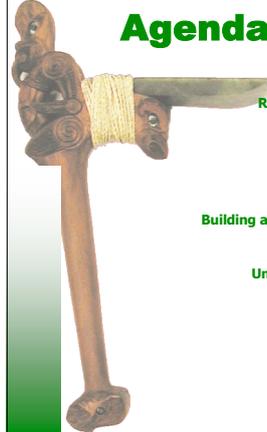




Greenstone: Open source software for building digital library collections

Ian H. Witten and Kathy Don
 Computer Science Department
 Waikato University
 New Zealand
<http://greenstone.org>
<http://nzdl.org>



Agenda

Overview
 What does Greenstone do?
 Greenstone facts; standards
 Reader's Interface: examples of collections

Librarian interface
 Build a collection in 30 sec (Hobbits)
 Build a multimedia collection (Beatles)
 Adding and using metadata
 Browsing classifiers, search indexes
 Building a collection manually (for masochists only)

Advanced stuff
 Under the hood: collection configuration file
 Customizing with macros
 Personalizing your home page
 Different interface languages
 Examples of what others have done

Reaching out
 Serving and acquiring OAI
 DSpace and METS
 Greenstone3

What we wanted

Greenstone turns a ragtag menagerie of documents in various formats into an easy-to-use collection that can run on a standalone laptop in a Ugandan village's information center



ALA 2002

What we wanted

- ❖ "Collections" of digital material
- ❖ Individualized, depending on metadata etc
- ❖ Up to several Gb of text ...
- ❖ ... + associated images, movies, whatever
- ❖ Fully searchable
- ❖ Served on WWW, or published on removable media
- ❖ Run anywhere, on any computer
- ❖ Fully internationalized
- ❖ Non-exclusive: documents and metadata in any format
- ❖ Non-prescriptive: standard and non-standard metadata

What we got: Greenstone

Access

- ❖ Accessible via any Web browser
- ❖ Server runs on anything (all Windows + Unix + Mac)
- ❖ Collections can be published on CD-ROM/DVD
- ❖ Trivial to install
- ❖ GUI interface for building and publishing collections

Searching/ browsing

- ❖ Collection-specific
- ❖ Full-text and fielded search
- ❖ Flexible browsing facilities
- ❖ Metadata-based (Dublin Core recommended)
- ❖ Creates all access structures automatically

Extensible

- ❖ Plugins — new document, metadata formats
- ❖ Classifiers — new metadata browsers

Multi-*

- ❖ Multilingual: Documents *and* interfaces
- ❖ Multimedia: image, video, audio collections exist
- ❖ Multiformat: Documents *and* metadata

UNESCO: Distributing Greenstone DL software

Sustainable development

*"Give a man a fish, feed him for a day
 Teach a man to fish, feed him for life"*

Greenstone software on CD-ROM

- ❖ GNU licensed
- ❖ Fully documented ... in English/French/Spanish/Russian
- ❖ Language interfaces ... Arabic Chinese Czech ... Thai Turkish
- ❖ Unix/Windows/Mac OS-X
- ❖ Trivial to install
- ❖ GUI interface for gathering, enriching, building ...
- ❖ Serve collections on Web or write them to CD-ROM
- ❖ Document formats: HTML, Word, PDF, PS, plain text, e-mail
- ❖ Metadata formats: XML, DC, OAI, MARC, ...

download from <http://greenstone.org>





Greenstone facts

- Distribution**
 - ❖ Open source: Gnu GPL
 - ❖ Distributed via SourceForge since: Nov 2000
 - ❖ Average downloads: 5000/month since: Nov 2000
 - ❖ Humanitarian CD-ROMs produced: 30
 - ❖ Distribution for each one: 5000/year
- International**
 - ❖ Languages for interface: 38
 - ❖ Languages for full software + manual: 38
 - ❖ Countries represented on email lists: 60
 - ❖ UNESCO training courses in:
 - Bangalore, Almaty, Dakar, Suva, ...
- UN Agencies**
 - ❖ UNESCO, Paris ("Information for All")
 - ❖ FAO, Rome (Info Management Resources)
 - ❖ UNU, Japan (CD-ROM collections of UN)
- Technical centers**
 - ❖ University of Waikato, New Zealand
 - ❖ Indian Institute of Sciences, Bangalore
 - ❖ University College, London
 - ❖ University of Cape Town, South Africa
 - ❖ University of Lethbridge, Canada

Arabic
Armenian
Bosnian
Croatian
Deutsch (German)
English
Español (Spanish)
Farsi
Finnish
Français (French)
Galician
Georgian
Hebrew
Hindi
Indonesian
Italiano
Kannada
Māori
Nederlands (Dutch)
Serbian
Thai
Turkish
Ukrainian
português-BR (Brasil)
português-PT (Portugal)
Česky (Czech)
Ελληνικά (Greek)
Русский (Russian)
Қазақ (Kazakh)
日本語 (Japanese)
汉语 (Chinese)

Sample collections at greenstone.org

<p>International</p> <ul style="list-style-type: none"> Argentina Human Rights Commission Tasmania State Library Peking University Digital Library Gresham College, London University of Applied Sciences, Stuttgart Association of Indian Labour Historians, Delhi Indian Institute of Management, Kozhikode Indian Institute of Science, Bangalore Vimercate Public Library, Milan, Italy Netherlands Institute for Scientific Information Services Philippine Government Information Network Mari El Republic, Russia Slavonski Brod Public Library, Slovenia Vietnam National University Welsh Books Council 	<ul style="list-style-type: none"> Argentina Australia China England Germany India India India Italy Netherlands Philippines Russia Slovenia Vietnam Wales
<p>U.S.</p> <ul style="list-style-type: none"> Auburn University, Alabama Detroit Public Library Hawaiian Electronic Library ibiblio project, University of North Carolina Illinois Wesleyan University LeHigh University, Pennsylvania New York Botanical Garden University of California at Riverside University of Chicago Library University of Illinois Texas A&M University Washington Research Library Consortium 	

Standards

- Metadata**
 - ❖ Can use any metadata set, Dublin Core supplied
 - ❖ Plugins for

XML	Refer
MARC	OAI
CDS/ISIS	METS (subset)
ProCite	DSpace
BibTex	
 - ❖ METS can be used as Greenstone's internal representation
- Serving**
 - ❖ Web
 - ❖ Can publish Greenstone collections on CD-ROM
 - ❖ Can publish Greenstone collections on OAI
 - ❖ Export collections to METS
 - ❖ Export collections to DSpace (ready for DSpace's batch import program)
- Documents**
 - ❖ Plugins for

PDF	ZIP	Images
PostScript	Excel	(any format: GIF, JPEG, TIFF ...)
Word, RTF	PPT	MP3
HTML	Email	Ogg Vorbis
Plain text	Source code	UnknownPlug
Latex		(e.g. for audio, MPEG, Midi)

What is open-source software?

- ❖ "The **basic idea behind open source** is very simple: When programmers can read, redistribute, and modify the source code for a piece of software, the software evolves. People improve it, people adapt it, people fix bugs. And this can happen at a speed that, if one is used to the slow pace of conventional software development, seems astonishing."
 - from www.opensource.org
- ❖ Anyone can redistribute the software, even for a fee
- ❖ Source code must always be available

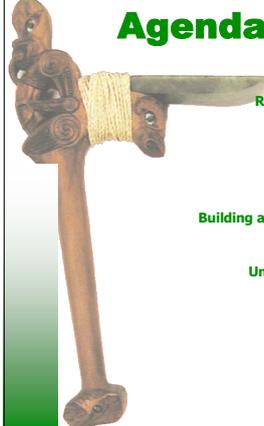
The power of open source: Greenstone uses ...

- ❖ **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**
- ❖ **wwWare** **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 ...

