An Alternative Method of Title Authority Control: The Shifted-Authority Control Model for Chinese Classics

Maiko Kimura
Keio University

Follow this and additional works at: https://scholarsarchive.byu.edu/jeal

Part of the Cataloging and Metadata Commons

BYU ScholarsArchive Citation
Available at: https://scholarsarchive.byu.edu/jeal/vol2019/iss169/3

This Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Journal of East Asian Libraries by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
An Alternative Method of Title Authority Control: The Shifted-Authority Control Model for Chinese Classics

Cover Page Footnote
This work was partially supported by Konosuke Matsushita Memorial Foundation Research Grants 2016, JSPS KAKENHI Grant Number 17J40023 and 18K18329. In the process of constructing KWMA-san, the Definition Table for Kanji Unification Index developed by the NII, working together with Yasunori Wakayama of the Reitaku University Library, was partially adopted.
An Alternative Method of Title Authority Control: The Shifted-Authority Control Model for Chinese Classics

Maiko Kimura

School of Library and Information Science, Keio University

Background

Catalogs and authority control for Chinese classics

The Functional Requirements for Bibliographic Records (FRBR) and the IFLA Library Reference Model (IFLA LRM) (International Federation of Library Associations and Institutions 2017), which consolidated FRBR and its related conceptual models Functional Requirements for Authority Data (FRAD) and Functional Requirements for Subject Authority Data (FRSAD), have proposed a new model for library-cataloging. In this model, several related bibliographic records are connected and presented together in groups. Works, the top-level concept of the hierarchy of the bibliographic data, are presented by including groups of several materials with the same content but different titles and/or related materials. This model forms the basis of Resource Description and Access (RDA) (Canadian Library Association and American Library Association, Joint Steering Committee for Development of RDA 2017), a cataloging code adopted by many libraries worldwide.

Applying the work-expression-manifestation-item (WEMI) hierarchy shown in FRBR and IFLA LRM can also be suitable for cataloging rare materials or classic works; Attig (2003) stated that although catalogers of rare materials must generally be flexible in order to address the special characteristics of such materials and their uses, standards and guidelines are still important, as they provide benefits regarding the global sharing of bibliographic information. However, Fell and Lapka (2016) highlighted that "the FRBR assessment does not accurately account for the importance of a number of data elements for the description of rare materials," while Martín and Prada (2016) noted insufficiencies in RDA in terms of instructions for describing special features of rare materials.

Special measures are required for cataloging rare materials, as they have special characteristics. Therefore, since 1981, sets of instructions concerning cataloging standards for rare materials have been advocated by the Bibliographic Standards Committee, Rare Book and Manuscript Section (RBMS), of the Association of College and Research Libraries (ACRL) and the Library of Congress. Further, a series of manuals
entitled *Descriptive Cataloging of Rare Materials* (DCRM), which relate to several types of rare materials, were published between 2007 and 2016. After several years of efforts, an RDA-compatible set of guidelines for cataloging rare materials, entitled *RBMS Policy Statement to RDA* (RBMS PS), was submitted by the DCRM Task Force of the ACRL/RBMS at the 2017 American Library Association Annual Meeting (American Library Association 2017). Thus, RBMS PS is tasked with addressing the abovementioned issues raised by previous researchers concerning cataloging rare materials. Meanwhile, regarding rare Chinese books, which mainly relate to Chinese works published before 1796 (i.e., by the end of the Qianlong era (1736-1795) of the Qing Dynasty, covering a part of the Chinese classics published before 1912), the Council on East Asian Libraries’ Committee on Chinese Materials has been assigned the task of revising the *Cataloging Guidelines for Creating Chinese Rare Book Records in Machine-Readable Form* (published in 2000 and revised in 2009), to ensure that RDA can be applied to rare Chinese materials (Council on East Asian Libraries, 2017). Thus, cataloging and bibliographic control, including for Chinese classics, is clearly trending toward the use of FRBR.

Several projects have applied FRBR or RDA to catalogs of rare materials. Katic (2001) tested FRBR by using this style to present bibliographic records for rare books in Latin and their German and Italian translations, concluding that structuring records using the WEMI hierarchy is sufficient for meeting users’ needs. Meanwhile, RDA was implemented in a project concerning rare materials that was undertaken by researchers from the University of Kent (Caplehorne and Dickinson, 2015), and also in a project conducted by 16 organizations from German-speaking countries (Aliverti and Behrens, 2016). Further, MacDonald and Lawrence (2016) stated that the RDA approaches for collected works, translations, and works issued together are potentially suitable for the cataloging of rare materials. Demonstrating this, Tokita et al. (2012) investigated 158 major Japanese classics and, as each work featured several expressions and a large number of manifestations, they concluded that applying the FRBR model to cataloging such classics is beneficial.

