IDRC
CRDI
CIID

CANADA

NEW HORIZONS IN AGRICULTURAL INFORMATION MANAGEMENT

PROCEEDINGS

OF AN INTERNATIONAL SYMPOSIUM

MARCH 13-16, 1991

BEIJING, CHINA

The International Development Research Centre is a public corporation created by the Parliament of Canada in 1970 to support research designed to adapt science and technology to the needs of developing countries. The Centre's activity is concentrated in six sectors: agriculture, food and nutrition sciences; health sciences; information sciences; social sciences; earth and engineering sciences; and communications. IDRC is financed solely by the Parliament of Canada; its policies, however, are set by an international Board of Governors. The Centre's headquarters are in Ottawa, Canada. Regional offices are located in Africa, Asia, Latin America, and the Middle East.

Le Centre de recherches pour le développement international, société publique créée en 1970 par une loi du Parlement canadien, a pour mission d'appuyer des recherches visant à adapter la science et la technologie aux besoins des pays en développement; il concentre son activité dans six secteurs : agriculture, alimentation et nutrition; information; santé; sciences sociales; sciences de la terre et du génie et communications. Le CRDI est financé entièrement par le Parlement canadien, mais c'est un Conseil des gouverneurs international qui en détermine l'orientation et les politiques. Établi à Ottawa (Canada), il a des bureaux régionaux en Afrique, en Asie, en Amérique latine et au Moyen-Orient.

El Centro Internacional de Investigaciones para el Desarrollo es una corporación pública creada en 1970 por el Parlamento de Canadá con el objeto de apoyar la investigación destinada a adaptar la ciencia y la tecnología a las necesidades de los países en desarrollo. Su actividad se concentra en seis sectores: ciencias agrícolas, alimentos y nutrición; ciencias de la salud; ciencias de la información; ciencias sociales; ciencias de la tierra e ingeniería; y comunicaciones. El Centro es financiado exclusivamente por el Parlamento de Canadá; sin embargo, sus políticas son trazadas por un Consejo de Gobernadores de carácter internacional. La sede del Centro está en Ottawa, Canadá, y sus oficinas regionales en América Latina, Africa, Asia y el Medio Oriente.

This series includes meeting documents, internal reports, and preliminary technical documents that may later form the basis of a formal publication. A Manuscript Report is given a small distribution to a highly specialized audience.

La présente série est réservée aux documents issus de colloques, aux rapports internes et aux documents techniques susceptibles d'être publiés plus tard dans une série de publications plus soignées. D'un tirage restreint, le rapport manuscrit est destiné à un public très spécialisé.

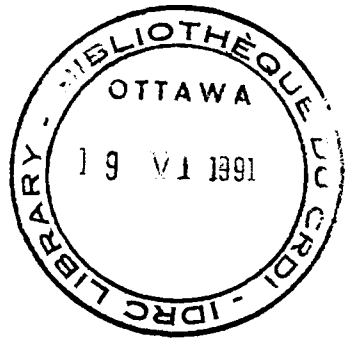
Esta serie incluye ponencias de reuniones, informes internos y documentos técnicos que pueden posteriormente conformar la base de una publicación formal. El informe recibe distribución limitada entre una audiencia altamente especializada.

IDRC-MR293e
May 1991

New Horizons in Agricultural Information Management

Proceedings of an International Symposium,

March 13-16, 1991, Beijing, China



Compiled and Edited by

Gary K. McCone

IDRC
CANADA

SDIC

Sponsored by

International Development Research Centre

Organized by

Scientech Documentation and Information Centre

Chinese Academy of Agricultural Sciences

Organizing Committee

Main Organizing Committee

Chairman	WANG Xianfu
Vice Chairman	JIAO Bin
	HE Chunpei
Secretary	MIAO Zhuoran
Members	HAN Ling
	JIA Shangang
	HUANG Xuegao
	GUO Dianrui
	ZHAO Huaying
	PAN Shuchun
	LI Kaiyang

