



Adding metadata— and using it



Course material prepared by

Greenstone Digital Library Project
University of Waikato, New Zealand

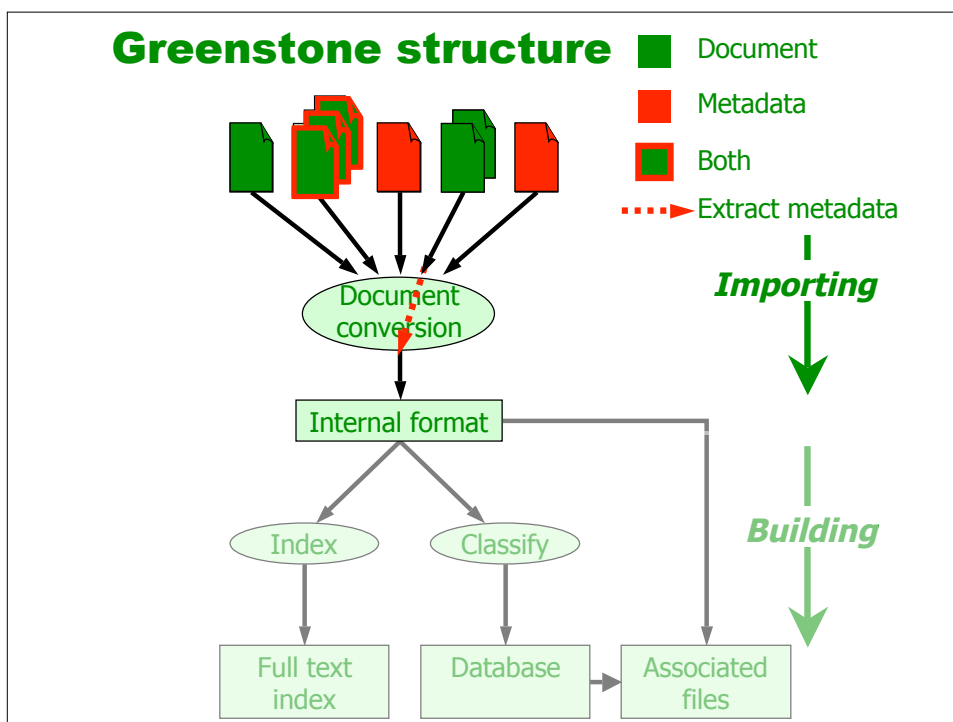


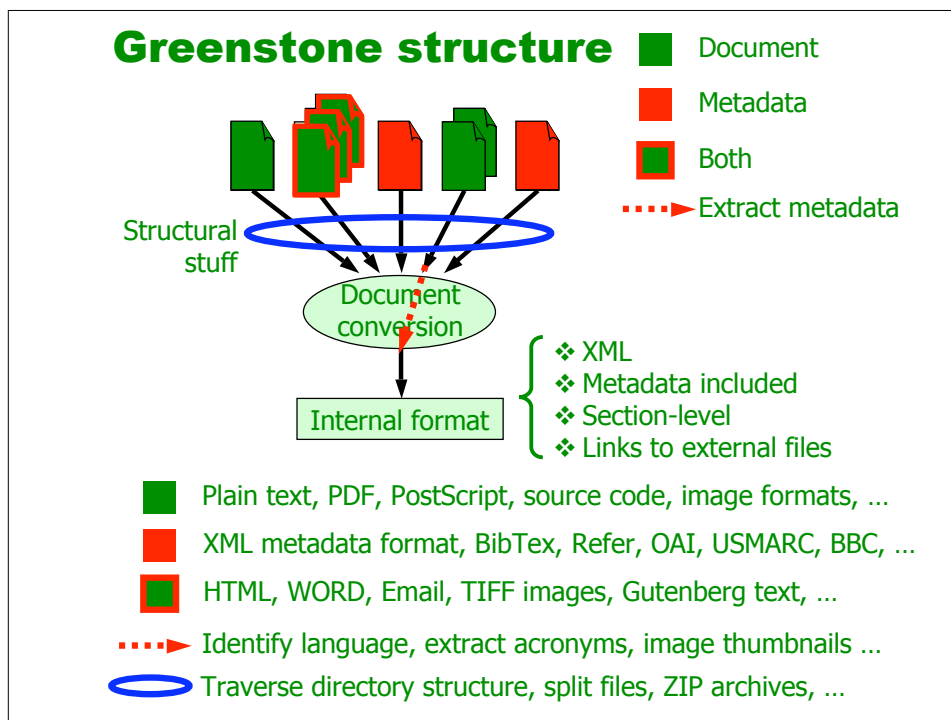
Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Collection building process

- ❖ Input: Set of documents + metadata
 - Various formats, e.g. Word, PDF, HTML, images ...
- ❖ Import: Conversion to common internal (XML) format
 - Uses third party tools
- ❖ Build: Create indexes, browsing structures, metadata database
- ❖ Output: Greenstone collection





Agenda

A wooden staff with a carved head and a yellow cord is shown on the left. A green arrow points from the staff towards the agenda list on the right.

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Metadata Sets

❖ Standard metadata sets

- Librarian added
- Good quality, but high cost
- Many types
 - ❖ Dublin Core (dc)
 - ❖ Development Library Subset (dls)
 - ❖ New Zealand Government Locator Service (nzgls)

❖ Greenstone extracted metadata (ex)

- Automatically extracted
- Non-editable
- May be poor quality, but cheap

Dublin Core metadata

Metadata	Tag	Definition
Title	dc.Title	Name given to the resource
Creator	dc.Creator	Entity primarily responsible for making the content of the resource
Subject and keywords	dc.Subject	Topic of the content of the resource
Description	dc.Description	Account of the content of the resource
Publisher	dc.Publisher	Entity responsible for making the resource available
Contributor	dc.Contributor	Entity responsible for making contributions to the content of the resource
Date	dc.Date	Date associated with an event in the life cycle of the resource
Resource type	dc.Type	Nature or genre of the content of the resource
Format	dc.Format	Physical or digital manifestation of the resource
Resource identifier	dc.Identifier	Unambiguous reference to the resource within a given context: this is the object identifier or OID
Source	dc.Source	Reference to a resource from which the present resource is derived
Language	dc.Language	Language of the intellectual content of the resource
Relation	dc.Relation	Reference to a related resource
Coverage	dc.Coverage	Extent or scope of the content of the resource
Rights management	dc.Rights	Information about rights held in and over the resource

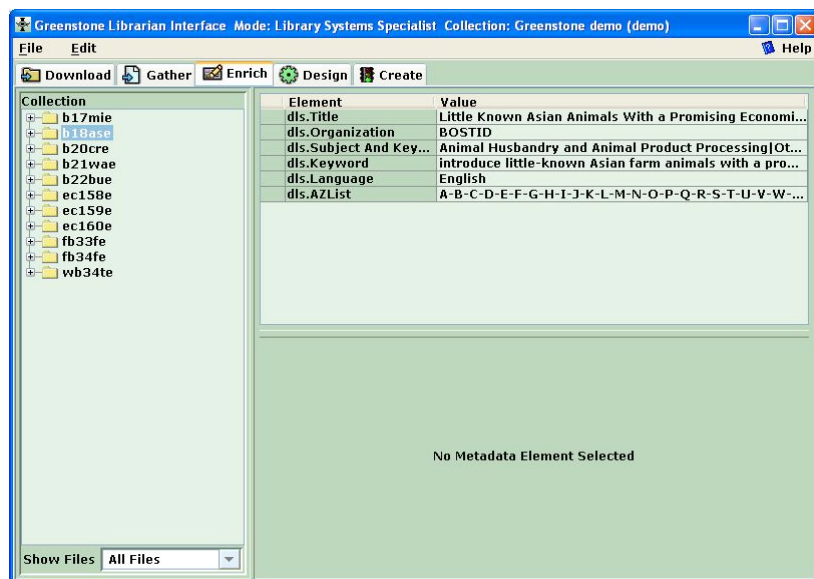
DLS metadata

Metadata

Definition

dls.Title	Title
dls.Keyword	Keyword (originally "How To")
dls.Organization	Organization that produced the document
dls.Language	Language of the document
dls.Subject and Keywords	Hierarchical subject metadata
dls.AZList	A-Z range into which this document's title falls

Example of DLS metadata



Example of DLS metadata

The screenshot shows the Greenstone Librarian Interface (Mode: Library Systems Specialist) for the 'Greenstone demo (demo)' collection. The interface displays a list of collections on the left, including b17mie, b18ase, b20cre, b21wae, b22bue, ec158e, ec159e, ec160e, fb33fe, fb34fe, and wb34te. The main panel shows the DLS metadata for the selected collection, b18ase. The metadata is organized into a table with two columns: Element and Value.

Element	Value
dls.Title	an Animals With a Promising Economic Future (b18ase)
dls.Organization	BOSTID
dls.Subject And Key...	Animal Husbandry and Animal Product Processing OT...
dls.Keyword	introduce little-known Asian farm animals with a pro...
dls.Language	English
dls.AZList	A-B-C-D-E-F-G-H-I-J-K-L-M-N-O-P-Q-R-S-T-U-V-W-...

Below the table, there is a section titled 'Existing values for dls.Title' which lists several related titles and their corresponding collection codes:

- Butterfly Farming in Papua New Guinea (b22bue)
- Development in practice: Toward Gender Equality (wb34te)
- Farming snails 1: Learning about snails; Building a pen; Food and shelter plant
- Farming snails 2: Choosing snails; Care and harvesting; Further improvement
- Little Known Asian Animals With a Promising Economic Future (b18ase)**
- Managing Tropical Animal Resources - Crocodiles as a Resource for the Tropics
- Microlivestock - Little-Known Small Animals with a Promising Economic Future
- The Courier - N°158 - July - August 1996 Dossier Communication and the med
- The Courier - N°159 - Sept- Oct 1996 Dossier Investing in People Country Rep
- The Courier - N°160 - Nov - Dec 1996 - Dossier Habitat - Country reports: Fiji
- The Water Buffalo: New Prospects For An Underutilized Animal (1984) (b21w

Example of DLS metadata

The screenshot shows the Greenstone Librarian Interface (Mode: Expert) for the 'Greenstone demo (demo)' collection. The interface displays a list of collections on the left, including b17mie, b18ase, b20cre, b21wae, b22bue, ec158e, ec159e, ec160e, fb33fe, fb34fe, and wb34te. The main panel shows the DLS metadata for the selected collection, b18ase. The metadata is organized into a table with two columns: Element and Value.

Element	Value
dls.Title	Little Known Asian Animals With a Promising Econo...
dls.Organization	BOSTID
dls.Subject And Key...	Animal Husbandry and Animal Product Processing O...
dls.Keyword	introduce little-known Asian farm animals with a pro...
dls.Language	English
dls.AZList	A-B-C-D-E-F-G-H-I-J-K-L-M-N-O-P-Q-R-S-T-U-V-W-...

Below the table, there is a message: 'This piece of metadata is inherited from a folder above it, and cannot be edited. You can visit the actual folder containing the metadata by clicking on the folder icon at the start of the row.'

Extracted metadata

Metadata

Definition

ex.Title	Title extracted from HTML, Word, PDF
ex.Source	Name of source file, e.g. Apache.html
ex.Language	Language of document, e.g. en, sp, fr
ex.Encoding	Encoding of document, e.g. iso_8859_1

...

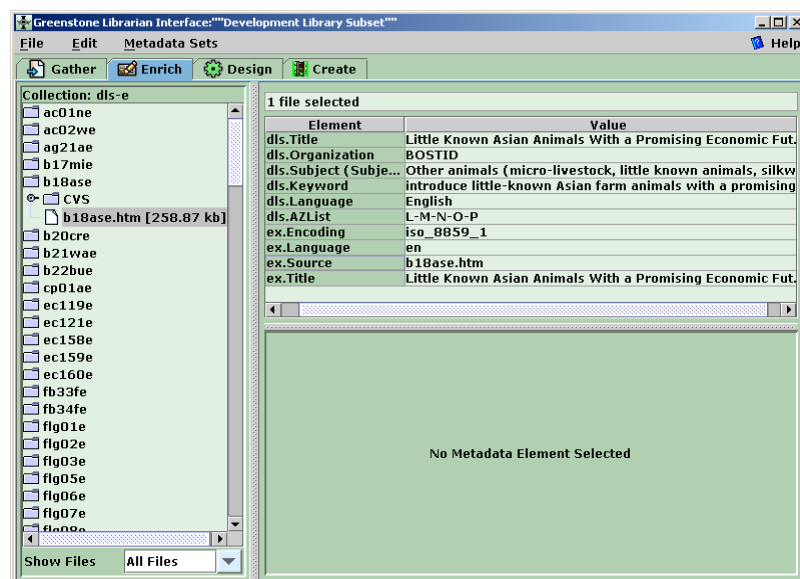
...

ex.Acronym	Acronyms that appear in the document
------------	--------------------------------------

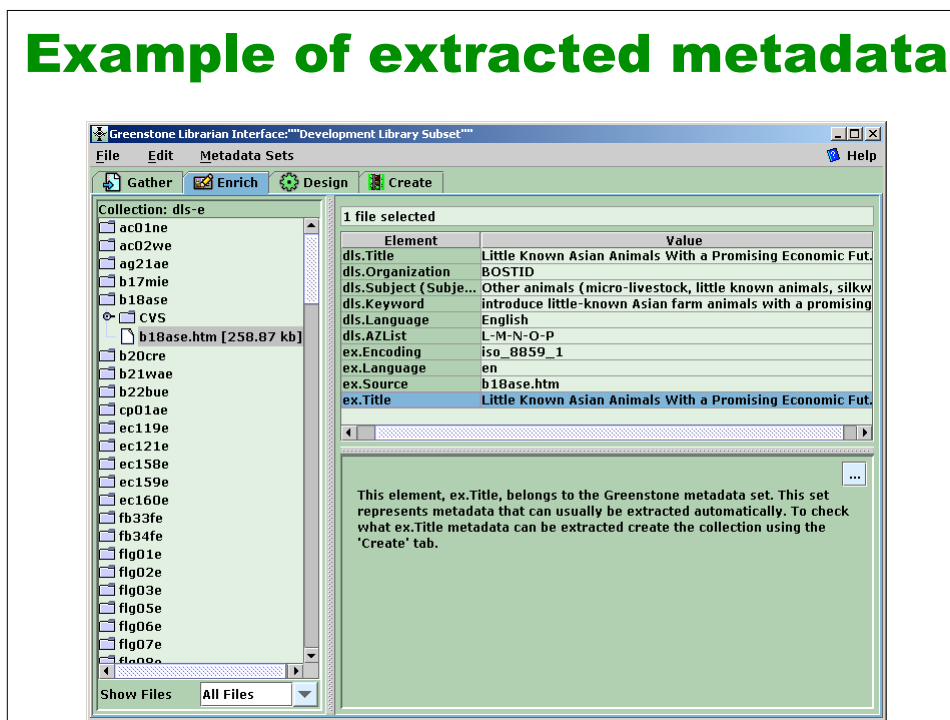
...

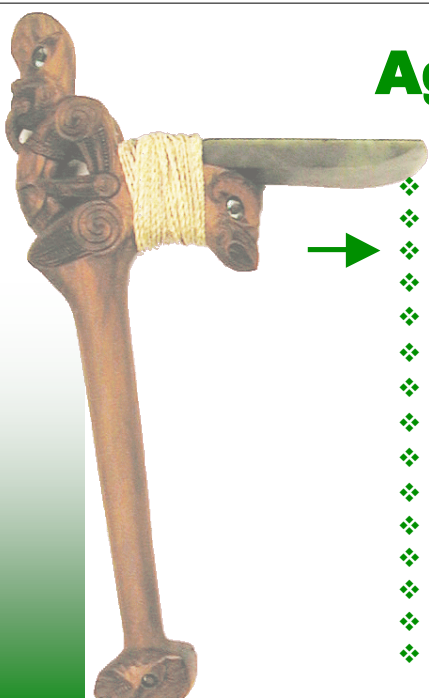
...

Example of extracted metadata



Example of extracted metadata





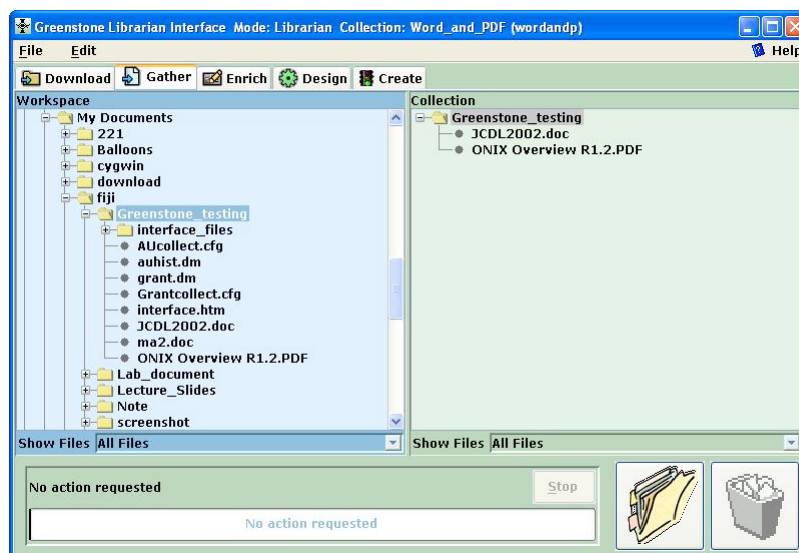
Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

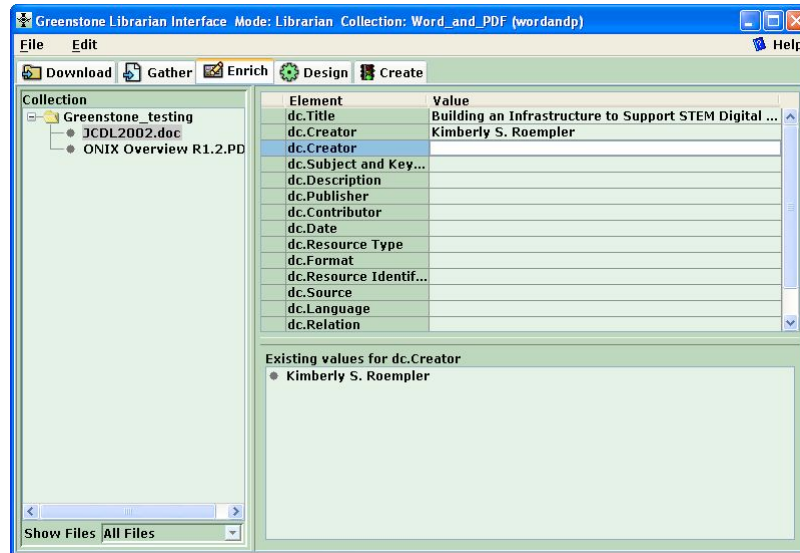
GLI: Adding metadata

- ❖ Enrich pane
- ❖ Folder level or document level
- ❖ Specified metadata set
- ❖ Can view extracted metadata but not edit

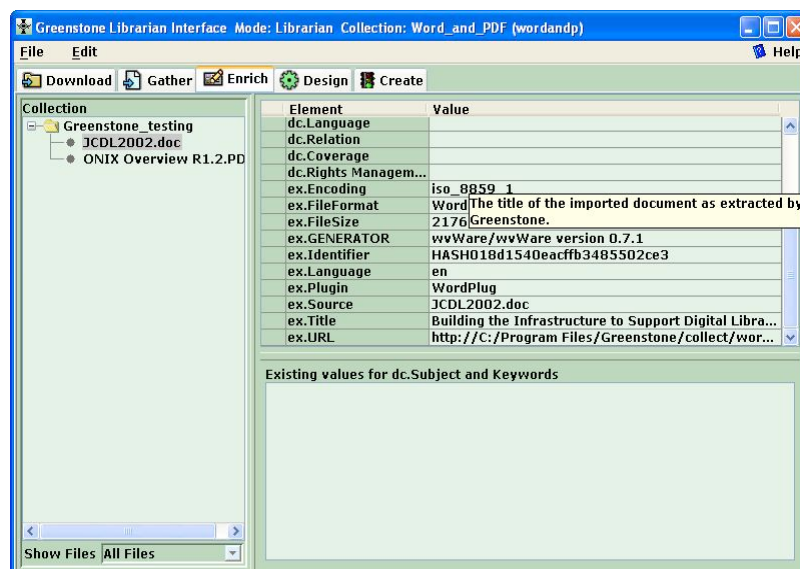
GLI: adding metadata



GLI: adding metadata



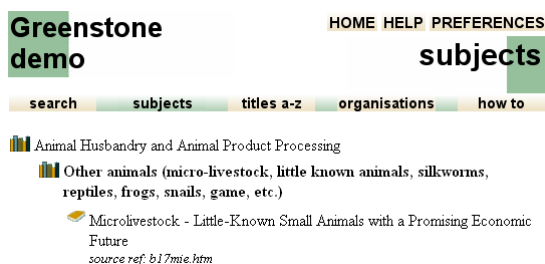
Extracted Metadata



Greenstone Librarian Interface demo: adding metadata

Hierarchical Metadata

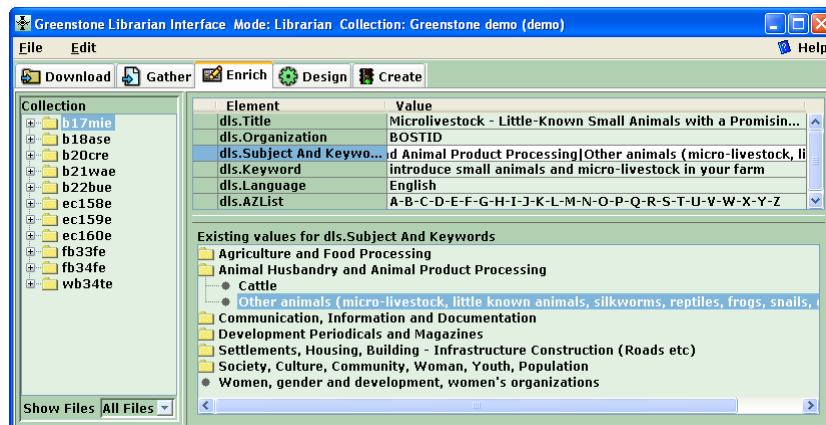
❖ Metadata is often hierarchical:



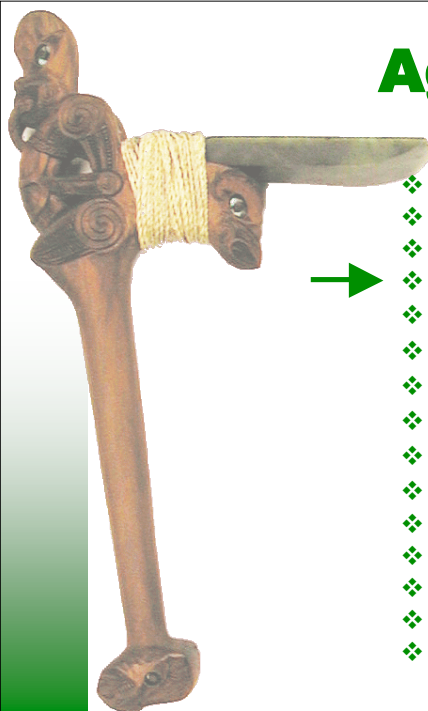
❖ How is this done in Greenstone?

