figure 9) is used to construct a collection based on the documents and assigned metadata. The brunt of this work is borne by the Greenstone code itself. The user controls this external process through a series of separate interaction screens, each dealing with the arguments provided to a certain stage of the creation process. The user observes the building process through a window that shows not only the text output generated by Greenstone's importing and index-building scripts, but also progress bars that indicate the overall degree of complete of each script.

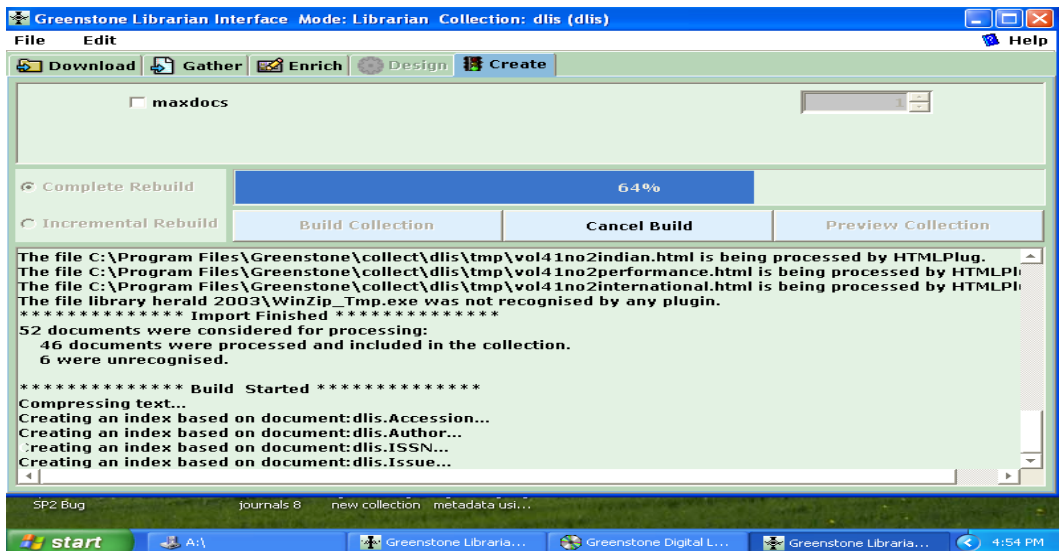


Figure 9. Getting ready to create new collection

(Figure 9) shows the create view. At the top are shown some options that can be applied during the creation process. The user selects appropriate value for the options. This figure illustrates a popup "tool tip" that is available throughout the interface to explain the function of each argument. When satisfied with the arguments, the user clicks Build Collection.

7.6.7. Previewing

The preview collection button (Figure 9) is used to view the collection that has been built. Clicking this button launches a web browser showing the home page of the collection (Figure 10).

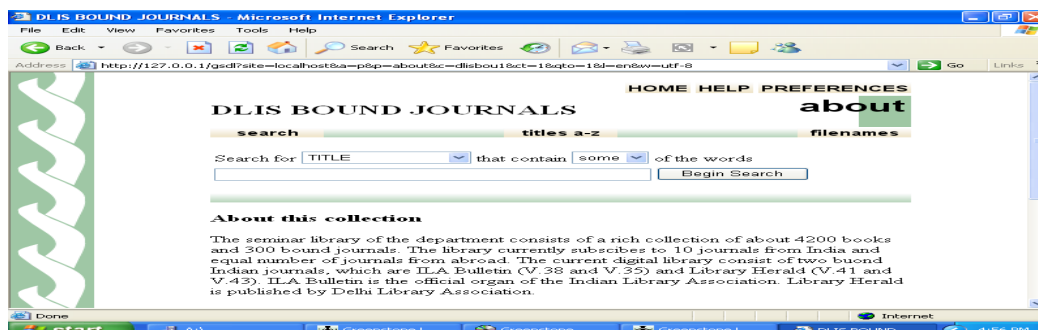


Figure 10. Previewing the newly built collection

In practice, previewing often shows up deficiencies in the collection design, or in the individual metadata values, and the user frequently returns to earlier stages to correct these. This button becomes active once the collection has been created the newly created collection will also have been installed on Greenstone home page as one of the regular collection.

