CJK Searching in HOLLIS

This document offers tips for searching Chinese, Japanese, and Korean (CJK) materials in the HOLLIS Catalog. CJK materials may be searched in HOLLIS using either romanized or vernacular (CJK script) input.

Your computer workstation must be set up to support CJK input. Consult these documents for assistance:

- Inputting CJK scripts: http://hul.harvard.edu/ois/systems/aleph/docs/hcat/cjk2.pdf

What is an input method editor (IME)? An IME is a program that allows computer users to enter complex characters and symbols, such as Chinese/Japanese/Korean characters, using a standard western (non-Asian) keyboard.

What are the basic software requirements to view and input CJK characters? You will need a web browser (Internet Explorer 5.x or higher; Netscape 6.2 or higher) configured to support Unicode-based characters. Windows users will need Windows 95 or higher (Windows 2000 is recommended) with appropriate Global IMEs (input method editors) installed. Macintosh users will need OS9.x with the appropriate language packs installed or OS X. See the document CJK software setup for detailed instructions.

Can I view/input CJK characters from computers in the library? Ask a reference librarian about the availability of CJK-capable computer workstations in your Harvard library. While some library workstations run on Windows 2000, many more run on Windows NT. The Global IME running on Windows NT has certain limitations, particularly for Chinese input. Users accustomed to Global IME on Windows 2000 or XP may see differences in how searches are input when in NT.

Contents

1. Searching by romanization .................................................................2
   Advantages of romanized searching ..................................................2
   Select a romanization system .........................................................4
   Tips on romanized entry: ..............................................................4

2. Searching by vernacular .................................................................5
   CJK browse searches ........................................................................5
   CJK browse display ..........................................................................6
   CJK keyword searches .....................................................................11

January 30, 2003
© President and Fellows of Harvard College
1. Searching by romanization

Advantages of romanized searching
Select a romanization system
Tips on romanized entry

Advantages of romanized searching

Romanized searching has a number of advantages:

- It is often the fastest way of searching, because there is no need to activate an IME, to select characters, to adapt the keyboard, etc.
- Not all records for materials in CJK languages are catalogued in CJK scripts. This is especially true for older materials and smaller collections. Romanized searching allows you to retrieve these records.
- All CJK entries have a parallel romanized entry, so you can retrieve CJK records with a romanized search.

Records retrieved through romanized searches will display CJK information

For example, a romanized browse for the Japanese author Miwa Yoshio looks like this:
The Romanized browse for author Miwa Yoshio produces a romanized browse list:

Click on entry to retrieve the record or list of relevant records

The browse list entry for Miwa Yoshio retrieves a record with both CJK and romanized information.

Distributed by the HUL Office for Information Systems
Select a romanization system

To enter a search directly in romanized form, please use the following systems:

<table>
<thead>
<tr>
<th>Language</th>
<th>Romanization system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>Pinyin</td>
</tr>
<tr>
<td>Japanese</td>
<td>Hepburn (macrons not available; use regular Latin vowel)</td>
</tr>
<tr>
<td>Korean</td>
<td>McCune-Reischauer (diacritics not necessary)</td>
</tr>
</tbody>
</table>

Tips on romanized entry

Chinese:

- For pinyin searches, compound personal, place, administrative, and organizational names are generally written as one word (e.g. Ouyang 歐陽, Zedong 则栋, Taiwan 台湾, Beijingshi 北京市, Zhonghuarenminggongheguo 中华人民共和国, zizhiqu 自治区). Other compounds (e.g. lishi 歷史 are entered as separate words. If you are having trouble finding items, check the spacing; it is not entirely consistent. Alternately, switch to a vernacular search and enter the characters.

  **Example:** Normal pinyin transcription: Zhongguo shehui lishi

  HOLLIS entry: Zhongguo she hui li shi

- ü may be searched using u (e.g. a search for Yulin will retrieve Yülin)

- In general, romanized searches for Chinese-language materials should be entered in **Hanyu pinyin**. All authorized Library of Congress headings (e.g. title, author, subject, etc.) have been converted from Wade-Giles to **Hanyu pinyin**. However, some instances of Wade-Giles or other transliterations may remain in bibliographic descriptions or in personal or place names.

