

CHINESE-JAPANESE-KOREAN AUTOMATION NEWS

TALIS

Tamkang University, in a joint venture with IBM Taiwan, has developed an integrated Chinese library automation system based on the DOBIS/LIBIS version 1.4 library system. DOBIS/LIBIS can handle all major library functions. It runs on a mainframe IBM 370 or higher with IBM 5550 Chinese workstations as terminals and includes Chinese-English display monitor, standard size Chinese-English keyboard and Chinese-English printer.

Although the 3-byte code for Chinese characters was preferred, the 2-byte code was adopted for reasons of compatibility; development of a 3-byte code version may still take place. Chinese input methods are Tsang Chieh and Pinyin; Japanese is also supported. Records can be retrieved on a Chinese keyword search in author, title, subject or abstract fields, and the date of publication can be in the Gregorian, Chinese, Japanese or Buddhist calendars. Chinese, Japanese or English text can be mixed in any field and is searchable among mixed text. In addition, LC MARC and OCLC MARC tapes can be loaded through a conversion program.

From: Shen, I-Ming. "The development and implementation of TALIS - an integrated Chinese library automation system."

教育資料與圖書館學

= Journal of Educational Media & Library Sciences, v. 26(1)
Autumn 1988; pp. 41-57.

See also:

淡江大學發展 DOBIS/LIBIS/TALIS之經驗

= The development of DOBIS/LIBIS/TALIS: the Tamkang University Experience."

教育資料與圖書館學

= Journal of Educational Media & Library Sciences, v. 25(4)
Summer 1988; pp. 468-488.

Readers may be interested in the Research Libraries Information Network (RLIN) record format and in details of RLIN's new microcomputer-based Multiscript Workstation.

RLIN(™) NON-ROMAN BIBLIOGRAPHIC RECORDS

(text from RLIN circular of October 1988, courtesy of RLG)

Since September 1983, a number of East Asian libraries or library departments, together with the Library of Congress, have been entering records into the Research Libraries Information network (RLIN) database that contains Chinese, Japanese and/or Korean characters.

In 1986, Cyrillic script was added -- specifically, the modern alphabets of the Russian, Belorussian, Bulgarian, Macedonian, Serbian, and Ukrainian languages -- and, in 1988, full Hebraic script support was added. In addition to these three RLIN enhancements, RLG is investigating supporting the script used to write Arabic and other languages.

The purpose of these developments is two-fold: (1) to extend the advantages of automation to important special collections and their users; and (2) to integrate information about these collections into a mainstream of research information. To do this, The Research Libraries Group (RLG) has developed the capacity to enter, store, transmit, search, and display bibliographic records that contain non-Roman alphabets.

This paper provides a simple description of how non-Roman scripts are treated in RLIN.

A non-Roman record can exist in any of the RLIN bibliographic files and has the same basic structure as any other RLIN record: it contains tagged fields with indicators and delimited subfields. It differs from other records, however, in having one or more non-Roman fields. A non-Roman field contains at least one character in a non-Roman script. Non-Roman fields often contain a mixture of Roman and non-Roman script languages.

In a non-Roman record, identical information exists in pairs of fields, where the romanised version of a field is followed immediately by the non-Roman version of the same field. (Note that the fields paired with non-Roman fields contain the romanisation of the script, not a translation.) This parallelism is necessary to give all RLIN users, even users whose terminals cannot display non-Roman scripts, access to the content of non-Roman records.

Only bibliographic fields from 100 (main entry - personal name) to 899 (local series added entry - uniform title heading) in the

USMARC formats may contain non-Roman data. Like other RLIN records, non-Roman records also require bibliographic control data plus holdings, archival control, or acquisitions information. All such information is entered in the Roman alphabet (usually in English).

Depending on their contents, bibliographic fields in RLIN records fall into three categories:

Description - the most basic information about the title, imprint, and series that usually appears in the central block of a catalogue card. For non-Roman materials, this data is often entirely in non-Roman script and is entered into core fields that must be paired with their romanised equivalents.

Access points - indexed information such as subject headings or added entries that can be used to retrieve records. When non-Roman access points are included, a record or record cluster is retrievable by a non-Roman search. This is particularly useful when the romanised form of an access point is uncertain.

Other, unindexed fields - additional information about an item, such as contents notes. Unindexed fields, too, may be in Roman or non-Roman scripts, or a mixture of both.

With a few exceptions, the following core fields must occur in both romanised and non-Roman form whenever they are present in a non-Roman record:

- Title statement (USMARC 245 field)
- Edition statement (250 field)
- Imprint statement (260 field)
- Series statement(s) (4xx fields)

Because the basic information in RLIN records must be accessible to all users, whether or not they have a terminal capable of displaying non-Roman scripts, the system does not allow the presence of a non-Roman core field without its romanised counterpart. However, core fields may appear only in English (or another Roman-alphabet language) when that is how the information appears on the material being catalogued. The system warns users when they enter non-Roman records with any solitary Roman alphabet core fields, to ensure that the single occurrence of the field is intentional and not accidental.

