‘INSTITUTIONAL REPOSITORY’

1.1 Introduction

A digital/online archive where a university or college community’s intellectual work is made accessible and preserved for posterity is Institutional Repository (IR). The concept of IR suggests the tantalizing possibility of greater library influence over the full cycle of scholarly communication on campus, from research through publication, collection, and preservation. Libraries are performing lead role in shaping institutional digital repositories all over the world.

1.2 Objectives

The institutional repositories are the digital collections meant for capturing and preserving the intellectual output of a single or multi-faculty institution or a university. Each institution defines its own policies dealing with access to and use of materials in repositories. Not all materials can be made available freely. Copyrighted materials may carry a variety of restrictions. Non-exclusive publisher licenses would increase availability to these materials and place the publishers in the open access arena. Major objectives for creating Institutional Repository were:

- Utilization of Digital Library Software.
- Creation of Institutional Repository of G.S.College, Nagpur
- To make in-house accessibility to the literature.
- To digitize institution's academic and co curricular reports, events, etc.
- To make old syllabus and question papers accessible to student community in digital format.
- To store learning materials
- To preserve digital materials for the long term

1.3 Type of contents included

- Published, Peer-reviewed literature by faculty.
- Learning Materials
- Institutional Records
- Conference Proceedings
- Old/New Syllabus and
- Old/New Question Papers
- Photographs of the activities held at institution.

1.4 Relevance of the work

The building of an Institutional Repository for any organization is needed in the present scenario of digital world because of the following certain changes:

- Rapid changes in the technology;
- Significant increase in the overall volume of research;
- Change in government policies in Higher education.
- Increasing need of archival and access to unpublished information bearing objects;
- Increasing demand to access knowledge objects from anywhere at anytime;

1.5 Methodology

Detailed Project Planning Steps

- Developing a Service Definition;
- Identifying literature to be digitized;
- Scanning of documents/converting to GREENSTONE DIGITAL LIBRARY SOFTWARE readable format.
- Installation of hardware and software systems;
- Batch load existing collections;

- Launching a Service;
- Orientation for users;
- Running a Service;
- Long term tasks – running the system, growing the service, etc.

1.6 Preparation before installation

For the present work a branded desktop, scanner with necessary softwares has been acquired and installed at the work place.

1.7 The Infrastructure

The institution library has provided with internet connection. The internet access is essentially of great help for the present work.

1.8 Benefits

i. For the contributor

- **Greater citation**: Articles freely available on the Internet are cited more often than their paper counterparts.
- **Google Scholar** gives preferential treatment to materials in IRs; a paper picked up from an IR would appear higher up on the Google results list.

- **Speed:** Faculty members can self-publish their preprints immediately, with the possibility of receiving immediate feedback.
- **Organization:** It contain all of the scholarly work by one faculty member, including pre-prints, post -prints, presentations, classroom materials. Instead of being scattered about in different databases, servers, or computer hard drives, this material can be browsed easily in one place by the user, and reused easily by the contributor.
- **Preservation:** Depositing a file into an institutional repository means that the burden of ensuring the file can be opened is placed on the curator of the institutional repository, and not on the owner.
- **Ease of use**: Although self-submission is possible in our institutional repository, it's much more likely that all uploading will be done by the library. All that is needed are files to upload and permission to upload it.
- **Permanent place**: Depositing an item into an institutional repository means that it stays in one place and maintains the same URL.
- **Feedback and commentary** from users. Authors are able to receive and respond to commentary on 'pre-prints'.

- Also more sponsors of funded research now have mandates for authors to deposit their articles and other research outputs as a condition for funding. Some policies promote Open Access for funded research. These requirements are intended to increase readership, re-use and dissemination of research outputs. The message to researchers is that research is incomplete until the output is widely disseminated.

ii. For the institution

- The scholarly material produced by the institution is available in one place, reflecting the intellectual achievements of the institution, and serves as a valuable marketing tool.
- Increased visibility and prestige.
- Documents reflecting the institutional history of the university, both scholarly and non-scholarly, are preserved for future use, much like a traditional archive preserves paper material.
- Material that is not traditionally published is included in the repository, including drafts of unpublished articles or book chapters, unpublished research, student works, learning objects, and creative works.

- Breaking down of publishers' costs and permissions barriers.

iii. For the user

- Material in an institutional repository can be found through a search engine.
- There is no charge to access this material, and there are no subscription fees.
- Repository contains material that is best displayed in its original digital format, such as audio files, video files, animations, and data sets.
- Gray literature, material not easily found through conventional means, will be actively recruited for the repository.

iv. Individual Benefits

- Wider distribution
- Showcase
- Safekeeping
- Lower technology barrier
- Time

v. Other Benefits

1. Increased visibility to the Library

2. Complete customization of policies and user

3. Responsiveness to local user needs and

4. Increased contact with constituents

5. Showcase and preserve scholarly output and historic documents

6. Archive post-prints, preprints

7. Support teaching and learning

8. Provide curatorial stewardship for disorganized and scattered digital materials

9 No need of maintaining server or back up.

1.9 Software for Institutional Repositories

There are various types of Digital Library software are available e.g.,

1. DSpace

2. Greenstone Digital Library Software

3. eprint Archive,

4. Fedora

5. AGES Software,

GREENSTONE DIGITAL LIBRARY SOFTWARE

2.1 About Greenstone Digital Library Software:

Greenstone 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. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source software, available from http://greenstone.org under the terms of the GNU General Public License.

Greenstone runs on different platforms and different configurations. The distribution includes ready-to-use binaries for all versions of Windows, and for Linux. It also includes complete source code for the system, which can be compiled using Microsoft C++ or gcc. Greenstone works with associated software, which is also freely available: the Apache Web server and PERL. The user interface uses a Web browser: typically Netscape Navigator or Internet Explorer (version 4.0 or above in both cases).

Greenstone is based upon a search engine called MG. All search engines turn words into numbers for speed and in the case of MG these numbers

are also used to improve the comparison of the index which means that in turn the collection takes up much less space on the hard drive.

Greenstone provides two separate windows binary programs of the CD-ROM: the local Library and Web Library.

2.2 Why Greenstone Digital Library Software?

As Greenstone Digital Library Software installation is very easy and stores all types of data like Ph.D theses, faculty publications, lecture notes, student's dissertations, learning objects, PG level & NET/ SET question papers, links to open knowledge objects, project reports, gray literature, unpublished theses, necessary photographs etc. successfully and enables the upload from every terminal with fantastic user interface so the researcher has selected Greenstone Digital Library Software for digitization project.

i. Local Library

It offers a complete, self-contained, web-serving capability. The Local library is intended for use on standalone computers or that do not already have web server software. It contains a small built-in web server so that other computers on the same network can also access the library.

ii. Web Library

This enables any computer with an existing web server to serve pre-built Greenstone collection with small changes in the configuration of the server setup.

Greenstone is internally separated into two components: “the collection server” which provides services on one side, and a “receptionist” which access the services through an interface. This has made it particularly adaptable to support both traditional web-based access and rich graphical environments from one-server program.

2.3 Workflow in GSDL SOFTWARE

The Collector

The structure of each collection is determined at set up. This includes specifying the format (or formats) of source documents, deciding how to display the documents on the screen, determining what the source of metadata will be, choosing what full-text searching and browsing facilities should be provided, and outlining how the search and browsing results should be displayed. Once a collection is in place, new documents in the same format can be added automatically.

The Greenstone "Collector" is an interactive subsystem for managing and accessing collections. The Collector can be used to:

- create a new collection with structure existing
- create a new collection with a different structure;
- add new material to an existing collection;
- modify the structure of an existing collection delete a collection;
- publish a CD-ROM of an existing collection

Collections can also be built through command mode. The collection building process is not feasible but to bring radical and effective changes, to create collections with completely new structures, the command mode is used.

Creating a New Collection

On logging on to the Collector, it displays a sequence of steps involved in collection building. They are:

1. Collection information

It specifies the collection's name to identify the collection. The email address is used for diagnostic reports in case any problems arise with the collection. A few lines are entered under *About this collection*.

2. Source data

It defines where the source data will come from. The collection can be either completely new or a “clone" of an existing one. Which can be selected from a pull-down menu?

Boxes are provided to indicate where the source documents are located. Any number of input sources can be specified. Specifications can be:

- a directory name on the Greenstone server system (beginning with "file://")
- an address beginning with "http://" for files to be downloaded from the Web an address beginning with "ftp://" for files to be downloaded using FTP.

In each case of "file://" or "ftp://" the collection will include all files in the specified directory, any directories it contains, any files and directories they contain, and so on. If a filename is specified, that file alone is included. For "http://" the collection will mirror the specified Web site.

3. Configuring the Collection

It tailors the configuration options. The construction and presentation of all collections is controlled by specifications in a configuration file. Depending on the collection clone selected, the configuration file will be different for different collections. We can add assign metadata to

classify/index a collection by adding the required lines. Plugins can be added depending on the format of the documents in the collection. Many other configuration settings can be implemented. The path of collection icons can be specified.

4. Building the Collection

The system makes all the indexes and gathers all information required to make the collection operate.

First, an internal name is chosen for the collection, based on the title that has been supplied. Then a directory structure is created that includes subdirectories to receive, index and present the source documents. A recursive file system copy command is issued to retrieve source documents already on the file system; for offsite files a web mirroring package is used to copy the specified site along with any related image files. Next, the documents are converted into a standard XML form. Appropriate plugins to perform this operation must be specified in the collection configuration file. Then, the copied files are deleted: the collection can always be rebuilt from the information stored in the XML files.

Then, the full-text searching indexes and browsing structures specified in the collection configuration file are created. Finally, the result of the building process is moved to the area for active collections. This

precaution ensures that if a version of this collection already exists, it continues to be served right up until the new one is ready. The software assigns a global, persistent identifier to each document to ensure that the changeover is almost always invisible to users.

The building stage is potentially time-consuming and it depends on the size and file format of the file in the collection.

