

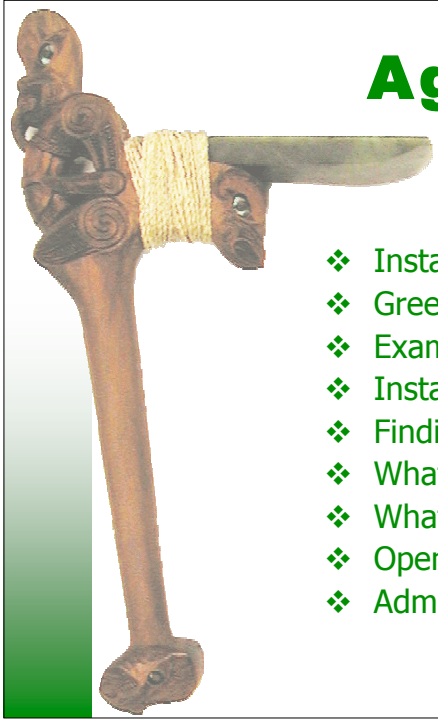
Platforms, installation, configuration; accessing example collections



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

- Installing a pre-built collection
- Greenstone: platforms and versions
- Example collections
- Installing and troubleshooting
- Finding Greenstone and GLI
- What's on the CD-ROM?
- What's on my disk?
- Open source
- Admin pages

Pre-built collection: Installation demo

(lab exercise)

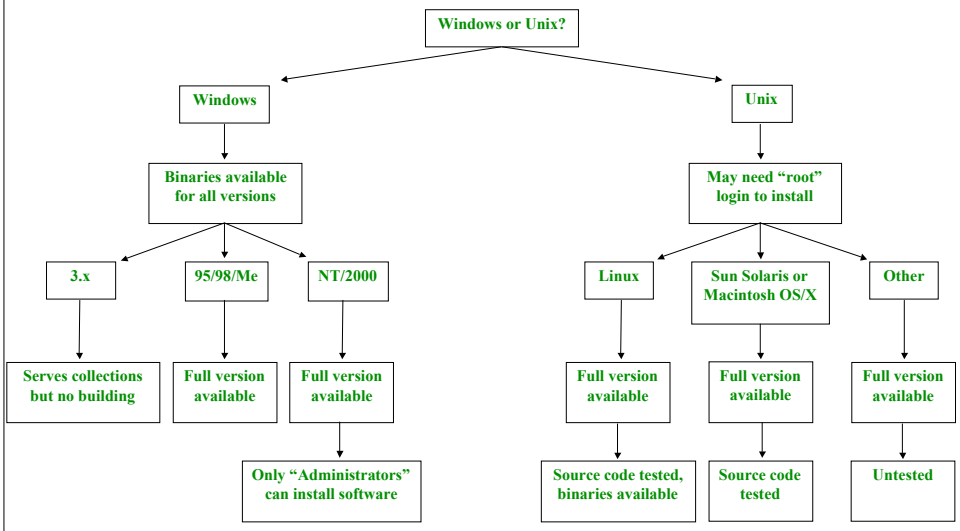
Platforms

- Operating system:
  - Windows (any version)
  - Linux (any version)
  - Unix (most versions, e.g. Solaris)
  - Mac OS X (some problems with GLI interface)
- Restrictions:
  - No longer runs under Windows 3.1/3.11
  - For Librarian interface (GLI), need Java— which is 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
- Install by compiling source code
  - Windows
  - Linux
  - Unix
  - Mac