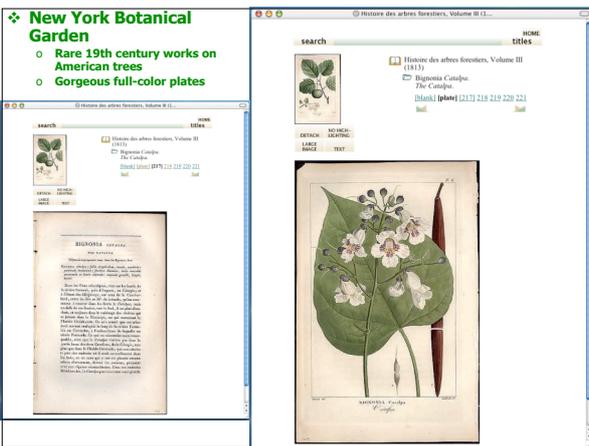
- ❖ **MG** Creates compressed full-text indexes and performs searches
- ❖ **GDBM** Database used for metadata etc
- ❖ **wget** Downloading pages from the Web when creating collections
- ❖ **YAZ** Client and server implementation of Z39.50
- ❖ **Stemmer** English language stemmer

- ❖ **GCC** C/C++ compiler
- ❖ **CVS** Version control system
- ❖ **Perl** Used for plugins etc
- ❖ **Apache** Web server used by many Greenstone installations



Agenda

- Overview**
 - What does Greenstone do?
 - Greenstone facts; standards
 - Reader's Interface: examples of collections
- Librarian interface**
 - Build a collection in 30 sec (Hobbits)
 - Build a multimedia collection (Beatles)
 - Adding and using metadata
 - Browsing classifiers, search indexes
 - Building a collection manually (for masochists only)
- Advanced stuff**
 - Under the hood: collection configuration file
 - Customizing with macros
 - Personalizing your home page
 - Different interface languages
 - Examples of what others have done
- Reaching out**
 - Serving and acquiring OAI
 - DSpace and METS
 - Greenstone3



New York Botanical Garden

- o Rare 19th century works on American trees
- o Gorgeous full-color plates



University of Chicago Library

Chopin Early Editions

Allegro de concert - pour le piano - op. 46 /

Page 22



Pukapuka 1, Nama 1:1

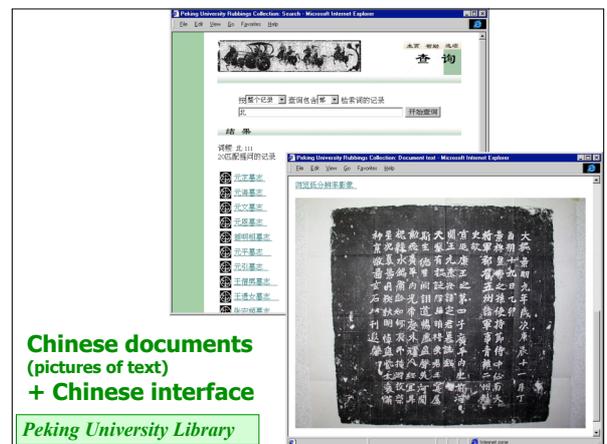
niupepa: maori newspapers

Te Waka Maori

O NIU TIRANI

"KO TE TIKI, KO TE PONO, KO TE AROHA"

VOL. I, TURANGA, WENHREI, AKUHATA 21, 1878. [No. 1.]



Peking University Library

Chinese documents (pictures of text) + Chinese interface

Agenda

- Overview**
 - What does Greenstone do?
 - Greenstone facts; standards
 - Reader's Interface: examples of collections
- Librarian interface**
 - Build a collection in 30 sec (Hobbits)
 - Build a multimedia collection (Beates)
 - Adding and using metadata
 - Browsing classifiers, search indexes
 - Building a collection manually (for masochists only)
- Advanced stuff**
 - Under the hood: collection configuration file
 - Customizing with macros
 - Personalizing your home page
 - Different interface languages
 - Examples of what others have done
- Reaching out**
 - Serving and acquiring OAI
 - DSpace and METS
 - Greenstone3

The Greenstone Librarian Interface (GLI)

- ❖ Building collections
- ❖ Interactive Java program
- ❖ Runs on anything
- ❖ Build a collection on the computer you are on
- ❖ ... plus new applet version
- ❖ Includes metadata editor

Download Gather Enrich Design Create

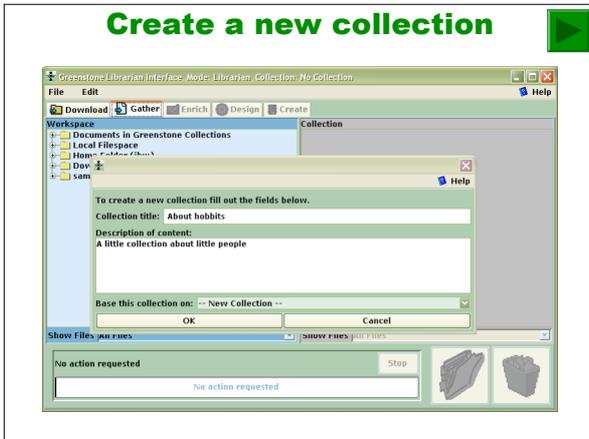
- ❖ Caveat: cannot deal with such huge collections as Greenstone can (particularly of metadata)

(Tutorial exercise #5: small collection of HTML files)

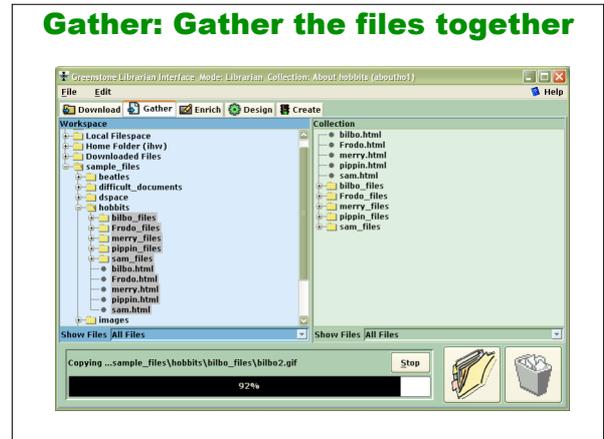
Invoke GLI: build a small collection of HTML files

- ❖ Gather
- ❖ Create
- ❖ Look at extracted metadata
- ❖ Set up shortcut in the Librarian interface

