

计算机汉字键盘输入技术

汉字部件笔顺码

李建黎 著



电子科技大学出版社

77.9834

LJZ

C-1

计算机汉字键盘输入技术

汉字部件笔顺码

李建黎 著

电子科技大学出版社

内 容 提 要

部件笔顺码是一种好学易记、实用高效的计算机汉字键盘输入技术。本书全面介绍了部件笔顺码的技术指标、编码规则及训练方法，书中还分别列出了以编码类别排序和以拼音排序的部件笔顺编码表。

本书适合广大专业、非专业汉字键盘输入人员使用。

汉字部件笔顺码

李建黎 著

*

电子科技大学出版社出版

(成都建设北路二段四号) 邮编610054

四川省金堂新华印刷厂印刷

新华书店经销

*

开本 787×1092 1/16 印张 6 字数 100 千字

版次 1997 年5月第一版 印次 1997 年5月第一次印刷

印数 1—10000 册

ISBN 7-81043-708-9/TP·287

定价:9.00 元

前 言

汉字部件笔顺码是本人发明的一种计算机汉字键盘输入技术。本书以简洁的形式介绍了部件笔顺码的技术指标、编码规则及训练方法。

客观地讲,近20年来,中文信息处理技术是国内最活跃的研究领域之一,尤其在汉字编码法方面,更是成果林立、新码不断。然而,令各界人士均感满意的汉字编码法一直没有出现,以至于人们不得不在难度与效率的共生矿中寻找适合自己使用的汉字编码法。于是,对汉字编码法形成了“好学的不好用,好用的不好学”的看法。

我在优化部件笔顺码的过程中追求的目标是:“重码率低,词汇量大,记忆点少”。现在可以认为,相对于其它各类流行的汉字编码法,部件笔顺码最接近上述目标。例如,部件笔顺码的单字编码重码率为3.7%;对50 000余条字词编码的重码率为10.4%;学习记忆点为56个。在此我诚挚地向国内所有计算机使用者推荐部件笔顺码,其目的不仅仅是为了参与当前汉字编码技术领域的学术讨论,为汉字编码的百花园增添一朵小花,更是为了让专业与非专业汉字录入人员都掌握这种好学易记、简便高效的汉字键盘输入技术,从计算机汉字键盘输入的“瓶颈”中走出来。

尽管部件笔顺码在优化的过程中,不可避免地涉及到了许多理论和技術上的难题,但是,展现在公众面前的部件笔顺码已经不具有(或者说不应该再有)刻板的理论面孔了。既然是交给大众使用的技术,理当平白如话,清澈如水。现在,部件笔顺码已经做到了小学生学得懂,专业汉字录入人员喜欢用。假如在本书的某一段落上读者读起来费劲的话,实在是因为本人文字表达能力有限,而不是部件笔顺码难懂。当然,部件笔顺码有56个学习记忆点,这仍需要花一、两天的时间才能消化掌握。

电子科技大学党委组织部的各位同志、校内许多领导、教师、学生以及我的家人对我的发明研究工作给予了热情地鼓励和支持;电子科技大学出版社的领导 and 编辑人员对本书的出版做了大量默默无闻的工作。在此一并表示衷心的感谢!

本书的成书时间较短,错漏之处在所难免,敬请专家、学者和读者批评指正,我的通讯地址是:成都市建设北路·电子科技大学党委组织部 邮编610054 电话:(028)3202247; Email address: jlli@uestc.edu.cn

李建黎

1997. 5. 于成都

目 录

第一章 概述	1
§ 1.1 时代的呼唤	2
§ 1.2 汉字键盘输入法的主要技术指标	3
§ 1.3 部件笔顺码的技术优势	6
§ 1.4 部件笔顺码的版本	8
第二章 编码规则	10
§ 2.1 笔画与部件	11
§ 2.2 基本概念及汉字的分类	13
§ 2.3 单字编码规则	14
§ 2.4 词组编码规则	18
§ 2.5 拆字的原则	19
§ 2.6 标点、部首及查询	20
第三章 规则详解	21
§ 3.1 笔画代码的记忆特点	22
§ 3.2 部件代码的记忆特点	23
§ 3.3 子部件的拆分	27
§ 3.4 单字编码详解	29
§ 3.5 词组与简码	35
第四章 训练重点	39
§ 4.1 键盘指法	40
§ 4.2 单字编码分类练习	42
附 录 拼音排序编码表	65

