

(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
2016 年 10 月 6 日 (06.10.2016)



(10) 国际公布号
WO 2016/155386 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/099580
- (22) 国际申请日: 2015 年 12 月 29 日 (29.12.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510148638.4 2015 年 3 月 31 日 (31.03.2015) CN
- (71) 申请人: 北京奇虎科技有限公司 (BEIJING QIHOO TECHNOLOGY COMPANY LIMITED) [CN/CN]; 中国北京市西城区新街口外大街 28 号 D 座 112 室 (德胜园区), Beijing 100088 (CN)。奇智软件 (北京) 有限公司 (QIZHI SOFTWARE (BEIJING) COMPANY LIMITED) [CN/CN]; 中国北京市朝阳区酒仙桥路 6 号院 2 号楼 B 座 2 层、3 层 301-306 室, Beijing 100015 (CN)。
- (72) 发明人: 王智广 (WANG, Zhiguang); 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。魏少俊 (WEI, Shaojun); 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。

- (74) 代理人: 北京市隆安律师事务所 (BEIJING LONGAN LAW FIRM); 中国北京市朝阳区建国门外大街 21 号北京国际俱乐部 188 室, Beijing 100020 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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)。

[见续页]

(54) Title: METHOD AND DEVICE FOR DETERMINING WHETHER WEBPAGE COMPRISES POINT OF INTEREST (POI) DATA

(54) 发明名称: 用于确定网页页面中包括兴趣点 POI 数据的方法及装置

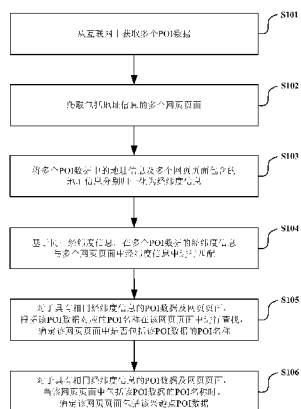


图 1a

S101 Acquire multiple pieces of POI data from the Internet
S102 Crawl multiple webpages comprising address information
S103 Separately normalize address information in the multiple pieces of POI data and the address information comprised in the multiple webpages to be longitude and latitude information
S104 Match longitude and latitude information of the multiple pieces of POI data with that of the multiple webpages according to same longitude and latitude information
S105 For the POI data and the webpages having same longitude and latitude information, search in the webpages according to POI names corresponding to the POI data, so as to determine whether the webpages comprise the POI names of the POI data
S106 For the POI data and the webpages having same longitude and latitude information, when the webpages comprise the POI names of the POI data, determine that the webpages comprise the point of interest (POI) data

(57) Abstract: A method and device for determining whether a webpage comprises point of interest (POI) data. The method comprises: acquiring multiple pieces of POI data from the Internet (S101); crawling multiple webpages comprising address information (S102); separately normalizing address information in the multiple pieces of POI data and the address information comprised in the multiple webpages to be longitude and latitude information (S103); matching longitude and latitude information of the multiple pieces of POI data with that of the multiple webpages (S104); for the POI data and the webpages having same longitude and latitude information, searching in the webpages according to POI names corresponding to the POI data, so as to determine whether the webpages comprise the POI names of the POI data (S105); and when the webpages comprise the POI names of the POI data, determining that the webpages comprise the point of interest (POI) data (S106). The method and the device help to subsequently determine the accuracy of collected POI data according to the accuracy of content recorded by the webpages, thereby subsequently providing a reliable guarantee for collecting accurate POI data in the Internet in a large scale.

(57) 摘要: 用于确定网页页面中包括兴趣点 POI 数据的方法及装置, 包括: 从互联网中获取多个 POI 数据(S101); 爬取包括地址信息的多个网页页面(S102); 将多个 POI 数据中的地址信息及多个网页页面包含的地址信息分别归一化为经纬度信息(S103); 在多个 POI 数据的经纬度信息与多个网页页面中经纬度信息中进行匹配(S104); 对于具有相同经纬度信息的 POI 数据及网页页面, 根据该 POI 数据对应的 POI 名称在该网页页面中进行查找, 确定该网页页面中是否包括该 POI 数据的 POI 名称(S105); 当该网页页面中包括该 POI 数据的 POI 名称时, 确定该网页页面包括该兴趣点 POI 数据(S106)。该方法及装置有利于后续根据网页页面所记载内容的准确性, 来确定收集到的 POI 数据的准确性, 为后续为大批量地收集互联网中准确 POI 数据提供可靠保证。

WO 2016/155386 A1

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

用于确定网页页面中包括兴趣点 POI 数据的方法及装置

技术领域

本发明涉及计算机技术领域，具体而言，本发明涉及一种用于确定网页页面中包括兴趣点 POI 数据的方法及装置。

5

背景技术

在地理信息系统中，一个 POI（Point Of Interest，兴趣点）可以是一栋房子、一个商铺、一个邮筒、一个公交站等。POI 数据包括地址信息和 POI 名称。

10 传统的 POI 数据采集方法，需要技术人员采用精密的测绘仪器去获取每个 POI 的经纬度信息，然后再标记下来，这种方法比较费时费力，导致通过采集得到的 POI 数据的数量很少，地理信息系统很难根据数量很少的 POI 数据来提供高水平的服务。

互联网上存在着大量的 POI 数据，如果能从互联网上收集包含 POI 数据的网页，从收集的网页中提取出这些 POI 数据供地理信息系统使用，则会大大节省人力和时间。但是互联网上充斥着大量虚假的 POI 数据，比如博客网页内容中包含“原文地址：http://xxx.xxx.xxx/xxx”，虽然包含“地址”字样，但该地址是网络地址或者说是 URL（Uniform Resource Locator，统一资源定位器），并不是 POI 数据中的地理地址信息；从而导致收集到的 POI 数据中
20 虚假的 POI 数据的比例较高。

发明内容

本发明针对现有技术的缺点，提出一种用于确定网页页面中包括兴趣点 POI 数据的方法和装置，用以解决现有技术存在的收集较多虚假的 POI 数据
25 问题。

本发明根据一个方面，提供了一种用于确定网页页面中包括兴趣点 POI 数据的方法，包括：

从互联网中获取多个 POI 数据；

爬取包括地址信息的多个网页页面；

将所述多个 POI 数据中的地址信息及所述多个网页页面包含的地址信息分别归一化为经纬度信息;

基于同一经纬度信息,在所述多个 POI 数据的经纬度信息与多个网页页面中经纬度信息中进行匹配;

- 5 对于具有相同经纬度信息的 POI 数据及网页页面,根据该 POI 数据对应的 POI 名称在该网页页面中进行查找,确定该网页页面中是否包括该 POI 数据的 POI 名称;

当该网页页面中包括该 POI 数据的 POI 名称时,确定该网页页面包括该兴趣点 POI 数据。

- 10 本发明根据另一个方面,还提供了一种用于确定网页页面中包括兴趣点 POI 数据的装置,包括:

POI 数据获取模块,用于从互联网中获取多个 POI 数据;

网页页面爬取模块,用于爬取包括地址信息的多个网页页面;

- 15 经纬度信息归一化模块,用于将所述多个 POI 数据中的地址信息及所述多个网页页面包含的地址信息分别归一化为经纬度信息;

经纬度信息匹配模块,用于基于同一经纬度信息,在所述多个 POI 数据的经纬度信息与多个网页页面中经纬度信息中进行匹配;

- 20 网页页面包括 POI 名称确定模块,用于对于具有相同经纬度信息的 POI 数据及网页页面,根据该 POI 数据对应的 POI 名称在该网页页面中进行查找,确定该网页页面中是否包括该 POI 数据的 POI 名称;

网页页面包括 POI 数据确定模块,用于当该网页页面中包括该 POI 数据的 POI 名称时,确定该网页页面包括该兴趣点 POI 数据。

- 25 根据本发明的又一个方面,提出了一种计算机程序,包括计算机可读代码,当所述计算机可读代码在计算设备上运行时,导致所述计算设备执行上文所述的用于确定网页页面中包括兴趣点 POI 数据的方法。

根据本发明的再一个方面,提出了一种计算机可读介质,其中存储了上述的计算机程序。

- 30 本发明的技术方案中,将地址信息归一化为经纬度信息,可以滤除非地理的地址信息,由于经纬度的唯一性,基于经纬度信息的匹配结果的准确性,远高于现有的基于文本信息的匹配结果的准确性,从而有利于后续避免收集

到虚假地址信息的 POI 数据; 在 POI 数据的经纬度信息与网页页面中的经纬度信息相匹配的基础上, 进一步确定网页页面中是否包括 POI 数据的 POI 名称, 来准确判断 POI 数据是否被包含在同一网页页面中, 有利于后续根据网页页面所记载内容的权威性和准确性, 来确定收集到的 POI 数据的准确性, 进而为大批量地收集互联网中的准确度较高的 POI 数据提供可靠保证。

本发明附加的方面和优点将在下面的描述中部分给出, 这些将从下面的描述中变得明显, 或通过本发明的实践了解到。

附图说明

10 本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解, 其中:

图 1a 为本发明实施例的用于确定网页页面中包括兴趣点 POI 数据的方法的流程示意图;

图 1b 为本发明实施例的包括多个 POI 数据的网页的示意图;

15 图 2 为本发明实施例的用于确定网页页面中包括兴趣点 POI 数据的装置的内部结构的框架示意图;

图 3 为本发明实施例的 POI 数据获取模块的内部结构的框架示意图;

图 4 示意性地示出了用于执行根据本发明的方法的计算设备的框图; 以及

20 图 5 示意性地示出了用于保持或者携带实现根据本发明的方法的程序代码的存储单元。

具体实施例

下面详细描述本发明的实施例, 所述实施例的示例在附图中示出, 其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的, 仅用于解释本发明, 而不能解释为对本发明的限制。

本技术领域技术人员可以理解, 除非特意声明, 这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是, 本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步

骤、操作、元件和/或组件，但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。应该理解，当我们称元件被“连接”或“耦接”到另一元件时，它可以直接连接或耦接到其他元件，或者也可以存在中间元件。此外，这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的措辞“和/或”包括一个或更多个相关联的列出项的全部或任一单元和全部组合。

本技术领域技术人员可以理解，除非另外定义，这里使用的所有术语（包括技术术语和科学术语），具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是，诸如通用字典中定义的那些术语，应该被理解为具有与现有技术的上下文中的意义一致的意义，并且除非像这里一样被特定定义，否则不会用理想化或过于正式的含义来解释。

图 1a 为本发明中用于确定网页页面中包括兴趣点 POI 数据的方法的流程示意图。

S101: 从互联网中获取多个 POI 数据；S102: 爬取包括地址信息的多个网页页面；S103: 将多个 POI 数据中的地址信息及多个网页页面包含的地址信息分别归一化为经纬度信息；S104: 基于同一经纬度信息，在多个 POI 数据的经纬度信息与多个网页页面中经纬度信息中进行匹配；S105: 对于具有相同经纬度信息的 POI 数据及网页页面，根据该 POI 数据对应的 POI 名称在该网页页面中进行查找，确定该网页页面中是否包括该 POI 数据的 POI 名称；S106: 当该网页页面中包括该 POI 数据的 POI 名称时，确定该网页页面包括该兴趣点 POI 数据。

本发明的上述用于确定网页页面中包括兴趣点 POI 数据的方法，将地址信息归一化为经纬度信息，可以滤除非地理位置的地址信息，由于经纬度的唯一性，基于经纬度信息的匹配结果的准确性，远高于现有的基于文本信息的匹配结果的准确性，从而有利于后续避免收集虚假地址信息的数据；在 POI 数据的经纬度信息与网页页面中的经纬度信息相匹配的基础上，进一步确定网页页面中是否包括 POI 数据的 POI 名称，来准确判断 POI 数据是否被包含在同一网页页面中，有利于后续根据网页页面所记载内容的权威性和准确性，来确定收集到的 POI 数据的准确性，进而为大批量地收集互联网中的准确度较高的 POI 数据提供可靠保证。

下面具体介绍流程示意图如图 1a 所示的用于确定网页页面中包括兴趣点 POI 数据的方法，包括如下步骤：

S101：从互联网中获取多个 POI 数据。

具体地，利用网络爬虫类的程序，从互联网中爬取多个包括 POI 数据的网页；随后从多个包括 POI 数据的网页中提取多个 POI 数据。POI 数据包括地址信息和 POI 名称；优选地，POI 数据还可以包括联系方式、邮编和网络标签等等。

本发明的发明人发现，在互联网中存在这样一些网页，它们中每个网页的内容包含有一个或者多个 POI 数据，POI 数据中的地址信息包括“地址”等字样的地址关键词；并且这些网页的页面结构特征 URL 格式，以及 POI 数据在网页中的位置和格式是有规律性的。也就是说可以通过一种统一的方法快捷地从这些网页上提取出 POI 数据。

较佳地，可以从互联网中，爬取包括“地址”等地址关键词的多个网页对应的多个 URL（Uniform Resource Locator，统一资源定位器）；对爬取得到的多个 URL 进行 pattern 聚类，将具有相同结构特征的 URL 聚类为同一 pattern 集合。

更优地，对于众多包括地址关键词的网页中，只包括一个 POI 数据的网页，获取所有只包括一个 POI 数据的网页的 URL；对获取的所有 URL 进行 pattern 聚类，将具有相同结构特征的 URL 聚类为同一 pattern 集合。

例如，众多包括地址关键词的网页中，URL 为 <http://www.aibang.com/detail/1537772035-1606201508> 的网页中只包括“爱普生（中国）有限公司”这一 POI 数据、URL 为 <http://www.aibang.com/detail/152928073-419169481> 的网页中只包括“北京王府中西医结合医院”这一 POI 数据，这两个 URL 具有相同的结构特征 www.aibang.com/detail/*，其中*为通配符表示任意字符；因此，可以将这两个 URL 聚类到同一 pattern 集合中；也就是说，该 pattern 集合中所有的 URL 都具有同一结构特征 www.aibang.com/detail/*。