Hierarchical Metadata

- ❖ In the GLI's Enrich pane, use the pipe character (|) to separate the levels



- ❖ Preview the hierarchy in the value tree

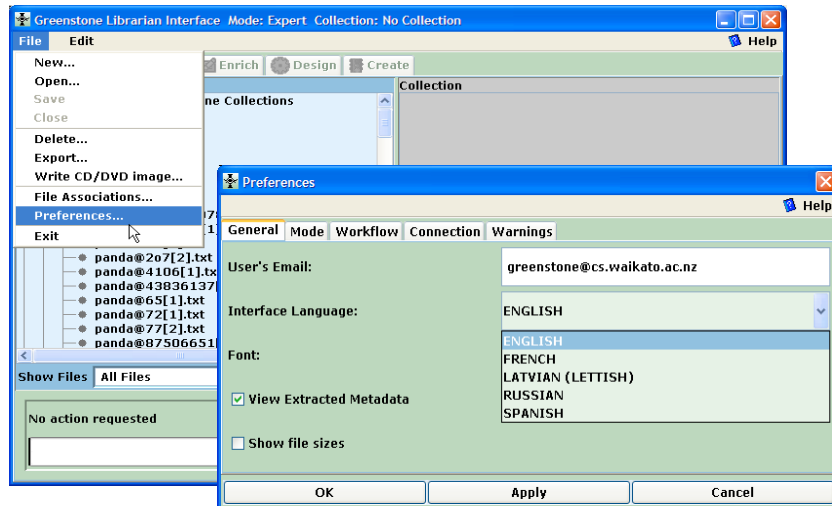


Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

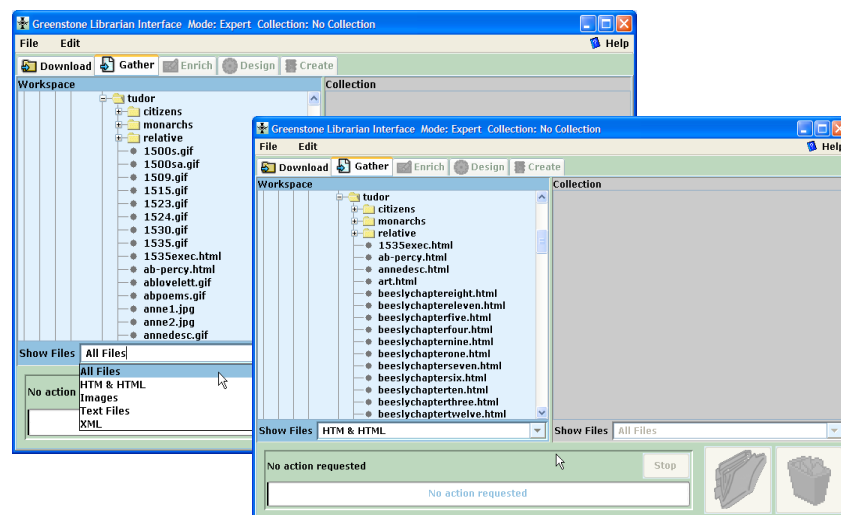
GLI Tricks

❖ General Preferences



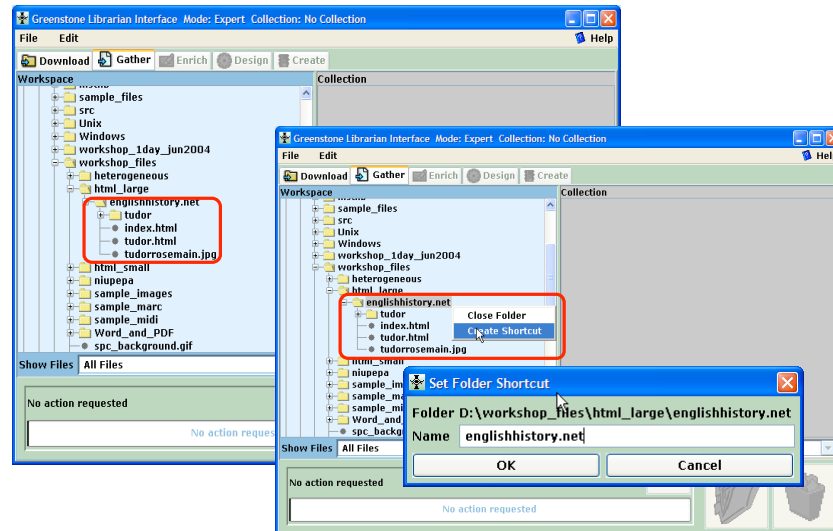
GLI Tricks

❖ File Filtering

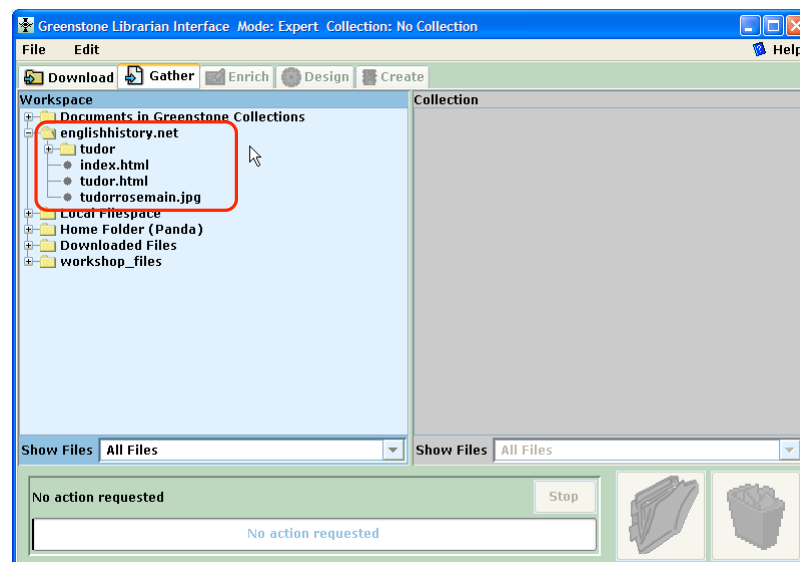


GLI Tricks

❖ Create file folder shortcut

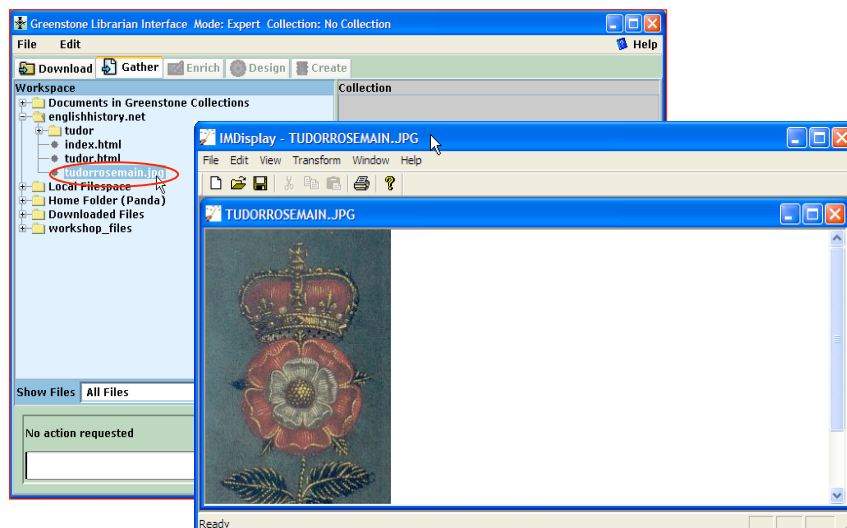


GLI Tricks



GLI Tricks

❖ Viewing files: Double-clicking



Agenda

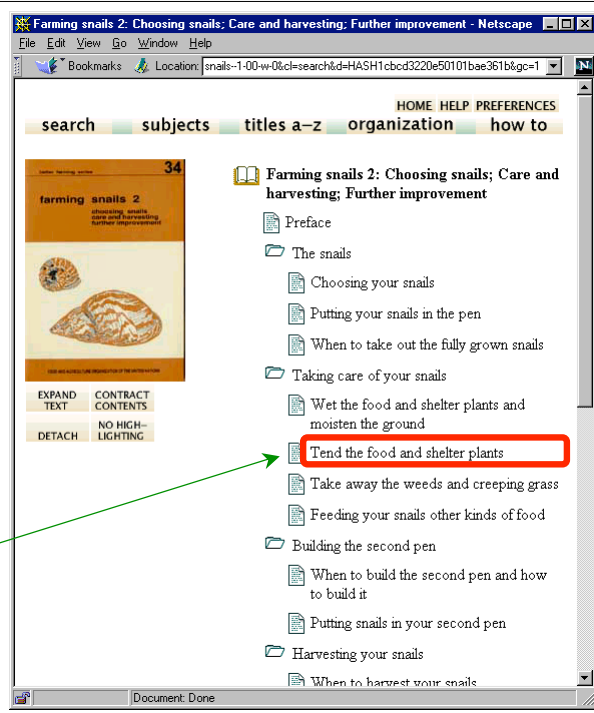


- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Hierarchical document model

❖ Metadata specified at any level

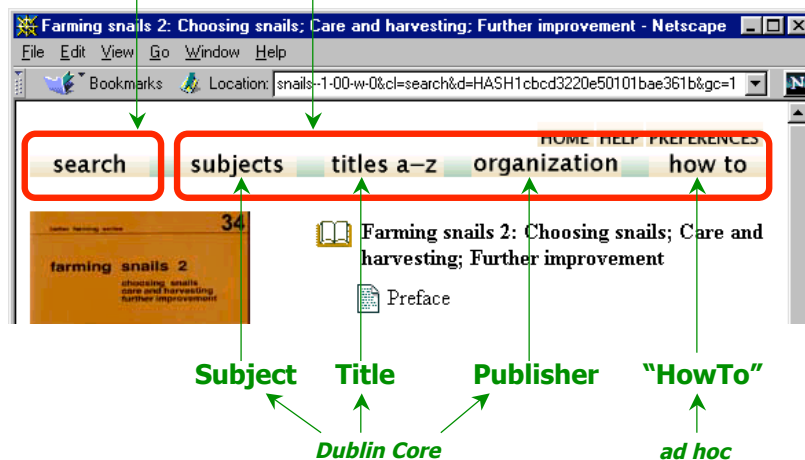
Title metadata



Searching and browsing

❖ Searching

❖ Metadata-based browsing



Multiple search indexes

text

metadata

Greenstone demo

HOME HELP PREFERENCES about

search subjects titles a-z organisations how to

Search for chapters that contain some of the words

chapters
chapters
section titles
entire documents

Begin Search

About this collection

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

How to find information in the Greenstone demo collection

There are 5 ways to find information in this collection:

- search for particular words
- access publications by subject
- access publications by title
- access publications by organization
- access publications by "how to" listing

You can search for particular words that appear in the text from the "search" page. This is the first page that comes up when you begin, and can be

Collection-dependent

metadata

computer science bibliographies

HOME HELP PREFERENCES search

search browse

Search for references which contain some of the words

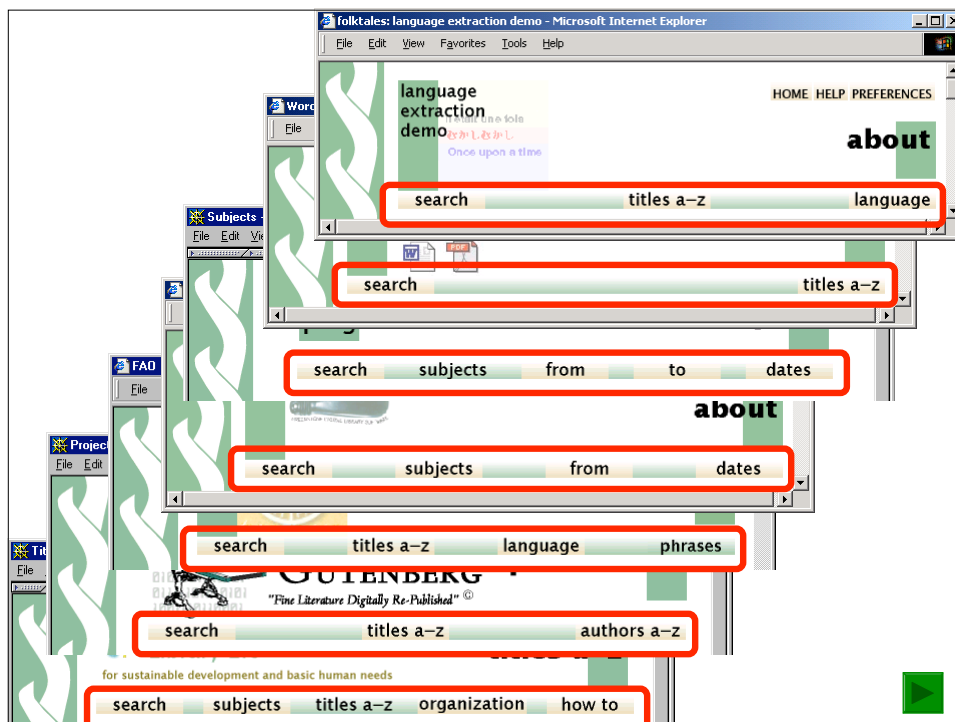
Honkala titles
authors
conference/publication titles
references
references and abstracts

Begin Search

results

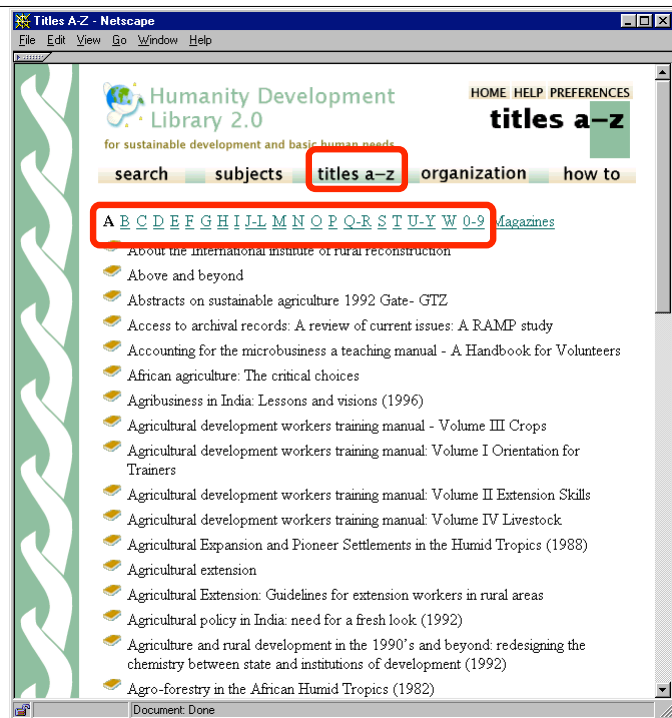
Word count: 1714 words
4 documents matched the query.

- Juha Honkala (1994) "On generalized DTOL systems and their fixed points" *Theoretical Computer Science* 127(2), 269-286. From [Bibliography for the journal "Theoretical Computer Science"](#)
- Juha Honkala (1993) "On DOL systems with immigration" *Theoretical Computer Science* 120(2), 229-245. From [Bibliography for the journal "Theoretical Computer Science"](#)
- Juha Honkala (1991) "On algebraic generalized zeta functions of formal power series" *Theoretical Computer Science* 79(1), 263-273. From [Bibliography of "Theoretical Computer Science"](#)
- Juha Honkala and Werner Kuich (1997) "On Lindenmayerian algebraic power series" *Theoretical Computer Science* 183(1), 113-142. From [Bibliography of "Theoretical Computer Science"](#)