Warnings are issued if any of the following occur:

- non-existent input files or URLs are requested,
- there is no plugin that can process a file, or
- associated files -- such as images embedded in html documents -- are missing.

5. Viewing the Collection

The new collection is built and installed to be viewed.

2.4 Working with existing collections

Four additional facilities are provided when working with existing collections: adding new material, modifying the collection structure, deleting the collection, and printing it on a CD-ROM.

Add New Data - New data can be added to an existing collection. It gets copied and converted to XML, joining any existing imported material.

Edit the Collection Configuration - the structure of existing collections can modify by editing their configuration file.

Delete Collection – A collection can be selected and deleted after confirmation. Only collections built with the collector can be removed other than collections created through command line.

Export Collection - to write an existing collection to a CD-ROM, select the collection and it is automatically massaged into a disk image in a standard directory using a standard CD-writing utility. Upto 150,000 pages can be indexed on one CD. Every CD in turn can become an Internet Server, a self-installing Greenstone CD-ROM for Windows. The exported collection directory contains four files related to the installation process and three subdirectories that contain the complete collection and software.

2.5 Document Types

Source documents come in a variety of formats, and are converted into a standard XML form for indexing by "plugins." Plugins distributed with Greenstone process plain text, HTML, WORD and PDF documents, and Usenet and E-mail messages. Greenstone generally uses the filename to determine document format. “GML” is the name of the internal XML document format of the source files after digitizing.

Collections can contain text, pictures, audio and video. Non-textual material is either linked into the textual documents or accompanied by

textual descriptions (such as figure captions) to allow full-text searching and browsing. Compression technology is used throughout to ensure best use of storage.

Plugins are specified in the collection configuration file. Some of them are as follows:

TEXTPlug to interpret a plain text file as a simple document and adds title metadata based on the first line of the file.

HTMLPlug It processes HTML files. It extracts metadata based on the title tag and has many other options.

WORDPlug It imports Word documents. Greenstone uses the program wvWare to convert Word files to HTML. It does not work with RTF documents.

PDFPlug It imports documents in PDF. It uses pdf to html program to convert PDF files to HTML.

EMAILPlug It imports files containing E-mail and deals with common E-mail formats such as used by Netscape, Eudora and Unix mail readers. The plugin extracts a Subject, To, From and Date metadata.

ZIPPlug It handles compressed and archived input formats. It is disabled on Windows.

RecPlug It expands subdirectories and pours their contents into the plugin list, thereby traversing arbitrary directory hierarchies.

GMLPlug - It processes previously imported documents.

2.6 Administration

An administrative facility is included with Greenstone. It presents configuration information about the installation and allows it to be modified. It facilitates examination of error logs that record internal errors, and the user logs that record usage. It enables a specified user i.e. administrator to authorize others to build collections and add new material to existing ones. All collections are listed here for there may be private collections also which are not accessible through the Greenstone Home page.

Configuration Files

There are two configuration files that control Greenstone's operation

- Site configuration file, Greenstone Digital Library Softwaresite.cfg
- Main configuration file, main.cfg

The Greenstone Digital Library Softwaresite.cfg file is used to configure the Greenstone software for the site where it is installed. It is designed for keeping configuration options that are particular to a given site. Examples include the name of the directory where the Greenstone software is kept,

the HTTP address of the Greenstone system, and whether the fast cgi facility is being used.

The main.cfg file contains information that is common to the interface of all collections served by a Greenstone site. It includes the E-mail address of the system maintainer, whether the status and collector pages are enabled, whether logs of user activity are kept, and whether Internet "cookies" are used to identify users.

Logs

Greenstone generates three kinds of logs.

- Usage Log
- Error Log
- Initialization Log

All user activity-every page that a user visits is recorded. Logging should be enabled in the main system configuration file. The log cgiargs line turns logging on and off. By activating usecookies a unique identification code is assigned to each user, which enables individual user's interaction to be traced through the log file.

Each line in the user log records, IP address of the user's computer, the timestamp, CGI arguments and the name of the user's browser. The log file usage.txt is placed in the etc directory in the Greenstone file structure.

User Management

Greenstone incorporates an authentication scheme which is used to control access to certain facilities like the Collector and administrative functions. Authentication is done by requesting user name and password. Users having administrative privileges can add new users. Each user can belong to different groups. By default there are two groups, admin and colbuilder. An admin user can create new users and passwords for the colbuilder group. The colbuilder group can only build collections.

Search/Browse Features

Greenstone uses MG to index and retrieve documents. Information collections built by Greenstone combine full-text search with browsing indexes based on different metadata types. There are several ways for users to find information, although they differ between collections depending on the metadata available and the collection design.

The default search interface is simple. Advanced searching allows Boolean expressions, phrase searching and case and stemming control. These can be enabled from the Preferences page.

Searching is full-text, and-depending on the collection's design-the user can choose between indexes built from different parts of the documents, or from different metadata. Some collections have an index of full documents, an index of sections, an index of paragraphs, an index of

titles, and an index of section headings, each of which can be searched for particular words or phrases.

Browsing involves data structures created from metadata that the user can examine: lists of authors, lists of titles, lists of dates, hierarchical classification structures, and so on. Data structures for both browsing and searching are built according to instructions in a configuration file, which controls both building and serving the collection.

Structures for both searching and browsing are specified by instructions in the configuration file, and can be rebuilt entirely automatically.

Each document can be hierarchically organized into logical sections, each of which comprises paragraphs. Metadata such as author, title, date, keywords, may be associated with documents, or with individual sections. This is the raw material for indexes. It must either be provided explicitly (for example, in an accompanying spreadsheet) or be derived automatically from the source documents. Metadata is stored with the document for internal use.

The various search options accessible are

- Search for particular words
- Access publications by subject
- Access publications by title
- Access publications by organization

- Access publications by "how to" listing etc..

A document within a collection can be detached to open in a new browser window. If the document is reached through a search, then the search terms are highlighted. The highlighting button can be made on or off. Apart from this, text and contents can be expanded in documents having a hierarchical structure.

The Collections could be searched for particular words, subjects (based on Universal Decimal Classification,Dewey Classification, Library of Congress classifications,etc), Organization, Titles, Keywords, Topics (author, type of publication), publication-dates, Publication-numbers-or-codes, Countries etc. Punctuation in between search terms are ignored.

There are two different kinds of query.

- Queries for all the words. These look for documents (or chapters, or titles) that contain all the words you have specified. Documents that satisfy the query are displayed.
- Queries for some of the words. Just list some terms that are likely to appear in the documents you are looking for. Documents are displayed in order of how closely they match the query. When determining the degree of match,

- more search terms a document contains, the closer it matches;
- rare terms are more important than common ones;
- short documents match better than long ones.

Advanced Search Features

These are accessible from the Preferences page.

Case sensitivity and stemming

When you specify search terms, you can choose whether upper and lower case must match between the query and the document: this is called "case sensitivity." You can also choose whether to ignore word endings or not: this is called "stemming." Generally case differences and word endings should be ignored unless you are querying for particular names or acronyms.

Phrase searching

If your query includes a phrase in quotation marks, only documents containing that phrase, exactly as typed, will be returned. Phrases are processed by a post-retrieval scan.

Advanced query mode

It can be selected on the Preferences page,the queries for all of the words, described above, are actually Boolean queries. They consist of a list of

terms joined by logical operators & (and), | (or), and ! (not). Absent operators between search terms are interpreted as & (and): thus a query without any operators returns documents that match all the terms.

If the words AND, OR, and NOT appear in your query they are treated as ordinary search terms, not operators. For operators you must use &, |, and !. In addition, parentheses can be used for grouping.

Using Search History

This feature on the Preferences page will show the last few searches, along with a summary of how many results they generated.

Collection Preferences

Some collections comprise several subcollections, which can be searched independently or together, as one unit. If so, one can select which subcollections to include in searches on the Preferences page.

Language Preferences

Each collection has a default presentation language, but you can switch to a different language if you like. You can also alter the encoding scheme used by Greenstone for output to the browser

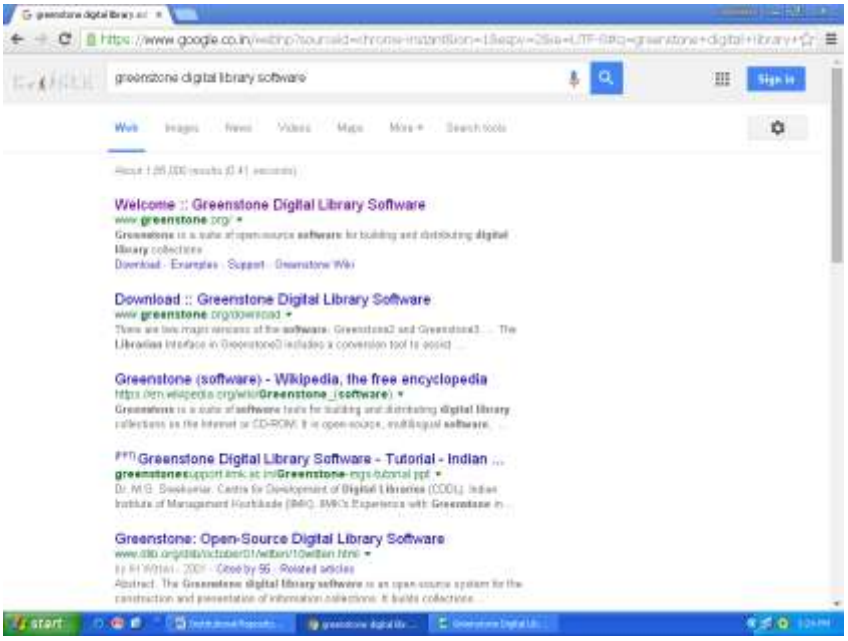
Presentation Preferences

Depending on the collection, one can set options to control the presentation.

