

ISSUES AND SOLUTIONS IN AUTOMATING EAST ASIAN LIBRARIES:
THE DYNIX EXPERIENCE

David Yeo, Dynix Australia Pty. Ltd.,
175 Fullarton Road, Dulwich, S.A. 5065.

Dynix software has been developed by library specialists. The simplicity and thoroughness of the application software are well recognised by librarians. In the April issue of *Library Journal* each year, a survey of installations of automated library systems by vendor in the United States and worldwide is published. The survey delineates growth and trends in the selection, marketing and application of library automation systems. The authors, Robert A. Walton and Frank R. Bridge (library automation consultants) noted that in 1989 Dynix held a "clear lead" in worldwide installations by "capturing a 26 percent marketshare." This performance, they noted, was an increase over the 21 percent record Dynix had set just a year earlier. "The world view", they said, "continues to be strongly focused on Dynix."

Dynix, with a view to supporting more international libraries, has been focusing on enhancing the software application to handle foreign languages, diacritics, and graphic character sets such as Chinese. Dynix, in enhancing its software to support Chinese graphic characters, has also included Japanese and Korean character sets. The Chinese, Japanese and Korean (CJK) version of Dynix was developed concurrently based on the enhancements incorporated in the Dynix foreign language system.

Some issues in supporting CJK vernacular characters in
automated library systems

There are several generic problems that have to be overcome in developing a CJK-based library automation system. The major problems can be categorised as follows:

CJK internal code structure

With a single byte internal code structure it is not possible to represent the several thousand CJK characters required by libraries in cataloguing their CJK collections. A single byte structure affords only 256 possible unique combinations for character representation. Therefore a multibyte per character representation is required to accommodate the several thousand CJK characters needed for cataloguing East Asian collections. Dynix supports a 3 byte

per CJK character structure together with a leading byte to differentiate the 3 byte strings from normal single byte ASCII characters.

Application of MARC format in cataloguing East Asian collections

The MACHine Readable Catalogue (MARC) format was introduced in the 1960s to facilitate the interchange of bibliographic records between automated libraries. The MARC format often takes on national characteristics. Hence the evolution (from USMARC) of various national MARC formats (over 20), such as AUSMARC and NZMARC. The MARC format is machine independent, and primarily concerned with communication of bibliographic records rather than the content of machine-readable records. In using the standard record structure of USMARC 2 ISO-2709, several countries have based their communication format on this broadly defined standard. However, the assignment of tag, indicator and subfield codes are so varied that it is not possible to exchange records without an intermediary "ReMARC" conversion.

The International Federation of Library Associations (IFLA) set up a working committee to look into the standardisation of content designators. This resulted in the development of IFLA MARC, which was later renamed UNIMARC. The UNIMARC format was designed as a carrier format where specifications for tags, indicators and subfields were assigned to machine readable format in order that data could be easily sent and received by different libraries of various countries. Each country could, in theory, have its own national format but in exchanging records, each national format should be translatable into a common MARC format, namely, UNIMARC.

Chinese MARC was developed to expedite library automation in Taiwan and to facilitate the interchange of Chinese language bibliographic records. The Chinese MARC record structure is based on ISO-2709 documentation while content designators and coded elements are based on UNIMARC. The Wade-Giles romanisation schedule is used as the standard transliteration for bibliographic data. Wade-Giles, while being useful in the United States of America, is less applicable in some countries such as Hong Kong and Singapore, where Pin-Yin romanisation is used for transliteration.

Selecting a CJK character code

In Chinese (or East Asian) library automation systems, one important issue that has to be considered carefully is whether the character set employed is large enough to

accommodate present and future cataloguing of library materials. Another issue is whether the character code used adheres to international standards.