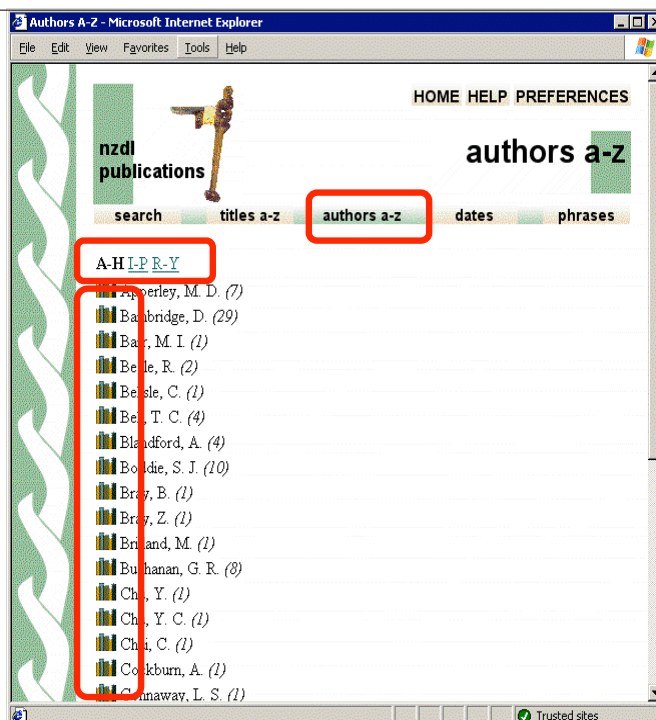


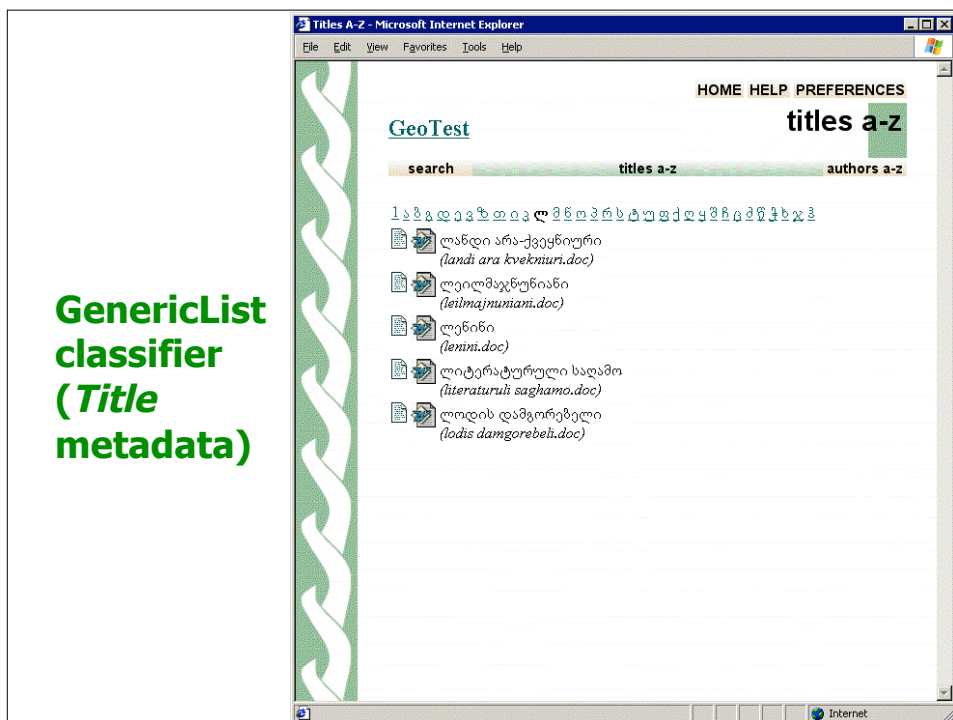
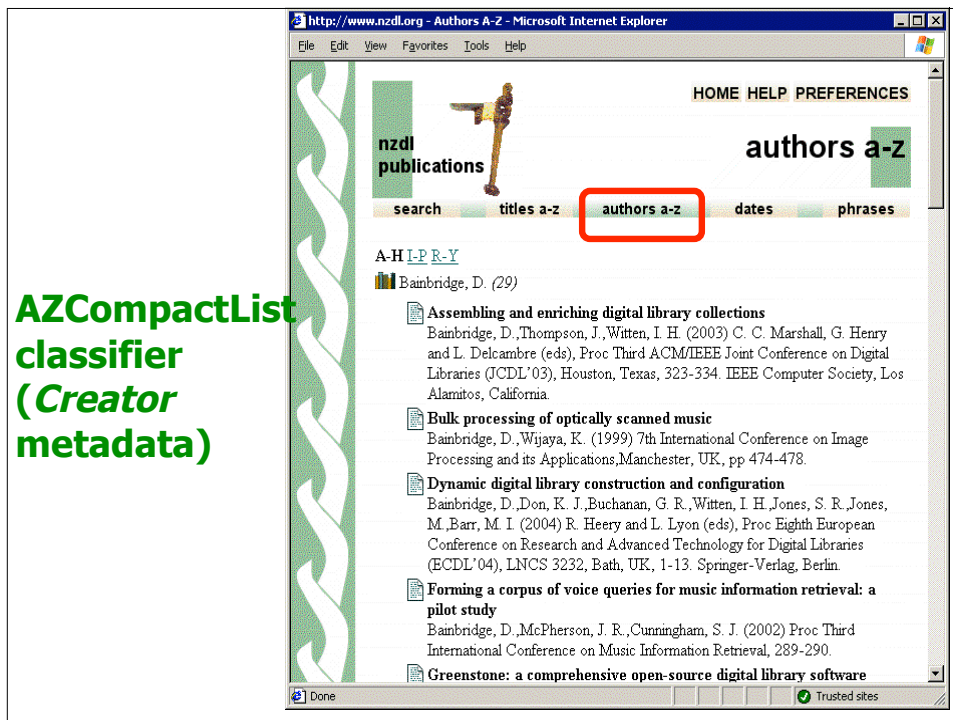
Browsing using classifiers

**AZList
classifier
(Title
metadata)**

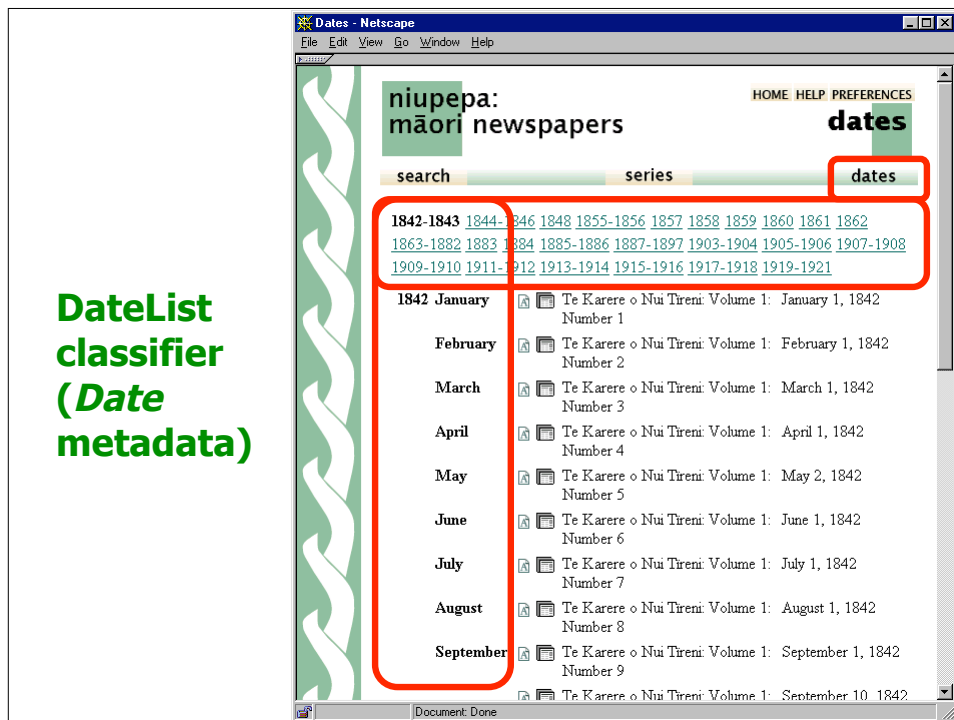


**AZCompactList
classifier
(Creator
metadata)**

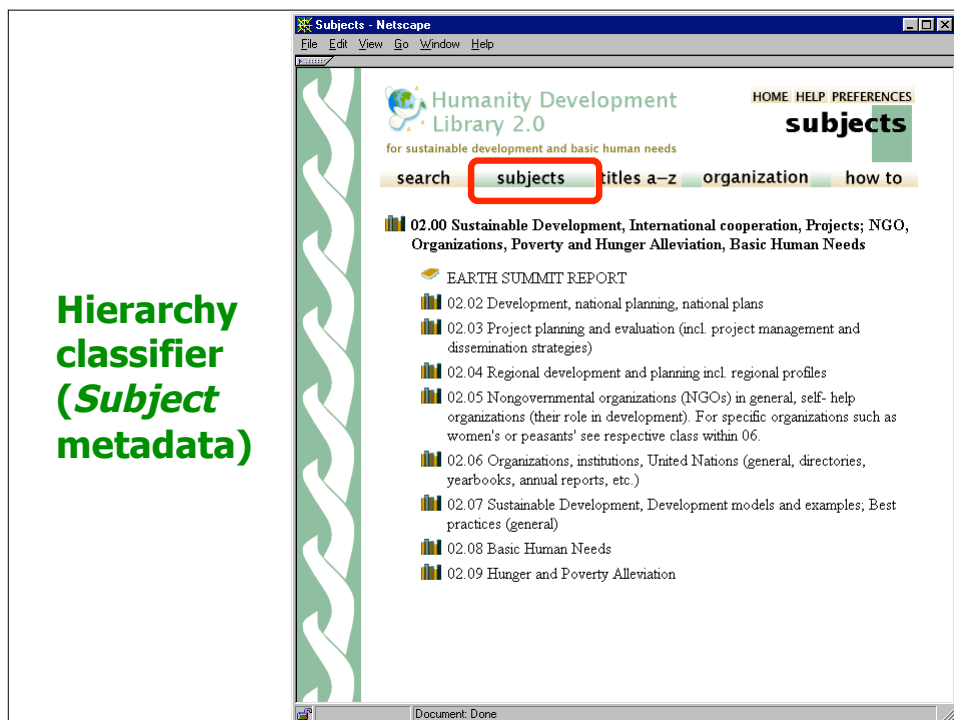


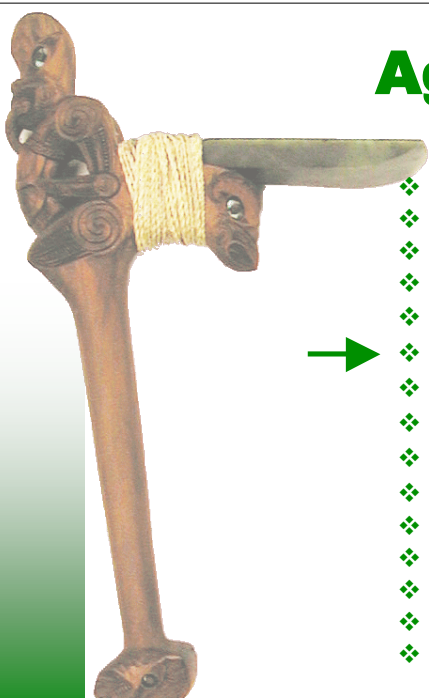
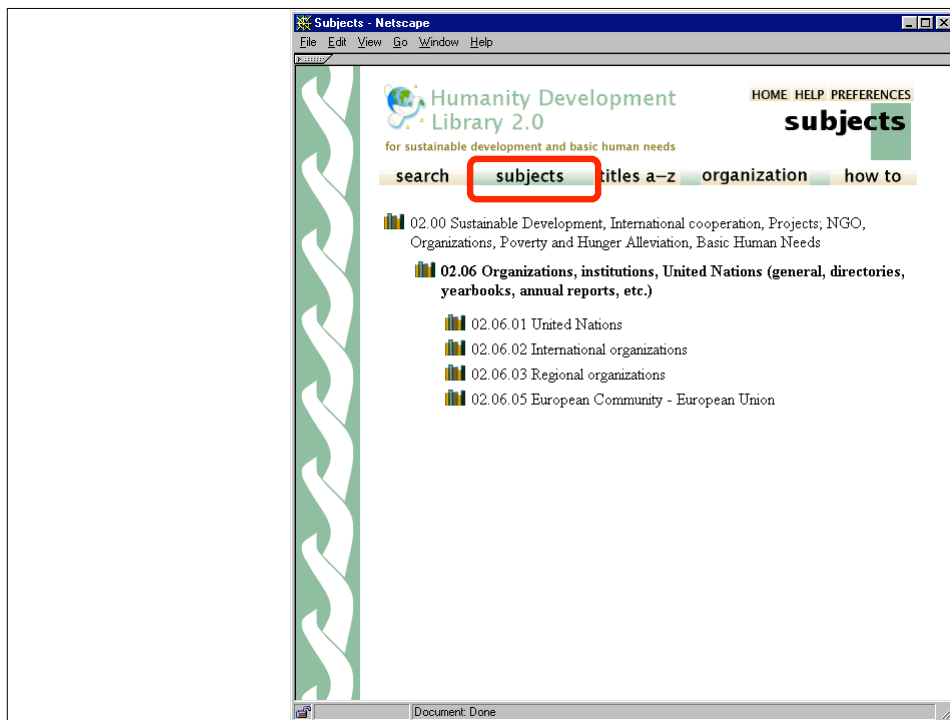


**DateList
classifier
(Date
metadata)**



**Hierarchy
classifier
(Subject
metadata)**





Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ➔ ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

General Options

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: Greenstone demo

Collection folder: demo

URL to 'about page' icon: _httpprefix_/collect/demo/images/en/de Browse...

URL to 'home page' icon: _httpprefix_/collect/demo/images/en/de Browse...

☒ This collection should be publicly accessible

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

greenstone demo - Netscape

File Edit View Go Window Help

Back Forward Reload Home Search Guide Print Security Stop

Bookmarks Location: http://www.nzdl.org/fast-cgi-bin/libray?a=p&p=about&c=demo

demo collection 1

HOME HELP PREFERENCES

about

search subjects titles a-z organization how to

Search for chapters which contain some of the words

Begin Search

About this collection

This is a demonstration collection for the Greenstone digital library software. It contains a small subset (11 books) of the Humanity Development Library

How to find information

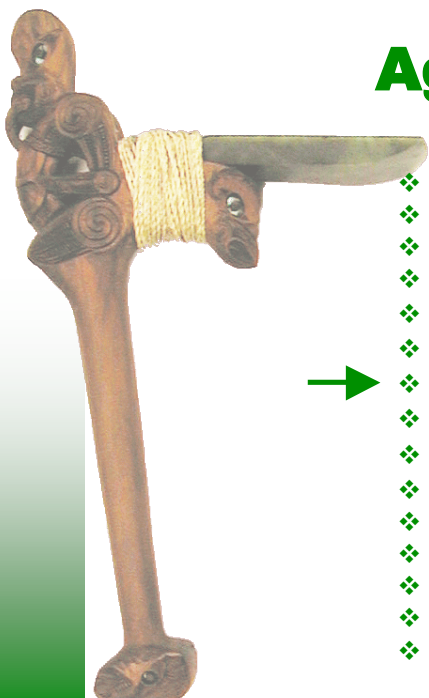
There are 5 ways to find information in this collection:

- search for particular words
- access publications by subject
- access publications by title
- access publications by organization
- access publications by how to listing

You can search for particular words that appear in the text from the "search" page. This is the first page that comes up when you begin, and can be reached from other pages.

demo collection 2

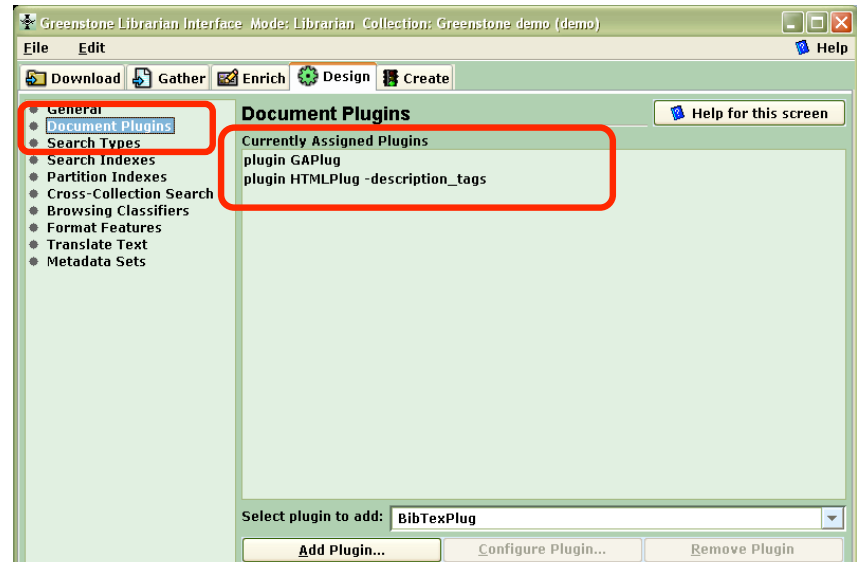
Help text generated automatically



Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ➔ ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Plugins



Greenstone Librarian Interface. Mode: Librarian. Collection: Greenstone demo (demo)

File Edit Help

Download Gather Enrich Design Create

General
Document Plugins
Search Types

Search Indexes
Partition Indexes
Cross-Collection Search
Browsing Classifiers
Format Features
Translate Text
Metadata Sets

Document Plugins

Help for this screen

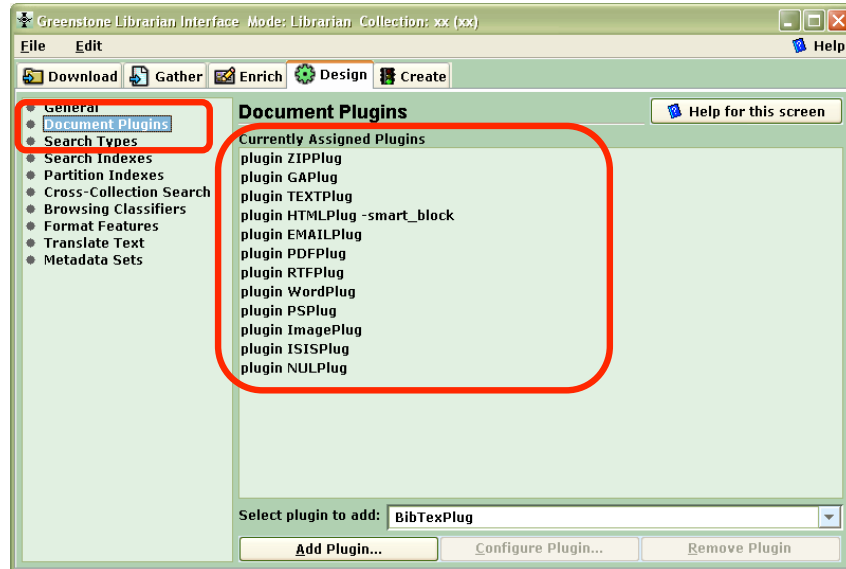
Currently Assigned Plugins

- plugin GAPlug
- plugin HTMLPlug -description_tags

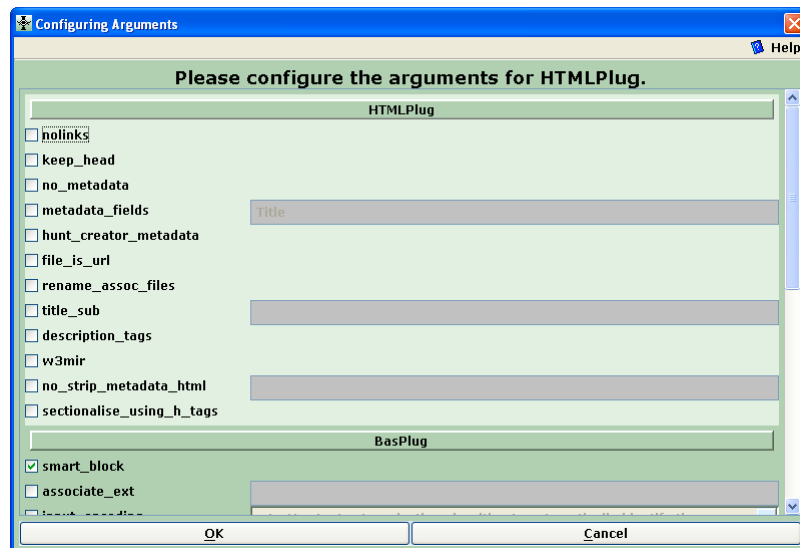
Select plugin to add: BibTexPlug

Add Plugin... Configure Plugin... Remove Plugin

Plugins



Plugin Options



Plugins

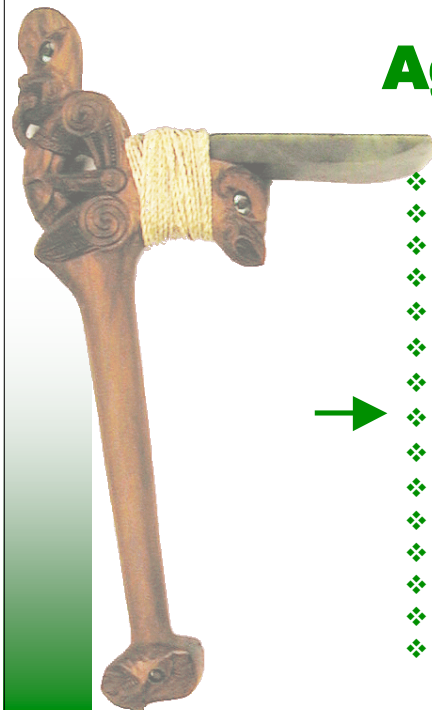
Used by collection-building software to accomplish format-specific parsing of source documents

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
ImagePlug
PPTPlug
ISISPlug
TCCPlug ...

- ❖ GAPug processes doc.xml files generated during *import*
- ❖ ArcPlug processes filelist in *archives.inf*
- ❖ RecPlug recurses through a directory structure

Agenda



- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Search Indexes

Greenstone Librarian Interface Mode: Librarian Collection: Greenstone demo (demo)

Index Selection

Here you choose what searchable indexes the collection will have.
 To add a new index, select what material is to be indexed, choose the level of the index, and click 'Add Index'.
 To remove an index, select it from the assigned indexes list and click 'Remove Index'.
 To set the default index, select it from the assigned indexes list and click 'Set'.

