

用汉字打印检索信息的微型计算机 医学情报数据库及其检索软件

广西计算中心 符华儿 黄新新

提 要

1981年5月,我们在C. M. Technologies B档微型计算机上用CBASIC语言建立了部分医学文献情报资料数据库,研制了其检索软件,成功地实现了对部分医学情报的检索。当时使用的是汉语拼音字母,使用、阅读不便。1982年10月,我们在BCMⅡ微型计算机上,采用BASIC80语音,成功地用中文明码(即电报代码)建立了部分医学情报数据库,用明码检索,检索到的情报信息由计算机在打印机上用汉字打印输出,使用、阅读方便,很有发展前途。

一、概 况

当前科学技术发展迅速,文献资料的数量急剧增加,及时地了解、学习和研究国内外科学技术发展的最新动态和内容,是实现四个现代化必不可少的手段和条件。要做到这一点,光靠传统的人工检索办法已不能满足需要。因此,用现代化手段装备情报机构已势在必行。电子计算机技术的发展为情报检索的自动化开辟了广阔的前景。1981年5月,我们根据广西医学情报研究所提出的要求,从医学情报存贮和检索角度,在C. M. Tehnologies B档微型计算机上建立了部分医学文献情报资料数据库,研制了其检索软件,成功地实现了对部分医学情报的检索。有关这件工作的文章资料曾于1981年在杭州召开的全国Z-80微型计算机学术会议上交流。当时使用的是汉语拼音字母,使用、阅读不太方便。1982年10月,我们吸取了1981年的经验,克服汉语拼音使用不便的缺点,在带有汉字库的BCMⅡ微型计算机上,采用BASIC80语言(该机只有这种语言能连接打印汉字),成功地用中文明码。即电报代码)建立了部分医学情报数据库,用明码检索,检索到的情报信息由计算机在打印机上用汉字打印输出,使用、阅读方便,很有发展前途。1982年10月在南宁市召开的中南地区医学情报会议上,我们交流了此汉字输出检索的文章资料,还为与会代表表演了此情报检索软件,受到代表们的好评。

二、数据库设计方法简介

BCMⅡ微型计算机使用八寸单密度磁盘,每盘容量为512KB。由于情报资料有关内容

长短不一，难于规定固定长度，加上本微型计算机只有 64K 内存，除去系统占用的外，用户能用的大约只有 32K，容量有限，为节约存贮单元，我们选用了流结构的文件形式（顺序文件）来建立医学情报数据库，文件中的每个记录包括主病名、明细病名、文章名、作者名、期刊名、期号、发表日期、页号及内容摘要等八个项目。为克服顺序文件检索时间长的缺点，我们将全部数据库按主病名的类别划分成多个（目前有近四十个）类别文件，将有限的主病名按医学情报上习惯的先后次序，依序用自然数码作为其类别号，一个数码建立一个类别文件，并用这个数码作该文件名。例如：

<u>主 病 名</u>	<u>类 别 文 件 号</u>
检查	1
消化道出血	2
消化性溃疡	3
胃炎	4
肠炎	5
腹水	6
黄疸	7
肝病	8
⋮	⋮
⋮	⋮
肿瘤诊断	29
肿瘤治疗	30
⋮	⋮
⋮	⋮

