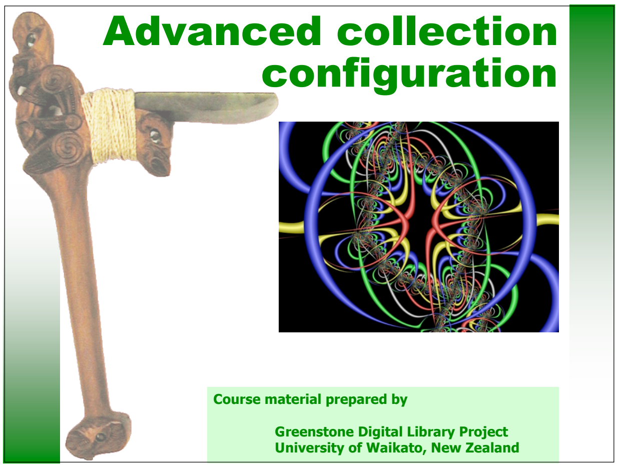


Advanced collection configuration

Course material prepared by

Greenstone Digital Library Project
University of Waikato, New Zealand

Agenda

- Formatting
- Formatting extended metadata
- Changing metadata sets
- Collection configuration file
- GLI modes
- PDF documents
- PPT documents
- Exploding metadata databases
- Cross collection searching
- Translate text
- Full-text tagging
- Creating a CD-ROM

Format Statements: Components

HTML	
`[Text]`	*displays document text*
`[Title],[Howto] …`	*displays metadata*
`[link] … [/link]`	*links to document*
`[srclink] … [/srclink]`	*links to original file*
`[icon], [srcicon]`	*page/book/bookshelf* *source icons*
If and *Or* statements	*conditional processing*

Format Statements

`{If}{test, if true, if false}`

- Test can be:

`[metadata]`	exists
`[metadata] eq ‘value’`	equals
`[metadata] ne ‘value’`	not equals

- Examples

```
{If}{[ex.Source],(<i>[ex.Source]</i>)}
{If}{[numleafdocs],[ex.Title],[dc.Creator]}
{If}{[ex.FileFormat] eq PDF, [srclink] PDF
  document [/srclink]}
```

Format Statements

`{Or}{[metadata1],[metadata2],…`

- Chooses the first metadata that exists
- Last item can be plain text
- Examples

```
{Or}{[dls.Title],[dc.Title],[ex.Title],Untitled}
{Or}{[ex.thumbicon],[ex.srcicon]}
```

HTML page

```
<html>
  <head>
    <title>The page title</title>
  </head>
  <body formatting-attributes>
    Page content goes here
  </body>
</html>
```

HTML elements

`<b>Bold</b>, <i>italics</i>, <u>underline</u>`

`<br/> a line break`

`<p>a paragraph</p>`

`<table><tr><td>cell content</td></tr></table>`

`<a href="link address">link text</a>`

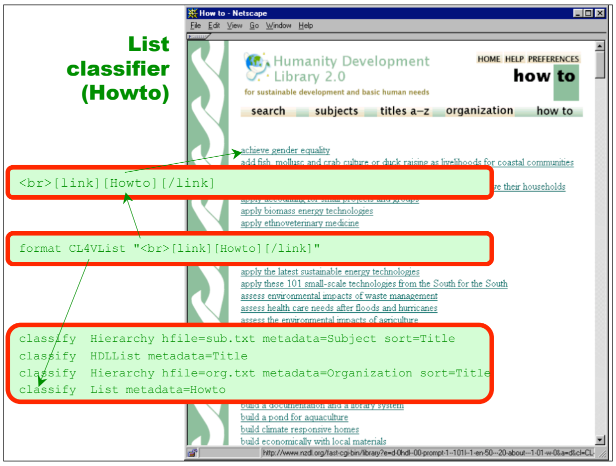
`<img src="path to image"/>`

Format statements

- Defaults provided to give reasonable results for most cases
- Interpreted at run-time – no need to rebuild
- More information: FAQ

  http://www.greenstone.org/cgi-bin/library?a=p&p=faqcustomize#customizeformat

List classifier (Howto)


How to - Netscape
File Edit View Go Window Help
Humanity Development Library 2.0
for sustainable development and basic human needs
HOME HELP PREFERENCES
how to
search
subjects
titles a–z
organization
how to

achieve gender equality
add fish, mollusc and crab culture or duck raising as livelihoods for coastal communities
…ve their households
apply accounting for small projects and groups
apply biomass energy technologies
apply ethnoveterinary medicine
apply the latest sustainable energy technologies
apply these 101 small-scale technologies from the South for the South
assess environmental impacts of waste management
assess health care needs after floods and hurricanes
assess the environmental impacts of agriculture
build a documentation and a library system
build a pond for aquaculture
build climate responsive homes
build economically with local materials

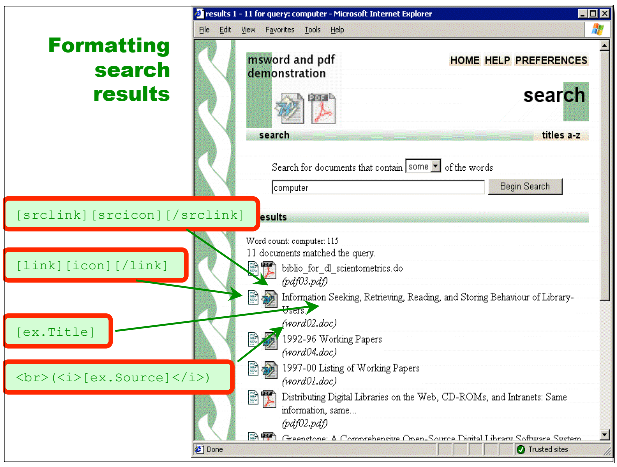
http://www.nzdl.org/fast-cgi-bin/library?e=d-0hdl--00-prompt-1--101l--1-en-50---20-about---1-01-w-0&a=d&cl=CL

```
<br>[link][Howto][/link]
```

```
format CL4VList "<br>[link][Howto][/link]"
```

```
classify  Hierarchy hfile=sub.txt metadata=Subject sort=Title
classify  HDLList metadata=Title
classify  Hierarchy hfile=org.txt metadata=Organization sort=Title
classify  List metadata=Howto
```

Formatting search results


results 1 - 11 for query: computer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
msword and pdf demonstration
HOME HELP PREFERENCES
search
search
titles a-z
Begin Search
esults
Done
Trusted sites

Search for documents that contain some of the words

computer

Word count: computer: 115
11 documents matched the query.
biblio_for_dl_scientometrics.do
(pdf03.pdf)
Information Seeking, Retrieving, Reading, and Storing Behaviour of Library-Users.
(word02.doc)
1992-96 Working Papers
(word04.doc)
1997-00 Listing of Working Papers
(word01.doc)
Distributing Digital Libraries on the Web, CD-ROMs, and Intranets: Same information, same...
(pdf02.pdf)
Greenstone: A Comprehensive Open-Source Digital Library Software System

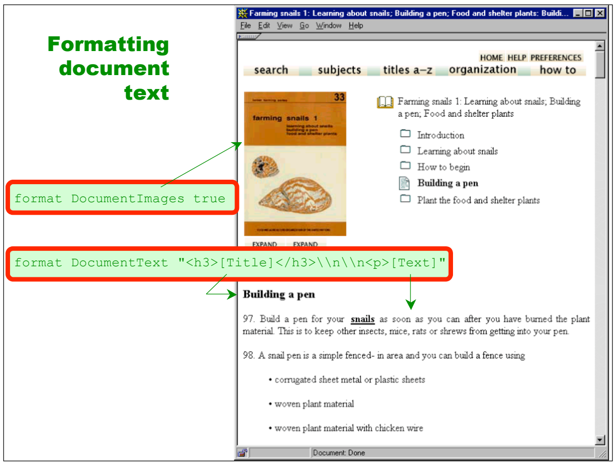
```
[srclink][srcicon][/srclink]
```

```
[link][icon][/link]
```

```
[ex.Title]
```

```
<br>(<i>[ex.Source]</i>)
```

Formatting document text


Farming snails 1: Learning about snails; Building a pen; Food and shelter plants: Buildi...
File Edit View Go Window Help
HOME HELP PREFERENCES
search
subjects
titles a–z
organization
how to
33
farming snails 1
EXPAND
EXPAND
Document: Done

Farming snails 1: Learning about snails; Building a pen; Food and shelter plants

- Introduction
- Learning about snails
- How to begin
- **Building a pen**
- Plant the food and shelter plants

Building a pen

97. Build a pen for your **snails** as soon as you can after you have burned the plant material. This is to keep other insects, mice, rats or shrews from getting into your pen.