Assigned Indexes

- section:text "chapters"
- section:ex.Title "section titles"
- document:text "entire documents"

Index Name: section titles

Build index on:

- ☐ ex.Language
- ☐ ex.Plugin
- ☐ ex.Source
- ☒ ex.Title
- ☐ ex.URL
- ☐ text

At the level: section

Buttons: Add Index, Replace Index, Remove Index

Annotations:

- name on menu (points to 'section titles')
- "text", or any metadata (points to 'ex.Title')
- document, section, or paragraph (points to 'section')

Greenstone demo HOME HELP PREFERENCES about

search subjects titles a-z organisations how to

Search for chapters that contain some of the words

chapters
section titles
entire documents

About this collection

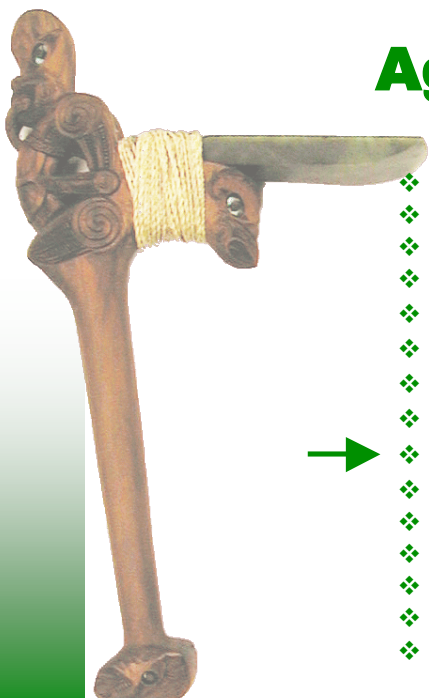
This is a demonstration collection for the Greenstone digital library software. It contains a small subset (11 documents) of the Humanitarian and Development Libraries.

How to find information in the Greenstone demo collection

There are 5 ways to find information in this collection:

- search for particular words
- access publications by subject
- access publications by title
- access publications by organization
- access publications by "how to" listing

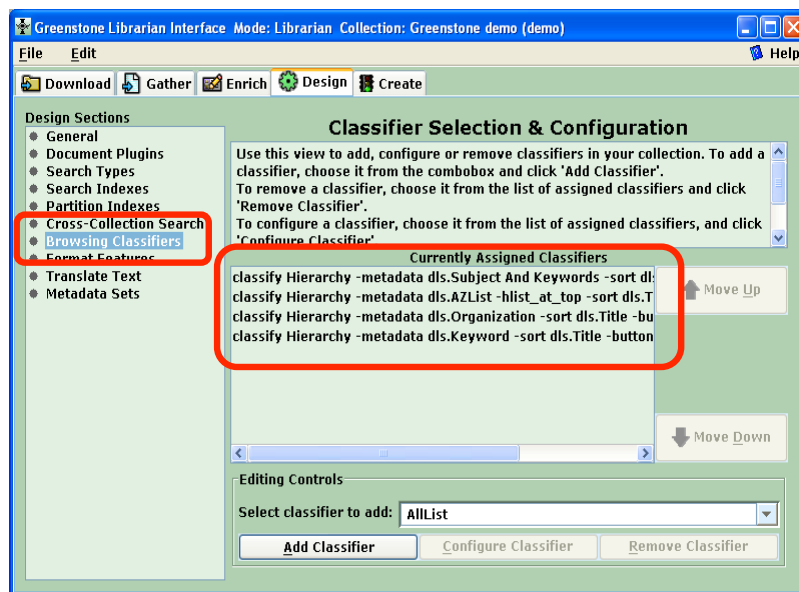
You can search for particular words that appear in the text from the "search" page. This is the first page that comes up when you begin, and can be

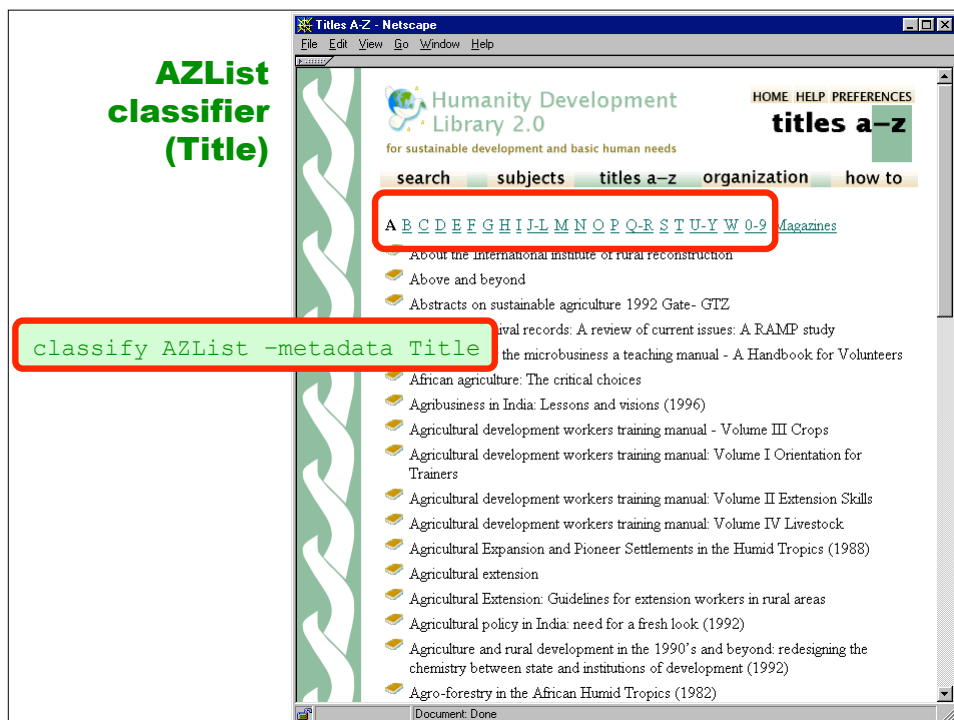
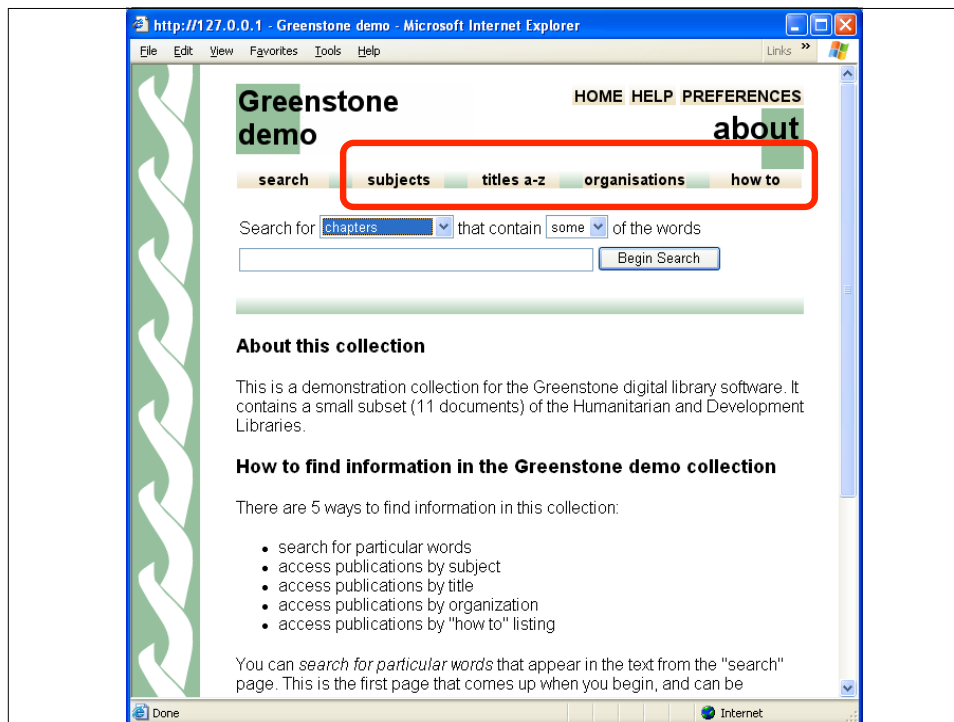


Agenda

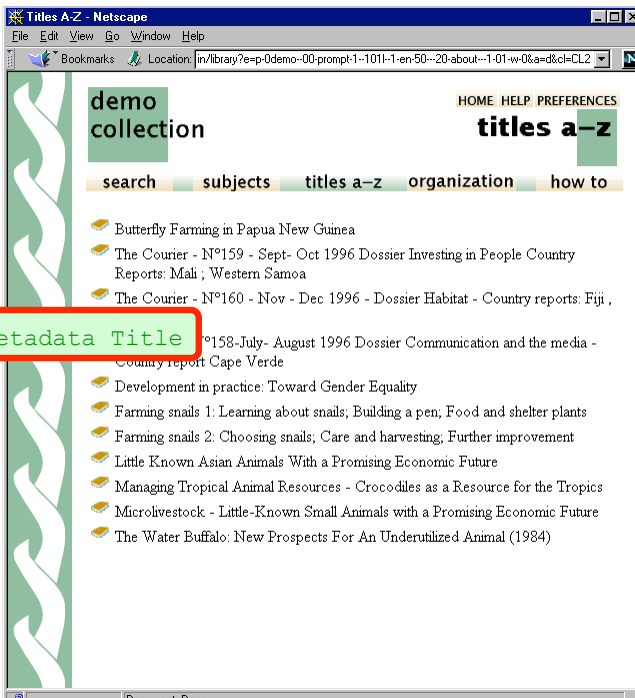
- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Browsing Classifiers





AZList classifier (Title) for the demo collection



demo collection

HOME HELP PREFERENCES

titles a-z

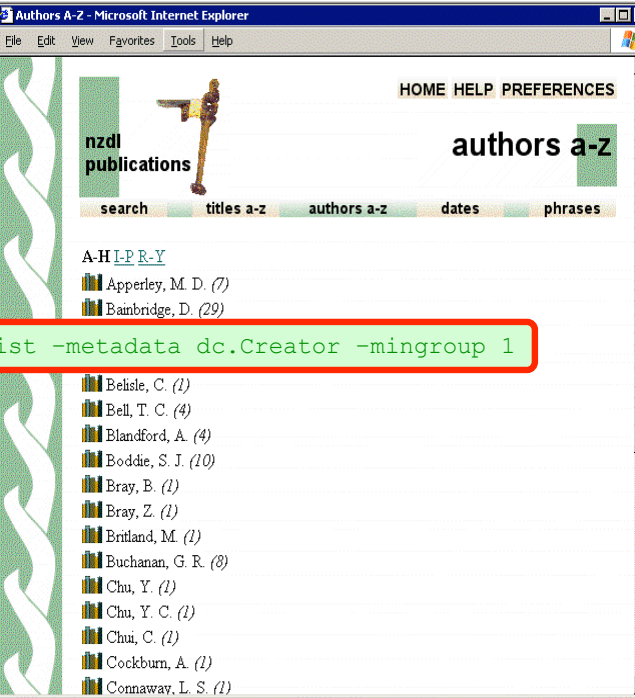
search subjects titles a-z organization how to

- Butterfly Farming in Papua New Guinea
- The Courier - N°159 - Sept- Oct 1996 Dossier Investing in People Country Reports: Mali ; Western Samoa
- The Courier - N°160 - Nov - Dec 1996 - Dossier Habitat - Country reports: Fiji ,
- 158-July- August 1996 Dossier Communication and the media - Country report Cape Verde
- Development in practice: Toward Gender Equality
- Farming snails 1: Learning about snails; Building a pen; Food and shelter plants
- Farming snails 2: Choosing snails; Care and harvesting; Further improvement
- Little Known Asian Animals With a Promising Economic Future
- Managing Tropical Animal Resources - Crocodiles as a Resource for the Tropics
- Microlivestock - Little-Known Small Animals with a Promising Economic Future
- The Water Buffalo: New Prospects For An Underutilized Animal (1984)

Document: Done

```
classify AZList -metadata Title
```

AZCompactList classifier (Creator)



nzdl publications

HOME HELP PREFERENCES

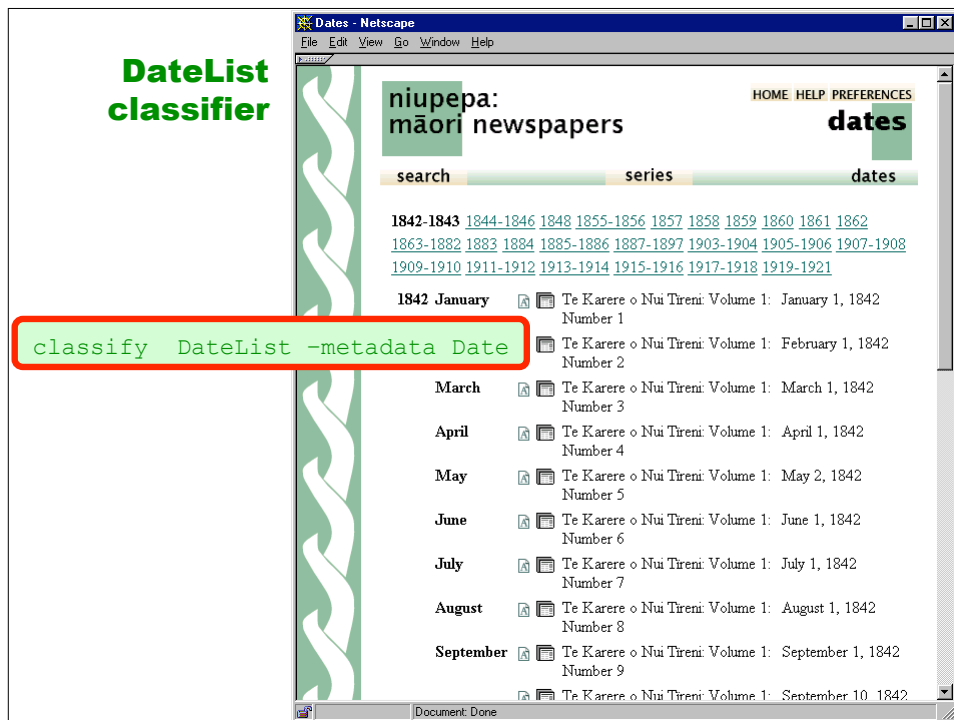
authors a-z

search titles a-z authors a-z dates phrases

A-H-I-P-R-Y

- Apperley, M. D. (7)
- Bainbridge, D. (29)
- Beksle, C. (1)
- Bell, T. C. (4)
- Blandford, A. (4)
- Boddie, S. J. (10)
- Bray, B. (1)
- Bray, Z. (1)
- Britland, M. (1)
- Buchanan, G. R. (8)
- Chu, Y. (1)
- Chu, Y. C. (1)
- Chui, C. (1)
- Cockburn, A. (1)
- Connaway, L. S. (1)

```
classify AZCompactList -metadata dc.Creator -mingroup 1
```



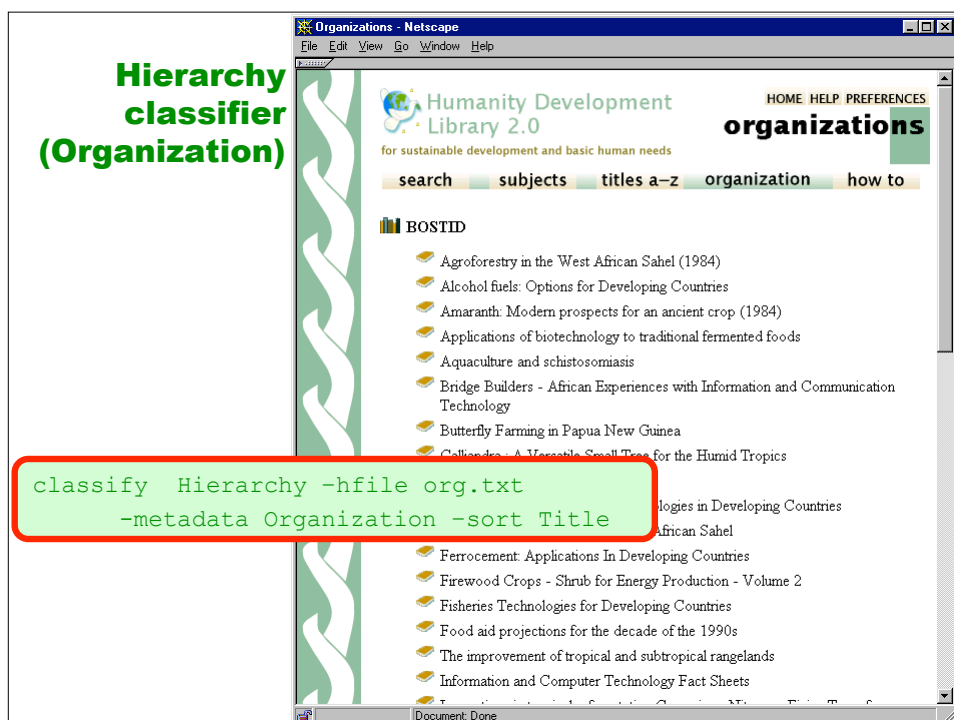
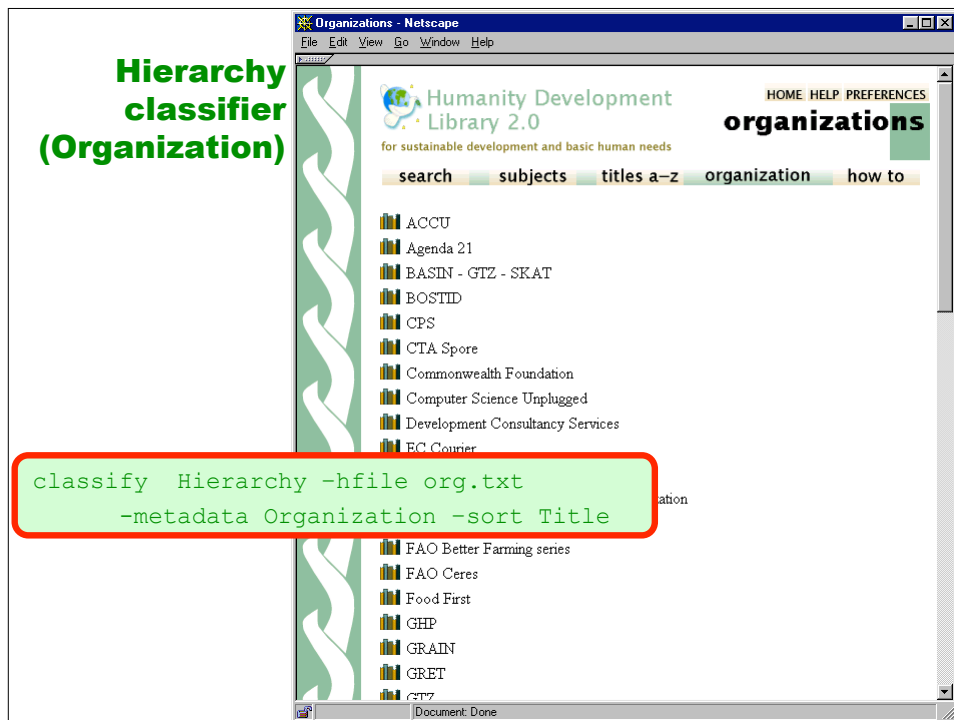
search subjects titles a-z **organization** how to

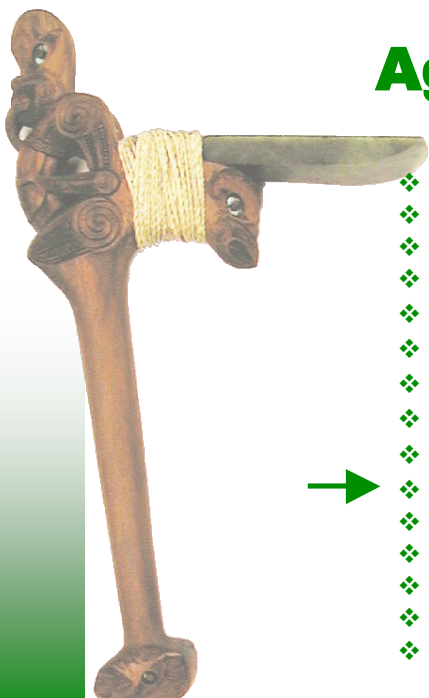
`classify Hierarchy -hfile org.txt
 -metadata Organization -sort Title`

etc/org.txt

identifier (matches *Organization* metadata value)
 position in the hierarchy that the browser implements
 title to be displayed in hierarchy browser

accu	1	ACCU
ag21	2	"Agenda 21"
bgz	3	"BASIN - GTZ - SKAT"
bostid	4	BOSTID
cps	5	CPS
cpas	6	"CTA Spore"
cf	7	"Commonwealth Foundation"
csu	8	"Computer Science Unplugged"
dcs	9	"Development Consultancy Services"
ecc	11	"EC Courier"
ecdg8	12	"EC DG8"
echo	13	"Educational Concerns for Hunger Organization"
fao	14	FAO
faobfs	15	"FAO Better Farming series"
ceres	16	"FAO Ceres"
ff	17	"Food First"

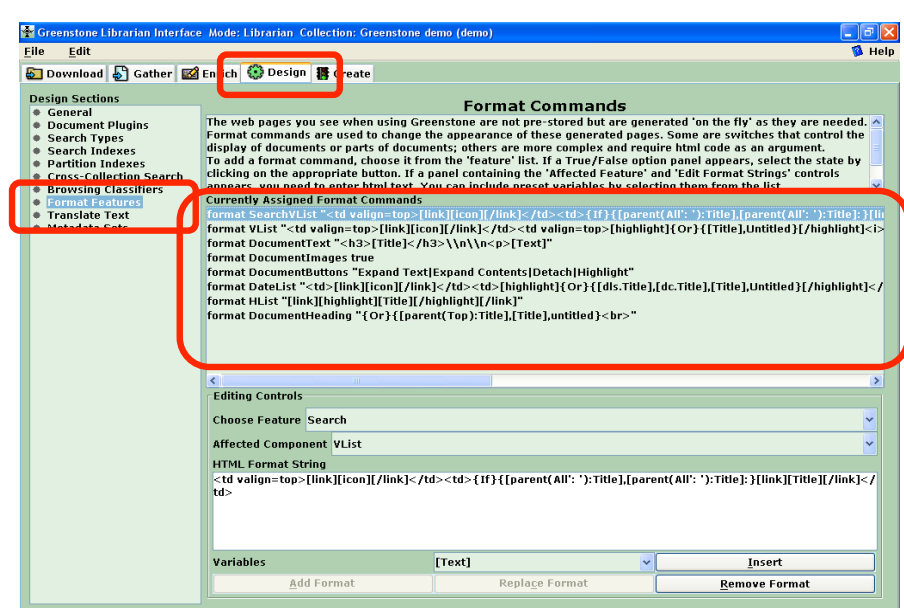




Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Format Features



Greenstone Librarian Interface Mode: Librarian Collection: Greenstone demo (demo)

Format Commands

The web pages you see when using Greenstone are not pre-stored but are generated 'on the fly' as they are needed. Format commands are used to change the appearance of these generated pages. Some are switches that control the display of documents or parts of documents; others are more complex and require html code as an argument. To add a format command, choose it from the 'feature' list. If a True/False option panel appears, select the state by clicking on the appropriate button. If a panel containing the 'Affected Feature' and 'Edit Format Strings' controls appears, you need to enter html text. You can include `parent` variables by selecting them from the list.

