



Greenstone Digital Library Software

Presented by:
Dynal Patel (dpatel@cs.uct.ac.za)

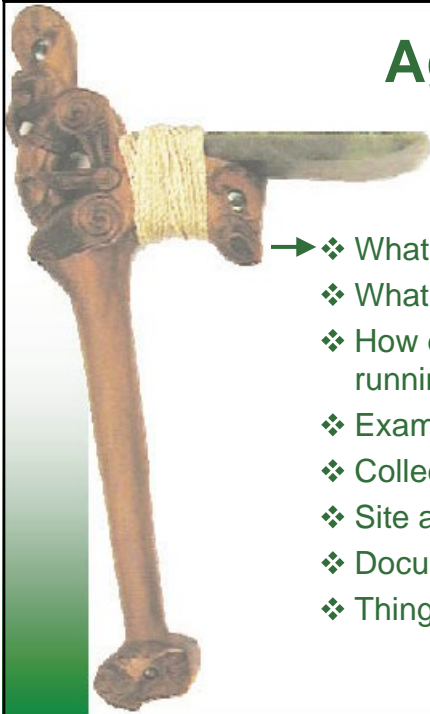
Course material prepared by:

- *African Digital Library Centre*
- *New Zealand Digital Library Group*
- *National Centre for Science Information*

Workshop materials

- ❖ Workshop notes
 - Presentation slides, lab instructions, lab screenshots, test files
- ❖ Software
 - Greenstone V2.4.1 (includes Librarian)
 - Java 1.4.1 run time environment
 - Community Development Library CD
- ❖ Greenstone Help Documentation
 - Installers, users, and developers guide
 - Papers
 - From Paper to Collection
 - Customizing Greenstone
- ❖ Evaluation form

Agenda



- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ How do we get Greenstone up and running?
- ❖ Example Greenstone Collections.
- ❖ Collection Building.
- ❖ Site and Collection Customization.
- ❖ Documentation and help.
- ❖ Things to look forward to.

What is a digital Library (DL)?

- ❖ A digital library is like a “regular” library except: contents digitized, services computerized, and clientele delocalized.
- ❖ It offers access to not only to text, but to just about anything that can digitized: music, film clips, rare paintings, x-rays, photographs, ancient manuscripts and more.
- ❖ It provides methods for access and retrieval, selection, organisation, security and maintenance.

DL Software Key Requirements

- ❖ Document types (book, journal,...)
- ❖ Document formats (text, word, pdf, ps,...)
- ❖ Content acquisition (online and offline)
 - metadata description, content tagging,
 - content uploading
- ❖ Indexing and retrieval
 - Structured and full text indexing
 - Automatic metadata extraction
- ❖ Storage
 - Data compression
 - Efficient storage of data and metadata

DL Software Key Requirements (cont)

- ❖ Access and delivery
 - Structured search, browse, hierarchical browsing
 - CDROM distribution
- ❖ Scalability – for large collections
- ❖ Multilingual support
- ❖ Security and access management
- ❖ Usage Monitoring and reporting
- ❖ Standards compliance
 - XML, Dublin Core, Unicode
- ❖ Interoperation
 - OAI, Z39.50 compliance, MARC, CDS/ISIS,...



Agenda

- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ How do we get Greenstone up and running?
- ❖ Example Greenstone Collections.
- ❖ Collection Building.
- ❖ Site and Collection Customization.
- ❖ Documentation and help.
- ❖ Things to look forward to.

What is the Greenstone Software?

- ❖ Software suite for building, maintaining and distributing digital library collections
- ❖ Developed by the New Zealand Digital Library Group at the University of Waikato
- ❖ Open source project
- ❖ Distribution Partners
 - UNESCO
 - Human Info NGO, Belgium
 - NCSI, Bangalore, UCT, Dakar, Senegal, Almaty, Kazakhstan,...
 - You!

Example: Human Development Library

for sustainable development and basic human needs

- ❖ 160,000 pages
- ❖ 30,000 images
- ❖ 800 books
- ❖ 430 magazines
- ❖ 340 kg
- ❖ US\$20,000
- ❖ CD-ROM
- ❖ US\$6
- ❖ Win3.1x upwards
- ❖ Standalone
- ❖ Provides an intranet server
- ❖ Web browser user interface

Why Greenstone?