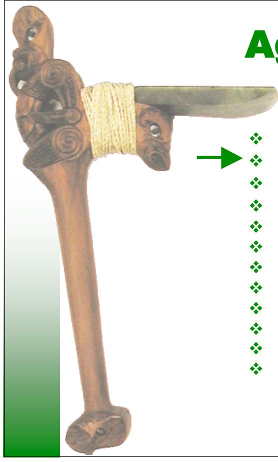
98. A snail pen is a simple fenced- in area and you can build a fence using

- corrugated sheet metal or plastic sheets
- woven plant material
- woven plant material with chicken wire

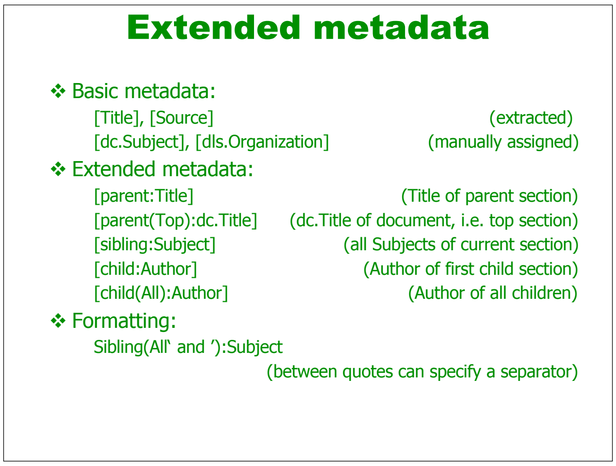
```
format DocumentImages true
```

```
format DocumentText "<h3>[Title]</h3>\\n\\n<p>[Text]"
```

Agenda



- Formatting
- Formatting extended metadata
- Changing metadata sets
- Collection configuration file
- GLI modes
- PDF documents
- PPT documents
- Exploding metadata databases
- Cross collection searching
- Translate text
- Full-text tagging
- Creating a CD-ROM



Extended metadata

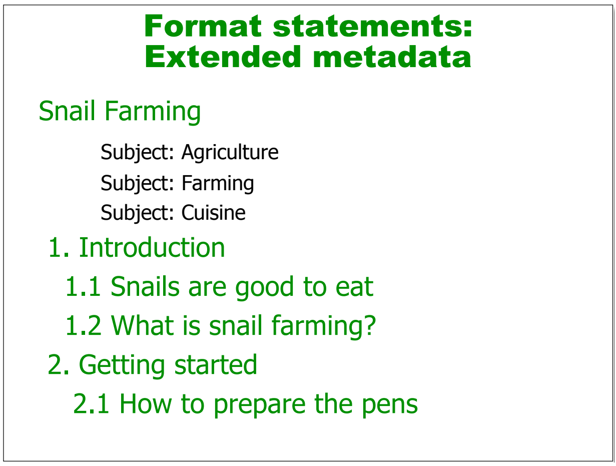
- Basic metadata:

[Title], [Source]	(extracted)
[dc.Subject], [dls.Organization]	(manually assigned)

- Extended metadata:

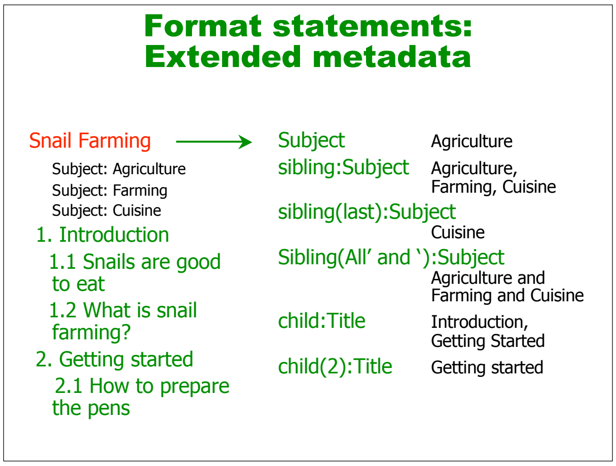
[parent:Title]	(Title of parent section)
[parent(Top):dc.Title]	(dc.Title of document, i.e. top section)
[sibling:Subject]	(all Subjects of current section)
[child:Author]	(Author of first child section)
[child(All):Author]	(Author of all children)

- Formatting:
  Sibling(All‘ and ’):Subject
  (between quotes can specify a separator)



Format statements: Extended metadata

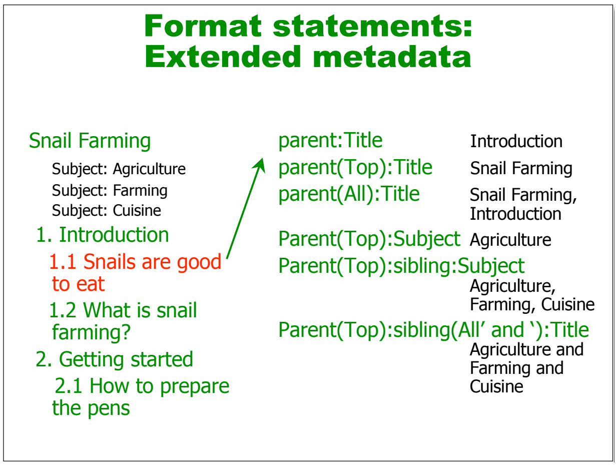
Snail Farming
  Subject: Agriculture
  Subject: Farming
  Subject: Cuisine
1. Introduction
  1.1 Snails are good to eat
  1.2 What is snail farming?
2. Getting started
  2.1 How to prepare the pens



Format statements: Extended metadata

Snail Farming →
  Subject: Agriculture
  Subject: Farming
  Subject: Cuisine
1. Introduction
  1.1 Snails are good to eat
  1.2 What is snail farming?
2. Getting started
  2.1 How to prepare the pens

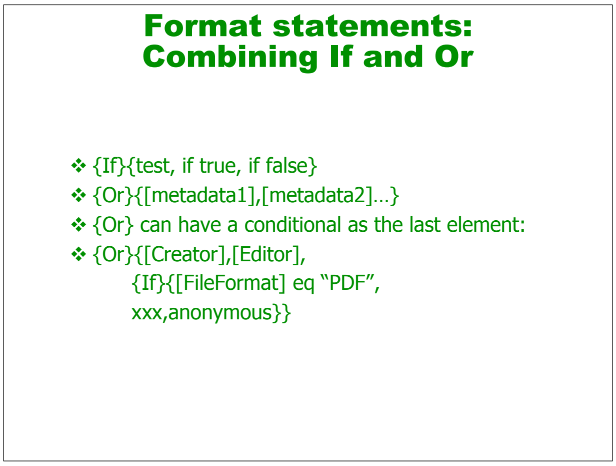
Subject	Agriculture
sibling:Subject	Agriculture, Farming, Cuisine
sibling(last):Subject	Cuisine
Sibling(All’ and ‘):Subject	Agriculture and Farming and Cuisine
child:Title	Introduction, Getting Started
child(2):Title	Getting started



Format statements: Extended metadata

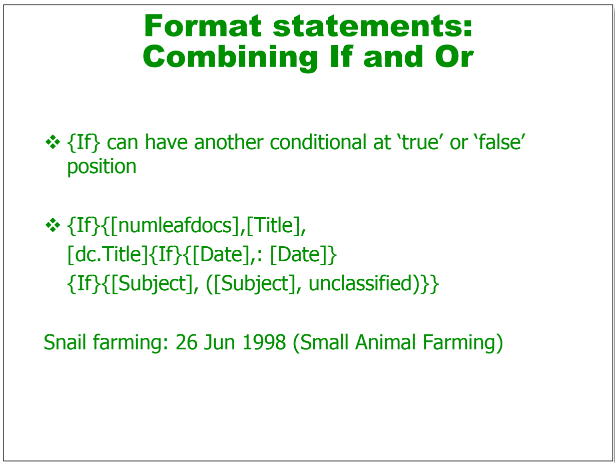
Snail Farming
  Subject: Agriculture
  Subject: Farming
  Subject: Cuisine
1. Introduction
  1.1 Snails are good to eat
  1.2 What is snail farming?
2. Getting started
  2.1 How to prepare the pens

parent:Title	Introduction
parent(Top):Title	Snail Farming
parent(All):Title	Snail Farming, Introduction
Parent(Top):Subject	Agriculture
Parent(Top):sibling:Subject	Agriculture, Farming, Cuisine
Parent(Top):sibling(All’ and ‘):Title	Agriculture and Farming and Cuisine



Format statements: Combining If and Or

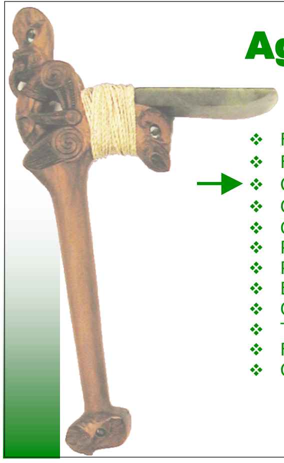
- {If}{test, if true, if false}
- {Or}{[metadata1],[metadata2]…}
- {Or} can have a conditional as the last element:
- {Or}{[Creator],[Editor],
  {If}{[FileFormat] eq “PDF”,
  xxx,anonymous}}



Format statements: Combining If and Or

- {If} can have another conditional at ‘true’ or ‘false’ position
- {If}{[numleafdocs],[Title],
  [dc.Title]{If}{[Date],: [Date]}
  {If}{[Subject], ([Subject], unclassified)}}

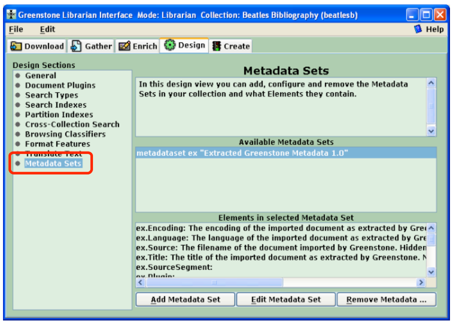
Snail farming: 26 Jun 1998 (Small Animal Farming)



Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- → ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

Metadata Sets



Greenstone Librarian Interface Mode: Librarian Collection: Beatles Bibliography (beatlesb)

Design Sections
- General
- Document Plugins
- Search Types
- Search Indexes
- Partition Indexes
- Cross-Collection Search
- Browsing Classifiers
- Format Features
- Translate Text
- Metadata Sets

Metadata Sets

In this design view you can add, configure and remove the Metadata Sets in your collection and what Elements they contain.