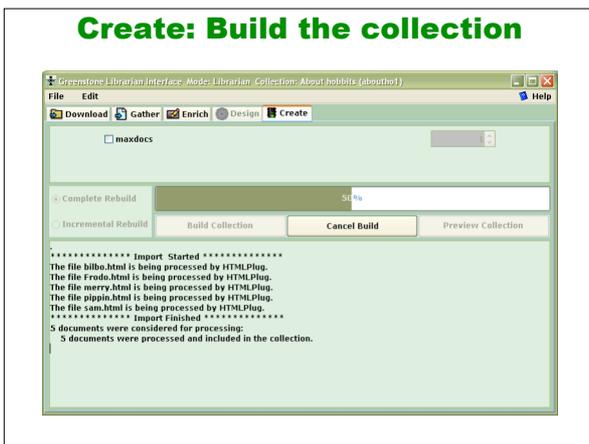
Create a new collection



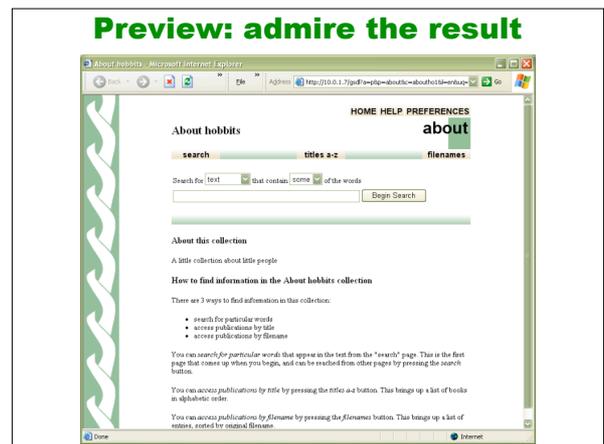
Gather: Gather the files together



Create: Build the collection



Preview: admire the result



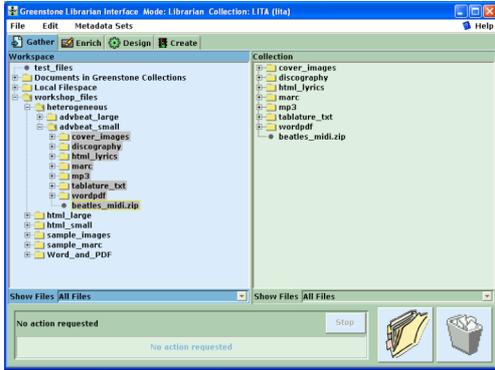
An example: Beatles collection



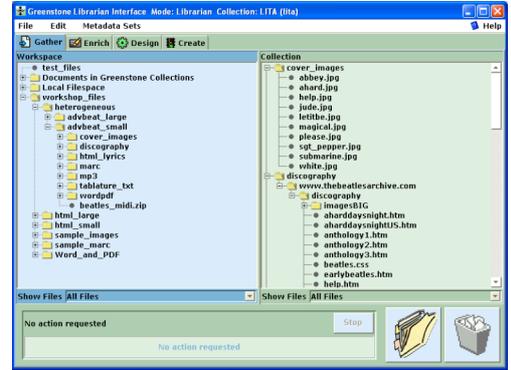
Building the Beatles collection



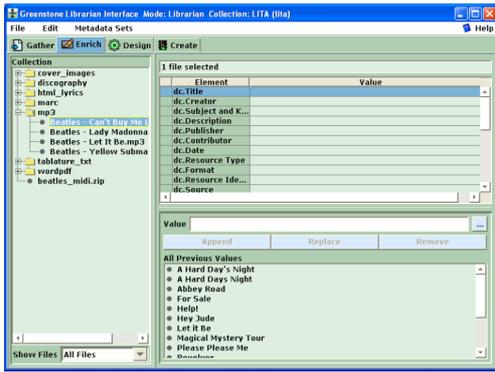
Gather: Gather the files together



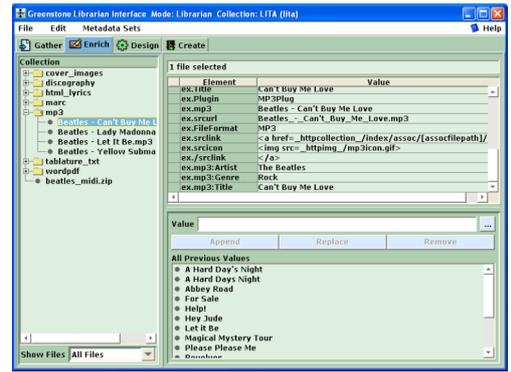
The ragtag managerie of documents



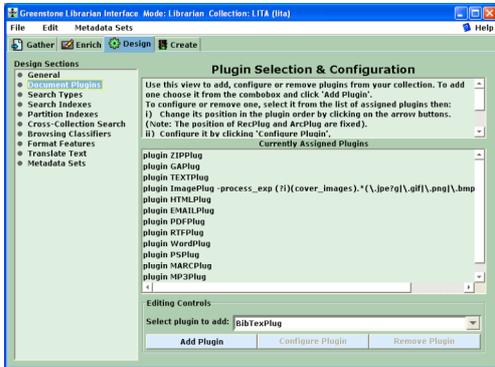
Enrich: Add metadata (if you like)



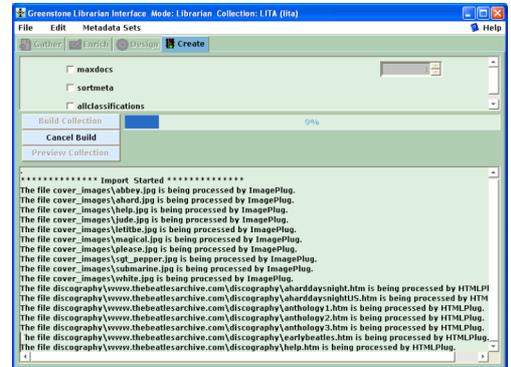
Enrich: Extracted metadata from MP3Plug



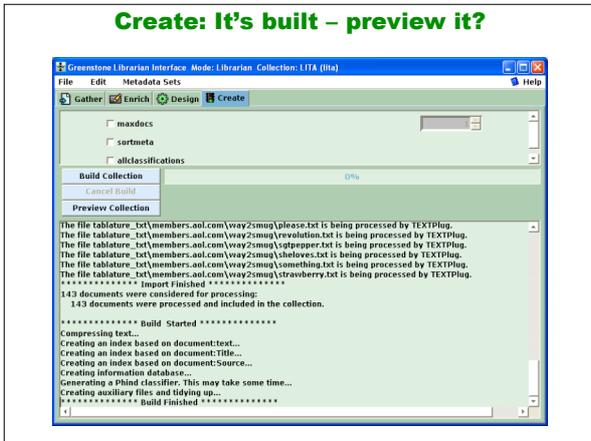
Design: Here are the plugins (and much more)



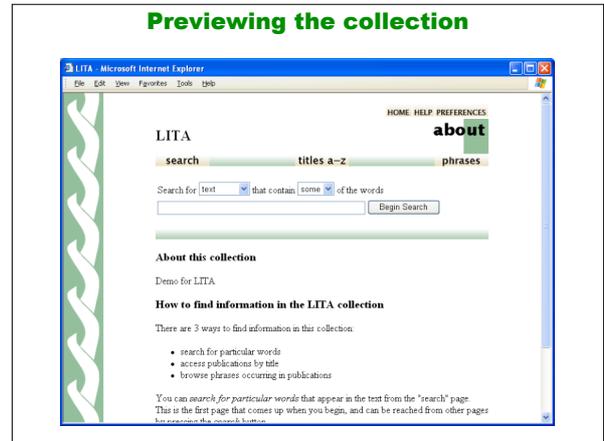
Create: Building the collection



Create: It's built – preview it?



Previewing the collection



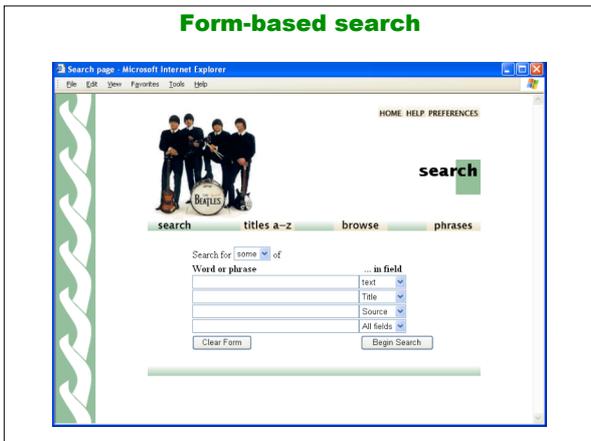
Export the collection to CD-ROM?



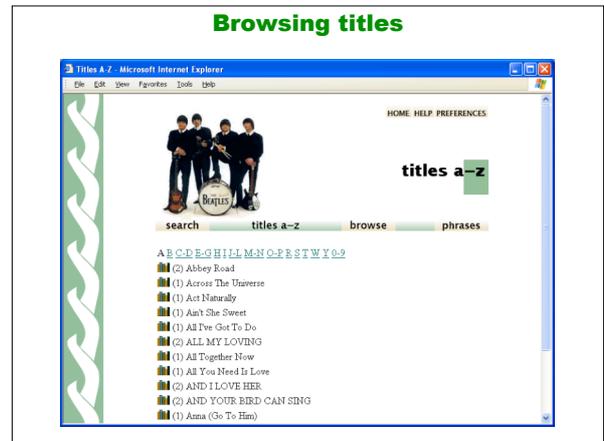
Full-text search



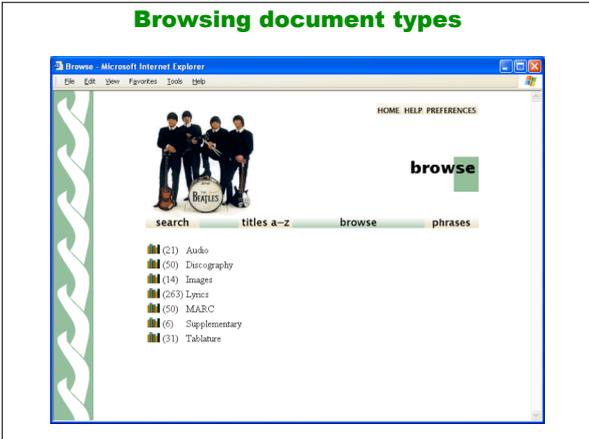
Form-based search



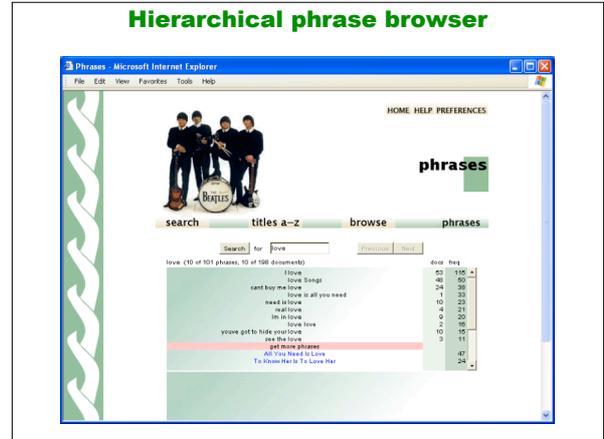
Browsing titles



Browsing document types



Hierarchical phrase browser



News flash: Applet version of GLI

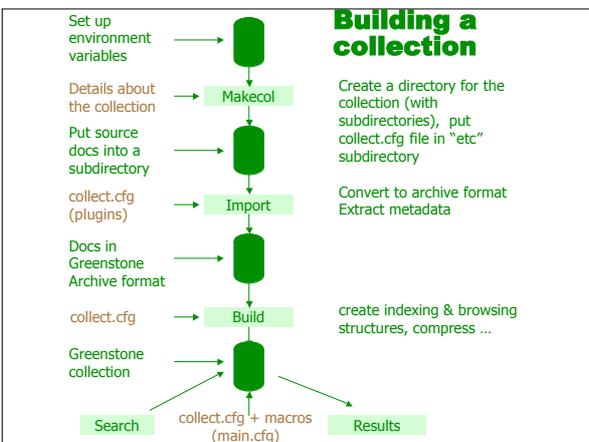
- ❖ Collection on remote Greenstone server
- ❖ Edit it locally
 - gather
 - enrich
 - design
 - create
- ❖ Authentication
 - uses admin facility
- ❖ Locking



