



上载日期: 2005-09-29

## 十码汉字输入法

高永贤 [gaoyx3721@126.com](mailto:gaoyx3721@126.com)

### 简介

汉字是象形表意文字，结构复杂，偏旁部首多，同音字也多。在《新华字典》中，有129个一至四声的同音字：“yi”。中国的方言差异很大，不懂普通话的人有很多。摆在拼音智能输入面前的难题是，使用者拼不准拼音。钱伟长教授曾说：“好的汉字编码还没出来。好的汉字编码应该是一种文字，或者是一种准文字。”

十码汉字输入法基本符合这一要求。它是一种字形码，以国家规定的五种基本笔画为基础，根据汉字部件中起笔笔画分布不均匀，笔形是动态变化的特点，把汉字的笔画、笔画组合、偏旁部首分成十类，对汉字进行拆分和编码。有简便易学易记的特点。如：“一七、二十、三、八、九”七个数字是以横起笔、以撇起笔的五组分类的特征；汉字的编码短，码长1—4个；找字容易，重码大多在4个以下；而且可以对词组编码。只要会写字，就能很快学会打字。很好地解决了人们用计算机以后对汉字日益生疏的问题，可以说“十码汉字输入法”找到了中文输入数字化之门的钥匙，可能成为“中文信息处理技术”的突破口。

26个字母键的汉字输入法都没解决易学易记的问题，十码汉字输入法行吗？

第一，“十码汉字输入法”尽量让汉字编码平均分配，少出现空码。按理想模式，每码一个字，输入3个码时，T9五个笔画输入3个码的个数为 $5 \times 5 \times 5 = 75$ ，空码19个，空码率近25%；T9拼音八个笔画输入3个码的个数为 $8 \times 8 \times 8 = 512$ ，空码406个，空码率近80%；“十码汉字输入法”十个笔画输入3个码的个数为 $10 \times 10 \times 10 = 1000$ ，对《新华字典》中的7693个字编码，空码31个，空码率近3%。（PC键盘容量太大，没法统计比较，故用数字键盘的统计资料）

第二，按理想模式，每码一个字，十码汉字输入法

输入1个码时， $10 \times 1 = 10$

输入2个码时， $10 \times 10 \times 1 = 100$

输入3个码时， $10 \times 10 \times 10 \times 1 = 1000$

输入4个码时， $10 \times 10 \times 10 \times 10 \times 1 = 10000$

用10个数字编码，码长1—4，每码一个字，字库的最大容量为11110字，每码二个字，字库的最大容量为22220字，每码三个字，字库的最大容量为33330字……国标一级字库有3755字，国标二级字库有3008字，共6763个字。《新华字典》中的简体字、繁体字总共才近万字，即使一半以上的码是空的，用10个数字给汉字编码也是可行的。

“十码汉字输入法”用同一分类表对繁体字、简体字编码。不须死记硬背，记住第一笔与第二笔的结构，多用几次就能记住。十组分类按顺序归纳成一句话“一七、二十、三厂、4口、5日、6一点、7多点（二、三、四点）、八一撇、九多撇（两撇、三撇）、0折”。十码汉字输入法使汉字实现有序化成为了现实。

# 说 明 书

本发明属于计算机中文信息处理领域。随着科技的进步，信息时代的到来，汉字输入成为了社会信息化，数字化进程的“瓶颈”。笔画输入方便易学，但汉字是拼形文字，大多以部首开头，一定长度的部首笔画，增加了区分汉字的码长，降低了汉字的输入速度。本发明针对数字设备上输入汉字速度上存在的不足，提供一种输入速度有些改进的数字化汉字输入方案。

汉字的造字历史延续了几千年，经历了从甲骨文到仿宋体的演变；繁体字到简化字的演变；有篆、隶、草、行、楷等多种书体；说明汉字也有一个不断发展完善的过程。构成汉字的笔画有几十种，而且是有规律的。在楷书中，笔画分为八种基本类型：点、横、竖、撇、捺、钩、折、提（挑），每个基本类型又分出许多变化形态。如“点”就有“左点、右点、长点、短点”等。国家语委制定的笔画规范规定：汉字有“横、竖、撇、点、折”五种基本笔画，其中规定“提”放“横”中；“竖钩”放“竖”中；“捺”放“点”中；横向钩、右向的各种钩归于“折”中。

