DOCUMENT RESUME

ED 337 180 IR 053 754

AUTHOR Li, Tze-chung
TITLE Library Automation in the Republic of China: Practical Aspects and Perspectives.
PUB DATE May 91
NOTE 26p.; Paper presented at the International Conference on the New Frontiers in Library and Information Services (Taipei, Taiwan, May 8-12, 1991).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Access to Information; *Bibliographic Databases; Evaluation Criteria; Foreign Countries; Information Dissemination; *Library Automation; *Online Catalogs; Online Searching; *Online Systems; Records Management; User Satisfaction (Information)
IDENTIFIERS *Taiwan

ABSTRACT

This paper evaluates major library automated systems in the Republic of China, explores their strengths and weaknesses, and provides suggestions for further consideration. Nine representative systems were selected for this purpose: the National Central Library Automated Information Service, Science and Technology Information Center of the National Science Council, Agricultural Science Information Center, Legislative Information System, the Academia Sinica databases, National Defense Medical College, Tamkang University, National Taiwan Institute of Technology, and Chung Koh Library Products Company CAS system. The strengths and weaknesses are evaluated on the bases of their ease of use, speed of searching, and precision of results. In general, the library automated systems have been remarkably designed and well implemented. However, the lack of standardization in coding, data input, and machine readable record formats are major problems and constitute a barrier to implementing an effective networking system. It is expected that further progress will come with the standardization of Chinese coding and the simplification of Chinese input. (25 references) (Author/MAB)

Library Automation in the Republic of China:

Practical Aspects and Perspectives

Tze-chung Li
Visiting Professor
National Taiwan University
Department of Library Science

Paper presented at the
International Conference on the New Frontiers
in Library and Information Services
May 8-12, 1991
National Central Library
Taipei, Taiwan, Republic of China

Library Automation in the Republic of China:

Practical Aspects and Perspectives*

Tze-chung Li
Visiting Professor
National Taiwan University
Department of Library Science

Abstract: This paper is intended to evaluate major library automated systems in the Republic of China, explore their strengths and weaknesses, and provide suggestions for further consideration. Nine representative systems were selected for this purpose. The strengths and weaknesses are evaluated on the bases of their ease of use, speed of searching, and precision of results. In general, the library automated systems have been remarkably designed and well implemented. However, the lack of standardization in coding, data input, machine readable record formats are major problems and constitute a barrier toward implementing an effective networking system. It is expected that further progress will come with the standardization of Chinese coding and the simplification of Chinese input.

Development of library automation in the Republic of China can be traced to the early 1970s. The first use of automation began with an IBM 1120 computer for the book catalog of Western materials at the Department of Physics, National Tsing-hua University in 1972.[1] For the materials in Chinese, an

* Revised version of a paper delivered at the International Conference on the New Frontiers in Library and Information Services, May 8-12, 1991 in Taipei, Republic of China

integrated library system of acquisitions, cataloging, and circulation was experimented on a Wang MVP minicomputer in 1975.[2] In 1979, the Universal Database Access Service (UDAS) was introduced to provide access to major database vendors in the world. The year 1980 proved pivotal when the Chinese Character Code for Information Interchange (CCCII) was completed, and a working group was established to develop the Chinese MARC.[3]

Since the mid-1985s, progress in library automation has advanced steadily. The National Cheng-chi University began in 1984 its automated circulation system designed by the First International Computer which later developed a FIC integrated library system. In 1986, the National Cheng-kung University main library developed, though rudimentary, its own integrated system for circulation, cataloging, and acquisitions. A survey of library automation released in 1988 indicates that the National Central Library, the Academic Sinica, nine universities, four colleges, twelve special libraries and two producers have implemented, developed, or tested their library automated systems in varying degrees.[4]

Automated library systems have been introduced from abroad or developed domestically. The systems introduced from abroad include DOBIS/Leuven, URICA, UTLAS , and recently INNOPAC. In addition to programs developed by libraries themselves, the following programs commercially available may be mentioned: FIC (First International Computer) Integrated Library System, TOTALS (Technology Opulent TRANSTECH Automated Library System), R.P.T. Intergroups International Integrated Library System, and Chung Koh Library Products Company's CAS.

