

Intermediate Report: Fundamental Topics in Digital Humanities

Christian Wittern

2014-09-27

Aim of the project

- ▶ Support text based research in East-Asian Studies through a new electronic infrastructure
- ▶ Implementation should support critical textual scholarship
- ▶ Let the Researcher/User be in control, do not require central authority
- ▶ Collect existing data and put them in a research environment to make best use of them

Schedule & output

- ▶ April 2013 to March 2016
- ▶ Sketching of methods, data format, infrastructure
- ▶ Port existing tools to enable the research community to use it
- ▶ Desired output:
 - ▶ Data and programs
 - ▶ A user community
 - ▶ A handbook with explanations of how to use the system
 - ▶ Researchers' testimony: reports on using the system for specific research projects

Background(1): Text

- ▶ Most texts have a long history and come in many versions and editions
- ▶ Over the centuries many commentaries have been written and emendations or annotations proposed
- ▶ Scholars want to trace and evaluate these connections and discover new ones
- ▶ No single publication/institution can solve this problem -> collaboration is needed

Background(2): Metadata

- ▶ A key to understanding a text lies in the information (data) about the texts
- ▶ Traditionally, this type of information has been collected in catalogs and commentaries
- ▶ This information should be combined with the texts themselves to provide new insights
- ▶ We want also tap into the long tradition of 目錄学 which will provide fruitful insights

How can this be realized?

- ▶ We do not yet have definite answers, this project is trying one way
- ▶ We are experimenting with a new infrastructure, that can be used for a network of text repositories:
 - ▶ Server side: Kanseki repository (Gitlab)
 - ▶ Client side: Mandoku, Kanripo Website

Use case(1): Close reading

- ▶ Reading/translating/annotating of a text
 - ▶ requires frequent lookup of words in the text database and dictionaries
 - ▶ collecting of annotations, references
 - ▶ writing of translation, analysis, study

Use case(2): Thematic study

- ▶ topical reading in many different texts
 - ▶ requires creation of specific lists of texts of interest
 - ▶ keeping notes in relation to the texts
 - ▶ link from text location to notes
- ▶ publish the results / research data

Use case(3): Browsing and searching

- ▶ Researching of phrases and words in various contexts
 - ▶ Browser or Mandoku
 - ▶ Research notes are kept in Word or other programs
- > What is needed to best support this?
- ▶ Other use cases?

For casual browsing: Kanripo Website

- ▶ <http://www.kanripo.org>
- ▶ **Very early stage** of development
- ▶ Searching and browsing of text and catalog
- ▶ more to come:
 - ▶ what do we need?

For close reading: Mandoku

- ▶ A specialized environment and interface to the repository
- ▶ Quick lookup of phrases in the text database
- ▶ reference dictionaries incorporated
- ▶ preview available -> see separate document

Technical digression

- ▶ Mandoku is based on Emacs and Org-Mode:
 - ▶ Why Emacs?
 - ▶ Is there a better tool for this purpose?
 - ▶ What is Org-Mode? (<http://orgmode.org/>)
 - ▶ An extension of Emacs for planning and organizing
 - ▶ Also used for research assistance and writing of academic papers, presentations etc.
 - ▶ Used also for note taking, blog articles etc.
- ▶ Viable perspective? Other possibilities?

Technical details: Repository

- ▶ The texts are kept in the repository under a version control system
- ▶ This allows
 - ▶ to keep multiple corresponding versions of a text
 - ▶ to keep a record of what has been edited (by whom)
 - ▶ updating and collating of multiple copies in different locations

Gitlab

- ▶ A clone of the popular www.github.com Github code sharing site
- ▶ Set up at <http://gl.kanriipo.org>
- ▶ 8951 texts have been processed and set up in repositories
- ▶ Provides some functionality for creating research groups, user interaction etc.
- ▶ Editing on-site is possible, but awkward
- ▶ Not specially suited for working with texts -> additional/other access point is needed

Available data

- ▶ Buddhist texts digitized and annotated by CBETA
 - ▶ Buddhist catalogs from the Jinglu project
- ▶ 正統道藏
- ▶ 四庫全書 (partly)
 - ▶ cf 文淵閣四庫全書電子版 (commercial CDROM etc) Digital Heritage Publishing
 - ▶ Wikisource
 - ▶ 360doc
 - ▶ other source now not available anymore
 - ▶ Digital facsimile at Archive.org (文瀾閣本):
 - ▶ 四庫全書
 - ▶ 四庫全書薈要