



100033

北京市西城区金融街 19 号富凯大厦 B 座 11 层 中原信达知识产权代
理有限责任公司
韩黎捷((8610)66576688) 郭晗((8610)66576688)

发文日:

2023 年 01 月 03 日



申请号或专利号: 202110910630.2

发文序号: 2022122801083030

申请人或专利权人: 王中琪

发明创造名称: 汉字排序、汉字检索和汉字插入的方法和装置

发明专利申请公布通知书

上述专利申请, 经初步审查, 符合专利法实施细则第 44 条的规定。根据专利法第 34 条的规定, 该申请在 38 卷 5201 期 2022 年 12 月 27 日专利公报上予以公布。

提示:

1. 发明专利申请人可以自申请日起 3 年内提交实质审查请求书、缴纳实质审查费, 申请人期满未提交实质审查请求书或期满未足额缴纳实质审查费的, 该申请被视为撤回。

2. 专利费用可以通过网上缴费、邮局或银行汇款缴纳, 也可以到国家知识产权局面缴。

网上缴费: 电子申请注册用户可登陆 <http://cponline.cnipa.gov.cn>, 并按照相关要求使用网上缴费系统缴纳。

邮局汇款: 收款人姓名: 国家知识产权局专利局收费处, 商户客户号: 110000860。

银行汇款: 开户银行: 中信银行北京知春路支行; 户名: 国家知识产权局专利局; 账号: 7111710182600166032。

汇款时应当准确写明申请号、费用名称(或简称)及分项金额。未写明申请号和费用名称(或简称)的视为未办理缴费手续。了解更多详细信息及要求, 请登陆 <http://www.cnipa.gov.cn> 查询。

缴费人可通过专利收费电子票据交付服务系统 <http://pjonline.cnipa.gov.cn> 及支付宝、微信的电子票夹小程序查询、下载相应电子票据。

3. 申请人可以访问国家知识产权局政府网站(www.cnipa.gov.cn), 在专利检索栏目中查询公布文本。如果申请人需要纸件申请公布单行本的纸件, 可向国家知识产权局请求获取。

4. 申请文件修改格式要求:

对权利要求修改的应当提交相应的权利要求替换项, 涉及权利要求引用关系时, 则需要将相应权项一起替换补正。如果申请人需要删除部分权项, 申请人应该提交整理后连续编号的部分权利要求书。

对说明书修改的应当提交相应的说明书替换段, 不得增加和删除段号, 仅只能对有修改部分段进行整段替换。如果要增加内容, 则只能增加在某一段中; 如果需要删除一个整段内容, 应该保留该段号, 并在此段号后注明: “此段删除” 字样。段号以国家知识产权局回传的或公布/授权公告的说明书段号为准。

对说明书附图、摘要、摘要附图修改的应当提交相应的说明书附图、摘要、摘要附图替换页。

同时, 申请人应当在补正书或意见陈述书中标明修改涉及的权项、段号、页。

审 查 员: 自动审查

审查部门: 专利局初审及流程管理部

联系电话: 010-62356655

210305
2021.1

纸件申请, 回函请寄: 100088 北京市海淀区蓟门桥西土城路 6 号 国家知识产权局专利局受理处收
电子申请, 应当通过电子专利申请系统以电子文件形式提交相关文件。除另有规定外, 以纸件等其他形式提交的文件视为未提交。



(12) 发明专利申请

(10) 申请公布号 CN 115525728 A

(43) 申请公布日 2022. 12. 27

(21) 申请号 202110910630.2

G06F 40/216 (2020.01)

(22) 申请日 2021.08.09

G06F 16/383 (2019.01)

(30) 优先权数据

17/304,849 2021.06.27 US

(71) 申请人 王中琪

地址 美国华盛顿州贝威尔市西萨马米什湖
路东南527号

(72) 发明人 王中琪

(74) 专利代理机构 中原信达知识产权代理有限
责任公司 11219

专利代理师 韩黎捷 郭晗

(51) Int.Cl.

G06F 16/31 (2019.01)

G06F 40/126 (2020.01)

G06F 40/237 (2020.01)

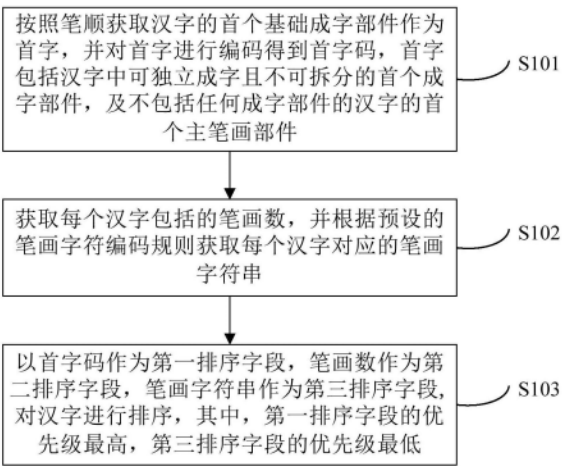
权利要求书5页 说明书21页 附图19页

(54) 发明名称

汉字排序、汉字检索和汉字插入的方法和装置

(57) 摘要

本发明公开了一种汉字排序、汉字检索和构建字典的方法和装置,涉及计算机技术领域。该方法的一具体实施方式包括:按照笔顺获取汉字的首个基础成字部件作为首字,并对首字进行编码得到首字码,首字包括汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;以首字码作为第一排序字段,笔画数作为第二排序字段,笔画字符串作为第三排序字段,对汉字进行排序,其中,第一排序字段的优先级最高,第三排序字段的优先级最低。该实施方式能够解决由于部首的定义及位置不固定等带来的排序和检索困难的问题。



1. 一种依据首个基础成字部件的汉字排序方法,其特征在于,包括:

接收用户利用终端设备所发来的汉字排序请求;

对接收到的汉字排序请求进行分析处理,并将汉字排序结果反馈给终端设备;其中,分析处理包括:

按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码,所述首字包括所述汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;

获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;

以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低。

2. 根据权利要求1所述的汉字排序方法,其特征在于,按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码包括:

对汉字的基础部件进行编码得到基础部件双字符码;

将汉字拆分为基础部件,并根据所述基础部件双字符码得到汉字双字符码;

按照笔顺获取汉字的首个基础成字部件作为首字,建立所述基础部件与所述首字的关联关系,并生成基础部件双字符码与首字码的第一匹配表;

根据所述汉字双字符码和所述第一匹配表确定汉字的首字码。

3. 根据权利要求2所述的汉字排序方法,其特征在于,对汉字的基础部件进行编码得到基础部件双字符码包括:

根据《信息处理用GB13000.1字符集汉字部件规范》获取汉字的基础部件;

根据基础部件的笔画数对所述基础部件进行第一排序;

对第一排序后的基础部件中笔画数相同的基础部件,按照基础部件包括的笔画顺序进行第二排序;

将第二排序后的基础部件保存到表中生成有序码表以得到基础部件双字符码,其中,所述有序码表的横行和纵列分别以不同类型的字符进行编码。

4. 根据权利要求2所述的汉字排序方法,其特征在于,对汉字的基础部件进行编码得到基础部件双字符码包括:

根据《信息处理用GB13000.1字符集汉字部件规范》获取汉字的基础部件;

根据基础部件是否为成字部件将所述基础部件分为成字部件和非字部件,并进行第一排序;

根据基础部件的笔画数,分别对所述成字部件和所述非字部件进行第二排序;

对第二排序后的成字部件和非字部件中笔画数相同的基础部件,分别按照基础部件包括的笔画顺序进行第三排序;

将第三排序后的基础部件保存到表中以生成有序码表以得到基础部件双字符码,其中,所述有序码表的横行和纵列分别以不同类型的字符进行编码。

5. 根据权利要求2所述的汉字排序方法,其特征在于,将汉字拆分为基础部件,并根据所述基础部件双字符码得到汉字双字符码包括:

根据《信息处理用GB13000.1字符集汉字部件规范》将汉字拆分为基础部件；

根据所述基础部件双字符码，获取每个汉字包括的基础部件的双字符码；

对每个汉字，根据《GB13000.1字符集汉字笔顺规范》，按照所述汉字包括的基础部件在所述汉字中的笔画顺序，将所述汉字包括的基础部件的双字符码进行组合得到汉字双字符码。

6. 根据权利要求2所述的汉字排序方法，其特征在于，所述首字码包括成字首字码和笔画首字码；

根据所述汉字双字符码和所述第一匹配表确定汉字的首字码包括：

获取所述汉字双字符码包括的基础部件双字符码；

根据基础部件双字符码的顺序，依次检查每个基础部件双字符码所对应的所述第一匹配表中的首字码；

对每个基础部件双字符码，若所述基础部件双字符码所对应的所述第一匹配表中的某个首字码为成字首字码，则将该成字首字码作为汉字的首字码；

否则，根据所述第一匹配表，获取所述汉字双字符码包括的第一个基础部件双字符码对应的首字码作为汉字的首字码。

7. 根据权利要求1-6中任一所述的汉字排序方法，其特征在于，若某汉字的第一个成字部件为“口”，而且该汉字还包括不是“口”的成字部件，则将该汉字的第一个不是“口”的成字部件确定为该汉字的首字。

8. 根据权利要求1-6中任一所述的汉字排序方法，其特征在于，根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串包括：

为每个笔画定义一个对应的字母作为笔画的字符编码；

对每个汉字所包括的笔画，按照笔顺将所有笔画对应的字符编码拼接得到每个汉字对应的笔画字符串。

9. 根据权利要求2-6中任一所述的汉字排序方法，其特征在于，根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串包括：

为每个笔画定义一个对应的字母作为笔画的字符编码；

对每个基础部件所包括的笔画，按照笔顺将所有笔画对应的字符编码拼接得到每个基础部件对应的笔画字符串；

建立基础部件双字符码与基础部件笔画字符串的关联关系并生成第二匹配表；

对每个汉字，根据所述汉字的汉字双字符码得到所述汉字双字符码包括的基础部件双字符码；

对每个基础部件双字符码，从所述第二匹配表中查找对应的基础部件笔画字符串；

按照基础部件双字符码在汉字双字符码中的顺序，将每个基础部件双字符码对应的基础部件笔画字符串拼接得到所述汉字的笔画字符串。

10. 根据权利要求2-6中任一所述的汉字排序方法，其特征在于，所述汉字双字符码、所述基础部件双字符码和所述首字码为使用大小写字母编成的双字母码；并且，在根据所述首字码进行排序时，按照所述双字母码的字符顺序依次比较进行排序；在根据所述笔画字符串进行排序时，按照所述笔画字符串的字符顺序依次进行比较排序。

11. 根据权利要求1所述的汉字排序方法，其特征在于，所述主笔画部件包括“一”，

“丨”、“丿”、“丶”和“乙”；并且，所述方法还包括：

根据《GB13000.1字符集汉字折笔规范》，将笔画“提”归并为首字“一”，将笔画“竖钩”归并为首字“丨”，将笔画“捺”归并为首字“丶”，将所有“折”笔画归并为首字“乙”。

12. 一种依据首个基础成字部件的汉字排序装置，其特征在于，包括：

用于接收用户利用终端设备所发来的汉字排序请求的单元；

用于对接收到的汉字排序请求进行分析处理，并将汉字排序结果反馈给终端设备的单元；其中，分析处理通过以下模块实现：

首字处理模块，用于按照笔顺获取汉字的首个基础成字部件作为首字，并对所述首字进行编码得到首字码，所述首字包括所述汉字中可独立成字且不可拆分的首个成字部件，及不包括任何成字部件的汉字的首个主笔画部件；

笔画处理模块，用于获取每个汉字包括的笔画数，并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串；

多级排序模块，用于以所述首字码作为第一排序字段，所述笔画数作为第二排序字段，所述笔画字符串作为第三排序字段，对所述汉字进行排序，其中，所述第一排序字段的优先级最高，所述第三排序字段的优先级最低。

13. 一种基于权利要求1-11中任一项所述的依据首个基础成字部件的汉字排序方法进行汉字插入的方法，其特征在于，包括：

按照笔顺获取汉字的首个基础成字部件作为首字，并对所述首字进行编码得到首字码，所述首字包括所述汉字中的首个成字部件，及不包括任何成字部件的汉字的首个主笔画部件；

获取每个汉字包括的笔画数，并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串；

以所述首字码作为第一排序字段，所述笔画数作为第二排序字段，所述笔画字符串作为第三排序字段，对所述汉字进行排序，其中，所述第一排序字段的优先级最高，所述第三排序字段的优先级最低；

对排序后的每个汉字，获取所述每个汉字在字典正文中的存储位置，并根据排序后的每个汉字及所述每个汉字对应的存储位置生成检字表；

根据笔画数和笔顺对所述首字进行排序以生成首字表；

通过所述首字表、所述检字表和所述字典正文构建字典；

在首字码表中找到汉字含有的首字以及该首字相应的首字码；

计算该汉字含有的笔画数，写出该汉字的笔画，并且将全部笔画转换为一个笔画字符串；

找出该汉字在字典正文中的页数；

将该汉字按照它的首字码、笔画数及笔画字符串的先后顺序与检字表辅助排序表中的字比较，然后按顺序将该汉字插入到相应的行。

14. 根据权利要求13所述的进行汉字插入的方法，其特征在于，所述检字表具有页码，并且，根据笔画数和笔顺对所述首字进行排序以生成首字表包括：

根据笔画数由小到大对所述首字进行第一次排序；

对第一次排序后的首字，将笔画数相同的首字按笔画顺序进行第二次排序；

根据第二次排序后的首字和每个首字在所述检字表中的页码生成首字表。

15. 根据权利要求13所述的进行汉字插入的方法,其特征在於,在对所述首字进行编码得到首字码之前,还包括:

根据字典的大小以及所包括的字,对获取的首字进行调整。

16. 根据权利要求15所述的进行汉字插入的方法,其特征在於,根据字典的大小以及所包括的字,对获取的首字进行调整包括:

根据字典的大小判断是否要对获取的首字进行调整;

若是,则统计获取的首字中每个成字部件的出现概率;

根据设定的阈值,将出现概率小于所述阈值,且删除后不会引起混淆的成字部件从所述获取的首字中删除;

或将包含相同成字部件的成字部件进行合并。

17. 一种基于权利要求1-11中任一项所述的依据首个基础成字部件的汉字排序方法进行汉字插入的装置,其特征在於,包括:

首字处理模块,按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码,所述首字包括所述汉字中的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;

笔画处理模块,获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;

多级排序模块,以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低;

检字表生成模块,用于对排序后的每个汉字,获取所述每个汉字在字典正文中的存储位置,并根据排序后的每个汉字及所述每个汉字对应的存储位置生成检字表;

首字表生成模块,用于根据笔画数和笔顺对所述首字进行排序以生成首字表;

字典构建模块,用于通过所述首字表、所述检字表和所述字典正文构建字典;

用于在首字码表中找到汉字含有的首字以及该首字相应的首字码的模块;

用于计算该汉字含有的笔画数,写出该汉字的笔画,并且将全部笔画转换为一个笔画字符串的模块;

用于找出该汉字在字典正文中的页数的模块;

用于将该汉字按照它的首字码、笔画数及笔画字符串的先后顺序与检字表辅助排序表中的字比较,然后按顺序将该汉字插入到相应的行的模块。

18. 一种使用如权利要求13-16中任一项所述的方法进行汉字检索的方法,其特征在於,包括:

