

国际数据集团
IDG
INTERNATIONAL DATA GROUP

爱奇

信息丛书—7

汉字输入技术实用手册



国际数据集团
IDG
INTERNATIONAL DATA GROUP

爱奇

高技术(北京)有限公司
IDG HIGH TECH COMPANY LTD - BEIJING

316

5

汉字输入技术实用手册

何 平 边长海 肖调尧 编
李文良 刘保元 张景生

为您的成功提供全面服务

君有奇志 必爱信息

IDG 爱奇高技术(北京)有限公司是经北京市人民政府于1989年12月正式批准并登记注册的高科技信息开发服务公司。本公司是美国IDG国际数据集团成员。该集团创建于1964年,总部设在美国波士顿,是电子信息技术领域内在市场研究、技术预测、软件开发和情报服务等方面均享有世界信誉的软科学公司。其服务对象是以电子信息技术为中心的高新技术领域。其业务经营范围包括:市场研究咨询、技术发展预测、资料报刊出版、科技交流展览、专业培训教育及软件开发推广,在44个国家和地区拥有广泛的国际联系,形成一个快速而全面的全球信息网。

本公司是中外科技交流的理想桥梁和信息交换的最佳通道。本公司下设研究部、软件部、交流部、编辑部、咨询部和服务中心,愿为贵单位提供全面的技经贸信息咨询和服务:

1. 为企业开拓国际市场扩大出口提供技贸信息服务;
2. 承接高新技术开发研究和应用软件工程施工业务;
3. 为引进技术和资金、兴办三资企业、寻找国外合作伙伴,提供咨询服务;
4. 接受委托、承办国内外市场调查及技术预测的调研课题;
5. 编辑和印刷产品说明书、公司刊物、技术手册、培训教材、维修指南等资料;
6. 翻译和印刷出口产品的技术宣传资料并承办引进项目和来华展示的翻译业务;
7. 组织和承办产品发布会、技术报告会、交流展示会、用户培训班等并收集反馈信息;
8. 在《新产品世界》月刊和有关集团刊物上提供专栏报导;
9. 承办技术资料、软件及有关产品的代销、邮购和国外样品的代购业务;
10. 承办组织对口企业出国参加展览、贸易洽谈及管理培训等服务。

我们将为您的事业成功提供尽善尽美的服务,欢迎来函来电联系,谢谢。

IDG 爱奇高技术公司

地址:北京市复兴路15号821室(彩电中心西侧)

邮政编码:100038 传真:801.4073 电挂:3725

电话:8015544-2821;2822 电传:20079 ISTIC CN

快速而全面的 IDG 全球信息网

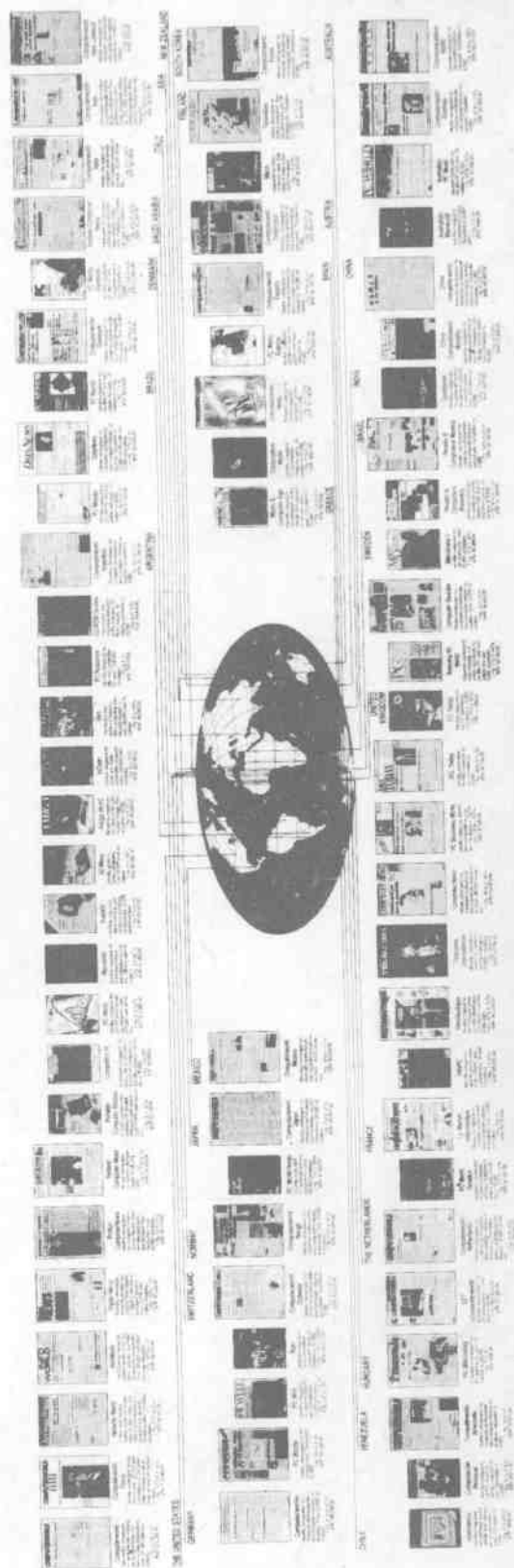
美国 IDG 国际数据集团公司 (International Data Group) 是电子信息技术领域内在市场研究、技术预测和情报服务等方面均享有世界声誉的软科学公司。集团公司创建于 1964 年,总部设在美国波士顿。其业务经营范围包括:市场研究咨询、技术发展预测、资料报刊出版、科技交流展示、专业培训教育及软件开发推广等。