Japanese:

- It is not necessary to enter macrons to search for words with long vowels. Simply use the regular Latin vowel (e.g. a search for Bukkyo will retrieve Bukkyō).

Korean:

- It is not necessary to enter diacritics. Simply use the regular Latin vowel (e.g. a search for Il-song will retrieve Il-sŏng).
2. Searching by vernacular

Vernacular CJK searches may be performed as browse or keyword searches.

- **Browse Search**: Searches headings of titles and authors. CJK entry cannot be used for subject browses.
- **Keyword Search**: Searches for keywords in records as well as keywords in specific indexes such as titles, authors, subjects, etc.

CJK browse and keyword searches have some special features.

- CJK browse searches
- CJK browse display
- CJK keyword searches

**CJK browse searches**

CJK browse searches are limited to particular indexes that are specific to a CJK language. These are:

- Chinese: Author
- Chinese: Title beginning with…
- Japanese: Author
- Japanese: Title beginning with…
- Korean: Author
- Korean: Title beginning with…

You may find these indexes by scrolling down the browse search box.

To enter a vernacular browse search, you must select the appropriate index and enter the search term using the appropriate IME for that specific language (Chinese input method for Chinese browse, Japanese input method for Japanese browse, Korean input method for Korean browse). Then click on Browse.

1. Scroll through the browse indexes and select the appropriate choice.
2. Type the vernacular search term(s) using the appropriate input method.
3. Click on the Browse button to perform the search.
To search other (non-CJK) indexes on the browse search menu you must use romanized input. Search terms entered in vernacular for these other indexes (e.g. Title beginning with…) will retrieve irrelevant results. For example, selecting Title beginning with… and entering "敦煌" retrieves meaningless results:

![CJK browse display](image)

**CJK browse display**

CJK browse searches produce ordered, vernacular results lists. Scroll through these lists by clicking on the Previous Page and Next Page buttons on the right side of the screen. Click on any of the names or titles to see the relevant records.

Examples:

- Single entries lead directly to a record with English and CJK entry. [See example.]

- Multiple records for one entry in a CJK browse produce romanized brief lists, which in turn link to bilingual records. [See example.]

- Japanese and Korean browse results will sort by Unicode value. This will be different than the phonetic order you may be used to seeing in other applications. [See example.]

- Chinese browse results will sort in pinyin order. Simplified and traditional character results will sort separately. Traditional character results do not necessarily follow their simplified counterparts. To work around this, click on the author or title in one of the records to link to an authorized heading that will produce a complete list of relevant works. [See example.]

Distributed by the HUL Office for Information Systems
Single record example

When a browse results in a single entry, click to select that entry and display a record with CJK and romanized data.

A search for the Japanese author 三輪 良雄 produces a single result:

Clicking on the entry leads directly to a record with both CJK and romanized information:

Distributed by the HUL Office for Information Systems
Brief list example

Multiple records for one entry in a CJK browse produce romanized brief lists, which in turn link to bilingual records.

A search for the Japanese author 三家英治 produces multiple records:

Clicking on the entry produces a brief list in romanized form:

Each individual record displays in both CJK and romanization:
Browse order: Japanese and Korean

Japanese and Korean browse results will sort by Unicode value. This will be different than the phonetic order you may be used to seeing in other applications.

A Japanese: Author browse search for 加藤 displays in Unicode value order:

<table>
<thead>
<tr>
<th>No. of Recs</th>
<th>Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>加持武雄, 1928-</td>
</tr>
<tr>
<td>1</td>
<td>加藤 善夫, 1949-</td>
</tr>
<tr>
<td>1</td>
<td>加藤 幸雄, 1947-</td>
</tr>
<tr>
<td>1</td>
<td>加藤 昌, 1952-</td>
</tr>
<tr>
<td>1</td>
<td>加藤 留一, 1960-</td>
</tr>
<tr>
<td>1</td>
<td>加藤 勝男, 博士 遺著 記念 総集, 総集 翻行会.</td>
</tr>
<tr>
<td>1</td>
<td>劉備, 購易販售會.</td>
</tr>
<tr>
<td>2</td>
<td>勝田 錦夫, 1912-</td>
</tr>
<tr>
<td>1</td>
<td>北京 (Tokyo, Japan).</td>
</tr>
</tbody>
</table>

