

# 计算机汉字输入方案的选择

唐 玲

(总参三十一研究所)

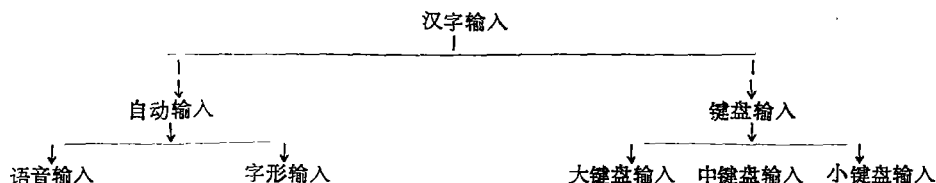
**提 要:** 本文对国内使用的汉字编码输入方案, 进行综合分析, 从中挑选了四种汉字输入方案, 即“层次四角”码、“大众”码、“汉语拼音正词法”和“五笔字型”, 供办公自动化中的不同人员使用。

所谓汉字输入就是将汉字通过一定的方式输入到计算机。如何将数以万计的方块字通过26个字母键输入到计算机。这一问题是汉字信息处理的“瓶颈”问题。汉字与西文不同, 字形复杂, 字数众多等, 因此为汉字输入计算机带来了许多困难。

自五十年代末开始, 我国便开始了对汉字输入的研究。但由于受当时计算机存储容量的限制许多问题没有解决。到七十年代由于集成电路和磁盘存储器的出现, 使汉字库的存储成为可能。这给汉字输入提供了物质条件, 由此汉字输入技术有了较大发展。特别是近几年来, 汉字编码输入有了迅速的发展。据不完全统计, 目前已有七百多种汉字编码输入方案, 已有文字根据的也有一百多种, 如何利用现有资源为办公自动化服务, 现对现存的汉字输入方案加以剖析。

## 1 汉字输入的几种方法

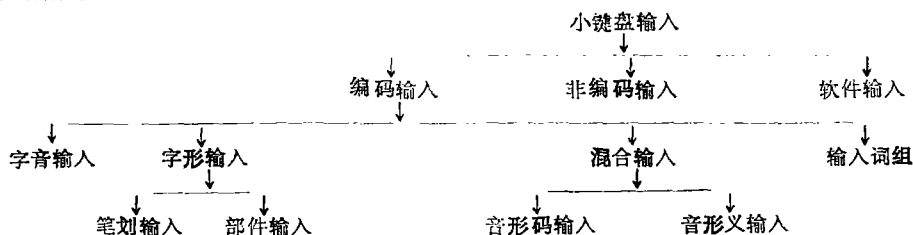
根据汉字输入时所用设备不同, 可以分为以下两类: 自动输入方式和键盘输入方式。如下图所示。



自动输入方式包括语音输入和字形输入两种形式。这种输入汉字的方法是最近几年才发展起来的, 目前受到国内外的广泛重视。尽管还有许多问题尚未解决, 但随着计算机硬件的发展, 它将使汉字输入有很大突破。

目前使用最为普遍的是利用小键盘来输入汉字。其优点是设备价格低廉, 与西文字符的输入设备有很好的兼容性, 可以实现盲打, 具有广泛的推广意义。专职操作人员, 在某些汉字输入方式下能达到比西文输入更快的速度。

现行的绝大多数汉字输入方案都是以小键盘为对象进行研究的。到目前为止各方案的基本思想都是对汉字进行编码, 然后进行输入。编码的方法因人而异, 从而使得汉字输入方案数以百计, 但是我们仍然可以归结为以下几种类型。



从这些现存的资源中要挑选出适于办公自动化使用的方案。就必须对这些方案加以比较进行优选。

## 2 众多方案的比较

### 2.1 比较的基点:

目前还没有统一的标准去衡量现存汉字输入方案的优劣,只能借助一些辅助手段。86年举行了全国首届汉字输入比赛,从比赛的结果中反映出了许多东西。例如:方案本身编码的优劣;静态重码的多少;输入字元的分布是否合理等,只有当这些问题都解决得比较合理时,整个汉字输入的速度才能提高。那么我们可以在这场“比赛”的基础上,同时考虑到其它因素来进行比较,从而挑选出适用的方案来。

