

第一章 计算机汉字处理简述

汉字输入是计算机中文信息处理的关键。随着计算机技术的普及和推广,计算机的应用已经深入到企业、公司、机关、学校及家庭等各个领域。计算机汉字输入早已不仅仅是少数专业录入人员的工作,而且成为千千万万个普通干部、技术人员、职工、作家、记者、教师、学生不可缺少的日常工作。如果不能高效率地解决他们的汉字输入问题,要真正实现中文信息处理的现代化将是不可能的。

1.1 计算机的汉字处理

1.1.1 汉字的输入与汉字外码

汉字处理的首要问题就是解决输入这个瓶颈。虽然,让汉字进入计算机的方法很多,例如:语音转换输入、光电输入、联机手写输入和键盘输入等许多方法。但是目前最成熟、最简便易行和实用化的还是键盘输入。

所谓键盘输入,就是通过计算机的键盘,把全部汉字或者各种符号输入计算机。我们知道,计算机键盘上标出的都是英文字母或符号,每当我们敲下一个键时,在屏幕上就会显示出这个键上标出的字母或符号。那么怎样才能使得我们在键盘上敲下字母或符号键时,在屏幕上显示出汉字呢?这正是多年来我国许多计算机专家孜孜不倦奋发解决的问题。

西文是一种符号文字,因为它只由 26 个字母组成,所以由键盘直接输入字母和符号而完成整体文字的组合,将是轻而易举的。但汉字是一种象形文字,仅常用汉字就有 6 千多个,无法直接从键盘输入。于是人们就用符号来表示汉字(比如电报码用四个数字表示一个汉字),通过在键盘上输入符号,并经转换而间接地输入汉字目的。

因为汉语拼音字母和英文字母十分相似,自然人们也就想到了用汉语拼音来输入汉字。我们把英文字母看成是汉语拼音的字母,把汉语拼音中的特殊字符用相应字母替代(如韵母 ü 用字母 v 替代)。于是在计算机键盘上就可以用拼音输入汉字了,这就是汉字输入的拼音码。

由于许多汉字的拼写太长,输入一个汉字的拼音需要多次操作键盘按键,并且汉语中同音异字繁多,还由于我国地域广大人口众多,各地方言千差万别,掌握拼音也很难准确,并且遇到生字无法读音。所以,人们又把汉字按笔划的构成拆成许多基本部件,把每个部件对应到一个固定的键位上,输入汉字时再用不同的部件组成各种汉字,这就是汉字输入的字形码。

由于不同目的和环境的要求,人们根据汉字各种特性,制定不同的规则,便可以产生许许多多不同的汉字输入编码规范。也就是说,可以有许许多多不同的汉字编码系统和

汉字输入法。这些编码规范都是人和计算机交换信息时使用的;我们把它们通称为汉字外码。

1.1.2 汉字的编码

因为汉字的特点是字量大、字形复杂,所以任何汉字处理系统中必须解决汉字的输入、输出和存储问题。

汉字是记录汉语的图形符号,利用光电扫描的方法直接识读汉字,或者利用语音识别技术,把汉字直接读入计算机,或者使用手写输入方式输入汉字都是很好的汉字输入方式。但是这些汉字输入技术的难度都比较大,成本也都比较高。目前,这类汉字模式识别技术的汉字输入系统还处于初级不太成熟的阶段。目前最成熟,最简便易行和实用化的还是键盘输入法。

我们知道,计算机只能识别处理数字符号 0 和 1。所以输入计算机的任何数据信息都必须转换成各种不同的 0 和 1 的组合后才能被计算机接受。把各种不同的数据信息分别用各种不同的 0 和 1 的符号串表示,这些 0 和 1 的符号串在计算机中就代替了其对应的数据信息。我们把这一替代过程叫做编码,把这些 0 和 1 的符号串叫做代码。

1. 汉字的内码

由于不同的汉字输入码(外码)差异很大,有两字节的、三字节的,有等长码的、不等长码的。因此,为了节省存贮空间,简化和统一处理的方便,汉字在计算机内处理时都采用了汉字内码技术。

汉字的机内码(或叫汉字内码)是计算机系统内部处理和存储汉字时使用的由 0 和 1 符号串组成的代码。由于英文处理比较简单,所以其内码就是 ASCII 码。ASCII 码用八位二进制数表示一个字符,其中第一位是奇偶校验位。而汉字在我国大陆范围内,基本上都采用国家标准的 GB2312《信息交换用汉字编码字符集——基本集》作为内码。这种国标码采用两个字节表示一个汉字,每个字节只使用了后七位。这种设计方案虽然使得汉字与英文字符能够完全兼容,但是英文字符与汉字混合存储时还是很容易发生冲突,或混淆不清,所以人们实际上总是把国标码的两个字节高位置 1,或者把其中一个字节的高位置 1 后再作为汉字的内码使用。这样一来,汉字内码才能即兼容英文 ASCII 码,又不与 ASCII 码产生二义性,同时汉字内码还和国标码具有很简单的对应关系。