“一、二、三、四、五、六、七、八、九、十”十个数字，是最常用的汉字，也可能是先祖最先造的汉字。其中以横起笔的就有六个，对《新华字典》中的汉字进行分析，以横起笔汉字多，折笔的变化多。为了部件间的相互协调，字形美观，有些部件的笔形会随所处的位置不同而发生变化。如：“己”字在右边构字时，最后一笔为“乚”，在左边构字时，最后一笔为“レ”。

十码汉字输入法是一种字形码，以国家规定的五种基本笔画为基础，根据汉字部件中起笔笔画分布不均匀，笔形是动态变化的特点，把汉字的笔画、笔画组合、偏旁部首分成十类，对汉字进行拆分和编码。

## 一、 分类情况

十码输入法以起笔笔画的类型及笔画组合分类，以横起笔的分成三组，以竖、撇、点起笔的各分成两组，以折起笔的不分，共十组。记住每组的特征和附注中的注解，就能简便而快速对汉字拆分和编码。

有些笔画组合字库中没有，汉字组字时用到，可以用文字描述其结构特征。例如：“一”+“丨”可以组成多种笔画组合“十 乚 丁……”，这些组合可以用一句话概括，即“第一笔为横，第二笔为竖的笔画相连组合”。下面为具体的分类表：

键位	特征	偏旁和部首	附 注
1	一七	一七车車寸丁雨 扌匚	一横、提，“一”+“丨、折”的偏旁部首。
2	二十	二于大十士土走革 升艹 ㄥ	二横，“一”+出头的“丨ノ”的笔画组合和偏旁部首（“木”除外）。
3	三厂	三王馬厂石歹耳西木 丰厂丁 冫邑𠂇	三横、木，“一”+不出头的“丨ノ”的笔画组合和偏旁部首（“雨車”除外）。
4	凵口	口田目貝貝巾門 凵冂𠂇	二竖（包括凵），以竖起笔的有半包围和全包围框的偏旁部首（“日虫足”除外）。
5	丨日	日虫足止山 丨虍足	一竖、竖钩
6	丶	广立方言之冂 丶冫讠ネ疒	一点、捺、以“一点”起笔的笔画组合（冫冫）和偏旁部首（“心”除外）。

7	ㄚ ㄚ	米火心 ㄚ ㄚ ㄚ ㄚ ㄚ ㄚ	二点、三点，四点、心，以“二点”起笔的的偏旁部首。
8	八ノ	八人金白又 ノイ ㄚ 食 ㄚ	一撇，撇加“竖、点(捺)”的偏旁部首。
9	九厂	九儿几月匕鱼魚舟牛毛禾气 ㄚ ㄚ ㄚ ㄚ ㄚ ㄚ ㄚ ㄚ ㄚ ㄚ	二撇、三撇，撇加“横、撇、折”的笔画组合和偏旁部首(“ㄚ ㄚ”除外)。
0	折	了又女马弓尸 マ ㄚ ㄚ ㄚ ㄚ ㄚ ㄚ	除“竖钩”外的各种折笔

说明：

1、第一笔为“一”，第二笔为“丨、レ、ㄚ、ㄚ”的笔画相连组合，与“丁七”相似，拆分时当作“丁七”，编码为“1”。

2、“冉、禺”中“冂”+“丨”组合，拆分时当作“巾”，编码为“4”。

3、“脊”的最前四笔、“隶”的最后四笔，拆分时当作“四点”；编码为“7”。

4、第一笔为“ノ”，第二笔为“冂、レ、ㄚ、ㄚ”的笔画组合，与“勺儿匕”相似，拆分时当作“勺儿匕”，编码为“9”。

5、“四横”拆分成“三横”+“一横”。

## 二、 汉字的拆分和编码

在汉字集中，笔画、笔顺一样，下笔的位置、长短不同，会产生字形、字义不同的字。如：“九儿，人入八又，太犬……”，要使汉字编码无重码，几乎不可能。

汉字是由多种笔画组合而成的方块形式，对于同一汉字，各人拆分方法不同。十码输入法尽可能减少码长和重码，对汉字的拆分和编码作以下规定：

1) 按照国家规范规定的笔形标准和笔顺拆分和编码。

如：与→一ㄚ一→101

万→一冂ノ→108

义→、又→68

比→一レ匕→19 (一、レ相连当作“七”)

