Future of the Network for Research on Japanese Classical Books

Junko Koyama (presenter)

National Institute of Japanese Literature
Research Department
Associate Professor

prepared by Center for Collaborative Research on Pre-modern Texts,
National Institute of Japanese Literature
1. Current situation of the project

Project for advancing the frontiers of large-scale academic research efforts

**Project to Build an International Collaborative Research Network for Pre-modern Japanese Texts**

- **Term:** 2014-2023
- **Goals:**
  1. To make publically available on the internet a collection of 300,000 digitized Japanese classic books from various fields, along with their bibliographies (71,000 books will be digitized by March 2016).
  2. To promote collaborative international research projects through the use of digitized materials.
  3. To build networks for collaborative international research.

A new public image viewing system starts in April, 2017.
2. Collaborative institutions

20 universities

National Institute for the Humanities
- National Museum of Japanese History
- National Institute of Japanese Language and Linguistics
- International Research Center for Japanese Studies

Foreign institutions with collaborative agreements
- Collège de France
- Korea University
- University of Florence
- Beijing Foreign Studies University
- Leiden University
- University of Venice
- University of Naples
- Sapienza University of Rome
- Vatican Apostolic Library
- University of Cambridge
- University of British Columbia
- Columbia University
- C. V. Starr East Asian Library at University of California

Other institutions
Digitization of classic books in other institutions
Cooperation in building databases by the National Institute of Informatics.

Collaborative efforts will be strengthened and broadened as this project progresses.
3. Two points of view for classic studies

- **How to read a classic**
  - Training in deciphering cursive scripts
  - Confirming the meaning of words
  - Developing programs to help decipher cursive scripts
  - The Hentaigana App
  - KuLA

- **How to gather intelligence data for the study**
  - Portal site of the image
  - Improvement of search functions
  - Stable search results
  - SMART-GS
  - JapanKnowledge
  - Classical Books OCR

- User-friendly search
- Cross-searching
- Tagging
  - TEI-conformant texts
- Embedding DOI

+ Tool development support:
  - Open Data Ideathon • hackathon
4. Portal site of Classical Books

Improvement of information and image availability
Not individually but comprehensively

Cross-searching
via NIJL’s authorized-management site, not through individual sites.

encouraging unexpected findings

NIJL (Platform for Japanese classics)

Partner

Partner

Partner

Partner

and others

and

others
5. The reason why NIJL has taken on this project

《To arrive at necessary information》

- Each library has developed a catalog of its collection of classic books.
- NIJL has authorized management regarding the catalogs and micro films of classic books.

**Advantage of NIJL:**
authorized management of classics

【Renewed URL of Integrated Catalogue of Early Japanese Books】
http://base1.nijl.ac.jp/~tkoten/

- NIJL has been developing a comprehensive bibliographical database of Japanese classics, which includes the "Integrated Catalogue of Pre-modern Japanese Books", an index of classic books from all fields.
6. Improved search functions

Addition of tags to individual images

Developing a tagging tool on the Web
7. Improved search functions②

Social tagging will commence in April, 2017 (initially by researchers)
8. Improved search functions

Application:
Locate a range of images which share similar shapes in a given image

⇒ *semi-automated tagging*

Algorithm for word spotting


Advantages:
No need to manually plot a range of tags on a given image.
Feasible even for cursive writing.
Applicable for various writing systems.
9. Improved search functions

■ Reduced labor for reprinting classical Japanese books

http://doi.org/10.1241/johokanri.58.819

=A character which the software was unable to decode
◯A character which the software decoded incorrectly

■ Efficient transcription of necessary books

■ TEI-conformant texts are currently under consideration.
Embedding **digital object identifiers (DOI)** into each digitalized image of a classical book, and recommendation to cite the DOIs in academic papers.

This makes it easy for readers to refer back to digital images of books referred to in academic papers.

This process was very difficult in previous research.

⇒ Research should be verifiable.

**Toward verifiable humanities**

Most work in the humanities has depended on drawing upon fragments and citations of other researchers. However, we must now work to verify these fragments for ourselves.
Towards the development of new frontiers via unexpected findings

We aim to promote original research via extraction of various information from the database.

Search functions mentioned above

Finding something significant in unexpected books

Search for the word "aurora"

Digitized images

Tags

The Database

Ex.

Enkouan zuikan zue
12. Encouraging further interaction with Japanese classic books

■ To encourage public engagement

Published digitized images and bibliographic information under CC BY-SA: available for everyone.

【NIJL】
As an experiment, 350 digitized classic books have been made public under CC BY-SA to promote open data searching. (In the near future, much more of NIJL’s collection will be published.)

http://www.nijl.ac.jp/pages/cijproject/data_set_list.html

■ From Ideathon to Hackathon

An ideathon was held by NIJL in Kyoto

Developing an idea suggested in the ideathon
13. Encouraging further interaction with Japanese classic books

To encourage utilization not only by researchers but also by the public

A user-friendly viewer for open data searches has been released by Dr. Nagasaki

http://www2.dhii.jp/nijl_opendata/openimages.php
Wrap up: Goals of the project

(1) Promote integrated searching and collaboration as a platform for further research

(2) Encourage various institutions to engage in this sharing of information
   ※ Assisting, collaborating with, and forming frameworks for various projects
   ※ Supporting sustainability of their results

To encourage interaction with Japanese classic books in mutually beneficial relationships