25 年来,集团公司已发展成为全球性的跨国公司,在全世界 44 个国家和地区建有子公司和分公司,拥有 3800 名高水平的研究专家和编辑人员,采用电子邮政网、数据库、电传及传真通讯以及在线服务等现代化信息处理和传递手段,建立了快速而全面的世界性信息网络。

每年发表数百篇市场研究报告和技术发展预测报告,用 18 种语言出版 140 余种高科技报刊杂志,举办各种国际性和地区性的高新技术学术报告会、市场分析会和产品展示会,提供各种定题和定向的咨询服务。

IDG 集团现在包括的国家和地区为:

美国、加拿大、阿根廷、巴西、智利、墨西哥、委内瑞拉、奥地利、丹麦、芬兰、法国、西德、希腊、意大利、荷兰、挪威、西班牙、瑞典、英国、匈牙利、苏联、捷克、波兰、保加利亚、日本、中国、沙特阿拉伯、澳大利亚、印度、南朝鲜、新西兰、新加坡及香港和台湾省。



目 录

§ 1 键盘输入	(1)
1. 1 直接输入方式	(1)
1. 2 间接输入方式	(1)
§ 2 汉字识别自动输入	(2)
§ 3 语音输入	(2)
§ 4 汉字编码方案的评估和选择方案的要点	(3)
§ 5 标准汉字字符集, 标准编码及字体、字形、字号规范	(5)
§ 6 实用汉字输入方法	(6)
6. 1 五笔字型输入法	(6)
6. 2 声数编码输入法	(16)
6. 3 汉语拼音输入法	(19)
6. 4 区位码输入法	(20)
6. 5 电报码输入法	(20)
6. 6 国标码输入法	(20)
6. 7 首尾码输入法	(21)
6. 8 汉字联想系统与拼音联想输入法	(21)
6. 9 自然码输入法	(25)
.....	(33)
.....	(39)
.....	(43)
.....	(43)
.....	(57)
7. 3 容易写错的字	(62)
7. 4 校对时容易混的字	(70)
7. 5 容易用错的词	(73)
7. 6 校对时容易出错的词	(77)
7. 7 校对时可以不改动的词	(79)
附录 I: 区位码、国标码、机内码、五笔字型等编码索引	(81)
附录 II: 全国计算机中文信息处理展览会汉字 (语) 输入技术信息	(130)

纵观汉字输入技术,大致可分为:键盘输入、识别输入、语音输入这三大类。

§1 键盘输入

键盘输入是目前汉字输入方式中技术最成熟、应用最普及、研制成果最多、使用又是最早的技术。据统计仅编码方案就多达七百多种,形式各异,都体现了中国方块汉字音、形、义的特征。键盘输入可分为直接输入方式、间接输入方式和人机对话输入方式。

1.1 直接输入方式

是在字盘上选择所需的汉字或点触字盘直接输入汉字,其典型代表是汉字整字键盘和各种笔触式字盘。汉字整字键盘是使用最早的一种,它分为全键式和多段移位式两种。全键式整字键盘是采用一字一键的方式,键盘上的每个键都与特定的汉字相对应。它的基本操作就是选字和击键,击一次键输入一个汉字。多段移位式汉字键盘类似于字母、数字合用键,用移位键来区分字母和数字,即在一个键位上定义多个汉字,用相应的移位键来切换键位上的各个汉字。这两种键盘一般盘面收容字量在 2000 以上,特点是直观便于操作,经训练后输入速度可达每分钟 100 字左右,但设备较大,价格较贵,输入速度较慢,难记忆,属早期技术。