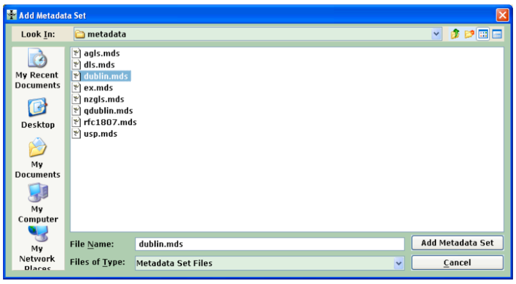
Available Metadata Sets

metadataset ex "Extracted Greenstone Metadata 1.0"

Elements in selected Metadata Set

ex.Encoding: The encoding of the imported document as extracted by Gree
ex.Language: The language of the imported document as extracted by Gre
ex.Source: The filename of the document imported by Greenstone. Hidden
ex.Title: The title of the imported document as extracted by Greenstone. N
ex.SourceSegment:

Add Metadata Set



Add Metadata Set

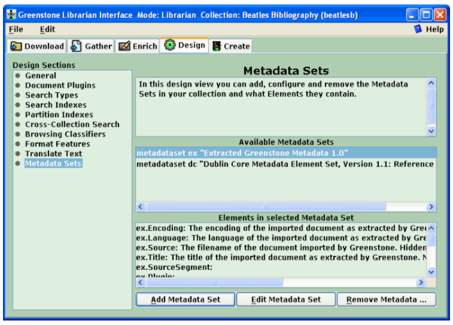
Look In: metadata

- agls.mds
- dls.mds
- dublin.mds
- ex.mds
- nzgls.mds
- qdublin.mds
- rfc1807.mds
- usp.mds

File Name: dublin.mds

Files of Type: Metadata Set Files

Metadata Sets



Greenstone Librarian Interface Mode: Librarian Collection: Beatles Bibliography (beatlesb)

Design Sections
- General
- Document Plugins
- Search Types
- Search Indexes
- Partition Indexes
- Cross-Collection Search
- Browsing Classifiers
- Format Features
- Translate Text
- Metadata Sets

Metadata Sets

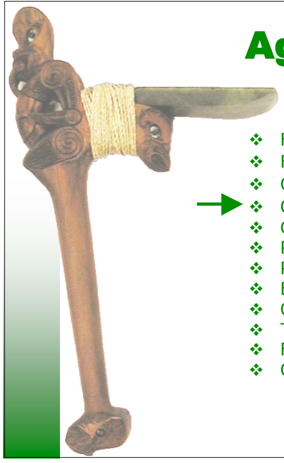
In this design view you can add, configure and remove the Metadata Sets in your collection and what Elements they contain.

Available Metadata Sets

metadataset ex "Extracted Greenstone Metadata 1.0"
metadataset dc "Dublin Core Metadata Element Set, Version 1.1: Reference

Elements in selected Metadata Set

ex.Encoding: The encoding of the imported document as extracted by Gree
ex.Language: The language of the imported document as extracted by Gre
ex.Source: The filename of the document imported by Greenstone. Hidden
ex.Title: The title of the imported document as extracted by Greenstone. N
ex.SourceSegment:



Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- → ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

Collection configuration file

- ❖**name, icon, etc**
- ❖**description**
- ❖**email of creator**
- ❖**search indexes**
- ❖**plugins**
- ❖**classifiers**

how to format

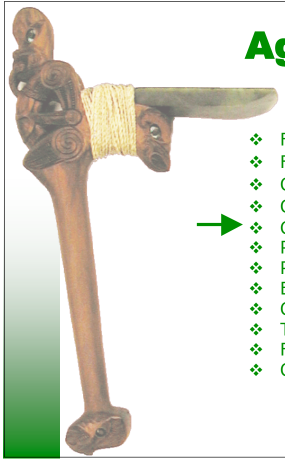
- ❖**documents**
- ❖**query results**
- ❖**classifiers**

```
creator        sjboddie@cs.waikato.ac.nz
maintainer     sjboddie@cs.waikato.ac.nz
public         true
beta           true

indexes        section:text section:Title document:text
defaultindex   section:text

plugin         GAPlug
plugin         ArcPlug
plugin         RecPlug

classify       Hierarchy hfile=sub.txt metadata=Subject sort=Title
classify       HDLList metadata=Title
classify       Hierarchy hfile=org.txt metadata=Organization sort=Title
classify       List metadata=Howto

format         SearchVList "<td valign=top>[link][icon][/link]</td>
               <td>{If}{[parent(All': '):Title],[parent(All': '):Title]: }
               [link][Title][/link]</td>"
format         CL4VList     "<br>[link][Howto][/link]"
format         DocumentImages true
format         DocumentText "<h3>[Title]</h3>\\n\\n<p>[Text]"

collectionmeta collectionname  "greenstone demo"
collectionmeta collectionextra "This is a demonstration collection for the
               Greenstone digital library software.\nIt contains a small
               subset (11 books) of the Humanity Development Library"
collectionmeta iconcollectionsmall "/gsdl/collect/demo/images/demosm.gif"
collectionmeta iconcollection      "/gsdl/collect/demo/images/demo.gif"
collectionmeta .section:Title      "section titles"
collectionmeta .document:text      "entire books"
collectionmeta .section:text       "chapters"
```

Documentation and Help

- *User's Guide* (user.pdf)

  Includes substantial sections on the GLI (Sections 3.1 and 3.2, 36 pp)
- Tooltips
- On-line help

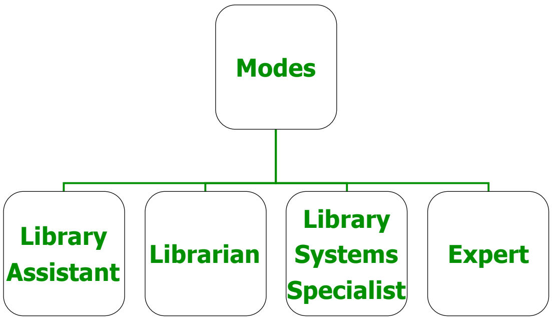


Agenda

- Formatting
- Formatting extended metadata
- Changing metadata sets
- Collection configuration file
- → GLI modes
- PDF documents
- PPT documents
- Exploding metadata databases
- Cross collection searching
- Translate text
- Full-text tagging
- Creating a CD-ROM

GLI Modes

- **File->Preferences->Mode**



Modes

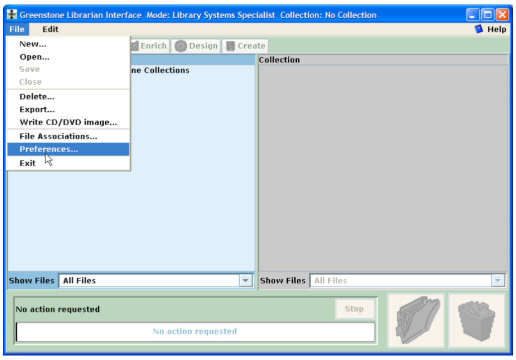
Library Assistant

Librarian

Library Systems Specialist

Expert

GLI Mode Setting

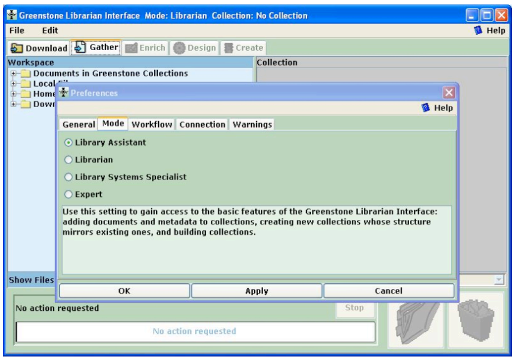


Greenstone Librarian Interface Mode: Library Systems Specialist Collection: No Collection

No action requested

No action requested

GLI Mode Setting



Greenstone Librarian Interface Mode: Librarian Collection: No Collection

● Library Assistant
○ Librarian
○ Library Systems Specialist
○ Expert

Use this setting to gain access to the basic features of the Greenstone Librarian Interface: adding documents and metadata to collections, creating new collections whose structure mirrors existing ones, and building collections.

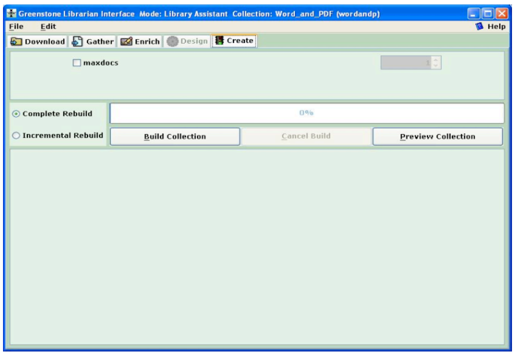
No action requested

No action requested

GLI Modes

- **Library Assistant**
  - Access to basic features: creating new collections; adding documents and metadata; building collections
  - No design functions are available
- **Librarian**
  - Basic features + Design

Library Assistant Mode



Greenstone Librarian Interface Mode: Library Assistant Collection: Word_and_PDF (wordandp)

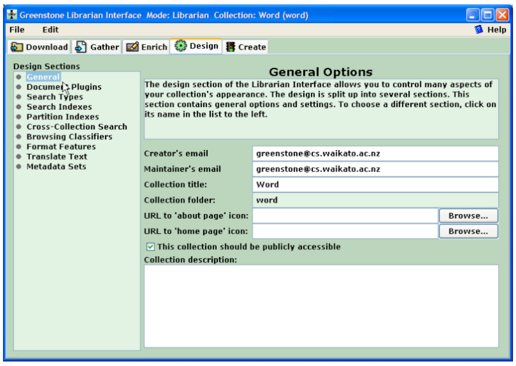
☐ maxdocs 1

● Complete Rebuild

○ Incremental Rebuild

0%

Librarian Mode



Greenstone Librarian Interface Mode: Librarian Collection: Word (word)

Design Sections
- General
- Document Plugins
- Search Types
- Search Indexes
- Partition Indexes
- Cross-Collection Search
- Browsing Classifiers
- Format Features
- Translate Text
- Metadata Sets

General Options

The design section of the Librarian Interface allows you to control many aspects of your collection's appearance. The design is split up into several sections. This section contains general options and settings. To choose a different section, click on its name in the list to the left.

