

Librarian Interface (GLI): Building collections



Course material prepared by

Greenstone Digital Library Project
University of Waikato, New Zealand

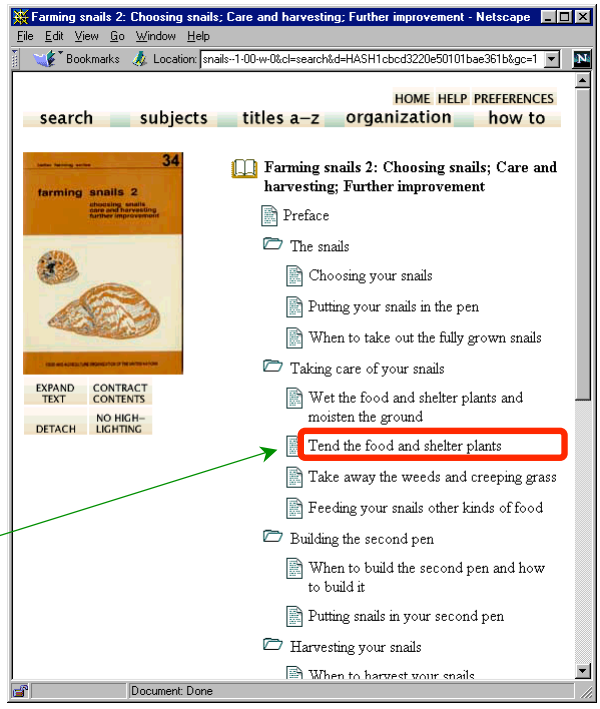
and **National Centre for Science Information,**
Indian Institute of Science, Bangalore



Agenda

- Review: Searching and browsing
- Plugins
- Collection configuration
- Greenstone Librarian Interface: building collections

