

(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
2020 年 3 月 19 日 (19.03.2020)



(10) 国际公布号
WO 2020/052067 A1

(51) 国际专利分类号:
G06F 16/242 (2019.01)

(21) 国际申请号: PCT/CN2018/116342

(22) 国际申请日: 2018 年 11 月 20 日 (20.11.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811060981.3 2018年9月12日 (12.09.2018) CN

(71) 申请人: 北京字节跳动网络技术有限公司 (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市石景山区实兴大街30号院3号楼2层B-0035 房间, Beijing 100041 (CN)。

(72) 发明人: 何轶 (HE, Yi); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。李磊 (LI, Lei); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。宗显子 (ZONG, Xianzi); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。汤颢 (TANG, Hao); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。郑光果 (ZHENG, Guangguo); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。

(74) 代理人: 北京英赛嘉华知识产权代理有限公司 (INSIGHT INTELLECTUAL PROPERTY

(54) Title: INFORMATION SEARCH METHOD AND DEVICE

(54) 发明名称: 用于搜索信息的方法和装置

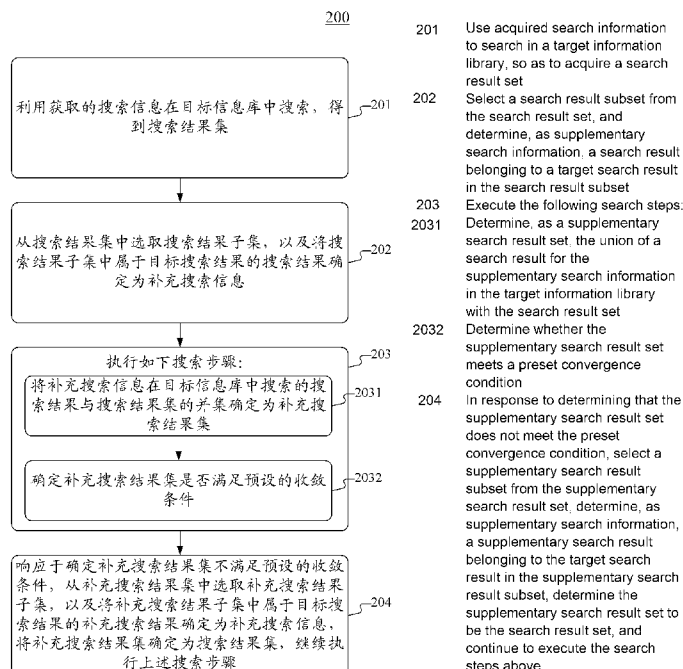


图 2

(57) Abstract: An information search method and device. The method comprises: using search information to search in a target information library, so as to acquire a search result set (201); selecting a search result subset from the search result set, and determining, as supplementary search information, a search result which is a target search result in the search result subset (202); executing the following search steps: determining, as a supplementary search result set, the union of a search result for the supplementary search information in the target information library with the search result set (2031); and determining whether the supplementary search result set meets a

LIMITED); 中国北京市海淀区知春路甲48号盈都大厦A座19A, Beijing 100098 (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, 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))。

preset convergence condition (2032) (203); and in response to determining that the supplementary search result set does not meet the preset convergence condition, selecting a supplementary search result subset from the supplementary search result set, determining, as supplementary search information, a supplementary search result belonging to a target search result in the supplementary search result subset, determining the supplementary search result set to be the search result set, and continuing to execute the search steps (204). The method increases the number of search results.

(57) 摘要: 一种用于搜索信息的方法、装置, 该方法包括: 利用搜索信息在目标信息库中搜索, 得到搜索结果集(201); 从搜索结果集中选取搜索结果子集, 将搜索结果子集中为目标搜索结果的搜索结果确定为补充搜索信息(202); 执行如下搜索步骤: 将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集(2031); 确定补充搜索结果集是否满足预设的收敛条件(2032)(203); 响应于确定补充搜索结果集不满足预设的收敛条件, 从补充搜索结果集中选取补充搜索结果子集, 将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息, 将补充搜索结果集确定为搜索结果集, 继续执行搜索步骤(204)。该方法实现了增大搜索结果的数目。

