

(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
2018 年 7 月 26 日 (26.07.2018)



(10) 国际公布号
WO 2018/133648 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2017/119820

(22) 国际申请日: 2017 年 12 月 29 日 (29.12.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710042949.1 2017 年 1 月 20 日 (20.01.2017) CN

(71) 申请人: 北京三快在线科技有限公司 (BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD) [CN/CN]; 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。

(72) 发明人: 杨荣权 (YANG, Rongquan); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080

(CN)。覃婷立 (QIN, Tingli); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409 室, Beijing 100085 (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, 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,

(54) Title: SEARCH METHOD AND APPARATUS, AND NON-TEMPORARY COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 搜索方法、装置及非临时性计算机可读存储介质

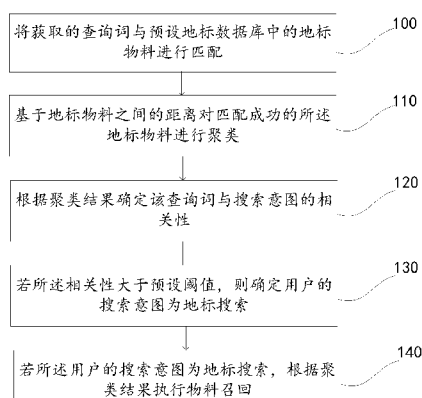


图 1

- 100 Match an obtained query term with landmark materials in a preset landmark database
- 110 Cluster, according to distances among the landmark materials, the landmark materials that have successfully matched
- 120 Determine a correlation between the query term and a search intention according to the clustering result
- 130 If the correlation is greater than a preset threshold, determine that a search intention of a user is a landmark search
- 140 If the search intention of the user is the landmark search, execute a material recall according to the clustering result

(57) Abstract: A search method, a search apparatus, and a computer-readable storage medium. In an example, the method comprises: matching an obtained query term with first landmark materials in a preset landmark database, and determining multiple second landmark materials matching the query term; clustering the multiple second landmark materials according to physical distances among the geographical positions corresponding to the multiple second landmark materials, so as to obtain a clustering result; determining a correlation between the query term and a search intention according to the clustering result; and recalling matching third landmark materials according to the search intention.

(57) 摘要: 一种搜索方法、搜索装置和计算机可读存储介质。在一示例中, 将获取的查询词与预设地标数据库中的第一地标物料进行匹配, 确定与所述查询词相匹配的多个第二地标物料; 基于所述多个第二地标物料各自对应的地理位置之间的物理距离, 对所述多个第二地标物料进行聚类, 得到聚类结果; 根据所述聚类结果, 确定所述查询词与搜索意图之间的相关性; 根据所述搜索意图, 召回相匹配的第三地标物料。

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)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

搜索方法、装置及非临时性计算机可读存储介质

相关申请的交叉引用

5 [01]本专利申请要求于2017年1月20日提交的、申请号为201710042949.1、发明名称为“一种搜索方法及装置、电子设备”的中国专利申请的优先权，该申请的全文以引用的方式并入本文中。

技术领域

10 [02]本申请涉及搜索技术领域，特别是涉及搜索方法、装置及非临时性计算机可读存储介质。

背景技术

[03]在搜索技术领域，可获取到查询词，搜索引擎会首先根据查询词确定用户的搜索意图，然后，根据用户的搜索意图选择相应的搜索策略执行搜索操作。

15 发明内容

[04]本申请提供一种搜索方法，解决现有技术中存在的无法识别与查询词匹配的搜索意图而导致的搜索结果不准确的问题。

[05]为了解决上述问题，第一方面，本申请实施例提供了一种搜索方法，包括：

20 [06]将获取的查询词与预设地标数据库中的第一地标物料进行匹配，确定与所述查询词相匹配的多个第二地标物料；

[07]基于所述多个第二地标物料各自对应的地理位置之间的物理距离，对所述多个第二地标物料进行聚类，得到聚类结果；

[08]根据所述聚类结果，确定所述查询词与搜索意图之间的相关性；

[09]根据所述搜索意图，召回相匹配的第三地标物料。

25 [10]第二方面，本申请实施例提供了一种搜索装置，包括：

[11]处理器和非临时性计算机可读存储介质，

[12]所述非临时性计算机可读存储介质存储有能够被所述处理器执行的机器可执行指令，所述处理器被所述机器可执行指令促使：

[13]将获取的查询词与预设地标数据库中的第一地标物料进行匹配，确定与所述查询词相匹配的多个第二地标物料；

[14]基于所述多个第二地标物料各自对应的地理位置之间的物理距离，对所述多个第二地标物料进行聚类，得到聚类结果；

[15]根据所述聚类结果，确定所述查询词与搜索意图之间的相关性；

[16]根据所述搜索意图，召回相匹配的第三地标物料。

[17]第三方面，本申请实施例提供了一种非临时性计算机可读存储介质，存储有机机器可执行指令，在被网络设备的处理器调用和执行时，所述机器可执行指令促使所述处理器执行：

