LAB 1:

Greenstone: Installing, browsing, building

Participants are encouraged to read the following Greenstone Manuals from the Manuals Section of the Programme CD:

- **Greenstone Digital Library Installer's Guide**
- **Greenstone Digital Library User's Guide**
- **Greenstone Digital Library Developer's Guide**
- **Greenstone Digital Library: From paper to collection**
- **Other Documentations of Greenstone**

N.B: *CD-ROMs with Greenstone version 2.62 or earlier also include the* ***Greenstone Language Pack****, which gives reader's interfaces in many languages (currently about 40). This has its own installer which you have to invoke separately, after you have installed Greenstone. CD-ROMs with version 2.70 or later now come with reader's interfaces in all available languages. Textual images have been removed from the interface; they are now done using CSS (Cascading Style Sheets). The Greenstone Language Pack is no longer needed. Instead, these CD-ROMs come with the* ***Classic Interface Pack****, which contains the old text images for use with a backwards compatibility macro file.*

There are various ways of getting Greenstone:

1. From a UNESCO CD-ROM (version 2.70) (or FAO IMARK CD-ROM, but this is an earlier version 2.51)

   These CD-ROMs contain the **Greenstone software**, plus **documented example collections**, four **language interfaces** (English French Spanish Russian), the **Export to CD-ROM** package, the **ImageMagick** graphics package, the **Java runtime environment**, and an **installer** that installs all of these.

2. From the IITE Digital Libraries in Education CD-ROM, or a Greenstone workshop CD-ROM

   In addition to all the above software, these CD-ROMs contain the tutorial exercises and a set of ***sample files*** *to be used for these exercises. CD-ROMs with Greenstone version 2.62 or earlier also include the* ***Greenstone Language Pack****, which gives reader's interfaces in many languages (currently about 40). This has its own installer which you have to invoke separately, after you have installed Greenstone. CD-ROMs with version 2.70 or later now come with reader's interfaces in all available languages. Textual images have been removed from the interface; they are now done using CSS (Cascading Style Sheets). The Greenstone Language Pack is no longer needed. Instead, these CD-ROMs come with the* ***Classic Interface Pack****, which contains the old text images for use with a backwards compatibility macro file.*

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

All these CD-ROMs contain the full Greenstone software, which allows you to view collections and build new ones. They are not the same as CD-ROMs that contain a pre-packaged Greenstone collection, which only allow you to view that collection.

3. You can also get Greenstone and its associated software from http://www.greenstone.org as well as at http://greenstonesupport.iimk.ac.in.

   Most people download the Windows distribution from http://www.greenstone.org, which contains the latest version of Greenstone. There are several optional modules that must be downloaded separately (to avoid a single massive download): **documented example collections**, the **Export to CD-ROM** package (Greenstone 2.70 and earlier), the **Language Pack** (Greenstone 2.62 and earlier) and **Classic Interface Pack** (Greenstone 2.63 and later). There is also the set of **sample files** used in these exercises. (To reduce the download size the documented example collections are distributed in unbuilt form and need to be built.)

   To work with image collections, you need **ImageMagick** (from http://www.imagemagick.org).

If Greenstone has been installed on your computer before, you should completely remove the old version before installing a new one. (However, you need not remove any pre-packaged collections that you may have installed.) To do this, see **Updating a Greenstone installation**.

Here is what you need to do to install Greenstone. Older versions of the installer follow much the same sequence but use slightly different wording.

1.1. Installing Greenstone

Installing Greenstone on a Windows system

A graphical tool for greenstone collection building called the Greenstone Librarian Interface (GLI) which requires 1.5.0 or later version of the Java Runtime Environment (JRE). So you need to install JRE also, before installing Greenstone for getting the best out of the software.

I. Installing Java Run-time Environment (JRE) Version 1.5.0-05

You need **Java** to run Greenstone. You might already have it; otherwise download it from http://java.sun.com or install if from the Programme CD.

