



国际标准汉字全形检索 电脑输入法及其字集

谭国政 陈永明 谭文武 著

浙江大学出版社

C J K
汉字

国际标准汉字全形检索 电脑输入法及其字集

谭国政 陈永明 谭文武 著

自从计算机技术开始进入了全新的信息时代以来，人类社会就取得了空前的进步与文明。语言文字与电脑及通讯的结合，是现代信息社会的一个重要特征，也是人们掌握使用电脑的重要前提。电脑对使用字母文字的人们来说，是个简单的任务，而对使用汉字的人们来说却是个严重的挑战。如何使博大精深的古老汉字能够享受神奇智能的全新电脑技术，关键在于解决电脑汉字输入问题，而最常用的电脑键盘输入的关键在于解决汉字编码问题。

纵观电脑的各种汉字部件编码，主要问题出在部首选择和汉字的拆分上，究其原因是对汉字编码的基本理论缺乏系统的研究。面对新的国家标准即国际标准的中日韩三国汉字的出现，深入开展汉字编码理论的研究，建立一个统一的汉字编码就显得更为重要和紧迫。这本著作就展示了作者在这方面的研究成果，尤其是建立汉字部首的统一性原则，实现汉字检索与电脑输入一体化的思想，确是独特创新之处，也是汉字形码的重大突破。为实现电脑教育从娃娃抓起，为电脑的普及推广，为面向21世纪电脑的新发展都有重要作用。

向社会建议，向公众推荐，尽快推广应用国际标准汉字全形检索电脑输入法，将会使你在电脑世界游刃有余，运用自如，乐趣盎然。是为序。

浙江大学出版社

820×1168 32开 9印张 245千字

1992年8月第1版 1992年8月第1次印刷

ISBN 7-308-01587-4/TP·145 定价：15.00元

地址：杭州大学路28号 浙江出版集团 浙江出版集团 浙江出版集团

电话：0571-2301024

邮编：311104

云种网商：城数

国际通用汉字全形检索 及其输入法

著 者 谭国政 陈永明 谭文武

国际标准汉字全形检索

电脑输入法及其字集

谭国政 陈永明 谭文武 著

责任编辑 俞妙送

* * *

浙江大学出版社出版

浙江大学出版社电脑排版中心排版

杭州富阳何云印刷厂印刷

浙江省新华书店发行

* * *

850×1168 32开 9印张 242千字

1995年8月第1版 1995年8月第1次印刷

印数 0001—1000

ISBN 7-308-01587-4/TP·142 定价:12.00元

如发现书中有缺页、倒页和破页,请持此证到杭州富阳何云印刷厂调换
地址:富阳何云 邮编:311404 电话:0571-3201054

序 言

自从 20 世纪 40 年代出现第一台电脑以来,人类开始进入了全新的信息时代,电脑以其飞速发展和神奇功能给社会带来了空前的进步与文明。语言文字与电脑及通讯的结合,是现代社会信息化的一个重要特征,也是人们掌握使用电脑的重要前提。电脑对使用字母文字的人们来说是得心应手的事情,而对使用方块象形文字的我们来说却是个严重的挑战,如何使博大精深的古老汉字能够享受神奇智能的全新电脑技术,关键在于解决电脑汉字输入问题,而最常用的电脑键盘输入的关键在于解决汉字编码问题。

纵观电脑的各种汉字部件编码,主要问题出在部首选择和汉字的拆分上,究其原因是对汉字编码的基本理论缺乏系统的研究,面对新的国家标准即国际标准的中日韩三国汉字的出现,深入开展汉字编码理论的研究,建立一个统一的汉字大厦就显得更为重要和紧迫。这本著作就展示了作者在这方面的研究工作成果,尤其是建立汉字部首的统一性原则,实现汉字检索与电脑输入一体化的思想,确是独特创新之处,也是汉字形码的重大突破。为实现电脑教育从娃娃抓起,为电脑的普及推广,为面向 21 世纪电脑的新发展都有重要作用。

