Greenstone:
Open source Software for Building Digital Library Collections

Dr. M.G. Sreekumar
UNESCO Coordinator, Greenstone Support, South Asia
Librarian & Head, CDDL, IIM Kozhikode
The Wonderful Open Source World
What is open source software?

- In the phrase *open source*, *source* refers to source code, the human-readable computer code which is the origin, or source, of the computer application.
- *Open* refers to the terms of access to that computer source code.
- So *open source* software is software for which the source code is freely available.
General Platforms / Applications

- OSs
  - Linux, BSD...
- Web Servers
  - Apache, Java apps..
- Server-Side Scripting
  - LAMP, WAMP, MAMP
- Web Services
  - Tomcat, Cocoon, Ant, SRU/SRW…
- RDBMS
  - MySQL, PostgreSQL, mSQL…
- Prgm Langs
  - Perl, PHP, Python…
- Mobile Computing
  - WML, WAP, WiFi…
- Net/Web Security
- OAI
- Img Processing
  - IMageMagick…
- Applications
  - Emacs, grep, sendmail, ssh…
- MLs
  - 1992+
IM/KM Platforms/Applications

- **CMS**
  - Joomla, Drupal, MediaWiki...

- **ILMSs**
  - KOHA, NewGenLib, Evergreen, OpenBiblio...

- **Search Engines**
  - Cheshire, Pears, dbWiz...

- **Open IRs**
  - DSpace, Eprints, GSDL...

- **Interoperability**
  - Z39.50, OAI-PMH...

- **Open DLs**
  - GSDL, DSpace, Eprints...

- **Portals**
  - MyLibrary, Wordpress...

- **Link Resolving**
  - GODOT...

- **Citation Mngr**
  - PKP’s Citation Manager...

- **Serials Mngr**
  - CUFTS...

- **DBMSs**
  - CDS/ISIS, Win/ISIS, GenISIS...

- **OAI DPs / SPs**
  - PHP OAI DP...

**IM/KM Platforms/Applications**

- **Open IRs**
  - DSpace, Eprints, GSDL...

- **Interoperability**
  - Z39.50, OAI-PMH...

- **Search Engines**
  - Cheshire, Pears, dbWiz...

- **Open DLs**
  - GSDL, DSpace, Eprints...

- **Portals**
  - MyLibrary, Wordpress...

- **Link Resolving**
  - GODOT...

- **Citation Mngr**
  - PKP’s Citation Manager...

- **Serials Mngr**
  - CUFTS...

- **DBMSs**
  - CDS/ISIS, Win/ISIS, GenISIS...

**IM/KM Platforms/Applications**

- **Open IRs**
  - DSpace, Eprints, GSDL...

- **Interoperability**
  - Z39.50, OAI-PMH...

- **Search Engines**
  - Cheshire, Pears, dbWiz...

- **Open DLs**
  - GSDL, DSpace, Eprints...

- **Portals**
  - MyLibrary, Wordpress...

- **Link Resolving**
  - GODOT...

- **Citation Mngr**
  - PKP’s Citation Manager...

- **Serials Mngr**
  - CUFTS...

- **DBMSs**
  - CDS/ISIS, Win/ISIS, GenISIS...
Advantages and Disadvantages

Advantages
- Mostly issued under an internationally accepted License
- Access to source code and ability and right to modify it
- Right to redistribute modifications to benefit wider community
- Free
- Excellent support networks
- Large and enthusiastic user base

Disadvantages
- Limited or no accountability
- Informal and unaccountable support channels
About Greenstone

Greenstone is a suite of software for building and distributing digital library 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 developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software, issued under the terms of the GNU General Public License. Read the Greenstone Factsheet for more information.

The aim of the Greenstone software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. Digital libraries are radically redefining how information is disseminated and acquired in UNESCO's partner communities and institutions in the fields of education, science and culture around the world, and particularly in developing countries. We hope that this software will encourage the effective deployment of digital libraries to share information and place it in the public domain. Further information can be found in the book How to build a digital library, authored by two of the group's members.

The complete Greenstone interface, and all documentation, is available in English, French, Spanish, Russian and Kazakh. Greenstone also has interfaces in many other languages. We are looking for volunteers to add new language interfaces and help maintain existing ones.