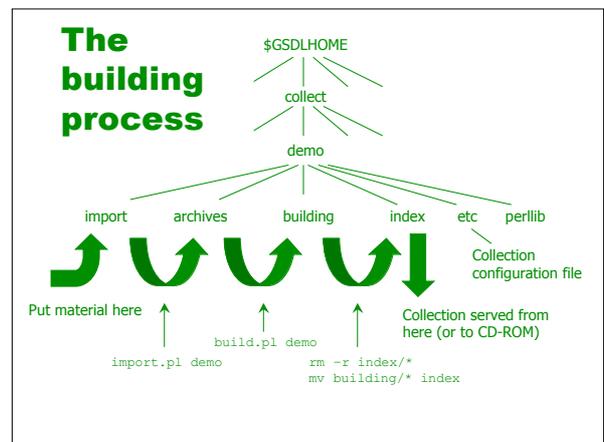
Agenda

- Overview**
 - What does Greenstone do?
 - Greenstone facts; standards
 - Reader's Interface: examples of collections
- Librarian interface**
 - Build a collection in 30 sec (Hobbits)
 - Build a multimedia collection (Beatles)
 - Adding and using metadata
 - Browsing classifiers, search indexes
 - Building a collection manually (for masochists only)
- Advanced stuff**
 - Under the hood: collection configuration file
 - Customizing with macros
 - Personalizing your home page
 - Different interface languages
 - Examples of what others have done
- Reaching out**
 - Serving and acquiring OAI
 - DSpace and METS
 - Greenstone3

Building a collection



The building process



The building process

```

Microsoft Windows [Version 5.00.2195]
(C) Copyright 1985-2000 Microsoft Corp.

C:\> cd "C:\Program files\gsdl"

C:\Program files\gsdl> setup

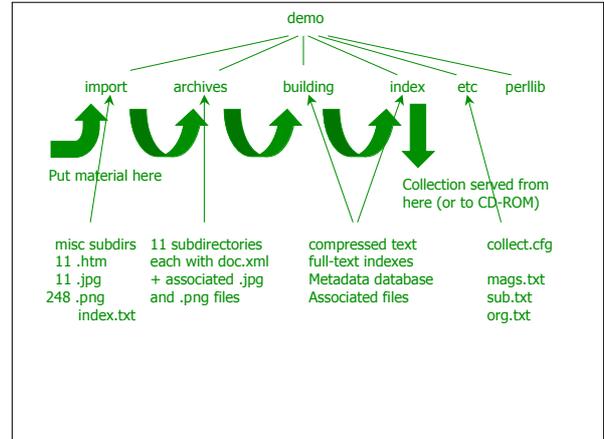
C:\Program files\gsdl> perl -S mkcol.pl
    -creator me@here colname

Copy source into collect\colname\import

C:\> perl -S import.pl -removeold colname

C:\> perl -S buildcol.pl colname

Rename the "building" directory to "index"
    
```



Agenda

- Overview**
 - What does Greenstone do?
 - Greenstone facts; standards
 - Reader's Interface: examples of collections
- Librarian interface**
 - Build a collection in 30 sec (Hobbits)
 - Build a multimedia collection (Beatles)
 - Adding and using metadata
 - Browsing classifiers, search indexes
 - Building a collection manually (for masochists only)
- Advanced stuff**
 - Under the hood: collection configuration file
 - Customizing with macros
 - Personalizing your home page
 - Different interface languages
 - Examples of what others have done
- Reaching out**
 - Serving and acquiring OAI
 - Dspace and METS
 - Greenstone3

Under the hood: Collection configuration file

- ❖ name, icon, etc
- ❖ description
- ❖ email of creator
- ❖ search indexes
- ❖ plugins
- ❖ 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
            SOLList -metadata Title
classify    Hierarchy -hfile org.txt -metadata Organization -sort Title
classify    List -metadata Howto

format      SearchWList "<td valign=top>[[link]][icon]/[[link]]</td>
                        <td>[[parent(All): ')]Title],[parent(All): ')]Title: ]
                        [[link]][Title]/[[link]]</td>"
            CLAVList:  "<td>[[link]]Howto/[[link]"
format      DocumentImages true
format      DocumentText "<h3>[[Title]]</h3>\n\n<pre>[[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/demo.gif"
collectionmeta iconcollection "/gsdl/collect/demo/images/demo.gif"
collectionmeta .section:title "section titles"
collectionmeta .document:text "entire books"
collectionmeta .section:text "chapters"
    
```

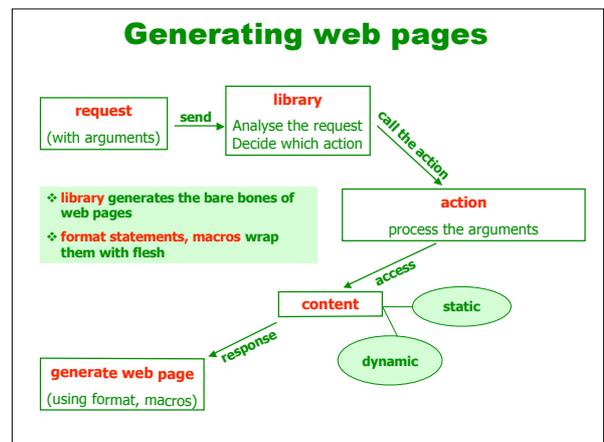
- ❖ documents
- ❖ query results
- ❖ classifiers

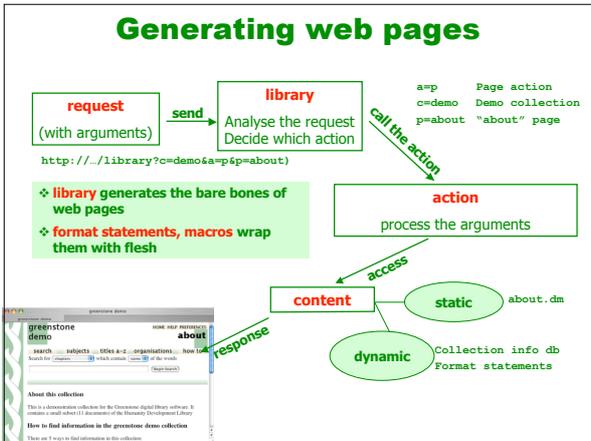
how to format

Alter configuration

- ❖ Add full-text index of titles
- ❖ ... or authors
- ❖ Add alphabetic author browser
- ❖ Include Word documents
- ❖ Include PDF documents
- ❖ Separate index for each language
- ❖ Extract acronyms and add list
- ❖ Import OAI metadata
- ❖ Extract phrase hierarchy and add browser
- ❖ Alter the format of any of the above
- ❖ Restrict collection's interface langs
- ❖ Change default interface language

indexes	document:Title
indexes	document:Creator
classify	AZList -metadata Creator
plugin	WordPlug
plugin	PDFPlug
languages	en fr es
plugin	PDFPlug -extract_acronyms
plugin	OAIPlug
classify	Phind
format ...	
format	PreferenceLangs en fr es
cgiaarg	shortcode=1 argdefault =fr





Customizing with macros

- let you customize presentation
- present pages in different languages
- print variables into the page text (e.g. number of search hits)
- ❖ **Macro files**
 - stored in `gsdl/macros` folder
 - each file defines one or more "packages" (A "package" is a group of macros)
 - loaded on startup (note difference between Local and Web Library)
 - listed in `etc/main.cfg`
- ❖ **Collection-specific macros**
 - Stored in `gsdl/collect/mycol/macros/extra.dm`
 - Or include argument `[c=collectionname]` for each macro

Personalizing your home page

C:\Program Files\gsdl\etc\main.cfg change *home.dm* to *yourhome.dm*

yourhome.dm