按照笔顺获取待检索汉字的首个基础成字部件作为首字;若所述汉字包括成字部件,则所述首字为所述汉字中的首个成字部件;若所述汉字不包括任何成字部件,则所述首字为所述汉字的首个主笔画部件;

从首字表中查找到所述首字,并得到所述首字对应的检字表页面;

从所述检字表页面中查找所述汉字。

19. 根据权利要求18所述的汉字检索的方法,其特征在於,从首字表中查找到所述首

字,并得到所述首字对应的检字表页面包括:

从首字表中查找到所述首字,并获取所述首字在所述检字表中的页码;
根据所述页码获取所述首字对应的检字表页面。

20.一种使用如权利要求13-16中任一项所述的方法进行汉字检索的装置,其特征在于,包括:

首字获取模块,用于按照笔顺获取待检索汉字的首个基础成字部件作为首字;若所述汉字包括成字部件,则所述首字为所述汉字的首个成字部件;若所述汉字不包括任何成字部件,则所述首字为所述汉字的首个主笔画部件;

首字表查找模块,用于从首字表中查找到所述首字,并得到所述首字对应的检字表页面;

检字表查找模块,用于从所述检字表页面中查找所述汉字。

21.一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-11和/或13-16和/或18-19中任一所述的方法。

22.一种计算机可读介质,其上存储有计算机程序,其特征在于,所述程序被处理器执行时实现如权利要求1-11和/或13-16和/或18-19中任一所述的方法。

汉字排序、汉字检索和汉字插入的方法和装置

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种汉字排序、汉字检索和构建字典的方法和装置。

背景技术

[0002] 汉字的排序与检索分成形序法和音序法两大类。形序法根据汉字的形状,主要有部首法、笔画与笔顺法和号码法。音序法根据汉字的发音,现在主要是汉语拼音法。

[0003] 自从一千八百多年前的许慎的《说文解字》建立了部首法来排检汉字以来,一直到几十年前汉字拼音法的出现,部首法一直占据统治地位。所谓部首法,就是将汉字归纳为几百个不同的部(也就是组),每部选出一个字或符号作为代表,称为部首。检索一个字时,先确定该字所在的部及其部首的位置,然后就可以在该部首后面相近的地方找到该字了。如果一个部内的字比较多,它还可以采用笔画数与笔顺来辅助检索部内的字。《说文解字》定义了540个部,《康熙字典》定义了214个部,《新华字典》用了189个部,《词源》用了250个部,《汉字部首表》定义了201个部。总之,《说文解字》的部首定义得比较细,比较全面,而以后的部首则有很多归类合并。

[0004] 笔画与笔顺法根据字的笔画数与笔画顺序排序。该法先根据笔画数排序,然后再根据笔顺。号码法将字按照字的形状给予一个数字来排序。音序法按照字的声音来排序。汉语拼音法采用了罗马字母来表示声音和进行排序。

[0005] 然而,传统的、至今广泛使用的部首法的主要缺点在于部首的定义及位置模模糊糊,缺乏严格的规律性。确定一个字的部首是什么,有时凭位置、有时凭声音、有时凭形状、有时凭习惯、有时凭偏旁与字的关系与含义、有时硬性规定,很多的时候让人无所适从。部首可以是独体字,可以是偏旁,也可以是一个形体结构。往往查几个字就会碰上一个需要花费很多时间,甚至于最终不得不放弃的情况。部首法使得从字典中查一个不认识的字这样一个简单的事情变成一个需要凭运气和概率的事情。

[0006] 笔画与笔顺法的同笔画字太多,查找不便。它一般作为其它方法的辅助方法。号码法所适用的字数要么有限,要么太多,都不好记忆。拼音法相同音的字多,不认识或不知读音的字无法检索。

发明内容

[0007] 有鉴于此,本发明实施例提供一种汉字排序、汉字检索和构建字典的方法和装置,能够解决传统的部首检字法由于部首的定义及位置不固定、缺乏严格的规律性等带来的排序和检索困难的问题,打破了需要用到本身不是字的偏旁之类形体的传统,具有系统性,严谨性,实用性,高效易学易用的特点。

[0008] 为实现上述目的,根据本发明实施例的一个方面,提供了一种汉字排序方法。

[0009] 一种依据首个基础成字部件的汉字排序方法,包括:按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码,所述首字包括所述汉字中可独

立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低。

[0010] 可选地,按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码包括:对汉字的基础部件进行编码得到基础部件双字符码;将汉字拆分为基础部件,并根据所述基础部件双字符码得到汉字双字符码;按照笔顺获取汉字的首个基础成字部件作为首字,建立所述基础部件与所述首字的关联关系,并生成基础部件双字符码与首字码的第一匹配表;根据所述汉字双字符码和所述第一匹配表确定汉字的首字码。

[0011] 可选地,对汉字的基础部件进行编码得到基础部件双字符码包括:根据《信息处理用GB13000.1字符集汉字部件规范》获取汉字的基础部件;根据基础部件的笔画数对所述基础部件进行第一排序;对第一排序后的基础部件中笔画数相同的基础部件,按照基础部件包括的笔画顺序进行第二排序;将第二排序后的基础部件保存到表中生成有序码表以得到基础部件双字符码,其中,所述有序码表的横行和纵列分别以不同类型的字符进行编码。

[0012] 可选地,对汉字的基础部件进行编码得到基础部件双字符码包括:根据《信息处理用GB13000.1字符集汉字部件规范》获取汉字的基础部件;根据基础部件是否为成字部件将所述基础部件分为成字部件和非字部件,并进行第一排序;根据基础部件的笔画数,分别对所述成字部件和所述非字部件进行第二排序;对第二排序后的成字部件和非字部件中笔画数相同的基础部件,分别按照基础部件包括的笔画顺序进行第三排序;将第三排序后的基础部件保存到表中以生成有序码表以得到基础部件双字符码,其中,所述有序码表的横行和纵列分别以不同类型的字符进行编码。

[0013] 可选地,将汉字拆分为基础部件,并根据所述基础部件双字符码得到汉字双字符码包括:根据《信息处理用GB13000.1字符集汉字部件规范》将汉字拆分为基础部件;根据所述基础部件双字符码,获取每个汉字包括的基础部件的双字符码;对每个汉字,根据《GB13000.1 字符集汉字笔顺规范》,按照所述汉字包括的基础部件在所述汉字中的笔画顺序,将所述汉字包括的基础部件的双字符码进行组合得到汉字双字符码。

[0014] 可选地,所述首字码包括成字首字码和笔画首字码;根据所述汉字双字符码和所述第一匹配表确定汉字的首字码包括:获取所述汉字双字符码包括的基础部件双字符码;根据基础部件双字符码的顺序,依次检查每个基础部件双字符码所对应的所述第一匹配表中的首字码;对每个基础部件双字符码,若它所述基础部件双字符码所对应的所述第一匹配表中的某个首字码为成字首字码,则将该成字首字码作为汉字的首字码;否则,根据所述第一匹配表,获取所述汉字双字符码包括的第一个基础部件双字符码对应的首字码作为汉字的首字码。

[0015] 可选地,若某汉字的第一个成字部件为“口”,而且该汉字还包括不是“口”的成字部件,则将该汉字的第一个不是“口”的成字部件确定为该汉字的首字。

[0016] 可选地,根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串包括:为每个笔画定义一个对应的字母作为笔画的字符编码;对每个汉字所包括的笔画,按照笔顺将所有笔画对应的字符编码拼接得到每个汉字对应的笔画字符串。

[0017] 可选地,根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串包括:为每个笔画定义一个对应的字母作为笔画的字符编码;对每个基础部件所包括的笔画,按照笔顺将所有笔画对应的字符编码拼接得到每个基础部件对应的笔画字符串;建立基础部件双字符码与基础部件笔画字符串的关联关系并生成第二匹配表;对每个汉字,根据所述汉字的汉字双字符码得到所述汉字双字符码包括的基础部件双字符码;对每个基础部件双字符码,从所述第二匹配表中查找对应的基础部件笔画字符串;按照基础部件双字符码在汉字双字符码中的顺序,将每个基础部件双字符码对应的基础部件笔画字符串拼接得到所述汉字的笔画字符串。

[0018] 可选地,所述汉字双字符码、所述基础部件双字符码和所述首字码为使用大小写字母编成的双字母码;并且,在根据所述首字码进行排序时,按照所述双字母码的字符顺序依次比较进行排序;在根据所述笔画字符串进行排序时,按照所述笔画字符串的字符顺序依次进行比较排序。

[0019] 可选地,所述主笔画部件包括“一”,“丨”、“丿”、“丶”和“乙”;并且,所述方法还包括:根据《GB13000.1字符集汉字折笔规范》,将笔画“提”归并为首字“一”,将笔画“竖钩”归并为首字“丨”,将笔画“捺”归并为首字“丶”,将所有“折”笔画归并为首字“乙”。

[0020] 根据本发明实施例的另一方面,提供了一种汉字排序装置。

[0021] 一种依据首个基础成字部件的汉字排序装置,包括:首字处理模块,用于按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码,所述首字包括所述汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;笔画处理模块,用于获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;多级排序模块,用于以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低。

[0022] 根据本发明实施例的另一方面,提供了一种构建字典的方法。

[0023] 一种基于本发明实施例所述的依据首个基础成字部件的汉字排序方法构建字典的方法,包括:按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码,所述首字包括所述汉字的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低;对排序后的每个汉字,获取所述每个汉字在字典正文中的存储位置,并根据排序后的每个汉字及所述每个汉字对应的存储位置生成检字表;根据笔画数和笔顺对所述首字进行排序以生成首字表;通过所述首字表、所述检字表和所述字典正文构建字典。

[0024] 可选地,所述检字表具有页码,并且,根据笔画数和笔顺对所述首字进行排序以生成首字表包括:根据笔画数由小到大对所述首字进行第一次排序;对第一次排序后的首字,将笔画数相同的首字按笔画顺序进行第二次排序;根据第二次排序后的首字和每个首字在所述检字表中的页码生成首字表。

[0025] 可选地,在对所述首字进行编码得到首字码之前,还包括:根据字典的大小以及所

包括的字,对获取的首字进行调整。

[0026] 可选地,根据字典的大小以及所包括的字,对获取的首字进行调整包括:根据字典的大小判断是否要对获取的首字进行调整;若是,则统计获取的首字中每个成字部件的出现概率;根据设定的阈值,将出现概率小于所述阈值,且删除后不会引起混淆的成字部件从所述获取的首字中删除;或,将包含相同成字部件的成字部件进行合并。

[0027] 根据本发明实施例的又一方面,提供了一种构建字典的装置。

[0028] 一种基于本发明实施例所述的依据首个基础成字部件的汉字排序方法构建字典的装置,包括:首字处理模块,按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码,所述首字包括所述汉字中的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;笔画处理模块,获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;多级排序模块,以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低;检字表生成模块,用于对排序后的每个汉字,获取所述每个汉字在字典正文中的存储位置,并根据排序后的每个汉字及所述每个汉字对应的存储位置生成检字表;首字表生成模块,用于根据笔画数和笔顺对所述首字进行排序以生成首字表;字典构建模块,用于通过所述首字表、所述检字表和所述字典正文构建字典。

[0029] 根据本发明实施例的又一方面,提供了一种汉字检索方法。

[0030] 一种使用本发明实施例构建的字典进行汉字检索的方法,包括:按照笔顺获取待检索汉字的首个基础成字部件作为首字;若所述汉字包括成字部件,则所述首字为所述汉字中的首个成字部件;若所述汉字不包括任何成字部件,则所述首字为所述汉字的首个主笔画部件;从首字表中查找到所述首字,并得到所述首字对应的检字表页面;从所述检字表页面中查找所述汉字。

[0031] 可选地,从首字表中查找到所述首字,并得到所述首字对应的检字表页面包括:从首字表中查找到所述首字,并获取所述首字在所述检字表中的页码;根据所述页码获取所述首字对应的检字表页面。

[0032] 根据本发明实施例的又一方面,提供了一种汉字检索装置。

[0033] 一种使用本发明实施例构建的字典进行汉字检索的装置,包括:首字获取模块,用于按照笔顺获取待检索汉字的首个基础成字部件作为首字;若所述汉字包括成字部件,则所述首字为所述汉字中的首个成字部件;若所述汉字不包括任何成字部件,则所述首字为所述汉字的首个主笔画部件;首字表查找模块,用于从首字表中查找到所述首字,并得到所述首字对应的检字表页面;检字表查找模块,用于从所述检字表页面中查找所述汉字。

[0034] 根据本发明实施例的又一方面,提供了一种电子设备。

[0035] 一种的电子设备,包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现本发明实施例所提供的汉字排序、汉字检索和构建字典的方法。

[0036] 根据本发明实施例的再一方面,提供了一种计算机可读介质。

[0037] 一种计算机可读介质,其上存储有计算机程序,所述程序被处理器执行时实现本发明实施例所提供的汉字排序、汉字检索和构建字典的方法。

[0038] 上述发明中的一个实施例具有如下优点或有益效果：通过按照笔顺获取汉字的首个基础成字部件作为首字，并对首字进行编码得到首字码，首字包括所述汉字中的首个成字部件，及不包括任何成字部件的汉字的首个主笔画部件；获取每个汉字包括的笔画数，并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串；以首字码作为第一排序字段，笔画数作为第二排序字段，笔画字符串作为第三排序字段，对汉字进行排序，其中，第一排序字段的优先级最高，第三排序字段的优先级最低。本技术方案将首个基础成字部件（首字）作为汉字排序和检索的标准，解决了传统的部首检字法由于部首的定义及位置不固定、缺乏严格的规律性等带来的排序和检索困难的问题，打破了需要用到本身不是字的偏旁之类形体的传统，具有系统性，严谨性，实用性，高效易学易用的特点。

[0039] 上述的非惯用的可选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

附图说明

[0040] 附图用于更好地理解本发明，不构成对本发明的不当限定。其中：

[0041] 图1是根据本发明实施例的汉字排序方法的主要步骤示意图；

[0042] 图2是本发明一个实施例的汉字基础部件双字符码表示意图；

[0043] 图3是本发明一个实施例的首字码编码表示意图；

[0044] 图4是本发明一个实施例的确定汉字的首字码的实现过程示意图；

[0045] 图5是本发明一个实施例的笔画字符编码表示意图；

[0046] 图6是本发明另一个实施例的获取汉字对应的笔画字符串的实现过程示意图；

[0047] 图7是本发明一个实施例的检字表辅助排序表示意图；

[0048] 图8是本发明另一个实施例的检字表辅助排序表示意图；

[0049] 图9是根据本发明实施例的构建字典的方法的主要步骤示意图；

[0050] 图10是本发明一个实施例的检字表示意图；

[0051] 图11是本发明一个实施例的首字表示意图；

[0052] 图12是本发明一个实施例的样品字典首字码示意图；

[0053] 图13是根据本发明实施例的汉字检索方法的主要步骤示意图；

[0054] 图14是本发明一个实施例的电子字典首字表示意图；

[0055] 图15是本发明一个实施例的电子字典检字表示意图；

[0056] 图16是本发明一个实施例的电子字典汉字正文内容示意图；

[0057] 图17是根据本发明实施例的汉字排序装置的主要模块示意图；

[0058] 图18是根据本发明实施例的构建字典的装置的主要模块示意图；

[0059] 图19是根据本发明实施例的汉字检索装置的主要模块示意图；

[0060] 图20是本发明实施例可以应用于其中的示例性系统架构图；

[0061] 图21是适于用来实现本发明实施例的终端设备或服务器的计算机系统的结构示意图。

具体实施方式

[0062] 以下结合附图对本发明的示范性实施例做出说明，其中包括本发明实施例的各种

细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本发明的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0063] 为了解决现有技术中存在的问题,本发明提供了一种简单、快速、准确、不模糊、依据首个基础成字部件的汉字排序、检索法,可用于字典等应用,简称首字法。本发明解决了千百年来部首检字法模糊、不规律、靠运气的问题,以及笔画与笔顺法检索不便的问题。