- ❖ FREE - Software, documentation, and user support
- ❖ Well funded project
- ❖ Designed for developing nations
- ❖ Extensive documentation and Fast User Support
- ❖ Fully documented
- ❖ Multilingual support for documents and interfaces
- ❖ Multiplatform - Windows/ Unix / Mac OS-X
- ❖ Multiformat – documents (e.g. HTML, word, pdf, ps, plain text, email,...) and metadata (XML, DC, OAI, MARC,...)
- ❖ Stable and reliable, has been tested under many configurations

Why Greenstone? (cont)

- ❖ Scalability – Can support large collections up to several Gb of text, associated images, movies, whatever...
- ❖ Trivial to Install
- ❖ Provides tool for gathering, enriching and building and updating collections
- ❖ Serve collections on the Web or write them to CD-ROM
- ❖ Boasts some of the worlds best compression and searching algorithms
- ❖ Has a number of novel ways of browsing information (e.g. Phrase browsing)
- ❖ Provides support for access management, site management, usage monitoring and error reporting

What are the alternatives?

- ❖ Develop a local web-based application
 - Development costs + Maintenance Costs
- ❖ Use a commercial DL solution
 - Initial costs, licensing costs, upgrade costs
- ❖ Adapt other existing open source software
 - Eprints, DSpace, Koha, CDS/ISIS
 - Need to consider development cost, long term project support, usability, documentation and user support
 - Unseen problems may occur when adapting software that wasn't designed for ones intended application



Agenda

- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ How do we get Greenstone up and running?
- ❖ Example Greenstone Collections.
- ❖ Collection Building.
- ❖ Site and Collection Customization.
- ❖ Documentation and help.
- ❖ Things to look forward to.

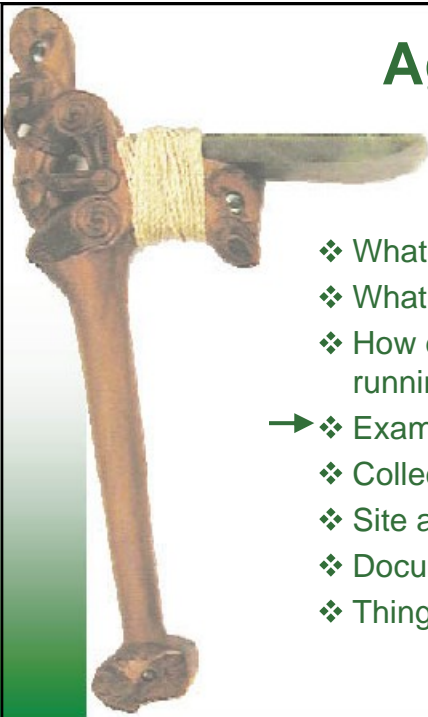
Where do I to get a copy?

- ❖ Download Greenstone v2.41
<http://www.greenstone.org>
- ❖ Download Installers Guide (install.pdf)
<http://www.greenstone.org>
 - Installation procedure, Greenstone collections, setting up the web server, configuring your site, personalising your installation

Installation issues

- ❖ What operating system?
 - windows, linux, unix, mac os x
- ❖ Disk Space
 - 50 MB for binary installation
 - 215 MB for optional example collections
 - 5 MB for online documentation
 - 25 MB for export to CD function
- ❖ Local Library vs. Web Library
 - Local library version includes a built in web server to serve collections on a standalone PC and to others on the same network
 - Web library requires you to setup an external web server (e.g. apache (free) or Microsoft PWS/ IIS (available with windows))
- ❖ Librarian interface needs Java

Agenda



- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ How do we get Greenstone up and running?
- ❖ Example Greenstone Collections.
 - ❖ Collection Building.
 - ❖ Site and Collection Customization.
 - ❖ Documentation and help.
 - ❖ Things to look forward to.

Example Greenstone Collections

- ❖ Has been rapid growth in use
 - International – China, Germany, India, UK, USA, Russia, Brazil, etc...
 - Increasing activity on Greenstone mailing lists
 - Promotion by UNESCO
- ❖ Wide variety of DL collections have been developed in several languages
 - Historical, educational, cultural, and research

What does a Greenstone
collection look like?