Installing Greenstone

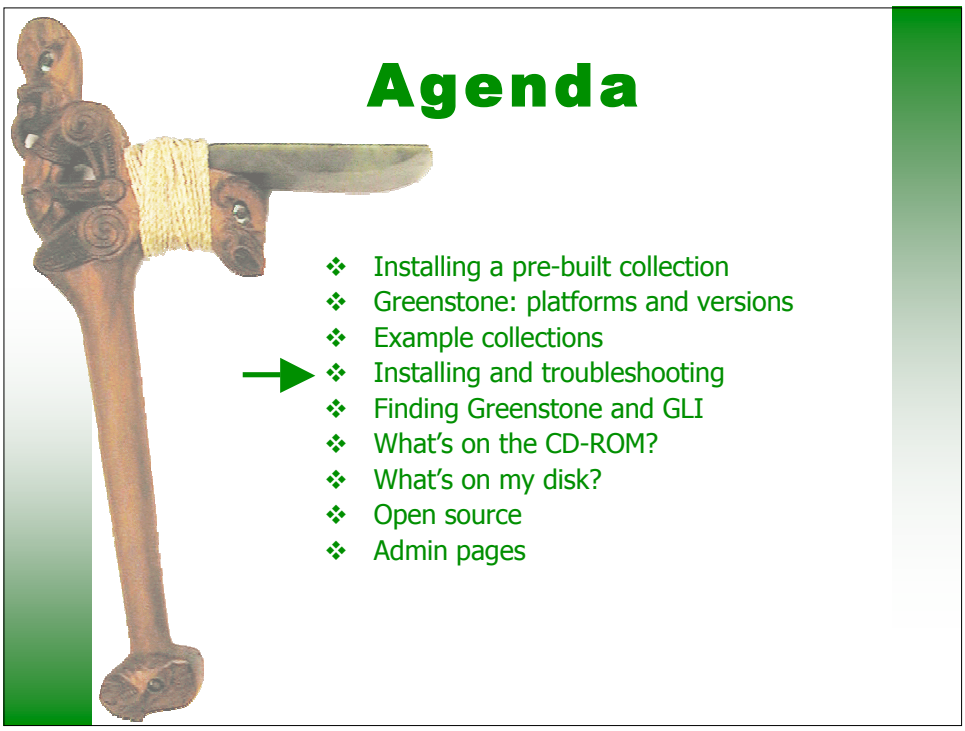


Example collections

development library subset	The DLS collection has the same structure as the Greenstone demo collection. 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, or 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	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).

Example collections

OAI example	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.
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.
authentication and formatting 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 demo	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.
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.


Agenda
Installing a pre-built collection
Greenstone: platforms and versions
Example collections
Installing and troubleshooting
Finding Greenstone and GLI
What's on the CD-ROM?
What's on my disk?
Open source
Admin pages


Local library:
Installation demo
(lab exercise)

Local Library: remember

- Installation starts automatically when CD is inserted … or else
  … double-click "My Computer" on the desktop and double-click again on the CD-ROM drive
- Default installation directory:
  C:\>\program files\gsdl
- Starting up Greenstone:
  Start → Greenstone digital library
- Starting up Librarian interface (GLI):
  Start → Greenstone digital library

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/10: Web library only**
  - **Use Apache (or other web server)**
  - **Linux binaries supplied**
  - **Tested on SUN Solaris, Mac OS/10**
  - **Need GDBM (standard on Linux)**
- **Greenstone developed on Linux**

Web Library: Installation demo

Web Library: remember

- Installation starts when CD is inserted
- Default installation directory:
  C:\>\program files\gsdl
- Must configure the Web server (geeky job?)
- Accessing Greenstone:
  http://localhost/cgi-bin/library.exe
  or http://hostname/cgi-bin/library.exe
  or http://xxx.xxx.xx.xx/cgi-bin/library.exe
- Starting up Librarian interface (GLI):
  Start → Greenstone digital library

Configuring the Web server (PWS/IIS)

1. Select *Advanced* to get the *Advanced Options* screen.

2. Select *Home* and click *Add*. Fill out the fields as follows:

Directory field:	*C:\Program Files\gsdl*
Alias field:	*gsdl*
Access permissions:	*Read*
Application permissions:	*None*
Click *OK*	

   This makes Greenstone files accessible to the webserver.

3. Back in *Advanced Options*, select *gsdl* and click *Add*. Fill out the fields as follows:

Directory field:	*C:\Program Files\gsdl\cgi-bin*
Alias field:	*cgi-bin*
Access permissions:	*None*
Application permissions:	*Execute*
Click *OK*	

   This allows the Greenstone program *library.exe* to be executed by the webserver.