GREENSTONE DIGITAL LIBRARY SOFTWARE

INSTALLATION

GREENSTONE DIGITAL LIBRARY SOFTWARE is available free on its website. If we search for the term Greenstone Digital Library Software following screen appears:



If we select www.greenstone.org following website is displayed.

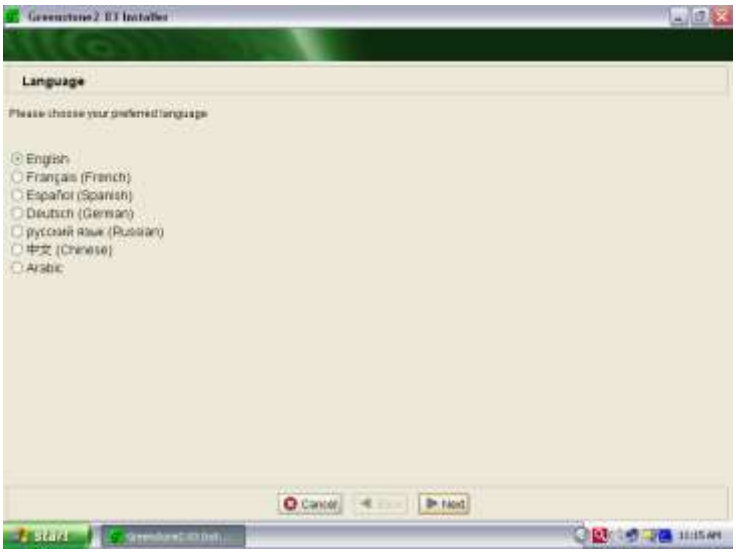
www.greenstone.org

Steps for Installation of Software

The following steps are required for installation:

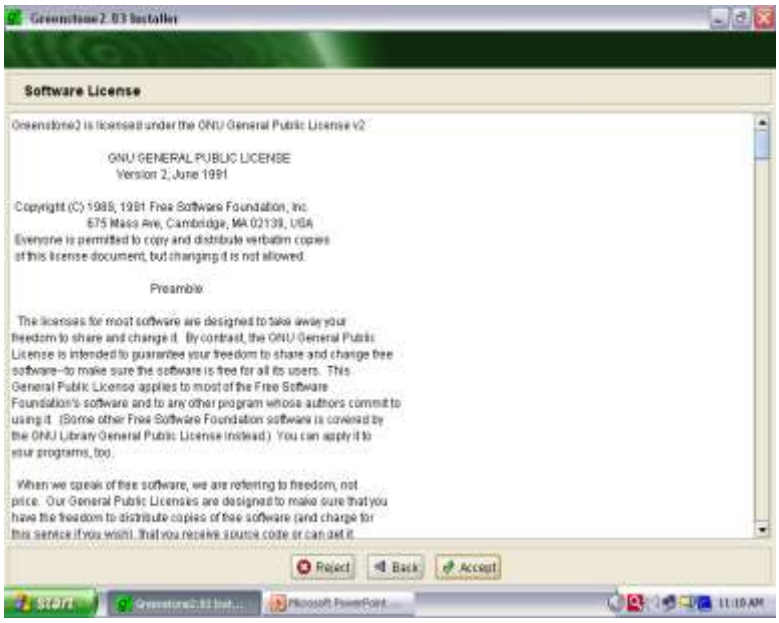
1. Select GREENSTONE DIGITAL LIBRARY SOFTWARE folder choose setup Greenstone Digital Library Software 2.83
2. Choose the type of installation (Local Library).
3. Set the admin password.

For the simplest installation, just accept the default at each point by clicking the Next button. Greenstone is installed.

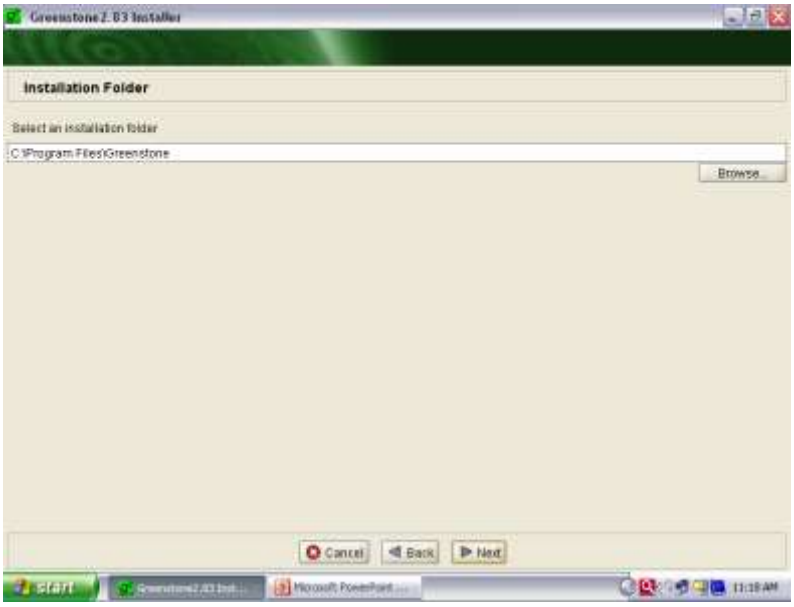


Choose setup Language. English (US)is the default.

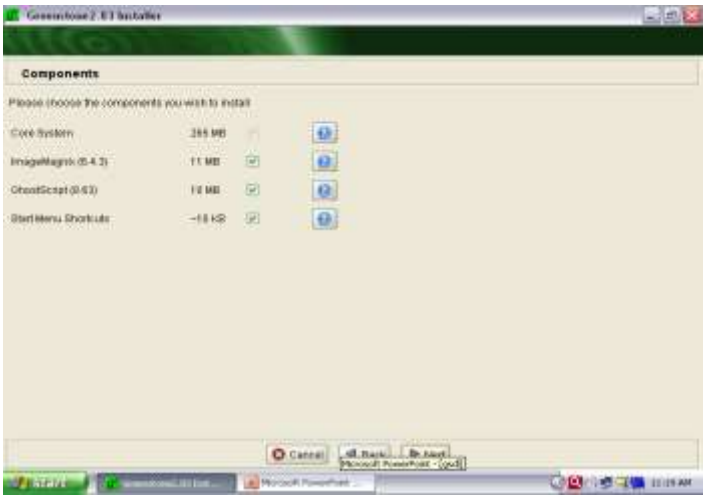
The Install Shield Wizard will begin the installation of GREENSTONE DIGITAL LIBRARY SOFTWARE software. Click <next>.

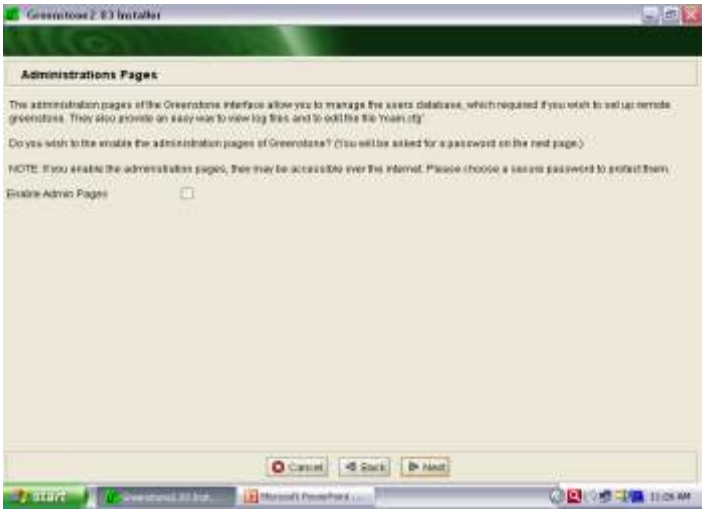


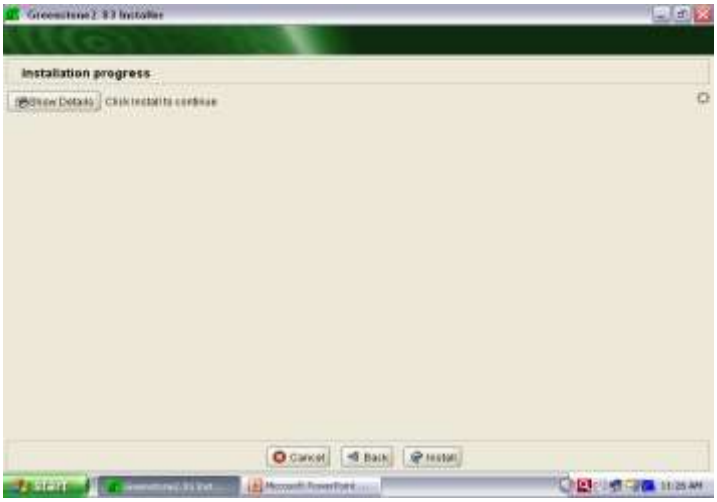
Accept all the terms of license agreement by clicking on <yes> button.

