

AUTOMATION OF EAST ASIAN COLLECTIONS IN THE U.S.: CURRENT STATE AND PROBLEMS

*Kwok-Kan Tam**

ABSTRACT

One challenge facing East Asian librarians and computer specialists is the efficient input of East Asian materials in their original non-roman languages. It is difficult to design a character set that will suit the different needs of Chinese, Japanese and Korean. Each language has developed an independent character set tailored to the needs of its community.

The over 200 plans and methods for automating Chinese language materials fall into four major categories: the phoenetic encoding method, the numerical encoding method, the individual character encoding method, and the component-parts (radicals) encoding method. Besides the technical aspects of automation of East Asian collections in the United States, there are the additional concerns involving costs, personnel, and re-organization of library administration.

If library services are to meet the public's increasing information needs, the automation system should perform at least three functions: circulation, cataloging, and bibliographic searching. Today, RLG/RLIN's CJK (Chinese-Japanese-Korean) terminal can perform the first two of these three functions. The introduction of the RLG/RLIN system, and possibly in the near future an OCLC one, will surely revolutionize work flow and practices in East Asian library service.

* Kwok-Kan Tam, post-doctoral Fellow, East-West Center, Institute of Culture and Communication, Honolulu, Hawaii.

Library automation has now become a trend beyond any doubt. Its many advantages have persuaded even traditional librarians to believe that computers, in many cases, can often perform better than human beings, and in the long run can help reduce the cost of library service. What has been puzzling to East Asian librarians and computer experts is how to input East Asian materials in an efficient way and in their original non-roman languages. Traditional computer terminals are designed for the use of roman alphabets. The three most popular and important languages in East Asia—Chinese, Japanese, and Korean, however, are non-roman systems that cannot be processed by conventional computers.

Facing the technological changes in library and information services, East Asian librarians are eager to seek a break through in their conservative situation. As Professor Ching-chih Chen has pointed out:

It is clear, then, that community-based information services must be problem- and need-oriented. Libraries, in their current mode of information delivery, should realize that they are neither the only nor the most important source of information in our society. We must realize, and realize soon, that the library of the future will be more than a repository of printed work. If the library is to survive, it must be the public's main access link to a network of knowledge containing all types of information in all types of formats available through all types of information providers. These are the challenges to libraries in the 80's. (7)

In the past ten years, most East Asian libraries in the United States have mainly functioned as warehouses of books rather than as active information service organizations, not to mention as joint information networks. This is chiefly due to the lack of a computer system that can handle East Asian materials in a satisfactory manner. To paraphrase Professor Chen, if East Asian collections are to survive the challenge of the computer age, they

have to become a member of the information society.

Today conventional computers can only be used as a help-mate in the circulation of East Asian materials. For instance, at the University of Illinois Library at Urbana-Champaign, the third largest American university library system, the LCS (Library Computer System), a conventional computer with a design based on roman alphabets, can do nothing about inputting East Asian materials in their vernacular. To solve the problem, a new kind of computer has to be devised. In the past five years, computer experts have been working on the possibility of inputting and outputting East Asian language materials in their own forms. There has been very slow and unsatisfactory progress, mainly because the inputting of Chinese characters, which is an indispensable component of both the Japanese and Korean languages, cannot be done on a keyboard which is simple enough for non-specialists to handle.

To complicate the matter, the Chinese character has been simplified in Japan and on the Chinese mainland, while in Taiwan the traditional form is preferred. Even the same Chinese character may be written in various ways. Hence, it becomes difficult to design a character set that will suit the different needs of the three East Asian languages. In the past several years, China, Japan, and Taiwan have been experimenting with their own systems. They have developed their character sets independently, and have tailored them for use within their own communities.

Over 200 plans for automating Chinese language materials have emerged. All of these different plans and methods, however, can be subsumed into four major categories, each representing the use of a particular method of encoding the Chinese character for computer use: the phonetic encoding method, the numerical encoding method, the individual character encoding method, and the component-parts (radicals) encoding method.

The phonetic encoding method is still under vigorous research. Yet, so far there is no satisfactory result. Its basic idea

is to transliterate the different East Asian languages into roman alphabets so that they can be put into a conventional computer. But because there are so many words in Chinese that bear the same sound, it is unwise to transcribe an ideogram into a phonetic symbol. Chinese characters are pictorial in their peculiar way of signification. Once they are transcribed into phonetic symbols they do not reveal their meanings.

The numerical encoding method is also not a good means to input Chinese characters. Its fundamental principle is to divide the physical set-up of Chinese characters into basic component parts and designate them with arabic numerals. Unfortunately, there are too many Chinese radicals to designate, and the transcription of Chinese characters into strings of numerals will be a tedious task. The individual character encoding method is even more difficult, for it requires an enormously large keyboard that can contain more than 5,000 characters.

