

Electronic Submission and Management of Publications for ICDO

Travail de diplôme réalisé en vue de l'obtention du diplôme HES

par : **Hany BOTROS**

Conseiller au travail de diplôme :
(David Billard, Professeur IG)

Genève, 13.05.2009
Haute École de Gestion de Genève (HEG-GE)
Filière Informatique de gestion

Déclaration

Ce travail de diplôme est réalisé dans le cadre de l'examen final de la Haute école de gestion de Genève, en vue de l'obtention du titre Bachelor d'informaticien de gestion. L'étudiant accepte, le cas échéant, la clause de confidentialité. L'utilisation des conclusions et recommandations formulées dans le travail de diplôme, sans préjuger de leur valeur, n'engage ni la responsabilité de l'auteur, ni celle du conseiller au travail de diplôme, du juré et de la HEG.

« J'atteste avoir réalisé seul le présent travail, sans avoir utilisé des sources autres que celles citées dans la bibliographie. »

Fait à Carouge, le 13.05.09

Hany Botros

.....

Thanks

On behalf of Mr. David BILLARD, professor HEG Genève, and I would to thank all the members of the ICDO, especially Mr. Nawaf EI SEILEBI, Secretary-General and Mr. Rayidh ATWAN, head of the department of research and development, for their help and collaboration in realizing this project and hope that the digital library improve the services offered by ICDO to its members.

Special thanks to Mr. Alexander Boder, Professor HEG Genève, for his help and advices regarding the Indexing and cataloguing.

Last and not least, thanks to the IT centre team of the HEG, that has spared no effort to install Greenstone Digital Library on Linux test server.

Synopsis

The International Civil Defence Organization is looking forward to offer its services of library to its 50 member states, by submitting and managing its publications

This project is the beginning of bilateral collaboration between ICDO, represented by Mr. Nawaf EL SEILIBI the General-Secretary, and HEG Geneva, represented by Mr. Mr David BILLAD professor HEG IG section.

The project goals expressed by ICDO:

- Save time and effort of the ICDO in answering the requested subjects of members, institutions, non-members and individuals
- Offer the latest human knowledge about the civil defence and civil protection
- Distribute the information about the civil defence and civil protection and their functions for the wellbeing of humanity
- Help researchers, planners and decision makers by providing periodic publications about the latest of human knowledge in the field of civil protection

During the period of the project HEG team collaborated with Mr. Riyadh ATWAN head of ICDO Research and Development department, the HEG team consisted of:

Head of the project: Mr. David BILLARD.

Diploma Thesis: Mr. Hany BOTROS

The following stages were undertaken by the project team to achieve the entrusted mission:

- Project planning and management
- Business analysis
- Use and test Greenstone's different features
- Research on library Workflow (Catalogue & index)
- Installing and configuring Greenstone (test sever)
- Staff training

The identified library building processes and the deployment of Greenstone Digital library will permit the ICDO to make his first step in the way towards the civil protection knowledge revolution that ICDO seeks.

Members, non-members, searchers and individuals all over the world will finally be able to follow and consult the latest researches, reports and recommendations of ICDO through the on-line library services.

Contents

Déclaration.....	i
Thanks	ii
Synopsis	iii
Contents.....	v
List of Table and Figure.....	vi
Introduction	1
1.Library building processes for ICDO.....	3
1.1 Selection	5
1.2 Indexing and Cataloguing	6
1.3 Metadata and Dublin Core.....	6
1.3.1 Dublin Core.....	7
1.4 Digitisation.....	9
2. Greenstone Digital Library.....	11
2.1 Historic and examples of use	13
2.2 Technical Requirements for deployment	17
2.3 Greenstone Installation & configuration of web server for it	20
2.4 Librarian Interface	22
3. Needs of ICDO and Greenstone Digital Library	26
4. Project Progress and problems met	28
4.1 Project Progress.....	29
4.2 Problems met.....	32
Conclusion.....	33
References:.....	34
Appendix Librarian interface User manual	35

List of Tables and Figures

Table 1	Steps of the project	28
Figure 1	Digital library building processes	5
Figure 2	Metadata element.....	7
Figure 3	Dublin Core	9
Figure 4	Digital Library building	13
Figure 5	Greenstone and web server	19

Introduction

The International Civil Defence Organisation (ICDO) is an inter governmental organisation whose objective is to contribute to the development by states of structures ensuring the protection and assistance of populations and safeguarding property and the environment from natural or man-made disasters.

These structures are generally known as civil protection, civil defence, civil safety and emergency management. The ICDO federates these structures with the aim of favouring cooperation and mutual solidarity between them.

To achieve their mission the ICDO has decided to establish a centre for Documentation and Information Studies, collection, compilation and organization of scientific knowledge related to civil defence and civil protection at the level of Organization and member states, regarding it as a complement to the workshops of service provided to member states, in order to provide all forms of knowledge and scientific support as possible for those countries or institutions or individuals, in the field of scientific research on the civil defence and civil protection.

According to the agreement between the HEG team and the ICDO, the specifications of the project are:

- Members states bodies send their publications to the ICDO, either in digital or printed format
- ICDO transforms the publications in digital format and associates keywords and attributes for each publication, to that effect , ICDO creates and maintains a keywords corpus and a list of categories into which the publications can be fitted
- The publications are loaded to a server
- ICDO can manage the publications, keywords and categories on the server
- ICDO can delegate to some/all member states bodies the right to upload publications on the server
- The publications will be available on the internet through a link on the ICDO website

The team purpose of the project is analyzing and implementing a simple library system for the publications of ICDO and its members states

This report is trying to redraw the steps and the progress of the project, starting by analyzing the needs of ICDO library processes, using Greenstone Digital Library as a tool or library system, passing by a global research on some processes of documents live cycle in a library system.

The final section of this report handles the progress of the project and the met problems

The appendix describes the main librarian interface of Greenstone.

1. Library building processes for ICDO

The first step of the project was the business analysis and to do it the following questions had to be answered

- What are the business goals or why outsiders will demand the services offered, what are the benefits of doing this and what are their expectations and needs?
- How the business is done to fulfill the needs or how the work is internally organized.

About the first point, the answer is easy to find, why a net surfer would come to CDO's digital library? Evidently the user of ICDO digital library will look for a record (document, book...etc) talking about certain subject concerning the civil protection.

To answer the second question I have tried to identify the

- Business processes or the mains processes witch are directly related to the business goals
- The auxiliary processes witch are not directly related to it,

After analysing the needs of the ICDO through several meetings with Mr. ATWAN head of the department of research and development also in charge of the future Digital library, we noticed the following points:

- The actual collection of documents of ICDO is almost 1200 documents. They are published in several languages including English, French, Arabic, Russian and others , they consist of books, reports, periodical publications of ICDO and its members, these publications are organised in several ways, type of document, language, country, date of creation ... etc.
- These publications are generally serial edited. They are issued yearly, quarterly or biannually.
- The increase of number of documents is about 40 per year
- A gallery of 2metres by 4metres is the only reserved place for the documents of ICDO
- Only Mr. Atwan is in charge of organising the ICDO's publications

To offer information or records for net surfer, these records should be organized or classified in a certain way to help user finding the right subject, in other word a digital library or digital collections should be based on a real collections, yet the publications of the ICDO are not clearly classified, some publications are classified by language others by country publisher, others by serial numbers or date

The conclusion about the above points is the leak of a determined structure, strategy and workflow for the library, also human and materials resources should be decided in function of the needs of the project

Our suggestion is that ICDO should implement different processes for cataloguing and organising its documents because a digital collection is generally based on a physical one.

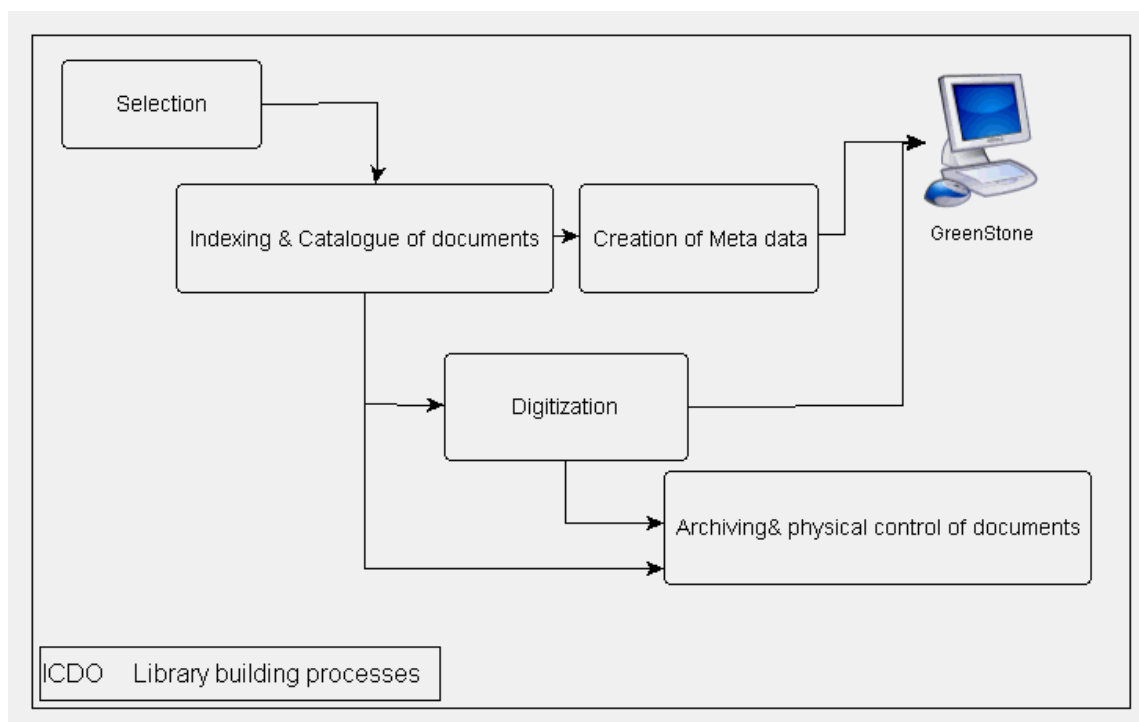
A simple practical way to start preparing for the digital library is the following 2 steps:

- A list of documents containing the bibliographic details should be created, at least an Excel file listing these documents, their physical position and metadata about them
- Classification process is the most difficult because the title or the introduction does not always mention the subjects handled in the document so a document overview should be done to reveal the subjects handled by it, and another list of topics, subtopics...etc should be created with a number (decimal number) assigned to each item of this list, In other word this corresponds to a handmade thesaurus, each document's metadata will contain an item to denote the category of the document that will greatly facilitate the transition to the digital library

After these steps comes the selection process that is treated below.