Currently Assigned Format Commands

```
format SearchVList "<td valign=top>[link][icon]/[link]</td><td>{if}{parent(All:):Title},{parent(All:):Title}</td>[link][icon]/[link]</td><td>[highlight]{Or}{[Title],Untitled}}[/highlight]</td>"
format DocumentText "<h3>[Title]</h3>\n\n<p>[Text]"
format DocumentButtons "Expand Text|Expand Contents|Detach|Highlight"
format DateList "<td>[link][icon]/[link]</td><td>[highlight]{Or}{[dls.Title],[dc.Title],[Title],Untitled}}[/highlight]</td>"
format HList "[link][highlight][Title]/[highlight]/[link]"
format DocumentHeading "{Or}{parent(Top):Title],[Title],untitled}<br>"
```

Editing Controls

Choose Feature: Search

Affected Component: VList

HTML Format String

```
<td valign=top>[link][icon]/[link]</td><td>{if}{parent(All:):Title},{parent(All:):Title}</td>[link][icon]/[link]</td><td>[highlight]{Or}{[Title],Untitled}}[/highlight]</td>
```

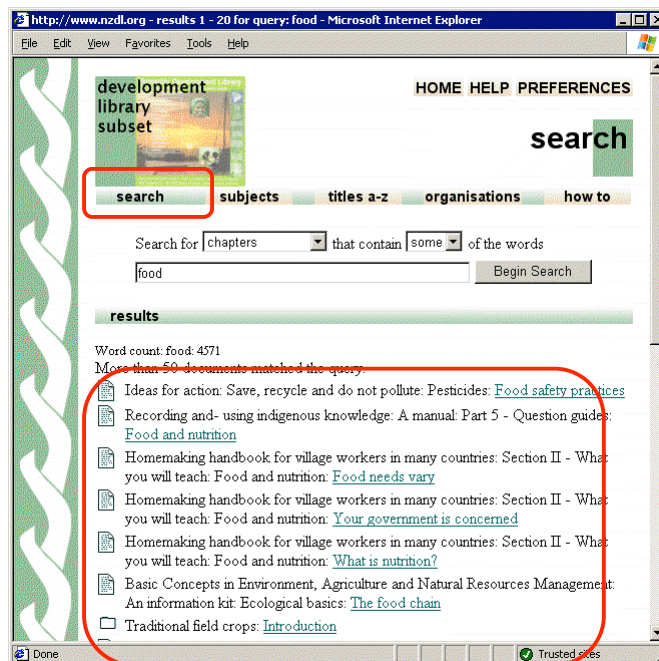
Variables: [Text] [Insert]

[Add Format] [Replace Format] [Remove Format]

Format Statements

- ❖ **VList** – vertical lists: search, classifiers, document table of contents
- ❖ **HList** – horizontal lists: classifiers
- ❖ **SearchVList**, **CL1VList** – individual vertical lists
- ❖ **DocumentImages/DocumentHeading/DocumentContents** – document header display
- ❖ **DocumentText** – document display

Search
VList



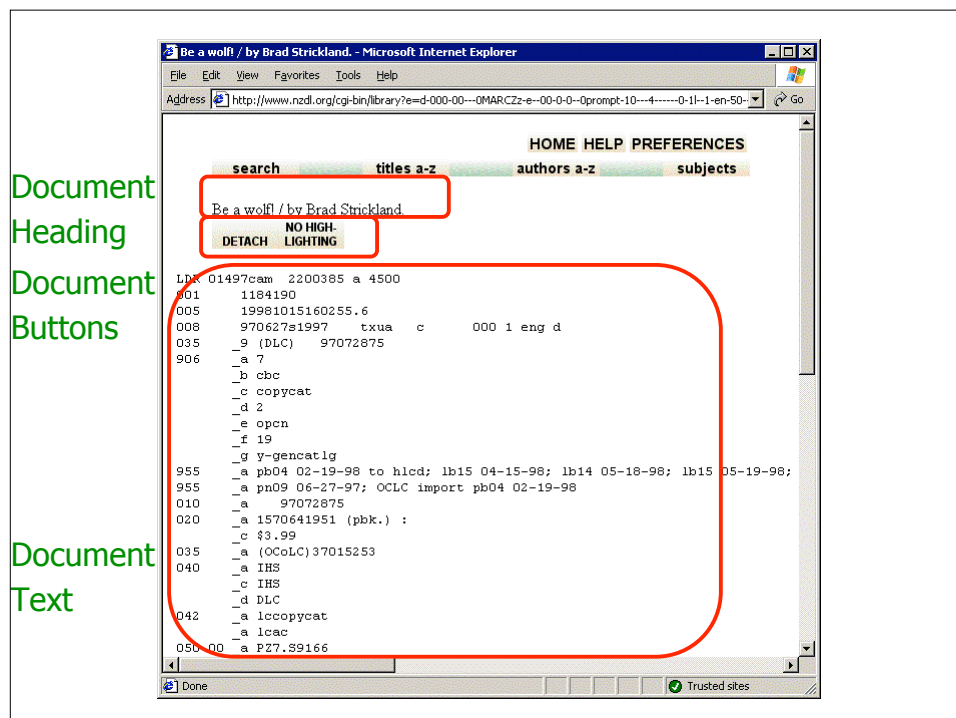
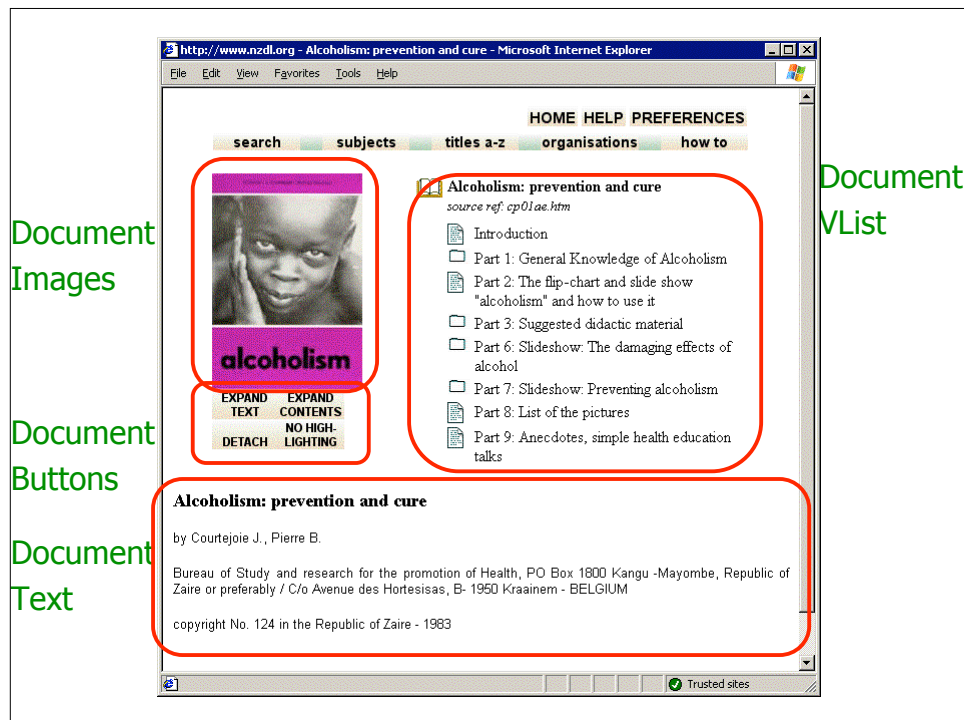
CL1
VList

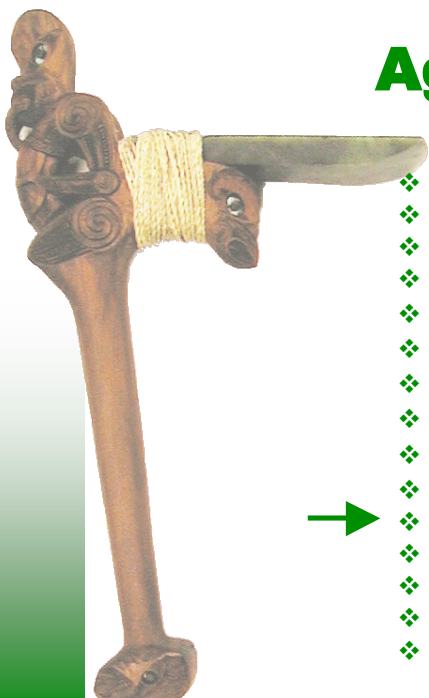


CL2
HList

CL2
VList







Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

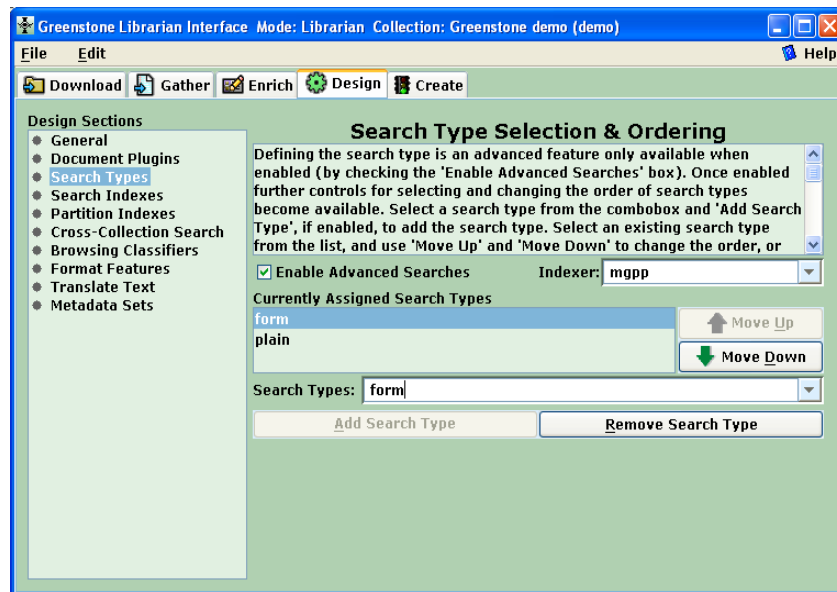
Search Types

- ❖ “Enable search types” – allows fielded searching
- ❖ MGPP/Lucene
- ❖ Form/plain – can switch on preferences
- ❖ Different index specification

Default Search Indexes



Search Types



Form Search Indexes: Indexes

Greenstone Librarian Interface Mode: Librarian Collection: Greenstone demo (demo)

File Edit Help

Download Gather Enrich Design Create

Design Sections

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

Index And Level Selection

Here you choose what searchable indexes the collection will have. To add a new index, enter a unique name for the index, select material/metadata is to be indexed, and click 'Add Index'. If you wish to add all of the available sources so as to have indexes built on them, then click 'Add All' (the name for each index will default to be the name of the source). Note that this is subtly different from adding the

Manage Indexes **Manage Levels**

Currently Assigned Indexes

text "text"	Move Up Move Down
ex.Title "Title"	

Index Name:

Index Source: allfields

Add Index Add All Replace Index Remove Index

Form Search Indexes: Levels

Greenstone Librarian Interface Mode: Librarian Collection: Greenstone demo (demo)

File Edit Help

Download Gather Enrich Design Create

Design Sections

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

Index And Level Selection

Here you choose what searchable indexes the collection will have. To add a new index, enter a unique name for the index, select material/metadata is to be indexed, and click 'Add Index'. If you wish to add all of the available sources so as to have indexes built on them, then click 'Add All' (the name for each index will default to be the name of the source). Note that this is subtly different from adding the

Manage Indexes **Manage Levels**

Currently Assigned Levels

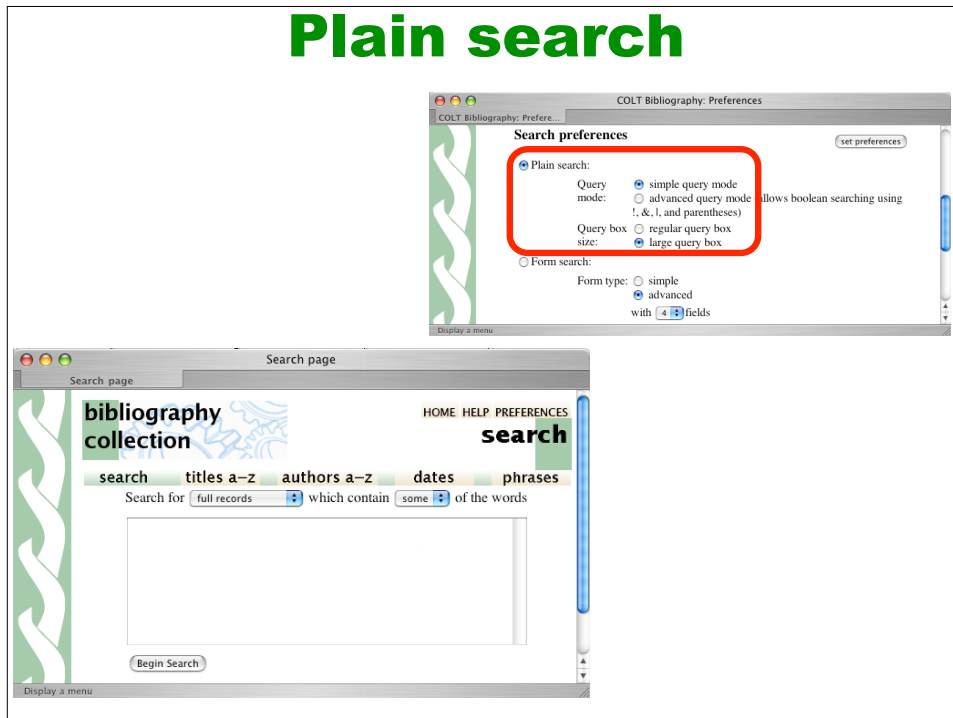
document "document"	Move Up Move Down
section "section"	

Level Name: section

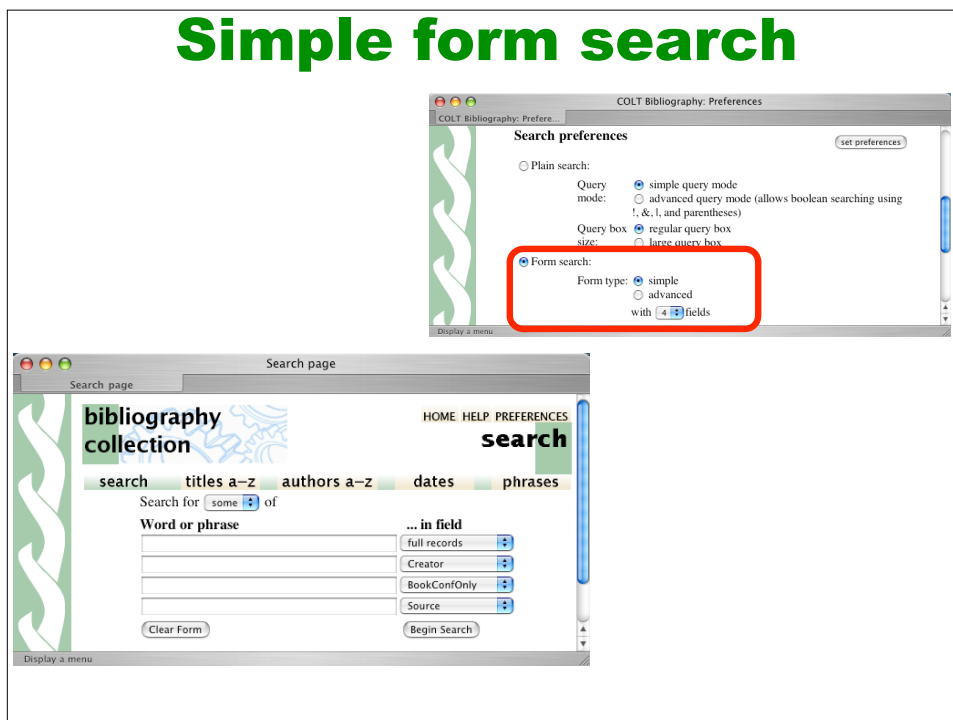
Level Source: section

Add Level Remove Level

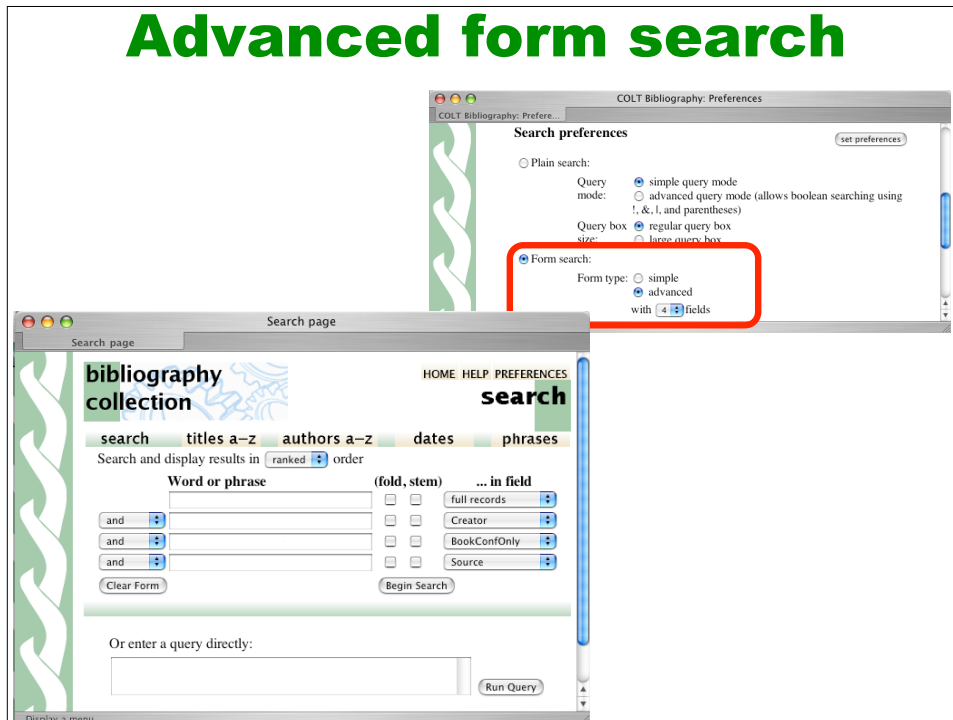
Plain search



Simple form search



Advanced form search



Advanced form search

The image shows the 'Search page' window with search results displayed. The search form is filled with the query: `[gori#i]:CR & [problem#i]:TI & [1992]:DA`. The 'Run Query' button is highlighted with a red rectangle. Below the search form, the results are shown, including a word count and a list of documents.