[18]将获取的查询词与预设地标数据库中的第一地标物料进行匹配，确定与所述查询词相匹配的多个第二地标物料；

[19]基于所述多个第二地标物料各自对应的地理位置之间的物理距离，对所述多个第二地标物料进行聚类，得到聚类结果；

[20]根据所述聚类结果，确定所述查询词与搜索意图之间的相关性；

[21]根据所述搜索意图，召回相匹配的第三地标物料。

[22]本申请实施例公开的搜索方法，通过将获取的查询词与预设地标数据库中的地标物料进行匹配，然后基于地标物料之间的距离对匹配成功的所述地标物料进行聚类，根据聚类结果确定匹配成功的所述地标物料的相关性，若所述相关性大于预设阈值，则确定用户的搜索意图为地标搜索，并进一步根据聚类结果执行物料召回，解决了现有技术中存在的无法准确识别与查询词匹配的搜索意图的问题从而导致的搜索结果不准确的问题。通过结合文本匹配与聚类方法，确定用户的搜索意图，可以准确地识别用户搜索意图，并在采用其他既有搜索策略无法召回搜索结果的情况下，根据地标物料之间的距离确定召回聚类中心点的地标物料，提高了召回搜索结果的准确性。

附图说明

[23]为了更清楚地说明本申请实施例的技术方案，下面将对实施例或现有技术描述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性的前提下，还可以根据这些附图获得其他的附图。

5 [24]图1是本申请示例的搜索方法的流程图；

[25]图2是本申请示例的搜索逻辑的结构示意图；

[26]图3是本申请示例的搜索逻辑的结构示意图；

[27]图4是本申请示例的搜索装置的硬件结构示意图。

10 具体实施方式

[28]下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

15 [29]可将搜索意图对应的搜索物料存储在数据库中。在搜索过程中，可根据查询词与数据库中的搜索物料的文本相关性确定该查询词对应的搜索意图。但，为了保证识别的准确性，可能会设置严格的文本相关准则，导致无法识别某些用户的搜索意图，无法召回搜索结果。

[30]本申请示例公开了一种搜索方法，如图1所示，该方法包括：步骤100至步骤140。

[31]步骤100，将获取的查询词与预设地标数据库中的地标物料进行匹配。

20 [32]具体实施时，预设地标数据库中可包括多条第一地标物料，每条第一地标物料可至少包括：地标名称、地标的地理位置。其中，地标的地理位置通常由地标的经纬度坐标表示。为了提高文本匹配的鲁棒性，地标名称还可包括对应相应的简称、全称、别名、汉字名称、数字名称等。例如，地标数据库中的地标名称可包括“北京市第十八中学（18中）”、“北京大学（Peking University）”等形式的地标。

25 [33]查询词可以为用户通过搜索平台的输入界面手动输入的查询词，也可以是用户点击搜索平台页面上的链接后由页面程序提取的该链接对应的关键词，或者是用户通过搜索平台的搜索频道选择输入的关键词，例如，商家名称、地标名称等。本申请对获取查询词的方式不做

限定。

[34]在获取到查询词之后，搜索引擎会根据获取的查询词在预设地标数据库中执行匹配操作。例如，将获取的所述查询词分别与所述预设地标数据库中每条第一地标物料的名称（可包括简称或全称、别名、汉字名称、数字名称等）进行模糊匹配，选择文本相关性满足预设条件的地标物料名称对应的第一地标物料作为与该查询词相匹配的第二地标物料。在一示例中，经过模糊匹配，将可获取到与所述查询词相匹配的多条第二地标物料。

[35]步骤 110，基于第二地标物料地理位置之间的物理距离，对与查询词相匹配的所述第二地标物料进行聚类。

[36]根据第二地标物料的地理位置，对获取到的与所述查询词匹配的多条第二地标物料进行聚类，将距离相近的第二地标物料聚集到一个簇中，可以得到多个簇。每个簇中可包括多个第二地标物料。具体实施时，可以采用：K-MEANS 算法、K-MEDOIDS 算法、CLARANS 算法等聚类算法对与查询词匹配的第二地标物料按照地理位置进行聚类。

[37]在一示例中，基于第二地标物料之间的距离对所述第二地标物料进行聚类的具体方法如下：以与所述查询词匹配的第二地标物料作为聚类样本，以预设的地标物料之间的距离阈值 D_{th} 作为约束，迭代计算并判断任意两个聚类样本之间的距离，直至所有聚类样本聚集到至少一个簇。

[38]输入：与所述查询词匹配的第二地标物料；

[39]特征：两个第二地标物料之间的距离；

[40]具体算法：计算聚类样本（第二地标物料）中两两样本之间的距离，取其中的最小距离 D_{min} 。如果所述最小距离 D_{min} 在预设的距离阈值 D_{th} 范围内，合并所述最小距离 D_{min} 对应的两个样本，如样本 A 和 B，即根据最小距离 D_{min} 对应的两个样本 A 和 B 生成一个聚类样本 C，并删除最小距离 D_{min} 对应的两个样本 A 和 B。在根据最小距离 D_{min} 对应的两个聚类样本 A 和 B 生成一个聚类样本 C（地标物料）时，取两个聚类样本的地理位置的中间点的经纬度坐标作为生成的聚类样本 C 的地理位置。

[41]重复上述计算距离和样本合并的过程，直到所有的聚类样本都聚集到一个簇，或者最近的两个聚类样本之间的距离大于预设的距离阈值 D_{th} 。

[42]经过前述聚类过程，将得到所述聚类样本对应的至少一个簇，其中，每个簇中包括至少一个地理位置，所述地理位置可为第二地标物料的地理位置，或者根据第二地标物料的地理位置生成的聚类样本的地理位置。根据聚类得到的每个簇中的地理位置，可以确定每个簇对