Similar to Japanese classics, there can be many different expressions and manifestations of a work in Chinese classics, each possessing a different title (Downing, 2014). Moreover, it is common for an item to have different titles in different areas (e.g., first page of the main text, center column, and inner cover). Therefore, to meet the demands of users who wish to retrieve all items relating to an intended work, without omission, several access points must be provided in the relevant title authority record; namely, a title authority record must include the authorized title and variant titles of the work.
Utilizing title authority control, a part of the WEMI hierarchy can be applied to the field of Chinese classics. Each bibliographic record is a surrogate of a Chinese classic item, which is "somehow derived from a common progenitor" (Smiraglia and Leazer, 1999), and an authorized title of the work to which the item belongs is also recorded. Further, various bibliographic records relating to the same work (i.e., the progenitor) can be grouped under the authorized title of the work. Even if there are no records of expressions or manifestations, users can at least find Chinese classics at the work-level. An example of the correspondence relationship between the WEMI hierarchy and a Chinese classic is shown in Figure 1. This figure is an example to help readers easily understand the idea of this study. Numerous exceptions must exist, especially in the expression- and manifestation-levels. A strict way to determine these levels for Chinese classics has not been developed yet and should be investigated in future studies.

Figure 1. An example of a Chinese classic with WEMI hierarchy.¹

Current situation of title authority control in Japan and China

Few title authority records exist in Japan and China. The National Diet Library (NDL) in Japan, which possessed 4,678 title authority records (not limited to classical works) as of

March 2017, only conducts title authority control for Japanese classics (National Diet Library, n.d.; 2012a). Further, NACSIS-CAT, an online shared-cataloging system for academic libraries in Japan, operated by the National Institute of Informatics (NII), had 36,078 title authority records (not limited to classical works) as of March 2017 (National Institute of Informatics, 2017). Although Chinese classics are included in the NACSIS-CAT’s title authority files, it is assumed that there are few authority records for Chinese classics, since conducting authority control and creating authority records are not mandatory for member libraries (Miyazawa and Koto, 2012). Further, NACSIS-CAT cataloging guidelines of Japanese and Chinese classics (NII, 2011) have not yet established any rule to construct authorized access points for Chinese classics.

In October 2017, the National Library of China had approximately 60,000 title authority records (Niu, 2017). Further, according to the author’s personal correspondence, in January 2018, the China Academic Library and Information System (CALIS), which is an academic library consortium funded primarily by the Chinese government, possessed 51,079 Chinese title authority records (mainly concerning series titles). Moreover, as part of a project by the Joint University Librarians Advisory Committee, which is a consortium of eight academic libraries in Hong Kong, the Hong Kong Chinese Authority Name Project operates a union database for Chinese records (Joint University Librarians Advisory Committee, n.d.), and in November 2016, this database held over 54,000 title authority records (Xu, 2016). Finally, according to an interview conducted by the author, in July 2013, the National Central Library in Taiwan had 18,252 title authority records (the abovementioned figures are not limited to classical works).

Since 2004, CALIS has especially maintained a database for Chinese classics, which also partially includes images of each item (CALIS, n.d.). However, it seems that the database is not authority-controlled.

Compared to the Library of Congress, which had 526,623 title authority records on June 1, 2013 (Virtual International Authority File 2018), the numbers of title authority records in the abovementioned libraries in Japan and China are very low. Unfortunately, it is impossible to know the exact number of title authority records for Chinese classics in these databases because there is no identifier on the authority records proving that the record is for a Chinese classic.

Instead of title authority control, book catalogs of Chinese classics, in which the works are organized in a special classified order, are used. In addition, a reference book entitled Tong shu yi ming hui lu [The list of different titles for the same book] (Du and Wang
which is designed to facilitate the checking of different titles concerning Chinese classics, has been revised several times and is widely used. However, without title authority data, it is impossible to either conduct comprehensive title authority control for Chinese classics or apply the WEMI hierarchy to such works.

**Features of Chinese classics**

In contrast to modern Chinese materials, Chinese classics have four distinguishing features:

1. *They do not increase.* Although there are a few exceptions (Center for Informatics in East Asian Studies 2009), Chinese classics have primarily been defined as works written by Chinese people in the Chinese language before approximately 1912 (Standardization Administration of the People’s Republic of China, 2008; Japan Library Association, 2006). This is because the Republic of China was established after the Qing dynasty ended in 1911. Therefore, unless new works are discovered, the number of Chinese classics cannot increase anymore.

   There are no statistics on how many Chinese classics exist in Japan. According to Hao (2014), China has about 50 million physical volumes of Chinese classics. Wu (2010) stated that there are about 30 million physical volumes. The Chinese government has implemented the “Chinese ancient books census” project and the exact figure will be known in 2020 (Ministry of Culture of the People’s Republic of China 2017).