第一章

概 述

- 时代的呼唤
 - 汉字键盘输入法的主要技术指标
 - 部件笔顺码的技术优势
 - 部件笔顺码的版本

本章简述了当前汉字编码技术的状况，讲述了汉字编码方案的主要技术指标，并根据这些指标将部件笔顺码与五笔字型作了对比。读者可以通过本章的学习，从应用的层面上把握评价各类汉字编码方案的基本方法，认识部件笔顺码的技术优势，为选用部件笔顺码作为工作、学习的必备技能建立信心。急于掌握部件笔顺码的读者也可跳过本章，直接学习第二章的内容。

§ 1.1 时代的呼唤

自20世纪70年代末以来，个人计算机以其无与伦比的高速度渗透到社会的各个方面并带动了电子信息技术领域的巨大革命。伴随着个人计算机的加速发展过程，各类系统软件和应用软件争相斗妍、层出不穷，在人类与精细密集的硅片电路之间架起了宽阔的桥梁。人们借助于各种软件成功地将自己的思维活动转化成了在无生命的硬件结构中奔流的电子信息。以汉语作为自己思维外壳的中国人，在驾驭电子计算机的过程中，无不面临将方块汉字从键盘敲入计算机的计算机汉字键盘输入技术。由于东西方文化的天然差异，脱胎于西方文化的计算机键盘多年来一直未能与浓缩了东方文化精髓的方块汉字和睦相处。在中国人脑海中奔腾激荡、轻灵敏捷的汉语意识流一触及到冰凉漠然的西文键盘，顿时显得艰难晦涩，变成了幽咽难续的结巴语。

面对计算机汉字键盘输入的狭窄通道，众多不甘低头、奋力创新的中国人纷纷进入了汉字编码技术领域。人们刻苦钻研、反复摸索，力图消除“瓶颈”、降低“门槛”、开辟汉字进入计算机的宽阔大道。自70年代末以来，许多颇具特色的汉字计算机输入编码方案

相继问世。如，拼音编码方案、首尾码、仓颉码以及80年代面世的五笔字型、表形码、自然码等。进入90年代以后，汉字编码技术领域更是迎来了“万码”奔腾的时代：认知码、傻瓜码、书童码、智能双拼码、天然码、全息码、快三码、天然元声码、飞两笔码……。新码不断，林林总总，令人眼花缭乱、目不暇接。据估计，截至1996年底，已有近千种汉字编码方案向中国专利局申请了专利，上百种编码方案已被编成商品软件进入了市场。

在上述众多的汉字编码方法中，拼音编码方案和优化五笔字型编码法是最具代表性的两种。前者利用了汉字的基本音素，接近人脑的思维习惯，简便易学，缺点是重码率太高；后者利用了汉字丰富的字形信息，重码率较低，有利于实现盲打，缺点是其编码规则脱离了大众通常所具有的汉字知识，使非专业录入人员难于掌握。而其它各种汉字编码方法虽各具优点，但仍未能很好地兼顾低重码率和简单易学这两项基本要求。

汉字部件笔顺码根据汉字的组字部件和构字笔画本身所具有的天然特点，主要采用音托，并辅之以形托、义托、形义托来表征部件和笔画；用部件和笔画依笔顺构字；用音形码方式组词。有效地发掘了音码和形码各自的优点，既符合人脑思维的习惯，实现了编码方案的好学易记，又充分拓展了编码空间，使字词输入重码率大大降低。

日新月异的电子信息时代呼唤出了部件笔顺码，部件笔顺码以其优秀的技术指标无愧于多姿多彩的时代。用部件笔顺码对国标信息交换用汉字编码字符集GB2312—80(简称国标GB2312字库)中的全部6 763个汉字进行单字编码时，单字编码重码率低于3.7%，若采用简码技术，单字编码重码率低于1%；将国标GB2312字库中的单字与45 000条常用词组混合编码时，字词混合重码率低于10.4%。部件笔顺码以其低重码率、大词汇量和规则的简易性等各项优异指标领先于前述各种汉字键盘输入方法，成为继拼音和五笔字型输入法以后的又一种优秀的汉字编码方案。

