



Examples: Multimedia and scanned images



Course material prepared by

Greenstone Digital Library Project
University of Waikato, New Zealand



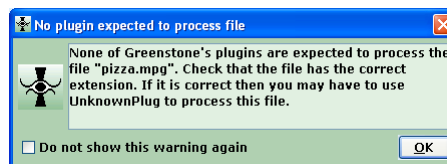
Agenda

- ❖ Multimedia and UnknownPlug
- ❖ Scanned images

Multimedia and UnknownPlug

- ❖ Greenstone has plugins for these multimedia file types:
 - Images (.bmp, .gif, .jpg, .png, .tif, ...): ImagePlug
 - Audio (.mp3, .ogg): MP3Plug, OggVorbisPlug
 - Video (.rm): RealMediaPlug

- ❖ What about files that Greenstone doesn't have a plugin for?



- Use UnknownPlug, a plugin that will process *any* type of file

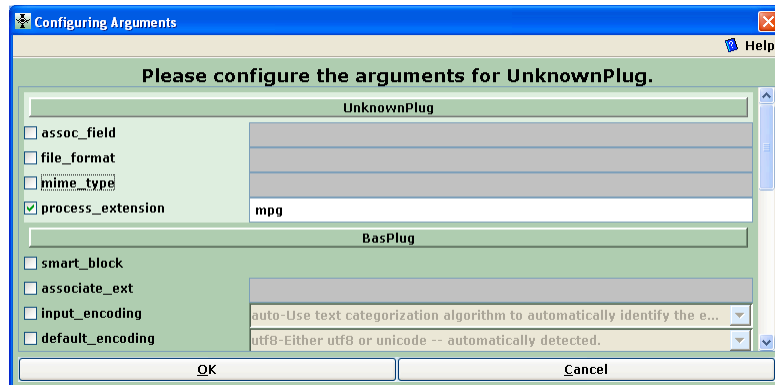
Multimedia and UnknownPlug

- ❖ UnknownPlug knows nothing about any type of file, so it can only extract *very* basic metadata: file name and file size
 - Metadata must be manually assigned

- ❖ UnknownPlug is also unable to extract any text, so the files cannot be searched
 - Files are accessed by browsing or searching on assigned metadata

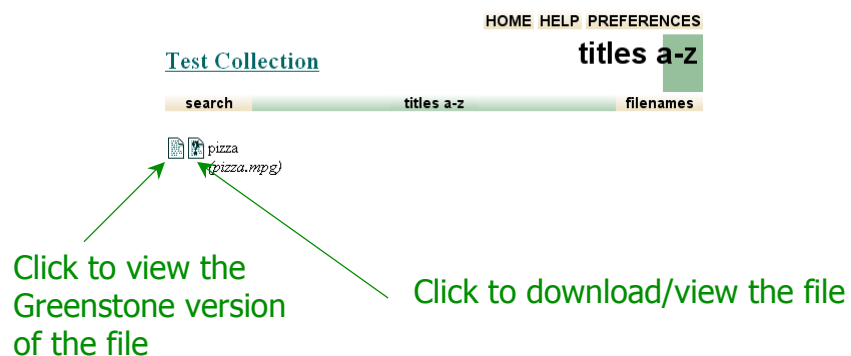
Multimedia and UnknownPlug

- ❖ In the GLI, add UnknownPlug to the collection
 - Use the “process_extension” option to tell UnknownPlug which type of files to process



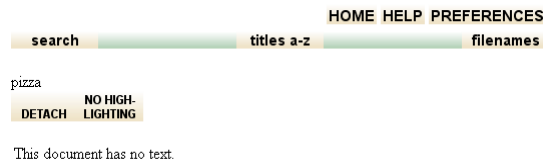
Multimedia and UnknownPlug

- ❖ When the collection is built, the file will now be included in the collection:

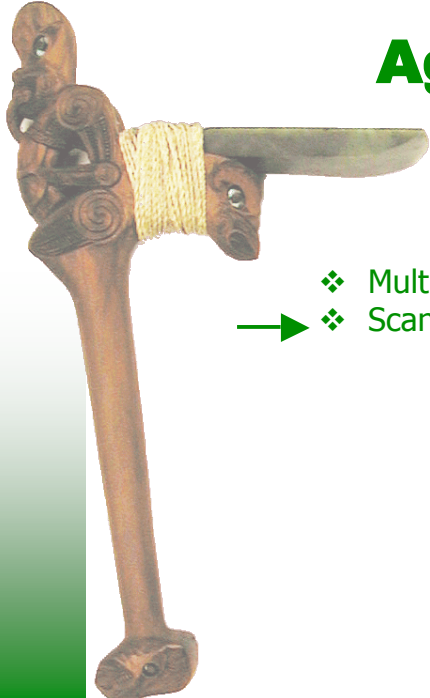


Multimedia and UnknownPlug

- ❖ Since UnknownPlug cannot extract any text, the Greenstone version of the document is very boring!



