JHTI (Japanese Historical Text Initiative) Project:

A Collaborative Research between Japan and the US on Designing and Constructing the Full Text Coordinated Retrieval System of Japanese Historical Documents using the Internet

Ikuo Oketani*1 Delmer Brown*2 Shigeo Kikuchi*1 Chizuko Saito*2

*1: Faculty of Human Sciences, Osaka International University
*2: University of California, Berkeley
The purposes of this full-text retrieval system (1)

Mainly, about the research of the Japanese history and Japanese literature

- Research of Japanese ancient history
- Research of Japanese literature
- Research of Japanese spiritual life
- For assisting foreign researchers in studying Japanese classics
- For assisting general studies in Japanese culture
- Promotion of international collaboration
The purposes of this full-text retrieval system (2)

- Collaboration with the ECAI project

- Project name: JHTI (Japanese Historical Text Initiative)

The goal of JHTI is to digitize the twenty-five volumes of Japanese classic texts (discussed later on) and to create a database.
The purposes of this full-text retrieval system (3)

(1) To develop a retrieval method focusing on the structure of texts and the method of historical documentation
(2) To develop the cross-cutting (integrated) retrieval function
(3) To quantitatively analyze historical documents
(4) To expand the *kanji* attribute database, targeting *gaiji*
(5) To correlate the present system with the GIS, or the Geographic Information System
The purposes of JHTI (Japanese Historical Text Initiative) project

Four points of our presentation

1. The purposes of this full-text retrieval system
2. The characteristics of our full-text retrieval system
3. The non-standard kanji characters (Gaiji) and the variant characters (Iji)
4. Activation of an English Translation Assisting Program
5. Promotion of the Collaborative Research between Japan and the U.S.
6. The future expansion of this full-text retrieval system
   - database management system
   - XML-tagged text system
   - Correlation with the GIS
Document names which we are planning to digitize in near future
(*: able to search, **: work in progress)

Text 1: Kojiki (古事記) *
Text 2: Nihon Shoki (日本書紀) *
Text 3: Shoku Nihongi (続日本紀) *
Text 4: Izumo Fudoki (出雲風土記)**
Text 5: Kogoshui (古語拾遺)
Text 6: Engi Shiki (延喜式) *
Text 7: Eiga Monogatari (栄華物語)**
Text 8: Okagami (大鏡) **
Text 9: Azuma Kagami (吾妻鏡)
Text 10: Gukansho (愚管抄) *
Text 11: Jinno Shotoki (神皇正統記) *
Text 12: Taiheiki (太平記)
Text 13: Daijingu Jin'iiki (大神宮神威記)
Text 14: Dokushi Yoron (読史余論)**
Text 15: Meiji igo Shukyo kankei Horei *
(明治以降神社関係法令史料)
Text 16: Kokutai no Hongi (国体の本義)**
Text 17: Tenri-kyo (天理教)
Text 18: Kurozumi-kyo (黒住教)
Text 19: Konko-kyo (金光教)
Text 20: Omoto-kyo (大本教)
Text 21: Itto-en (一燈園)**
Text 22: Tensho Kotai Jingu-kyo
(天照皇太神宮教)
Text 23: Rissho Kosei-kai (立正佼成会)
Text 24: Tsubaki Ookami Yashiro (椿大神社)
Text 25: Manyousyu (万葉集)**
The characteristics of our full-text retrieval system (1)

We explain as an example of "Nihon-shoki."

About the Nihon-shoki
- The Imperial Chronicle of Japan
- The chronicle consist of 30 volumes
- The chronicle was edited in A. D. 720
- The chronicle covers the period from the mythological age to the end of the reign of Empress Jito, 697

* This sort of electronic joint work on Japanese ancient chronological literatures is important for researchers of ancient Japanese history in order to make Japanese ancient history, the foundation and the structure of ancient Japan, and folkloristic aspects of ancient Japan electronically easier to access.
Our Full Text Retrieval System includes four types of Documents. These four documents are as follows:

- **English translation document**

- **Japanese (Chinese writing style) document**
  The original text of "Shoki Shuge" (Kawamura, Hidene)著, 1785, 30 volumes.

- **The image of Japanese Nihonshoki document**
  The original text of "Shoki Shuge" (Kawamura, Hidene)著, 1785, 30 volumes. (stored in UCB East Asian Library)

- **Romanized document**
The characteristics of our full-text retrieval system (4)

- Three retrieval methods of this system

(1) **Keyword retrieval** (We will describe later)
(2) **Subject retrieval**
   
   For example, we can retrieve paragraphs by typing God’s name, name of a place, ritual, and shrine’s name. This function will work efficiently when we insert the SGML tags into this text in the future.

