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.

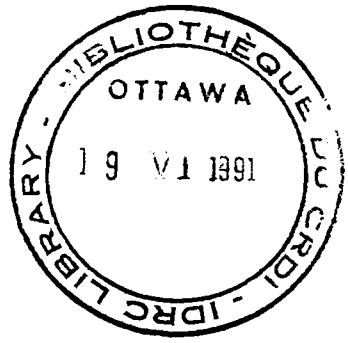
Esta serie incluye ponencias de reuniones, informes internos y documentos técnicos que pueden posteriormente conformar la base de una publicación formal. El informe recibe distribución limitada entre una audiencia altamente especializada.

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

Realization and Application of Large Capacity Chinese Character Disk Operating System (LCCDOS)

NIU Zhan Liang BAI Juping LIU Huifang

Scientech Documentation and Information Centre
Chinese Academy of Agricultural Sciences
Beijing, China

Abstract

The development of LCCDOS and the Chinese word processing software, Chinese WordStar (CWS) that is supported by LCCDOS are discussed based on an analysis of how to enable the microcomputer disk operating system (DOS) to process data in the Chinese language.

Preface

At present, the Chinese Character DOS (CCDOS) for use in microcomputers only supports more than seven thousand Chinese characters and other characters. Because the internal code of CCDOS is composed of two bytes, Chinese characters are expressed by the highest bit set, the maximum number of Chinese characters and other characters that can be expressed is more than 17,000 besides the Chinese character mark bit and the control characters. For Chinese information processing, particularly in comprehensively managing agricultural information resources, this number of Chinese characters is not sufficient. So, Large-capacity Chinese Character Disk Operating System (LCCDOS) has been developed by the Scientech Documentation and Information Centre, Chinese Academy of Agricultural Sciences. It can support 30,000 Chinese characters as well as other characters. At the same time, the function of the word processing software Chinese WordStar (CWS) has been expanded in order to have it supported by LCCDOS.

I. DOS Sinicizing Analysis

CCDOS for use in microcomputers is developed based on the Western language MS-DOS. The aim of sinicizing is to enable DOS to process both the Western languages and Chinese language at the same time. In order to achieve this aim, the main problems that must be solved are Chinese character expression and storage in a computer, their input and output, etc. In brief, we have done the following based on DOS:

1. Expression of Chinese characters in computers. Chinese character designation adopts a two-byte Chinese character internal code. A Chinese character is expressed by the highest bit of a two-byte set. This kind of Chinese character internal code can express about 9,000 Chinese characters and other characters.

2. Image storage of Chinese characters. Since the Chinese character image is needed when displaying or printing Chinese characters, the image information of each Chinese character must be provided. CCDOS adopts the dot matrix method to store the Chinese character image information, and all dot matrixes of Chinese characters are put in one file forming the Chinese character library for use in displaying and printing.

3. Input of Chinese characters. The small keyboard is designed for Western language input and there are no Chinese characters on it. In order to input Chinese characters on a small keyboard, we have designed several input methods in CCDOS, such as zone bit and phonetic transcription, etc. When a Chinese character has been inputted in one method, we can change the input code into an internal code through a corresponding table or algorithm.

4. Display of Chinese characters. For displaying the Chinese characters, the display processing program has been modified and expanded in CCDOS based on the Western DOS. On one hand, the character display format has been changed into an image format. On the other hand, the identifying and processing of the Chinese character internal code have been added. We have adopted the method of reading out the dot matrix from the Chinese character library and then putting it into the screen refresh area to display the Chinese characters.

5. Printing of Chinese characters. For printing the Chinese characters, some modification and expansion of the original print program have been done in CCDOS. The main task was to add the identifying Chinese character internal code and the link to the appropriate Chinese character library.

II. Selection of Schemes for Expanding Chinese Character Processing Capacity

The following three main problems exist in expanding the Chinese character processing capacity:

1. Internal codes.

From the above analysis we know that a key problem of expanding our capacity to process Chinese characters is the selection of the Chinese character internal code. From the viewpoint of expansion, the following are three feasible schemes in the selection of the internal code:

1) Two-byte internal code. For the first byte, the highest bit is set. For the second byte, there are no restrictions on the highest bit.

