RUBILA --- A CONCEPT IN RULE-BASED
INTEGRATED LIBRARY AUTOMATION†

Karl K. Lo*  Zy-Kaan Ding**

ABSTRACT

The current trend in library automation systems is to emulate long established manual operations of dividing the library tasks into subsystems of acquisitions, cataloging, circulation, serials control, data management and so forth. To enter each of these subsystems requires human intervention and a different command to logon. Within each subsystem there may also be a number of functional modules, such as card printing and accounting, that require human decisions to activate. Many of these human interceptions in the automation process may be eliminated through the use of rules built into the system's software, thus achieving a higher degree of library automation.

Beginning with the 1950's, librarians have been tapping the awesome power of the computer to perform some of their more routine tasks. In the earlier years, prices for hardware and software were high, and the operation of the computer was restricted solely to the programmer and the system analyst. Large and wealthy libraries often had a team of these specialists in a system office to control the computer operations. Machine time was

* Karl K. Lo, University of Washington, Seattle, Washington, U.S.A.
** Zy-Kaan Ding, Academia Sinica, Taiwan, R.O.C.
more expensive than a smaller library could afford. Automation in those days belonged to the elite. Products of those earlier days were simple, being mostly hardcopy printouts from 80-column punched cards sorted in different ways. Simple though the products were, the programs to produce them were complex. Programming and production were done only by people in the system office.

Today, computers are cheaper in price and are a million times more powerful than the early vacuum-tube dinosaur. A computer with power comparable to the dinosaur of the 1950's is compact enough to sit on top of a desk. Computers are prolific in number. They can be found in homes, offices and, of course, the library. People are no longer satisfied with a library computer system but want a library system of computers. The library has come a long way in computerization. However, computerization is not necessarily automation. The use of computers in the library still remains largely at the stage of word processing — namely, producing the three-by-five cards in more and newer formats.

The biggest progress in the last thirty years in the use of the computer has been in two areas: 1) faster speed in retrieving records; and, 2) sharing data with people in a larger geographic area. Library automation has not yet achieved in the way the manufacturing industry has automated their productions. Almost any output from a library computer requires a human action to trigger it.

While the library system is not highly automated, neither is it very easy to use. Except for the online catalog, all library subsystems require intensive training to use. Operators trained on one subsystem may not know how to use another. For people who have not received the proper training, assistance by trained staff members is always required.

The frequent human interventions are a legacy of the library's manual operations. The manual modes of various operations are
faithfully reflected in today's design of library computer systems. A system is typically subdivided into circulation, online catalog, acquisitions, cataloging, interlibrary loan, serials control, and so forth—just like the old-style organizational chart. Each subsystem has its human manager and a staff to maintain the data. The communication and exchange of data among subsystems and among staff members are often by batches rather than interactive. When a library patron fails to find a title in the online catalog, he may walk to the interlibrary loan office to search for a copy in another library on the ILL terminal. If not found in another library, he will then walk to the acquisitions department to place an order on a third terminal. On each search, he is likely to need a librarian's assistance. The reason why, the patron cannot utilize one terminal to perform any or all of the three tasks lies in the complex and divided library organization which depends more on human intuitive judgement than on clearly defined rules and procedures.

In spite of the complex library organization, the authors contend that a good library system should be simple to use, and that the system's data and functions should be as open as possible to as many people as possible. This simplicity and openness should be a goal of any library system.

Our contention is based on two assumptions:
1. Most library operations are rational and can be reduced to rules, and
2. Most of the data required by the rules and the rules themselves can be stored in and processed by the computer.

The functions of a traditional library seem to be too complex to be condensed into rules, especially for those of us who have accepted the functions of acquisitions, cataloging, circulation, etc., as natural, fixed and unchangeable. But these functions are not natural, fixed or are they unchangeable. The whole organization may be looked at from a different perspective.
We propose to look at the library universe as a network of
e of roads that link together many players: patrons, staff members,
library materials, book dealers, book binderies and other physical
locations. In this universe books are being moved from location to
location. The integrated library system has two main functions:

1. Receiving data from the user or in response to internal
changes of data to activate the movements of books
and other media, and

2. Updating data as a consequence of each movement of a
medium. Should the change of certain data trigger an-
other medium movement, the system will then perform
Function 1.

In other words, the system should know where each book is
at any given time; know what is required for a book to move from
one location to another, and what changes will happen in the data-
base when a book changes location. To be highly automated, the
system has to be able to respond to changes according to rules as
much as possible and in response to human commands as little as
possible, thereby minimizing the number of human interventions
required.

The challenge to this concept of automation is then two-fold:

1. Can the total number of possibilities of movements and
their consequences on the changes of data, and vice versa,
be anticipated?

2. Can the machine process a large set of complex rules?

The answer to the first question is relatively simple. Yes,
many actions in the library universe can be anticipated. The
degree of automation will be directly limited by the degree of the
librarian's ability to anticipate the actions.

