

汉字编码方法

申请号: [95113713.1](#)

申请日: 1995-07-31

申请(专利权)人 [尹宗谋](#)

地址 710077陕西省西安市沣镐路1号65分号空军电讯工程学院电路室

发明(设计)人 [尹宗谋](#)

主分类号 [G06F3/023](#)

分类号 [G06F3/023](#)

公开(公告)号 1142076

公开(公告)日 1997-02-05

专利代理机构 [西安市专利事务所](#)

代理人 [徐平](#)

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷

G06F 3/023

[12] 发 明 专 利 说 明 书

[21] ZL 专利号 95113713.1

[45]授权公告日 2001 年 7 月 11 日

[11]授权公告号 CN 1068444C

[22]申请日 1995.7.31 [24]颁证日 2001.4.12

[21]申请号 95113713.1

[73]专利权人 尹宗谋

地址 710077 陕西省西安市沣镐路 1 号 65 分号
空军电讯工程学院电路室

[72]发明人 尹宗谋

[56]参考文献

CN 931055881 1993.11.17 G06F3/023

CN 931184584 1995.4.12 G06F3/023

审查员 张静海

[74]专利代理机构 西安市专利事务所

代理人 徐 平

权利要求书 5 页 说明书 12 页 附图页数 0 页

[54]发明名称 一种汉字编码输入方法

[57]摘要

一种汉字编码输入方法,其将汉字的部件归纳为 26 部,定义了 26 个部首,部首与英文字母一一对应,以部首作为编码码元,以英文字母为部首的代码,将汉字的结构归纳为六种字型,依据字型选取编码部件,用部件所在部的部首代码组成部首码。本发明部件分类合理,部首定义明确,编码规则自然,易学、易记,简单、有效,易于实现部首、声母、韵母、英文字母四位一体,本发明可用于汉字键盘输入和汉字检索。

ISSN 1008-4274

知识产权出版社出版

权 利 要 求 书

1、一种汉字编码输入方法,包括确定部件和码元的对应关系及编码规则, 其特征在于: 所述的部件和码元的对应关系是按照部件的形、音或义的特征, 将部件分为26部, 每部定义一个最能代表该部特征的简单部件作为该部的部首, 并给出名称, 部首即为编码码元, 按照部首与英文字母形或音的关联关系, 将26个部首与26个英文字母一一对应, 英文字母即为部首的代码; 所述的编码规则是按照汉字的组成和结构选取编码部件, 将选取的部件所在部的部首代码按笔顺排列出, 构成汉字的部首码 ; 所述的26部的部首、名称、代码的对应关系如下:

部首	名称	代码	部首	名称	代码
亻	人	A	口	口	O
八	八	B	丿	撇	P
艹	草	C	丶	点	Q
刀	刀	D	日	日	R
厄	厄	E	氵	水	S
手	手	F	土	土	T
弓	弓	G	又	又	U
横	横	H	丷	折	V
竖	竖	I	木	木	W
交	交	J	乂	乂	X
卡	卡	K	月	月	Y
立	立	L	彡	彡	Z
门	门	M			
牛	牛	N			

所述按照汉字的组成和结构选取编码部件是将汉字结构按块归纳为独体结构、左右结构、上下结构、内外结构、左中右结构和上中下结构六种字型，其中左右、上下、内外为二分结构，左中右、上中下为三分结构，字型中的各块为字块，字块是单个部件或部件的组合；所述的部首码为一部首码、二部首码、三部首码或四部首码，其标准码长分别为1、2、3和4；所述的编码规则是：

① 按照汉字的组成和结构确定取码方式，即确定取码字块和字块取码数，字块数等于码长时，各字块各取一码；字块数大于码长时，按照字型结构确定取码字块，各取码字块分别取一码；字块数小于码长时，按照多部件字块拆分，单部件字块不拆分的原则或笔顺在先的字块拆分，笔顺在后的字块不拆分的原则拆分某字块，拆分的字块至少取二码，不拆分的字块取一码；

② 按照字块的组成和结构确定各字块取码部件，字块部件数等于取码数时，取该各部件；部件数大于取码数时，按字块的字型结构确定取码部件；部件数小于取码数时，按照级别高的部件拆分或笔顺在先的部件拆分的原则，拆分某部件，拆分的部件至少取二个较低级的部件，不拆分的部件取该部件；