- ❖ Let's remove the link to this version
 Edit the VList format statement, and replace
`[link][icon][link]`
 with
`{If}{[ex.Plugin] ne "UnknownPlug",[link][icon][link]}`



Agenda

- ❖ Multimedia and UnknownPlug
- ➔ ❖ Scanned images

Scanned images

- ❖ Need to be converted to an electronic form – scanning produces a set of images
- ❖ To add each page as an individual image, process using ImagePlug
- ❖ To group them into a single document, process using PagedImgPlug
 - Requires an 'item' file which lists all the pages and gives additional metadata

Scanned images

- ❖ Can add metadata to the images to enable searching
- ❖ If full text searching is desired, use OCR (Optical Character Recognition) to generate an electronic version of the text
- ❖ Alternatively, if the documents are small and few, manually type the text into a file
- ❖ Text files can be included with the images in the item file

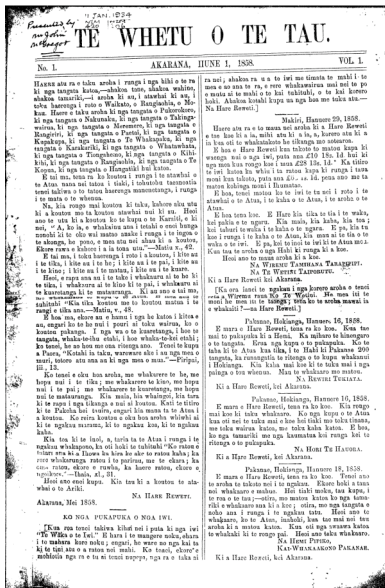
Sample document

4 newspaper page images and their scanned text

9 files:

- 10_1_1.item
- images/10_1_1_1.gif
- images/10_1_1_2.gif
- images/10_1_1_3.gif
- images/10_1_1_4.gif
- text/10_1_1_1.txt
- text/10_1_1_2.txt
- text/10_1_1_3.txt
- text/10_1_1_4.txt

10_1_1_1



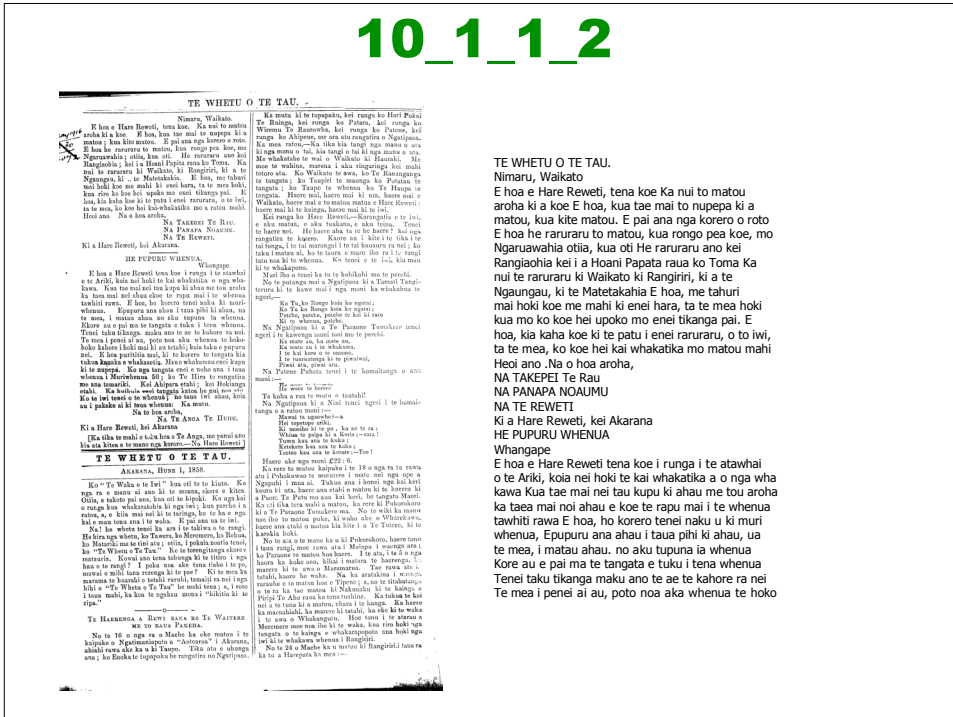
TE WHETU O TE TAU.
No. 1.
AKARANA, HUNE 1, 1858.
VOL. 1.

