

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022 年 5 月 19 日 (19.05.2022)



(10) 国际公布号
WO 2022/100154 A1

(51) 国际专利分类号:
G06F 17/00 (2019.01)

(21) 国际申请号: PCT/CN2021/109461

(22) 国际申请日: 2021 年 7 月 30 日 (30.07.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202011278707.0 2020 年 11 月 16 日 (16.11.2020) CN

(71) 申请人: 深圳壹账通智能科技有限公司(ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。

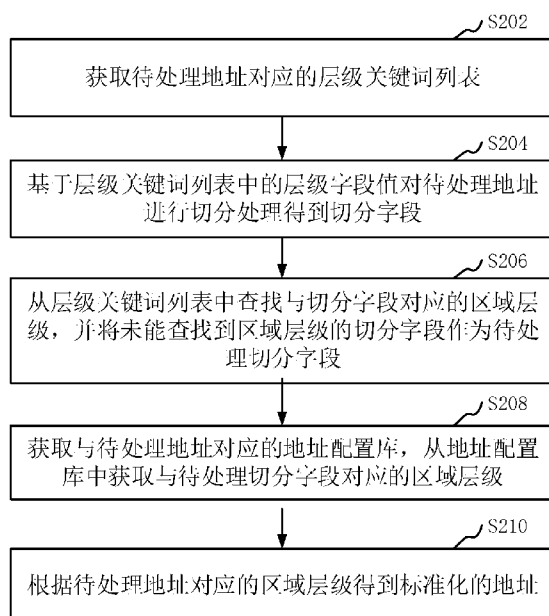
(72) 发明人: 李超(LI, Chao); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。 徐国强(XU, Guoqiang); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司(ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路 6 号 4501 房 (部位: 自编 01-03 和 08-12 单元) (仅限办公用途), Guangdong 510623 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT,

(54) Title: ARTIFICIAL INTELLIGENCE-BASED ADDRESS STANDARDIZATION METHOD AND APPARATUS, DEVICE AND STORAGE MEDIUM

(54) 发明名称: 基于人工智能的地址标准化方法、装置、设备和存储介质



S202 ACQUIRE A HIERARCHICAL KEYWORD LIST CORRESPONDING TO AN ADDRESS TO BE PROCESSED
S204 SEGMENT THE ADDRESS ON THE BASIS OF HIERARCHICAL FIELD VALUES IN THE HIERARCHICAL KEYWORD LIST TO OBTAIN SEGMENTED FIELDS
S206 SEARCH FOR AREA LEVELS CORRESPONDING TO THE SEGMENTED FIELDS FROM WITHIN THE HIERARCHICAL KEYWORD LIST, AND USE A SEGMENTED FIELD THE AREA LEVEL OF WHICH CANNOT BE FOUND AS A SEGMENTED FIELD TO BE PROCESSED
S208 ACQUIRE AN ADDRESS CONFIGURATION LIBRARY CORRESPONDING TO THE ADDRESS, AND ACQUIRE AN AREA LEVEL CORRESPONDING TO THE SEGMENTED FIELD TO BE PROCESSED FROM WITHIN THE ADDRESS CONFIGURATION LIBRARY
S210 OBTAIN A STANDARDIZED ADDRESS ACCORDING TO THE AREA LEVEL CORRESPONDING TO THE ADDRESS

图 2

(57) Abstract: An artificial intelligence-based address standardization method, which relates to the technical field of big data, and comprises: acquiring a hierarchical keyword list corresponding to an address to be processed (S202); segmenting the address on the basis of hierarchical field values in the hierarchical keyword list to obtain segmented fields (S204); searching for area levels corresponding to the segmented fields from within the hierarchical keyword list, and using a segmented field the area level of which cannot be found as a segmented field to be processed (S206); acquiring an address configuration library corresponding to the address, and acquiring an

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

area level corresponding to the segmented field to be processed from within the address configuration library (S208); and obtaining a standardized address according to the area level corresponding to the address (S210).

(57) 摘要: 一种基于人工智能的地址标准化方法, 涉及大数据技术领域, 包括: 获取待处理地址对应的层级关键词列表 (S202); 基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段 (S204); 从层级关键词列表中查找与切分字段对应的区域层级, 并将未能查找到区域层级的切分字段作为待处理切分字段 (S206); 获取与待处理地址对应的地址配置库, 从地址配置库中获取与待处理切分字段对应的区域层级 (S208); 根据待处理地址对应的区域层级得到标准化的地址 (S210)。

基于人工智能的地址标准化方法、装置、设备和存储介质

相关申请的交叉引用

本申请要求于 2020 年 11 月 16 日提交中国专利局，申请号为 202011278707.0，申请
5 名称为“地址标准化方法、装置、计算机设备和存储介质”的中国专利申请的优先权，其
全部内容通过引用结合在本申请中。

技术领域

本申请涉及大数据技术领域，特别是涉及一种基于人工智能的地址标准化方法、装置、
10 设备和存储介质。

背景技术

随着物流技术的发展，快递在我们生活中越来越普遍，在寄快递的过程中都需要填写
地址。而在一些情况下，地址是由人工填写，存在地址填写不规范的问题。

15 在传统方式中，可以通过人工的方式对不规范的地址进行标准化处理。

但是，发明人意识到，通过人工进行地址标准化的处理，使得地址的标准化处理效率
低下。

发明内容

20 根据本申请公开的各种实施例，提供一种基于人工智能的地址标准化方法、装置、设
备和存储介质。

一种基于人工智能的地址标准化方法，包括：

获取待处理地址对应的层级关键词列表；

基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段；

25 从层级关键词列表中查找与切分字段对应的区域层级，并将未能查找到区域层级的切
分字段作为待处理切分字段；

获取与待处理地址对应的地址配置库，从地址配置库中获取与待处理切分字段对应的
区域层级；及

根据待处理地址对应的区域层级得到标准化的地址。

30 一种基于人工智能的地址标准化装置，包括：

第一获取模块，用于获取待处理地址对应的层级关键词列表；

切分模块，用于基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到
切分字段；

查找模块，用于从层级关键词列表中查找与切分字段对应的区域层级，并将未能查找

到区域层级的切分字段作为待处理切分字段；

第二获取模块，用于获取与待处理地址对应的地址配置库，从地址配置库中获取与待处理切分字段对应的区域层级；及

标准化模块，用于根据待处理地址对应的区域层级得到标准化的地址。

- 5 一种计算机设备，包括存储器和一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取待处理地址对应的层级关键词列表；

基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段；

- 10 从层级关键词列表中查找与切分字段对应的区域层级，并将未能查找到区域层级的切分字段作为待处理切分字段；

获取与待处理地址对应的地址配置库，从地址配置库中获取与待处理切分字段对应的区域层级；及

根据待处理地址对应的区域层级得到标准化的地址。

- 15 一个或多个存储有计算机可读指令的计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

获取待处理地址对应的层级关键词列表；

基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段；

- 20 从层级关键词列表中查找与切分字段对应的区域层级，并将未能查找到区域层级的切分字段作为待处理切分字段；

获取与待处理地址对应的地址配置库，从地址配置库中获取与待处理切分字段对应的区域层级；及

根据待处理地址对应的区域层级得到标准化的地址。

- 25 上述基于人工智能的地址标准化方法、装置、计算机设备和存储介质，获取待处理地址对应的层级关键词列表；基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段；层级关键词列表中查找与切分字段对应的区域层级，并将未能查找到区域层级的切分字段作为待处理切分字段；获取与待处理地址对应的地址配置库，从地址配置库中获取与待处理切分字段对应的区域层级；根据待处理地址对应的区域层级得到标准化的地址。通过预先配置地址配置库以及层级关键词列表，然后在具体实施中就可以直接
30 根据预配置库获取对应的信息，进而根据获取到的信息实现对待处理地址的切分处理，进而得到标准化地址，提高了对地址的标准化处理效率。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

- 35 为了更清楚地说明本申请实施例中的技术方案，下面将对实施例中所需要使用的附图

作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其它的附图。

图 1 为根据一个或多个实施例中基于人工智能的地址标准化方法的应用环境图。

图 2 为根据一个或多个实施例中基于人工智能的地址标准化方法的流程示意图。

5 图 3 为根据一个或多个实施例中提供了一种基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段的流程示意图。