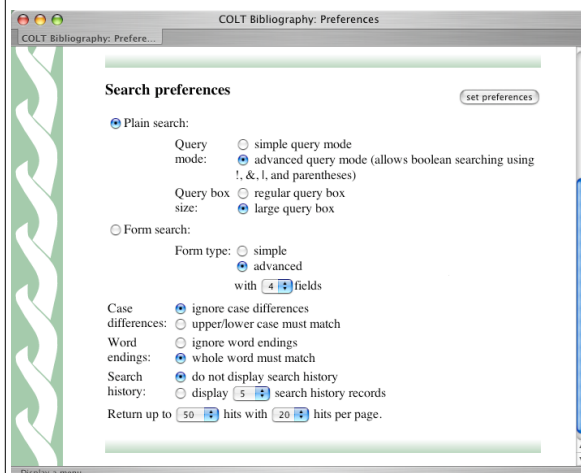
Fielded searching
 [terms]:Field
 Specifies searching within a particular field

Term modifiers
 #icus /x
 Stemming, casefolding and weighting

Fields
 TI title
 CR creator
 SU subject
 TX text
 ...

Word count: gori: 2, problem: 35, 1992: 181
 1 document matched the query.
 On the Problem of Local Minima in Backpropagation - Marco Gori and Alberto Tesi - 1992

Advanced plain search



❖ *Regular vs large query box*

❖ *Boolean operators*

& AND

| OR

! NOT

❖ *Phrase searching ("...")*

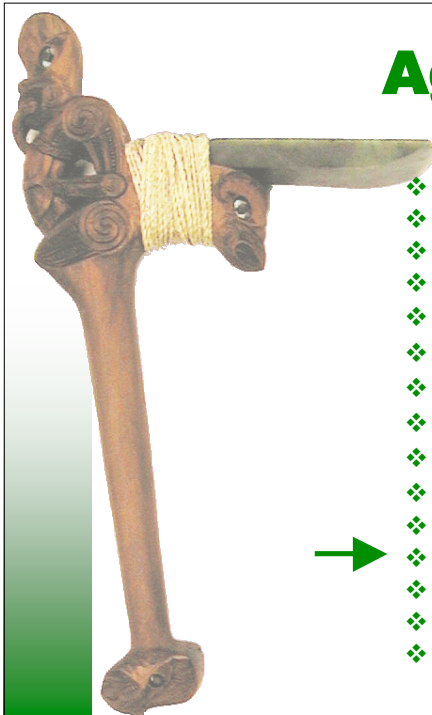
❖ *Case folding*

❖ *Stemming (word endings)*

❖ *Search history*

❖ *Number of returned results*

Agenda

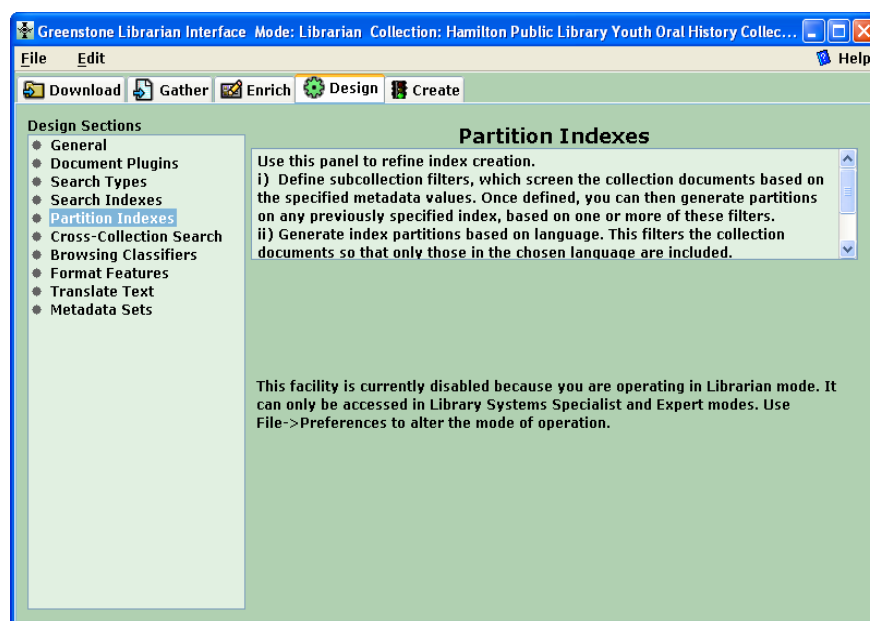


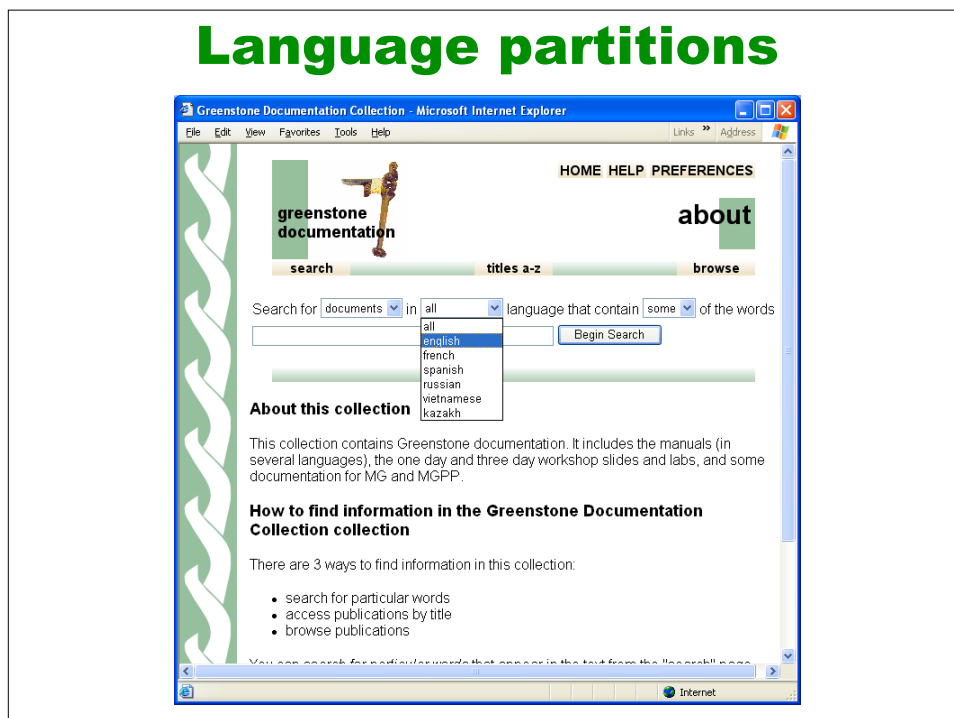
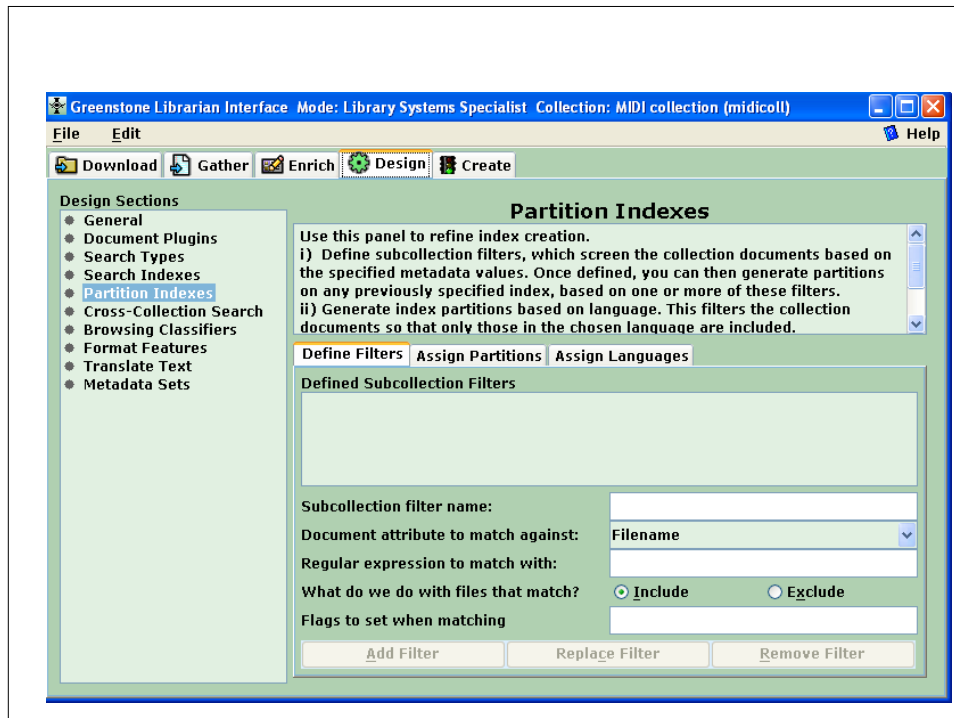
- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Partition Indexes

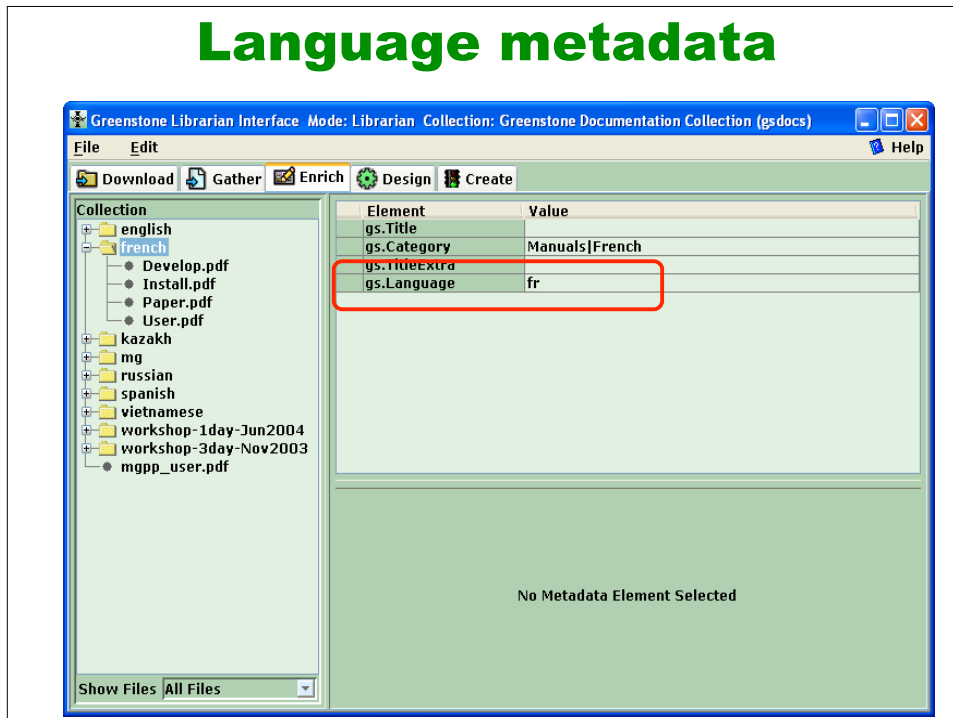
- ❖ Searching within a single collection
- ❖ The indexes are split based on subcollection definitions
 - E.g. language, file types, metadata
- ❖ Browsing classifiers contain all the documents
- ❖ Library Systems Specialist mode

