

新一代拼音输入法

——扩展拼音



汤南华 编著

同济大学出版社

内容提要

本书介绍扩展拼音输入法的特点和扩展拼音输入法的编码规则；系统的安装、运行；如何使用练习程序；重点介绍了扩展拼音在兼容普通拼音的基础上所扩展的汉字体系固有的七种基本字型和组字词组等功能。

扩展拼音输入法学习简单，不用记忆众多繁琐的字根，只要有一点汉语拼音基础即可，非常适合办公人员、行政管理人员，财会人员和其他科技人员使用。

责任编辑 胡兆民

封面设计 潘向葵

新一代拼音输入法

——扩展拼音

汤南华 编著

同济大学出版社出版

(上海市四平路 1239 号 邮编 200092)

青浦任屯印刷厂印刷

新华书店上海发行所发行

开本：850×1168 1/32 印张：5 字数：140 千字
1998 年 5 月第 1 版 1998 年 5 月第 1 次印刷

印数：1-6000 定价：8.00 元

ISBN7-5608-1892-7/TP·203

目 录

第一章 拼音输入法功能的扩展	(1)
§ 1.1 概述	(1)
1.1.1 普通拼音的缺点	(1)
1.1.2 扩展拼音的主要特点	(1)
1.1.3 组字词组的概念和用途	(2)
1.1.4 扩展拼音的动态词组及单键码的实现	(5)
1.1.5 动态词组单键码的编码样本	(6)
§ 1.2 扩展拼音输入系统编码规则	(10)
1.2.1 扩展拼音的五条基本编码规则	(10)
1.2.2 不同层次用户的输入样本	(13)
§ 1.3 扩展拼音和标准拼音的区别	(16)
1.3.1 初学者使用扩展拼音时的要点	(16)
1.3.2 用扩展拼音作为字典查询时的要点	(19)
1.3.3 提高输入速度的要点	(20)
第二章 扩展拼音的安装和启动	(23)
§ 2.1 系统文件的组成	(23)
2.1.1 程序清单	(23)
2.1.2 系统安装过程	(24)
2.1.3 系统的启动	(24)
§ 2.2 系统的运行环境	(25)
§ 2.3 练习程序的启动和使用	(26)
2.3.1 组字词及推荐码演示练习程序	(26)
2.3.2 型声码演示及电子字典	(27)
第三章 汉字体系的特点	(29)

§ 4.4	汉字的字型键盘图	(60)
4.4.1	字型键盘图说明	(60)
4.4.2	型声码字型键盘图	(63)
4.4.3	汉字的体系结构	(65)
第五章	扩展拼音输入系统使用细则	(66)
§ 5.1	系统概况	(66)
5.1.1	使用规则速查	(66)
5.1.2	特殊键说明	(67)
5.1.3	编码样本	(68)
§ 5.2	输入方式说明	(69)
5.2.1	简拼说明	(69)
5.2.2	全拼说明	(72)
5.2.3	双拼说明	(73)
5.2.4	三拼说明	(74)
5.2.5	声型码说明	(75)
5.2.6	型声码说明	(76)
5.2.7	二拼说明	(76)
附录一	偏旁部首名称表(笔划,字型,笔顺序)	(78)
附录二	部分汉字编码摘录	(83)
附录三	汉字的扩展拼音序(字型拼音序)	(93)
附录四	汉字的字型笔划序	(135)

第一章 拼音输入法功能的扩展

§ 1.1 概述

1.1.1 普通拼音的缺点

汉语拼音是国家制定的汉字读音的规范;汉语拼音输入法是目前使用最广泛的汉字输入法;这是谁都承认的事实。一个好的输入法应该符合国家文字规范,与文化基础教育相一致,汉字输入编码应该和字典查询码相一致……,这也是大家对汉字输入编码的希望。汉语拼音完全符合这些条件,可是新的汉字输入法依然不断涌现,目前光是见报的已达数百种之多,这是为什么呢?因为普通的拼音输入法至少有三大缺点:

1. 不认识的汉字无法用普通拼音输入(因为不知它的读音);
2. 相同读音的汉字在普通拼音中实在太多(重码多、不能盲打、影响输入速度);
3. 读音不准的汉字很难用普通拼音输入(平舌音,卷舌音;前鼻音,后鼻音分不清);

1.1.2 扩展拼音的主要特点

那么有没有一种编码规则既能兼容普通拼音,又能解决普通拼音的这些缺点呢?回答是肯定的。那就是我们在这儿要介绍的扩展拼音系列码编码规则。

扩展拼音系列码的编码规则可简单地归纳为下面三条:

1. 单个汉字按声、韵、型、调的次序顺序进行编码,其中的韵、型、调可分别省略或全部省略(如果仅仅省略“型”和“调”就是普通拼音)。

如:日:RI 得:DE 的:D 地:DI

2. 多字词的编码按每个单字的编码顺序连接而成(多字词的概念包括字数不限的常用词和字数限制在三个以内的特定的“组字词”)。

如:开放 → KAIFANG → KAIF → KF…(常用词组)

组 → 组彡且 → ZUJQ → ZJQ…(组字词组)

3. 卷舌音(ZH, CH, SH)中的 H 可以省略;后鼻音拼写符号中的 G 也可以省略。也即 ZH, CH, SH 可用 Z, C, S 代替;ANG, ENG, ING 可用 AN, EN, IN 代替。

如:生命 → SHENGMING \ → SENGMING → SENMIN…

上面规则 1 中的字型概念读者可看《扩展拼音使用规则》,这儿暂不介绍,读者只需简单地将其在编码中省略即可。这儿让我们先来解释一下“组字词”的概念,然后看一下扩展拼音是如何解决普通拼音的三大缺点的。

1.1.3 组字词组的概念和用途

我们知道,绝大多数的汉字都是可以很方便地一拆为二的。我们将该汉字连同它一拆为二而形成的偏旁字统称为该汉字的“组字词组”。

譬如:绝 → 绝彡色 → JUEJIAOSE,多 → 多夕夕 → DUOXIXI…等。其中“绝彡”和“绝彡色”都是“绝”字的“组字词组”,而“多夕”和“多夕夕”则是“多”字的“组字词组”。扩展拼音码允许将一个汉字的“组字词组”作为它输入时的特定编码,从而达到输入该汉字的目的。要注意的是,在利用组字词组的概念输入汉字时,我们虽然输入了两个或三个汉字的拼音,但后两个汉字的拼音是用来修饰指定第一个汉字的,是第一个汉字的“组字词”。也就是说,

是为了解决普通拼音的重码而额外增加的编码。因此,当我们利用“组字词”的概念输入汉字时,虽然输入了两个或三个汉字的拼音(三个汉字的扩展拼音编码一般只需三个声母即可),但最终只转换成一个汉字,而不像常用词组那样,有多少个汉字的拼音编码就转换成多少个汉字。

譬如:码 → 码石马 → MASIMA(MASHM, MASM, MSM…) → 码;

输 → 输车俞 → SUCEYU(SHUCY, SHCHY, SCY…) → 输;

一、不认识的汉字用万能声母键“/”输入

为了用扩展拼音输入不认识的汉字,我们还得引进一个万能声母符号“/”。因为不认识的汉字一般连它的声母也是不知道的,这时我们只要用万能声母符“/”来表示它的声母,然后省略它的韵母、字型、和声调即可。

譬如:魑 → 魑鬼离 → /GUI LI → 魑;

骹 → 骹骨介 → /GU JIE → 骹;

从上面的例子我们可以很清楚地看到,只要掌握了组字词组的概念,我们就有可能利用扩展拼音来输入不认识的汉字了。当然,要熟练地使用“组字词组”来输入汉字,还需熟悉它们的“组字词”以及常用偏旁的读音。这些都可以在系统编码表里查到(用户可参见附录一及附录二的部分汉字编码摘录)。