§ 1.2 汉字键盘输入法的主要技术指标

评价汉字键盘输入法的技术指标可分为定性指标和定量指标

两大类。定性指标可以概括表达为：简便易学、实用高效。定量指标包括：码长、码元数量、记忆点、单字编码的重码率、字词混编重码单字数、词组重码率、词汇容量、字词混编重码率、重码的重数。

码元数量

码元是构成汉字编码的键位符号，如，a、b、c、d、1、2、3、4、；、/、* 等都可以在特定的编码方案中被当作码元。码元数量即是被用作码元的键位符号的数量。

在码长确定时，码元数量越多，编码空间越大，重码的可能性也越小。但是，如果码元数量太多，手指在敲击键盘时换位跨度大，会增大输入的难度、减慢输入的速度。同时，码元数量过多也会增大记忆的难度。一般认为，在现有的西文键盘上，选用26个英文字母所在的三排键位（即中央区键位）符号作为码元，是比较适当的选择。

码长

码长是指编码的长度，码长的单位是键。例如，在拼音编码方案中，“中”的编码为zhong，其码长为5键；“个”的编码为ge，其码长为2键。

在码元数量确定时，码长越长，编码空间越大，重码的可能性越小；码长越短，编码空间越小，重码的可能性越大。当然，如果码长太长，会增加敲键的次数，导致输入速度降低。关于码长、码元数量与编码空间的关系，可以用下面的例子来说明：

如果某汉字编码方案的码元数量为30个、码长为3键，则

$$\text{编码空间} = 30 \times 30 \times 30 = 27\,000 \text{ (条)}$$

显然，该例子中的编码空间太小，无论编码规则如何设置，最大限度也只能做到对27 000条字词无重码。超过27 000条后，每增加一条词组，便增加一个重码。

如果保持该例中的码元数量不变，码长增至4键，则

$$\text{编码空间} = 30 \times 30 \times 30 \times 30 = 810\,000 \text{ (条)}$$

考虑到汉语常用字词的数量不多于20万条，许多汉字编码方案的发明者都认为，在码长为4键的指标下，寻找简单易学且无重码的编码方案，是汉字编码追求的最高目标。

记忆点

在汉字编码方案中，每一个超出现行小学语文教学大纲内容的

条目均被看作是一个记忆点。

在一个编码方案中，每一条编码规则均是一个记忆点，码元所对应的某些字根或部件也是记忆点。在全拼方案中，记忆点为零。在双拼方案中，所有二字母组成的声母和二字母以上的韵母都是记忆点。在各类形码方案中，凡是用音托作为代码的部件（或字根），都不作为记忆点，以形托、义托、形义托作为代码的每一部件（或字根）都应算作一个记忆点。

一般说来，记忆点越多，编码方案的记忆难度越大。

单字编码重码率

在不考虑词组编码的情况下，具有与其它单字相同编码的单字的数目与被编码单字的总数之比即是单字编码重码率（注：本书所述各类重码率指标均以国标GB2312字库中的全部6763个汉字作为被编码总字数）。

字词混编重码单字数

将单字与词组混合编码时，具有与其它字、词相同编码的单字的数目被称为字词混编重码单字数。该类重码单字数越少，从屏幕上选字的次数越少，输入效率也越高。

词组编码重码率

具有与其它词组相同编码的词组的数目与被编码词组的总数之比即是词组编码重码率。

词汇容量

在词组重码率一定时，编码字词库所包含的词组的数量即是词汇容量。词汇容量越大，输入的速度越快。

字词混编重码率

将单字与词组混合编码时，重码字与重码词的总和与全部字、词的总和之比即是字词混编重码率。

重码的重数

指某一重码所对应的重码字词数。例如，在拼音编码方案中，编码kuan对应的单字有三个：宽、款、髓，则该条重码的重数为3；又如，在部件笔顺码中，编码jdqt对应的字词有两条：基地、经得起，则该条重码的重数为2。

重码的重数这一指标可以衡量用人工方式或计算机智能方式鉴别选择重码词的难易程度。一般情况下，重码的重数越高，鉴别

选择重码字、词的难度越大。

评价一套汉字编码方案的优劣，除了采用上述的定性指标和定量指标以外，还应有一些基本要求。如，有利于采用现有西文键盘、尽量不用数字键作为码元、手指敲击键盘的跨距小，等等。