所有的类别文件就构成了整个医学情报数据库。类别文件起了一级索引的作用。假如整个数据库划分为 M 个类别文件，每个文件中的记录数都为 N，那么整个数据库有 MN 个记录。设单位访问时间为 T，如果不划分类别文件，要检索该数据库中的最后一条记录，按顺序检索需要时间为 $M \cdot N \cdot T$ ，数据库中每条记录的平均访问时间为 $\frac{1+MN}{2} T$ 。按类别文件检索，假如这条记录在某个类别文件的最后，则顺序检索时间是 $N \cdot T$ ，各个类别文件的每个记录的平均访问时间为 $\frac{1+N}{2} T$ 。可见，各个类别文件每个记录的平均查找时间是整个数据库中每个记录的平均查找时间的 $\frac{1+N}{1+M \cdot N} \approx \frac{N}{MN} = \frac{1}{M}$ 。（当 $N \gg 1$ 时），类别文件数 M 越大，查找得越快。由于采用这种“类别文件+顺序文件”的存贮结构，在检索时，先访问类别文件，找出相关的类别文件后再进行顺序查找，这样，通过一级索引（类别文件），使得顺序文件的访问方式变成随机——顺序访问方式，加快了查找速度（同时，对有限的明细病名也按这种原则赋予自然数码，按序存放）。这样，也简化、方便了数据库的维护。例如，若主病名为胃炎的情报资料要增加新的资料，只要在以该主病名对应的自然数码 4 命名的类别文件 4 末尾添上新的记录即可；若主病名为胃炎的资料要删除、修改某个记录，只要在以该主病名对应的自然数码 4 命名的类别文件 4 中删除、修改该记录即可，不牵涉数据库中的其他类别文件。

三、检索功能

1. 已知记录中主病名、明细病名、文章名、作者名、期刊名、期号、用户检索文章的时间范围：最初发表时间、最终发表时间等八个项目中的一项、二项，或八项中多个项的任意组合，均可检索出满足条件的所有资料，并按用户的要求打印出所检索到的资料的主病名、文章名、作者名、期刊名、期号、发表日期、页号及内容摘要等八个项目中的全部或部分项目。

2. 在检索过程中，可以控制转盘或换盘，使存贮着医学情报数据的所有磁盘按序轮流地被自动检索。由于多个磁盘能交替更换、轮流地被检索，因此医学情报数据库的规模可以不断扩大，其容量可认为是无穷的，它弥补了微型计算机容量小的缺陷。

检索时，若已知主病名，其对应的数据是类别文件名，只要把此类别文件所在的磁盘放到驱动器中就能迅速被检索。另外，由于文件名、明细病名依序编号，当机器检索到数据库中的文件名或某记录的明细病名号大于所要检索的文章的文件号或明细病名号时，就自动停止检索，表示此后无满足此条件的文章，这样就减少了不必要的无效检索工作量。

四、应用举例

若要检索主病名为“胃炎”，作者名为张学庸的文章，则先查主病名与类别文件名对应表，找到“胃炎”对应的类别文件4，把文件4所在的磁盘放到磁盘驱动器A（在我们的程序中，文件号大于等于9的文件所在的磁盘放到驱动器B检索），然后查“张学庸”对应的电报明码为1728 1331 1661，于是运行情报检索程序，使用人机对话方式，对屏幕上显示的提问信息，依次回答如下：

屏幕上用汉语拼音显示的信息：	用户从键盘回答的信息
文件名	4 ✓
明细病名号	0 ✓
有其他条件吗？	Y ✓ (有)
知文章名吗？	0 ✓ (不知)
知作者名吗？	1728 1331 1661 ✓
⋮	(张学庸)
⋮	0 ✓ (不知)
要打印全部检索信息吗？	Y ✓ (要)