2. 汉字的交换码

实际上,各地的汉字内码是不统一的。台湾的汉字系统、日本的汉字系统、韩国的汉字系统以及东南亚地区的汉字系统都各不相同。甚至在我国大陆,不同的汉字系统中汉字的内码也不完全相同。统一汉字内码是人们久已期待的愿望,但是这个目标在短时间内还不可能实现。为了使装有不同汉字内码的系统之间能够实现正确的汉字信息交换,必须制定统一的汉字交换码。

目前国内计算机所采用的标准信息处理交换码是根据 GB1988《信息处理交换用七位编码字符集》和 GB2311《信息处理交换用七位编码字符集的扩充方法》标准制定的。因此

汉字交换码应当与 GB1988 兼容,并根据 GB2311 规定的方法编制。由于汉字数量远大于七位编码的容量,所以一个汉字必须用两个或多个字节表示。

根据这些要求,我国首先对六千多个常用汉字制定了交换码的国家标准,即 GB2312《信息交换用汉字编码字符集——基本集》。其中每个汉字都用对应于 GB1988 的两个七位码来表示。这是一种标准汉字交换码,它含有 6763 个汉字和 682 个字符,每个汉字或字符都用两个字节的 0 和 1 的串表示。

3. 汉字控制功能码

能够直接影响汉字数据的格式处理、传输方式和解释执行的代码叫做汉字控制功能码。它总是和汉字数据代码混合组成汉字数据流,并且贯穿于汉字输入、机内处理、汉字输出等整个汉字处理的全过程。

1.1.3 汉字的输出

所谓汉字的输出,就是在计算机输出设备(显示器、打印机等)上根据需要输出汉字的字形。

汉字字形存放在汉字库中,而汉字库又可以固化在 ROM 芯片中,或存放在软、硬磁盘中。固化有汉字库的 ROM 芯片若安装在打印机中就是汉字打印机,若安装在显示器中就是汉字显示器或汉字终端。在计算机硬盘中安装了汉字库及汉字处理系统就可以使一般的西文计算机系统具有汉字输出能力。

1. 汉字库

在计算机内处理的是汉字的内码,而在显示器、打印机上要输出的是汉字的字形,所以必须事先把全部的汉字字形都放在计算机内,并通过汉字处理系统的专门处理程序自动把要输出的汉字内码转换成对应的汉字字形后输出。

全部汉字字形的集合叫做汉字字形库(简称汉字库)。汉字的字符集很大,不包括罕见字、古字,仅现用汉字就有一万多个。目前我国的汉字系统一般都使用 GB2312 规定的两级常用汉字建立汉字库。其中一级汉字有 3755 个,二级汉字有 3008 个,再包括 700 个左右的字母、数字和图形符号;一共有大约有 8000 多个汉字字符。

汉字字形的表现形式主要有:铅字形式和点阵形式。铅字形式用于打字机和印刷机,而点阵形式特别适合于计算机的汉字处理。所谓点阵形式,就是把字符图形放在网状方格内用黑白点表现;在网格内,凡是有笔画通过的网点为黑点,否则为白点,例如图1-1和

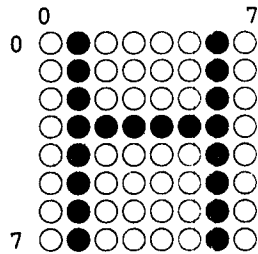


图 1-1

图 1-2 所示。

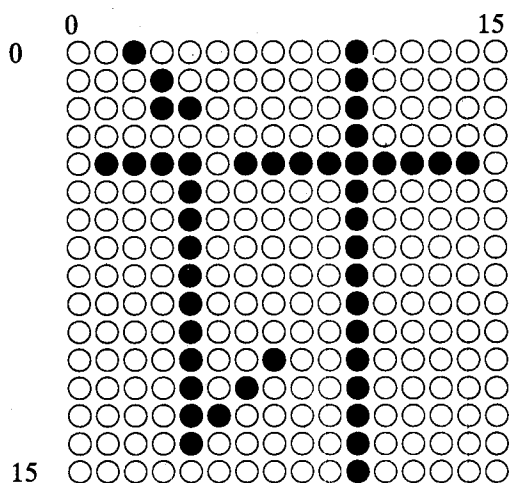


图 1-2

其中图 1-1 表现的是英文字符 H 的 8 * 8 点阵形式, 图 1-2 是汉字“计”的 16 * 16 点阵形式。在这里, 表现图形字符的每个黑白点是字符图形的最小元素, 我们叫做位点。

采用点阵形式表示汉字, 在字形的数字化和存储时都比较容易; 因为每个位点都有黑白两种状态, 正好对应了二进制数的 1 和 0 两个数字符号。比如图 1-1 的英文字符, 每行可看作一个字节, 用 8 个字节就可以表示了。而图 1-2 的汉字字符, 每行要用 2 个字节, 一共要用 $2 * 16 = 32$ 个字节来表示。

一般说, 表现汉字时使用的点阵数越高, 则汉字字形的质量也越好。当然, 每个汉字点阵的存储量也越大。通用型汉字字形点阵常分为三种, 它们的存储量可由下表看出。

GB2312 汉字的点阵数和存储量

	点 阵	存储量(字节)
简易型	16 × 16	261,696
普通型	24 × 24	588,816
	32 × 32	1,046,784
	48 × 48	2,355,264
精密型	64 × 64	4,187,136
	96 × 96	9M
	128 × 128	16M
	256 × 256	64M