图 4 为根据一个或多个实施例中基于人工智能的地址标准化装置的结构框图。

图 5 为根据一个或多个实施例中计算机设备的内部结构图。

具体实施方式

10 为了使本申请的技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

本申请提供的基于人工智能的地址标准化方法,可以应用于如图 1 所示的应用环境中。其中,终端 102 通过网络与服务器 104 进行通信。服务器 104 获取待处理地址对应的层级关键词列表;基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字
15 段;从层级关键词列表中查找与切分字段对应的区域层级,并将未能查找到区域层级的切分字段作为待处理切分字段;获取与待处理地址对应的地址配置库,从地址配置库中获取与待处理切分字段对应的区域层级;根据待处理地址对应的区域层级得到标准化的地址。并将标准化地址推送至终端 102。其中,终端 102 可以但不限于各种个人计算机、笔记本
20 电脑、智能手机、平板电脑和便携式可穿戴设备,服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

在其中一个实施例中,如图 2 所示,提供了一种基于人工智能的地址标准化方法,该基于人工智能的地址标准化方法具体可以应用于计算机设备,其中计算机设备具体可以是服务器或者终端,可以包括以下步骤:

25 步骤 202,获取待处理地址对应的层级关键词列表。

其中,待处理地址是需要进行标准化处理的地址,标准化处理是指获知待处理地址中包括的区域层级的处理过程,如将待处理地址中的从属于不同区域层级的地址字段进行切分处理得到切分单元,并得到每一个切分单元对应的区域层级的过程。其中,区域层级是行政区域对应的行政等级,如可以是“省”、“市”或者“县”等,区域层级还可以包括“region”、
30 “unit”等。

其中,层级关键词列表与区域层级对应。具体地,可以预先为每一个区域层级构建一个对应的层级关键词列表。

在其中一个实施例中,计算机设备获取待处理地址,并将待处理地址与层级关键词列表中的区域层级进行匹配,并将匹配成功的区域层级对应的层级关键词列表作为与待处理
35 地址对应的层级关键词列表。需要说明的是,当待处理地址中包括一个或者多个区域层级

时,此时计算机设备可以从待处理地址中获取与每一个区域层级分别对应的层级关键词列表。例如,当待处理地址为“山东省日照市东港区****”时,此时待处理地址中包括的区域层级为“省”、“市”以及“区”,并且,计算机设备可以获取与每一个区域层级分别对应的层级关键词列表。

- 5 在本实施例中,层级关键词列表中可以包括区域层级对应的层级信息,层级信息中可以包括但不限于该区域层级对应的层级字段值、层级关键词、以及层级位置信息中的一个或者多个。

10 在本实施例中,计算机设备获取待处理地址对应的区域层级,并根据区域层级获取对应的层级关键词列表,以在层级关键词列表中查找区域层级对应的层级字段值、层级关键词以及层级位置关系等。

15 例如,一个具体的待处理地址为:“FLT 11 22/F BLK 33 AAA HSE BBB EST CCC STREET SHAM SHUI PO KLN”,该待处理地址中的区域层级包括“region”、“unit”。故而计算机设备可以分别获取与“region”以及“unit”对应的层级关键词列表。具体可以在“region”关键词列表中获取到区域层级“region”的层级字段值为“KLN”,并且对应的层级关键词为“KLN”,对应的层级位置信息是“位置同时在前后”。计算机设备从“unit”关键词列表中获取到的区域层级“unit”的层级字段值为“FLT 11”。其中,层级关键词为“FLT”,对应的层级位置信息是“位置在前”,其他同理。

20 在其中一个实施例中,层级关键词列表中表示区域层级“unit”的在前关键词有:“FLAT”、“FLT”、“RM”、“ROOM”、“SHOP”、“UNIT”等,其他关键词无。表示区域层级“street”的在后关键词有:“AVENUE”、“LANE”、“RD”、“ROAD”、“ST”、“STREET”、“TERRACE”等,其他关键词无。表示区域层级“floor”的在前关键词有:“FLOOR”,在后关键词有:“/F”,其他关键词无。表示区域层级“region”的同时在前后的关键词有“KOWLOON”、“KLN”、“KOWLOON”、“NT”等,其他关键词无。需要说明的是,计算机设备可以基于每一个区域层级分别构建与每一个区域层级分别对应的区域层级关键词列表,这样,得到的区域层级关键词列表的数量与区域层级的数量是一致的。计算机设备还可以将所有的区域层级构建在同一个区域层级关键词列表中,在此不作限制。

25 步骤 204,基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段。

30 其中,切分字段是待处理地址中的需要进行标准化处理的字段,即是需要获取区域层级信息的字段。

具体地,计算机设备可以将待处理地址中的地址字段与层级关键词列表中的层级字段值进行匹配,根据匹配结果对待处理地址进行切分处理,得到切分字段。例如,可以将匹配成功的地址字段从待处理地址中切分出来,并作为切分字段。

35 在本实施例中,切分字段的数量为一个或者多个,在此不做限制。一般来说,每一个切分字段都可以对应一个或者多个区域层级。需要说明的是,当待处理地址对应多个层级

关键词列表时，还可以包括根据多个层级关键词列表对待处理地址分别进行切分处理后，得到的多于一个的切分字段。

步骤 206，从层级关键词列表中查找与切分字段对应的区域层级，并将未能查找到区域层级的切分字段作为待处理切分字段。

5 在本实施例中，层级关键词列表中可以包括区域层级与层级字段值的对应关系。由于切分子段是与层级字段值相对应的，故而也能在层级关键词列表中查找到与切分字段对应的区域层级。

在其中一个实施例中，计算机设备得到的每一个切分字段都有对应的层级字段值，此时可以基于层级关键词列表查找到与每一个切分字段对应的区域层级。

10 在其中一个实施例中，计算机设备得到的切分字段中存在未与层级字段值对应的字段，此时存在无法从层级关键词列表中查找到与该切分字段对应的区域层级。

步骤 208，获取与待处理地址对应的地址配置库，从地址配置库中获取与待处理切分字段对应的区域层级。

15 在本实施例中，地址配置库中可以包括待处理地址对应的地址标识，进而可以根据地址标识从地址配置库中获取与地址标识对应的配置信息。

具体地，地址配置库是预先构建的，可以构建对应不同省份的地址配置库，并将对应的省份与地址配置库进行关联，以在具体实施中可以快速地获取与对应省份关联的地址配置库，并根据对应的地址配置库获取对应省份的地址信息。

20 可以理解的是，在其他实施例中，还可以将多个省份的地址信息构建到同一个地址配置库中，可以在同一个地址配置库中查找与对应省份的地址配置信息。

需要说明的是，由于在国内邮寄快递的地址中，是以省份为最高级别的行政单位，故而，可以以省份为最小单元构建每一个省份对应的地址配置库。

25 进一步，不论是网上购物、寄快递等，都需要填写地址。在一些情况下，地址由用户自己填写，免不了会有缺失和错误，即便完全填写正确，如何明确知道地址的各个字段也非易事。而除了中文地址，还有英文地址等其他语言的地址。对于“xxx 省 xx 市”这样的中文地址，关键词都在后面，且有比较明确的行政区划，地址的标准化相对简单一些。而对于类似“FLT 11 22/F BLK 33 AAA HSE BBB EST CCC STREET SHAM SHUI PO KLN”这样的英文地址（香港），则没有特别好的方法。

30 在其中一个实施例中，还可以包括构建香港行政区对应的地址配置库，具体以香港英文地址为例进行说明。

35 具体地，首先确定地址配置库对应的一个或者多个字段名，其中字段名可包括字段标识（ID）、地址标识、区域层级、层级深度以及不同地区之间的层级关系等。具体地，香港的地址标识可同时或者分别为地址英文名“Hong Kong”，以及地址中文名“香港”。由于香港在地理行政划分中是最高级的行政区域，故而，将香港的上一个区域层级的级别记作“”，即记作为空，香港的区域层级可以记作“region”，香港的层级深度记作“1”，并

且，由于香港为最高层级的行政区域，故而可以将香港在地址配置库中的字段标识（ID）自动记作为 1。需要说明的是，在其他实施例中，还可以通过其他形式来标记香港行政区域对应的字段值。

在本实施例中，区域层级代表了行政区域的级别，对应的区域层级的等级越高（如 1 的等级大于 2），说明对应的行政区域对应的地理范围更大，以及对应的地址范围也越大。

在本实施例中，地址配置库中预先配置了不同地址标识对应的地址信息，地址信息中具体可以包括区域层级，从而可以根据地址标识从地址配置库中获取与其对应的区域层级。需要说明的是，待处理地址中可包括多个的行政区域，如可包括层级深度为 1 的行政区域，还可以包括层级深度为 2 的行政区域等，并且不同层级深度对应的区域层级是不同的。例如，以香港为例，香港对应的层级深度为 1，对应的区域层级为“region”，香港岛对应的层级深度为 2，对应的区域层级为“district”，以及九龙对应的层级深度为 2，对应的区域层级为“district”。