2. *Each book is different from others on the item-level.* Chinese classics are, and should be, handled as individual items of a manifestation, with each item being unique, even though they have been printed using the same printing woodblocks. A set of printing woodblocks for a manifestation could be preserved and used many times. With time, a part of a woodblock or even a part of the set woodblocks might become broken, deleted, re-carved, corrected, or altered by another woodblock. Such differences exist among items printed around the same time, implying that each item may have unique texts. In addition, each item has its own item-specific carrier characteristics, such as handwritten marginalia, seals, and binding information. Such information is extremely important and might be the subject of East-Asian studies. Therefore, a bibliographic record has been created for each item in Japan (Japan Library Association (JLA) 2006). Although two items may resemble each other, and may even have no item-specific characteristics, their bibliographic records are maintained separately. This is a common feature among rare materials from other cultural spheres (Attig, 2003).
3. **Chinese classics were also produced in Japan.** Chinese classics, both content-wise and physically, are widespread across the entire cultural sphere of Chinese characters: China, Japan, Korea, and Vietnam. In Japan, for example, many Chinese classics were imported and republished during the Edo period (1603–1867) (Zhou, 2012). As a result, Chinese classics in Japan include not only items imported from China, but also a massive amount of items republished in Japan.

4. **Including numerous variant Chinese characters.** Even in Mainland China and Japan, where people currently do not use traditional Chinese characters in their daily lives, bibliographic data for Chinese classics is customarily recorded in traditional Chinese characters (SAC 2008; CIEAS 2009). This is because Chinese classics must be written in traditional Chinese characters. However, in some cataloging systems, variant characters, such as Japanese *Kanji*, remain in some records. For example, the Universal multiple-octet coded Character Set (UCS) was adopted by NACSIS-CAT in 2000. Before this, when a cataloger could not input a Chinese character to the system because the character set did not cover them, the cataloger used corresponding Japanese *Kanji* characters instead (Yonezawa, 2001). Such use of Japanese *Kanji* characters may cause retrieval omissions.

In addition, variant characters exist among traditional Chinese characters themselves. For example, both “鑒(U+9452)” and “鑳(U+9373)” are traditional Chinese characters with the same meaning and pronunciation, but they are different characters (their Unicode code points (U+XXXX) are different). Chinese classics include numerous variant characters (Li, 2015). Authority control for titles of Chinese classics is very important because each of the several items belonging to the same work sometimes has almost the same title, with a variant Chinese character. Without authority control, this also may cause retrieval omissions.

To avoid such retrieval omissions, the Chinese-Japanese-Korean (CJK) unified index has been introduced to NACSIS-CAT (Miyazawa, 2006). Applying the index, the system automatically broadens the search target and variant characters can also be retrieved. As Miyazawa (2006) mentioned, although it may cause inappropriate results (i.e., noises) because the index tends to unify as many variant characters as possible, the index is valuable because users do not need to worry about variant characters. A similar index is also applied to the Zenkoku Kanseki Database, a dominant database used for retrieving Chinese classics in Japan (Yasuoka, 2002). Currently, at least in Japan, each retrieval system applies its own index with different ranges of variant characters (Nagasaka, 2014). The NDL Search is an integrated search engine of the catalogs and digital resources of Japanese public libraries, the National Diet Library, and other academic institutes (National Diet Library, 2012b). Nagasaka (2014) also pointed out that
the NDL Search only has an index for a small number of variant characters. In addition, since the number of variant Chinese characters is uncountable, such indexes cannot be perfect. There is always the fear of omission or noise.

5. **Having difficulties of identification in manifestation-level.** Items printed from one set of printing woodblocks may be assumed to belong to one manifestation. However, identifying a manifestation is very difficult without sophisticated expertise. For example, a set of newly-carved printing woodblocks are simulant of an older set of woodblocks and sometimes bear a striking resemblance to the older set (Zhou, 2016). In such a case, it is not easy to determine whether two sets belong to the same manifestation or not. Further, it is very difficult to decide whether a part of an item printed from altered woodblocks should be considered as a part of the original manifestation, or if they should be assumed as a new manifestation altogether. Therefore, at least in Japan, it is commonly understood that identification of manifestation is a professional task that should be performed by experts, such as researchers and bibliographers, and not by library catalogers.

Taking the above into consideration, the author proposes a new authority control model for Chinese classics in Japan. Usually, authority control requires links between authority records and bibliographic records. However, even if we could successfully create comprehensive authority data for the titles of Chinese classics, linking them to each bibliographic record retrospectively would be cumbersome. Further, adding the access points for all existing bibliographic records is also nearly impossible without a large labor force. Thus, the proposed model, named the shifted-authority control (SAC) model, pursues improved recall with less expense, allowing users to find more relevant items than ever. The SAC can improve recall more effectively than the current situation in Japan, as it utilizes title authority data, and has less expense than a fully authority-controlled model, as the SAC does not require links between authority and bibliographic records.

Specifically, the purpose of this study is to investigate the effectiveness of the SAC model. Although this study limits its scope to applying the SAC model to bibliographic records in Japan, the model itself and title authority data that the author prepares can also be applied to bibliographic databases of Chinese classics in other regions.