1. Locate The file **jre-1_5_0_05-windows-i586-p.exe** on the CD
2. Start installation by double clicking on this "**setup" icon**

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

3. Click on **<Yes>** to accept license agreement
4. Click on **<next>** to select default setup type, which is **'Typical'**
5. JRE gets installed to default directory, **c:\program files\java**
6. Test The installation by executing the following command in MS_DOS prompt

   C:\Java –version

The above command generates three lines of output indicating JRE and its version.

II. Installing Greenstone Version 2.83 on Windows

Insert the Programme CD to the Drive and select "**Software**" button from the **Menu** in the popping up page, and click on the "**Install**" button of the Greenstone 2.83.

- Select the language for this installation. We choose **English**
- Welcome to the Greenstone Digital Library Software Installer. It is recommended that you uninstall any previous installations of Greenstone2 before running this installer Click <**Next**>
- License Agreement. Click <**Accept**>
- Choose location to install Greenstone. Leave at the default and click <**Next**>
- Components. Click the question mark button on the right of each component will display the description of this component in a popup window. Leave at the default (all components are selected) and click <**Next**>
- (For older installers you must now select collections. Leave at the default, Documented

  Example Collections, and click <**Next**>)
- Enable administration pages. Read the description on this page, if you check to enable, click <**Next**> to set admin password. Choose a suitable password and click <**Next**> (If your computer will not be serving collections online, the password doesn't matter)
- Click <**Install**> to start the installation. Click <**Show Details**> to show the details of this installation
- Files are copied across
- Installation is complete.
- We have changed to a new installer in 2008. This installer will automatically install ImageMagick and GhostScript.
- If you are installing from a CD-ROM, the installer will offer to install ImageMagick (see below), and Java, if necessary.

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

To invoke the Greenstone Reader's interface, go to the *Greenstone Digital Library Software* item under *Programs* on the Windows *Start* menu and select *Greenstone Digital Library*. To invoke the Librarian interface, go to the same item and select *Greenstone Librarian Interface*.

III. Installing Apache Webserver on Windows

- Note: Stop other web servers like **IIS**, if running.
- Run file ***apache_2.2.9-win32-x86-no_ssl-r2.msi*** from Programme CD
- Double Click the file and start installation
- Agree License Agreement
- For local installation give Network Domain as **localdomain** and server name as **localhost**
- Select **Typical** installation and install to the default location
- Finish the installation and ensure the Apache2 service is running.

Configuring Apache Webserver for Greenstone 2.83

To use the Greenstone Web Library you will need to have a webserver installed. The following information is intended to help you configure your webserver for use with Greenstone. Particular attention is given to the Apache webserver which may be downloaded free from www.apache.org. Greenstone will work with other webservers too if that is your preference.

Greenstone has been set up to expect C:/Program Files/Greenstone to be accessible from your webserver at the URL http://localhost/gsdl (localhost will of course be the usual web address of your machine if it has one). Likewise, the C:/Program Files/Greenstone/cgi-bin directory must be accessible as a cgi executable directory.

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

ScriptAlias /gsdl/cgi-bin "C:/Program Files/Greenstone/cgi-bin"
<Directory "C:/Program Files/Greenstone/cgi-bin">
Options None
AllowOverride None
</Directory>
Alias /gsdl "C:/Program Files/Greenstone"
<Directory "C:/Program Files/Greenstone">
Options Indexes MultiViews FollowSymLinks
AllowOverride None
Order allow,deny
Allow from all
</Directory>

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

In addition to this, in order to get the webserver to work with Greenstone, you will have to edit the gsdlsite.cfg file in the cgi-bin subdirectory. Rename gsdlsite.cfg.in to gsdlsite.cfg and edit the file. It is advised to take a back up of the file before renaming. Ensure that the gsdlhome parameter points to the Greenstone home directory (e.g. gsdlhome**"C:\Program Files\greenstone"**). See that the path is given in double quotes without fail. Also set the httpprefix and httpimg parameters as default if the webserver's DocumentRoot is set to Greenstone home directory. Otherwise, give the parameters as "**httpprefix /gsdl**" and "**`httpweb/gsdl/web`**" if the alias for the home directory is gsdl. Remove the "#" from "httpweb/gsdl/web" also