在本实施例中，计算机设备可以从地址配置库中查找切分字段对应的区域层级。由于地址配置库中存在预先配置的区域层级信息，故而可以根据地址配置库获取切分字段对应的区域层级。通过将地址配置库以及层级关键词列表进行结合，以确定切分字段对应的区域层级，使得区域层级的获取更加精准以及快速。

步骤 208，根据待处理地址对应的区域层级得到标准化的地址。

具体地，计算机设备根据预先构建的地址配置库以及层级关键词列表，从地址配置库中查找每一个切分字段对应的区域层级，进而实现对待处理地址中的待标准化的待处理字段的标准化处理，得到待处理字段对应的区域层级，以实现获取待处理地址中的字段对应的区域层级。

上述实施例中，通过预先配置地址配置库以及层级关键词列表，然后在具体实施例中就可以直接根据预配置库获取对应的信息，进而根据获取到的信息实现对待处理地址的切分处理，得到标准化地址，以提高对地址的标准化处理的效率。

在其中一个实施例中，获取待处理地址对应的层级关键词列表，可以包括：获取历史地址数据，对历史地址数据中的区域层级进行标注得到层级标注值；从历史地址数据中获取与层级标注值对应的层级字段值、层级字段值对应的层级关键词以及层级关键词的层级位置信息；根据每一个区域层级对应的层级标注值、层级字段值、层级关键词以及层级关键词的层级位置信息，建立每一个区域层级对应的层级关键词列表，所述层级关键词列表存储于区块链中。

在本实施例中，考虑到地址信息的不规范，在不同的地址中，对同一个区域层级的表达方式存在区别，本申请方案还可以包括通过在计算机设备中获取大量的历史地址数据，通过对历史地址数据进行分析，得到不同区域层级对应的层级字段值、层级关键词以及层级关键词对应的层级位置信息，并根据层级字段值、层级关键词以及层级位置信息中的至少一个构建层级关键词列表。

在其中一个实施例中，计算机设备随机选择大约 200 份地址数据，并标注出地址里面“region”、“district”、“street”、“estate”、“building”、“block”、“floor”、“unit”八个区域层级的层级字段值、层级字段值对应的层级关键词以及层级关键词对应的层级位置信息（位置在前、位置在后、位置同时在前或无）。在本实施例中，计算机设备可以根据每一个区域层级建立对应的层级关键词列表，也可以将多个区域层级的层级关键词列表形成一个列列表文档，在此不作限制。

需要强调的是，为进一步保证上述层级关键词列表的私密性和安全性，上述层级关键词列表还可以存储于一区块链的节点中。

上述实施例中，通过获取大量的地址数据构建层级关键词列表，使得层级关键词列表中包含的数据信息更加全面，进一步提高了后续根据层级关键词列表确定待处理地址对应的区域层级的准确性以及效率性。

在其中一个实施例中，如图 3 所示，提供了一种基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段的流程示意图。具体地，基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段，可以包括：

步骤 302，将待处理地址中的地址字段与层级关键词列表中的层级字段值进行匹配，将匹配成功的层级字段值提取为切分字段值，将匹配成功的地址字段提取为待切分字段。

具体地，层级关键词列表中包括对应区域层级的层级标注值、层级字段值、层级关键词以及层级关键词的层级位置信息。待处理地址中可以包括一个或者多个的地址字段。

在本实施例中，计算机设备可以将待处理地址中的地址字段与各层级关键词列表中的层级字段值分别进行匹配，并将匹配成功的层级字段值提取为切分字段值，将匹配成功的地址字段提取为待切分字段。

步骤 304，从层级关键词列表中获取与切分字段值对应的层级关键词，以及层级关键词对应的层级位置信息。

可以理解的是，待切分字段是指在层级关键词列表中具有相匹配的配置信息的字段，故而计算机设备可以根据待切分字段从对应的层级关键词列表中获取对应的配置信息。

在本实施例中，配置信息可以包括但不限于与切分字段值对应的层级标注值、层级字段值以及层级关键词的层级位置信息等。

步骤 306，获取层级位置信息对应的层级位置替换符，根据层级位置替换符对待切分字段进行替换处理得到待替换地址。

具体地，计算机设备可以预先为不同的层级位置信息配置对应的层级位置替换符，如为将层级位置对应为在前关键词的位置替换符设置为“| 在前关键词”，为在后关键词设置的位置替换符为“在后关键词|”等。当计算机设备根据层级关键词列表确定对应的信息为在前关键词时，执行将“在前关键词”替换为“| 在前关键词”，当为在后关键词时，执行将“在后关键词”替换为“在后关键词|”，当为同时在前或后关键词时，执行将“同时在前或后关键词”替换为“| 同时在前或后关键词|”的步骤，进而得到待替换地址。

进一步地，在待处理地址中替换关键词并得到待替换地址之后，计算机设备还可以将连续的多个的“|”替换为单个的“|”，以及将首尾的“|”进行去除处理，并根据位置替换符“|”将待替换地址进行切分处理，即可得到待处理地址对应的切分字段。

步骤 308，根据待替换地址中的位置替换符对待替换地址进行切分处理，得到多个切分字段。

例如，待处理地址为“FLT 11 22/F BLK 33 AAA HSE BBB EST CCC STREET SHAM SHUI PO KLN”，根据位置替换符“|”对待处理字段进行替换处理得到待替换地址为“FLT 11 22/F | BLK 33 AAA HSE | BBB EST | CCC STREET | SHAM SHUI PO | KLN”，然后以“|”字符作为切分字符，对待处理地址进行切分处理，得到包含各切分字段的地址为[“FLT 11 22/F”，“ BLK 33 \n AAA HSE”，“ BBB EST”，“ CCC STREET”，“ SHAM SHUI PO”，“ KLN”]。其中，每一个引号中的地址信息作为一个切分字段。

上述实施例中，根据层级关键词列表实现对待处理地址的切分处理，处理过程简单易行，可以提高待处理地址的切分效率。并且，层级关键词列表中的配置信息是可以随时维护的，也使得根据层级关键词列表进行切分处理可以更加灵活，适用于更多的场景。

在其中一个实施例中，基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段之后，还可以包括：查找切分字段中对应的层级关键词的个数；当层级关键词的个位为一个时，从层级关键词列表中查找与切分字段对应的区域层级；当层级关键词的个数为多于一个时，继续对待处理字段进行切分处理，直至得到包括一个层级关键词的切分字段。

具体地，对于切分字段中出现两个层级关键词及以上的切分字段，可以做进一步的处理。

在其中一个实施例中，区域层级“unit”和“block”对应的层级字段值中若包括两个切分字段，这两个切分字段之间会通过“，”、“\n”等标点符号分开。利用这一特性，可以把待处理地址里面的“，”、“\n”等标点符号都替换为“|”，这时类似包括多个层级关键词的层级字段值“BLK 33 \n AAA HSE”也被进一步进行切分处理了，得到“BLK 33”以及“AAA HSE”，进而使得一个切分字段中只包括一个层级关键词以及一个区域层级。

在其中一个实施例中，对于仍然存在两个层级关键词及以上的切分字段时，还可以通过判断切分字段中是否包含“unit”和“block”这种只由两个单词组成的字段，并根据对应层级关键词的层级位置信息（如在前/后的情况）通过正则进行替换。例如对于切分字段“FLT 11 22/F”，可以通过正则表达式“r”%s.+?”%（在前关键词），先提取“FLT 11”并替换其为“FLT 11 |”，关键词在后的情况同理。这时就得到了只包含一个层级关键词的切分字段，如[“FLT 11”，“22/F”，“BLK 33”，“AAA HSE”，“BBB EST”，“CCC STREET”，“SHAM SHUI PO”，“KLN”]。

需要说明的是，上述实施例中以香港英文地址为例，但思路对于其他地址的标准化依然适用，在此不作限制。

上述实施例中，通过检测切分字段中包括的层级关键词个数确定对切分字段的处理方式，使得包括不同个数层级关键词的切分字段都能匹配到对应的处理方式，并根据对应的处理方式对切分字段进行处理，使得对切分字段处理的方式更加灵活、快速。