International Links:
Greenstone support for Southern Africa
Greenstone support for South Asia
Grupo de Usuarios de Habla Hispana de Greenstone
Lista De Usuários de Língua Portuguesa do Greenstone
Eling de la communauté francophone
Greenstone, Libraries and Open Access

“The aim of the software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. Digital libraries are radically reforming how information is disseminated and acquired in UNESCO's partner communities and institutions in the fields of education, science and culture around the world, and particularly in developing countries. We hope that this software will encourage the effective deployment of digital libraries to share information and place it in the public domain.”

—from www.greenstone.org
What is the Greenstone software?

- Software suite for building, maintaining, and distributing digital library collections
- Comprehensive, open-source
- Developed by New Zealand Digital Library Project at the University of Waikato
- Distribution and promotion partners:
  - UNESCO
  - Human Info NGO, Belgium
  - NCSI, Bangalore; UCT, Cape Town; Dakar, Senegal; Almaty, Kazakhstan; …
  - You!
Humanity Development Library
for sustainable development and basic human needs

- 160,000 pages
- 30,000 images
- 800 books
- 430 magazines
- 340 kg
- US$20,000
- CD-ROM
- US$1
- Fully searchable
- Win3.1x upward
- Stand-alone + intranet server
- Web browser user interface

Global Help Project, Antwerp (+ UN agencies)
Features of Greenstone

- Open Source Philosophy
- Interfacing & Content Delivery via Web
- Multi S/W Platform
- Multi Lingual Support
- Multi Formats
- Structured Metadata in XML using DC
- Metadata Extraction
- Searching & Browsing
- Plug-ins for Documents
- Full-text mirroring
- Text Level Penetration
- Data Compression
- Password protection
- Administrative Functions
- Concurrent & Dynamic Content Development
- Uniform Presentation
- Publishing on CDROMs
- International Presence
Greenstone Features contd...

- Easy Installation
- Easy Maintenance
- Content Development (3 alternate ways)
- Predominantly GLI now - since (V. 2.41)
- Hierarchy Structure
- Interface Customization
  - Front Page Design, Header for the Digital Library, Collection Icon, Cover Images
- Collection Configuration (Collect.cfg) File
- Scalability, Flexibility
- Interoperability (Crosswalk), OAI Compliance
- Lifeline: Listserv / E-Group / Archives / Wiki
UNESCO: Distributing Greenstone DL software

Sustainable development

“Give a man a fish, feed him for a day
Teach a man to fish, feed him for life”

Greenstone software on CD-ROM

- GNU licensed
- Fully documented ... in English/French/Spanish/Russian
- Language interfaces: Arabic Chinese Czech ... Thai Turkish
- Unix/Windows/Mac OS-X
- Trivial to install
- GUI interface for gathering, enriching, building ...
- Serve collections on Web or write them to CD-ROM
- Document formats: HTML, Word, PDF, PS, plain text, e-mail
- Metadata formats: XML, DC, OAI, MARC, ...

download from http://greenstone.org
What we wanted

- “Collections” of digital material
- Individualized, depending on metadata etc
- Up to several GB of text ...
- ... + associated images, movies, whatever
- Fully searchable
- Served on WWW, or published on CD-ROM
- Multi-platform (Unix + all Windows + Mac)
- Multi-format documents and metadata
- Multi-lingual: documents and interfaces
- Multimedia
- Metadata: standard and non-standard
Greenstone DL Software

Access
- Accessible via any Web browser
- Server runs on Windows and Unix
- Collections can be published on CD-ROM

Searching/browsing
- Full-text and fielded search
- Flexible browsing facilities
- Metadata-based (Dublin Core)
- Collection-specific
- Hierarchical phrase browsing supported
- Creates all access structures automatically

Extensible
- Plugins — new document, metadata formats
- Classifiers — new metadata browsers

Multilingual
- Documents and interfaces
- Chinese, Arabic, Maori, Russian, etc (+ European, Indian...)

Multimedia
- Video, Audio...
The power of open source: Greenstone uses ...