由于高点阵汉字库要存储的信息量太大, 所以在一般的汉字系统中常常采用压缩技术来减少存储量。常见的压缩技术主要有三种:

- (1) 只收汉字字根的字根式压缩法;
- (2) 利用向量组字法产生的矢量字库;
- (3) 使用 N 元信息(霍夫曼树型)压缩法。

目前使用最多的是矢量法,它的压缩比大,字形也美观。但是使用汉字库时,必须要用特定的算法,通过软件生成汉字,所以输出速度较慢。

一般说,大尺寸的汉字要用高点阵的汉字输出,但是汉字库的存储量太大了,在一个汉字系统中,不可能具备各种点阵的汉字库。当我们需要输出大小不同尺寸的汉字时,总是采用汉字放大、缩小技术,使用同一个汉字库产生大小不同的汉字来。

当汉字放大倍数较高时,其中斜的比划会产生失真。这是因为汉字的比划是由点组成的,当点放大时,点间的间隙也放大了;于是就产生了台阶。好的汉字系统常采用“汉字平滑”技术加以弥补。

2. 输出设备

汉字输出设备的好坏对汉字输出的质量也有很大的影响。在一般打印输出时,高精度的激光打印机的效果最好,其次是喷墨打印机,最差的是针式打印机。但是打印机的价格和打印成本却正好相反,最便宜的是针式打印机,其次是喷墨打印机、激光打印机。

1.1.4 汉字的处理方式

1. 字符方式

在处理汉字时,直接用内码传送和处理的方式叫做字符方式。使用这种方式处理汉字时,计算机传送到输出缓存区的汉字信息总是汉字内码,直到系统最后需要显示或打印输出时,再由专门的程序从字库中提取相应的字模点阵输出。这种处理方式速度较快。

2. 图形方式

在显示或打印汉字时,直接把汉字的字模点阵信息传送到输出设备输出的方式叫做图形方式。这种方式处理汉字时,在计算机内存和输出缓存器中存放的都是汉字的图形点阵,无需转换就可以在输出设备上直接输出汉字。这种处理方式一般速度较慢。

3. 全角和半角

在计算机处理信息时常常有半角字符和全角字符之分。

所谓半角字符是指在存储和输出时占用一个标准字符位(一个字节)的字符。全部ASCII码表中的英文字母及符号都是半角字符。在计算机系统内,半角字符是作为基本代码处理的,所有的命令字符和参数都使用半角字符。

全角字符要占两个标准字符位,全部汉字和汉字国标码表内的图形符、英、日、俄文等都是全角字符。

1.2 汉字输入法概述

1.2.1 输入方法的分类

在我国,汉字输入方法的发展历史上,曾经出现过一字一键的大键盘和用偏旁部首组合的中键盘;虽然它们操作简单、不需学习就可使用,但是因为造价高,输入速度慢,目前

已经被淘汰。现在普遍使用的是各国通用的计算机小键盘,利用键盘上的字母、数字和符号键,按照编码规则输入汉字。

我们使用计算机输入汉字时,由于环境或目的不同,可能会有许多种方式,比如:

- (1) 看打(看着文稿输入)
- (2) 想打(想着中文内容输入)
- (3) 听打(听别人口授输入)
- (4) 说打(自己口授输入)

一般来说,好的汉字输入方法应该能适用于各种不同的输入方式。

到目前为止,键盘汉字输入方法已经有几百种之多,并且还不时有新的输入方法出现。但是,使用较普遍,影响面较大的主要也就是十几种。而众多的键盘汉字输入方案最常见的又可以归纳为三种类型。

1. 拼音方案(音码)

用汉语拼音作为汉字的编码,以输入拼音字母而实现汉字的输入。

这种方法对于学习过汉语拼音的人来说,一般不需要经过专门的训练就可以掌握。而且,它可以适用于各种不同的输入方式。但是对于不会拼音,或不会讲普通话的人来说,使用拼音方法输入汉字显然是很困难的。而且,用拼音方法输入汉字同音字多,需要选字。读不出音的生字也无法输入。

目前拼音输入法又可以分为三种方式:

1) 全拼音

如果需要完整地输入汉字拼音的每一个字母才能实现汉字的输入,这种方式叫做标准拼音方案,简称“全拼”方案。

2) 压缩拼音

规定用一至二个英文字母代替某些较长(多字符)的汉语拼音声韵母,从而压缩汉字全拼音的字符个数,实现汉字输入的方法,叫做压缩拼音方案,简称“简拼”方案。

3) 双拼音

规定任何汉字不论其全拼音字符的长短,统一用两个字符来替代,这种汉字输入方法叫做双拼音方案,简称“双拼”方案。

2. 字形方法(形码)

字形方法又分为字根法和笔划法,它们把一个汉字拆成若干偏旁、部首及字根,或者拆成若干种笔划,使偏旁、部首、字根或笔划与键盘上的键对应编码,按字形敲键输入汉字。

字形方法输入汉字重码率低、速度快,只要能看见字形就可以拆分输入。因此,字形方法是最受专业录入员欢迎的。但是,任何人都必须重新学习,并大量记忆编码规则和汉

字拆分原则。同时,形码很难适用于“想打”、“听打”和“说打”。

人们最常见的形码方案有:五笔字型码、表型码、郑码、太极码等。

3. 音形方法(结合码)