在其中一个实施例中，将未能查找到区域层级的切分字段作为待处理切分字段之后，
5 还可以包括：在待处理地址中获取与待处理切分字段相邻的相邻切分字段；获取相邻切分字段对应的相邻区域层级，以根据相邻区域层级确定待处理切分字段的区域层级；根据各切分字段对应的区域层级得到待处理地址对应的标准化地址。

在本实施例中，当在对应的层级关键词列表中匹配到切分字段对应的区域层级时，可以记录匹配到的切分字段（可以将每一个切分字段以一个地址行文本的形式进行显示）所在区域层级。例如，对于包括多个切分字段的地址[“FLT 11”，“22/F”，“BLK 33”，“AAA HSE”，“BBB EST”，“CCC STREET”，“SHAM SHUI PO”，“KLN”]，根据对应的层级关键词列表可以获取到切分字段“AAA HSE”之后的切分字段（地址行文本）对应的区域层级，即可得切分字段“KLN”对应的区域层级为“region”，切分字段“SHAM SHUI PO”对应的区域层级为“district”等。

15 但是，当在对应的层级关键词列表中无法匹配到切分字段对应的区域层级时，还可以将未能确定区域层级的切分字段作为待处理切分字段，并按照其他方式对待处理切分字段进行再次处理，以确定待处理切分字段对应的区域层级。

在其中一个实施例中，可以包括推导剩余的待处理切分字段所在的区域层级。具体的，当还存在待处理切分字段时，还可以根据位置信息确定待处理切分字段对应的区域层级，例如，待处理切分字段“A B C”，A的区域层级为“unit”，C的区域层级为“block”，那么可以认为B的区域层级为“floor”。

在其中一个实施例中，对于剩下的待处理切分字段以地址行的形式进行展示，若待处理切分字段是连续2行以上的情况，还可以通过确定剩下的地址行在总的地址行的上半部分还是下半部分，然后采取从上而下/从下而上推理的方法来确定，直到所有地址行都确定所在区域层级，实现对待处理地址的标准化处理。在其他实施例中，由于有的待处理地址不一定规范，或者行政区划可能会存在更新的情况，或者预配置的数据库存在收录信息不完全的情况，则可能存在只匹配到最后的“KLN”的情况。

上述实施例中，通过预先配置地址配置库以及层级关键词列表，然后在具体实施例中就可以直接根据预配置库获取对应的信息，进而根据获取到的信息实现对待处理地址的切分处理，进而得到标准化地址，提高了对地址的标准化处理效率。

35 在其中一个实施例中，获取与待处理地址对应的地址配置库，可以包括：获取地址标识对应的网页数据，地址标识是根据最大级别的层级深度对应的行政区域确定；从网页数据中查找与地址标识对应的子行政区域，以及地址标识对应的区域层级、各子行政区域对应的子区域层级；根据地址标识、子行政区域、区域层级以及子区域层级构建地址标识对应的地址配置库。

其中，子行政区域是地址标识对应的行政区域的下级区域。如当地址标识对应的行政区域为香港时，那么子行政区域为香港岛、九龙以及新界。

具体地，通过访问网页上的“香港”的公开介绍，可以得知香港包括三个行政区，“香港岛”、“九龙半岛”以及“新界”，并继续分别访问这三个行政区的介绍之后可继续得知这三个行政区的下级行政区，以及还可以继续访问下级行政区，直至达到地址数据获取需求为止，停止访问，并将访问到的数据继续存入数据库中，得到地址配置库。

在其中一个实施例中，以“香港岛”为例，字段标识（ID）在地址配置库中通过自增长的方式实现。具体地，在地址配置库中将香港岛的字段标识自增长为“2”，地址标识对应的地址中文名记作“香港岛”，地址标识对应的地址英文名记作“Hong Kong Island”，上一个层级深度记作“1”（香港的层级深度为“1”），区域层级记作“district”，层级深度记作“2”。同理，“九龙半岛”的字段标识也以ID自增长的方式，并记作“3”，地址标识中的地址中文名记作“九龙半岛”，地址标识中的地址英文名记作“Kowloon Peninsula”，以及上一个层级深度记作“1”，区域层级记作“district”，层级深度记作“2”。并且，继续通过访问下一级行政区的公开介绍，并将访问到的数据按照以上存储的方式将数据存入数据库中，即可建立行政区划数据库即包括地址信息的地址配置库。

需要说明的是，以上的数据存储方式中，其中区域层级就像“省/直辖市”、“地级市”、“县级市/区”一样，理应结合国家的行政区划来记录，这里因为香港地址没有明确的层级划分，所以具体实施中可以结合地址标准化的实际需求，人为地分为“region”、“district”、“street”、“estate”、“building”、“block”、“floor”、“unit”八个区域层级。在其他实施例中，还可以通过设置其他规则来进行区域层级的确定。

在本实施例中，地址配置库中可以包括一个或者多个行政区域的相关信息，如行政区域具体可以是某一个或者多个省份、市区以及县等。相关信息具体可以是对应的行政区域所属的区域层级、层级深度以及不同行政区域之间的层级关系。

在本实施例中，区域层级可以分为“省”“市”“县”“镇”“村”等多个层级，层级深度可以是对区域层级的量化值，具体可用于表征对应区域层级的级别大小，如可以将对应为“省”的区域层级的层级深度设置为1，将对应为“市”的区域层级的层级深度设置为2，并且设置1的级别大于2。层级关系用于表征不同行政区域之间的级别关系，如层级深度为1的行政区域为层级深度为2的行政区域的上一个级别。

在其中一个实施例中，当地址配置库对应为山东省的配置库时，可知山东省为行政区类别“省”、山东省包括多个市区，并且每一个市区对应的行政区类别为“地级市”，以及地级市中还包括多个县，并且县对应的行政区类别为“县”。具体的，根据行政区划分规则可知，“省”的级别大于“地级市”，地级市的级别大于县，故而可以设定“省”的层级深度为1，地级市的层级深度为2，以及县的层级深度为3。以及，可知山东省的地级市包括济南市、青岛市、菏泽市等，菏泽市的下辖县包括曹县、巨野以及定陶等，故而，还可以设定山东省的区域层级设置为“省”、济南市、青岛市、菏泽市等区域层级设置为“市”、

曹县、巨野以及定陶等区域层级设置为“县”。进一步，还可以设置不同区域层级之间的层级关系，如设置区域层级为“省”的地地区的上一个区域层级为“”（为空），设置区域层级值为“市”的地地区的上一个区域层级为 1（为省的区域层级），设置区域层级为“县”的地地区的上一个区域层级为 2（为地级市的区域层级）。

5 在其中一个实施例中，在构建香港行政区对应的地址配置库时，首先以最高行政级别的香港为例，在数据库中首先存入第一条数据，各字段如下：字段标识（ID）通过自增长的方式确定，首先确定为“1”，并且地址标识中的中文名记作“香港”，地址标识中的地址英文名记作“Hong Kong”，上一个层级深度记作“”，区域层级记作“region”，以及层级深度记作“1”。然后继续获取香港对应的地级区域数据，并根据不同地级区域数据之间以及地级区域数据与香港之间的关系，继续配置数据库中的数据，以得到最终的地址配置库。通过预先配置规则的方式，然后在具体实施中可以按照预设的规则自动生成地址配置库，提高了地址配置库的生成效率，以及由于网站上的数据都是实时更新的，故而通过自动从网站中爬取数据的方式也保证了地址配置库中的数据是最新的数据，以及提高了地址配置库的准确性。

15 在其中一个实施例中，上述方法还可以包括：获取地址配置库，从地址配置库中提取个性化配置信息；按照预设规则对个性化配置信息进行通用化处理，以将地址配置库中的个性化配置信息转换为通用配置信息。

具体地，还可以优化行政区划数据库。结合实际情况，可以通过人工对部分数据进行修改。例如“九龙半岛”，在实际应用中更多地使用“九龙”，英文一般也写成“Kowloon”而非“Kowloon Peninsula”。

20 在其中一个实施例中，上述方法还可以包括：获取区域更新信息，根据区域更新对地址配置库进行更新处理。具体可以实现在地址配置库中补充未赋值的区域层级，或者对地址配置库中的地址进行纠错处理或者补全处理等。例如，在真实的地址书写中，可能存在漏写或者错写“district”等区域层级的情况，导致即使是标准化后的地址中也会存在区域层级信息的缺失。