The answer to the second question is somewhat fluid because
of the rapid change of technology and, more importantly, the
price of technology. As of today, computers already can process
rules with a high degree of efficiency. Technologically, computers
are large enough to meet these needs and give the second question a positive answer. The question is only whether the library is willing to pay a high price for the hardware and software. As the library community begins to afford newer and more powerful technology in parallel processing, expert systems, artificial intelligence, and larger computers, there is no doubt that a rule-based library system can be built.

If we proceed with the concept of the library universe mentioned earlier, we may agree that answers to both questions are easily but expensively affirmative. To simplify the management and use of data in the new library system, library information can be organized into five sets of data:

1. The USERS, including library patrons, staff members, book dealers, and other libraries, are the locations on the network. Attributes of these users include their names, addresses, status, privileges, etc.

2. The MEDIA which are the books and other materials that contain information. There are two kinds of media:
   a) The PRIMARY MEDIA which are the books that the system is to manage: the electronic text of books, music scores, etc. Attributes include identity number, i.e. the call number, for each medium and its contents, and
   b) The SECONDARY MEDIA, the media that the system generates to contain data for the management of the primary media. Secondary media include the descriptive catalog, holdings files, overdue notices, orders to dealers, ledger, pay-checks and so forth. The attributes includes the creator of the record, the receiver, the message, etc.

3. The CHANNELS are the roadways on which data and media travel. Attributes include types of communication mode, restrictions in access, speed of transmission, etc.
4. TIME is the fourth dimension of the universe.
5. The RULES are a set of conditions that regulate the changes and the utilizations of the other four kinds of data.

Most of today's integrated library systems are very sophisticated in the management of bibliographic data. Some are also good in handling other secondary data such as those for acquisitions and circulation. For primary data, the information industry is now emerging with the CD-ROM and other computer technologies that are very promising in providing primary data in video text, and graphic and audio data. Time, the fourth dimension, is factored into some, but not many, library functions. Few of the current integrated systems take the channels of communication and the delivery mechanism into calculation, nor is there any system on the market that has a recognizable rule base in operation.

The lack of fuller exploitation of data for channels, time and rules is the barrier to achieving a higher degree of automation. We propose a different concept, which we shall call RUBILA®, to overcome this barrier. It is a concept that will utilize these three factors more fully in library system design. With this proposed concept the user does not begin the interaction with the system by signing on with a specific subsystem. Instead, the user will typically answer five questions:
1. Who is he?
2. Which medium is to be activated?
3. Where should the medium be moved to?
4. By when?
5. Is the move accomplished already or to be accomplished?

If the user cannot answer any of the questions, the machine can provide the user a series of choices to lead to a specific answer. For example, when a user does not know what specific books he/she wants, a null answer would bring up the bibliographic sub-
system for the user to search by different access methods. Once all of the questions are answered, the machine will check through a set of rules to see what actions are needed. Typically, the system will retrieve from the database all the relevant attributes of the user, the media, and the destination. All of these attributes will have been previously stored in the computer like the bibliographic data that is stored in the database of the bibliographic utilities. Then the system software will go to the rules once again to match the retrieved attributes against the rules to see what secondary data should be updated, and to calculate whether the task as requested, including when and how, can be accomplished. Any possible options will be presented to the user for selection. In instances where the machine does not have enough data or rules to determine proper actions, the machine will ask for human intervention. There is no need for a library patron or a staff member to spend a lot of time understanding how the tasks are performed. Answers to the five questions are all that are needed from the user in order for the system to take action. Even the lack of an answer to a question would prompt the machine to aid the user to find the answer.

The RUBILA® concept can be implemented by overlaying existing subsystems of an integrated library system with a set of rules to govern different operations. Once a request is received the rules will direct the actions to proper subsystems and update the proper data as illustrated by the following diagrams.
The library universe:

channels of communication and transportation
The RUBILA® machine:
A RUBILA® system has two objectives. The first is to minimize human intervention in the operation of a library system so that when human decision is required, it is only an exception that was not anticipated by the system designer. The second objective is to stream-line library operations. RUBILA is a faster and more consistent way to maintain library data in ways that were not possible before the large computer.

If we take a moment to reflect upon the history of library organization, we can see that it has been evolved from the basis of card catalog thinking. As the card file grew bigger, it was divided into acquisitions, circulation, cataloging, reference and other card files. People were hired to maintain these card files. Library organizational structure grew with them into departments and divisions, each having a card catalog as the center of operation. The traditional library organization is focused on card files. People walk from catalog to catalog for data so that operations can be related.

Today, much of a library's work is still in the updating of the different data either in paper or machine files. Although the computer can link all files for people to consult from one terminal, traditional division of labor still prevails in the library in ways that inhibit the cross reference and calculation of data. The RUBILA® concept suggests that not only the library records can be integrated but that the organization can be integrated in a new way, a way to reflect the efficiency of the computer. We have not had enough time to go into further discussion on the computer's impact on library organization, and only wish to leave this thought with you for a future dialogue.