这是拼音和字形方法相结合的一种汉字编码方案。一般以音为主以形为辅,音形结合取长补短。

这类方法因为兼顾了音码和形码的优点,即降低了重码率,又不需要做大量的记忆,所以不但使用简单方便,而且输入速度快、效率高。它也适用于各种不同的输入方式。

最常见的音形码方案有:自然码、栗码等。

1.2.2 汉字键盘输入的发展过程

一般把汉字键盘输入技术的发展过程划分为三个阶段。不论是哪种汉字输入方案,都顺应了这三个阶段。

1. 单字输入

最初的汉字输入技术的特点是以字为基本输入单位。70年代末、80年代初提出的键盘汉字编码输入方案大多是属于这一阶段。由于汉字字形复杂,同音字又多,所以这时汉字输入效率的高低主要取决于汉字编码的优劣。

这期间汉字输入技术的功绩在于它使计算机处理中文成为可能,缺点在于汉字输入法和人们的日常使用习惯差距较大,无论那种输入法,要么难学,要么输入速度慢,至今无法提供一种既容易学又输入快的输入方法,应用面也窄得多。在使用计算机输入汉字的几种方式上,字输入法仅比较适用于零散文件的看打输入形式。

2. 词输入

为解决第一代汉字输入技术存在的“容易学的输入慢,输入快的不容易学”及重码太多等问题,出现了以词为基本输入单位的第二代技术,并在80年代后期开始流行。

由于词输入方式利用了汉语的语法关系、语义关系、常用词搭配关系等,所以汉字的输入效率已经由过去主要依赖编码设计转向依赖处理软件的设计。同样的编码在不同的软件支持下会有明显的差异。

词输入方式比字输入方式所需要的输入信息少,当然易学性和输入速度一般也都比字输入方式优越。但是,词输入方式不能保证完全准确,当词库中词汇量多时则词的重码也多,词汇量少时又容易产生输入差错,同时还要再用单字输入,这样使得输入速度降低。另外,词输入方式输入时,往往需要用户人工分词,给用户加重了负担。人工判别字或词的工作方式妨碍了输入操作的协调,降低了输入速度。

3. 语句输入

这是80年代末、90年代初开始出现的一种汉字输入方式。其特点是以短语或句子为基本输入单位。从理论上讲,语句输入方式是明显优于字词输入方式的。因为首先从

操作心理学上看,操作人员倾向于按有一定意义的短语或句子为单位进行短时记忆、持续输入。其次,语句输入方式比字词输入法需要更少的输入信息。

由于语句输入方式并不十分看重每个单字的特征,以及重码率的高低,又不需用户分词;所以用户可以把精力集中在输入的内容上,不会因为分词打断思路。是一种简单易学、输入快速的理想汉字输入方式。

语句输入方式的最主要缺点就在于,输入语句在识别时很难做到百分之百的正确,需要再修正。语句输入法的研究才刚刚开始,还有很长的道路要走。

第二章 自然码输入法基础

自然码汉字输入法(简称自然码)以字为基础、词为主导,利用句子、文章前后相关联系做智能处理;汉字编码简单易学,输入以双拼音为主,形义为辅,音形义结合、使用方便、顺应自然。这正是当前汉字输入发展的方向。它的主要优点如下:

(1) 简单易学、效率高。特别适合普通办公室人员、管理人员、干部、作家、记者及广大群众学习使用。

(2) 具有南方音输入功能,不论用户的发音是否准确,使用起来都会得心应手。

(3) 在输入过程中,可以随机实时造词、自动存词,这是自然码特有的功能。

(4) 智能字词相关技术使系统自动记忆字词间搭配关系,动态调整重码,使重码率大大降低,从而使汉字输入的准确率和速度明显提高。

(5) 有独特的全模糊检字检词、叠字叠词、中文标点、表格功能。

(6) 具有很强的适应性,可以自由挂接到各类机型和各类汉字操作系统下使用。

目前,最新版本的自然码汉字输入系统是 5.6 版本。其中含有 5 万国标词汇和 1 万多其它词汇。

2.1 自然码输入法的启动与退出

自然码输入法总要挂接在汉字系统下使用。挂接的方式有两种:一种是已经含有自然码输入法的汉字系统,只要启动汉字系统后,自然码输入法就已经挂接好了。我们把这类系统叫做内部挂接系统。另一种是不含自然码输入法的汉字系统,应当在启动汉字系统后,再单独启动并挂接自然码输入法。我们把这类系统叫做外部挂接系统。

2.1.1 在“自然码中文集成环境”中启动

有许多厂家已经把自然码输入法做在了自己的汉字系统中,启动汉字系统后,就可以直接使用自然码汉字输入法。其中做的最完善、并能够及时升级的系统当然还是自然码发明人周志农所在的北京长城商务促销公司产品——“自然码中文集成环境”。