```

package home
_content {
<h2> Your own Greenstone home page</h2>
<ul>
<table>
<tr valign=top><td>Search page for the demo collection<br></td>
<td><a href="httpquery_d=c=demo">Click here</a></td></tr>
<tr><td>About page for the demo collection</td>
<td><a href="httppageabout_c=demo">Click here</a></td></tr>
<tr><td>Preferences page for the demo collection</td>
<td><a href="httppagepref_&c=demo">Click here</a></td></tr>
<tr><td>Home page</td>
<td><a href="httppagehome_">Click here</a></td></tr>
<tr><td>Help page</td>
<td><a href="httppagehelp_">Click here</a></td></tr>
<tr><td>Administration page</td>
<td><a href="httppagestatus_">Click here</a></td></tr>
<tr><td>The Collector</td>
<td><a href="httppagecollector_">Click here</a></td></tr>
</table>
</ul>
}
    
```

Macros used in home.dm

<code>_httppagehome_</code>	name of the home page
<code>_httppagehelp_</code>	... the help page
<code>_httppagestatus_</code>	... the administration page
<code>_httppagecollector_</code>	... the Collector page
<code>_httpquery_&c=demo</code>	search page for the demo collection
<code>_httppageabout_&c=demo</code>	about page for the demo collection
<code>_httppagepref_&c=demo</code>	preferences page for the demo collection
<code>_content_{ ... }</code>	defines a macro called <code>_content_</code> contains HTML, but <code>{</code> , <code>}</code> , <code>\</code> , and <code>'</code> must be escaped with a backslash
<code>_header_{ ... }</code>	HTML page header (contains squiry bar)
<code>_footer_{ ... }</code>	HTML page footer

`main.cfg` contains list of macros, replace *home.dm* by *yourhome.dm* and put it in the *macros* directory

about.dm

```

package about
#####
# about page content
#####
pagetitle {_collectionname_}
_content {
<center>
<navigationbar_
</center>
query:queryform_
<p> iconblankbar_
<p> textabout_
textsubcollections_
<h3> _help:textsimplehelpheading_ </h3>
help:simplehelp_
}
_textabout_{
<h3> _textabcol_ </h3>
_Global:collectionextra_
}
    
```

- ❖ can contain conditional statements
 - `_if_ ...`
- ❖ can take arguments
 - `a=p, l=d` etc.

Annotations in the image: 'Macro names begin and end with underscore' points to the curly braces in the macro definition; 'Content defined in curly brackets' points to the content between the curly braces; 'Content is plain text, HTML, other macros, Javascript, applet links' points to the HTML code within the macro definition.

Customization hints

- ❖ To change the look and feel of Greenstone
 - edit the *base* and *style* packages
- ❖ To change the query page
 - edit *query.dm*
- ❖ To change the Preferences page
 - edit *pref.dm*

Different interface languages

english.dm

```
#-----
# text macros
#-----

_textimageabout_ [l=zh] (About page)
_textimagehelp_ [l=zh] (Help page)
_textimagepref_ [l=zh] (Preferences page)
_textimagegreenstone_ [l=zh] (Greenstone Digital Library Software)

_textimagesearch_ [l=zh] (Search for specific terms)
_textimageCreator_ [l=zh] (Browse alphabetical list of authors)
_textimageTitle_ [l=zh] (Browse alphabetical list of titles)
_textmatches_ [l=zh] (Matches)
_textbeginsearch_ [l=zh] (Begin Search)

_texticonsmalltext_ [l=zh] (View this section of the text)
_texticondetach_ [l=zh] (Open this page in a new window)
_texticonexpandtoc_ [l=zh] (Expand table of contents)
_texticonexpandtext_ [l=zh] (Display all text)
_texticonhighlight_ [l=zh] (Highlight search terms)

_textlanguage_ [l=zh] (Interface language: )
_textsimplemode_ [l=zh] (simple query mode)
_textadvancedmode_ [l=zh] (advanced query mode (allows boolean searching u
```

chinese.dm

```
#-----
# text macros
#-----

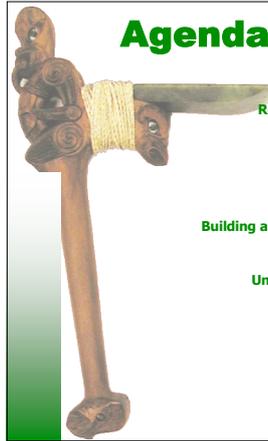
_textimageabout_ [l=zh] (主页)
_textimagehelp_ [l=zh] (帮助页)
_textimagepref_ [l=zh] (选项页)
_textimagegreenstone_ [l=zh] (绿石数字化图书馆软件)

_textimagesearch_ [l=zh] (特定范围检索)
_textimageCreator_ [l=zh] (按作者名字的阿拉伯字母顺序浏览)
_textimageTitle_ [l=zh] (按题名阿拉伯字母顺序浏览)
_textmatches_ [l=zh] (匹配)
_textbeginsearch_ [l=zh] (开始查询)

_texticonsmalltext_ [l=zh] (浏览该部分文本)
_texticondetach_ [l=zh] (打开一个新的窗口显示本页)
_texticonexpandtoc_ [l=zh] (扩展目录表)
_texticonexpandtext_ [l=zh] (显示所有文本)
_texticonhighlight_ [l=zh] (高亮显示搜索项)