③ 完全拆分为部首级部件后部件数仍然小于码元数的汉字，其部首码就由其全部部首级部件对应的代码组成，其码长小于标准码长；所述按照汉字字型确定取码方式的规则如下：

字型	一部首码	二部首码	三部首码	四部首码
独 体	取首部件	首末部件	首、次、末	一、二、三、末
左 右	取首部件	左右各一	左二右一	左二右二
			左一右二	左三右一
				左一右三
上 下	取首部件	上下各一	上二下一	上二下二
			上一下二	上三下一
				上一下三
内 外	取首部件	内外各一	外二内一	外二内二
			外一内二	外三内一
				外一内三
左中右	取首部件	左右各一	左、中、右	左二中一右一
				左一中二右一
				左一中一右二
上中下	取首部件	上下各一	上、中、下	上二中一下一
				上一中二下一
				上一中一下二

所述的按照字块的组成和结构确定取码部件的规则是：

1) 字块取一个部件时，一般取首部件，但

① 字块位于上下或上中下结构的字型之下，或内外结构、开口向下的字型之内，且字块本身为上下或上中下结构，或虽为整体结构，但可分离出位置在下的非单笔部件时，取该字块的下部件；

② 字块位于左右结构或左中右结构的字型之右，且字块本身为左右结构时，取该字块的右部件；

③ 字块位于上中下结构的字型之上或中或下，且本身为左中右结构时，字块取中部件；

2) 字块取二个部件的规则与汉字取二部首码的规则相同；

3) 字块取三个部件的规则与汉字取三部首码的规则相同。

说明书

一种汉字编码输入方法

本发明涉及一种汉字编码输入方法。

汉字编码是中文信息处理的关键技术之一，广泛用于中文输入和中文检索等领域。现有的汉字编码大致分为音码和形码，音码直接采用汉语拼音方案，简单易学，便于推广，但同音字和多音字较多，重码率高，输入速度慢。形码利用汉字的字形特征编码，其方法是选择若干个汉字部件，确定一组码元，建立部件与码元的对应关系，制定编码规则，由于汉字数量繁多，字型各异，加之很多形码以减少重码率作为首要目标，有些方法部件的选择和分类以及部件与码元的对应关系不合理，有些方法汉字拆分和编码规则不规范，因而，一些形码虽然重码率低，输入速度快，但却难学、难记，不完全符合汉字的结构和组成规律。例如五笔字型，它是将130多个字根分布在25个键位上，字根与字母的对应关系由各个字母在键盘上的位置来确定，而且同一字母所对应的多个字根除了起笔相同外，没有其它明显的共同特征，难以记忆；另外，五笔字型要将汉字完全拆分为所限定的字根，然后再取一、二、三、末字根，不符合汉字的结构特征和组字规律；五笔字型的码长为4，一个汉字一般要用4个字母编码，尽管常用字有一级、二级、三级简码，但需要大量记忆才能分辨之；五笔字型字根的顺序性不明显，因而不便于汉字检索。

本发明的目的在于避免上述现有技术中的不足之处，而提供一种部件与码元对应关系合理，汉字的拆分和编码规则规范，易于实现部首、声母、韵母及英文字母四位一体，输入效果高，简

单易学、易记，可用于汉字输入和汉字检索的汉字编码输入方法。

本发明的目的可通过以下措施来达到：

一种汉字编码输入方法，包括确定部件和码元的对应关系及编码规则，其特殊之处在于，所述的部件和码元的对应关系是按照部件的形、音或义的特征，将部件分为26部，每部定义一个最能代表该部特征的简单部件作为该部的部首，并给出名称，部首即为编码码元，按照部首与英文字母形或音的关联关系，将26个部首与26个英文字母一一对应，英文字母即为部首的代码；所述的编码规则是按照汉字的组成和结构选取编码部件，将选取的部件所在部的部首代码按笔顺排列出，构成汉字的部首码。

本发明26部的部首、名称、代码的对应关系如下：

部首	名称	代码	部首	名称	代码
亻	人	A	口	口	O
八	八	B	丿	撇	P
艹	草	C	丶	点	Q
刀	刀	D	日	日	R
厄	厄	E	シ	水	S
手	手	F	土	土	T
弓	弓	G	又	又	U
横	横	H	丷	折	V
竖	竖	I	木	木	W
交	交	J	乂	乂	X
卡	卡	K	月	月	Y
立	立	L	了	之	Z
门	门	M			
牛	牛	N			