“自然码中文集成环境”含有一个中文环境菜单、一个汉字操作系统和汉字库,我们可以先启动中文菜单,然后在菜单上完成所有操作;也可以直接进入汉字操作系统使用。下面我们来看看怎样从这套汉字系统进入自然码输入法。

1. 启动汉字系统

在硬盘中安装了“自然码中文环境”系统软件后,可以用以下命令启动汉字系统:

ZRHZ <回车>

在屏幕上首先显示如图 2-1 内容。

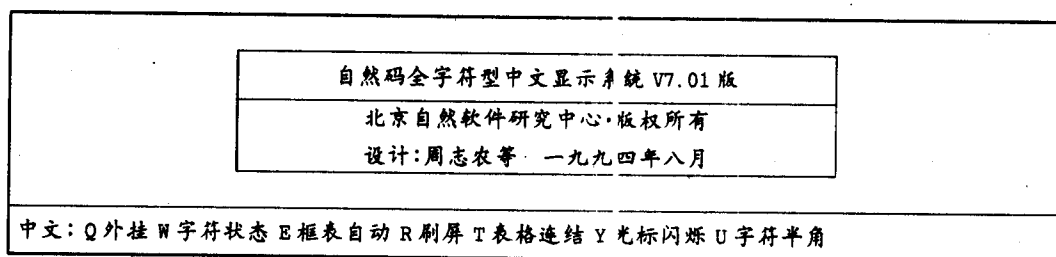


图 2-1

显示几秒钟以后,清屏并重新显示如图 2-2。

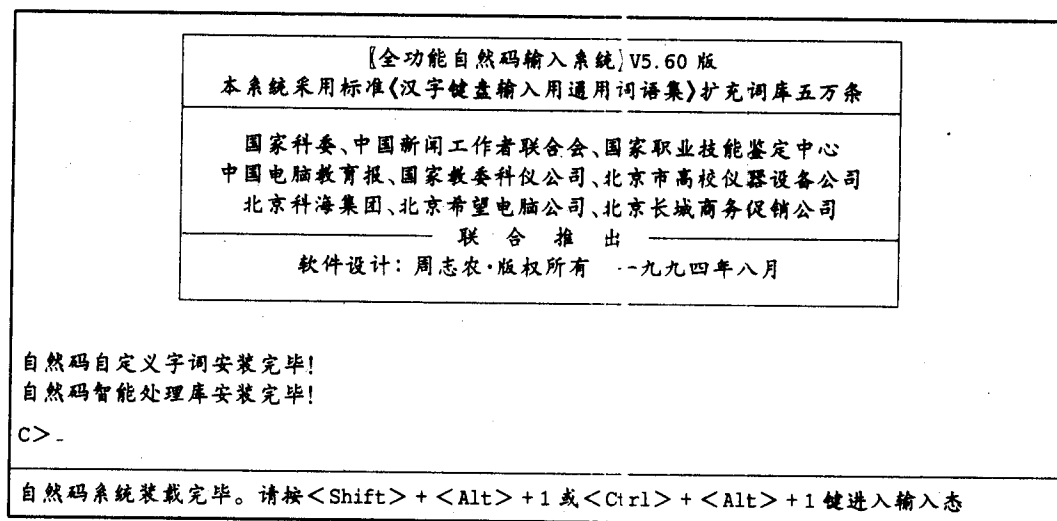


图 2-2

2. 启动自然码输入法

要进入自然码输入状态,应按照屏幕提示敲组合键【Shift + Alt + 1】或【Ctrl + Alt + 1】。系统进入自然码汉字输入状态后,屏幕提示行显示如图 2-3。

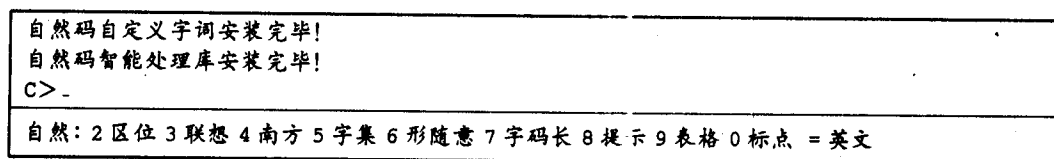


图 2-3

在自然码汉字输入状态下,只要使键盘处于小写状态,就可以用拼音输入汉字了。从提示行还可以看出,用命令键【Ctrl + Alt + 1~0】或【Shift + Alt + 1~0】可以做不同的功能选择。

2.1.2 在一般汉字系统下挂接启动

如果所选汉字系统中不含自然码输入法,就必须使用自然码输入系统挂接到汉字系

统下操作。假定我们已经把自然码输入系统装入硬盘子目录 ZRM 中,启动汉字系统以后,再敲命令:

ZRM <回车>

屏幕将显示如图 2-4 的信息。

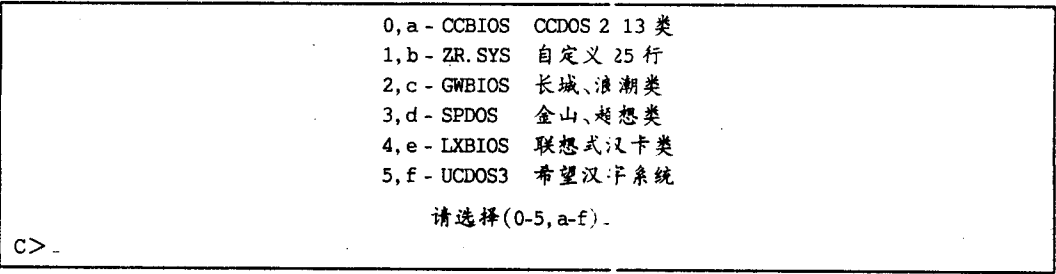


图 2-4

根据自己使用的汉字系统选择数字 0-5 或字母 a-f 以后,自然码汉字输入系统将挂接到当前汉字系统上;并且进入自然码输入状态,屏幕显示如图 2-5。

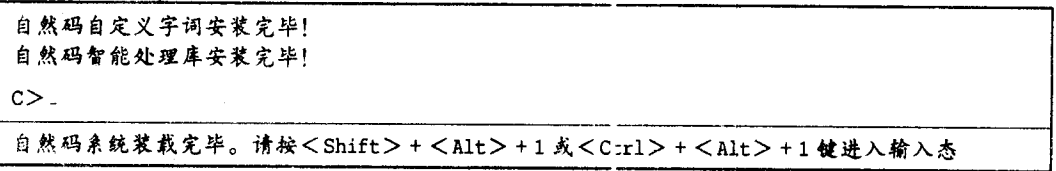


图 2-5

当用户第一次选择参数挂接好自然码系统后,每次启动本汉字系统不需要重新选择参数就会自动挂接好自然码输入法。若想修改已选参数时,可在 ZRM 子目录下使用挂接命令 ZRM ␣(␣表示空格)或将 ZR.INI 文件删除。

2.1.3 自然码输入法的退出

1. 暂时退出

在需要输入英文时,可以在自然码输入状态下,按下【Shift】敲英文字母,则可输入大写字母;若用【Caps Lock】键使键盘处于大写方式,则可以任意输入大写英文字母;此时若再按下键【Shift】,敲英文字母,则可输入小写字母。

2. 一般退出

如果敲命令键【Ctrl + Alt + =】或【Shift + Alt + =】则退出汉字输入状态,进入英文输入状态,如图 2-6。

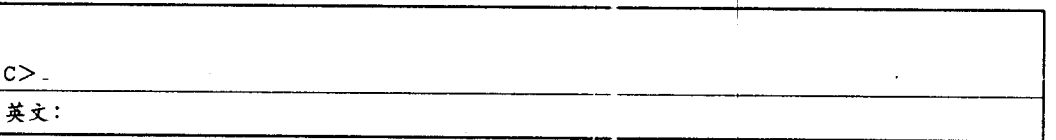


图 2-6

此时,可以任意输入大小写英文字母。

3. 用命令键退出驻留

如果连续三次按命令键[Ctrl + Alt + =]或[Shift + Alt + =],则可以关闭自然码输入法,并把自然码输入系统占用的内存空间彻底释放出来。屏幕将显示如图 2-7。

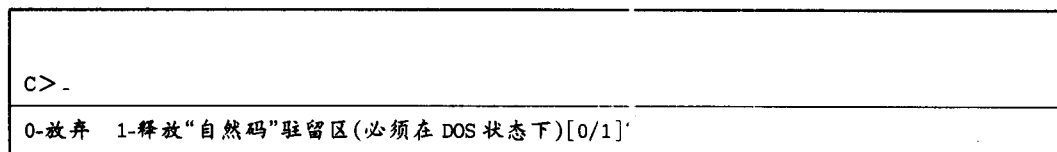
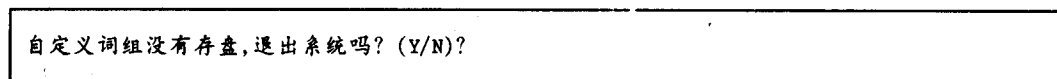


图 2-7

如果此时敲数字键“1”后,系统在退出内存驻留前有可能会提示如下:



这说明了在此之前用户定义了自造词组,而且:

- (1) 正在编辑状态,并且当前编辑文件还没有存盘;
- (2) 不在 DOS 命令状态;
- (3) 词组库文件 ZR.CZ 出错;
- (4) ZR.SYS 参数文件的定义不符。

这时应当先敲字母 N 放弃退出,关闭当前编辑文件,返回 DOS 命令状态,然后再试验一下;如果还出现以上提示,就应该考虑修改自然码参数文件(ZR.SYS)。这时用编辑软件打开文件 ZR.SYS,可以看到参数文件中指定了词组库文件的路径(参考附录 D),而这一路径很可能与实际不符,改正后存盘退出。

再连续三次按命令键[Ctrl + Alt + =]或[Shift + Alt + =],应当可以正常退出自然码输入法系统。否则就肯定是词组库出了问题,应当把原系统软盘上的 ZR.CZ 文件覆盖到硬盘上。虽然词组库文件破坏的机会不多,但是为了尽可能减少损失,用户应当经常用软盘备份该文件。

