



五笔字型 多功能 速查编码字典

○ 瞿霭堂 劲松 ○ 山西高校联合出版社



五笔字型多功能速查编码字典

瞿霭堂 劲松

山西高校联合出版社

(晋)新登字 8 号

责任编辑 文 琴
封面设计 胡建锋
社 长 刘长鼎
总 编

JS361/02

五笔字型多功能速查编码字典

墨霭堂 劲 松

*

山西高校联合出版社出版(太原并州北路 267 号)

新华书店经销 山西美术印刷厂印刷

*

开本:787×1092 1/32 印张:20.875 字数:451 千字

1995 年 4 月第 2 版 1995 年 4 月太原第 2 次印刷

印数:5701—14700 册

*

ISBN 7—81032—589—2

Z·10 定价: 平装:19.86 元
精装:35 元

五笔字型编码字典的设计和编纂*

(代序)

一、编码字典不能可有可无

编码字典是学习和使用编码的重要和不可或缺的工具。就像学习和使用文字离不开字典一样,在未学习编码以前,编码字典为你提供编码的概貌;学习编码时,它为你释疑解难;在使用编码时,它又备检备忘。王码是一种拼形编码法,就像写字,而且是按部件即字根装配的一种简化方式的写字,装配成一个字要认识字根,知道笔顺,记住规则。由于按形编码是对汉字的二次开发,会写汉字的人常认为只要学会了编码方法就能正确输入所有汉字。这实在是一种误解,就像学会了汉字还要查字典,因为你虽然认得汉字却认不得全部汉字,学会了编码方法能输入大部汉字也不能输入全部汉字。这是因为:第一,王码有两百个字根,这些字根在组成一个字时,形体的变化很大,常常使你一时难于识别,如“尧”中有一个“乚”的字根,“曳”字中有一个“匕”的字根等等;第二,有时候即使认识这个字根也会弄错次序和识别码,何况像“九、刀、力、匕、戏、我、成、藏”等字以及左下两面包围和四面包围的字,笔顺和末笔都有一些与习惯不同的规定,难免弄错;第三,有些字由于区别重码的需要,使用了一些特别的规定和方法,如“末、示、未”、“喜、嘉”、“去、云、支”、“藏、茂”等字和所谓加 L 码,即固定码的字和词,如“灭、丈、尤、朽”、“系列、增加、尽量、反应”等等,都要记忆,难免忘记,有的即使能用选字的办法输入,也大大影

* 本文获首届五笔字型学术研讨会优秀论文二等奖。

响了速度;第四,五笔字型的操作系统提供了两种编码的查询系统:一是语音查询,是一个完整的系统,但不仅六千多个汉字不能人人都读得出来,即使是常用词,方言区的人也常常会弄错;二是所谓万能键即起检索作用的 Z 键,能替代不认识的字根,但首先这是一个不完整的系统,只能查缺某个(或某些)位置字根的一部分字,其次,打一个 Z 键常常为选字要翻几页,多打几个 Z,要翻十几页乃至几十页,查字的效率很低;第五,对于词语来说,也不是知道了编码的方法就都能随意输入的,这不仅有上述的所谓固定码,像“带来、紧接着、从实际出发”等等是无法猜测的,即使是常用词语,因为容量有限,有哪些词和没有哪些词也很难猜测,仅凭打打试试看的办法,影响学习和使用,不言而喻。传统的五笔字型单字码表是以《国家标准信息交换汉字编码》,也即区位码表为基础,3755 个一级常用字按音序检索,3008 个二级常用字则按部首检索,方法不一,查检不便,而词语码表则至今阙如。因此,我们在长期教学实践中深感一本科学和完整的编码字典的重要性。随着王码的普及,使用对象从专业的录入人员深入到不同文化层次的一般使用者,使用的汉字从常见常用的字遍及比较偏僻的字,因此即使是专业打字员也难以每个字都打得出来,不免要查检,就遑论一般的使用者了。再说,一本好的编码字典也应是一本好的辅助教材,学习者在查检过程中不仅能查到正确的编码,也能熟悉和正确掌握编码的规则和方法。由上可见一本科学高效的编码字典不是可有可无,而是学习和使用王码的必备工具。从另一方面来看,编码的研究者都有一个共同的认识误区,为了强调一种编码学习的便捷性,减少学习者的负担,以达到迅速推广和普及的目的,常常忽视或不提供辅助性的工具书籍和有关资料,这是一种短期行为,在电脑尚不普及,编码的学习和使用只局限于专业打字