7. Searching Methods for access Digital Library

The digital database of the bibliographic information along with abstracts of the dissertations (1972-1987) and bound journals (ILA Bulletin and Library Herald) have been made available for searching to the users. It can be access by various searching techniques, which are as follows:

8.1. Finding Information

Greenstone digital library systems usually comprise several separate collections-for example, computer science technical reports, literary works, Internet FAQ's, magazines. There will be a home page for the digital library system, which allows user to access any publicly accessible collection; in addition, each collection has its own "about" page that gives information about how the collection is organized and the principles governing what is included in it. (Figure 11) shows a screenshot of the "DLIS Bound Journals" collection supplied with the Greenstone software; let's use it as an example to describe the different ways of finding information.

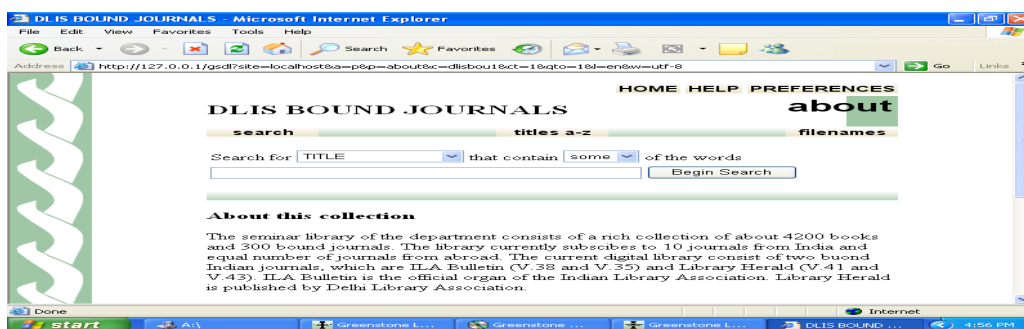


Figure 11. Using DLIS Bound Journals

The “search ... titles a-z ... filenames” bar underneath gives access to the searching and browsing facilities. The leftmost button is for searching, and the once to the right of it-four, in this collection-evoked different browsing facilities.

8.2. How to find information

A user can search for particular words (Figure 12) that appear in the text from the “search” page. The search page can be reached from other pages by pressing the search button. User can access publications by title by pressing the title a-z button (Figure 13).



Figure 12. Search by a particular word



Figure 13. Access collections by title

This brings up a list of books in alphabetic order. User can access publications by organization by pressing the organization button. This brings up a list of organizations.

9. Limitation of the Project

The major limitations of the present project are as follows:

- ◆ It provides only bibliographic information and abstracts of the dissertations from 1972 onwards; however it provides full text of the dissertations submitted from 2005 onwards.

-
- ◆ It provides database of contents of only two volumes of bound journals, Library Herald (V-43 and V-41) and ILA Bulletin (V-38 and V-35).

10. Conclusion

Libraries of twenty first century extend beyond the doors of one building. These will provide immediate access to a variety of print and electronic information resources when the users need them. The library of the future will be a place where students, faculty, researchers, scholars and decision makers will access the vast information resources of the world. The digital libraries assist scholars to what relevant information exists, anywhere in the world in a multimedia format.

The present project "Digitization of Dissertations and Other Collections of Department Library Using Greenstone" is a practical attempt to digitize the bibliographic information along with abstracts of the dissertations and some bound volume of the journals. The digital database also covers the full text of the collection of dissertations submitted from the year 2005 onwards. The digital database of dissertations and bound journals will be much benefited to students, research scholars and teachers of the department for their research, teaching and learning processes.

References

1. Jafar Iqbal. Digitization of the dissertations (1972-1987) and bound journals of seminar library using greenstone. Unpublished Dissertation. Deptt. of Library and Information Science, AMU, Aligarh, 2006.
2. Krishnamurthy, M. Digital library using greenstone: a study. SRELS Journal of Information Management, 2003, 40 (4), p. 313-320.
3. Krishnamurthy, M. Digital library of mathematics using dspace: a practical experience. SRELS Journal of Information Management, 2005, 42 (3), p. 245-256.
4. Greenstone, <http://www.greenstone.org/> (Accessed on 10/01/2008).

About Authors

Mr. Jafar Iqbal, Research Scholar, Department of Library & Information Science, Aligarh Muslim University, Aligarh-202002
Email: zfriqbal@gmail.com,

Dr. P M Naushad Ali, Reader, Department of Library & Information Science, Aligarh Muslim University, Aligarh-202002
E-mail : naushadali.amu@gmail.com