[0064] 图1是根据本发明实施例的汉字排序方法的主要步骤示意图。如图1所示,本发明实施例的汉字排序方法主要包括如下的步骤S101至步骤S103。

[0065] 步骤S101:按照笔顺获取汉字的首个基础成字部件作为首字,并对首字进行编码得到首字码,首字包括汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件。在本发明的实施例中,所出现的编码(例如:首字码、汉字双字符码、基础部件双字符码等)均以字母码的形式为例进行介绍。在具体实施时,可根据需要进行灵活设定编码形式,例如:字母与数字结合、字母与特殊符号结合等。

[0066] 在现代计算机技术研究成果的基础上,国家发布了《信息处理用 GB13000.1字符集汉字部件规范》以及所附的“汉字基础部件表”。该国家标准的发布为建立一套新的字形检索方法提供了基础。其中,“汉字基础部件表”中定义了560个基础部件。

[0067] 根据以上国家标准,部件是由笔画组成的具有组配汉字功能的构字单位。基础部件是最小的不再拆分的部件,分为基础成字部件(本身就是字,以下简称成字部件)和基础非成字部件(本身不是字,以下简称非字部件)。汉字都可以由基础部件构成。也就是说,GB 13000.1 字符集中包括繁体字和日韩汉字在内的20902个汉字都可以由这560 个基础部件构成。

[0068] 在这560个基础部件中,共有225个成字部件(根据一个部件是否包括在《现代汉语通用字表》中的7000个通用字中来决定是否为成字部件)。其余的335个为非字部件。这335个非字部件占了整个部件数的约60%。而且除了个别偏旁以外,大部分都是人们不熟悉且叫不出名字的形体结构。

[0069] 相反,225个成字部件基本上都是人们常用而且熟悉的独体字。在字典这类大众型的工具书中,能否采用这225个成字部件来检索全部汉字,而不要用到另外335个非字部件呢?

[0070] 通过对大约四千个常用汉字的拆解分析,发现虽然非字部件在汉字中出现的频率很高,用得很广,但是,汉字中不包含任何成字部件的情况却很少。因此,完全可能利用成字部件来检字。这样人们就可以在检索汉字时避开那些不熟悉又叫不出名称的形体符号了。

[0071] 笔画是汉字排检中常用到的一种辅助工具。用得恰当可以帮助用户。

[0072] 为了标记并且在计算机中存储上述560个基础部件中的每一个,考虑到像键盘打字这样的应用,如果能够以罗马字母来代表它们应该是很方便的。但是,一个字母从A至Z共26个,包括大小写总计只有 52个,必然不够用,而利用两个字母,则有 $26 \times 26 = 676$ 个不同组合,能够满足560个部件的需要,并且还有116个空间供扩容。双字母资源的利用率相当高,为83% (560/676)。

[0073] 通过对汉字基础部件进行编码可以更方便地确定每个基础部件。在本发明的一个实施例中,基于将成字部件与非字部件分开、部件按笔画排列和采用双字母进行编码的发

明构思,对汉字基础部件进行编码。具体地,对汉字的基础部件进行编码得到基础部件双字符码的步骤包括:根据《信息处理用GB13000.1字符集汉字部件规范》获取汉字的基础部件;根据基础部件是否为成字部件将基础部件分为成字部件和非字部件,并进行第一排序;根据基础部件的笔画数,分别对成字部件和非字部件进行第二排序;对第二排序后的成字部件和非字部件中笔画数相同的基础部件,分别按照基础部件包括的笔画顺序进行第三排序;将第三排序后的基础部件保存到表中以生成有序码表以得到基础部件双字符码,其中,有序码表的横行和纵列分别以不同类型的字符进行编码。

[0074] 下面结合附图,以双字母码为例,介绍本发明一个实施例的汉字基础部件双字符码的编码过程。图2是本发明一个实施例的汉字基础部件双字符码表示意图。如图2,该汉字基础部件的双字符码表中,按照汉字包括的笔画及笔画顺序将所有560个基础部件编码,每个基础部件用两个字母代表。

[0075] 在编制汉字基础部件双字符码表时,首先将560个基础部件中的成字部件排在前半部分,非字部件排在后半部分。然后,两个部分内部的基础部件排列以笔画数为标准,从小排到大。然后,将笔画数相同的基础部件按笔画顺序排列。最后,将所有以上基础部件按顺序依次填入一个25x24的方格中,从而生成一个有序码表,即根据汉字包括的基础部件的笔画笔顺构建出汉字基础部件双字符码表,其中,一个基础部件所在的行和列决定了该基础部件的双字符码。比如,“木”,处于D行e列,其双字符码即为De。

[0076] 根据本发明的另一个实施例,对汉字的基础部件进行编码得到基础部件双字符码的步骤还可以是:根据《信息处理用GB13000.1字符集汉字部件规范》获取汉字的基础部件;根据基础部件的笔画数对基础部件进行第一排序;对第一排序后的基础部件中笔画数相同的基础部件,按照基础部件包括的笔画顺序进行第二排序;将第二排序后的基础部件保存到表中生成有序码表以得到基础部件双字符码,其中,有序码表的横行和纵列分别以不同类型的字符进行编码。在构建基础部件双字符码表时,还可以根据图2,不将成字部件与非字部件分开,而是混合起来,统一进行排序编码。

[0077] 在汉字基础部件双字符码的基础上根据规范就可将汉字编码,得出汉字双字符码。在本发明的实施例中,将汉字拆分为基础部件,并根据基础部件双字符码得到汉字双字符码具体可以包括以下步骤:根据《信息处理用GB13000.1字符集汉字部件规范》将汉字拆分为基础部件;根据基础部件双字符码,获取每个汉字包括的基础部件的双字符码;对每个汉字,根据《GB13000.1字符集汉字笔顺规范》,按照汉字包括的基础部件在汉字中的笔画顺序,将汉字包括的基础部件的双字符码进行组合得到汉字双字符码。

[0078] 在具体实施时,首先根据《信息处理用GB13000.1字符集汉字部件规范》,将汉字拆分为基础部件。由于规范规定,基础部件不得再行拆分,相离和相接可拆,交重(交叉)不拆,因此拆字的步骤比较简单清晰。比如,“明”拆成“日”和“月”。“韭”拆分为“非”和“一”。当遇到个别的有两种或以上拆分方法的字时,拆分时,部件数要尽可能少,先拆相离,后拆相接,然后,第一个部件尽可能笔画多。

[0079] 将字拆分为基础部件以后,在汉字基础部件双字符码表中可以找到相应的基础部件双字符码。再根据《GB13000.1字符集汉字笔顺规范》,依照字的各个基础部件在字中的笔画顺序,将这些基础部件所对应的双字符码组合在一起,得到该字的双字符码。

[0080] 下面基于图2所示的汉字基础部件双字符码表进行举例说明。

[0081] “木”只含有一个基础部件“木”(部件双字符码De),因此“木”字的编码(即“木”字的双字符码)就是De。

[0082] “样”拆分为两个部件“木”(双字符码为De)和“羊”(双字符码为Hs)。依照两个基础部件的笔画顺序,“样”的编码为DeHs。

[0083] “品”拆成三个“口”(双字符码为Bp),所以“品”的编码为 BpBpBp。

[0084] “汉”拆成“讠”(双字符码为0h)和“又”(双字符码为As),所以,“汉”的编码为0hAs。

[0085] “顺”拆成“川”(双字符码为Bu)、“ㄣ”(双字符码为Kc)和“贝”(双字符码为Du),所以,“顺”的编码为BuKcDu。

[0086] 按照以上编码方法,对GB13000.1字符集中的所有20902个汉字及繁体字和日韩汉字编码,则可以得到完整的汉字双字符码。各个字的码长从单个基础部件的二个字母至多个字母不等。

[0087] 本发明实施例的汉字双字符码包括了汉字的所有基础部件,因而是完整的、全息的,具有单一性和双向性。每个字的编码是唯一的,没有重码,可以从字推出码,从码推出字。本发明实施例的汉字编码可以用于排序和检索等场合。

[0088] 本发明一个实施例的汉字双字符码表如表1所示,其中仅示例性地列出几个汉字和对应的汉字双字符码。

[0089] 表1

汉字	汉字双字符码
木	De
样	DeHs
品	BpBpBp
典	UeAj
汉	0hAs
字	0jCk
笔	UhEc
顺	BuKcDu
码	KcBpLnAa

[0091] 绝大多数汉字至少包含一个成字部件,但是有约一半的汉字并不以成字部件开头。人们在检字时是否乐意寻找不是排在一个字的开头的成字部件呢?通过多人测试,发现人们对这些成字部件太熟悉了,很喜欢用它们来检字,并不在乎它们是否在一个字的开头。因此,这样就引出了本发明的一个重要概念,首个基础成字部件,简称“首字”。对于包含有成字部件的汉字来说,汉字的“首字”也就是在一个汉字中按照笔画顺序所遇到的第一个成字部件,称为成字首字。比如“杆”中的“木”,“汗”中的“干”(“讠”不是成字部件,所以跳过)。

[0092] 然而,并非所有汉字都包括有成字部件,对于那些不包含任何成字部件的汉字,作为一种补充,本发明将选择字的第一个笔画并且归并成五个主笔画部件“一”、“乙”、“丨”、“丿”和“丶”作为首字。但是由于成字部件中已经包括了主笔画部件“一”和“乙”,因此,本发明实施例中将“乙”既作为成字部件,又作为主笔画部件;将“一”仅作为主笔画部件。其中,根据《GB13000.1字符集汉字折笔规范》,将笔画“提”归并为“一”,笔画“竖钩”归并为“丨”,笔画“捺”归并为“丶”,所有“折”笔画都归并为“乙”。这样,所有不含有成字部件的字将按照

字的第一个笔画也含有一个“首字”，称为笔画首字。笔画首字和成字首字即构成了本发明的“首字”。本领域技术人员应当理解，本发明中对笔画首字和成字首字的划分并不是唯一的，“一”和“乙”既可以作为成字首字，也可以作为笔画首字，具体可根据应用场景需求来灵活设定。

[0093] 在得到汉字的首字后，本发明为首字编制双字符码，并且命名为首个基础成字部件双字符码，简称首字码，相应地，首字码也包括成字首字码和笔画首字码。图3是本发明一个实施例的首字码编码表示意图。如图3所示，其中示出了在本发明的实施例中，使用大小写字母编成的双字母码为获取的228个首字进行编码，得到的首字码编码。由于非字部件“日”与成字部件“月”在字形上非常相近，而且经常互换使用，因此，在作为这个用途时，将它与“月”相同对待。类似地，将非字部件“口”（框）与“口”相同对待。

[0094] 在进行首字码编码时，具体是：首先，将所有首字先按笔画数排列，同笔画数再按笔顺排列；然后，按顺序将它们填入一个新建的表中，该表的横行以小写字母进行编码，且小写字母从“a”开始，不超过“y”，纵列以大写字母进行编码，且大写字母从“A”开始，不超过“Y”。在图3中，以首字“一”为例，其首字码为“Aa”。

[0095] 然而，在具体实施过程中，对于成千上万的汉字要逐一确定其首字码时，显然从上述的首字码编码表中进行逐个查询的效率较低，为了通过计算机程序来高效确定汉字的首字和首字码，在本发明的一个实施例中，步骤S101按照笔顺获取汉字的首个基础成字部件作为首字，并对首字进行编码得到首字码，具体可以包括以下步骤：对汉字的基础部件进行编码得到基础部件双字符码；将汉字拆分为基础部件，并根据基础部件双字符码得到汉字双字符码；按照笔顺获取汉字的首个基础成字部件作为首字，建立基础部件与首字的关联关系，并生成基础部件双字符码与首字码的第一匹配表；根据汉字双字符码和第一匹配表确定汉字的首字码。

[0096] 在前面的实施例介绍中，已介绍了如何得到基础部件双字符码和汉字双字符码，下面将结合表2和图4介绍构建的第一匹配表，以及基于汉字双字符码和第一匹配表确定汉字的首字码。如下表2中第一列和第二列示出了本发明一个实施例中第一匹配表所包括的内容。

[0097] 表2

[0098]

基础部件双字符码	首字码	基础部件笔画字符串
Aa	Aa	a
Ab	Ae	V
...
Aj	Am	eg
Bp	Cc	cAa
Bu	Ch	ecc
...
Hs	Jj	feaaac
Kc	Aa	ae
Lg	Ad	fe
Ln	Ae	AX

Oh	Ad	ffb
Oj	Ad	ffC
Ue	Ab	cAacca
Uh	Ac	eafeaf
...
Xh	Ao	ecAaaScAaaAaaAgea

[0099] 在构建第一匹配表时,要结合图2所示的汉字基础部件双字符码表和图3所示的首字码编码表来进行构建。对于图2中的每个基础部件,若该基础部件为成字部件,则其直接对应图3中的某个成字首字,例如:图2中的基础部件“八”,其对应的基础部件双字符码为“Aj”,由于其为一个成字部件,故而在图3可找到其对应的首字码为“Am”,由此即可建立基础部件双字符码与首字码的关联关系。另外一种情况,若该基础部件为非字部件,则获取该非字部件的首个笔画对应的主笔画部件,并将该主笔画部件作为该基础部件对应的首字,例如:图2 中的基础部件“丿”,其对应的基础部件双字符码为“0h”,由于其为一个非字部件,故而需获取该非字部件的首个笔画“丶”对应的主笔画部件“丶”,并建立该主笔画部件的首字码“Ad”与该基础部件双字符码的关联关系。由此,即可建立基础部件与首字的关联关系,并生成基础部件双字符码与首字码的第一匹配表。

[0100] 图4是本发明一个实施例的确定汉字的首字码的实现过程示意图。如图4所示,在本发明的一个实施例中,对每个汉字,根据汉字双字符码和第一匹配表确定汉字的首字码时,主要可以包括以下步骤:获取汉字双字符码包括的基础部件双字符码;根据基础部件双字符码的顺序,依次检查每个基础部件双字符码所对应的第一匹配表中的首字码;对每个基础部件双字符码,若该基础部件双字符码所对应的第一匹配表中的某个首字码为成字首字码,则将该成字首字码作为汉字的首字码;否则,根据第一匹配表,获取汉字双字符码包括的第一个基础部件双字符码对应的首字码作为汉字的首字码。具体地,可以根据汉字双字符码包括的基础部件双字符码的顺序,每次取出一个基础部件双字符码,并检查相应的第一匹配表中的首字码,若该首字码为成字首字码,则将该成字首字码作为首字码;否则,说明该汉字不具有成字部件,此时可获取汉字双字符码包括的第一个基础部件双字符码对应的首字码,并将其作为该汉字的首字码。

