

●袁国发 邹 华 赵淑君 袁立军

适用于 WINDOWS 环境的 “第二代中文报刊人机通用无号排序法”^{*}

摘 要 中文报刊题名汉语拼音首字母——全拼排序法,适应 Windows 操作环境,简单,方便,快捷,易掌握,人机两用,检索速度快。参考文献 4。

关键词 中文报刊 排序 检索

分类号 G254.132

ABSTRACT In this paper, the authors introduce the Chinese Serial Title Pinyin Initials/ Full Pinyin Sorting Method used in Windows environment. It is convenient, easy to be used, both for human beings and machines and quick in response. 4 refs.

KEY WORDS Chinese serials. Sorting. Retrieval.

CLASS NUMBER G254.132

计算机技术飞速发展,几年前还是 DOS 盛行,如今已是 Windows 天地。原来适用于 DOS 环境的“中文报刊人机通用无号排序法”能不能适用于 Windows 环境?考察显示,两种排序结果并不完全一样。要在 Windows 环境下的数据库管理系统采用原“中文报刊人机通用无号排序法”,就要使用 Microsoft Visual FoxPro 和 Microsoft Access 数据库管理系统与 FoxBASE 数据库管理系统之间的导入导出操作。为在 Windows 环境下的数据库管理系统中直接解决中文报刊人机通用无号排序的问题,我们研究设计出了纯粹适用于 Windows 环境的“第二代中文报刊人机通用无号排序法”。

1 第二代中文报刊人机通用无号排序法

它适用于 Windows 环境下的 Microsoft Visual FoxPro 数据库管理系统和 Microsoft Access 数据库管理系统的数据库操作。数据表可在 Microsoft Visual FoxPro 或 Microsoft Access 系统创建,也可由 DOS 系统导入。

适用范围:中文报纸数据表和中文期刊数据表的题名排序、检索和分类排序、检索。

它继承了原适用于 DOS 操作环境的“中文报刊人机通用无号排序法”最基本的排序方法,绝大部分报刊的排列顺序和原排序法相同,继承了它的主要优点,又适应了 Windows 这一新的操作环境,既克服了原“排序法”的不足,又避免了 Windows 环境数据

表单单纯题名字段排序中错误问题的出现,既适于计算机操作,又便于手工排检,它是 Windows 环境下的“中文报刊题名汉语拼音首字母——全拼排序法”,故称为“第二代中文报刊人机通用无号排序法”。

2 Windows 环境下的“中文报刊题名汉语拼音首字母——全拼排序法”

2.1 含义

“中文报刊题名汉语拼音首字母”是指中文报刊名称每个汉字的汉语拼音的首字母。“全拼”是指报刊题名汉语拼音的全音节,还包括它的声调及笔画数。在中文报刊数据表中,除设置“题名”、“出版地”、“分类号”等必要的字段外,还需增设一个长度为 18b 的名为“配”的字段,用于放置与题名相匹配的汉语拼音首字母。“中文报刊题名汉语拼音首字母——全拼排序法”是“中文报纸题名汉语拼音首字母——全拼排序法”和“中文期刊题名汉语拼音首字母——全拼排序法”的简称。

2.2 题名的选“配”

题名中每个汉字的汉语拼音首字母都在所取之列。汉语拼音字母或英文字母皆按汉语拼音字母对待,如《中国科学 A 辑》就配“ZGKXAJ”。题名中的阿拉伯数字则按汉字对待,如《21 世纪报》则配“EYSJB”。题名中的其他符号如标点符号等均不在“配”中表示。

^{*}本文系河北省社会科学规划项目的研究论文。

2.3 排序方法

当需用计算机编排名目时,不管原来建制的数据库是按什么顺序编排的,只要按“配”、“题名”字段的先后次序,利用“查询”进行组合排序即可。在 Microsoft Visual FoxPro 利用菜单栏“工具”下拉菜单“向导”级联菜单“查询”选项的“查询向导”,在“查询向导”,利用“步骤 4 - 排序记录”,依次选取“配”、“题名”字段,确定“完成”,即可“浏览”。在 Microsoft Access 利用菜单栏“插入”级联菜单“查询”选项的“新建查询”中的“设计视图”,进入“查询 1:选择查询”对话框,在该对话框进行“配”、“题名”字段的题名组合排序即可。在 FoxPro 命令窗口使用“SORT ON 配,题名 TO<题名目录库名>”命令也可。所排出的题名目录就是按题名汉字的汉语拼音首字母以字母表顺序排列的。首字母无法区别前后的相同首字母报刊,则从题名的第一个汉字开始,以汉语拼音的全音节相比较,按汉语拼音字母表顺序排列前后,多音字取常用的发音。第一个字的音节全相同者,再按它的声调排前后,同音同调的字则以该字的总笔画数由少到多排前后,同音同调同笔画者再按该汉字的笔形顺序排前后。当需要编制分类目标时,只需按“分类号”、“配”、“题名”字段的先后排列次序利用“查询”进行组合排序即可,在 FoxPro 命令窗口使用“SORT ON 分类号,配,题名 TO<分类目录库名>”命令亦可。这样排出的报刊分类目录将是按“分类号”、“配”、“题名”字段排序的分类目录,相同分类号的报刊将是按上述的题名排序法排序的。