Hierarchical document model

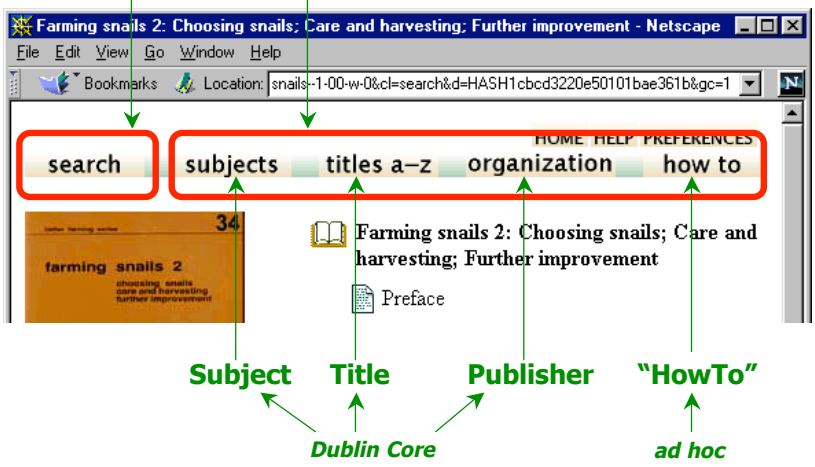


- Metadata specified at any level

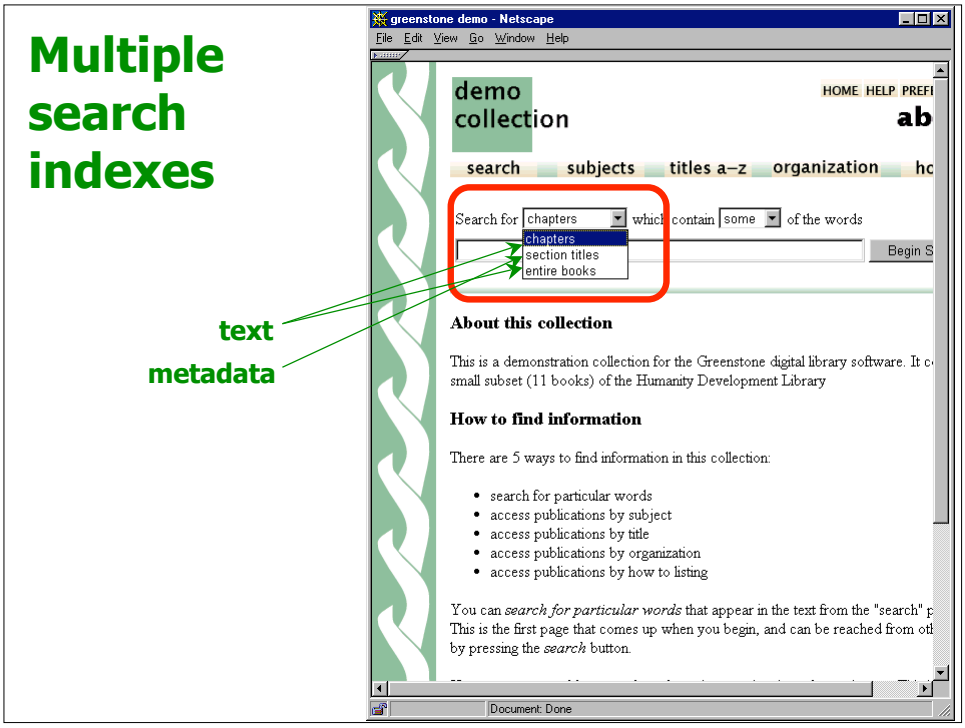
Title metadata

Searching and browsing

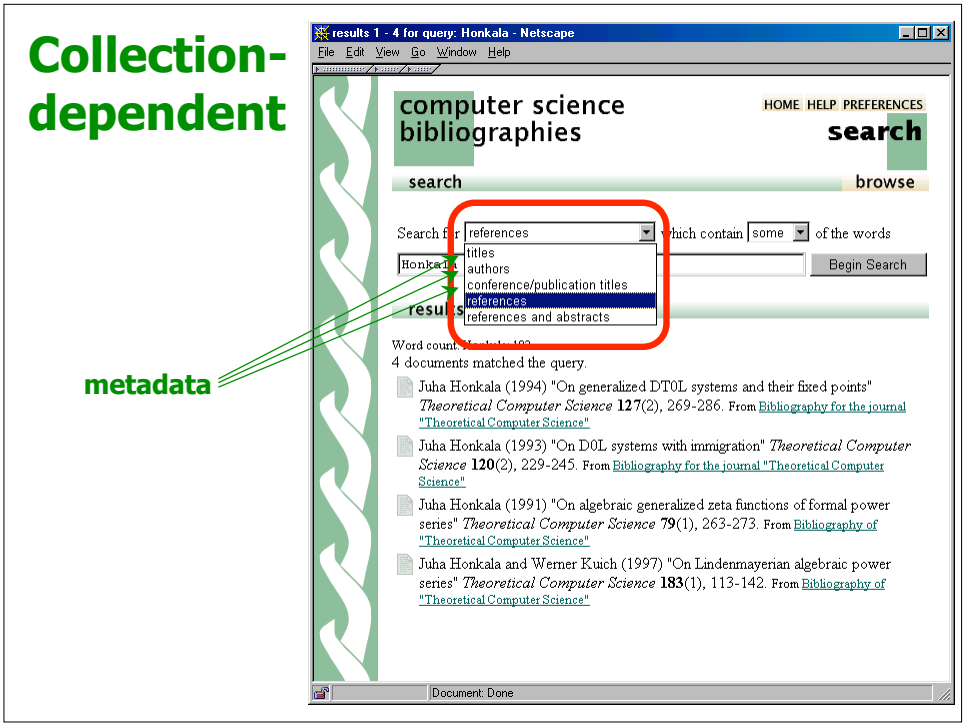
