Information Management Resource Kit

Module on Digitization and Digital Libraries

UNIT 6. EXAMPLE OF DIGITAL LIBRARY SOFTWARE: GREENSTONE

LESSON 3. USING GLI: DESIGNING, CREATING AND PREVIEWING A COLLECTION

NOTE

Please note that this is a reference guide covering the main steps for using GLI as described in the corresponding lesson. It provides an excellent means for additional practice and for printing after you have completed the course.

We recommend that you first take the lesson using the courseware environment, as it contains interactive features such as exercises with feedback, pop-ups, animations etc.



SCOPE NOTES

These practical exercises will help you build an operational digital library collection from a set of sample source documents. The collection will be built using the Greenstone Librarian Interface (GLI).

Building a new collection using GLI takes five steps:

- 1) Gathering
 - Copying the documents from the computers' file space, including existing collection/s, into the new collection. Any existing metadata remains attached to these documents. Documents may also be gathered from the web through a built-in mirroring facility
- 2) Enriching
 - Enrich the documents by adding further metadata to individual documents or groups of documents
- 3) Designing
 - Design the collection by establishing its appearance and the access facilities it will support
- 4) Creating
 - Build the collection
- 5) Previewing
 - Preview the newly created collection, which will be included in your Greenstone home page as one of the collections.

By using this document you will apply steps 3) to 5): Designing, Creating and Previewing.

The interface design mainly refers to defining collection specifications in terms of searching, browsing, and results display requirements. For the sample collection that you will be building, you will be guided in providing the following access features:

- Simple Searching by
 - o full text, author and title
- Advanced, form-based searching by

 full text, author and title
- Browsing by
 - o author and title

You have full control over the way the records are displayed after a search or while a collection is being browsed. The format you will be guided to use to display the search records is:

- Search results display: We shall use the default format which is: Link to HTML icon (for the HTML version of the document) followed by a link to the document type image icon (for the native format of the document), document title, and source file name.
- Title browse display: Default format described above.
- Author browse display: Grouping by individual authors using the bookshelf icon. Upon clicking the bookshelf icon, show document details as per the default format.

The General section

The Design step involves a series of separate interaction screens, each dealing with one aspect of the collection design. The design step is split up into several sections as shown in the figure below.

The first section contains general options and settings. In this section, values provided during the collection definition step (GLI step 1) can be modified if required.

Document Plugins your collection's appearance	General Options prarian Interface allows you to control many	🔞 Help
Design Sections General Document Plugins The design section of the Lil your collection's appearance	prarian Interface allows you to control many	
General The design section of the Lil Document Plugins your collection's appearance	prarian Interface allows you to control many	
Search Types Search Indexes Partition Indexes Cross-Collection Search	e. The design is split up into several sections ions and settings. To choose a different secti t.	. This
Browsing Classifiers Creator's email	franc@ncsi	
□ Format reatures Maintainer's email □ Translate Text □ Metadata Sets	franc@ncsi publicly accessible	
This collection is still under	er construction	
Collection title:	WordPDF	
Collection folder:	wordpdf	
URL to about page icon:	_httpprefix_/collect/wordpdf/images/	Browse
URL to home page icon:		Browse
Collection description: This collection demonstrate:	s Greenstone's ability to handle documents ir	n different fi

Collection Design (general section)

You can brand your collection using a suitable image. For example, you can use the wordpdf-e.gif (you will find it in the 'image' folder of the 'wordpdf-e' collection) and copy it to the 'image' folder within the 'wordpdf' collection folder created by Greenstone during collection definition (typically, c:\program files\gsdl\collect\wordpdf\images\ if you had installed Greenstone in the C: drive).

After this is done, go back to the GLI/General Options screen, use the 'Browse...' button associated with 'URL to about page icon:' and select this image file. GLI automatically generates the URL.

