Lab 1 (1 1/4 hours): Installing and accessing a pre-built collection; building some rudimentary collections

Part I – Installation and access of UNAIDS pre-built GREENSTONE collection

Follow these installation steps:

1. On inserting the **UNAIDS CD-ROM**, for many computers installation will begin automatically. If not, "auto-run"—a configurable setting under Windows—has been disabled on your computer and you need to double-click *setup.exe* on the CD-ROM.

    My Computer → UNAIDS20 → setup.exe

2. The InstallShield Wizard begins to install the UNAIDS pre-built collection. Select the English language.

3. Click the **<next>** button.

4. Choose **[Run from CD-ROM (standard)]** as the setup type. This is the default and is already selected. Then click **<next>**.

5. Click **<next>** again to install the UNAIDS collection in the default folder, which is **c:\program files\UNAIDS Library 2.0 [CD-ROM]**.

    [Installation Wizard copies the required files from CD-ROM to disk]

6. Click **<OK>** to confirm completion of UNAIDS collection (twice).

    [InstallShield quits—the UNAIDS Library is installed.]

Browse around UNAIDS collection:

7. Launch the prebuilt library by clicking:

    Start → All Programs → UNAIDS Library 2.0 (Standard Version).

    [Note: To access Greenstone through the Local Library Server, it is sometimes necessary to turn off the proxy settings of the browser if they are being used. Greenstone normally detects this and pops up a window alerting you to the problem.]

8. Click **<Enter Library>** in the dialog box and your browser (typically Internet Explorer by default) will display the Greenstone home page.

9. Within the Web browser, click **titles a-z** (in the centre of the navigation bar near the top of the page).

10. Access the **first book** in the list of titles by clicking the **book icon** next to the title: *About UNAIDS*.

11. Use the scroll bar to view the full length of the page.

12. In the table of contents near the top, click the "page" icon next to the heading *Guiding principles o UNAIDS* to view this section.

13. Click the page icon next to the heading *Global and local impact* to view the next section.

    [This style of interaction can be continued to further expand and contract folders and to switch to a different section.]

14. To fully expand the contents of this introduction chapter, click **Expand Document or Chapter** in the upper left portion of the page, under the picture of the document's front cover.

15. You can return to the currently selected page of document titles by clicking the **book icon** next to the title of the book at the top of the table of contents (this signifies closing the book). You also get to the document titles using **titles a-z** in the navigation bar, in this case to the titles beginning with A–D.

    [Note: If the table of contents is open at the top level—showing all the chapters—then clicking **Expand Document or Chapter** expands the full document. For long documents, which take some time to load in, Greenstone seeks confirmation for this action: clicking 'continue' loads the full document.]

16. Browse around and peruse some other documents in the collection.

Searching:

17. Access the search page by clicking **search** in the navigation bar.

18. In the query box under **Search chapters in any language which contain some of the words**, enter the term **gender** then click **<Begin Search>**.

    [After a short pause, the Web browser loads a fresh page showing the results of the search.]

19. Click the **page icon** for the **first matching document** in the result set (*Five Year Implementation Review of the Vienna Declaration and Programme of Action*) to view the document. Because the search was at the chapter level, you are taken directly to the matching chapter within the document.

20. Experiment further with searching, and with the interface in general. For example, there is a detailed Help page. It contains a Preferences section through which you can control some search settings.

The Preferences options in the UNAIDS collection are intentionally minimalist. Most collections have a separate Preferences button that offer more features—we will see an example in Lab 2 when browsing around the Demo collection that is supplied with the Greenstone software.

The home page of the UNAIDS library collection cycles through a sequence of front cover images, updated every 5 seconds or so. Clicking a particular image takes you directly to that document.

Leaving the UNAIDS Greenstone library:

21. There are two ways of leaving Greenstone:

    a. Exit from the Greenstone Software server. Click on the Greenstone Software in the task bar, then choose **Exit** from the **Browser Selection and Settings** menu (or click on the exit hotspot, the red cross at the top right). The Greenstone Software exits, but your web browser continues to run.

    b. Exit from your web browser. Leave your web browser in the usual way. The Greenstone server detects when you exit from the browser and generates a popup window that asks whether to close down the server as well. (The reason is that other people may be using Greenstone over the network, and should not be rudely terminated.)

Part II – Building your first collection with the Librarian Interface

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

22. 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.]

23. Start a new collection within GLI:

    File→New

24. 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>**.

25. Another window pops up, from which you select the metadata set (or sets) to use. We discuss this in Lab 2. For now, select **Dublin Core Metadata Element Set Version 1.1** followed by **<OK>**.

26. Next you must gather together the files that will constitute the collection. A suitable set has been prepared ahead of time in *workshop_files* 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 → workshop_files

    [This assumes that 'D' is the letter used for the drive where the workshop files are stored: this may vary depending on how the computer has been configured.]

27. 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.

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

29. 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.

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

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

32. 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). Look around the collection and learn about Hobbits!

33. Back in the Librarian Interface, 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 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 is set by an algorithm that Greenstone uses to analyse the document's text.

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

We now build a larger collection of HTML files based on the workshop's html_large folder:

35. 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.]

36. In the **Gather** panel, open the *html_large* folder in *workshop_files*.

37. Drag *englishhistory.net* from the left-hand side to the right to include it in your **tudor** collection. Note: this is quite a large set of documents. If you are pressed for time, navigate to one of the folders inside *englishhistory.net*, e.g. *Monarchs*, and use it instead.

38. 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:

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

40. 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.

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

Extra work

Similarly styled collections can be built from Word and PDF docs. We now experiment with a new collection based on these two formats:

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

43. Copy only the PDF files from *workshop_files→Word_and_PDF*.

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

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

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

Part III – Treasure hunt

Use the UNAIDS collection to answer the questions below.

1. How many publications are there in the collection?

2. How many documents are there that mention *Australia* in the title?

3. How many top-level subject categories are there?

4. What does AAVP stand for?

5. What does AIDS stand for?

6. Considering lower case variants only, how many times does the word "condom" appear in the collection? How many times for "condoms"?

7. If case sensitivity does not matter, how many times does the word "condom" appear in the collection? How many times for "condoms"?

8. If word endings are ignored, how many times does "condom" and variants such as "condoms" appear in the collection?

9. How many *documents* contain some variations of the word "condom"? Does this make it a useful search term?

10. What year saw the first reported case of AIDS in New Zealand?