

苹果机《友好》汉化 dBASE II 的设计和实现要点

王博文 侯济恭 姚慧娟

〔计算机科学(电脑)系〕

摘 要

本系统的目的在于为广大非计算机专业人员用户提供一性能价格比高、亲善性好的汉化大众化数据库管理工具。在实现上技巧地应用128k卡,可将Apple II plus机的内存“虚拟”地扩展到192k以上,实现了CP/M的造字功能。

一、系统的特点

本系统具有五个方面的特点:(1)中文功能体现在中文功能菜单、中文提示、中文错误信息及可存贮中、英文为内容的档案。(2)改造原来用户具有的汉卡,保持简体汉字的仓颉输入法,增加初学者最为熟悉的拼音输入。在实现上,留有增加新的输入法的余地。(3)具有造字功能。系统为用户开辟一个容量为100个字型的小字库。用户可造出特殊需要的字存放于小字库以供随后使用。(4)极为经济且响应速度快。所需基本硬件是Apple II plus硬件,带有二个软盘驱动器及一块64k RAM卡或128k RAM卡。由于不采用软字库,不需额外的软盘驱动器。硬字库技术使系统响应速度快。(5)保留英文版dBASE II 2.23B全部功能。当初学者通过菜单方式熟悉dBASE II以后,他可进一步通过键盘命令、命令文件使用dBASE II的全部功能。

二、多级菜单

1. 前导菜单

开机后出现在屏幕上的前导菜单是:

- | |
|-------------|
| 1. 使用中文功能菜单 |
| 2. CP/M |
| 3. 造字 |
| 请按需要输入1、2或3 |

本文1987年5月5日收到。

初学者可选择项 1, 通知系统提供下一级菜单。有经验的使用人员可不依赖于菜单而选择项 2, 进入 CP/M 操作系统后启动 dBASE II。本系统保留 CP/M 及 dBASE II 全部功能。当用户存贮案资料时, 其记录的内容可以是汉字的。

2. 第 1 级菜单

前导菜单在屏幕上出现后, 若按数字键“1”, 即可使第 1 级菜单出现, 其形式为:

- | |
|--------------|
| 1. 档案文件处理 |
| 2. 档案记录处理 |
| 3. 报表处理 |
| 4. 退回(到前导菜单) |

档案文件处理指的是档案的目录显示, 档案的建立、删除、重新命名、生成排序档案等。档案记录处理指的是对档案中的记录的查询、追加、插入、删除、求和等。

报表是数据处理的一个重要环节, 故菜单方式操作时也为用户提供报表处理的菜单。

3. 多个第 2 级菜单

(1) 档案级操作的菜单形式为:

- | |
|---------------|
| 1. 档案目录显示 |
| 2. 建立档案 |
| 3. 删除档案 |
| 4. 档案重命名 |
| 5. 生成排序档案 |
| 6. 退回(到菜单(1)) |

采用菜单方式使用本系统时, 用户不必关心其档案存放的物理盘号, 此时系统将用户的盘号初始订为 B 盘, 即其缺省值为 B; 以档案的建立为例, 若用户按数字键“2”后, 屏幕上会出现中文提示“请输入档案名”, 用户此时可打入不带盘号的档案名。系统接受档案名后, 会提示用户定义档案记录结构, 系统还会进一步让用户决定是否马上输入记录。

本级菜单上其它几个功能的工作过程和上例类同。用户在完成菜单上某功能的操作后, 可决定是否继续留在本级菜单工作或返回到上一级菜单。

(2) 第 2 级菜单关于档案中记录的处理的菜单形式是:

- | | |
|-----------|---------------|
| 1. 显示查询记录 | 2. 追加记录 |
| 3. 插入记录 | 4. 删除记录 |
| 5. 记录求和 | 6. 退出(到菜单(1)) |

用户利用菜单操作时, 系统会要求用户输入档案名。若是显示或删除记录, 系统还会要求用户指定操作的范围和条件。

(3) 报表处理: 由第 1 级菜单的选择项 3, 可进入报表处理。处理时, 系统会询问用户是否已定义好报表格式, 在没有定义好报表格式的情况下, 系统会要求用户对报表格式加以定义, 用户可在屏幕或打印机上得到所需的报表。