The Document Plugins section

The next section is 'Document Plugins'. Activate this section by clicking it.

le <u>E</u> dit <u>M</u> etadata Sets	,		🔞 Help
👌 Gather 🛛 🜃 Enrich 🛛 🔃	Design 🚺 Create		
Design Sections General Document Plugins Search Types Search Indexes Partition Indexes Cross-Collection Search Browsing Classifiers Format Features Translate Text Metadata Sets	Use this view to add, cor one choose it from the co To configure or remove o i) Change its position in (Note: The position of Re ii) Configure it by clickin plugin ZIPPlug	n Selection & Config offigure or remove plugins from ombobox and click 'Add Plugin one, select it from the list of as the plugin order by clicking o ccPlug and ArcPlug are fixed). og 'Configure Plugin', Currently Assigned Plugins	your collection. To add .signed plugins then: n the arrow buttons.
	plugin GAPlug plugin TEXTPlug plugin HTMLPlug plugin EMAILPlug plugin PDFPlug plugin RTFPlug plugin WordPlug plugin PSPlug		 Move To Top Move Up Move Down
		adata_files -show_progress	♥ Move To <u>B</u> ottom ♥
	Editing Controls		
	Select plugin to add:	BasPlug	•
	Add Plugin	Configure Selected	Remove Plugin

Collection Design (document plugins)

In this section you can add, configure or remove plugins to be used for your collection. Plugins extract text from source documents and convert these to the Greenstone Archive (GA) format. They also extract some of the metadata fields automatically.

For our source collection, which consists of PDF, Word, RTF, and Postscript documents, we do not require ZIPPlug, TEXTPlug, HTMLPlug, and EMAILPlug plugins. These can be deleted by selecting them one at a time and clicking on '<u>R</u>emove Plugin'. By removing unwanted plugin/s, the collection building process will be faster.

The source documents will be processed by the plugins in the same order as listed in the 'Document Plugin' section.

Greenstone Librarian Interface: W	ordPDF (wordpdf)		
<u>F</u> ile <u>E</u> dit <u>M</u> etadata Sets			🚺 Hel
長 Gather 🛛 🐼 Enrich 🤃	Design 🔣 Create		
Design Sections General Document Plugins Search Types Search Indexes Partition Indexes Cross-Collection Search Browsing Classifiers Format Features Translate Text Metadata Sets	Use this view to add, con one choose it from the co To configure or remove o i) Change its position in (Note: The position of Re ii) Configure it by clickin plugin PDFPlug plugin RTFPlug plugin WordPlug plugin GAPlug plugin GAPlug	n Selection & Config figure or remove plugins from mbobox and click 'Add Plugin' ne, select it from the list of as the plugin order by clicking o cPlug and ArcPlug are fixed). g 'Configure Plugin', Currently Assigned Plugins	your collection. To add signed plugins then: n the arrow buttons.
	Editing Controls		
	Select plugin to add:	BasPlug	-
	<u>A</u> dd Plugin	<u>C</u> onfigure Selected	<u>R</u> emove Plugin

The following figure shows the revised 'Document Plugin' section:

The Search Types section

The search types specify what kind of search interface should be provided for the collection:

- form for field level searching; and/or
- plain for regular searching.

The ordering is important, as the first one will be used for the initial search page by default.

<mark>∀Greenstone Librarian Interface: ₩</mark> ile Edit Metadata Sets	/ordPDF (wordpdf)	
	Design 🛛 🔀 Create	
Design Sections General Document Plugins Search Types Search Indexes Partition Indexes Cross-Collection Search Browsing Classifiers Format Features Translate Text Metadata Sets	Search Type Selec Defining the search type is an advanced fee checking the 'Enable Advanced Searches' b selecting and changing the order of search search type from the combobox and 'Add S search type. Select an existing search type 'Move Down' to change the order, or 'Remo Demonstration Searches Currently Assigned Search Types form	ature only available when enabled (by ox). Once enabled further controls for types become available. Select a earch Type', if enabled, to add the from the list, and use 'Move Up' and
	Search Types: form <u>A</u> dd Search Type	 <u>R</u> emove Search Type

Collection Design (search types)

• Check the 'Enable Advance Searches' box.

