Installing, browsing, building

Course material prepared by

Greenstone Digital Library Project
University of Waikato, New Zealand

Agenda

- ❖ Documentation and help
- ❖ Installing pre-built collections
- ❖ Greenstone: platforms
- ❖ What's on the CD-ROM?
- ❖ Installing Greenstone
- ❖ Running Greenstone
- ❖ Collection building
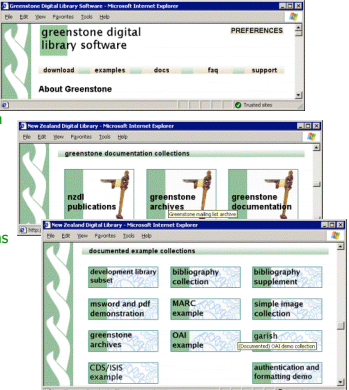- ❖ Downloading HTML
- ❖ WORD documents

Documentation and help

- ❖ Available at: www.greenstone.org
  - – Software
  - – Demo collections
  - – FAQ
  - – Tutorial materials
- ❖ Documentation:
  - – Installer's Guide, User's Guide, Developer's Guide, From Paper to Collection
- ❖ Mailing lists:
  - – Greenstone Users List
  - – Greenstone Developers List
- ❖ Greenstone Documentation Wiki

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

- ❖ 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

Mailing Lists
- – *Greenstone Users List*
  - For people installing and using standard Greenstone
  - Join at: 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
  - Mail to: greenstone-devel@list.scms.waikato.ac.nz

Mailing List Archives
- A Greenstone collection of mail from both mailing lists
- http://www.nzdl.org/gsarchives

Agenda

- ❖ Documentation and help
- → ❖ Installing pre-built collections
- ❖ Greenstone: platforms
- ❖ What's on the CD-ROM?
- ❖ Installing Greenstone
- ❖ Running Greenstone
- ❖ Collection building
- ❖ Downloading HTML
- ❖ WORD documents

Pre-built collection:
Installation demo

Installing pre-built collections

- ❖ Pre-built collections install on Windows only
- ❖ Installer will start automatically when the CD is put in the drive
- ❖ Option to install all the files, or leave the bulk on the CD
- ❖ New item is added to Start -> Programs

Agenda

- ❖ Documentation and help
- ❖ Installing pre-built collections
- → ❖ Greenstone: platforms
- ❖ What's on the CD-ROM?
- ❖ Installing Greenstone
- ❖ Running Greenstone
- ❖ Collection building
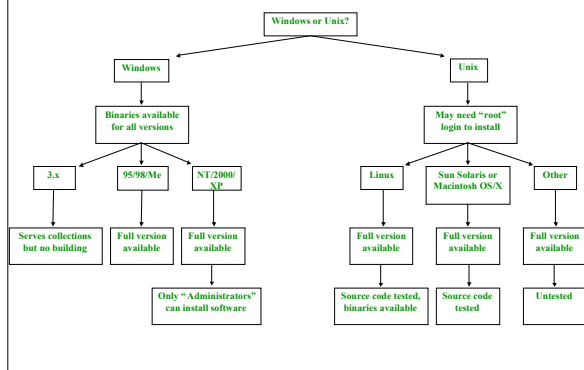- ❖ Downloading HTML
- ❖ WORD documents

Greenstone: Platforms

- ❖ Operating system:
  - Windows (any version)
  - Linux (any version)
  - Unix (most versions, e.g. Solaris)
  - Mac OS X
- ❖ Restrictions:
  - No longer installs on Windows 3.1/3.11
  - Need Java for Librarian interface (GLI)
    - ❖ No longer supported on Windows 95
- ❖ Disk space
  - 50 MB for a binary installation
  - 215 MB for the example collections (optional)
  - 5 MB for online documentation
  - 25 MB for "export to CD" function

Methods of Installation

- ❖ Install from binary (executable)
  - Windows
  - Linux
  - Mac OS X
- ❖ Install by compiling source code
  - Windows
  - Linux
  - Unix
  - Mac OS X

Installing Greenstone

Windows or Unix?

