



[12] 发明专利申请公开说明书

[21]申请号 94117376.3

[51]Int.Cl⁶

G06F 3/023

[43]公开日 1995年8月30日

[22]申请日 94.11.1

[71]申请人 彭志春

地址 444200湖北省远安县茧丝绸公司

[72]发明人 彭志春

说明书页数:

附图页数:

[54]发明名称 汉字音形编码

[57]摘要

汉字音形编码属语言文字类,该发明解决汉字编码中存在的编码长,重码多,难掌握,速度慢等难题。其主要技术特征是:首先确定60多个人们容易接受,使用频度高,又具有一定意义的字根,然后依据字根将汉字划分为字数和使用频度依次减少的四类字。编码时先确定汉字拼音的首字母,再按照汉字类别分别编字形码。由于汉字给人的第一印象是读音、字义和大致字形特征,而非笔画细节,因而使用该编码进行汉字输入简单、迅速。

权 利 要 求 书

汉字音形编码设置有任意码，对拼音判断不明时使用。对于重码，采用屏幕显示，用数字加以选择的办法，还可以通过联想前后字加以区别。拼音码采用汉字拼音的首字母，字形码根据汉字的结构和笔画编码。对于常用词组和短语，也可采用拼音缩写码表示。其主要特征是，选择60多个人们容易接受，组字能力强，具有一定含义的字根，根据字根将汉字分类，编码时首先确定汉字拼音的首字母，再按照汉字类别编字形码。汉字分类是按使用频度和字数来依次排列，并且不交差，容易区别，从而使编码迅速。该发明的权利要求如下：

1. 根据汉字音形编码的主要特征申请的其它专利，及由此专利产生的其它权利。

2. 根据汉字音形编码制定具体实施方案，及以此实施方案编制汉字输入系统，制造、销售和使用该产品。

3. 对汉字音形编码进行必要的补充和修改。

4. 以赢利为目的，利用宣传工具和其他传播途径对汉字音形编码及具体实施方案直至最后产品进行解释、说明和介绍。

5. 汉字音形编码除汉字输入外的其它用途。

6. 法律保护专利的其它权利。

汉 字 音 形 编 码

该发明属语言文字领域。目前，汉字编码技术已有数百种，大体分为字形码和拼音码两类，具有代表意义的分别是五笔字型 and 智能ABC系统，这两种编码技术应用最广，特别是前者已为人们所熟知。另外一种区位码，一般不常用。发明汉字音形编码的目的是使汉字输入简单易学，快速方便。

汉字音形编码是考虑汉字拼音和字形两方面的特点。所谓字形，不仅仅是考虑汉字字根的笔画，而且充分地考虑到字根的含义，以此为标准将汉字区分为四类：

I类：字中含有22个字根：讠，卩，亻，尸，辶，卜，扌，彳，彡，亻，彳，彡，月，王，广，山，心，艹，竹，心，礻，车，耳。这些字根组字能力最强，在句中作偏旁部首使用，除“月”“王”之外本身都不能独立成字，但都具有约定的意义，并且此意义和其所组成的字有内在的联系。

II类：字中含有23个字根：土，口，山，女，马，心，力，木，走，车，日，贝，火，石，目，田，禾，鸟，米，耳，虫，雨，鱼。这些字根使用频度很高，本身都能独立成字，都具有一定的意义，并且此意义和其所组成的字有密切的联系。

III类：字中含有22个字根：又，刀，广，门，五，士，大，人，十，寸，小，巾，尸，弓，子，户，欠，皿，页，羽，见，厂。这些字根本身都能独立成字，使用频度高，都具有一定意义，但所组成的字与此意义联系不大。

IV类：尽量归入上述三类后的其它汉字，这类汉字数量很少，并且大都是笔画简单的常用字，容易按照笔画书写顺序熟练编码。

以上分类也是按照汉字的频度依次考虑，为减少重码，不允许同一个汉字有不同的分类编码，但允许同一个汉字在同一分类中有不同的编码。具体制定以下几项分类原则：

1. 前类优先 如果某个汉字可同时归入I、II、III类，则应依次归入到前面的类别。

2. 尽量归类 如果某个汉字仅含有Ⅰ类或Ⅱ类中的字根, 不具备其它特征, 也应归入Ⅰ类或Ⅱ类。

3. 偏旁部首优先 如果某个汉字同时含有Ⅰ类和Ⅱ类(或Ⅲ类)中的字根, Ⅰ类中的字根不作偏旁部首, 而Ⅱ类(或Ⅲ类)中的字根是偏旁部首, 则该汉字归入Ⅱ类(或Ⅲ类)。