Secretariat

Chairman	MIAO Zhuoran
Members	CHEN Junying
	HU Jia
	YU Fenghui
	TAI Weidong
	FANG Baoqin

Accomodation & Transport

Chairman	HUANG Xuegao
Members	ZHAO Huaying
	QIN Juanjuan

Conference Site

Chairman	HE Chunpei
Members	PAN Shuchun
	ZHANG Rongchang

Papers

Chairman	JIA Shangang
Members	LI Kaiyang
	LIANG Suzhen
	WANG Zhenjiang
	GUO Jian

Table of Contents

Keynote Address

Session I: Management and Development of National Agro-Information Systems

Session II: Information Management and New Technology Application

Session III: Management and Development of Regional Agro-Information Systems

Session IV: Scientech Information and Productivity

Session V: Development and Utilization of Agro-Information Resources

Appendix 1: Supporting Papers

Appendix 2: Symposium Participants

Appendix 3: Author Index

The Close Associations between Indexing and Microcomputer Software Maintenance

BI Jinping

Institute of Scientech Information
Liaoning Academy of Agricultural Sciences
Shenyang, China

Abstract

The close associations between document pre-processing and microcomputer software maintenance are studied based on the experience of setting up the Comprehensive Database of Chinese Agricultural Bibliographies using microcomputers. The suggestion to combine indexing with inputting to exploit the potential of applied software and the feedback of the suggestion of doing the indexing and data input simultaneously is made. It is necessary to develop application software which has a definite objective in view and to study the effects of indexing quality on information retrieval. The author of this paper also suggests that training a contingent of qualified T-type personnel is very important for assuring the set up of a database.

In recent years, the vast amount of scientific and technical documents increasing dramatically all over the world along with the full speed development of modern science and technology has resulted in more and more barriers for users to search books and reference materials.

Information retrieval is the guide book by which readers gain access to a huge treasure house of information, and searching by computer is rapid and effective. One necessary condition for using computers to do information retrieval, however, is to first set up a database of sufficient size. For this purpose, a lot of work in both the pre-processing (including the selection of subject material, registration and indexing) and the microcomputer operation (such as selecting suitable software, building up a set of file specifications required for building a database applying this software, inputting the pre-processed data, and debugging and maintaining the software) would be done. The objectives of this paper are to study the close relationships, inter-permeated and inter-stimulated, between the pre-processing of the data and software maintenance based on the experience of setting up the Comprehensive Database of Chinese Agricultural Bibliographies using a microcomputer.

SIMULTANEOUS INDEXING AND INPUT TO INCREASE EFFICIENCY AND DATABASE QUALITY

At the beginning of 1989, we (N.E. Sub-centre) took part in the project to set up the microcomputer-based database, sponsored by the Information Centre of CAAS. Generally, doing this work would demand several staff who would be in charge of

indexing, software maintenance and inputting data, respectively. In our experience, however, we found that data input and indexing can be done simultaneously. By doing so, the efficiency would be increased and the quality of the database would be improved and a lot of paper would be saved.

At first, we used to fill in the items on the worksheet one by one. Every record must be filled in on a worksheet (16 cm sides). Later, because the indexing and inputting processes had to be completed by the same person, these indexing items were only filled in on the worksheet and the descriptive elements could be omitted. The temporary numbers were inputted into the computer but the assigned one-up serial numbers of the titles were indexed, so the final serial number would exactly match the total number of records for a serial by the end of indexing, which made it very convenient to count up the total titles of each serial issue. The following fields were directly inputted for each title:

Type of Document, Language, Author, Other Authors, Title of Serial or Book, Sponsor or Author, Publication Year-Volume-Issue, Annotation, and so on. Therefore the pre-processing efficiency could be raised by about 40%.