§ 1.3 部件笔顺码的技术优势

本节将全面介绍部件笔顺码的各项技术指标。为了便于读者在比较中鉴别优劣，本节对目前尚保持较高使用率的五笔字型的有关指标也作一对比介绍。读者从中将不难看出部件笔顺码具有的强大技术优势。

选取五笔字型作为对比技术的理由主要有两条：一是五笔字型属于形码输入法，与部件笔顺码有较强的可比性；二是在部件笔顺码推出之前，五笔字型的多项技术指标一直居于众多编码方案的前列，五笔字型本身可以作为部件笔顺码对比的高起点。

部件笔顺码和五笔字型的各项技术指标均已列入表1-1中。需要说明的是，部件笔顺码的统计指标来自于部件笔顺码BJ3.2版，五笔字型的统计指标来自于UCDOS 5.0版中的五笔字词库(WB.IMG)。

表1-1 部件笔顺码的技术优势

技术指标 \ 码类	部件笔顺码	五笔字型
码长	≤ 4	≤ 4
码元数量	30个	25个
记忆点	56个	239个
单字编码重码字数	251个	422个
单字编码重码率	3.7%	6.2%
字词混编重码字数	408个	815个
字词混编重码率	10.4%	15.4%
词汇容量	45 000条	13 285条

注：1) 所用汉字库均为国标GB2312字库；

2) 五笔字型的编码库取自于UCDOS 5.0。

从表1-1可以清楚看到，部件笔顺码和五笔字型的最大码长相

同，均为4键。部件笔顺码比五笔字型多5个码元。即，“Z”、逗号(,)、英文句号(·)、分号(;)和除号(/)。由于部件笔顺码多出的5个码元均位于键盘中央区的键位上，故与五笔字型相比不会增加手指敲键的移动跨度。

部件笔顺码的记忆点为56个，五笔字型的记忆点为239个。前者的记忆点仅为后者的24%。这两种编码方案记忆点的具体计算方法如下：

按照§1.2节所述，每一个超出现行小学语文教学大纲内容的条目应算作汉字编码的一个记忆点；因此，凡是与汉语拼音无联系的部件（字根、笔画）代码均应看作记忆点。在部件笔顺码中，有45个部件是根据形托或形义托确定的代码，有3种笔画是根据形托确定的代码，有8条单字编码规则，其总的记忆点数量为56个。在五笔字型中，有215个非音托字根、5种非音托笔画、15个交叉识别码、4条单字编码规则，其总的记忆点数量为239个。

考虑到形托、形义托具有的助记功能以及记忆难度随记忆点数量的指数变化等因素，可以认为部件笔顺码的记忆难度约为五笔字型的1/10。

从表1-1中还可看出，部件笔顺码的单字编码重码率为3.7%，重码字数为251个（若利用简码技术，则重码数可降至60个）；五笔字型的单字编码重码率为6.2%，重码字数为422个。在其它各项重码指标上，部件笔顺码也明显优于五笔字型，在此不一一赘述。

词汇容量也是衡量汉字编码方案的一项重要指标。人们通常希望自己使用的汉字编码方案包含尽可能多的词汇编码，以利于用整词输入的方式来提高输入的速度。然而，任何一种编码方案的编码空间都是有限的。在同一个汉字输入软件中，编入的词汇越多，在词汇之间、词汇与单字之间形成的重码也就越多。重码率增大到一定值后，即使采用整词输入的方式，输入速度也会大大降低。为了获得较快的输入速度，各种输入软件只好兼顾词汇编码量与重码率这两项重要指标。即，将词汇编码控制在一定数量内，以保证重码率不至于过高。

从表1-1中可以看出，UCDOS 5.0中的五笔字型仅编入13 285条词组，字词混编重码率就达到15.4%；而部件笔顺码（BJ3.2）编入45 000条词组（包括《现代汉语词典》中90%的词条），字词混编