本发明26部的部件可分别分为三级，其级别由低至高依次是，部首和变形部首为部首级部件，简单部件为1级部件，其余部件为2级部件，所述的26部部件的分级表如下：

代码	部首级部件	1级部件	2级部件
A	イ 人	入	金食佳
B	ハ 儿	火米 𠂇	父羊𦍋𦍋𦍋
C	𦍋 井	𦍋𦍋𦍋	草其 𦍋
D	勺 刀	夕 𠂇	欠鸟鱼角 𠂇
E	𠂇 厂 丁 𠂇	𠂇 兀 石 耳 𠂇	豕页而辰
F	𠂇 寸 手 𠂇 爪	毛寸	求
G	コ 匚 𠂇 工	弓尸巳巳己ヨ 𠂇	巴尸 𦍋 𦍋 𦍋
H	一	王	二西雨酉歹瓦豆牙
I	丨	𠂇 𠂇 𠂇 𠂇 山	中申
J	十 𠂇	𦍋 𦍋 𦍋 𦍋 𦍋	尤龙文韦专甫
K	ト 上 𠂇	𠂇 止 𠂇 𠂇 𠂇	业长
L	𠂇 立	亡方文六	盲衣辛亦 𦍋
M	𠂇 𠂇 𠂇	巾贝皿穴 𠂇 见	骨
N	𠂇 牛 𠂇	𦍋 𦍋 𦍋 𦍋	矢生 𦍋
O	口 𠂇	虫	足 𦍋
P	ノ	𠂇 𠂇 𠂇 𠂇 𠂇	千斤白白舟身豕自采血鬼
Q	、	𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇	户斗头鹿麻
R	日 𠂇	目田四 𠂇	由甲
S	𠂇 𠂇 水 𠂇 𠂇	小川𦍋 𠂇 𠂇	𦍋
T	土 𠂇 𠂇	𦍋 𦍋	走
U	又	女大	夫 𦍋
V	𠂇 一 乙 𠂇 𠂇	𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇	母母习 𦍋 𦍋
W	木	𦍋	未未
X	乂 𦍋 𦍋 𦍋 九	𦍋 𦍋 𦍋 𦍋 𦍋 𦍋	皮𦍋
Y	月 𠂇 𠂇 𠂇 𠂇	且	
Z	𠂇	𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇	乃及 𦍋 𦍋

本发明按照汉字的组成和结构选取编码部件是将汉字的结构按块归纳为独体结构、左右结构、上下结构、内外结构、左中右结构和上中下结构六种字型，其中左右、上下、内外为二分结构，左中右、上中下为三分结构，字型中的各块为字块，字块是单个部件或部件的组合。

本发明部首码可为一部首码、二部首码、三部首码或四部首码，其标准码长分别为1、2、3和4。

本发明的编码规则是

① 按照汉字的组成和结构确定取码方式，即确定取码字块和字块取码数，字块数等于码长时，各取码字块各取一码；字块数大于码长时，按照字型结构确定取码字块，各字块分别取一码；字块数小于码长时，按照多部件字块拆分，单部件字块不拆分的原则或笔顺在先的字块拆分，笔顺在后的字块不拆分的原则拆分某字块，拆分的字块至少取二码，不拆分的字块取一码；

② 按照字块的组成和结构确定各字块取码部件，字块部件数等于取码数时，取该各部件；部件数大于取码数时，按字块的字型结构确定取码部件；部件数小于取码数时，按照级别高的部件拆分或笔顺在先的部件拆分的原则，拆分某部件，拆分的部件至少取二个较低级的部件，不拆分的部件取该部件；