说明书

用于搜索信息的方法和装置

本专利申请要求于 2018 年 9 月 12 日提交的、申请号为 201811060981.3、申请人为北京字节跳动网络技术有限公司、发明名称为“用于搜索信息的方法和装置”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申请中。

技术领域

本申请实施例涉及计算机技术领域，具体涉及用于搜索信息的方法和装置。

背景技术

在搜索过程中，一般是根据搜索算法直接选择出与输入的搜索信息对应的搜索结果。这种方式对于针对搜索信息进行多次调整之后的信息可能无法直接搜索得到，因此，得到的搜索结果可能不全面。

发明内容

本申请实施例提出了用于搜索信息的方法和装置。

第一方面，本申请实施例提供了一种用于搜索信息的方法，该方法包括：利用获取的搜索信息在目标信息库中搜索，得到搜索结果集；从搜索结果集中选取搜索结果子集，以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息；执行如下搜索步骤：将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集；确定补充搜索结果集是否满足预设的收敛条件；响应于确定补充搜索结果集不满足预设的收敛条件，从补充搜索结果集中选取补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行上述搜索步骤。

在一些实施例中，上述搜索步骤还包括：响应于确定补充搜索结果集满足上述收敛条件，从补充搜索结果集中确定目标搜索结果集。

在一些实施例中，将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息，包括：获取搜索结果子集中的搜索结果的标注信息，其中，标注信息用于表示搜索结果是否是目标搜索结果；根据标注信息，确定搜索结果子集中属于目标搜索结果的搜索结果。

在一些实施例中，在利用搜索信息在目标信息库中搜索，得到搜索结果集之后，还包括：确定搜索结果集中的搜索结果的关联度，其中，关联度用于表示搜索结果与搜索信息的关联程度。

在一些实施例中，针对搜索结果集中的搜索结果，该搜索结果的关联度通过如下步骤确定：确定该搜索结果与搜索信息的相似度；根据相似度，确定该搜索结果的关联度，其中，相似度与关联度成正比。

在一些实施例中，在将补充搜索信息在目标信息库中的搜索结果与搜索结果集的并集确定为补充搜索结果集之后，还包括：针对补充搜索结果集中的补充搜索结果，响应于该补充搜索结果存在于搜索结果集中，按照预设的关联度提升算法，更新该补充搜索结果的关联度，以及将更新后的关联度作为该补充搜索结果的关联度；响应于该补充搜索结果不存在于搜索结果集中，确定该补充搜索结果的关联度。

在一些实施例中，上述确定该补充搜索结果的关联度包括：确定该补充搜索结果与搜索信息的相似度；根据相似度，确定该补充搜索结果的关联度，其中，相似度与关联度成正比。

在一些实施例中，上述从补充搜索结果集中确定目标搜索结果集，包括：按照关联度从大到小的顺序，从补充搜索结果集中选取目标数目个补充搜索结果作为目标搜索结果，得到目标搜索结果集。

在一些实施例中，该方法还包括：按照预设的目标数目个关联度区间，将搜索结果集划分为对应的目标数目个搜索结果子集；以及从搜索结果集中选取搜索结果子集，包括：从目标数目个搜索结果子集中分别选取搜索结果，得到搜索结果子集。

在一些实施例中，该方法还包括：确定目标数目个搜索结果子集中的搜索结果子集的准确度，其中，准确度用于表示搜索结果子集中

目标搜索结果所占的比例。

在一些实施例中，在将补充搜索信息在目标信息库中的搜索结果与搜索结果集的并集确定为补充搜索结果集之后，还包括：按照上述关联度区间，将补充搜索结果集划分为对应的目标数目个补充搜索结果子集；以及从补充搜索结果集中选取补充搜索结果子集，包括：从目标数目个补充搜索结果子集中分别选取补充搜索结果，得到补充搜索结果子集。