_textlanguage_ [l=zh] (界面语言: )
_textsimplemode_ [l=zh] (简单查询方式)
_textadvancedmode_ [l=zh] (高级查询方式(可以使用 !, &, | 和序号来布尔查询))
```

Agenda



Overview

What does Greenstone do?
Greenstone facts; standards
Reader's Interface: examples of collections

Librarian interface

Build a collection in 30 sec (Hobbits)
Build a multimedia collection (Beatles)
Adding and using metadata
Browsing classifiers, search indexes
Building a collection manually (for masochists only)

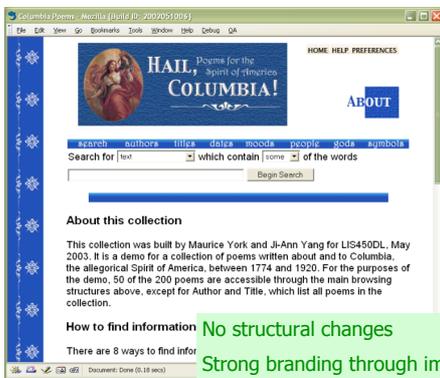
Advanced stuff

Under the hood: collection configuration file
Customizing with macros
Personalizing your home page
Different interface languages
Examples of what others have done

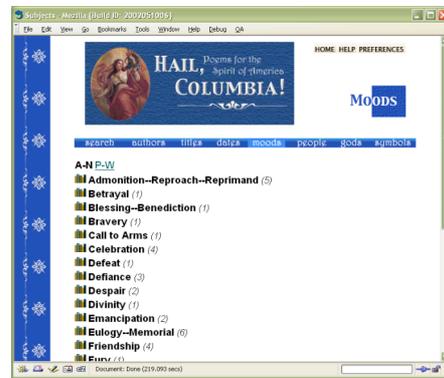
Reaching out

Serving and acquiring OAI
DSpace and METS
Greenstone3

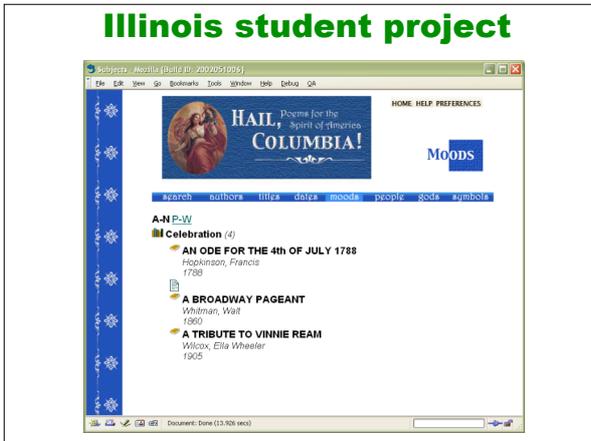
Illinois student project



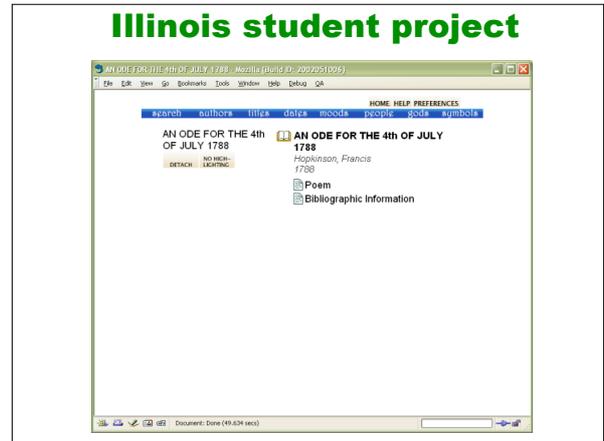
Illinois student project



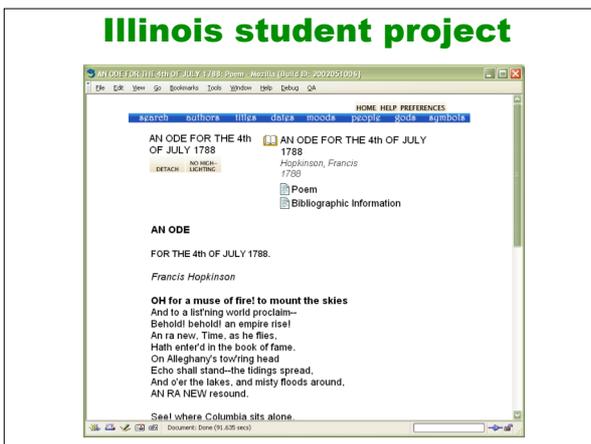
Illinois student project



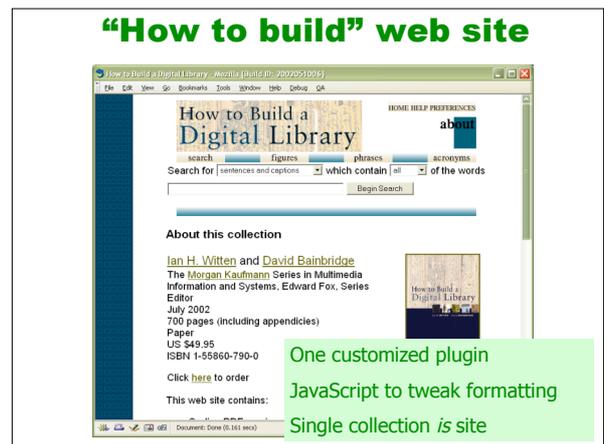
Illinois student project



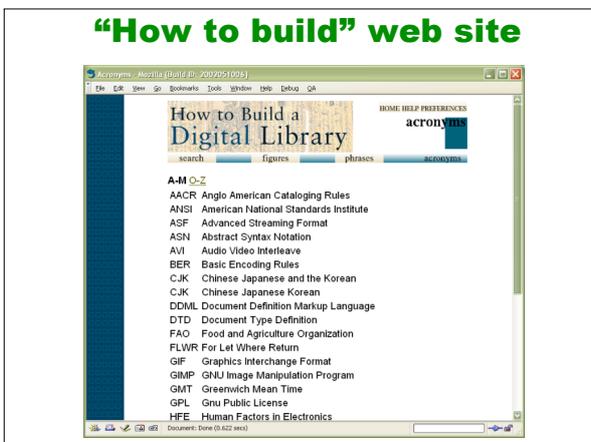
Illinois student project



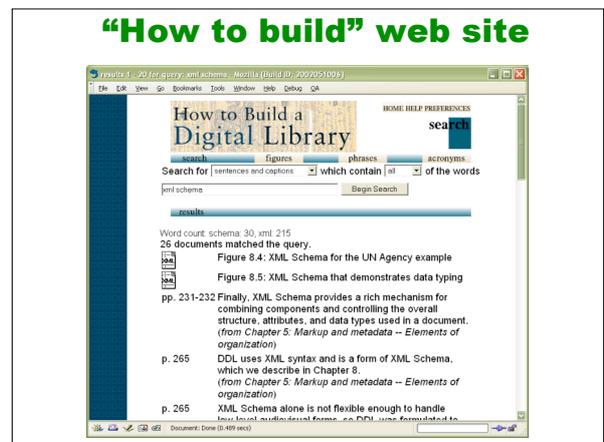
“How to build” web site



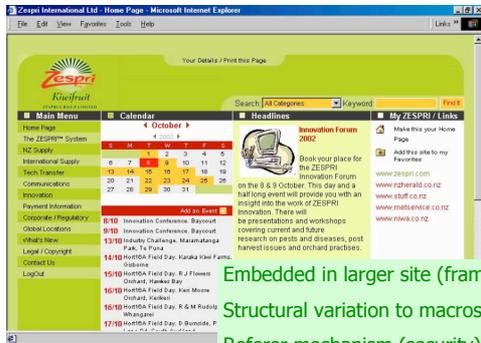
“How to build” web site



“How to build” web site

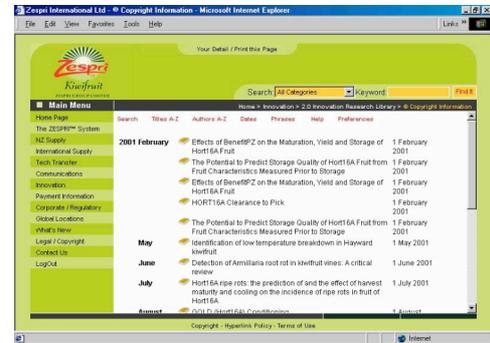


Zespri innovation library



Embedded in larger site (frames)
Structural variation to macros
Referer mechanism (security)

Zespri innovation library



Zespri innovation library

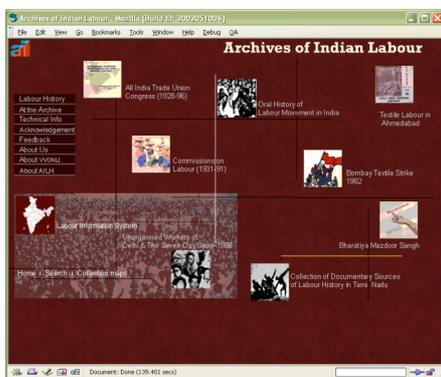


Archives of Indian labour



Login security to whole site
Home page customized

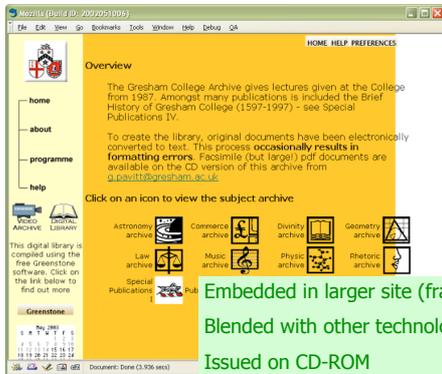
Archives of Indian labour



Archives of Indian labour

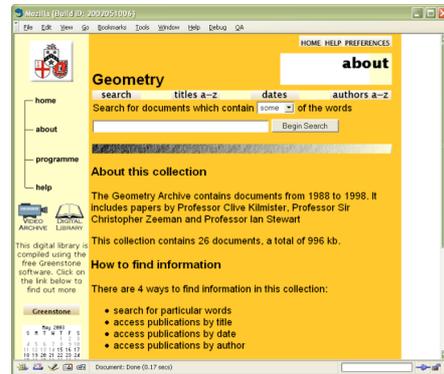


Gresham College archive

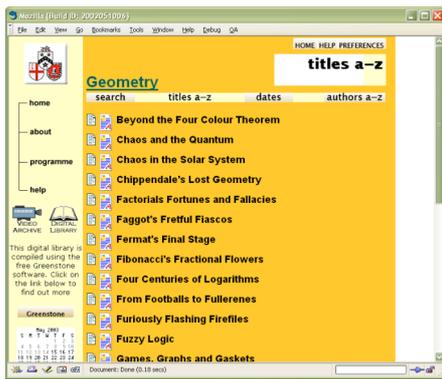


Embedded in larger site (frames)
Blended with other technologies