- Searching
- Metadata-based browsing



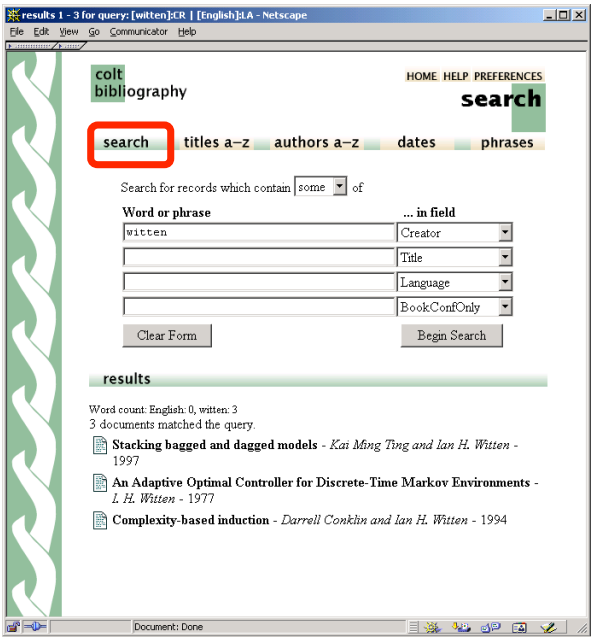
Multiple search indexes



Collection-dependent

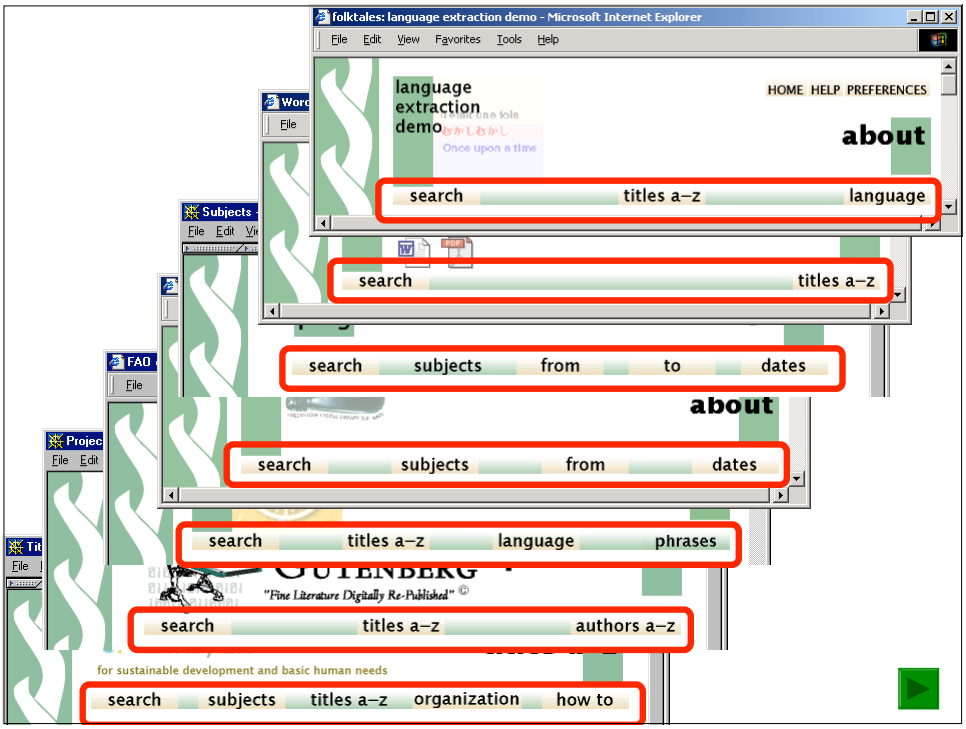


Form search