更优的，对于众多的包括地址关键词的网页中，包括多个 POI 数据的网页，获取所有包括多个 POI 数据的网页的 URL；对获取的所有 URL 进行 pattern 聚类，将具有相同结构特征的 URL 聚类为同一 pattern 集合。

例如，URL 为 www.dianping.com/topic/s_c_2_120_r14_x540/p7 的网页，如图 1b 所示，该网页中包括 POI 标题为“boy london”、“COACH（悠唐折扣店）”和“米兰店（三里屯店）”等的多个 POI 数据；URL 为 www.dianping.com/topic/s_c_2_120_r14_x540/p6 的网页中也包括多个 POI 数据；获取所有结构特征符合 www.dianping.com/topic/* 的 URL，其中*为通配符表示任意字符；对获取的所有 URL 进行 pattern 聚类，聚类得到的同一 pattern 集合中的 URL 都具有结构特征 www.dianping.com/topic/*。

从多个 pattern 集合中筛选出包括多个包括 POI 数据的网页的 pattern 集合，并从该 pattern 集合中提取多个包括 POI 数据的网页。

10 较佳地，从多个包括 POI 数据的网页中提取多个 POI 数据具体可以包括：

基于属于同一 pattern 集合中多个 URL 对应多个包括 POI 数据的网页的页面结构特征，生成与该 pattern 集合相应的 POI 数据提取模板。具体地，对于属于同一 pattern 集合中每个 URL，根据该 URL 对应的各网页中 POI 数据的格式和位置，生成与该 pattern 集合相应的 POI 数据提取模板。

基于 POI 数据提取模板，从多个包括 POI 数据的网页中提取多个 POI 数据。具体地，对于同一 pattern 集合中每个 URL，针对该 URL 对应的网页，根据生成的 POI 数据提取模板中的 POI 数据的格式、以及多个 POI 数据在网页中的位置，从该网页中提取多个 POI 数据。

20 S102：爬取包括地址信息的多个网页页面。

具体地，利用网络爬虫类的程序，从互联网中爬取包括地址关键词的多个网页页面。

提取多个网页页面中与地址关键词相关联的多个文本信息。

具体地，对于一个网页页面，提取该网页页面的文本内容，在文本内容
25 中查找“地址”、“位于”或“坐落于”等可能包括地址信息的地址关键词；提取地址关键词附近的文本片段；根据设定的分隔符以及片段长度对文本片段进行分割，比如文本片段距离地址关键词的文本长度大于设定的阈值、和/或文本片段出现设定的分隔符（比如空格、逗号、句号等），则对文本片段进行分割；将分割结果中，分割处（例如分隔符处）与地址关键词之间的文
30 本片段，作为该网页页面中与地址关键词相关联的文本信息。

从多个文本信息中提取相应网页页面的地址信息。

具体地，对于提取自网页页面中的每个文本信息，从该文本信息中提取出地址信息，作为该网页页面的地址信息。

S103：将多个 POI 数据中的地址信息及多个网页页面包含的地址信息
5 分别归一化为经纬度信息。

预先获取包括全国的省、市、县（区）、乡镇、道路等的地址信息、经纬度信息，以及地址信息与经纬度信息之间的对应关系的地理信息库。其中，地理信息库中的地址信息，可以包括表示同一地理地址的多种表达形式的地址信息；例如，“朝阳区酒仙桥路 6 号”、“北京市朝阳区酒仙桥 6 号”和“朝
10 阳区酒仙桥 6 号”等多个地址信息，都表示同一地理地址。

具体地，将多个 POI 数据中的地址信息分别归一化为多个 POI 数据的经纬度信息。例如，对于每个 POI 数据中的地址信息，从预先获取的地理信息库中查找出该地址信息所对应的经纬度信息，将查找出的经纬度信息确定为该 POI 数据的经纬度信息。

15 S104：基于同一经纬度信息，在多个 POI 数据的经纬度信息与多个网页页面中经纬度信息中进行匹配。

具体地，对于每个 POI 数据，判断各网页页面中，是否存在经纬度信息与该 POI 数据的经纬度信息相一致的网页页面，若是，则确定出该 POI 数据与该网页页面相匹配，即确定出该 POI 数据与该网页页面具有相同经纬度信
20 息，否则，忽略该 POI 数据。

由于经纬度的唯一性，基于经纬度信息的匹配结果的准确性，远高于现有的基于文本信息的匹配结果的准确性，从而后续可以根据更准确的匹配结果收集到更为准确的 POI 数据。而且，基于经纬度信息进行匹配，相当于基于该经纬度信息所对应的多个地理信息分别进行匹配，扩大了匹配的范围，
25 有利于后续收集到更多的 POI 数据。

S105：对于具有相同经纬度信息的 POI 数据及网页页面，根据该 POI 数据对应的 POI 名称在该网页页面中进行查找，确定该网页页面中是否包括该 POI 数据的 POI 名称。

具体地，对于具有相同经纬度信息的 POI 数据及网页页面，从该网页页
30 面中查找出所有的名称信息；对于查找出的每个名称信息，判断该名称信息

是否与该 POI 数据中的 POI 名称相匹配：若是，确定出该网页页面中包括该 POI 数据的 POI 名称；否则，忽略该 POI 数据。

较佳地，对于具有相同经纬度信息的 POI 数据及网页页面，若该网页页面中的名称信息与该 POI 数据中的 POI 名称，虽然在文字表达上不完全一致，
5 但是实质上表示同一 POI，可以确认为该 POI 数据中的 POI 名称与该网页页面中的名称信息相匹配，从而确定出该网页页面中包括该 POI 数据的 POI 名称。

例如，对于具有相同经纬度信息的 POI 数据及网页页面，该 POI 数据中的 POI 名称为“奇虎 360”，而该网页页面中的名称信息为“北京奇虎科
10 技有限公司”，可以确认为该网页页面中包括该 POI 数据的 POI 名称。

优选地，对于具有相同经纬度信息的 POI 数据及网页页面，当判断出该网页页面中包括多个名称信息时，分别计算多个名称信息与该网页页面的地址信息之间的文本距离。将最小的文本距离所对应的名称信息，确定为与该网页页面中地址信息相对应的名称信息。其中，文本距离可以是名称信息与
15 地址信息之间的字符的数量。

对于具有相同经纬度信息的 POI 数据及网页页面，将该 POI 数据对应的 POI 名称与该网页页面中地址信息相对应的名称信息进行比对。比对一致时，确定该网页页面中包括该 POI 数据的 POI 名称。