The other items in the screen get activated. For our collection let us use both the plain and form searches. To add a search type:

• select it from the 'Search types' list and click 'Add Search Type'.

For our collection let us have the plain search as the default search type. To achieve this:

• click on 'plain' and then click on 'Move Up'.

GLI STEP 3: DESIGNING THE COLLECTION

🛉 Greenstone Librarian Interface: ¥	VordPDF (wordpdf)	
<u>F</u> ile <u>E</u> dit <u>M</u> etadata Sets		🐞 He
	Design Image: Create Defining the search type is an advanced fe checking the 'Enable Advanced Searches' I selecting and changing the order of search search type from the combobox and 'Add S search type. Select an existing search type 'Move Down' to change the order, or 'Remeted Move Down' to change the order, or 'Remeted Search Search Types ✓ Enable Advanced Searches Currently Assigned Search Types plain form Search Types: plain Add Search Type	eature only available when enabled (by box). Once enabled further controls for h types become available. Select a Search Type', if enabled, to add the e from the list, and use 'Move Up' and
	Collection Design (search typ	pes)

The Search Indexes section

The next step in the Design Section is 'Search Indexes'. Indexes specify what parts of the collection are searchable. The following figure shows a screen shot of this section.

🔊 Gather 🏾 📝 Enrich 🛛 🔅	Design 🔀 Create	
Constant Plugins General Document Plugins Search Types Search Indexes Partition Indexes Cross-Collection Search Browsing Classifiers Format Features Translate Text	Index And Le Here you choose what searchable indexes To add a new index, enter a unique name is to be indexed, and click 'Add Index'. If you wish to add all of the available sout them, then click 'Add All' (the name for ea the source). Note that this is subtly differ Manage Indexes Currently Assigned Indexes text "text" ex.Title "Title"	for the index, select material/metadata rces so as to have indexes built on ach index will default to be the name of
_) Metadata Sets	ex.Source "Source"	Move Down
	Index Source: al A unique name	for the new index
	<u>A</u> dd Index	Add A <u>l</u> l
	Repla <u>c</u> e Index	Remove Index

Collection Design (search indexes)

By default Greenstone builds indexes and supports searching on automatically extracted metadata (title and source file names) and words extracted from document text. If you wish to support searching using the metadata set elements (in our case Dublin Core), you need to provide the required indexing and search specifications here. For the WordPDF collection that we are building, Greenstone has created three indexes by itself. These indexes are displayed in the 'Currently Assigned Indexes' box. A brief description of each of these three indexes follows.

The first index is text and is called 'text'. It is a full text index. This implies is that this index has been built using all the words from the documents of the WordPDF collection.

The second index is ex.Title and is named as 'Title". This index will be built by extracting the titles of the source documents. The titles are automatically extracted by the plugins, which handle the respective file types. The titles thus extracted are not reliable (as Greenstone may have difficulty in correctly extracting this information from some documents). The prefix, 'ex' in ex.Title indicates that the titles are automatically extracted by Greenstone from the source documents (remember the 12 documents that we have added?). \

The third index is ex.Source named as 'Source'. This is an index of source file names.

We will delete the ex.title and ex.Source indexes. As mentioned above, ex.Title is not reliable and ex.Source (source file names) may not be of much use as an access point, for our collection. To delete these two indexes, select them one at a time and click 'Remove Index'. Once you have removed these two indexes only the 'text' index will remain.

Let us continue with our collection building process by adding two more indexes to our collection. The two indexes that we will be adding are dc.Title and dc.Creator.

Adding Title (dc.Title) index:

Provide an Index name. Let us name the dc.Title index 'Document Title'.

From the Index Source box, select dc.Title. Click on <u>A</u>dd Index to add the dc.Title index to the Currently Assigned Indexes list.