在其中一个实施例中，预先在地址配置库中存储了地址的配置信息，故而可以根据地址配置库获取待处理地址中未填写的地址信息。如可以通过查询地址配置库，以补充/更正这些区域层级的层级字段值。

30 具体地，对于任一个行政区域，可以通过在地址配置库中不断查询该行政区域的上一级的 ID，进而可以获取当前行政区域所属的所有上级的行政区域。并且，地址配置库的数据是分别独立的，使得在具体的查询过程中就像一个行政区划树一样，查询效率很高。如只要“building”在待处理地址中出现，且在地址配置库中可以查询到，那么从“building”往上的每一级行政区划（到“region”）都是确定的，那么就可以进行补全和纠错。而对于同名的情况，也只需要多判断一个原始地址的其他字段是否存在于当前行政区划即可。

35 在其中一个实施例中，可以首先设置地址配置库，地址配置库中预先存储了对应行政

区域以及子行政区域分别对应的区域层级、层级深度以及层级关系，进而可以通过地址配置库直接获取对应的区域层级等数据，提高数据获取的效率。

在本实施例中，通过预先配置层级关键词列表，使得可以直接根据预配置的层级关键词列表获取对应的配置信息，从而可以根据获取到的配置信息实现对待处理地址的切分处理得到切分字段，并得到每一个切分字段对应的区域层级，进而得到包含区域层级信息的标准化地址，提高了对地址的标准化处理效率。

上述实施例中，采用地址配置库和层级关键词列表相结合的方式，可以比较精确地实现地址标准化，同时地址配置库的构建简单、快捷。并且，极少采用正则表达式，便于代码修改和维护。进一步，对于填写不是特别完整的地址，也可以实现标准化，具有一定的鲁棒性，以及同时还支持地址补全和纠错。

具体地，本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中，人工智能（Artificial Intelligence，AI）是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、机器人技术、生物识别技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

应该理解的是，虽然图 2-3 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 2-3 中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

在其中一个实施例中，如图 4 所示，提供了一种基于人工智能的地址标准化装置，包括：

第一获取模块 402，用于获取待处理地址对应的层级关键词列表。

第一切分模块 404，用于基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段。

查找模块 406，用于从层级关键词列表中查找与切分字段对应的区域层级，并将未能查找到区域层级的切分字段作为待处理切分字段。

第二获取模块 408，用于获取与待处理地址对应的地址配置库，从地址配置库中获取与待处理切分字段对应的区域层级。

标准化模块 410，用于根据待处理地址对应的区域层级得到标准化的地址。

在其中一个实施例中，第一获取模块 402 还用于获取历史地址数据，对历史地址数据

中的区域层级进行标注得到层级标注值；从历史地址数据中获取与层级标注值对应的层级字段值、层级字段值对应的层级关键词以及层级关键词的层级位置信息；根据每一个区域层级对应的层级标注值、层级字段值、层级关键词以及层级关键词的层级位置信息，建立每一个区域层级对应的层级关键词列表。

- 5 在其中一个实施例中，第一切分模块 404 还用于将待处理地址中的地址字段与层级关键词列表中的层级字段值进行匹配，将匹配成功的层级字段值提取为切分字段值，将匹配成功的地址字段提取为待切分字段；从层级关键词列表中获取与切分字段值对应的层级关键词，以及层级关键词对应的层级位置信息；获取层级位置信息对应的层级位置替换符，根据层级位置替换符对待切分字段进行替换处理得到待替换地址；根据待替换地址中的位置替换符对待替换地址进行切分处理，得到多个切分字段。

10 在其中一个实施例中，装置还包括第二切分模块，第二切分模块用于查找切分字段中对应的层级关键词的个数；当层级关键词的个数为一个时，从层级关键词列表中查找与切分字段对应的区域层级；当层级关键词的个数为多于一个时，继续对待处理字段进行切分处理，直至得到包括一个层级关键词的切分字段。

- 15 在其中一个实施例中，装置还包括第三切分模块，第三切分模块用于在待处理地址中获取与待处理切分字段相邻的相邻切分字段；获取相邻切分字段对应的相邻区域层级，以根据相邻区域层级确定待处理切分字段的区域层级；根据各切分字段对应的区域层级得到待处理地址对应的标准化地址。

- 20 在其中一个实施例中第二获取模块 408 还用于获取地址标识对应的网页数据，地址标识是根据最大级别的层级深度对应的行政区域确定；从网页数据中查找与地址标识对应的子行政区域，以及地址标识对应的区域层级、各子行政区域对应的子区域层级；根据地址标识、子行政区域、区域层级以及子区域层级构建地址标识对应的地址配置库。

- 25 关于基于人工智能的地址标准化装置的具体限定可以参见上文中对于基于人工智能的地址标准化方法的限定，在此不再赘述。上述基于人工智能的地址标准化装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

- 30 在其中一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器或者终端，其内部结构图可以如图 5 所示。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性或易失性存储介质、内存储器。该非易失性或易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储地址标准化数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种基于人工智能的地址标准化方法。
- 35

本领域技术人员可以理解，图5中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体地计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

一种计算机设备，包括存储器和一个或者多个处理器，存储器中存储有计算机可读指令，计算机可读指令被该处理器执行时，使得一个或者多个处理器执行以下步骤：获取待处理地址对应的层级关键词列表；基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段；从层级关键词列表中查找与切分字段对应的区域层级，并将未能查找到区域层级的切分字段作为待处理切分字段；获取与待处理地址对应的地址配置库，从地址配置库中获取与待处理切分字段对应的区域层级；及根据待处理地址对应的区域层级得到标准化的地址。

在其中一个实施例中，该处理器执行计算机可读指令时还用于实现以下步骤：获取历史地址数据，对历史地址数据中的区域层级进行标注得到层级标注值；从历史地址数据中获取与层级标注值对应的层级字段值、层级字段值对应的层级关键词以及层级关键词的层级位置信息；及根据每一个区域层级对应的层级标注值、层级字段值、层级关键词以及层级关键词的层级位置信息，建立每一个区域层级对应的层级关键词列表。

在其中一个实施例中，该处理器执行计算机可读指令时还用于实现以下步骤：将待处理地址中的地址字段与层级关键词列表中的层级字段值进行匹配，将匹配成功的层级字段值提取为切分字段值，将匹配成功的地址字段提取为待切分字段；从层级关键词列表中获取与切分字段值对应的层级关键词，以及层级关键词对应的层级位置信息；获取层级位置信息对应的层级位置替换符，根据层级位置替换符对待切分字段进行替换处理得到待替换地址；及根据待替换地址中的位置替换符对待替换地址进行切分处理，得到多个切分字段。

在其中一个实施例中，该处理器执行计算机可读指令时还用于实现以下步骤：查找切分字段中对应的层级关键词的个数；当层级关键词的个位为一个时，从层级关键词列表中查找与切分字段对应的区域层级；及当层级关键词的个数为多于一个时，继续对待处理字段进行切分处理，直至得到包括一个层级关键词的切分字段。

在其中一个实施例中，该处理器执行计算机程序时还用于实现以下步骤：在待处理地址中获取与待处理切分字段相邻的相邻切分字段；获取相邻切分字段对应的相邻区域层级，以根据相邻区域层级确定待处理切分字段的区域层级；及根据各切分字段对应的区域层级得到待处理地址对应的标准化地址。

在其中一个实施例中，该处理器执行计算机可读指令时还用于实现以下步骤：获取地址标识对应的网页数据，地址标识是根据最大级别的层级深度对应的行政区域确定；从网页数据中查找与地址标识对应的子行政区域，以及地址标识对应的区域层级、各子行政区域对应的子区域层级；及根据地址标识、子行政区域、区域层级以及子区域层级构建地址标识对应的地址配置库。

在其中一个实施例中，该处理器执行计算机可读指令时还用于实现以下步骤：获取地

址配置库,从地址配置库中提取个性化配置信息;及按照预设规则对个性化配置信息进行通用化处理,以将地址配置库中的个性化配置信息转换为通用配置信息。

一个或多个存储有计算机可读指令的计算机可读存储介质,计算机可读指令被一个或多个处理器执行时,使得一个或多个处理器执行以下步骤:获取待处理地址对应的层级关键词列表;基于层级关键词列表中的层级字段值对待处理地址进行切分处理得到切分字段;从层级关键词列表中查找与切分字段对应的区域层级,并将未能查找到区域层级的切分字段作为待处理切分字段;获取与待处理地址对应的地址配置库,从地址配置库中获取与待处理切分字段对应的区域层级;及根据待处理地址对应的区域层级得到标准化的地址。