Windows

Unix

Binaries available for all versions

May need "root" login to install

3.x | 95/98/Me | NT/2000/ XP

Linux | Sun Solaris or Macintosh OS/X | Other

Serves collections but no building

Full version available

Full version available

Full version available

Full version available

Full version available

Only "Administrators" can install software

Source code tested, binaries available

Source code tested

Untested

Agenda

- ❖ Documentation and help
- ❖ Installing pre-built collections
- ❖ Greenstone: platforms
➡ ❖ What's on the CD-ROM?
- ❖ Installing Greenstone
- ❖ Running Greenstone
- ❖ Collection building
- ❖ Downloading HTML
- ❖ WORD documents

What's on the CD-ROM?

- ❖ Greenstone v2.70w
- ❖ Greenstone Librarian Interface v2.70w
- ❖ Greenstone Classic Interface Pack v2.63
- ❖ Workshop materials
- ❖ 14 documented example collections
- ❖ 4 manuals (Each in English, French, Spanish and Russian)
- ❖ Java (Required for running the GLI)
- ❖ Apache web server
- ❖ ImageMagick (Required for building image collections)
- ❖ Ghostscript (Required for advanced PDF/PPT conversion)

Example collections

Development Library Subset collection — Like the Greenstone Demo, this is a subset of the HDL -- but much larger. It contains 25 publications -- books, reports and magazines -- in various areas of human development (the full HDL contains 1,230 publications). It has the same structure as the Greenstone Demo. It's fairly complex, and if you're just starting out you might prefer to look at some other collections first (e.g. MSWord and PDF demonstration, the Greenstone Archives, or the simple image collection).

MSWord and PDF demonstration — This contains a few documents in PDF, MSWord, RTF, and Postscript formats, demonstrating the ability to build collections from documents in different formats. The collection configuration file is very simple.

Greenstone Archives collection — A collection of email messages from the Greenstone mailing list archives, this uses the Email plugin, which parses files in email formats. The collection configuration file is very simple.

Bibliography collection — With about 4,000 bibliography entries, this collection incorporates a form-based search interface that allows fielded searching. It is fairly complex.

Bibliography supplement — This tiny collection of 10 bibliography entries illustrates the "supercollection" facility which searches several collections together, seamlessly. It operates together with the bibliography collection, and its configuration file is almost the same.

MARC example — Based on some MARC records from the Library of Congress, this is a simple collection (and does not allow form-based searching).

OAI demo — Using the Open Archive Protocol and the Import-From feature, this retrieves metadata from an archive and builds a collection from the records. In this case they are images, so both the OAI and Image plugins are used.

Example collections

Simple image collection — This very basic image collection contains no text and no explicit metadata -- which makes it rather unrealistic. The configuration file is about as simple as you can get.

Paged image Example — This collection contains some newspapers from the Niupepa Maori Newspaper collection. Each newspaper consists of a series of scanned images and (optionally) the corresponding OCR text. PagedImgPlug is used to group the images into documents, and extended formatting is used to switch between image and text versions of each page.

Formatting and authentication demo — With the same material as the original Greenstone demo collection, this shows off two independent features: non-standard document formatting, and controlled access to the documents via user authentication.

Garish version of demo collection — This collection also contains the same material as the Greenstone demo. Its appearance has been altered to show how the pages generated can be set out differently. It relies on a non-standard macro file that is supplied with Greenstone.

Cascading Style Sheets Demo collection — This collection demonstrates Greenstone's use of Cascading Style Sheets (CSS) for visual formatting in web browsers. (Greenstone uses CSS instead of HTML tables since version 2.63.) On every page, you can change the style-sheet in effect, to modify that page's appearance.

CDS/ISIS example — This collection is built from a CDS/ISIS database of about 150 bibliography entries. It uses the ISISPlug plugin, which reads the standard ISIS .mst and .fdt files and converts them to Greenstone metadata.

DSpace example — This collection is based on records exported from a DSpace institutional repository.