在一些实施例中，该方法还包括：确定目标数目个补充搜索结果集中的补充搜索结果子集的准确度，其中，准确度用于表示补充搜索结果子集中的目标搜索结果所占的比例。

在一些实施例中，上述从目标数目个补充搜索结果子集中分别选取补充搜索结果，包括：针对目标数目个补充搜索结果子集中的补充搜索子集，确定该补充搜索结果子集所在的关联度区间对应的搜索结果集的准确度与预设的准确度阈值之间的差值的绝对值；根据绝对值和该补充搜索结果子集包含的补充搜索结果的数目，确定从该补充搜索结果子集中选取的补充搜索结果的数目，其中，绝对值和选取的补充搜索结果的数目成反比，包含的补充搜索结果的数目和选取的补充搜索结果的数目成正比。

在一些实施例中，搜索结果集中的搜索结果的数目和补充搜索结果子集中的补充搜索结果的数目相同。

在一些实施例中，收敛条件包括：目标关联度区间对应的搜索结果子集和补充搜索结果子集的准确度的差值小于预设的差值阈值。

在一些实施例中，收敛条件包括：搜索结果集和补充搜索结果集的差集所包含的搜索结果的数目小于预设的数目差阈值。

第二方面，本申请实施例提供了一种用于搜索信息的装置，该装置包括：搜索单元，被配置成利用获取的搜索信息在目标信息库中搜索，得到搜索结果集；选取单元，被配置成从搜索结果集中选取搜索结果子集，以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息；补充搜索单元，被配置成执行如下搜索步骤：将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定

为补充搜索结果集；确定补充搜索结果集是否满足预设的收敛条件；确定单元，被配置成响应于确定补充搜索结果集不满足预设的收敛条件，从补充搜索结果集中选取补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行上述搜索步骤。

在一些实施例中，上述补充搜索单元进一步被配置成：响应于确定补充搜索结果集满足上述收敛条件，从补充搜索结果集中确定目标搜索结果集。

在一些实施例中，上述选取单元进一步被配置成：获取搜索结果子集中的搜索结果的标注信息，其中，标注信息用于表示搜索结果是否是目标搜索结果；根据标注信息，确定搜索结果子集中属于目标搜索结果的搜索结果。

在一些实施例中，该装置还包括：关联度确定单元，被配置成确定搜索结果集中的搜索结果的关联度，其中，关联度用于表示搜索结果与搜索信息的关联程度。

在一些实施例中，上述关联度确定单元进一步被配置成：确定该搜索结果与搜索信息的相似度；根据相似度，确定该搜索结果的关联度，其中，相似度与关联度成正比。

在一些实施例中，上述关联度确定单元进一步被配置成：针对补充搜索结果集中的补充搜索结果，响应于该补充搜索结果存在于搜索结果集中，按照预设的关联度提升算法，更新该补充搜索结果的关联度，以及将更新后的关联度作为该补充搜索结果的关联度；响应于该补充搜索结果不存在于搜索结果集中，确定该补充搜索结果的关联度。

在一些实施例中，上述关联度确定单元进一步被配置成：确定该补充搜索结果与搜索信息的相似度；根据相似度，确定该补充搜索结果的关联度，其中，相似度与关联度成正比。

在一些实施例中，上述补充搜索单元，进一步被配置成：按照关联度从大到小的顺序，从补充搜索结果集中选取目标数目个补充搜索结果作为目标搜索结果，得到目标搜索结果集。

在一些实施例中，该装置还包括划分单元，被配置成：按照预设

的目标数目个关联度区间，将搜索结果集划分为对应的目标数目个搜索结果子集；以及上述选取单元，进一步被配置成：从目标数目个搜索结果子集中分别选取搜索结果，得到搜索结果子集。

在一些实施例中，该装置还包括：准确度确定单元，被配置成确定目标数目个搜索结果子集中的搜索结果子集的准确度，其中，准确度用于表示搜索结果子集中目标搜索结果所占的比例。

在一些实施例中，上述划分单元进一步被配置成：按照上述关联度区间，将补充搜索结果集划分为对应的目标数目个补充搜索结果子集；以及上述确定单元进一步被配置成：从目标数目个补充搜索结果子集中分别选取补充搜索结果，得到补充搜索结果子集。

在一些实施例中，准确度确定单元进一步被配置成：确定目标数目个补充搜索结果集中的补充搜索结果子集的准确度，其中，准确度用于表示补充搜索结果子集中的目标搜索结果所占的比例。