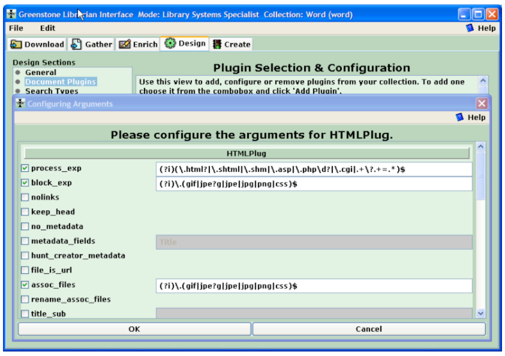
Creator's email: greenstone@cs.waikato.ac.nz
Maintainer's email: greenstone@cs.waikato.ac.nz
Collection title: Word
Collection folder: word
URL to 'about page' icon:
URL to 'home page' icon:
☑ This collection should be publicly accessible
Collection description:

GLI Modes

- ❖ **Library Systems Specialist**
  - Full use of GLI
  - Formulate regular expressions to make use of formatting features
    - ❖ For example: HTML block expression

      `q^(?i)\.(gif|jpe?g|jpe|png|css)$^`
  - Partial options for import and building

Library Systems Specialist Mode



Greenstone Librarian Interface Mode: Library Systems Specialist Collection: Word (word)

Design Sections
- General
- Document Plugins
- Search Types

Plugin Selection & Configuration

Use this view to add, configure or remove plugins from your collection. To add one choose it from the combobox and click 'Add Plugin'.

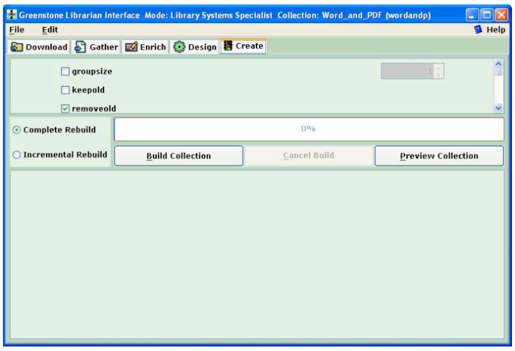
Configuring Arguments

Please configure the arguments for HTMLPlug.

HTMLPlug

☑ process_exp `(?i)(\.html?|\.shtml|\.shm|\.asp|\.php\d?|\.cgi|.+\?.+=.*)$`
☑ block_exp `(?i)\.(gif|jpe?g|jpe|jpg|png|css)$`
☐ nolinks
☐ keep_head
☐ no_metadata
☐ metadata_fields Title
☐ hunt_creator_metadata
☐ file_is_url
☑ assoc_files `(?i)\.(gif|jpe?g|jpe|jpg|png|css)$`
☐ rename_assoc_files
☐ title_sub

Library Systems Specialist Mode



Greenstone Librarian Interface Mode: Library Systems Specialist Collection: Word_and_PDF (wordandp)

☐ groupsize 1
☐ keepold
☑ removeold

● Complete Rebuild

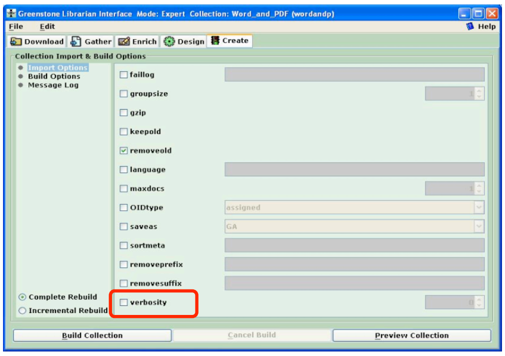
○ Incremental Rebuild

0%

GLI Modes

- ❖ **Expert**
  - All features are enabled
  - Recommended for experienced users
  - Perform troubleshooting tasks
  - Options to control the import and build processes
  - Shows the output from the processes

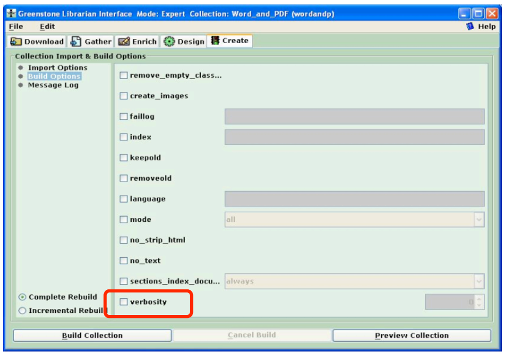
Expert Mode—Import Options



- ☐ failog
- ☐ groupsize 1
- ☐ gzip
- ☐ keepold
- ☑ removeold
- ☐ language
- ☐ maxdocs 1
- ☐ OIDtype assigned
- ☐ saveas GA
- ☐ sortmeta
- ☐ removeprefix
- ☐ removesuffix
- ☐ verbosity 0

● Complete Rebuild
○ Incremental Rebuild

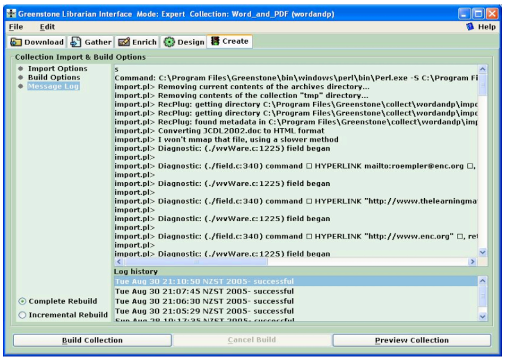
Expert Mode—Build Options



- ☐ remove_empty_class...
- ☐ create_images
- ☐ failog
- ☐ index
- ☐ keepold
- ☐ removeold
- ☐ language
- ☐ mode all
- ☐ no_strip_html
- ☐ no_text
- ☐ sections_index_docu... always
- ☐ verbosity 0

● Complete Rebuild
○ Incremental Rebuil

Message Log—Expert Mode

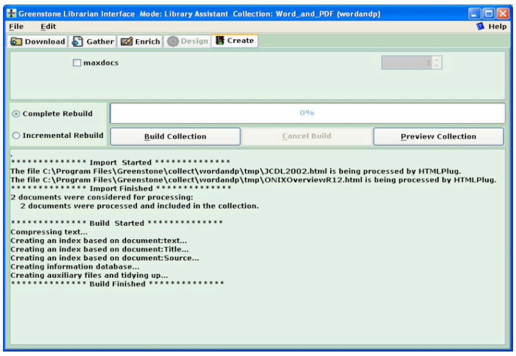


```
s
Command: C:\Program Files\Greenstone\bin\windows\perl\bin\Perl.exe -S C:\Program Fi
import.pl> Removing current contents of the archives directory...
import.pl> Removing contents of the collection "tmp" directory...
import.pl> RecPlug: getting directory C:\Program Files\Greenstone\collect\wordandp\impo
import.pl> RecPlug: getting directory C:\Program Files\Greenstone\collect\wordandp\impo
import.pl> RecPlug: found metadata in C:\Program Files\Greenstone\collect\wordandp\imp
import.pl> Converting JCDL2002.doc to HTML format
import.pl> I won't mmap that file, using a slower method
import.pl> Diagnostic: (./wvWare.c:1225) field began
import.pl>
import.pl> Diagnostic: (./field.c:340) command □ HYPERLINK mailto:roempler@enc.org □,
import.pl>
import.pl> Diagnostic: (./wvWare.c:1225) field began
import.pl>
import.pl> Diagnostic: (./field.c:340) command □ HYPERLINK "http://www.thelearningma
import.pl>
import.pl> Diagnostic: (./wvWare.c:1225) field began
import.pl>
import.pl> Diagnostic: (./field.c:340) command □ HYPERLINK "http://www.enc.org" □, re
import.pl>
import.pl> Diagnostic: (./wvWare.c:1225) field began
```

Log history

```
Tue Aug 30 21:10:50 NZST 2005- successful
Tue Aug 30 21:07:45 NZST 2005- successful
Tue Aug 30 21:06:30 NZST 2005- successful
Tue Aug 30 21:05:29 NZST 2005- successful
Sun Aug 28 10:17:35 NZST 2005- successful
```

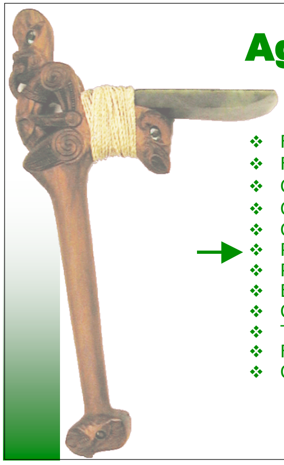
Building Message—Others



☐ maxdocs 1

● Complete Rebuild
○ Incremental Rebuild

```
.
************** Import  Started **************
The file C:\Program Files\Greenstone\collect\wordandp\tmp\JCDL2002.html is being processed by HTMLPlug.
The file C:\Program Files\Greenstone\collect\wordandp\tmp\ONIXOverviewR12.html is being processed by HTMLPlug.
************** Import Finished **************
2 documents were considered for processing:
  2 documents were processed and included in the collection.

************** Build  Started **************
Compressing text...
Creating an index based on document:text...
Creating an index based on document:Title...
Creating an index based on document:Source...
Creating information database...
Creating auxiliary files and tidying up...
************** Build Finished **************
```



Agenda

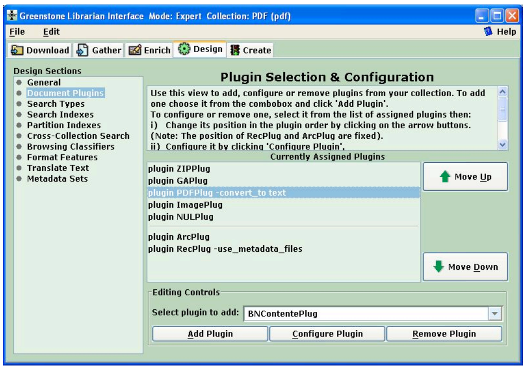
- Formatting
- Formatting extended metadata
- Changing metadata sets
- Collection configuration file
- GLI modes
- → PDF documents
- PPT documents
- Exploding metadata databases
- Cross collection searching
- Translate text
- Full-text tagging
- Creating a CD-ROM