A Korean: Author browse search for 01 displays in Unicode value order:

<table>
<thead>
<tr>
<th>No. of Recs</th>
<th>Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>윤한용.</td>
</tr>
<tr>
<td>1</td>
<td>이석현, 국회의회의회 전향헌전 헌의회.</td>
</tr>
<tr>
<td>2</td>
<td>이규덕.</td>
</tr>
<tr>
<td>1</td>
<td>이규영, 1945-</td>
</tr>
<tr>
<td>1</td>
<td>이기용.</td>
</tr>
<tr>
<td>1</td>
<td>이범덕, 1920-</td>
</tr>
<tr>
<td>1</td>
<td>이덕철.</td>
</tr>
<tr>
<td>1</td>
<td>이도길.</td>
</tr>
<tr>
<td>1</td>
<td>이범석, 1959-</td>
</tr>
<tr>
<td>1</td>
<td>이상일.</td>
</tr>
</tbody>
</table>
Browse order: Chinese

Chinese browse results will sort in **pinyin** order. Simplified and traditional character results will sort separately. Traditional character results do not necessarily follow their simplified counterparts. To work around this, click on the author or title in one of the records to link to an authorized heading that will produce a complete list of relevant works.

A **Chinese: Title beginning with…** browse search for 敦煌 produces a list like this:

<table>
<thead>
<tr>
<th>No. of Recs</th>
<th>Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>敦煌: 学术论集</td>
</tr>
<tr>
<td>3</td>
<td>敦煌: 佛教</td>
</tr>
<tr>
<td>1</td>
<td>敦煌: 题目</td>
</tr>
<tr>
<td>1</td>
<td>敦煌: 建议</td>
</tr>
<tr>
<td>1</td>
<td>敦煌: 研究集</td>
</tr>
<tr>
<td>1</td>
<td>敦煌: 纪念敦煌藏经阁发现一百周年</td>
</tr>
<tr>
<td>1</td>
<td>敦煌: 最大的文化宝藏</td>
</tr>
<tr>
<td>1</td>
<td>敦煌: 纪念</td>
</tr>
<tr>
<td>2</td>
<td>敦煌: 英译, Chinese</td>
</tr>
<tr>
<td>1</td>
<td>敦煌: 建议 题目</td>
</tr>
</tbody>
</table>

Click on the author or title in one of the records to link to an authorized heading that will produce a complete list of relevant works.
CJK keyword searches

Below are tips for performing CJK keyword searches in HOLLIS. As with browse searches, click on a specific title in the Brief List of search results to retrieve a record with both CJK and romanized information.

- CJK keyword searches entered as **characters** retrieve results across **all three CJK languages**.

  For example, a "Keyword Anywhere" search for 民 produces results in all CJK languages:

  ![Image showing search results]

- Terms may be combined using **Boolean operators** (and, or, not) to expand or focus searches by combining characters, kana/hangul, and romanized terms.

**Examples:**

1) A keyword search for Dunhuang or 敦煌 produces a fuller range of results, since it captures records which contain the characters 敦煌 but where the romanization might be different.

2) A keyword search for 蒙古 or モンゴル retrieves a fuller range of results on Mongolia.

3) A keyword search for 北京 and Architecture combines vernacular keyword and Library of Congress subject term search for more focused results.

- Character adjacency is the default, meaning that results will produce records with the desired characters next to each other and in left to right order.

**Note:** If you do not wish the characters to be necessarily adjacent, join them with the Boolean operator "and." E.g. 中國經濟 vs. 中國 and 經濟.

- Default sort for vernacular keyword search results is in descending order by year and within year by author (the same as for non-vernacular keyword searches).