应的地标物料。具体实施时，可以遍历聚类得到的每个簇中的地理位置，取与该地理位置最近的第二地标物料，作为每个簇对应的预设地标数据库中的地标物料。

[43] 步骤 120，根据聚类结果确定该查询词与搜索意图的相关性。

[44] 聚类结果的聚集程度反映了查询词与地标搜索的相关性。具体实施时，通过聚类得到的最大簇中对应的第二地标物料的数量和与查询词相匹配的第二地标物料的数量比例表示聚类结果的聚集程度，作为查询词与搜索意图（例如，地标搜索）的相关性。所述查询词与所述搜索意图之间的相关性为所述聚类结果中所述多个第二地标物料之间的相关性。在一示例中，可根据聚类结果直接确定该查询词与地标搜索的相关性。聚类得到的最大簇中对应的地标物料的数量越多，说明匹配成功的地标物料的聚集程度越高，相关性越强。

[45] 具体实施时，聚类得到的每个簇中包含的每一个地理位置对应一个地标物料。根据聚类结果，确定查询词与地标搜索的相关性，包括：确定聚类得到的最大簇中地标物料的数量和与查询词匹配成功的第二地标物料的数量比值；将所述比值作为查询词与地标搜索的相关性。

[46] 步骤 130，若所述相关性大于预设阈值，则确定用户的搜索意图为地标搜索。

[47] 在一示例中，若所述聚类结果中匹配成功的多个第二地标物料之间的相关性大于预设阈值，确定所述搜索意图为地标搜索。

[48] 具体实施时，所述预设阈值可以为小于 1 的数值，如 70%。如果确定的相关性大于预设阈值，则确定用户的搜索意图为地标搜索，否则，认为用户的搜索意图不是地标搜索。例如，如果聚类结果中，最大簇中包含的地理位置对应的地标物料的数量和与查询词相匹配的第二地标物料的数量比例大于 70%，则确定用户的搜索意图为地标搜索，否则，认为用户的搜索意图不是地标搜索。

[49] 在一示例中，该阈值可以设置为 60%至 90%之间的数值。如果预设阈值设置为较低的数值，表示放宽了聚类结果的判断条件，则搜索的准确率会相应降低。如果预设阈值设置为较高的数值，表示严格聚类结果的判断条件，则搜索的准确率会相应提高，则可能导致匹配的搜索结果较少。

[50] 具体实施时，如果聚类得到的最大簇，即包含最多地理位置的簇所包含的地理位置（对应于地标物料）的个数占用于聚类的总地理位置的总数（相当于聚类样本总数）的比例大于 70%，认为地标的聚集性很高，确定用户的搜索意图为地标搜索。然后，可以取与聚类得到的最大簇的中心点距离最近的地标物料，作为为用户查询到的地标。

[51] 步骤 140, 若所述用户的搜索意图为地标搜索, 根据聚类结果确定第三地标物料, 并召回该第三地标物料。

[52] 具体实施时, 确定第三地标物料, 包括: 确定聚类得到的最大簇的地理位置中心点; 将距离所述最大簇的地理位置中心点最近的地标物料确定为第三地标物料。

5 [53] 经过前述聚类过程, 将得到所述聚类样本对应的多个簇, 每个簇中包括多个地标物料, 每个地标物料对应一个地理位置, 所述地理位置为地标物料的原始地理位置, 或者根据地标物料的地理位置生成的地理位置。根据聚类得到的每个簇中的地理位置, 可以确定每个簇对应的地标物料。具体实施时, 首先确定聚类得到的最大簇中的地理位置中心点; 然后, 遍历聚类样本, 确定聚类样本中与所述地理位置中心点距离最近的第二地标物料, 将该地标物料
10 作为为用户查询到的第三地标物料召回。确定聚类得到的最大簇中的地理位置中心点的过程, 是确定该最大簇中多个地理位置的中心点的过程。确定与所述地理位置中心点距离最近的聚类样本的过程, 即是计算地理位置中心点分别与多个地理位置之间的距离, 并确定最小距离的过程。

[54] 若确定所述用户的搜索意图非地标搜索, 则采用默认的搜索策略执行物料召回。

15 [55] 本申请实施例公开的搜索方法, 通过将获取的查询词与预设地标数据库中的地标物料进行匹配, 然后基于地标物料之间的距离对匹配成功的所述地标物料进行聚类, 根据聚类结果确定查询词与地标搜索的相关性, 若所述相关性大于预设阈值, 则确定用户的搜索意图为地标搜索, 若所述用户的搜索意图为地标搜索, 则根据聚类结果执行物料召回。通过结合文本匹配与聚类方法, 确定用户的搜索意图, 可以准确地识别用户搜索意图, 并根据地标物料之
20 间的距离确定召回聚类中心点的地标物料, 提高了召回搜索结果的准确性。

[56] 基于如图 1 所述的方法, 在本申请示例还公开的一种搜索方法, 在该方法中, 所述将获取的查询词与预设地标数据库中的第一地标物料进行匹配, 包括: 基于文本相关性, 将获取的查询词与预设地标数据库中每个第一地标物料进行模糊匹配。

[57] 具体实施时, 在将获取的查询词与所述预设地标数据库中每条第一地标物料的名称基于
25 文本相关性进行匹配时, 预先设置第一文本相关性判断阈值和第二文本相关性判断阈值。其中, 第一文本相关性判断阈值为判断查询词与地标名称是否匹配的文本相关性判断阈值; 第二文本相关性判断阈值为在搜索策略 (例如, 商家策略、地标策略) 中判断查询词与数据库中的搜索物料是否匹配的文本相关性判断阈值。例如, 第一文本相关性判断阈值低于第二文本相关性判断阈值。

