Creating and Customizing Digital Library Collections with the Greenstone Librarian Interface

Ian H. Witten
New Zealand Digital Library Project
Department of Computer Science
University of Waikato, New Zealand
ihw@cs.waikato.ac.nz

Abstract

The Greenstone digital library software is a comprehensive system for building and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet. This paper describes how digital library collections can be created and customized with the new Greenstone Librarian Interface. Its basic features allow users to add documents and metadata to collections, create new collections whose structure mirrors existing ones, and build collections and put them in place so for users to view. More advanced users can design and customize new collection structures. At the most advanced level, the Librarian Interface gives expert users interactive access to the full power of Greenstone, which could formerly be tapped only by running Perl scripts manually.

Keywords: librarian interface, creating digital library collections, customization, Greenstone digital library software

1. Introduction

Digital libraries are organized, focused collections of information. They concentrate on a particular topic or theme—and good digital libraries will articulate the principles governing what is included. They are organized to make information accessible in particular, well-defined, ways—and good ones will include a description of how the information is organized [3].

The Greenstone digital library software is a comprehensive system for building and distributing digital library collections [4]. It provides a new way of organizing information and publishing it on the Internet (or on CD-ROM). It is widely used in a very large number of countries: see www.greenstone.org for many example sites.

This paper focuses on the Librarian Interface, a subsystem of Greenstone intended to help librarians, and other people who build information collections, construct and organize digital information collections very quickly. Only a few minutes of the user's time are needed to set up a collection based on a standard design and initiate the building process, assuming that documents and metadata are already available.

More than a few minutes may be required to actually build the full-text indexes and browsing structures that comprise the collection, and compress the text. Some collections contain Gbytes of text; millions of documents. Additionally, even larger volumes of information may be associated with a collection—typically audio, image, and video, with textual metadata. Once initiated, the mechanical process of collection-building may take from a few moments for a tiny collection to several hours for a multi-Gbyte one that involves many full-text indexes.

Naturally, customized collections that have their own idiosyncratic requirements—as most substantial collections do—take longer to set up, and the design and debugging process can take days, weeks if iterative usability testing is involved. The Greenstone designers wholeheartedly endorse Alan Kay's maxim that "simple things should be simple, complex things should be possible" [1].

The facilities that Greenstone provides, and the user interface through which library readers access them, are highly customizable at many different levels. Even librarians who need to produce new collections in just a few minutes can dictate what document formats (e.g. HTML, Word, PDF, PostScript, PowerPoint, Excel) or image formats (e.g. TIFF, GIF, PNG, JPEG) will be included, what forms of metadata (e.g. MARC records, OAI archives, BibTex or Refer files, CDS/ISIS databases) are available, what searchable indexes will be provided (e.g. full text, perhaps partitioned by language or other features, and selected metadata such as titles or abstracts), and what browsing structures will be constructed (e.g. list of authors, titles, dates, classification hierarchy). Advanced collection-builders can control the presentation of items on the screen, personalizing each and every page that Greenstone serves up. All these facilities can be controlled through the Librarian Interface.

There are many additional features of Greenstone that lie outside the Librarian Interface. Users can translate the interface into different natural languages. If they know HTML they can hook into Greenstone widgets like the full-text search mechanism or browsers from their own Web pages. If they know JavaScript they can incorporate browsing mechanisms such as image maps, and using Perl they can add entirely new browsing facilities, such as stroke-based or Pinyin-based browsing for Chinese. Some new requirements are best met by altering the Greenstone "receptionist" program, written in C++, to add new facilities at runtime.

The Greenstone Librarian Interface is targeted at four different levels of user.

Assistant Librarians gain access to the basic features of the Librarian Interface: adding documents and metadata to existing collections, creating new collections whose structure mirrors existing ones, and rebuilding collections to reflect changes.

Librarians, the regular or default users of the Librarian Interface, perform all the Assistant Librarian tasks above, and can also design new collections—adding, for example, new document types, new full-text indexes, and new metadata browsing features. They typically design a new collection by identifying an existing one that closely matches their needs and adapting its structure as necessary.

Library Systems Specialists can perform all the functions of Librarians, and in addition customize collections in more complex ways, such as those that involve defining and using regular expressions—for example, partitioning collections based on filename or directory structure.

Expert users are those who are experienced with Greenstone and are familiar with running Perl scripts and examining their output. These users can access all features of the Greenstone Librarian Interface.