Agenda

- ❖ Documentation and help
- ❖ Installing pre-built collections
- ❖ Greenstone: platforms
- ❖ What's on the CD-ROM?
➡ ❖ Installing Greenstone
- ❖ Running Greenstone
- ❖ Collection building
- ❖ Downloading HTML
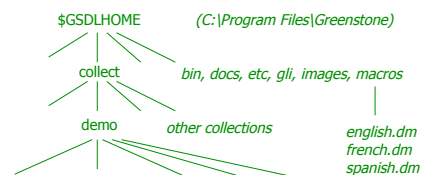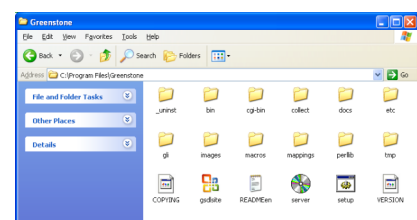- ❖ WORD documents

Local library:
Installation demo

Installing Greenstone

❖ Installation starts automatically when CD is inserted
  – If it doesn't, double-click "My Computer" on the desktop and double-click again on the CD-ROM drive
❖ 10 mouse clicks and you're done!

Local library vs Web library

❖ **Local library: stand-alone**
  – **Serves collections on a standalone PC ...**
  – **...and on others on the same network**
  – **Includes built-in Web server**
❖ **Web library: uses external web server**
  – **Apache, Microsoft PWS/IIS**

❖ **Windows: Both local library and web library**
  – **All versions: (3.1, 3.11), 95, 98, NT, 2000, ME, XP**
  – **Binaries supplied**
  – **Normally use local library (else must set up server)**
  – **Web library works with Microsoft PWS, IIS**
❖ **Unix, Mac OS X: Web library only**
  – **Use Apache (or other web server)**
  – **Linux binaries supplied**
  – **Tested on SUN Solaris, Mac OS X**
  – **Need GDBM (standard on Linux)**

What's on my hard drive?

$GSDLHOME   (C:\Program Files\Greenstone)

collect   bin, docs, etc, gli, images, macros

demo   other collections   english.dm
                                      french.dm
                                      spanish.dm

Agenda

❖ Documentation and help
❖ Installing pre-built collections
❖ Greenstone: platforms
❖ What's on the CD-ROM?
❖ Installing Greenstone
➜ ❖ Running Greenstone
❖ Collection building
❖ Downloading HTML
❖ WORD documents

Running Greenstone

❖ A new menu is added to the Programs list in the Start menu:
  Start ➜ Programs ➜ Greenstone digital library software
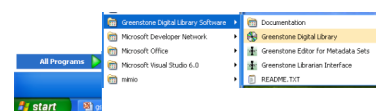
❖ Run Greenstone, GLI, and GEMS from here

Local Library Troubleshooting

Problem	Try this
When I start Greenstone my computer asks me to dial up my Internet Service Provider.	Push the *Cancel* button in the dialog box. This usually solves the problem.
When I point my browser at the digital library, it can't find that page.	Check your Internet Proxy settings and turn proxies off (use *Edit Preferences* on Netscape or *Internet Options* on Explorer).

Documentation and Help

❖ *Installer's Guide* (install.pdf)
  Versions of Greenstone, installation procedure, Greenstone collections, setting up the web server, configuring your site, personalizing your installation

❖ FAQ: sections on
  – Obtaining Greenstone
  – Installing Greenstone
  – Running Greenstone

Agenda

❖ Documentation and help
❖ Installing pre-built collections
❖ Greenstone: platforms
❖ What's on the CD-ROM?
❖ Installing Greenstone
❖ Running Greenstone
→ ❖ Collection building
❖ Downloading HTML
❖ WORD documents

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

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"

Collection building in GLI

❖ Start a new collection
    File -> New
❖ Gather some documents
❖ Enrich with metadata
❖ Design (configure) the collection
❖ Create (build) the collection

Greenstone Librarian Interface demo: building a simple collection

Agenda

❖ Documentation and help
❖ Installing pre-built collections
❖ Greenstone: platforms
❖ What's on the CD-ROM?
❖ Installing Greenstone
❖ Running Greenstone
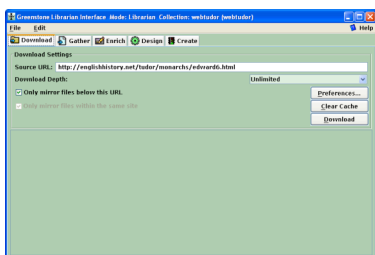❖ Collection building
➡ ❖ Downloading HTML
❖ WORD documents