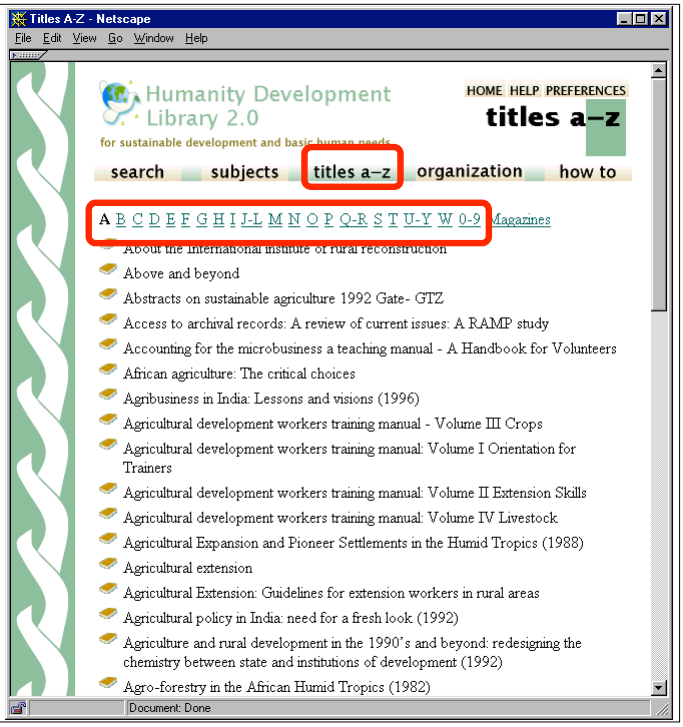
Multilingual searching



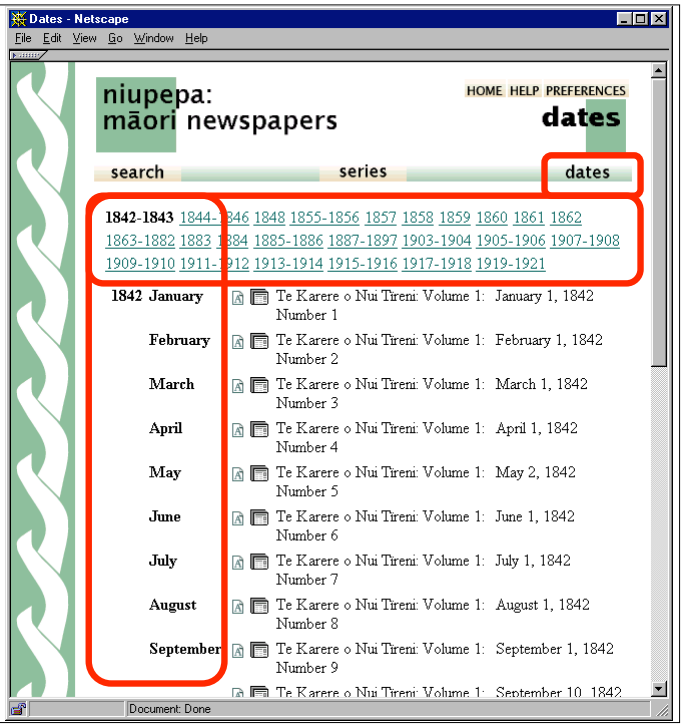


Browsing using *classifiers*

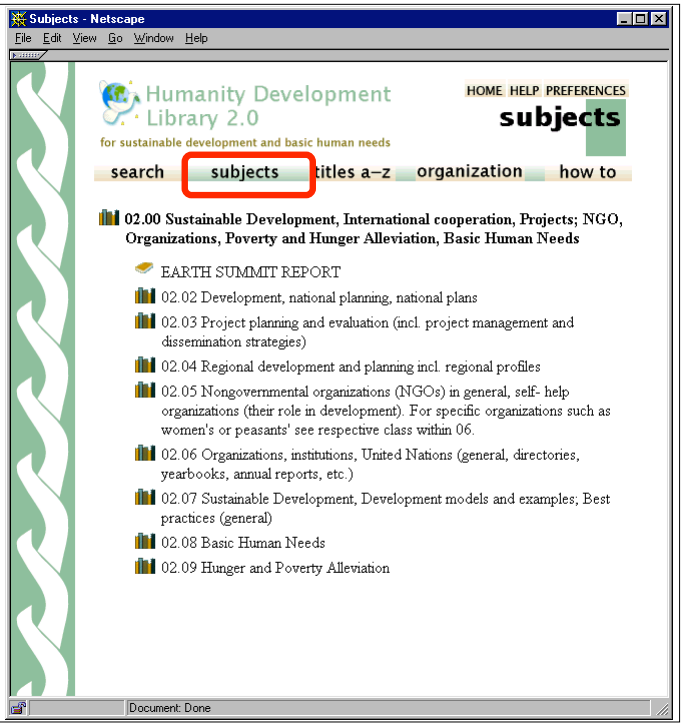
AZList classifier (*Title* metadata)