PDF Document

- PDF conversions in Greenstone
  1. Text only for Unix system
  2. HTML
     - use_sections option
     - complex option
  3. Image
     - ImageMagick needs to be installed
     - For advanced conversions, GhostScript must be installed
     - Use of convert utility
     - Convert_to

       pagedimg_jpg
       pagedimg_gif
       pagedimg_png

PDF -> Text



Plugin Selection & Configuration

Use this view to add, configure or remove plugins from your collection. To add one choose it from the combobox and click 'Add Plugin'.
To configure or remove one, select it from the list of assigned plugins then:
i) Change its position in the plugin order by clicking on the arrow buttons.
(Note: The position of RecPlug and ArcPlug are fixed).
ii) Configure it by clicking 'Configure Plugin',

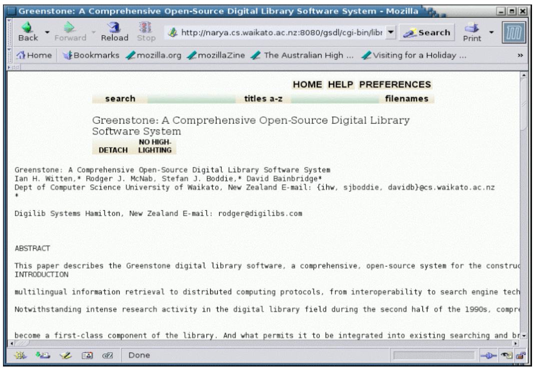
Currently Assigned Plugins

plugin ZIPPlug
plugin GAPlug
plugin PDFPlug -convert_to text
plugin ImagePlug
plugin NULPlug

plugin ArcPlug
plugin RecPlug -use_metadata_files

Select plugin to add: BNContentePlug

PDF: Text Document Display

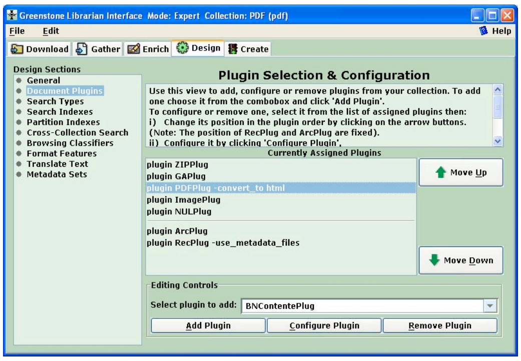


http://narya.cs.waikato.ac.nz:8080/gsdl/cgi-bin/libr

Greenstone: A Comprehensive Open-Source Digital Library Software System

```
Greenstone: A Comprehensive Open-Source Digital Library Software System
Ian H. Witten,* Rodger J. McNab, Stefan J. Boddie,* David Bainbridge*
Dept of Computer Science University of Waikato, New Zealand E-mail: {ihw, sjboddie, davidb}@cs.waikato.ac.nz
*

Digilib Systems Hamilton, New Zealand E-mail: rodger@digilibs.com



ABSTRACT

This paper describes the Greenstone digital library software, a comprehensive, open-source system for the construc
INTRODUCTION

multilingual information retrieval to distributed computing protocols, from interoperability to search engine tech

Notwithstanding intense research activity in the digital library field during the second half of the 1990s, compre


become a first-class component of the library. And what permits it to be integrated into existing searching and br
```

PDF -> HTML



Plugin Selection & Configuration

Use this view to add, configure or remove plugins from your collection. To add one choose it from the combobox and click 'Add Plugin'.
To configure or remove one, select it from the list of assigned plugins then:
i) Change its position in the plugin order by clicking on the arrow buttons.
(Note: The position of RecPlug and ArcPlug are fixed).
ii) Configure it by clicking 'Configure Plugin',

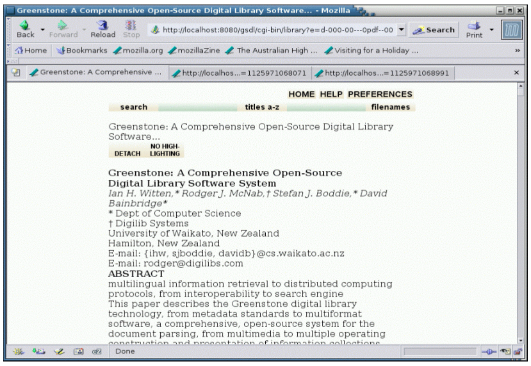
Currently Assigned Plugins

plugin ZIPPlug
plugin GAPlug
plugin PDFPlug -convert_to html
plugin ImagePlug
plugin NULPlug

plugin ArcPlug
plugin RecPlug -use_metadata_files

Select plugin to add: BNContentePlug

PDF: HTML Document Display 1

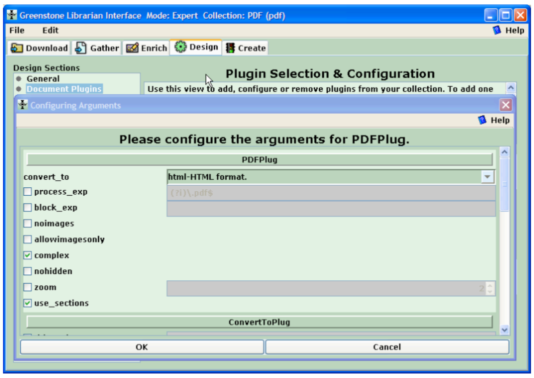

Greenstone: A Comprehensive Open-Source Digital Library Software... - Mozilla
Back
Forward
Reload
Stop
Search
Print
Home
Bookmarks
mozilla.org
mozillaZine
The Australian High ...
Visiting for a Holiday ...
Greenstone: A Comprehensive ...
http://localhos...=1125971068071
http://localhos...=1125971068991
HOME
HELP
PREFERENCES
search
titles a-z
filenames
DETACH
NO HIGH-LIGHTING
Done

http://localhost:8080/gsdl/cgi-bin/library?e=d-000-00---0pdf--00

Greenstone: A Comprehensive Open-Source Digital Library Software...

Greenstone: A Comprehensive Open-Source Digital Library Software System
Ian H. Witten, Rodger J. McNab,† Stefan J. Boddie,* David Bainbridge**
* Dept of Computer Science
† Digilib Systems
University of Waikato, New Zealand
Hamilton, New Zealand
E-mail: {ihw, sjboddie, davidb}@cs.waikato.ac.nz
E-mail: rodger@digilibs.com
ABSTRACT
multilingual information retrieval to distributed computing protocols, from interoperability to search engine
This paper describes the Greenstone digital library
technology, from metadata standards to multiformat
software, a comprehensive, open-source system for the
document parsing, from multimedia to multiple operating
construction and presentation of information collections

PDF: use_sections



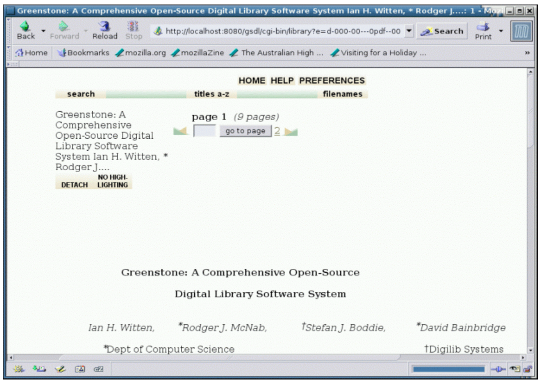
Plugin Selection & Configuration

Use this view to add, configure or remove plugins from your collection. To add one

Please configure the arguments for PDFPlug.

convert_to: html-HTML format.
☐ process_exp (?i)\.pdf$
☐ block_exp
☐ noimages
☐ allowimagesonly
☑ complex
☐ nohidden
☐ zoom 2
☑ use_sections

PDF: HTML Document Display 2



http://localhost:8080/gsdl/cgi-bin/library?e=d-000-00---0pdf--00

Greenstone: A Comprehensive Open-Source Digital Library Software System Ian H. Witten, * Rodger J....

page 1 *(9 pages)* 2

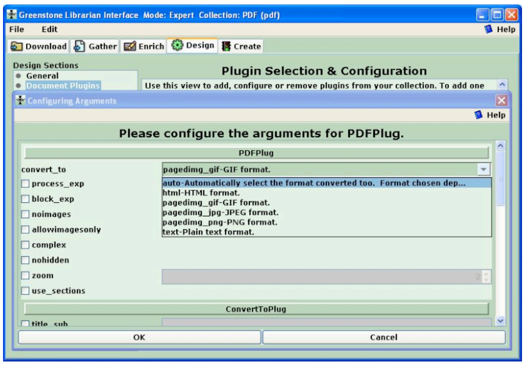
Greenstone: A Comprehensive Open-Source

Digital Library Software System

Ian H. Witten, **Rodger J. McNab,* *†Stefan J. Boddie,* **David Bainbridge*

*Dept of Computer Science †Digilib Systems

PDF -> Image



Plugin Selection & Configuration

Use this view to add, configure or remove plugins from your collection. To add one

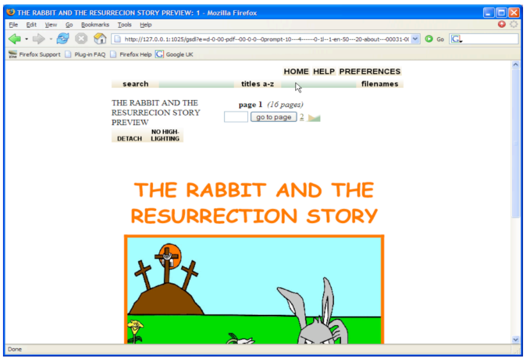
Please configure the arguments for PDFPlug.

convert_to: pagedimg_gif-GIF format.

auto-Automatically select the format converted too. Format chosen dep...
html-HTML format.
pagedimg_gif-GIF format.
pagedimg_jpg-JPEG format.
pagedimg_png-PNG format.
text-Plain text format.

