

汉字输入一日通

—— 肖码易学快速中文输入法

● 肖 水 清 著



HANZISHURUYIRITONG
XIAOMAYIXUEKUAISU
ZHONGWENSHURUFA

7 qgk uai Z	8 b p o - /	9 j uan /
4 f hsi aao -	5 w c u -	6 l nm iou \
1 y r i ,	2 eaoz e -	3 s x an
0 tdlin in ,		

北京经济学院出版社

汉字输入一日通

肖码易学快速中文输入法

肖水清 著

北京经济学院出版社

1993年1月

(京)新登字 211 号

汉字输入一日通

Hanzi Shuru Yitong

肖水清 著

北京经济学院出版社出版

(北京市朝阳区红庙)

北京通县永乐印刷厂印刷

新华书店北京发行所发行

787×1092 毫米 16 开本 4 印张 102 千字

1993 年 1 月第 1 版 1993 年 10 月第 1 版第 2 次印刷

印数:12 001—17 000

ISBN7-5638-0389-0/TP·6

定价:5.00 元

前言

九十年代是计算机的年代,电子计算机已渗透到国民经济各个领域。中国的“八五”计划和十年规划把电子信息产业列为重点发展的支柱产业。《人民日报》指出:现在哪一个领域如果没有使用上计算机,它就决不能被认为是先进的。以前,不识字的被称为“文盲”,今后不会使用计算机的人将是信息时代的“文盲”。

在中国,普及计算机的最大障碍是如何快速输入中文的问题。八十年代使用的中文输入方法主要有两大类:一类是容易学但输入速度慢的方法,以汉语拼音输入为代表;另一类是输入速度快但难学的方法,以“五笔字型”(简称王码)输入为代表。为了解决亿万人民群众能用计算机快速输入中文这一难题,经有关专家充分论证后,1988年,国家科委将“普及型汉字编码方案研究”正式列为中国“七五”重点攻关项目。该项目把重点放在“普及”二字,要求把编码方案与中小学的语文教育结合起来,使今后的学生在毕业后,基本上不用进行编码的再教育,即可使用计算机快速地输入中文,这样的编码方案一旦研究成功,将作为标准方案在全中国推广使用,这是关系到子孙后代的一件大事,其社会效益与经济效益都是无法估量的。

“无理序号数字编码法及其键盘”简称“肖码”,就是根据上述中国“七五”重点攻关项目的要求,而研究出来的一种普及型汉字编码方案,肖码把重点放在易学上,充分利用汉语拼音和笔画的知识,并且只用10个数字编码,做到码长最短,无重码,可高速盲打,达到快速与易学的统一。

肖码从1988年开始研究,1990年研究成功,1992年获得了中国发明专利。肖码软件首批获得中国机电部颁发的计算机软件著作权登记证书。1990年7月,中国国家科委在北京举办“全国民办科技机构科技成果展览交流会”,肖码计算机软件由桂林市科委选送参展,受到好评。1991年3月人民日报社为此向中央进行汇报,中共中央办公厅亲自来电话,询问肖码的详细情况,并调肖码技术资料 and 软件产品去北京。1991年5月1日,中国最权威的计算机行业报刊《计算机世界》以《快速与易学统一的肖码》为题,对肖码进行了介绍,肯定了肖码的先进性。1991年5月3日和5月12日,新华社国内部和国际部分别播发了关于肖码研究成功的电传稿,国内外许多报刊、电台都分别进行报导,如《人民日报》、《广西日报》、《深圳特区报》、香港《大公报》等。1991年6月在“1991中文信息通联世界学术论文评选会”上,肖码论文被中国中文信息学会的最高学术专家们评为优秀论文,这是来自国内外学者的200多篇论文中唯一获奖的汉字编码方案论文。1991年9月3日《人民日报》海外版以《中文电脑打字的新“擂主”》为题,及1991年9月9日《劳动者报》都详细报导了肖码的研究、开发、推广过程。1992年5月,广西电视台、广西广播电视大学联合摄制了肖码教学录像片,并在广西电视台播出,这是在几百种汉字输入方法中,唯一由省级电视台摄制成教学录像片并播出的汉字输入方法。1992年10月肖码教学片获全国优秀电视教育节目三等奖。1992年5月19日至25日,国务院电子信