2) 汉字的编码长度为1—4个，分类表中的部件不拆分，码长为1个；当拆分的部件数为2—4个时，编码按顺序有多少，取多少；当拆分的部件数多于4个时，前3码按顺序取、第4个码取最后一个部件的编码。

如：春→三人日→385

佳→イ ㄚ 丰→863

堡→イ 口木土→8432

愉→ト 人冂二 ㄚ→7814

3) 有的偏旁部首在书写时不是连续笔画，除“匚、口、心”可以不拆，其它的不是连续笔画的部件都要拆分。

如：巨→匚冂一→101

国→口王、→436

必→心ノ→78

黑→口ㄚ丨二 ㄚ→4757

4) “冂、口、日、目”外粘连有“ノ”的，拆分为“ノ”加“冂、口、日、目”。

如：向→ノ冂口→844

囟→ノ口又→848

白→ノ日→85

自→ノ目→84

5) 有些字有多种拆分的方法，有的字取笔画少的部件，部件数或重码会少一些，或者更符合习惯。

规定“开→一升、未→一木、正→一止、缶→一山、亚→一川ノ一”，字的下面有“升、木、止、山”的比照拆分。

如：井→一升→12                                      末→一木→13  
戒→一升匕ノ、→1206                              朱→一木→93  
正→一止→15    武→二止匕ノ、→2506  
出→一止丨山→055                                      卸→一止丨→9155  
严→一川ノ厂→1473                                      缶→一山→915

6) “弋”的下面有“横”的，与“弋”的“一”是连续笔画，拆分时“弋”的“横”单独编码。

如：式→一丁一匕ノ、→1316                              贰→一二贝匕ノ、→1246

7) 除上面4、5、6例举的几种情况的部件外，其余的汉字按笔顺遵循“有多不取少”的拆分原则，即拆分时先取笔画多的部件，不用笔画少的部件。

如：曹→一冫一曰→1425                              东→七丨八→158  
曹→一口冫日→1425                                      臣→一丨一冫一丁→1503  
牙→一丨一ノ→1058                                      瓦→七乙ノ、→106  
共→一冫一八→218                                      其→一冫三八→238  
直→一丨一冫一→2431                                      干→二丨→25  
酉→西一→31    青→一丰一冫二→342  
由→一口十→42      也可以                              曲→一口冫→42                              也可以  
由→一冫土→42    曲→一冫一→421  
见→一冫儿→49    亡→一丨一→60  
手→一ノ二丨→825    丘→一丨一丁一→831  
豸→一ノ二点一勾二撇→8799                              及→一ノ折捺→806  
生→一牛一→91    己己己→一冫一七→01  
民→一冫一七七→011                                      聿→一冫一三十→032                              等等

### 三、词组的编码

1) 对于两字词组、取每个字的前两个码，当词组中有分类表中的部件时，就只取一码。

如：词组→6004    车辆→111

2) 对于三字词组，取前两字的第一个码，第三字的前两码，当第三个字为分类表中的部件时，就只取一码。

如：太平洋→2177    心连心→717

3) 对于四字及以上字的词组，取每个字的第一个码。

如：三言两语→3616    岗位责任制→58389

十码输入法可以对《新华字典》中所有的字编码。对于更大的汉字集一样能够编码，只是增加了一些重码。

字的编码(略)

繁体字的编码(略)



井田汉字，独一无二的汉字结体构形理论，能够科学地解决数码时代汉字所面临的问题！

推荐：井田汉字、汉字书同文研究、中文虚拟学校、WPL语言文学网、汉字编码设计学、《现代语文》、《中文》、百度、谷歌

湘ICP备05008125号 语言文字网 YYWZ.COM©版权所有