(3) **Browsing function**
   
   We have also made a “browsing function.” We assume that this function is very useful for the beginners of the Japanese history study. When we use this browsing function, we are able to display the texts from the beginning to the end consecutively.
Diagram 1. Home Page of the JHTI Project

Welcome to Japanese Historical Text Initiative, a web site for helping scholars and students experience the wonders of electronic research.

Objectives

Since this initiative is on-going and collaborative, we appreciate receiving recommendations for the addition of other texts, suggestions for the improvement or addition of translations and references, and offers to publish Japanese historical studies electronically. These should be addressed to Co-Principal Investigator Delmer Brown of UC Berkeley 11B09@socrates.berkeley.edu.
Diagram 2. A sample of entering a key word (大日 %u5b41;) on a screen

http://sunsite.berkeley.edu/JHTI/
Diagram 3. A screen display of the retrieval result of this system
Diagram 4. A display of both English and Japanese paragraphs which correspond to each other.
Diagram 5. A full screen display of the paragraph number 5
Promotion of the Collaborative Research between Japan and the U.S.

We are currently preparing to conclude an agreement as follows;

Agreement between Grant-in-Aid for Scientific Research (S) : International Collaboration for Japanese Study (A Project of The National Institute of Japanese Literature)

and

The Japanese Historical Text Initiative (A Project of The International and Area Studies of The University of California, Berkeley)
Diagram 6. The concept of the confirmation system

- Registration System
- ALL Users (Under Graduates)
- JHTI Databases (24 Document-α)
- Nihon Koten Bungaku Taikei Databases
- NIJL (Kokubun)
- Nihon Koten Bungaku Taikei

Japanese Researchers

Researchers

Post-graduates

Nihon Koten …: The Anthology of Japanese Classical Literature

Japanese Historical Text Initiative

Certification

JHTI: Japanese Historical Text Initiative

NIJL: National Institute of Japanese Literature
JHTI (Japanese Historical Text Initiative)
Database Structure
Activation of an English Translation Assisting Program

About the Shoku Nihongi:
- The Imperial Chronicle written in Chinese
- The chronicle consists of 40 volumes
- The chronicle was edited as the second of the National Histories in Enryaku 16 (797) (following Nihon Shoki)
- The chronicle covers the period from the Emperor Monmu’s reign (697) to the Emperor Kanmu’s reign (791)

- English translation: only volumes one to six (by J.B. Snellen)

- The implementation of English translation-assisting program on the Internet
  - Use of the characteristics of Shoku Nihongi as a chronicle the years, seasons, months, and dates -> automatically insertion
  - To refer to other texts, such as Kojiki and Nihon Shoki, etc.

http://sunsite.berkeley.edu/JHTI/nihongi.html
Diagram 7. The input screen of the translation-assisting program
The reference screen of an English translation-assisting program

English

Date: 17th day

Paragraph: An imperial Edict was issued as follows: Hearken! all ye assembled August Children, Princes, Nobles, Officials of the Realm under Heaven to the Word which he speaks even as the Word of the Sovereign that is a manifest God ruling over the Great Land of Many Islands b)
The non-standard kanji characters (*Gaiji*) and the variant characters (*Iji*)

(1) *Gaiji* Input Method

Parts of Chinese Characters and specific tags

*Gaiji*: put “%u” and “;(semicolon)” just before and at the end of its number.

Example: 樺 %u6a34; Unicode; 樺 %u6f83; Unicode; 雙 %uf1a2; Not in Unicode, Creation;
The non-standard kanji characters (Gaiji) and the variant characters (Iji)

(2) Gaiji Retrieval method

We can search the non-standard Kanji characters by the following items.

- Radical name
- Radical strokes
- Total strokes
- Japanese (Chinese) pronunciation
- Unicode
- Daikanwa (morohashi) code

(3)(4) Gaiji Display and Forwarding Method

Gif formatted file (image file) : via the Internet
Diagram 1. The total number and a variety of forms of *gaiji* appeared in each document

<table>
<thead>
<tr>
<th></th>
<th>Nihon-shoki</th>
<th>Kojiki</th>
<th>Shoku-Nihongi</th>
<th>Jinno-Shotoki</th>
<th>Shinroyojo</th>
<th>Godaikunen</th>
<th>Jinno-Shotoki</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Gaiji</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in Unicode</td>
<td>7,512</td>
<td>969</td>
<td>545</td>
<td>1,340</td>
<td>153</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td><em>Number of making Gaiji</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in Daikanwa</td>
<td>5,517</td>
<td>131</td>
<td>235</td>
<td>118</td>
<td>11</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Not in Daikanwa</td>
<td>5,398</td>
<td>48</td>
<td>8</td>
<td>86</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Kinds of Gaiji</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in Unicode</td>
<td>317</td>
<td>305</td>
<td>83</td>
<td>109</td>
<td>54</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td><em>Number of making Gaiji</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in Daikanwa</td>
<td>116</td>
<td>75</td>
<td>29</td>
<td>15</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Not in Daikanwa</td>
<td>88</td>
<td>31</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