Adding Creator (dc.Creator) index:

Provide an Index name. Let us name the dc.Creator index as 'Authors'. From the Index Source box, select dc.Creator. Click on <u>A</u>dd Index to add the dc.Creator index to the Currently Assigned Indexes list. The 'search indexes' screen will appear as shown in this figure:

File Edit Metadata Sets	Design 🕌 Create]		🔞 Helj
Design Sections General Document Plugins Search Types Search Indexes Partition Indexes Cross-Collection Search Browsing Classifiers Format Features Translate Text Metadata Sets	To add a new index, is to be indexed, and If you wish to add a them, then click 'Add	at searchable indexes enter a unique name I click 'Add Index'. Il of the available sou d All' (the name for ea at this is subtly differ Manage Levels Indexes	evel Selection s the collection will have. for the index, select material/met rces so as to have indexes built on ach index will default to be the name ent from adding the 'AllFields' sour Move Up Move Down	ne of
	Index Name: Index Source:	Authors dc.Creator		
	Add	Index ce Index	Add A <u>l</u> l Remove Index	
		esign (search ind		

GLI STEP 3: DESIGNING THE COLLECTION

The Partition Indexes section

The next Design Section is 'Partition Indexes'. A collection index can be partitioned into sub collections based on metadata values and languages. This will facilitate controlling the search space. This partition can be achieved by using the 'Partition Indexes' feature. For our collection we will not create any Partition Indexes.

The Cross-Collection Search Configuration section

The next section is the '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. Let us not make any changes here. The default view of this section is shown in this figure:

ile	<u>E</u> dit	<u>M</u> etadata 9	Sets		🔞 Hel
Ð	Gather	🛃 Enrich	😳 Design	📳 Create	
	Search T Search Ir Partition Cross-Co	t Plugins ypes ndexes Indexes llection Sear (Classifiers eatures PText	collecti a searc current WARNI subcolle V Word	ollection sea ons, as if all h by clicking collection w NG: If the in	
					ss-collection search configuration.)

The Browsing Classifiers section

The next section is classifier selection and configuration.

This section explains how to assign 'classifiers'. Classifiers are used to provide browsing functionality to a collection. For this collection, Greenstone has provided us with two classifiers, namely, AZList for ex.Title and ex.Source as shown in the figure below.

Greenstone Librarian Interface: W	'ordPDf (wordpdf)	
<u>F</u> ile <u>E</u> dit <u>M</u> etadata Sets		🔞 Help
🔄 Gather 🛛 🐼 Enrich 🔅	Design 🔀 Create	
Design Sections General Document Plugins Search Types Search Indexes Partition Indexes Cross-Collection Search Browsing Classifiers Format Features Translate Text Metadata Sets	Classifier Selection 4 Use this view to add, configure or remove cla classifier, choose it from the combobox and c To remove a classifier, choose it from the list 'Remove Classifier'. To configure a classifier, choose it from the li 'Configure Classifier'. <u>Currently Assigned</u> Classify AZList -metadata ex.Title classify AZList -metadata ex.Source	assifiers in your collection. To add a Alick 'Add Classifier'. cof assigned classifiers and click st of assigned classifiers, and click
	Editing Controls	
	Select classifier to add:	.ist
	Add Classifier Configure Cla	ssifier Remove Classifier

Collection Design (classier selection and configuration)

What this means is that Greenstone has already provided an alphabetical browsing list for extracted titles and extracted source file names for our collection. Let us remove the default classifiers and add our own classifiers for the Title and Creator metadata fields.

To remove the default classifiers:

- select them one at a time and
- click 'Remove Classifier'.

Let us now add two classifiers, namely, AZList for dc.Title and AzcompactList for dc.Creator, for our collection.

To add a classifier:

- select one from the 'Select classifier to add' drop-down list and
- click on Add Classifier.

Let us now add an AZList classifier for our collection. Select AZList from the 'Select classifier to add' dropdown list and click on Add Classifier. A pop-up screen as shown in figure 17 should come up. Select dc.Title metadata for this classifier and provide a buttonname by checking the 'buttonname checkbox. Let the name be 'Title'.