4. Go to the URL *http://localhost/cgi-bin/library.exe*.

   Note: you need to specify the *.exe* file extension with PWS and IIS.

Troubleshooting

	Problem	Try this
LOCAL LIBRARY (WINDOWS ONLY)	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 start Greenstone my computer *still* asks me to dial up my Internet Service Provider.	Choose the "Restricted version" when you run Greenstone. This version only works with Netscape.
	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).
WEB LIBRARY (WINDOWS AND UNIX)	When I start Apache, it quits immediately.	Add a *ServerName localhost* directive to the Apache configuration file (see Section 4.1).
	When I point my browser at the digital library, it displays garbage—a binary file.	Check the *ScriptAlias* directive in the Apache configuration file, making sure it comes before the *Alias* directive (see Sections 4.2 and 4.3).
	I get the Greenstone home page (Figure 2), but the Demo collection icon does not appear.	Run the program *library* (in the cgi-bin directory) from the DOS (or shell) prompt to generate debugging information that will help you locate the problem.

Troubleshooting

BOTH VERSIONS

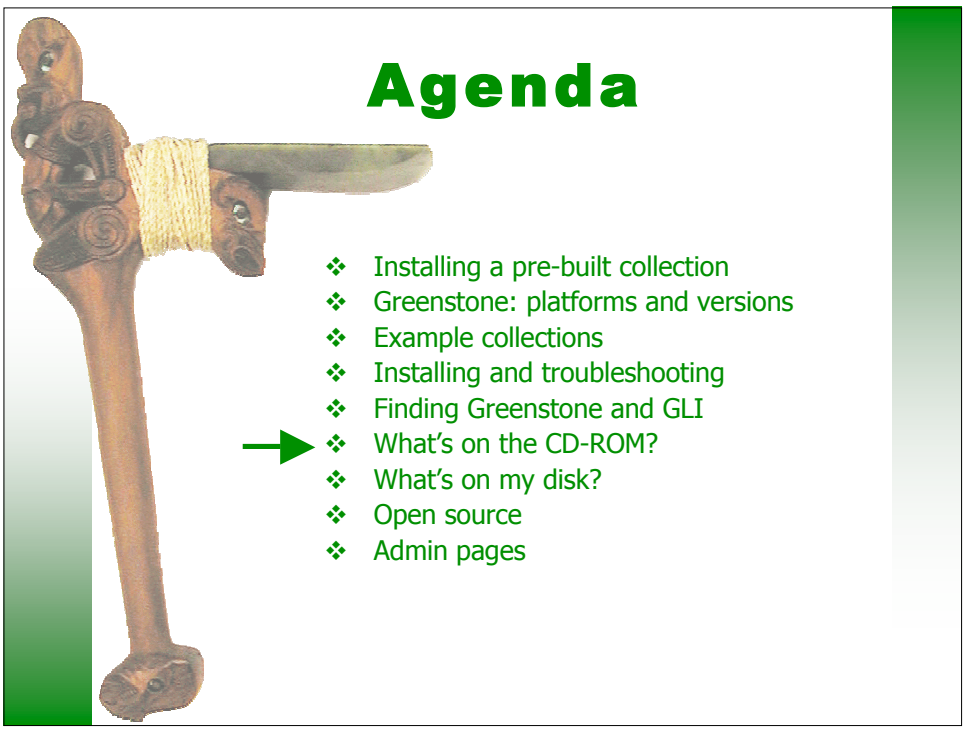
Problem	Try this
When I point my browser at the digital library, it can't find that page.	Try using 127.0.0.1 in place of *localhost*. This reserved IP number is defined to be a "loopback" to your local computer.
My browser complains that it can't find *main.cfg*.	Check that the Greenstone files exist and are world-readable. If you are using the Web library, try running the *library* program from the command line. If it runs OK, the problem is with file permissions (see Section 5.1). If not, the *gsdlhome* variable is probably set incorrectly in the *gsdlsite.cfg* configuration file (see Section 5.2).
I'm having trouble using the Collector.	Read the *Greenstone Digital Library User's Guide*, Section 3.
I've added a new user but they can't seem to log in.	Check that the directory *C:\Program Files\gsdl\etc* and all its contents are globally writeable (see Section 5.1).