具体地，对于具有相同经纬度信息的 POI 数据及网页页面，判断该 POI
20 数据对应的 POI 名称，与该网页页面中地址信息相对应的名称信息是否一致：若是，则确定该网页页面中包括该 POI 数据的 POI 名称；否则，确定该网页页面中不包括该 POI 数据的 POI 名称。

S106：对于具有相同经纬度信息的 POI 数据及网页页面，当该网页页面中包括该 POI 数据的 POI 名称时，确定该网页页面包括该兴趣点 POI 数
25 据。

具体地，对于具有相同经纬度信息的 POI 数据及网页页面，当在上述步骤 S105 中确定出该网页页面中包括该 POI 数据的 POI 名称时，在本步骤中确定该网页页面包括该兴趣点 POI 数据，具体地确定该网页页面包括该兴趣点 POI 名称和地址信息。

30 基于上述用于确定网页页面中包括兴趣点 POI 数据的方法，本发明提供

了用于确定网页页面中包括兴趣点 POI 数据的装置, 该装置的内部结构的框架示意图如图 2 所示, 包括: POI 数据获取模块 201、网页页面爬取模块 202、经纬度信息归一化模块 203、经纬度信息匹配模块 204、网页页面包括 POI 名称确定模块 205 和网页页面包括 POI 数据确定模块 206。

5 其中, POI 数据获取模块 201 用于从互联网中获取多个 POI 数据。

网页页面爬取模块 202 用于爬取包括地址信息的多个网页页面。

具体地, 网页页面爬取模块 202 从互联网中爬取包括地址关键词的多个网页页面; 提取多个网页页面中与地址关键词相关联的多个文本信息; 从多个文本信息中提取相应网页页面的地址信息。

10 经纬度信息归一化模块 203 用于将多个 POI 数据中的地址信息及多个网页页面包含的地址信息分别归一化为经纬度信息。

具体地, 经纬度信息归一化模块 203 将多个 POI 数据中的地址信息分别归一化为多个 POI 数据的经纬度信息。较佳地, 对于每个 POI 数据中的地址信息, 从预先获取的地理信息库中查找出该地址信息所对应的经纬度信息, 15 将查找出的经纬度信息确定为该 POI 数据的经纬度信息。其中, 预先获取的地理信息库中包括全国的省、市、县(区)、乡镇、道路等的地址信息、经纬度信息, 以及地址信息与经纬度信息之间的对应关系。

以及, 经纬度信息归一化模块 203 将多个网页页面包含的地址信息分别归一化为多个网页页面的经纬度信息。较佳地, 对于每个网页页面包含的地址信息, 从预先获取的地理信息库中查找出该地址信息所对应的经纬度信息, 20 将查找出的经纬度信息确定为该网页页面的经纬度信息。

经纬度信息匹配模块 204 用于基于同一经纬度信息, 在多个 POI 数据的经纬度信息与多个网页页面中经纬度信息中进行匹配。具体地, 经纬度信息匹配模块 204 对于每个 POI 数据, 判断各网页页面中, 是否存在经纬度信息 25 与该 POI 数据的经纬度信息相一致的网页页面, 若是, 则确定出该 POI 数据与该网页页面相匹配, 即确定出该 POI 数据与该网页页面具有相同经纬度信息, 否则, 忽略该 POI 数据。

网页页面包括 POI 名称确定模块 205 用于对于具有相同经纬度信息的 POI 数据及网页页面, 根据该 POI 数据对应的 POI 名称在该网页页面中进行 30 查找, 确定该网页页面中是否包括该 POI 数据的 POI 名称。具体地, 网页页

面包括 POI 名称确定模块 205 对于具有相同经纬度信息的 POI 数据及网页页面，从该网页页面中查找出所有的名称信息；对于查找出的每个名称信息，判断该名称信息是否与该 POI 数据中的 POI 名称相匹配：若是，确定出该网页页面中包括该 POI 数据的 POI 名称；否则，忽略该 POI 数据。

- 5 网页页面包括 POI 数据确定模块 206 用于当该网页页面中包括该 POI 数据的 POI 名称时，确定该网页页面包括该兴趣点 POI 数据。

优选地，POI 数据获取模块 201 的内部结构的框架示意图如图 3 所示，包括：网页爬取单元 301 和 POI 数据提取单元 302。

其中，网页爬取单元 301 用于从互联网中爬取多个包括 POI 数据的网页。

- 10 具体地，网页爬取单元 301 从互联网中爬取包括地址关键词的多个网页对应的多个 URL；对多个 URL 进行 pattern 聚类，将具有相同结构特征的 URL 聚类为同一 pattern 集合；从多个 pattern 集合中筛选出包括多个包括 POI 数据的网页的 pattern 集合，并从该 pattern 集合中提取多个包括 POI 数据的网页。

- 15 POI 数据提取单元 302 用于从多个包括 POI 数据的网页中提取多个 POI 数据。

- 具体地，POI 数据提取单元 302 具体用于基于属于同一 pattern 集合中多个 URL 对应多个包括 POI 数据的网页的页面结构特征，生成与该 pattern 集合相应的 POI 数据提取模板；基于 POI 数据提取模板，从多个包括 POI 数据的网页中提取多个 POI 数据。
- 20

优选地，如图 2 所示，本发明的用于确定网页页面中包括兴趣点 POI 数据的装置，还包括：网页页面中名称信息确定模块 207。

- 网页页面中名称信息确定模块 207 用于当判断该网页页面中包括多个名称信息时，分别计算多个名称信息与该网页页面的地址信息之间的文本距离；将最小的文本距离所对应的名称信息，确定为与该网页页面中地址信息相对应的名称信息。
- 25

此时，网页页面包括 POI 名称确定模块 205 还用于将该 POI 数据对应的 POI 名称与该网页页面中地址信息相对应的名称信息进行比对；比对一致时，确定该网页页面中包括该 POI 数据的 POI 名称。

- 30 本发明实施例的技术方案中，将地址信息归一化为经纬度信息，可以滤

除非地理的地址信息，从而有利于后续避免收集地址信息虚假的 POI 数据；在 POI 数据的经纬度信息与网页页面中的经纬度信息相匹配的基础上，进一步确定网页页面中包括 POI 数据的 POI 名称，有利于后续避免收集一些经纬度信息或者 POI 名称无法匹配的 POI 数据，而无法匹配的 POI 数据的准确性往往较低，从而可以便于后续根据确定出的网页页面中包括的兴趣点 POI 数据，收集到更为准确的 POI 数据。

本技术领域技术人员可以理解，本发明包括涉及用于执行本申请中所述操作中的一项或多项的设备。这些设备可以为所需的目的地而专门设计和制造，或者也可以包括通用计算机中的已知设备。这些设备具有存储在其内的计算机程序，这些计算机程序选择性地激活或重构。这样的计算机程序可以被存储在设备（例如，计算机）可读介质中或者存储在适于存储电子指令并分别耦联到总线的任何类型的介质中，所述计算机可读介质包括但不限于任何类型的盘（包括软盘、硬盘、光盘、CD-ROM、和磁光盘）、ROM（Read-Only Memory，只读存储器）、RAM（Random Access Memory，随机存储器）、EPROM（Erasable Programmable Read-Only Memory，可擦写可编程只读存储器）、EEPROM（Electrically Erasable Programmable Read-Only Memory，电可擦可编程只读存储器）、闪存、磁性卡片或光线卡片。也就是，可读介质包括由设备（例如，计算机）以能够读的形式存储或传输信息的任何介质。