4. 用命令退出驻留

可以在 DOS 状态或批命令文件中用命令退出自然码输入法系统的驻留。

命令格式:ZRMOUT

注意:应当指名命令文件 ZRMOUT.COM 存放的路径。

该命令的作用与效果和键盘命令[Ctrl + Alt + =]或[Shift + Alt + =]的效果是完全一样的。但是它可以放在程序中调用,所以更加灵活方便。

2.2 单字的输入

虽然单字输入法不是自然码系统的主要输入方法,但却是各种输入方法的基础。我

们弄明白了单字的输入方法后,学习词组输入法是很容易的。

2.2.1 简码字输入法

为了方便用户,自然码系统把最常用的一些汉字分别用键盘上的 26 个英文字母相对应。这 26 个简码字可以直接敲它们对应的字母键加空格键(也可以是数字“1”)选出。它们之间的对应关系可以从图 2-8 中看出。

Q 七	W 我	E 二	R 人	T 他	Y 一	U 是	I 出	O 欧	P 片
A 按	S 三	D 的	F 非	G 个	H 和	J 及	K 可	L 了	
Z 在	X 小	C 次	V 着	B 不	N 你	M 没			

图 2-8

比如要输入“的”字,敲字母“d”后,显示如图 2-9。

C>_
自然: d 1 的 2 第; 3 党' 4 党 5 但 6 是 7 队 8 员 9 等 0 于

图 2-9

再敲数字 1 或空格键可以选择输入“的”字,显示如图 2-10。

C>的_
自然: d 1 的 2 第; 3 党' 4 党 5 但 6 是 7 队 8 员 9 等 0 于

图 2-10

在以上提示行中,第一个字是一级简码字,最好用空格键输入,这样可以有效地提高输入的速度。在一级简码字后面的字为次级简码字,可以敲“;”或数字“2”键输入。余下的一些常用字和常用词,可用可不用。

常用简码字有三种:

- (1) 一级简码字,输入方法:“声母 + 空格键”
- (2) 次级简码字,输入方法:“声母 + ;”
- (3) 附加简码字,输入方法:“声母 + '”

我们看上面的提示行,第二个字“第”和第三个字“党”的后面,分别标有“;”号和“'”号。这是它们输入符提示,虽然可以用对应的数字键选字,但为了提高输入速度,最好用“;”号和“'”号选择它们。

用简码也可以输入一些常用词组。比如图 2-9 敲字母“d”键后,在提示行中;我们选

敲数字键“6”，出来的是“但是”这个词，显示如图 2-11。

C>但是_
自然: d 1 的 2 第; 3 党' 4 党 5 但 6 是 7 队 8 员 9 等 0 千

图 2-11

选敲数字键“8”，则输入词组“队员”，选敲数字键“0”，则输入词组“等于”。一般说，在简码字后面用偶数数字选择时系统给出简码词。

如果在第一页提示中找不到所要的字，还可以用“[”键和“]”键做前后翻页查找，最后用相应的数字键选择输入。

2.2.2 标准码输入法

自然码输入法采用的标准外码是双码拼音方案，其中每个汉字都固定用两个字符表示。它是在 CCDOS 压缩拼音的基础上设计的，除了把汉字声母“zh”改放在“V”键上，并且又压缩了一些长的汉字韵母，使得每个由多字符组成的汉字韵母都能定义在一个相应的字母键上外，基本保持了 CCDOS 压缩拼音的原有设计。于是所有的汉字拼音都可以用声韵两个键输入，这样虽然比 CCDOS 压缩拼音要多记忆一些码，但是编码统一、规矩，输入和处理的效率较高。

1. 编码规则

自然码拼音输入法的编码规则如图 2-12。在本书最后附有此图放大样插页，读者可以剪下，来贴在机器边随用随看。

Q iu	W ua ia	E e	R uan üan	T üe	Y uai ing	U sh I ch u i	O o uo	P un ün	← 声母 ← 韵母
A a	S iong ong	D uang iang	F en	G eng ng	H ang	I ai	K ao	L ai	
Z ei	X ie	C iao	V zh ui ü	B ou	N in	M ian			

图 2-12

从上图可以看出，其中与原 CCDOS 压缩编码相同的地方有：

压缩码	u	i	s	f	g	h	j	k	l	y	v
拼音	sh	ch	ong	en	eng	ang	an	ao	ai	ing	ü

需要重新记忆的编码有：

压缩码	v	q	w	r	t	y	o	p	s	d	g	z	x	c	b	n	m
拼音	zh ui	iu	ua ia	uan üan	ve	uai	uo	ur ür	longuang iang	ng	ei	ie	iao	ou	in	ian	

例如：

汉 字	王	之	试	桥	窗	软	平	冲	浪
全拼音	wang	zhi	shi	qiao	chuang	ruan	ping	chong	lang
双拼音	wh	vi	ui	qc	id	rr	py	is	lg

自然码在三个特殊的双字母的声母上,为了便于记忆,安排如下:

- (1) 声母 zh(读“枝”音)安排在字母键 V 上,其形状与“树枝”相像;
- (2) 声母 ch(读“尺”音)安排在字母键 I 上,其形状像“尺子”;
- (3) 声母 sh(读“试”音)安排在字母键 U 上,其形状很像“试管”。