Once your webserver and greenstone are configured in this way you can access Greenstone by pointing your web browser at http://localhost/gsdl/cgi-bin/library.cgior http://IPaddressOfgsdlServer/gsdl/cgi-bin/library.cgi

For the Lab Exercises we will work with Local Library Mode of Greenstone.

1.2. Updating a Greenstone installation

These tutorial exercises assume that you are using Greenstone 2.60 or above.

Before updating to a new version of Greenstone, ensure that the computer is not running the Greenstone Librarian Interface or the Greenstone local library server. Normally, quitting your web browser, or quitting the Librarian Interface, also quits the server.

Removing Greenstone from a Windows system

1. Ensure that you are not running Greenstone.
2. If the installed Greenstone version is 2.81 and above, to remove the old version, go to the Greenstone home directory (eg. C:\Users\<username>\Greenstone2 by default, where <username> is your user name) and click **Uninstall.bat**. Otherwise, if the version is lower than 2.81, remove the old version by going to the Windows Control Panel (from the *Settings* item on the *Start* menu). Click **Add or Remove Programs**, select **Greenstone Digital Library Software**, and **Remove** it. (To do this you may need Windows "Administrator" privileges.)
3. For version 2.81 and above, the uninstaller has an option for keeping all your Greenstone collections, leave it at default as selected. For versions lower than 2.81, at the end of the uninstallation procedure you will be asked whether you would like all your Greenstone collections to be removed: you should probably say *No* if you wish to preserve your work

Occasionally, problems are encountered if older Greenstone installations are not fully removed. To clean up your system, move your Greenstone collect folder, which contains all your collections, to the desktop. Then check for the folder C:\Program Files\gsdl or C:\Program Files\Greenstone or C:\Users\<username>\Greenstone2 for version 2.81 and

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

above, which is where Greenstone is usually installed, and remove it completely if it exists.

Reinstalling Greenstone on a Windows system

4. The reinstallation procedure is exactly the same as the original installation procedure, described in **Installing Greenstone**. If you already have ImageMagick, you do not need to install it again.

Amalgamating different Greenstone collections

5. If you have previously installed the Greenstone Digital Library software in a non-standard place, you should amalgamate your collections by moving them from the *collect* folder in the old place into the folder *Program Files\Greenstone\collect*.
6. If you have installed collection from pre-packaged Greenstone CD-ROMs, they reside in a different place: *C:\GSDL\collect*. To amalgamate these with your main Greenstone installation, move them into the folder *Program Files\Greenstone\collect*. The mini version of Greenstone that is associated with the pre-packaged collections is no longer necessary. To uninstall it, select *Uninstall* on the Greenstone menu of the Windows *Start menu*

Installing the Greenstone language pack (2.62 and earlier)

*If you go to the Preferences page of any Greenstone collection, and look at the **Interface language** menu, you will probably find that only English, Spanish, French and Russian interfaces are installed.*

7. Locate the Greenstone Language Pack (glp-x.xx.exe/glp-x.xx-linux.bin/gli-x.xx-macOSx.command). This may be on the CD-ROM from which you installed Greenstone, or you may have to download it from http://www.greenstone.org.

8. Run the executable file (double click it on Windows); this will start the installer. Accept all the defaults
9. Restart the Greenstone Digital Library and look at the interface language menu again. Now you should see about 40 different languages.

Enabling other languages (2.63 and later)

If you have downloaded Greenstone from the web, then all the languages will be enabled by default. However, if you have installed Greenstone from a UNESCO CD-ROM, then only English, French, Spanish and Russian will be enabled.