所谓拼音, 是指汉字拼音的首字母, 包括: b. p. m. f. d. t. n. l. g. k. h. j. q. x. z. c. s. r 18个声母和 a. e. o. y. w 5个首写韵母。

将25个高频字和Ⅰ. Ⅱ. Ⅲ类字根分别定义在英文字母键上, 制定如下的程序辨识英文字母键。

1. 英文字母键+空位键, 英文字母键表示25个高频字。

2. 字母键+字母键, 前一个字母键表示拼音, 后一个字母键表示Ⅰ类字根。

3. 字母键+转换键+字母键, 前字母键表示拼音, 后字母键表示Ⅱ类字根。

4. 字母键+转换键两次+字母键, 前字母键表示拼音, 后字母键表示Ⅲ类字根。

如果按照拼音首字母+字形键对汉字进行编码, 显然, 对于高频字, Ⅰ类、Ⅱ类、Ⅲ类汉字, 分别依照如上所述的1. 2. 3. 4种程序编码, 而对于Ⅳ类汉字, 按照汉字笔画的四个方向“一”“丨”“/”“\”设置四个笔画键, 编码时考虑四个方向的笔画书写顺序, 编码为: 字母键+首笔画键+次笔画键+……, 直至能选择出该汉字为止。

考虑到实际运用的方便, 作以下几点补充规定:

1. 对于只知道字形不会拼音的字, 可设一任意码(五笔字型为Z码)代替拼音码, 然后按照汉字音形编码中的字形码部分编码, 并依靠笔画键选择出该汉字。

2. 只会拼音而不知字形的字, 编码时连续三次该汉字的首字母, 再接全拼音法编剩下的音码。

3. 本身不成字的组字单元, 按照笔画的书写顺序, 直接由四个方向的笔

画键选择出该组字单元。

汉字音形编码充分利用汉字中的形音字占最大比例的特点，根据汉字的拼音和字形进行最抽象的概括，从而确定汉字。由于编码也是按照汉字的使用频度进行分类，并且分类不重叠，这样出现在优先分类中的高频字将不在以后的分类中出现，因而减少了重码。并且对于在汉语中大多数出现的词组很少有相同的拼音码和字形码，例如对两个字的词组就至少可区分： $23 \times 22 \times 23 = 11638$ 个，这样可进一步减少重码，汉字音形编码吻合人们长期以来对汉字形成的思维习惯和书写规则，从而简单易学，编码迅速。这些都是五笔字型 and 智能ABC系统所缺少的。

与五笔字型相比，汉字音形编码具有如下优点：

1. 避免了五笔字型将汉字进行机械的分拆，从而导致字根多，重码多，输入速度慢，难学。五笔字型几乎是对汉字的重新认识。
2. 汉字音形编码可以实现不看字形进行盲打。
3. 汉字音形编码码长短，常用字大多数只有两个或三个码。频度越高的字编码越短。

智能ABC系统主要是根据拼音编码，考虑字形时也是将汉字机械地拆解为无意义的零散笔画，从而导致编码长，重码多，输入慢等缺点。

总之，该发明码长短，重码少，简单易学，快速方便，是比五笔字型和智能ABC系统更为先进的汉字输入技术。

汉字音形编码主要用于汉字输入，实现该发明可编制汉字音形编码程序软件，用于目前的汉字输入系统中。考虑到兼容性，并结合人们已养成的打字习惯和实践经验，对键盘要作相应的定义。例如将26个高频字和Ⅰ、Ⅱ、Ⅲ类字根分别定义在不同的字母键上，将光标键定义为笔画键等等。在软件开发完成之后，操作人员就可以根据汉字音形编码快速地输入汉字。