This paper is intended to evaluate the major automated systems in the Republic of China, explore their strengths and weaknesses, and suggest directions for further consideration. Nine representative library automated systems were selected for this purpose. They are: National Central Library Automated Information Service, Science and Technology Information center of the National Science Council, Agricultural Science Information Center, Legislative Information System, the Academia Sinica databases, National Defense Medical College, Tamkang University, National Taiwan Institute of Technology, and Chung Koh Library Products Company CAS system.

Representative Library Automated Systems

National Central Library Automated Information Service (NCLAIS)

National Central Library Automated Information Service (NCLAIS) is the most ambitious library project in the Republic of China. In 1980, the National Central Library and the Library Association of China jointly developed the Chinese Library Automation Planning Project with four main objectives: (1) to develop the Chinese MARC format; (2) to organize the data processing system for Chinese language materials; (3) to create a database for Chinese publications and introduce foreign databases; and (4) to establish a national information network.

The National Central Library has developed four main databases on the Wang SV series computer: acquisitions, catalog, serials control, and index to Chinese periodical literature. Of the four, three are currently in use: acquisitions, catalog, and

index. The acquisitions database consists of basically four sub-files, namely acquisitions, budget control, publishers/suppliers, and statistics. The statistics file provides statistics data on the quantity, cancellation, budgeting, performance of suppliers, and average price. There is no breakdown by disciplines.

The index database indexes both Chinese periodicals and gazettes published since 1983. About 1,000 items of periodicals and gazettes are regularly reviewed for inclusion. The database is using two-byte for the Chinese character. Conversion from two-byte to three-byte is in a testing stage. The index database is not yet available for public access. Its print counterparts are *Index to Chinese Periodical Literature* and *Index to Chinese Official Gazettes*.

The best and currently fully implemented database is the on-line catalog. The database is using the Chinese MARC which has been well developed by the National Central Library. The Chinese MARC format was structured on the basis of UNIMARC with modifications to meet local requirements. Full descriptions of bibliographic records in the Chinese language conform with the Chinese Cataloging Rules and, to a large extent, with the International Standard Bibliographic Descriptions and the second edition of the Anglo-American Cataloging Rules.[5] The print version of the Chinese MARC, first published in 1982, is now in its third edition.[6] The on-line catalog database provides the source for publishing the *Chinese National Bibliography*, a monthly with annual and multi-year cumulations.

The on-line catalog of the National Central Library provides a dozen or so access points, such as title, author, Chinese

subject heading, English subject heading, Chinese classification, Library of Congress classification, Chinese printed card number, SuDoc number, ISBN, ISSN, Library of Congress card number, and accession number. In addition, four qualifiers (language, publisher, publication date, and format) are available for restricted searching. Exact set of Chinese characters can be searched with Boolean logical operators. The response time is, however, slow. An average of twenty seconds is required for a simple search. For a search statement that includes more than one set of Chinese characters, search results frequently require more than two-minutes. Four methods can be used for inputting Chinese: three-corner, *chu-yin* (phonetic symbols), *ts'ang-chieh* (radical index), and simplified *ts'ang-chieh*.

The National Central Library is currently developing cooperative on-line cataloging on Tandem, purchased in 1989, as a centralized processor connected with work stations in participating libraries. The participating libraries will convert their new book bibliographic data into the centralized processor and make the National Central Library the national bibliographic utility.[7] It may be noted that the Tandem source code has not been made available to the National Central Library, thus delaying the transfer of data from Wang to Tandem for co-operative on-line cataloging.

Science and Technology Information Center Network (STICNET)

In 1973, the National Science Council established the Science and Technology Center (STIC) with an aim at providing a favorable environment for science and technology development and

research in the Republic of China.[8] STIC maintains a collection of monographs, periodicals, proceedings, theses, patents, and other materials; is responsible for centralizing purchase of publications with the National Science grants to over forty-six universities, colleges and research centers; publishes monographs; and provides abstracting and indexing services. Thirteen or so indexes and abstracts are currently published by STIC. They include *Sci-Tech Policy Review*, *Sci-Tech Briefs* in eight areas, *Ongoing Research Projects of the ROC*, *Index of Sci-Tech Journal of the ROC*, *Abstracts of Sci-Tech Journals of the ROC*, *Index of Domestic and International Sci-Tech Conferences and Seminars*, *Index of Sci-Tech Research Reports of the ROC*, *Abstracts of Sci-Tech Research Reports of the ROC*, *Abstracts of the Doctoral Dissertations and Index of Master's Theses*, *Union List of Non-Chinese Sci-Tech Serials in the ROC*, and *SDI Bulletin*.