Figure 1 will show the suggested processes of building the library

Figure 1
Digital library building processes



1.1 Selection

The process of selecting a collection to build is the most important process, the following points should be considered to take the decision of choosing a collection to build:

1. The collection should support the objectives of the project, ICDO library team should define clearly the goals of the digital library, inquire the future users about the collections or subjects that might be interesting for them and finally collect documents that will compose every collection
2. Copy rights is not an easy subject to handle because many digitized documents are not accompanied by a clear statement of ownership or copy rights, so ICDO has to decide either to publish a collection to public or keep its access restricted to its members

3. Staff, space and budget should be considered when creating a new collection because creating and maintaining a collection for a long period of time is resources consuming tasks

1.2 Indexing and Cataloguing

As neither the team nor the ICDO's responsible of researches development department are expert in this domain we have asked consultation and advice of Mr. Alexander Boder Liberian consultant

According to Mr Boder concern the cataloguing and classification:

Classification is the process of attributing a certain category to a book or document in function of the subjects treated in it, most of the important library uses a universal classification system like DDC (Dewey decimal classification) or AACR but according to Mr Boder and the analysis of the number of documents managed by the ICDO and the limited number of subjects treated in these documents a universal system would be too rich for library needs of the ICDO.

The ICDO will adopt a personalized classification system that they would have elaborate themselves, so they have to determine their own procedure and rules for classification

For cataloguing they will use the Dublin Core as a base for their cataloguing and they will add or remove metadata in function of their needs

1.3 Metadata and Dublin Core

Metadata consists of statements we make about records to help us find, identify, use, manage, evaluate, and preserve them.

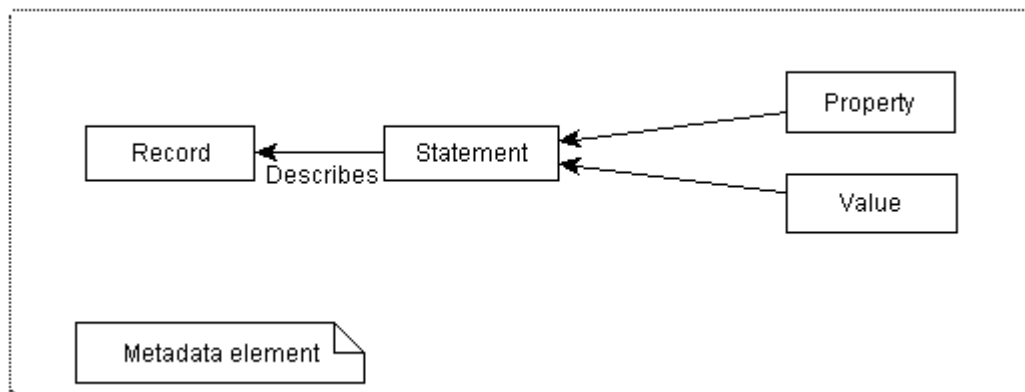
So metadata can be summarized as:

1. The basic unit of metadata is a statement.

2. A statement consists of a property (element) and a value.
3. Metadata statements describe records.

Figure 2 describes metadata

Figure 2
Metadata element



1.3.1 Dublin Core

Dublin core is a set of metadata describing a document or a record; it came to existence in 1995 due to:

- Increase in the number of document-like records on the net
- Slow improvement in indexing services made records hard to discover
- Belief that descriptive metadata would improve discovery
- Perceived need for a descriptive standard that was simple to apply (by non-professionals)

The basic Dublin core set describes 3 categories and consists of the following 15 elements:

Content:

- Title: The name given to the record by the creator or publisher
- Subject: The topic of the record

- Type: The nature or genre of the content of the record
- Description: A textual description of the content of the source
- Source: A reference to a second record from which the present record is derived
- Relation: A reference to a related record, and the nature of its relationship

Responsibility:

- Creator: The person responsible for the intellectual content of the record
- Publisher: The entity responsible for making the record available
- Contributor: A person or organization (other than the Creator) who is responsible for making significant contributions to the intellectual content of the record
- Rights: Information about rights held in the record

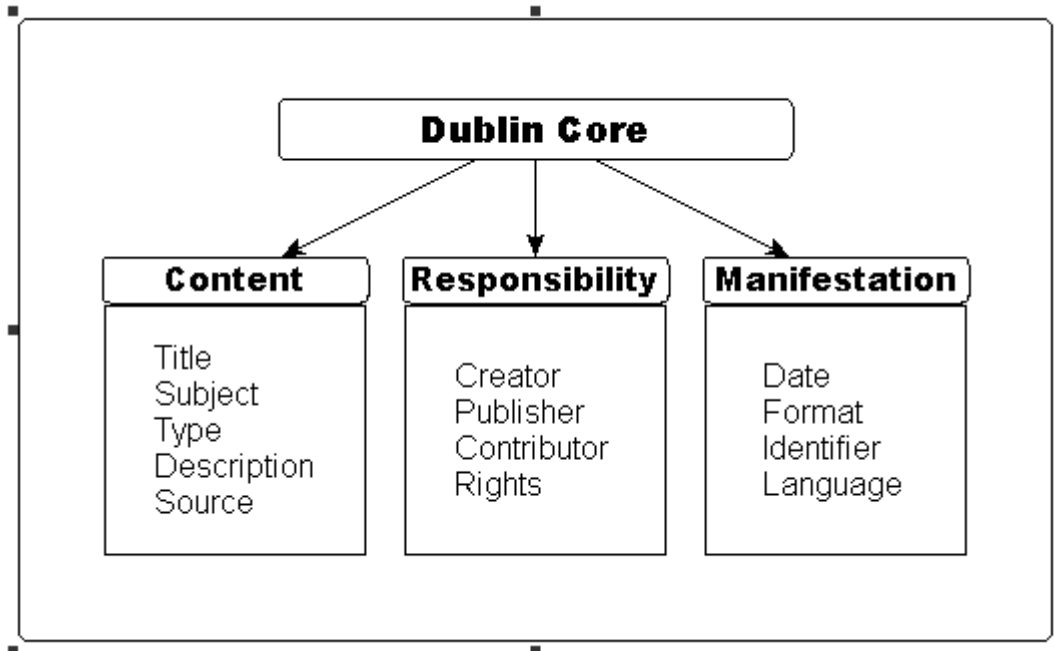
Manifestation:

- Date: A date associated with the creation or availability of the record
- Format: The physical or digital manifestation of the record
- Identifier: An unambiguous reference that uniquely identifies the record within a given context
- Language: The language of the intellectual content of the record

Each element is optional and can be repeated.

Figure 3 shows Dublin Core elements

Figure 3
Dublin Core



1.4 Digitisation

Building of digital library is based on:

- a- Digital format documents like word, PDF, jpeg, gif ... etc
- b- Image resulted from scanning documents

Digital formats are supported by Greenstone's plug-ins (see section 5.2 for more details about supported file formats).

When migrating from document to a digital format, there are two methods:

- a- Scan the document to obtain an image file
- b- Scan the document and decode the obtained image by an OCR (Optical Character Recognition) to extract the text contained in the scanned image

The process of scanning a document to get an image is much easier than decoding the image to obtain a text file, because one has to use an OCR to achieve the task.

ICDO's documents are published in several languages (English, French, Arabic, Russian, Asian languages) so an OCR tool has to decode these languages, these are some soft wares that dominate this market:

[FineReader](#)

[OmniPage](#)

[ReadIRIS](#)

[TextBridge](#)

[SimpleOCR](#)

Only ReadIRIS is capable of decoding special languages like Arabic and Asian languages, but even under the best conditions: a high resolutions scanner, black and white text document; The user of such software has to be aware that this kind of tools is not 100% reliable, so every text decoded by the OCR has to be compared carefully to the original one and corrected manually. Furthermore the text page obtained after the decoding process has to be submitted to a page setup process.

2. Greenstone Digital Library

According to Greenstone home page and factsheet

Greenstone is a suite of software for building and distributing digital library collections. It provides way of organizing information and publishing it on the Internet or on CD-ROM. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software, issued under the terms of the GNU General Public License.

Platforms:

Greenstone runs on all versions of Windows, and Unix/Linux, and Mac OS-X. For the default Windows installation absolutely no configuration is necessary, and end users routinely install Greenstone on their personal laptops or workstations. Institutional users run it on their main web server, where it interoperates with standard web server software (e.g. Apache).

Interoperability:

Greenstone is highly interoperable using contemporary standards, it incorporates a server that can serve any collection over the Open Archives Protocol for Harvesting (OAI-PMH), and Greenstone can harvest documents over OAI-PMH and include them in a collection. Any collection can be exported to METS (Encoding and Transmission Standard) and Greenstone can ingest documents in METS form.