更加具体的使用方法可参看后面 1.3.2 怎样用扩展拼音作为字典查询软件中关于冷僻字输入的方法,这儿就不具体介绍了。

二、增加编码长度自然就没有重码了

下面我们再来介绍一下,怎样利用扩展拼音中的组字词组概念来解决普通拼音的重码问题。

我们知道,普通拼音对每个汉字一般只能输入声母和韵母两个信息(声调一般不用)。在汉语中,声母和韵母的各种组合若不包括声调的话,总共不过 400 种左右,而常用的汉字一般有二三千个。这样,普通拼音的重码自然就多了。

而在扩展拼音里,一个汉字可以用它的组字词组来编码,这时

输入的信息量就大大增加了,自然就不会有重码了。但另一方面,如果输入的信息量太多的话,编码长度就太长了。因此我们在用扩展拼音编码时,一方面要利用它的组字词组信息量大的优势,增加输入的信息量;但另一方面又要利用扩展拼音可以省略韵、型、调的规则,来缩短编码长度。

譬如:码 → 码石马 → MASIM → 码;

输 → 输车俞 → SUCE → 输;……

在扩展拼音系统的中文输入状态中,每按一个键,在提示行里都会有相应的中文字或词组显示。编码长度越长,则提示行里的重码就越少。像上面的“码”,当我们输入第一个编码 M 时,提示行里将显示 9 个常用的重码字,而事实上如果用翻页键 0 继续查看的话,我们将发现有 300 多个重码字呢。当我们继续输入“码”字的编码时,我们就会发现,在编码输入到 MASIM 时,提示行里就已经没有重码了,因此后面的编码就没有必要再输入了。

同样,“输”的编码在输入到 SUCE 时,就已经没有重码了。从这两个例子我们就可以看到,各个汉字的重码分布情况是各不相同的。如果我们统一采用普通拼音的“声韵 声韵 声韵”的次序来编码输入的话,就将导致编码长度很长,而最后的编码又往往是多余的。为了达到既简单易记,又能提高编码效率的目的,一般我们按照“声韵 声声”的规则进行编码,这就是附录二,扩展拼音编码表里的推荐码。因此,上面两个汉字的推荐码就是:

码 → MASM ; 输 → SUCY ;

三、读音不准可用模糊输入法

最后,对于读不准音的汉字,只要按照上面所讲的第三条规则办就是了。事实上,对于吃不准读音的汉字,我们也可以省略它的整个声母(用万能声母键/代替)或韵母,因为这本来就是扩展拼音里的最基本的规则。

譬如:神仙 (SHENXIAN) → SENXIAN

风车 (FENGCHE) → FENCE …

至此,我们可以看到,扩展拼音在兼容普通拼音的基础上圆满地解决了普通拼音在输入单个汉字时的三大缺点。当然,具体的细节,我们还要参照下文和系统的编码表。

1.1.4 扩展拼音的动态词组及单键码的实现

在概述中我们已经介绍了扩展拼音的基本功能,并着重介绍了它利用“组字词组”的概念来解决不认识的汉字无法用拼音输入的方法。这儿我们要进一步介绍扩展拼音输入系统中的动态词组的概念。这是扩展拼音在输入汉字时超出普通拼音的又一个重要功能。用户在输入大量文字材料时,就更能体会到它的方便和实用。

我们知道,不同的汉字在使用中出现的频率是很不一样的。譬如最常用的汉字“的”字,它的使用频度就远远大于任何其他汉字(约为普通汉字使用频率的 100 倍以上)。一般的汉字输入系统都将最常用的汉字放在提示行的最前面,以提高输入速度。如:的、一、是、为……等。但在实际使用中,某些原先并不常用的汉字有可能在某篇文章中却用得特别多,这就需要临时将这些汉字作为常用字来处理。譬如,在我们这篇文章中,“字”这个一般很少遇到的字就变成了常用字,这就需要有用常用字的动态定义功能。