Of particular importance, STIC provides information services that include literature search, translation, international exchange and sale, referral, document delivery, and support of research projects services. In 1986, it developed STICNET, a nationwide sci-tech information system, which went on-line on December 28, 1988. The purpose of STICNET is to purchase and create databases to be shared by participating libraries. At present, it consists of five foreign databases (BIOSIS PREVIEWS, CA SEARCH, COMPENDIX, EPIC, INSPECT, and NTIS) and seven Chinese databases (UNION LIST OF NON-CHINESE SCI-TECH SERIALS IN THE REPUBLIC OF CHINA, UNION CATALOG OF NON-CHINESE SCI-TECH BOOKS IN THE REPUBLIC OF CHINA, SCI-TECH JOURNAL ARTICLES OF THE REPUBLIC

OF CHINA, ONGOING RESEARCH PROJECTS OF THE REPUBLIC OF CHINA, SCI-TECH RESEARCH REPORTS OF THE REPUBLIC OF CHINA, SCI-TECH BRIEFS, and NSC AWARDED RESEARCH PAPERS ABSTRACTS). STIC is using the Wang VS series computer which can be accessed by IBM PC or its compatible as a terminal via one of three telecommunication channels: PACNET, leased line, and dial-up. More than eighty institutions are now subscribing to STICNET.

STICNET is basically a menu driven system. Documents can be searched through limited access points, such as author, title, abstract, subject, and publication date.[9] In addition, free-text, proximity, and truncated searching are available.

STICNET may be considered the database of databases for sharing resources at a fee. It is more economical than connecting with the individual database directly. However, STICNET is not without problems. First, the response time is unreasonably slow. For instance, it took more than five minutes to retrieve postings in the ERIC file on the subject of "library science" and "automation." Second, searching can not be aborted when in process, although the system will not proceed search operation when a searched term appears in more than 60,000 postings. In searching Chinese materials, the user may use any one of the three input methods: three-corner, *chu-yin* (phonetic spelling), and *ts'ang-chieh* (radical index).

Agricultural Science Information Center Databases

The Agricultural Science Information Center (ASIC) of the Agriculture Association of China was created in 1973 with the purpose of providing information on agricultural science and

technology. The Center has two major functions: (1) photo-reproduction of abstracts, and (2) setting up the Agricultural Science and Technology Information Management System (ASTIMS). The following abstracts are available for reproduction: *CAB Abstracts*, *Food Science and Technology Abstracts (USTA)*, *Aquatic Science and Fisheries Abstracts (ASFA)*, *Biological Abstracts*, *Chemical Abstracts: Biochemistry Section*, *Biotechnology Abstracts*, and *Pollution Abstracts*.

The ASTIMS consists of four databases: (1) Agri-Thesaurus,[10] (2) Files for Agricultural Science and Technology Personnel (FASTEP), (3) Files for Agricultural Science and Technology Research Projects (FASTEJ), and (4) ASIC MARC Bibliographic System. The Agri-Thesaurus, frequently updated, contains about 20,000 terms in both Chinese and English. FASTEP contains about 16,000 personnel data of the agricultural labor force involved in policy-making, administration, research, and education in the Republic of China. In FASTEJ, nearly 24,000 records from 1973 to the present have been entered.

The latest and the largest database is the ASIC MARC Bibliographic System that began its function in 1983. The database contains over 70,000 citations of books, periodical articles, proceedings, theses and monographic contents. It regularly indexes over 500 periodicals. The system also serves as a union catalog of the Council of Agriculture Library and the Asian Pacific Food and Fertilizer Information Center.