A new technique for automatically forming indexing entries by using nine functional symbols, was developed in Sept. 1989 by Wang Huaihui working in the Information Centre of CAAS. This advanced method has improved the building of the database. The efficiencies of indexing and inputting were raised by 35% more or less, and indexing was more simple when using these functional symbols. Merging the two steps of indexing and inputting into one not only has high efficiency but has also improved the quality of work. This also allows indexing mistakes to be conveniently corrected at the time of input. For example, we may think we have a paper concerned with sheep diseases when the Class Number is input as S858.26, but if we then find out it has been assigned a Subject Term for swine disease, at that point we could change the Class Number "S858.26" to "S858.28" immediately. Under normal circumstances, the workers are vigorous and sober-minded when the contents of the worksheets are being input one by one into the computer, and then having the satisfaction of seeing one's final results is possible. On the other hand, if the process of indexing and inputting is preformed by two different persons, ideal effects could not be obtained. In addition, the special input staff often do not understand the indexing process and the indexing staff usually takes much time as they proofread the final proofs, so some mistakes and waste of manpower might be occurring.

DEVELOPMENT OF APPLIED SOFTWARE BASED ON THE NEEDS FOR CREATING A DATABASE

The application software commonly used for setting up databases is Micro CDS/ISIS, the operating system is CCS3.3L-LX Chinese Card. Since this operating system is dependent on phonetic association, the inputting speed might be influenced without the good Five-Stroke Structure System. Although two steps of indexing and input merged into one step could have raised the efficiency and quality of setting up the

database, the input process would fall into the same plight again without a good operating system. Therefore we have tried to run Micro CDS/ISIS under the Five-Stroke Structure System and we have been successful. In this case, the functions of browsing, printing and searching could be run successfully except for the following two shortcomings: 1. The first menu of the database might disappear too soon; 2. Some names of fields on the work sheet could not be displayed. But both these deficiencies are no problem to a skilled person. In order to help the novice, first, we printed the first level menu of the Micro CDS/ISIS system under the CCS3.3L operating system as the supplementary screen menu; second, the names of fields that did not display on screen could be found from the work sheet. As stated above, after practicing again and again, the new hand could work skillfully. According to our experience, after matching the Five-Stroke Structure System with Micro CDS/ISIS software, input speed has been increased by five times. Over 300 records could be inputted within two work-days each month. The speed would be more rapid if operated by a special staff.

CONDUCTING AN INFORMATION SEARCHING TEST TO MODIFY INDEXING MISTAKES

The information searching test has not only served the readers but is also a means through which the data of an information database could be modified, coordinated and united.

At the core of indexing work is subject analysis, but usually, the results of subject analysis vary with the person and work time. This kind of case frequently crops up, especially because we lack a normal thesaurus at the present time. So the terms selected from the Free Word field could be divided into Normal and Abnormal Free Words. The so-called Normal Free Words are those words indexed according to the Chinese Thesaurus, the Agriculture Thesaurus and other special handbooks; and the so-called Abnormal Free Words are truly just free text words. In these cases, the indexing errors made by a single operational person would be uncountable in spite of the errors made by several operational persons. So it seems that the key is how to minimize individual errors over time to maximize the consistency of indexing. The method we took to solve this was to conduct the information searching tests frequently and definitively to decrease indexing mistakes and improve the quality of indexing.

Example 1. Normally, "Chickens Marek's Disease" was regarded as a pre-coordinated term. In order to check whether this word was treated in this form in all records after indexing over 800 records, a Boolean search was conducted and the searching formula is "Chickens * Marek's Disease." The results showed that in record no. 298 it was treated as a post-coordinated word, and then it was modified immediately.

Example 2. In addition, the rapid inquiry would be conducted to determine if some words were regarded as Use term (U) or replaced by Used for combination (UFC). For instance, Vaccine inoculation is generally treated as a UFC and Immunity Inoculation as the Use term. Four wrong records were discovered this way and then modified at once.