- Ghostscript: Interpreter for Adobe Postscript documents (Postscript plugin)
- Kea: Keyphrase extraction program (to generate metadata)
- pdf2html: Converter for PDF documents (PDF plugin)
- rtf2html: Converter for RTF documents (RTF plugin)
- TextCat: Detects languages and document encodings
- wvWare: Converter for Word documents (Word plugin)
- Xhtml: Converter for Excel/Powerpoint documents (plugins)
- XML::Parser: Parses XML documents, used to read and write Greenstone’s internal XML document format
and ...

- **MG**: Creates compressed full-text indexes and performs searches
- **GDBM**: Database used for metadata etc
- **wget**: Downloading pages from the Web when creating collections
- **YAZ**: Client and server implementation of Z39.50
- **Stemmer**: English language stemmer

---

- **GCC**: C/C++ compiler
- **CVS**: Version control system
- **Perl**: Used for plugins etc
- **Apache**: Web server used by many Greenstone installations
- **OAI-PMH**: OAI Performance
Example Greenstone collections

- Rapid growth in use
- International – Many Countries…China, Germany, India, UK, USA, Russia, Malaysia, Singapore… – Almost all countries/Continents
- Increasing activity on Greenstone mailing list
- Promotion by UNESCO – “deployment of DL’s for sharing public domain information”
- Wide variety of DL collections have been developed in several languages
  - historical, educational, cultural, and research
New York Botanical Garden

- Rare 19th century works on American trees
- Gorgeous full-color plates
Allegro de concert : pour le piano : op. 46 /

Title: Allegro de concert : pour le piano : op. 46 /
Composer: Chopin, Frédéric, 1810-1849.
Dedicatee: M. J.-A. Streicher, Friederike
Publication Info.: Leipzig : Breitkopf & Härtel ; Paris : Chez M. Schlesinger, [1841]
Description: 23 p. of music ; 34 cm.
Notes: Engraved. Price: 1 Thlr. 6 Ngr.
Also available on the Internet.
Subject: Piano music.
Plate No.: 6651
URL: http://pl.lib.uchicago.edu/1001/dig/chopin/017
Call No.: M25.C54 A41
Ko te kahere o nui tiri nei.

He tuitakinga tangata.

Ka taea tawhiti te renga o Tawera, ka ngaro e tahi o nga whetu maori, ka oti te hurate kalaha mangungu te po, kia kitera nga kapua mai i te marangai (Te Ita) ka rea mea te tangata "ka takiri te tara" me reira ka ara kororia mai i runga i tona torona whero, te rangatira o te ao; na, ka rangi. Otu aua anaau e poka kore ka marika kai, ka moiri te ra i runga kau pe, ka mea toku whakaaro maranga i te moe te tahi tangata haere mai ana tetahi i te huia tutaki pua raua, ka noho ki raro.

Ko "Pai-Maori" tetahi o taa hanga, te tangata poto, he moko i runga kupu, tu-a-kumatarua, he mawhata kai, engaipo, i parahia te ihu, takataka ngahakohi, nga kanohi, e tokii ana nga paparinui. Ka matotori ana nga ngutu, keokeo nga tukimata, a i hua nui nga kai, he moko e tetenei katoa. Ko nga
Chinese documents (pictures of text) + Chinese interface

Peking University Library
红楼梦

第一回 甄士风尘怀闺秀
第二回 贾母演说荣国府
第三回 贾母抛父进京都
第四回 薛姨母混世罗衣

诗曰：
陈国汉户数光名，千金女儿误门立。
稀疏天助星前，罗绮遥遥相映起。
面道今年初见人，倩影摇摇长随身。
思得侍儿为道意，后室一笑相相亲。

话说王婆拿细子出门，便向妇人道而下笑来，说道：“老身去那街上看买东西，不料领着王婆出去，有白娘子相待官人坐一坐。盖了有酒，便叫二两，且和大官人吃着，老身再喜县东街，那里有好酒买一两来，有好酒我儿多。”妇人听了一说：“好姐休走，我酒不多用。”婆子便道：“你姐，大官人又不是别人，没处相陪，有好酒，不怕怎的。”妇人说不用了”。坐者却不知身，婆子一把门上，用索儿拴了，便关住二人在屋里。当路坐了，一头搂着铁，
وصف الجامع