10. To enable a new language, edit the file *greenstone →etc →main.cfg*. Look for the appropriate "Language" line, and uncomment it (i.e. remove the # from the start). Check that the required encoding is also enabled.

    For example, suppose that we want to enable Turkish. The "Language" line for Turkish

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

looks like:

```
#Language shortname=tr longname=Turkish default_encoding=windows-1254
```

To enable it, we remove the #, i.e. make it look like:

```
Language shortname=tr longname=Turkish default_encoding=windows-1254
```

The default encoding for Turkish is windows-1254. So we look for the windows-1254 Encoding line:

```
Encoding shortname=windows-1254 "longname=Turkish (Windows-1254)" map=win1254.ump
```

This is already enabled (no # at the start) so we don't need to do anything else.

Installing the Classic Interface Pack (2.63 and later)

Greenstone now comes with all languages enabled. The generated HTML uses text + CSS rather than images for navigation bar, home, help, preferences buttons etc. The classic interface pack is not needed if you want to use Greenstone in another language. It is only needed if you want to revert back to the old style HTML with text images. This may be useful if you have customized your Greenstone, or if you require compatibility with Netscape 4.

11. Locate the Classic Interface Pack (gcip-x.xx.zip). This may be on the CD-ROM from which you installed Greenstone, or you may have to download it from http://www.greenstone.org.

12. The classic interface pack is a zip file containing the old text images, such as classifier buttons. Unzip the zip file into the images directory of your Greenstone installation.

13. Enable the use of the old-style macros by editing *greenstone → etc → main.cfg*: replace "nav_css.dm" with "nav_ns4.dm" in the "macrofiles" list.
14. Restart the Greenstone Digital Library. It should now be using the old text images.

IV. Building Collection using the GLI tool

The Greenstone Librarian Interface (GLI) is an easy-to-use front-end to Greenstone's collection-building functionality. It provides a graphical, point-and-click interface that allows you to gather files for your digital library collection, assign metadata to them, and then design, customize and build your collection. The Librarian Interface comes as standard in all. It is installed in a subdirectory of your Greenstone installation called "\gli", and requires a recent version of Java (JRE – Java Run-Time Environment) to run.

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

Accessing the Greenstone Librarian Interface (GLI)

Start > **Programs** > **Greenstone-2.83** >**Greenstone Librarian Interface (GLI)**

1. Wait for a while – it takes a few seconds to get the module ready.
2. From **GLI**, select **File** > **New**

   A window will pop up. Give appropriate values

   Collection title

   Description of Content

   Leave the settings for ***Base this collection on:*** at its default ***New collection***

   And click **<*OK*>**

3. You need to now **gather file/s** that will constitute the **collection**. The Sample Files are available at the Programme CD.
4. **Drag and Drop the required file/s OR folder/s** from the **Workspace** to the **Collection Area.**
5. You can see the file contents by double clicking on the file in the **Collection Area**.
6. Now go to Create Panel by clicking the **Create Tab**.
7. To start building the collection, simply click **<Build Collection>** at the panel.
8. Once the collection has built successfully, a window pops up. To confirm this, Click **<OK>**

10. Click on **Preview Collection** button to look at the end result.

Features of the GLI (The 'Gather', 'Enrich', 'Design', 'Create' and Format Panels)

The '**Gather**' Panel facilitates putting the relevant files from the 'workspace' to the 'collection building' area. The **'Enrich'** Panel explains how metadata is created, edited, assigned and retrieved, and how to use external metadata sources. Help for this is provided in the GLI Interface. The **'Design'** Panel facilitates customising your interface, once your files are marked up with metadata. Using the Gather Panel, you can specify the fields that are searchable, allow browsing through the document, facilitate the languages that are supported, and provide the buttons that are to appear on the page. Help for this is provided in the GLI Interface. The **'Create'** Panel facilitates creation of your collection. The

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

Format Panel facilitate the display of documents in the collections, branding collection with images, collection specific customizations etc.