The ASIC MARC Bibliographic System provides multiple access to its records. There are twenty-one access points that include personal name, corporate name, conference name, title, series

title, uniform title, variant titles, author/title, title keywords, AGRI-THESAURUS keywords, LC subject heading, Sears subject heading, CODEN, ISBN, ISSN, and so on.[11] Nineteen qualifiers are also provided for restricted searching. These include publication date, record type, abstract, thesis, biography, government publication, conference, index, language, location, target audience, chronological coverage, illustration, contents, physical medium, frequency, regularity, serial type, and index availability. Other features of searching include, for instance, EXPAND to list AGRI-Thesaurus terms, HELP and H/E to display command tables, SA to search for authority records, and SB to search for bibliographic records. A summary of search statements can be displayed too. There are, however, shortcomings with the system. Free-text, truncation, and string or proximity searching are not available, Boolean logical operators cannot be used to combine controlled vocabularies, and a search operation can not be aborted.

Three commands are used to display the search result: to review current records (bibliographic citations), to display full bibliographic records, and to print full bibliographic records either in card or MARC format. Three-corner, *chu-yin* (phonetic spelling), and *ts'ang-chieh* (radical index) methods are used to input Chinese. The user may use either one of the three.

The Legislative Information System

The Legislative Information System (LEGISIS), developed in 1985, consists of four databases: interpellation statistical data, interpellations, Chinese code, and legislative periodical

literature.[12] The statistical database contains data since 1984. Data are divided according to session, interpellator, type, category (such as defense and education), person who responds, and so on.

The interpellation database contains some 20,000 items since 1984. The records can be searched in a number of access points such as subject, reference, meeting and session number, interpellator, category, nature of interpellation, the person who replies, date of reply, format of reply, and so on.

The Chinese Code database, updated weekly, consists of four sub-categories: text, legal terms, main points, and classified catalogue. It may be noted that the classified catalogue is a simple listing of the contents of three major compilations: *Chung-hua-min-kuo fa-lu wei-pien mo-lu (A Compilation List of Laws in the Republic of China)*, *lu-fa chuan-shu (A Complete Work of Six Laws)*, and *Chung-hua-min-kuo hsien-hsin fa-kwei wei-pian (A Compilation of Current Laws in the Republic of China)*.

The legislative periodical literature database, updated weekly since 1985, periodically reviews 95 titles for inclusion. It is an index to about 10,000 items. Each record is divided by fields and all fields can be searched, such as subject, title, author, keyword, category, journal title, and date. There are no abstracts.

The Legislative Information System uses both menu and command searches.[13] There are nine commands to retrieve data. Essential ones for a simple search include FIND for searching, EXPAND for listing multiple-words; RECALL to give a summary of search statements; HELP for on-line help; S for combining set

statements; and QUIT to quit the system. A number of access points are provided, as, for example, in the legislative periodical literature database, a user may search data by subject, author, title, journal name, date of publication and so on. Boolean logical operators and truncation are available and the number of truncated words can be specified. But in free-text searching, multiple words or a set of words not single word must be used. Boolean logical operators are executed according to the sequence of entering. Nesting must be used for priority of execution. The system is not capable of aborting search process, but the user can stop searching by exiting to the main menu.

There are three result procedures: BROWSE for displaying bibliographical records, DISPLAY for listing full bibliographical record; and PRINT. PRINT has three formats: full record, bibliographical record, and title.

The Legislative Information System is a well-developed system. It will be a significant reference tool if its Chinese Code database could include abstracts of the Supreme Court precedents, interpretations of the Judicial Yuan, and the uniform interpretations of the Council of Grand Justices.

Academia Sinica Databases

The Academia Sinica began operating its Computing Center in 1986. The Center has developed a number of databases and is engaged in nine joint on-going projects of databases, including, for instance, AUTOMATIC CHINESE SENTENCE ANALYSIS, MING-CHING DOCUMENTS AND RECORDS, TZU-PU AND CHI-PU, SHIH-SAN-CHING AND ITS NOTES, and TAIWAN-TI-FANG-CHIH.[14] At present, databases

available for public use include full-text databases, master's and doctoral theses in the Republic of China, and works of researchers in the Academia Sinica.[15]

Of the full-text databases, two deserve particular mention: ELECTRONIC DICTIONARY and HISTORY OF TWENTY-FIVE DYNASTIES. ELECTRONIC DICTIONARY or KUO-YU RHY-PAO TIEN-TSE TS'U-TIAN, a counterpart of the printed version, *Kuo Yu Daily News Dictionary*, contains about 40,000 entries of vocabulary, pronunciation, and definition. Syntactic analyses including morpheme, part of speech, attribute, description, and cross references are provided for each entry.[16] One welcome feature is its capability of semantic browse of related terms.