息系统推广应用办公室、国家科委等单位在北京展览馆联合举办“第二届全国电子信息应用展览会”，这次展览会展出的项目，都是“七五”期间在信息系统建设和电子信息应用方面所取得的获得显著经济效益和社会效益的重大成果，肖码科技成果由广西政府选送参展，并受到重视。《中国电子报》在1992年5月27日头版中心版面刊登了肖码展出的大幅照片。

肖码发明后，得到了各级政府、社会团体的大力支持。1992年6月19日，广西人民政府龙川副主席在办公室接见肖码发明人肖水清，听取汇报后欣然题词：“大力开发肖码，为经济建设作出更大的贡献。”1992年6月8日至10日，国家科委、中国科协、全国工商联、全国青联及中国民协在北京联合召开全国青年民办科技企业代表座谈会，全国50名代表参加。肖水清作为广西壮族自治区的唯一代表出席了这次会议，并受到会议重视。1992年6月9日《科技日报》在头版头条报导了这次会议，其中对肖水清进行了重点介绍。

目前肖码软件已拥有包括中共中央办公厅、国家科委、人民日报社、新华社、机电部、国家统计局等一批国家机关在内的用户。南至广西，北至黑龙江，许多中央机关、省、市政府都在办肖码学习班，培训机关工作人员。

肖码不仅在国内得到广泛应用，而且已经冲出亚洲、走向世界。1992年12月，“第四届美国风险投资年会及发明展览会”在美国波士顿举行，中国由中国民办科技企业家协会陈绳武理事长为团长，第一次组团参加该展览会。肖水清随中国代表团参加了展览，受到美国的热烈欢迎。12月8日，美国马萨诸塞州州长威廉·F·威尔德接见了代表团全体成员。12月10日《波士顿先驱报》刊登了肖码展出的大幅照片。由波士顿世界贸易中心、麻省理工学院企业技术评审委员会、新英格兰电话公司联合组成的评审委员会授予了肖水清先生计算机技术新发明杰出成就奖。

肖码技术资料分册编印：

第一册 基本理论和方法：介绍肖码的基本理论、编码方法和方案、上机操作说明。通过本册的学习，您能掌握肖码的理论和方法，掌握肖码的输入方法和修改字、词编码的方法。

第二册 操作练习：介绍英文指法、肖码指法练习，常用字、常用词组的输入练习。此册内容储存在肖码软盘 read.me 文件中。

第三册 汉字编码表：介绍国标 6763 个汉字的编码。

第四册 普通词组编码表：介绍普通词组的编码。

第三、四册内容存在肖码软盘中，可用肖码程序 xiao.exe 调出。

从第五册起，分专业介绍专业词组的编码。

肖码已获得中国发明专利，肖码软件已由北京大学出版社正式出版，文字资料和计算机软件均不准非法复制，违者必究。凡磁盘标签上无本部印章者，均为非法复制。不经肖码发明人同意，擅自盗用肖码技术者，必追究法律责任。请广大用户向本部举报，对举报者，本部将予物质奖励，并根据案情大小，给予最高 10000 元的奖金。

肖码汉字编码方案已配置到四通 2406 打字机、日本卡西欧打字机等各种打字机、计算机，已生产出肖码软件、汉卡，可挂接到各种汉字系统。已开发出肖码个人中英文计算机、肖码中文记事器。不久即可投放市场。

本书在以上技术资料的基础上编写而成。内容包括肖码设计的理论和方法、肖码的编码方案、肖码的有关程序以及常用字词操作练习等，是一本全面介绍肖码的专著，同时也是有志于学习肖码的广大读者的一本入门读物。本书的出版必将进一步扩大肖码的普及和影响，使越来越

越来越多的人认识到肖码是跨越汉字输入难关的最佳途径和桥梁。可以相信,肖码必将为广大群众掌握汉字输入方法,为作为文化概念的计算机应用技术在中国的普及,以及为使汉字在 21 世纪发扬光大并走向世界作出自己应有的贡献。

肖水清

1992 年 12 月

目 录

前 言	1
第一章 基本理论和方法	1
第一节 概论	1
第二节 中文信息数字化	2
第三节 音形码编码方法	7
第四节 图形符号的编码	14
第五节 形码编码方案	17
第二章 操作说明	23
第一节 第一版中文输入程序	23
第二节 码表维护程序	26
第三节 第二版中文输入程序	30
第三章 肖码的应用	36
第四章 常用字词操作练习	39