在其中一个实施例中,计算机可读指令被处理器执行时还用于实现以下步骤:获取历史地址数据,对历史地址数据中的区域层级进行标注得到层级标注值;从历史地址数据中获取与层级标注值对应的层级字段值、层级字段值对应的层级关键词以及层级关键词的层级位置信息;及根据每一个区域层级对应的层级标注值、层级字段值、层级关键词以及层级关键词的层级位置信息,建立每一个区域层级对应的层级关键词列表。

在其中一个实施例中,计算机可读指令被处理器执行时还用于实现以下步骤:将待处理地址中的地址字段与层级关键词列表中的层级字段值进行匹配,将匹配成功的层级字段值提取为切分字段值,将匹配成功的地址字段提取为待切分字段;从层级关键词列表中获取与切分字段值对应的层级关键词,以及层级关键词对应的层级位置信息;获取层级位置信息对应的层级位置替换符,根据层级位置替换符对待切分字段进行替换处理得到待替换地址;及根据待替换地址中的位置替换符对待替换地址进行切分处理,得到多个切分字段。

在其中一个实施例中,计算机可读指令被处理器执行时还用于实现以下步骤:查找切分字段中对应的层级关键词的个数;当层级关键词的个位为一个时,从层级关键词列表中查找与切分字段对应的区域层级;及当层级关键词的个数为多于一个时,继续对待处理字段进行切分处理,直至得到包括一个层级关键词的切分字段。

在其中一个实施例中,计算机可读指令被处理器执行时还用于实现以下步骤:在待处理地址中获取与待处理切分字段相邻的相邻切分字段;获取相邻切分字段对应的相邻区域层级,以根据相邻区域层级确定待处理切分字段的区域层级;及根据各切分字段对应的区域层级得到待处理地址对应的标准化地址。

在其中一个实施例中,计算机可读指令被处理器执行时还用于实现以下步骤:获取地址标识对应的网页数据,地址标识是根据最大级别的层级深度对应的行政区域确定;从网页数据中查找与地址标识对应的子行政区域,以及地址标识对应的区域层级、各子行政区域对应的子区域层级;及根据地址标识、子行政区域、区域层级以及子区域层级构建地址标识对应的地址配置库。

在其中一个实施例中,计算机可读指令被处理器执行时还用于实现以下步骤:获取地址配置库,从地址配置库中提取个性化配置信息;及按照预设规则对个性化配置信息进行

通用化处理，以将地址配置库中的个性化配置信息转换为通用配置信息。

本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain)，本质上是一个去中心化的数据库，是一串使用密码学方法相关联产生的数据块，每一个数据块中包含了一批次网络交易的信息，用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory, ROM)、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器(Random Access Memory, RAM)或外部高速缓冲存储器。作为说明而非局限，RAM可以是多种形式，比如静态随机存取存储器(Static Random Access Memory, SRAM)或动态随机存取存储器(Dynamic Random Access Memory, DRAM)等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种基于人工智能的地址标准化方法，包括：

获取待处理地址对应的层级关键词列表；

基于所述层级关键词列表中的层级字段值对所述待处理地址进行切分处理得到切分字段；

5 从所述层级关键词列表中查找与所述切分字段对应的区域层级，并将未能查找到区域层级的所述切分字段作为待处理切分字段；

获取与所述待处理地址对应的地址配置库，从所述地址配置库中获取与所述待处理切分字段对应的区域层级；及

根据所述待处理地址对应的区域层级得到标准化的地址。

10 2、根据权利要求1所述的方法，其中，所述获取待处理地址对应的层级关键词列表，包括：

获取历史地址数据，对所述历史地址数据中的区域层级进行标注得到层级标注值；

从所述历史地址数据中获取与所述层级标注值对应的层级字段值、所述层级字段值对应的层级关键词以及所述层级关键词的层级位置信息；及

15 根据每一个所述区域层级对应的所述层级标注值、所述层级字段值、所述层级关键词以及所述层级关键词的层级位置信息，建立每一个区域层级对应的层级关键词列表，所述层级关键词列表存储于区块链中。

3、根据权利要求2所述的方法，其中，所述基于所述层级关键词列表中的层级字段值对所述待处理地址进行切分处理得到切分字段，包括：

20 将所述待处理地址中的地址字段与所述层级关键词列表中的所述层级字段值进行匹配，将匹配成功的所述层级字段值提取为切分字段值，将匹配成功的所述地址字段提取为待切分字段；

从所述层级关键词列表中获取与所述切分字段值对应的层级关键词，以及所述层级关键词对应的层级位置信息；

25 获取所述层级位置信息对应的层级位置替换符，根据所述层级位置替换符对所述待切分字段进行替换处理得到待替换地址；及

根据所述待替换地址中的所述位置替换符对所述待替换地址进行切分处理，得到多个切分字段。

30 4、根据权利要求3所述的方法，其中，所述基于所述层级关键词列表中的层级字段值对所述待处理地址进行切分处理得到切分字段之后，所述方法还包括：

查找所述切分字段中对应的层级关键词的个数；

当所述层级关键词的个位为一个时，从所述层级关键词列表中查找与所述切分字段对应的区域层级；及

当所述层级关键词的个数为多于一个时，继续对所述待处理地址进行切分处理，直至

得到一个包括一个层级关键词的切分字段。

5、根据权利要求 1 所述的方法，其中，所述并将未能查找到区域层级的所述切分字段作为待处理切分字段之后，所述方法还包括：

在所述待处理地址中获取与所述待处理切分字段相邻的相邻切分字段；

5 获取所述相邻切分字段对应的相邻区域层级，以根据所述相邻区域层级确定所述待处理切分字段的区域层级；及

根据各所述切分字段对应的区域层级得到所述待处理地址对应的标准化地址。

6、根据权利要求 1 所述的方法，其中，所述获取与所述待处理地址对应的地址配置库，包括：

10 获取地址标识对应的网页数据，所述地址标识是根据最大级别的层级深度对应的行政区域确定；

从所述网页数据中查找与所述地址标识对应的子行政区域，以及所述地址标识对应的区域层级、各所述子行政区域对应的子区域层级；及

15 根据所述地址标识、所述子行政区域、所述区域层级以及所述子区域层级构建所述地址标识对应的地址配置库。

7、根据权利要求 1 所述的方法，其中，还包括：

获取地址配置库，从所述地址配置库中提取个性化配置信息；及

按照预设规则对所述个性化配置信息进行通用化处理，以将所述地址配置库中的个性化配置信息转换为通用配置信息。

20 8、一种基于人工智能的地址标准化装置，包括：

第一获取模块，用于获取待处理地址对应的层级关键词列表；

切分模块，用于基于所述层级关键词列表中的层级字段值对所述待处理地址进行切分处理得到切分字段；

25 查找模块，用于从所述层级关键词列表中查找与所述切分字段对应的区域层级，并将未能查找到区域层级的所述切分字段作为待处理切分字段；

第二获取模块，用于获取与所述待处理地址对应的地址配置库，从所述地址配置库中获取与所述待处理切分字段对应的区域层级；及

标准化模块，用于根据所述待处理地址对应的区域层级得到标准化的地址。

30 9、一种计算机设备，包括存储器和一个或者多个处理器，所述存储器存储有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取待处理地址对应的层级关键词列表；

基于所述层级关键词列表中的层级字段值对所述待处理地址进行切分处理得到切分字段；

35 从所述层级关键词列表中查找与所述切分字段对应的区域层级，并将未能查找到区域

层级的所述切分字段作为待处理切分字段；

获取与所述待处理地址对应的地址配置库，从所述地址配置库中获取与所述待处理切分字段对应的区域层级；及

根据所述待处理地址对应的区域层级得到标准化的地址。

- 5 10、根据权利要求 9 所述的计算机设备，其中，所述处理器执行所述计算机可读指令时所实现的所述获取待处理地址对应的层级关键词列表，包括：

获取历史地址数据，对所述历史地址数据中的区域层级进行标注得到层级标注值；

从所述历史地址数据中获取与所述层级标注值对应的层级字段值、所述层级字段值对应的层级关键词以及所述层级关键词的层级位置信息；及

- 10 根据每一个所述区域层级对应的所述层级标注值、所述层级字段值、所述层级关键词以及所述层级关键词的层级位置信息，建立每一个区域层级对应的层级关键词列表，所述层级关键词列表存储于区块链中。