☐ process_exp
☐ block_exp
☐ noimages
☐ allowimagesonly
☐ complex
☐ nohidden
☐ zoom
☐ use_sections
☐ title_sub

PDF: Image Document Display

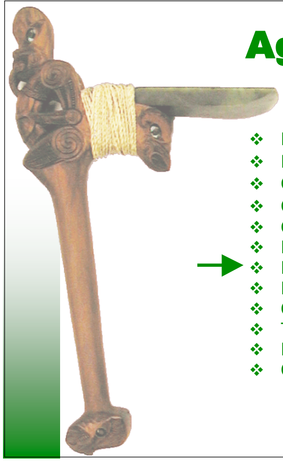


http://127.0.0.1:1025/gsdl?e=d-0-00-pdf--00-0-0--0prompt-10---4------0-1l--1-en-50---20-about---00031-0(

THE RABBIT AND THE RESURRECION STORY PREVIEW

page 1 (16 pages) 2

Agenda



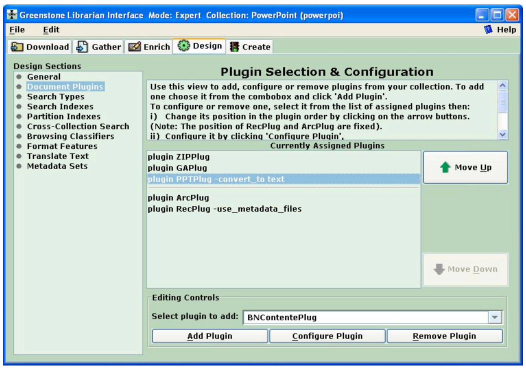
- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- → ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

PowerPoint Document

- ❖ PPT conversions in Greenstone

1. Text
   - ❖ use_strings option
2. HTML
3. Image (JPEG, GIF, PNG)
   - ❖ windows_scripting option
   - ❖ convert_to
     - pagedimg_jpg
     - pagedimg_gif
     - pagedimg_png

PPT -> Text



Plugin Selection & Configuration

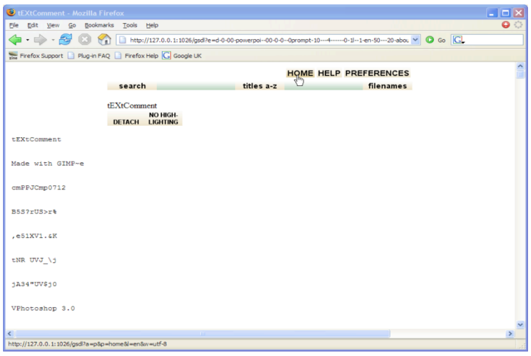
Use this view to add, configure or remove plugins from your collection. To add one choose it from the combobox and click 'Add Plugin'.
To configure or remove one, select it from the list of assigned plugins then:
i) Change its position in the plugin order by clicking on the arrow buttons.
(Note: The position of RecPlug and ArcPlug are fixed).
ii) Configure it by clicking 'Configure Plugin'.

Currently Assigned Plugins

plugin ZIPPlug
plugin GAPlug
plugin PPTPlug -convert_to text
plugin ArcPlug
plugin RecPlug -use_metadata_files

Select plugin to add: BNContentePlug

PPT: Text Document Display



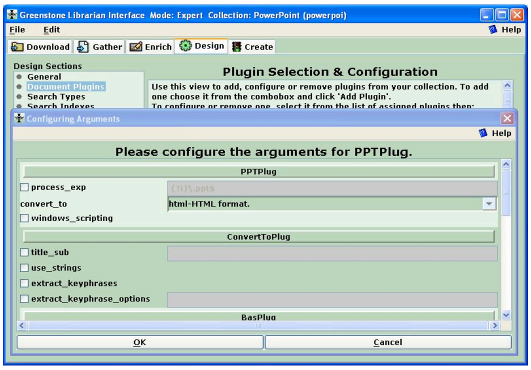
http://127.0.0.1:1026/gsdl?e=d-0-00-powerpoi--00-0-0--0prompt-10---4------0-1l--1-en-50---20-abou

tEXtComment

```
tEXtComment

Made with GIMP~e

cmPPJCmp0712

B5S?rUS>r%

,e51XV1.&K

tNR UVJ_\j

jA34"UV$j0

VPhotoshop 3.0
```

http://127.0.0.1:1026/gsdl?a=p&p=home&l=en&w=utf-8

PPT -> HTML



Greenstone Librarian Interface Mode: Expert Collection: PowerPoint (powerpoi)

Plugin Selection & Configuration

Use this view to add, configure or remove plugins from your collection. To add one choose it from the combobox and click 'Add Plugin'.
To configure or remove one, select it from the list of assigned plugins then:

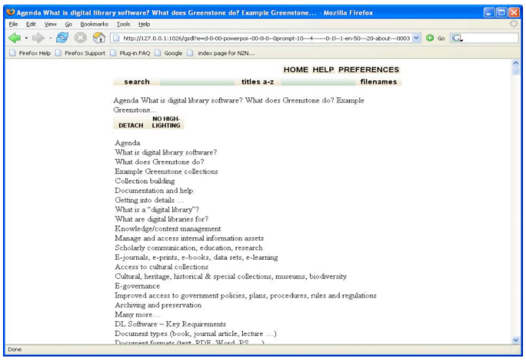
Please configure the arguments for PPTPlug.

PPTPlug
☐ process_exp (?i)\.ppt$
convert_to html-HTML format.
☐ windows_scripting

ConvertToPlug
☐ title_sub
☐ use_strings
☐ extract_keyphrases
☐ extract_keyphrase_options

BasPlug

PPT: HTML Document Display



Agenda What is digital library software? What does Greenstone do? Example Greenstone... - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

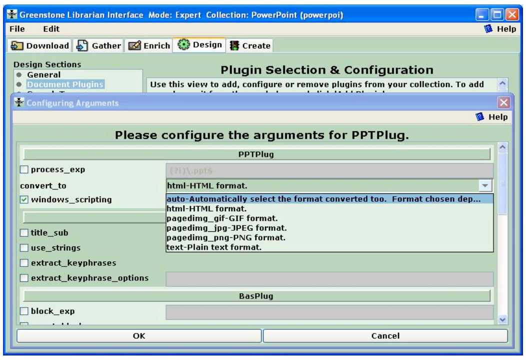
http://127.0.0.1:1026/gsdl?e=d-0-00-powerpoi--00-0-0--0prompt-10---4------0-1l--1-en-50---20-about---0003

Go
Firefox Help Firefox Support Plug-in FAQ Google index page for NZN...
HOME HELP PREFERENCES
search titles a-z filenames

Agenda What is digital library software? What does Greenstone do? Example Greenstone...

Agenda
What is digital library software?
What does Greenstone do?
Example Greenstone collections
Collection building
Documentation and help
Getting into details …
What is a "digital library"?
What are digital libraries for?
Knowledge/content management
Manage and access internal information assets
Scholarly communication, education, research
E-journals, e-prints, e-books, data sets, e-learning
Access to cultural collections
Cultural, heritage, historical & special collections, museums, biodiversity
E-governance
Improved access to government policies, plans, procedures, rules and regulations
Archiving and preservation
Many more…
DL Software – Key Requirements
Document types (book, journal article, lecture …)
Document formats (text, PDF, Word, PS, …)

PPT -> Image



Greenstone Librarian Interface Mode: Expert Collection: PowerPoint (powerpoi)

Plugin Selection & Configuration

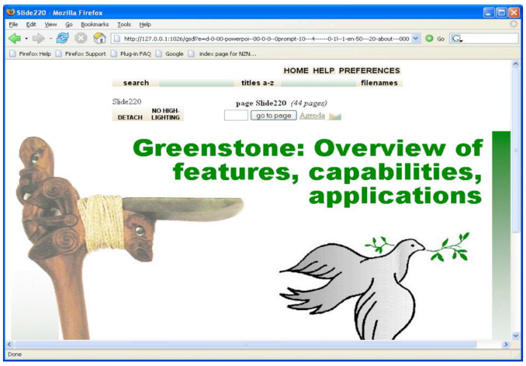
Use this view to add, configure or remove plugins from your collection. To add

Please configure the arguments for PPTPlug.