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

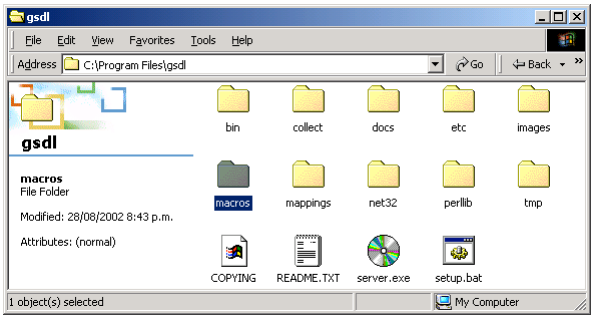
- FAQ: sections on
  - Obtaining Greenstone
  - Installing Greenstone
  - Running Greenstone

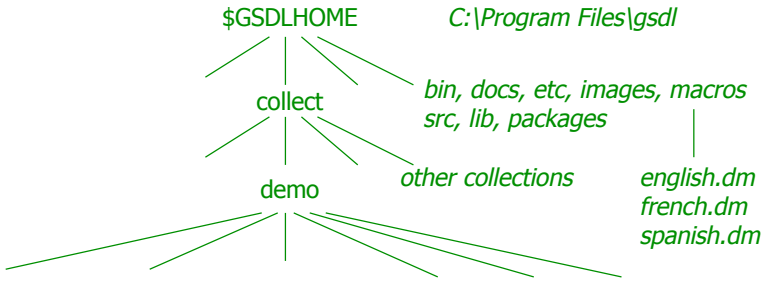


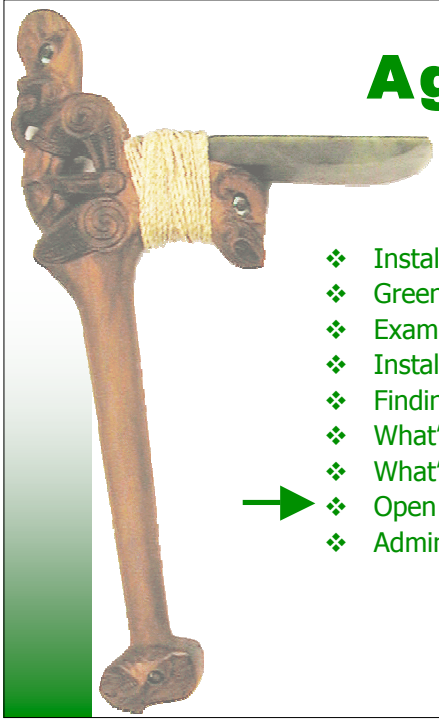
What's on the CD-ROM?

What's on my disk?

What's where?







Agenda

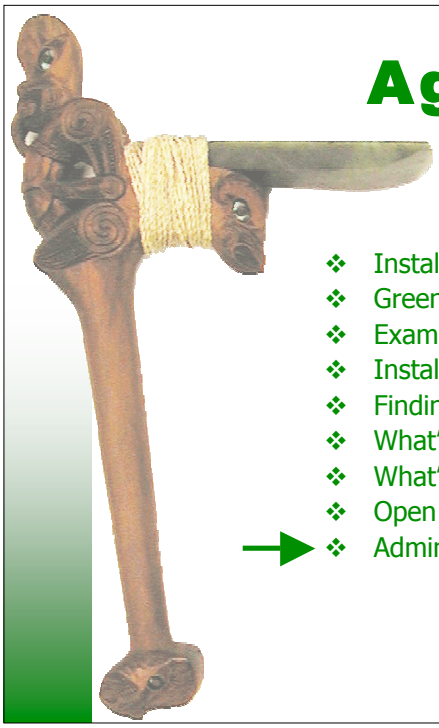
- Installing a pre-built collection
- Greenstone: platforms and versions
- Example collections
- Installing and troubleshooting
- Finding Greenstone and GLI
- What's on the CD-ROM?
- What's on my disk?
- Open source
- Admin pages

The power of open source: Greenstone uses ...