- 15 11、根据权利要求 10 所述的计算机设备，其中，所述处理器执行所述计算机可读指令时所实现的所述基于所述层级关键词列表中的层级字段值对所述待处理地址进行切分处理得到切分字段，包括：

将所述待处理地址中的地址字段与所述层级关键词列表中的所述层级字段值进行匹配，将匹配成功的所述层级字段值提取为切分字段值，将匹配成功的所述地址字段提取为待切分字段；

- 20 从所述层级关键词列表中获取与所述切分字段值对应的层级关键词，以及所述层级关键词对应的层级位置信息；

获取所述层级位置信息对应的层级位置替换符，根据所述层级位置替换符对所述待切分字段进行替换处理得到待替换地址；及

根据所述待替换地址中的所述位置替换符对所述待替换地址进行切分处理，得到多个切分字段。

- 25 12、根据权利要求 11 所述的计算机设备，其中，所述处理器执行所述计算机可读指令时还实现以下步骤：

查找所述切分字段中对应的层级关键词的个数；

当所述层级关键词的个位为一个时，从所述层级关键词列表中查找与所述切分字段对应的区域层级；及

- 30 当所述层级关键词的个数为多于一个时，继续对所述待处理地址进行切分处理，直至得到包括一个层级关键词的切分字段。

13、根据权利要求 9 所述的计算机设备，其中，所述处理器执行所述计算机可读指令时还实现以下步骤：

在所述待处理地址中获取与所述待处理切分字段相邻的相邻切分字段；

- 35 获取所述相邻切分字段对应的相邻区域层级，以根据所述相邻区域层级确定所述待处

理切分字段的区域层级；及

根据各所述切分字段对应的区域层级得到所述待处理地址对应的标准化地址。

14、根据权利要求 9 所述的计算机设备，其中，所述处理器执行所述计算机可读指令时所实现的所述获取与所述待处理地址对应的地址配置库，包括：

5 获取地址标识对应的网页数据，所述地址标识是根据最大级别的层级深度对应的行政区域确定；

从所述网页数据中查找与所述地址标识对应的子行政区域，以及所述地址标识对应的区域层级、各所述子行政区域对应的子区域层级；及

10 根据所述地址标识、所述子行政区域、所述区域层级以及所述子区域层级构建所述地址标识对应的地址配置库。

15、一个或多个存储有计算机可读指令的计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取待处理地址对应的层级关键词列表；

15 基于所述层级关键词列表中的层级字段值对所述待处理地址进行切分处理得到切分字段；

从所述层级关键词列表中查找与所述切分字段对应的区域层级，并将未能查找到区域层级的所述切分字段作为待处理切分字段；

获取与所述待处理地址对应的地址配置库，从所述地址配置库中获取与所述待处理切分字段对应的区域层级；及

20 根据所述待处理地址对应的区域层级得到标准化的地址。

16、根据权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行时所实现的所述获取待处理地址对应的层级关键词列表，包括：

获取历史地址数据，对所述历史地址数据中的区域层级进行标注得到层级标注值；

25 从所述历史地址数据中获取与所述层级标注值对应的层级字段值、所述层级字段值对应的层级关键词以及所述层级关键词的层级位置信息；及

根据每一个所述区域层级对应的所述层级标注值、所述层级字段值、所述层级关键词以及所述层级关键词的层级位置信息，建立每一个区域层级对应的层级关键词列表，所述层级关键词列表存储于区块链中。

30 17、根据权利要求 16 所述的存储介质，其中，所述计算机可读指令被所述处理器执行时所实现的所述基于所述层级关键词列表中的层级字段值对所述待处理地址进行切分处理得到切分字段，包括：

将所述待处理地址中的地址字段与所述层级关键词列表中的所述层级字段值进行匹配，将匹配成功的所述层级字段值提取为切分字段值，将匹配成功的所述地址字段提取为待切分字段；