1.3 Building a small collection of HTML files

You will need some HTML files, such as those in the simple_html folder in sample_files.

Running the Greenstone Librarian Interface

1. Start the Greenstone Librarian Interface:

   Start → All Programs → Greenstone 2.83 → Librarian Interface (GLI)

   After a short pause a startup screen appears, and then after a slightly longer pause the main Greenstone Librarian Interface appears. (A command prompt is also opened in the background.)

Starting a new collection

2. Start a new collection within the Librarian Interface:

   File → New...

3. You will create a collection based on a few HTML web pages from the Tudor collection.

   A window pops up. Fill it out with appropriate values—for example,

   Collection title: Small HTML Collection
   Description of content: A small collection of HTML pages.

   Leave the setting for **Base this collection on:** at its default: **-- New Collection --**, and click **<OK>**.

4. Next you must gather together the files that will constitute the collection. A suitable set has been prepared ahead of time in *sample_files → simple_html*. Using the left-hand side of the Librarian Interface's **Gather** panel, interactively navigate to the *sample_files* folder.

Adding documents to the collection

5. Now drag the *simple_html* folder from the left-hand side and drop it on the right. The progress bar at the bottom shows some activity. Gradually, duplicates of all the files will appear in the collection panel.

   You can inspect the files that have been copied by double-clicking on the folder in the right-hand side.

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

6. Since this is our first collection, we won't complicate matters by manually assigning metadata or altering the collection's design. Instead we rely on default behaviour. So pass directly to the **Create** panel by clicking its tab.

Building the collection

7. To start building the collection, click the **<Build Collection>** button.

8. Once the collection has built successfully, a window pops up to confirm this. Click **<OK>**.

9. Click the **<Preview Collection>** button to look at the end result. This loads the relevant page into your web browser (starting it up if necessary).

Viewing the extracted metadata

10. Back in the Librarian Interface, click the **Enrich** tab to view the metadata associated with the documents in the collection.

11. Presently there is no manually assigned metadata, but the act of building the collection has extracted metadata from the documents. Double click the *simple_html* folder to expand its content. Then single-click *aragon.html* to display all its metadata in the right-hand side of the panel. The initial fields, starting "dc.", are empty. These are Dublin Core metadata fields for manually entered data.

12. Use the scroll bar on the extreme right to view the bottom part of the list. There you will see fields starting "ex." that express the extracted metadata: for example **ex.Title**, based on the text within the HTML Title tags, and **ex.Language**, the document's language (represented using the ISO standard 2-letter mnemonic) which Greenstone determines by analyzing the document's text.

13. Close the collection by clicking **File → Close**. This automatically saves the collection to disk.

Viewing the internal links and external links

14. Hyperlinks in a Greenstone collection work like this. If the link is to a document that is also in the collection, clicking it takes you to that document in the collection. If the link is to a document that is *not* in the collection, clicking it takes you to that document on the web.

    Open *boleyn.html* and look for the link to *Katharine of Aragon* (in the 5th paragraph of the *Biography* section). This links to a document inside the collection--*aragon.html*. View this document by clicking the link. For an external link, click *letters written by Katharine* (in the *Primary Sources* section). This takes you out on to the web. If you want

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

a warning message to be displayed first, you can open *Greenstone → etc → main.cfg* file and uncomment the line `cgiarg shortname=el argdefault=prompt`.

Setting up a shortcut in the Librarian interface

15. To set up a shortcut to the source files, in the **Gather** panel navigate to the folder in your local file space that contains the files you want to use—in our case, the *sample_files* folder. Select this folder and then right-click it, and choose **Create Shortcut** from the menu. In the **Name** field, enter the name you want the shortcut to have, or accept the default *sample_files*. Click **<OK>**. Close all the folders in the file tree in the left-hand pane, and you will see the shortcut to your source files.

1.4 A collection of Word and PDF files – Part A

You will need some source files like those in the sample_files → Word_and_PDF folder.

