

Greenstone tutorial exercise

By Ms. K.T. Anuradha, NCSI, India

[Back to index](#)

Sample files: multi-sort

Worked out for Greenstone version: 2.80

NCSI-DSIR Workshop on GSDL December 2007

Lab session 1: Multilingual collection

This exercise explores

1. *Installation of IME on Windows XP system*
 2. *Configuring web browser to view different Charsets*
 3. *Interface Language Customization in GSDL*
 4. *Building a Hindi collection with different file formats*
 5. *Building a Hindi Unicode collection through GLI using sort option enabled*
 6. *Searching and Browsing Unicode content in GSDL*
-

1. Installation of IME on Windows XP system

Start → Settings → Control Panel → Regional & Language Options → Languages → Supplement Language Support → Check (Install Complex Scripts and right-to-left languages including Thai)

Adding language Keyboard layouts:

Start → Settings → Control Panel → Regional & Language Options → Languages → Details → Installed Services → Add → Select your Language & Say OK

2. ***Configuring web browser to view different Charsets*** (most of the cases browser will detect Charset of an Web pages, if it fails you have to make the following settings)

For Internet Explorer

For Unicode Pages – utf-8 Charset: View → Encoding → utf-8

For other charset: View → Encoding → Respective charset

For Netscape

View → Character Encoding → Respective Charset

3. Interface Language Customization in GSDL

- Make a copy of *english.dm* and rename it as *yourlanguage.dm* (In the name of required language ex. *hindi.dm*)
- Pass ISO 639 two letter language standard in front of each macro with in [] bracket, ex.

[l=hi] for Hindi

- Translate Macros value into your language (you have to copy encoding of your translation { } bracket)
- Copy translated Macro file in to *macros* folder of Greenstone
- Include your translated macro file name in macro files list of *etc/main.cfg*
- Configure your language in *etc/main.cfg* i.e. *Language shortname=hi longname=Hindi default_encoding=utf-8 for Hindi*
- Know you can see your language interface by changes interface language option to your language in preference page

4. Building a Hindi collection with different file formats

- Start a new collection called **hindi1**. Fill out the fields with appropriate information.
- In the **Gather** panel, on the workshop CD locate the folder *sample_files* → *Unicode-collection for lab* → *xxx* . Drag this folder into the collection and drop it there.

5. Building a Hindi (UTF-8) collection with sort option enabled:

- Copy allkeys.txt from /resources/allkeys.txt to %Greenstone%\bin\windows\perl\lib\Unicode\Collate directory, where %Greenstone% is the directory where Greenstone is installed. The allkeys.txt file is available at: <http://www.unicode.org/Public/UCA/latest/allkeys.txt>. This file essentially defines how to sort Unicode data. For your convenience the file is kept in the /software/resources directory of the workshop CD.
- Start a new collection called **hindi2**. Fill out the fields with appropriate information.
- In the **Gather** panel, on the workshop CD locate the folder *sample_files* → *hindi-unicode-sort* → *xxx* . Drag this folder into the collection and drop it there.
- Create a collection with metadata details entered in the language of the original file. Give only title that is enough i.e. 'Indhan' as title entered in Hindi. Enter the title in UTF-8 format using any editor which supports this facility and cut and paste it to GLI panel.
- Use GenericList classifier with -sort_using_unicode_collation option. In the collect.cfg an entry will look like: `classify GenericList -metadata title -sort_using_unicode_collation`

6. Searching and Browsing Unicode content files in GSDL

- Change interface to your language through preference page
- For Searching: Require IME or Virtual Key board (Unicode supported)
- Browser Settings: As said above
- Enter search terms and try different combination

For Non-Unicode mapped content follow the same procedure by enabling the required encoding in main.cfg