三、汉字的拼音输入及造字功能

最能为用户接受的汉字输入法是拼音法，因为不需要事先训练即可使用拼音法。在汉字输入量不大的情况下最容易使用，输入速度不见得快的拼音法是最优的选择之一。

本系统提供的拼音法也是目前各类机型采用的“关联式”拼音法，即用户只要打入想要输入汉字的前几个拼音字母，甚至首个拼音字母，就可从出现在屏幕下端的汉字串中选择汉字。例如输入“铜字，”可打入拼音码TON或只打入拼音码T。在仅打T的情况下，出现在屏幕下端的汉字串为：

0 塌 1 他 2 它 3 她 4 塔 5 獾 6 挞 7 榻 147

用户可按数次“→”键，使铜字出现在屏幕的下端的汉字串里。汉字串中每个汉字串左边的小数码称为该汉字的选择码，汉字串最后一个汉字右边的数字表示还有若干个关联的汉字未显示出来。在打入TON拼音码的情况下，屏幕下端的汉字串为：

0 通 1 桐 2 酮 3 瞳 4 同 5 铜 6 彤 7 童 005

用户这时只要按数字键“5”，即可输入“铜”字。

拼音法输入汉字时，所打入的拼音码会出现在屏幕下端，若发现刚打入的拼音码不对，可按“←”键加以删除，一旦确信所打入的拼音码正确无误，按空格键可使关联的汉字串出现在屏幕下端供用户选择输入。系统也提供一定的手段供用户删除刚输入的汉字，其操作与删除英文字符操作一致。从经济上考虑，本系统仍采用用户原已配备的汉卡（佳佳、松台、益生），但需对该卡加以改造。这样，系统为用户提供了简体汉字的仓颉码输入。改造好的汉卡需要：（a）是简体硬汉字库；（b）保持原有的为DOS 3.3相连的功能又增加为CP/M相连的功能，包括汉字库的字型能显示在屏幕上又能打印到打印机上。

用户采用拼音法或仓颉法输入汉字时，系统可供使用的汉字数达4800，占日常使用汉字约99.99%，系统所未能提供用户特别需要的汉字式字型，用户可通过系统的造字功能自己制造并存库。

由于系统采用仓颉码作为汉字内码，从汉字的拼音码到仓颉码的转换需要通过trie树进行。本系统设计的退化trie树的节点构造为7个字节，分4个部分：

拼音码	不相等的查找指计	相等时的查找指计	关联汉字数
-----	----------	----------	-------

除了第1部分为1个字节外，其余了部分均为2个字节。当拼音码部分为20H时，表示某有效拼音码串的查找结束，其它部分为：

20H	下一查找指计	内码表相对地址	关联汉字数
-----	--------	---------	-------

系统有一个汉字内码表，按照汉字的拼音码顺序存放汉字内码。一旦用户输入的拼音码串在

退化trie 树查找正常结束,即可得到关联汉字串的内码表相对地址及关联汉字数。系统根据得到的内码表相对地址,将关联的汉字串显示到屏幕下端供用户选择输入。需要造字的用户可从系统的前导菜单选择造字功能、造字菜单形式为:

1. 造字	2. 改错字
3. 显示小字库	
4. 返回(到前导菜单)	

系统为用户目前提供汉字的三种输入,即拼音、仓颉、小字库码。CP/M 命令或 dBASE II 命令只能在非汉字输入态下打入。非汉字输入态及三种汉字输入态均可自由切换,即可用 ↑F、↑k、↑L、↑O这四组键对实现。

四、64k RAK 卡的使用

实现本系统最为关键的是内存空间问题。由于CP/M 常驻内存为7k, dBASE II 常驻内存占41k, Apple II I/O占4k, 用于汉字显示的高分辨率区占8k, 6502其它系统程序及数据区占4k, 故该微机64k地址空间已告占用完毕。汉字库即使压缩也要占几十 k 空间, 另外拼音码到内码的转换trie树、内码表也要占几十k。其它的本系统中的各种处理程序都要解决其空间问题。采用软字库、利用软盘必然为汉字的输入带来速度慢、设备寿命短、实用性差的毛病, 利用64k RAM卡或128k RAM卡可以有效地解决这一问题。需要解决的是 Apple II 机硬件结构带来的困难。