DEMO

Russian

Mari El Republic
<http://gov.mari.ru/gSDL>

Региональные
документы

Марийский
фольклор

ЭДВЕНТА
ИНТЕРУБЛАНКИ

Федеральные
документы

Spanish

PAHO, WHO



French

UNESCO, Paris



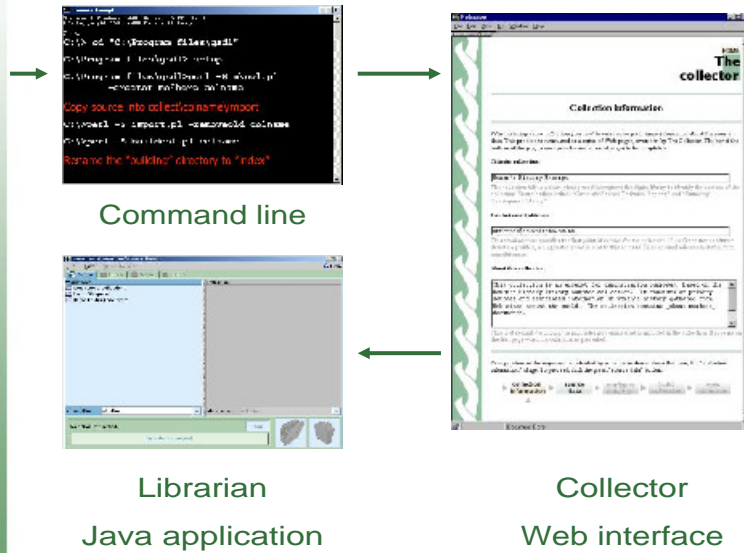




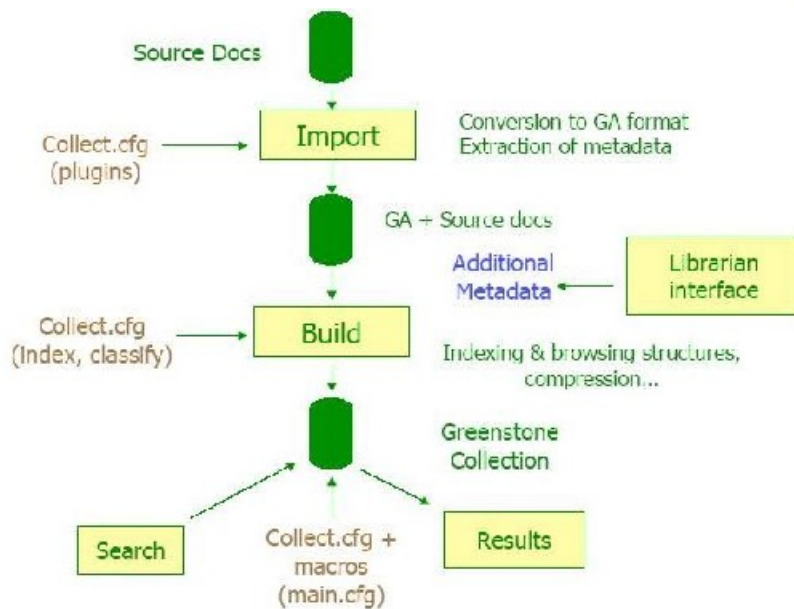
Agenda

- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ How do we get Greenstone up and running?
- ❖ Example Greenstone Collections.
- ➔ ❖ Collection Building.
- ❖ Site and Collection Customization.
- ❖ Documentation and help.
- ❖ Things to look forward to.

Collection Building



Basic Collection Building Process





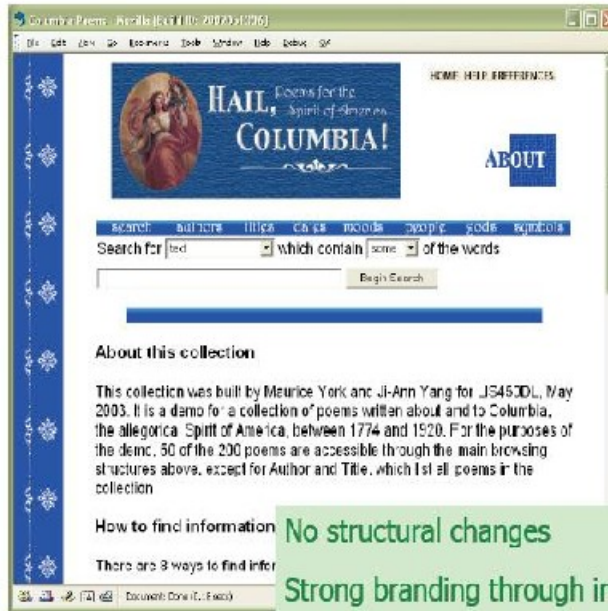
Agenda

- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ How do we get Greenstone up and running?
- ❖ Example Greenstone Collections.
- ❖ Collection Building.
- ❖ Site and Collection Customization.
- ❖ Documentation and help.
- ❖ Things to look forward to.

Site and Collection Customization

- ❖ You may need to alter
 - Collection configuration files (collect.cfg)
 - Macro files (*.dm)
 - Let you customize presentation
 - Present pages in different languages
 - Images
 - Site specific configuration files (main.cfg)
- ❖ Unfortunately still quite low level and requires librarians to know about Greenstone Visualisation mechanism.
- ❖ Good news: Developing a tool to simplify customization and provide complete abstraction from Greenstones low level architecture.