Fields outside the set of core fields may be entered in romanised form without paired non-Roman fields, but a record's value is

much greater when parallelism is maintained as far as possible. The paired fields provide better record access and enhanced record displays.

If all the indexed fields are given in both romanised and non-Roman form, users can enter search requests in Roman or non-Roman characters for any of the following:

title words; title phrases; related titles;
personal names; corporate names/words; corporate
names/phrases; subject phrases; subject division

This increased access particularly aids scholars and reference librarians. The benefits are especially clear in the case of Chinese and Korean names. Often, in their original form, these names use distinctly different characters, which RLIN indexes; but they are romanised in very similar ways. For example, at least eighteen different Chinese surnames are romanised as "Yu", making a Roman-alphabet search containing this term cumbersome at best. A search by specific character ensures that only records containing that character will be retrieved.

In addition, users of RLIN's CJK^(TM) (Chinese-Japanese-Korean) capability can use a set of unique Chinese character indexes to retrieve records containing character variants; for example, records containing either the traditional or the PRC simplified form of the same character.

RLIN allows users to display portions of records in different formats, depending on the information wanted for various purposes -- collection development, cataloguing, reference searching, etc. The non-Roman fields in a record affect its presentation on these formatted displays to users at terminals able to handle non-Roman scripts. The "Multiple results", "Primary cluster member", "Partial data", and "Long" (catalogue-card-like) displays show only the non-Roman form of fields when these exist. The "Full" display, which alone shows the entire record with each field preceded by its RLIN/MARC tag, is the only one that presents both the fields in a pair. Users at terminals that cannot display non-Roman scripts will, of course, see only the romanised fields in a record.

In the RLIN database, records held by different users for the same edition of a work are clustered together. This means that such records are retrieved as a group in response to search requests, and the user has the option of displaying and comparing the various records entered for an edition. Non-Roman records are clustered in the same way as all other records: they are grouped together if they have the same or similar values (depending on the field) for LC card number, international standard book number, title, publisher, publication date, pagination, and edition statement. The matching is based on the

romanised, rather than non-Roman fields. Thus, records containing non-Roman scripts can cluster together with completely romanised records for the same edition.

Users entering records containing non-Roman scripts employ the same commands and procedures as all other RLIN technical processing users:

"create" to do new entity cataloguing; that is, to add records for titles or editions not already represented in the database;

"derive" to add records based on records belonging to other institutions;

"update" to add more holdings, to convert acquisitions date to full cataloguing, or to otherwise update a record.

Being able to derive cataloguing copy from Roman-alphabet records is a tremendous advantage in non-Roman cataloguing, since users need only add the non-Roman paired fields to produce their own records. When users at non-Roman terminals start from records that already have non-Roman data in them, they are saved even this amount of effort.

At present, catalogue cards printed at RLG's central offices do not include any non-Roman scripts. Users with terminals that can display non-Roman scripts can also use local non-Roman printers to produce hard copy of non-Roman records. These can be printed out in any of the RLIN record display formats preferred. Users can also request their non-Roman records on magnetic tape in MARC communications format, where the non-Roman data is so tagged that it can be used or ignored by programs processing such tapes.



RLG MULTISCRIPPT WORKSTATION
(from circular rev. 02/7/89, courtesy of RLG)

The RLG Multiscript Workstation (MSW) supports bibliographic control and information retrieval on RLIN using Chinese, Japanese, Korean, Hebrew and, Cyrillic scripts.

The MSW is based on an AT-class personal computer configured with:

- one 40 MB hard disk drive, one 360K floppy disk drive, and one 1.2 MB floppy disk drive;
- a Wyse-700 high-resolution monitor/display system;
- a CJK character-generator read-only memory board developed by JHL Research Inc., of Anaheim, California;
- SDLC communications adaptor;
- an ASCII extended keyboard for use when CJK input is not required;
- a CJK keyboard, produced by Transtech International Corporation of Natick, Massachusetts, with a simplified Chinese character component input method;
- RLIN PC terminal software, including the Cyrillic enhancements RLG implemented in May 1986 and the Hebrew script support implemented in January 1988;
- an Epson LQ 850 printer (optional).

Features:

MSW users may retrieve and display any of the approximately 500,000 RLIN bibliographic records containing Chinese, Japanese or Korean scripts and use them as the basis for copy cataloguing, as well as consult the RLIN CJK^(TM) Thesaurus, an on-line, multi-indexed dictionary showing information on all the East Asian characters currently supported. The RLIN CJK Thesaurus may be consulted while creating or updating a record to look up Chinese characters by radical, stroke count, Wade-Giles or Pinyin romanisation, kana readings, modified Hepburn romanisation, or McCune-Reischauer romanisation.

The optional "CJK PC keyboard" also allows users to "build" CJK

characters from component parts engraved on the keys, in the order normally used to write the character. A character selection display window pops up on the screen as soon as a Chinese or Korean component is pressed; the user may select the desired character from the display window or enter more components until the character is completed.