③ 完全拆分为部首级部件后部件数仍然小于码元数的汉字，其部首码就由其全部部首级部件对应的代码组成，其码长小于标准码卡。

本发明按照汉字字型确定取码方式的规则如下：

字型	一部首码	二部首码	三部首码	四部首码
独 体	取首部件	首末部件	首、次、末	一、二、三、末
左 右	取首部件	左右各一	左二右一	左二右二
			左一右二	左三右一
				左一右三
上 下	取首部件	上下各一	上二下一	上二下二
			上一下二	上三下一
				上一下三
内 外	取首部件	内外各一	外二内一	外二内二
			外一内二	外三内一
				外一内三
左中右	取首部件	左右各一	左、中、右	左二中一右一
				左一中二右一
				左一中一右二
上中下	取首部件	上下各一	上、中、下	上二中一下一
				上一中二下一
				上一中一下二

本发明按照字块确定取码部件的规则是

1) 字块取一个部件时，一般取首部件，但

① 字块位于上下或上中下结构的字型之下，或内外结构、开口向下的字型之内，且字块本身为上下或上中下结构，或虽为整体结构，但可分离出位置在下的非单笔部件时，取该字块的下部件；

② 字块位于左右结构或左中右结构的字型之右，且字块本身为左右结构时，取该字块的右部件；

③ 字块位于上中下结构的字型之上或中或下，且本身为左中右结构时，字块取中部件；

2) 字块取二个部件的规则与汉字取二部首码的规则相同；

3) 字块取三个部件的规则与汉字取三部首码的规则相同。

下面将结合实施例对本发明作进一步详述：

本发明的编码方法主要是将汉字的部件归纳为26部，定义26个部首作为码元，部首与26个英文字母一一对应，将汉字结构归纳为六种字型，依据字型选取编码部件，用部件所在部的部首代码组成部首码。部首码有一部首码、二部首码、三部首码、四部首码等不同码制。下面是三部首码编码的实施方案：

1、将汉字拆分为字块，按照组成和结构确定取码方式：

1) 独体结构的汉字

① 与部首相同的部首字，取部首对应的部首代码；

② 非部首字，先拆分为部件，拆分后为二个部件的按二分结构取码，为三个部件的按三分结构取码，超过三部件的按汉字笔顺首、次、末三部件依次各取一代码；

2) 二分结构的汉字

① 两字块均为部首的双部首字，按笔顺依次取对应的两个部首代码；

② 一字块为部首，另一字块为非部首，部首字块取一个代码，非部首字块取二个代码；

③ 一字块为1级部件，另一字块为2级部件，1级部件取一个代码，2级部件取二个代码；

④ 一字块为部件，另一字块为非部件，部件取一个代码，非部件字块取二个代码；

⑤ 其余二分结构的汉字，笔顺在先的字块取二个代码，笔顺在后的字块取一个代码；

3) 三分结构的汉字

三个字块各取一个代码；

2、字块选取二个部首代码的方法：

1) 独体结构的字块

先将其拆分为部件，拆分后为二个部件的两部件各取一个代码，超过两个部件的按笔顺首、末二部件各取一个代码；

2) 二分结构的字块

二个字块各取一码；

3) 三分结构的字块

① 左中右结构的字块，左、右两字块各取一码；

② 上中下结构的字块，上、下两字块各取一码。

3、字块选取一个部首代码的方法

1) 独体结构的字块

① 字块为部件的，取部件所在部的部首代码；

② 字块为非部件的，拆分按笔顺取其首部件或主体部件所在部的部首代码；

2) 二分结构的字块

① 左右结构的字块，当该字块位于左右结构或左中右结构的汉字之右时，取右部件所在部的部首代码；其余则取左部件所在部的部首代码；

② 上下结构的字块，当该字块位于上下结构或上中下结构的汉字之下，或内外结构的汉字之内，且外结构开口向下时，取下部件所在部的部首代码；其余则取上部件所在部的部首代码；