各种笔触式,是近年发展较快的输入方式,它和图形输入方式很相似,一般由坐标盘、字盘、接触笔、控制器等部分组成。字盘上印有按矩阵排列的汉字,将字盘复盖在坐标盘上,使字盘上的汉字与坐标盘上的 X、Y 扫描线的交叉点一一对应。这些交叉点就是文字的位置检测点,它也是一种目测检字输入装置。其特点:操作直观、简便,一般操作员较容易接受。但是只能单手用输入笔逐字进行输入,而且背记字盘,所以速度不会高,熟练的操作人员也只能达到每分钟输入 60 个字左右。因此,输入速度慢,字盘物理尺寸限制了收容的字数,盘外输入问题突出,影响其发展与应用。

1.2 间接输入方式

采用国际上标准的计算机键盘,提取汉字的某些特征信息,经汉字编码来实现汉字输入,它是目前国内最普及的汉字输入方式。这种方式输入汉字主要取决于编码方案。汉字编码设计,是从分析汉字结构或发音入手,抽取一些共同的基本特征,并用适当的字符或数字去表示这些特征,从而达到对汉字进行编码的目的。这个过程包括选取一定数量的代表元(即字元、字根、字素、部件或构件等),确立规则,进行键盘设计、程序设计等环节。目前社会上较流行的有五笔字型、层次四角、前三末一、五十字元、二维三码和拼音输入法等编码方案,可归结为拼音输入和拼形输入。以拼音方式输入汉字的编码称拼音码,以拼形方式输入汉字的编码叫拼形码。

(1) 拼音输入法的研究起步较早,应用也最为普遍,凡乎在所有能处理汉字的机器上(计算机、四通、科理等文字处理机),都装有拼音输入法。利用汉字的字音属性来对汉字进行编码,其实质是把汉字的拼音作为汉字代码来使用。汉字音节是声母和韵母两部分构成的,由于声母、韵母、声调的总数量约 60 个,可把它们直接配置在字母数字键的键位上,操作人员只需知道汉字的拼音,便可按规定的顺序击键,从而完成汉字的输入。缺点是:重码多,输入速度慢,方言使拼音输入法受到一定限制。

由于汉语的特点是小音节集,同音字多,采用拼音对实现“字处理”应该说是低效的。但随着软件技术及人工智能技术的发展,拼音输入已经从“字处理”进入“词处理”,大大提

高了输入效率。

(2) 拼形输入是利用汉字的字形信息给汉字编码,一般分为笔形编码法和字根编码法。笔形编码的原则就是把汉字分解为笔型,在不同的笔型编码法中对“笔型”的归类是不同的,因此有“五笔型编码法”、也有“八笔型编码法”。其次是不同的笔型编码法中对“笔型”的取码顺序不同,有按汉字的书写顺序取笔型的,也有按笔型的高低左右位置不同取码。字根编码的基本思想是:把汉字看作一个拓扑图形,这个图形是由若干单元(字根)按照一定的结构形式组成,根据构成汉字字根的信息特征和字型的结构形式确定汉字的编码。对于上述两类拼形输入方案,从使用角度看:笔形编码原则简单,易于掌握,但为区分重码,对汉字的取码较长,输入效率较低;字根编码规则相对笔形编码法较为复杂,需要使用人员经过一定的学习过程才能掌握,但对汉字的取码短,输入效率较高。笔形输入法解决了前期易学易掌握的问题,未解决如何提高后期输入效率的问题。字根输入法则解决了后期效率的问题,但未解决前期易学易掌握的问题。还有角形代码、音与形相结合的代码和自然码(音、形、义结合的)等编码方案,各有长短。

§2 汉字识别自动输入

汉字识别自动输入,就是用计算机自动识别印刷在纸上或人写在纸上的文字,使汉字字形自动高速地转换成计算机内部的编码,完成信息处理。

汉字识别包括:1、联机手写汉字识别,即用笔在图形输入板上写字,人一面写,机器一面认。它是汉字识别中最简单的一种。2、单体印刷体汉字识别。3、多体(宋、仿宋、楷、黑体)印刷体汉字识别。4、手写印刷体汉字识别,即识别人写在纸上的符合若干限定的规整汉字。5、特定人手写汉字识别,是手写印刷体汉字识别的一个特例。除第一种联机识别外,其它都属脱机识别。主机一般采用微机,输入装置多为光电扫描式。

汉字识别的基本原理:光电扫描纸上的文字,产生模拟电信号,以每毫米若干点取样,经模数转换成带灰度值的数字信号,再进行行、字切化、二值化、细化或抽取轮廓、平滑、规范化等。预处理后,汉字图形形成规范的二值数字点阵,如“1”反映了笔划部分,“0”是文字空白背景。对输入的汉字点阵抽取特征,和存储在机器字典中已知的标准特征用一定准则进行图式比较,该特征对应的标准特征字就是识别结果。