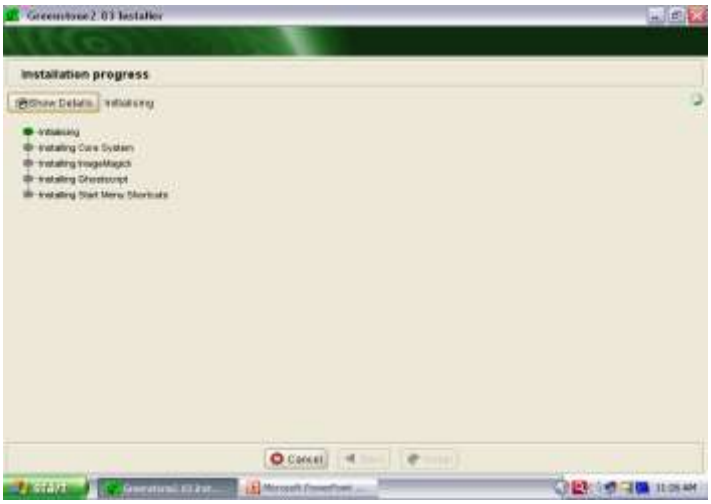


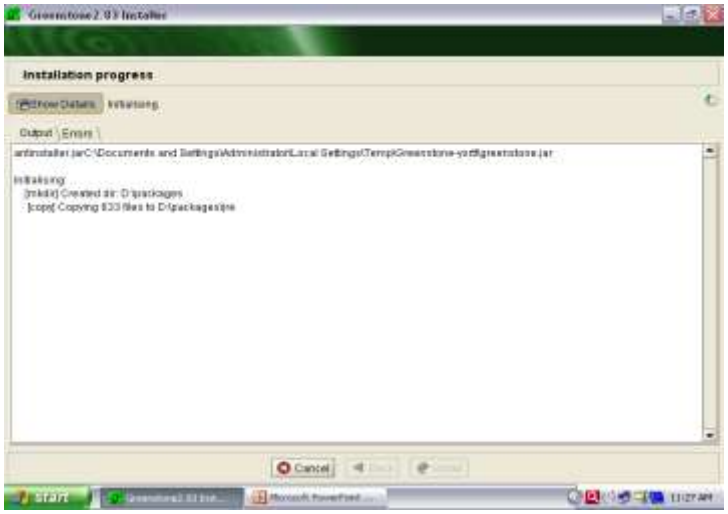
Select location C: Program files

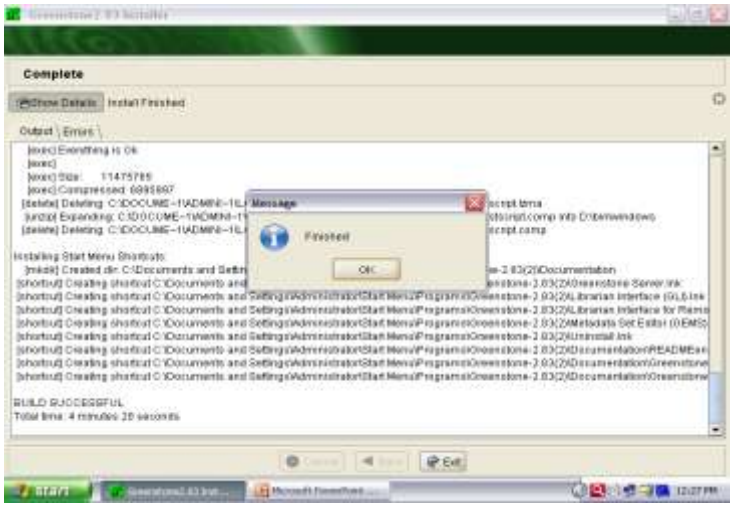

Components
Please choose the components you wish to install
Core System
ImageMagick (6.4.2)
GhostScript (8.63)
Start Menu Shortcuts
Cancel
Back
Next


Administrations Pages
Enable Admin Pages
Cancel
Back
Next


Installation progress
Show Details
Cancel
Back
Install

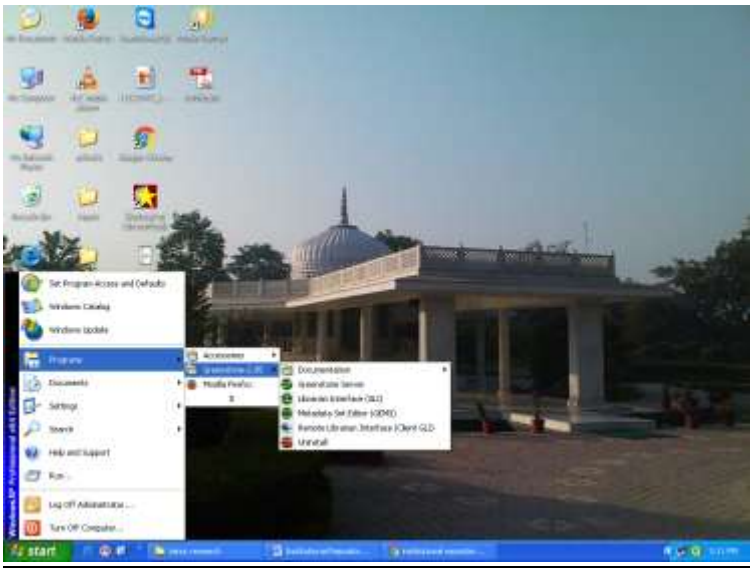

Installation progress
Show Details
Initialising
Cancel


Installation progress
Show Details
Initialising
Output
Errors
Cancel


Complete
Show Details
Install Finished
Output
Errors
Message
Finished
OK
BUILD SUCCESSFUL
Exit

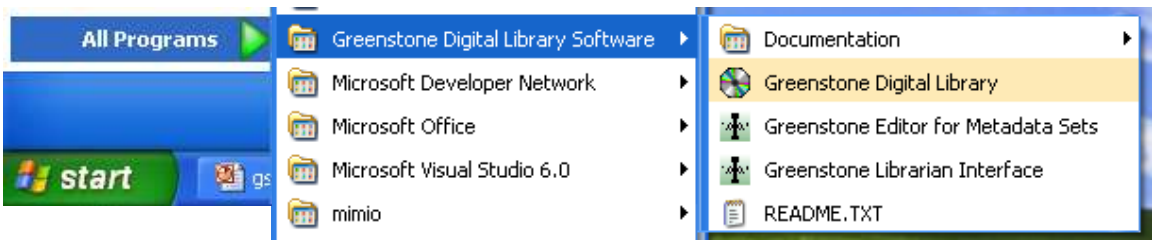
Once installation is completed, to start Greenstone system,

- Click on the Start button,



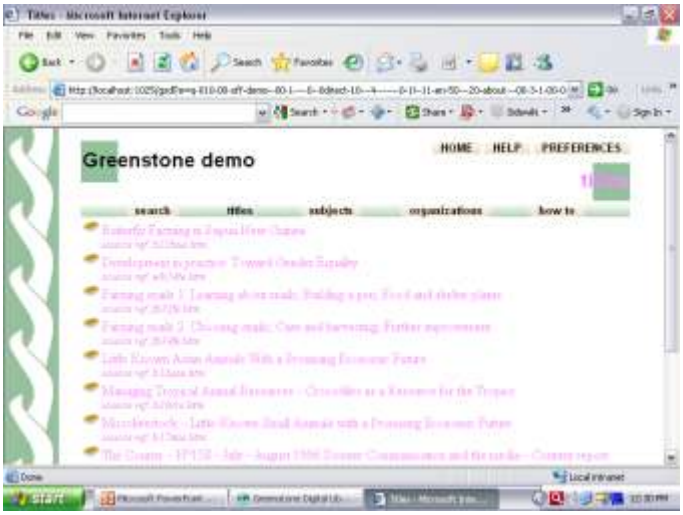
- Open the Program menu,
- and select Greenstone Digital Library.
- This brings a dialogue box: just click Enter Library.

This automatically starts Internet browser and loads the Greenstone Digital Library home page.




GREENSTONE DIGITAL LIBRARY


greenstone digital
library software
HELP
PREFERENCES
Greenstone
demo


Greenstone demo
HOME
HELP
PREFERENCES
search
titles
subjects
organizations
how to

INSTITUTIONAL REPOSITORY CREATION

A typical digital library built with Greenstone can contain many collections, individually organized, and bear a strong family resemblance. The collections are easy to maintain and can be augmented and rebuilt automatically. A flexible process structure allows different collections to be served by different computers and yet presented to the user as part of the same digital library -- and even, seamlessly, as part of the same collection. Existing collections can be updated and new ones brought on-line at any time, without bringing the system down -- the interface process checks periodically and automatically adds new collections to the list presented to the user.

Unicode, is a standard scheme for representing the character sets used in the world's languages, is used throughout Greenstone. This allows any language to be processed and displayed in a consistent manner. Collections have been built containing Arabic, Chinese, English, French, Mäori and Spanish. Multilingual collections embody automatic language recognition, and the interface is available in all the above languages.

Implementation

This part deals with the various hardware and software environment under which this software was implemented. Internet Explorer was used as the user interface. The Windows source of Greenstone source code occupies 50 Mb of disk space, but to compile it needs about 90 Mb. To compile the source on Windows it needs, the Microsoft Visual C++ compiler.

The default setup of Greenstone DL Software was chosen. It gets installed in the directory C:\Program Files\Greenstone Digital Library Software. First, the Local Library version of Greenstone was tried. It is a restricted version of Greenstone with an inbuilt webserver software. It was tried to create few test collections with different document formats. The disadvantage in the Local Library version is that the server needs to be restarted all the time whenever the browser is shut down. Later, the Web Library version of Greenstone was chosen to avoid port assignment conflicts. It is a standard version and to run the Web Library version, webserver software is essential.

The Collector was used to build collections. It requires Perl to run which gets installed along with Greenstone DL Software. Before creating proper collections, small test collections were built using 10-12 documents of

different document formats like text, HTML, Word, PDF, RTF etc to understand the collection building process. Plugins were added depending on the file formats chosen for the collection. Any metadata can be specified but it should correspond with the GML documents. Collection meta entry should correspond with the indexes entry. A few lines were added from the Word-Pdf demo configuration file to a collect.cfg file to display the document icons for Word and PDF documents.

The RTF Converter was used to convert RTF files to .html format. The Collection Organizer was used to create the browsing structures for few collections. This helped to create collections with hierarchical browsing structures. Each book or document gets identified with a job number which is unique.