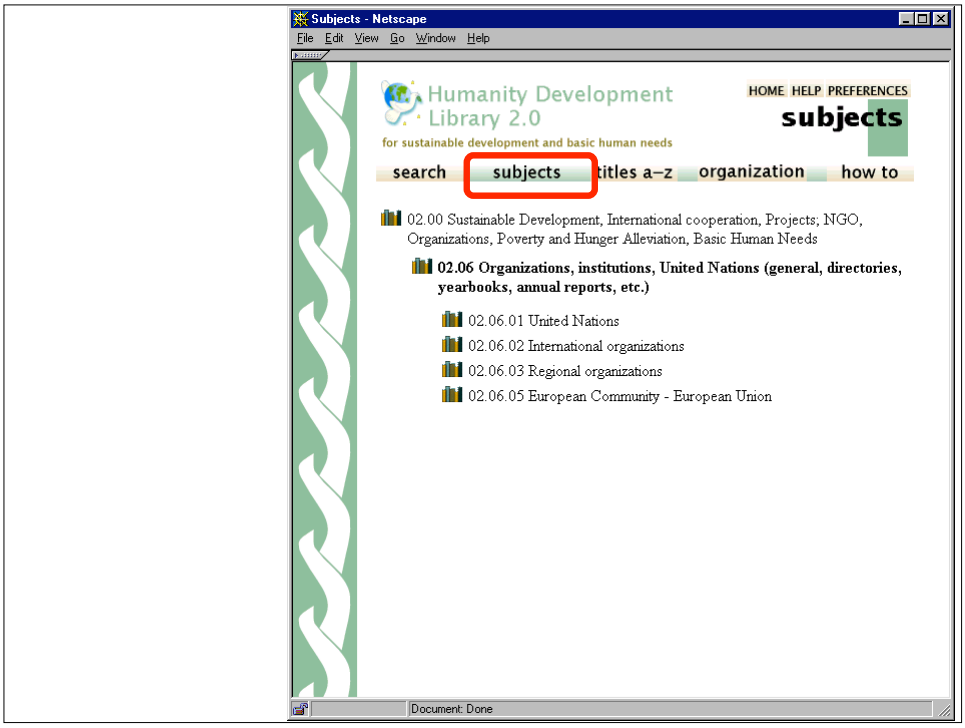


DateList classifier (*Date* metadata)