**Methods**

**SAC Model**

Figure 2 shows the current authority control situation in Japan. Under the current situation in Japan, although the bibliographic record ID 54321 is the relevant record, the user cannot find it because it does not include the search word and the title is not
authority-controlled. The title “九章重差” (Jiu zhang chong cha) is a search word used by the user in Figure 2. The work known under the title “九章重差” (Jiu zhang chong cha) is also known under a variant title “海岛算经” (Hai dao suan jing). The two titles are completely different in their Chinese characters, pronunciations, and meanings.

On the other hand, in the normal authority control model shown in Figure 3, a bibliographic record has an authorized access point (AAP) for the work, which is linked to the work’s corresponding authority record. The AAP is assigned to only one authorized title among several different titles for a work. In the authority record, the AAP is recorded. Further, the work’s variant access points (VAPs), which are variant titles that are not selected as the AAP, are also recorded along with the authors’ names, sources of information, etc. In this model, even when a user searches for a title that does not appear in the bibliographic records, the system can suggest it or automatically retrieve an appropriate authority record from the authority database. Through the link between the authority and bibliographic records, the user can obtain all relevant bibliographical records which have the AAP for the work, including the bibliographic record ID 54321 in Figure 3. The normal model (Figure 3) needs the AAP for the work to be added to each existing bibliographic record with links to the work authority records, as well as the creation of authority records. Although some systems work out authority control without links between bibliographic and authority records, AAPs controlled by authority records must be added to bibliographic records. Since titles for Chinese classics had not been controlled in Japan, achieving the normal model requires an enormous amount of labor; this seems almost impossible in Japan.
Figure 3. The normal authority control model.

The SAC model proposed in this study is shown as model A in Figure 4. For comparison, the author also set a model B, the authority records of which are easier to construct, but the omissions in its retrieval results seem to improve.

Model B, named as the independent authority records model, is a more realistic approach for work title authority control for Chinese classics in Japan. Although the SAC
model (model A) needs the construction of authority records, AAPs are not assigned to every bibliographic record. When a user searches for a title, the system searches the authority database first to find the work's AAP and VAP. Then, the system automatically searches the bibliographic database using the AAP and the VAP as search words. In case of Figure 4, the final search results obtained by the user include bibliographic record ID 54321. However, the results may also include some records that are not relevant to the search query. This is because bibliographic records do not have controlled AAP and the retrieval has no choice but to run a keyword search of the entire bibliographic database.

Besides the increasing noise, an anticipated problem with model B is retrieval omissions. Chinese classics may contain variant Chinese characters in their titles as explained in the fourth feature. In case a bibliographic record is recorded in variant Chinese characters that are different from both AAP and VAPs, the record may not be retrieved.

In the SAC model (model A), determining the link between bibliographic records and authority records is not required, and AAPs are not assigned to every bibliographic record, as well as model B. In this model, however, each title authority record contains the several item titles assigned to each work. When a user searches for a title, the system searches the authority database first to find the work's AAP, VAPs, and item titles. Then, the system automatically searches the bibliographic database using those access points and item titles as search words.

Compared to model B, there are two advantages of model A. First, since various item titles and their variant titles are recorded in authority files, it can prevent retrieval omissions caused by variant Chinese characters, even for a system that does not implement an index for variant Chinese characters. Second, since the SAC model refers to the FRBR (now IFLA LRM) model and adopts work and item-level records, users can also directly search the authority database and obtain the representative item titles of each work. Although bibliographic records and authority records are not linked in both models A and B, model A's approach can reveal relationships between the works and items. This is beneficial to users, including librarians, who may seek exact item titles belonging to a particular work.

It is assumed that in both models A and B, once the authority database is constructed, catalogers in libraries will not be required to create or maintain authority records again; this is because the number of Chinese classics cannot increase any further, as noted in the previous section.
Data construction

For the purpose of investigating the effectiveness of the SAC model, first, title authority data were constructed. Specifically, the data construction process was as follows:

1. Choosing an AAP for each work.

   The title authority data used were based on *Tong shu yi ming hui lu* (Du and Wang, 2000). Although *Tong shu yi ming hui lu* provides different titles for each Chinese classic, it does not specify whether titles should be AAPs or VAPs. Therefore, the author selected an AAP for each work, applying the following order of priority. (a) The title that returns the most search results in the Zenkoku Kanseki Database (CIEAS, n.d.) is preferred; (b) a title that is not used in other works is preferred; (c) a title that appropriately represents the subject of the work is preferred. To distinguish between different works sharing the same title, for some works the second priority takes precedence over the first; this removes the risk of AAP duplication.

2. Recording the AAP and VAPs for each title authority record.

   At this stage, a prototype of a title authority database for Chinese classic works, *KWMA-san* (Kimura, 2017), was constructed, and a title authority record was created for each work. Each record contained an AAP selected by the author, while non-preferred titles were recorded as VAPs. The works listed in *Tong shu yi ming hui lu* are written in simplified Chinese characters; however, the author entered AAPs, VAPs, authors’ names, and other information into the database using traditional Chinese characters.