一、单个汉字输入时重码显示的动态调整

在扩展拼音输入系统中,设置了单字重码提示行显示的动态调整功能,可以由系统自动调整,也可手动调整。刚启动扩展拼音输入系统时,系统的默认状态是自动调整状态,可以用 F5 键进行切换。

对于一般用户而言,输入时并不需要高速盲打,因此,系统总是将用户已经用过的单字显示在提示行的前几个,避免了频繁使用翻页键查找的麻烦。对于需要高速盲打的专业输入人员来说,由于自动调整功能经常改变单字重码显示的前后次序,因而反倒显得不方便,此时,可用 F5 键,将重码显示的调整功能切换到手

动定义状态。

常用字的手动定义也是非常方便的,用户只需在输入这个汉字的编码时,不要省略该字的字型编码,则以后该汉字就被定义成常用汉字了。比如,当我们发现,“字”应该定义成常用字时,只需在碰到该字时,在编码中包含它的字型信息,即用编码 ZI② 来输入它,其中 ② 表示该汉字是上下型汉字(字型编码的输入方法是按住 SHFT 键后,再按相应的数字键)。以后当你要输入“字”这个字时,只要键入:ZI 则“字”就将出现在提示行的第一个了。

二、多字词组的动态定义

单个汉字提示行显示的动态调整功能虽然很实用,但单个汉字的输入效率毕竟是不高的。因此在实用中,我们总是尽量利用多个汉字形成的词组来输入汉字。但汉语的词组数量非常大,一般的汉字输入系统中只能收入一些最常用的词组,而大量的非通用性词组就只能一个一个地用单个汉字来输入了。

为了解决这个问题,扩展拼音输入系统还专门开辟了一个动态词组表,专门用来记录最新输入的汉字词组。这样,在用户输入时不管遇到了什么词组,如人名、地名等,只要输入一次后,就被自动地记录在系统的动态词组表里了。以后再遇到这些人名,地名时,就可很方便地按扩展拼音中最简短的编码法(即每个汉字只输入一个声母)来输入它们了。

1.1.5 动态词组单键码的编码样本

下面就是利用扩展拼音的动态定义功能进行编码输入的样本。编码样本中第一行编码是刚使用扩展拼音的人用全拼输入法进行编码输入的样本;第二行则是熟练后,记住扩展拼音的几个简化键后简拼输入法的编码样本(主要就是提示行中的 6 个鼻音简化键);第三行则是重复输入该样本时,利用动态词组功能的一字一键编码输入。为清晰起见,有些用数字选择键输入的汉字,在编码样本中用空格来代替了。

国家教育委员会
GUOJIA JIAOYU WYH
GUOJIA JIYU WYH
GJJYWYH

国家语言文字工作委员会
GJ YUYAN WENZI GONGZUO WYH
GJ YUY[W;ZI G,ZUO WYH
GJYYWZ GZWYH

关于公布《汉语拼音正词法基本规则》
GUANYU GONGBU 《HANYU PINYI ZENHUA CYS FSQ JIBEN GUIZE 》
GU[YU G, BU 《H[YU P'Y Z;HZ CYS FSQ JIB;GUIZE 》
GYGB《H[YPYZCF JIBGZ 》

的联合通知
D LIANHE TONGZI
D LI[HE T,ZI
D LI[HTZ

(1988年7月1日)
(1988NIAN 7YUE \ 1RI)
(1988NI[7YUE \ 1RI)

《汉语拼音方案》自1985年公布以来，
《HYPY FVAN 》ZI \ 1985NIAN GONGBU YI \ L,
《HYPY FV[》ZI \ 1985NI[G, BU YI \ L,
《H[YPYFV 》ZI \ 1985 NI[G, BYL;