The history of twenty-five dynasties database, based on the printed version published by Chung-hua Book Company in mainland China, contains over seven million words of *The Historical Records*, *The History of the Han Dynasty*, *The History of Later Han*, *The History of the Three Kingdoms*, *The Economic Records of the Twenty Four Books of History*. The history database is based on the Chinese Text Processor (CTP) system which features both free-text and controlled vocabulary searching.[17] The controlled vocabulary refers to a single key-word set that includes the name of person, the name of place, the name of official titles, the name of books, the expression of time in terms of an era, a dynasty or an emperor and ten classes of special terminologies.[18]

Both the dictionary and history databases can be searched by using individual Chinese characters. However, searching of words in the electronic dictionary database is limited to the field only.[19] One helpful feature is that the search operation can be

aborted before it is completed.

In the history database, after the user enters characters for searching, the system displays on the screen such information as seconds spent, number of characters searched, number of sections and terms, and asks whether the user wants to see the search result.[20] For Chinese input, two methods are available: *Chu-yin* (phonetic spelling) and *ts'ang-chieh* (radical index).

National Defense Medical College Library

The Library began plans for automation in 1985 and in 1989, on-line cataloging and circulation became fully operative to the public.[21] The system uses UTLAS with modifications for Chinese characters. Automated acquisitions and serial control are to be included in the near future. The on-line public access catalog (OPAC), updated daily, is using LC MARC. It contains more than 30,000 titles in English and Chinese of monographs, serials, and A-V materials.

The OPAC provides seven access points: title, author, subject, corporate author, series, key-word, and individual borrower.[22] Free-text searching is available, but in logical operations, only AND can be used and it cannot be executed in one search statement. For instance, the user asks the system to search for "psychiatry" and "patient," the system will display postings for each term. The user has to combine the two search statements, if he wants to do so. It must be noted that in OPAC, only a term, i.e., a set of characters, except if it is a controlled vocabulary, can be searched. The user cannot abort the search operation. However, if an overflow occurs, the system will not

proceed but inform the user to limit the search. The result procedure involves three levels: brief - a simple citation, medium - bibliographical citation plus series note, and full bibliographical record. The full bibliographical record includes related subjects, general note, ISBN, etc., in addition to the medium display.

Tamkang University TALIS System

If an integrated automated system is considered to includes at least the handling of acquisitions, cataloging, serials control, and circulation, Tamkang's TALIS (Tamkang Automated Library Integrated System) is perhaps the first, fully operative system in Taiwan. Implemented in 1986, TALIS is operated by the DOBIS/Leuven program modified for Chinese characters. Data structure is based on LC MARC. TALIS is basically a menu driven system. It displays three choices: (1) search the catalog; (2) display your borrow record; and (3) send a note to the library staff, after it is logged on.[23] If the user selects number 1 (search the catalog), the system displays fifteen numbered choices, such as author, title, subject, publisher, classification number, ISBN/ISSN, accession number, call number, and so on. The system is using the KWIC approach. For instance, if the term "library science" is searched, the system displays all titles in alphabetical order that contain the words. Boolean operation is currently not provided for public use, but it can be installed together with qualifiers for publication date and type. TALIS uses both ts'ang-chieh (radical index) and chu-yin (phonetic spelling) for Chinese input.

National Taiwan Institute of Technology

The institute is a relatively new one established in 1974. In July 1990, it adopted TOTALS (Technology Opulent TRANSTECH Automated Library System) for its circulation and cataloging processes. TOTALS commercially available in 1989 can be used for Chinese, English, Japanese, and Korean languages. It is a menu driven system. There are eight access points: title, author, subject, keyword, call number, command, year, and language. Some fields, such as the year and language fields, can not be used alone; they must be used in combination with other fields. The command access point is used to combine numbered search results, and, as such Boolean logical operators can be used. However, the system does not have the free-text search capability. If the postings exceed 500, the system will alert the reader to modify his strategy to narrow the search.