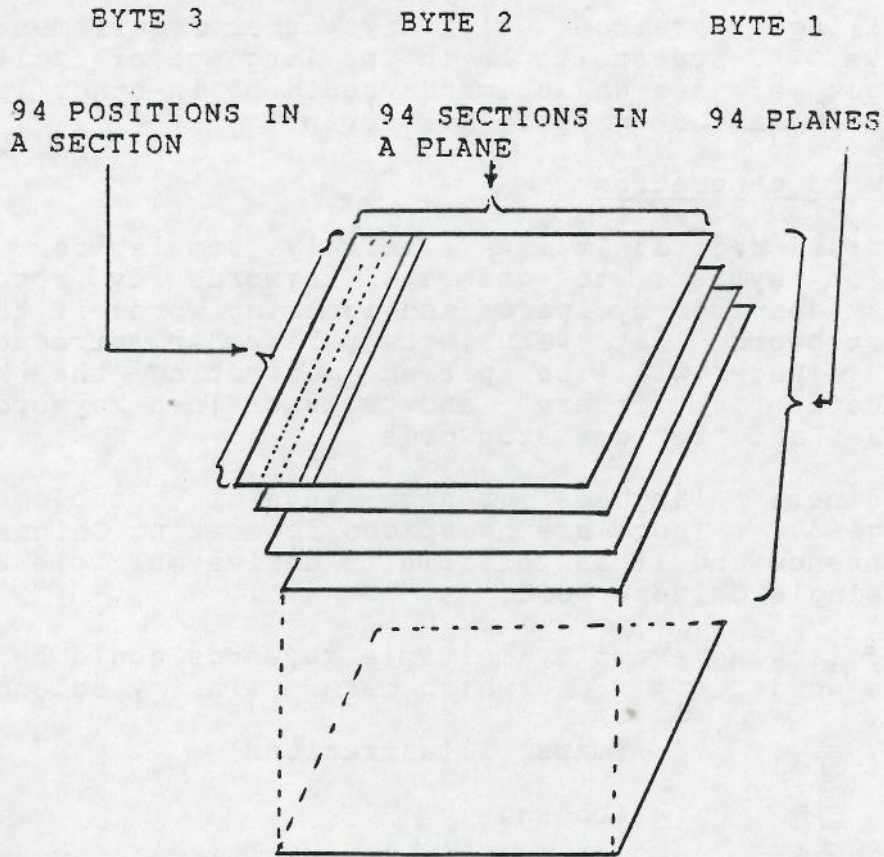
Some characters may exist in the Chinese traditional and simplified form, as a variant Chinese character as well as a Japanese or Korean character. An efficient linkage method must exist to link all these forms one to the other.

Conformity to Anglo-American Cataloguing Rules (AACR) dictates that characters appearing on the title page should be transcribed as is, and not replaced by a more modern or variant form, even if the characters are no longer used. This translates to the requirement that there must be a sufficient number of East Asian characters supported by cataloguing terminals. However, how many characters are sufficient? No authority knows with absolute certainty how many East Asian characters there are.

Implementing a three byte plus one leading byte internal code structure in Dynix

Dynix has adopted the East Asian Character Code (EACC) structure based on CCCII. The leading byte "~" is implemented in front of each EACC character in order that the Dynix software can recognise that the three bytes form an EACC character and that special EACC software routines are invoked. The first byte denotes the position within a section, while the second and third byte denotes the section of a plane and the plane respectively. There are 94 planes with 94 sections in each plane, with 94 positions in each section, giving a three dimensional 94x94x94 structure with a limit of 830,584 unique three byte characters. In the diagram below, byte 1 specifies the plane, byte 2 specifies the section within the plane, and byte 3 specifies the position within the section.

Different planes (byte 1) can be designated for groupings of CJK characters such as traditional Chinese characters, People's Republic of China (PRC) simplified Chinese characters, variants of traditional Chinese characters, Japanese Industrial Standard (JIS) characters, and Korean Information Processing System (KIPS) Hangul characters. Chinese (traditional and simplified forms), Japanese, Korean and variant characters having the same meaning can thus be easily linked by being mapped on different planes (byte 1), but with identical sections and positions (bytes 2 and 3). The internal logic of this 3 byte structure is the basis for powerful indexing and searching.