Interfaces:

Greenstone has two separate interactive interfaces, the Reader interface and the Librarian interface. End users access the digital library through the Reader interface, which operates within a web browser. The Librarian interface is a Java-based graphical user interface (also available as an applet) that makes it easy to gather material for a collection (downloading it from the web where necessary), enrich it by adding metadata, design the searching and browsing facilities that the collection will offer the user, and build and serve the collection.

Metadata formats:

Users define metadata interactively within the Librarian interface. These metadata sets are predefined:

Dublin Core (qualified and unqualified)

RFC 1807

NZGLS (New Zealand Government Locator Service)

AGLS (Australian Government Locator Service)

New metadata sets can be defined using Greenstone's metadata Set Editor. "Plug-ins" are used to ingest externally-prepared metadata in different forms, and plug-ins exist for: XML, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI, DSpace, METS

Document formats:

Plug-ins are also used to ingest documents. For textual documents, there are plug-ins for: PDF, PostScript, Word, RTF, HTML, Plain text, Latex, ZIP archives, Excel, PPT, Email (various formats), source code. For multimedia documents, there is plug-ins for: Images (any format, including GIF, JIF, JPEG, and TIFF), MP3 audio, Ogg Vorbis audio, and a generic plug-in that can be configured for audio formats, MPEG, MIDI, etc.

Languages:

One of Greenstone's unique strengths is its multilingual nature. The reader's interface is available in the following languages: Arabic, Armenian, Bengali, Catalan, Croatian, Czech, Chinese (both simplified and traditional), Dutch, English, Farsi, Finnish, French, Galician, Georgian, German, Greek, Hebrew, Hindi, Indonesian, Italian, Japanese, Kannada, Kazakh, Kyrgyz, Latvian, Maori, Mongolian, Portuguese (BR and PT versions), Russian, Serbian, Spanish, Thai, Turkish, Ukrainian, Vietnamese
The Librarian interface and the full Greenstone documentation (which is extensive) is in: English, French, Spanish, and Russian

Collection building:

- Input to GSDL collection building is a set of source documents
- GSDL 'imports' these collections and converts them to GAF format (XML/HTML)
- GSDL 'builds' indexes and browsing structures using the GAF files
- 'collect.cfg' determines content conversion, extraction and building of indexes and browsing structures: Indexes, classifiers, and plug-in

- Presentation of search/browse results and collection interface is determined by 'format' strings in 'collect.cfg' and 'macros' in 'main.cfg'

Greenstone used to have three modes of collection building, Command Line, Web Interface and the GLI (Greenstone Librarian Interface).

Progressing with version 2.4x the GLI got strengthened as well as popularized Web Interface mode has been withdrawn temporarily.

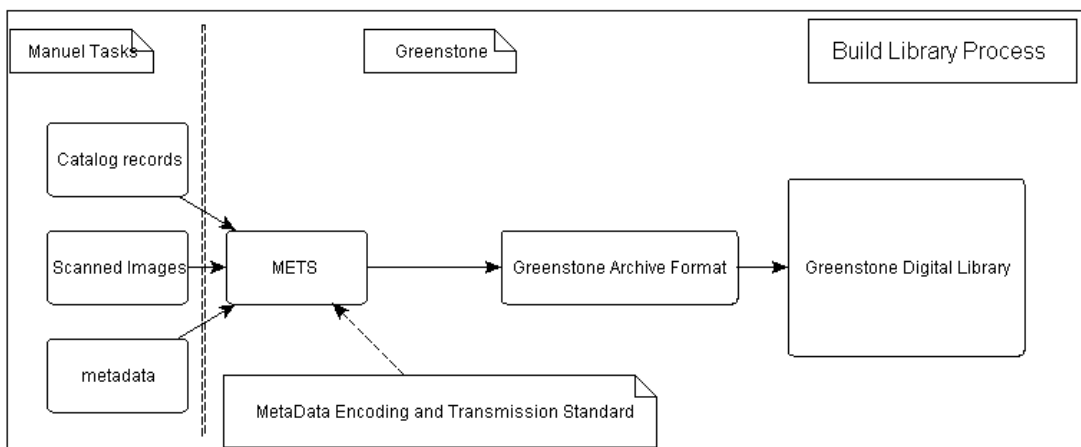
The GLI based collection building is a quite simple method and recommended for beginners.

Collection developers can activate the GLI software and use the "Gather", "Enrich", "Design", and "Create" panel for making collection.

The following Figure shows the library building process with Greenstone

Figure 4

Digital Library building



2.1 Historic and examples of use

“The New Zealand Digital Library Project was established in 1995, beginning with a collection of 50,000 computer science technical reports downloaded from the Internet (Witten et al., 1995). At the time several research groups in computer science departments collecting technical reports and making them available on the web

The New Zealand Digital Library project grew out of research on text compression (Bell et al., 1990) and, later, index compression (Witten et al., 1994).

In 1997 we began to work with Human Info NGO to help them produce fully-searchable CD-ROM collections of humanitarian information.

We became acquainted with UNESCO through Human Infos long-term relationship with them. Although they supported Human Infos goal of producing humanitarian CD-ROMs and distributing them in developing countries

From the outset, UNESCO's goal was to produce CD-ROMs containing the entire Greenstone software (not just individual collections plus the run-time system, as in Human Infos products), so that it could be used by people in developing countries who did not have ready access to the Internet

At the time of writing (January 2007) Greenstone—a versatile open source multilingual digital library environment with over a decade of pedigree—has a user base hailing from over 70 countries, is downloaded 4,500 times a month, runs on all popular operating systems (even the iPod!), and has a reader's interface in over 40 languages. How did this software project and the research team behind it reach this point? Team members often give anecdotal stories about life behind the scenes at conferences and workshops; this article gives a more definitive and coherent account of the project”

Ian H. Witten and David Bainbridge

University of Waikato, Hamilton, New
Zealand

Greenstone is actually used by over 60 organizations all over the world; we will mention some examples in the following section

Oxford Digital Library

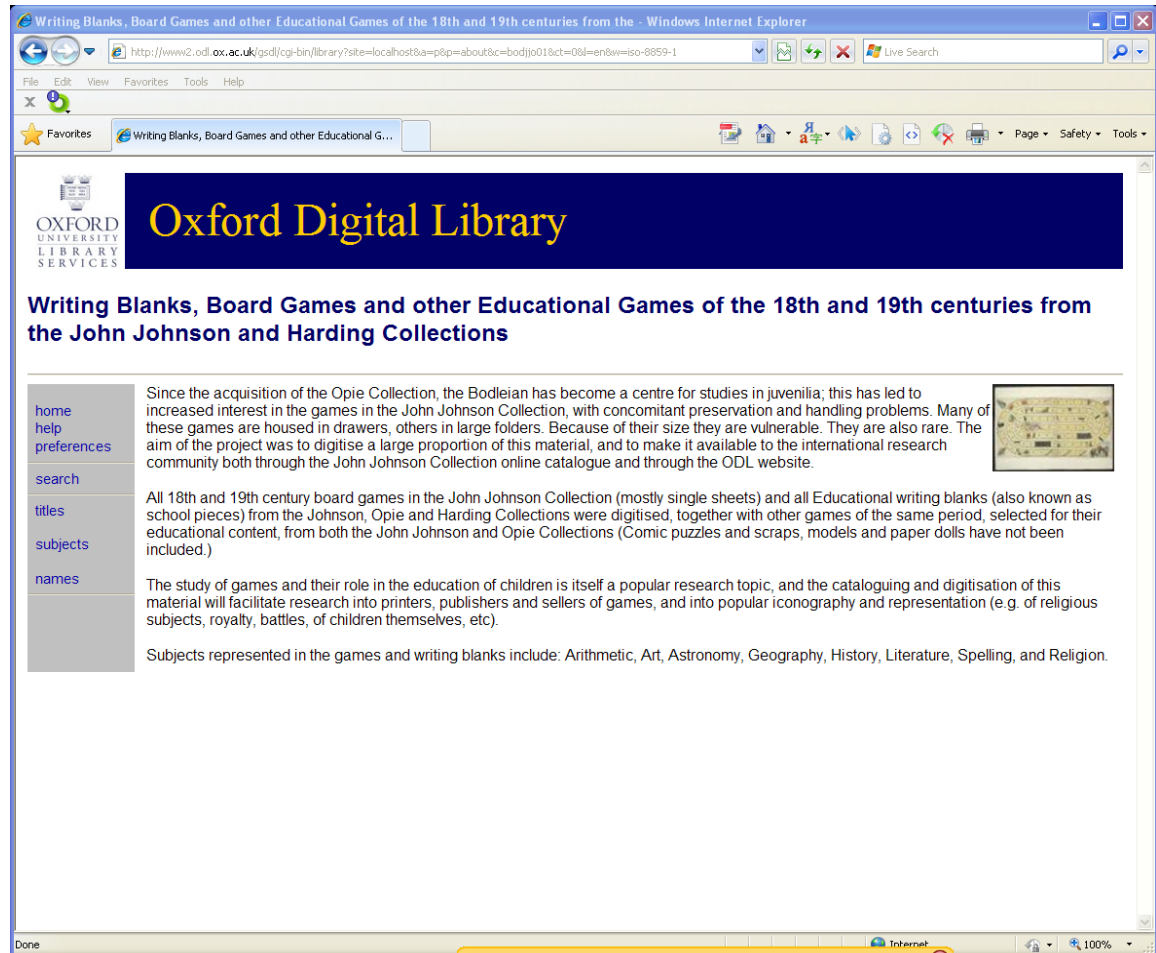
According to Oxford Digital Library home page:

The Oxford Digital Library (ODL) is a core service of Oxford University Library Services. This web site offers central access to digital collections of Oxford libraries and informs you about ODL services, funding activities, digital library technology and developments. The Oxford Digital Library has a major role in co-ordinating and stimulating digitization activities in the University.