2.4 检索方法

在 Microsoft Visual FoxPro 利用菜单栏“工具”下拉菜单“向导”级联菜单“查询”选项的“查询向导”。在“查询向导”,利用“步骤 3 - 筛选记录”,选取“配”字段,输入题名首字母,或选取“分类号”字段,输入分类号,即可实现报刊题名或分类的检索。在 FoxPro 命令窗口,输入“LIST[字段名列表]FOR 配 = ‘<配>’”(单引号中的<配>为所要检索的报刊题名的首字母)命令,或使用“DISP[字段名列表]FOR 分类号 = ‘<分类号>’”命令也可。在 Microsoft Access 利用菜单栏“插入”级联菜单“查询”选项的“新建查询”中的“设计视图”,进入“查询 1:选择查询”对话框,在“准则”行“配”字段输入所要检索的题名首字母,或在“准则”行“分类号”字段输入所要检索的刊物类号。然后,单击“查询”下拉菜单的“运行”选项,即可显示检索结果。

2.5 其他文种报刊和少数民族语言报刊的排序

这类报刊在多种报刊征订目录、征订单及报刊大全之类的书籍中,其题名的使用很不一致,一般使用中文译名,随后注以文种,但有个别的其他文种的中国报刊是用的外文题名。这类报刊如需按中文报刊进行排序,在数据表中应把中文译名作为“题名”对待,随后注以文种。

3 Windows 环境“中文报刊题名汉语拼音首字母—全拼排序法”的特点

(1)简单易掌握。只要使用者熟悉汉语拼音和所使用的分类法就行。绝大部分题名的排序是单题名首字母,根本不用去逐字考虑它们的确切发音,一般不会出现错误判断。

(2)其参加排序的题名首字母是根据题名的标准发音专配的,不会因题名含多音字而错排位置。即便在相同首字母报刊的题名全拼排序中偶尔遇到多音字题名的错误排序,错排的位置也仅在临近的相同首字母题名的范围之内,最多不超过 7 条记录。

(3)这种排序法既可用于书本式报刊目录的排序,又可用于卡片式报刊目录的排序,还可用于中文报纸、期刊的现刊排架及合订本的排架。而这诸多方面的排序,除编制分类目录需配给每种报刊一分类号外,均不用另外配备种次号或其他辅助符号,也不会出现异刊重位和一刊多位的现象。这样,报刊合订本分类时就无须再配种次号,报刊合订本及报刊合订本登录卡上也均不需再著录种次号,报刊合订本分类人员也就省去了编制“种次号记录卡”的麻烦。报刊合订本管理人员和读者在取刊和放刊时,只要按刊物的类号及其题名寻找位置即可,无须考虑每一种报刊的索刊号。

(4)重名报刊易区别前后。为解决重名报刊的排序问题,可在组合排序字段的后边再增加 1 个“出版地”字段,它的排列顺序即是按“首字母 - 拼音 - 声调 - 笔画 - 出版地”编排。分类排序则使用“分类号、配、题名、出版地”4 个字段组合,排列顺序为“分类号 - 首字母 - 拼音 - 声调 - 笔画 - 出版地”。FoxPro 利用“查询”排序,只能使用 3 个字段组合。为解决“同名报刊”的前后排序问题,所使用的数据表应是首先利用“配”、“题名”、“出版地”排序过的,再用“分类号”、“配”、“题名”3 个字段组合排序,就可达到使用“分类号、配、题名、出版地”4 个字段组合排序的效果。否则,只能在命令窗口使用“LIST ON 分类号,配,题名,出版地 TO<分类库名>”的排序命令

排序。在 Access 可使用 4 个字段组合排序,达到按“分类号-首字母-拼音-声调-笔画-出版地”排序的效果。其中,重名报刊将是按刊物“出版地”第 1 个汉字的汉语拼音排序。手工排序也照此办理。

(5) 在计算机文字符号的输入中,有半角、全角之分。在数据表题名的输入时,由于题名中所包含的符号的输入方式及表示方法的不同,就有可能造成检索时输入的题名和建库时输入的题名不一致,因而造成查漏。使用题名首字母检索就不用考虑这些问题,无论原来是哪种方式输入的,都可以检索出。