   To obtain item titles for each work, CiNii Books was searched using the corresponding AAPs and VAPs. Then, appropriate bibliographic records were selected and imported into *KWMA-san* through the OpenSearch API of CiNii Books (NII, 2015). This suggests that each work possessed representative item titles. When CiNii Books returned many records, those with titles that differed from other records were selected; however, when CiNii Books did not return results, the item title could not be recorded. Each authority record had between one and 22 unique item titles, with the average being five.

   With regard to the third step, the reasons for choosing CiNii Books as a source of item titles are as follows:
(a) Although there is no authoritative cataloging code for Chinese classics in Japan, usually only one title is chosen and recorded in such catalogs, even if there are various titles for an item. However, because of the cataloging guidelines for Japanese and Chinese classics of NACSIS-CAT (NII, 2011), which is a union-catalog database system for academic libraries in Japan and the source of bibliographic records for CiNii Books, bibliographic records for Chinese classics tend to have various titles in variant title fields. Therefore, it was easy to obtain several item titles from CiNii Books records.

(b) The Zenkoku Kanseki Database usually does not provide variant titles. In addition, it does not provide an API that would facilitate the utilization of its bibliographic records as sources.

(c) NDL Search integrates bibliographic records of Japanese public libraries, the National Diet Library, and other academic institutes. Bibliographic records from public libraries, however, are usually very simple and not suitable for obtaining item information.

(d) The author could not find any other dominant bibliographic database for retrieving Chinese classics in Japan.

*Tong shu yi ming hui lu* has over 13,500 titles (Du and Wang 2000), listed in the order of the stroke counts of the first Chinese character in each title. In this list, each work has at least two titles, and some have more than five; the author has not yet calculated the total number of featured works. Data imputing began with the first page of *Tong shu yi ming hui lu* and, as of January 2018, 270 title authority records have been inputted into KWMA-san.

The models between the WEMI hierarchy and the SAC’s are shown in Figure 5. As IFLA LRM and FRBR are conceptual models, the author merely adopts work and item entities for the data model of the SAC. While the author also assumes that the manifestation entity is useful for identifying Chinese classics, manifestation has not been adopted in this implementation because 1) bibliographic records for Chinese classics are created at the item-level, and not in the manifestation-level in Japan, 2) manifestations of Chinese classics are too difficult to identify for librarians in Japan. Figure 6 presents an example of a title authority record in KWMA-san.
Retrieval experiments

In this study, the author substituted three types of search queries for models A, B, and the current situation in Japan (Figure 2). As model A in Figure 4 shows, in the SAC approach,
when the user searches the bibliographic database, the system automatically retrieves the authority database first and obtains the AAP, VAPs, and item titles of the work needed. Then, the system returns to the bibliographic database and searches it with those access points. Instead of constructing a retrieval system for the SAC model, the author runs a Boolean OR search for the AAP, VAPs, and all item titles (query iii) for 80 works.

CiNii Books, NDL Search, and the Zenkoku Kanseki Database were searched using the following query types: (i) AAP only; (ii) a Boolean OR search for the AAP and all VAPs. Query i represents the current situation in Japan (Figure 2) and query ii represents model B in Figure 4. Two queries are used for comparison with query iii.

Retrievals were conducted for 80 works, which were selected in the order they appear in Tong shu yi ming hui lu. However, works that had no item title (i.e., those that did not return any results from CiNii Books during the third step of the data construction process) were not included in those 80 works, as it was supposed that search queries ii and iii would return identical findings. Similarly, if all item titles for a work corresponded to the work’s AAP or VAPs, the work was also excluded from this experiment, because both query ii and iii would again return identical results.

After each retrieval, the author examined whether individual bibliographic records in the search results for the given work could be considered items of the work. In other words, if an item represented in a bibliographic record belonged to, or was derived from, the progenitor “work” that intended to retrieve, the bibliographic record would be identified as a correct record. When a bibliographic record represents an item that corresponds to various works, the record is judged to be the correct record for each of these works, under the condition that the title representing each work appears in the title, variant title, or note field of the bibliographic record. The correct bibliographic records identified for each work through this process are assumed to be linked to each authority record in the normal authority control model (Figure 3).

By comparing the search results of search query i and ii, the effectiveness of the SAC model could be investigated. Search query i represents a situation in which there is no title authority control, and where a user has searched for a work using only one already-known title. Meanwhile, query ii represents model B (the independent authority records model) in Figure 4.

The AAPs and VAPs presented in traditional Chinese characters were used for retrieval. Item titles, on the other hand, can be recorded in various types of characters in KWMA-san, not limited to traditional Chinese characters because NACSIS-CAT allows all
types of Chinese characters (including Japanese Kanji) for their records. Therefore, item titles searched using query iii were inputted as they appear in the bibliographic records in CiNii Books.