Ghostscript	**Interpreter for Adobe Postscript documents (Postscript plugin)**
Kea	**Keyphrase extraction program (to generate metadata)**
pdftohtml	**Converter for PDF documents (PDF plugin)**
rtftohtml	**Converter for RTF documents (RTF plugin)**
TextCat	**Detects languages and document encodings**
wvWare	**Converter for Word documents (Word plugin)**
Xlhtml	**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



Agenda

- Installing a pre-built collection
- Greenstone: platforms and versions
- Example collections
- Installing and troubleshooting
- Finding Greenstone and GLI
- What's on the CD-ROM?
- What's on my disk?
- Open source
- Admin pages

Admin support

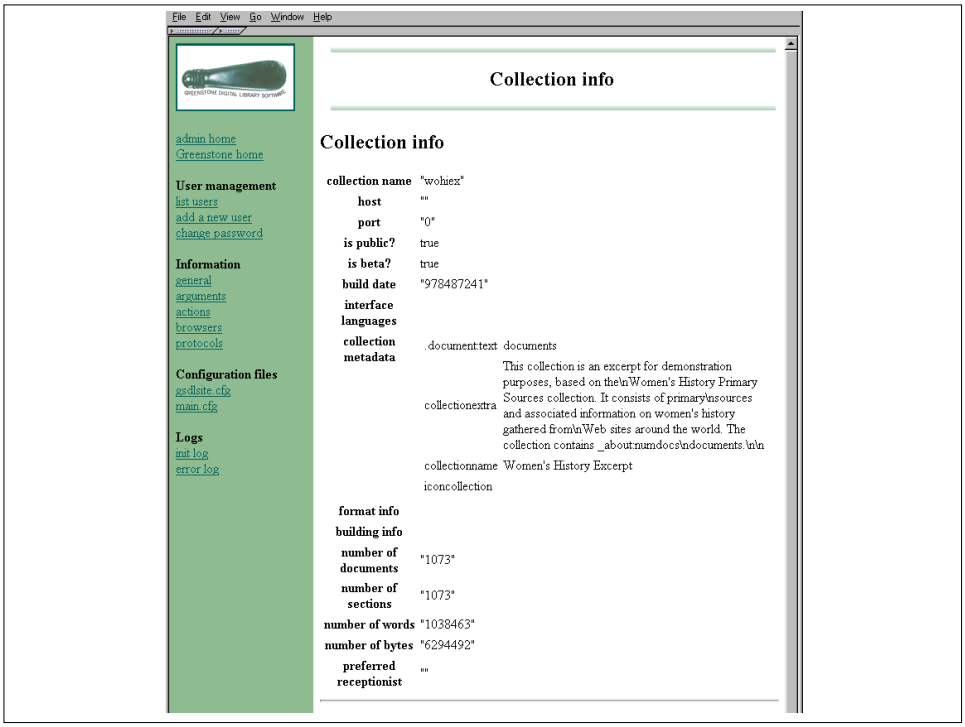
- Add new user with … privileges
  - E.g. admin, collection-building
- Check what collections are available
  - Including "private" ones not on home page
- Check summary info about a collection
  - build date
  - collection metadata
  - interface language preferences
  - number of docs/sections/words/bytes
- Logging
  - switch on or off
  - check user logs—every page access is logged
  - check system logs—errors are logged
- Notify whenever a new collection is built
- Browse technical info about the installation