(6) 检索速度快。现在,用计算机通过题名检索报刊,多采用题名全拼检索法。而汉字注音每字平均由 3.2 个字母组成。要检索 5 个字的题名,不包括确认和换行选字的击键次数就要 16 次。再加上确认和换行选字的击键次数,就不只 16 次了。而我们所设计的排序法,检索题名时只需敲入题名的首字母,几个字的题名敲几次键就行。我们曾做过抽样调查,首字母比全拼的速度要平均高出近 3 倍。使用计算机搞题名检索,还有一个更快的方法,就是在使用全题名首字母检索的基础上,以减少所输入的首字母来进一步提高题名的检索速度。一般来说,报纸只需输入前 3 个首字母,期刊只需输入前 4 个首字母。

(7) 人机两用。使用它,手工排检简单方便易操作,又能利用计算机方便地排序和打印目录。该排序法既可用手工排检,又可用于计算机排检,而且可以保证手工排序和计算机排序的绝对统一。

(8) 可实现中文报刊分类排序的统一。我国每一种正式发行的报刊都有一个国内统一刊号,期刊刊号的末尾都配有该刊的中图分类号,我们可以利用该分类号来实现每种期刊分类号的统一。结合本排序法的使用,就可达到期刊分类排序的统一。报纸统一刊号中目前还没有附带的分类号,但报刊司内部也为每种报纸配有一个类别代码,如果报刊司能把他们的数据库发送到网上,我们也就可方便地解决报纸分类号的统一问题,可达到报纸分类排序的统一。

4 对 Windows 环境“中文报刊题名汉语拼音首字母-全拼排序法”的考察

4.1 对报纸的考察

2906 种被考察的报纸,有 2273 种单使用报纸名称汉语拼音首字母就可排出前后顺序,占总测数的 78.22%,和 DOS 系统排序的结果一样。再使用汉语拼音全音节排序就可排出前后顺序的有 474 种,占总测数的 16.31%,比 DOS 系统排序多 0.62%。再继续使用声调排序能排出前后顺序的有 75 种,占总

测数的 2.58%,比 DOS 系统排序多 0.27%。使用以上 3 步排序可保证原排序法 96.22% 的报纸名称排列顺序的不变。需继续使用笔画区别前后的共 60 种,占总测数的 2.06%,比 DOS 系统使用笔顺排序的多 0.10%。而特殊排序的报纸仅有 2 种,占总测数的 0.07%,比 DOS 系统排序少 1.00%,该特殊排序的报纸就是首字母同为 NJRB 的内江日报和南京日报,它是由不规则发音的多音字“内”引起的,使“内江日报”排在了“南京日报”的前边。“内”在汉字排序中发 n à 音,是 Windows 操作系统数据库汉字字段排序中多音字发音的一个特例。原来在 DOS 系统排序出现的 8 组 31 种因第二级不常用汉字造成的特殊排序的报纸已全部变成了其他方式的排序,其中 5 组 20 种已变成了拼音排序,2 组 8 种已变成了声调排序,1 组 3 种已变成了笔画排序。

4.2 对期刊的考察

7839 种被考察的期刊,在数据库排序中有 6941 种单使用刊名汉语拼音首字母就可排出前后顺序,占总测刊的 88.45%,和 DOS 系统排序的结果一样。再使用汉语拼音全音节排序就可排出前后顺序的有 699 种,占总测刊的 8.92%,比 DOS 系统排序多 0.09%。再继续使用声调排序能排出前后顺序的有 129 种,占总测刊数的 1.65%,和 DOS 系统的排序结果相同。使用以上 3 步排序可保证原排序法 99.22% 的刊名排列顺序不变。需继续使用笔画区别前后的有 40 种,占总测刊数的 0.51%,和 DOS 系统使用笔顺排序的期刊一样。这 40 种期刊共由 16 组相同首字母刊组成,其中 15 组 37 种刊均按刊名的笔画数排出了前后,只有 1 组 3 种重首字母刊,出现了两个同音同调同笔画数的汉字,它们就是首字母为 FZ 的相同首字母刊“发展”、“法制”、“法治”。特殊排序的期刊共 2 组 6 种,仅占总测刊数的 0.08%,比 DOS 系统排序少 0.09%。原来在 DOS 系统排序时出现的 5 组 13 种特殊排序的相同首字母刊,其中的 2 组 4 种,因刊名第 1 个字是第二级不常用汉字,排序时已归入了拼音排序;1 组 3 种已归入了声调排序。只留下 2 组 6 种因多音字打头引起的特殊排序刊,它们是首字母为 SH 的“参花”、“山花”、“社会”、“收获”,和首字母为 ZW X 的“传记文学”、“湛江文学”。它们分别是由两个多音字造成,那就是发 c ān、c ān、sh ān 音的“参”字和发 chu án、zhu àn 音的“传”字。它们的常用发音分别为 c ān 和 chu án。因此,在首字母为 SH 的相同首字母刊中,使文学类期刊“参花”排在了“山花”、“社会”、“收获”三刊的前边,使首字母为 ZW X 的相同首字母刊“传记文学”