我国汉字识别技术虽然从70年代末展开,但近几年的进展使这一技术摆脱了对国外技术的模仿,在印刷体和联机手写体方面已达到世界先进水平。目前,我国已鉴定的印刷体汉字识别软件和系统有几十家,输入设备主要采用传真机、专用CCD、摄像机、图文扫描器等,识别率高达99.9%,识别速度在286微机上达到9~20字/秒。手写汉字识别存在问题较多,离普及使用有较大距离,仍处于方案探讨和规模实验、小范围使用阶段。

§3 语音输入

语音输入就是人用口述的方式将信息送给计算机。

通过人口述方式所表达出来的话音语言,计算机根据发音传递的声学信号波形提取与字或词有关的参数来辨认所说的字或词,然后把这些字和词连接起来进行分析理解,这个过程称为语音识别。

语音识别技术研究在世界上十分活跃,美、日研制出了适合于语音识别用的专用信号处理片,如日本 NEC 公司的 PD7720,美国 TI 公司的 TWS230 系列等。

我国语音识别研制工作发展迅速。如 90 年元月 22 日北京市四通技术开发中心研制的四通 252 型汉字语音系统通过了北京市民政局主持的鉴定,该系统由一支定向话筒、一个高保真音箱、一块每秒能运算一千万次乘法的专用语音数据处理卡和一套软件构成。系统安装在普通微机上,就可以用语音来输入国标一、二级汉字库的全部汉字。它的四声识别正确率为 100%;一次识别正确率为 89.1%;带 5 个候选音节识别正确率为 99.5%;识别速度为每分钟 30~60 个汉字。语音识别技术赋予计算机以新的使命,但该技术还有待于探索改进,达到真正普及实用还需要艰苦努力,不断优化。

§ 4 汉字编码方案的评估和选择方案的要点

汉字编码方案象雨后春笋一样层出不穷,我们如何选择方案,尤其是用哪些方法选择出易学、易记、易用的方案,是大家所关心的。

4.1 汉字编码方案的评估

评测编码输入方案国内已进行过多次。通过评估可以促进编码方案的优化,加速择优汰劣,使用户在选用方面掌握主动权,减少人、财、物力的重复浪费。评估参考要点:

(1) 平均输入字数; (2) 平均击键次数; (3) 误字率; (4) 错字率; (5) 每字击键均值; (6) 组字重复率; (7) 词语速率; (8) 编辑操作; (9) 操作错码率; (10) 记忆强度; (11) 键盘设计合理否; (12) 操作者的素质。人的因素不可忽视,对前期和后期训练效果应呈上升趋势。

4.2 汉字编码输入方案选择要点

- (1) 编码规则要简易明确、易学易记,便于熟练掌握。
- (2) 编码容量要大。至少应复盖 GB2312-80 基本集中规定的全部汉字 (6763 个),并易于扩充。在必要的情况下,应能复盖辅助汉字集中所包括的汉字。
- (3) 编码要有唯一性,重码率要低,并且有简便的区分重码的方法。
- (4) 避免编码方法的二义性,同一个汉字只能产生一个确定的编码。
- (5) 尽可能使用标准字符键盘作为汉字输入的设备,键位的分布要合理。
- (6) 平均每个汉字的输入击键次数要少,尽量缩短编码长度,以提高编码效率。
- (7) 不同的汉字编码若是等长码,有利于增加输入操作的规律性。
- (8) 输入速度要高。

要认真分析其客观因素,如键位分布、码长、重码率以及代表元、规则等,因为这些因素决定着编码方案的质量,影响着方案的易学程度和输入速度。

汉字编码输入方案,是目前最通用的,普及率也最高。一般文字处理机都装有两种以上方案,选用哪一种方案输入汉字,使用者可根据自己的实际情况择优选用。请注意了解如下几点:

(1) 代表元 (即字元、部件等)

人们总是希望以最少的代表元去完全描述全部汉字,这是不太容易办到的。一般代表元数量受到多方面因素的影响,如码长、重码率等。在层次四角方案中,减少一个代表元,其它条件不变,最好的情况也得增加四个重码。因此它是衡量方案的重要指标。五笔

字型”有代表元约 130 个,“层次四角”103 个,“前三末一”只有 80 个,而“五十字元”仅有 50 个,作为我们使用者希望每个代表元在键盘输入中出现的概率尽可能地平衡。如果代表元出现概率过多,会增加重码的机会;概率太小,则需要更多的代表元。

确定了代表元,应优先考虑选取我们熟悉的部首或部件,即字的偏旁部首、笔划等。这一点有利于大家学习和掌握汉字编码输入方案。

(2) 拆字规则

拆字一般应尽量简单、直观、使我们能容易地理解和接受。如五笔字型、前三末一、二维三码等汉字方案中的拆字规则,主要是根据汉字的书写顺序来确定的;而层次四角、五十字元等方案,主要是根据汉字的方块结构来划分的。