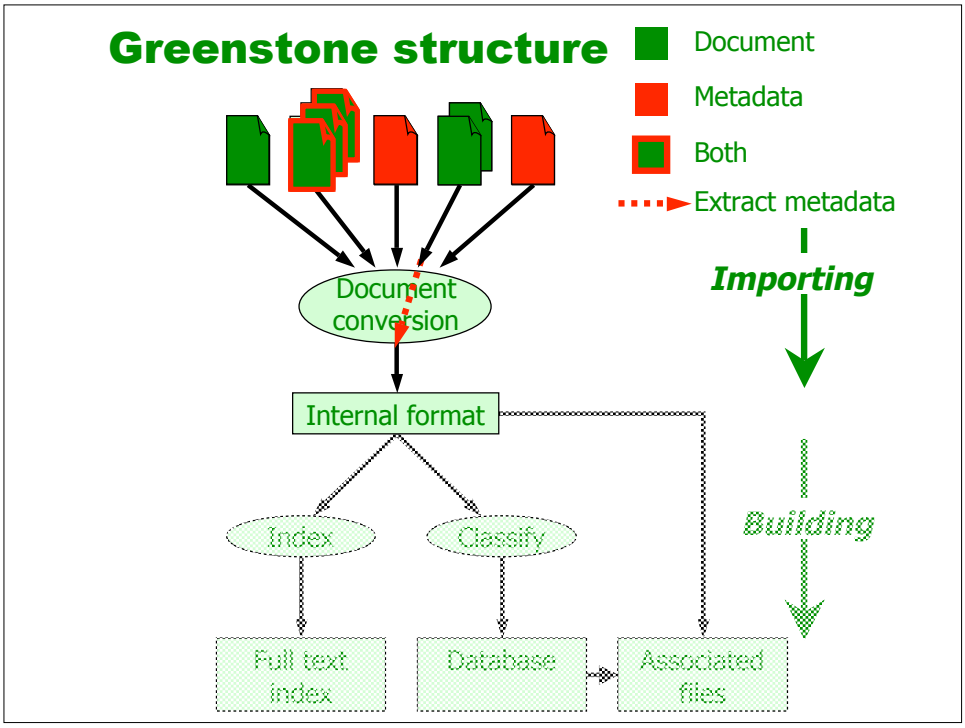


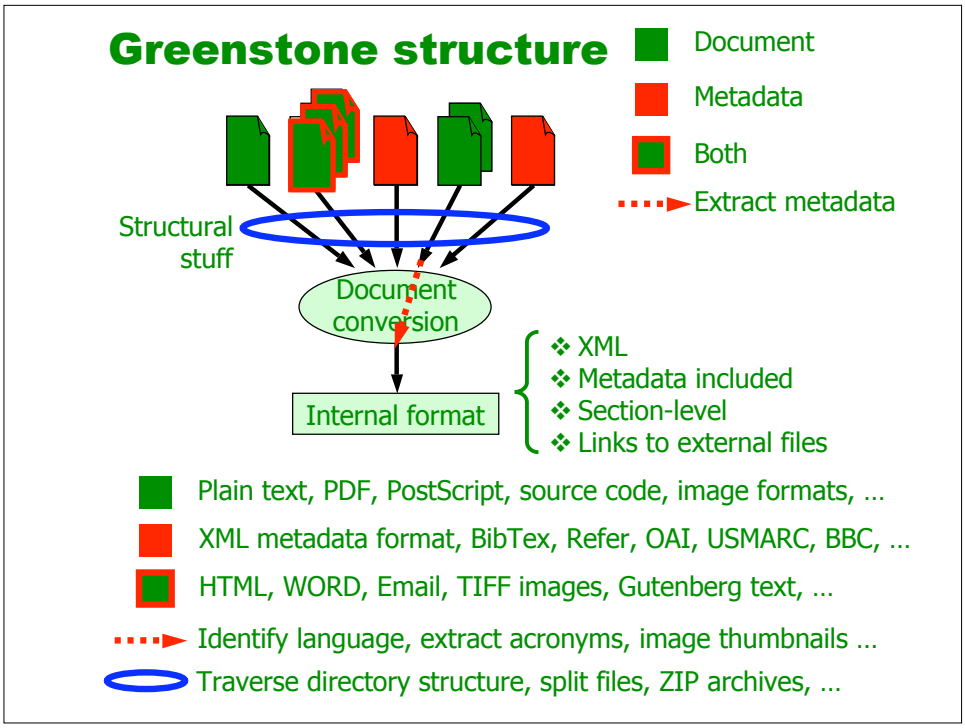
Hierarchy classifier (*Subject* metadata)




Subjects - Netscape
File Edit View Go Window Help
Humanity Development Library 2.0
for sustainable development and basic human needs
HOME HELP PREFERENCES
subjects
search
subjects
titles a–z
organization
how to
02.00 Sustainable Development, International cooperation, Projects; NGO, Organizations, Poverty and Hunger Alleviation, Basic Human Needs
02.06 Organizations, institutions, United Nations (general, directories, yearbooks, annual reports, etc.)
02.06.01 United Nations
02.06.02 International organizations
02.06.03 Regional organizations
02.06.05 European Community - European Union
Document: Done


Agenda
Searching and browsing
Plugins
Collection configuration
Greenstone Librarian Interface: building collections


Greenstone structure
Document
Metadata
Both
Extract metadata
Importing
Document conversion
Internal format
Index
Classify
Building
Full text index
Database
Associated files


Greenstone structure
Document
Metadata
Both
Extract metadata
Structural stuff
Document conversion
Internal format
XML
Metadata included
Section-level
Links to external files
Plain text, PDF, PostScript, source code, image formats, ...
XML metadata format, BibTex, Refer, OAI, USMARC, BBC, ...
HTML, WORD, Email, TIFF images, Gutenberg text, ...
Identify language, extract acronyms, image thumbnails ...
Traverse directory structure, split files, ZIP archives, ...