Online Public Access Catalogue (OPAC) service for a multilingual patron community

OPAC services are an expensive resource that has to be employed efficiently by a library to ensure that as many query transactions as possible are put through the system. The library's OPAC system has to be user friendly and offer multiple access points with powerful features, such as keyword searching, keyword truncation, Boolean operators, related terms, *see* and *see also* references, related works and search limiters. These features would be important in single language OPAC services. The provision of OPAC service to a multilingual patron community poses several other challenges for automated library systems.

A monolingual patron using OPAC services will demand that OPAC screen and commands are presented in the language that the patron is familiar with. An English-only OPAC service

will reduce usage of OPAC terminals in a predominantly Italian, French or a generally non-English patron community.

A multilingual patron will have other requirements. Not only have OPAC screens to be in the language of familiarity, but input of search commands could be in other languages which could include graphic CJK script.

CJK keyword generation

In romanised records it is relatively simple for library automation systems to generate keywords by recognising keywords separated by spaces and ignoring words if they are in a stopword list. Thus, the phrase "An introduction to library science" will be parsed generating the keywords "introduction", "library" and "science" in a keyword index, while "an" and "to" are stopwords.

The Chinese language poses several problems (or challenges!). There are no spaces separating Chinese words in a sentence and it is possible to derive multiple keywords from a single Chinese word.

In the following example, multiple keywords could be derived from the word "圖書館學" which means "library science":

圖	(maps, illustrations)
圖書	(books)
圖書館	(libraries)
圖書館學	(library science, librarianship)

The need to derive several keywords from a single Chinese phrase is a requirement that English library automation systems cannot handle with existing keyword generation and indexing techniques. One possible "solution" that has been adopted by some library automation vendors in Taiwan is to have the cataloguer manually define each keyword that could possibly be used by patrons in keyword searching, a solution not favoured by cataloguers of East Asian material.

The Dynix solution to CJK library automation

The Dynix CJK system is more than just a CJK cataloguing system. It is a fully integrated library automation system. There are limited benefits that can be derived in just cataloguing East Asian materials without an accompanying OPAC service or the ability to process circulation transactions for CJK items. All functions that are

currently supported in the English version of Dynix are also supported by the CJK version of Dynix for the following modules:

- * Cataloguing
- * Circulation
- * Online Public Access Catalogue

The other Dynix modules (Acquisitions and Serials Control) are currently being enhanced to support CJK script for general release in the near future.

Some of the features that the Dynix CJK system currently supports are as follows:

CJK Terminals:

Dynix CJK terminals can be configured with three levels of CJK character support:

Level 1:	#22,000 CJK characters
Level 2:	#35,000 CJK characters
Level 3:	#57,000 CJK characters

Dynix CJK terminals can either be dedicated terminals or CJK workstations using microcomputers with CJK cards and a Hercules display card. In Taiwan, there are specialised CJK terminals for different functions. For example, there is a 19 inch, high resolution 32x32 colour monitor that can be used for OPAC services, normal 12 inch monochrome terminals for cataloguing, acquisitions and serials control processing, and a 9 inch terminal (similar to bank counter terminals) for the circulation counter.

The keyboard used for CJK character creation is identical to the standard QWERTY keyboard with CJK basic strokes etched onto each key.

Internal code structure

Dynix CJK East Asian Character Code (EACC) is based on the Chinese Character Code for Information Interchange (CCCII) 3 byte structure plus an additional leading byte. The leading byte is implemented to enable the system to differentiate normal ASCII data correctly. When no leading byte is detected, the system will process data based on the standard Dynix system programs.

Differentiating CJK data from normal ASCII-based data simply by recognition of a leading byte, allows Dynix to process CJK and romanised records using standard Dynix software with

additional CJK processing software.