以上右边用圆括号括起的汉字是注解，没有括号的内容才是用户从键盘上输入的，打印出的检索卡内容为：

广西医学情报研究所文摘检索卡

胃 炎：

慢性萎缩性胃炎专题笔谈 张学庸等 实用内科杂志 1 (2) : 57, 1981

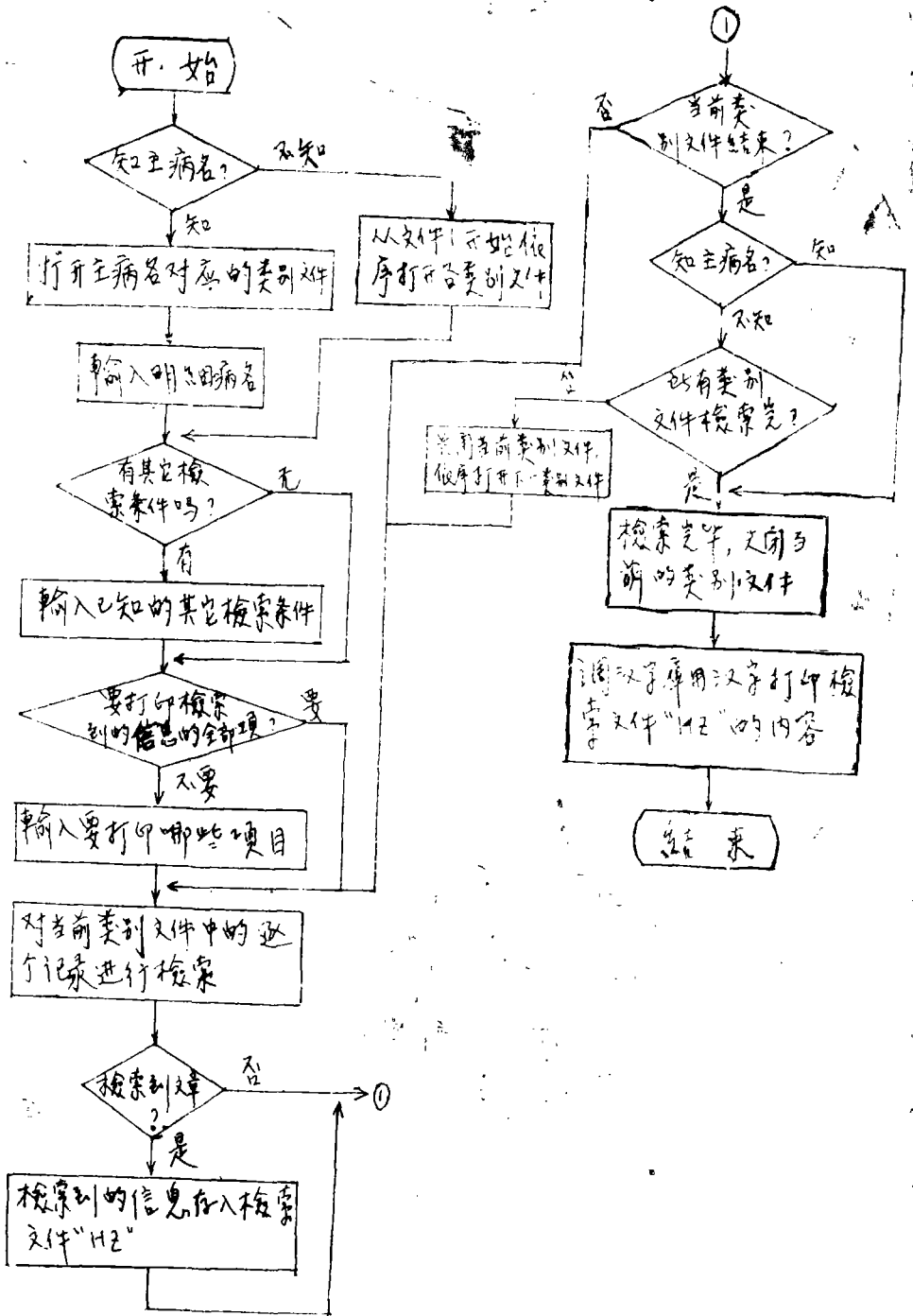
五、优 缺 点

1. 在我们的检索软件中, 检索到的情报信息用汉字打印, 与人们的习惯一致, 用户乐于使用, 便于推广。

2. 微型计算机体积小、价钱便宜、安全可靠、维护方便、使用简单、不受环境限制, 适用于各单位。

3. 微型计算机运行速度较慢, 特别是打印汉字时要调汉字库, 使速度受到一定影响。用微型计算机进行情报检索对于文献情报资料不太大的单位尤其合适, 在一个情报资料专业性较强的业务范围较单一的单位, 微型机的检索速度是足够满足用户要求的。例如, 我们在 BCM I 型微型计算机上检索完四十条资料要30秒—40秒, 这是人工所望尘莫及的。

六. 程序框图



注：本课题参加者：广西计算中心 黄新新、符华儿、何红波、兰红星、李小清、韦南茹
 广西医学情报所 关令华