The Alliance of Digital Humanities Organisations

The Association for Literary and Linguistic Computing

The Association for Computers and the Humanities

The Society for Digital Humanities – Société pour l'étude des médias interactif

Digital Humanities 2010

Conference Abstracts

King's College London, London
July 7 – 10, 2010

The 22nd Joint International Conference of the Association for Literary and Linguistic Computing and Association for Computers and the Humanities

and

The 3rd Joint International Conference of the Association for Literary and Linguistic Computing, the Association for Computers and the Humanities and the Society for Digital Humanities – Société pour l'étude des médias interactif

(Textological Study). Ulaanbaatar: Centre for Mongol Studies, NUM.

Ernst-Gerlach, A. and Fuhr, N. (2007). 'Retrieval in text collections with historic spelling using linguistic and spelling variants'. *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL2007).* Vancouver, Canada, June 2007, pp. 333-341.

Garmaabazar, Kh. and Maeda, A. (2008). 'Developing a Traditional Mongolian Script Digital Library'. *Proceedings of the 11th International Conference on Asia-Pacific Digital Libraries (ICADL2008).* Bali, Indonesia, December 2008, pp. 41-50.

Kimura, F. and Maeda, A. (2009). 'An Approach to Information Access and Knowledge Discovery from Historical Documents'. *Conference Abstracts of the Digital Humanities 2009 (DH09).* College Park, MD, June 2009, pp. 359-361.

Tsevel, Y. (1966). *Mongol helnii tovch tailbar toli.* Ulaanbaatar (Mongolian).

Tungalag, D. (2005). *Mongol ulsiin undesnii nomiin san dahi Mongoliin tuuhiin gar bichmeliin nomzuin sudalgaa.* Ulaanbaatar (Mongolian). V. 1.

Notes

1. `http://www.dl.is.ritsumei.ac.jp/tmsdl`

2. `http://toli.query.mn/`

A Digital Archive of Buddhist Temple Gazetteers

Bingenheimer, Marcus
m.bingenheimer@gmail.com
Dharma Drum Buddhist College

Hung, Jen-jou
Dharma Drum Buddhist College

Temple gazetteers are a subset of the Chinese genre of gazetteers (difang zhi 地方志). Chinese gazetteers, or local histories, are composite texts containing descriptions, bibliographies, poems and other material pertaining to the history of a location or region. The temple gazetteers produced since the 16th century are important for the study of Chinese Buddhist history. They are especially relevant for the last three hundred years, but contain much older material on the history of Buddhist sites.

The archive is currently being constructed at the Dharma Drum Buddhist College, Taiwan, and for the first time opens up a large amount of this material for the study of Buddhism, which so far has been available only with great difficulty. The project website is at: `http://buddhistinformatics.ddbc.edu.tw/fosizhi/`.

The poster briefly introduces the content and then focusses on the technical realization of the "Digital Archive of Buddhist Temple Gazetteers".

Though mainly conceived as an online image database, the project includes 12 large gazetteers as full-text. These are marked up in TEI for names and dates and connected to the Buddhist Authority Database Project (`http://authority.ddbc.edu.tw`), to serve information to the interface. The project is an example for the growing trend to present full text next to digital facsimiles and the poster will show one way this can be done.

As of autumn 2009, the image database consists of more than 100,000 archival-quality images in TIFF format. The images were scanned in 8-bit greyscale at a resolution of 400dpi. From the digital master we produce watermarked JPEG files for use in the interface, and these are made freely available under a Creative Commons license. The quality of the JPEG files is sufficient to read and research the material.

The main limitation on the facsimile value of these images is that they are scans of copies of the original prints or manuscripts and therefore cannot achieve the same verisimilitude as facsimiles taken directly

from the originals. This deficit is mitigated by the fact that the material itself is unproblematic. Mostly text, it consists of black and white woodblock prints or brush writing. Due to the large character-size and the high image resolution the actual facsimile quality is high in the "Excellent Readability" range of the Quality Index (QI) benchmark for printed text (Kenney & Chapman 1995). Eventually, 237 gazetteers will be digitized and made available in the image database.

The gazetteer project includes various kinds of meta-data. We use MIX (Meta-data for still Images in XML) to record technical information about the image files, and meta-data for the TEI full-text files is kept in the teiHeader section of each file. While the MIX data is mainly generated automatically, the TEI meta-data is created as part of the mark-up process.

There are two other important datasets. First is the bibliographic data, which contains important information about the location of the temples, the relationship between the two printed collections, the edition history and chapter order of individual gazetteers, and additional bibliographic information on gazetteers collected from other sources. All this is kept in TEI files to allow for seamless integration with the other textual data.

Second, we have collected semantic data pertaining to each image file, i.e. each page. This data includes the image filename and the page number from the printed edition, which is required for the interface, especially for the majority of gazetteers which have not been digitized as full-text. This semantic image meta-data also records the first three characters on each page and, importantly, all title headings that appear on the page. This results in a database of all title headings, which in turn allows for shallow searches across the whole archive.

Although for production purposes it was useful to create distinct meta-data sources for the project, for archival purposes we integrate all of these different resources in METS wrappers.

The interface is based on the the ExtJS JavaScript library (http://extjs.com/). ExtJS was chosen because it is one of the more advanced JS libraries available. It provides many vital interface functions 'out-of-the-box' and is used in for a number of other projects at the Library and Information Center. The full-texts are kept in a native XML database called eXist (http://www.exist-db.org/), which stores and retrieves the TEI source files.

The interface is geared to enable convenient reading of the gazetteers online. The design is based on the assumption that digital archives will increasingly want to present electronic text in conjunction with and alongside electronic facsimiles of the original source. Among the challenges we solved here was how to have images and text move in unison. Another problem we faced was how to give the user an idea about where she was in the structure of the text, and it was decided to offer a detailed navigation panel, which contains detailed titles of every section to address this need. Finally, for better and more convenient accessibility of the image files we have included a magnifier to aid reading. A sophisticated search function lets users choose to search for strings in the fulltext, metadata or both, over the whole archive or one single gazetteer.

References

Bingenheimer, Marcus (forthcoming). 'Bibliographical notes on Buddhist temple gazetteers and some remarks on their use for the study of Chinese Buddhist history'. *Oslo Studies in Language.*

Brook, Timothy (2002). *Geographical Sources of Ming-Qing History.* Michigan monographs in Chinese Studies. Ann Arbor: Univ. of Michigan, Center for Chinese Studies V. 58.

Eberhard, Wolfram (1964). *Temple Building Activities in Medieval and Modern China.* Monumenta Serica. V. 23, pp. 264-318.

Zhongguo Fosi Shizhi Huikan (1980-1985). 中國佛寺史志彙刊. Compiled by Du Jiexiang 杜潔祥 (ed.). Taipei 110 vols.

Zhongguo fosizhi congkan (2006). 中国佛寺志叢刊. Compiled by Zhang Zhi 张智 (ed.). Hangzhou: Guangling shushe 广陵书社. 130 vols.