(3) 键盘设计

键盘设计是关键,不但要考虑键位分布,还需考虑码长、重码率、记忆量等问题。

键盘分布,尽可能地只用 26 个英文字母来进行编码,这样可以缩小手指的移动范围和缩短运动距离,提高击键速度,减轻劳动强度。对 10 个指头键位分配应注意其灵活性、与高频键的配置、各自承受的负担等。

码长:就是选用多少键位(即字母、数字键)完成所有汉字的编码。码长对编码的影响,主要表现在速度上,这对专业操作员很重要。

重码率是由于汉字的复杂性带来拆分时的多义性所致。重码有两种,一种是同形字的重码,如层次四角中的“本、末”等;另一种是键盘设计上的重码,如五笔字型中的“器、露”编码都是“KKDK”。重码就带来一些字必须强行记忆,影响输入速度和输入的准确性。

(4) 记忆量

记忆包括键位、拆字法、重字、特殊字、代表元等,这些是影响汉字编码输入应用的“关键”,有些方案将需要死记硬背的部分转化为理解性记忆和联想记忆。记忆量与速度、正确性有很大关系。

(5) 疲劳性

对专业录入人员来讲关心的是输入速度和疲劳性。对疲劳性来说指:眼、脑、手三者,如果键盘设计的合理,手指运动范围小,各指击键的频率符合各自的能力,要领和姿势正确,不会影响眼、脑、手的协调配合,劳动强度不高,疲劳不会提前出现。

如果用户要求方案易学、且输入速度不高时,请注意以下几点即可:

(1) 所需记忆强度小。(2) 与人的识字心理过程相符合。(3) 初学期的学习曲线的斜率要小。(4) 能否靠自学说明书掌握方案。(5) 达到规定指标的时间是否短。(6) 键入过程中的眼、手、脑等的疲劳度是否低。(7) 重码应为最低。

另外介绍一下二维三码和自然码。

二维三码是以笔型结构作为编码的基本部件,并按二维坐标的方式与键位对应,只用 27 个基本键位就将全部字码的最大码长压缩到 3 键;且动态重码率低于千分之一,其规则简单易学,用户不需背字根,也不用拆根,只要掌握横向规则和纵向规则就可进行输入。一般经两天培训后即可进入盲打。

自然码是以音为主,形义为辅,声代形义并具有高度灵活性的开放式编码方案。它分为表音(声部及韵部)表意(义部)和表形(形部和尾部)三个部分。输入一般技术类文

章时,若不使用三字以上词组,平均码长大约为 2.3 键/汉字(包括所有键),利用多字词和定义性词组,平均码长可为 1.6 键/汉字,熟练操作员一般每分钟可输入约 200 个汉字。自然码充分利用了汉字的表音、表义以及字形结构上的特点,并着重考虑汉字编码方案中的易学、高效、反应效果等因素。

§ 5 标准汉字字符集、标准编码及字体、字形、字号规范

5.1 汉字字符集和标准编码

(1) 国家标准《信息交换用汉字编码字符集—基本集》,即 GB2312-80,其中收集了使用频度较高的 6763 个汉字,分为两集,第一集包括 3755 字;第二集包括 3008 字。一般情况下使用时,基本集中所包括的汉字能满足 99.99% 的复盖率。但是,当涉及某些行业和应用时,例如文史、考古,特殊的科技,公安户籍等领域,基本集所收集的汉字已明显地不能满足需要,必须使用辅助汉字集。

(2) 一九八六年,我国已制订出两个辅助汉字集,其名称为:《信息交换用汉字编码字符集 第二辅助集》,《信息交换用汉字编码字符集 第四辅助集》。它们都是基本集的增补图形字符集,各收 7426 个汉字。

(3) 《基本集》、《第二辅助集》、《第四辅助集》中所收的汉字属简化汉字,以繁体字型表现的对应汉字集分别称作《第一辅助集》、《第三辅助集》、《第五辅助集》。

(4) 《基本集》中汉字国际交换码的表示,是以七单位编码的字符 ASCII 码中,除去 34 个控制码以外的 94 个图形字符代码,按双字节的结构组成。在基本集的汉字码表中,横坐标由 94 个位构成,纵坐标由 94 个区构成,相对的纵横坐标值构成国标汉字的区位码。对于汉字代码来说,这种双字节的编码,是最紧凑有效的编码方式。《第二辅助集》、《第四辅助集》扩充了汉字字数,编码方式仍未变,只是用一批新的辅助集汉字替代了原有的汉字。