Similarly add AZCompactList for dc.Creator. Let the button name be 'Creator'. Furthermore, to group all publications by the same author and generate a bookshelf icon, select the 'mingroup' checkbox and increase this number to 1. The screen shot is shown in this figure:

Distom Arguments:				
	<u>ı</u>	AZList		_
metadata	dc.Title			
🗹 buttonname	Title			
removeprefix				
removesuffix				
builddir				
outhandle				
ignore_namespace				
<u>0</u>	ĸ		<u>C</u> ancel	
<u>0</u> 1		n (adding a class	— —	
		n (adding a class	— —	
		n (adding a class	— —	×
Configuring Arguments	Collection Desig		- ifier - Title)	× Melp
Configuring Arguments Please	Collection Desig		— —	
Configuring Arguments Please	Collection Desig		- ifier - Title)	
Configuring Arguments Please	Collection Desig		- ifier - Title)	
Configuring Arguments Please ustom Arguments:	Collection Desig	arguments fo	- ifier - Title)	
Configuring Arguments Please ustom Arguments: metadata	Collection Desig	arguments fo	- ifier - Title)	Help
Configuring Arguments Please ustom Arguments: metadata v buttonname mingroup	Collection Desig configure the dc.Creator	arguments fo	- ifier - Title)	
Configuring Arguments Please ustom Arguments: metadata v buttonname mingroup	Collection Desig configure the dc.Creator	arguments fo	- ifier - Title)	Help
Configuring Arguments Please ustom Arguments: metadata ☑ buttonname ☑ mingroup ☐ removeprefix ☐ removesuffix	Collection Desig configure the dc.Creator	arguments fo	- ifier - Title)	Help
Configuring Arguments Please ustom Arguments: metadata ☑ buttonname ☑ mingroup ☐ removeprefix ☐ removesuffix	Collection Desig configure the dc.Creator	arguments fo	- ifier - Title)	Help
Configuring Arguments Please ustom Arguments: metadata buttonname mingroup removeprefix removesuffix minnesting mincompact	Collection Desig configure the dc.Creator	arguments fo	- ifier - Title)	Help
Configuring Arguments Please ustom Arguments: metadata ✓ buttonname ✓ mingroup □ removeprefix □ removesuffix □ minnesting □ mincompact □ maxcompact	Collection Desig configure the dc.Creator	arguments fo	- ifier - Title)	Help
Configuring Arguments Please ustom Arguments: metadata buttonname mingroup removeprefix removesuffix minnesting	Collection Desig configure the dc.Creator	arguments fo	- ifier - Title)	

er adding the two classifiers	the "browsing classifiers	s' section looks as show	vn in the figure below.
🚰 Greenstone Librarian Interface: ¥	VordPDF (wordpdf)		
<u>F</u> ile <u>E</u> dit <u>M</u> etadata Sets			🔞 Help
🔄 Gather 🛛 🐼 Enrich 🔅	Design 🔠 Create		
Gather Metadata Sets	Classifie Use this view to add, config classifier, choose it from th To remove a classifier, cho 'Remove Classifier'. To configure a classifier, ch 'Configure Classifier'.	e combobox and click 'Add C ose it from the list of assign oose it from the list of assig urrently Assigned Classifier . Title -buttonname Title	your collection. To add a A Classifier'. ed classifiers and click med classifiers, and click
	Editing Controls		 Move <u>D</u>own ★ Move To <u>B</u>ottom ★
	Select classifier to add:	AZCompactSe	ectionList
	Add Classifier	Configure Classifier	Remove Classifier
C	ollection Design (classifi	er selection and design))

The Format Features section

The next section is the 'Format Features'. Format commands control the structure and appearance of search and browse results and the display of selected documents. Format commands are not easy to develop. You should read chapter 2 of the Greenstone developer's guide. We are not making any changes in this section as far as our collection is concerned.

The Translate Text section

The 'Translate Text' section is used to review and assign translations of text fragments in a collection. We are not going to do any translation for our collection.