1. Start a new collection called **reports** (**File → New...**) and base it on **-- New Collection --**.

2. Copy all the .doc, .rtf, .pdf and .ps files from *sample_files → Word_and_PDF → Documents* into the collection. There are 9 files in all: you can select multiple files by clicking on the first one and shift-clicking on the last one, and drag them all across together. (This is the normal technique of multiple selection.)

3. Switch to the **Create** panel, and **build** and **preview** the collection.

Viewing the extracted metadata

4. Again, this collection contains no manually assigned metadata. All the information that appears—title and filename—is extracted automatically from the documents themselves. Because of this the quality of some of the title metadata is suspect.

5. Back in the Librarian Interface, click the **Enrich** tab to view the automatically extracted metadata. You will need to scroll down to see the extracted metadata, which begins with "ex.".

6. Check whether the **ex.Title** metadata is correct for some of the documents by opening them. You can open a document from the Librarian Interface by double clicking on it.

7. The extracted Title metadata for some documents is incorrect. For example, the Titles for *pdf01.pdf* and *word03.doc* (the same document in different formats) have missed out the second line. The Title for *pdf03.pdf* has the wrong text altogether.

In exercise 2.1 we correct some of this incorrect metadata by manually adding Dublin Core Title metadata.

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

1.5 A large collection of HTML files—Tudor

You will need the files in the sample_files → tudor folder.

1. Invoke the Greenstone Librarian Interface (from the Windows *Start* menu) and start a new collection called **tudor** (use the **File** menu), based on the default **-- New Collection --**.

2. In the **Gather** panel, open the *tudor* folder in *sample_files*.

3. Drag *englishhistory.net* from the left-hand side to the right to include it in your **tudor** collection. (This material is from Marilee Hanson's Tudor England Collection at http://englishhistory.net/tudor.html, distributed with her permission.)

4. Switch to the **Create** panel and click <**Build Collection**>.

5. When building has finished, **preview** the collection.

Extracting more metadata from the HTML

6. The browsing facilities in this collection (*Titles* and *Filenames*) are based entirely on extracted metadata. Return to the **Enrich** panel in the Librarian Interface and examine the metadata that has been extracted for some of the files.

7. Many HTML documents contain metadata in `<meta>` tags in the `<head>` of the page. Open up the *englishhistory.net → tudor → monarchs → boleyn.html* file by navigating to it in the tree on the left hand side, and double clicking it. This will open it in a web browser. View the HTML source of the page (**View → Source** in Internet Explorer, **View → Page Source** in Mozilla). You will notice that this page has *page_topic*, *content* and *author* metadata.

8. By default, **HTMLPlugin** only looks for Title metadata. Configure the plugin so that it looks for the other metadata too. Switch to the **Design** panel and select the **Document Plugins** section. Select the **plugin HTMLPlugin** line and click <**Configure Plugin...**>. A popup window appears. Switch on the **metadata_fields** option, and set the value to

   ```
   Title,Author,Page_topic,Content
   ```

   Click <**OK**>.

9. Switch to the **Create** panel and **rebuild** the collection. Go back to the **Enrich** panel and look at the extracted metadata for some of the HTML files in *englishhistory.net → tudor → monarchs*. The new metadata should now be visible.

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia

Looking at different views of the files in the Gather and Enrich panels

10. Switch to the **Gather** panel and in the right-hand side open *englishhistory.net → tudor*.

11. Change the **Show Files** menu for the right-hand side from **All Files** to **HTM & HTML**. Notice the files displayed above are filtered accordingly, to show only files of this type.

12. Change the **Show Files** menu to **Images**. Again, the files shown above alter.

13. Now return the **Show Files** setting back to **All Files**, otherwise you may get confused later. Remember, if the **Gather** or **Enrich** panels do not seem to be showing all your files, this could be the problem.

****

Copyright © Greenstone Digital Library Project, University of Waikato New Zealand & Greenstone Support Network for South Asia