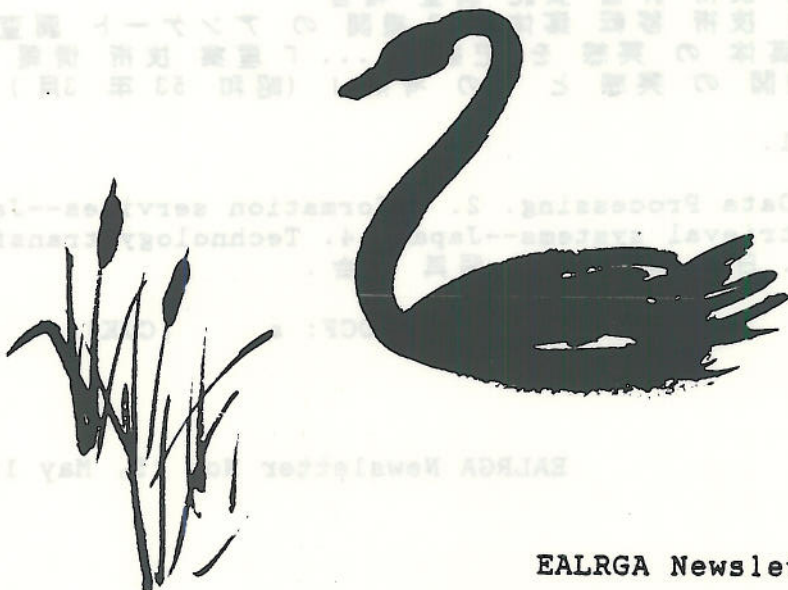
The MSW comes with a standard PC keyboard, with which users can retrieve, display, and copy from CJK records. Blocks of text can be stored and copied into other fields or records, subject to no limitation other than system memory. Users may download records (including CJK and other non-Roman scripts) onto disk or directly into a local system. The MSW can also be used for stand-alone functions when it is not being used as an RLIN terminal.

A CJK version of PC-Write for the MSW that supports Chinese, Japanese and Korean scripts as well as the ALA-extended Roman character set, developed under contract by JHL Research Inc., is available. A CJK interface with other software packages such as dBase III and Lotus 1-2-3 will be announced separately.

Cost:

The cost of the complete Multiscript Workstation -- system unit, ASCII keyboard, CJK PC keyboard, CJK ROM board, RLIN terminal software, and Epson LQ850 printer -- costs US\$6,066 FOB Natick, Massachusetts. (An MSW for reference use or copy cataloguing, without CJK PC keyboard, costs US\$4,566.)

For more information, please contact the RLIN Information Center, The Research Libraries Group, Inc., 1200 Villa Street, Mountain View, Ca 94041-1100, USA, phone 800-537-RLIN.



CJK RECORD EXAMPLES USING RLIN MULTISCRIP WORKSTATION

(courtesy of RLG)

TEMP Serials LON CSU087-S111 Search CRLG-
FIN SP CHINA - ECONOMIC CONDITIONS# ALS LG CHI ALS GEN A - Cluster 7 of 8
+B

政治 成績 統計 .

Feb.-Mar. 1931, June 1932, Mar. 1933-Dec. 1935, May-July 1936.

-- 南京 : 中國 國民黨 中央 統計處 , 1931-1936.

12 v. : tables ; 20 x 27 cm. ; 28 cm.

Monthly.

At head of title: Chung-kuo kuo min tang chih tao hsia chih
Microfilm. Stanford, Ca. : Hoover Institution, 1987. 3 reels.

1. China--Statistics. 2. China--Politics and government--1927-1937--
Periodicals. 3. China--Economic conditions--1927-1937--Periodicals. 4. China--
Foreign relations--1927-1937--Periodicals. I. 中國 國民黨 . 中央 統計處 .

035: (CSt-H)SN00051

035: (CSt-H)SN00052

035: (CSt-H)SN00053

ID: CSU087-S111

CC: 9114

DCF: a

[CJK]

CALL: 4030.1\1352

TEMP Books LON HAU086-B1100 Search CRLG-
FIN SP TECHNOLOGY TRANSFER#JAPAN ALS LG JPN - Cluster 6 of 11
+B

産業 技術 情報 システム の 研究 : 調査 結果 の 総括 と 技術 移転 促進策 .
東京 : 日本 産業 技術 振興 協会 , 昭和 54 [1979]

iii, 171 p. : ill. ; 26 cm.

Cover title.

Errata slips inserted.

"昭和 53 年度 工業 技術院 技術 評価 委託 調査 報告"

"昭和 52 年では, 官民の 技術 移転 媒体 118 機関 の アンケート 調査 および
面接 調査 から 技術 移転 媒体 の 実態 を 把握し ... 「産業 技術 情報 システム
の 研究 -- 技術 交流 仲介 機関 の 実態 と その 考察」 (昭和 53 年 3 月) を
とりまとめた"

Bibliography: p. 167-171.

1. Corporations--Japan--Data Processing. 2. Information services--Japan. 3.
Information storage and retrieval systems--Japan. 4. Technology transfer--Japan
I. Japan. 工業 技術院 . II. 日本 産業 技術 振興 協会 .

ID: HAU086-B1100

CC: 9114

DCF: a

[CJK]

CALL: HD2907.S3642 1979