PPTPlug
☐ process_exp (?i)\.ppt$
convert_to html-HTML format.
☑ windows_scripting
auto-Automatically select the format converted too. Format chosen dep...
html-HTML format.
pagedimg_gif-GIF format.
pagedimg_jpg-JPEG format.
pagedimg_png-PNG format.
text-Plain text format.
☐ title_sub
☐ use_strings
☐ extract_keyphrases
☐ extract_keyphrase_options

BasPlug
☐ block_exp

PPT Image: Image View



Slide220 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://127.0.0.1:1026/gsdl?e=d-0-00-powerpoi--00-0-0--0prompt-10---4------0-1l--1-en-50---20-about---000

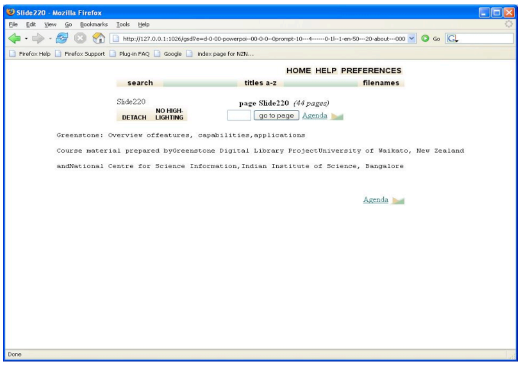
Go
Firefox Help Firefox Support Plug-in FAQ Google index page for NZN...
HOME HELP PREFERENCES
search titles a-z filenames

Slide220

page Slide220 *(44 pages)*

Greenstone: Overview of features, capabilities, applications

PPT Image: Text View



Slide220 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://127.0.0.1:1026/gsdl?e=d-0-00-powerpoi--00-0-0--0prompt-10---4------0-1l--1-en-50---20-about---000

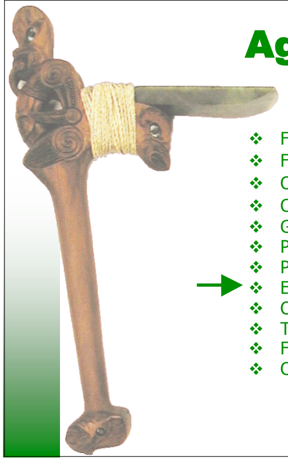
Go
Firefox Help Firefox Support Plug-in FAQ Google index page for NZN...
HOME HELP PREFERENCES
search titles a-z filenames

Slide220

page Slide220 *(44 pages)*

```
Greenstone: Overview offeatures, capabilities,applications

Course material prepared byGreenstone Digital Library ProjectUniversity of Waikato, New Zealand

andNational Centre for Science Information,Indian Institute of Science, Bangalore
```

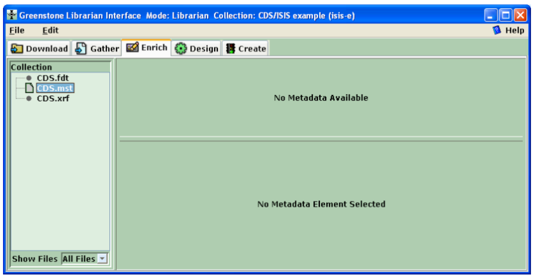
Agenda



- Formatting
- Formatting extended metadata
- Changing metadata sets
- Collection configuration file
- GLI modes
- PDF documents
- PPT documents
- → Exploding metadata databases
- Cross collection searching
- Translate text
- Full-text tagging
- Creating a CD-ROM

Exploding metadata databases

- ❖ The GLI does not show metadata extracted from bibliographic files: CDS/ISIS, MARC etc.



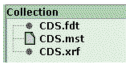
No Metadata Available

No Metadata Element Selected

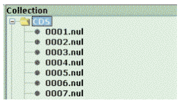
Why not?

- ❖ The GLI treats these files just like any other: it shows extracted and assigned metadata, but not the file's contents
  - – Double-click the file to open it in the normal editor for this file type eg. WinISIS for CDS/ISIS files
- ❖ In rare cases you might want to convert these files into Greenstone format, allowing the metadata records to be visible and editable from the GLI
- ❖ This is irreversible: there's no going back!
  - – Keep a backup of the original file

Exploding Metadata Databases

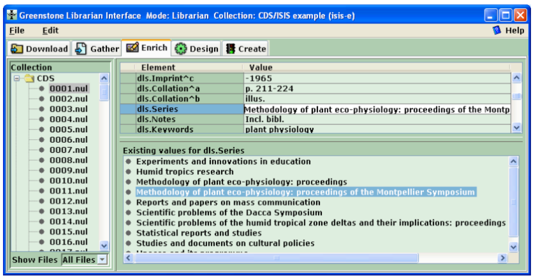


- ❖ Files that can be exploded are shown with a green icon in the GLI
- ❖ Right-click on the file and choose "Explode Metadata Database"
- ❖ You will be prompted to map the metadata into a metadata set
- ❖ When complete, the file will be replaced with a folder containing a file for each record



Exploding Metadata Databases

- ❖ The metadata can now be viewed and edited:



Element	Value
dls.Imprint^c	-1965
dls.Collation^a	p. 211-224
dls.Collation^b	illus.
dls.Series	Methodology of plant eco-physiology: proceedings of the Montp
dls.Notes	Incl. bibl.
dls.Keywords	plant physiology

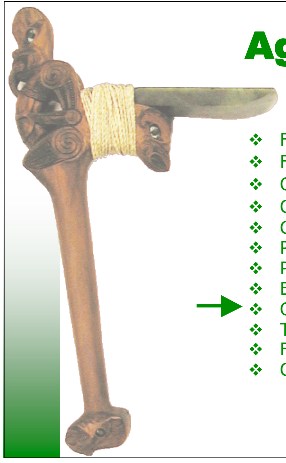
Existing values for dls.Series

- Experiments and innovations in education
- Humid tropics research
- Methodology of plant eco-physiology: proceedings
- Methodology of plant eco-physiology: proceedings of the Montpellier Symposium
- Reports and papers on mass communication
- Scientific problems of the Dacca Symposium
- Scientific problems of the humid tropical zone deltas and their implications: proceedings
- Statistical reports and studies
- Studies and documents on cultural policies

- ❖ Note: the GLI is *not* a real database system
  - – Only feasible with a small number of records

Exploding Metadata Databases

- ❖ Change classifiers, index specifications and format statements to use the namespaced metadata elements
- ❖ When importing the collection, the files will now be processed by NULPlug rather than the plugin for the original file (eg. ISISPlug)



Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- → ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

Cross Collection Searching

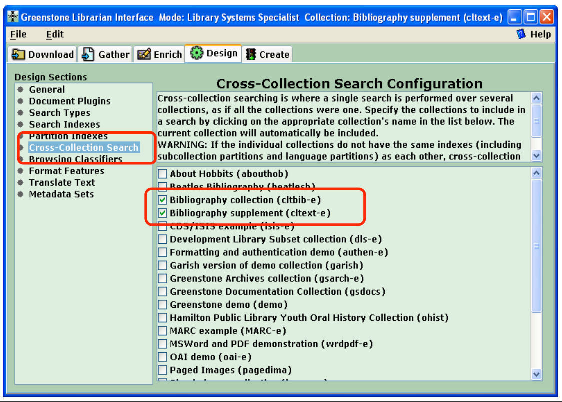
- ❖ Select a list of collections
- ❖ Collections need the same indexes
- ❖ User can select which collections to search on Preferences page
- ❖ Format statements applied from original collection

Bibliography collections

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.

Cross collection search

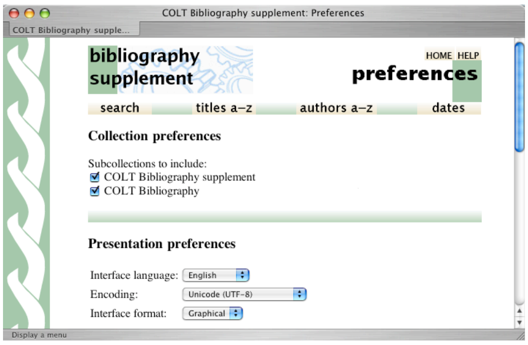


Cross-Collection Search Configuration

Cross-collection searching is where a single search is performed over several collections, as if all the collections were one. Specify the collections to include in a search by clicking on the appropriate collection's name in the list below. The current collection will automatically be included.
WARNING: If the individual collections do not have the same indexes (including subcollection partitions and language partitions) as each other, cross-collection

- ☐ About Hobbits (abouthob)
- ☐ Beatles Bibliography (beatlesb)
- ☑ Bibliography collection (cltbib-e)
- ☑ Bibliography supplement (cltext-e)
- ☐ CDS/ISIS example (isis-e)
- ☐ Development Library Subset collection (dls-e)
- ☐ Formatting and authentication demo (authen-e)
- ☐ Garish version of demo collection (garish)
- ☐ Greenstone Archives collection (gsarch-e)
- ☐ Greenstone Documentation Collection (gsdocs)
- ☐ Greenstone demo (demo)
- ☐ Hamilton Public Library Youth Oral History Collection (ohist)
- ☐ MARC example (MARC-e)
- ☐ MSWord and PDF demonstration (wrdpdf-e)
- ☐ OAI demo (oai-e)
- ☐ Paged Images (pagedima)

Bibliography supplement



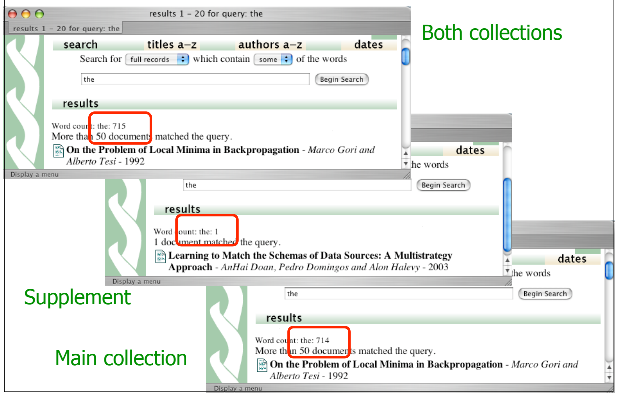
Collection preferences

Subcollections to include:
☑ COLT Bibliography supplement
☑ COLT Bibliography

Presentation preferences

Interface language: English
Encoding: Unicode (UTF-8)
Interface format: Graphical

Cross-collection searching ...