*Nihon-shoki* : The Imperial Chronicle of Japan
*Kojiki* : The Imperial Records of Ancient Matters
*Shoku-Nihongi* : Chronicle of Japan, continued, from 697-791A.D.
*Jinno-Shotoki* : A Chronicle of Gods and Sovereigns
## Samples of *Gaiji* which appears in Kojiki (in Unicode)

<table>
<thead>
<tr>
<th>Number</th>
<th>Pronunciation</th>
<th>Num.</th>
<th>Strokes</th>
<th>Radical</th>
<th>Daikanwa</th>
<th>Unicode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>修</td>
<td>修</td>
<td>0</td>
<td>3</td>
<td>(toolbar)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>修</td>
<td>修</td>
<td>0</td>
<td>3</td>
<td>(toolbar)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>修</td>
<td>修</td>
<td>0</td>
<td>3</td>
<td>(toolbar)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>修</td>
<td>修</td>
<td>0</td>
<td>3</td>
<td>(toolbar)</td>
<td></td>
</tr>
</tbody>
</table>

![Images of Gaiji samples in Kojiki](Image)
Samples of *Gaiji* which appears in *Kojiki*  
(Not in Unicode)

<table>
<thead>
<tr>
<th>Strokes</th>
<th>Pronunciation</th>
<th>Num.</th>
<th>Radical</th>
<th>Num.</th>
<th>Radical</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>しゅく</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>そこ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ひ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ひ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>た</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>と</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>しゅく</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>そこ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ひ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ひ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>た</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>と</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>しゅく</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>そこ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ひ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ひ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>た</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>と</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Gaiji retrieval method “Using the number of strokes of radicals”
The result of retrieval “Using the number of strokes of radicals” (radical: 女)
Radical’s Name

Gaiji Character of おんな・おんなへん

20 瑠
13 嫔
9 妫
7 妻

Back to Previous Page
Interactive Searching of Nihon Shoki
— in Process —

MODE: Retrieval

Find | Complete Sentence
Find word or phrase: 大日Ku5641.

If you want to find a combination of words, you should insert + (meaning “or”) or * (meaning “and”) between the words. Do not insert space or blank.

Word(s) retrieval: Which version? Japanese

If you want to show more than 10 results: 10

Submit  Clear
Document names which we are planning to digitize in near future
(*:able to search, **:work in progress)

Text 1: Kojiki (古事記) *
Text 2: Nihon Shoki (日本書紀) *
Text 3: Shoku Nihongi (続日本紀) *
Text 4: Izumo Fudoki (出雲風土記) **
Text 5: Kogoshui (古語拾遺)
Text 6: Engi Shiki (延喜式) *
Text 7: Eiga Monogatari (栄華物語) **
Text 8: Okagami (大鏡) **
Text 9: Azuma Kagami (吾妻鏡)
Text 10: Gukansho (愚管抄) *
Text 11: Jinno Shotoki (神皇正統記) *
Text 12: Taiheiki (太平記)
Text 13: Daijingu Jin'iki (大神宮神威記)
Text 14: Dokushi Yoron (読史余論) **

Text 15: Meiji igo Shukyo kankei Horei *
(明治以降神社関係法令史料)
Text 16: Kokutai no Hongi (国体の本義) **
Text 17: Tenri-kyo (天理教)
Text 18: Kurozumi-kyo (黒住教)
Text 19: Konko-kyo (金光教)
Text 20: Omoto-kyo (大本教)
Text 21: Itto-en (一燈園) **
Text 22: Tensho Kotai Jingu-kyo
(天照皇大神宮教)
Text 23: Rissho Kosei-kai (立正佼成会)
Text 24: Tsubaki Ookami Yashiro (椿大神社)
Text 25: Manyousyu (万葉集) **
The future expansion of this full-text retrieval system

Our Three Immediate Tasks

- Database Management System : OpenText

- XML or SGML-tagged text system
  We will improve our present tagging system into XML (Extensible Markup Language), or SGML (Standard Generalized Markup Language) and the documents will be stored in the Database Management System (OpenText).

- Correlation with the GIS (Geographic Information System)
  We will correlate the present system with the GIS (Geographic Information System), in order to use the information of the geographic distribution of Shinto shrines for our system more efficiently.
Appendix(1)

http://sunsite.berkeley.edu/JHTI/

http://sunsite.berkeley.edu/JHTI/nihongi.html