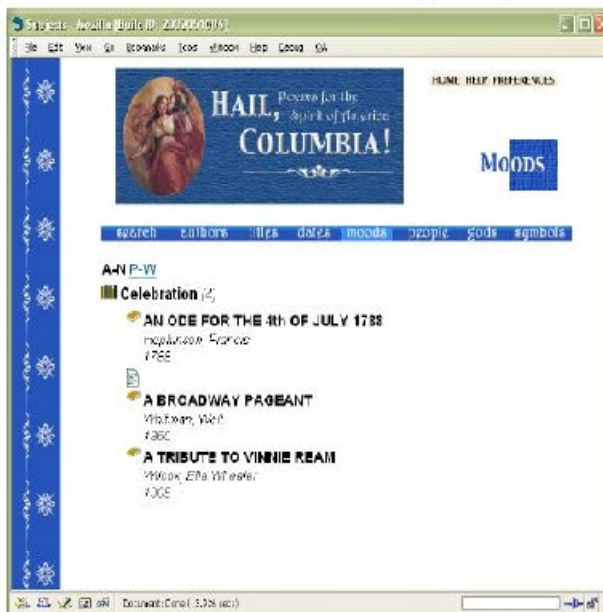
Illinois student project



No structural changes

Strong branding through images

Illinois student project



Zespri innovation library

The screenshot shows the Zespri innovation library website. The main content area features a calendar for October 2002. The calendar highlights several events, including the 'Innovation Forum 2002' on October 2nd and 3rd, and the 'Zespri Innovation Library' on October 4th. The website also includes a search bar, a navigation menu, and a list of links.

Embedded in larger site (frames)

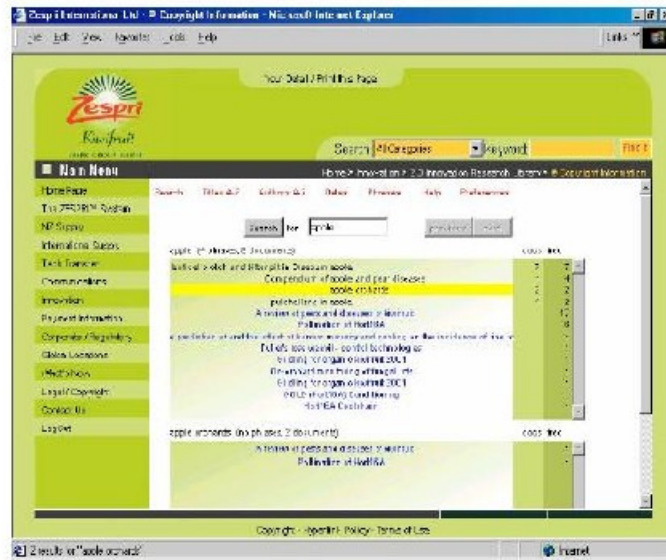
Structural variation to macros

Referer mechanism (security)

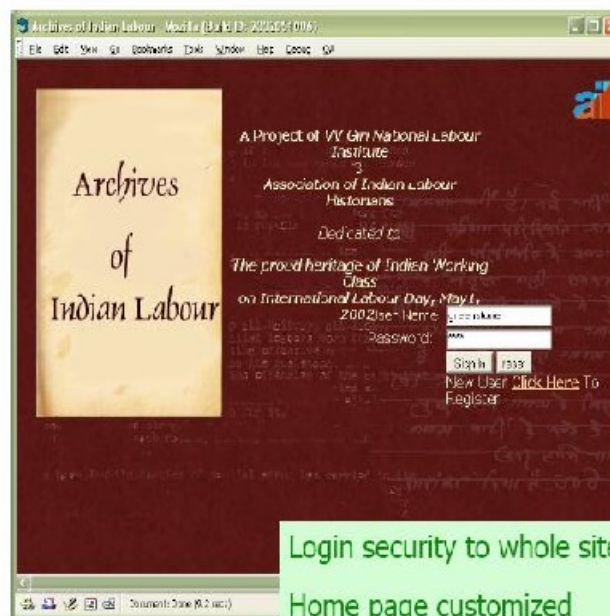
Zespri innovation library

The screenshot shows the Zespri innovation library website. The main content area features a list of events for February 2001. The events include 'Effects of Benet 12 on the Malaria, Field and Storage of Hort EA 1-11', 'The Potential to Predict Storage Quality of Hort EA 1-11 from First Characteristic Measure from Storage', 'Effects of Benet 12 on the Malaria, Field and Storage of Hort EA 1-11', 'Hort EA 1-11 Clearance to PIC', 'The Potential to Predict Storage Quality of Hort EA 1-11 from First Characteristic Measure from Storage', 'Identification of storage quality breakdown in Hort EA 1-11', 'Detection of storage quality breakdown in Hort EA 1-11', and 'Hort EA 1-11: the impact of the effect of storage quality and control on the incidence of storage quality breakdown'. The website also includes a search bar, a navigation menu, and a list of links.