The Oxford Digital Library started operationally in July 2001 and will be under continuous development in the coming years.

ODL uses Greenstone and consists of dozen or so collections on an eclectic variety of subjects: ornithology and entomology, prints and maps, educational games from the 18th and 19th century; classic geological literature, medieval science, and many others

Link to Oxford Digital Library: <http://www2.odl.ox.ac.uk/gsd/cgi-bin/library>



United Nations Digital Library Islamabad

According to Islamabad Digital Library:

The United Nations Digital Library Islamabad, is an Open-Access, online searchable repository containing full-text of documents, reports, publications and other public information items produced by the country offices of United Nations Organizations, Programs and Funds in Pakistan.



Visit United Nations Digital Library Islamabad on <http://library.un.org.pk/gsd/cgi-bin/library>

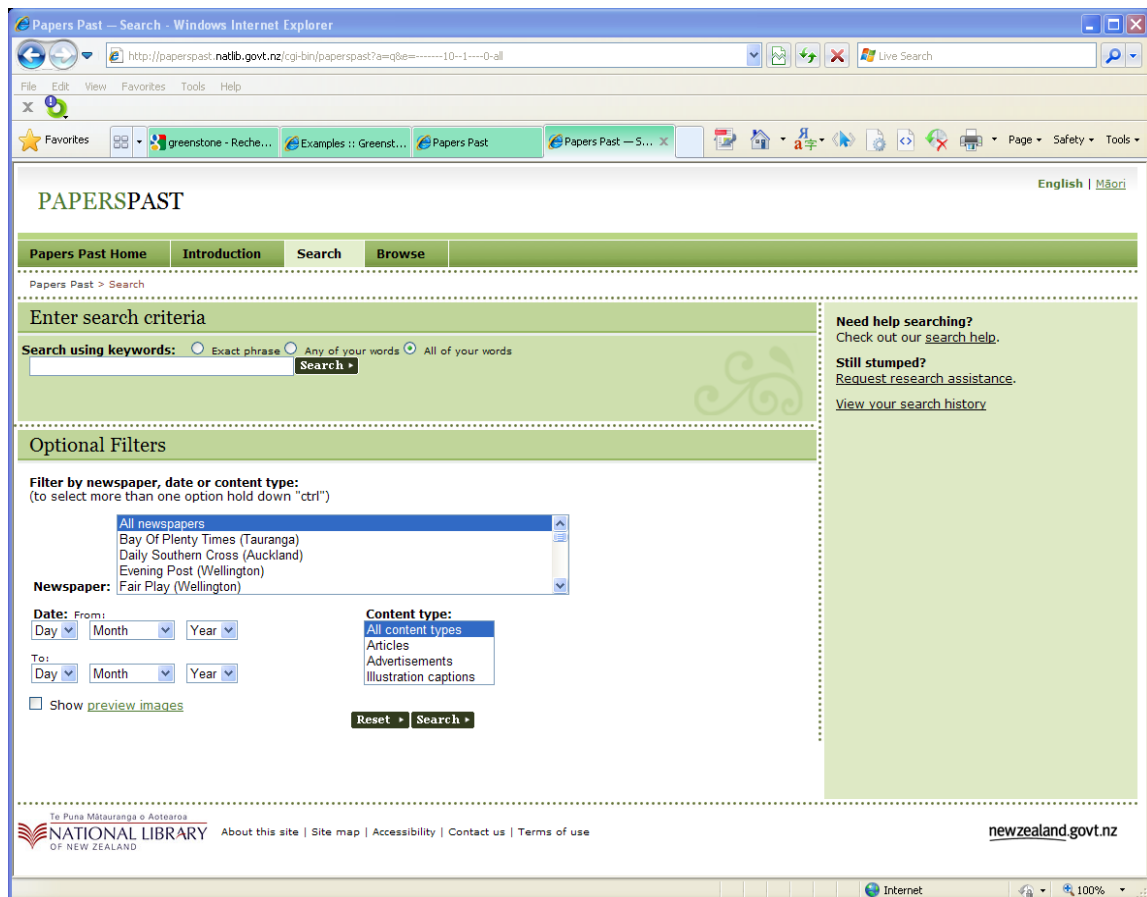
PAPERSPAST

According to Paperpast Digital Library:

Papers Past contains more than one million pages of digitised New Zealand newspapers and periodicals. The collection covers the years 1840 to 1915 and includes publications from all regions of New Zealand.

There are two main ways to find information in Papers Past: searching and browsing. Browsing lets you look at all the newspapers, starting with a year, a region, or a newspaper title. Searching lets you enter a query term and retrieves articles that contain that term. All the newspaper titles on the site can be browsed. Currently, only some of the titles can be searched. An asterisk (*) beside the name of a newspaper indicates that it can be searched

Find PAPERPAST on <http://paperspast.natlib.govt.nz/cgi-bin/paperspast>



2.2 Technical Requirements for deployment

Installing and using Greenstone requires the following dependences

- Perl: several libraries of Perl is used in the Greenstone kernel
- Java: used for running the GLI(Greenstone Librarian Interface) and some Java classes are used by Greenstone, from the version 2.4X further on and version 3.X is written mainly using Java

Installing Greenstone executable version on Windows works directly without configuration (web server, Perl, and JVM are included), but installing it on UNIX like systems requires a web server, the installation on Linux is described later in this chapter

After the installation of Greenstone the folder created for the installation will contain the following folders:

Folder Name	Contents
<i>bin</i>	Executable code, including binaries in the directory with your O/S name.
<i>bin/script</i>	Perl scripts used for creating and building collections (for example <i>import.pl</i> and <i>buildcol.pl</i>). To get a description of any of these programs, type their name at the command prompt.
<i>perllib</i>	Perl modules used at import and build time (plugins, for example).
<i>perllib/plugins</i>	Perl code for document processing plugins.
<i>perllib/classify</i>	Perl code for classifiers (for example the AZList code that makes a document list based on the alphabetical order of some attribute).
<i>cgi-bin</i>	All Greenstone CGI scripts, which are moved to the system cgi-bin directory.
<i>tmp</i>	Directory used by Greenstone for storing temporary files.
<i>etc</i>	Configuration files, initialization and error logs, user authorization databases.
<i>src</i>	C++ code used for serving collections via a web server.
<i>src/colservr</i>	C++ code for serving collections—answering queries and the like.
<i>src/recpt</i>	C++ code for getting queries from the user interface and formatting query responses for the interface.
<i>packages</i>	Source code for non-Greenstone software packages that are used by Greenstone.
<i>packages/mg</i>	The source code for mg, the compression and indexing software used by Greenstone.
<i>mappings</i>	Unicode translation tables (for example for the GB Chinese character set).
<i>macros</i>	The macro files used for the user interface.
<i>collect</i>	Collections being served from this copy of Greenstone
<i>lib</i>	C++ source code used by both the collection server and the

	receptionist.
<i>images</i>	Images used in the user interface.
<i>docs</i>	Documentation.

On line digital library requires a web server, for our project we will use Apache but using another web server is possible.

The web sever will use Greenstone's CGI "Common Gateway Interface" to show collections and the other pages of your digital library

Web clients (browsers) communicate with web servers using HTTP protocol, the form of information exchanged is html, when a web server receive a request for a client, it will analyzes the request and acts in function of the requested object, two kinds of object may be contained in a request:

File: in this case the server will return the contents of the file coded en html

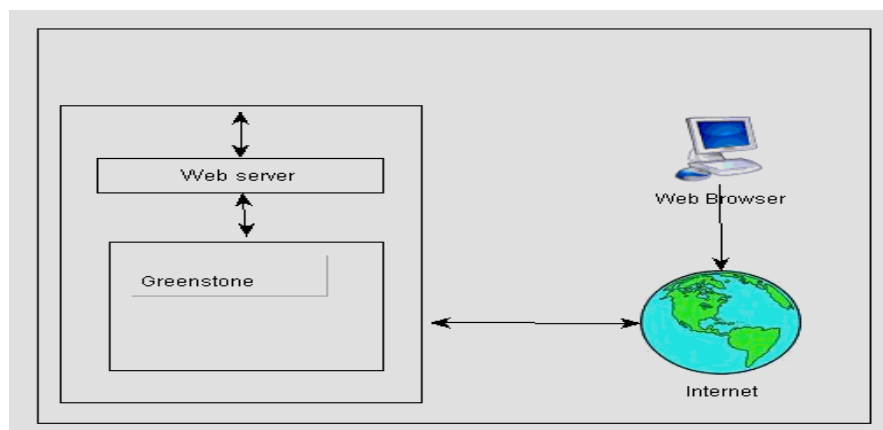
A program or executable: the server will run the program and return back the output coded in html

Greenstone is a program executed by the server or CGI, so when you search the title that contains "toto" in your collection, your browser will send a request for the server asking it to execute the search program of Greenstone with "toto" as argument; the server formats the output in html and sends it to the client

A simple figure below shows this process

Figure 5

Greenstone and web server



2.3 Greenstone Installation & configuration of web server for it

The installation of Greenstone on a Linux server needs some technical skills because you have to install Greenstone, configure the web server (apache) and modify permissions of Greenstone and its subfolders, so you have to know how to manage users and permissions under Linux to make Greenstone web sever accessible for internet users

- 1- Download Greenstone-2.81-linux.bin from the Greenstone web site
- 2- Go to the folder where you have downloaded the binary file

Execute the binary with the following command line:

```
./Greenstone-2.81-linux.bin
```

You may need to make the file executable with the next command

```
chmod +x Greenstone-2.81-linux.bin
```

Of course you have to be either a root user or a sudoer user if the bin file is not in the tree of your home folder

- 3- Follow the instructions
- 4- The proposed location for installation is /usr/local/GreenStone2, the installation in this folder will not work if you are neither root nor have the permission to write in it, a solution is changing the location to your home folder or create a new user for Greenstone and install the Greenstone Digital Library in the home folder of this new user
- 5- To execute the ./Greenstone-2.81-linux.bin as a normal user you have to:
 - a. Change the property of the binary file with the following command:

```
$ chmod +x Greenstone-2.81-linux.bin
```

- b. Have the permission to write in the installation folder
- 6- To avoid changing a lot of permissions and make the installation complicated the better way is to execute the binary file as root and leave the proposed folder /usr/local/GreenStone2.