向社会建议,向诸公推荐,尽快推广应用国际标准汉字全形检索电脑输入法,将会使你在电脑世界纵横驰骋,任用自如,乐趣盎然。是为序。

第四章 选择汉字构字部件	17
第一节 分析部首	17
第二节 选择部首	18
第三节 选择复笔组字单元	20
第四节 字部首与组字单元的说明	21

浙江省科委副主任、高级工程师
昌金铭

目 录

第一章 汉字部首优选原则	1
第一节 从查字典得到的启示	1
第二节 部首统一性原则的实现	2
第三节 非左右结构字部首统一性原则	2
第四节 选取部首办法	4
第二章 汉字的结构规律	6
第一节 汉字的基本笔画	6
第二节 汉字的结构	7
第三节 方块结构的分析	9
第四节 汉字的部件	10
第三章 电脑汉字分解原则	11
第一节 从电脑键盘谈起	11
第二节 汉字的认知习惯和书写方式	12
第三节 汉字规范化	13
第四节 电脑汉字的切分问题	15
第四章 选择汉字构字部件	17
第一节 分析部首	17
第二节 选择部首	18
第三节 选择复笔组字单元	20
第四节 字部首与组字单元的说明	21

第五章 键盘安排与难拆字分解	23
第一节 字部首的安排	23
第二节 组字单元的安排	24
第三节 难拆字分解举例	26
第四节 拆字的几点说明	32
第六章 汉字全形编码	34
第一节 编码原则与实施	34
第二节 全形码表与口诀	38
第三节 难拆字编码举例	40
第四节 几点注意	46
第七章 总结与展望	48
第一节 变革性的思路	48
第二节 展望未来	49
第八章 中日韩国际标准汉字全形码字集	52
第九章 简化汉字集的全形编码	217
第一节 简化汉字集全形码索引	217
第二节 简化汉字的词汇编码	271
第十章 全形码电脑操作指南	273
第一节 系统文件清单及简介	273
第二节 在 UC DOS 下使用全形码	273
第三节 在中国龙系统下使用全形码	274
第四节 在金山 SPDOS 系统下使用全形码	274
第五节 全形码在 WINDOWS 环境下的安装与使用	274

第一章 汉字部首优选原则

在中国汉字字典或词典中,一般都有部首检字表,它包括三个内容:一是部首目录,二是检字表,三是难检字索引。这说明检索汉字不能像查西文那样一次完成。在字典检字法中把汉字中包含有共同形体偏旁的归为一部,这个共同的形体偏旁就是这一部汉字的部首。传统部首从《说文解字》中的 540 部减少到康熙字典的 214 部。最近颁布的国际标准汉字 ISO/IEC 10646(即国家标准汉字 GB 13000)所用的部首也是 214 部,而新华字典又减少到 189 部。这不仅大大减轻了人脑记忆的负担,同时也说明部首处于不断发展变化之中。为了实现检字法与编码法一体化,必须对汉字部首进一步优选,使其具有统一性。如何优选,其原则是什么,就是本章要讨论的问题。

第一节 从查字典得到的启示

在每种汉字字典中都有部首检字法。人们在查汉字时,总是先查这个字的部首。在确定了这个字的部首之后,再按笔画多少在这个部首的汉字中去找这个汉字。这就是我们在儿时学会的查字法,也是如今所说的汉字二次查找法。

人们在经常查字中发现,绝大部分的部首都在字的左边、上方、左上方或外框上,这类字要占九成以上。同时也发现,有少量字其部首所在部位不定,上下、左右、内外都有,因而有些字是很难查到的。现在,新华字典在编排上采用一个字在多个部首中出现的办法并列出难查字表,即使这样仍难以解决少数字难查的问题。这里要提出的问题是,能否对这些字的部首也限制在字的左边、上方、左上方或外框上。这就是我们所说的部首的统一性原则。