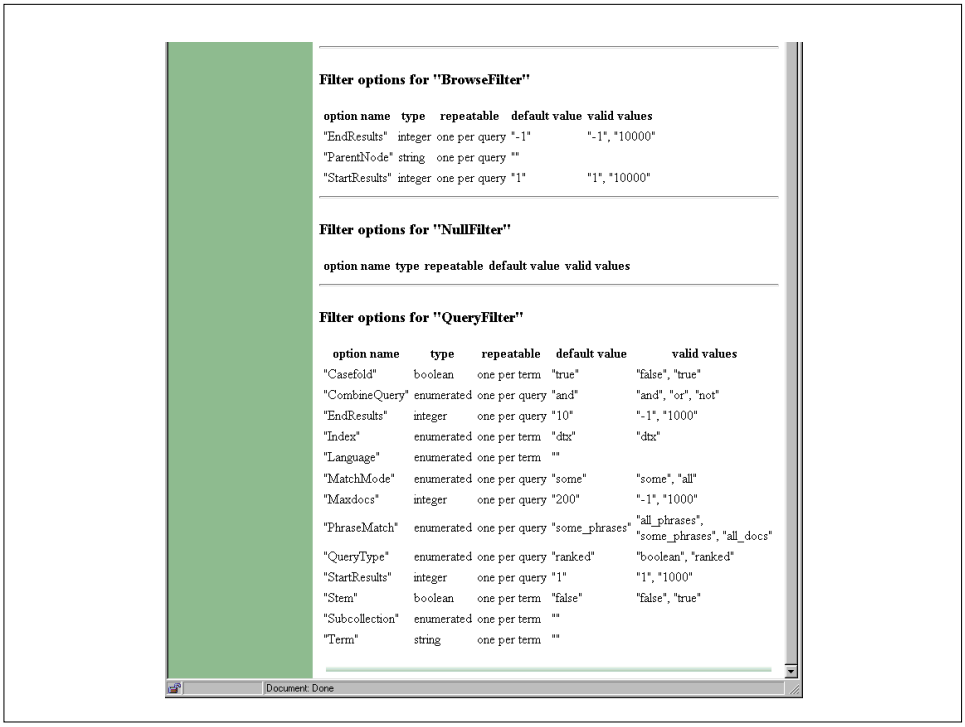
DL admin

≠

computer systems admin




File Edit View Go Window Help
admin home
Greenstone home
User management
list users
add a new user
change password
Information
general
arguments
actions
browsers
protocols
Configuration files
gsdlsite.cfg
main.cfg
Logs
mut log
error log
Collection info
Collection info
collection name "wohiex"
host ""
port "0"
is public? true
is beta? true
build date "978487241"
interface languages
collection metadata .document:text documents
collectionextra This collection is an excerpt for demonstration purposes, based on the\nWomen's History Primary Sources collection. It consists of primary\nsources and associated information on women's history gathered from\nWeb sites around the world. The collection contains _about:numdocs\ndocuments.\n\n
collectionname Women's History Excerpt
iconcollection
format info
building info
number of documents "1073"
number of sections "1073"
number of words "1038463"
number of bytes "6294492"
preferred receptionist ""


Filter options for "BrowseFilter"
option name type repeatable default value valid values
"EndResults" integer one per query "-1" "-1", "10000"
"ParentNode" string one per query ""
"StartResults" integer one per query "1" "1", "10000"
Filter options for "NullFilter"
option name type repeatable default value valid values
Filter options for "QueryFilter"
option name type repeatable default value valid values
"Casefold" boolean one per term "true" "false", "true"
"CombineQuery" enumerated one per query "and" "and", "or", "not"
"EndResults" integer one per query "10" "-1", "1000"
"Index" enumerated one per term "dtx" "dtx"
"Language" enumerated one per term ""
"MatchMode" enumerated one per query "some" "some", "all"
"Maxdocs" integer one per query "200" "-1", "1000"
"PhraseMatch" enumerated one per query "some_phrases" "all_phrases", "some_phrases", "all_docs"
"QueryType" enumerated one per query "ranked" "boolean", "ranked"
"StartResults" integer one per query "1" "1", "1000"
"Stem" boolean one per term "false" "false", "true"
"Subcollection" enumerated one per term ""
"Term" string one per term ""
Document: Done

Questions?
Comments?
Discussion?
Feedback form!