The Metadata Sets section

The 'Metadata Sets' is the last part of the Design Section. In this read-only design view you can review the Metadata Sets in your collection, identify what Elements they contain, and see how they will appear in the collection configuration file. We are now all set to create and build the 'WordPDF' collection!

GLI STEP 4: CREATE – BUILD THE COLLECTION

After having defined a new collection, collected documents for the collection, annotated it with metadata and designed its appearance, you are now ready to produce the collection using Greenstone. Do not get confused with Greenstone and GLI. GLI is just one of the tools which can be used to build a collection using Greenstone.

Click the 'Create' tab. The create screen as shown in the following figure comes up.

🔊 Gauler 🔯 Enrich i	🔅 Design 🔄 Export 🐻 Create	💋 He
maxdocs		
🗆 sortmeta	dc.Creator	
🗆 əlicləssific	,	
Build Collection	0%	
<u></u> ancel Build		
Preview Collection		

GLI STEP 4: CREATE – BUILD THE COLLECTION

Let us not make any changes to 'maxdocs', 'sortmeta' and 'allclassifications' options. Click on 'Build Collection' to import and then build the 'WordPDF' collection. A progress bar will indicate the status of importing and building. A few messages will be displayed in the text area. The following figure shows a screen shot of the collection building process.

	Man I Car I Microsta	
Sather Sector	😳 Design 🔄 Export 🦉 Create	
□ maxdoc		
🗌 🗆 sortmet	a dc.Creator 🗾	
🗖 allclass	fications	
Build Collection	0%	
Cancel Build		
Preview Collection	-	
Flexiew conection		
***************** Ir Of the 12 'documents	nport Started ************************************	
******************* Ir Of the 12 'documents 12 were recognised t ********************* Compressing text Creating an index bas Creating information	nport Finished ************************************	
*************** Ir Of the 12 'documents 12 were recognised t ************************************	nport Finished ************************************	

Build Collection

The time taken for collection building process depends on the number and types of source documents, and of course, on your PC's configuration. If you have been correctly following the GLI collection building procedure, a small pop-up window as shown in this figure should come up once the collection has successfully been built:

Collection Creation Result
The collection has been built and is ready for previewing.
Do not show this warning again

Collection Creation Resul	t
---------------------------	---

Did it? Congratulations! You have successfully built your fist Greenstone collection using GLI. You can now preview the collection.

GLI STEP 5: PREVIEW THE COLLECTION

6. Example of Digital Library Software: Greenstone - 3. Using GLI: designing, creating and previewing a collection – page 16

The Preview step is used to view the collection that has been built. In practice, previewing often shows deficiencies in the collection design or in the individual metadata values. The user frequently returns to earlier stages to correct these. The preview button and the 'Create' tab become active once the collection has been created, and take you to the collection in your regular Greenstone installation, as shown in the figure:



You can now search and/or browse the WordPDF collection. Two types of searching are supported – plain and form. Click on the 'Search' button to do a plain search.

No collection can ever be an ideal collection. There is always a scope for further refinement or improvement in terms of additional indexes, appearance, or navigation. For the WordPDF collection, we created two additional indexes, namely 'Titles' and 'Creators'. These indexes correspond to the titles and author/s of the source collection you have keyed-in during the 'Enrich' section.

The WordPDF collection will have three indexes as shown in the following figure. The first index, namely, text was automatically created by Greenstone. This index corresponds to full text. Depending on our requirements, we have the option of either retaining the default indexes or not. This can be controlled during the design section (search indexes) stage.