Input methods

Dynix CJK terminals allow several input methods. The Tsang-chieh input method is very efficient, allowing quick input of CJK records. However, it is not an easy input method to learn and requires a substantial amount of practise to learn how to create CJK characters. This method is not suited to the novice CJK terminal user and is not friendly for patrons using a CJK OPAC service. OPAC users who are familiar with the romanised form of Chinese will find that it is easy to use the Pin-Yin input method for searching Chinese records. Some other input methods supported by the Dynix CJK terminals are: Jian-yi, Chu-yin fu-hao, Phrase retrieval, Three Corner Code method, Romaji Input system (romanisation of Japanese kana) and Romanisation of Korean Hangul (McCune-Reischauer).

User-defined language selection

Dynix now supports the capability of allowing users to switch online between languages. The SLC command in Dynix allows users to choose the language of the screen display. The attached screens are examples of the Dynix Cataloguing Main Menu (MCAT) screen in Chinese, English and French (please refer to Appendix 1).

As it is time consuming for users to input systems commands in Chinese, all system commands are in English or numeric codes. For example, from MCAT it is easier to input either "1" or "UBR" (for Update Bibliographic Record) as a system command compared with

增補書目資料

Multilingual bibliographic database

Dynix supports multiple languages in a single bibliographic database. Therefore English and romanised bibliographic records can co-exist with CJK records in a single database. Within each bibliographic record, both romanised and CJK vernacular characters will be indexed correctly. For example, the title "BASIC programming" can be catalogued in Chinese as "BASIC 程序" where "BASIC" and "程序" are indexed as keywords.

Full Dynix Public Access Catalogue (OPAC) searches in CJK

All Dynix OPAC features are supported in the CJK Dynix System. Some of these features are:

- * Help Screens and Bulletin Board
- * Alphabetical searches (see Appendix 2)
- * Keyword with Boolean operators AND, OR
- * Search limiters with the following operators:
 - =
 - >
 - <
 - <=
 - >=
 - <>
- * Keyword truncation (Dynix ? and ?? commands) (see Appendix 3)
- * Keyword Title and Author combinations
- * Review Search Help function
- * Synonym list

CJK support in the Dynix cataloguing system

The Dynix CJK cataloguing system supports MARC cataloguing. Some MARC formats that can be supported are USMARC, AUSMARC and UNIMARC. The Dynix CJK cataloguing system has the unique capability of having a system-generated CJK keyword list to assist cataloguers define keywords that are appropriate in building a keyword index. As a Chinese sentence is read in context, it is not possible to have keywords automatically generated by applying normal English language parsing rules of spaces defining separate keywords. The Dynix CJK system "learns" Chinese phrases as CJK records are input and keywords defined. The keywords have to be defined only once and are learned by the system. Should a similar phrase be input again, the keywords previously defined are suggested to the cataloguer for acceptance or rejection. In the example in Appendix 4:

For a圖書館和文化的發展 |e圖書館學研究的前途|

the following Chinese keywords are suggested automatically by the system:

Suggest: 1)圖書館/2)文化/3)發展/4)圖書館學/5)研究/6)前途

Likewise the system suggests: 1)書館學/2)圖書館

as keywords for a|圖書管學

With the Dynix CJK system, as more catalogue records are created, the system becomes "smarter" and is able to suggest an increasing number of CJK keywords automatically. The intelligent generation of CJK keywords is a useful and unique feature of Dynix CJK cataloguing. This compares favourably with systems that do not generate a CJK keyword index or that expects the cataloguer to define every single keyword that is to be built into the existing keyword index.