第一章 基本理论和方法

第一节 概 论

汉字的使用频率是极不均匀的,根据国家汉字信息处理系统工程对两千多万字词资料的统计,1000个常用字的使用频率为0.908。将汉字按其使用频率从大到小排序,其序号就叫无理序号。无理序号数字编码法就是按照无理序号越小,即越常用的字击键越少、击键越容易的原则,用谐音法把汉字音和形的信息转化为数字,用数字对汉字进行编码。

电子计算机汉字输入方法,一直是在中国普及计算机的“瓶颈”问题。国内外许多专家进行了大量的研究,提出了500多种汉字输入方法。这些输入方法可分为六大类:

一、音码:即根据汉字的读音来编码。如汉语拼音码、山东省唐樊宽的“中文声数码”、北京市刘卫民的“双音码”。音码的优点是容易学、直观。缺点是码长较长,重码率高,不会读的字无法输入。

二、形码:即根据汉字的字形来编码。如河南省王永民的“五笔字型”、陈国斌的“层次四角”、江西省万仁芳的“前三未一”。形码的优点是重码率低,码长较短,不会念的字也能输入。缺点是有一套拆分汉字进行编码的规则,要记忆字根在键盘上的分布位置,较难学。

三、音形码:即同时根据汉字的读音和字型来编码。如四川省陈代于的“大众码”、北京市郭淑珍的“声韵声声”、扶良文的“智能码”、周志农的“自然码”。根据汉字“同音不同形、形似不同音”的规律,音形码的重码率大为降低。但是仍然存在较难学、不会念或念不准的字难输入的问题。

四、数码:即用数字进行编码,如区位码、电报码。数字编码的优点是只用10个数编码,容易操作、无重码。缺点是用一般的学习方法极难记忆。

五、声控输入:即读出汉字的音,就可输入汉字。如北京四达技术中心开发的声控卡。声控输入的优点是,不用对汉字编码。缺点是错误率高,对同音字难以区分。

六、手写输入:即写出汉字的形,就可输入汉字。如哈尔滨工业大学开发的手写识别系统。手写识别的优点是,不用键盘,适合文秘人员用笔的习惯。缺点是输入速度慢,每字平均10画,远远大于每个汉字击键的次数,错误率高,形相近的字难识别。

无理序号数字编码法简称肖码,是在研究、总结了几百种汉字输入方法后提出来的,它集百家编码之精华,扬长避短:

- 1、扬拼音码易学之长,避其重码率高之短;
- 2、扬形码快速之长,避其难学之短;

3、扬数码只用数字编码,无重码之长,避其难记忆之短;

.....

肖码包括音形码和形码两类,每一类又包括 10 个数字输入法和 3×10 个数字输入法。

第二节 中文信息数字化

本发明将汉语拼音的声母、韵母、汉字的笔画、部首各分成 10 类,编码 0—9,分别叫声母码、韵母码、笔画码、笔画修改码、部首码。见表 1。

中文信息数字转化表

表 1

数字	声 母 码	韵 母 码	笔 画 码	笔画修改码
0	d t lín líng	in íng ian iang iong	、点左右型	、点其它型
1	y r	i	、点其它型	、点左右型
2	e a o z zh	e er ei en eng	丨竖其它型	丨竖左右型
3	s sh x (sì shì 除外)	an ang	丨竖左右型	丨竖其它型
4	f h sì shì	a ai ao	一横左右型	一横其它型
5	w c ch	u ü	一横其它型	一横左右型
6	l n m (lín líng 除外)	iou ie ia iao iu	ㄋ 捺	乙 钩
7	q g k	uai ua uo uei ue	乙 钩	ㄋ 捺
8	b p	ou ong	丿撇左右型	丿撇其它型
9	j	uan uang uen ueng ün üan	丿撇其它型	丿撇左右型

声 母 表

表 2

声 母	b p m f d t n l g k h j q x z c s zh ch sh r y w
读 音	玻 坡 摸 佛 得 特 纳 勒 哥 科 喝 基 欺 希 资 雌 思 知 蚩 诗 日 衣 乌

韵母分类表

表 3

单 韵 母	复 韵 母	鼻 韵 母
a 啊	ai ao 哀 熬	an ang 安 昂
o 喔	ou 欧	ong 轰的韵母
e 鹅	ei er 诶 儿	en eng 恩 亨的韵母
i 衣	ia ie iu iao iou=iu 呀 耶 淤 腰 忧	ian iang in íng iong 烟 央 因 英 雍
u ü 乌 迂	ua uo uai uei=ui üe 蛙 窝 歪 威 约	uan uang uen=un ueng üan ün 弯 汪 温 翁 冤 晕

肖码是用谐音法将汉语拼音转化成为数字,比如要记忆圆周率 $\pi=3.14159$,转化成:山崩一寺一壶酒。就很容易记忆了。

一、声母码:汉语拼音包括 23 个声母,见表 2。

数字 0、1、2、3、4、5、6、7、8、9,有多种读音,如:

- 1、按普通话正常读音,如 0 零、1 一、2 二.....
- 2、电讯读音,如 0 洞、1 妖、7 拐.....