Search page - Microsoft I File Edit View Favorites			
-	🛃 🏠 🔎 Search 🤺 Favorites 📢 Media 🧭 😒 -	😓 🔟 • 🔜 🦓	
-	dl?e=p-0-00-wordpdf00-1-0-0prompt-1040-111-en-5020-abo		💌 ラ Go 🛛 Links 🌺
Aguress enttp://iocainost/gsi	a/e=p-u-uu-woraparuu-1-u-uprompc-1u4u-111-en-su2u-abo	ut00031-001-1-00tr22-8-00&a≕q	
	MandDDE	HOME HELP PREFERENCES	
	WordPDF	search	
	الله 🗭 📂	Search	
		s a-z authors a-z	
		hich contain some 💌 of the words	
	text Document Title	Begin Search	
	Authors		
			-
ë –			Scolar intranet

Enter a search term, 'witten', and click on <u>Begin Search</u>. Titles of the matched records with the corresponding filenames will be displayed as shown in this figure:

ile <u>E</u> dit <u>V</u> iew	Favorites <u>T</u> ools <u>H</u> elp		
🔁 Back 👻 🕥 -	🗸 🖹 🙆 🏠 🔎 Search 👷 Favorites 🜒 Media 🤕 😒 🖓	🍃 🗹 🕘 🦓	
ddress 🙆 http://lo	calhost/gsdl?a=q&r=1&hs=1&e=q-0-00-wordpdf00-1-0-0prompt-1040-1 1-er	-5020-about00031-001-1-0utfZz-8-00&fqf=TX&t=1&q=witten	💌 芛 Go 🛛 Links 🌺
			_
	WordPDF	HOME HELP PREFERENCES	
		search	
	search titles a	authors a-z	
	Search for text 💌 whic	h contain some 💌 of the words	
	witten	Begin Search	
	results		
	Word count: witten: 69		
	11 documents matched the query.		
	📓 彛 1997-00 Listing of Working F	apers	
	(word01.doc)		
	Image: Scalable browsing for large c (pdf04.pdf)	ollections: a case study	
7	(2 5 2 57	mpression Models	
	(text cat.ps)		
	Clustering with finite data from	n semi-parametric mixture	
	distributions		
	(cluster.ps)		
	Software System	e Open-Source Digital Library	
	(pdf01.pdf)		
	📓 🐼 1992-96 Working Papers		
Dana			Local intranet
Done	Search Re	, , , ,	

You can restrict the search to a specific field.

The same search term, 'witten', when restricted to the 'Creators' index retrieves only four documents as shown in the following figure:

	en - Microsoft Internet Explorer Iools <u>H</u> elp		
		• 🛬 🖬 • 🗔 🦓	
Address E http://localhost/gsdl	?a=q&r=1&hs=1&e=q-0-00-wordpdf00-1-0-0prompt-10-TX40-	1 1-en-5020-about-witten00031-001-1-0utfZz-8-00&fqf=DR&t=1	1&q=wit 💌 🔁 Go 🛛 Links '
			ŕ
	WordPDF	HOME HELP PREFERENCES	
		search	
	<u>></u> 22		
	search title	authors a-z	
	Search for Authors 💽 w	hich contain some 🗾 of the words	
	witten	Begin Search	
	,		
	results		
	Word count: witten: 7		
	7 documents matched the sucre		
	Clustering with finite data i	from semi-parametric mixture	
	distributions	r	
	(cluster.ps)		
	📓 🔥 Using language models for	r generic entity extraction	
	(langmodl.ps)		
	📓 Scalable browsing for larg	e collections: a case study	
	(pdf04.pdf)		
		es on the Web, CD-ROMs, and	
	(pdf02.pdf)	n, same look-and-feel, different media	
		Compression Models	
	Text Categorization using (text cat.ps)	Compression models	
		nsive Open-Source Digital Library	
		sive open source bigin thorny	2
🛃 Done			New Yorking Contract

6. Example of Digital Library Software: Greenstone - 3. Using GLI: designing, creating and previewing a collection - page 20

You can do a form-based search which facilitates searching across fields. During the design section (search types), we chose 'plain search' as the default search type. To do a form based search, click on 'PREFERENCES". Check the 'Form Search' option and check the 'advanced' 'Form Type'. Also, ensure that 'ignore case differences' and 'ignore word endings' are checked. Save the changes by clicking on 'set preferences'. A partial view of this screen is shown in the figure below.

