Lab 2 (1 3/4 hours): Building collections with the Librarian interface

Part I – Review and additional features

Familiarise yourself with the Demo collection, another example of a pre-built collection:

1. Launch the Greenstone local library server by clicking:

   Start → Programs → Greenstone Digital Library → Greenstone Digital Library

2. Click **<Enter Library>** in the dialog box; your browser will display the Greenstone homepage.

3. Re-familiarise yourself with the features of Greenstone by browsing and searching the Demo collection, which is a small subset of the *Humanity Development Library*.

Additional features:

4. Access this collection's Preferences page by clicking the **Preferences** button at the upper-right corner of the page.

5. Change the interface language to **French** using the pull-down menu adjacent to **Interface language**. The Preferences page will reload automatically in the new language, which will also be used for all subsequent pages.

6. Change the interface language back to **English**.

7. Switch the presentation format from **graphics to textual** by changing the pull-down menu next to **Interface format**. The current page will automatically reload as text only, and subsequent pages will be text only too.

8. Change the interface format back to **Graphics**.

9. Set the **query mode** to **advanced** and the **query box** to **large**; then click **<Set Preferences>**.

10. Access the query page by clicking the **search** button in the navigation bar.

11. Enter a **Boolean** query that searches for **entire documents** on the topic of crocodile or alligator farming.

    [Note: if required, the end of this section shows exactly how to formulate this query.]

12. Peruse the search results list. Do all the documents seem to pertain to the desired topic? Remember, you can load the entire document using the 'expand text' feature.

If you have time, consider the following:

13. Experiment with the use of **search history**, controlled through the **Preferences** page.

    To learn more about searching and browsing, see the on-line, collection-specific, help.

14. If always having to click on <**Enter Library**> after starting the library program seems tedious, you can automate this step through the Greenstone Digital Library Software window. Select **File➔Settings**, switch on **Enter library automatically on start up**, and click <**OK**>. Next time, when you next run the library program the web browser will launch automatically.

    You can use the same **Settings** panel to specify which Web browser to launch for the library, and control which port the server runs on.

For item 11 above, the necessary Boolean search can be entered as:

    (crocodile | alligator) & farming

Part II – Building your first collection with the Librarian Interface

The Librarian Interface is the tool used to create, develop and maintain collections:

1. First make sure you are not already running the Greenstone Local Library server. If you are, quit it first. [**File➔exit** in the Greenstone Digital Library Software window; see Lab 1 for more details.]

2. Start the Greenstone Librarian Interface:

    Start➔All Programs➔Greenstone Digital Library➔Greenstone Librarian Interface

    [After a short pause a startup screen appears, and then after a slightly longer pause the main Greenstone Librarian Interface (GLI) appears.]

3. Start a new collection within GLI:

    File➔New

4. You will create a collection based on a few HTML web pages that describe some Hobbits in *Lord of the Rings*.

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

Collection Title: About Hobbits
Description of Content: A collection about hobbits.

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

5. Another window pops up, from which you select the metadata set (or sets) to use. We discuss this in Lab 3. For now, select **Basic Dublin Core 1.1 (dc)** followed by **<OK>**.

6. Next you must gather together the files that will constitute the collection. These are on the workshop CD-ROM in the folder *html_small*. Using the left-hand side of the Librarian Interface's **Gather** panel, interactively navigate to just above this directory:

Local FileSpace → D → test_files

[This assumes that 'D' is the letter used for the CD-ROM/DVD: on other systems it may be different.]

7. Now drag the *html_small* 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 right-hand panel.

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

9. 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 the **Create** tab.

10. To start building the collection, click **<Build Collection>** at the bottom of the panel.

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

12. Back in the Librarian Interface, 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).

13. Click the **Enrich** tab to view the metadata associated with the documents in the collection.

Presently there is no manually assigned metadata, but the act of building the collection has generated some extracted metadata. Double click the *html_small* folder to expand its content. Then single-click *bilbo.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 (we asked you to include this metadata set when the collection was initially formed) for manually entered data.

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 Title, based on the text within the HTML Title tags, and the document's language (represented using the ISO standard 2-letter mnemonic) which is set by an algorithm that Greenstone uses to analyse the document's text.

14. Close the collection by clicking **File➔Close**. This automatically saves the collection to disk.

We now build a larger collection of HTML files from the CD-ROM's html_large folder:

15. Start a new collection called **tudor** which will contain documents about this period of English history. Fill out the pop-up dialog with appropriate values and choose Dublin Core as the metadata set.

    [The act of starting a new collection or opening an existing one will automatically close the current collection, saving it to disk, so the previous step of closing the collection was not strictly necessary.]

16. In the **Gather** panel, open the *html_large* folder in *test_files* on the CD-ROM:

17. Drag *englishhistory.net* from the left-hand side to the right to include it in your **tudor** collection.

18. Switch to the **Create** panel and click **<Build Collection>**. This collection has more files and takes longer to build than the hobbit one. When it has finished building, preview the collection.

Now look at different views of the files in the Gather and Enrich panels:

19. Switch to the **Gather** panel and open *englishhistory.net➔tudor*.

20. 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.

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

22. Click in the **Show Files** menu, delete the text that is there, replace it with ***.gif**, and click **<enter>**. Now the files above show only the GIF images—the ones whose filename ends in '.gif'.

    Similar behaviour can be applied to the **left-hand** side of the **Gather** panel, and to the **Enrich** panel. Note that the right-hand side of the **Gather** panel is exactly the same as the left-hand side of the **Enrich** panel: any changes you make in one are automatically reflected in the other.

We now experiment with a new collection based on PDF and Word documents:

23. Start a new collection called **reports**, fill out appropriate fields for it, and choose Dublin Core as the metadata set.

24. Copy only the PDF files from *test_files→Word_and_PDF* on the CD-ROM

    [We do this to practice multiple file selection and file filters.]

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

26. Now return to the **Gather** panel and copy the **remaining Word documents** into the collection, then rebuild and preview.

Part III – Extra work

The extra work in this and subsequent labs focuses on the theme of tourism in your country. You will be asked to locate some appropriate content on the web, export or save it locally, and build a Greenstone collection from it. The document types, and what you do in the Librarian Interface, reflect the focus of the current lab.

Find your own content:

1. Download from the Web some HTML and PDF files about tourism in your country, build a collection.

Explore other features of the Librarian Interface in the File menu:

2. File association; preferences.