3、音符读音,如4发、6拉……

在以上读音的基础上,我们把声母作如下归类:(见表1)

0——因为电讯读音“洞”的声母为d,t是与d成对的声母,所以把d,t归到0。

1——因为读音“一”的声母为y,r的读音与y相近,所以把y,r归到1。

2——因为读音“二”无声母,以韵母e开头,能位于读音首字母的韵母还有a,o,所以把e、a、o开头的汉字称为无声母汉字,其声母码规定为2。

3——因为读音“三”的声母为s,sh,x的读音与s相近,所以把s,sh,x归到3。

4——因为音符“发”的声母为f,h的读音与f相近,所以把f,h归到4。

5——因为读音“五”的声母为w,所以把w归到5。

6——因为读音“六”的声母为l,n,m的读音与l相近,所以把l,n,m归到6。

7——因为读音“七”的声母为q,电讯读音“拐”的声母为g,k是与g成对的声母,所以把q,g,k归到7。

8——因为读音“八”的声母为b,p是与b成对的声母,所以把b,p归到8。

9——因为读音“九”的声母为j,所以把j归到9。

还剩z,zh,c,ch没有归类,考虑到把汉字均匀地归到0~9,本发明把z,zh归到2,c,ch归到5。

如果声母归类与数字读音相抵触,则以数字读音为准,即零的声母为l,四的声母为s,但根据零、四的读音,零归到0,四归到4。

二、韵母码:汉语拼音有37个韵母,包括6个单韵母,15个复韵母,16个鼻韵母。首先根据韵母首字母a,o,e,i,u(包括ü),把韵母分成5大类,每一类又分成单韵母、复韵母、鼻韵母。共分成15类,见表3。其中e,ei,en等的使用频率很小,把使用频率小的韵母合为一类,计有e,ei,en合为一类,o,ou,ong合为一类,a,ai合为一类。这样共分成10类,按照数字0~9的韵母读音,用0~9进行编码,见表1。具体是:

0——因读音“零”的韵母为ing,所以以i开头的鼻韵母归到0。

1——因读音“一”的韵母为i,所以单韵母i归到1。

2——因读音“二”的韵母为er,所以以e开头的韵母归到2。

3——因读音“三”的韵母为an,所以以a开头的鼻韵母归到3。

4——因音符读音“发”的韵母为a,所以以a开头的单韵母、复韵母归到4。

5——因读音“五”的韵母为u,所以单韵母u归到5。

6——因读音“六”的韵母为iou,所以以i开头的复韵母归到6。

7——因电讯读音“拐”的韵母为uai,所以以a开头的复韵母归到7。

以o开头的韵母和u开头的鼻韵母没有对应的数字读音,数字8,9尚未编码,故规定以o开头的韵母归到8,以u开头的鼻韵母归到9。

汉语拼音声母和韵母组成的汉字音节共有400多个,构成的声母码、韵母码组合有87个,还剩13个声母码、韵母码组合是构不成汉字的空音节,见表4。

表 1

汉字	字母	音节	节	编	码	表
----	----	----	---	---	---	---

[illegible]

三、笔画码:根据笔画的读音声母码和汉字字型结构,把笔画分成10类,编码0—9,具体编码是:

汉字的字型可分为左右型、其它型。能把汉字分成左右两部分或左、中、右三部分的字型称为左右型,如“汉、结、撇”。除了左右型以外的其它字型称其它型,如“字、型、因”。汉字的笔画可分为横、竖、撇、捺、点(包括提)、钩(包括折、折钩)。其中横、竖、撇、点的使用频率很大,进一步细分为左右型、其它型的横、竖、撇、点。各类的编码见表1。具体是:左右型的笔画取笔画名的声母码、其它型的笔画紧挨在左右型的笔画旁。