Search for full records which contain some of the words

the

Word count: the: 715
More than 50 documents matched the query.
On the Problem of Local Minima in Backpropagation - *Marco Gori and Alberto Tesi* - 1992

Word count: the: 1
1 document matched the query.
Learning to Match the Schemas of Data Sources: A Multistrategy Approach - *AnHai Doan, Pedro Domingos and Alon Halevy* - 2003

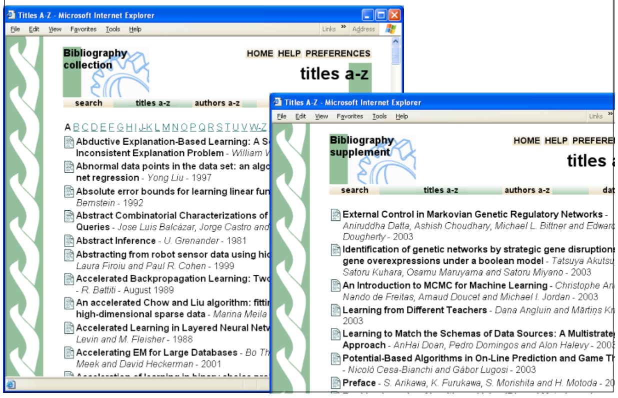
Word count: the: 714
More than 50 documents matched the query.
On the Problem of Local Minima in Backpropagation - *Marco Gori and Alberto Tesi* - 1992

Both collections

Supplement

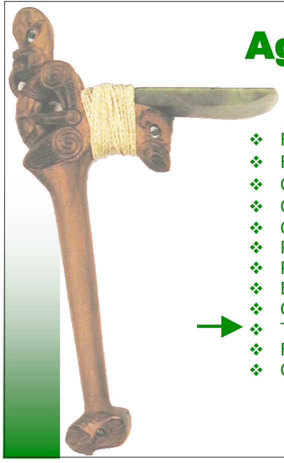
Main collection

... but not browsing



Abductive Explanation-Based Learning: A S
Inconsistent Explanation Problem - *William W*
Abnormal data points in the data set: an algo
net regression - *Yong Liu* - 1997
Absolute error bounds for learning linear fun
Bernstein - 1992
Abstract Combinatorial Characterizations of
Queries - *Jose Luis Balcázar, Jorge Castro and*
Abstract Inference - *U. Grenander* - 1981
Abstracting from robot sensor data using hid
Laura Firoiu and Paul R. Cohen - 1999
Accelerated Backpropagation Learning: Two
- *R. Battiti* - August 1989
An accelerated Chow and Liu algorithm: fittin
high-dimensional sparse data - *Marina Meila*
Accelerated Learning in Layered Neural Netw
Levin and M. Fleisher - 1988
Accelerating EM for Large Databases - *Bo Th*
Meek and David Heckerman - 2001

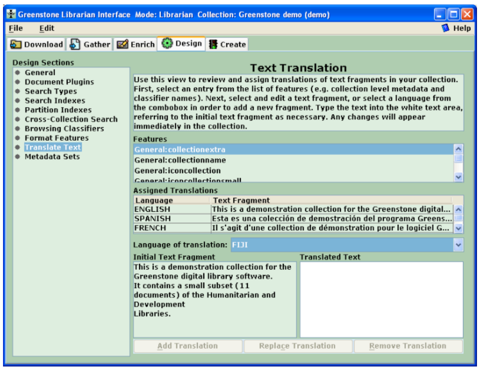
External Control in Markovian Genetic Regulatory Networks - *Aniruddha Datta, Ashish Choudhary, Michael L. Bittner and Edward Dougherty* - 2003
Identification of genetic networks by strategic gene disruptions gene overexpressions under a boolean model - *Tatsuya Akutsu, Satoru Kuhara, Osamu Maruyama and Satoru Miyano* - 2003
An Introduction to MCMC for Machine Learning - *Christophe An Nando de Freitas, Arnaud Doucet and Michael I. Jordan* - 2003
Learning from Different Teachers - *Dana Angluin and Mārtiņš Kr* 2003
Learning to Match the Schemas of Data Sources: A Multistrateg Approach - *AnHai Doan, Pedro Domingos and Alon Halevy* - 200
Potential-Based Algorithms in On-Line Prediction and Game Th - *Nicolò Cesa-Bianchi and Gábor Lugosi* - 2003
Preface - *S. Arikawa, K. Furukawa, S. Morishita and H. Motoda* - 20



Agenda

- Formatting
- Formatting extended metadata
- Changing metadata sets
- Collection configuration file
- GLI modes
- PDF documents
- PPT documents
- Exploding metadata databases
- Cross collection searching
- → Translate text
- Full-text tagging
- Creating a CD-ROM

Translate Text



Text Translation

Use this view to review and assign translations of text fragments in your collection. First, select an entry from the list of features (e.g. collection level metadata and classifier names). Next, select and edit a text fragment, or select a language from the combobox in order to add a new fragment. Type the text into the white text area, referring to the initial text fragment as necessary. Any changes will appear immediately in the collection.

Features

General:collectionextra
General:collectionname
General:iconcollection
General:iconcollectionsmall

Assigned Translations

Language	Text Fragment
ENGLISH	This is a demonstration collection for the Greenstone digital...
SPANISH	Esta es una colección de demostración del programa Greens...
FRENCH	Il s'agit d'une collection de démonstration pour le logiciel G...

Language of translation: FIJI

Initial Text Fragment

This is a demonstration collection for the Greenstone digital library software.
It contains a small subset (11 documents) of the Humanitarian and Development
Libraries.

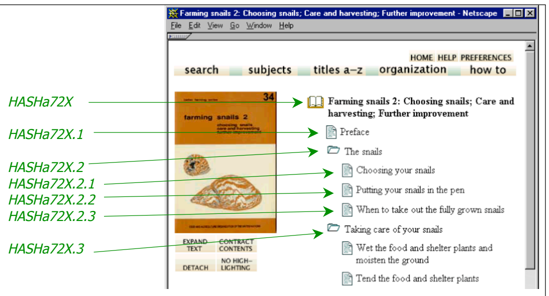
Translated Text

Agenda

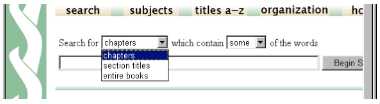
- Formatting
- Formatting extended metadata
- Changing metadata sets
- Collection configuration file
- GLI modes
- PDF documents
- PPT documents
- Exploding metadata databases
- Cross collection searching
- Translate text
- → Full-text tagging
- Creating a CD-ROM

Full Text Tagging

- While creating large digital collections:
  - the collection must be organized
  - the larger the collection the greater the need for organization
  - the larger the documents the greater the need for sections/subsections
- Greenstone lets you tag the full text of documents
- Then you can read them hierarchically …
- … and search them by section



Farming snails 2: Choosing snails; Care and harvesting; Further improvement
Preface
The snails
Choosing your snails
Putting your snails in the pen
When to take out the fully grown snails
Taking care of your snails
Wet the food and shelter plants and moisten the ground
Tend the food and shelter plants



Search for chapters which contain some of the words

Full Text Tagging…

To show the hierarchical structure, tag the source files like this:

```
<!--
 <Section>
 <Description>
 <Metadata name="Title">Realizing human rights for
      poor people: Strategies for achieving the
      international development targets</Metadata>
 </Description>
 -->
      (text of section goes here)
 <!-
 </Section>
 -->
```

Full Text Tagging…

- ❖ Section tags define a hierarchical structure
- ❖ Sections can be nested within other sections
- ❖ All sections must be nested within a single enclosing section that encompasses the entire document
- ❖ In the collection configuration file, put

```
HTMLPlug -description_tags
```

- ❖ Mainly for HTML, but can be used in Word and PDF documents.

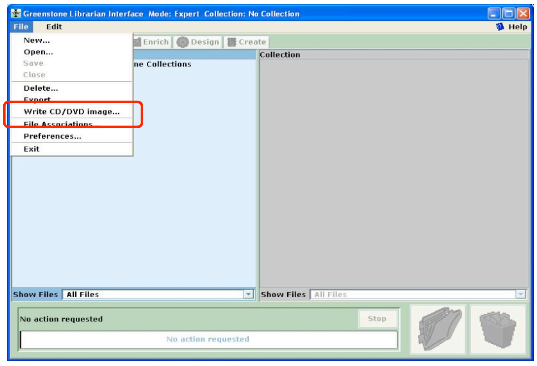
Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- → ❖ Creating a CD-ROM

Creating a CD-ROM

- ❖ Export any Greenstone collection as a CD-ROM
- ❖ Self-installing
- ❖ Windows only (sorry!)
- ❖ Full Installation of Greenstone
- ❖ In the Librarian
  File -> Write CD/DVD image
- ❖ C:\Program Files\Greenstone\tmp\exported_xxx

Exporting to CD-ROM

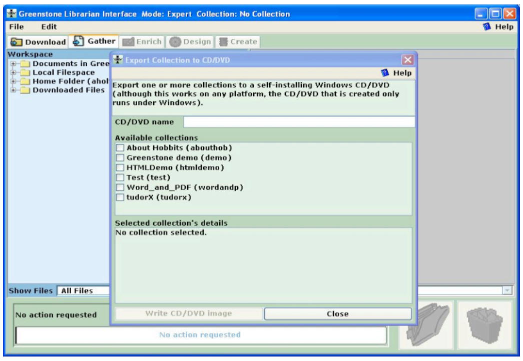


Greenstone Librarian Interface Mode: Expert Collection: No Collection

No action requested

No action requested

Exporting to CD-ROM



Greenstone Librarian Interface Mode: Expert Collection: No Collection

Export Collection to CD/DVD

Export one or more collections to a self-installing Windows CD/DVD (although this works on any platform, the CD/DVD that is created only runs under Windows).

CD/DVD name

Available collections

- ☐ About Hobbits (abouthob)
- ☐ Greenstone demo (demo)
- ☐ HTMLDemo (htmldemo)
- ☐ Test (test)
- ☐ Word_and_PDF (wordandp)
- ☐ tudorX (tudorx)

Selected collection's details

No collection selected.

No action requested

No action requested

Note

- ❖ CD-ROM's created this way have not been tested extensively under different Windows configurations
- ❖ But they should work on all Windows platforms …
- ❖ … except 3.1/3.11 (is this a problem?)

Creating a CD-ROM: demo