Email plugin



Agenda

- Searching and browsing
- Plugins
- Collection configuration
- Greenstone Librarian Interface: building collections

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"
```

Alter configuration

Add full-text index of titles	**indexes document:Title**
... or authors	**indexes document:Creator**
Add alphabetic author brows	**classify AZList –metadata Creator**
Include Word documents	**plugin WordPlug**
Include PDF documents	**plugin PDFPlug**
Separate index for each lang	**languages en fr es**
Extract acronyms and add lis	**plugin PDFPlug –extract_acronyms**
Import OAI metadata	**plugin OAIPlug**
Extract phrase hierarchy and browser	**classify phind**
Alter the format of any of the	**format ...**
Restrict collection's interface	**format PreferenceLangs en|fr|es**
Change default interface lang	**cgiarg shortname=1 argdefault =fr**

Collection configuration file

```
Creator     sjboddie@cs.waikato.ac.nz
Maintainer  sjboddie@cs.waikato.ac
public      true
beta        true

indexes      section:text
defaultindex section:text

plugin  RECPlug
plugin  WORDPlug
Plugin  ZIPPlug
plugin  PDFPlug
plugin  MARCPlug

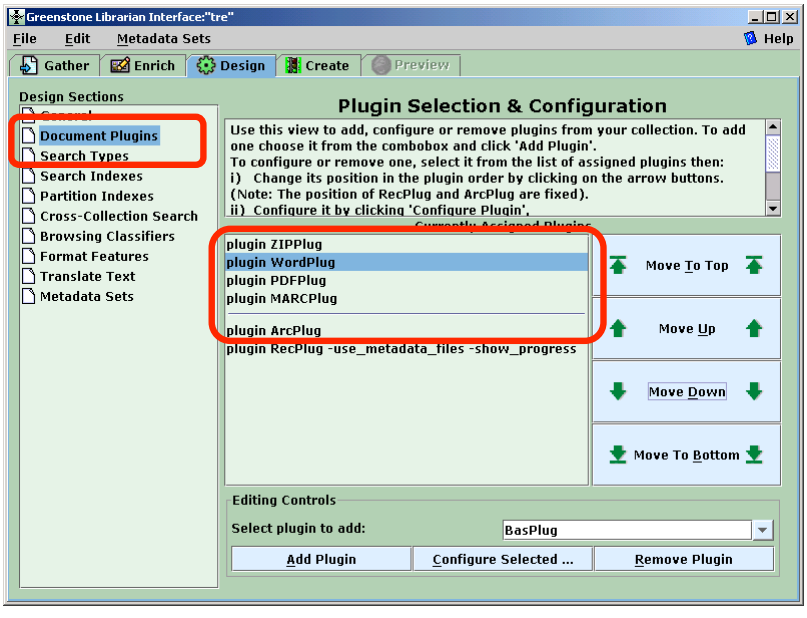
classify  Hierarchy hfile=
classify  HDLList metadata=
classify  Hierarchy hfile=org
classify  List metadata=Howto

format    SearchVList "<td valign=to
          <td>{If}{[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"
```

```
plugin  RECPlug
plugin  WORDPlug
plugin  ZIPPlug
plugin  PDFPlug
plugin  MARCPlug
```

Librarian Interface (GLI)



Plugins

Used by collection-building software to accomplish format-specific parsing of source documents

Plugin pipeline: files are passed to each plugin in turn until one is found that can process it

TEXTPlug
HTMLPlug
EMAILPlug
WORDPlug
RTFPlug
PDFPlug
PSPlug
FoxPlug
PrePlug
GBPlug
TCCPlug ...

- GAPlug — processes .ga files generated during *import*
- ArcPlug — processes .ga filelist in *archives.inf*
- RecPlug — recurses through a directory structure



Agenda

- Searching and browsing
- Plugins
- Collection configuration
- Greenstone Librarian Interface: building collections

Greenstone Librarian Interface (GLI): building collections

Invoke GLI: build a small collection of HTML files. Gather and Build panels.

- **Look at extracted metadata**
- **Build larger collection of HTML files**
- **Look at different views of the files (filters)**
- **Build collection of Word/PDF files**

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

Questions?

Comments?

Discussion?

Feedback form!