0——因为点的声母码为0,所以左右型的点归到0。

3——因为竖的声母码为3,所以左右型的竖归到3。

4——因为横的声母码为4,所以左右型的横归到4。

8——因为撇的声母码为8,所以左右型的撇归到8。

6——因为捺的声母码为6,捺用得较少,不再分型,都取6。

7——因为钩的声母码为7,钩用得较少,不再分型,都取7。注意一切拐弯的笔画都叫钩,包括折、折钩等。

1——其它型的点紧挨在0的旁边,取1。

2——其它型的竖紧挨在3的旁边,取2。

5——其它型的横紧挨在4的旁边,取5。

9——其它型的撇紧挨在8的旁边,取9。

四、笔画修改码:笔画修改码由0~9组成,其它型的点、竖、横、撇依次为0、3、4、8,左右型的点、竖、横、撇依次为1、2、5、9,钩取6,捺取7。笔画修改码与笔画码有互换关系,见表1。

在一个汉字中,左右型的笔画码不能与其它型的笔画码共存,如0不能与1共存,这样就浪费了码名,增加了重码。设立笔画修改码,就解决了这个问题,充分利用了码名,离散了重码,笔画修改码只位于字、词编码的最后一位。对于初学者,看屏幕提示,可不必强行记忆哪些字的最后一键是笔画码或笔画修改码,熟练后自然就记住了。

6还作为笔画码和笔画修改码的通用码,可以代表任何一种笔画,当笔画不易判断时,都可取6,特别是位于右下方末笔的点不易与捺区分时,可都取6。

五、部首声码:根据《新华字典》收录的189个部首,对部首进行编码。编码方法是:取部首名称的声母码,无名称的部首取部首首笔笔画的声母码,分不清部首的汉字归于单笔部首,取汉字首笔笔画的声母码。为了离散重码,对4个构字能力极强的部首作特殊规定,它们是扌——3、讠——5、艹——8、木——9。《新华字典》收录的各部首编码,见表5。

部首声码表

表5

	横起笔	竖起笔	撇起笔	点起笔	钩起笔
0	大 土 歹 豆	丨 田 冂(念同)	丨 夕	丶 斗 讠	刀
1	尤 弋 页 酉 雨	日 曰 业	人(入) 月 用 鱼	讠(言) 音 讠 扌(衣) 羊(美) 羊	又 幺 羽 己(己)
2	耳 走	止 足	儿 爪(㇀) 竹 自 舟	讠(讠) 讠	冂 卩 巳 子(子)

	横起笔	竖起笔	撇起笔	点起笔	钩起笔
3	西 扌	丨 𠂇 山	彳 夕 手 舌 血 身 鼠 系	忄 心 辛	小 灬 水 糸
4	一 十 土 示 石 豕 影	四 𠂇 (念虎) 黑	𠂇 (食) 父 风 矢 禾 缶	灬 方 火 户	尸
5	厂 开 寸 王 韦 瓦 车 臣 赤 辰	支 虫 齿	久 夊 采	文 彳	
6	龙 耒 老 麦	目 皿 鹵 里	牛 (𠂇 𠂇) 毛 鸟	冂 立 米 麻 鹿	力 女 马 母 矛
7	王 犬 戈 青 革	口 口 骨	彳 彳 气 欠 瓜 谷 鬼	广	乙 弓 厶 冂 冂 五 𠂇 𠂇 𠂇 𠂇 𠂇
8	比 𠂇	卜 (𠂇) 贝	丿 八 片 白 鼻 勺 彳 彳 佳	宀 穴 宀 宀 宀	疋 疋 皮
9	冂 (念匠) 木	巾 见	几 (几) 斤 𠂇 金 角		纟 纟 (念建)

特殊规定:提三水五草八木九。

六、部首数码:部首声码的编码是取部首的声母码,对于不会说普通话的用户难以使用,为此我们又设计了部首数码。编码方法是:取部首的笔画数,7画以上的部首都归到7。分不清部首的字归入单笔部首,取1。例:一——1、彳——2、扌——3、王——4、目——5、耳——6、赤——7。数字0、8、9尚未编码,一些部首的使用频率很大,构字能力很强,为了离散重码,选出33个构字能力很强的部首分配到0、8、9。例:彳——0、扌——8、土——9。见表1、《新华字典》收录的189个部首,见表6。