### 2.2 方案挑选的范围:

首届汉字输入比赛共有34种方案参加,有11种被评为A类优秀方案。这11种方案是:“大众汉字拼形编码”,“五十字元多能电脑输入系统”,“中文声数编码”,“宏观字形编码”,“《层次四角》编码”,“前三末一汉字输入编码”,“部形码”,“笔形编码法”,“联合45—3汉字拼形输入法”,“GKG码”,“JDL无间隔中文输入方案”。这些方案不仅有输入速度快,误码率低的特点,更主要的是这些方案已有了众多的用户,并且对其方案本身都有较好的评价。

除了以上这些方案外,还有一些发明较早的汉字输入方案,由于历时较长和广泛的宣传作用,到目前为止已拥有许多用户。另外还有一些拼音类方案,因为有些用户在今天挑选方案时就已考虑将来中国文字的发展问题,所以选用了这类方案。

当然优秀的编码方案它的用户不一定就多,反之,结论也成立。由于当今是信息社会,怎么组织信息的流通,信息流通时间的长短对用户都起着决定性的影响作用。

### 2.3 综合考虑挑选出四种汉字输入方案:

办公自动化主要针对两类不同人员而言,第一类是专职操作人员,第二类是机关办公人员。对于这两类不同的人员,他们对汉字输入方案的要求是不同的。前者要求使用的方案,经过一定时间的学习训练之后,能够达到较高的输入速度,以便提高工作效率。而对后者,输入速度不是主要问题,主要的是使用方法要简单,易学,易记,方便临时性的使用。根据这些要求,同时还考虑到汉字输入方案本身具有的特性,及办公自动化将来向综合方向发展,而汉字输入方案必须要适应这种发展的需要等,这众多方面的因素,综合起来认为将以下四种方案同时配在机器上,让用户在使用时有挑选的权力,这些方案是《层次四角》编码方案,《大众》汉字编码,《汉语拼音正词法电脑输入系统》,《五笔字型》汉字输入。

### 3 四种汉字输入方案简介

#### 3.1 《层次四角》编码方案:

纵观所有的汉字输入方案,目前输入速度最快的汉字输入方案是《层次四角》编码。这是一种纯字形编码方案。该方案选用了103个部件,其中有93个部件是新华字典中出现的偏旁部首,有10个是自己引入的,这些部件按使用频次合并压缩在26个不同的字母键上。

该方案对于一般汉字输入的取码规则是:依次提取每个字的左上、右上、左下、右下角的各部件,每取一个部件则移去此部件,在所余笔划中再提取,最多提取四个部件。例如:输入汉字“颁”,依次取它的四个角部件,分别是八、一、ノ、贝,然后找到对应的键,顺次击键输入即完毕。取码规则中还规定每取完一个角,应将取过的部件移走,在剩下的部分取其它角。其含意可用下列字的输入来说明,例如输入“癌”字,四个角应拆分为广、口、山、口,按照以上规则,若有些字取不够四码已无部件可取时,按字结构分为两种情况,对于左右结构的字应补以左右结构标志“∞”,若还不够四码,应再补一空格键做为结束标志,若是其它结构的字则只补空格键做结束标志。该编码方案还编有26个高频字,这些字直接分布在不同的26个键,当要输入高频字时,只需击该字所在的键,再输入一空格键就行了。

方案对高频字的重码也进行了处理,另外还装有2300条常用词汇,方案还为用户割下了300多个词汇空间,以备用户使用。有了输入规则及高频字使用等手段,大大提高了输入速度,平均码长为2.49。据86年5月全国汉字编码评测委员会在专职操作人员表演现场测试,该方案5分钟可打1150个字(即每分钟230个汉字)。一般人经过学习后也能输入100—130字/分。

该方案编码时没有完全解决重码问题,经采取各种措施使重码率为0.77。当输入了有重码的编码时,机器响笛提示,同时在屏幕的最下面按使用概率由大到小的顺序依次出现重码字,供选择。这种编码方案便于掌握,只要经过短期训练,输入速度会迅速提高。特别适合专职录入员使用。

#### 3.2 《大众》汉字拼形输入:

该方案选定了172个字元,其中包括128个基本字元及44个联想字元。该方案为了便于学习掌握,采用了“模拟字记忆法”。即将172个字元的大部分归并成26个模拟字,对这些字贯以读音,按其发音的第一字母分别定位于26个不同的键上,还有少量没有归并的字元叫某个基本字元的联想字元,列在相应模拟字的下方。有了这种模拟字记忆法,只需记住26个模拟字及读音就能输入汉字了。组成模拟字的字元分布是经统计按频次规律分布在不同的键位上的。