Issued on CD-ROM

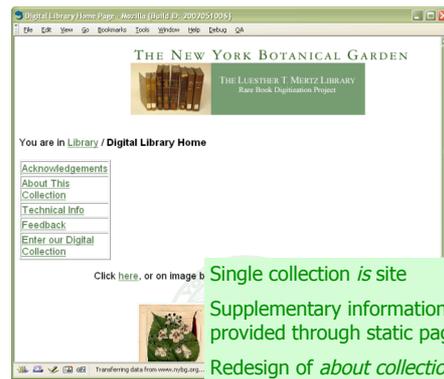
Gresham College archive



Gresham College archive

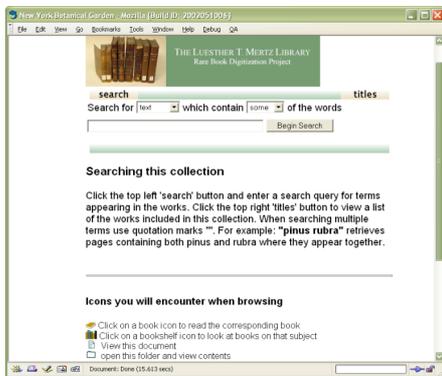


The NY Botanical Gardens



Single collection *is* site
Supplementary information provided through static pages
Redesign of *about* collection page

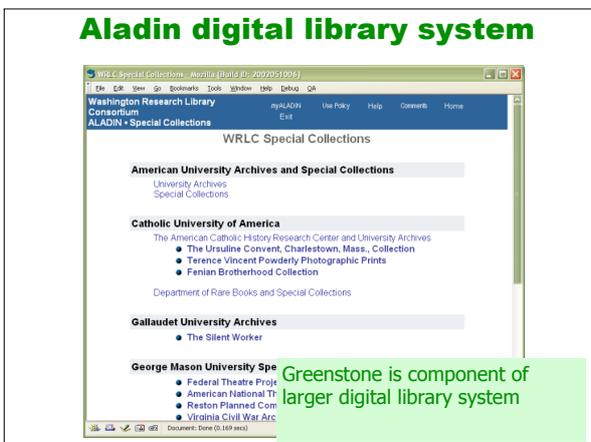
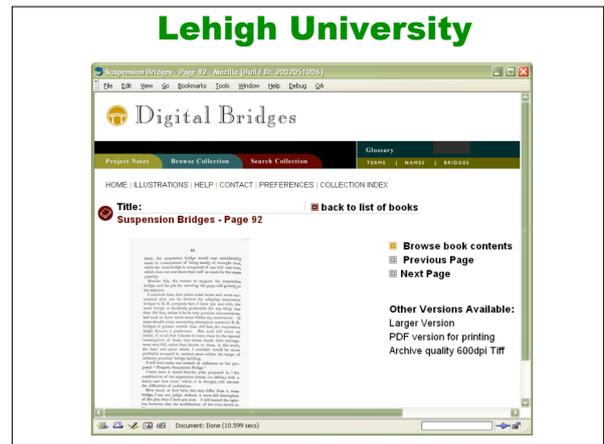
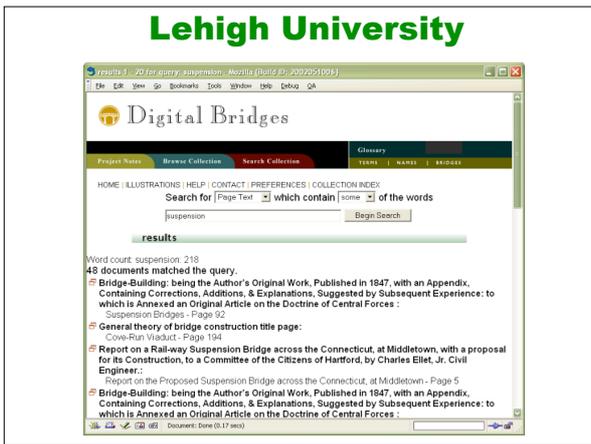
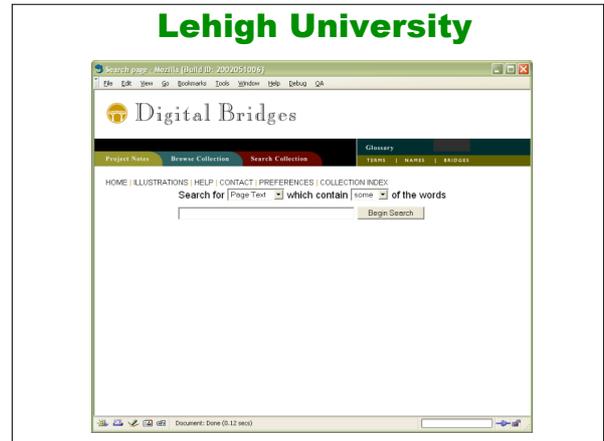
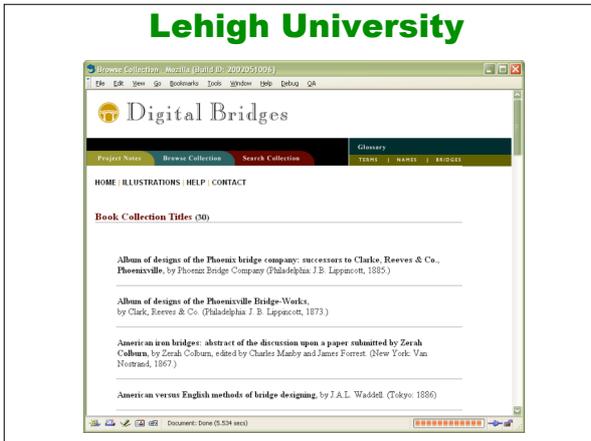
The NY Botanical Gardens



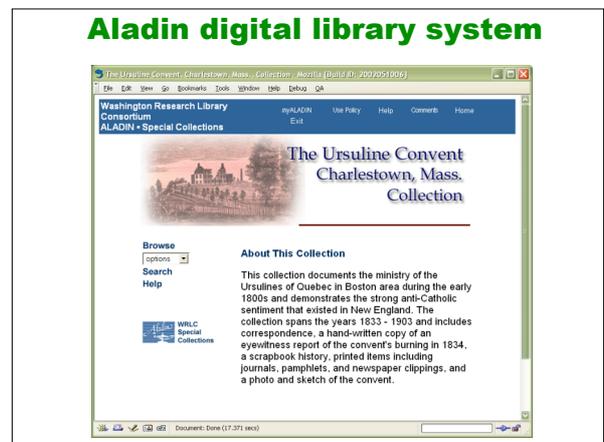
Lehigh University



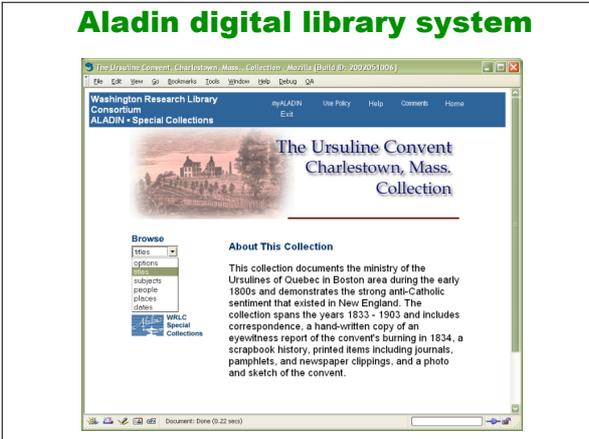
Single collection *is* site
Strongly branded
Search feature embedded into essentially static site



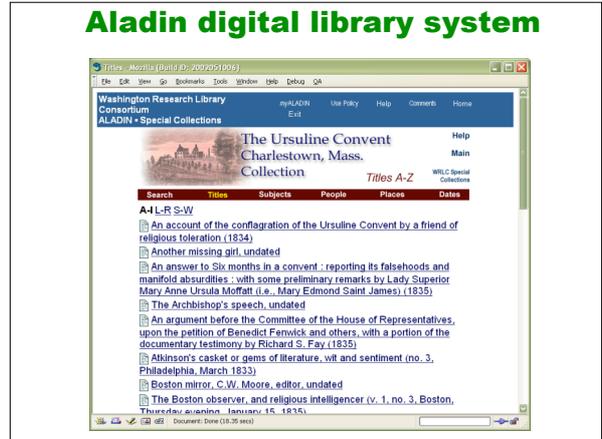
Greenstone is component of larger digital library system



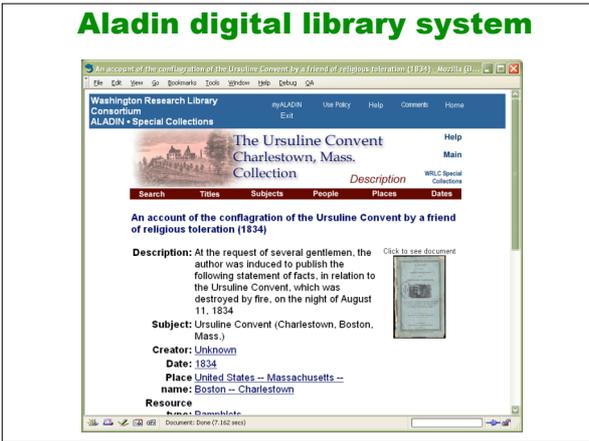
Aladin digital library system



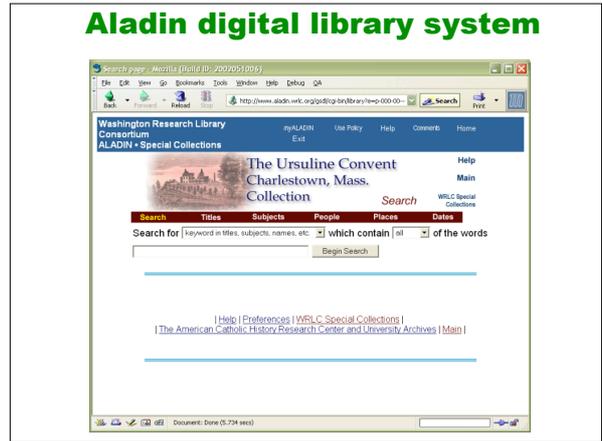
Aladin digital library system



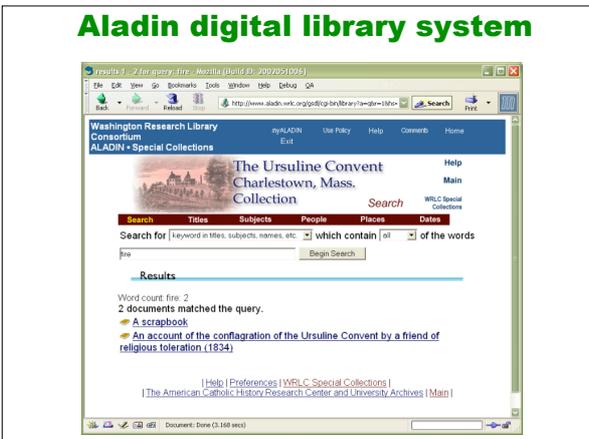
Aladin digital library system



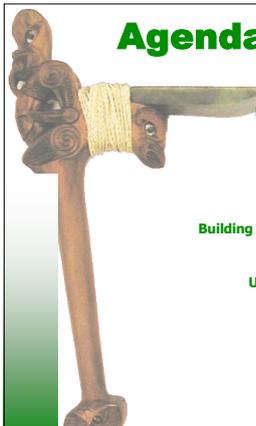
Aladin digital library system



Aladin digital library system



Agenda



- Overview**
 - What does Greenstone do?
 - Greenstone facts; standards
 - Reader's Interface: examples of collections
- Librarian interface**
 - Build a collection in 30 sec (Hobbits)
 - Build a multimedia collection (Beatles)
 - Adding and using metadata
 - Browsing classifiers, search indexes
 - Building a collection manually (for masochists only)
- Advanced stuff**
 - Under the hood: collection configuration file
 - Customizing with macros
 - Personalizing your home page
 - Different interface languages
 - Examples of what others have done
- Reaching out**
 - Serving and acquiring OAI
 - DSpace and METS
 - Greenstone3

The Greenstone OAI server

- ❖ Runs as a CGI program called *oaiserver*

- Greenstone installation
- OAI server