2. The role and structure of metadata

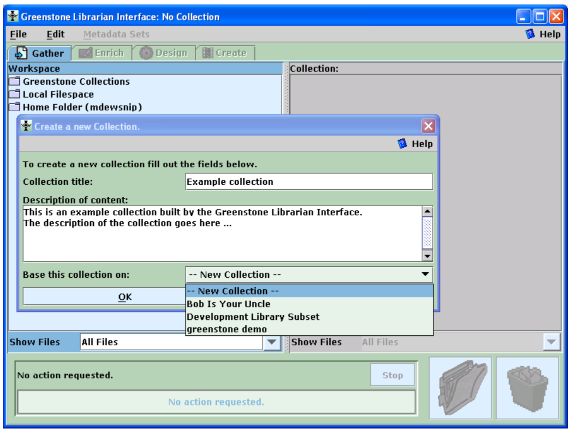
A digital library's organization is reflected in the interface it presents to users. Much of the organization rests on metadata—structured information about the resources (typically documents) the library contains. Metadata is the stuff in the traditional card catalogs of bricks-and-mortar libraries (whether computerized or not). It is "structured" in that it can be meaningfully manipulated without necessarily understanding its content. For example, given a collection of source documents, bibliographic information about each document would be metadata for the collection. The structure is made plain, in terms of which pieces of text represent author names, which are titles, and so on. The notion of "metadata" is not absolute but relative: it is only really meaningful in a context that makes clear what the data itself is [2]. For example, given a collection of bibliographic information, metadata might comprise information about each bibliographic item, such as who compiled it and when.

The use of metadata as the raw material of organization is really the defining characteristic of digital libraries: it is what distinguishes them from other collections of online information. It is metadata that allows new material to be sited within a library and hooked into existing structures in such a way that it immediately enjoys first-class status as a member of the library. Adding new material to ordinary online information collections requires manually linking it in with existing material, but the only manual work needed when adding new items to a digital library is to determine metadata values for each one. If a standard metadata scheme is used, even that may be unnecessary: the information may already be available from another source.

In Greenstone, one or more metadata sets are associated with each collection. There are a few pre-prepared sets, of which Dublin Core is one, and the Librarian Interface allows new sets to be defined—usually by adding a few additional elements to an existing set. One important set is the *extracted* metadata set, which contains information extracted automatically from the documents themselves (e.g. HTML *Title* tags, *meta* tags, or built-in Word author and title metadata). This is always present behind the scenes, though it may be hidden from the user.

The system keeps metadata sets distinct using namespaces. For example, documents can have both a Dublin Core *Title* (*dc.Title*) and an extracted *Title* (*ex.Title*); they do not necessarily have the same value. Behind the scenes, metadata in documents, and metadata sets themselves, are represented in XML.

In order to expedite manual assignment of metadata, the Librarian Interface allows metadata to be associated with document folders as well as with individual documents. This means that users can take advantage of document groupings to add shared metadata in one operation. Within the interface users can organize the document hierarchy by dragging items around and creating new sub-hierarchies, which may expedite joint metadata assignment. Metadata values assigned to a folder remain with that folder and are inherited by all files nested within it. If the user subsequently selects a file and changes



Greenstone Librarian Interface: No Collection

Create a new Collection.

To create a new collection fill out the fields below.

Collection title: Example collection

Description of content:
This is an example collection built by the Greenstone Librarian Interface.
The description of the collection goes here ...

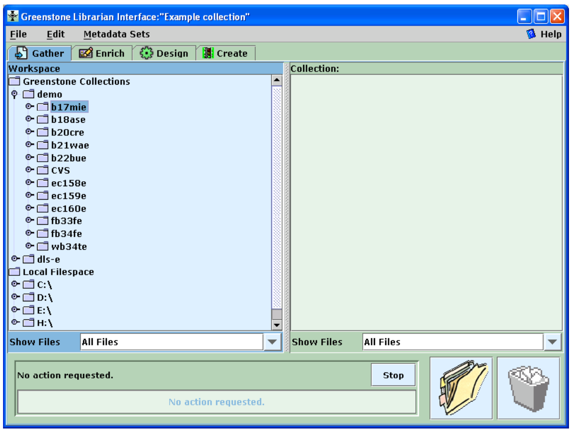
Base this collection on: -- New Collection --

Show Files: All Files

No action requested.

No action requested.

Figure 1. Starting a new collection



Greenstone Librarian Interface:"Example collection"

Show Files: All Files

Show Files: All Files

No action requested.

No action requested.

Figure 2. Exploring the local file space

an inherited metadata value, a warning appears that doing so will override the inherited value.

Metadata in Greenstone can be a simple text string (e.g. title, author, publisher). Or it can be hierarchically structured, as with hierarchical classification values, in which case new values can be placed in the classification tree. In addition, it is multivalued: each element can have more than one value. This is used, for example, for multiple authors. The Librarian interface allows existing metadata values to be reused where appropriate, encouraging consistency in metadata assignment by eliminating the need to retype duplicate values.

3. Working with the Librarian Interface

Within the Librarian Interface, users collect sets of documents, import or assign metadata, and build them into a Greenstone collection. It is an interactive platform-independent Java application that runs on the same computer that operates the Greenstone digital library server. It is closely coupled to the server, and tightly integrated with Greenstone's collection design and creation process. It incorporates various open-source packages for such tasks as file browsing, HTML rendering, web mirroring, and efficient table sorting.

The Librarian Interface supports five basic activities, which can be interleaved but are nominally undertaken in this order:

1. Bring documents into a collection—whether to populate a new collection or update an existing one. Metadata files may also be brought in. Users browse the the computer's file space to find documents to include, and drag and drop them into place. Any documents imported from existing collections come with existing metadata attached.