东汉许慎在《说文解字》中说独体为文,合体为字。他把汉字分为

独体与合体两种。如何使每个汉字都具有部首统一性原则,同时做到拆字有规定性,在此基础上设计出一个合理实用的编码方法,编出一本一次查找汉字的字典,这应该是我们追求的目标,也是使电脑走进千家万户和实现“电脑教育从娃娃抓起”的保证。

第二节 部首统一性原则的实现

1. 不管是独体字还是合体字,如果其部首已位于该字的左边、上方、左上方或外框外,这些部首基本不变。

2. 对左右或左中右结构的字部首如不具有统一性原则,可分三种情况处理:一是如走之旁(辶)与建之旁(廴)之类的部首,它们位于字的左下方,而且在汉字结构中有传统的称谓,我们就把它们看成字的左边部首而选用。二是有的字部首可左可右,如“艰”、“叔”中的“又”,“陆”、“邓”中的“阝”,“鸵”、“鸽”中的“鸟”等,我们对它们位于左边的部首保留,如“艰”取“又”部,“陆”取“阝”部,“鸵”取“鸟”部。我们对它们位于右边的部首换成左边或左上方的部首,如“叔”取“扌”,“邓”取“又”部,“鸽”取“人”部。三是有的字部首只能取在右边,对这些字的部首只能取消,一律换成左边或左上方的部首作为该字的部首。如“巾”、“彡”、“斗”、“犬”、“攴”、“攵”、“欠”、“攴”、“见(見)”、“页(頁)”等部都不能再作为部首,如“刑”、“形”、“顶”都取“一”部,“斜”取“人”部,“献”、“故”都取“十”部,“观”、“欢”都取“又”部,“殺”取“×”部,“敵”取“一”部。通过以上的处理,对于左右结构字的部首就可以达到统一了。

3. 非左右结构字部首的统一性问题比较复杂,涉及到汉字结构及拆字原则,内容丰富,我们在下一节中详细叙述。

第三节 非左右结构字部首统一性原则

我们要研究的拆字方法必须具有合理解释和易于教学的特点,

要为中外人士学习汉语着想。说文解字古已有之，而且新说不断，如老一辈革命家在启发工人觉悟时，指出“工人”为“天”就是一例。从上下结构说“天”也可以拆成“一”、“大”，意说天为第一大了。这就是会意造字。从笔顺上也可拆成“二”、“人”，这当然和“夫”字的笔顺拆字一样了。从上下结构说“夫”也可拆成“土”、“人”，意为“丈夫是土地的主人”，这和“男人是田地的劳动力”一样去理解。也许有人说“夫人”不是指女性吗？其实“夫人”应理解为：“丈夫的人”，这当然是指女性了。至于说到“夫”为什么上下可断，这可以从“人”字的结构特点来说明。在《编辑部的故事》主题歌中说：人字的结构就是相互支撑。这一说法颇富哲理，也就说明人字的结合点很重要，因而将“夫”字上下分拆为“土”、“人”，就是自然的事情了。又如字典中将“重”归入“里”部，本身就将“重”字拆成“丿”、“十”、“里”了，如果取消“里”部，那“重”字就可拆成“丿”、“十”、“田”、“土”了。这样做也是很好理解的。推广到：“未”拆成“土”、“个”，“末”拆成“土”、“个”，“耒”拆成“土”、“木”，“来”拆成“艹”、“一”、“个”，“夹”拆成“艹”、“一”、“人”等，也是情理之中的事情。以上说的是一些上方交叉、形非左右结构字部首统一的意见。

还有一类非左右结构字部首需要更换。如字典查“看”字就要查“目”部，显然不符合部首统一性原则。从“看”字内涵说，这是一个象形字，用手放在目上就是看的动作形象，因此把“看”字归于“手”部也未尝不可。再如字典查“我”字就远不如查“你”、“他”方便，“我”与“你”、“他”同为人称代词，而要查“戈”部。而这“戈”是隐含在字内的，能不能与“你”、“他”统一取“亻”部呢？这应该可以办到，因为“我”取“戈”部已经拆字了，现在取“亻”旁并不和原拆字取“戈”部相矛盾，只不过“亻”旁取在“我”字的左上方而已。这样取法在字的意义上也统一了。以此类推，如“戢”取“口”部，而原取“戈”部；“栽”取“十”部，而原取“戈”部；“千”取“丿”部，而原取“十”部；“垂”取“丿”部，而原取“土”部；等等。总之，对于一些可上可下的部首，我们总是保留字上方的部首，而把字下方的部首替换成字上方的部首。如“寿”由“寸”部换成“耂”部，“垫”由“土”部换成“扌”部，“忘”由“心”部换成“忄”部，