HAERE atu ra e taku aroha i runga i nga hihi o te ra ki nga tangata katoa,—ahakoa tane, ahakoa wahine, ahakoa tamarii,—i aroha ki au, i atawhai ki au, i toku haerenga i roto o Waikato, o Rangiaohia, o Mokau. Haere e taku aroha ki nga tangata o Pukorokoro, ki nga tangata o Nakunaku, ki nga tangata o Takingsawairua, ki nga tangata o Meremere, ki nga tangata o Rangiriri, ki nga tangata o Paetia, ki nga tangata o Kupakupa, ki nga tangata o Te Whakapaku, ki nga tangata o Karakariki, ki nga tangata o Whatawhata, ki nga tangata o Tioungahemo ki nga tangata o Kihikhi, ki nga tangata o Rangiaohia ki nga tangata o Te Kopua, ki nga tangata o Hangatiki hui katoa.

E tai ma, tena ra ko koutou i runga i te atawhai o te Atua nana nei tatou i tiaki, i tohutohu taeanoaia teni takiwa o te tatou haerenga manenetanga, i runga i te mata o te whenua.

Na, kia rongo mai koutou ki taku, kahore aku utu ki a koutou mo ta koutou atawhai nui ki au. Heoi ano te utu ki a koutou ko te kupu o te Karaiti, e ki nei, "A, ko ia, e whakainu ana i tetahi o enei hunga nonohi ki to oku wai matao anake i runga i te ingoa o te akonga, he pono, e mea atu nei ahau ki a koutou, Kore rawa e kahore i a ia tona utu.—Matu x 42.

10_1_1_2



TE WHETU O TE TAU.
 Nimiru, Waikato
 E hoa e Hare Reweti, tena koe Ka nui to matou aroha ki a koe E hoa, kua tae mai to nupepa ki a matou, kua kite matou. E pai ana nga korero o roto E hoa he raruraru to matou, kua rongu pea koe, me Ngaruawahia otia, kua oti He raruraru ano kei Rangiaohia kei i a Hoani Papata raua ko Toma Ka nui te raruraru ki Waikato ki Rangiriri, ki a te Ngunaugu, ki te Matetakahia E hoa, me tauri mai hoki koe me mahi ki enei hara, ta te mea hoki kua mo ko koe hei upoko mo enei tikanga pai. E hoa, kia kaha koe ki te patu i enei raruraru, o to iwi, ta te mea, ko koe hei kai whakatika mo matou mahi Heoi ano. Na o hoa aroha,
 NA TAKEPEI TE RAU
 NA PANAPA NOAUMU
 NA TE REWETI
 Ki a Hare Reweti, kei Akarana
 HE PUPURU WHENUA
 Whangape
 E hoa e Hare Reweti tena koe i runga i te atawhai o te Ariki, kua nei hoki te kai whakatika o nga whaka-
 tika. Kua tae mai nei tau kupu ki ahau me tou aroha ka taea mai noi ahau e koe te rapu mai i te whenua tawhiti rawa E hoa, ko korero tenei naku u ki muri whenua, Epuuru ana ahau i taua pihiri ahau, ua te mea, i matau ahau. no ako tupuna ia whenua Kore au e pai ma te tangata e tuku i tena whenua Tenei taku tikanga maku ano ae te kahore ra nei Te mea i penei ai au, potu noa aha whenua te hoko