[58]以查询词为“人大西门”为例，假设既有搜索物料中包含“人大”的搜索物料，并设置一个文本相关性判断阈值并根据搜索策略（例如，商家策略、地标策略等）判断“人大西门”和“人大”两个词的文本相关性。由于该文本相关性判断阈值设置较严格，如将文本相关性判断阈值设置为文本相关性得分高于90分，因此，导致查询词“人大西门”无法与搜索物料“人大”匹配成功。

[59]本实施例中，设置了较宽松的第一文本相关性判断阈值，如将第一文本相关性判断阈值设置为文本相关性得分高于80分。当用查询词“人大西门”与预设地标数据库中的“人大”、“人民大学西门烧烤店”等地标物料进行匹配时，由于设置了较宽松的第一文本相关性判断阈值，因此，查询词“人大西门”与预设地标数据库中的“人大”、“人民大学西门烧烤店”等地标物料均可以匹配成功。

[60]具体实施时，还可以通过对查询词进行预处理，如提取核心词的方式，将查询词与地标物料进行模糊匹配。以查询词为“人大西门”为例，可以丢弃不重要的词“西门”，提取核心词“人大”与预设地标数据库中的地标物料进行匹配，那么地标物料“人大宿舍一号楼”也可以匹配成功。

[61]基于文本相关性，将获取的查询词与预设地标数据库中每个地标物料进行模糊匹配能够保证召回的地标物料满足文本相关，保证基本的用户体验。

[62]以上对本申请提供的方法进行了描述。下面对本申请提供的装置进行描述：

[63]参见图4，图4为本申请一些实施例提供的搜索的硬件结构图。该搜索装置40可包括处理器41以及非临时性计算机可读存储介质42。其中，处理器41和非临时性计算机可读存储介质42可经由系统总线43通信。并且，通过读取并执行非临时性计算机可读存储介质42中存储的与搜索逻辑20对应的机器可执行指令，处理器41可执行上文所述的搜索方法。

[64]本文提到的非临时性计算机可读存储介质42可以是任何电子、磁性、光学或其他物理存储装置，可以包含或存储信息，如可执行指令、数据，等等。例如，所述非临时性计算机可读存储介质92可以是RAM（Random Access Memory，随机存取存储器）、易失存储器、非易失性存储器、闪存、存储驱动器（如硬盘驱动器）、固态硬盘、任何类型的存储盘（如光盘、DVD等），或者类似的存储介质，或者它们的组合。

[65]如图2所示，从功能上划分，所述搜索逻辑20可包括：

[66]文本匹配模块200，用于将获取的查询词与预设地标数据库中的第一地标物料进行匹配，确定与所述查询词相匹配的多个第二地标物料；

[67] 聚类模块 210, 用于基于所述多个第二地标物料各自对应的地理位置之间的物理距离, 对所述多个第二地标物料进行聚类, 得到聚类结果;

[68] 相关性确定模块 220, 用于根据所述聚类结果, 确定所述查询词与地标搜索之间的相关性;

5 [69] 物料召回模块 240, 用于若所述相关性大于预设阈值, 根据所述聚类结果确定第三地标物料, 并召回所述第三地标物料。

[70] 在一示例中, 所述搜索装置进一步包括意图识别模块 230。

[71] 所述意图识别模块 230 用于若所述相关性大于预设阈值, 则确定用户的搜索意图为地标搜索。

10 [72] 所述物料召回模块 240 用于若用户的搜索意图为地标搜索, 根据所述聚类结果确定第三地标物料, 并召回所述第三地标物料。

[73] 在一示例中, 如图 3 所示, 所述相关性确定模块 220 包括:

[74] 比例确定单元 2201, 用于确定所述聚类结果中的最大簇对应的地标物料的数量与所述第二地标物料的数量比值;

[75] 相关性确定单元 2202, 用于将所述比值作为所述查询词与地标搜索之间的相关性。

15 [76] 可选的, 所述文本匹配模块 200 具体用于: 基于文本相关性, 将所述查询词与所述地标数据库中每个第一地标物料进行模糊匹配。

[77] 在一示例中, 如图 3 所示, 所述物料召回模块 240 包括:

[78] 中心点确定单元 2401, 用于确定所述聚类结果中的最大簇的地理位置中心点;

20 [79] 地标物料召回单元 2402, 用于将距离所述最大簇的地理位置中心点最近的地标物料作为所述第三地标物料。

[80] 在一示例中, 在对所述多个第二地标物料进行聚类时, 所述第二地标物料作为聚类样本。

[81] 所述聚类模块 210 用于计算所述聚类样本中每两个聚类样本之间的距离, 得到最小距离; 如果所述最小距离小于所述距离阈值, 根据该在最小距离对应的两个聚类样本生成新的聚类样本, 并删除该最小距离对应的两个聚类样本, 其中, 该新生成的聚类样本对应的地理位置为所述两个聚类样本地理位置的中间点; 当剩余的聚类样本聚集到一个簇, 或者每两个剩余的聚类样本之间的距离大于所述距离阈值, 所述聚类完成; 否则, 再次执行计算所述聚类样本中每两个聚类样本之间的距离。