独体与合体两种。如何使每个汉字都具有部首统一性原则,同时做到“泰”由“水”部换成“𣎵”部,“贵”由“贝”部换成“中”部,“拿”由“手”部换成“人”部,“禁”由“示”部换成“木”部,“盂”由“皿”部换成“一”部,“袋”由“衣”部换成“亻”部,“民”由“氏”部换成“口”部,表明民众吃饭、说话都要用口。通过以上替换的办法,就可以把这一类非左右结构字的部首都统一到字的上方或左上方。

另外,还有一类由封闭框或半封闭框所构成的非左右结构字。根据部首统一性原则,这类字的部首要取在外框上。如“回”取“口”部,“同”取“冂”部,“问”取“门”部,“风”取“几”部等。对于“匚”、“冂”、“凵”为部首的字都要取消原部首,因为它们无法实现部首统一性原则,都要替换成上方的部件。如“区”取“一”部,“寻”取“冂”部,“凶”取“乂”部(即“凵”形只能在字下部),还有类似的“义”字,也只能取“丶”部。通过以上处理,这类字的部首也就统一了。

还有几个特殊交叉结构的字,如“弗”、“夷”、“秉”、“隶”、“曲”、“尹”、“事”等字,它们原来部首分别为“弓”、“大”、“禾”、“丰”、“臼”、“乚(尸)”、“一(丨)”。要实现部首统一,我们只能采用机械方式将它们上下分割,把它们的部首分别变成“丿”、“ナ”、“ノ”、“丨”、“丨”、“𠂇”、“十”。另外还有几个特殊嵌套结构的字,如“乘”、“幽”、“幽”、“爽”、“办”、“夾”、“來”等字,它们的部首只好看成上下为同一个字,分别取“禾”、“山”、“山”、“大”、“力”、“大”、“木”,特点是这些字与其它部件都不相交。经过从一般到特殊的处理,这样就实现了部首的统一性。

最后指出几个字与基本笔画交叉的情况,如“九”、“及”、“及”、“本”、“必”等字。它们的部首分别取作“ノ”、“丿”、“丿”、“木”、“心”部。综合上述这一系列的处理,我们就完全实现了非左右结构字部首的统一性原则。

第四节 选取部首办法

在实现汉字部首统一性以后,查字典部首还是按照原来的办法

不变,即按部首笔画多少对号入座去查找。每个字取部首时要按最大结构优先原则,才能保证每个字部首的唯一性。如“焦”取“隹”部,就不能取“亻”部;“寿”取“耂”部,而不能取“土”或“十”部。这里的最大结构一般都可按笔画多少去对号入座查找,但有的字还有一个如何切分的问题,如“武”字,部首是“一”还是“二”;“栽”字,部首是“十”还是“土”;等等。要分析这个问题,首先要看我们所选的部首有多少,是否包含这些结构。第二要看我们所定的拆字原则,若规定字上方相交笔画不可拆的话,则“武”只能取部首“一”,“栽”只能取部首“十”;若规定字上方相交笔画可拆,则“武”只能取部首“二”,“栽”只能取部首“土”。根据上节规定“夫”拆成“土”、“人”的原则来看,我们已规定字上方相交笔画不可拆,所以“武”只能取部首“一”,“栽”只能取部首“十”。还有字典上“甲”、“由”、“申”、“电”四字,同时在“丨”(按首笔画)与“田”(按最大结构)两部首内出现。现在按部首统一性来处理,“甲”应取“田”或“口”部首,“由”、“申”、“电”应取“丨”部首。另有“尹”可以在字典的“丿”部和“尸”部内查到,现在按部首统一性处理,“尹”应取“丿”部首。上面处理的前提是把封闭框当作独立单元来对待。当然,具体如何选择部首和如何规定分解原则,以后章节中还要详细论述。