- 7- Choose the password for admin user, this password is used when using the Administration page on Greenstone home page
- 8- Once you have the message « Successfully built» you can pass to configuring the web server (e.g. Apache)

Configuring web Server (Apache)

To use the Greenstone Web Library you will need to have a web server installed (you will also need perl if you want to build your own collections). The following information is intended to help you configure your web server for use with Greenstone. Particular attention is given to the Apache web server, which may be downloaded free from www.apache.org. Greenstone will work with other web server too if that is your preference.

Greenstone has been set up to expect `$GSDLHOME/GS2` to be accessible from your web server at the URL `http://localhost/gSDL` (localhost will be the web address of your machine if it has one) Likewise, the `$GSDLHOME/cgi-bin` directory must be accessible as a cgi executable directory.

For the Apache web server this means adding the following directives to your `httpd.conf` configuration file:

Edit `/etc/apache2/httpd.conf` and add: the following directives to `httpd.conf`

```
ScriptAlias /gSDL/cgi-bin "$GSDLHOME/cgi-bin"
```

```
<Directory "$GSDLHOME/cgi-bin">
```

```
    Options None
```

```
    AllowOverride None
```

```
</Directory>
```

```
# The ScriptAlias directive is used to point at the Greenstone's cgi
```

```
Alias /gSDL "$GSDLHOME"
```

```
<Directory "$GSDLHOME">
```

```
    Options Indexes MultiViews FollowSymLinks
```

```
    AllowOverride None
```

Order allow,deny

Allow from all

```
<FilesMatch "^\.(gif|jpe?g|png|css|mov|mpeg|ps|pdf|doc|rtf|jar|class)$">
```

Order allow,deny

Allow from all

```
</FilesMatch>
```

```
</Directory>
```

The Alias directive is used as DocumentRoot for Greenstone without this directive the collections created will not be accessed on Greenstone homepage

<FilesMatch > is used for security issue so the librarian will be able to load files with extensions that match the regex (regular expression) in the directive

Once your web server is configured, you can access Greenstone by pointing your web browser at `http://localhost/gsd/cgi-bin/library` or `http://your.server.com/gsd/cgi-bin/library`

Notes

- \$GSDLHOME is the folder where you have installed Greenstone for example if it is installed in /home/GreenStone2 so you will replace \$GSDLHOME by /home/GreenStone2 in the ScriptAlias and Alias of httpd.conf
- You have to restart Apache to test it, as root user or using sudo type the command

```
/etc/init.d/apache2 restart
```

2.4 Librarian Interface

The librarian interface is the main entry for creating or manipulating existing collections, assigning metadata and configuring the classifiers and indexes for collections

There is a librarian interface on your library site but the web-based librarian interface is not supported since the version 2.4 of Greenstone, the main control interface for

organizing your collections is gli or Graphical Librarian Interface, which needs java (JVM) installed on the server

A complete documentation for the librarian interface is available on Greenstone's web site at <http://www.greenstone.org/manuals/gsd12/>, but I will try to give a simple description of using it with focusing on the important points.

The procedure to launch the librarian interface using local windows version is to choose start-->programs--> Greenstone--> librarian interface (gli)

Using librarian interface on a remote server is described below

The connection to the remote server hosting Greenstone

The users of ICDO will have Greenstone installed on a remote server, mainly Linux server, so they have to use a secured shell protocol (ssh) to login to the sever and launch the librarian interface

A user has to be created previously on the server; this user has to be permitted to write in Greenstone installation folder

If you use Linux to make a connection with the server use the following command:

```
$ ssh {servername} -l {username} -p {port number}
```

A password will be asked, type your password and now you are connected to the server

To activate the librarian interface you have to type the following commands

```
cd /$Greenstone/gli/
```

Where \$Greenstone is the path for Greenstone installation folder

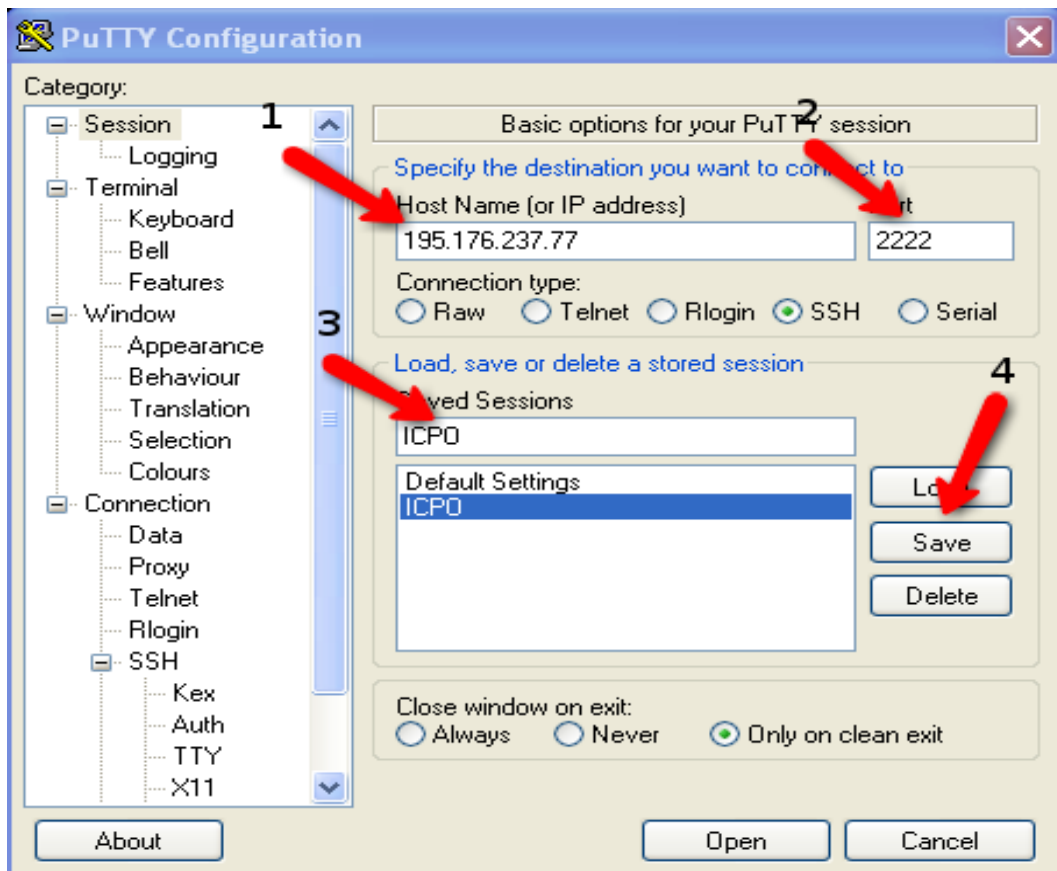
```
./gli.sh
```

If you are using Windows to the connection with the remote server you will need three additional tools to use gli on remote Linux server

- 1- ssh connection tool the recommended free one is Putty downloadable on <http://www.putty.org/>. that you will use for the connection to the server and launching the gli (graphical librarian interface)
- 2- A graphical X window server for MS Windows, a good tool is Xming downloadable on <http://sourceforge.net/projects/xming> this client application is for the communication with the X window server on the Linux server
- 3- File transfer tool like FileZila that you will use to transfer files needed, for building your collections, from your pc to a file on the server

The following steps describe how to launch the Librarian interface on the server from your pc:

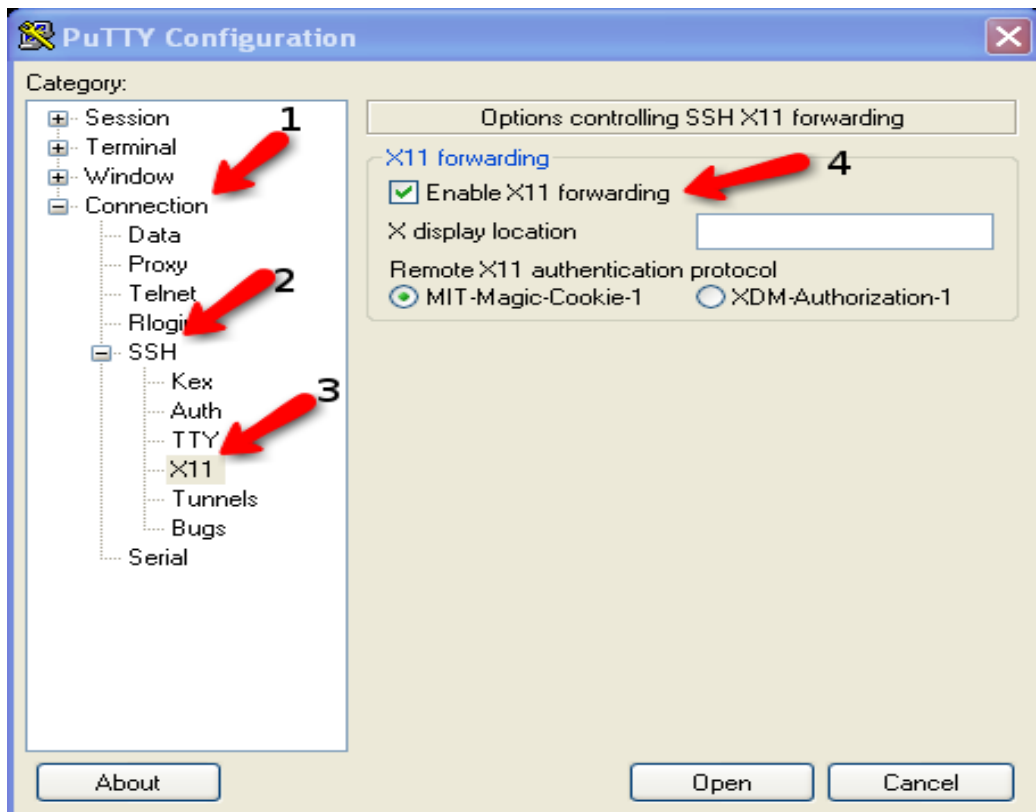
- 1- Install and launch Xming on your pc
- 2- Launch putty and configure a session using the server IP or name and the port number (default 22 for SSH protocol), you also have to enable the X11 forwarding to Graphical client, which will communicate with the X window server on the remote Linux server.



The

screen shots describing Putty configuration

The following screen shots shows how to enable X11 forwarding



After the configuration of a session, get back to Session category and save it for a next usage and click open to see a terminal, where you will have to type the user name and then the password.

- 3- Once you got the command line you have to change the working folder to Greenstone/gli folder; to do it type the command

```
cd /Greenstone/gli/
```

Then execute the gli.sh as follow

```
./gli.sh
```

Now you can start using Librarian interface to modify or build a new collection.

After launching the Librarian interface you can count on annex 1 for the details of using Greenstone to build your library

3. Needs of ICDO and Greenstone Digital Library

According to the specifications the ICDO's expectations of Greenstone are:

- Manage publication's keywords and categories on the server
- Delegate to some members the right to upload publications o the server
- Make publications available on internet

Greenstone Digital library is very good choice for the needs of the ICDO. During the training phase in collaboration with Mr. ATWAN we tested a locally installed version of Greenstone; we realized that the features of Greenstone digital library cover largely the specifications of the ICDO.

Loading publications on a server

The librarian interface is the main method for loading publications on it, also it is possible to do it using the java applet form website of your library but this feature of Greenstone still needs enhancements because it is not yet 100 % operational. We hope that would be resolved by the Greenstone team with the incoming version.

Manage publication's keywords and categories on the server

A special attention of Greenstone developers was given to metadata because generally the richer the metadata available for a collection the more options there are for building your digital library

A complete system of metadata sets management already exists with Greenstone 2.81, the default set of metadata is the Dublin core and it is possible to add or remove elements to an exiting set or creating a new fully personalized set of metadata.

Each metadata set has a namespace, (e.g. dc for Dublin core),when creating a new set of metadata you have to choose a namespace for your set, this name space will always proceed the metadata element's name (e.g. dc.Title for the metadata Title)

While creating a collection, for each one of the records included in the collection Greenstone will try to collect metadata values that the user would have typed, if an element has no value Greenstone will try to extract metadata values from the record

itself, for this reason after building a collection we may find additional metadata proceeded by the name space ex for extracted

Delegate to some members the right to upload publications o the server

ICDO library team can delegate the task of loading files or creating collections on the server by supplying these members the procedure of the connection to the server to use the “Librarian interface”.

Make publications available on internet

The library web site of Greenstone can be customized and basic customization of the default Greenstone “look and feel” could be done like changing the logo or the homepage text, but for advanced customizations, technical skills are needed and a Greenstone special macros language should be used.

While configuring a collection we have the choice to make it public for net user or keep it private which is recommended for commercial records with copy rights

Project Progress and problems met

The project started the 3rd of March 2009 and ends the 19th may 2009 after the final presentation of it.

Reserved hours for the project:

- Monday 13:00 to 16:00
- Tuesday 08:00 -12:00 and 13:00 -17:00
- Wednesday 08:00 -12:00 and 13:00 -17:00
- Thursday 08:00 to 12:00
- Friday 08:00 -12:00 and 13:00 -17:00

The table 1 shows the steps of the project, the duration for each step:

Table 1

Steps of the project

Task	Hours	Contacted person	Milestone?
Project plan and management (Check points)	26	Mr. David BILLARD	non
Business analysis	80	Mr. Rayidh ATWAN	yes
Use and test Greenstone's different features	60		yes
Research on library workflow (Catalogue & index)	40	Mr. Alexander BODER	yes
Installing and configuring Greenstone(test sever)	14	Mr. Cyril Déchelette	yes
Staff training	80	Mr. Rayidh ATWAN	yes
Redaction and presentation	60		Yes

3.1 Project Progress

Project planning and management (Check points):

The planning of the project and the management of its steps were made under the supervision of Mr. David BILLARD the professor responsible of the project.

Almost always a meeting was held with Mr. BILLARD every Tuesday during the project. The meetings with Mr. ATWAN were held almost every Wednesday, Thursday and Friday during the project. Two meetings were held with Mr. BODER, the 14th and 21st April 2009.

Business analysis:

During the first three weeks of the project I tried to analyze the ICDO business, which means dividing the business into processes, sub processes, and finally tasks, during the meeting with Mr. ATWAN I tried to understand the workflow applied by ICDO library.

In parallel I started to read about and test Greenstone Digital Library, for that purpose I installed a ready to use Windows version of Greenstone on my Laptop.

The conclusion after these three weeks is:

- The library service or activity should be a new one in ICDO as no workflow noticed during the business analysis; to implement this activity in ICDO a library workflow is required.

Research on library workflow (Catalogue & index)

The following step was reading about library workflow and contacting a specialist in this domain, who was Mr. BODER

Two meetings were organized with Mr. BODER whose explanations and advices about library workflow were greatly precious

During the first meeting the following points were discussed:

- Document life cycle
- Document type

- Selection of documents
- Acquisition of documents
- Cataloguing, indexation, the thesaurus
- A method for analyzing the needs of ICDO library

The second meeting was held in presence of Mr. ATWAN and the discussed points were:

- Cataloguing and indexing
- A personalized classification system that Mr. ATWAN had made
- Business analysis of the ICDO
- The possibility of making reports with Greenstone
- Plug-in Greenstone
- Thesaurus

The conclusion after this step was chapter 4 of this report or “Library building processes of ICDO”

Use and test Greenstone’s different features

Generally building a collection under Greenstone is based on metadata set, Dublin Core is default set upon which the software is based.

During that step the following tasks were identified and studied:

- Installing Greenstone on Linux server and configuring the server, which makes Greenstone collections available for on-line consultation
- Metadata and the basic Dublin Core
- Creating collection and managing the metadata sets
- Configuring indexes and classifiers which will be shown as options for the final user of Greenstone
- Personalizing Greenstone on line interface, the home text and the metadata to be shown.

Installing and configuring Greenstone (test sever)

I installed the software on Linux version (Ubuntu), the installation has gone without problems and Greenstone was operational in a short time.

But the real test server is administrated by HEG IT center, the installation was done by Mr. Cyril Déchelette, the procedure of installation started the second week of the project but some administrative procedures (security issues) have delayed the resolutions of some problems that appeared during the installation, finally these problems were resolved after many e-mail and after five meetings with Mr. Déchelette

The test sever is operational as of 5th Mai 2009

Staff training

The problems met with the server did not disrupt the staff training step thanks to Greenstone

The training of Mr. ATWAN started with library workflow step (03.04.09), a Windows version was installed on a computer at ICDO, training continues to the end of the project and he is already trained on:

- Document scanning and OCR process
- Join documents to a collection and organizing documents in the collections
- Metadata and Dublin Core and inherited metadata
- Creating of collection and managing the metadata sets
- A simple customizations of the final interface
- Configuring indexes and classifiers
- Connection to the test server and creating a collection on it

3.2 Problems met

1 - The absence of the business structure, because the ICDO is trying to create a new service or activity.

The main problem that I have met during the project and particularly at the beginning is the absence of the business that I would have to analyze.

The method applied at the beginning of the project was identifying the business Use-cases; use-case is a kind of textual form of the work procedure or workflow, as the library workflow was not implemented at ICDO I had to restart the analysis from the beginning after the 3rd week of the project, the continuation was identifying the library processes that the ICDO library service will use.

2 - A problem of delay for setting up Greenstone on a test server at the HEG site.

A request of setting up Greenstone on a Linux server was communicated by Mr. David Billard, the professor in charge of this project, to the IT centre of the HEG at the first week of March, the Greenstone installation became operational only the 5th of May, 2009, fortunately Greenstone has an out of box Windows version that we used for testing and the training of Mr. ATWAN

Conclusion

We demonstrated and setup a whole digital library infrastructure, using Greenstone as library software, the scanning of paper, books and publications of ICDO and the treatment of these images by an OCR (Optical Character Recognition) was demonstrated. The ICDO now possesses all the technical tools and necessary knowledge in order to manage efficiently their publications and make the information about civil defence available on-line.

The choice of Greenstone Digital Library was a real success for several reasons, Greenstone is: open source and free tool, practical and easy to use, support almost all kind of digital files, multiplatform, Multilanguage, can be configured to harvest metadata from other server or databases through OAI-PMH (Open Archives Protocol), and finally Greenstone developers continue to enhance and support the tool.