Downloading in GLI

❖ Can download, or "mirror", web pages and web sites to local disk

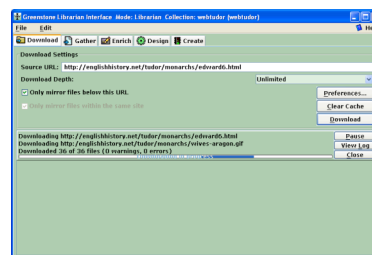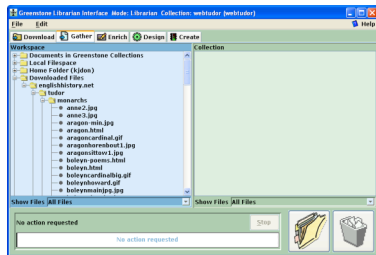❖ Options: within URL, within site, depth of links to follow

❖ Can be added into collection

Download panel

Setting up a download

Downloading in progress

Downloaded files

Behind a firewall?

If you are behind a firewall or proxy server
then you need to set this information in
File -> Preferences -> Connection
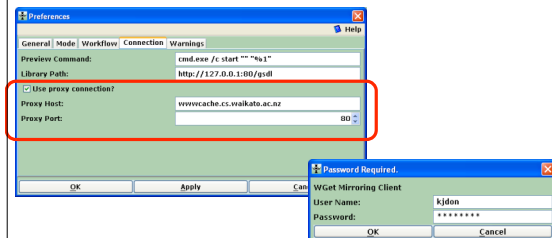
Downloaded files

❖ File hierarchy preserves site structure

❖ `file_is_url` option to HTMLPlug adds URL
metadata based on the file hierarchy

❖ `[weblink][webicon][/weblink]` links to
original if URL metadata has been set

❖ So you can download web sites to index,
then link back to the originals

Agenda
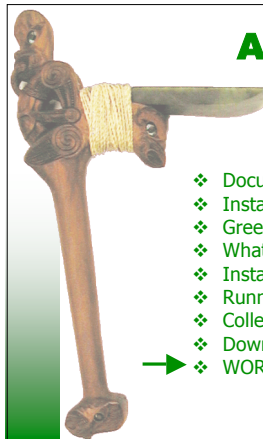
❖ Documentation and help
❖ Installing pre-built collections
❖ Greenstone: platforms
❖ What's on the CD-ROM?
❖ Installing Greenstone
❖ Running Greenstone
❖ Collection building
❖ Downloading HTML
➔ ❖ WORD documents

Word Document

❖   Word conversions in Greenstone

1. Text
    ❖ Unix *strings* command
    ❖ `use_strings` option

2. Flat format HTML => wvWare

3. Styled format HTML => VB script
    ❖ `windows_scripting` option
    ❖ Heading setting
       <Heading 1>, <Heading 2>, <Heading 3>……
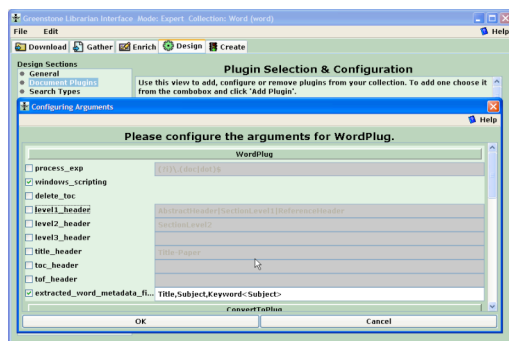   User-defined heading style