اتهم جزءًا من هذه المجموعة الفنية التي تطور الشريعة ويتكون من قاعدة متميزة مكونة من أربع دعائم مربعة بها أربعة أعمدة رخامية مكسوة من الخارج بخراشة دقيقة مطروعة بالتصغروف وكان من الاعتدام والأحجام من جميع أنحاء مجموعة على شكل مربعة والمثلثات ت تكون من ثلاثة طوابق واحدة، وتحتوي على شكل مربع على مرتفع من ثمانية أعمدة رخامية مكسوة من الخارج بألوان متعددة وتم تحديد المثلثات على بدناصر محمد بن قلاوون عام 1300 ميلادي وتوجد المدرسة المنصورية أمام باب القبة ومتممة الأبواب الشرقية له. وتم تجديد المدرسة المنصورية أمام باب القبة وبمجرد الابتدائية الشرقية يبدأ المدرسة بين عامي 1340 ميلادي ويتكون من ثلاثة أروقة، أكبرها أسماء المدينة بالاسم العثماني وكونها مدرسة على أعمدة رخامية تعلوها عقود حلي، هي السياجات المستديرة فوفقاً بالزخارف الجصية، أما البيمارستان لم يبق شيء من العناوين.:
UNESCO, Paris

SAHEL point DOC
ANTHOLOGIE du Développement au Sahel

Agriculture, pêche, élevage
Environnement
Développement social et humain
Coopération, intégrations, marchés
Culture, information, communication

Documentaires par : AMITA, Banque Mondiale,
Centre de la Banque, CFA, BCD, GEMET,
Institut du Sahel, Editions Massonnette et Larose, YMLD.

UNESCO - SAHEL point DOC - CD-ROM 1.0

Ce CD-Rom SAHEL point DOC, première publication de la collection PUBLICA de l'UNESCO, met à la disposition du public un choix de documents d'informations utiles et disponibles en langue française sur le développement du Sahel. Plus de 12.000 pages de documents y sont réparties en cinq thèmes que vous pourrez découvrir sous le bouton "sujets".

Comment trouver les informations ?

AIDE. Pour de l'aide à tout moment, activez le bouton "Aide".

Cliquez une couverture pour accéder au livre

Il y a cinq manières d'accéder aux informations contenues dans ce CD-ROM :

• rechercher des mots particuliers ou une phrase
• accéder aux publications par sujet
• accéder aux publications par titre dans l'ordre alphabétique
• accéder aux publications par organisation
• accéder aux publications par le bouton

Le Projet de la Collection PUBLICA de l'UNESCO
Collections: on the Web

nzdl.org

(demo, not service)
Greenstone & Associated Softwares

- Greenstone 2.85 (http://www.greenstone.org)
- Java Runtime Environment (JRE) (http://java.sun.com)
- ImageMagick (http://www.imagemagick.org)
- Ghostscript (http://www.cs.wis.edu/~ghost/)
- Module for CD-ROM Publishing (http://www.greenstone.org)
- Additional Language Pack (http://www.greenstone.org)
Installing Greenstone

Softwares/Files Required

File and Folder Tasks
- Make a new folder
- Publish this folder to the Web
- Share this folder

Other Places

Details
GSDL-Software
File Folder
Date Modified: Today, July 10, 2012, 4:06 PM

Greenstone-2.85-windows.exe
jre-1_5_0_05-windows-i586-p...
Sequence of Installation

1. Java Runtime Environment (JRE) ([http://java.sun.com](http://java.sun.com))
2. ImageMagick ([http://www.imagemagick.org](http://www.imagemagick.org))
Installing... Java Runtime Environment (JRE)

Step 1. Check and Remove any Java Presence

Step 2. Locate the jre-1_5_0_05-windows-i586-p and Click to Install
Installing... Java Runtime Environment (JRE)

**Installing**

The program features you selected are being installed.

**Status:**

Please wait while the Install Wizard installs J2SE Runtime Environment 5.0 Update 5. This may take several minutes.

**Installation Completed**

The Install Wizard has successfully installed J2SE Runtime Environment 5.0 Update 5. Click Finish to exit the wizard.
Greenstone2
Digital Library Software

Extracting JAR installer
Greenstone 2.81rc installer

Greenstone2 is software for building and sharing digital libraries.