《大众》汉字输入的拆拼规则为:对所有汉字,一律按书写顺序和大元优先准则取其第一、二、三、尾最多4个字元。输入汉字时首先应注意字的书写顺序。而“大元优先准则”是指小于整字的最大可能字元,即笔划数尽可能多的字元。要输入汉字时,按照以上规则,把它拆成一串最多四个的字元,再找到每个字元所对应的键然后顺次敲击这些键。例如:输入汉字“就”,应提取输入字元为二、口、小、尤,所以只需敲这四个字元分别在的键就能把汉字“就”输入。若有的字拆分不足四个字元,输入时就追击一次空格键。

该方案在编码时编入了24个高频字,它们分别在24个不同的键上,输入时只需要击高频字所在的键,然后再输入一个空格键即可。

《大众》编码输入,采用高频简码字后,平均输入汉字的击键次数为3.42键/字。该方案还存在一定的重码,重码使用概率为0.379%。重码的处理方法与“层次四角”编码相同。有了以上措施,使方案的输入速度也名列前茅。据测专职操作人员在单字方式下输入速度可达120字/分。在词汇方式下可达150字/分。

该方案具有造词造字的功能,还允许用户自定义词汇,具有一定的输入速度,易学,易记。作为机关办公人员使用是很合适的。在全国第二次工业普查中各省、市、自治区已广泛采用了《大众》码输入方案。

### 3.3 汉语拼音正词法

前面介绍的《层次四角》编码及《大众》编码,它们都采用了汉字拼形输入方式。如果想让“中文电脑”较完全地发挥功能,光有字形编码键入法是不够的。有许多场合需要一种更接近语言,思维自然的键入法才能胜任。拼音键入法比起任何字形编码键入都更接近语言自然,思维自然。从长远来说,它是理想的标准键入法,据有关部门分析,目前中文汉字正处于发展之中,正向拼音文字方向发展。鉴于这种发展情况及目前汉字输入方案使用单位的反映,在这里介绍一种拼音输入方式——“汉语拼音正词法”。

这种输入方式的主要特点是:采用了国际通用字符键盘。将声母和单韵母按通用的26键排列;复韵母和鼻韵母以齐齿呼,开口呼,合口呼和撮口呼的顺序,分别安排在下、中、上三排键上;零声母取首字母,声调用数字键表示。整个输入系统包含了GB2312—80的6763个汉字,并且还约有一万条词语,按双字、三字、四字以上的结构排列。输入方式采用声韵相拼的原则。输入速度对熟练的操作人员,在词汇方式下,能达到141字/分。

这个系统,符合现代汉语的规律和中文信息处理工程的需要,具有良好的扩充性能,是国内首创。其键面布阵,根据语言文学的原理,采用了语言信息压缩技术,合理使用了国标通用字符键盘。这种输入方式对于具有拼音知识的人,稍加学习便能使用。

目前使用这种输入方式的单位很多。据有关使用单位反映,该输入方式是拼音类方案中最佳方案之一。在将来的文字发展中,它作为计算机的一种汉字输入方式有较强的生命力。

以上三种方案分别满足办公自动化中各种情况的要求,同时方案都具有词汇输入,这样不仅提高了输入速度,更主要的是降低了误码率。

### 3.4 “五笔字型”编码输入:

这里再简单说一下大家都比较熟悉的“五笔字型”编码输入。这种汉字输入方式是较早的社会产品,在各行业都有一定的影响,目前采用这种输入方案的已有160多个部委级。在86年的首届汉字比赛中,由于主观和客观的因素,这种方案没有被评为A类方案。但就方案本身来说也有它的独特性。例如,输入速度较快,可达160个汉字/分。方案设计从用户角度出发,经过几年的系统完善,已产生出了专门的汉卡与小容量机器相配置。是一个具有硬件,软件较完善的汉字输入系统。它有“五笔画”,“五笔字型”两种输入法,另外具有灵活的“词汇处理”,智能化“联想输入”等有效设计。这种输入方案的发明专利,已由美国的DEC公司购买,今后,联合国的所有计算机的中文输入都将采用该输入方案。

根据上述原因,在目前国内汉字输入方案中,挑选出了“层次四角”编码,“大众”码,“汉语拼音正词法”,“五笔字型”输入四种汉字输入方案,供有关用户的不同人员参考使用。