本技术领域技术人员可以理解，可以用计算机程序指令来实现这些结构图和/或框图和/或流程图中的每个框以及这些结构图和/或框图和/或流程图中的框的组合。本技术领域技术人员可以理解，可以将这些计算机程序指令提供给通用计算机、专业计算机或其他可编程数据处理方法的处理器来实现，从而通过计算机或其他可编程数据处理方法的处理器来执行本发明公开的结构图和/或框图和/或流程图的框或多个框中指定的方案。

本技术领域技术人员可以理解，本发明中已经讨论过的各种操作、方法、流程中的步骤、措施、方案可以被交替、更改、组合或删除。进一步地，具有本发明中已经讨论过的各种操作、方法、流程中的其他步骤、措施、方案也可以被交替、更改、重排、分解、组合或删除。进一步地，现有技术中的具有与本发明中公开的各种操作、方法、流程中的步骤、措施、方案也可以被交替、更改、重排、分解、组合或删除。

以上所述仅是本发明的部分实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

本发明的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本发明实施例的用于确定网页页面中包括兴趣点 POI 数据的装置中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序（例如，计算机程序和计算机程序产品）。这样的实现本发明的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下下载得到，或者在载体信号上提供，或者以任何其他形式提供。

例如，图4示意性地示出了用于执行根据本发明的方法的计算设备的框图。该计算设备传统上包括处理器410和以存储器420形式的计算机程序产品或者计算机可读介质。存储器420可以是诸如闪存、EEPROM（电可擦除可编程只读存储器）、EPROM、硬盘或者ROM之类的电子存储器。存储器420具有用于执行上述方法中的任何方法步骤的程序代码431的存储空间430。例如，用于程序代码的存储空间430可以包括分别用于实现上面的方法中的各种步骤的各个程序代码431。这些程序代码可以从一个或者多个计算机程序产品中读出或者写入到这一个或者多个计算机程序产品中。这些计算机程序产品包括诸如硬盘，紧致盘（CD）、存储卡或者软盘之类的程序代码载体。这样的计算机程序产品通常为如参考图5所述的便携式或者固定存储单元。该存储单元可以具有与图4的计算设备中的存储器420类似布置的存储段、存储空间等。程序代码可以例如以适当形式进行压缩。通常，存储单元包括用于执行根据本发明的方法步骤的计算机可读代码431'，即可以由例如诸如410之类的处理器读取的代码，这些代码当由计算设备运行时，导致该计算设备执行上面所描述的方法中的各个步骤。

30 应该注意的是上述实施例对本发明进行说明而不是对本发明进行

限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包括”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

10 此外，还应当注意，本说明书中使用的语言主要是为了可读性和教导的目的而选择的，而不是为了解释或者限定本发明的主题而选择的。因此，在不偏离所附权利要求书的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。对于本发明的范围，对本发明所做的公开是说明性的，而非限制性的，本
15 发明的范围由所附权利要求书限定。

本发明可以应用于计算机系统/服务器，其可与众多其它通用或专用计算系统环境或配置一起操作。适于与计算机系统/服务器一起使用的众所周知的计算系统、环境和/或配置的例子包括但不限于：个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人
20 电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境，等等。

计算机系统/服务器可以在由计算机系统执行的计算机系统可执行指令（诸如程序模块）的一般语境下描述。通常，程序模块可以包
25 括例程、程序、目标程序、组件、逻辑、数据结构等等，它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施，分布式云计算环境中，任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中，程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

30 本文中所称的“一个实施例”、“实施例”或者“一个或者多个

实施例”意味着，结合实施例描述的特定特征、结构或者特性包括在本发明的至少一个实施例中。此外，请注意，这里“在一个实施例中”的词语例子不一定全指同一个实施例。

权 利 要 求

1. 一种用于确定网页页面中包括兴趣点 POI 数据的方法，其特征在于，包括：

5 从互联网中获取多个 POI 数据；

爬取包括地址信息的多个网页页面；

将所述多个 POI 数据中的地址信息及所述多个网页页面包含的地址信息分别归一化为经纬度信息；

10 基于同一经纬度信息，在所述多个 POI 数据的经纬度信息与多个网页页面中经纬度信息中进行匹配；

对于具有相同经纬度信息的 POI 数据及网页页面，根据该 POI 数据对应的 POI 名称在该网页页面中进行查找，确定该网页页面中是否包括该 POI 数据的 POI 名称；

15 当该网页页面中包括该 POI 数据的 POI 名称时，确定该网页页面包括该兴趣点 POI 数据。

2. 根据权利要求 1 所述的方法，其特征在于，所述从互联网中获取多个 POI 数据的步骤进一步包括：

从互联网中爬取多个包括 POI 数据的网页；

从所述多个包括 POI 数据的网页中提取多个 POI 数据。

20 3. 根据权利要求 1-2 任一项所述的方法，其特征在于，所述从互联网中爬取多个包括 POI 数据的网页的步骤进一步包括：

从互联网中爬取包括地址关键词的多个网页对应的多个 URL；

对所述多个 URL 进行 pattern 聚类，将具有相同结构特征的 URL 聚类为同一 pattern 集合；

25 从多个 pattern 集合中筛选出包括多个包括 POI 数据的网页的 pattern 集合，并从该 pattern 集合中提取多个包括 POI 数据的网页。

4. 根据权利要求 1-3 任一项所述的方法，其特征在于，所述从所述多个包括 POI 数据的网页中提取多个 POI 数据的步骤进一步包括：

30 基于属于同一 pattern 集合中多个 URL 对应多个包括 POI 数据的网页的页面结构特征，生成与该 pattern 集合相应的 POI 数据提取模板；

基于所述 POI 数据提取模板，从所述多个包括 POI 数据的网页中提取多个 POI 数据。

5. 根据权利要求 1-4 中任一项所述的方法, 其特征在于, 所述爬取包括地址信息的多个网页页面的步骤进一步包括:

从互联网中爬取包括地址关键词的多个网页页面;

提取所述多个网页页面中与所述地址关键词相关联的多个文本信息;

5 从所述多个文本信息中提取相应网页页面的地址信息。

6. 根据权利要求 1-5 中任一项所述的方法, 其特征在于, 该方法还包括:

当判断该网页页面中包括多个名称信息时, 分别计算所述多个名称信息与该网页页面的地址信息之间的文本距离;

10 将最小的文本距离所对应的名称信息, 确定为与该网页页面中地址信息相对应的名称信息;