It is recommended that you uninstall any previous installations of Greenstone2 before running this installer.
Greenstone2 is licensed under the GNU General Public License v2

GNU GENERAL PUBLIC LICENSE
Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc.
675 Mass Ave, Cambridge, MA 02139, USA
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your
freedom to share and change it. By contrast, the GNU General Public
License is intended to guarantee your freedom to share and change free
software--to make sure the software is free for all its users. This
General Public License applies to most of the Free Software
Foundation's software and to any other program whose authors commit to
using it. (Some other Free Software Foundation software is covered by
the GNU Library General Public License instead.) You can apply it to
Installation Folder

Select an installation folder

C:\Program Files\Greenstone

Question

The folder does not exist, create it?
C:\Program Files\Greenstone

Yes  No
Please choose the components you wish to install

- Core System: 179 MB
- Start Menu Shortcuts: ~10 KB
- ImageMagick with JPEG2000 (6.4.3): 11 MB
- GhostScript (8.63): 10 MB

[Checkboxes for each component]
Please set the Admin Password

In order to access certain parts of the Greenstone interface you need a password. A user with the username 'admin' will be created for you with the password you provide. The password must be between 3 and 20 characters long.

Admin Password: admin
Installation progress

Show Details  Click Install to continue
Installation progress

Show Details  Click Install to continue

Output  Errors
Installation progress

Installing Core System

Output

Install native library C:\Documents and Settings\bijul.ov4n\orangevolt-4n-1.1.1.dll
antinstaller.jar C:\Documents and Settings\bijul\Local Settings\Temp\Greenstone-ohgub\greenstone.jar

Looking For Previous Installation:

Installing Core System:
C:\Documents and Settings\bijul\Start Menu\Programs\Greenstone-2.81rc
[echo] basedir: C:\DOCUME~1\bijul\LOCALS~1\Temp\Greenstone-ohgub\antinstall
[echo] installDir: D:\Program Files\Greenstone2
[echo] antinstaller.jar: C:\Documents and Settings\bijul\Local Settings\Temp\Greenstone-ohgub\greenstone
[echo] startmenu.path: C:\Documents and Settings\bijul\Start Menu\Programs\Greenstone-2.81rc
[echo] Creating Installation directory
[unzip] Expanding: C:\Documents and Settings\bijul\Local Settings\Temp\Greenstone-ohgub\greenstone.jar
[echo] Installing Core Component
[unzip] Expanding: C:\Documents and Settings\bijul\Local Settings\Temp\Greenstone-ohgub\greenstone.jar
Complete

- Looking For Previous Installation
- Installing Core System
- Installing Start Menu Shortcuts
- Installing ImageMagick
- Installing Ghostscript

[Buttons: Show Details, Install Finished, Cancel, Back, Exit]
Greenstone’s Interfaces

Digital Library (User + Librarian)

Librarian Interface (GLI)
Invoking Greenstone on Browser
Opening Greenstone on Browser

Digital Library Server

Greenstone Digital Library

GREENSTONE DIGITAL LIBRARY
Version 2.85

Enter Library
Press this button to begin using the library.
Opening Greenstone on Browser

Greenstone Digital Library

greenstone digital library software

Collections

select a collection

Greenstone demo

The Librarian Interface
Helps you create new collections, modify or add to existing ones, or delete collections

About Greenstone
Tells you about the Greenstone software and the New Zealand Digital Library Project where it originated

Greenstone Documentation
Greenstone manuals

Greenstone is a suite of software for building and distributing digital library 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 developed 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.

The aim of the software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. Digital libraries are radically reforming how information is disseminated and acquired in UNESCO's partner communities and institutions in the fields of education, science and culture around the world, and particularly in developing countries. We hope that this software will encourage the effective deployment of digital libraries to share information and place it in the public domain.

This software is developed and distributed as an international cooperative effort established in August 2000 among three parties.

New Zealand Digital Library Project at the University of Waikato
Greenstone software grew out of this project, and this initiative has been endorsed by the Communication Sub-Commission of the New Zealand National Commission for UNESCO as part of New Zealand’s contribution to UNESCO’s programme.
Opening the GLI
Opening the GLI
GLI Functions

- Establish new collection (or work on old)
- Select files to include in collection (Gather)
- Enrich files with metadata (Enrich)
- Select Plugins, Indexes, Classifiers (Design)
- Build Collection (Create)
- Format and Control Display (Format)
- Customize Appearance
- Preview Collection
GLI

- Building collections
- Interactive Java program
- Runs on anything
- Build a collection on the computer you are on
- ... plus new applet version
- Includes metadata editor

Caveat: cannot deal with such huge collections as Greenstone can (particularly of metadata)

- Invoke GLI: build a small collection of HTML files
- Gather
- Create
- Look at extracted metadata
- Set up shortcut in the Librarian interface
Collection Building...