For Chinese coding, it uses REACC (RLIN East Asian Character Code) that is close to CCCII (Chinese Character Code for Information Interchange) noted earlier. It is capable to covert REACC to CCCII. Both US MARC and Chinese MARC can be used. *Chu-yin* or TRANSTECH radicals is used for inputting Chinese. One drawback of TOTALS is its bulky keyboard with some 160 keys that is not convenient to use. It has been reported that a standard keyboard will be marketed soon.

The Chung Koh Library System

Chung Koh Library Products Company has developed a library integrated system for small and medium-sized libraries. The system consists of three sub-systems: CCS-100 for circulation, CAS-200 for acquisitions, and CCS-300 for cataloging and public

access catalog. The three sub-systems became fully operative in 1987. Cataloging structure is based on Chinese MARC. The Chung Koh system uses *ping-yin* (phonetic spelling) for Chinese input. Its on-line public access catalog is a menu driven system. A user may search by either author, title, subject, or call number. Free-text searching is not available.

Evaluation of Library Automated Systems

Evaluation of the effectiveness of a library automated system adheres to four basic guiding principles: (1) ease of use, (2) speed of search operation, (3) precision of search results, and (4) quality control of data. Ease of use concerns primarily the methods of information retrieval. All systems are menu driven. A few provide limited command search features. The menu approach provides a streamlined display of pre-determined menu items for the user to choose. The user simply follows the displayed instructions to take a next step, yet the menu approach involves repeated steps and lacks the flexibility and volatility of free-text searching. Since each system was developed with consideration to local requirements, a user will encounter a diversity of menu instructions.

The difficulty of grasping the Chinese input poses another problem. Currently, there are a number of programs that offer different methods of input. Already mentioned are *ts'ang-chieh* using radicals; three-corner, analyzing three corners of a character; *chu-yin*, using phonetic symbols; and *ping-yin* (phonetic English alphabets). Other input systems include, for example, TianMa, Wei Wu Hanzi, EasyWord, and Ideogramem. TianMa, produced by International Geosystems Corporation in Canada, uses

Chinese ping-yin for Chinese text processing. Wei Wu Hanzi, developed by Kam-fu Wong and marketed through Star Paging (Chinese Computer) of Hong Kong, uses a combination of shape and stroke as well as *ping-yin* for Chinese input. EasyWord, invented by Ron Tomas of Australia, relies upon eight strokes to input Chinese character. The eight strokes are dot, horizontal stroke, vertical stroke, left sloping stroke, right sloping stroke, horizontal angle stroke, vertical angle stroke, and left sloping angle. But its writing sequence does not always conform with the Chinese tradition. Hsin-chu Lee's Ideogramem, patented but not yet marketed, employs eighteen strokes for inputting Chinese. The strokes, equivalent to the English alphabet, are used on the basis of writing sequence to form the Chinese character. Lee claims that his method can uniquely display a Chinese word with the exception of two characters. Each input method requires considerable time and effort to learn to use.

Speed is a main factor for considering automation. It is generally considered that a simple search result should not take more than five seconds. As noted earlier, a simple search at the National Central Library can take twenty seconds, and it can take more than two minutes to search two terms with a Boolean logical operator. The STIC system requires more than five minutes to complete a simple search statement with Boolean logical operators. The slow processing is a definite drawback. In some systems, aborting a search is not an option.

The computer's processing speed will result in large recall. The precision of search is a primary concern to any

user. The computer's query language enables the user to communicate with the computer. The user receives feedback from which to modify or restructure the strategy, change approaches, give new directions, and refine questions as the search proceeds. An effective information retrieval system enables the user to broaden or narrow his search, to search any word in the text, or to restrict search by form, date, language, and other conditions. All these features are not seen in the library systems noted above. As well pointed out by Lucy Lee, the use of already developed programs with modifications to meet the local requirements will be more economical and less time-consuming.[24] Although it is preferable to have a system that meets the particular needs of a library, setbacks, delays, and problems can be expected from programs independently developed from scratch.