在一些实施例中，上述确定单元进一步被配置成：针对目标数目个补充搜索结果子集中的补充搜索子集，确定该补充搜索结果子集所在的关联度区间对应的搜索结果集的准确度与预设的准确度阈值之间的差值的绝对值；根据绝对值和该补充搜索结果子集包含的补充搜索结果的数目，确定从该补充搜索结果子集中选取的补充搜索结果的数目，其中，绝对值和选取的补充搜索结果的数目成反比，包含的补充搜索结果的数目和选取的补充搜索结果的数目成正比。

在一些实施例中，搜索结果集中的搜索结果的数目和补充搜索结果子集中的补充搜索结果的数目相同。

在一些实施例中，收敛条件包括：目标关联度区间对应的搜索结果子集和补充搜索结果子集的准确度的差值小于预设的差值阈值。

在一些实施例中，收敛条件包括：搜索结果集和补充搜索结果集的差集所包含的搜索结果的数目小于预设的数目差阈值。

第三方面，本申请实施例提供了一种电子设备，该电子设备包括：一个或多个处理器；存储装置，用于存储一个或多个程序；当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

第四方面，本申请实施例提供了一种计算机可读介质，其上存储有计算机程序，该计算机程序被处理器执行时实现如第一方面中任一实现方式描述的方法。

本申请实施例提供的用于搜索信息的方法和装置，通过在直接得到搜索结果之后，可以采样部分搜索结果并从中选取属于目标搜索结果的搜索结果再次进行搜索，之后，根据再次搜索的搜索结果判断是否继续采样进行搜索，从而实现了增大搜索结果的数目，提升搜索结果的覆盖率。

附图说明

通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本申请的其它特征、目的和优点将会变得更明显：

图 1 是本申请的一个实施例可以应用于其中的示例性系统架构图；

图 2 是根据本申请的用于搜索信息的方法的一个实施例的流程图；

图 3 是根据本申请的用于搜索信息的方法的又一个实施例的流程图；

图 4 是根据本申请的用于搜索信息的方法的又一个实施例的流程图；

图 5 是根据本申请实施例的用于搜索信息的方法的一个应用场景的示意图；

图 6 是根据本申请的用于搜索信息的装置的一个实施例的结构示意图；

图 7 是适于用来实现本申请实施例的电子设备的计算机系统的结构示意图。

具体实施方式

下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发

明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

图 1 示出了可以应用本申请的用于搜索信息的方法或用于搜索信息的装置的实施例的示例性架构 100。

如图 1 所示，系统架构 100 可以包括终端设备 101、102、103，网络 104 和服务器 105。网络 104 用以在终端设备 101、102、103 和服务器 105 之间提供通信链路的介质。网络 104 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。

终端设备 101、102、103 通过网络 104 与服务器 105 交互，以接收或发送消息等。终端设备 101、102、103 上可以安装有各种客户端应用，例如网页浏览器应用、购物类应用、搜索类应用、即时通信工具等。

终端设备 101、102、103 可以是硬件，也可以是软件。当终端设备 101、102、103 为硬件时，可以是支持数据存储和数据交换的各种电子设备，包括但不限于智能手机、平板电脑、电子书阅读器、膝上型便携计算机和台式计算机等等。当终端设备 101、102、103 为软件时，可以安装在上述所列举的电子设备中。其可以实现成多个软件或软件模块（例如用来提供分布式服务的多个软件或软件模块），也可以实现成单个软件或软件模块。在此不做具体限定。

服务器 105 可以是提供各种服务的服务器，例如根据终端设备 101、102、103 发送的搜索信息在目标信息库中进行搜索，并向终端设备 101、102、103 返回搜索结果的搜索服务器。

需要说明的是，上述搜索信息和目标信息库也可以直接存储在服务器 105 的本地，服务器 105 可以直接提取搜索信息，并在本地所存储的目标信息库中进行搜索，此时，可以不存在终端设备 101、102、103 和网络 104。

需要说明的是，本申请实施例所提供的用于搜索信息的方法一般

由服务器 105 执行，相应地，用于搜索信息的装置一般设置于服务器 105 中。

还需要指出的是，终端设备 101、102、103 中也可以安装有信息搜索类应用，此时，用于处理图像的方法也可以由终端设备 101、102、103 执行，相应地，用于处理图像的装置也可以设置于终端设备 101、102、103 中。此时，示例性系统架构 100 可以不存在服务器 105 和网络 104。