- Greenstone used to have three modes of collection building, viz., **Command Line, Web Interface and the GLI** (Greenstone Librarian Interface)
- Progressing with version 2.4x., the GLI got strengthened as well as popularized
- Web Interface mode has been withdrawn temporarily.
- The GLI based collection building is quite easy and simple a method.
- Collection developers can activate the GLI software and use the ‘Gather’, ‘Enrich’, ‘Design’, ‘Format’ and ‘Create’ panel for making collection
Collection Building

• Input: a set of source documents
• Possibly in many different formats
• Greenstone “imports” these documents and converts them to its own internal (GA) format
  – Extracts as much metadata as possible
• Greenstone “builds” indexes and browsing structures using the GA files
• Start with a few documents, get the design right, then add the bulk of the documents
Building a New Collection

In GLI, Go to File, Select New and Say “Multimedia” and base it to New
Building a New Collection

In Gather, Browse Files From Workspace & Drag-Drop to Collection Area
<table>
<thead>
<tr>
<th>Complete Rebuild</th>
<th>Build Collection</th>
<th>Cancel Build</th>
<th>Preview Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal Rebuild</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The collection has been built and is ready for previewing.
multimedia

About this collection

This collection contains 709 documents. It was last built 0 days ago.

How to find information in the Multimedia collection

There are 3 ways to find information in this collection:

- search for particular words that appear in the text by clicking the Search button
- browse documents by Title by clicking the Titles button
- browse documents by Filename by clicking the Filenames button
greenstone digital library software

select a collection

Audios
Collection1
Greenstone demo
isis-to-greenstone
multi-format collection

Image-Collection

Multimedia
my-references

The Librarian Interface
Helps you create new collections, modify or add to existing ones, or delete collections

Administration Page
Allows you to add new users, summarizes the collections in the system, gives technical information on the Greenstone installation

About Greenstone
Tells you about the Greenstone software and the New Zealand Digital Library Project where it originated

Greenstone Documentation
Greenstone manuals

Greenstone is a suite of software for building and distributing digital library 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 developed 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.

The aim of the software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. Digital libraries are radically reforming how information is disseminated and acquired in UNESCO’s partner communities and institutions in the fields of education, science and culture around the world, and particularly in developing countries. We hope that this software will encourage the effective deployment of digital libraries to share information and place it in the public domain.
<table>
<thead>
<tr>
<th>Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.Title</td>
<td></td>
</tr>
<tr>
<td>dc.Creator</td>
<td></td>
</tr>
<tr>
<td>dc.Subject and Keywords</td>
<td></td>
</tr>
<tr>
<td>dc.Description</td>
<td></td>
</tr>
<tr>
<td>dc.Publisher</td>
<td></td>
</tr>
<tr>
<td>dc.Contributor</td>
<td></td>
</tr>
<tr>
<td>dc.Date</td>
<td></td>
</tr>
<tr>
<td>dc.Resource Type</td>
<td>Discography</td>
</tr>
<tr>
<td>dc.Format</td>
<td></td>
</tr>
<tr>
<td>dc.Resource Identifier</td>
<td></td>
</tr>
<tr>
<td>dc.Source</td>
<td></td>
</tr>
<tr>
<td>dc.Language</td>
<td></td>
</tr>
<tr>
<td>dc.Relation</td>
<td></td>
</tr>
<tr>
<td>dc.Coverage</td>
<td></td>
</tr>
</tbody>
</table>