Cornerstone of library system was identified and presented to ICDO during the analysis of their needs; they can continue to improve the proposed library processes to attain high quality library service, by joining to the library service a qualified staff and furnishing it by needed materials (documents ...etc.)

The project team hopes that sooner the ICDO library system become an essential source of knowledge and information and a reference in the world of on-line libraries, for the interested entities in civil defence subjects.

We hope that the collaboration between ICDO and HEG progress more and more and give its fruits continually.

References:

- 1) Basic Dublin Core Semantics DC 2006 Tutorial 1, October 2006 Marty Kurth
Head of Metadata Services Cornell University Library
- 2) Cataloguing and organizing digital resources 2004 : a how -to-do-it manual for
librarians / Anne M. Mitchell and Brian E. Surratt; no 139
- 3) Guide de l'information numérique Comment traiter les données lisibles par
machine et les documents numériques ; Office des publications des
Communautés européennes 1997.
- 4) IIMK's Experience with Greenstone in Building Digital Library Collections,
tutorial by Dr. M.G. Sreekumar Centre for Development of Digital Libraries
(CDDL) Indian Institute of Management Kozhikode (IIMK)
- 5) http://wiki.greenstone.org/wiki/gsdoc/others/Greenstone_history.htm
- 6) <http://www.greenstone.org/>
- 7) <http://dublincore.org/>
- 8) <http://paperspast.natlib.govt.nz/cgi-bin/paperspast>
- 9) <http://www2.odl.ox.ac.uk/gsdll/cgi-bin/library>
- 10) <http://library.un.org.pk/gsdll/cgi-bin/library>

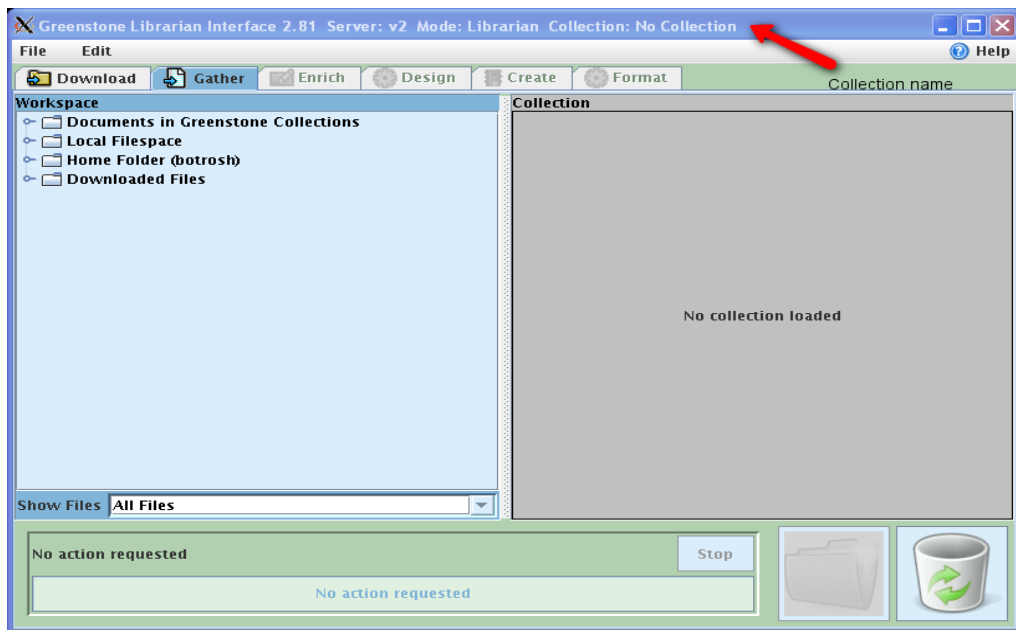
Appendix

Librarian interface User manual

The first step of using the Librarian interface is to choose a collection to modify or to create a new one; the window name will show the actual opened collection

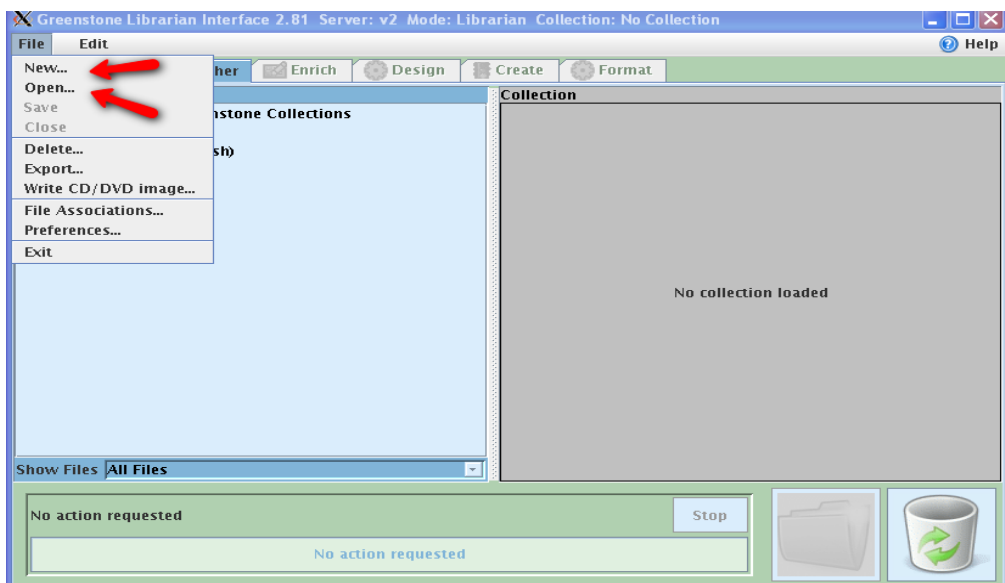
Figure 1 shows this step

Figure 1 collection loading



choose "File" in main menu then "New" or "Open" for new collection or open an exiting one, this is shown on figure 2

Figure 2 start work with a collection



After opening or creating a collection you can use the main window in which we have five tabs “Download”, ”Gather”, ”Enrich”, “Design” ,”format”

Download tab

It is possible to harvest metadata about records with the “Download tab”, the protocols used for communications and pumping metadata from other servers are:

- *OAI*: This protocol is based on Dublin Core and XML
- *Z39.50*: The protocol Z39.50 refers to the standards ISO 23950 and ANSI / NISO Z39.50. It acts as a protocol of IT communication customer-server, which allows sending queries through a computer network in various databases. This protocol is mainly used by libraries to question several catalogues simultaneously
- *SRW*: Search and Retrieve Web Service, it uses CQL (Common Query Language), it is the evaluation of z39.50, maintained by the Library of the Congress and intended to facilitate the search on Internet
- *Web*: Downloads from ftp or http server
- *MediaWiki*: Downloads from a Wiki based technology server

This feature of Greenstone is rarely used because most of the above mentioned protocol is under enhancement

Use to the options of this tab:

- Download records to a folder on your local file system and use these records in your collection
- Extract data from other servers and this feature is an advanced stage of the digital library that needs a lot of preparations and agreements with each server’s administrator

Gather tab

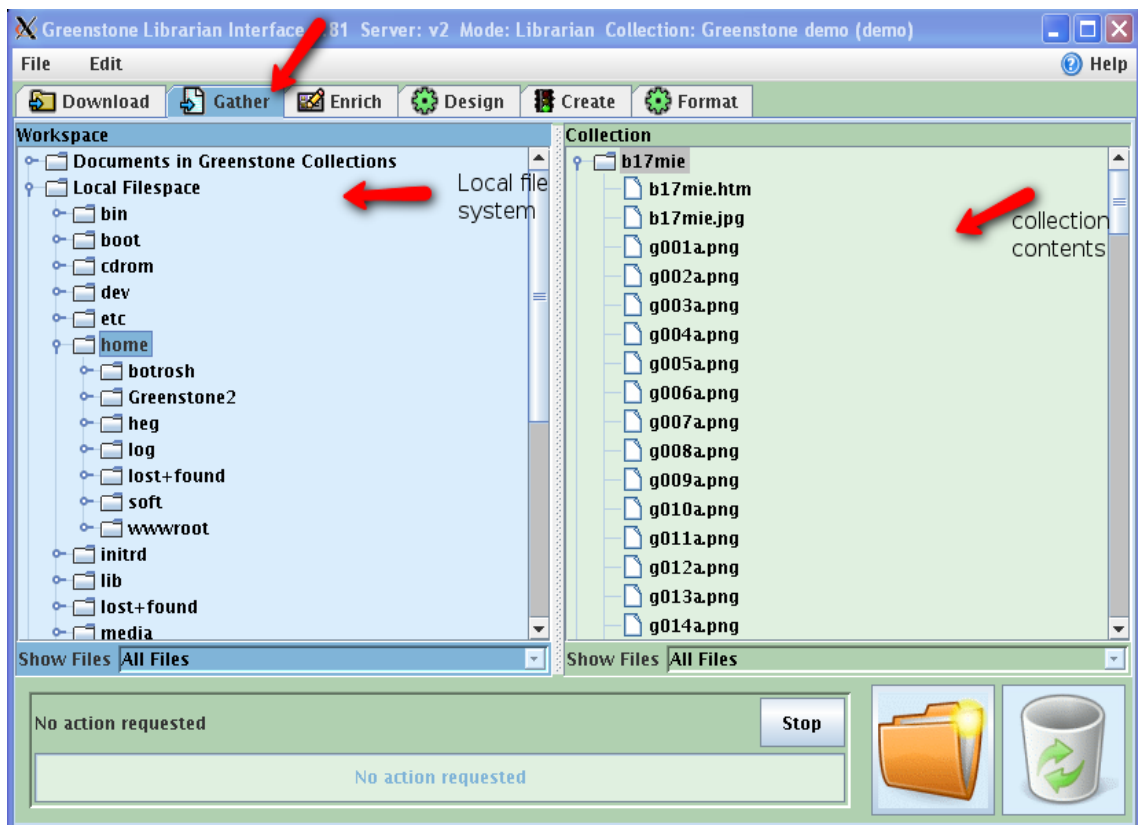
Using this window will enable the user to gather the records that will compose the opened collection

On the right side of the tab you have the contents of the collection, you can drag & drop files from the left hand side, the file system or the user work space, to put it in your collection.