[0101] 根据本发明的其中一个实施例,若某汉字的第一个成字部件为“口”,而且该汉字还包括不是“口”的成字部件,则将该汉字的第一个不是“口”的成字部件确定为该汉字的首字。这是由于首字“口”参与组字的合体字比较多,与其他首字参与组合的字不均衡,因此,降低一个字含有首字“口”的机会,有利于字典字在各个首字的均匀分配,利于快速排序和检索。

[0102] 根据上述介绍的步骤S101,即可按照笔画顺序获取汉字的首个基础成字部件作为首字,并编码得到首字码。

[0103] 步骤S102:获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串。具体地,在据预设的笔画字符编码规则获取每个汉字对应的笔画字符串时,本发明的一个实施例中为每个笔画定义一个对应的字母作为笔画的字符编码;对每个汉字所包括的笔画,按照笔顺将所有笔画对应的字符编码拼接得到每个汉字对应的笔画字符串。图5是本发明一个实施例的笔画字符编码表示意图。如图5,对属于同一个

首字的字之间根据《GB3000.1 字符集汉字折笔规范》，按照笔画数与笔顺排序。为了方便笔画之间的比较，每个笔画定义了一个对应的字母作为字符编码。字母靠前，则笔画排序靠前。以“十”字为例，它有两个笔画“一”和“丨”。根据图5，“一”的对应的字符编码为字母“a”，“丨”对应的字符编码为字母“c”，所以，“十”的笔画字符串为“ac”。

[0104] 在本发明的另一个实施例中，根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串具体可以包括步骤：为每个笔画定义一个对应的字母作为笔画的字符编码；对每个基础部件所包括的笔画，按照笔顺将所有笔画对应的字符编码拼接得到每个基础部件对应的笔画字符串；建立基础部件双字符码与基础部件笔画字符串的关联关系并生成第二匹配表；对每个汉字，根据汉字的汉字双字符码得到汉字双字符码包括的基础部件双字符码；对每个基础部件双字符码，从第二匹配表中查找对应的基础部件笔画字符串；按照基础部件双字符码在汉字双字符码中的顺序，将每个基础部件双字符码对应的基础部件笔画字符串拼接得到汉字的笔画字符串。

[0105] 在该实施例中，将汉字拆分为基础部件，并预先根据笔画字符编码规则生成基础部件对应的笔画字符串，即可通过将基础部件对应的笔画字符串进行拼接得到汉字对应的笔画字符串，可以更为快速、方便地得到汉字的笔画字符串。在本发明的实施例中，汉字包括的笔画数，即为汉字的笔画字符串中包括的字符个数。

[0106] 在该实施例中，生成基础部件对应的笔画字符串的过程与前述实施例的生成汉字的笔画字符串的过程类似，此处不再赘述。下面将结合表2和图6介绍构建的第二匹配表，以及基于基础部件笔画字符串确定汉字对应的笔画字符串。在本发明的实施例介绍中，将第一匹配表和第二匹配表整合在前述的表2中进行展示。表2中的第一列和第三列示出了本发明另一个实施例中第二匹配表所包括的内容。

[0107] 在构建第二匹配表时，要结合图2和图5来进行构建。根据基础部件包括的笔画和笔顺，以及图5，即可得到基础部件笔画字符串。例如：图2中的基础部件“八”，其对应的基础部件双字符码为“Aj”，根据其笔画和笔顺，可从图5中得到其笔画字符串为“eg”，如此，即可建立基础部件双字符码和基础部件笔画字符串之间的关联关系。

[0108] 图6是本发明另一个实施例的获取汉字对应的笔画字符串的实现过程示意图。如图6所示，其示出了获取汉字对应的笔画字符串的实现过程，对于每个汉字，首先获取汉字双字符码包括的基础部件双字符码，然后从第二匹配表中查找每个基础部件双字符码的基础部件笔画字符串，具体地，可以每次取出一个基础部件双字符码，并从第二匹配表中查找其对应的基础部件笔画字符串，直至所有基础部件双字符码对应的基础部件笔画字符串都找到后，按照基础部件双字符码在汉字双字符码中的顺序，将每个基础部件双字符码对应的基础部件笔画字符串拼接得到汉字对应的笔画字符串，并将汉字对应的笔画字符串的长度作为汉字包括的笔画数。

[0109] 步骤S103：以首字码作为第一排序字段，笔画数作为第二排序字段，笔画字符串作为第三排序字段，对汉字进行排序，其中，第一排序字段的优先级最高，第三排序字段的优先级最低。其中，在根据首字码进行排序时，按照双字符码的字符顺序依次比较进行排序；在根据笔画字符串进行排序时，按照笔画字符串的字符顺序依次进行比较排序。

[0110] 在对汉字进行排序时，首先新建一个汉字排序表，如表3所示，表中第一列为排序的字，第二列为该字的首字码，第三列为该字的笔画数，第四列为该字的笔画字符串。然后，

将所有首字插入到表3中,具体地,以首字对应的首字码为最高优先级排序字段,笔画数为第二优先级排序字段,笔画字符串为最低优先级排序字段进行排序后,将所有首字插入表3;最后,对于任一待排序的汉字,根据其首字码、笔画数和笔画字符串,与表3中已有的字进行逐个比较,以确定其插入位置。将所有待排序字段均插入表3之后,即可得到排序结果。

[0111] 表3

[0112]	字	首字码	笔画数	笔画字符串
	一	Aa	1	a
[0113]	丨	Ab	1	c
	丶	Ad	1	f
	二	Af	2	aa
	十	Ag	2	ac
	丁	Ah	2	ad
	厂	Ai	2	ae

[0114] 图7是本发明一个实施例的检字表辅助排序表示意图。在本发明的实施例中,以将汉字进行排序并加入字典为例进行介绍。图7示出了将首字进行排序并插入检字表辅助排序表后的一个页面的示意图。在这个阶段,每一行有一个首字,首字之间有一格空行。每行的第一列是首字或以后要加入字典的字,第二列是该字在字典正文中的页数,第三列是该字所含的首字码,第四列是该字的笔画数,第五列是该字用字母符号代表的完整的笔画字符串。

[0115] 一个字在该表中的次序由每行后面三列的内容决定。先比较第三列,然后第四列,最后第五列。对于第三列,在根据首字码进行排序时,按照双字符码的字符顺序依次比较进行排序;对于第四列,根据数字由小到大进行排序;对于第五列,根据笔画字符串进行排序时,按照笔画字符串的字符顺序依次进行比较排序,此处字母中小写字母排在大写字母前面。按照这个规则,一个字只要定义了它自己的这三个参数,就知道应该加入到一个已有的检字表辅助排序表中的什么位置了。

[0116] 图8是本发明另一个实施例的检字表辅助排序表示意图。图8示出了将待排序汉字进行排序并插入检字表辅助排序表后的一个页面的示意图。对于任意一个汉字,根据前述的排序方法即可插入到检字表辅助排序表中。

[0117] 根据上述的步骤S101至步骤S103,即可将首个基础成字部件(首字)作为汉字排序的标准,解决了传统的部首检字法由于部首的定义及位置不固定、缺乏严格的规律性等带来的排序和检索困难的问题,打破了需要用到本身不是字的偏旁之类形体的传统,具有系统性,严谨性,实用性,高效易学易用的特点。

[0118] 根据本发明的另一个方面,提供了一种构建字典的方法。本发明中的字典可以是电子字典,也可以是常用的实体字典。

[0119] 图9是根据本发明实施例的构建字典的方法的主要步骤示意图。如图9所示,本发

明实施例的构建字典的方法主要包括如下的步骤S901 至步骤S906。

[0120] 步骤S901:按照笔顺获取汉字的首个基础成字部件作为首字,并对首字进行编码得到首字码,首字包括汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;

[0121] 步骤S902:获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;

[0122] 步骤S903:以首字码作为第一排序字段,笔画数作为第二排序字段,笔画字符串作为第三排序字段,对汉字进行排序,其中,第一排序字段的优先级最高,第三排序字段的优先级最低;

[0123] 步骤S904:对排序后的每个汉字,获取每个汉字在字典正文中的存储位置,并根据排序后的每个汉字及每个汉字对应的存储位置生成检字表;

[0124] 步骤S905:根据笔画数和笔顺对首字进行排序以生成首字表;

[0125] 步骤S906:通过首字表、检字表和字典正文构建字典。

[0126] 在上述步骤中,步骤S901至步骤S903与步骤S101至步骤S103 的实现方式相同,此处不再赘述。

[0127] 在步骤S904中,检字表用来查一个字在字典正文的具体存储位置,按首字分为许多群,每群以首字开始,包括所有包含该首字的字,按照笔画顺序排列。检字表包括字典里的所有字。用户通过首字表和检字表来查一个字在字典正文中所在的存储位置。对于电子字典来说,一个字在字典正文中的存储位置可能是数据库的某一个表中的某一个页面;对于实体字典来说,一个字在字典正文中的存储位置可能是字典中的某一页。图10是本发明一个实施例的检字表示意图,其中示出了本发明的检字表的其中几页,根据检字表即可查询到一个字在字典正文的具体存储位置。

[0128] 在步骤S905中,根据首字的笔画数和笔顺对首字进行排序,并用于进行首字检索的表本发明称为首字表。这是用户查字时首先要查的表格,是查字的总表格。它告诉用户所有的首字及其在检字表中所在的页数。从这个表,用户找到一个字的首字以及该首字在检字表中的存储位置,然后再到检字表中查找具体的字。对于电子字典来说,首字通过数据位置指针指向检字表,因此,首字表中无需包括首字在检字表中的页面;对于实体字典来说,检字表具有页码,并且,根据笔画数和笔顺对首字进行排序以生成首字表时,是根据笔画数由小到大对首字进行第一次排序;然后对第一次排序后的首字,将笔画数相同的首字按笔画顺序进行第二次排序;之后,根据第二次排序后的首字和每个首字在检字表中的页码生成首字表。图11是本发明一个实施例的首字表示意图,其中,首字表中每个首字下方标示了其所在检字表中的页面,在首字前方的方框内的数字表示了首字的笔画数。

[0129] 另外,根据本发明的一个实施例,在具体生成字典过程中,在对首字进行编码得到首字码之前,还可以根据字典的大小以及所包括的字,对获取的首字进行调整。具体地,根据字典的大小判断是否要对获取的首字进行调整;若是,则统计获取的首字中每个成字部件的出现概率;根据设定的阈值,将出现概率小于阈值,且删除后不会引起混淆的成字部件从获取的首字中删除;或,将一个包含另一个的两个基础部件或两个非常相似的基础部件进行合并。

[0130] 图12是本发明一个实施例的样品字典首字码示意图。本发明的样品字典含有约四

千个字,它的首字码如图12所示,仅含有171个首字。通常情况下,若一个字典字数比较少时,则首字的数目也可以少一点,可对获取的首字进行调整,具体地,首先要统计获取的首字中每个成字部件的出现概率,若出现概率小于预先设定的某个阈值,则说明该首字所构成的字比较少,该首字可认为不是必须的,可以去掉。也可以将一个包含另一个的两个基础部件或两个非常相似的基础部件进行合并,例如:将“犬”合并至“大”,将“方”合并至“万”,将“户”合并至“尸”,将“玉”合并至“王”,将“孑”合并至“子”,等等。另外,对于出现概率较低的某些成字部件,在本字典中用得很少并且去掉后不容易引起混淆,则可直接将其删除。但是,对于个别的成字部件,比如“四”,虽然没有作为首字用在四千字的其它字中,但是为了避免混淆,还是将它保留了。

[0131] 下面结合一个具体的实施例介绍本发明的字典构建方法。在本发明的实施例中,构建字典的过程主要包括以下步骤:

[0132] 1、对构建字典的字进行首字提取和分析,以确定该字典包括的首字,并获取首字码如图12所示,每个首字对应的笔画数、笔画字符串及其在字典正文中的页数;

[0133] 2、将首字插入到如图7所示的检字表辅助排序表中;

[0134] 3、将字典中的每一个字按照以下步骤插入到图7的表中:

[0135] 第一步:在图12的首字码表中找到该字含有的首字以及该首字相应的首字码。注意成字部件不能拆分。比如,如果含有的第一个基础成字部件是“自”,则必须选“自”,不能选“目”;

[0136] 第二步:计算该字含有的笔画数;

[0137] 第三步:写出该字的笔画,并且按照图5将全部笔画转换为一个笔画字符串;

[0138] 第四步:找出该字在字典正文中的页数;

[0139] 第五步:将该字按照它的首字码、笔画数及笔画字符串的先后顺序与图5中的字比较,然后按顺序将该字插入到相应的行;

[0140] 4、将字典中的所有字按以上步骤都插入到图7所示的检字表辅助排序表以后,表中的前两列的字和字典正文的页数就是检字表所需要的内容,将其摘出来,放在一起,并为这些内容进行分页写上页数,检字表就制成了。为了方便阅读,表中每个首字前都空了一行;

[0141] 5、制作完检字表以后,将检字表中的首字的页数填回到图12的首字下方,并且去掉字母码,则制成了图11所示的首字表。表中各首字下方的数字表明该首字在检字表中的页数;

[0142] 6、根据首字表、检字表和字典正文即可构建字典。

[0143] 其中,上述实施例中,如果一个字所含的第一个成字部件为“口”,而且它的后面还有不是“口”的成字部件,则不选“口”,而是选后面的第一个不是“口”的成字部件作为首字,以使得首字均匀分配,提高检索效率。比如,“吓”字,选“下”,不选“口”。

[0144] 另外,如果一个字不包括纯笔画首字以外的任何首字,则根据该字的第一个笔画选择一个主笔画首字,根据《GB13000.1字符集汉字折笔规范》,将笔画“提”归并为首字“一”,将笔画“竖钩”归并为首字“丨”,将笔画“捺”归并为首字“丶”,将所有“折”笔画归并为首字“乙”。

[0145] 由于成字部件本身大都是最简单最常用的独体字,来源于最初的象形字,它们的

造字功能最强,它们在一个合体字中代表的字义成分很大,它们在一个合体字中代表的声音成分也很重,以它们为核心聚集起来的字群就很有鲜明的汉字的群体特征,而且各群字数的差别不像部首法那么明显,分布相对比较平衡,利于快速检字。

[0146] 根据本发明的又一个方面,提供了一种汉字检索方法。

[0147] 图13是根据本发明实施例的汉字检索方法的主要步骤示意图。如图13所示,本发明实施例的汉字检索方法是使用前述所构建的字典进行汉字检索的方法,主要包括如下的步骤S1301至步骤S1303。

[0148] 步骤S1301:按照笔顺获取待检索汉字的首个基础成字部件作为首字;若汉字包括成字部件,则首字为汉字中的首个成字部件;若汉字不包括任何成字部件,则首字为汉字的首个主笔画部件;

[0149] 步骤S1302:从首字表中查找到首字,并得到首字对应的检字表页面;