最后把“夫、未、末、来、寿”与“甲、由、申、电、尹”两组字的部首统一起来看,按字上方或左上方取最大结构为部首,就有字上方交叉单笔画与封闭框都不能拆分的要求。这个规定很重要,它既是区分与其它编码的不同点,也是我们构造拆字的基础之一。

第二章 汉字的结构规律

传说自炎黄上古时代仓颉造字以来,汉字已有悠悠 5000 年历史,是世界文字中唯一的“活化石”。汉字在秦以前称为古文字阶段。从公元前 3 世纪秦统一文字,实现了书同文,称为篆书,经过隶书,草书到楷书的演变,直到今天。经历几千年,尽管汉字在形体上发生了很大变化,然而,汉字的演变始终遵循了方块字形和偏旁部首的发展道路,而没有向字母文字转变,充分反映了汉字的博大精深和旺盛的生命力,表明了古老汉字能够不断改革更新,永远独立于世界民族文化之林的气概。我们在这章中,通过对汉字结构的研究,进一步学习汉字造字的历史知识,作为研究汉字结构的基础,同时也为拆字原则提供依据。

第一节 汉字的基本笔画

基本笔画是汉字的基础。在一般字典上汉字基本笔画也都是部首,如常用的新华字典的前五个部首:点(丶)、横(一)、竖(丨)、撇(丿)、折(乙,包括乚、乚等笔形)。它们除“一”、“乙”又是单笔字以外,都是基本笔画。按照过去传统的说法,基本笔画是构成字的最小单元,但由于方块汉字的笔画形态多变,再加上简化字引入的一笔多折,根据上海交通大学出版社出版的《写字式汉字输入法》所统计的基本笔画就有 36 种之多。这样看来,五种基本笔画的划分就显得名不符实,很有重新划分的必要了。

在儿时写字中还有两种基本笔画:提(㇀亦称挑)与捺(㇏)。现在一般都把“提”笔归纳到“横”笔内。这是可以接受的,因为它们的运笔走向基本上是一致的。还有人把“捺”、“点”归于一类,这是不可接受的。其一,“捺”与“点”是两个大类的笔画,如“捺”有平捺、斜捺、

竖捺之分,点有左右、长短之分。其二,根据部首统一性原则,如“曳”字只能取捺部,“之”字只能取点部,这两部如果混淆起来,人们是难以接受的。

对于基本笔画“折”的情况就比较复杂了,它包含“折”的数量与“折”的方向,具体说来有四种:左折(乚、ㄣ、ㄤ),右折(ㄥ、ㄣ、ㄤ、ㄣ、ㄣ),钩(亅、ㄣ),多折(ㄣ、ㄣ、ㄣ、ㄣ)。当然还可以细分,但很难掌握原则区别。提醒注意,这里的多折有一个特点,不但折的数量多,而且折的方向变。字典上作为折的代表字“乙”也是属于“多折”的范畴,但它是个真正的方块字。

综上所述,基本笔画可分为两类:

一笔不转折:横、竖、点、撇、捺。

一笔又转折:左折、右折、钩、多折。

以上两类我们分别称为:基本笔画Ⅰ,基本笔画Ⅱ。

这样分析之后,我们坚持了汉字的最小单元是基本笔画的传统,又给基本笔画进行了科学的分类。以后我们简称基本笔画Ⅰ为单笔画,基本笔画Ⅱ为折与钩。

第二节 汉字的结构

汉字的方块结构是区别于字母文字的基本特征。以字母构成的文字只有两层结构,单独字母构成线性的单词,而线性的单词只有长短的区别和从左到右的排列,我们称其一维文字。而汉字除“一”既表示横又表示字外,所有的汉字都是方块结构的,我们称它为二维文字。它有如下三层结构:

表层结构(整字)

汉字结构 { 中层结构(部件)

底层结构(笔画)

这里底层结构的笔画就是指基本笔画,包括单笔画与折、钩,这在第一节中已分析过了。

“史”在表层结构整字分析之前，我们先学点有关造字的知识。根据语言学家研究，东汉许慎在《说文解字》中归纳出汉字的造字原理是以“六书”为依据的，即象形、指事、会意、形声、转注、假借。其中又以形声字最多，约占80%左右。形声字一旁表音，一旁表义，如“讠”、“讽”、“论”、“说”；“价”、“值”、“仟”、“亿”；“江”、“河”、“湖”、“泊”；“忙”、“忤”、“惊”、“慌”；“迷”、“逃”、“选”、“逊”；“抓”、“拍”、“拾”、“指”；“坦”、“坂”、“城”、“坑”；“吵”、“哄”、“咬”、“吓”；“岭”、“峰”、“峭”、“峻”；“姐”、“妹”、“姑”、“娘”；“驷”、“驹”、“驻”、“骠”；“杭”、“校”、“标”、“桶”；“衬”、“裤”、“袍”、“袖”；“旺”、“昭”、“晚”、“晖”；“肝”、“胆”、“肚”、“肠”。以上都是左旁表义，右旁表音。还有一些左旁表音，右旁表义的字，如“功”、“和”、“救”、“朗”等。指事字如“上”、“中”、“下”及“一”到“十”的10个数字都属于这一类，它们的组成全然属于符号的性质。如“上”、“中”、“下”中的一长横或一长竖都看作一基线来理解，就明白这三个字的意义了。象形字与会意字在第一章中已介绍过。最后两种造字法（转注与假借）为数很少，说法也不统一，就略去不说了。总之，要了解汉字体系和历史，学点六书知识是必要的。但六书毕竟是汉字发展过程中的产物，就是说先有汉字，然后有六书，不是先有六书后有汉字，而且汉字所经过的历史长，所通行的地域广，已经形成一个复杂而奥妙的文字体系，所以企图说明这个体系的六书造字方法也就更加复杂奥妙，众说纷云。然而下面结论是肯定的，汉字的字形是从象形、指事、会意、形声等发展起来的。一般来说，象形、指事两者为独体字，会意、形声两者为合体字。时代越古远，它的形音义的联系越紧密。经过几千年发展变化，特别在隶变楷化、创造新字之后，形音义脱节的字就越来越多了。但是，从历史上和总体上看，汉字仍是形音义的统一体。

通过上面对造字的分析，一般地讲，每个汉字都有形、有音、有义，也就是说汉字结构层次中的表层结构（整字）是使用单位。有的字本身就作词或句使用，这在汉字信息字典中有大量的材料。但也有的字有形、有音，没有意义，如“蚯”、“蚓”；“垃”、“圾”；“咖”、“啡”；“囫”、

“囧”、“志”、“忑”等一类字,大约有 560 个,它们只能两两组合在一起才有意义。

第三节 方块结构的分析

一维线段只有左右两个方向,二维方块就有左右、上下、左上、左下、右上、右下等八个方向,还有内外包含问题。

从上节知道,形声字一旁表音,一旁表义,字数最多,约占总字数的 80%,也就是说左右结构的汉字最多。对于左中右结构的汉字,可看成二次左右结构,如

湖	[古 月
树	[木 对 又 寸

二次以上左右结构的字,也可照此处理。

对非左右结构的汉字还可以分为上下结构、包围结构、混合结构三种。现分述如下:

上下结构:分列上下(或左上、右下有一为基本笔画)两个部分有明显的界限,如“圣”、“古”、“穴”、“夕”、“么”、“幺”、“文”、“各”、“上”、“下”、“夫”等字,我们称为上下夫形连接,用 F 表示。