2. Enrich the documents by adding metadata to them manually. Documents can be grouped into folders (it is easy to retain existing folder structures when dragging documents in under step 1), and any metadata assigned to folders is inherited by all documents nested within them.

3. Design the collection by determining its appearance and the access facilities that it will support: full-text search indexes, browsing structures, the format of items on the pages that Greenstone generates, etc. This design facility is not available in Assistant Librarian mode.

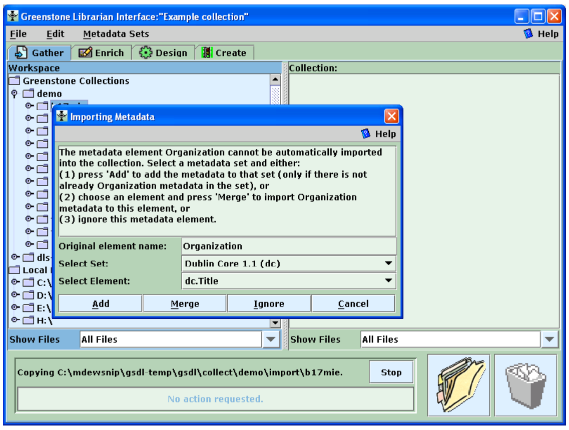
4. Build the collection using Greenstone. This work is done by the computer; users are presented with a progress bar. This is the point where Expert users might examine the output of Perl scripts, which are presented in a scrolling window, to determine if anything is going wrong.

5. Pass the newly-created collection to the Greenstone digital library server for previewing. The collection is automatically installed as one of those in the user's personal digital library, and a web page is opened showing the collection's home page.

To convey the operation of the Librarian Interface we work through a small example. Figures 1 to 12 are screen snapshots at various points during the interaction. This example uses documents in the Humanity Development Library Subset collection, which is distributed with Greenstone. For expository purposes, the walkthrough takes the form of a single pass through the steps listed above. A more realistic pattern of use, however, is for users to switch back and forth through the various stages as the task proceeds.

Assembling source material

To commence, users either open an existing collection or begin a new one. Novice users ("Assistant Librarians") generally work with existing collections, adding documents and/or



Greenstone Collections
demo
dls
Local
C:\
D:\
E:\
H:\

The metadata element Organization cannot be automatically imported into the collection. Select a metadata set and either:
(1) press 'Add' to add the metadata to that set (only if there is not already Organization metadata in the set), or
(2) choose an element and press 'Merge' to import Organization metadata to this element, or
(3) ignore this metadata element.

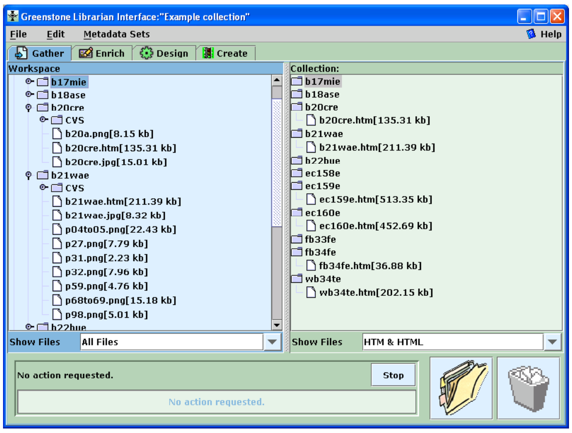
Original element name: Organization
Select Set: Dublin Core 1.1 (dc)
Select Element: dc.Title

Show Files: All Files
Show Files: All Files

Copying C:\mdewsnip\gsdl-temp\gsdl\collect\demo\import\b17mie.

No action requested.

Figure 3. Importing existing metadata



b17mie
b18ase
b20cre
CVS
b20a.png[8.15 kb]
b20cre.htm[135.31 kb]
b20cre.jpg[15.01 kb]
b21wae
CVS
b21wae.htm[211.39 kb]
b21wae.jpg[8.32 kb]
p04to05.png[22.43 kb]
p27.png[7.79 kb]
p31.png[2.23 kb]
p32.png[7.96 kb]
p59.png[4.76 kb]
p68to69.png[15.18 kb]
p98.png[5.01 kb]
b22hue

b17mie
b18ase
b20cre
b20cre.htm[135.31 kb]
b21wae
b21wae.htm[211.39 kb]
b22hue
ec158e
ec159e
ec159e.htm[513.35 kb]
ec160e
ec160e.htm[452.69 kb]
fb33fe
fb34fe
fb34fe.htm[36.88 kb]
wb34te
wb34te.htm[202.15 kb]

Show Files: All Files
Show Files: HTM & HTML

No action requested.

No action requested.

Figure 4. Filtering the file trees

metadata. However, they can begin a new collection by copying the structure of an existing one, effectively creating an empty shell exactly like an existing collection, and adding documents and metadata to it. Collection design involves more advanced skills.