Word - Text
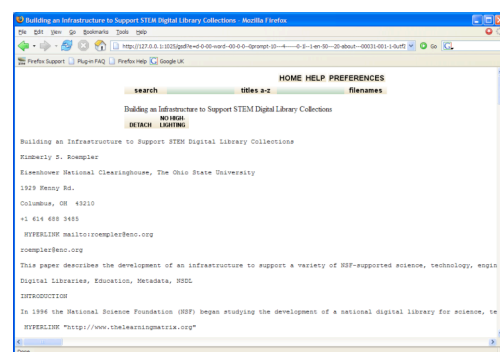
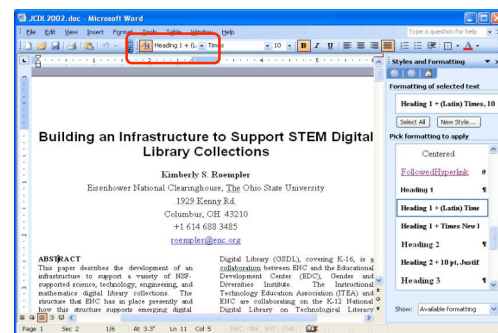## Word - HTML (wvWare)
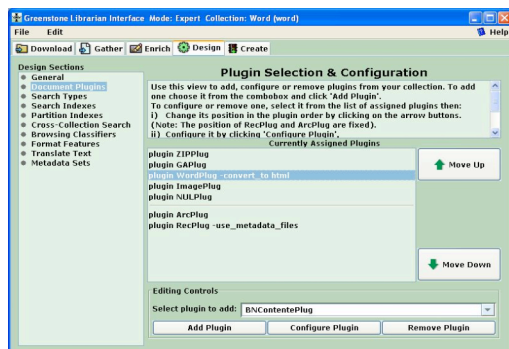
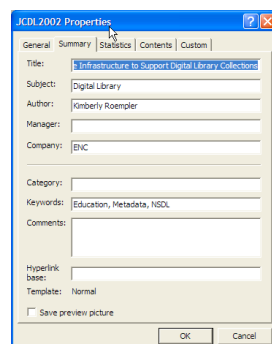## Word: Flat HTML format
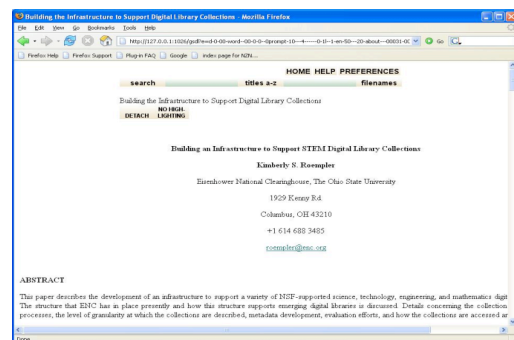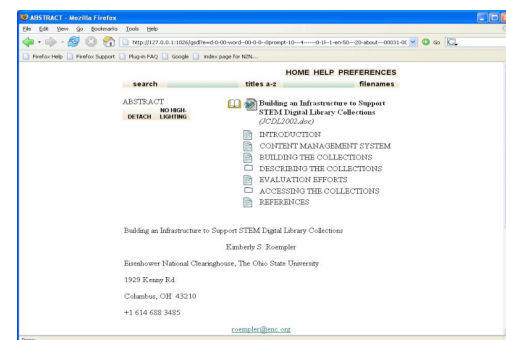

Word - HTML (Windows Scripting)

Word Document

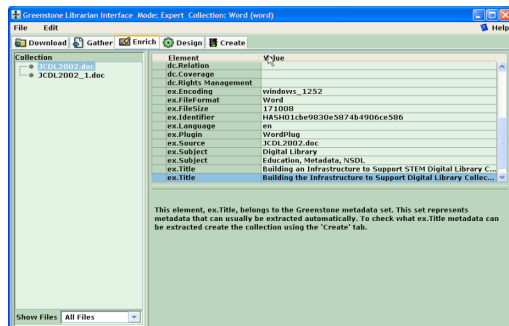## Word Document Properties

File -> Properties

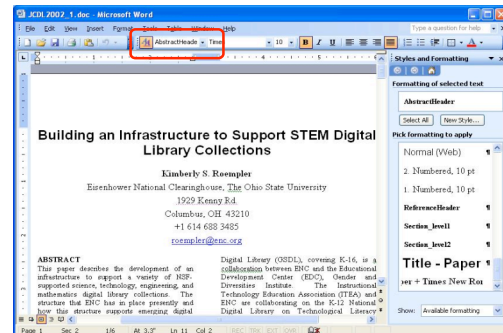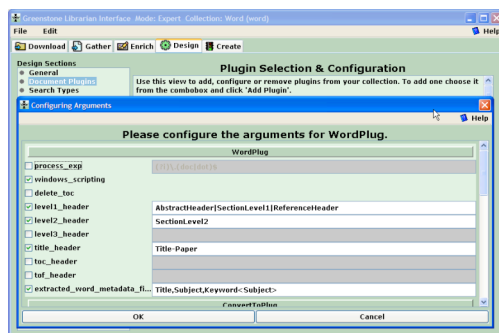## Word: Hierarchical HTML format

Extracted Word Document Properties

User-defined Style Formatting

WordPlug – User-defined Style

Word: Hierarchical HTML Format