After identification, recall, precision, and F-measure, which is a weighted harmonic mean of recall and precision, were calculated based on the correct records, as follows:

\[
\text{Recall} = \frac{\text{the number of correct records}}{\text{the number of correct records obtained through query iii}}
\]

\[
\text{Precision} = \frac{\text{the number of correct records}}{\text{the number of records retrieved}}
\]

\[
F - \text{measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}
\]

The results obtained by applying query iii were assumed to include the largest number of correct records for the searches in question. Therefore, the recall of query iii was always 100%, and the recall of queries i and ii were determined by comparing their number of correct recalls to those of query iii. It was also assumed that the recall of query iii (100%) is the same as the recall of model A.

Ideally, for each bibliographic database queries i to iii should be identical; however, the Zenkoku Kanseki Database does not allow Arabic numbers, colons, semicolons, and periods in search queries. Therefore, when applying query iii, Arabic numbers that appeared in item titles were changed to numerical figures written in Chinese characters. This change had no practical impact on the search results of CiNii Books and NDL Search since, in both databases, there was no change in the number of retrieved results. Further, for the Zenkoku Kanseki Database, colons in item titles were also deleted in search queries, while semicolons and periods were simply changed to the “+” character, which means Boolean OR search. Additionally, NDL Search does not allow spaces in search
queries, except on both sides of Boolean operators; therefore, spaces were deleted in queries for NDL Search.

Results and Discussion

Summary of retrievals and identifications

Each bibliographic database was searched for each work using the three query types, and then correct records were identified by the author. If any of the three queries returned no records or none that were identified as correct, the recall or the precision of the query in question could not be calculated. Consequently, in such cases all results for the work in question were excluded from the overall results. Therefore, although all 80 works were searched, the numbers of works (n) shown in Tables 1 and 2 differ depending on the bibliographic databases.

As evident in Table 1, the number of retrieved/correct records differs for each work. For example, in CiNii Books, one work returned 1,485 bibliographic records, of which 1,314 were identified as correct. On the other hand, 10 works returned only one bibliographic record, with each being correct.
### Table 1. Summary of retrievals and identifications.

<table>
<thead>
<tr>
<th></th>
<th>Query</th>
<th>$n$</th>
<th>$M$</th>
<th>$SD$</th>
<th>$Mdn$</th>
<th>MAD</th>
<th>MIN</th>
<th>MAX</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Records Retrieved, CiNii Books</td>
<td>78</td>
<td>40.10</td>
<td>172.37</td>
<td>8</td>
<td>8.90</td>
<td>1</td>
<td>1485</td>
<td>7.65</td>
<td>60.50</td>
<td>19.52</td>
</tr>
<tr>
<td></td>
<td>Correct Records</td>
<td></td>
<td>34.35</td>
<td>152.18</td>
<td>7</td>
<td>7.41</td>
<td>1</td>
<td>1314</td>
<td>7.71</td>
<td>61.28</td>
<td>17.23</td>
</tr>
<tr>
<td>2</td>
<td>Records Retrieved, NDL Search</td>
<td>55</td>
<td>41.40</td>
<td>200.33</td>
<td>3</td>
<td>2.97</td>
<td>1</td>
<td>1482</td>
<td>6.76</td>
<td>45.63</td>
<td>27.01</td>
</tr>
<tr>
<td></td>
<td>Correct Records</td>
<td></td>
<td>26.93</td>
<td>113.50</td>
<td>3</td>
<td>2.97</td>
<td>1</td>
<td>830</td>
<td>6.46</td>
<td>42.60</td>
<td>15.30</td>
</tr>
<tr>
<td>3</td>
<td>Records Retrieved, Zenkoku Kanseki DB</td>
<td>73</td>
<td>35.52</td>
<td>46.84</td>
<td>21</td>
<td>26.69</td>
<td>1</td>
<td>267</td>
<td>2.65</td>
<td>8.50</td>
<td>5.48</td>
</tr>
<tr>
<td></td>
<td>Correct Records</td>
<td></td>
<td>33.25</td>
<td>45.42</td>
<td>18</td>
<td>22.24</td>
<td>1</td>
<td>267</td>
<td>2.91</td>
<td>10.20</td>
<td>5.32</td>
</tr>
</tbody>
</table>

Note: The table continues with similar entries for the remaining columns.
**Statistical analysis**

Table 2 presents the means for recall, precision, and F-measure, which were calculated from search results and identifications. In addition, a repeated one-way analysis of variance (ANOVA) was used to compare the mean values between the three query types.