As in the English version of Dynix, all current cataloguing functions are also supported in CJK. Some of these functions are:

- * MARC record maintenance
- * Full MARC edit features
- * Tag, indicator and sub-field CJK prompts
- * Authority control in Subject, Author, Series and Uniform title
- * Authority maintenance
 - # Update authorities
 - # Global and Global Global changes
 - # Create *See* and *See also* references
 - # Created Related Terms
 - # Merge Authorities

CJK circulation system

All features that currently exist in the English version of Dynix will be supported in the Dynix CJK system. Some features are:

- * CJK patron name and address
- * Display of CJK titles, patron names and messages at Check-out screens
- * Printing of CJK overdue, reservation and recall notices
- * Support of all integrated functions at Check-out and Check-in screens:
 - #.S Search access points
 - #.H Holds (reservations)
 - #.HL List of Holds unfilled
 - #.L List of Items checked out to patron
 - #.B List patron blocks

The Dynix CJK version has been made possible, due to the conscious efforts made by Dynix Australia Pty. Ltd. to enhance the standard version of the Dynix software to handle CJK vernacular characters.

The capability of the software to handle CJK script was not developed in Asia but in Dynix Inc., located in Provo, Utah, USA. There is a "natural endowment" of Asian staff comprising Hong Kongers and Taiwanese within the programming staff of Dynix Inc. Dynix Australia Pty. Ltd. employed and seconded a Taiwanese with experience in CJK library automation development as project leader for development of the Dynix CJK system.

The ability of Dynix programming staff to appreciate the demands of East Asian languages on library automation assisted in the development of features in the Dynix CJK system.

The Dynix CJK project coincided with and has become part of the Dynix foreign language project. The foreign language project is the result of Dynix entering the non-English library automation marketplace with the sale and installation of the Dynix system at the Bibliothèque Municipale D, Angers in France.

[The Chinese text was added with the Duke Chinese Typist word processing software by our editorial staff and printed on a 9-pin printer - Editor]



28 AUG 90

DYNIX 圖書館自動化系統
Cataloging

09:21AM

Select Language Code:

Default Language = EN

Total = 3

CODES

1. EN
2. FR
3. CH

DESCRIPTIONS

English
French
CJK processing

Quit: Up <Cr>:

28 AUG 90

DYNIX 圖書館自動化系統
Cataloging

09:20AM

MENU - CATALOGING (MCAT)

1. UBR Update Bibliographic Records
2. UHR Update Holdings Records
3. BRM Bibliographic Record Merge
4. MAUTH Menu - Authorities Maintenance
5. DHR Discard Holdings Records
6. USI Update Status of Items
7. HSE Bibliographic Short Entry
8. TFM Training on File Maintenance
9. MAIL MAIL - send or review
10. DB Display Bulletin board
11. UC Operator Change
12. LOGTO LOG TO another module

ENTER SELECTION:

28 AUG 90

DYNIX 圖書館自動化系統
編目系統

09:19AM

Menu - 編目 (MCAT - CH)

1. UBR 增補書目資料
2. UHR 增補向編記錄
3. BRM 合併書目記錄
4. MAUTH 權威維護
5. DHR 刪除向編記錄
6. USI 增補向編現況
7. HSE 簡易書目資料輸入
8. TFM FM 使用練習
9. MAIL 電子郵件收發
10. DB 電子佈告欄顯示
11. UC 更換操作人員
12. LOGTO 轉入其他作業模組

請輸入作業項目:

28 AUG 90

DYNIX 圖書館自動化系統
Catalogage

09:20AM

MENU - CATALOGAGE (MCAT - FR)

1. MJB Mise a Jour des notices Bibliographiques
2. MJE Mise a Jour des notices Exemplaires
3. FNB Fusion des Notices Bibliographiques
4. MAUT Menu - Autorites Maintenance
5. SNE Suppression des Notices Exemplaires
6. MJSE Mise a Jour Statut des Exemplaires
7. NBA Notices Bibliographiques Abregees
8. FFM Formation sur Fichier Maintenance
9. COUR Courrier - envoyer ou consulter
10. AF Afficher Flash actualite
11. CO Changement codes d'Operateur
12. CON Connexion a un autre module