员或录人员等少数人时,或者说在编码推广的初期,或许能起到一定的作用,随着电脑的普及,编码使用者范围的扩大,层次的提高,使用的深入,对学习和使用编码的科学性和准确性会提出更高的要求,一种好的编码就必须提供必要的系列教材和辅助性工具书籍,把推广深入一步,才能实现真正的普及。

二、怎样才算一本好的编码字典

编码字典就像一般字典一样,关键问题是排序和检索。第一是科学,排序和检索的原则和方法系统合理,涵盖全面;第二是效率,快速查检,准确无误;第三是简便,好学易记,使用方便;第四是实用,也就是多功能,一书多用,能查检和学习兼顾。传统的排序和检索主要是音序和部首。音序排序检索,看似简明,科学合理,但它有一个前提,就是要读得出这个字的音,而六千多个汉字却不是人人都能读得出来的,不说那些难读的字,如沆瀣(hàng xiè)、魑魅(chīmè)、踉跄(liàngqiàng)、饕餮(tāotiè)、倜傥(tītǎng),就是一些常见常用的字,也未必人人都读得出来,如偿(cháng)、黜(chù)、辍(chuò)、俸(fèng)、咖(kā)、鳅(qiū)、袂(mèi)、憩(qì)、稔(rěn)、歃(shà)、舐(shì)、圩(wéi),有些字连北方方言区的人甚至北京人都常常错读,如蓓(bèi)、庇(bì)、畸(jī)、嚼(jué)、赂(lù)、塑(sù)、呷(xiā)、贮(zhù)等,遑论说南腔北调方言的人了。这就是为什么音序检索科学合理而使用却有很大局限性的原因。正因为音序检索的局限性,部首检索还有相当的市场,但部首检索无论从使用和学习来说,比音序检索更是等而下之:第一,部首不统一,各家不一,一个字典一个样,加上汉字简化、印刷字体变更、改并原则不一,部首的形状和数目都不一样;第二,二百个左右的

部首(尚未计变形)内部没有任何形、音的关联,要靠死记硬背,学习一种部首检索不下于学习一种编码的功夫;第三,检索原则混乱,难检字成堆,越是常用字越难检,随便抄录《现代汉语词典》一些难检字,就可见一斑。如“乙、七、丁、九、了、之、三、干、于、下、与、义、上、么、及、久、也、为、天、夫、长、本、以、开”等等。目前王码唯一的检索码表是以《国家标准信息交换汉字编码》,也即区位码表为基础,3775个一级常用汉字按音序检索,3008个二级常用字则按部首检索,方法不一,编排不当,查检不便,是一种非正规的检索方法,可以说集两种方法之短,用作学习和检索的工具已远远不能适应王码飞速发展的需要,设计和编制一本新的编码字典已势在必行。

三、新构想和新方法

一种科学高效的编码字典期望达到的高境界主要有三点:

- (1)排序、检索原则与编码原则一致。音码按音序和语音,形码按形序和字形,统一处理原则,标准划一,系统规范。
- (2)弹性检索。跳出音序、形序按第一位置的单一检索方法,扩大检索范围,加快检索速度。
- (3)不需专门学习。节省学习时间,便于推广普及。

第一种境界称“高”原本有一点滑稽,因为这是文字和信息处理的一条基本原则,比如世界拼音文字无不以字母为序检索,是拼音按音序检索;汉字以部首为序检索,是拼形按形序检索,都是顺理成章的。但是目前各种拼形编码却都以音序或部首检索,与编码本身拆分的形体剥离,形成了重复学习和上述的各种弊病;达到第二种境界,对音码来说,就是模糊检索,说方言的人拼错一个或几个音照样可以