Figure 1 shows the process of starting a new collection. Having selected *New* from the file menu, the user fills out general information about the collection—its name and a brief description of the content—in the popup window shown. The name is a short phrase used to identify the collection throughout the digital library: existing collections have names like *Food and Nutrition Library*, *World Environmental Library*, and so on. The description is a statement about the principles that govern what is included in the collection, and appears under the heading *About this collection* on the collection's home page.

At this point, the user decides whether to base the new collection on an existing one, selecting from the menu pulled down in Figure 1, or design a new one. In this example we will design a new collection, and now it is necessary to select one or more metadata sets for it. We choose Dublin Core from a popup menu (not shown in the Figure).

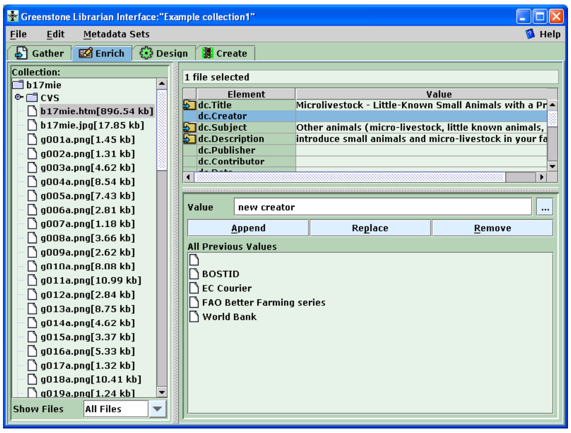
At this point the remaining parts of the interface, which were grayed out before, become active. The *Gather* panel, selected by the eponymous tab near the top of Figures 1–4, is active initially. It allows the user to explore the local file space and existing collections, gathering selected documents into the new collection. The panel is divided into two sections, the left for browsing existing file structures and the right for organizing the documents in the collection.

At this stage, users navigate the existing file structure hierarchy in the usual way. They can select files or directories, drag them into the collection on the right, and drop them there. Entire file hierarchies can be dragged and dropped, and files can be multiply selected in the usual way. Users can navigate around the collection on the right too, adjusting the file hierarchy by dragging items around, creating new sub-hierarchies, and deleting files if necessary.

Another source of documents is the web itself. The Librarian Interface has a *Mirror* panel, which can optionally be activated. Through it, users interact with a mini web browser and select certain pages, or sites, for mirroring. There are many options: mirroring depth, automatically download embedded objects like images, only mirror from the same site, etc. The actual download operation is accomplished by a widely-used open-source mirroring utility. The resulting files appear as another top-level folder on the left-hand side of the *Gather* panel.

In Figure 2 the interactive file tree display is being used to explore the local file system. At this stage the collection on the right is empty; the user populates it by dragging files of interest from the left-hand panel and dropping them into the right-hand one. Such files are copied rather than moved, so as not to disturb the original file system.

Existing collections are represented by a subdirectory on the left called "Greenstone Collections," which can be opened and explored like any other directory. However, the documents therein differ from ordinary files because they already have metadata attached, which the Librarian Interface preserves when it moves them into the new collection. Conflicts may arise because their metadata may have been assigned according to a different metadata set from the one attached to the new collection, and the Librarian Interface helps the user resolve these.



Collection:
- b17mie
  - CVS
  - b17mie.htm[896.54 kb]
  - b17mie.jpg[17.85 kb]
  - g001a.png[1.45 kb]
  - g002a.png[1.31 kb]
  - g003a.png[4.62 kb]
  - g004a.png[8.54 kb]
  - g005a.png[7.43 kb]
  - g006a.png[2.81 kb]
  - g007a.png[1.18 kb]
  - g008a.png[3.66 kb]
  - g009a.png[2.62 kb]
  - g010a.png[8.08 kb]
  - g011a.png[10.99 kb]
  - g012a.png[2.84 kb]
  - g013a.png[8.75 kb]
  - g014a.png[4.62 kb]
  - g015a.png[3.37 kb]
  - g016a.png[5.33 kb]
  - g017a.png[1.32 kb]
  - g018a.png[10.41 kb]
  - g019a.png[1.24 kb]

Show Files: All Files

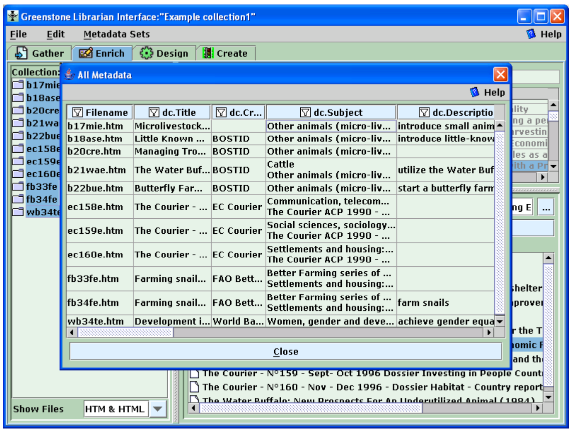
1 file selected

Element	Value
dc.Title	Microlivestock - Little-Known Small Animals with a Pr
dc.Creator	
dc.Subject	Other animals (micro-livestock, little known animals,
dc.Description	introduce small animals and micro-livestock in your fa
dc.Publisher	
dc.Contributor	