需要说明的是，服务器可以是硬件，也可以是软件。当服务器为硬件时，可以实现成多个服务器组成的分布式服务器集群，也可以实现成单个服务器。当服务器为软件时，可以实现成多个软件或软件模块（例如用来提供分布式服务），也可以实现成单个软件或软件模块。在此不做具体限定。

应该理解，图 1 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。

继续参考图 2，其示出了根据本申请的用于搜索信息方法的一个实施例的流程 200。该用于搜索信息的方法包括以下步骤：

步骤 201，利用获取的搜索信息在目标信息库中搜索，得到搜索结果集。

在本实施例中，用于搜索信息的方法的执行主体（如图 1 中的服务器 105）可以先从本地或其他存储设备获取搜索信息，然后在目标信息库中进行搜索以得到搜索结果集。其中，目标信息库可以是预先由用户指定的信息库，也可以是根据搜索信息确定的与搜索信息匹配的搜索库，还可以是根据实际应用需求设置的信息库。搜索过程所使用的搜索算法可以是预先设置好的。

步骤 202，从搜索结果集中选取搜索结果子集，以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息。

在本实施例中，可以对搜索结果集先进行采样，选取至少一个搜索结果得到搜索结果子集。采样的比例可以是随机设置的，也可以是由用户或技术人员预先设置的，也可以是根据搜索结果集的大小确定

的（例如选取搜索结果集包含的搜索结果的数目的百分之十作为搜索结果子集的数目）。

在本实施例中，目标搜索结果可以指满足用户搜索需求的搜索结果。例如可以通过计算搜索结果子集中的各个搜索结果与搜索信息的相似度来验证各个搜索结果是否是目标搜索结果。

可选地，也可以先获取搜索结果子集中的搜索结果的标注信息，其中，标注信息用于表示搜索结果是否是目标搜索结果。然后，根据标注信息，确定搜索结果子集中属于目标搜索结果的搜索结果。

具体地，可以由人工对搜索结果子集中的各个搜索结果进行验证，并将验证结果输入为标注信息，然后发送至上述执行主体。

步骤 203，执行如下搜索步骤：

步骤 2031，将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集。

在本实施例中，可以先利用补充搜索信息在目标信息库中进行搜索，得到搜索结果。由于使用了补充搜索信息，因此，和上述搜索结果集相比，本次的搜索结果中可能包含有新的搜索结果。因此，可以将本次的搜索结果与上述搜索结果集的并集确定为补充搜索结果集。

步骤 2032，确定补充搜索结果集是否满足预设的收敛条件。

在本实施例中，收敛条件可以是由用户或技术人员根据实际的搜索需求具体设置的。例如，收敛条件可以是补充搜索结果集的数目大于预设的阈值。又例如，收敛条件也可以是补充搜索结果集和上述搜索结果集的差集所包含的搜索结果的数目小于预设的阈值。

步骤 204，响应于确定补充搜索结果集不满足预设的收敛条件，从补充搜索结果集中选取补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行上述搜索步骤。

在本实施例中，在补充搜索结果集不满足预设的收敛条件时，可以按照上述步骤 202 类似的方法，可以先从补充搜索结果集中进行采样，得到补充搜索结果子集，然后将补充搜索子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜

索结果集，继续执行上述步骤 203，进行迭代搜索。

可选地，响应于确定补充搜索结果集满足上述收敛条件，从补充搜索结果集中确定目标搜索结果集。例如，可以直接将补充搜索结果集作为目标搜索结果集。

本申请的上述实施例提供的用于搜索信息的方法通过对搜索信息在目标信息库中搜索得到的搜索结果集进行采样，并将采样的搜索结果子集中属于目标搜索结果的搜索结果作为补充搜索信息在目标信息库中再次进行搜索，从而能够补充新的搜索结果。之后还可以进一步判断搜索结果集加上新的搜索结果得到的补充搜索结果集是否满足收敛条件，若不满足，还可以按照上述采样选取补充搜索信息的方法在目标信息库中迭代搜索，从而不断增加新的搜索结果，直到得到的所有的搜索结果满足收敛条件，从而提升了搜索结果的覆盖率，避免了直接一次的搜索可能会出现、遗漏许多搜索结果的情况。