10_1_1.item

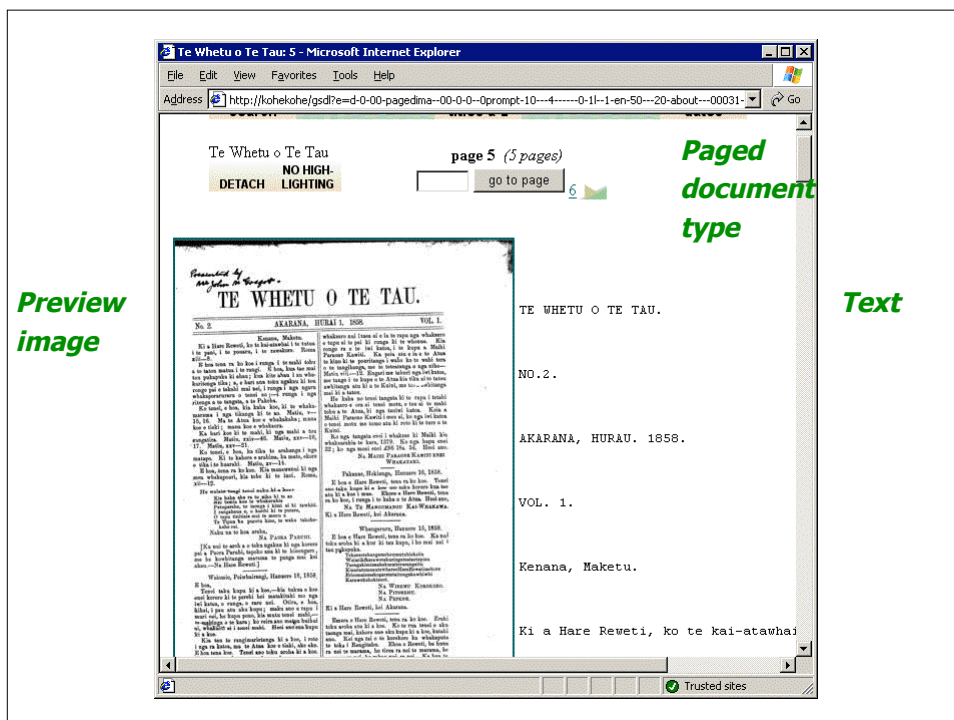
```

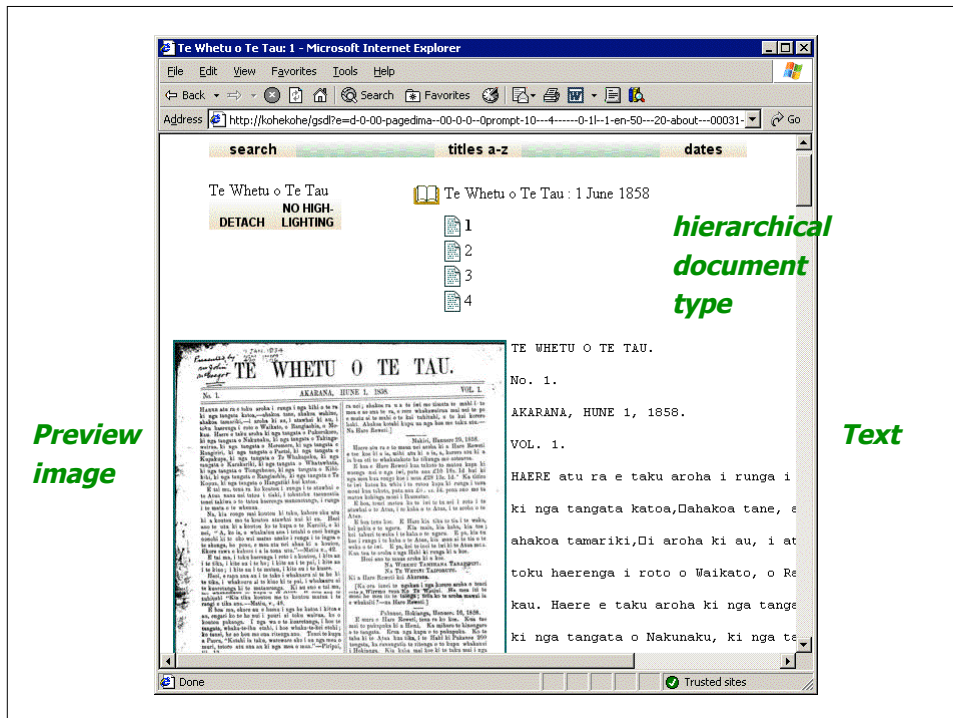
<Title>Te Whetu o Te Tau
<Date>18580601
1: images/10_1_1_1.gif: text/10_1_1_1.txt:
2: images/10_1_1_2.gif: text/10_1_1_2.txt:
3: images/10_1_1_3.gif: text/10_1_1_3.txt:
4: images/10_1_1_4.gif: text/10_1_1_4.txt:
  
```

↑ Page number ↑ Image file ↑ Text file

PagedImgPlug

- ❖ Processes item files and their corresponding image and text files
- ❖ Options:
 - screenview (screenviewsize, screenviewtype) produce a preview image
 - thumbnail (thumbnailsize, thumbnailtype) produce a thumbnail image
 - documenttype paged or hierarchical





Extended item format

```

<PagedDocument>
  <Metadata name="Title">The Title of the entire
    document</Metadata>
  <Metadata name="Subject">A Document level
    Subject</Metadata>
  <Page pagenum="1" imgfile="image1.jpg"
    txtfile="pagel.jpg">
    <Metadata name="Title">The Title of this
      page</Metadata>
    ... more metadata
  </Page>
  ... more pages
</PagedDocument>
  
```