It may be noted that in free text searching, not a single character but a set of characters can be searched in some systems. There are pros and cons about this approach. One argues that the meaning of a single Chinese word differs from the meaning of a combination of the word with another word.[25] For instance, the term "information education" is represented by four Chinese characters: *tse*, *hsin*, *chiao*, and *yu*. Each of the four Chinese characters has more than one meaning. *Tze* and *hsin* have seven different meanings each, whereas *chiao* has five and *yu* has six meanings. Only the four characters grouped together into two sets mean information education, that is *tse-hsin* (information) and *chiao-yu* (education). It is suggested that to search a set of Chinese characters is more meaningful than to search a single character. The Legislative Information System is using sets of

characters, instead of individual character, for search terms. The Agriculture Science databases do the same.

The argument for using sets of characters is persuasive and has its merits. However, different views exist. In the first place, many single Chinese characters have unique meanings, particularly in the Chinese classic. Text what contains a character of interest will prove elusive, if only sets of characters in the text can be searched. Secondly, all sets of characters with same and related meanings have to be entered for large recall. For instance, the English word "law" has its Chinese equivalents in many sets of characters, such as *fa-lu*, *fa-kwei*, *fa-ti*, *fa-lin*, *fa-chieh*, *fa-tiao*, and so on. If any one of these has not been designated as a set of characters, it will be missed in search. The ideal situation would be that in which both single characters or sets of characters can be searched with proximity searching capabilities to ensure recall relevance.

Concerns with data quality are perennial. There are data errors and data inconsistency that include document errors, editorial errors, record errors, and inconsistency in indexing.[27] It appears that each database needs to provide a mechanism for name authority, title authority, error detection, number correction, and arithmetic checking.

Suggestions for Further Consideration

In the last decade, we see an imposing array of achievements in library automation in the Republic of China. Some systems have been well implemented; others are still in the stage of testing and experimenting. The four objectives for developing

Chinese library automation formulated in 1981 by the National Central Library and the Library Association of China, noted earlier, have been largely fulfilled. Indeed, the National Central Library has played a leading role in the development of library automation. When completed, its planned national information network will be a milestone towards bibliographical control and resources sharing. Improvements still can be made in the technical areas of searching, such as free-text searching, positional operators, speed in response, and capability of aborting search. There remain two main problems that require immediate attention: diversity of Chinese coding and variety of Chinese input methods.

The diversity of data coding appears to be a roadblock for library networking. Ta-I Huang reviewed several coding systems that include the Japanese industry standard X0208-1983, the Chinese mainland GB 2312-80, Big-Five, Standard of Popular Chinese Character Interchange Code, Unicode and so on.[26] The author concluded that the best coding system is the Chinese Character Code for Information Interchange (CCCII).[27] Unfortunately, CCCII, though developed in 1981 and later used in the Chinese MARC, has never been accepted as the Chinese coding standard.

A benefit from diversity is the competition for excellence. But, in many aspects, it serves no meaningful purpose. After more than a decade of developing and implementing the Chinese coding, it is time for uniformity and standardization. If voluntary cooperation proves fruitless for addressing the issue, the government should take the helm to achieve standardization.

The problem of various Chinese input methods may be solved by developing a non-keyboard method. The user may use methods of putting data into a computer other than the use of a keyboard. For instance, a user writes Chinese characters with a stylus on a special slate. The handwriting would be translated into the Chinese coding and sent to the computer. The user need not learn the use of different input systems.

In the library automation systems, our colleagues in the Republic of China have made some remarkable achievements, such as the development of Chinese MARC. It is expected that further progress will come with the standardization of Chinese coding and the simplification of Chinese input.

Notes

1. Lucy Te-chu Lee, "A study of the automated library information system development in Taiwan, R.O.C." *Library Association of China Bulletin*, pp. 107- 23, 120 (1988).
2. Margaret C. Fung. *On Library and Information Science* (Taipei: Student Book Company, 1982), p. 135.
3. Ching-chun Hsieh and others, "The design and application of the Chinese Character Code for Information Interchange (CCCII)." *Journal of Library and Information Science*, 7(2): 129-43 ((1981); Lucy Te-chu Lee and others, "Chinese MARC: its present status and future development." *Journal of Library and Information Science*, 7(1): 1-18 (1981).
4. Lucy Te-chu Lee, "A study of the automated..."
5. Lucy Te-chu Lee and others, "Chinese MARC: its present

status..."; Lucy te-chu Lee, *Library Automation and Information Science: A Collection of Essays* (Taipei: Sino-American Publishing, 1985), pp. 20.