排在了“湛江文学”的前边。

4.3 考察结果

适用于 Windows 环境的“第二代中文报刊人机通用无号排序法”和适用于 DOS 环境的原“中文报刊人机通用无号排序法”在报刊排序中,利用题名首字母、汉语拼音全音节、声调 3 步排序,可保证绝大部分报刊题名排列顺序的不变。在相同首字母报刊中,由于 Windows 系统数据表汉字字段排序中不存在一、二级汉字之分,题名全部实现了拼音排序,特殊排序的报刊仅存在于极少数的多音字打头的题名之中,不足全部报刊的 0.1%,更接近于传统的题名拼音排序。

5 结论

“第二代中文报刊人机通用无号排序法”既适应了更先进的 Windows 操作环境,又继承了原“中文报刊人机通用无号排序法”的基本排序方法,继承了原“中文报刊人机通用无号排序法”简单、方便、快捷、易掌握、人机两用、检索速度快等优点,又摆脱了原“中文报刊人机通用无号排序法”使用中第二级不常

用汉字对拼音排序的干扰,比原“中文报刊人机通用无号排序法”更科学,更适用。

参考文献

- 1 袁国发等.中文期刊人机通用无号排序法.大学图书馆学报,2000(6)
- 2 袁国发等.可实现中文期刊统一排序的人机通用无号排序法.图书情报工作,2000(12)
- 3 袁国发等.中文期刊刊名的计算机快速检索.中国图书馆学报,2002(1)
- 4 袁国发等.中文报纸人机通用无号排序法.图书情报工作,2002(7)

袁国发 副研究馆员。通讯地址:石家庄市裕华东路 186 号,河北科技大学图书馆。邮编 050018。

邹 华 河北科技大学图书馆馆员。通讯地址同上。

赵淑君 河北科技大学图书馆馆员。通讯地址同上。

袁立军 石家庄乐人堂医药经营公司工作。通讯地址:石家庄市中山东路 161 号。邮编 050011。

(来稿时间:2002-07-10)

专题讲座暨学术研讨会征文通知

技术、经济、社会的飞速进步决定了图书情报事业的发展要发生深刻变化,变革促成了图书情报事业在理论和实践上的研究热点层出不穷:信息资源管理概念的提出、知识管理理念的引入、数字图书馆的建设、图书馆发展机制的变革、读者服务方式的转变、信息技术的广泛应用、专业教育的发展等。为了能够透过多方视点,探讨学科热点问题,发表业界各家观点,决定由中国科学院文献情报中心主办,图书情报工作杂志社承办,中国图书馆学会专业图书馆分会协办,于 2003 年 5 月中旬在北京举办大型专题讲座暨学术研讨会,现面向社会各界广泛征文。

一、会议主题:热点、视点、观点——图书情报事业面面观。

二、会议形式:1. 邀请国内外著名专家学者作专题报告;2. 组织代表进行论文交流。

三、征文内容:

1. 图书情报基础理论研究;
2. 图书情报学科体系的变革与教育发展;
3. 图书情报事业的发展战略;
4. 图书情报网络和资源共享;
5. 图书情报服务(侧重现代化电子文献信息服务);
6. 图书情报机构的管理;
7. 图书情报从业人员的资格认证与素质培养;
8. 图书情报事业与现代信息技术;
9. 图书情报事业发展中的法律问题;
10. 其它。

四、征文要求:1. 来稿主题明确,论据充分,数据可靠。字数在 3000~5000 字为宜。2. 论文格式:中文标题、中文文摘(300 字以内)、中文关键词(3 个以上)、正文、参考文献、作者简介(包括姓名、性别、年龄、职称、职务、工作单位、电话、通讯地址、邮政编码)。3. 请在来稿第 1 页左上角注明“热点征文”字样。4. 来稿不退,请自留底稿。

五、参会方式:投稿参会;提交论文摘要参会;直接报名参会。

六、时间要求:提交论文、论文摘要或报名截至到 2003 年 4 月 10 日。会议具体时间待定。

七、联系方式:地址:北京中关村北四环西路 33 号 614 室(100080)。联系人:常彤。

电话:010-82623933 010-82626611-6614 传真:010-82621460

E-mail: journal@mail.las.ac.cn