2) Three-byte internal code.

3) Internal code of combining two bytes with three bytes. The basic Chinese character set uses a two-byte internal code. The expanded Chinese character set adopts a three-byte internal code.

The first method has the following advantages: the shortest internal code, storage consistent with display and the Chinese language can be compatible with the Western language in processing. But the number of Chinese characters can only be expanded to about 17 thousand because of the limit to the length of the internal code. This can't meet our requirements, so we can only select one of the last two schemes. The third method is more complex than the second in processing, but the third one maintains a compatibility between LCCDOS and CCDOS. So we have selected the third one.

2. Organization and storage of the Chinese character library (CCLIB)

If all Chinese characters could be coded onto a computer card or stored in external or expanded memory which is beyond the 640k limit of internal memory, this problem would not exist. But at present, we can not attain either of these, so we must seek other alternatives to solve this problem.

For the organization of the CCLIB, there are two schemes, that is, the single CCLIB and multiple CCLIBs. We selected the latter. Besides the basic CCLIB, we have added three expanded CCLIBs whose size is the same as the basic CCLIB.

The basic CCLIB is stored in the internal storage area and the other three expanded CCLIBs are stored in the external storage area, that is all expanded Chinese characters are stored on a hard disk.

3. Input of Chinese characters.

Although all types of input in the original CCDOS can be used in LCCDOS directly, they are only suitable to inputting basic Chinese characters. In order to solve the problem of inputting expanded Chinese characters, we use the following two ways:

1) Modifying the existing zone bit input to adapt inputting the expanded Chinese characters. The key point of this is to expand the expression scope of the zone bit code. We can build a corresponding relationship between each expanded zone bit code and each expanded Chinese character.

2) Designing one or more new kinds of input. All these new input methods can use existing input principles. The main task to be done is to design the input codes for the expanded Chinese characters based on the coding regulations, and build the correspondence table between input code and internal code.

We have adopted the first way mentioned above in LCCDOS, so only the zone bit input method can be used to input the expanded Chinese characters in LCCDOS at present.

III. Realization of LCCDOS

LCCDOS is developed by modifying the original CCDOS. There are the following principal aspects that have been modified:

1. **Internal codes.** CCDOS adopts two-byte internal codes, but LCCDOS uses an internal code combining two bytes with three bytes. LCCDOS adopts a two-byte internal code for basic Chinese characters following that of CCDOS. So this can keep LCCDOS compatible with CCDOS. The expanded Chinese characters use a three-byte internal code, of which the first byte is the mark byte, the last two bytes have the same value range as the basic Chinese character internal codes. The mark byte is used to mark the Chinese character library to which the Chinese characters belong. Its value range is hexadecimal FC to FE, which represent the three expanded CCLIBs, that is, if the value of the mark byte of a Chinese character internal code is FC, then this Chinese character belongs to the first CCLIB, and so on.

2. **Keyboard entry.** Since the first CCLIB of LCCDOS is the same as the CCLIB of CCDOS, and the internal code of this part of the Chinese characters doesn't change, LCCDOS can use any kind of CCDOS input to input the basic Chinese characters. But with the expanded Chinese characters, the original input ways don't work. In order to input the expanded Chinese characters, the original zone bit input method has been modified in LCCDOS. A zone bit code is expressed by four hexadecimal numbers instead of by four decimal numbers in the former. The maximum value of the zone bit code is changed into 255 instead of 94. The total number of Chinese characters and other characters that the zone bit code can express is 65,025. The range of corresponding zone bit codes of each CCLIB is:

The first CCLIB: 0101-5B5E
The second CCLIB: 015F-5BBC
The third CCLIB: 5F01-B95E
The fourth CCLIB: 5F5F-B9BC

After a Chinese character has been input by zone bit code, the basic Chinese characters and the expanded Chinese characters are processed separately. The basic Chinese characters are directly transformed into internal codes, the internal code count is 2. The mark bytes of expanded Chinese characters are added in the course of input processing, the internal code count is 3.