[0150] 步骤S1303:从检字表页面中查找汉字。

[0151] 下面结合附图14-16介绍如何使用首字法在电子字典上查找“班”字。图14是本发明一个实施例的电子字典首字表示意图;图15是本发明一个实施例的电子字典检字表示意图;图16是本发明一个实施例的电子字典汉字正文内容示意图。首先,在图14的首字表中选择“班”字的首字“王”,则可得到图15所示的“王”的检字表;然后,在图15的检字表中选择“班”字,即可得到图16所示的“班”字的详细解释。整个过程非常简单快速。

[0152] 在本发明的另一个实施例中,若字典为实体字典,则从首字表中查找到首字,并得到首字对应的检字表页面的步骤具体可以是:从首字表中查找到首字,并获取首字在检字表中的页码;根据页码获取首字对应的检字表页面。具体地,在一部以首字表和检字表排序的字典中,相应的检字方法如下:

[0153] 1、查一个字时,先按照笔顺,在首字表中找出该字含有的首字及它在检字表中的页数;

[0154] 2、根据找到的首字在检字表中的页数,在检字表中的相应页数找到该首字的位置;

[0155] 3、紧挨着以上首字所在位置的后面,按照所找汉字的笔画数和笔顺寻找该字。

[0156] 由于一般的首字所包含的字数都不太大,因此这个检字过程一目了然。少数情况下需借助笔画数与笔顺。如果找不到,则该字典不包括该字,绝不模糊。

[0157] 在具体实施过程中,如果一个字的第一个成字部件为“口”,而且它的后面还有不是“口”的成字部件,则找出后面出现的第一个不是“口”的成字部件为首字,并获取该首字在检字表中的页数以进行检字。如果一个字中不包括任何成字部件,则按该字的第一个笔画找出一个笔画首字及该笔画首字在检字表中的页数以进行检字。

[0158] 下面是使用图11的首字表和图10的检字表进行汉字检索的一些实例:

[0159] 查“杆”字,根据笔顺,“木”是该字包含的首字,然后在首字表(图11)中找到“木”及其在检字表中的页数10。然后在检字表(图10)中找到第10页及该页中的首字“木”。“杆”就在其下方,数字显示其在字典正文的页数为第55页。到正文第55页就可以找到该字的详细信息了。

[0160] 查“汗”字。该字的第一个基础部件是“三点水”,是非成字部件,需要跳过它,选择首字“干”。然后在首字表中找到“干”以及其在检字表中的页数5,然后在检字表中找到第5

页及“干”的位置。“汗”就在其下方,数字显示其在字典正文中的页数为第56页。

[0161] 查“叮”字,需要跳过“口”查“丁”这个首字。根据首字表,“丁”在检字表的第2页,然后在检字表的第2页查到首字“丁”及“叮”。

[0162] 查“勾”字,因为该字不包含任何成字部件,所以在首字表中查该字的第一个笔画首字“丿”。根据首字表,“丿”在检字表的第1页,然后在检字表第1页找到“丿”及“勾”字。

[0163] 根据本发明的其中一个实施例,还可以将字的字典详细内容直接放在该字在检字表中的相应位置,从而不必再用数字指向字典正文所在的页。在完成检字表辅助排序以后,保持字的排列顺序不变,但是将每行字的内容都去掉,换成该字在字典中想要写的全部内容即可。然后将检字表改名为字典的正文,原来首字表指向检字表的页数就是首字在字典正文的位置,一步到位。

[0164] 另外,由于繁体字和日韩汉字同简体字使用的成字部件是相同的,因此,本发明同样适用于繁体字和日韩汉字,并且不需要作任何改动。

[0165] 根据本发明的又一个方面,提供了一种汉字排序装置。图17是根据本发明实施例的汉字排序装置的主要模块示意图。如图17所示,本发明实施例的汉字排序装置1700主要包括首字处理模块1701、笔画处理模块1702和多级排序模块1703。

[0166] 首字处理模块1701,用于按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码,所述首字包括所述汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;

[0167] 笔画处理模块1702,用于获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;

[0168] 多级排序模块1703,用于以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低。

[0169] 根据本发明的一个实施例,首字处理模块1701还可以用于:对汉字的基础部件进行编码得到基础部件双字符码;将汉字拆分为基础部件,并根据所述基础部件双字符码得到汉字双字符码;按照笔顺获取汉字的首个基础成字部件作为首字,建立所述基础部件与所述首字的关联关系,并生成基础部件双字符码与首字码的第一匹配表;根据所述汉字双字符码和所述第一匹配表确定汉字的首字码。

[0170] 根据本发明的另一个实施例,首字处理模块1701还可以用于:根据《信息处理用GB13000.1字符集汉字部件规范》获取汉字的基础部件;根据基础部件的笔画数对所述基础部件进行第一排序;对第一排序后的基础部件中笔画数相同的基础部件,按照基础部件包括的笔画顺序进行第二排序;将第二排序后的基础部件保存到表中生成有序码表以得到基础部件双字符码,其中,所述有序码表的横行和纵列分别以不同类型的字符进行编码。

[0171] 根据本发明的又一个实施例,首字处理模块1701还可以用于:根据《信息处理用GB13000.1字符集汉字部件规范》获取汉字的基础部件;根据基础部件是否为成字部件将所述基础部件分为成字部件和非字部件,并进行第一排序;根据基础部件的笔画数,分别对所述成字部件和所述非字部件进行第二排序;对第二排序后的成字部件和非字部件中笔画数相同的基础部件,分别按照基础部件包括的笔画顺序进行第三排序;将第三排序后的基础部件保存到表中以生成有序码表以得到基础部件双字符码,其中,所述有序码表的横行和

纵列分别以不同类型的字符进行编码。

[0172] 根据本发明的又一个实施例,首字处理模块1701还可以用于:根据《信息处理用GB13000.1字符集汉字部件规范》将汉字拆分为基础部件;根据所述基础部件双字符码,获取每个汉字包括的基础部件的双字符码;对每个汉字,根据《GB13000.1字符集汉字笔顺规范》,按照所述汉字包括的基础部件在所述汉字中的笔画顺序,将所述汉字包括的基础部件的双字符码进行组合得到汉字双字符码。

[0173] 根据本发明的又一个实施例,所述首字码包括成字首字码和笔画首字码;首字处理模块1701还可以用于:获取所述汉字双字符码包括的基础部件双字符码;根据基础部件双字符码的顺序,依次检查每个基础部件双字符码所对应的所述第一匹配表中的首字码;对每个基础部件双字符码,若所述基础部件双字符码所对应的所述第一匹配表中的某个首字码为成字首字码,则将该成字首字码作为汉字的首字码;否则,根据所述第一匹配表,获取所述汉字双字符码包括的第一个基础部件双字符码对应的首字码作为汉字的首字码。

[0174] 根据本发明的又一个实施例,若某汉字的第一个成字部件为“口”,而且该汉字还包括不是“口”的成字部件,则将该汉字的第一个不是“口”的成字部件确定为该汉字的首字。

[0175] 根据本发明的又一个实施例,笔画处理模块1702还可以用于:为每个笔画定义一个对应的字母作为笔画的字符编码;对每个汉字所包括的笔画,按照笔顺将所有笔画对应的字符编码拼接得到每个汉字对应的笔画字符串。

[0176] 根据本发明的又一个实施例,笔画处理模块1702还可以用于:为每个笔画定义一个对应的字母作为笔画的字符编码;对每个基础部件所包括的笔画,按照笔顺将所有笔画对应的字符编码拼接得到每个基础部件对应的笔画字符串;建立基础部件双字符码与基础部件笔画字符串的关联关系并生成第二匹配表;对每个汉字,根据所述汉字的汉字双字符码得到所述汉字双字符码包括的基础部件双字符码;对每个基础部件双字符码,从所述第二匹配表中查找对应的基础部件笔画字符串;按照基础部件双字符码在汉字双字符码中的顺序,将每个基础部件双字符码对应的基础部件笔画字符串拼接得到所述汉字的笔画字符串。

[0177] 根据本发明的又一个实施例,所述汉字双字符码、所述基础部件双字符码和所述首字码为使用大小写字母编成的双字母码;并且,在根据所述首字码进行排序时,按照所述双字母码的字符顺序依次比较进行排序;在根据所述笔画字符串进行排序时,按照所述笔画字符串的字符顺序依次进行比较排序。根据本发明的再一个实施例,所述主笔画部件包括“一”,“丨”、“丿”、“丶”和“乙”;并且,汉字排序装置1700还包括笔画归并模块(图中未示出),用于:根据《GB13000.1字符集汉字折笔规范》,将笔画“提”归并为首字“一”,将笔画“竖钩”归并为首字“丨”,将笔画“捺”归并为首字“丶”,将所有“折”笔画归并为首字“乙”。

[0178] 根据本发明的又一个方面,提供了一种构建字典的装置。图18是根据本发明实施例的构建字典的装置的主要模块示意图。如图18所示,本发明实施例的构建字典的装置1800主要包括首字处理模块1701、笔画处理模块1702、多级排序模块1703、检字表生成模块1804、首字表生成模块1805和字典构建模块1806。其中,此处的首字处理模块1701、笔画处理模块1702、多级排序模块1703与汉字排序装置1700中包括的各个模块相同。

[0179] 首字处理模块1701,按照笔顺获取汉字的首个基础成字部件作为首字,并对所述

首字进行编码得到首字码,所述首字包括所述汉字中的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;

[0180] 笔画处理模块1702,获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;

[0181] 多级排序模块1703,以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低;

[0182] 检字表生成模块1804,用于对排序后的每个汉字,获取所述每个汉字在字典正文中的存储位置,并根据排序后的每个汉字及所述每个汉字对应的存储位置生成检字表;

[0183] 首字表生成模块1805,用于根据笔画数和笔顺对所述首字进行排序以生成首字表;

[0184] 字典构建模块1806,用于通过所述首字表、所述检字表和所述字典正文构建字典。

[0185] 根据本发明的一个实施例,所述检字表具有页码,并且,首字表生成模块1805还可以用于:根据笔画数由小到大对所述首字进行第一次排序;对第一次排序后的首字,将笔画数相同的首字按笔画顺序进行第二次排序;根据第二次排序后的首字和每个首字在所述检字表中的页码生成首字表。

[0186] 根据本发明的另一个实施例,构建字典的装置1800还包括首字调整模块(图中未示出),用于在对所述首字进行编码得到首字码之前,根据字典的大小以及所包括的字,对获取的首字进行调整。

[0187] 根据本发明的又一个实施例,首字调整模块(图中未示出)还可以用于:根据字典的大小判断是否要对获取的首字进行调整;若是,则统计获取的首字中每个成字部件的出现概率;根据设定的阈值,将出现概率小于所述阈值,且删除后不会引起混淆的成字部件从所述获取的首字中删除;或,将一个包含另一个的两个基础部件或两个非常相似的基础部件进行合并。

[0188] 首字调整模块的具体实施是通过调整部件双字符码/首字码匹配表中的匹配来完成的。只要换掉一个部件所对应的首字,首字调整模块将把整个汉字排序装置重新运行一遍,得到一个新的字典排列结果,并且给出统计结果供参考。由于动一个首字可能会影响到很多汉字,因此计算量很大。如果没有首字调整模块,这样的调整是不可能做到科学准确的。以下是几个例子。例如:若要把“广”与“厂”合并,只要将部件“广”对应的首字改为部件“厂”所对应的同一个首字码就行了;若要把某个成字部件“文”不作为首字,只需要把该部件对应的首字码改为该部件的第一个笔画“丶”所对应的笔画首字“丶”就行了。总之,非常方便高效。

[0189] 根据本发明的又一个方面,提供了一种汉字检索装置。图19是根据本发明实施例的汉字检索装置的主要模块示意图。如图19所示,本发明实施例的汉字检索装置1900主要包括首字获取模块1901、首字表查找模块1902和检字表查找模块1903。

[0190] 首字获取模块1901,用于按照笔顺获取待检索汉字的首个基础成字部件作为首字;若所述汉字包括成字部件,则所述首字为所述汉字中的首个成字部件;若所述汉字不包括任何成字部件,则所述首字为所述汉字的首个主笔画部件;

[0191] 首字表查找模块1902,用于从首字表中查找到所述首字,并得到所述首字对应的

检字表页面；

[0192] 检字表查找模块1903,用于从所述检字表页面中查找所述汉字。

[0193] 根据本发明的一个实施例,首字表查找模块1902还可以用于:从首字表中查找到所述首字,并获取所述首字在所述检字表中的页码;根据所述页码获取所述首字对应的检字表页面。

[0194] 根据本发明实施例的技术方案,通过按照笔顺获取汉字的首个基础成字部件作为首字,并对首字进行编码得到首字码,首字包括所述汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;以首字码作为第一排序字段,笔画数作为第二排序字段,笔画字符串作为第三排序字段,对汉字进行排序,其中,第一排序字段的优先级最高,第三排序字段的优先级最低。这个技术方案将首个基础成字部件(首字)作为汉字排序和检索的标准,解决了传统的部首检字法由于部首的定义及位置不固定、缺乏严格的规律性等带来的排序和检索困难的问题,打破了需要用到本身不是字的偏旁之类形体的传统,具有系统性,严谨性,实用性,高效易学易用的特点。

[0195] 图20示出了可以应用本发明实施例的汉字排序、汉字检索和构建字典的方法或汉字排序、汉字检索和构建字典的装置的示例性系统架构2000。

[0196] 如图20所示,系统架构2000可以包括终端设备2001、2002、2003,网络2004和服务器2005。网络2004用以在终端设备2001、2002、2003 和服务器2005之间提供通信链路的介质。网络2004可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0197] 用户可以使用终端设备2001、2002、2003通过网络2004与服务器2005交互,以接收或发送消息等。终端设备2001、2002、2003上可以安装有各种通讯客户端应用,例如电子字典应用、网页浏览器应用、搜索类应用、输入法类应用、虚拟键盘类应用等(仅为示例)。

[0198] 终端设备2001、2002、2003可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

[0199] 服务器2005可以是提供各种服务的服务器,例如对用户利用终端设备2001、2002、2003所发来的汉字排序请求提供支持的后台管理服务器(仅为示例)。后台管理服务器可以对接收到的汉字排序请求等数据进行分析等处理,并将处理结果(例如汉字排序结果--仅为示例)反馈给终端设备。

[0200] 需要说明的是,本发明实施例所提供的汉字排序、汉字检索和构建字典的方法可以由服务器2005或终端2001,2002,2003执行,相应地,汉字排序、汉字检索和构建字典的装置也可以设置于服务器2005 或终端2001,2002,2003中。

[0201] 应该理解,图20中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器的。

[0202] 下面参考图21,其示出了适于用来实现本发明实施例的终端设备或服务器的计算机系统2100的结构示意图。图21示出的终端设备或服务器仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0203] 如图21所示,计算机系统2100包括中央处理单元(CPU) 2101,其可以根据存储在只读存储器(ROM) 2102中的程序或者从存储部分 2108加载到随机访问存储器(RAM) 2103中的程序而执行各种适当的动作和处理。在RAM 2103中,还存储有系统2100操作所需的各种程