Zespri innovation library



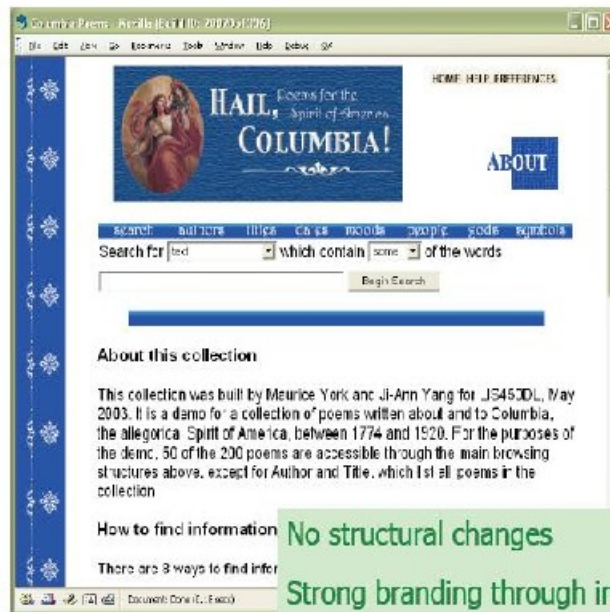
Archives of Indian labour



Login security to whole site

Home page customized

Illinois student project



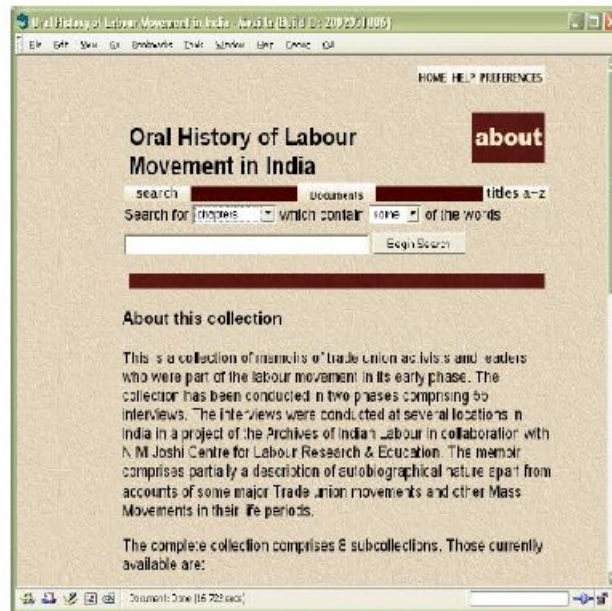
No structural changes

Strong branding through images

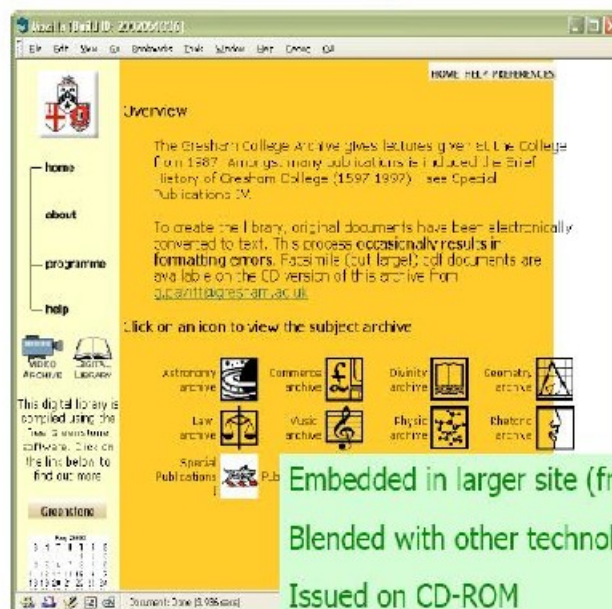
Archives of Indian labour



Archives of Indian labour

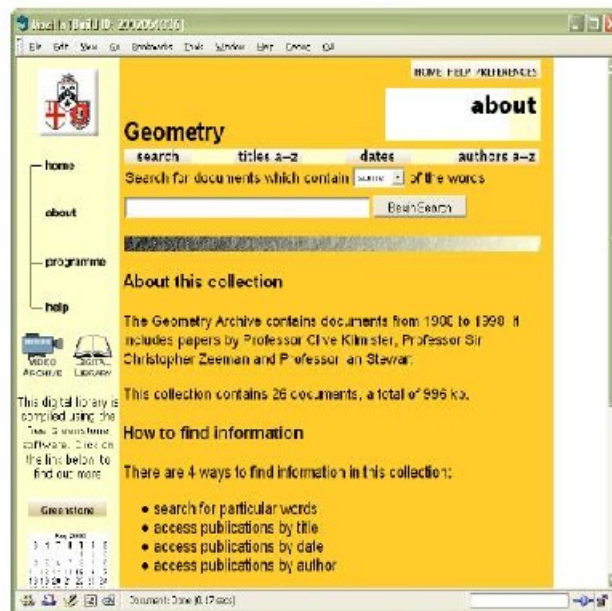


Gresham College archive

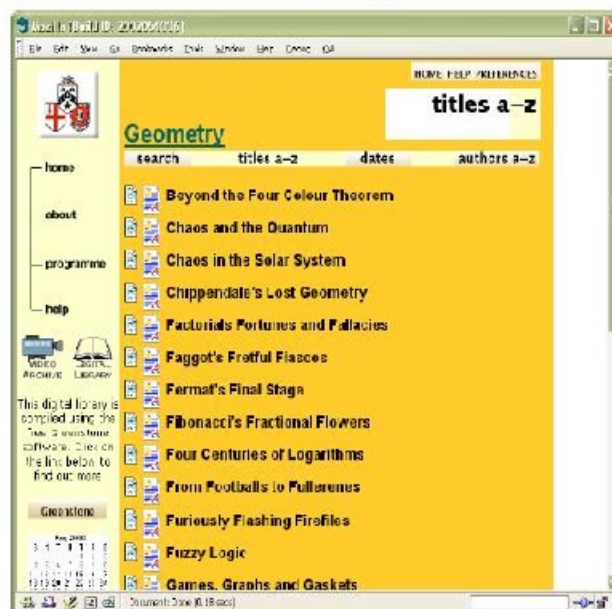


Embedded in larger site (frames)
Blended with other technologies
Issued on CD-ROM