部首数码表

表6

编码	横起笔	竖起笔	撇起笔	点起笔	钩起笔
0	木 𠂇	口 口	彳 竹	彳 彳	纟 尸
1	一	丨	丿	丶	乙 (→ 冂 𠂇)
2	厂 十 冂	𠂇 卜 (𠂇) 冂	𠂇 久 几 (几) 几	彳 彳 一	冂 (𠂇) 冂 刀 力 厶 又 纟 巳
3	工 寸 开 尤 戈	山 巾 小 (𠂇)	彳 彳 彳 夕 冬 𠂇	广 冂 扌	冂 (𠂇 𠂇) 巳 己 中 弓 马 𠂇 𠂇
4	王 韦 犬 歹 车 戈 比 瓦	比 爻 贝 见	牛 (𠂇 𠂇) 手 毛 片 斤 爪 (𠂇) 父 欠 风 夊 女	文 方 斗 灬 户 心	𠂇 水 月
5	示 石 龙	业 目 田	矢 禾 白 瓜 用 鸟	立 穴 扌 扌	疋 (𠂇) 皮 矛 母 𠂇
6	耒 老 耳 臣 西 页	𠂇 虫	𠂇 舌 白 自 血 舟	衣 羊 (𠂇 𠂇) 米	艮 羽
7	麦 走 赤 豆 酉 辰 豕 青 其 雨 革 影	𠂇 里 鹵 龟 骨 黑	身 采 谷 彳 角 佳 食 鱼 鬼 鼠 鼻	辛 鬼 音 麻 鹿	
8	扌 大	足 (𠂇) 𠂇 四	𠂇 人 (入) 八 (𠂇)	𠂇 (言) 𠂇	女 糸
9	土 土	日 日	月 气	火 火	𠂇 子

第三节 音形码编码方法

一、字编码:全码码长5位,为:

声母码+韵母码+首笔笔画码+尾笔笔画码+部首声码

其中首笔、尾笔指整个汉字的首笔、尾笔。

例:依——11869(yi 丿 ㇏)、诣——11041(yi、一 丨)

按照无理序号从小到大,对最常用字、常用字、次常用字、稀用字分别取一位、二位、三位、四位简码。

最常用字一位:声母码,共有10个,它们是:的——0(d)、也——1(y)、在——2(z)、说——3(s)、是——4(s)、我——5(w)、了——6(l)、个——7(g)、把——8(b)、就——9(j)。

常用字二位:声母码+韵母码,例:应——10(ying)、以——11(yi)。

次常用字三位:声母码+韵母码+首笔笔画码。例:议——110(yi 丶)、亿——118(yi 丨)。

稀用字四位:声母码+韵母码+首笔笔画码+尾笔笔画码或笔画修改码。例:翼——1171(yi 乙、)、谊——1104(yi、一)。

对字取码时,首先按简码取码,有重码时,重码字按五位全码取码。

冷僻字的编码:对于中国中小学教材中没有的冷僻字,一般都不会读,对这些冷僻字的编码如下:

1、部首声码+首笔笔画码+尾笔笔画码或修改码。例:汨——504(丨、一)、汨——505(丨、一)。

2、部首声码+首笔笔画码+尾笔笔画码+余部第一笔笔画码或修改码。例:橄——9467(木一 ㇏)、棣——9466(木一 ㇏)。

3、部首声码+首笔笔画码+尾笔笔画码+余部第一笔笔画码+余部第二笔笔画码或笔画修改码。例:族——85615(尸一 ㇏、一)、蕤——85614(艹一 ㇏、一)。

4、部首声码+首笔笔画码+尾笔笔画码+余部第一笔笔画修改码+余部第二笔笔画码或笔画修改码。例:鲩——18053(鱼 ㇏、一 丨)、鲃——18058(鱼 ㇏、一 丨)。

其中首笔、尾笔为整个字的首笔、尾笔,部首指《新华字典》规定的部首,无部首的汉字以93取代部首声码。当部首位于起笔时,余部第一笔、第二笔指除部首外的第一笔、第二笔,当部首不在起笔或无部首时,余部第一笔、第二笔指整个字的第二笔、第三笔。

前述13个不能构成汉字音节的2位数组合,用来构成一些特殊的字,见表7。编码方法是:

(一)特殊字:有些音节构成的汉字很少,为了离散重码,将前述13个空音节和23个构字很少的稀用音节用来对36个构词能力很强的字编码,这些字称为二位特殊字,字码的第一位是特殊字的声母码,第二位是不规则码(子例外,取94)。一码字和二码字的编码见表8,表中带()的字就是二位特殊字。请读者熟记这110个一码字和二码字。

捺的笔画码为6,在汉字中捺不可能位于首笔,因此在声母码+韵母码+首笔笔画码的单字编码方案中,第三位不可能取6。这些以6结束的三位数组合,用来对一些常用字编码,这些字称为三位特殊字。编码方法是:声母码+韵母码+6。例:日——116、顶——006、特——026。详见表9。