```

http://.../cgi-bin/gsd1
http://.../cgi-bin/oaiserver?verb=Identify
http://.../cgi-bin/oaiserver?verb=ListSets
http://.../cgi-bin/oaiserver?verb=ListIdentifiers&set=xxx
http://.../cgi-bin/oaiserver?verb=ListIdentifiers&set=xxx&metadataPrefix=oai_dc
http://.../cgi-bin/oaiserver?verb=ListRecords&set=xxx&metadataPrefix=oai_dc
http://.../cgi-bin/oaiserver?verb=GetRecordIdentifier&set=xxx&metadataPrefix=oai_dc
  
```

OAI 1.1
OAI 2.0

- ❖ Requires a full webserver (not “local library” version)
- ❖ Configuration file: `etc/oai.cfg` in the Greenstone filesystem
 - repository name and version (OAI 1.1 or 2.0)
 - collections to be made accessible to OAI clients
 - metadata mapping file into DC (server only supports DC)



Acquiring OAI metadata + docs

Using OAI-PMH, build a Greenstone collection based on metadata exported from an OAI server

1. Use Greenstone’s `importfrom.pl` command to acquire data from the JCDL01 collection at `rocky.dlib.vt.edu`
2. Use Greenstone’s `import.pl` and `buildcol.pl` commands to build a service provider based on the acquired metadata (and documents).

OAI Collection: acquisition

Use `importfrom.pl` to acquire metadata from the external data provider:

```

gsdl% importfrom.pl oai-e
OAI Acquire: from rocky.dlib.vt.edu/~jcdlpix/cgi-
bin/OAI1.1/jcdlpix.pl
Requesting list of identifiers ...
... Done.
Downloading metadata record for oai:JCDLPICS:200101dla1.oai
Getting document
http://rocky.dlib.vt.edu/~jcdlpix/pictures/200104dla/01dla1.jpg
Downloading metadata record for oai:JCDLPICS:200101dla2.oai
Getting document
http://rocky.dlib.vt.edu/~jcdlpix/pictures/200104dla/01dla2.jpg
Downloading metadata record for oai:JCDLPICS:200101dla3.oai
Getting document
http://rocky.dlib.vt.edu/~jcdlpix/pictures/200104dla/01dla3.jpg
...
Number of documents processed: 81
  
```

OAI Collection: acquisition

Excerpts from Greenstone collection configuration file.
Used by `importfrom.pl`, `import.pl` and `buildcol.pl`

```

acquire OAI -src rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI1.1/
jcdlpix.pl -getdoc
#...

plugin OAIPlug -input_encoding iso_8859_1 -default_language en
plugin ImagePlug -screenviewsize 300
plugin GAPLug
plugin ArcPlug
plugin RecPlug -use_metadata_files -show_progress
#...

classify AZCompactList -metadata Subject -doclevel top
classify AZCompactList -metadata Description -buttonname Captions
#...

format VList "<td>[link] [thumbicon] [/link]</td> \
<td valign=middle><i>[Description]</i></td>"
  
```

OAI Collection: building

Using OAI-PMH, build a Greenstone collection based on metadata exported from an OAI server

1. Use Greenstone’s `importfrom.pl` command to acquire data from the JCDL01 collection at `rocky.dlib.vt.edu`
2. Use Greenstone’s `import.pl` and `buildcol.pl` commands to build a service provider based on the acquired metadata (and documents).

(or just look at `oai-e`)

- ❖ Save Greenstone collection as DSpace, METS
- ❖ Ingest DSpace documents (DSpacePlug)

MODS: XSLT mapping

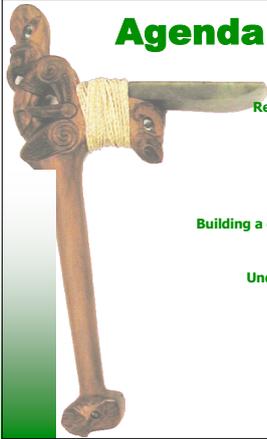
MODS mapping into Greenstone using XSLT
(specify as argument to OAIPPlug):

```

...
<xsl:template match="mods:topic">
  <dc:Subject><xsl:value-of select="." /></dc:Subject>
</xsl:template>

<xsl:template match="mods:titleInfo">
  <dc>Title>
    <xsl:value-of select="mods:title" />
    <xsl:if test="mods:subTitle">
      <xsl:text> </xsl:text>
      <xsl:value-of select="mods:subTitle" />
    </xsl:if>
  </dc>Title>
</xsl:template>
...

```



Agenda

- Overview**
 - What does Greenstone do?
 - Greenstone facts; standards
 - Reader's Interface: examples of collections
- Librarian interface**
 - Build a collection in 30 sec (Hobbits)
 - Build a multimedia collection (Beatles)
 - Adding and using metadata
 - Browsing classifiers, search indexes
 - Building a collection manually (for masochists only)
- Advanced stuff**
 - Under the hood: collection configuration file
 - Customizing with macros
 - Personalizing your home page
 - Different interface languages
 - Examples of what others have done
- Reaching out**
 - Serving and acquiring OAI
 - DSpace and METS
 - Greenstone3 ←

Greenstone 3

Greenstone 2 was designed about 7 years ago. Since then, new software and library standards have emerged ...

- ❖ **Easier to modify interface: generates data (XML) rather than presentation (HTML)**
 - XML can be transformed using XSLT, rather than HTML via macro files
- ❖ **More dynamic: more can be done at runtime**
 - e.g. sort search results on the fly, generate new classifiers at runtime
- ❖ **METS foundation: multiple hierarchies possible**
 - e.g. chapter-section-subsection hierarchy can coexist with page header-contents-footer hierarchy
- ❖ **Distributed modules, using SOAP**
 - interface can be on different computer to collection, can talk to several different Greenstone sites
- ❖ **Java basis: platform independent**
 - Greenstone2 operates on different platforms, but at a high cost to implementors
- ❖ **Better software technology**
 - easier for people (e.g. our students!) to modify the software to make it do radically different things