此外,自然码在韵母键位的设计上,除考虑了合理分布的原则以外,特别让它与英文字母相一致,以便于记忆。其关系如下:

- (1) 拼音 uai(读“外”音)安排在字母键 Y 上;
- (2) 拼音 ui(读“微”音)安排在字母键 V 上;
- (3) 拼音 iu(读“优”音)安排在字母键 Q (近似于读 keiu)上;
- (4) 拼音 ei(读“诶”音)安排在字母键 Z (近似于读 zei)上;
- (5) 拼音 ua(读“瓦”音)安排在与电工符号“瓦特”的英文缩写“W”字母上。

如果用户只认识汉语注音字母,自然码输入法系统还提供了一套注音字母编码,其对照码表如图 2-13。

Q< X	WX XY Y	E< 儿	RE X马 马	T女 世	Y X男 Z	UP X	I个 	OZ X乙	P女 X乙 马
AY	SZ Z XZ	D力 X九 九	FC 乙	G< Z	H厂 力	J力 男	K乙 么	L力 男	
	ZP 乙	XT 世	C乙 么	V世 X乙 	B乙 X	N乙 乙	M乙 马		

图 2-13

除汉语注音字母对照表外,自然码输入法系统还提供了一套用同音汉字注示的编码表,其码表如图 2-14。

Q 七 优	W 污 挖 呀	E 恶 二	R 日 弯 冤	T 物 约	Y 衣 外 英	L 诗 污	I 吃 衣	O 喔 握	P 坡 温 晕
A 啊	S 思 拥 翁	D 德 汪 央	F 佛 思	G 哥 (亨)	H 喝 昂	J 机 安	K 科 凹	L 勒 哀	
	Z 资 族	X 西 椰	C 词 腰	V 支 微 余	B 播 欧	N 呐 音	M 模 烟		

图 2-14

2. 同音字的选择

用拼音法输入单字,常会出现许多同音字,需要人工选择。比如要输入“工”字,当我们敲入字符“gs”以后,屏幕状态行显示内容见图 2-15。

C>_	
自然: gs 1 共 c 2 工 a 3 公 s 4 供 r 5 功 l 6 攻 w 7 巩 g 8 贡 b 9 宫 b 0 弓 q	公司'

图 2-15

可以敲 2 选择“工”。如果要输入同音字“躬”,因为当前页上没有,则可敲“[”键向后翻页,屏幕显示内容见图 2-16。

C>_	
自然: gs 1 躬 u 2 恭 x 3 汞; 4 拱 f 5 冀 l 6 蚣 i 7 珙 w 8 宏 o 9 觥 j	

图 2-16

敲数字键 1,或空格键选择。

3. 多音字的输入

在汉语中有很多字是有多种读音的,还有一些字又很容易读错。为方便用户,自然码拼音法中已经尽可能多地考虑到了这些问题;在输入时不必像 CCDOS 拼音那样必须使用唯一的拼法。

比如“重”字可以读“chong”,也可以读“zhong”。如果我们按“chong”音敲字母“is”后,屏幕显示内容见图 2-17。

C>重_	
自然: is 1 充. 2 冲. 3 虫 k 4 宠 b 5 崇 e 6 重/ 7 忡 x 8 怱 x 9 茺 c 0 觥 j	出色'

图 2-17

敲数字键 6 来选择。如果我们按“zhong”音敲字母“vs”后, 屏幕显示如图 2-18。

C>重
自然: vs 1 中, 2 种 h 3 重 / 4 众 r 5 终 s 6 钟 j 7 忠 x 8 肘 o 9 仲 r 0 种 v

图 2-18

用数字键 2 选择。

类似的常用多音字还有: 长 = ih, vh 率 = lv, uy 朴 = pu, pc 等。

还有些字是容易读混的, 比如“波”字一般读做“bo”, 但是如果读做“po”时, 也可以找到。

如果读做“bo”, 敲字母“bo”后, 屏幕显示内容见图 2-19。

C>_
自然: bo 1 波; 2 玻 w 3 播 f 4 拨 f 5 博 t 6 勃 l 7 剥 d 8 膊 o 9 搏 f 0 牌 o

图 2-19

如果读做“po”, 敲字母“po”后, 屏幕显示内容见图 2-20, 选 7 即可。

C>_
自然: po 1 迫 z 2 坡 t 3 婆 n 4 破. 5 魄 b 6 泼; 7 波; 8 泊; 9 颇 a 0 粕 n

图 2-20

还有“械”字应当读做“xie”, 但是如果读做“jie”也可以找到。常见的这类字还有一些: 塑 = su so 束 = uu su 约 = yt yk。

2.3 词的输入

单字输入虽然是基础, 但输入速度实在太慢。词输入法是自然码系统的基本输入方式, 也是目前最有效的快速汉字输入方式。

2.3.1 简码词输入法

简码词是指可用两个声母作词组代码的极常用的双字词组。简码词通常总是显示在提示行的最末端。如没有其它字或词的话, 也同时显示在第一位。当在第一位时, 也可用空格选取, 否则必须用“'”选取。

输入方式: 声 + 声 + '

比如要输入“可以”一词, 应当敲“可”字的声母“k”和“以”字的声母“y”, 屏幕显示内容见图 2-21。