非汉字音节数字组合字词编码表

表 7

组合	二码字	三 ~ 五 码 字	二码词	三 ~ 五 码 词
16	一	图形符,如★	一些	一打头,如一样、一致、一心
20	中	日文平假名,如あ	中央	中打头,如中心、中性、中华
26	自	数字符,如①	自己	自打头,如自费、自动、自然
40	实	日文片假名,如ア	实际	实打头,如实现、实验、实质
46	发	汉语拼音符,如u	发展	发打头,如发动、发达、发扬
50	文	希腊字母,如α	文章	文打头,如文化、文艺、文物
56	无	部首符,如?	无论	无打头,如无法、无疑、无限
87	不	标点符号,如《	不仅	不打头,如不但、不要、不足
89	变	制表符,如r、t	变化	变打头,如变动。制表词组,如+-
92	十	中文数字0零~九拾	五四	中文数组,如以一二、二三
93	家	部首不清的冷僻字	十亿	中文数词,如三万、五千
94	子	用户自定义常用字	一日	中文数量词,如四年、五次
98	教	俄文字母,如б	教育	教打头,如教师。地名,如北京

一、二位字编码表

表 8

二 位 一 位		0	1	2	3	4	5	6	7	8	9
0	的	定	地	(天)	当	打	(同)	(电)	对	通	(大)
1	也	应	以	人	严	(热)	如	(一)	(有)	优	原
2	在	(中)	指	这	安	照	主	(自)	最	重	专
3	说	向	(下)	生	上	(相)	数	小	学	收	选
4	是	(实)	时	分	反	好	(合)	(发)	火	(失)	混
5	我	(文)	(成)	为	产	外	出	(无)	(万)	从	创
6	了	明	理	美	南	那	(面)	(内)	(老)	(来)	(年)
7	个	清	起	可	看	高	(前)	(工)	国	公	全
8	把	平	比	本	半	保	部	(百)	(不)	(报)	(变)
9	就	建	机	(十)	(家)	(子)	(进)	加	(经)	(教)	(交)

说明:()内为二位特殊字

三位特殊字编码表

第三位取 6

表 9

二 位 一 位	0	1	2	3	4	5	6	7	8	9
0	顶	敌	特	但	套	土	调	推	动	断
1	银	日	任	掩	要	于		月	邮	远
2		只	正	仗	载	祖		作	钟	转
3	心	细	舍	山	沙	书	朽	水	手	顺
4		使	和	放	法	复		花	红	换
5		吃	策	长	茶	初		错	重	传
6	良	你	能	满	劳	木	列	落	农	论
7	庆	气	根	抗	开	去	求	却	构	光
8	凭	皮	培	办	怕	步	表		破	
9	奖	急				局	结	决		军

(二)中文数字〇~九,零~拾:92+数字声母码。如有重码,则加数字韵母码。例:二——922、三——923、叁——9233。

(三)用户自定义常用字:有些字在普通文章中是冷僻字,但对某些专业却是常用字。专业用户可将这些字定义为专业常用字,编码方法是:94+声母码,如有重码,则加韵母码、首笔笔画码。如矸——947(g)、锂——9461(li)。

(四)部首不清的冷僻字:有些冷僻字一般人不会读,又不易判断部首。对这些字的编码是:93+首笔笔画码。如有重码,则加尾笔笔画码、第二笔笔画码或笔画修改码。如果连首笔、尾笔笔画也不易判断,则不易判断的笔画用6表示。例如:凸——932、凹——9325、丰——936、甬——9367。

二、词编码:码长1~5位,分二字词、三字词、四字词、五字以上词。

二字词:以二码字开头的二字词,其编码是:首字码+尾字声母码。如有重码,则加尾字韵母码。13个以空音节二位特殊字开头的最常用字取特殊字的二位码,详见表7。例如:一些——16、地点——010(did)、地图——0105(ditu)。

其它二字词按词的使用频率从大到小,对最常用词、常用词、次常用词、稀用词依次取一位、二位、四位、五位。

最常用词一位:首字声母码,共10个。它们是:他们——0(t)、已经——1(y)、知道——2(z)、什么——3(s)、方法——4(f)、我们——5(w)、没有——6(m)、情况——7(q)、必须——8(b)、技术——9(j)。