其中, 所述根据该 POI 数据对应的 POI 名称在该网页页面中进行查找, 确定该网页页面中是否包括该 POI 数据的 POI 名称的步骤进一步包括:

将该 POI 数据对应的 POI 名称与该网页页面中地址信息相对应的名称信息进行比对;

15 比对一致时, 确定该网页页面中包括该 POI 数据的 POI 名称。

7. 一种用于确定网页页面中包括兴趣点 POI 数据的装置, 其特征在于, 包括:

POI 数据获取模块, 用于从互联网中获取多个 POI 数据;

20 网页页面爬取模块, 用于爬取包括地址信息的多个网页页面;

经纬度信息归一化模块, 用于将所述多个 POI 数据中的地址信息及所述多个网页页面包含的地址信息分别归一化为经纬度信息;

经纬度信息匹配模块, 用于基于同一经纬度信息, 在所述多个 POI 数据的经纬度信息与多个网页页面中经纬度信息中进行匹配;

25 网页页面包括 POI 名称确定模块, 用于对于具有相同经纬度信息的 POI 数据及网页页面, 根据该 POI 数据对应的 POI 名称在该网页页面中进行查找, 确定该网页页面中是否包括该 POI 数据的 POI 名称;

网页页面包括 POI 数据确定模块, 用于当该网页页面中包括该 POI 数据的 POI 名称时, 确定该网页页面包括该兴趣点 POI 数据。

30

8. 根据权利要求 7 所述的装置, 其特征在于, 所述 POI 数据获取模块具体包括:

网页爬取单元，用于从互联网中爬取多个包括 POI 数据的网页；

POI 数据提取单元，用于从所述多个包括 POI 数据的网页中提取多个 POI 数据。

5 9. 根据权利要求 7-8 任一项所述的装置，其特征在于，

所述网页爬取单元具体用于从互联网中爬取包括地址关键词的多个网页对应的多个 URL；对所述多个 URL 进行 pattern 聚类，将具有相同结构特征的 URL 聚类为同一 pattern 集合；从多个 pattern 集合中筛选出包括多个包括 POI 数据的网页的 pattern 集合，并从该 pattern 集合中提取多个包括 POI
10 数据的网页。

10. 根据权利要求 7-9 任一项所述的装置，其特征在于，

所述 POI 数据提取单元具体用于基于属于同一 pattern 集合中多个 URL 对应多个包括 POI 数据的网页的页面结构特征，生成与该 pattern 集合相应的 POI 数据提取模板；基于所述 POI 数据提取模板，从所述多个包括 POI
15 数据的网页中提取多个 POI 数据。

11. 根据权利要求 7-10 中任一所述的装置，其特征在于，

所述网页页面爬取模块具体用于从互联网中爬取包括地址关键词的多个网页页面；提取所述多个网页页面中与所述地址关键词相关联
20 的多个文本信息；从所述多个文本信息中提取相应网页页面的地址信息。

12. 根据权利要求 7-10 中任一所述的装置，其特征在于，还包括：网页页面中名称信息确定模块；

25 所述网页页面中名称信息确定模块用于当判断该网页页面中包括多个名称信息时，分别计算所述多个名称信息与该网页页面的地址信息之间的文本距离；将最小的文本距离所对应的名称信息，确定为与该网页页面中地址信息相对应的名称信息；以及

所述网页页面包括 POI 名称确定模块还用于将该 POI 数据对应的
30 POI 名称与该网页页面中地址信息相对应的名称信息进行比对；比对一致时，确定该网页页面中包括该 POI 数据的 POI 名称。

13、一种计算机程序，包括计算机可读代码，当所述计算机可读代码在计算设备上运行时，导致所述计算设备执行根据权利要求 1-6 中的任一项所述的用于确定网页页面中包括兴趣点 POI 数据的方法。

14、一种计算机可读介质，其中存储了如权利要求 13 所述的计算机程序。

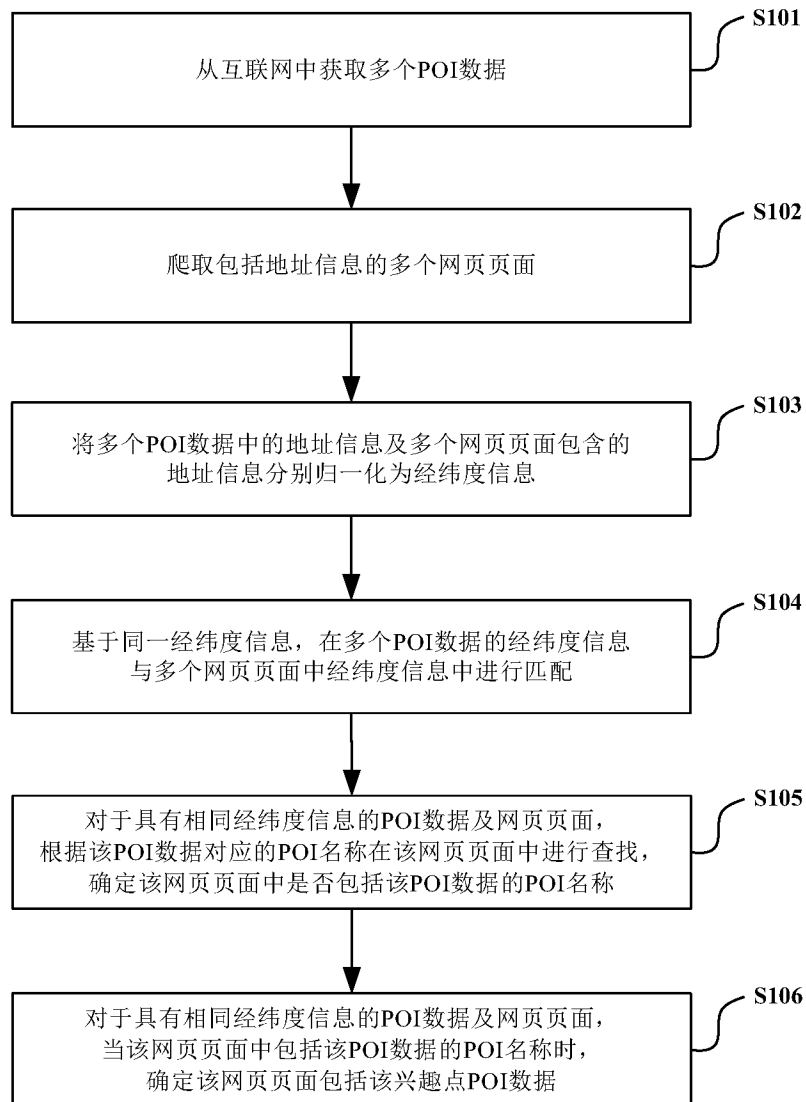
附图

图 1a

boy london	
地址: 朝阳区三里屯路33号3.3大厦一楼门店	
标签: 服饰鞋包 三里屯	
COACH(悠唐折扣店)	
地址: 朝阳区三丰北里悠唐购物中心2期1楼2-132,133室 57478856	
标签: 服饰鞋包 朝外大街	
米兰站(三里屯店) 分店	
地址: 朝阳区三里屯北路19号三里屯太古里南区S1-16 64169929	
标签: 服饰鞋包 三里屯	
Stradivarius(远洋未来广场店) 分店	
地址: 朝阳区北四环东路73号远洋未来广场2楼 84445633	
标签: 服饰鞋包 亚运村 小营	
Li(侨福芳草地店) 分店	
地址: 朝阳区东大桥路9号侨福芳草地大厦LG1-01号	
标签: 服饰鞋包 朝外大街 国贸	