**Table 2.** The means for recall, precision, and F-measure regarding the search results.

<table>
<thead>
<tr>
<th>Bibilographic database</th>
<th>CiNii Books (n = 78)</th>
<th>NDL Search (n = 55)</th>
<th>Zenkoku Kanseki DB (n = 73)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>query</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>i</td>
<td>0.764</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td>0.945</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>iii</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>F(2, 154) = 26</td>
<td>F(2, 108) = 20.5</td>
<td>F(2, 144) = 39.6</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>η²</td>
<td>0.338</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>1-β (α = 0.10%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>query</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>i</td>
<td>0.922</td>
<td>0.907</td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td>0.787</td>
<td>0.795</td>
</tr>
<tr>
<td></td>
<td>iii</td>
<td>0.667</td>
<td>0.658</td>
</tr>
<tr>
<td></td>
<td>F(2, 154) = 26</td>
<td>F(2, 108) = 15.7</td>
<td>F(2, 144) = 52.5</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>η²</td>
<td>0.125</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>1-β (α = 0.10%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>F-measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>query</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>i</td>
<td>0.793</td>
<td>0.841</td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td>0.814</td>
<td>0.813</td>
</tr>
<tr>
<td></td>
<td>iii</td>
<td>0.727</td>
<td>0.703</td>
</tr>
<tr>
<td></td>
<td>F(2, 154) = 3.4</td>
<td>F(2, 108) = 5.5</td>
<td>F(2, 144) = 18.5</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>0.036</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>η²</td>
<td>0.026</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>1-β (α = 0.10%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.36</td>
<td>0.90</td>
<td>1</td>
</tr>
</tbody>
</table>
The ANOVA showed that there was a strong significant difference between the mean values for recall, precision, and F-measure in all bibliographic databases, except for the F-measures of CiNii Books and NDL Search (p < 0.05 and p < 0.01, respectively). Measuring the strength of a relationship ($\eta^2$) is one of the measures of effect size facilitated by an ANOVA; according to Cohen (1988), effect-size magnitudes are small, medium, and large when $\eta^2 > 0.01$, $\eta^2 > 0.06$, and $\eta^2 > 0.14$, respectively. Consequently, it was found that the F-measures of CiNii Books and NDL Search had relatively small magnitudes.

Further, when statistical power (1-\(\beta\)) is less than 0.8, the probability of making a Type II error (i.e., obtaining a false negative) exists (Cohen 1992). It was consequently determined that the F-measure for CiNii Books contained this risk.

**Table 3. Results of pairwise comparisons (Holm’s method)**

<table>
<thead>
<tr>
<th></th>
<th>CiNii Books</th>
<th>NDL Search</th>
<th>Zenkoku Kanseki DB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recall</strong></td>
<td>query ii</td>
<td>query iii</td>
<td>query ii</td>
</tr>
<tr>
<td>query i</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>query ii</td>
<td>-</td>
<td>&lt; 0.001</td>
<td>-</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td>query i</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>query ii</td>
<td>-</td>
<td>&lt; 0.001</td>
<td>-</td>
</tr>
<tr>
<td><strong>F-measure</strong></td>
<td>query i</td>
<td>0.491$^a$</td>
<td>0.243$^a$</td>
</tr>
<tr>
<td>query ii</td>
<td>-</td>
<td>0.02</td>
<td>-</td>
</tr>
</tbody>
</table>

$^a$ Not statistically significant

Next, Holm’s sequential Bonferroni procedure for pairwise comparisons was applied. As Table 3 shows, this procedure also returned a high number of significant differences (p < 0.001) between most of the queries in regard to recall and precision. However, there was no statistical difference between the F-measures of queries i and ii (p = 0.49) and i and iii (p = 0.24) in CiNii Books, queries i and ii (p = 0.50) in NDL Search, and queries i and ii (p = 0.12) in the Zenkoku Kanseki Database.
Factors affecting the results

Since recalls between queries i and iii and queries ii and iii had significant differences, it could be concluded that the proposed SAC model, which is represented by query iii, increases the number of results returned. On the other hand, it also lacks a certain degree of precision. For some query pairs, the differences concerning the F-measures were not significant.

There were more item titles with a semicolon or a period (i.e., items without collective titles) than the author expected (e.g., the second item title (Item id: 610) in Figure 6). In fact, of the 80 works searched, 32 had colons or periods in their item titles. Since the Zenkoku Kanseki Database does not allow semicolons or periods in search queries, the individual titles in these items were Boolean OR searched. This is considered to have decreased the precision of the search results. Further, Welch’s t-tests revealed that for query iii, the difference between the precision rates for searches of 73 general works and of 42 works that did not feature items without collective titles was statistically significant ($t(91) = 2.588, p = 0.011$); in fact, among the latter works the mean precision of query iii in terms of the Zenkoku Kanseki Database increased to 0.763.

In some bibliographic records in CiNii Books for items without collective titles, the second and subsequent titles were recorded in variant title fields. Such a practice is unnecessary, although the cataloging guidelines of NACSIS-CAT do not prohibit it. An item without a collective title usually corresponds to various works that do not share a common progenitor, implying that because of the inclusion of CiNii Books’ records, the authority records in KWMA-san may have included irrelevant item titles. Naturally, such irrelevant item titles decrease the precision of searches of CiNii Books and NDL Search; however, eliminating such irrelevant titles from KWMA-san is impossible, as CiNii Books records do not have any designator to differentiate them from normal variant titles.