25

[82] 本实施例中公开的搜索装置的各模块的具体实施方式，参见实施例一和实施例二的相关部分，此处不再赘述。

5 [83] 本实施例中公开的搜索装置，通过将获取的查询词与预设地标数据库中的地标物料进行匹配，然后基于地标物料之间的距离对匹配成功的所述地标物料进行聚类，并根据获得的聚类结果确定匹配成功的所述地标物料的相关性，最后若所述相关性大于预设阈值，则确定用
10 户的搜索意图为地标搜索，并根据聚类结果执行物料召回，解决了现有技术中存在的无法准确识别与查询词匹配的搜索意图，无法召回搜索结果的问题。通过结合文本匹配与聚类方法，确定用户的搜索意图，可以准确地识别用户搜索意图，并在采用其他既有搜索策略无法召回搜索结果的情况下，根据地标物料之间的距离确定召回聚类中心点的地标物料，提高了召回
15 搜索结果的准确性。

[84] 本申请还公开了一种电子设备，包括存储器、处理器及存储在所述存储器上并可在处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现本申请实施例一和实施例二所述的搜索方法。所述电子设备可以为 PC 机、移动终端、个人数字助理、平板电脑等。

15 [85] 本申请还公开了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本申请实施例一和实施例二所述的搜索方法的步骤。

[86] 本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。
20

[87] 以上对本申请提供的一种搜索方法及装置进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

25 [88] 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件实现。基于这样的理解，上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如 ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行各个实施

例或者实施例的某些部分所述的方法。

[89] 本说明书中描述的主题及功能操作的实施例可以在以下中实现：数字电子电路、有形体现的计算机软件或固件、包括本说明书中公开的结构及其结构性等同物的计算机硬件、或者它们中的一个或多个的组合。本说明书中描述的主题的实施例可以实现为一个或多个计算机程序，即编码在有形非暂时性程序载体上以被数据处理装置执行或控制数据处理装置的操作的计算机程序指令中的一个或多个模块。可替代地或附加地，程序指令可以被编码在人工生成的传播信号上，例如机器生成的电、光或电磁信号，该信号被生成以将信息编码并传输到合适的接收机装置以由数据处理装置执行。计算机存储介质可以是机器可读存储设备、机器可读存储基板、随机或串行存取存储器设备、或它们中的一个或多个的组合。

[90] 本说明书中描述的处理及逻辑流程可以由执行一个或多个计算机程序的一个或多个可编程计算机执行，以通过根据输入数据进行操作并生成输出来执行相应的功能。所述处理及逻辑流程还可以由专用逻辑电路——例如 FPGA（现场可编程门阵列）或 ASIC（专用集成电路）来执行，并且装置也可以实现为专用逻辑电路。

[91] 适合用于执行计算机程序的计算机包括，例如通用和/或专用微处理器，或任何其他类型的中央处理单元。通常，中央处理单元将从只读存储器和/或随机存取存储器接收指令和数据。计算机的基本组件包括用于实施或执行指令的中央处理单元以及用于存储指令和数据的一个或多个存储器设备。通常，计算机还将包括用于存储数据的一个或多个大容量存储设备，例如磁盘、磁光盘或光盘等，或者计算机将可操作地与此大容量存储设备耦接以从其接收数据或向其传送数据，抑或两种情况兼而有之。然而，计算机不是必须具有这样的设备。此外，计算机可以嵌入在另一设备中，例如移动电话、个人数字助理（PDA）、移动音频或视频播放器、游戏操纵台、全球定位系统（GPS）接收机、或例如通用串行总线（USB）闪存驱动器的便携式存储设备，仅举几例。

[92] 适合于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、媒介和存储器设备，例如包括半导体存储器设备（例如 EPROM、EEPROM 和闪存设备）、磁盘（例如内部硬盘或可移动盘）、磁光盘以及 CD ROM 和 DVD-ROM 盘。处理器和存储器可由专用逻辑电路补充或并入专用逻辑电路中。

[93] 虽然本说明书包含许多具体实施细节，但是这些不应被解释为限制任何发明的范围或所要求保护的发明，而是主要用于描述特定发明的具体实施例的特征。本说明书内在多个实施例中描述的某些特征也可以在单个实施例中被组合实施。另一方面，在单个实施例中描述的各种特征也可以在多个实施例中分开实施或以任何合适的子组合来实施。此外，虽然特征可

以如上所述在某些组合中起作用并且甚至最初如此要求保护，但是来自所要求保护的组合中的一个或多个特征在一些情况下可以从该组合中去除，并且所要求保护的组合可以指向子组合或子组合的变型。

[94] 类似地，虽然在附图中以特定顺序描绘了操作，但是这不应被理解为要求这些操作以所示的特定顺序执行或顺次执行、或者要求所有例示的操作被执行，以实现期望的结果。在某些情况下，多任务和并行处理可能是有利的。此外，上述实施例中的各种系统模块和组件的分离不应被理解为在所有实施例中均需要这样的分离，并且应当理解，所描述的程序组件和系统通常可以一起集成在单个软件产品中，或者封装成多个软件产品。

[95] 由此，主题的特定实施例已被描述。其他实施例在所附权利要求书的范围以内。在某些情况下，权利要求书中记载的动作可以以不同的顺序执行并且仍实现期望的结果。此外，附图中描绘的处理并非必需所示的特定顺序或顺次顺序，以实现期望的结果。在某些实现中，多任务和并行处理可能是有利的。

[96] 需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[97] 以上对本发明实施例所提供的方法和装置进行了详细介绍，本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