The only method that has been intriguing computer experts is the radicals encoding method. It is in this direction that both the RLG/RLIN (Research Libraries Group/Research Libraries Information Network) and the OCLC (Online Computer Library Center) are working. RLG/RLIN's approach involves the creation of a video terminal with 179 keys, development of a 15,600-character online dictionary, and extensions to RLIN programs for record creation, indexing, retrieval, display, modification, and output. OCLC will be marketing a radically different system, developed by Asiagraphics Corporation (Mt. Sinai, NY). The OCLC system allows users to retrieve transliterated Chinese records from the OCLC database, and, using software developed by Asiagraphics, display on the terminal screen or print-out these records in pictographic characters using a dot-matrix printer attached to the OCLC M300 workstation (a modified IBM PC). The software will support both the Pinyin and the Wade-Giles schemes of transliteration as well as full and simplified character sets of the Chinese language. ("Automation Update: RLG vs.

OCLC," 8-9). The OCLC system for East Asian languages, which is otherwise perfect, is not yet ready for use, whereas the RLG/RLIN has taken the lead with nearly twenty large East Asian libraries as members.

The radical encoding method used by the RLG/RLIN is based on the principle of dividing Chinese characters into radicals. Inputting is made possible by combining the radicals to form Chinese characters, full or simplified. Although RLG/RLIN is leading in automating East Asian language materials, their system is not without fault. Its basic design is for use by persons who are well trained in East Asian languages. In other words, the system needs a specialized operator, who knows the language and also the computer commands well enough not to waste money in computer sharing. It is evident that only people with native command of the language will be able to handle it. In this case, it will be almost impossible to popularize its use with general library users.

On the other hand, the OCLC system, though falling behind the RLG/RLIN system, has a better potential for development and for popular use, for "OCLC users will be able to retrieve Chinese language records from an OCLC M300 workstation without a special Chinese-character keyboard." ("Automation Update," 9) A terminal operator will be able to use the system after a few hours' training, and will not have to learn to use a Chinese character keyboard with 179 keys as one does with the RLG/RLIN system.

In view of the growth of information services in libraries and the public's increasing needs, it is expected that the automation system, if perfect, will be able to perform at least three functions: circulation, cataloging, and bibliographic searching by the patrons. Today, with the RLG/RLIN's CJK (Chinese-Japanese-Korean) terminal, only the first two of these three functions can be performed. If there is no breakthrough in the inputting of Chinese characters, it will be difficult for East Asian libraries to keep pace

with the general developments in an information society. It remains to be seen if the OCLC system will have the functions it promises.

Besides the technical aspects of automation of East Asian collections in the United States, there are also other problems involved in costs, personnel, and re-organization of library administration. At present the RLG/RLIN terminal costs around US\$40,000, and the one-time training fee costs around US\$3,500 to 4,000. For small-sized collections, these are enormous figures. As for personnel support, specially trained terminal operators will be able to perform the job of searching more efficiently than other librarians. This will require the addition of new staff members, or re-training of original staff. In the past year, some large East Asian collections, such as Harvard and Columbia, have recruited public service/reference librarians who have to handle computerized bibliographic and non-bibliographic searching. The biggest problem that East Asian libraries will face, in the long run of automation, is the re-organization of administrative and staff responsibilities.

Most medium- or large-sized East Asian libraries in the United States have a self-sufficient, half-autonomous structure of administration with their own acquisition, cataloging and public service units. The introduction of the RLG/RLIN system, and possibly in the near future the OCLC one, will surely revolutionize the work flow and practices in East Asian library service. The most obvious effect will be seen in cataloging, which the present development of computer terminals has begun to affect. With the establishment of the RLG/RLIN, less and less original cataloging is required. Hence, some catalogers might have to switch or re-orient themselves to other duties. In this case, there might be a possibility that more manpower could be spent in information/reference services.

Although automation of East Asian collections in the United States has fallen behind that of roman languages, it is optimistic

that they are heading in a right direction. East Asian libraries will survive in the 1990s and will not be excluded from the fast growing information networks.

References

- Agenbroad, James E. "Character Sets: Current State and East Asian Prospects." *Journal of Library Automation* 13 (1) (March 1980): 18-35.
- "Automation Update: RLG vs. OCLC." *Chinese American Librarians Association/Midwest Chapter Newsletter* (April 1984): 8-9.
- Becker, Joseph. "Libraries, Society and Technological Changes." *Library Trends* 27 (13) (Winter 1979). 409-417.
- Chen, Ching-chih. "Golden Opportunities in the 80's for Information Professionals." *Journal of Library and Information Science* 8 (1) (April 1982): 1-19.
- Miller, R. Bruce. "Nonroman Scripts and Computer Terminal Developments." *Information Technology and Libraries* 1 (2) (June 1982): 143-148.
- Morita, Ichiko. "Japanese Character Input: Its State and Problems." *Journal of Library Automation* 14 (1) (March 1981): 6-23.
- Perushek, Diane E. and Donovan, Maureen. "East Asian Library Automation: General and Aspects." *Committee on East Asian Libraries Bulletin* 70/71 (Feb./June 1983): 60-63.
- Tucker, Alan. "The East Asian Project of the Research Libraries Group." *Committee on East Asian Libraries Bulletin* 69 (Oct. 1982): 12-20.
- Wang Chi and Dunn, Robert. "Chinese Character Computer Encoding Methods and Library Automation: A Bibliographic Survey." *Library of Congress Information Bulletin* 41 (3) (March 1982): 95-100.
- Zais-Gabbert, Harriet. "Library of Congress RLG/CJK Project Update." *Committee on East Asian Libraries Bulletin* 72 (Oct. 1982): 58-62.