(5) 若增加两个辅助汉字集仍不够使用的情况,可以用增加《扩充汉字集》,这种情况较少。

5.2 汉字字体、字形及字号规范

(1) OA 系统中使用的汉字字体以宋体汉字为主,某些文件中标题也须使用黑体汉字。两种字体的数字化字模按照国家通用的汉字字模标准制作。

(2) 点阵式汉字字模的规格,国家标准局已组织制订了几种点阵规格的数字化学模,包括:

15×16 点阵(宋体)供显示和电子式中文打字机用;

24×24 点阵(宋体、黑体),供显示和针式打印机使用;

32×32 点阵(宋体、黑体),供简易激光印字机使用;

48×48 点阵(宋体、黑体),供简易印刷系统使用。

(3) 字号变倍的规范

目前一般所采用的字号加倍方法受技术上的限制,不能满足要求。例如,采用简单的软件放大的方法,将字形横向和纵向各扩大一倍,这样经放大后的尺寸不符合印刷行业所规定的字号尺寸规范。印刷行业通用的部分汉字字号标准如下:

由于印字机印字由分辨率固定, 例如 240 线/寸 (相当于 9.45 线/毫米), 300 线/寸 (相当于 11.81 线/毫米)。因此, 要在实用中推广使用具有连续变倍功能 (即能放大或缩小包括小数点在任意的任意倍数) 的变倍技术, 才能产生符合上述标准字号尺寸的汉字字模。特别是办公自动化系统用的简易印刷系统, 应尽可能向上述字号标准靠近。

在微机上进行汉字输入，关键是微型机必须配备一个汉字操作系统（包括一个汉字汉字操作系统有各种型号和版本。汉字库是把常用的汉字聚集在一起，以备进行汉字输入输出处理。目前，在国内使用的计算机汉字库一般都是按国标 GB2312-80 中规定的 7445 个字符组成。汉字库在微型机内部建立的方法有两种：软汉卡和硬汉卡。软汉卡是把汉字库放在硬盘或软盘上，启动汉字操作系统时调入微型机内部。例如：IBM-PC/AT、IBM-PC/XT 等微型机。硬汉卡是在生产微型机时（硬汉卡也可以单独安装）把汉字库直接装在微型机内部，例如：长城 0520-C、长城 0520-CH、长城 0520-DH 等微型机。不论是哪种方法，都必须和汉字操作系统联合使用。其中软汉卡是操作系统的启动，自动调入微型机内部。

(1) 概述

(2) 认识基本笔画

表 6-1

6 —

注意:

①横、竖、撇三种笔画中,都有相应的短笔画。

②提笔属于横。因为其笔画走向为从左到右。

③提和撇外形相同,但笔画走向不同。

④竖左钩属于竖,竖右钩属于折。

⑤点笔属于捺。

(3) 基本字根、字根键盘、键名

汉字可用几个基本的部分拼合而成。这些用来拼字的基本部分叫做“字根”,能组成尽可能多的常用字的字根,叫做“基本字根”。基本字根共有一百二十五种(一百二十种字根,五种基本笔画)。把这些基本字根安排在英文键盘上,就形成了“字根键盘”。五笔字型在键盘上安排字根的方式是:

①将A—Y24个英文字母键分成五个区,区号为1—5;每区5个键,每个键称为一个位,位号为1—5。如果将每个键的区号作为第一个数字,位号作为第二个数字,那么用两位数字(称为区位号)就可以表示一个键。

②将一百二十五种基本字根按首笔笔画并兼顾键位设计的需要划分作五大类,每大类各对应键盘上的一个区;每一大类又分作五小类,每小类各对应相应区内的一个位。这样键盘上的一个键就与一小类基本字根相对应。因此用一个键的区位号或字母就可以表示该键对应的一小类基本字根。

③从每个键所表示的基本字根中选出一个组字频度较高而形体上又有一定代表性的字根,作为“字根键盘”的键名,每一个键名也表示一个键。五笔字型键盘字根表见表6-2。

键盘字根总表将每个键对应的字根,每个键的区位号放在相应的键面上。每个键面上部为字根,下部为该键的区位号和英文字母;键面上的大写字根为主要字根,小写字根与其左边的大写字根同属一类。左上角的大写字根为该键的键名。下面是按区位顺序排列的键名。

五笔字型键名谱:

①区横起类:王土大木工;

②区竖起类:目日口田山;

③区撇起类:禾白月人金;

④区捺起类:言立水文之;

⑤区折起类:已子女又纟。(“纟”可读丝音)

表6-2 五笔字型键盘字根总表

金 钅 35 Q	人 亻 34 W	月 月 33 E	白 冫 32 R	禾 禾 31 T	立 立 42 U	水 水 43 I	火 火 44 O	之 之 45 P
工 工 15 A	木 木 14 S	大 大 13 D	土 土 12 F	王 王 11 G	目 目 22 J	口 口 23 K	田 田 24 L	;
弓 弓 55 X	弓 弓 54 C	弓 弓 53 V	弓 弓 52 B	弓 弓 51 N	山 山 25 M	<	>	?