上下结构还包括上下有两面或三面互被包围的字,如“凶”、“义”、“闭”、“太”、“介”、“公”、“凡”、“右”、“歹”、“历”、“冂”、“反”、“庆”、“方”、“尿”、“后”、“老”、“氏”、“丑”、“刃”、“寿”、“痒”、“旬”、“见”、“习”、“乃”等字,我们称为上下间连,用 J 表示。

以上所举例字仅限于只有上下两个部件的汉字。因为多于两个部件就难于分上下了。

包围结构:这里指的是全包围结构,就是说汉字的四面都有里外,称为回形结构,用 H 表示。如“回”、“国”、“囿”、“囫”等字。

混合结构:凡无法归纳于上述结构的字都归在这类。如上章讲的嵌套结构。对由三部分组成的字(指非左右结构),我们只指出三种结

构形式,一是如“晶”、“花”、“众”、“最”等字,我们称为d形结构。二是如“想”、“盟”、“坐”、“丛”等字,我们称为g形结构。上述两种结构特点是由三个部件组成的三角形或倒三角形的汉字。三是如“苦”、“菜”、“宁”、“劳”、“弓”等字,我们称为卡形结构,用K表示,其特点是由互不包含的上、中、下三个部件组成的字。

第四节 汉字的部件

汉字的中层结构由部件组成,按照汉字信息字典所说,提供该书所收 7785 个正体字的部件数为 694 个。部件由三部分组成:一是该书所收的独体字 258 个;二是由复笔构成的《辞海》部首 245 个;三是非字非《辞海》部首的复笔组字单元 191 个。前两者在第一章中已有了优选原则,至于如何具体选择部首到最少数目,这需要列出专门章节进行叙述。对 191 个非字非部首的复笔组字单元的优选问题,涉及到汉字分解的最小单元问题,如何处理,这也要作具体分析和讨论。但有一点必须做到,需要合理地选取最小最少组字单元来代替或组合成现在的部件。

这里所说的部件是汉字信息字典上的一个科学术语。它是整个汉字分解时的构字单元,常见的说法还有字根、字元、元件、码元等。为统一起见,也为充分利用汉字信息字典上的研究成果,我们把汉字中层结构单元统一称为部件。现在的问题是如何根据我们的选择原则,从 694 个部件中选取 100 多个部件,科学完整地建立起汉字结构大厦,完成 20000 多个汉字的查字与编码的统一性,实现汉字的一次查找和电脑的键盘输入。

第三章 电脑汉字分解原则

自从 20 世纪 40 年代出现第一台电脑以来,人类开始进入了全新的信息时代,电脑以其飞跃发展和神奇功能给人类社会带来了空前的进步与文明。语言文字与电脑的结合,是现代社会信息化的一个重要特征,也是人们掌握使用电脑的一个重要前提。由于电脑首先是外国人发明的,对使用字母文字的人们来说是不存在问题的,而对使用方块图形文字的我们来说是一个严峻的挑战。如何使博大精深的古老汉字能够享受神奇智能的全新电脑技术,关键是电脑汉字输入问题,也就是所谓汉字编码问题,而其前提又是电脑汉字分解问题。本章就专门论述电脑汉字分解的原则,它涉及到工程学、社会学、文字学、教育学等各个方面。

第一节 从电脑键盘谈起

电脑键盘是在外文打字机的基础上发明的。键盘上安排有英文字母与数字符号等键位,有很强的操作性。键盘输入是电脑输入的一种重要方式,具体如下:

1	2	3	4	5	6	7	8	9	0
Q	W	E	R	T	Y	U	I	O	P
A	S	D	F	G	H	J	K	L	
Z	X	C	V	B	N	M			

键盘上有 26 个字母,和我们采用的拉丁字母拼音方案是对应的,只是拼音中规定 V 只用来拼写外来语、少数民族语言和方言。这样一来,如果我们采用拼音方案输入好像没有问题,其实不然。因为汉字的同音字太多,加上不认识的字,拼音也不知道,就无法输入,再