Value: new creator

All Previous Values
- BOSTID
- EC Courier
- FAO Better Farming series
- World Bank

Figure 5. Assigning metadata using the *Enrich* view


Greenstone Librarian Interface:"Example collection1"
File
Edit
Metadata Sets
Help
Gather
Enrich
Design
Create
All Metadata
Help
Close
The Courier - N°159 - Sept- Oct 1996 Dossier Investing in People Count
The Courier - N°160 - Nov - Dec 1996 - Dossier Habitat - Country report

Collection:
- b17mie
- b18ase
- b20cre
- b21wa
- b22bue
- ec158e
- ec159e
- ec160e
- fb33fe
- fb34fe
- wb34te

Show Files: HTM & HTML

Filename	dc.Title	dc.Cr...	dc.Subject	dc.Descriptio
b17mie.htm	Microlivestock...		Other animals (micro-liv...	introduce small anim
b18ase.htm	Little Known ...	BOSTID	Other animals (micro-liv...	introduce little-know
b20cre.htm	Managing Tro...	BOSTID	Other animals (micro-liv...	
b21wae.htm	The Water Buf...	BOSTID	Cattle Other animals (micro-liv...	utilize the Water Buf
b22bue.htm	Butterfly Far...	BOSTID	Other animals (micro-liv...	start a butterfly farm
ec158e.htm	The Courier - ...	EC Courier	Communication, telecom... The Courier ACP 1990 - ...	
ec159e.htm	The Courier - ...	EC Courier	Social sciences, sociology... The Courier ACP 1990 - ...	
ec160e.htm	The Courier - ...	EC Courier	Settlements and housing:... The Courier ACP 1990 - ...	
fb33fe.htm	Farming snail...	FAO Bett...	Better Farming series of ... Settlements and housing:...	
fb34fe.htm	Farming snail...	FAO Bett...	Better Farming series of ... Settlements and housing:...	farm snails
wb34te.htm	Development i...	World Ba...	Women, gender and deve...	achieve gender equa

Figure 6. Viewing all metadata assigned to selected files

In Figure 3 the user has selected some documents from an existing collection and dragged them into the new one. The popup window explains that the metadata element *Organization* cannot be automatically imported, and asks the user to either select a metadata set and press *Add* to add the new element to that set, or choose a metadata set and element, and press *Merge* to effectively rename the old metadata element to the new one by merging the two. Metadata in subsequent documents will be imported in the same way automatically.

When large file sets are selected, dragged, and dropped into the collection, the copying operation may take some time—particularly if metadata must be converted too. The Librarian Interface indicates progress by showing which file is being copied and what percentage of files has been processed. The implementation is multi-threaded: users can proceed to another stage while copying is still in progress.

Special mechanisms are needed for dealing with large file sets. For example, the user can filter the file tree to show only certain files, using a dropdown menu of file types displayed underneath the trees. In the right-hand panel of Figure 4, only HTM and HTML files are being shown (and only these files will be copied by drag and drop). In fact, the left-hand panel is showing the same part of the file space without filtering, and you can see the additional *.png* and *.jpg* files that are present there.

Adding metadata to documents

The next phase of collection-building is to enrich the documents by adding metadata. This is where Librarian users spend most of their time: enhancing the collection by selecting individual documents and manually adding metadata. We have already discussed two features of the Librarian Interface that help with this task:

- Documents that are copied during the first step come with any applicable metadata attached.
- Whenever possible, metadata is extracted automatically from documents.

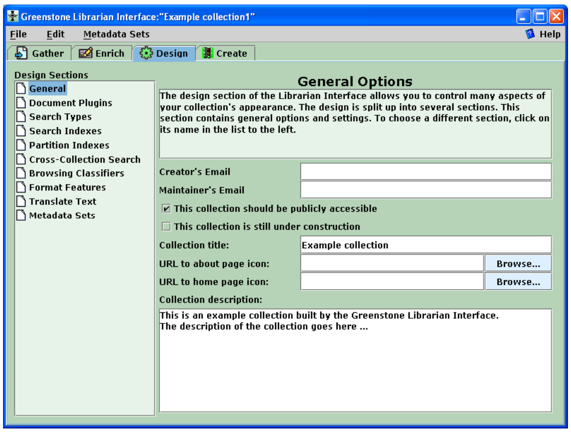
The Librarian implements two further features that expedite manual metadata assignment:

- Metadata values can be assigned to several documents at once, either by virtue of them being in a folder, or through multiple selection.
- Previously-assigned metadata values are kept around and made easy to reuse.

The *Enrich* tab brings up a panel of information (Figure 5). On the left is the document tree representing the collection, while on the right metadata can be added to individual documents, or groups of documents. Users often want to see the document they are assigning metadata to, and if they double-click a document in the pane on the left it is opened by the appropriate viewing program.