序和数据。CPU 2101、ROM 2102以及RAM 2103通过总线2104彼此相连。输入/输出 (I/O) 接口 2105也连接至总线2104。

[0204] 以下部件连接至I/O接口2105:包括键盘、鼠标等的输入部分2106;包括诸如阴极射线管 (CRT)、液晶显示器 (LCD) 等以及扬声器等的输出部分2107;包括硬盘等的存储部分 2108;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分2109。通信部分2109经由诸如因特网的网络执行通信处理。驱动器2110也根据需要连接至I/O接口 2105。可拆卸介质2111,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器2110上,以便于从其上读出的计算机程序根据需要被安装入存储部分2108。

[0205] 特别地,根据本发明公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本发明公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分2109从网络上被下载和安装,和/或从可拆卸介质2111被安装。在该计算机程序被中央处理单元 (CPU) 2101执行时,执行本发明的系统中限定的上述功能。

[0206] 需要说明的是,本发明所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器 (RAM)、只读存储器 (ROM)、可擦式可编程只读存储器 (EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器 (CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本发明中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本发明中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0207] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0208] 描述于本发明实施例中所涉及到的单元或模块可以通过软件的方式实现,也可以

通过硬件的方式来实现。所描述的单元或模块也可以设置在处理器中,例如,可以描述为:一种处理器包括首字处理模块、笔画处理模块和多级排序模块。其中,这些单元或模块的名称在某种情况下并不构成对该单元或模块本身的限定,例如,首字处理模块还可以被描述为“用于按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码的模块”。

[0209] 作为另一方面,本发明还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被一个该设备执行时,使得该设备包括:按照笔顺获取汉字的首个基础成字部件作为首字,并对所述首字进行编码得到首字码,所述首字包括所述汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;以所述首字码作为第一排序字段,所述笔画数作为第二排序字段,所述笔画字符串作为第三排序字段,对所述汉字进行排序,其中,所述第一排序字段的优先级最高,所述第三排序字段的优先级最低。

[0210] 根据本发明实施例的技术方案,通过按照笔顺获取汉字的首个基础成字部件作为首字,并对首字进行编码得到首字码,首字包括所述汉字中可独立成字且不可拆分的首个成字部件,及不包括任何成字部件的汉字的首个主笔画部件;获取每个汉字包括的笔画数,并根据预设的笔画字符编码规则获取每个汉字对应的笔画字符串;以首字码作为第一排序字段,笔画数作为第二排序字段,笔画字符串作为第三排序字段,对汉字进行排序,其中,第一排序字段的优先级最高,第三排序字段的优先级最低。这个技术方案将首个基础成字部件(首字)作为汉字排序和检索的标准,解决了传统的部首检字法由于部首的定义及位置不固定、缺乏严格的规律性等带来的排序和检索困难的问题,打破了需要用到本身不是字的偏旁之类形体的传统,具有系统性,严谨性,实用性,高效易学易用的特点。

[0211] 上述具体实施方式,并不构成对本发明保护范围的限制。本领域技术人员应该明白的是,取决于设计要求和因素,可以发生各种各样的修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

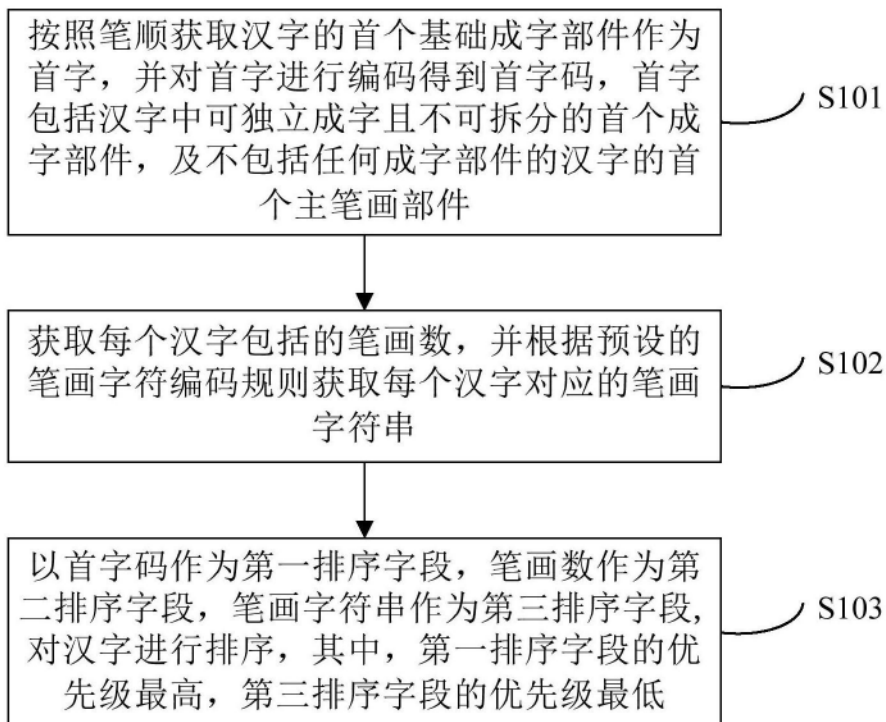


图1

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x
A	一	乙	二	十	丁	厂	七	匕	卜	八	人	乂	儿	入	刀	几	九	乃	又	了	力	刁	乜	三
B	干	于	工	土	士	下	才	寸	丈	兀	大	万	弋	上	门	口	山	巾	小	千	川	个	夕	及
C	久	广	亡	丫	之	尸	己	己	巳	弓	子	孑	女	乡	么	飞	也	习	丸	丰	王	井	夫	
D	韦	专	卅	甘	木	五	丐	长	不	犬	歹	瓦	车	牙	屯	戈	止	日	曰	中	贝	见	内	
E	牛	手	毛	气	片	斤	爪	父	氏	心	月	丹	文	方	火	为	户	尹	央	尺	丑	巴	书	
F	毋	月	未	末	玉	示	戈	甘	世	本	巨	石	戊	龙	东	业	且	目	申	甲	电	田	由	
G	史	四	皿	冉	凸	凹	乍	禾	丘	白	斥	瓜	乎	乐	必	册	立	永	民	弗	矛	皮	母	
H	耳	亚	吏	西	而	夷	曳	虫	曲	肉	缶	年	竹	自	白	舟	兆	衣	羊	米	州	聿	艮	
I	更	酉	束	甫	两	求	豕	里	串	我	身	彡	言	事	雨	果	非	垂	秉	臾	金	承	革	
J	禹	重	禹	食	鬼	戠	黑	象	熏	丨	丿	丿	、	、	冂	冂	フ	冂	冂	乚	乚	乚	乚	
K	丁	丁	フ	ナ	匚	匚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	
L	凡	月	勺	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	勹	
M	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	
N	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	
O	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	
P	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	彡	
Q	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	
R	衣	氏	月	鸟	步	予	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	
S	皿	毒	东	戊	马	步	未	未	未	未	未	未	未	未	未	未	未	未	未	未	未	未	未	
T	勹	鸟	疒	衤	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	丰	
U	卑	貝	史	史	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	曲	
V	長	車	車	曲	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	亞	
W	卑	長	亞	車	車	車	車	車	車	車	車	車	車	車	車	車	車	車	車	車	車	車	車	
X	畢	畢	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	癸	

图2

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
A	一	丨	丿	丶	乙	二	十	丁	厂	七	匕	卜	八	人	乂	儿	入	刀	几	九
B	乃	又	了	力	刁	乚	三	干	于	工	土	士	下	才	寸	丈	兀	大	万	弋
C	上	门	口	山	巾	小	千	川	个	夕	及	久	广	亡	丫	之	尸	己	巳	
D	弓	子	孑	孑	女	乡	么	飞	也	习	九	丰	王	井	夫	韦	专	卅	甘	木
E	五	丐	长	不	犬	歹	瓦	车	牙	屯	戈	止	日	曰	中	贝	见	内	水	牛
F	手	毛	气	片	斤	爪	父	氏	心	月	丹	文	方	火	为	户	尹	夫	尺	丑
G	巴	书	予	毋	片	未	末	玉	示	戈	甘	世	本	巨	石	戊	龙	东	业	且
H	目	申	甲	电	田	由	央	史	四	皿	冉	凸	凹	乍	禾	丘	白	斥	瓜	乎
I	乐	必	册	立	永	民	弗	矛	皮	母	耒	耳	亚	吏	西	而	夷	曳	虫	曲
J	肉	缶	年	竹	自	臼	舟	兆	衣	羊	米	州	聿	艮	臣	更	酉	束	甫	两
K	求	豕	里	串	我	身	豸	言	事	雨	果	非	垂	秉	臾	金	承	革	束	禺
L	重	禹	食	鬼	戢	黑	象	熏												

图3

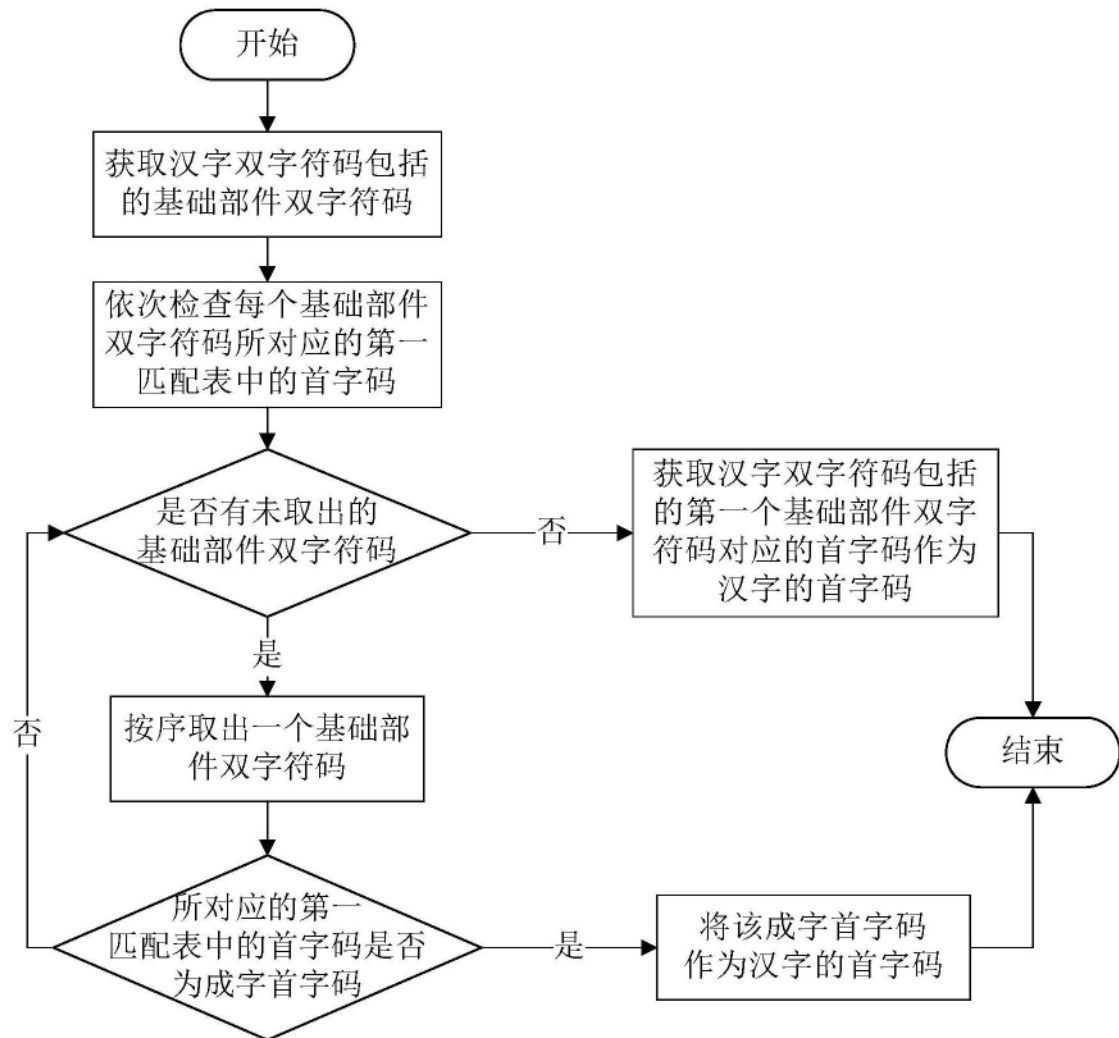


图4

笔画	字符	笔画	字符	笔画	字符
一	a	L	E	ㄥ	P
丿	b	ㄥ	F	ㄣ	Q
丨	c	厶(厶)	G	ㄣ(ㄣ、ㄣ)	R
J	d	<	H	L	S
J	e	ノ	I	ㄣ	T
、	f	丿	J	ㄣ	U
ㄥ	g	ㄥ	K	L(乙)	V
ㄥ(ㄣ)	A	L	L	ㄣ	W
ㄣ(ㄣ)	B	L	M	ㄣ(ㄣ)	X
ㄣ	C	ㄥ	N	ㄣ(ㄣ)	Y
L(L、L)	D	ㄣ(ㄣ)	O		

图5

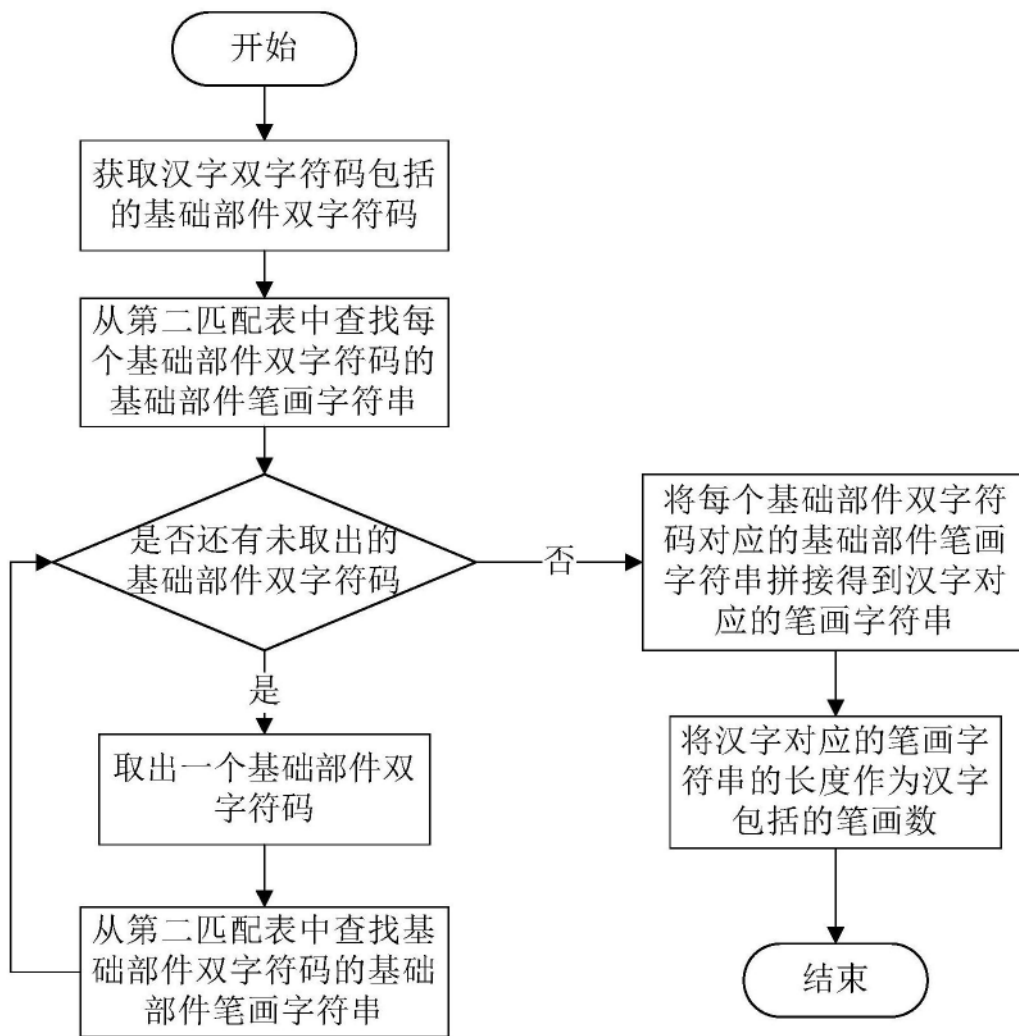


图6

汉字	在字典正文中的页数	首字码	笔画数	笔画字符串
一	167	Aa	1	a
[]		Ab	1	c
[丿]	65	Ac	1	e
[丶]		Ad	1	f
二	167	Af	2	aa
十	170	Ag	2	ac
丁	158	Ah	2	ad
厂	151	Ai	2	ae
乚	169	Aj	2	bS
匕	83	Ak	2	eS
卜	105	Al	2	cf
八	172	Am	2	cg
人	1	An	2	eg
乂	173	Ao	2	eg
入	83	Aq	2	eK
刀	85	Ar	2	Oe