③ 内外结构的字块，取外部件或首部件所在部的部首代码；

3) 三分结构的字块

① 左中右结构的字块，当该字块位于上中下结构的汉字之上或中或下时，取中部件所在部的部首代码；其余则取左部件所在部的部首代码；

② 上中下结构的字块，取上部件所在部的部首代码。

下面是一些汉字在不同码制时的三部首码编码举例：

例字	取码方式	取码部件	部首码	例字	取码方式	取码部件	部首码
人	部首	人	A	两	部首	一	HMA
门	部首	门	M	爽	次末	乂	UXX
手	部首	手	F	兼	次末	八	BGB
从	部首	人	AA	喜	次末	口	TOO
杏	部首	木	WO	糠	次末	广	BQG
明	部首	日	RY	糖	次末	米	BQO
又	部首	又	UQ	烯	左一	火	BXM
问	部首	门	MO	幢	左一	巾	MLT
生	部首	人	NT	撞	左一	立	FLH
目	部首	目	RH	槽	左一	木	WCR
山	部首	山	IG	殊	左一	夕	HNW
众	部首	人	AAA	谐	左一	艹	LVP
晶	部首	日	RRR	勤	左二	甘	CTX
投	部首	扌	FYU	勃	左二	十	JZX
在	部首	扌	JIT	疆	左二	弓	GTH
式	部首	戈	XQG	疑	左二	匕	XNV
兑	部首	儿	BOB	颜	左二	立	LPE
气	部首	一	NHV	赖	左二	木	WOD
荣	部首	艹	CMW	龄	左二	止	KGA
巴	部首	巳	GIV	剧	左二	尸	GOI
部	部首	立	LOZ	赫	左二	土	TBT
盟	部首	日	RYM	勒	左二	甘	CJX
馅	部首	夕	DDP	灯	左二	人	BAE
范	部首	夕	CSV	答	上一	人	NAO
摊	部首	扌	FUA	景	上一	日	RLS
陶	部首	扌	ZDN	荷	上一	艹	CAE
萝	部首	艹	CRD	篇	上一	竹	NQM
插	部首	扌	FPP	覆	上一	西	HPN
拆	部首	扌	FPQ	蒋	上一	艹	CQD
柴	部首	止	KXW	藻	上一	艹	CSO
亭	部首	止	FOG	蠹	上一	虫	JOO
频	部首	小	KSE	冠	上一	寸	MHF

下面以三部首码为例说明本发明汉字部首码的键盘输入技术。三部首码键盘输入采用三键输入，依次输入汉字的三个部首的代码，不足三码时以空格补充，则屏幕提示该码对应的汉字；若无重码字，或者提示行中第1个字是要输入的字，则直接输入下一个字的部首码；否则，用数字键选择重码字，然后再输入下一个汉字。按照一般规则得到的部首码称为原码，为了减少重码，可采用简码、补码及异码的方法重新定义常用字中的重码字，形成兼容码。为提高输入速度，还可采用词码、联想码等。词码的编码是将双字词看作二分结构，三字词看作三分结构，多字词当作多分结构来编码。

本发明与现有技术相比具有如下优点：

1、继承和革新了传统的部首概念，将部首减少为26个，并与英文字母一一对应，使得部首起到同部件的首部件、汉字的最小组成单元、汉字编码码元的作用。

2、由于部首与英文字母有着形或音相近的特征，同时，由部首很容易联想到本部的部件，因而本发明所建立的编码部件和码元的对应关系合理、自然，易学、难忘，便于掌握应用。

3、汉字拆分是先按字型拆分为字块，必要时再将字块拆分为部件，或再将较高级的部件拆分为较低级的部件，这种根据编码需要逐级拆分汉字的方法，以及按照字型结构和汉字组成确定取码方式，选取编码部件的编码规则，符合汉字的结构特征和组成规律，符合识字、写字和打字的规律。

4、部件分类和定义部首时，使各个键位的使用频率基本平衡，因而部首码的重码率与相同码长的其它形码的重码率相当或

更低，较好地解决了易学与高效的矛盾，三部首码原码的重码字在一级汉字范围内为15.8%，一、二级汉字范围内为21.51% (均指静态重码率)，容易学习掌握，且具有一定的速度，对于经常使用者来说，逐渐记住几百个常用重码字的兼容码后，可大大提高输入效率，甚至可实现盲打。

5、增加少量繁体字部件后，部首码可直接用于繁体字的编码和大汉字库的编码，编码方法基本不变，而且能区分出繁体字和简体字。

6、应用部首码进行检索，比现有的部首检字法简便、实用、有效，字典检字表的编排、人名的排序和检索、其它中文情报资料检索等领域可直接采用部首码。

7、采用本发明定义的部首，可使部首与声母、韵母及英文字母有机结合，四位一体，使得形码和音码相辅相成，形成三部首码、双拼一部码、双拼双部码、三部一声码等不同的码制，构成一个体系，可适应不同的需要。