权利要求书

1、一种搜索方法，包括：

将获取的查询词与预设地标数据库中的第一地标物料进行匹配，确定与所述查询词相匹配的多个第二地标物料；

5 基于所述多个第二地标物料各自对应的地理位置之间的物理距离，对所述多个第二地标物料进行聚类，得到聚类结果；

根据所述聚类结果，确定所述查询词与搜索意图之间的相关性；

根据所述搜索意图，召回相匹配的第三地标物料。

2、根据权利要求1所述的方法，其中，所述查询词与所述搜索意图之间的相关性为所述
10 聚类结果中所述多个第二地标物料之间的相关性，

所述方法进一步包括：

若所述聚类结果中匹配成功的多个第二地标物料之间的相关性大于预设阈值，确定所述搜索意图为地标搜索；

所述根据所述搜索意图召回相匹配的地标物料，包括：

15 若所述搜索意图为所述地标搜索，根据所述聚类结果确定所述第三地标物料，并召回所述第三地标物料。

3、根据权利要求1所述的方法，其中，根据所述聚类结果确定所述查询词与搜索意图之间的相关性，包括：

20 确定所述聚类结果中的最大簇对应的地标物料的数量与所述第二地标物料的数量之比；

将所述比值作为所述查询词与地标搜索之间的相关性。

4、根据权利要求1所述的方法，其中，所述将获取的查询词与预设地标数据库中的第一地标物料进行匹配的步骤，包括：

基于文本相关性，将所述查询词与所述地标数据库中每个第一地标物料进行模糊匹配。

25 5、根据权利要求2所述的方法，其中，根据所述聚类结果确定所述第三地标物料，包括：确定所述聚类结果中的最大簇的地理位置中心点；

将距离所述最大簇的地理位置中心点最近的地标物料作为所述第三地标物料。

6、根据权利要求1所述的方法，其中，在对所述多个第二地标物料进行聚类时，所述第二地标物料作为聚类样本；

30 对所述多个第二地标物料进行聚类，包括：

计算所述聚类样本中每两个聚类样本之间的距离，得到最小距离；

如果所述最小距离小于所述距离阈值，根据该在最小距离对应的两个聚类样本生成新的聚类样本，并删除该最小距离对应的两个聚类样本，其中，该新生成的聚类样本对应的地理位置为所述两个聚类样本地理位置的中间点；

5 当剩余的聚类样本聚集到一个簇，或者每两个剩余的聚类样本之间的距离大于所述距离阈值，所述聚类完成；

否则，再次执行计算所述聚类样本中每两个聚类样本之间的距离。

7、一种搜索装置，包括：

处理器和非临时性计算机非临时性计算机可读存储介质，

10 所述非临时性计算机非临时性计算机可读存储介质存储有能够被所述处理器执行的机器可执行指令，所述处理器被所述机器可执行指令促使：

将获取的查询词与预设地标数据库中的第一地标物料进行匹配，确定与所述查询词相匹配的多个第二地标物料；

15 基于所述多个第二地标物料各自对应的地理位置之间的物理距离，对所述多个第二地标物料进行聚类，得到聚类结果；

根据所述聚类结果，确定所述查询词与搜索意图之间的相关性；

根据所述搜索意图，召回相匹配的第三地标物料。

8、根据权利要求7所述的装置，其中，所述查询词与所述搜索意图之间的相关性为所述聚类结果中所述多个第二地标物料之间的相关性，

20 所述处理器进一步被所述机器可执行指令促使：

若所述聚类结果中匹配成功的多个第二地标物料之间的相关性大于预设阈值，确定所述搜索意图为地标搜索；

所述处理器被所述机器可执行指令促使：

25 若所述搜索意图为所述地标搜索，根据所述聚类结果确定第三地标物料，并召回所述第三地标物料。

9、根据权利要求7所述的装置，其中，所述处理器被所述机器可执行指令促使：

确定所述聚类结果中的最大簇对应的地标物料的数量与所述第二地标物料的数量之比；

将所述比值作为所述查询词与地标搜索之间的相关性。

30 10、根据权利要求7所述的装置，其中，所述处理器被所述机器可执行指令促使：

基于文本相关性，将所述查询词与所述地标数据库中每个第一地标物料进行模糊匹配。

11、根据权利要求 8 所述的装置，其中，所述处理器被所述机器可执行指令促使：

确定所述聚类结果中的最大簇的地理位置中心点；

将距离所述最大簇的地理位置中心点最近的地标物料作为所述第三地标物料。

5 12、根据权利要求 7 所述的装置，其中，在对所述多个第二地标物料进行聚类时，所述第二地标物料作为聚类样本；

所述处理器被所述机器可执行指令促使：

计算所述聚类样本中每两个聚类样本之间的距离，得到最小距离；

10 如果所述最小距离小于所述距离阈值，根据该在最小距离对应的两个聚类样本生成新的聚类样本，并删除该最小距离对应的两个聚类样本，其中，该新生成的聚类样本对应的地理位置为所述两个聚类样本地理位置的中间点；