重码率仅为10.4%，体现出了独特的大词汇容量的优势。

关心汉字计算机键盘输入技术前景的读者不禁会问：部件笔顺码的技术指标既然如此优良，它能否取代现存的其它编码方案呢？它下一步改进完善的方向是什么呢？

对于这两个问题，笔者在广泛听取意见、反复研究的基础上，形成了如下观点：

首先，部件笔顺码与当前流行的各种汉字编码方案相比，的确具有许多技术上的优势。随着这项技术的逐步推广，它必将被绝大多数华人计算机使用者所采用。然而，部件笔顺码是在汉字键盘输入这个应用层面上的一种形码，它不可能取代所有现存的其它汉字编码方案。估计部件笔顺码将与拼音方案、区位码在功能上相互依托，相互弥补，长期共存下去。

其次，进一步完善部件笔顺码的方向是实现整句输入的智能化。计算机汉字键盘输入智能化的着眼点在于降低误码率，而不在乎编码的重码率。大词汇量、低重码率的编码方案可以为实现字、词、句输入的智能化降低难度，使智能化键盘输入不需要太大的软、硬件开销。考虑到汉字智能化键盘输入的条件是整句输入，对于人名、古汉语、表格以及新词汇，“智能化”将无能为力。简言之，部件笔顺码还需要向整句输入智能化的目标迈进，智能化汉字键盘输入技术也亟待与部件笔顺码这类简便高效的汉字编码方案相结合。

§ 1.4 部件笔顺码的版本

为适应不同硬件配置的计算机和各类汉字操作系统，当前已开发出了部件笔顺码的BJ2.1版和BJ3.2版。这里将这两个版本的特点作一简要介绍。

BJ2.1是单字编码版本，它包含了国标GB2312字库中的全部6763个汉字的部件笔顺全码和重码字的部件笔顺简码。BJ2.1还包含了一些最常用的中文标点符号。如，句号（。）、顿号（、）。BJ2.1的存储规模很小，应用于DOS环境的版本规模为70KB，应用于WINDOWS环境的版本规模为90KB。BJ2.1不含词组编码，因而其输入重码率即是单字编码重码率，仅为3.6%。使用简码后，BJ2.1的输入重码率小于1%。

BJ3.2 是单字与词组混合编码的版本，它除了包含BJ2.1 中的单字和标点符号的编码外，还包含了45 000 条词组的编码。在已编码词组中，有37 000 余条取自于《现代汉语词典》（商务印书馆1981 年版），有约8 000 余条词组是笔者在广泛查阅80 年代以来的报刊、杂志和文字资料的基础上，挑选汇集起来的常用新词汇，如“移动通信”、“防守反击”、“头球攻门”、“操作系统”、“影碟”、“激光视盘”、“改革开放”、“知识产权”、“流行音乐”、“沪股”、“深股”、“打压”、“思想路线”，等等。由于BJ3.2 拥有丰富的词汇量，熟练掌握后能达到相当高的键盘输入速度。BJ3.2 的字词输入重码率为10.4%，它的存储规模在WINDOWS 环境中为400KB，在DOS 环境中为380KB。WINDOWS 对硬件配置的要求较高，至少应有4MB 内存的支持。

第二章

编码规则

- 笔画与部件
 - 基本概念及汉字的分类
 - 单字编码规则
 - 词组编码规则
 - 拆字的原则
 - 标点、部首及查询

本章所述内容是部件笔顺码的核心部分，涉及到笔画、部件、子部件等概念，完整叙述了部件笔顺码的编码规则。读者在阅读本章后，可结合第三、四章中的大量实例来进一步掌握部件笔顺码。

§ 2.1 笔画与部件

部件笔顺码规定，汉字的构字部件为228种，汉字的基本笔画为9种。用英文26个字母和逗号(,)、英文句号(·)、分号(;)、除号(/)共30个符号作为部件和笔画的代码，使构字部件和基本笔画映射到标准通用英文键盘中央区的相关键位上。在本书中作为笔画和部件代码的英文字母只表示键位，亦即同一英文字母的大、小写均表示同一键位。笔画的形状和代码见表2-1，部件的形状和代码见表2-2。

表2-1 笔画代码表

代码	笔画名	笔画形状
·	点	丶 丶 丶
H	横	一
L	竖	丨
P	撇	丿
N	捺	㇇
Z	折	㇇ ㇇ ㇇ ㇇
G	左钩	丨 丁
V	右钩	レ ㇇ ㇇ ㇇
T	提	㇇