Apple II 的64k卡或128k卡占用的内存区为B000 H—DFFF H, 共12k的区域。64k 卡或128k卡逻辑上分为若干个16k的子体, 每个16k子体又有A体B体之分(图1)。

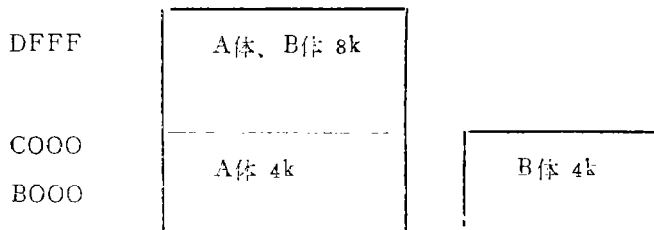


图1 某16k子体的内存对应关系

本系统在具体做法上做到了:(1)避免64k卡区域为CP/M区域的冲突;(2)在64k卡各子体上存取数据;(3)在各子体上存放程序并执行;(4)在各子体之间程序控制的转移;(5)在各子体之间数据块的传送。由于达到上述目标,无疑为Apple II 机的有限内存找到可行的扩充途径。

五、CP/M BIOS 的改造及dBASE II的修改

dBASE II 是在CP/M操作系统下运行的,需要对CP/M的最里层BIOS进行修改使之有汉字

输入输出能力。需要修改的包括：(1)从控制台接收一个字符子程序；(2)输出一个字符到控制台子程序；(3)输出一个字符到打印机子程序。修改后的这三个子程序就作为与dBASE II 相联的基础。对dBASE II 修改，使之能处理中文，需要做两方面繁重的工作。

1. 移动程序：Apple II 机内存高分辨区第二页3000H—4FFFH 要用来作为显示汉字的缓冲区，占据这一区域的dBASE II 程序和数据必须移往内存其它可利用的区域。我们采用半自动的机器码移动方法，具体做法是：(a)移动前对程序区及数据区加以区别；(b)移动程序区并对其中三字节指令作相应的修改；(c)对数据区作简单的移动，其内容不加修改；(d)夹杂的程序区中的个别数据由于移动程序时对其可能的自动修改，用手工的方法进行复原。

由于移动是采用程序的方法自动进行修改，而误修改的数据用手工方法复原，故称此法为半自动机器码移动法。

2. I/O模块及其它模块的修改：由于汉字内码的长度与一般字符的机内表示长度不一样，再加上本系统采用变长汉字内码，因而对dBASE II 中I/O模块应作较大的修改。其修改的结果使新版dBASE II 有存贮中英文档案的能力，也仍然保持了dBASE II 的全屏幕编辑能力，即在屏幕上移动光标对汉字进行插入和删除时，保持光标的正确位置及内存指针的正确位置。

为数不多的dBASE II 模块也作相应的修改。比如为使用户能打印报表和得到操作过程的硬拷贝，SET PRINT 命令和↑P的处理部分作了修改，增强了 SET PRINT 的功能。

参 考 文 献

- [1] 张寿堂，中文信息的计算机处理，国防工业出版社，(1985)。
- [2] 荣树熙、张开敬，6502微处理器及其应用，北京师范大学出版社，(1984)。
- [3] 赖翔飞等，dBASE II 在TRS-80(I)微机上的移动实现，计算机应用与软件，3, 1(1986)。
- [4] Clagbrook, B.G., File Management Techniques, Thompson Books, (1983)。

The Design and Implementation of a "Friendship" Chinese dBASE II on Apple II

Wang Bowen Hou Jigong Yao Huijun

Abstract

This newly implemented system aims to provide the public users a high performance/cost and friendly Chinese DBMS.

In implementation, Apple II plus's internal memory can be expanded virtually to over 192K by technically efficient use of 64K card or 128K card. The system implements the function of making Chinese characters under CP/M for the first time.