把这个字找出来;以形码来说,就是换位检索或换根检索,只要知道一个字中任何一个部件就可以把这个字找出来。至于不需学习的优点,不言而喻。我们的编码字典即以这三个境界为目标。

1. 以字根为序

王码作为拼形编码,以字根为基本单位,字根就像字母,以字根为序就像拼音文字以字母为序检索一样简捷方便。以字根为序却没有以部首为序而实际无序的毛病,因为王码有一个学习编码时人人必须会背的口诀,这个口诀包括了按区按位顺序排列的132个基本字根,就像一张字母表,使原来无序的字根变得有序,这是王永民教授对拼形编码有序化的一大发明。背诵这个口诀就像背诵英文字母表一样,成为实现字根检索的基础。现将这个口诀列述和解释如下:

口诀	所含基本词根
<u>王</u> 旁 <u>青</u> 头 <u>戈</u> (兼)五一,	王圭戈五一
土土二千十寸雨。	土土二千十寸雨
大犬三(羊)古石厂,	大犬三丰古石厂
木丁西,	木丁西
工戈 <u>草</u> 头 <u>右</u> 框七。	工戈艹匚七
目具上止卜虎皮,	目且上止卜户广
日早 <u>两</u> 竖与虫依。	日早𠂔虫
口与川字根稀,	口川
田甲 <u>方</u> 框四车力。	田甲口四车力
山由贝,下框几。	山由贝冂几
禾竹一撇双人立,	禾竹丩彳
<u>反</u> 文 <u>条</u> 头共三一。	攵夂
白手看头三二斤,	白手彳斤
月彡(衫)乃用 <u>家</u> 衣底。	月彡乃用豕衣

人和八，三四里，	人八
金勺缺点无尾鱼，	金勺角
犬旁留又儿一点夕，	犾又儿夕
氏无七(妻)。	𠂔
言文方广在四一，	言文方广
高头一捺谁人去。	𠂔 ㄣ 主
立辛两点六门广，	立辛 𠂔 六门广
水旁兴头小倒立。	水兴小 𠂔
火业头，四点米，	火业 𠂔 米
之字军盖建道底，	之 𠂔 𠂔 𠂔
摘 𠂔 (示) 𠂔 (衣)。	𠂔
已半已满不出已，	已巳已
左框折尸心和羽。	コ乙尸心羽
子耳了也框向上。	子耳了也 𠂔
女刀九白山朝西。	女刀九白 𠂔
又巴马，丢矢矣，	又巴马 𠂔
慈母无心弓和匕，	𠂔 𠂔 弓匕
幼无力。	么

为便于记忆，这首口诀①用了一些方法把原本没有关系的字串连起来，使得读起来顺口，有节奏感，意义上多少有些联系，以便联想。在上文口诀中，把这些起串连或说明作用的有关词语用下划线标出，以指出它们表明字根的作

①这个口诀不同的版本有一些小的出入，35“金”组的口诀有两种：“金勺缺点无尾鱼，犬旁留又儿一点夕，氏无七(妻)”、“金勺缺点无尾鱼，儿点夕氏无七(妻)”，我们取前者；45“之”组的口诀也有两种：“之宝盖，摘 𠂔 (示) 𠂔 (衣)”、“之字军头建道底，摘 𠂔 (示) 𠂔 (衣)”，我们取后者。所以取这两种口诀，是因为它们与132个基本字根相符。

用。这些表明字根的方法如下：

(1) 垫音方法：指加一个不起作用的字，只为增加节奏感，读起来顺口，如“旁”字，“王旁”、“犬旁”、“水旁”就是指“王、丩、水”。再如“一点”，“一点夕”就是指“夕”。