图7

汉字	在字典正文中的页数	首字码	笔画数	笔画字符串
絆	313	Hd	10	HHbffaaceg
永	273	He	5	fOBeg
咏	273	He	8	cAafOBeg
泳	273	He	8	ffbfOBeg
漾	273	He	14	ffbfeaacafoBeg
弗	414	Hf	5	AaXec
佛	414	Hf	7	ecAaXec
拂	414	Hf	8	adbAaXec
沸	414	Hf	8	ffbAaXec
绋	414	Hf	8	HHbAaXec
费	414	Hf	9	AaXeccEef
矛	58	Hg	5	Cfcde
茅	59	Hg	8	accCfcde
柔	59	Hg	9	Cfcdeaceg
揉	58	Hg	12	adbCfcdeaceg
商		Hg	12	CfcdecOcgcaA
皮	435	Hh	5	BceBg
披	435	Hh	8	adbBceBg
彼	435	Hh	8	eecBceBg
波	435	Hh	8	ffbBceBg
皱	504	Hh	10	eCAaaBceBg
疲	361	Hh	10	faeffBceBg
被	435	Hh	10	fBcffBceBg
菠	126	Hh	11	accffbBceBg
颇	469	Hh	11	BceBgaecEef
婆	436	Hh	11	ffbBceBgHea
母	203	Hi	5	DOfaf
拇	204	Hi	8	adbDOfaf
侮	205	Hi	9	ecaeDOfaf
海	204	Hi	9	fNaeDOfaf
毒	204	Hi	11	aacaaeDOfaf

图8

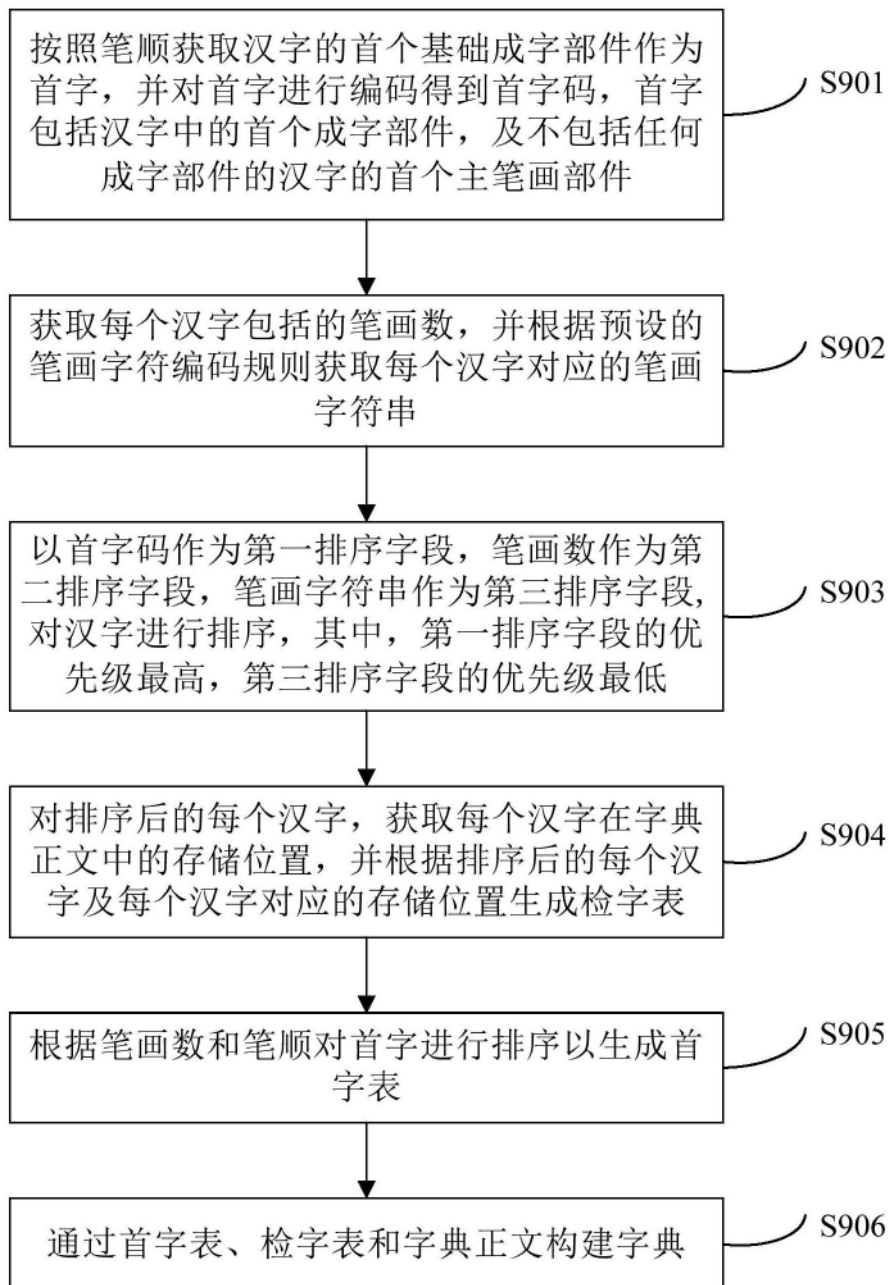


图9

教 36	姆 204	嫉 30	垫 207	宝 114
哮 36	姓 426		摅 207	珑 97
郭 432	始 161	么 193	热 428	玻 452
浮 6	要 30	幻 194		玷 106
淳 433	按 113	玄 194	丰 256	珍 352
猛 87	姜 270	幼 193	聿 378	玲 97
敦 433	姨 200	吆 193	契 197	珊 98
孰	娃 424	系 75	拜 295	珠 283
淳 433	姥 36	拗 4	艳 39	珽 402
锰	怒 372	呦 451	逢 379	班 97
塾	姻 197	兹 196	害 237	莹 483
孵 190	姝	素 76	彗 5	遑 401
熟 433	娇 497	畜 195	铎 379	逛 382
锰 451	婉 202	率 195	湃 443	球 303
孺 383	姚 189	滋 196	蓬 379	理 118
	娜 220	蓄 195	缝 379	琐 97
女 28	馁 356	搐	慧 5	琅 97
奴 372	凄 358	慈 196	篷 379	琉
奶 157	案 113	拌 196	豁 318	琴 407
妆 244	娛 201			琳 97
安 113	媧 496	也 23	王 59	琢 304
妇 407	恕 30	他 24	玉 96	琮 456
奸 56	娘 386	驰 24	主 90	斑 454
妈 452	绥 357	池 24	匡 382	筐 382
好 29	菇 127	拖 108	弄 96	善 270
如 29	婚 260		玛 452	瑟 407
妒 30	婢 448	习 136	玖 97	瑚 98
她 24	婉 359	羽 136	呈 400	璽 374
妥 356	媚 14	诩	住 90	瑞 382
妍	媒 217	寥 353	狂 381	瑙 452
妓 64	絮 29	翔 270	汪 382	碧 431
努 372	婿 215	谬 353	玨	瑰 97
妙 29	嫂 393	蓼 353	玩 285	璃 452
妖 201	媾 222	寥 353	玮	
姊 29	媳 21	翟	环 82	井 117
妣 41	媿 41	翠 297	现 468	讲 222
妨 336	嫌 45	翼 366	玫 97	阱 117
妻 358	嫁 29	戳 56	拄 90	进 377
妹 257	嫩 47		驻 90	
姑 29	嫡 71	丸 207	国 326	夫 200
弩 372	嬉 327	执 207	往 381	失 206
姐 99	嬲 441	势 428	注 90	夹 466
#####10#####				
芙 128	村 306	查 436	梳 28	楼 466

图10(a)

忍 296	九 7	拔 305	瘦 392	癍 363
招 453	仇 8	拨 502	瘫 363	
券 449	杂 416	艰 343		三 169
沼 453	旭 8	姦	了 26	兰 389
绍 454	抛 8	怪 226	辽 392	拦 389
涩 296	鸠 7	贤 470	亨	
筍 454	染 121	肾 471	疗 392	干 55
梁 297		受 12	丞 48	午 50
梁 297	乃 157	变 480	亟 168	刊 56
解 351	仍 157	叔 293	函 48	平 224
懈 20	扔 157	泼 502	拯 49	汗 56
蟹 131	孕 157	段	哼 451	许 50
	盈 157	蚤 293	烹 433	评 224
几 91	隼 419	叟 392	涵 48	莘 450
凡 86	携 419	侵 408	蒸 49	竿 55
冗 333	镌 419	471		萍 126
爻		爰 13	力 6	
亢 294	又 4	叛 313	为 148	于 223
讥 92	叉 293	译 473	办 68	迂 223
叽 92	友 5	择 473	另 320	芋 223
饥 95	双 5	驿 473	务 58	吁 125
鳧 91	劝 418	桑 120	加 328	宇 223
朵 415	邓 378	难 344	边 377	
设 72	支 46	紧 471	伪 148	工 84
投 4	仅 343	爰 10	劳 482	巧 224
抗 294	反 185	浸 408	苏 127	功 84
吭 22	凤 133	泽 473	别 320	左 5
役 111	发 305	绎 473	茄 450	式 261
沉 333	圣 401	授 13	拐 320	巩 85
没 443	对 246	假	咖 23	扛 85
咒 23	叹 343	缀 75	咖 22	江 442
刹	发 355	搔 293	勉 202	红 85
沿 316	汉 344	搜 392	驾 329	贡 85
段 434	戏 93	援 13	荔 127	汞 85
亮 116	观 417	骚 293	架 329	攻 84
疫 361	欢 418	缓 13	贺 329	巫 79
殷 289	扳 185	摊 344	捞 482	劲 481
铅 317	报 310	叠 139	捌 320	试 261
缎 434	鸡 198	噪 22	唠 482	经 480
凳 378	饭 185	滩 344	涝 482	项 468
锻 434	坚 471	寝 408	勇 254	拭 261
	返 185	敲 46	伤 3	差 463
#####5#####				
差 463	坯 265	致 241	趟 326	守 113

图10 (b)

泖 190	民 263	运 376	载 226	矗 281
洲 271	疋	尝 324	翅 137	
涕 415	出 215	偿 325	真 410	丁 158
冢	收 243	魂 364	葶 128	订 159
雀 418	丝 75		索 429	打 159
衰 82	纠 243	十 170	悖 428	叮 158
疼 361	约 239	玆 257	值 281	宁 104
准 228	巡 272	支 64	盾 3	行 110
兼 45	那 220	斗 54	读 476	钉 158
羔 270	玆 207	什 170	啤 451	顶 159
瓶 291	即 94	计 71	做 322	拧 105
流 28	肃 478	卉	商	狩 142
谁 72	隶 184	古 321	续 476	泞 105
档 478	录 503	叶 412	博 253	衍 111
淮 442	茎 480	汁 442	裁 226	亨 159
液 295	承 48	讯 258	韩 440	停 159
淀 211	径 481	协 7	戟	衍 111
宦 205	练 388	汛 257	朝 439	
敞 64	终 463	迅 257	率 74	厂 151
诵 254	降 208	克 286	葫 128	广 287
慨 95	既 94	孛 428	喷 380	厅 17
禅 182	甬 254	技 64	愤 380	历 399
溉 95	绑 256	抖 54	湖 322	厄
涌 254	剥 503	吱 450	渤 429	厉 285
谦 45	绊 313	针 282	献 91	扩 287
痛 255	陆 112	估 321	鄙 171	压 186
弊 65	逮 377	卖 476	嗦 23	仄 186
憔 20	乞 272	直 281	置 281	危 355
鹤 419	绿 503	苦 322	慎 410	庄 244
	维 134	咕 450	循 3	庆 146
乙 24	綻 76	固 321	截 56	产 455
刁 311	通 254	阜 258	兢 286	扼 4
马 143	吃 273	卑 298	摘 71	辰 234
亿 19	迄 273	质 233	喃 71	励 285
乡 194	疙 363	沽 321	滴 71	床 245
卫 487		贲 380	镇 410	库 484
飞 258	二 167	哉 225	德 281	序 192
与 238	亏 224	故 322	潮 439	庐 404
艺 428	云 282	胡 322	颠 410	庇 41
书 367	仁 167	南 429	翰 440	应 133
#####2#####				
沥 399	裤 81	唬 145	叭 23	翁 160
厓	渡 140	虚 151	只 316	颁 470
厕 394	搪 49	彪 145	穴 397	俱 15
庚 49	厦 9	滤 42	共 47	俊 162

图10(c)

开 92	豕 304	入 83	制 63	、
开 192	東 388	么 399	臾 222	丫
无 346	荆 192	勺 239	使 45	尤 37
专 500	蕩 505	片 243	爬 131	写 52
丐 155	药 464	升 55	卵 190	冯 463
互 238	拷 224	长 36	钙 451	礼 182
扎 25	挣 13	爪 5	钧 180	归 227
ㄥ 371	挽 203	印 360	钩 164	汇 442
尤 206	拼 291	勿 229	卸 51	农 502
瓦 256	挡 478	丹 117	笃	亦 295
节 309	拚 291	匀 180	追 227	并 291
戍	挪 220	勾 164	禹 129	州 271
邦 256	奉 269	乌 309	佣 251	弟 415
夷 200	泰 444	冬 463	角 351	汤 505
刑 191	髟 354	刂 503	伴 313	忧 9
邢 191	莺 483	印 310	饼 292	忱 37
吏 45	拥 195	乎 104	狰 142	弃 27
考 224	拌 313	乐 464	第 415	半 313
西 461	萧 479	犯 141	逸 148	判 313
芍 240	捷 4	甩 251	修 332	沛 460
扫 407	挽 149	匆 229	倦 449	沈 37
扬 505	推 228	册 100	谏 222	沟 221
麦 284	掖 295	钅	铐 224	宏 371
击 485	掸 78	年 286	铆 190	郎 385
形 191	救 303	夆 208	铛 478	夜 295
茺 346	棘 187	争 12	俑	疟 145
抚 346	捅 255	传 500	饯 149	瘍 362
扰 10	雄 371	优 9	馆 228	衫 154
抓	捧 269	免 202	卿 94	净 13
抑 360	蔽 64	仰 360	策 187	卷 448
求 303	蕉 128	伫 105	箴 124	剃 415
丽 146		卯 189	象 147	泻 52
走	龟 132	兔 148	煞 278	单 78
表 303	凹 154	攸 332	焦 393	官 227
事 44	凸 154	迎 360	御 50	定 211
茁 216	曲 154	犹 250	锥 410	冤 148
范 484	当 478	用 251	锭 211	迹 295
刺 186	罚 72	删 100	雏 504	养 269
枣 186	临 16	彤 154	熏 387	逆 35
拣 388	冕 202	秉 45	箫 479	递 415
拙 216	丿 65	钧 240	管 228	洒 146
#####1#####			像 148	
浓 503	忆 19	动 350	勃 429	戴 366
派 274	毋	污	栽 225	乾 25