Gresham College archive



Gresham College archive



The NY Botanical Gardens

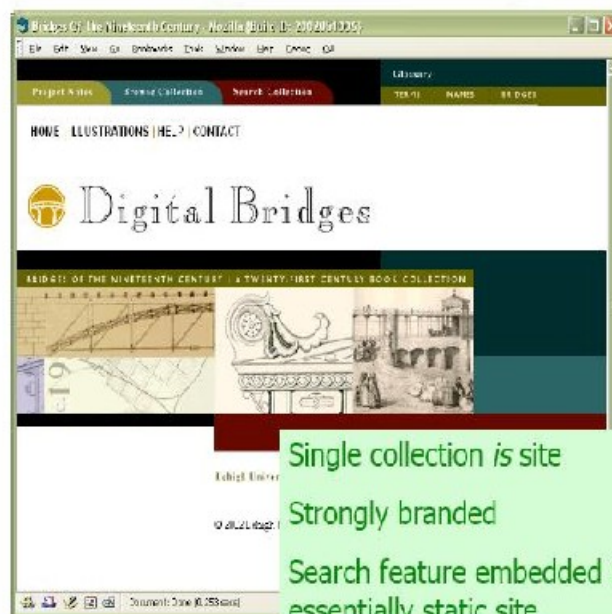


Single collection *is* site

Supplementary information provided through static pages

Redesign of *about collection* page

Lehigh University

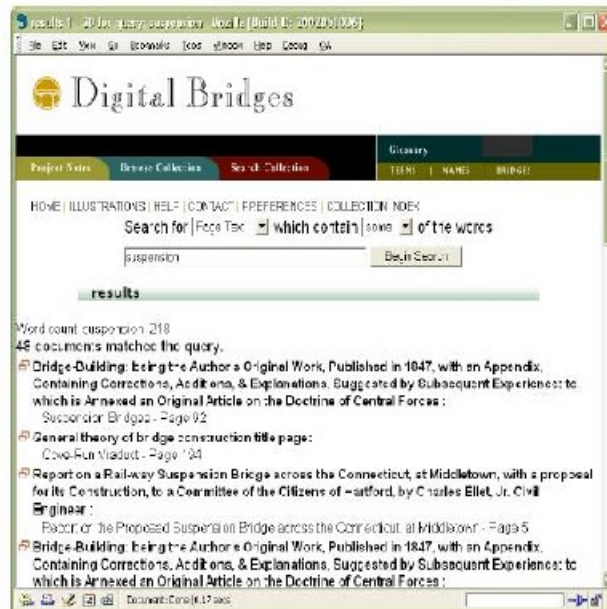


Single collection *is* site

Strongly branded

Search feature embedded into essentially static site

Lehigh University



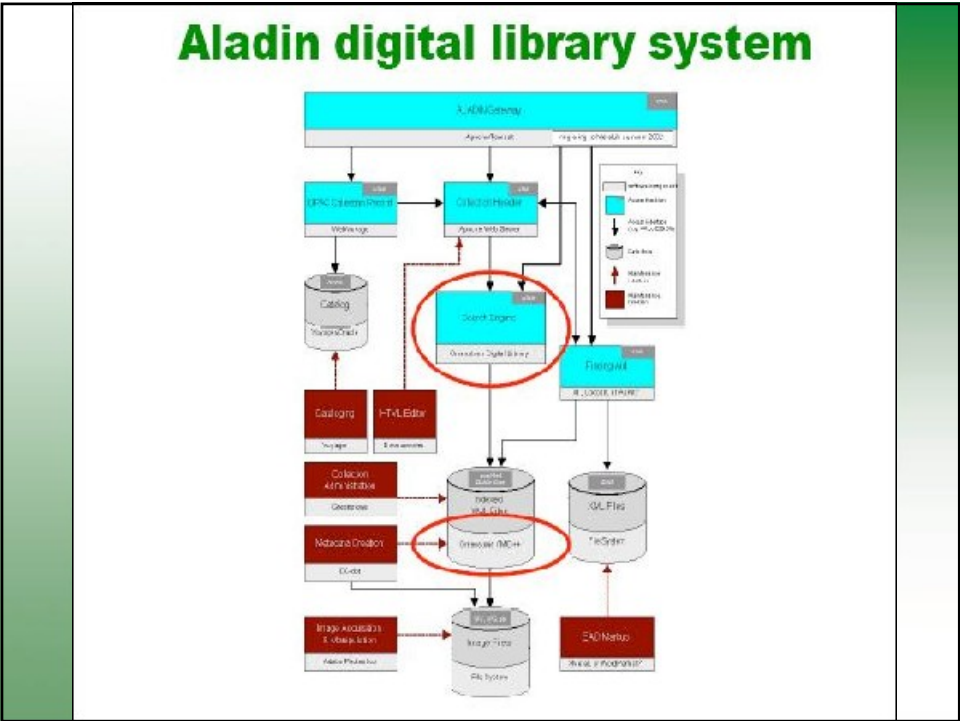
Lehigh University



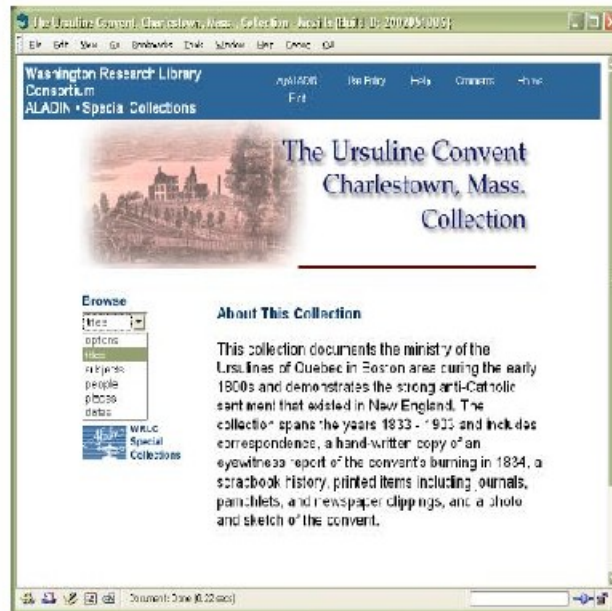
The screenshot displays the ALADIN (American Library of Ancient and Digital Information Network) web interface. The browser window title is "WRLC Special Collections: Aladdin Book ID: 2002040005". The address bar shows "http://www.wrlc.org/aladdin/". The main content area lists several digital collections:

- Washington Research Library Consortium**
 - ALADIN - Special Collections