- 11 王旁贯头戈(兼)五, 21 鼠具止卜虎段, 31 禾竹一撇双人立, 41 言文方广在四一, 51 巳半已清不出己,
- 12 土主二千十寸雨, 22 日早两竖与虫依, 32 反文条头共三一, 42 高头一撇挑人去, 52 子耳了也框向上,
- 13 大大三羊古石厂, 23 口与川, 字根稀, 33 由手看求三二斤, 43 立辛两点六门广, 53 决刀丸白山朝西,
- 14 木丁西, 24 田甲方框四车力, 34 月多(衫)乃用家衣底, 44 水旁兴头小侧立, 54 火业头, 四点米,
- 15 工龙草头右框七, 25 山由贝, 下框几, 35 金勾缺点头无钩金, 45 犬旁留义儿一点少, 氏无七(兼), 55 幼无旁,

《五笔字型》字根助记词

五笔字型编码方案字根总表，见表 6-3。

字根区总表中包括了每个键对应的区位号、字母、键名、笔形、基本字根、高频字。键名、笔形、基本字根统称为该键对应的基本字根。表中字体较大的字根为主要字根，其后字体较小的字根与它同属一种。

(4) 键名的键入

对二十五个键名的键入只要将键名对应的键连敲四下即可

如：王：GGGG

口：KKKK

纟：XXXX

(5) 成字字根及笔画的键入

成字字根：字根总表中，键名及笔画以外，本身即是汉字（包括“乚”、“一”等有国标的部首在内）的字根。其键入方法为：先打该字根所在键一下（俗称“报户口”），再打该字根的第一、第二及最后一个单笔画。不足三个笔画的字根则以空格结束。如：

五：G（报户口，字根五所在键）、G（第一笔横）、H（第二笔竖）、G（最后一笔）

贝：M（报户口，字根贝所在键）、H（第一笔竖）、N（第二笔折）、Y（最后一笔）

士：FGHG

文：YYGY

女：TTGY

笔画的键入：打笔画对应的键二下，再打两下 L 键。五种笔画的编码为：

一：GGLL

丨：HHLL

丿：TTLL

丶：YYLL

乙：NNLL

表 6-3 《五笔字型汉字编码方案》字根总表

区	位	代码	字母	键名	笔形	基本字根	高频字
1 横起类	1	11	G	王主	一	五戈	一地在要工
	2	12	F	土土	二	十干串寸雨	
	3	13	D	犬犬	三	石古厂	
	4	14	S	木木	平手彡	西丁	
	5	15	A	工工		匚乚乚戈片廿业	
2 竖起类	1	21	H	目且	丨丨丨	上止疋广广	上是中國同
	2	22	J	日目	日月日	早虫	
	3	23	K	口口	川		
	4	24	L	山山		甲口四	
	5	25	M	山山		由门贝儿	
3 撇起类	1	31	T	禾禾	ノ	厂竹彡文	和的有人我
	2	32	R	白	ㄣ	儿手扌手斤斤	
	3	33	E	月月	夕	用乃彡豕彡彡	
	4	34	W	人彡		八彡彡	
	5	35	Q	金彡		彡彡彡彡彡彡彡	
4 捺起类	1	41	Y	言讠	㇏	广文方圭	主产不为这
	2	42	U	立立	㇏	六立辛广门	
	3	43	I	水㇏	㇏	小彡	
	4	44	O	火火	㇏	米	
	5	45	P	之彡	㇏	彡彡	
5 折起类	1	51	N	已巳巳	乙	匚尸心彡	民了发以经
	2	52	B	子子	㇏	口丁耳	
	3	53	V	女女	㇏	刀九彡白	
	4	54	C	又又		厶巴马	
	5	55	X	彡彡		彡彡彡	

(6) 汉字的结构及字型

① 汉字的结构

基本字根在组成汉字时,按照它们之间的位置关系可以分成四类:

④ 单:如“口、木、竹”等。即基本字根本身就单独成为一个汉字。这种情况包括键名及成字字根。

⑤ 散:指构成汉字的基本字根之间可以保持一定的距离。如:汉、湘、结、别、字、意等。

⑥ 连:指一个基本字根连一单笔画。如“ノ”下连“目”成为“自”,“ノ”下连“十”成为“千”,“月”下连“一”成为“且”等。

另一情况是指“带点结构”。如:勺、术、太、主等。这种一个基本字根之前或后的孤立点,一律视作与基本字根相连,这样的字一律视作“连”的结构。

连的关系只存在于单笔画与字根之间，字根与字根之间不存在连的关系。

① 交：是指几个基本字根交叉套迭之后构成汉字。如“申”是由“日”与“丨”交，“夷”是由“艹”与“人”交叉构成的。

② 汉字的字型

④ 哪些汉字需要研究字型

五笔字型只研究由二个或三个字根组成的汉字的字型，由一个或多个三个字根组成的汉字都不用研究字型就可以直接输入。

⑤ 汉字的三种字型

汉字可以分成三种字型：左右型、上下型、杂合型。这些字型的代号分别为2、3，如表6-4所示。

表6-4

字型代号	字型	图 示	字 例
1	左右	田 田 田 田	汉 湘 筑 到
2	上下	日 目 日 日	字 室 花 型
3	杂合	回 回 回 回 回 回 回	困 凶 这 同 乘 本 重 天 旦

⑥ 一型：左右型汉字

如果一个汉字能分成有一定距离的左右两部分（其中任一部分可以由两个字根按散、连、交三种结构中任一种构成）或左中右三部分，则这个汉字称为一型 左右型汉字。如：肝、胡、理、胆、咽、侧、湘等。

⑦ 二型：上下型汉字

如果一个汉字能分成有一定距离的上下两部分（其中任一部分可以由两个字根按散、连、交三种结构中任一种构成）或上中下三部分，则这汉字称为二型 上下型汉字。如：字、节、意、想、花等。

⑧ 三型（杂合）：外内型汉字和单体型汉字

如果组成一个汉字的各部分之间没有简单明确的左右或上下型关系，则这个汉字称为三型 杂合型汉字。如：团、同、这、半、习、巫、天、册等。

⑨ 汉字的结构与字型的关系

① 基本字根单独成字，即汉字由一个字根组成时，输入时按照键名、成字字根或笔画的键入方法输入，不需要判断字型。

② 多于三个字根组成的汉字也不需要判断字型。

③ 属于“散”的汉字，才可以分为左右、上下型。

④ 属于“连”与“交”的汉字，一律属于第三型。

⑤ 不分左右、上下的汉字，一律属于第三型。

(7) 汉字的识别码

① 哪些字需要识别码？

与需要研究字型的汉字相同。

⑧ 为什么要识别码

“叭”与“只”，如果只输入它们的字根所对应的键 (KW)，是不能区分的，但这两个字的字型不同，因此，有必要采用字型信息。

对于“洒”、“沐”、“汀”三字，如果只输入它们的字根所对应的键 (IS)，也是不能区分的，再者，这三字的字型也一样，采用字型信息也无法区分。但它们末笔不同，因此，可用最后一个笔画加以区分。

将字型信息和末笔结合就构成了识别码。

⑨ 识别码的构成

将汉字的末笔代号作为十位，字型代号作为个位所构成的二位数，就称为汉字的末笔字型交叉识别码。由于汉字的笔画有五种，字型有三种，因此末笔字型交叉识别码共有 $5 \times 3 = 15$ 种，见表 6-5。

表 6-5

字 型	左右型	上下型	杂合型
笔 形	1	2	3
横	1	11 (一)	12 (二)
竖	2	21 (丨)	22 (丨丨)
撇	3	31 (丿)	32 (丿丿)
捺	4	41 (丶)	42 (丶丶)
折	5	51 (乙)	52 (乚)

• 举例:

汉: 末笔捺, 代号 4; 字型为左右型, 代号 1; 识别码为 41, Y 键。

字: 末笔横, 代号 1; 字型为上下型, 代号 2; 识别码为 12, F 键。

华: 末笔竖, 代号 2; 字型为上下型, 代号 2; 识别码为 22, J 键。

同: 末笔横, 代号 1; 字型为杂合型, 代号 3; 识别码为 13, D 键。

本: 末笔横, 代号 1; 字型为杂合型, 代号 3; 识别码为 13, D 键。

东: 末笔捺, 代号 4; 字型为杂合型, 代号 3; 识别码为 43, I 键。

说明:

① 注意识别码的使用条件。

② 为了有足够多的区分能力, 对于“进”、“连”等带“辶”的字, 它们的“末笔”规定为被包围部分末笔:

进: 末笔型, 代号 2; 字型为杂合型, 代号为 3; 识别二为 23, K 键。

③ 对于“九”、“刀”、“力”、“匕”四个字根, 当它们参加“识别”时一律规定用“折笔”作末笔。

花: 末笔折, 代号 5; 字型上下型, 代号 2; 识别二为 52, B 键。

④ “我”、“戈”、“成”等字的“末笔”, 遵从“从上到下”的原则, 撇“丿”应为末笔。

⑤ 单独点: “义”、“太”、“勺”等字中的“单独点”, 均被认为与附近字根“相连”, 故为三