No Metadata Element Selected
This new version of MG (MG plus plus) was developed by the New Zealand Digital Library Project. It does word level indexing, which allows fielded, phrase and proximity searching to be handled by the indexer. Boolean searches can be ranked. Only a single index file is created for a Greenstone collection: document/section levels and text/metadata fields are all handled by one index file. For collections with many indexes, this results in a smaller collection size than using MG. For large collections, searching may be a bit slower due to the index being word level rather than section level.
This is the original indexer used by Greenstone, developed mainly by Alistair Moffat and described in the classic book Managing Gigabytes. It does section level indexing, and searches can be boolean or ranked (not both at once). For each index specified in the collection, a separate index file is created. For phrase searching, Greenstone does an "AND" search on all the terms, then scans the resulting hits to see if the phrase is present. It has been extensively tested on very large collections (many GB of text).
Lucene was developed by the Apache Software Foundation. It handles field and proximity searching, but only at a single level (e.g. complete documents or individual sections, but not both). Therefore document and section indexes for a collection require two separate index files. It provides a similar range of search functionality to MGPP with the addition of single-character wildcards, range searching and sorting of search results by metadata fields. It was added to Greenstone to facilitate incremental collection building, which MG and MGPP can't provide.
Format Features

VList 

HList 

DocumentHeading 

DocumentText 

DocumentButtons

SearchTypes "plain,form"

Choose Feature: All Features

Affected Component: VList

HTML Format String:

<td align="top">[link][icon][/link]</td>
<td align="top">[ex.srclink]{Or}{[ex.thumbicon],[ex.srcicon]}[/link]</td>
<td align="top">[highlight]
{Or}{[dc.Title],[exp.Title],[ex.Title],Untitled}
[/highlight]
{If}{[ex.Source],<br><i>{[ex.Source]}</i>}</td>

Insert Variable...  Undo  Redo

Remove Format  Reset to Default
A (slightly) enhanced collection - Multimedia

Add plugin
- UnknownPlug, set to accept MIDI files

Add metadata
- for “browse” button (8 items)
- for image titles (14 titles)
- to correct misspelling (mistery) (1 item)

Add/modify classifiers
- modify to display dc.title or ex.title
- add one for “browse” button
- remove the one for filename
- add one for phrase index
- add regular expressions to clean up titles

Modify format statements
- show title only for cover images
- suppress text document icon for MP3/MIDI items
- make bookshelves show how many documents they contain

General
- assign collection icons
- assign icons for non-standard media types: lyrics, discography, etc

About this collection

Demonstration collection illustrating the use of heterogeneous documents. Source document are about The Beatles pop group in the following formats: HTML, TXT, JPEG, Word, PDF, MIDI, MP3, and MARC file formats.

How to find information in the dbbeat_large collection

There are 4 ways to find information in this collection:
Customization

- 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.
Customizing with macros

- Let you customize presentation
- Present pages in different languages
- Print variables into the page text
  (e.g. number of search hits)

- **Macro files**
  - stored in `greenstone2/macros` folder
  - each file defines one or more “packages”
    (A “package” is a group of macros)
  - loaded on startup
    (note difference between Local and Web Library)
  - listed in `etc/main.cfg`

- **Collection-specific macros**
  - Stored in `greenstone2/collect/mycol/macros/extra.dm`
  - Or include argument `[c=collectionname]` for each macro
Personalizing your home page

C:\Program Files\gs2\etc\main.cfg  change *home.dm* to *yourhome.dm*

Your own Greenstone home page

Search page for the demo collection  Click here
"About" page for the demo collection  Click here
Preferences page for the demo collection  Click here
Home page  Click here
Help page  Click here
Administration page  Click here
The Collector  Click here

http://localhost/gsdl/cgi-bin/library?e=p-00-00--------00-0-0prompt-1Document---1011--en-50---20-home
powerful searching capabilities, and offer rich and natural indexes, white browsing is based on various metadata; browsing facilities. Above all it must be easy to maintain support structures for both are created during the building and augment, which presents a significant challenge if any coercion. When new material appears it can be fully manual organization is involved, incorporated into the collection by rebuilding.