(2) 说明方法：用这种方法指出非成字字根和个别比较难区别的成字字根，一般通过与有关字比较的方法，如：

1) 青头、条头、看头、高头、兴头、字军头 指这些字的上半部分，即“主、久、丩、一、丩、一、一”。

2) 家衣底、建道底 指这些字的下半部分，即“豕、衣、丩、丩”。

3) 无尾鱼（鱼无尾巴）= 鱼、勺缺点（勺字缺一点）= 勺、一点夕（指带一点的夕，不同于不带点的夕和带两点的夕）= 夕、氏无七（氏字少了中间的七字）= 氏、小倒立（小字颠倒过来）= 小、母无心（母字无中间的笔划）= 母、幼无力（幼字没有右边的力字）= 幼、丢矢矣（矣字失去了矢）= 矣、谁人去（谁字去掉中间的“亻”）= 谁、圭框向上（框的缺口向上）= 圭、山向西（山字向西，即向左转 90°）= 山、摘丩丩（把丩和丩中的一点“丶”和两点“丶”去掉）= 丩、已半（已字的左框只封了一半）= 已、已满（已字的左框全部封住）= 已、不出己（己字的左框完全没有封住）= 己。

(3) 注音、注义方法：在字根后加括号注明读音或谐音的意义。如（羊）指这个字根读“羊”的音、乡（衫）指读“衫”的音、戈（兼）指读“兼”的音也表“兼”的义、示（示）指读“示”的音、衣（衣）指读“衣”的音、七（妻）指联想“妻”的意思。

(4) 取名方法：给一些非成字字根取个名，大多是习惯称呼。如“草头”= 艹、右框 = 匚、左框 = 匚、下框 = 冂、方框 = 口、一捺 = ㇏、两竖 = 𠄎、两点 = 冫、四点 = 灬、折 = 以“乙”代表的笔划。

唯有以“慈”字代表“纟”这个非成字键名字根，用的是一种不准确的半谐音的方法。把上面的解释代入口诀中，数一数正好是 132 个基本字根。

王码共有 200 个字根，132 个是基本字根，68 个是非基本字根。基本字根与非基本字根有三种关系：一是变形关系，即非基本字根与基本字根笔划相同、形体相异；二是变体关系，即非基本字根与基本字根笔划不同、形体相似；三是随机关系，即非基本字根与基本字根没有形体上的联系，也就是少数形体独立的非基本字根。如 ㄣ、ㄤ 是“厂”的变形字根，ㄥ 是“厂”的变体字根；ㄣ、皿 是“四”的变形字根，囧 是“四”的变体字根；而 县、夕、ㄣ、丨、ㄣ 等都是独立的非基本字根。我们以基本字根为主干按口诀顺序排序，把非基本字根按与基本字根的变形或变体关系，或同键关系排在相应的基本字根的后面，下面即为与字母表相似的包括 200 字根以 132 个基本字根排序的王码字根表：（表见 18 页）

王码字根的数目常常难以确定，这主要是由于对非基本字根，也即对变形或变体字根的不同取舍所致。我们确定为 200 个字根基本上根据王永民教授的《五笔字型键盘字根总图》，不作更改，因为这张图流传广，影响大，如果更改就会影响已经学会王码的人使用这本字典。其实，非基本字根是与相应的基本字根处于同一键盘的，因此只要记住了基本字根，与基本字根相应的变形和变体字根采用模糊记忆法就能记住，不需要专门背诵和记忆。需要记住的是少数独立非基本字根的键位，但大多独立非基本字根出现频率小，有的与所在区位关联，在使用过程中就能自然熟悉并记住。

以基本字根排序，达到了上述的两个高境界：排序、检索原则与编码原则一致，并且不需专门学习。它既克服了部首查字无序的缺点，又具有像部首检索一样不认得的字和不

会读音的字同样可以查检的优点。

2. 弹性检索

所谓弹性检索就是指一种互见互索的方法。简单来说,就是不像一般检索方法一样,只按第一位置的字根排序,而是根据王码的编码原则按每个字根可能出现的第一、二、三、末四种位置分别排序。比如第一字根“王”,先按“王”在第一位置排序(同位字根再按音序,下同):璦、班、斑、碧、玢、玻、璫、琛、璵、琮……,再按“王”在第二位置排序:癍、宝、呈、逞、盖、羔、羹、国、皇、遑……,再按“王”在第三位置排序:斑、埤、程、糕、掴、逛、蝓、煌、惶、凰……,再按“王”在末一位置排序:斑、癍、塵、嵒、醴、螭、玼、醴、望。这样,第一字根相同的字可以在不同的字根部中找到,第一字根不同的字也可以在同一字根部中找到。换句话说,几乎只要认得一个字中任何一个字根,就能查到这个字的编码,不需像传统的音序或部首查法必须按第一个字的语音或部首查阅。如查一个“何”字,在“亻、可、口”三个字根里都能查到。查“亻”时,它作为第一字根出现在“人”部的变形字根“亻”的第一个位置;查“丁”时,它作为第二个字根出现在“丁”部的第二位置;查“口”时,它作为第三个字根出现在“口”部的第三个位置。这种“互见互索”的弹性检索方法有什么好处呢?