In addition to the records added to your collection you can create a dummy folder for virtual record that will only contain the metadata, go to the “Enrich” tab to do it, or for some records like CDs or DVDs, also you can create folder for each group of files, this is practical for facilitate typing the metadata because when typing metadata for a folder, files inside this folder will inherit the metadata from the container folder so we can type common metadata for the contained files in the metadata fields of the container folder then type specific metadata for each contained file

Figure 3 shows the “Gather” tab

Figure 3 join documents to your collection

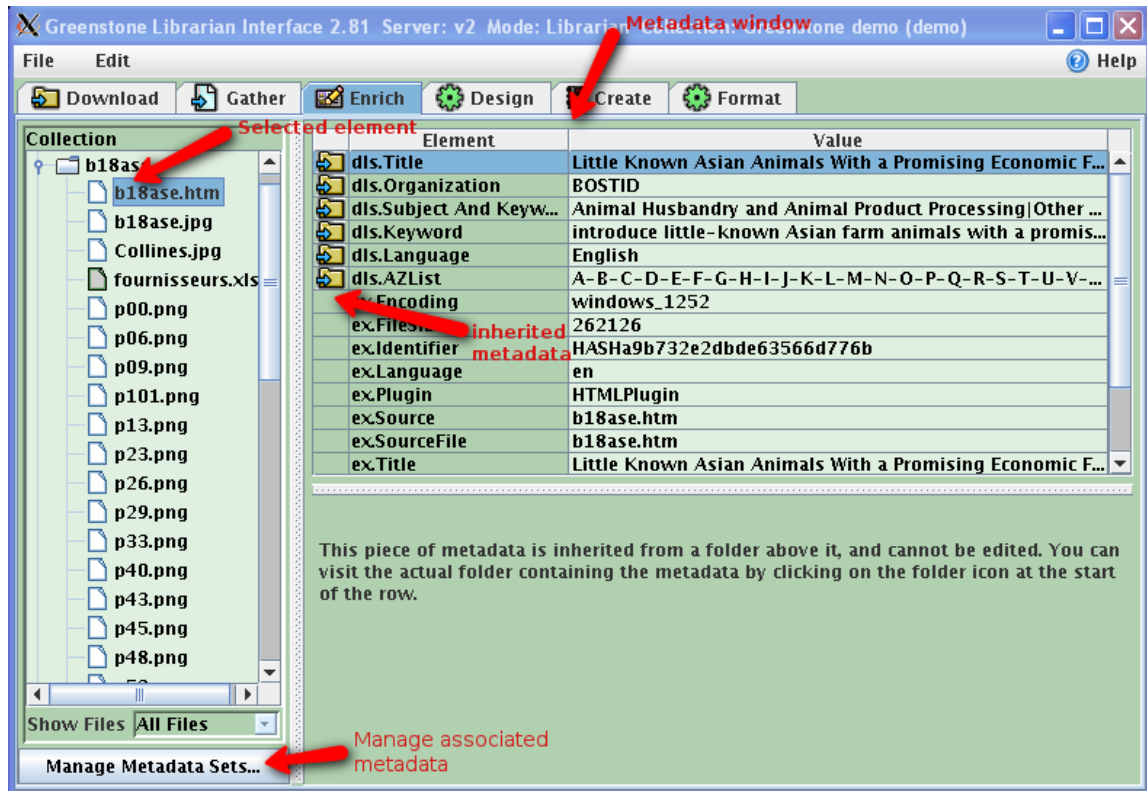


Enrich tab

After finishing gather all records needed for your collection pass to the “Enrich tab”. Here you can type the metadata for the files contained in your collection, to do it choose a file from the collection to activate it then metadata associated to the file will be shown on the right side of the window

Figure 4 shows “Enrich tab”

Figure 4 type and manage meta data



The “Enrich” step is the most important operation in the creation of your collection because the final web interface and the functionality of your digital library depend on what you type in the window of metadata, paying very attention to metadata will avoid you a lot of time wasted in finding what is not working in your collection

Other important item that is the name space, which is an abbreviation of the assigned metadata set for the selected record, for example:

“dc.Title” means the metadata “Title” of the metadata set Dublin Core.

The default set of metadata is Dublin core with its 15 elements, it is not recommended to delete this set of metadata but to create and personalize a new one

The value fields are optional but it is recommended to assign a value for the elements that you will use to construct indexes and classifiers on, if nothing is assigned to a value Greenstone will try to extract metadata value for this element, to do that Greenstone will analyses the contents of the file so errors may occurred and it is not recommended to build your indexes and classifiers on the extracted values.

A metadata set is automatically created after the building of collection, the default name for this set of metadata is “Extracted” and the name space is “ex”, this set can not be modified or deleted

A very interesting feature of Greenstone is the possibility to manage metadata sets
Figure 6 shows

By clicking on Manage Metadata Sets an applet (window) will be opened showing the assigned metadata set including the options "Add", "Modify" and "Remove"

By choosing the option "Add", another applet will show a list of available metadata sets on your server and the options "Add", "New", "Browse" and "cancel"

- Add: To assign the selected set to the actual collection
- New: To create a new set of metadata, this action will lead you to another window where you type name, name space and description for your new set of metadata, also you can base your new set on an existing one, then in a new window add/remove statements(elements) to this new set
- Browse: To import a set of metadata file (extension .msd)

Once a set of metadata assigned to a collection you will find the elements of the added set among the metadata list available for the collection.

Design tab

Using this tab will enable you to configure the research indexes this appear in the drop down list on the user interface, In other word the search options proposed to the user on your library interface.

For example if you chose the following elements as index

Title

Subject

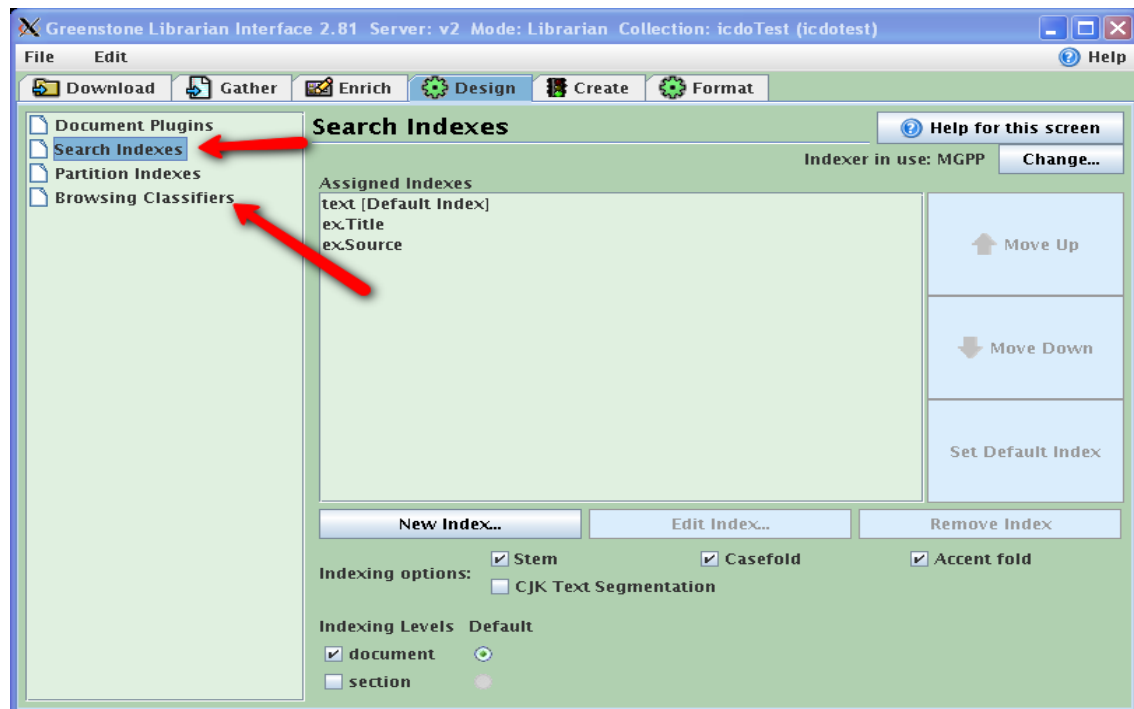
Author

If the user will chooses Subject of the above elements and type a word or a group of words, so Greenstone server will find the subjects that match the searched word(s).

You can configure indexes for all elements of the set of metadata affected to your collection

In this interface you will also find and configure the classifiers, the same thing like indexes you can configure a classifier for each element of the set of metadata affected to your collection. Figure 5 shows this step

Figure 5 indexes and classifiers



Create tab

After gathering the needed documents (gather tab), typing the metadata for each element (enrich tab) and configure the indexes and classifiers (design tab), now you can build your collection with the “create tab”

Two options are available here

Complete build or minimal build, choose the first the new collections and the seconde when adding documents to an existing collections

Format tab

This interface let you configure the format the display of the output or result of research but you need to have some html knowledge for basic modifications and special macro language (special language for Greenstone macro) for advanced modifications