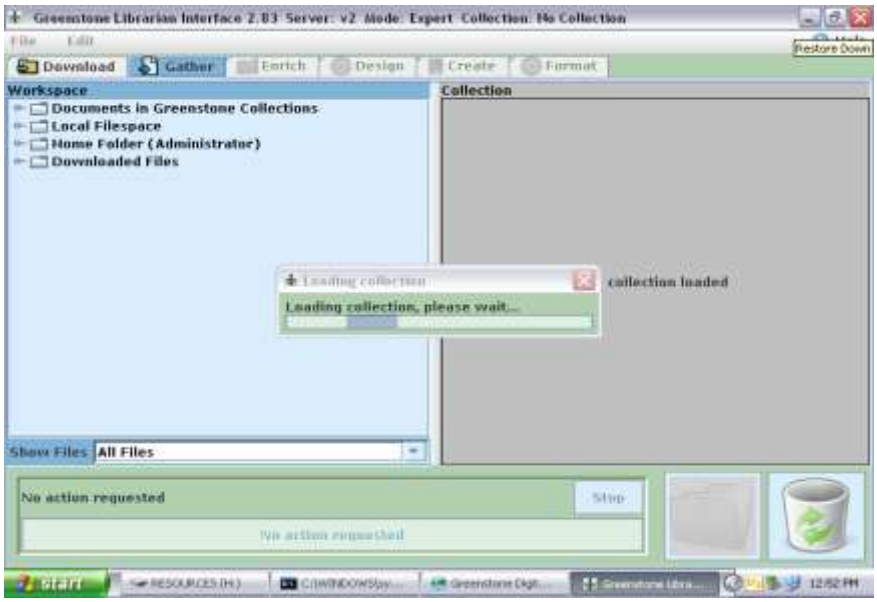
While creating any new collection, a short name gets assigned to the collection depending on the Collection name given by the user. A directory with this name gets created the Greenstone Digital Library Software/temp directory. This in turn holds five directories i.e, etc, import, images, index, perllib. If we are assigning collection icons then, the images should to copied in the images directory of this collection in the temp dir. In case the collection is cloned on the Greenstone Demo, then the metadata files i.e. the index.txt along with the Collection files should be placed in the Greenstone Digital Library Software's temp

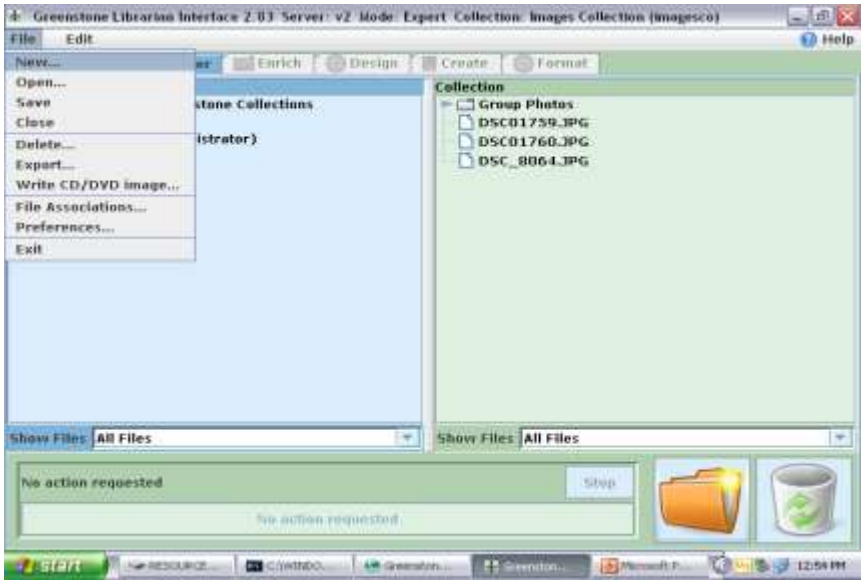
directory within the collection's import directory and the sub.txt and org.txt should be placed in the collection's etc directory. Actually behind the screen, the perl programs help in the whole collection building process. Plugins that extract metadata are written in Perl language. They are placed in the perllib/plugins directory. All source documents in Greenstone are converted into a format known as "Greentone Markup Language" or GML. It is an XML-compliant syntax that marks documents into sections and can also be used to store metadata at the document or section level. The Dublin Core metadata standard was used throughout Greenstone. Each document has an associated Object Identifier or OID. These are extended to identify sections and subsections.

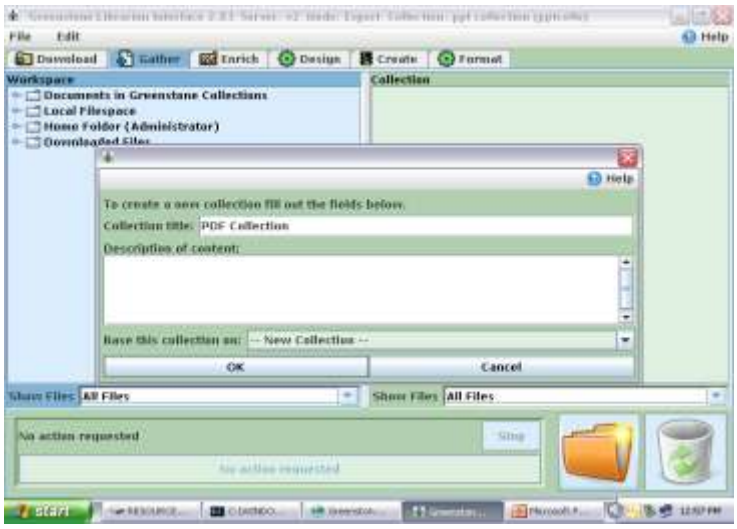
The source documents of different formats like text, HTML, Word, PDF and were well embedded with images and internal links within the file were chosen. The collection was cloned based on the Greenstone Demo Collection. The collection required essential metadata information to reflect the browsing structures. The Collection Organizer was used to generate the intermediate files used by Greenstone Digital Library Software to build the collections. Three subjects were identified and books/documents were added under each subject. The title, organization (i.e. author), year, language, no. of pages, job no., keywords etc were

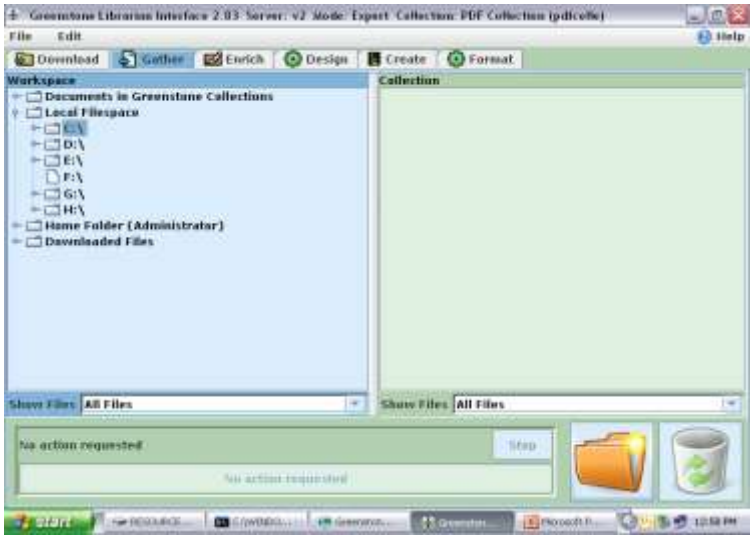
assigned for each document in the Collection Organizer. After creating the Collection's structure the three files index.txt, sub.txt and org.txt, which were generated were exported to the directory containing the documents to be imported. To access the collection, the searching and browsing facility was evoked with the help of the metadata information i.e the "search…..subject……titles a-z……authors a-z" icons. Subjects were identified as bookshelves. By clicking on the subject, one can browse through all the documents under that subject. Many of the documents had links to Power Point presentation files and these could be opened through the external link facility. The collection did not have "How to" metadata and so no classifier was built. To experiment the effectiveness of the Software, various queries were posed in the search interface and the advanced search preferences were tried. Later, the Collection was exported using the "Export Collection" facility. First, the collection gets stored in the temp directory with the collection's name and then we need to copy it on a CD using a CD-writing utility. The exported directory contains files related to installation process and three subdirectories containing the complete collection and software. This collection could now be viewed on any system though GREENSTONE DIGITAL LIBRARY SOFTWARE is not installed.

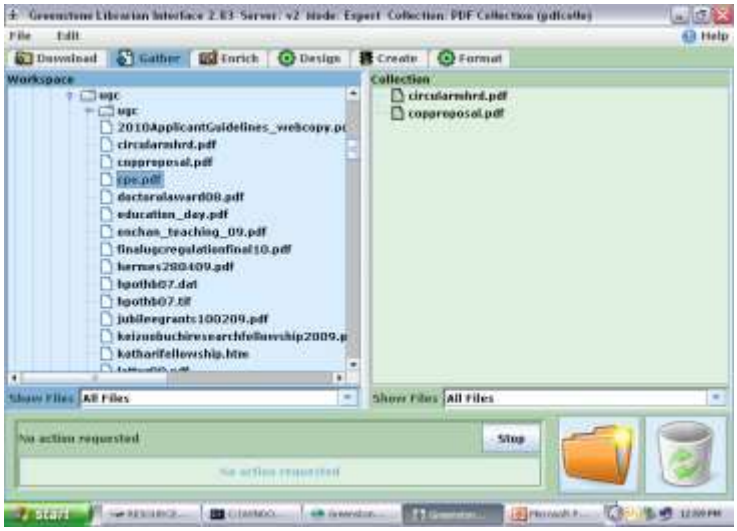
In this way, successfully collections could be built to understand the underlying mechanism in the Greenstone Software package.