- WRLC Special Collections**
- American University Archives and Special Collections**
 - University Archives
 - Special Collections
- Catholic University of America**
 - The American Catholic History Research Center and University Archives
 - The Ursuline Convent, Charlestown, Mass., Collection
 - Terence Vincent Powderly Photographic Prints
 - Ferrian Brotherhood Collection
 - Department of Rare Books and Special Collections
- Colgate University Archives**
 - The Silent Worker
- George Mason University Special Collections**
 - Federal Theatre Project
 - American National Theatre
 - Reston Planning Commission
 - Virginia Civil War Archives

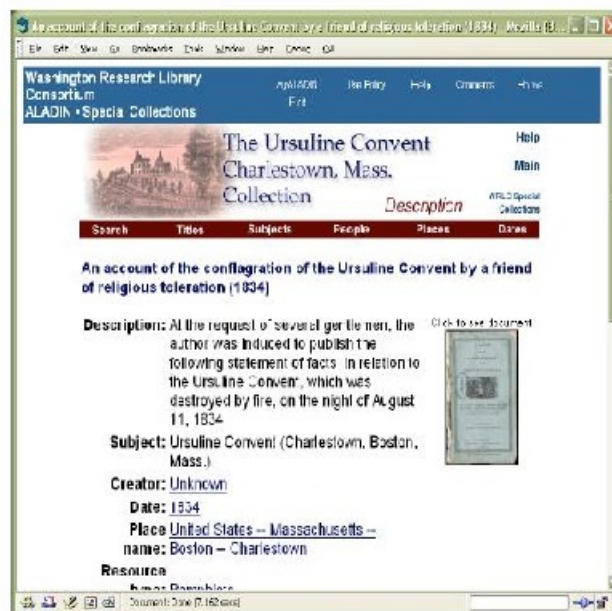
The interface includes a menu bar with options like "File", "Edit", "View", "Go", "Database", "Tools", "Search", "User", "Contact", and "Help". A status bar at the bottom indicates "Document: 2002040005".