VOTRE CHOIX:

APPENDIX 2

28 AUG 90

DYNIX 圖書館自動化系統
流通系統

09:41AM

Your Search: 圖書館

~~Enter a line number for more details~~
~~Alphabetical~~

1. 史奴比歷險記
2. 國會議員部份
- > 3. 圖書館和文化的發展
4. 圖書館員生活型態研究
5. 圖書館員的幻想之二
6. 圖書館在台大的利用情形調查
7. 圖書館學和化學作用

~~Enter a line number for more details~~

Commands: SU = Start Over, B = Back, P = Previous Screen,
<Enter> = Next Screen, ? = Help

(倉韻, C C C I I) 正

倉韻: 人戈十口口

28 AUG 90

DYNIX 圖書館自動化系統
流通系統

09:41AM

Call Number SALLES DE LECTURE ADULTES
020.34

Status: checked in
1 other copy

DYNIX # 10272
AUTHOR 林, 圖書
1) 林, 圖書

TITLE 圖書館在台大的利用情形調查

IMPRINT 台大圖書館編印 1989
COLLATION 321面 有台大圖書館平面圖

~~Press <Enter> to see Copy status~~

Commands: SU = Start Over, B = Back, RW = Related Works,
S = Select for reserve, ? = Help

(英數, C C C I I) 正

APPENDIX 3

28 AUG 90

DYNIX 圖書館自動化系統
流通系統

09:42AM

Your search: 圖書館

AUTHOR/TITLE (truncated)

		DATE
8.	林, 可愛 圖書館學在臺灣的前途	1989
9.	林, 圖書 圖書館和文化的發展	1989
10.	林, 圖書 圖書館在台大的利用情形調查	1989
11.	林, 想作 如果我不是一个圖書館員	1989
12.	林, 想作 如果我不是一个圖書館員	1989
13.	林, 有心 中文圖書館學的前途	1989
14.	林, 有心 圖書館員生活型態研究	1989

----- 20 titles, More on next screen -----

Enter a title number for more detail

Commands: SU = Start Over, B = Back, <Enter> = Next Screen,
P = Previous Screen, ? = Help

(英數, C C C I I) 正

APPENDIX 4

MARC 資料處理中

BIB 編號: 10284
 1. 101 1 achi
 2. 200 10 a圖書館和文化的發展|e圖書館學研究的前途|f林圖書著
 3. 205 a
 4. 210 a|c|d
 5. 215 a|c|d
 6. 225 2 a|v|i
 7. 300 a
 8. 410 t|x|v
 9. 503 1 a|m|n|j
 10. 600 1 a|b|x|y|z
 11. 606 a|x|y|z
 12. 607 a|x|y|z

-----Define CJK phrases-----

\$a圖書館和文化的發展\$a圖書館學研究的前途

Defined :
 NewDef. :
 Suggest : 1)圖書館/2)文化/3)發展/4)圖書館學/5)研究/6)前途/
 File, Quit, Add, Delete, <cr> : _

(倉頡, CCCII) 正

倉頡: 廿十大日

MARC 資料處理中

BIB 編號: 10284
 1. 101 1 achi
 2. 200 10 a圖書館和文化的發展|e圖書館學研究的前途|f林圖書著
 3. 205 a二版
 4. 210 a香港|c大林書局|d1989
 5. 215 a234向|c有圖
 6. 225 2 a圖書館學叢書
 7. 300 a附二十年來圖書館學書目
 8. 600 1 a杜威
 9. 606 a圖書館學
 10. 607 a|x|y|z
 11. 700 1 a林, 圖書
 12. 710 0 a

-----Define CJK phrases-----

\$a圖書館學

Defined : 1)圖書館學/2)圖書館/
 NewDef. :
 Suggest :
 File, Quit, Add, Delete, <cr> : _

(英數, CCCII) 正