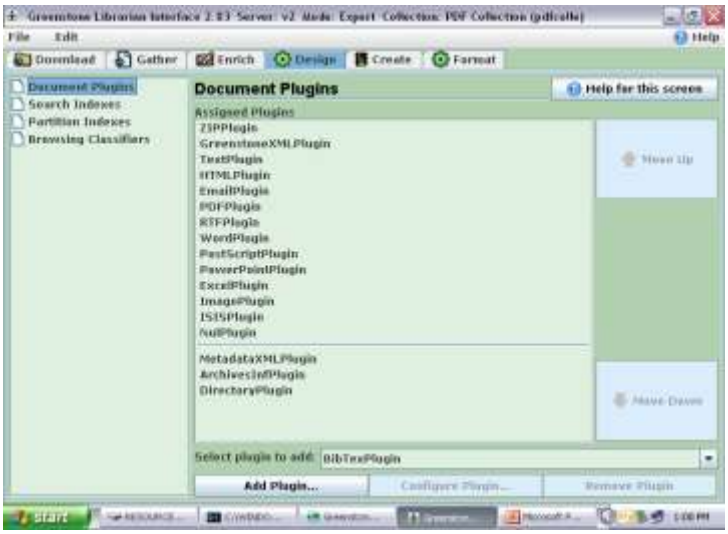


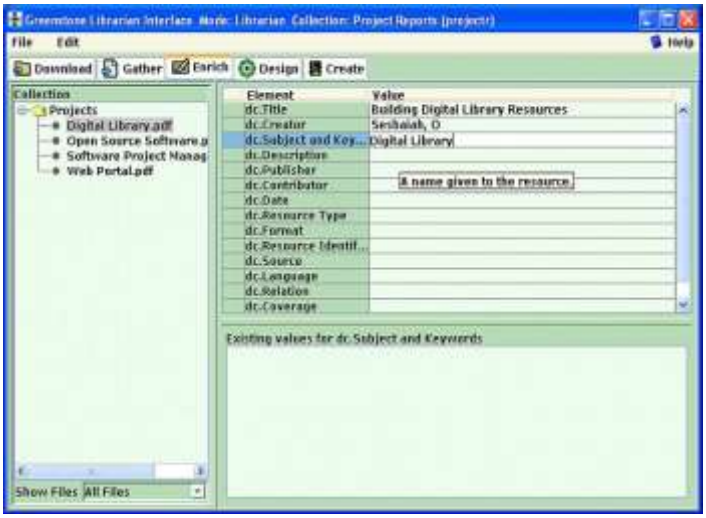


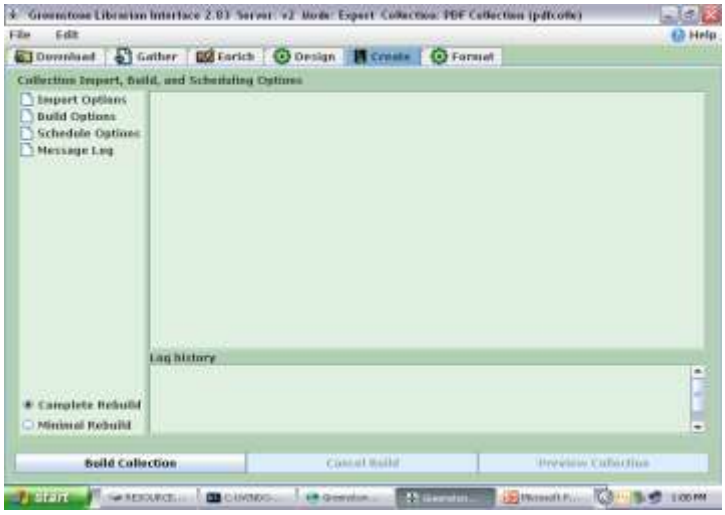

File Edit
Help
Download
Gather
Enrich
Design
Create
Format
Workspace
Collection
Documents in Greenstone Collections
Local Filespace
Home Folder (Administrator)
Downloaded Files
To create a new collection fill out the fields below.
Collection title:
PDF Collection
Description of content:
Base this collection on:
-- New Collection --
OK
Cancel
Show Files:
All Files
No action requested
Stop

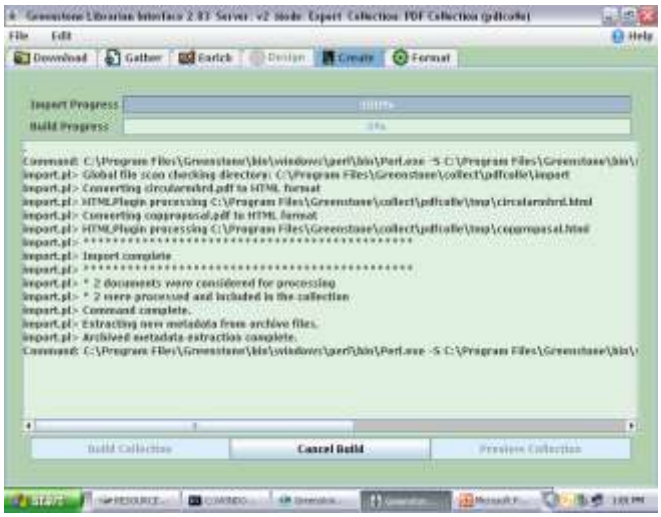

Greenstone Librarian Interface 2.83 Server: v2 Mode: Expert Collection: PDF Collection (pdfcolle)
File Edit
Help
Download
Gather
Enrich
Design
Create
Format
Workspace
Collection
Documents in Greenstone Collections
Local Filespace
C:\
D:\
E:\
F:\
G:\
H:\
Home Folder (Administrator)
Downloaded Files
Show Files:
All Files
No action requested
Stop

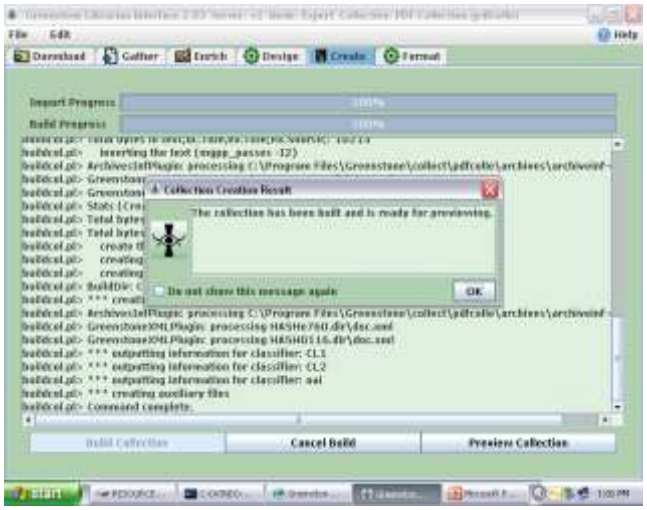

File Edit
Help
Download
Gather
Enrich
Design
Create
Format
Workspace
Collection
ugc
circularmhrd.pdf
copproposal.pdf
Show Files:
All Files
No action requested
Stop

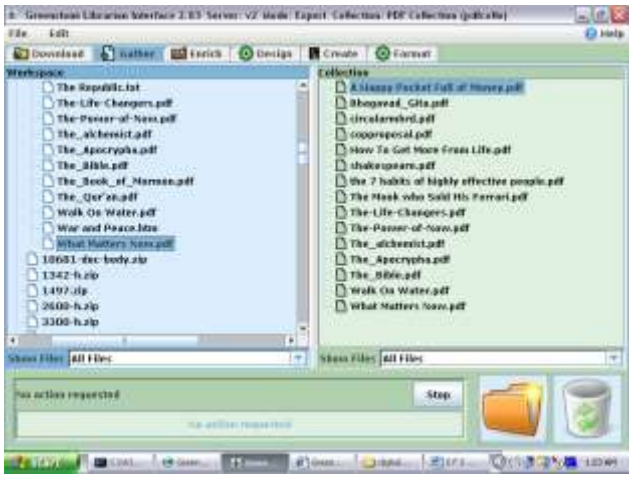

File
Edit
Help
Download
Gather
Enrich
Design
Create
Format
Document Plugins
Search Indexes
Partition Indexes
Browsing Classifiers
Document Plugins
Help for this screen
Assigned Plugins
ZIPPlugin
GreenstoneXMLPlugin
TextPlugin
HTMLPlugin
EmailPlugin
PDFPlugin
RTFPlugin
WordPlugin
PostScriptPlugin
PowerPointPlugin
ExcelPlugin
ImagePlugin
ISISPlugin
NulPlugin
MetadataXMLPlugin
ArchivesInfPlugin
DirectoryPlugin
Move Up
Move Down
Select plugin to add:
BibTexPlugin
Add Plugin...
Configure Plugin...
Remove Plugin

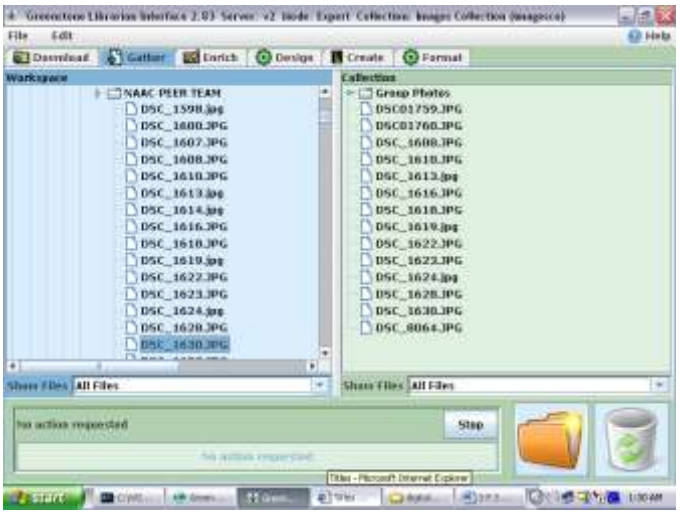

File
Edit
Help
Download
Gather
Enrich
Design
Create
Collection
Projects
Digital Library.pdf
Open Source Software.p
Software Project Manag
Web Portal.pdf
Show Files
All Files
Element
Value
dc.Title
Building Digital Library Resources
dc.Creator
Seshaiah, O
dc.Subject and Key...
Digital Library
dc.Description
dc.Publisher
dc.Contributor
dc.Date
dc.Resource Type
dc.Format
dc.Resource Identif...
dc.Source
dc.Language
dc.Relation
dc.Coverage
A name given to the resource.
Existing values for dc.Subject and Keywords