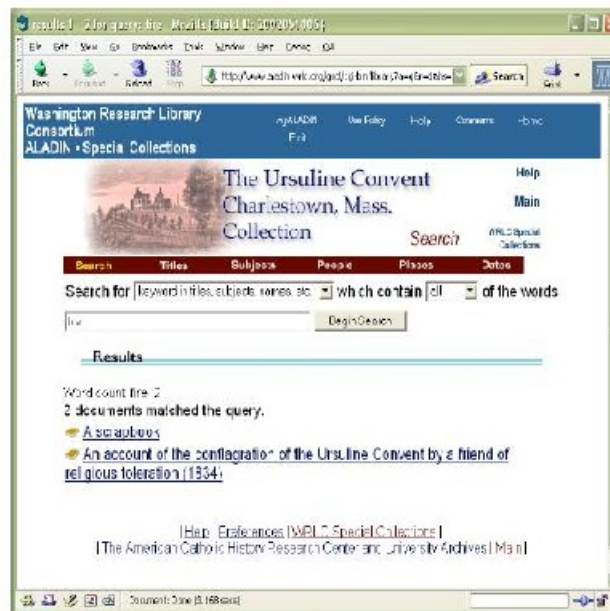
Aladin digital library system



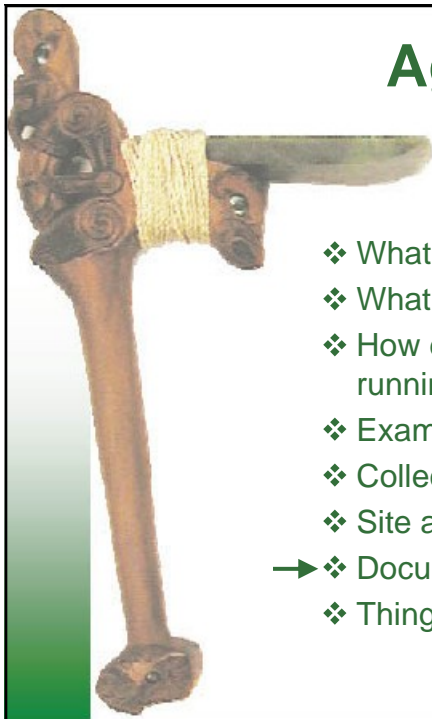
Aladin digital library system



Aladin digital library system



Agenda



- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ How do we get Greenstone up and running?
- ❖ Example Greenstone Collections.
- ❖ Collection Building.
- ❖ Site and Collection Customization.
- ➔❖ Documentation and help.
- ❖ Things to look forward to.

Documentation and help

- ❖ Available at www.greenstone.org
 - Software (Greenstone, Librarian, Export to CD package, The Organiser)
 - Example sites that have used Greenstone
 - Example Configuration Files
 - Documentation (Installer's Guide, User's Guide, Developer's Guide, From Paper to Collection, Instead Greenstone Collections, MGPP user guide, Customizing Greenstone Software)
 - Frequently asked questions
 - Mailing Lists
 - Greenstone User's List
 - Greenstone Developer's List
 - Additional Support – can submit query to support staff
- ❖ www.nzdl.org (New Zealand Digital Library Group)
 - Documented Example Collections
 - Information about the New Zealand Digital Project

Mirror Sites

- ❖ The African Digital Library Centre
University of Cape Town
<http://greenstone.cs.uct.ac.za>
- ❖ The Southern Alberta DL Centre
University of Lethbridge
<http://www.sadl.uleth.ca>



Agenda

- ❖ What is digital library software?
- ❖ What does Greenstone do?
- ❖ How do we get Greenstone up and running?
- ❖ Example Greenstone Collections.
- ❖ Collection Building.
- ❖ Site and Collection Customization.
- ❖ Documentation and help.
- ❖ Things to look forward to.

Things to look forward to

❖ Greenstone 3

- Offers everything Greenstone 2 does and much more
- Will be backwards compatible
- Improve modularity, flexibility and extensibility.
- Lower the bar for getting into Greenstone code with a view to understanding it and extending it
- Makes full use of existing XML related standards and software and Java
- Improves internationalisation, information browsing etc
- New content management features
- Operate on a scale ranging from personal desktop to corporate library
- Allows easy incorporation of text mining operations

❖ Interface Customisation tool for G2 and G3