常用词二位:首字声母码+首字韵母码。例:提高——01(ti)、得到——02(de)。一码、二码词的编码见表10。

一、二位词编码表

表10

二 位 一 位		0	1	2	3	4	5	6	7	8	9
0	他们	领导	提高	得到	但是	它们	突然	条件	脱离	东西	团结
1	已经	因为	意见	认识	研究	压力	于是	(一些)	若干	由于	愿意
2	知道	(中央)	只要	怎么	按照	造成	组织	(自己)	左右	终于	准备
3	什么	现在	希望	社会	商品	少数	许多	消息	所以	首先	迅速
4	方法	(实际)	世界	非常	方面	还是	妇女	(发展)	或者	后来	环境
5	我们	(文章)	此外	问题	完全	材料	除了	(无论)	错误	重新	存在
6	没有	民族	你们	能够	满足	劳动	目前	了解	灭亡	某些	论文
7	情况	强大	其它	革命	感谢	告诉	取得	切实	规律	共同	群众
8	必须	并且	批评	朋友	帮助	包括	补充	标准	(不仅)	玻璃	(变化)
9	技术	今天	继续	(五四)	(十亿)	(一日)	具有	解决	觉得	(教育)	均匀

说明:()内为以空音节二位特殊字开头的常用词。

次常用词四位:首字声母码+首字韵母码+尾字声母码+尾字韵母码。如果尾字是特殊字,则加取尾字第一码、第二码。例如:意义——1111(yiyi)、依然——1113(yiyan)、医院——1119(yiyuan)。

稀用词五位:首字声母码+首字韵母码+尾字声母码+尾字韵母码+尾字首笔笔画码或修改码。如果尾字是特殊字,则加取尾字第一码、第二码、第三码。例:议员——11192(yiyuan)

1)、意愿——11195(yiyuan 一)。

对于 13 个不能构成汉字音节的声母码、韵母码组合,一部分用来对地名、数量词及制表词组编码,见表 7。这样能有效地离散重码。构词方法是:

(一)地名:98+首字声母码+尾字声母码。如有重码,则加取尾字韵母码。例:桂林——9870、北京——9889、广西——9873。

(二)数组:92+首字声母码+尾字声母码。例:一三——9213、三二——9232、三三——9233。

(四)数词:93+首字声母码+尾字声母码。例:二万——9325、四千——9347、十亿——9301、三百——9338。注意第三位“十”取 0。若“两、全、半、每”能构成数量词,则“两”取 6,“全”取 7,“半”取 8,“每”取 9 或只取三码(下同)。

(五)数量词:94+首字声母码+尾字声母码。如有重码,则加尾字韵母码。例如:一点——9410、三月——943、五次——9455。数量词的第四~五位码见表 11。

数量词编码表

一、二位音形码取 93、94,形码取 62、63

表 11

数量词 第四码	数词	量 词 第 五 码											
		四	五	0	1	2	3	4	5	6	7	8	9
0	个	点	天					套	台	条		段	吨
1	亿	月		日	人								元
2	兆	章	丈			张						种	
3	卅	项			升								
4	十	分		时		行	号						
5	万	次		尺	位								寸
6	廿	年	名	米	类	两	厘	辆	秒	面	里		
7	千	克		起				区					
8	百	批	片	匹	倍		排	部		步			
9	京	件	斤	级	键				角	节			

(六)制表词组:89+制表符形状+制表符粗细+制表符长度。例:┐——89012、┌——89412。详见后面制表符的编码方法。

在二字词的编码中,特殊字起着重要的作用,二位特殊字码长 2 位,由构不成音节的二位数组成,首位是特殊字的声母码(子例外,为 9),第二位是不规则码。三位特殊字码长 3 位,由字的声母码+韵母码+6 组成。对于初学者来说,输入某字的声母码,屏幕提示二位特殊字的第二码,输入某字的声母码和韵母码,屏幕提示三位特殊字的第三码 6,根据提示可逐键输入特殊字,不用强记,熟练后,自然就记住特殊字的编码了。

二字词的平均码长 4 位,全部由一码字或二码字构成的二字词,两个字的全部码长 2~4 位,用词编码来输入这些二字词不能缩短码长,所以,对于全部由一码字或二码字组成的二字词不编词码,对这些二字词按单字编码输入。这样可大大减少二字词的重码。只要记住了 110 个一码、二码字的编码和 110 个一码、二码词的编码,就基本上可以按下述编码方法对常用词实现高速盲打:

1. 全部由一码字或二码字组成的二字词按单字输入。