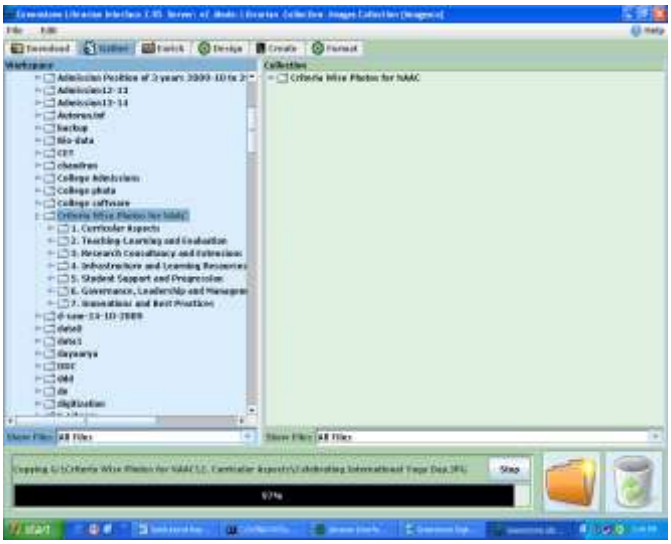

File
Edit
Help
Download
Gather
Enrich
Design
Create
Format
Collection Import, Build, and Scheduling Options
Import Options
Build Options
Schedule Options
Message Log
Log History
Complete Rebuild
Minimal Rebuild
Build Collection
Cancel Build
Preview Collection

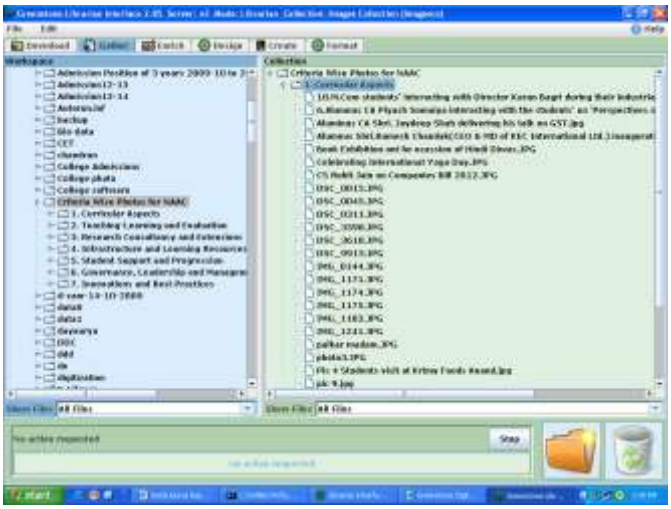

File Edit Help
Download Gather Enrich Design Create Format
Import Progress
Build Progress
Cancel Build

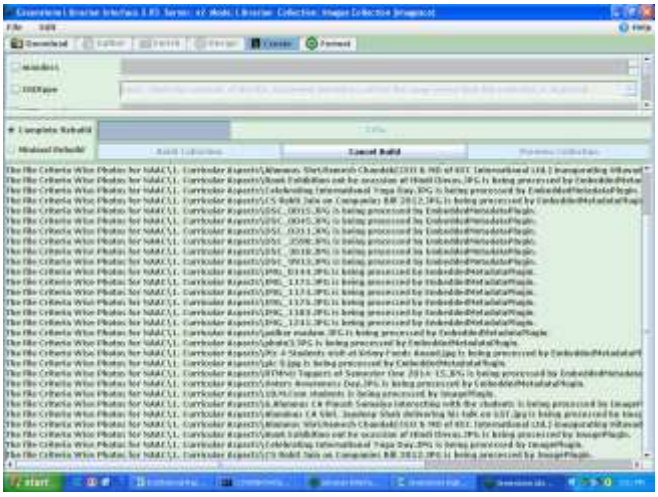

File Edit Help
Download Gather Enrich Design Create Format
Import Progress
Build Progress
Collection Creation Result
The collection has been built and is ready for previewing.
Do not show this message again
OK
Build Collection
Cancel Build
Preview Collection

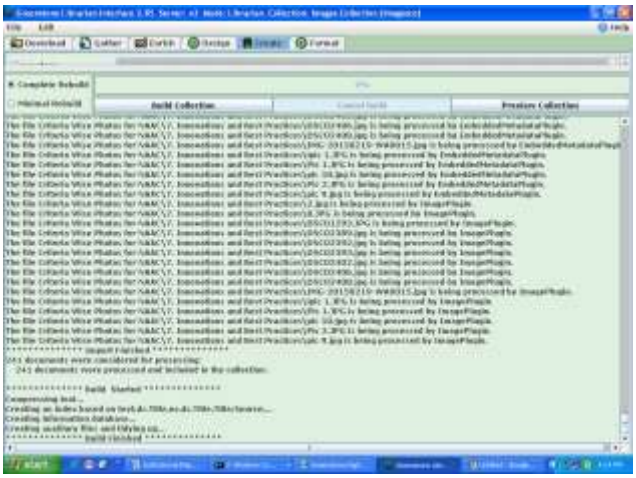

File Edit Help
Download Gather Enrich Design Create Format
Workspace
Collection
Show Files All Files
No action requested
Stop

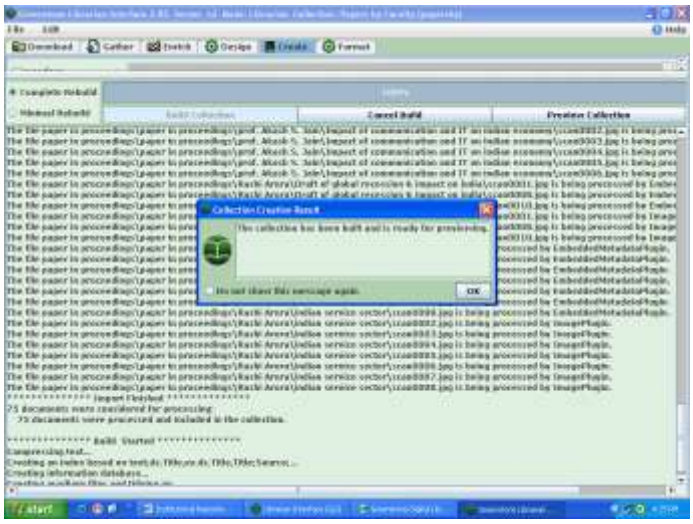






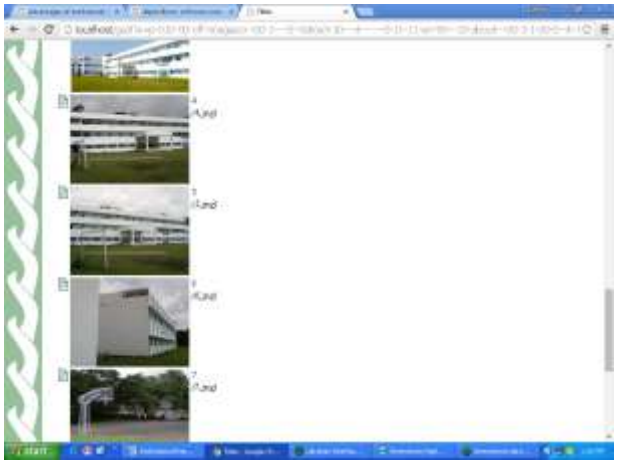






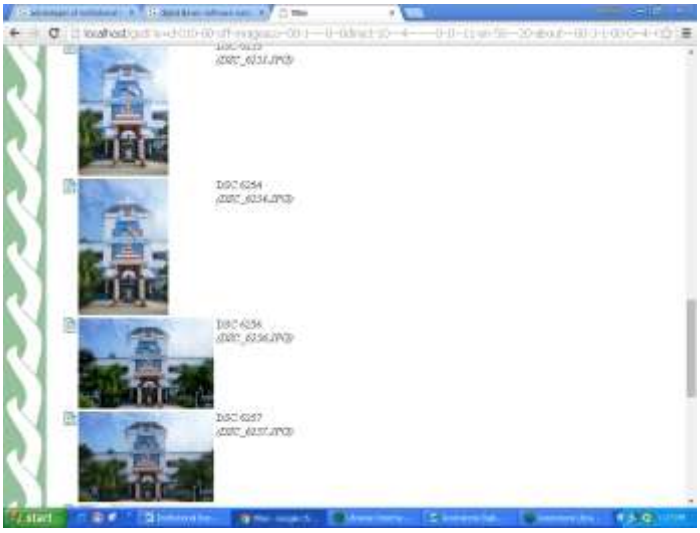

greenstone digital
library software
HELP
PREFERENCES
select a collection
Greenstone demo
Images Collection
Papers by Faculty
Refresher Course Collection
University Question Papers BA


Images Collection
HOME
HELP
PREFERENCES
titles



Images Collection

Images Collection



Images Collection

Papers by Faculty

Papers by Faculty

Scanned question papers.