The alternative is to automate these activities through To address the exceptionally broad demands of digital software tools. But the broad scope of digital library libraries, the system is public and extensible. It is issued requirements makes this a daunting prospect. Ideally the
Collection configuration

• Collection configuration file determines content conversion, extraction and building of indexes and browsing structures
  – indexes, classifiers, plugins

• Presentation of search/browse results and collection interface is determined by “format” strings and “macros”
Documentation and help

- Available at: www.greenstone.org
  - Software
  - Demo collections
  - FAQ
  - Tutorial materials

- Documentation:

- Mailing lists:
  - Greenstone Users List
  - Greenstone Developers List
Documentation and help

Manuals on the CD-ROM (docs)

– Installer’s Guide (install.pdf, 36pp)
  Versions of Greenstone, installation procedure, Greenstone collections, setting up the web server, configuring your site, personalizing your installation

– User’s Guide (user.pdf, 90pp)
  Overview of Greenstone, using Greenstone collections, the collector, administration, software features, glossary of terms

– Developer’s Guide (develop.pdf, 113pp)
  Understanding the collection building process, getting the most out of your collections, the Greenstone runtime systems, configuring your Greenstone site

– From Paper To Collection (paper.pdf, 30pp)
  Scanners and scanning, OCR, 3 examples – from 1,000 to 100,000 pages, Creating an electronic collection
Documentation and help

- **greenstone.org**
  - Download: software and tutorials
  - Example collections
  - Documentation
  - FAQ: general info section
  - support (+ join mailing list)
  - Configuration files for nzdl.org collections

- **nzdl.org**
  - Documentation collections
  - Documented
  - example collections

- **greenstonesupport.iimk.ac.in**
  - Download: software and tutorials
  - Example collections
  - Documentation
  - support (+ join mailing list)
Mailing Lists

– **Greenstone Users List**
  For people installing and using standard Greenstone
  Join at: [https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-users](https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-users)
  Mail to: greenstone-users@list.scms.waikato.ac.nz

– **Greenstone Developers List**
  For people customizing their version of Greenstone
  Join at: [https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-devel](https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-devel)
  Mail to: greenstone-devel@list.scms.waikato.ac.nz

– **Greenstone Support for South Asia** [http://greenstonesupport.iimk.ac.in](http://greenstonesupport.iimk.ac.in)
  Mail to: greenstonesupport@iimk.ac.in

Mailing List Archives

A Greenstone collection of mail from both mailing lists
[http://www.nzdl.org/gsarchives](http://www.nzdl.org/gsarchives)
Journal Content Page Service

Search Content Pages

Begin Search

Browse the Content Pages of Journals

- Accounting World
- American Behavioral Scientist
- American Economic Journal Applied Economics
- American Economic Journal Economic Policy
- American Economic Journal Micro Economics
- Analyst
- Bulletin for International Taxation
- Business Barons
- Canadian Journal of Economics
- Case Folio
- Chartered Accountant
- Chartered Secretary
- Chartered Secretary India
- Communications Today
- Crisil Rating Scan
- CTJ
- Current Science

Other Services
- Databases
- E-Journals
- Online Services
- Digital Library
- Ask a Librarian
- Leadership Compass
- INiK Institutional Repository
- News Paper Clippings

Online Catalogues
- Books
- Journals
DL - Hardships

- Copyright Issues
- Technology Complexities
- Infrastructure Issues
- Publications/Formats – Diverse Datastreams
- Digital Objects/Formats - Multiple
- Publishers’ Policies – Stringent, Inconsistent
Major Tasks

- Content identification (internal / external)
- Content Creation
- Content Collation/Signposts
- Organisation
- Updation
- Retrieval / Dissemination
- User Training
- Archiving
DIGITAL LIBRARY ARCHITECTURE
Greenstone Support for South Asia has been launched

The Greenstone Digital Library Software (http://www.greenstone.org) is an Open Source package for building and distributing Digital Libraries, which has been developed by the University of Waikato, New Zealand. Greenstone is supported by the UNESCO and the Human Info NGO Belgium for spreading the benefits of this software to developing countries. Greenstone is a powerful and flexible software which is of great potential interest to libraries and information centers and other public and private institutions in South Asia.

This support effort is coordinated by the CDDL at IIM Kozhikode in collaboration with the Greenstone development team in order to ensure effective promotion and support for Greenstone in South Asia.

Workshops Coordinated by Greenstonesupport, IIMK
Acknowledgement

- Prof. Ian Witten, Director, Greenstone Digital Library Project, University of Waikato, New Zealand
- Team Greenstone, New Zealand
- Greenstone Support South Asia
- IIM Kozhikode, India
- UNESCO
Thank you for your participation. If you have any questions, please feel free to ask.