第一、解决了不认得第一字根也能查检的问题,而且几乎只要认得一个字中任何一个字根即可查到这个字的编码。

第二、大大加快了查检的速度。总的来说,王码 200 个字根出现在第一位置超过 100 个字的只有 15 个字根,占 0.8%,超过 50 个字的只有 19 个,占 10.%,合起来不到 20%,也就是说,按字根检索,80% 以上的字根按第一位置查检都不影响速度,问题只在 20% 的字根上,因为这些字根大多是常用

字根,现将这些字根在不同位置上出现的字数列述如下:

一 (131、403、313、155)	土 (154、123、117、93)
木 (262、83、81、73)	ㄥ (372、115、43、26)
口 (397、223、342、224)	丿 (156、333、215、77)
扌 (320、31、8、0)	亻 (246、89、53、0)
车 (218、2、0、0)	讠 (143、7、3、0)
彳 (366、30、0、0)	宀 (131、112、17、0)
亻 (128、2、0、0)	女 (111、20、45、31)
纟 (141、9、1、0)	王 (87、28、44、9)
石 (96、6、7、4)	西 (63、38、7、0)
目 (58、36、43、13)	日 (55、122、124、4)
车 (50、29、13、5)	山 (82、37、36、35)
冂 (53、125、140、0)	禾 (66、26、9、7)
鱼 (66、0、0、0)	彡 (75、6、0、0)
夕 (74、79、24、0)	广 (68、41、5、0)
宀 (62、94、27、1)	疒 (99、2、0、0)
火 (79、23、31、11)	ㄣ (75、0、0、0)
尸 (54、79、32、2)	乚 (63、18、33、43)

从以上统计可见,一般情况下同一字根处在非第一位置的字要比处在第一位置的字少得多,因此遇到第一个字根字多的情况下,只要查检这个字的其他字根即可。比如查一个“槎”字,第一字根“木”的字有 262 个,逐字查找太麻烦,就可以换一个字根查,比如查末一字根“工”,即查“工(四)”(“工”在第四个位置上的字),一共只有 17 个字,“槎”在第一个字。当然,也可查第二个位置上的“ㄣ”或第三个位置上的“尸”。“ㄣ”在第二个位置上的字共有 65 个,“槎”在第四个字;“尸”在第三个位置上的字只有 7 个,“槎”在第一个字。由此可见,使用这种方法查字,不一定查第一字根,越查后面的字根越好查,特别是第一字根的字多

的时候效果更明显。这就大大地加快了查检的速度,达到上述第二个高速的境界。

第三、便于学习和教学。由于按字根和位置双轨的排列方法,同一字根的变形、变体和组字方式,不同字根在一个字中可能出现的位置一目了然。学习者可通过字典熟悉字根在不同位置时的变形和变体,也可以学习组字和析字的方法。教师则可方便地找到各种资料,编写出更加全面和完整的教材。因此这种字典不仅是快速的查检工具,也是绝好的教学辅助材料。

四、再进一步

达到了上述的三个境界不能说已经尽善尽美,还可以进一步提高字典的适应性和扩大字典的使用范围。

第一,编制一个音序检字表,使熟悉语音检索的人能按音序检索。

第二,编制一个部首检字表,使熟悉部首检索的人能按部首检索。

第三,编制一个难拆字音序索引,将字根形体变化大而难于辨认的字、键面字和不成字的部首等这些难拆和特殊的字单独列出,以备查检。

第四,编制一个词语码表,分双字、三字、四字和多字四类按音序排列,使用者可以通过词语码表查检词语编码,了解词语编码的分布情况,提高输入速度。

这样,这本编码字典集以字根为主的多种检索方式于一身,不仅包括单字而且包括词语,既使用于检索,还能辅助教学,因此是“多功能的”;由于按字根检索不需学习,查字也不以认得或会读这个字为条件,加上采用与位置双轨排列的“互见互索”方法,甚至不以认得需查字的第一个字根