在许多领域得到了应用。

Z XUDUO LINYU DEDAO L YIN \ Y。

Z XUDUO L'YU DEDAO L Y' \ Y。

Z XUDLY DEDLYY。

随着汉语拼音应用范围的日益扩大，

SUIZ HYPY YY FW D RIYI KUOD,

SUIZ HYPY YY FW D RITI KUOD,

SZHYPY Y'YFW D RIYKD,

文教、出版、信息处理及其他部门要求

WENJ、CUBAN、XINXI CULI JI'NN QITA BUM YAOQ

W;J、CUB[、X'XI CULI JI'NN QITA BUM YAOQ

W;J、CUB、X'XCLJ QITBM YAOQ

制订汉语拼音正词法的基本规则，

ZIDIN HYPYZCF D JBGZ,

ZID'HYPYZCF D JBGZ,

ZID HYPYZCF D JBGZ,

作为用汉语拼音方案拼写现代汉语的规范。

ZUOWEI YONG HYPYFV PINTI XIEPY XIAND HY D GUIF。

ZUOWEI Y, HYPYFV P'TI XIEPY XI[D HY D GUIF。

ZW Y, HYPYFV PXXDHY D GUIF。

为此，1982年3月原中国文字改革

WEIC, 1982NIAN 3YUE YUAN ZONGGUO WZ GG

WEIC, 1982NI[3YUE YU[Z, GUO WZ GG

WEIC, 1982NI[3YUE YU[ZGWZ GG

委员会决定成立由有关方面
WYH JUED CENLI YOU` YOUG FM \
WYH JUED C; LI YOU` YOUG FM
WYH JDCL YOUYGFM

的专家,学者组成的汉语拼音正词法委员会,
D ZUANJ, XUEZ ZUC D HYPYZCF WYH,
D ZU[J, XUEZ ZUC D HYPYZCF WYH,
D ZU[J, XUEZZCD HYPYZCF WYH,

拟订汉语拼音正词法基本规则;经多次
NID HYPYZCF JBGZ ; JINJL DUOC
NID HYPYZCF JBGZ ; J'JL DUOC
NID HYPYZCF JIBGZ ; J'DC

征求意见和讨论,修改,于1984年10月发表了
ZENQ YI \ J H TAOL, XIUG, YU1984NIAN10YUE FB L
Z; Q YI \ J H TAOL, XIUG, YU1984NI[10YUE FB L
ZQYJHTL, XIUG, YU 1984NI[10YUE FABL

《汉语拼音正词法基本规则(试用稿)》。
《HYPYZCF JBGZ (SIYONG GAO`)》。
《HYPYZCF JBGZ (SIY, GAO`)》。
《HPYZCF JBGZ (SYG)》。

...

从上面的样本中,我们可以看到,刚开始输入汉字时,编码长

度是比较长的,因为系统的动态词组表还是空的。那些第一次出现的不常用词组需要一个一个地用单个汉字编码的方法输入。譬如标题中的“正词法”的编码就是“ZENHZ CYS FSQ”,它们就是利用系统中的组字词组库“正一止 词 司 法 去”来编码输入的。其中的“正一止”中的“一”读音为“HENG”而不是“YI”,因为它是组字词中的笔划名称而不是单字。

又如标题中的“汉语拼音”的编码是“HANYU PINY”,它们是利用系统中现有的静态词库(常用词词库)来编码输入的。在输入的同时,这些词汇也被记入系统的动态词库里去了。这样,随着动态词组表中词汇量的增加,那些重复出现的长词汇就不再需要用单个汉字编码的方法来输入了。

像“汉语拼音”,“汉语拼音正词法”,“基本规则”等词汇在文章中多次出现,我们在第二次或以后碰到它们时,就只需输入这些词组中的每个汉字的声母了。比如,编码样本中最后两行的编码,就简单地将这些词组按每个字的声母来输入了。如,“HYPY”,“HYPYZCF”,“JBCZ”等。当然,如果出现重码时,就应再另外增加一些编码信息,如韵母、声调等,以区别重码。

在动态词组表中,词组的最大长度为7个字,超过7个字的话,我们可以分两次进行输入。

从编码样本中,我们可以清楚地看到,刚开始时,每个汉字的编码长度是比较长的,平均每个汉字的编码长度超过3键。而到了最后两行,则平均每个汉字的编码长度就不到2键了。特别是在重复输入时,每个汉字的编码长度基本上为一字一键了。