当剩余的聚类样本聚集到一个簇，或者每两个剩余的聚类样本之间的距离大于所述距离阈值，所述聚类完成；

否则，再次执行计算所述聚类样本中每两个聚类样本之间的距离。

15 13、一种非临时性计算机可读存储介质，存储有机器可执行指令，在被网络设备的处理器调用和执行时，所述机器可执行指令促使所述处理器执行：

将获取的查询词与预设地标数据库中的第一地标物料进行匹配，确定与所述查询词相匹配的多个第二地标物料；

20 基于所述多个第二地标物料各自对应的地理位置之间的物理距离，对所述多个第二地标物料进行聚类，得到聚类结果；

根据所述聚类结果，确定所述查询词与搜索意图之间的相关性；

根据所述搜索意图，召回相匹配的第三地标物料。

14、根据权利要求 13 所述的非临时性计算机可读存储介质，所述查询词与所述搜索意图之间的相关性为所述聚类结果中所述多个第二地标物料之间的相关性，

25 其中，所述机器可执行指令进一步促使所述处理器执行：

若所述聚类结果中匹配成功的多个第二地标物料之间的相关性大于预设阈值，确定所述搜索意图为地标搜索；

所述机器可执行指令促使所述处理器执行：

30 若所述搜索意图为所述地标搜索，根据所述聚类结果确定第三地标物料，并召回所述第三地标物料。

15、根据权利要求 13 所述的非临时性计算机非临时性计算机可读存储介质，其中，所述机器可执行指令促使所述处理器执行：

确定所述聚类结果中的最大簇对应的地标物料的数量与所述第二地标物料的数量比值；

5 将所述比值作为所述查询词与地标搜索之间的相关性。

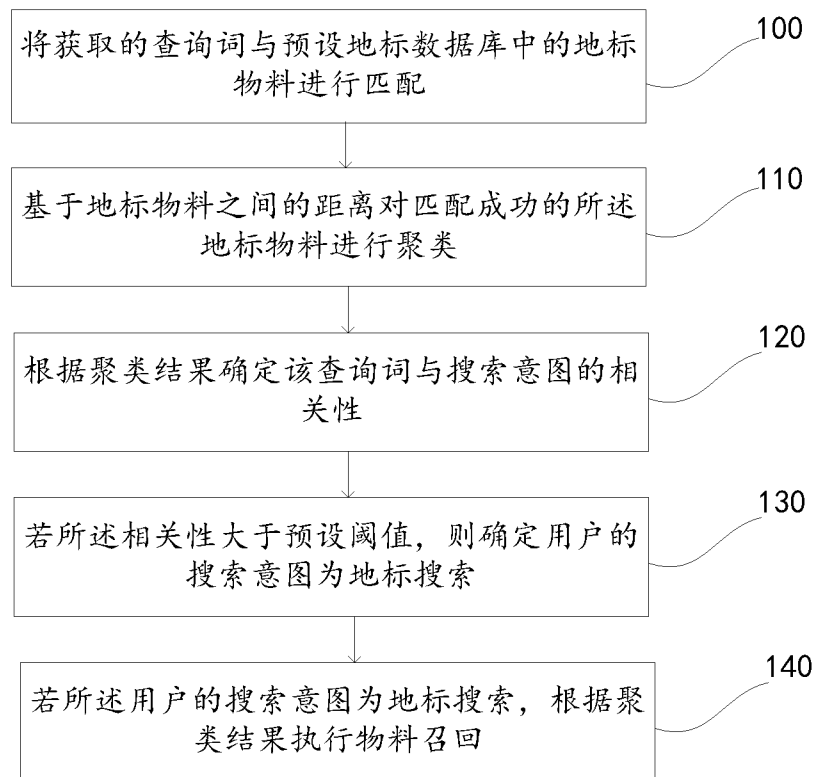


图 1

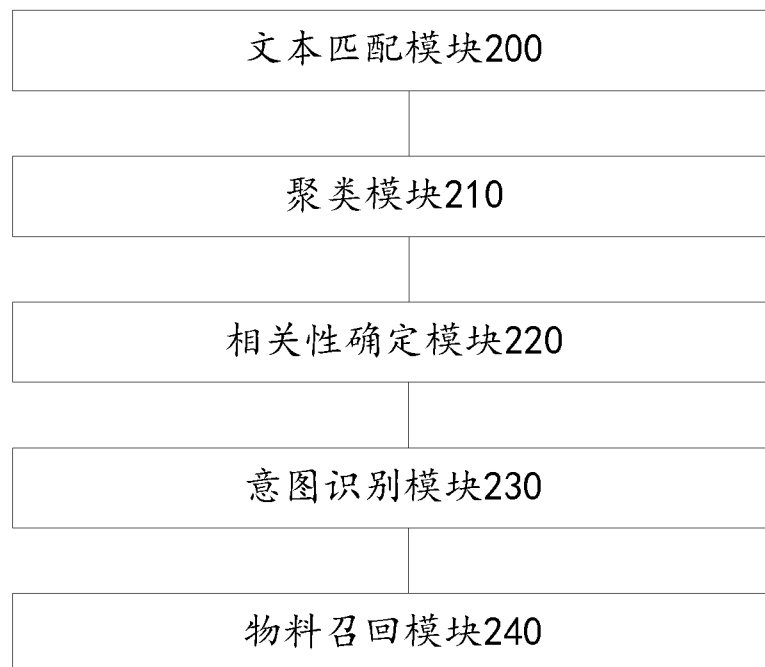


图 2



图 3

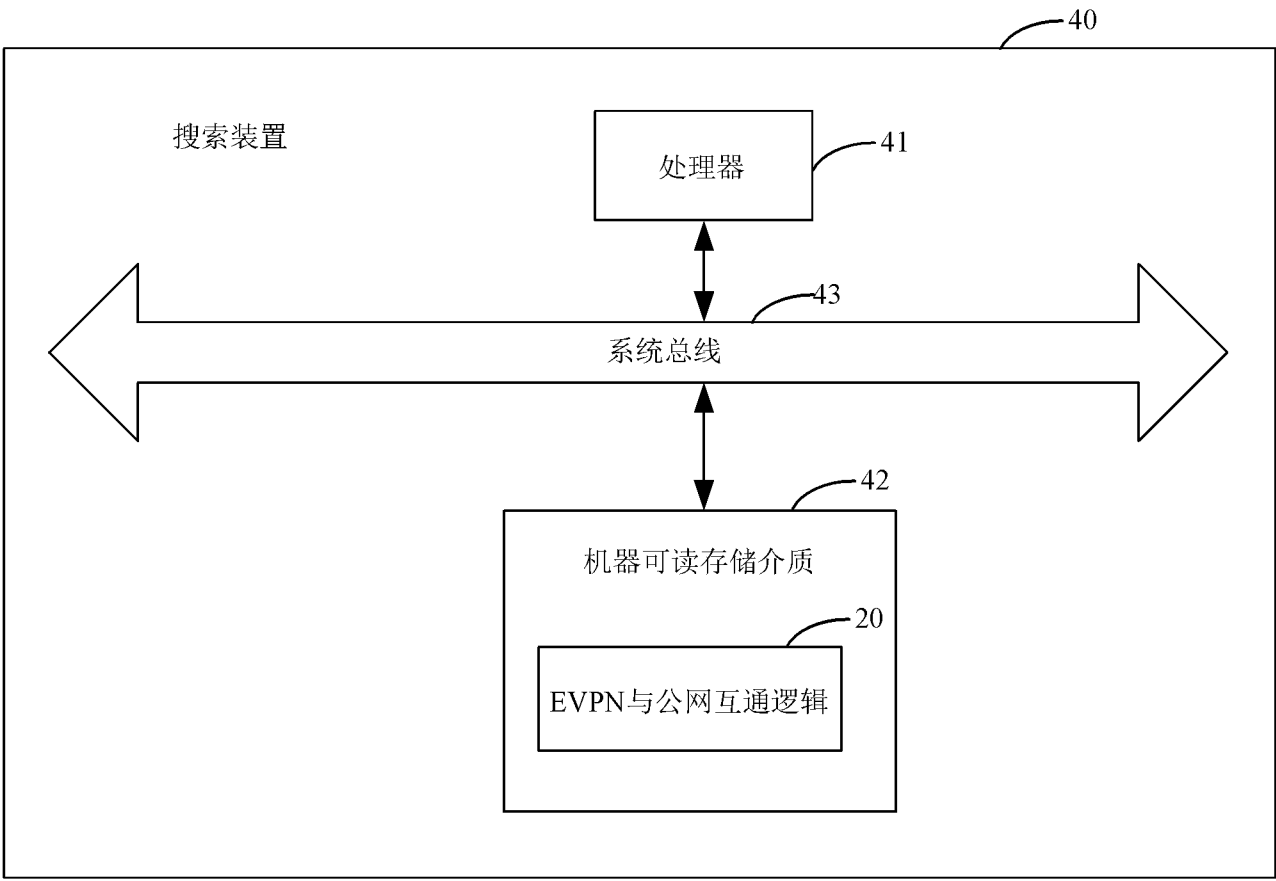


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/119820

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 G06Q

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; CNKI; WPI; EPODOC: 搜索, 查询, 关键词, 聚类, 相关性, 相似度, 检索, 距离, 匹配, search+, intention, cluster+, match+, correlation, keyword

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106933947 A (BEIJING SCIENCE AND TECHNOLOGY CO., THREE FAST ONLINE) 07 July 2017 (07.07.2017), claims 1-10	1-15
A	CN 103902694 A (HARBIN ENGINEERING UNIVERSITY) 02 July 2014 (02.07.2014), description, paragraphs [0005]-[0015] and [0021]-[0044]	1-15
A	CN 106095780 A (DATAGRAND TECH INC.) 09 November 2016 (09.11.2016), entire document	1-15
A	CN 104615620 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 13 May 2015 (13.05.2015), entire document	1-15