35 从所述层级关键词列表中获取与所述切分字段值对应的层级关键词，以及所述层级关

关键词对应的层级位置信息；

获取所述层级位置信息对应的层级位置替换符，根据所述层级位置替换符对所述待切分字段进行替换处理得到待替换地址；及

根据所述待替换地址中的所述位置替换符对所述待替换地址进行切分处理，得到多个切分字段。

18、根据权利要求 17 所述的存储介质，其中，所述计算机可读指令被所述处理器执行时还实现以下步骤：

查找所述切分字段中对应的层级关键词的个数；

当所述层级关键词的个数为一个时，从所述层级关键词列表中查找与所述切分字段对应的区域层级；及

当所述层级关键词的个数为多于一个时，继续对所述待处理地址进行切分处理，直至得到包括一个层级关键词的切分字段。

19、根据权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行时还实现以下步骤：

在所述待处理地址中获取与所述待处理切分字段相邻的相邻切分字段；

获取所述相邻切分字段对应的相邻区域层级，以根据所述相邻区域层级确定所述待处理切分字段的区域层级；及

根据各所述切分字段对应的区域层级得到所述待处理地址对应的标准化地址。

20、根据权利要求 15 所述的存储介质，其中，所述计算机可读指令被所述处理器执行时所实现的所述获取与所述待处理地址对应的地址配置库，包括：

获取地址标识对应的网页数据，所述地址标识是根据最大级别的层级深度对应的行政区域确定；

从所述网页数据中查找与所述地址标识对应的子行政区域，以及所述地址标识对应的区域层级、各所述子行政区域对应的子区域层级；及

根据所述地址标识、所述子行政区域、所述区域层级以及所述子区域层级构建所述地址标识对应的地址配置库。

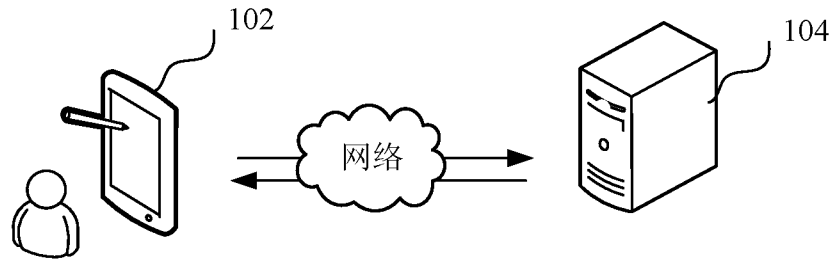


图 1



图 2

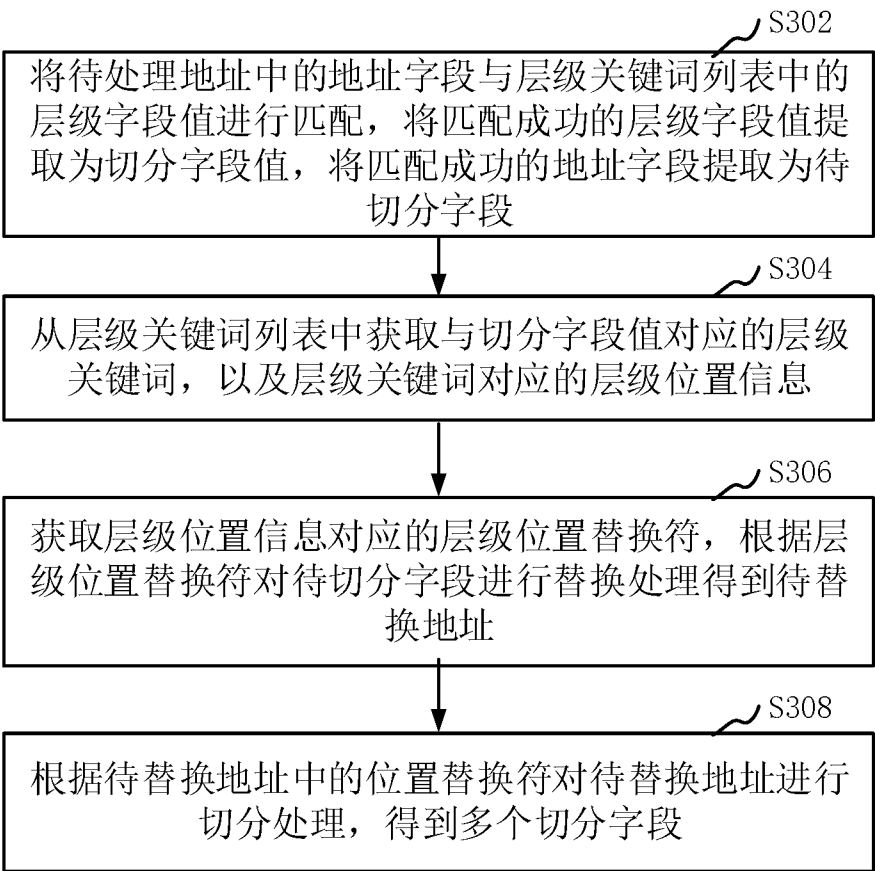


图 3

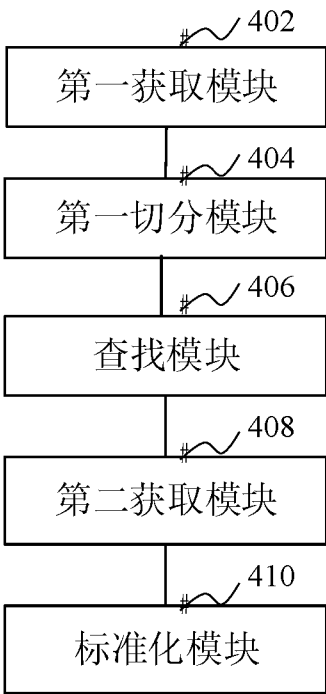


图 4

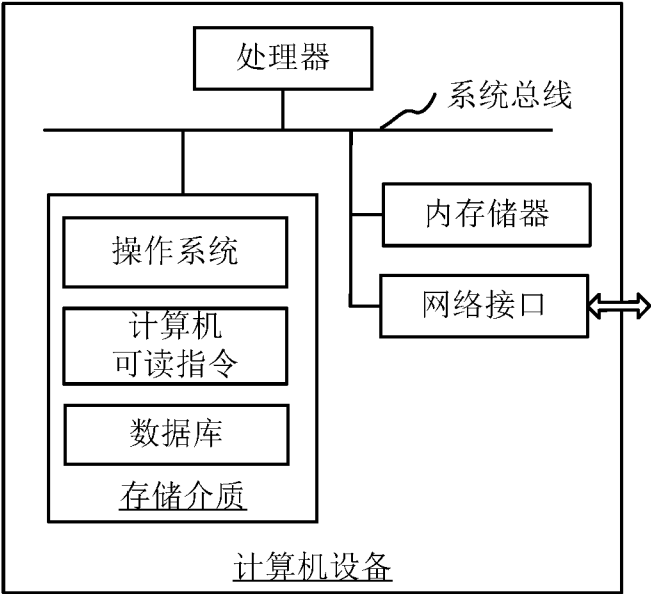


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/109461

A. CLASSIFICATION OF SUBJECT MATTER G06F 17/00(2019.01)i According to International Patent Classification (IPC) or to both national classification and IPC																							
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06F Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNPAT, WPI, EPODOC, CNKI, IEEE: 地址, 标准化, 规范化, 补全, 关键词, 字段, 层级, 切分, 匹配, 历史数据, 标注, address, standardization, normalization, completion, key word, field, hierarchy, segmentation, match, historical data, mark																							
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 112364114 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 12 February 2021 (2021-02-12) claims 1-10, description paragraphs [0004]-[0145]</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 109033086 A (CHINA UNIONPAY DATA SERVICES CO., LTD.) 18 December 2018 (2018-12-18) description paragraphs [0067]-[0122], [0154]-[0157]</td> <td>1, 2, 8-10, 15, 16</td> </tr> <tr> <td>A</td> <td>CN 111914557 A (SHANGHAI YANXI SOFTWARE INFORMATION TECHNOLOGY CO., LTD.) 10 November 2020 (2020-11-10) entire document</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 106970903 A (ALIBABA GROUP HOLDING LIMITED) 21 July 2017 (2017-07-21) entire document</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 109670122 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 23 April 2019 (2019-04-23) entire document</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2009182975 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 16 July 2009 (2009-07-16) entire document</td> <td>1-20</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	PX	CN 112364114 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 12 February 2021 (2021-02-12) claims 1-10, description paragraphs [0004]-[0145]	1-20	X	CN 109033086 A (CHINA UNIONPAY DATA SERVICES CO., LTD.) 18 December 2018 (2018-12-18) description paragraphs [0067]-[0122], [0154]-[0157]	1, 2, 8-10, 15, 16	A	CN 111914557 A (SHANGHAI YANXI SOFTWARE INFORMATION TECHNOLOGY CO., LTD.) 10 November 2020 (2020-11-10) entire document	1-20	A	CN 106970903 A (ALIBABA GROUP HOLDING LIMITED) 21 July 2017 (2017-07-21) entire document	1-20	A	CN 109670122 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 23 April 2019 (2019-04-23) entire document	1-20	A	US 2009182975 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 16 July 2009 (2009-07-16) entire document	1-20
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.																					
PX	CN 112364114 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 12 February 2021 (2021-02-12) claims 1-10, description paragraphs [0004]-[0145]	1-20																					
X	CN 109033086 A (CHINA UNIONPAY DATA SERVICES CO., LTD.) 18 December 2018 (2018-12-18) description paragraphs [0067]-[0122], [0154]-[0157]	1, 2, 8-10, 15, 16																					
A	CN 111914557 A (SHANGHAI YANXI SOFTWARE INFORMATION TECHNOLOGY CO., LTD.) 10 November 2020 (2020-11-10) entire document	1-20																					
A	CN 106970903 A (ALIBABA GROUP HOLDING LIMITED) 21 July 2017 (2017-07-21) entire document	1-20																					
A	CN 109670122 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 23 April 2019 (2019-04-23) entire document	1-20																					
A	US 2009182975 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 16 July 2009 (2009-07-16) entire document	1-20																					
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.																							
* Special categories of cited documents: “A” document defining the general state of the art which is not considered to be of particular relevance “E” earlier application or patent but published on or after the international filing date “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) “O” document referring to an oral disclosure, use, exhibition or other means “P” document published prior to the international filing date but later than the priority date claimed “T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art “&” document member of the same patent family																							
Date of the actual completion of the international search 14 October 2021		Date of mailing of the international search report 27 October 2021																					
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China Facsimile No. (86-10)62019451		Authorized officer Telephone No.																					

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/109461

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112364114	A	12 February 2021	None			
CN	109033086	A	18 December 2018	None			
CN	111914557	A	10 November 2020	None			
CN	106970903	A	21 July 2017	None			
CN	109670122	A	23 April 2019	None			
US	2009182975	A1	16 July 2009	WO	2009087226	A1	16 July 2009
				US	2014059321	A1	27 February 2014
				US	2012011341	A1	12 January 2012

国际检索报告

国际申请号

PCT/CN2021/109461

A. 主题的分类

G06F 17/00 (2019.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNPAT, WPI, EPDOC, CNKI, IEEE: 地址, 标准化, 规范化, 补全, 关键词, 字段, 层级, 切分, 匹配, 历史数据, 标注, address, standardization, normalization, completion, key word, field, hierarchy, segmentation, match, historical data, mark

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 112364114 A (深圳壹账通智能科技有限公司) 2021年 2月 12日 (2021 - 02 - 12) 权利要求1-10, 说明书第[0004]-[0145]段	1-20
X	CN 109033086 A (银联数据服务有限公司) 2018年 12月 18日 (2018 - 12 - 18) 说明书第[0067]-[0122]、[0154]-[0157]段	1、2、8-10、15、16
A	CN 111914557 A (上海燕汐软件信息科技有限公司) 2020年 11月 10日 (2020 - 11 - 10) 全文	1-20
A	CN 106970903 A (阿里巴巴集团控股有限公司) 2017年 7月 21日 (2017 - 07 - 21) 全文	1-20
A	CN 109670122 A (平安科技深圳有限公司) 2019年 4月 23日 (2019 - 04 - 23) 全文	1-20
A	US 2009182975 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2009年 7月 16日 (2009 - 07 - 16) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2021年 10月 14日

国际检索报告邮寄日期

2021年 10月 27日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国 北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

董显彬

电话号码 86-(10)-53961407

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/109461

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	112364114	A	2021年 2月 12日	无			
CN	109033086	A	2018年 12月 18日	无			
CN	111914557	A	2020年 11月 10日	无			
CN	106970903	A	2017年 7月 21日	无			
CN	109670122	A	2019年 4月 23日	无			
US	2009182975	A1	2009年 7月 16日	WO	2009087226	A1	2009年 7月 16日
				US	2014059321	A1	2014年 2月 27日
				US	2012011341	A1	2012年 1月 12日