In Figure 5 the user has selected a document and typed "new creator" as its *dc.Creator* metadata. The buttons for appending, replacing and removing metadata become active depending on what selections have been made. Values previously assigned to *Creator* metadata are shown in the pane labeled "All previous values."

Users can at any time view all the metadata that has been assigned to the collection. The popup window in Figure 6 shows the metadata in spreadsheet form. For large collections it is useful to be able to view the metadata associated with certain document types only, and if the user has specified a file filter as mentioned above, only the selected documents are shown in the metadata display.



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

Maintainer's Email

☑ This collection should be publicly accessible

□ This collection is still under construction

Collection title: Example collection

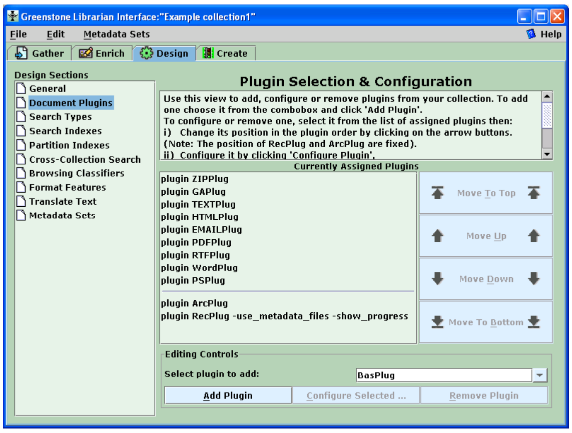
URL to about page icon:

URL to home page icon:

Collection description:

This is an example collection built by the Greenstone Librarian Interface.
The description of the collection goes here ...

Figure 7. Designing the collection



Plugin Selection & Configuration

Use this view to add, configure or remove plugins from your collection. To add one choose it from the combobox and click 'Add Plugin'.
To configure or remove one, select it from the list of assigned plugins then:
i) Change its position in the plugin order by clicking on the arrow buttons.
(Note: The position of RecPlug and ArcPlug are fixed).
ii) Configure it by clicking 'Configure Plugin',

Currently Assigned Plugins

plugin ZIPPlug
plugin GAPlug
plugin TEXTPlug
plugin HTMLPlug
plugin EMAILPlug
plugin PDFPlug
plugin RTFPlug
plugin WordPlug
plugin PSPlug

plugin ArcPlug
plugin RecPlug -use_metadata_files -show_progress

Editing Controls

Select plugin to add: BasPlug

Figure 8. Specifying which plug-ins to use

Designing a collection

All except “Assistant Librarian” users of the Librarian interface have the ability to design new collections, which involves specifying the structure, organization, and presentation of the collection being created. The result of this process is recorded in a “collection configuration file,” which is Greenstone’s way of expressing the facilities that a collection requires.

Collection design has many aspects. Users might review and edit collection-level metadata such as title, author and public availability of the collection. They might define what full-text indexes are to be built. They might create sub-collections and have indexes built for them. They might add or remove support for predefined interface languages. They will need to decide what document formats will be included. In Greenstone, document types are processed by modules called “plug-ins,” and each plug-in may need to be configured by specifying appropriate arguments. The collection designer will need to specify what browsing structures will be constructed—in Greenstone, these are built by modules called “classifiers,” which also have various arguments. It will also be necessary to specify the formatting of various items in the collection’s user interface. Sensible, generally-applicable defaults are supplied for all these features.

Users accomplish the design with the *Design* panel illustrated in Figures 7–10. It has a series of separate interaction screens, each dealing with one aspect of the collection design. In effect, it serves as a graphical equivalent to the process of editing the collection configuration file manually.

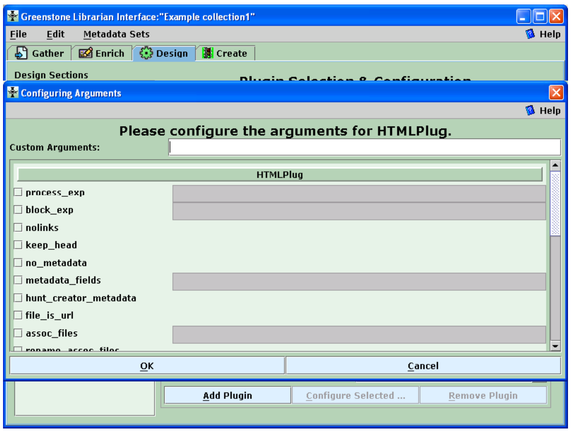
In Figure 7 the user has clicked the *Design* tab and is reviewing general information about the collection, which was entered when the new collection was created. On the left are listed the various facets that the user can configure: Document Plug-ins, Search Types, Search Indexes, Partition Indexes, Cross-Collection Search, Browsing Classifiers, Format Features, Translate Text, and Metadata Sets. For example, clicking the *Document Plug-in* button brings up the screen shown in Figure 8, which allows you to add, remove or configure plug-ins, and change the order in which the plug-ins are applied to documents.