Partition Indexes

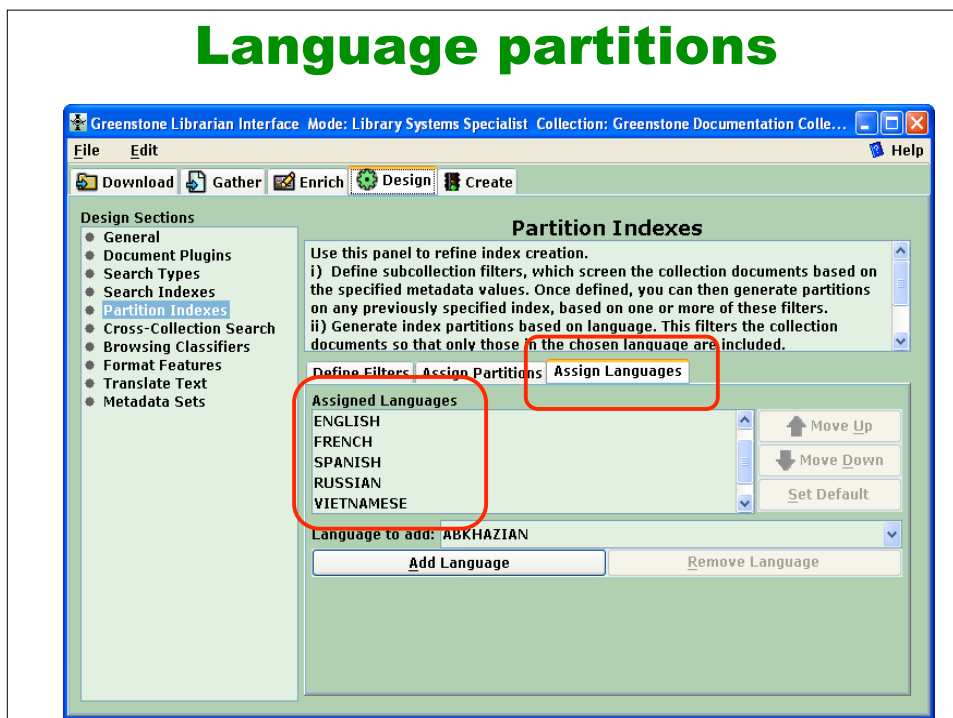




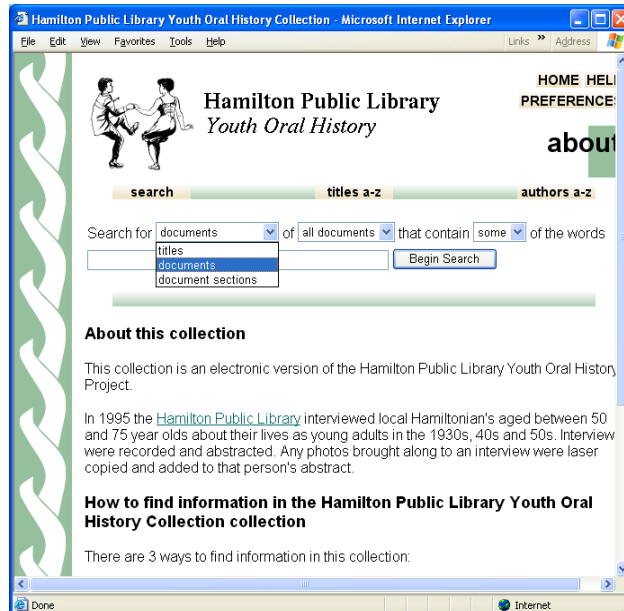
Language metadata



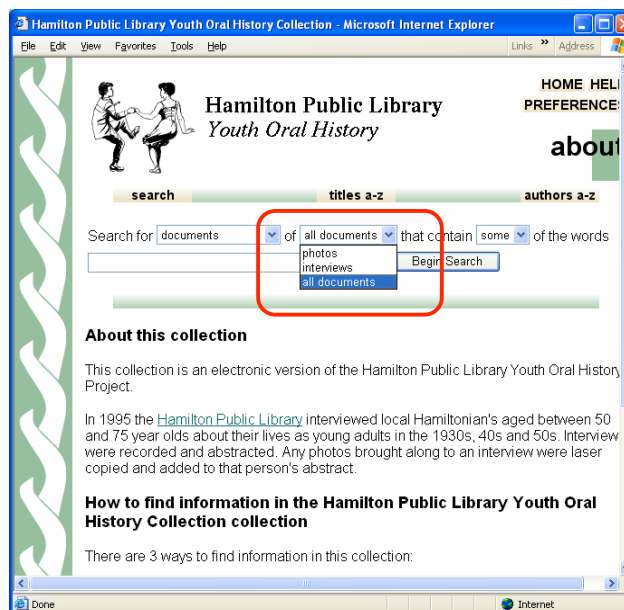
Language partitions



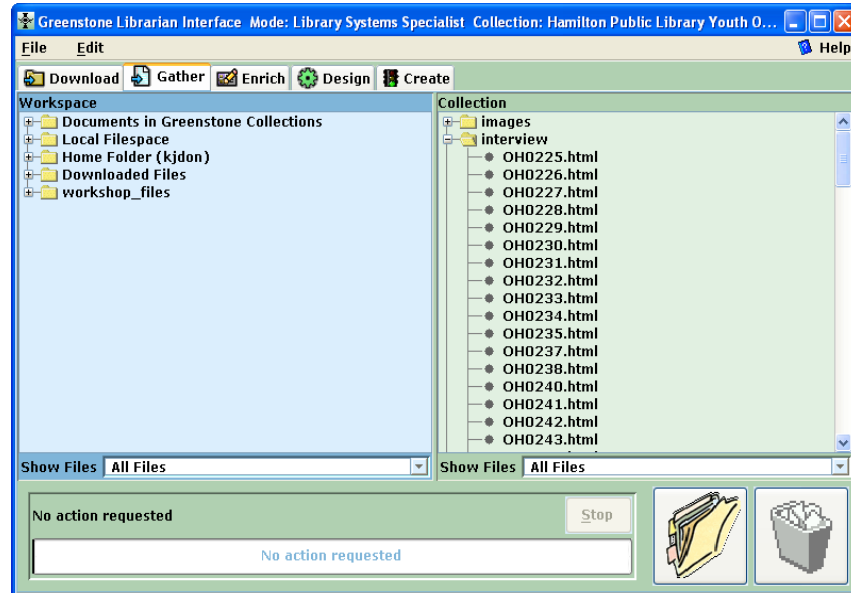
Document type partitions



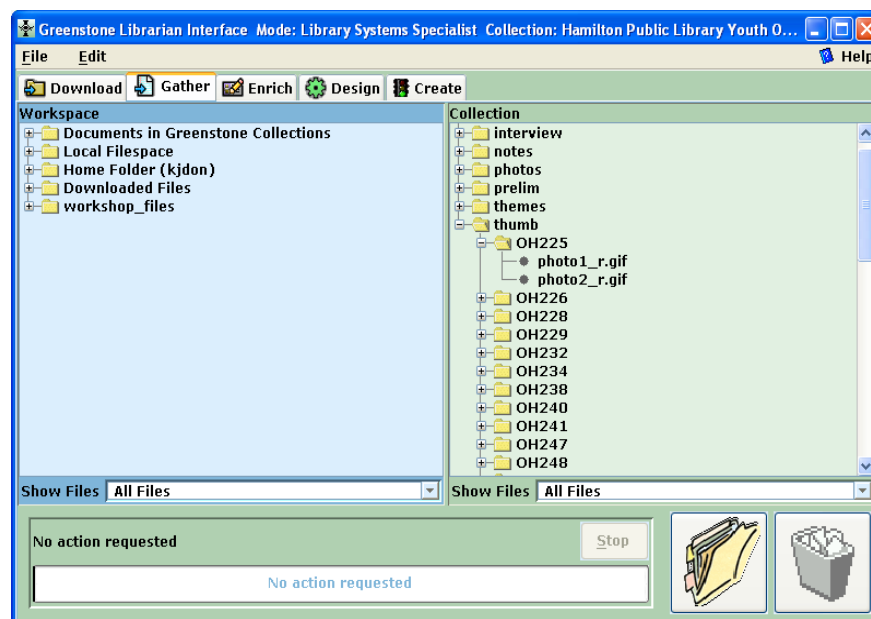
Document type partitions



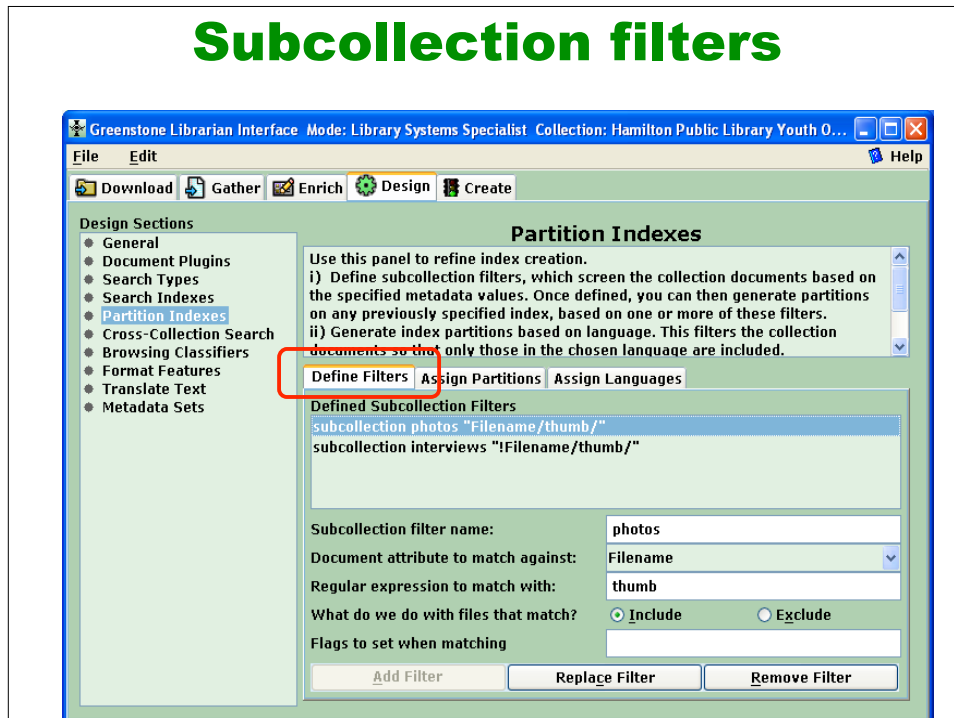
Collection documents



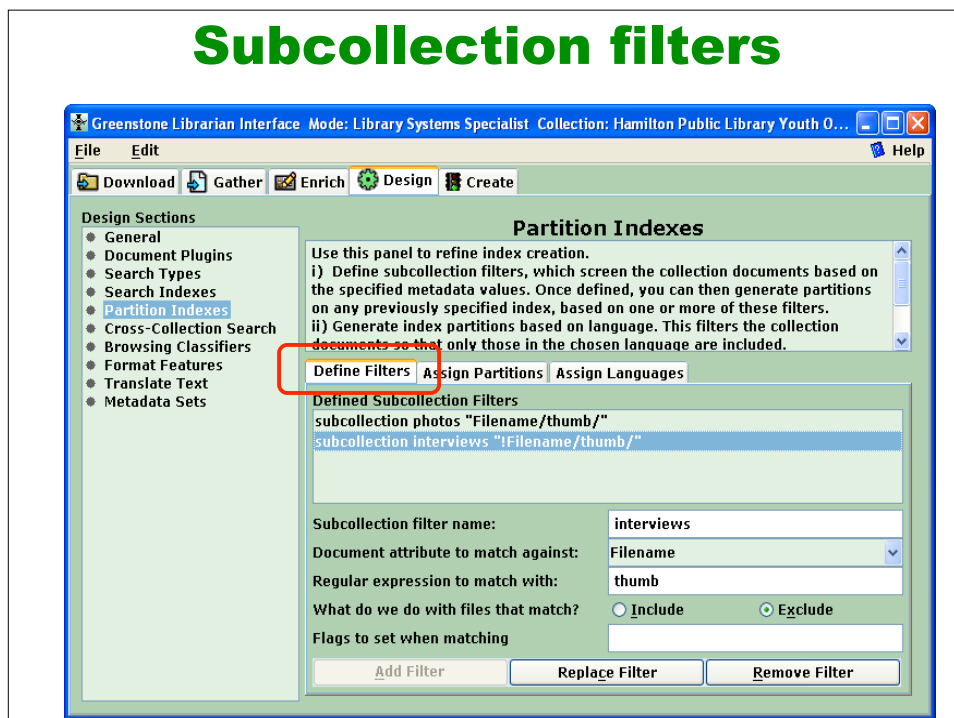
Collection documents



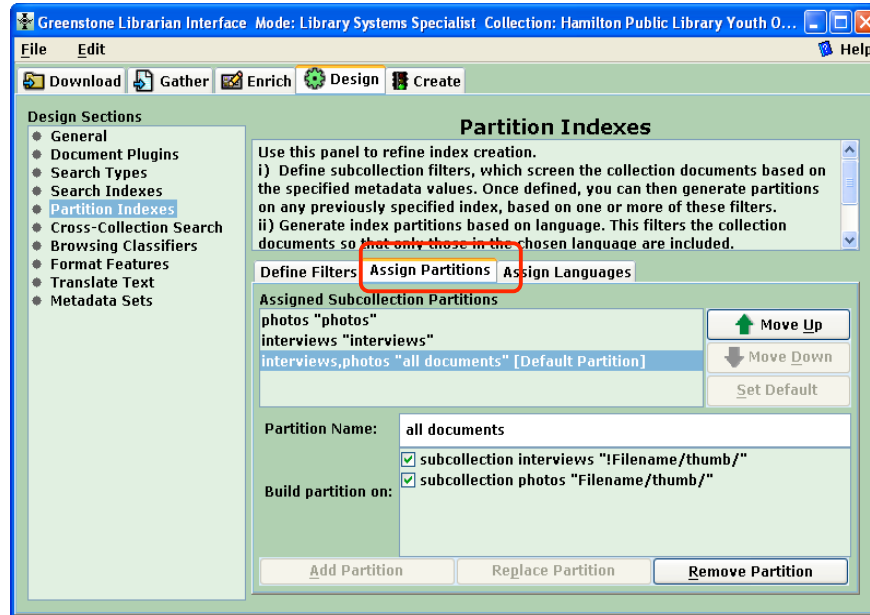
Subcollection filters



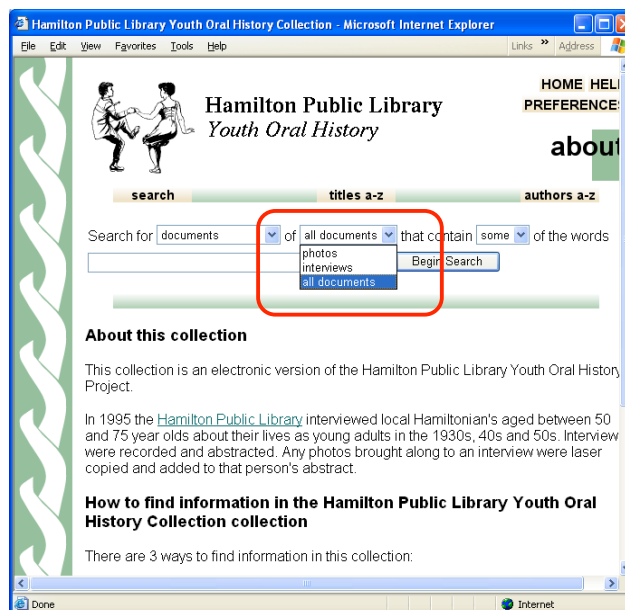
Subcollection filters



Assign partitions



Document type partitions

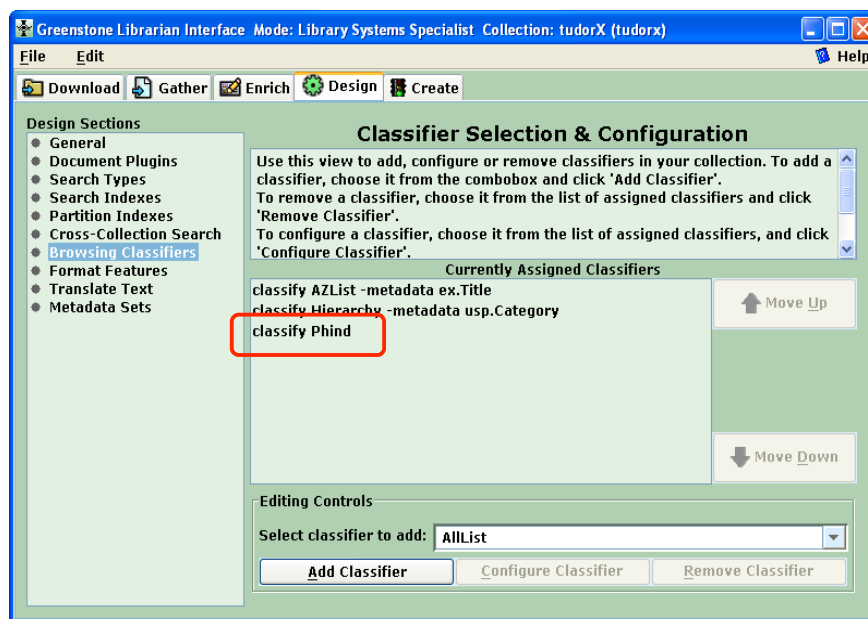




Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

Phind Phrase Index



PHIND

Hierarchical index of phrases

❖ What's in this collection?

❖ Is it any good?

❖ What coverage for topic X?

❖ My query returned too much/little, what now?

food and agriculture organization

HOME HELP PREFERENCES

phrases

search titles a-z language phrases

Search for locust Previous Next

food and agriculture organization

HOME HELP PREFERENCES

phrases

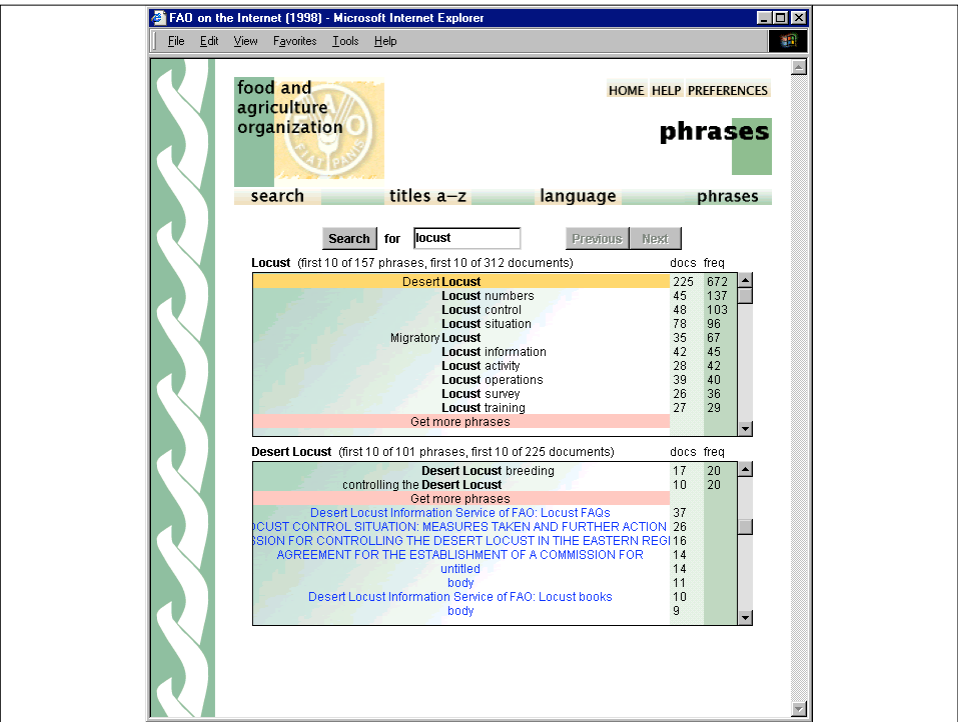
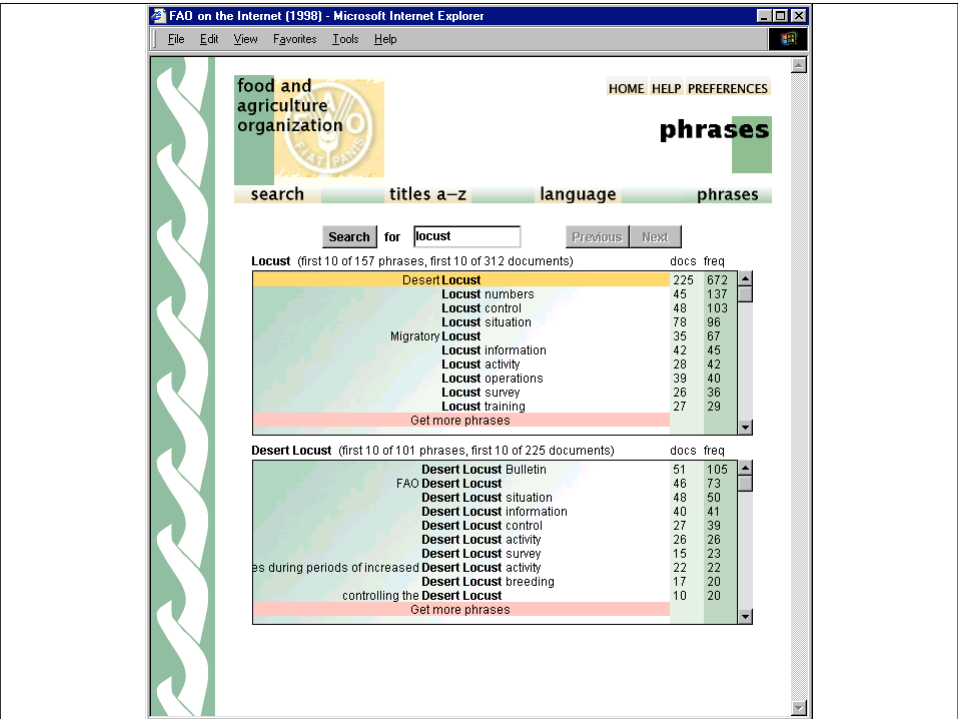
search titles a-z language phrases

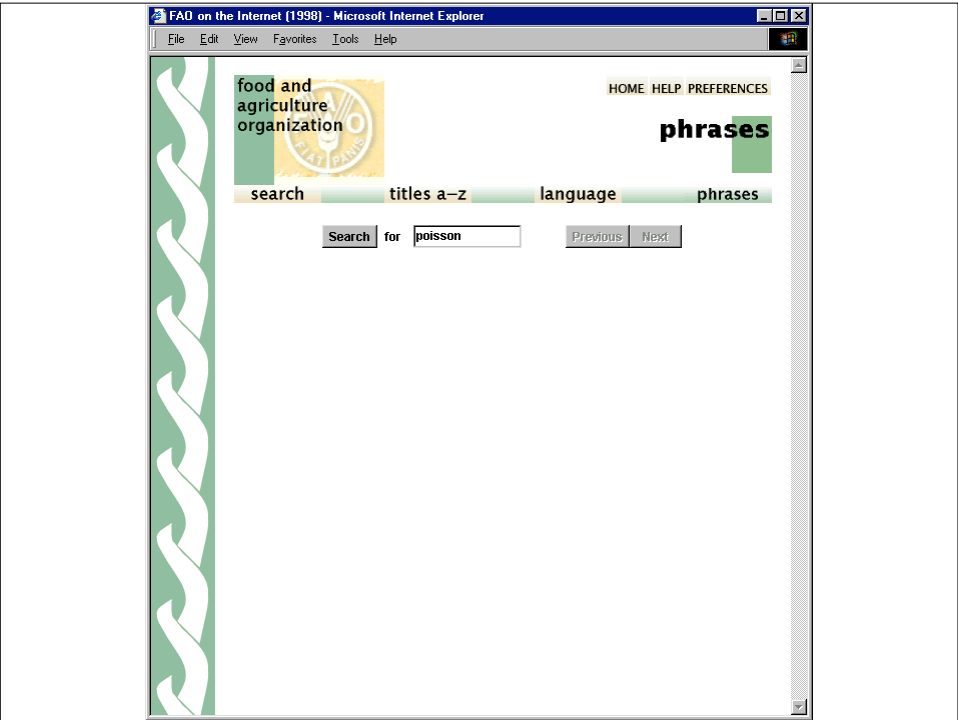
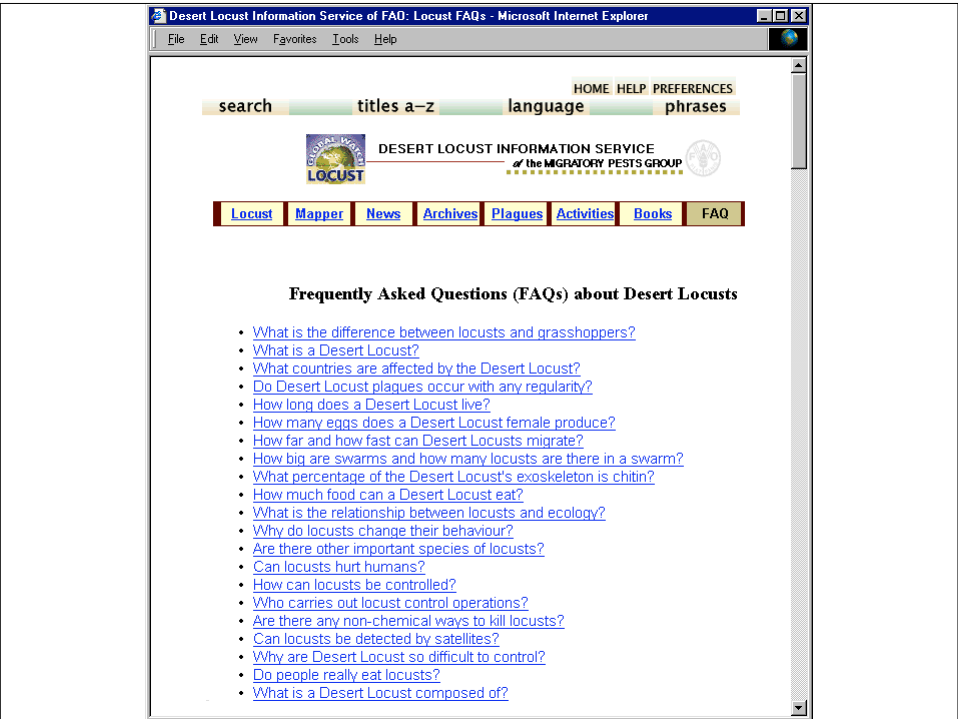
Search for locust Previous Next

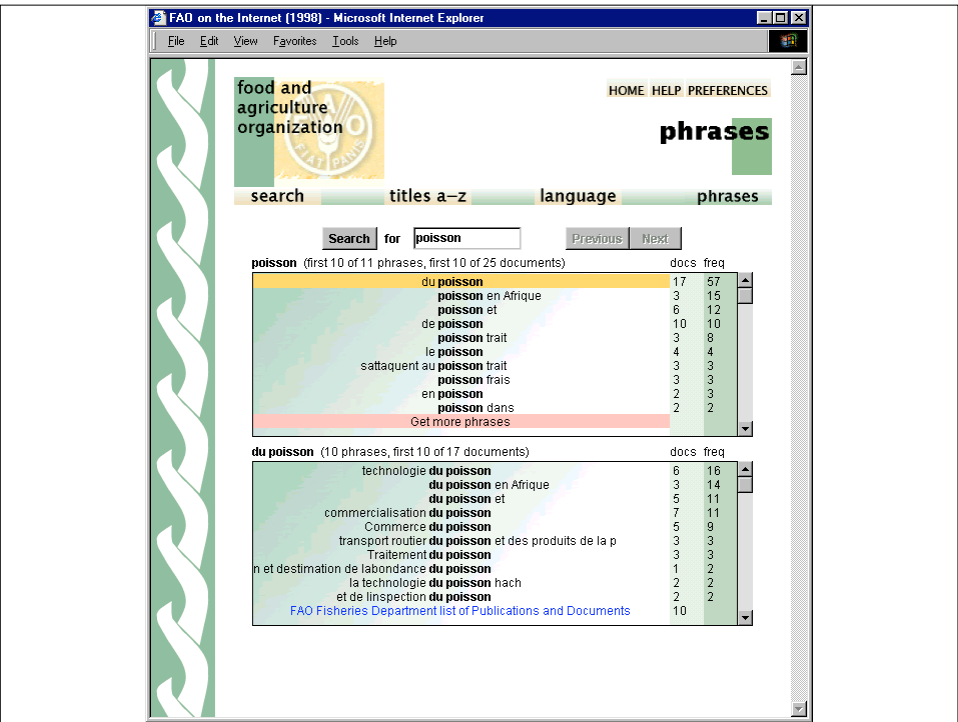
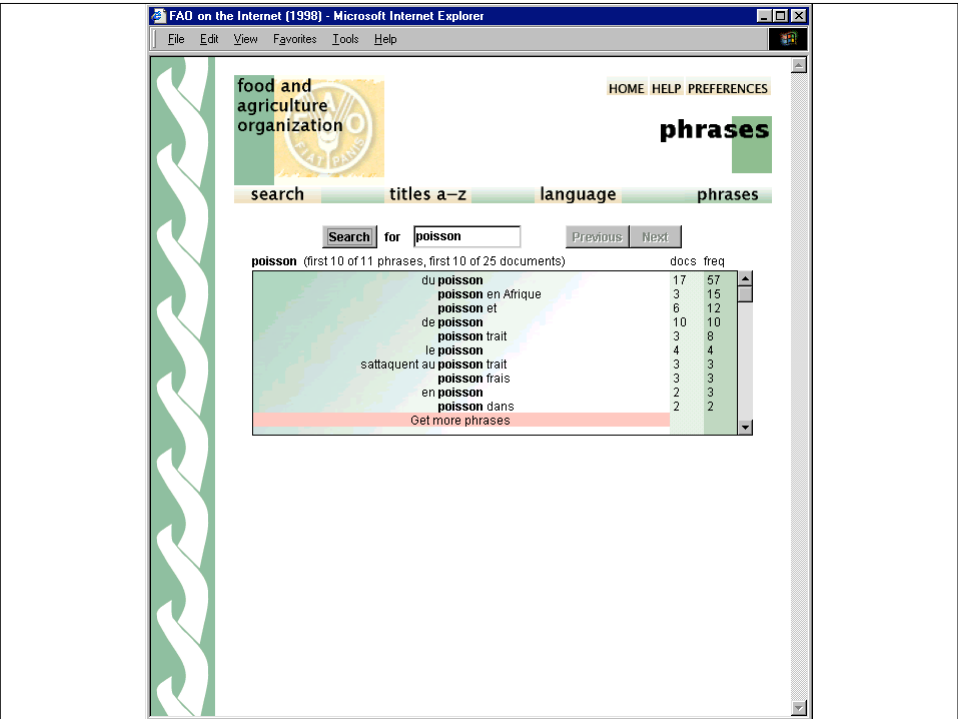
Locust (first 10 of 157 phrases, first 10 of 312 documents)