	ences - Microsoft Internet Explorer Favorites Iools Help	-
↔ → Back Forward	→ 🖄 🚰 🥘 🖬 🏈 🛃 → 🎒 🖸 · 🗐 Stop Refresh Home Search Favorites History Mail Print Edit Discuss	
Address 🕖 ompt-10-1	1X40-111-en-5020-about-%22ian+h+witten%22+%26+yang00031-001-1-0uttZz-8-00&a=p&p=preferences&qt=1 💌 🎓 G	o ∫Links '
	Interface language: English	4
	Encoding: Unicode (UTF-8)	
	Interface format: Graphical 💌	
	Search preferences	
	set preferences	
	C Plain search:	
	Query • simple query mode	
	mode: O advanced query mode (allows boolean searching	
	using !, &, , and parentheses)	
	Query box 🔹 or regular query box	
	size: O large query box	
	• Form search:	
	Form type: O simple	
	c advanced	
	with 4 🔽 fields	
	Case • ignore case differences	
	differences: O unnerlawer and must match	
🛃 Done	📄 📄 🔮 Internet	

🚰 Search page - Microsoft Interne	t Explorer		
<u>File Edit View Favorites T</u> ools			
🌀 Back 🝷 💮 🖌 🔀 🔮	🏠 🔎 Search 🤺 Favorites 📢 Media 🤕	🗟 • 😓 🖬 • 🔜 🦓	
Address 🕘 http://localhost/gsdl?e=p-	0-00-wordpdf00-1-0-0prompt-10-DR40-1 1-en-50-	20-preferences-witten01131-001-1-0utfZz-8-00&a=q	💌 芛 Go 🛛 Links 👌
	WordPDF	HOME HELP PREFERENCES	<u>-</u>
	🖗 😼	Jeulen	
		itles a-z authors a-z	
	Search and display results i	n ranked 💌 order	
	Word or phrase	(fold, stem) in field	
		Authors	
	and 🔽	Document Title	
	and 🔽	Document Title V	
	Clear Form	Begin Search	
	Clear Form	Degni Gearch	
	Or enter a query directly:		
		Run Query	
	1		

If you now click on the 'Search' button, you should get the advanced form based search interface as shown in the figure below.

A sample field-based search is shown in the figure below. In this search, the first search term 'witten' is searched for in 'Creators' and is combined, using the Boolean operator & (AND), with ' the search term 'cluster' in the 'titles' field. The search results in one document. Ensure that the 'fold' and 'stem' check boxes are checked.

File Edit View Favorites	vitten#si]:DR & [cluster#si]:DC - Microsoft Internet Explore : Tools Help		
-		3- 🗟 🗊 🛄 🦓	
			ter%2 🔻 ᠫ Go 🛛 Links 🎽
Aguress en http://iocalhost/gs	sare=q-u-uu-wordpdrUU-1-U-uprompt-1Uand%2cand%2cand-DR%2	cDC%2cTX%2cDC-0%2c0%2c0%2c0-4-0%2c0%2c0%2c0-witten%2cclus	tervozi 🗾 🖬 Go 🛛 Links **
		HOME HELP PREFERENCES	1
	WordPDF	search	
	🗭 🗭 📂	Search	
	search ti	tles a-z authors a-z	
	Search and display results in	ranked 💌 order	
	Word or phrase	(fold, stem) in field	
	witten	Authors	
	and 🔽 cluster	Document Title	
	and	text	
	and	Document Title	
	Clear Form	Begin Search	
	Or enter a query directly:		
	[witten#si]:DR & [cluster#si]		
	[witten#si]:DR & [Cluster#si]		
		Run Query	
	results		
	results		
	Word count: cluster: 1, witten: 7		
	1 document matched the query.		
	Clustering with finite data f	rom semi-parametric mixture distributions	
	(cluster.ps)		
Done			Nocal intranet
	Advanced Form	Based Searching	