图10 (d)

1	一	丨	丿	、	乙	2	二	十	丁	厂	七	匕	卜	八	人	乂	儿	刀
1	1	1	1	1	2	2	2	2	2	3	3	3	3	4	4	4	4	4
几	九	乃	又	了	力	3	三	干	于	工	土	士	下	才	寸	丈	兀	大
5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6
万	弋	上	门	口	山	巾	小	千	川	夕	及	久	亡	之	尸	己	巳	巳
7	7	7	7	7	8	8	8	8	9	9	9	9	9	9	9	9	9	9
弓	子	女	么	也	习	丸	4	丰	王	井	夫	韦	甘	木	五	不	歹	车
9	9	10	10	10	10	10	10	10	10	10	10	11	11	11	11	12	12	12
牙	屯	戈	止	日	曰	中	贝	见	内	水	牛	手	毛	气	斤	父	氏	心
12	12	12	12	12	13	13	13	13	13	13	13	14	14	14	14	14	14	14
月	文	火	尹	夬	尺	丑	巴	予	爿	5	未	未	示	戈	甘	世	本	巨
14	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
石	戊	龙	东	业	且	目	申	甲	电	田	由	央	史	皿	四	冉	乍	禾
15	15	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	17
丘	白	斥	瓜	必	立	永	弗	矛	皮	母	6	耒	耳	亚	而	虫	缶	竹
17	17	17	17	17	17	17	17	17	17	17	17	17	18	18	18	18	18	18
自	臼	舟	兆	衣	羊	米	聿	艮	7	臣	更	酉	束	甫	两	豕	里	串
18	18	18	18	18	18	18	19	19	19	19	19	19	19	19	19	19	19	19
我	身	豸	言	8	雨	果	非	垂	革	禺	重	鬼	黑					
19	19	19	19	19	19	19	19	19	19	20	20	20	20					

图11

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
A	一	丨	丿	、	乙	二	十	丁	厂	七	匕	卜	八	人	乂	儿	刀	几	九
B	乃	又	了	力	三	干	于	工	土	士	下	才	寸	丈	兀	大	万	弋	上
C	门	口	山	巾	小	千	川	夕	及	久	亡	之	尸	己	已	巳	弓	子	女
D	么	也	习	丸	丰	王	井	夫	韦	甘	木	五	不	歹	车	牙	屯	戈	止
E	日	曰	中	贝	见	内	水	牛	手	毛	气	斤	父	氏	心	月	文	火	尹
F	夬	尺	丑	巴	予	爿	未	未	示	戈	甘	世	本	巨	石	戊	龙	东	业
G	且	目	申	甲	电	田	由	央	史	皿	四	冉	乍	禾	丘	白	斥	瓜	必
H	立	永	弗	矛	皮	母	耒	耳	亚	而	虫	缶	竹	自	臼	舟	兆	衣	羊
I	米	聿	艮	臣	更	酉	束	甫	两	豕	里	串	我	身	豸	言	雨	果	非
J	垂	革	禺	重	鬼	黑													

图12

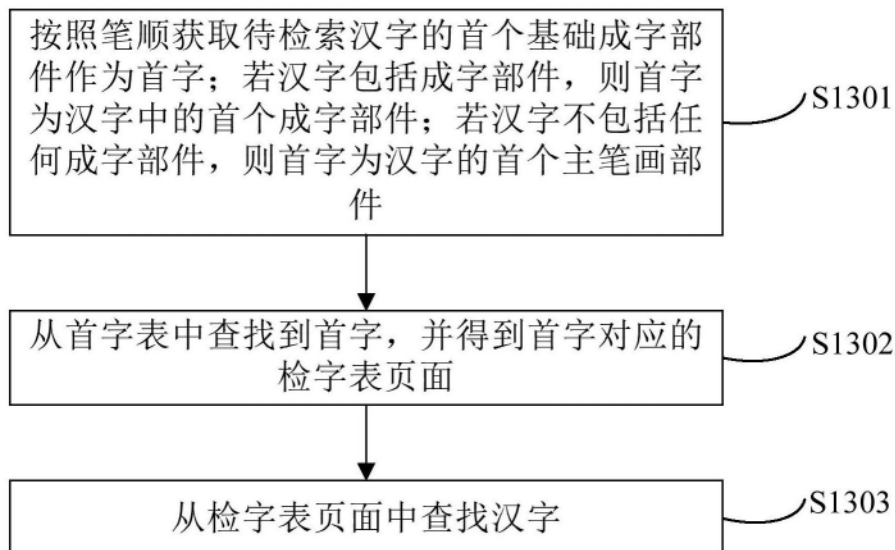


图13



图14



图15



图16

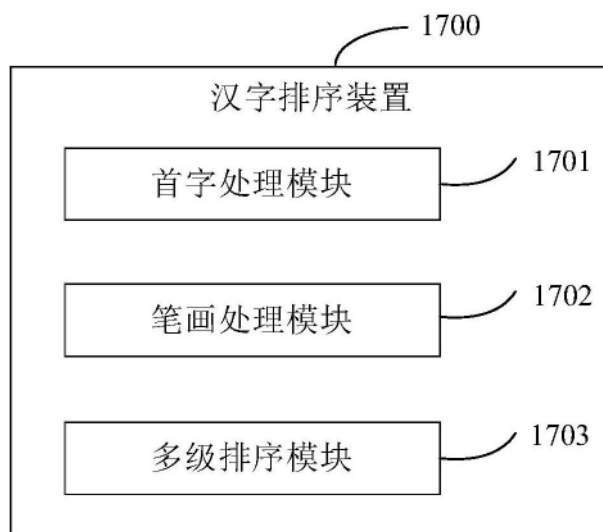


图17

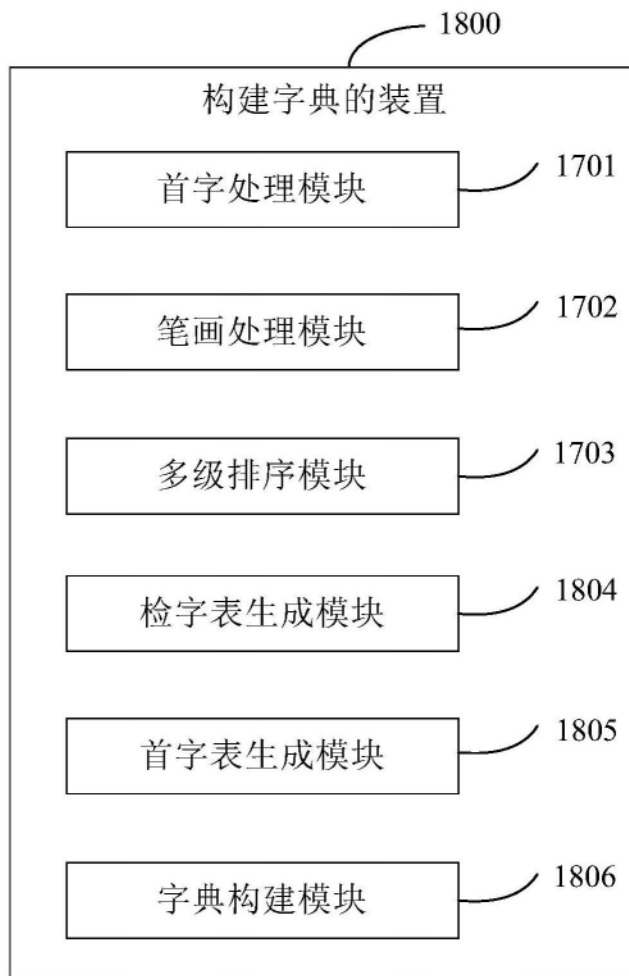


图18

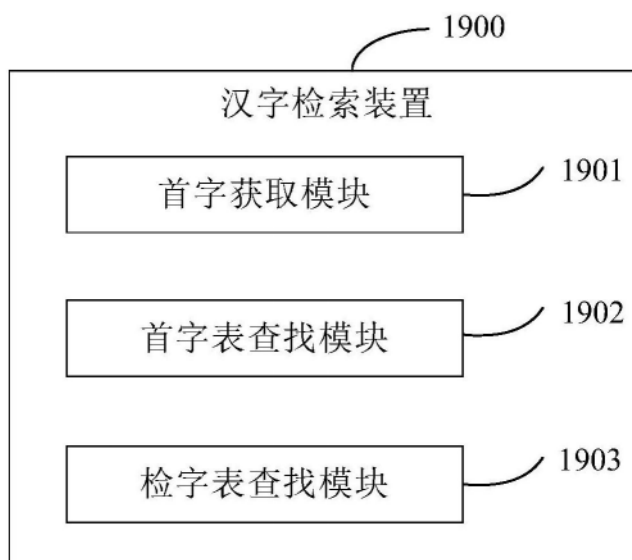


图19

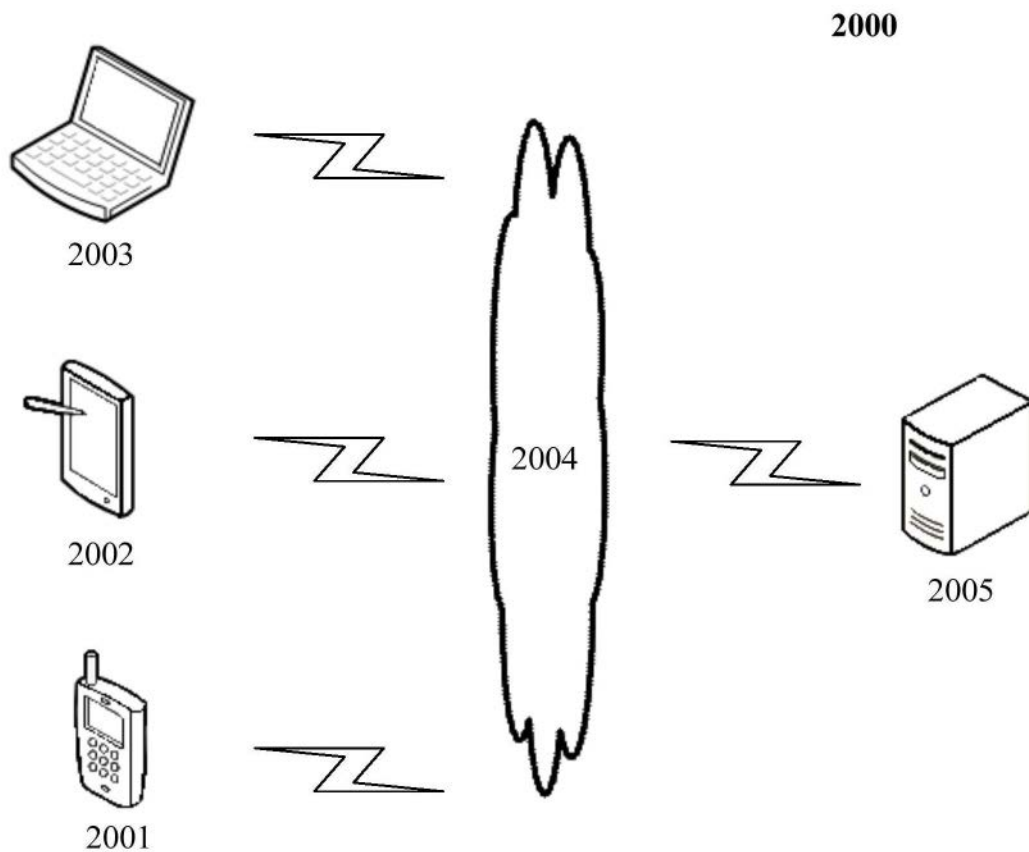


图20

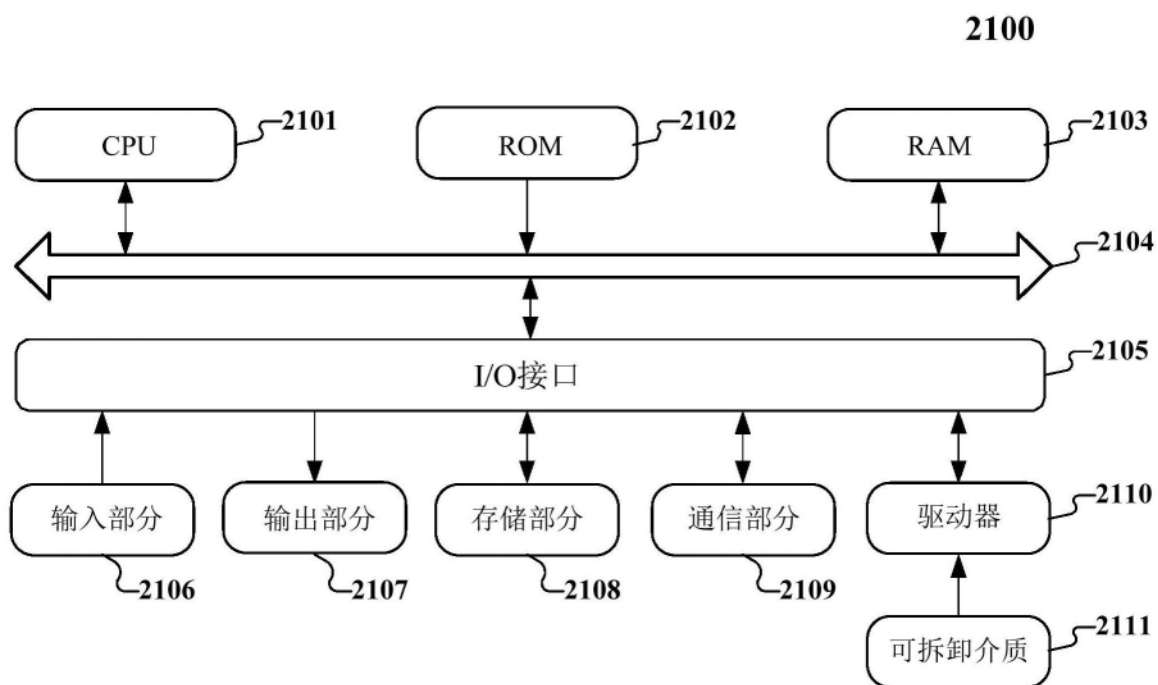


图21