☐ 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 24 March 2018	Date of mailing of the international search report 30 March 2018
Name and mailing address of the ISA 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 LIU, Man Telephone No. (86-10) 53961297

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/119820

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106933947 A	07 July 2017	None	
CN 103902694 A	02 July 2014	None	
CN 106095780 A	09 November 2016	None	
CN 104615620 A	13 May 2015	None	

国际检索报告

国际申请号

PCT/CN2017/119820

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 G06Q

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;CNKI;WPI;EPDOC:搜索, 查询, 关键词, 聚类, 相关性, 相似度, 检索, 距离, 匹配, search+, intention, cluster+, match+, correlation, keyword

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 106933947 A (北京三快在线科技有限公司) 2017年 7月 7日 (2017 - 07 - 07) 权利要求1-10	1-15
A	CN 103902694 A (哈尔滨工程大学) 2014年 7月 2日 (2014 - 07 - 02) 说明书第[0005]-[0015], [0021]-[0044]段	1-15
A	CN 106095780 A (达而观信息科技上海有限公司) 2016年 11月 9日 (2016 - 11 - 09) 全文	1-15
A	CN 104615620 A (腾讯科技深圳有限公司) 2015年 5月 13日 (2015 - 05 - 13) 全文	1-15

☐ 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

2018年 3月 24日

国际检索报告邮寄日期

2018年 3月 30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

受权官员

刘曼

传真号 (86-10)62019451

电话号码 (86-10)53961297

国际申请号
PCT/CN2017/119820

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	106933947	A	2017年 7月 7日	无	
CN	103902694	A	2014年 7月 2日	无	
CN	106095780	A	2016年 11月 9日	无	
CN	104615620	A	2015年 5月 13日	无	

表 PCT/ISA/210 (同族专利附件) (2009年7月)