Comparing the recalls of queries ii and iii with those of query i implies that the approaches of models A and B are always better than no authority control (Figure 2). Despite there being no link between bibliographic and authority records, the existence and utilization of an authority database can improve recall, and this may help users find more appropriate materials to use. The necessity of title authority control for Chinese classics in Japan should be re-evaluated and the current situation must be changed.

Conclusions

Table 4 shows the several aspects of authority control under the current situation in Japan (Figure 2), the normal authority control (Figure 3), the SAC model (model A in Figure 4), and the independent authority records model (model B in Figure 4). This study examined the efficiency of a new library-cataloging model, the SAC model. It was consequently found that this model could successfully obtain more results without requiring a cataloger to link between bibliographic records and authority records. Assuming that the normal authority control (Figure 3) can achieve nearly 100% recall without any noise, the precision of the SAC model is not preferable. However, the SAC model can avoid retrieval omissions as far as possible. This is highly important especially when a good index for variant Chinese
characters is not implemented as it expands the search results to the greatest extent possible. Further, the authority database with item titles of each work can provide work-item relationships to users, even without links between bibliographic and authority records.

Search results are dependent on bibliographic databases. As the Zenkoku Kanseki Database does not accept periods or semicolons, each part of the title of an item without a collective title had to be replaced with Boolean OR searched, consequently reducing searching precision. This suggests that the SAC model cannot be applied in an unrestricted manner; the search functions of each bibliographic database must be confirmed in advance. Unless the cataloging guidelines of NACSIS-CAT or the item title sources for KWMA-san change, increasing precision will be difficult. In other words, if tightly controlled bibliographic records are applied as item title sources, an increase in precision can be expected.

In RDA, the AAPs and VAPs representing works consist of (a) the AAPs for the agents (such as persons or corporate bodies) responsible for the works, if appropriate, and (b) the preferred titles of the works. This research, however, did not use agents’ AAPs. Contributors to Chinese classics (generally persons who were born before 1912) usually have several names, and those names are normally not authority-controlled in Japan. The difficulties in determining AAPs for Chinese personal names exceeds the scope of this research. Since bibliographic records for Chinese classics in Japan are currently based at item-level, collecting these items and forming manifestation- or expression-level records is also difficult. The author believes creating title authority records at work-level is the first essential step toward addressing these problems.

Taking into account that the precision of the SAC model is not optimal, the SAC model should be considered as merely an alternative to normal authority control and should be applied in situations where such authority control is insufficient. Although the scope of this research was limited to the field of Chinese classics, there is a possibility to apply the SAC approach to classics in other languages which share common features with Chinese classics. On the other hand, for modern materials, including reprints of Chinese classics that have modern covers and title pages, the author strongly recommends applying the normal authority control model.

<table>
<thead>
<tr>
<th>Features</th>
<th>Current</th>
<th>Normal</th>
<th>SAC</th>
<th>Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authority data</td>
<td>None</td>
<td>Need</td>
<td>Need</td>
<td>Need</td>
</tr>
<tr>
<td>Links between authority/bibliographic data</td>
<td>None</td>
<td>Need</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Item titles in authority data</td>
<td>None</td>
<td>None</td>
<td>Need</td>
<td>None</td>
</tr>
<tr>
<td>Recall</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Middle</td>
</tr>
<tr>
<td>Precision</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Middle</td>
</tr>
</tbody>
</table>
Glossary

Chinese classics  Books created by Chinese people, mainly before 1912, in the Chinese language. They include printed books and handwritten manuscripts.

CiNii Books  A union catalog that enables the searching of bibliographic records in academic libraries in Japan. It can be accessed at: https://ci.nii.ac.jp/books/?l=en. Its data source is NACSIS-CAT, and its bibliographic records for Chinese classics are at the item-level.

Item-level record  Bibliographic record created for each item. In Japan, bibliographic records for Chinese (and Japanese) classics are created at the item-level. This means that manifestation-level records are not created. Instead, bibliographic data that should be recorded at the manifestation-level are recorded in the item-level records. An item may consist of multiple physical volumes. In this case, an item-level record is created for multiple volumes. When a volume is a compilation, it includes multiple works. Even so, usually an item-level record is created for the volume. Analytical titles for the volume can be, but not must be, recorded in the note field of the item-level record.

Item title  All titles and variant titles that appears in an item. Item titles for a classic may be equivalent to the titles for the manifestation, although some item titles, such as handwritten titles, may only appear in the item. Item titles do not include analytical titles.

NACSIS-CAT  An online shared-cataloging system for academic libraries in Japan that is operated by the National Institute of Informatics (NII). Each bibliographic record for Chinese classics is created at the item-level in NACSIS-CAT.

References


Du, Xinfu, and Wang, Jian. *Tong shu yi ming hui lu* (The list of different titles for the same book). Nanjing: Jiangsu gu ji chu ban she.


Standardization Administration of the People’s Republic of China (SAC). 2008. *Bibliographical description for ancient Chinese books (GB/T 3792.7-2008).*

doi:10.1080/01639374.2012.681605