§ 1.2 扩展拼音输入系统编码规则

1.2.1 扩展拼音的五条基本编码规则

1. 每个汉字用大写字母顺序输入声母、韵母、字型、声调四个部分。其中韵母、字型和声调也可以分别或全部省略(四声用 - = ` \ 来表示)。通过这条规则可以实现对普通拼音的简化或压缩,故也可称为“简拼”,或“压缩拼音”。

2. 在扩展拼音中将全部汉字归纳为合体字和独体字两类,共7种基本结构:

其中合体字有4种基本字型:左右(改),上下(各),内外(国),对角(达)可分别用编码1,2,3,4来表示。

独体字有三种基本字型,它们是:附着(犬),交叉(本),粘连(巴)可分别用6,7,8来表示;加上偏旁部首中的合聚型用9来表示,它们又统称为独体结构,可统一用5来表示。字型的具体定义见§4.3汉字的字形信息。

如:计:JI1 - 算:SUAN2 \ 国:GUO3 = 这:ZHE4 \
乖:GUA16 - 子:ZI7 ` 山:SAN8 - 纟:JIAO9 =

在合体字的4种基本结构中,又可细化出许多复合字型。在扩展拼音中,目前最多可连着用三个基本字型编码来表示一个汉字的复合字型。

如:蔚:W271 靴:X121 - 烫:T216 骸:HU122
髁:KUAN122 -

【注】第一个“蔚”字的复合字型编码W271中,2表示“蔚”是上下型汉字;7表示上面的“艹”是交叉结构;1则表示“尉”是左右结构;其余的字可类推。详细解释请看4.3.3汉字的复合字型。另外,在后面的章节中,我们将用符号①~⑨来表示字型,以区别于数字键1~9。

3. 任何汉字的声母都可用万能声母键/来表示。

如:蔚:/EI271 烫:/ANG216 断:/148/J

又如:魑鬼离:/GUILI 骸骨介:/GUJIE 贱贝戈:JIANB/

4. 常用词组中每个单字都可按上述方式进行编码,并按字顺序输入。

如:改革→GAI1 = GE2 = →GAI\ GE = →GAIGE→GG→GGE→
 扩展拼音 标准拼音 普通拼音 压缩拼音
 GGE2 = →GGE2… →G182G277…
 压缩拼音…

在扩展拼音中,上面的各种编码都是双字词“改革”的合法的输入编码,它们也是采用以“常用词组”定字时,“改”字的合法编码,只不过此时应利用 tab 键而不是空格键来取字。

5. 每个汉字都可一拆为二,按先左后右,先上后下,先外后内,先大后小的次序,将所得的偏旁字附在原汉字后面,从而得到该字的“组字词组”。“组字词组”的编码形式和“常用词组”相同,但它是扩展拼音中对单个汉字的扩展编码,组字词组的编码长度可以短到三个,也可长到几十个,但无论编码多长都只代表了一个汉字。

如:男→男田力→NAN257TIANLI→NANTIANLI→
 组字词 扩展拼音 普通拼音
 NTL→NTLI→…
 压缩拼音…

又如:畜→畜帝口→/DI2 \ KOU8 = →/DI \ KOU =
 组字词 扩展拼音 标准拼音
 →/DIKOU →CDK…
 普通拼音 压缩拼音

通俗地说,所谓扩展拼音码就是这么一句话:允许以多字词为手段(包括常用词组和“组字词组”)来进行汉字的编码,其中每个单字可以按“声、韵、型、调”的次序顺序输入,也可将“韵”、“型”或“调”分别省略,或者全部省略。

举例来说,要输入“改”字,则可选用下面成千上万个编码中的任何一个:

GAI\ (- = \ 代表声调)

GAI1\ GAI18 GAI185 G185\… (“改”字的字型编码)