Example 3. Artificial Insemination of Domestic Animal usually belongs under the category of Reproduction in the light of China's National Classification Scheme for Books (CNCSB) but according to the AGRIS Classification Scheme, this descriptor belongs under the category of Animal Genetics and Breeding with a class number L10, and the latter classification only possesses Animal Physiology Reproduction with class number L53 other than the broader term of Reproduction. On the basis of CCB principles, all Artificial Insemination of Domestic Animal records were distributed into the class mark of L53. To check them, the search strategy (Artificial Insemination + Freezing Semen) * L53, was formed, and then six incorrect records were discovered, and subsequently changed their Subject Category Code of L53 into L10.

Example 4. Crop Tissue Culture was distributed in the category of Genetics Breeding in CNCSB, but it was located in the category of Plant Reproduction based on the AGRIS Classification Scheme with class number F02. In order to identify those records which were mixed in Plant Genetic and Breeding (class number F30) concerning tissue culture, the search strategy (Tissue Culture + Embryo Culture + Shoot Tip Culture + in Vitro Culture) * F30, was executed. Four incorrect records were then discovered and as a result, the F30 (Subject Category Code) was changed into F02.

In the light of our indexing experience, it seems that those Abnormal Free Words would be united whenever possible to avoid the repetition of synonymous words. This task still would be dependent on executing searches to detect incorrect records. All of the following words were classified as bound forms in this way, e.g., Grain Water Content, Grain Yield, etc. If they need to be modified in the future, these words would be changed into bound forms in the same way also.

It is very difficult to discover inconsistencies in indexing in records based on their serial number. In this case, inconsistencies in indexing within the same concept or category, could be apparent when the united searching and browsing had been conducted on special terms or categories, e.g., descriptor sequence, indexing depth, selection of heading word, and the Subject Index Entry. After indexing of a set of records had been completed, each record in a category would be searched and browsed. All problems (mistakes) displayed on the screen were written down and modified uniformly one by one. For convenience and consistency of indexing, the Free Word List was printed according to the inquirer, who checked the occurrence frequency of the words, compared them, and then decided which terms to choose.

THE RELATIONSHIP BETWEEN TRAINING QUALIFIED T-TYPE PERSONNEL AND BUILDING THE COMPREHENSIVE BIBLIOGRAPHIC DATABASE

The comprehensive database covers over twenty new subjects of agricultural sciences as well as its related fields, and nearly twenty types of documents. In fact, the database not only serves for automated searching of information, but also the publishing of reference books for manual searching. There are 32 fields for each record, eleven of

them being searchable fields. The reasonable structure of this database is designed with a great number of access points which are well-organized.

It requires a large amount of manpower to build the database. If we have enough people who are qualified, we can do the work of selection of subjects, description, indexing, input and software maintenance in assembly line fashion. That is, a specific person is assigned to each step of the work mentioned above, and it is coordinated by one person. Our sub-centre, however, is short of staff (only 2-3 workers) and we process our data independently under the unified command by the National Centre. So we can not set up an assembly line because we lack staff and we only have two persons, specifically, one person is in charge of indexing (including selecting the subject and registration) and another is in charge of inputting and software maintenance. One consequence is lower work efficiency, and some mistakes in linking occur which could influence the whole process of setting up the database. According to the actual schedule of every sub-centre, work procedures vary widely. Some sub-centre might be short of staff for software maintenance, so that data have not been inputted into the database in a timely manner, while another sub-centre might be short of indexers and the software experts are helpless. In view of the above mentioned facts, we would like to make the proposal that an excellent T-type qualified personnel contingent must be built up as soon as possible, so that 2-4 persons in each sub-centre could independently successfully operate the database by themselves. If this were the case, the process of setting up the database could be sure to run without a hitch, even though some persons may happen to leave. The other advantage of working independently is that people can keep informed on how the setting up of the database is progressing, and can grasp and solve the problems taking place in each link to improve the quality of the database.

In short, the various links are organically combined as well as dependent on each other during the whole course of setting up the database, and then it is very important to keep this close relationship, inter-permeated and inter-stimulated, between indexing and maintenance of microcomputer software. This paper sums up the preliminary experience of two years of practice, and undoubtedly many techniques will be discussed further on down the road.