Both plug-ins and classifiers have many different arguments or “options” that the user can supply. The dialog box in Figure 9 shows the user specifying arguments to a plug-in. The grayed-out fields become active when the user adds the option by clicking the preceding tick-box. Because Greenstone is a continually growing open-source system, the number of options tends to grow as developers add new facilities. To help cope with this, Greenstone has a “plug-in information” utility program that lists the options available for each plug-in, and the Librarian Interface automatically invokes this to determine what options to show. This allows the interactive user interface to automatically keep pace with developments in the software.

In Figure 10 the user is adding a new full-text-search index to the collection, in this case based on both *dc.Creator* and *dc.Description* metadata. In Figure 11 she is adding a “cross-collection search” capability so that other collections are searched whenever this one is.

Building the collection

The next step is to construct the collection formed by the documents and assigned metadata. The brunt of this work is borne by the Greenstone code itself. The user observes the building process though a window that shows not only the text output generated by Greenstone’s importing and index-

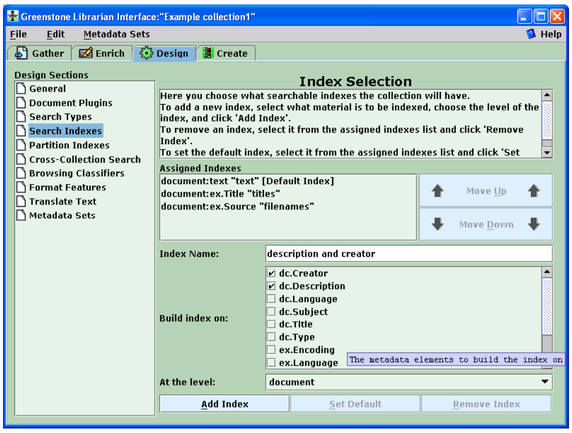


Please configure the arguments for HTMLPlug.

Custom Arguments:

□ process_exp
□ block_exp
□ nolinks
□ keep_head
□ no_metadata
□ metadata_fields
□ hunt_creator_metadata
□ file_is_url
□ assoc_files
□ rename_assoc_files

Figure 9. Configuring the arguments to a plug-in



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

document:text "text" [Default Index]
document:ex.Title "titles"
document:ex.Source "filenames"

Index Name: description and creator

Build index on:

☑ dc.Creator
☑ dc.Description
□ dc.Language
□ dc.Subject
□ dc.Title
□ dc.Type
□ ex.Encoding
□ ex.Language

The metadata elements to build the index on

At the level: document

Figure 10. Adding a full-text-search index

building scripts, but also progress bars that indicate the overall degree of completion of each script.

Figure 12 shows the *Create* view through which users control collection building. On the left are groups of options that can be applied during the creation process: Import, Build, and Message Log. The user selects values for the options if necessary, and clicks *Build Collection*. Greenstone programs continually print text that indicates progress, and the Librarian Interface shows these, along with a progress bar.

4. Beyond the Librarian Interface

Most of the customization that non-programming users perform in Greenstone takes place in the collection configuration file, which the Librarian Interface creates. It depends crucially on the availability of metadata, and the structures defined are only produced if appropriate metadata is provided. However, Greenstone has more advanced customization features. Our philosophy is to target the most common features and make them accessible to librarian-level users without particular training in computer science. But users who are prepared to dig deeper can accomplish more.

Macros

Greenstone incorporates a macro facility, expressed as an extension of HTML. It includes the ability to define macros and perform textual substitution. Currently, for example, there are interfaces in nearly thirty languages, from Arabic to Turkish, Bosnian to Ukrainian, Chinese to Vietnamese. To accommodate these variants, and to allow the language interfaces to be updated when new facilities are added, all web pages are passed through a macro expansion phrase before being displayed. This means that a new language can be added by providing a new set of language-specific macros, a task that has been performed many times by people with no expertise in Greenstone.

The digital library functionality is hooked into the user interface through "dynamic macros" whose expansions are determined by the system (in terms of other macros). For example, the search widget is generated by a dynamic macro. Users can incorporate this widget into their own Web pages, provided they go through the macro expansion phase. A total of about twenty dynamic macros provides access to Greenstone's full user interface functionality.

Users who work with Greenstone can capitalize on the macro system to radically alter the style of the pages generated, and some have produced attractive new designs for the Greenstone user interface [5].

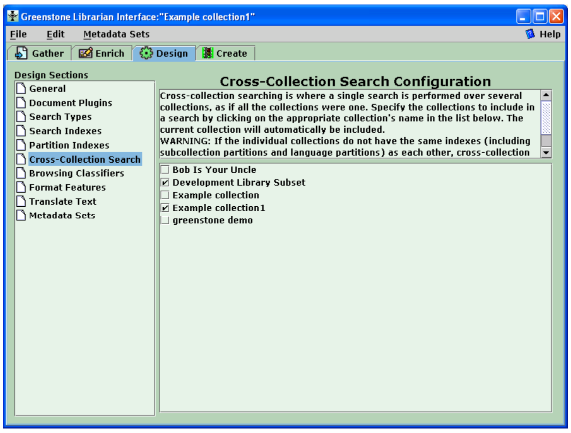
Altering the run-time system

The part of Greenstone that serves collections to users is called the "receptionist," and one sometimes has to resort to changing this program to achieve a desired level of customization. This rarely involves large changes, but creates software management difficulties in dealing with different parallel versions.