6. *Chung-kuo chi-tou pien-mo ké-shih (Chinese Machine Readable Cataloging Format)* 3d ed. (Taipei: National Central Library, 1989. 401p.); *Chung-kuo chi-tou pien-mo ké-shih fu-lu (Chinese Machine Readable Cataloging Format Supplement*. 3d ed. Taipei: National Central Library, 1989. 167p.)

7. Nancy Ou-lan Hu Chou, "The development of the national bibliographic database and network." *Journal of Library and Information Science*, 14: 195-207 (1988), 201-205.

8. For a brief description of STIC, see *Science and Technology Information Center* (Taipei: the Center, 1990 [pamphlet, no pagination].)

9. *Ko-chi-hsin chuan-kuo tse-hsin wung-luo shih-tung shou-ch'e (Science and Technology Nationwide Networking Operating Manual)*. 2d ed. (Taipei: STIC [no date; no pagination]

10.) Its print counterpart is *Lun-ye ko-chi so-yin (AGRI Thesaurus)* (Taipei: ASIC, 1988. 988p.)

11. *ASIC MARC shu-mo shih-tung pien-mo tse shih-tung shu-ju shou-ASIC MARC System Cataloging Input Manual)* (Taipei: ASIC, 1987 [no pagination].)

12. *Li-fa tse-hsin fu-wu chien-chieh (An Introduction to Legislative Information Service)* (Taipei: Legislative Yuan, Legal Materials Center, 1990. 13p.)

13. *Li-fa tse-hsin shih-tung shih-yung shou-t'se (Legislative Information System Operating Manual)* (Taipei: Legislative Yuan, Legal Information Center, 1988. 74p.) It contains a thesaurus of

legal terms (pp. 56-74).

14. For a good description of the Academia Sinica databases, refer to Margaret C. Fung, "The use of source materials enhanced by modern information technology," *Journal of Library and Information Science*, 14:158-77 (1988); *Computing Center: A Brief Sketch* (Taipei: Academia Sinica, Computing Center, 1990. 21p.)

15. *Computing Center...; Kuo-Yu ti tse-lai fan-cheh (Analysis of Chinese Terms)*. Rev. ed. (Taipei: Academia Sinica, Computing Center, 1989. 73p.)

16. Ching-chun Hsieh, "Full text processing of Chinese language." *Journal of Library and Information Science*, 11: 125-42 (1985).

17. Ibid., 131.

18. Ibid.

19. Huang Bai-p'ei. *Kuo-yu rhy-pao tien-tse ts'u-tien shih-yung shou-ts'e (Kuo-Yu Daily News Electronic Dictionary Operating Manual)* (Taipei: Academia Sinica, Computing Center, 1989. 104p.)

20. Tien Meng-lun. *Erh-shih-wu-shih chuan-wen tse-liao-kuo shih-yung shou-ts'e (History of Twenty-Five Dynasties Database Operating Manual)* (Taipei, Academia Sinica, Computing Center, 1989. 99p.)

21. *NDMC On-line Public Access Catalog* (Taipei: National Defense Medical College Chih-teh Medical Library [no date]. 35p.)

22. *Tsu-tung-hua chieh tse-hsin fo-wu chien-kai (A Brief Introduction to Automation and Information Service)* (Taipei: National Defense Medical College, Chih-teh Medical Library, 1989 [brochure].)

23. *Tamkang Automated Library Integrated System On-line Public*

Access Catalog User's Manual (Tamsui, Taipei: Tamkang University, Chueh Sheng Memorial Library, 1989. 187p.) 24. Lucy Te-chu Lee, "A study of the automated library information...," p. 122.

25. Min Ku, "A comparative study of word indexing between Chinese and English database." *Journal of Information and Micrographic Management*, 7: 8-11 (1987).

26. Anne P. Mintz, "Quality control and the zen of database production." *Online*, 14(6): 15-9 (1990); MaryEllen C. Sievert and Mark J. Andrews, "Indexing consistency in Information Science Abstracts." *Journal of the American Society for Information Science*, 42(1): 1-6 (1991).

27. Ta-i Huang. *Chung-wen tse-ma [Chinese Code]* (Taipei: Chang-song wen-hua shih-yeh Co., 1990), pp. 29-82.

28. Ibid., pp. 83-85.