OBSERVATIONS

Listed below are some of special features possessed by the Greenstone:

a) Accessible via web browser: Collections are accessed through a standard web browser (Netscape or Internet Explorer) and combine easy-to-use browsing with powerful search facilities

b) Full Text and Field Search: The user can search the full text of the documents, or choose between indexes built from different parts of the documents . For example, some collections have an index of full documents, an index of sections, an index of titles, and an index of authors, each of which can be searched for particular words or phrases. Results can be ranked by relevance or sorted by a metadata element.

c) Flexible browsing facilities: The user can browse lists of authors, lists of titles, lists of dates, classification structures, and so on . Different collections may offer different browsing facilities and even within a collection, a broad variety of browsing interfaces are available. Browsing and searching interfaces are constructed during the building process, according to collection configuration information.

d) Create access structures automatically: The Greenstone software creates information collections that are very easy to maintain. All searching and browsing structures are built directly from the documents

themselves. No links are inserted by hand, but existing links in originals are maintained. This means that if new documents in the same format become available, they can be merged into the collection automatically. Indeed, for some collections this is done by processes that wake up regularly, scout for new material, and rebuild the indexes—all without manual intervention.

e) Make use of available metadata: Metadata, which is descriptive information such as author, title, date, keywords, and so on, may be associated with each document, or with individual sections within documents. Metadata is used as the raw material for browsing indexes. It must be either provided explicitly or derivable automatically from the source documents. The Dublin Core metadata scheme is used for most electronic documents; however, provision is made for other schemes.

f) Plug-in extends system's capabilities: In order to accommodate different kinds of source document, the software is organized in such a way that "plug-in" can be written for new document types . Plug-in currently exist for plain text, html, Word, PDF, PostScript, E-mail, some proprietary formats, and for recursively traversing directory structures and compressed archives containing such documents.

g) Customization: The Greenstone allows customization of presentation of collection that are based on EXtensible Stylesheet Language

transformation (XSLT) and other agents that govern the definite functions of Digital library. The architecture of Greenstone purvey: a. A back end that provide services to manage documents and collections. b. A front end that provides a web based interface for searching and presentation of documents, collections.

h) Designed for Multi-gigabyte collection: Collections can contain millions of documents, making the Greenstone system suitable for collections up to several gigabytes.

i) Multilingual Support: Unicode is used throughout the software, allowing any language to be processed in a consistent manner. To date, collections have been built containing French, Spanish, Maori, Chinese, Arabic and English . On-the-fly conversion is used to convert from Unicode to an alphabet supported by the user's web browser.

j) Collections support multiple formats: Greenstone collections can contain text, pictures, audio and video clips . Most non-textual material is either linked in to the textual documents or accompanied by textual descriptions (such as figure captions) to allow full-text searching and browsing.

k) Administrative function provided: An “administrative” function enables specified users to authorize new users to build collections, protect documents so that they can only be accessed by registered users on

presentation of a password, examine the composition of all collections, and so on . Logs of user activity can record all queries made to every Greenstone collection

l) Collections can be published on the Internet or on CD-ROM: The software can be used to serve collections over the World-Wide Web. Greenstone collections can be made available, in precisely the same form, on CDROM. The user interface is through a standard web browser (Netscape is provided on each disk), and the interaction is identical to accessing the collection on the web—except that response times are more predictable. The CD-ROMs run under all versions of the Windows operating system.

5.1 Strengths of Greenstone Digital Library Software:

We can summarize the Strengths of Greenstone as follows:

- ***Widely accessible via Web*** - Collections can be accessed through a standard web browser.
- ***Multi-platform*** - Collections can be served on Windows and Unix, with an external Web server or a inbuilt server for Windows
- ***Flexible searching*** - Users can search the documents' full text, choosing between indexes built from different parts. Queries can

be ranked or Boolean; terms can be stemmed or unstemmed, case-folded or not.

- ***Flexible Browsing*** - Users can browse lists of authors, lists of titles, lists of dates, hierarchical classification structures, and so on. Different collections offer different browsing facilities.
- ***Zero Maintenance*** - All structures are built directly from the documents themselves. New documents in the same format can be merged into the collection automatically and can be accessed. No links need to be inserted by hand. The existing hypertext links in the original documents, leading both within and outside the collection, are preserved.

- ***Metadata-driven*** - Browsing and searching indexes are built from metadata. Metadata is associated with each document or with individual sections within documents. It can be provided explicitly or can be derived automatically from the source documents. The Dublin Core metadata scheme is used for most electronic documents.

- ***Extensible*** - The architecture is very extensible. Plugins can be written to accommodate new document types. Classifiers can be written to create new kinds of browsing indexes based on metadata.
- ***Phrases and key phrases*** **-** Standard classifiers create phrase and key phrase indexes of text or indeed any metadata.

- ***Multimedia*** - A collection can have source documents in different forms. Collections can contain pictures, music, audio and video clips.
- ***Large-scale*** - Collections containing millions of documents, and up to several gigabytes, have been built. Full-text searching is fast.

- ***Multi-language*** - Unicode is used throughout the software allowing any language to be converted 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.

- ***Compression*** **-** This reduces the size of the indexes and text. This increases the speed of the text retrieval.

- ***Security*** **-** An administrative function enables specified users to authorize new users to build collections, protect documents so that registered users on presentation of a password can only access them.
- ***Refreshing*** **-** Collections can be updated and new ones can be brought on-line.

- ***Sustained Operation*** **-** New collections can be installed without bringing the system down.
- ***CD-ROM option*** **-** Collections can be published on a self-installing CD-ROM. A multi-disk solution has been implemented for larger collections. Upto 150,000 pages can be indexed on one CD. Every CD can in turn become an Internet Server.

- ***Distributed Collections*** **-** Collections served by different computers can be presented to users as though they are part of the same library, through a flexible process structure.
- ***Z39.50 Compatible*** **-** The Z39.50 protocol is supported for accessing external servers and for presenting Greenstone collections to external clients.

And last but not least, because Greenstone is open-source software, it is easily modified.

5.2 Weaknesses of GREENSTONE DIGITAL LIBRARY SOFTWARE

- **Technological Obsolescence -** PDF files of earlier version could not digitized properly. Most of the PDF files having scanned content were displayed as images.
- **Content Refreshing -** The files had to be refreshed with the latest software so that it can be digitized and displayed on the browser interface properly linking to the content within. This consumes a lot of time.
- The html files in Word pdf demo are poorly formatted because of deficiencies in the programs that convert documents to HTML.
- The manuals are not very user friendly in decipicting the content to create or edit changes in the configuration files while creating a new collection and in using the Collection Organizer and its role in Greenstone Digital Library Software.
- The source code of the collections, being in executable(.exe) form, except for software developers, users can hardly made any changes in the search index buttons.

- PDF files take too much time in digitizing and sometimes depending on the size of the collection, it goes on for hours to digitize.
- PDF source files lose their images in conversion to HTML if the path has a space in it.
- There are no plugins that can handle MS-Excel format.
- RTF files fail to be handled in Word-Pdf demo type of collection.
- Sometimes, Internal links in the documents do not work properly.
- Greenstone does not allow deletion of individual documents within a collection.
- Greenstone provides no searching and browsing facility during the collection building process.
- Collections are not built properly when Norton Antivirus or any other virus protection software is running on the system.

CONCLUSION

Greenstone, being a open-source software is readily extensible, and benefits from the inclusion of GNU-licensed modules for full-text retrieval, database management, and text extraction from proprietary document formats.

Digital libraries will be ubiquitous in the future and will provide the basis for a very broad set of distributed living activities including computer-supported cooperative work, distance learning, electronic commerce and entertainment. The transition to an electronic information workplace has already begun in full force.

It provides a leadership role in the on-line development and application of worldwide access to digital library services. Development of this technology provides valuable fundamental research and supports the broader goal of research and education through improved means for collaboration and distance learning. We believe that digital libraries will significantly impact the quality of education and, indeed, the quality of life over the next decade. The development of digital libraries may be viewed as a fundamental contribution to research in all disciplines. Thus, through international cooperative efforts digital library softwares like

Greenstone should sufficiently become comprehensive to meet the world’s needs with the richness and flexibility that users deserve.

However, implementing an institutional repository is not as simple as just installing repository software and making the repository accessible to its potential users. The general idea that "if you build it, they will come" does not really reflect the reality of what happens when an academic institution establishes an institutional repository (IR). Adoption is usually slow, as users do not generally perceive the potential benefits that may arise from using such a system

The implementation of an institutional repository can be a Herculean task, not so much due to technical difficulties or unsustainable funding requirements, but mainly because institutional repositories interfere with the traditional practices of scholars and researchers. Nevertheless, as soon as an institutional repository is set up, all of the academy's research output is expected to be placed in the repository in order to increase the academy's visibility, usage and impact (among other things, such as constituting the long-term memory of the academy).

The task of convincing researchers to deposit their publications in the institutional repository is, by far, a repository manager's most demanding task. A great deal of research and imagination are needed to attempt to counter the initial reluctance of researchers to begin depositing their

research materials in the institutional repository. Simply breaking the barrier of inertia is probably the most difficult challenge of all. Once that obstacle has been surmounted, peer pressure and the awareness of the advantages of using such a system should be sufficient to maintain a healthy deposit rate from academic researchers.

BIBLIOGRAPHY

1. Bainbridge, D and others "Greenstone: A platform for distributed digital library applications." Proc European Digital Library Conference, Darmstadt, Germany; September 2001. (http://www.cs.waikato.ac.nz /~davidb/ ecdl01 /platform .ps
2. Ferreira**,** Miguel **"**Some Ideas on How to Create a Successful Institutional Repository**". In** D-Lib Magazine January/February 2008 Vol14, No 1/2 (http://www.dlib.org/dlib/january08/ferreira/01ferreira.html)
3. Greenstone Digital Library Installer's Guide (Install.pdf)
4. Greenstone Digital Library User's Guide (User.pdf)
5. Greenstone Digital Library Developer's Guide (Developer.pdf)
6. Greenstone "From Paper to Collection" Guide
7. Jain, P and Others "The Role of Institutional Repository in Digital Scholarly Communications" at (http://www.library.up.ac.za/digi/docs/jain_paper.pdf)
8. New Zealand Digital Library Project (http://www.nzdl.org/).
9. Tramboo , Shahkar and others . "A Study on the Open Source Digital Library Software's: Special Reference to DSpace, EPrints and Greenstone" **in** International Journal of Computer

Applications (0975 – 8887) Vol 59, No.16, December 2012 at http://arxiv.org/ftp/arxiv/papers/1212/1212.4935.pdf

10. Witten, Ian H and others " Greenstone: Open Source Digital Library Software", D-Lib Magazine, Oct 2001,Vol 7 No.10 (http://www.dlib.org /dlib/ October 01 /witten/10witten.html)