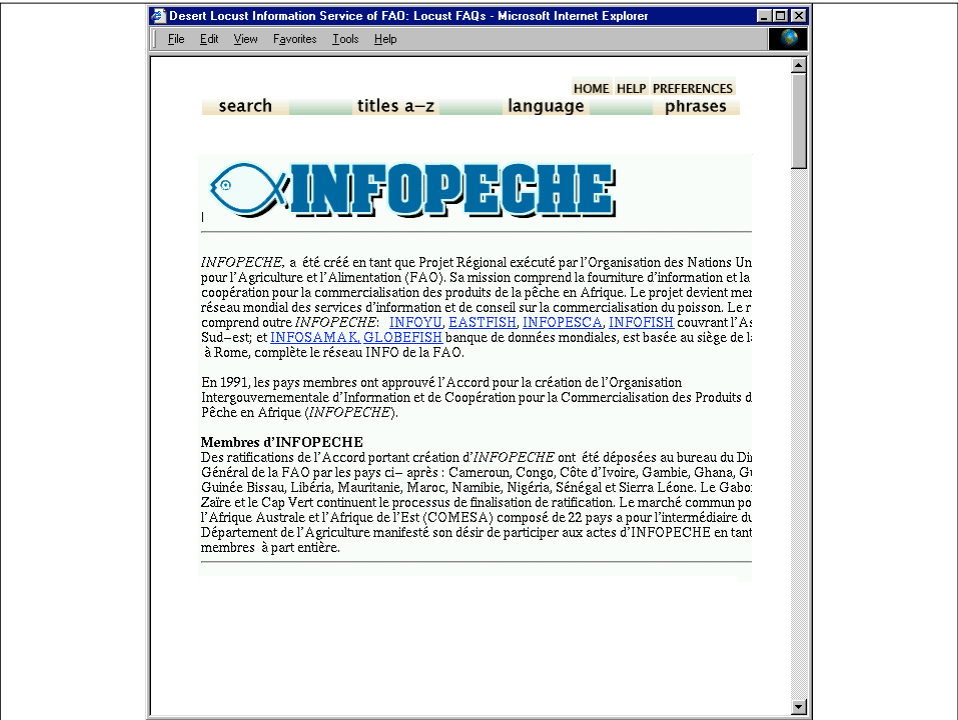
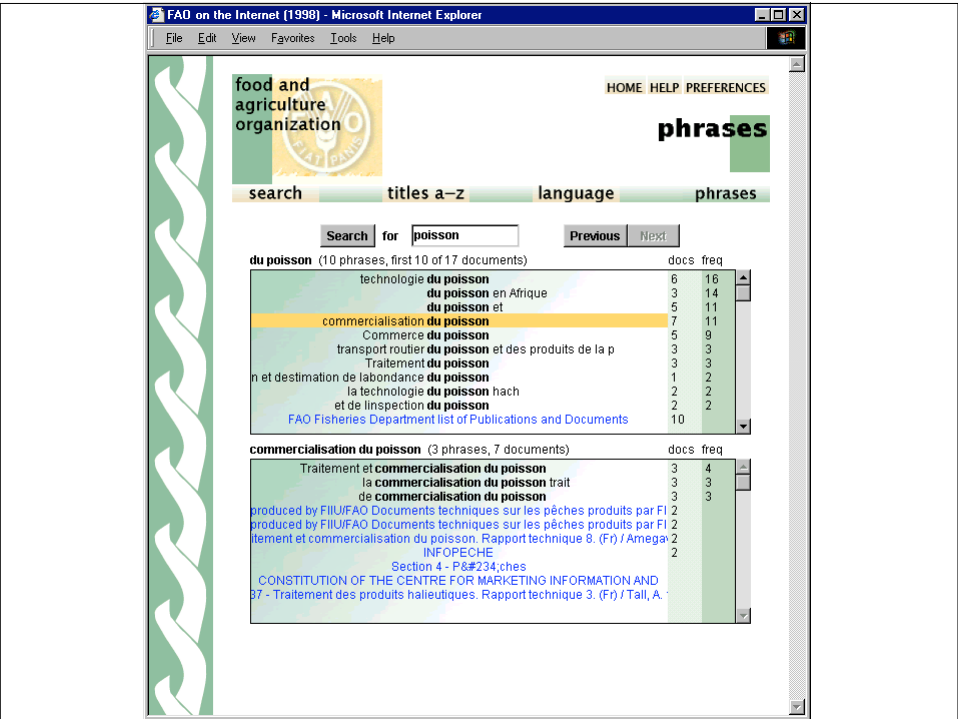
	docs	freq
DesertLocust	225	872
Locust numbers	45	137
Locust control	48	103
Locust situation	78	96
MigratoryLocust	35	67
Locust information	42	45
Locust activity	28	42
Locust operations	39	40
Locust survey	26	36
Locust training	27	29

Get more phrases











Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing



CDS/ISIS

- ❖ Bibliography collections are typically fairly complex:
 - Form searching
 - Customised query result and browse lists
 - Customised document display
- ❖ Let's work through creating a simple collection using a small CDS/ISIS database describing a set of film slides
(More information in the "Bibliography collection" and "CDS/ISIS" documented example collections)

CDS/ISIS

- ❖ Add the CDS/ISIS files to a new collection
- ❖ After building, let's view the collection:
 - No metadata searching is available:

search

ISIS collection

search titles a-z filenames

Search for that contain of the words

- The titles classifier is completely empty!

titles a-z

ISIS collection

search titles a-z filenames

CDS/ISIS

- ❖ More problems:
 - The filenames classifier is useless!

filenames

ISIS collection

search titles a-z filenames

Untitled
(slide.mst)

Untitled
(slide.mst)

Untitled
(slide.mst)

- The document display isn't very pretty:

search titles a-z filenames

untitled

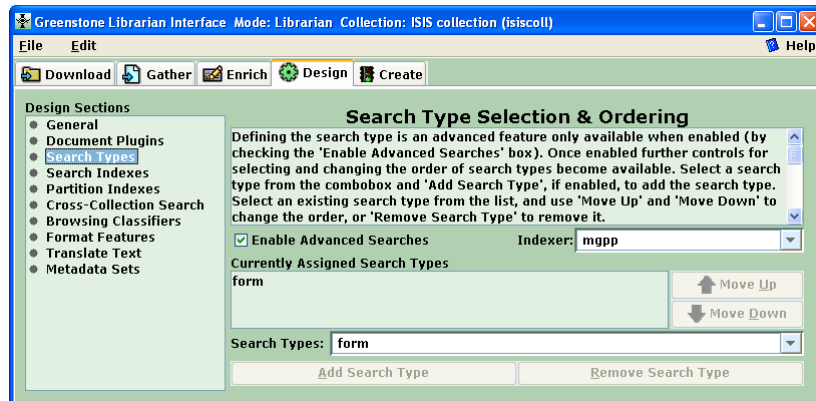
NO HIGH-LIGHTING

DETACH

tag=Year data=1997 tag=Month & Date data=Oct. tag=Notes data=contact plus negatives; Diwali photos tag=Program data=SECRETARIAT tag=Country data=<Fiji> tag=Photographer data=Woodward, P.

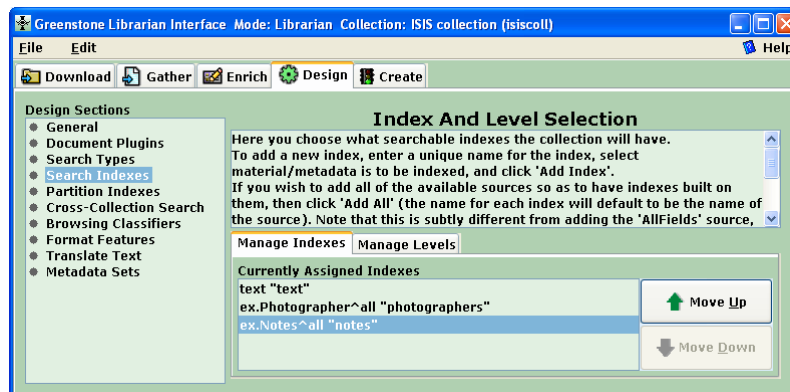
CDS/ISIS: Metadata searching

- ❖ To enable form searching, go to the "Search Types" area in the GLI's Design pane
 - Tick "Enable Advanced Searches" on
 - Add the "form" search type, and remove "plain"



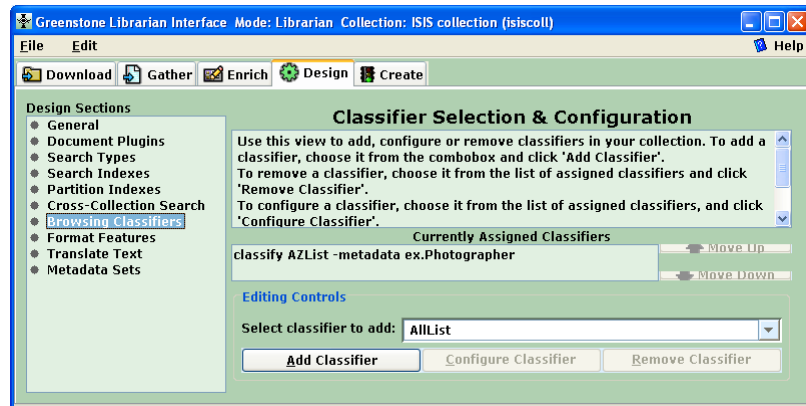
CDS/ISIS: Metadata Searching

- ❖ Add metadata indexes in the "Search Indexes" part of the GLI's Design pane
 - Add indexes for Photographer and Notes metadata
 - Remove the useless Source and Title indexes



CDS/ISIS: Better browsing

- ❖ Remove the existing (useless) classifiers for Title and Source metadata, and add a new one for Photographer



CDS/ISIS: Better browsing

- ❖ Change the VList format statement to display the Photographer and Notes metadata:



CDS/ISIS: Document display

- ❖ Next, let's change the DocumentText format statement to show the Photographer and Notes metadata:

```
<center><table width=_pagewidth ><tr><td>Photographer:
</td><td>[ex.Photographer^all]</td></tr><tr><td>Notes:
</td><td>[ex.Notes^all]</td></tr></table></center>
```

- ❖ Then, let's remove those annoying "Detach" and "Highlight" buttons by setting DocumentButtons to empty
- ❖ Lastly, clear DocumentHeading to remove the "untitled" at the top of the document

CDS/ISIS: Finished!

- ❖ Metadata searching now available:

[ISIS collection](#) **search**

search Photographer

Search for of

Word or phrase	... in field
	text
	photographers
	notes
	text

- ❖ Better browsing facilities:

[ISIS collection](#) **Photographer**

search Photographer

A-E [F](#) [G](#) [H](#) [I](#) [M](#) [R](#) [S](#) [T](#) [X](#)

Australia. Bureau of Mineral Resources
United States Geological Survey
Slides for presentations on the petroleum potential of the SOPAC region (New Ireland (PNG), Solomon Islands, Tonga and Vanuatu), obtained from N.F. Exon

Barclay, W.
10 col. slides of hydrocarbon potential of the Southwest Pacific

CDS/ISIS: Finished!

❖ Document display improved:

search [Photographer](#)

Photographer: Australia. Bureau of Mineral Resources
United States Geological Survey

Notes: Slides for presentations on the petroleum potential of the SOPAC region
(New Ireland (PNG), Solomon Islands, Tonga and Vanuatu), obtained
from N.F. Exon

❖ What could still be improved?

- More metadata indexes, classifiers
- Display all fields in the document display
- Nice images for classifiers
- ...?

Agenda



- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

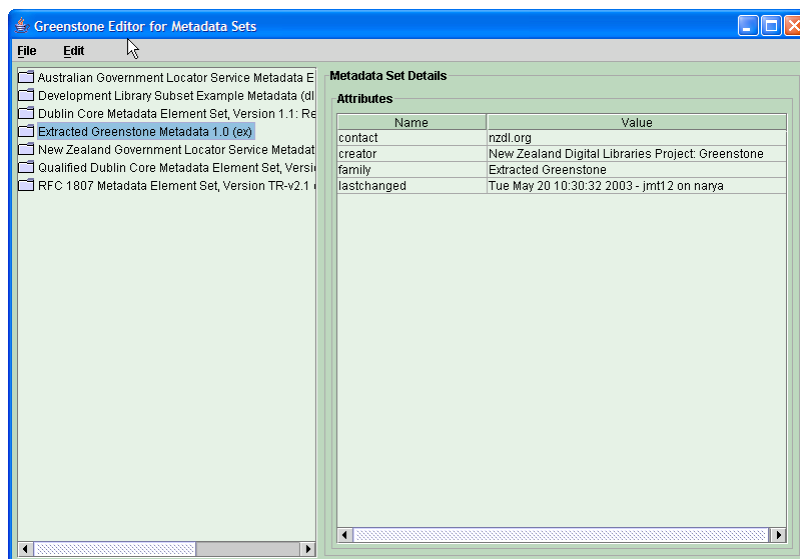


Greenstone Editor for Metadata Sets - GEMS

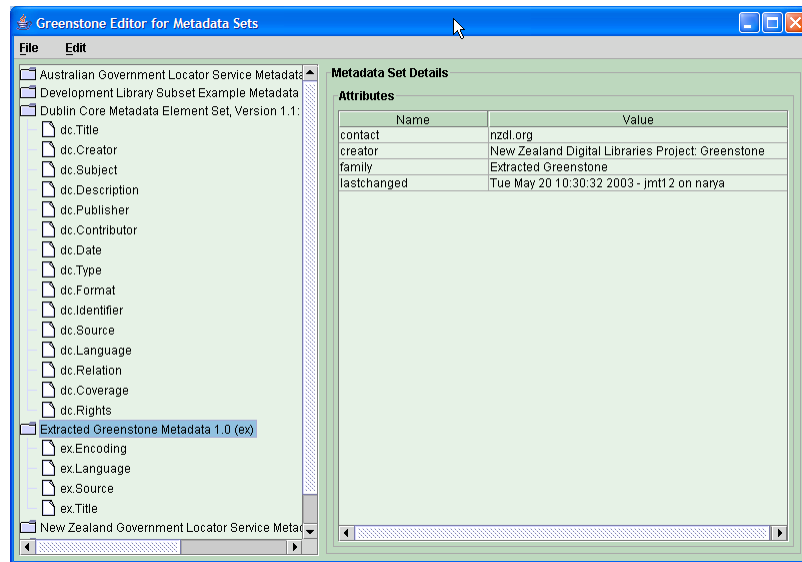
❖ What can GEMS do?

- Create a new metadata set
- Add metadata elements to an existing set
- Define and edit metadata attributes
- Inherit from an existing metadata set
- Import an existing metadata set
- Sets can be used in GLI

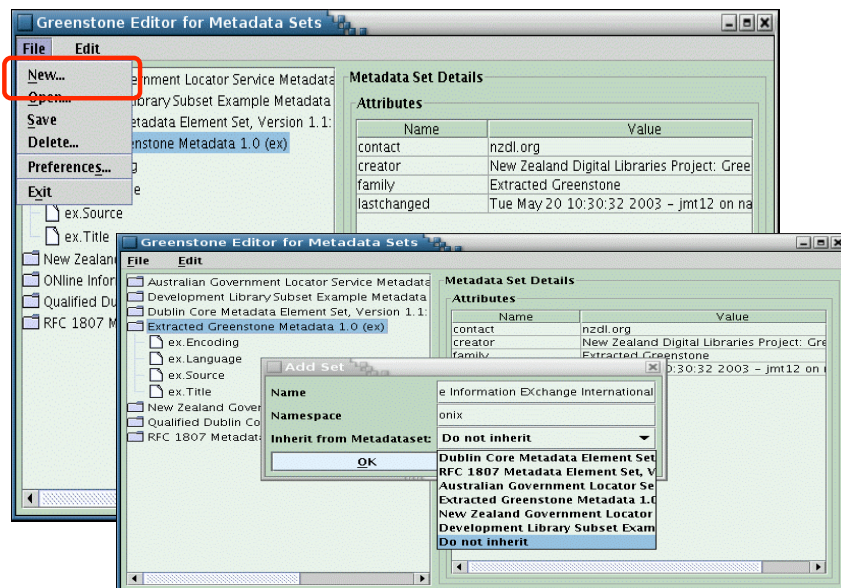
GEMS Interface



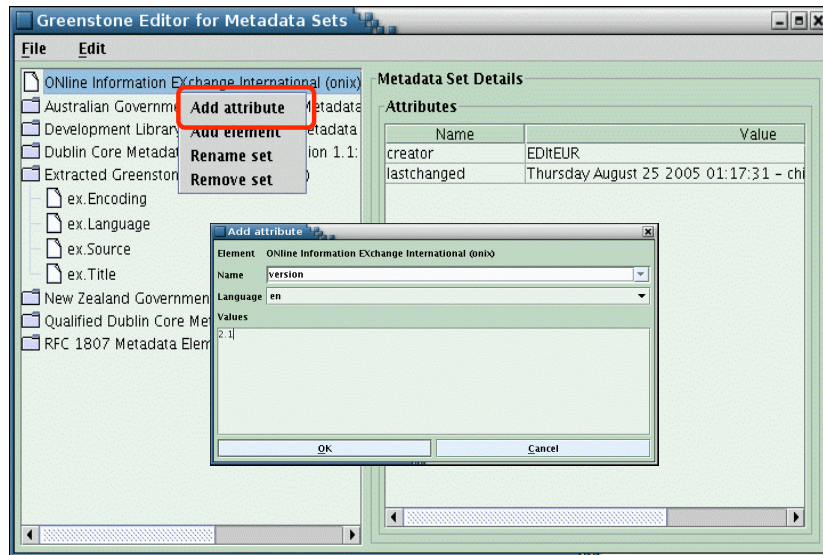
GEMS Interface



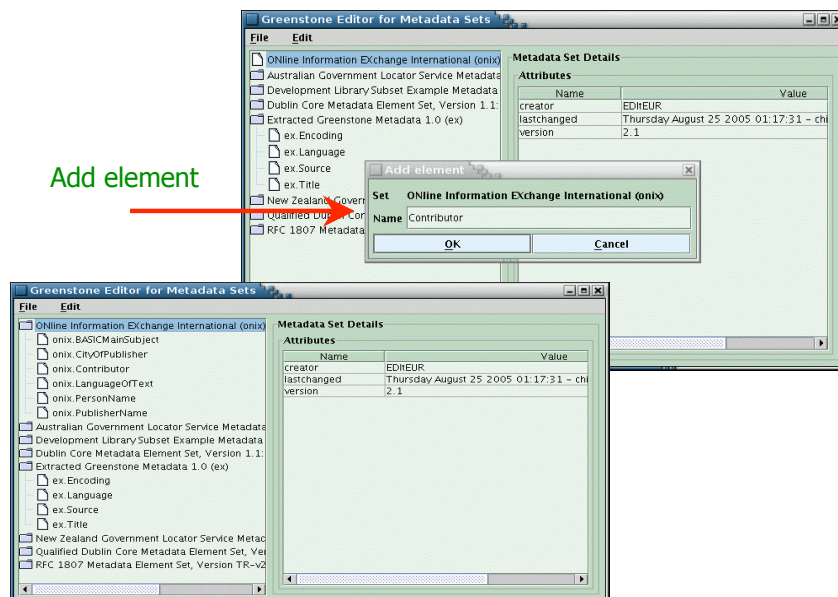
Adding a New Metadata Set



Adding Metadata Attribute



Adding Metadata Element



Using an existing metadata set

- ❖ Using a pre-defined metadata set
 - Must comply with Greenstone metadata XML file format
 - ❖ Write a script to convert?
 - Copy to /gsdl/gli/metadata directory