3. **Screen display.** In LCCDOS, we have added some judgments and processing for the three-byte internal codes. For the two-byte internal codes, we also use the same processing as CCDOS, that is to calculate the segment address in the internal memory of its dot matrix according to the internal code and then read the dot matrix of the Chinese character and display it. For the three-byte internal codes, we must calculate the storage sector of the corresponding dot matrix of the Chinese character in the hard disk according to the internal code, then read the data of the sector into internal memory using interrupt X'13', and finally fetch the dot matrix from memory. The other processing is the same as that of CCDOS.

4. **Printout.** The judgments and processing for the three-byte internal code have also been added. It is modified to be similar to that of screen displays.

Some application software supported by CCDOS would have a problem if they were operated under LCCDOS. For example, there will be a space before every expanded Chinese character when it is displayed on screen. This problem can be solved by modifying the application software.

IV. Establishment of CCLIB in LCCDOS

In order to build the CCLIBs for LCCDOS, we have developed a character-constructing software package. Besides the functions which common character-constructing software possesses, this software has its own characteristics:

1. The expression and range of zone bit code are different from common character-constructing software. The zone number and bit number are both expressed by hexadecimal numbers, whose maximum value is 255, not 94.

2. The fetch CCLIB and the store CCLIB are separated. The common character-constructing software only relates to one CCLIB, but this software relates to two CCLIBs; one is the fetch CCLIB specifically for fetching Chinese characters, the other is the store CCLIB specifically for storing the Chinese character that has been made. These two CCLIBs can be the same one.

The structure of the four CCLIBs is same, so the expanded CCLIBs can use the structure of the basic CCLIB. When we build the expanded CCLIBs, first we can copy a basic CCLIB, then each Chinese character in the duplicated CCLIB will be modified, so we can get a new CCLIB.

Because we modify the basic CCLIB to build the expanded CCLIBs, it is not necessary to build the Chinese characters in the order of the zone bit code and complete it at one time. We can complete it step by step.

V. Application of LCCDOS

In the realistic application of LCCDOS, it is necessary to modify the software which is supported by CCDOS. We have modified the software CWS which now has more comprehensive functions and wide applications.

From the above analysis we know that if the software supported by CCDOS is operated in LCCDOS, there will be a space before the expanded Chinese characters are displayed on screen. This space is caused by the mark byte of the expanded Chinese character internal code. This problem will affect our editing work. So we must solve it.

First of all, let's look at the CWS program. In these programs there are two buffers, one is the screen buffer, the other is the line buffer. The length of a line in both buffers is 80 bytes. Because the three-byte internal codes are introduced, the number of Chinese characters that can be displayed in a line will be reduced. From the point of eliminating the spaces before the expanded Chinese characters on screen and keeping the number

of Chinese characters displayed in a line unchanged, we think there are the following two modifying schemes:

1) Expanding the line buffer and the screen buffer to enable a line to put forty three-byte Chinese character internal codes, that is, the length of the line of the buffer is enlarged to 120 bytes from 80 bytes. When the contents of the line buffer are displayed, the place of the cursor must be controlled to ensure that each expanded Chinese character only takes up two rows on a screen.

2) Keeping the size of the line buffer and the screen buffer unchanged and trying to have the mark byte of the three-byte internal code stored in another place. That is, we establish another two buffers to hold the mark bytes of the expanded Chinese characters, and don't let them enter the line buffer and the screen buffer.

After analysis and comparison, we adopted the second scheme. In the course of processing, the mark byte is put into the mark buffer instead of the line buffer and the screen buffer. Only the last two bytes are put into the line buffer and the screen buffer. When the contents of the line buffer are displayed, the Chinese characters must be defined by a combination of the contents of the line buffer and the contents of the mark buffer. But the placement of the cursor is only based on the contents of the line buffer. For other software, the method of modification is the same as this one on the whole.

Summary

The successful development of LCCDOS not only enables the use of computers to manage agricultural information resources, but also provides possible conditions for the large information centers in using the computer systems. This LCCDOS can support thirty thousand Chinese characters and other characters. It can be expanded to support sixty thousand Chinese characters and other characters by using the same principle.