Our system development strategy is to accept the inevitability of occasionally having to build a special-purpose collection-dependent receptionist to achieve some desired features, and to note what is required with a view to incorporating it as an option within the standard Greenstone code.

5. Conclusions

A general-purpose digital library system like Greenstone must cater for a wide range of users. We have targeted the Librarian Interface at four different user levels: *assistant librarians*, who can add to

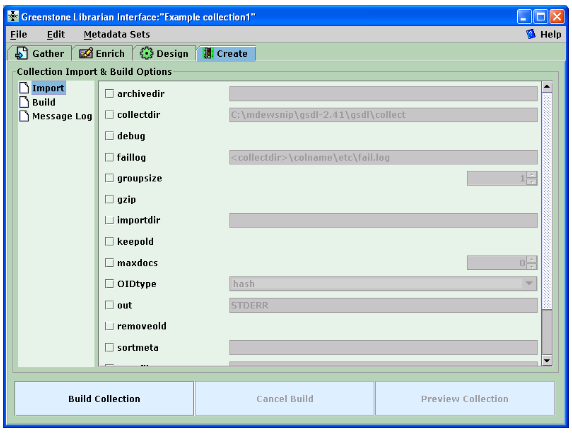


Cross-Collection Search Configuration

Cross-collection searching is where a single search is performed over several collections, as if all the collections were one. Specify the collections to include in a search by clicking on the appropriate collection's name in the list below. The current collection will automatically be included.
WARNING: If the individual collections do not have the same indexes (including subcollection partitions and language partitions) as each other, cross-collection

- ☐ Bob Is Your Uncle
- ☑ Development Library Subset
- ☐ Example collection
- ☑ Example collection1
- ☐ greenstone demo

Figure 11. Adding a cross-collection search facility



- ☐ archivedir
- ☐ collectdir C:\mdewsnip\gsdl-2.41\gsdl\collect
- ☐ debug
- ☐ faillog <collectdir>\colname\etc\fail.log
- ☐ groupsize 1
- ☐ gzip
- ☐ importdir
- ☐ keepold
- ☐ maxdocs 0
- ☐ OIDtype hash
- ☐ out STDERR
- ☐ removeold
- ☐ sortmeta

Figure 12. Getting ready to create the new collection

existing collections and create new ones with the same structure; *librarians*, who can, in addition, design new collections; *library systems specialists*, who can customize collections in more complex ways; and *expert users*, who can deal with every aspect of the system.

A digital library may be customized in a wide variety of different ways, and virtually every collection has its own idiosyncratic requirements. Although a basic Greenstone collection of new material with a standard look and feel can be set up in just a few minutes, most users want more personalization. As the number of collections grows and the variety of styles increases, it becomes more likely that some existing collection will match new requirements.

It is difficult to produce good, up to date, documentation for a richly functional software system. In fact, from a user's point of view the chief bottleneck in customization is documentation, not the facilities that are provided. Consequently collection builders need access to advice and assistance from others, in order to continue to learn how to tailor the software to meet ever-changing requirements. There is a lively Greenstone email discussion group; participants hail from over 40 countries.

Digital libraries have the advantage over other interactive systems that their user interfaces are universally based on metadata. Metadata is the glue that allows new documents to be added and immediately become first-class citizens. It is also the key to user interface customization, and Greenstone incorporates a range of mechanisms at different levels to capitalize on this.

Acknowledgements

The Greenstone Digital Library software has grown out of the stimulating research environment of the New Zealand Digital Library project, and I gratefully acknowledge the profound influence of all project members. In particular, David Bainbridge and John Thompson have been the prime movers behind the Librarian Interface described in this paper.

References

[1] Davidson, C. (1993) "The man who made computers personal." New Scientist, No. 1978, pp. 32–35; June.

[2] Lagoze, C. and Payette, S. (2000) "Metadata: Principles, practices and challenges." In *Moving theory into practice: digital imaging for libraries and archives*, edited by A.R. Kenney and O.Y. Rieger. Research Libraries Group, Mountain View, CA, pp. 84–100.

[3] Lesk, M. (1997) Practical digital libraries: Books, bytes, and bucks. Morgan Kaufmann, San Francisco.

[4] Witten, I.H. and Bainbridge, D. (2003) *How to build a digital library*. Morgan Kaufmann, San Francisco, CA.

[5] Zhang, A. (2003) "Customizing the Greenstone User Interface." Washington Research Library Consortium; August; http://www.wrlc.org/dcpc/UserInterface/interface.htm.