为条件,只要认得一个字中任何一个字根即可查检,还有其他检索方法备用,查字的速度要比一般的检索方式快得多,所以又是“速查”的;于是,这种编码字典终于成为一种真正的多功能速查编码字典。

瞿霭堂、劲松

写于中国人民大学语音实验室

1993年11月

凡 例

一 条目安排

1. 条目分单字和词语两部分:单字条目共收简体汉字 6763 个,包括一级汉字 3755 个,二级汉字 3008 个。词语条目共收 10944 个,包括二字词 7873 个,三字词 1532 个,四字词 1464 个,多字词 75 个。本表所谓的词除严格意义上的词外,还包括“不怕”、“带来”、“解放前”、“紧接着”、“北京时间”、“背道而驰”、“从实际出发”等词组、成语、熟语和习惯用法。

2. 单字条目双栏编排,每栏四项:第一项是汉字,第二项括号中是汉语拼音,第三项是五笔字型编码,第四项是区位编码。字音一律按汉语拼音方案的规则拼写。多音字只取一个常用音。声调符号用数字表示:阴平标作 1,阳平标作 2,上声标作 3,去声标作 4,轻声标作 5。词语条目四栏编排,每栏两项,第一项为词语,第二项是编码。按音序排列,每一个字母开始时,有题头字母注明。

3. 五笔字型的编码一律按全码列出,不足四个字母的编码是少根字,要按空格键补上。编码中的字母分大写和小写两种:大写表示必须输入的码,小写表示可以省略的码,不管省略几码必须按一次空格键代替。这是一种全码和简码共存的写法,从中既可以看出这个汉字的全码,也可看出简码。比如“拜”的编码是 RDFH,都是大写,说明这个字要用全码输入,没有简码;“白”的编码是 RRRr,三个字母大写,一个字母小写,说明这个字是一个三级简码,即既可按四下 R 键,用全码方式输入,也可以按三下 R 键和一下空格键,即按三级简码的方式输入;“瓣”的编码是 URcu,两

个字母大写,两个字母小写,说明这个字是一个二级简码,即既可以按四键用全码的方式输入,也可以按 UR 两键再加一个空格键,用二级简码的方式输入;“八”的编码是 WTY,只有三个大写字母,说明这个字是一个少根字,即不足四个字根的字,只要按这三个字母再加一个空格键即可输入。

二 排列方法

1. 单字不分级别,统一按基本字根次序排列。基本字根的次序按照口诀的顺序确定。五笔字型的编码共有 200 个字根,分基本字根和非基本字根两类:基本字根 132 个^①,由一个口诀组织起来,以便记忆。其余的 68 个非基本字根又分变形和变体字根两类:变形字根指与基本字根笔划相同、形体相异的字根;变体字根指与基本字根笔划不同、形体相似的字根。本字典以基本字根为条目,非基本字根按变形、变体的顺序用括号放在相应的基本字根的后面。如“厂(丂、𠂆、𠂇)”指“厂”这个基本字根包含“丂、𠂆”两个变形字根和“𠂇”一个变体字根。排列时先按口诀顺序排列基本字根,再按变形、变体的顺序排列与基本字根相应的非基本字根,非基本字根的题头字加括号表示。请参阅字根表和字根目录。

2. 同一字根的单字按“互见互索”的原则,再按字根在单字中的位置排列,字根及其位置都相同的字则按音序排

^①王永民教授在 1993 年出版的新教材中,将基本字根作了新的合并,由 132 个合并为 125 个,这种合并虽然加强了字根归类的齐整性,由于没有对口诀进行相应的调整,基本字根与口诀失去了联系,不利于记忆,我们仍用旧说。