图 1b

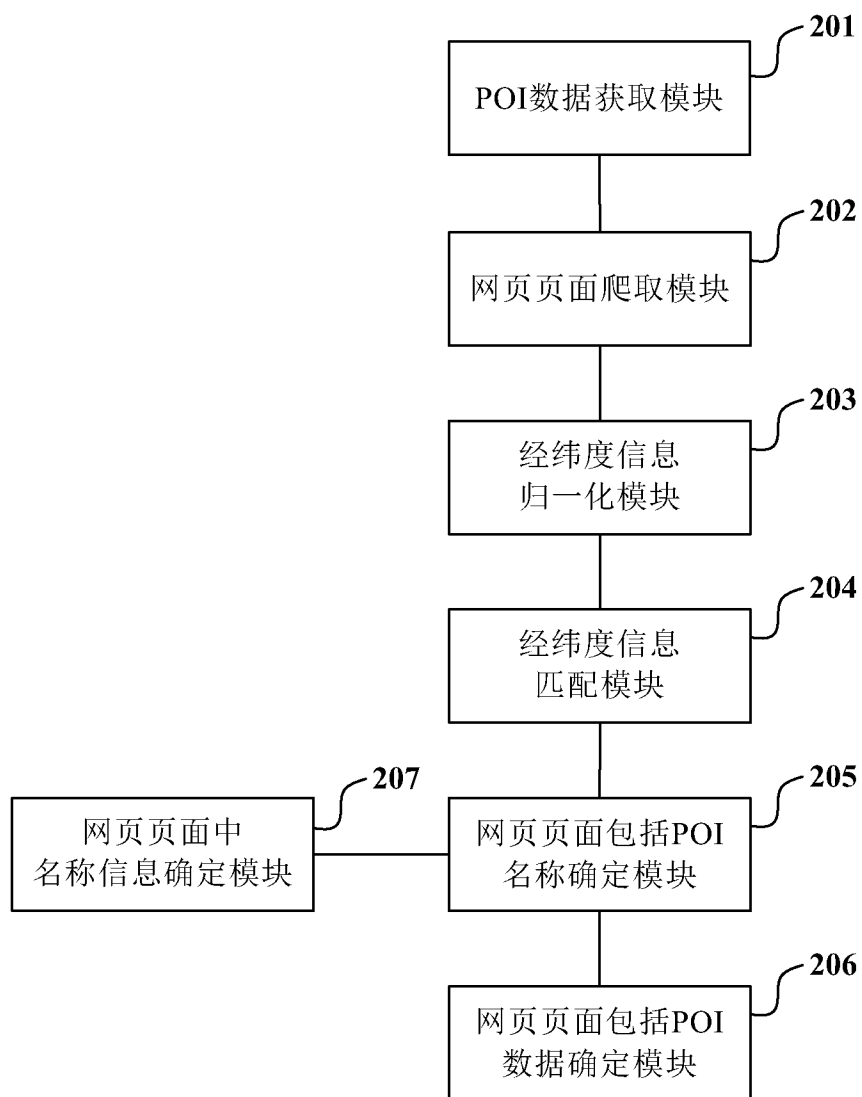


图 2

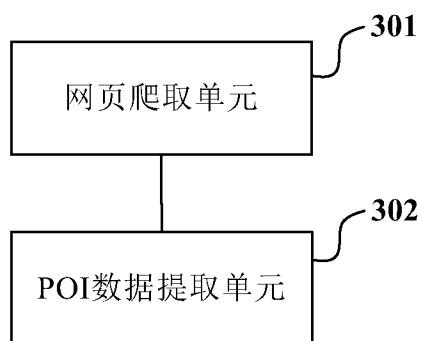


图 3

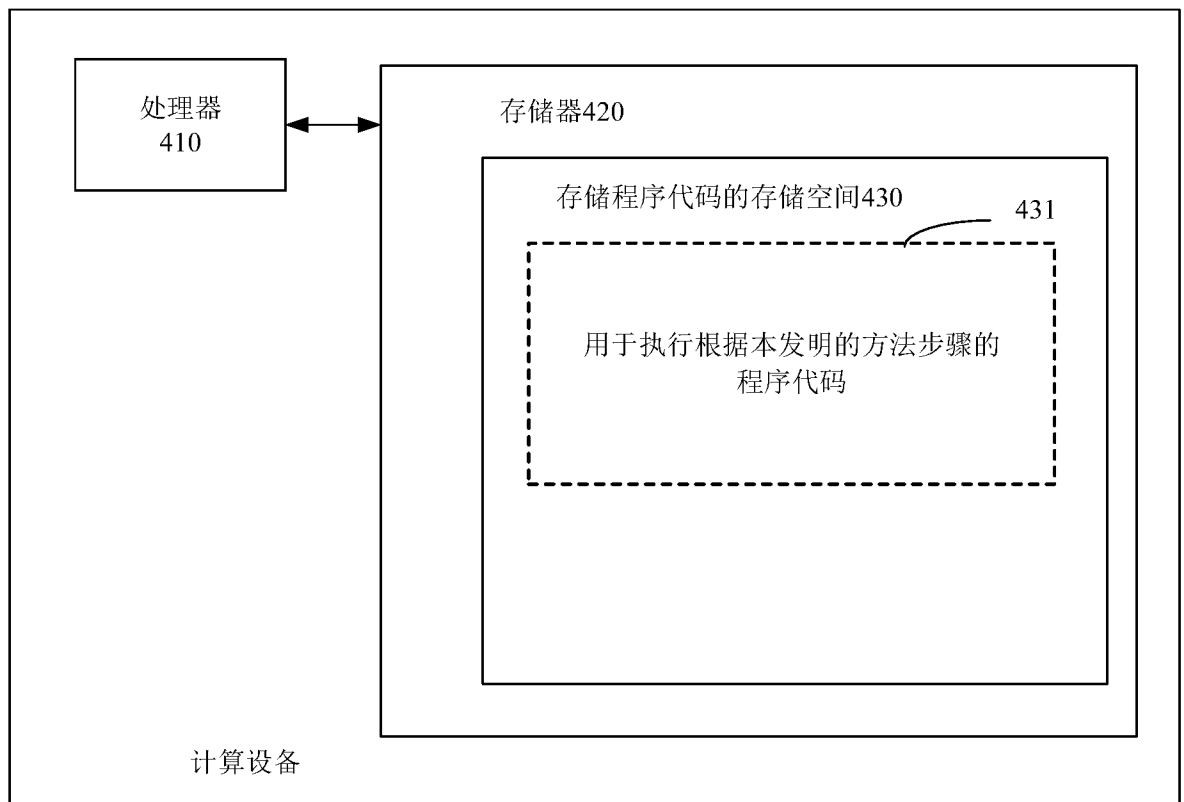


图4

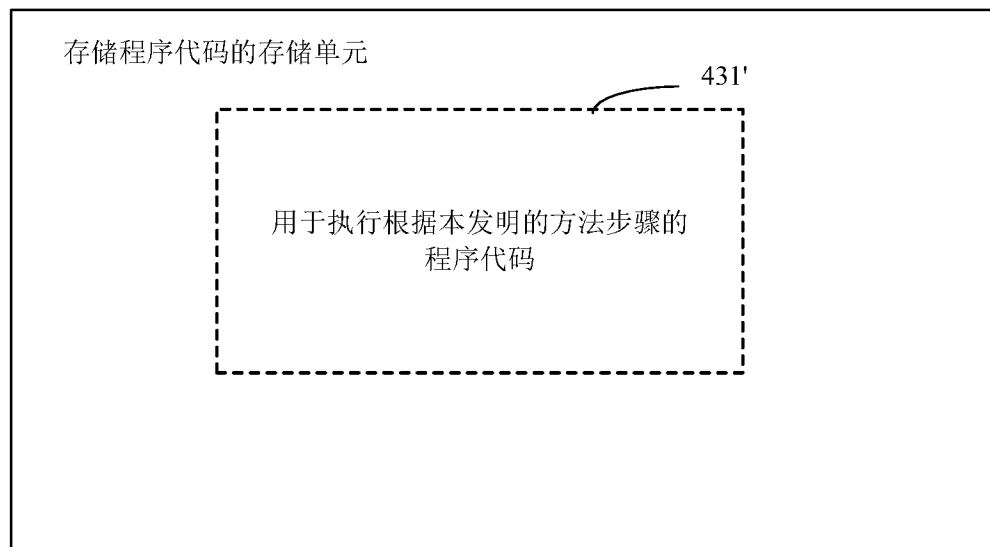


图5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/099580

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) 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, EPODOC, WPI, CNKI, GOOGLE: point of interest, POI, webpage, longitude and latitude, interest, point, page, website, address, name, longitude, latitude, match, find

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104699835 A (BEIJING QIHOO TECHNOLOGY CO., LTD. et al.), 10 June 2015 (10.06.2015), description, paragraphs [0006]-[0019] and [0087]-[0088]	1-14
A	CN 102142003 A (HUAWEI SOFTWARE TECHNOLOGY CO., LTD.), 03 August 2011 (03.08.2011), description, paragraphs [0004]-[0011] and [0063]-[0162], and figures 1-2	1-14
A	CN 101963962 A (AUTONAVI SOFTWARE CO., LTD.), 02 February 2011 (02.02.2011), the whole document	1-14
A	CN 103514234 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.), 15 January 2014 (15.01.2014), the whole document	1-14
A	US 2008040684 A1 (CRUMP RICHARD), 14 February 2008 (14.02.2008), the whole document	1-14

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 17 March 2016 (17.03.2016)	Date of mailing of the international search report 31 March 2016 (31.03.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer LI, Wanyi Telephone No.: (86-10) 62413992

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/099580

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104699835 A	10 June 2015	None	
CN 102142003 A	03 August 2011	CN 102142003 B	24 April 2013
CN 101963962 A	02 February 2011	CN 101963962 B	26 February 2014
CN 103514234 A	15 January 2014	None	
US 2008040684 A1	14 February 2008	US 2008040028 A1	14 February 2008
		US 2008040678 A1	14 February 2008

A. 主题的分类 G06F 17/30 (2006.01) 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, EPDOC, WPI, CNKI, GOOGLE: 兴趣点, POI, 网页, 页面, 地址, 名称, 经纬度, 匹配, 查找, interest, point, page, website, address, name, longitude, latitude, match, find																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 104699835 A (北京奇虎科技有限公司等) 2015年 6月 10日 (2015 - 06 - 10) 说明书第[0006]-[0019]段, 第[0087]-[0088]段</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 102142003 A (华为软件技术有限公司) 2011年 8月 3日 (2011 - 08 - 03) 说明书第[0004]-[0011]段, 第[0063]-[0162]段以及图1-2</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 101963962 A (高德软件有限公司) 2011年 2月 2日 (2011 - 02 - 02) 全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 103514234 A (北京百度网讯科技有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>US 2008040684 A1 (CRUMP RICHARD) 2008年 2月 14日 (2008 - 02 - 14) 全文</td> <td>1-14</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 104699835 A (北京奇虎科技有限公司等) 2015年 6月 10日 (2015 - 06 - 10) 说明书第[0006]-[0019]段, 第[0087]-[0088]段	1-14	A	CN 102142003 A (华为软件技术有限公司) 2011年 8月 3日 (2011 - 08 - 03) 说明书第[0004]-[0011]段, 第[0063]-[0162]段以及图1-2	1-14	A	CN 101963962 A (高德软件有限公司) 2011年 2月 2日 (2011 - 02 - 02) 全文	1-14	A	CN 103514234 A (北京百度网讯科技有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文	1-14	A	US 2008040684 A1 (CRUMP RICHARD) 2008年 2月 14日 (2008 - 02 - 14) 全文	1-14
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 104699835 A (北京奇虎科技有限公司等) 2015年 6月 10日 (2015 - 06 - 10) 说明书第[0006]-[0019]段, 第[0087]-[0088]段	1-14																		
A	CN 102142003 A (华为软件技术有限公司) 2011年 8月 3日 (2011 - 08 - 03) 说明书第[0004]-[0011]段, 第[0063]-[0162]段以及图1-2	1-14																		
A	CN 101963962 A (高德软件有限公司) 2011年 2月 2日 (2011 - 02 - 02) 全文	1-14																		
A	CN 103514234 A (北京百度网讯科技有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文	1-14																		
A	US 2008040684 A1 (CRUMP RICHARD) 2008年 2月 14日 (2008 - 02 - 14) 全文	1-14																		
☐ 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 2016年 3月 17日		国际检索报告邮寄日期 2016年 3月 31日																		
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (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 李婉怡 电话号码 (86-10) 62413992																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/099580

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	104699835	A	2015年 6月 10日	无	
CN	102142003	A	2011年 8月 3日	CN	102142003 B 2013年 4月 24日
CN	101963962	A	2011年 2月 2日	CN	101963962 B 2014年 2月 26日
CN	103514234	A	2014年 1月 15日	无	
US	2008040684	A1	2008年 2月 14日	US	2008040028 A1 2008年 2月 14日
				US	2008040678 A1 2008年 2月 14日