进一步参考图 3，其示出了用于搜索信息的方法的又一个实施例的流程 300。该用于搜索信息的方法的流程 300，包括以下步骤：

步骤 301，利用获取的搜索信息在目标信息库中搜索，得到搜索结果集。

本步骤的具体的执行过程可参考图 2 对应实施例中的步骤 201 的相关说明，在此不再赘述。

步骤 302，确定搜索结果集中的搜索结果的关联度。

在本实施例中，关联度可以用于表示搜索结果与搜索信息的关联程度。具体地，可以根据预设的关联度计算方法计算各个搜索结果分别与搜索信息的关联程度，也可以由人工来标注各个搜索结果与搜索信息的关联程度并返回至上述执行主体。

可选地，针对搜索结果集中的搜索结果，该搜索结果的关联度可以通过如下步骤确定：可以利用现有的一些相似度计算方法先确定该搜索结果与搜索信息的相似度，然后按照相似度与关联度成正比的关系，根据相似度确定该搜索结果的关联度。例如，可以直接将相似度确定为该搜索结果的关联度。

步骤 303, 从搜索结果集中选取搜索结果子集, 以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息。

本步骤的具体执行过程可参考图 2 对应实施例中的步骤 202 的相关说明, 在此不再赘述。

步骤 304, 执行如下搜索步骤:

步骤 3041, 将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集。

本步骤 3041 的具体执行过程可参考图 2 对应实施例中的步骤 2031 的相关说明, 在此不再赘述。

步骤 3042, 针对补充搜索结果集中的补充搜索结果, 通过如下方式确定该补充搜索结果的关联度:

步骤 30421, 响应于该补充搜索结果存在于搜索结果集中, 按照预设的关联度提升算法, 更新该补充搜索结果的关联度, 以及将更新后的关联度作为该补充搜索结果的关联度。

在本实施例中, 对于在搜索结果集中存在的补充搜索结果, 即搜索结果再次被搜索到, 可以提升该补充搜索结果的关联度。具体地, 可以由技术人员预先设置关联度提升算法用于提升关联度, 也可以预先设置关联度提升值, 在一个搜索结果再次被搜索到时, 将该搜索结果的关联度加上设置的关联度提升值, 并将得到的新的关联度作为该搜索结果的关联度。

步骤 30422, 响应于该补充搜索结果不存在于搜索结果集中, 确定该补充搜索结果的关联度。

在本实施例中, 可以利用现有的一些相似度计算方法先确定该补充搜索结果与搜索信息的相似度, 然后按照相似度与关联度成正比的关系, 根据相似度确定该补充搜索结果的关联度。例如, 可以直接将相似度确定为该补充搜索结果的关联度。

对于不存在与搜索结果集中的补充搜索结果, 即新增的搜索结果, 关联度的具体计算过程还可以参考上述步骤 302 的相关说明, 在此不再赘述。

步骤 3043, 确定补充搜索结果集是否满足预设的收敛条件。

本步骤 3043 的具体执行过程可参考图 2 对应实施例中的步骤 2032 的相关说明，在此不再赘述。

步骤 305，响应于确定补充搜索结果集不满足预设的收敛条件，从补充搜索结果集中选取补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行上述搜索步骤。

本步骤的具体执行过程可参考图 2 对应实施例中的步骤 204 的相关说明，在此不再赘述。

步骤 306，响应于确定补充搜索结果集满足上述收敛条件，按照关联度从大到小的顺序，从补充搜索结果集中选取目标数目个补充搜索结果作为目标搜索结果，得到目标搜索结果集合。

在本实施例中，目标数目可以是预先设置的，也可以是根据数目确定规则（如目标数目为补充搜索结果集所包含的补充搜索结果的数目的一半等）确定的。

从图 3 中可以看出，与图 2 对应的实施例相比，本实施例中的用于搜索信息方法的流程对于每次搜索之后的搜索结果，还可以确定每个搜索结果对应的关联度，并在迭代搜索过程中，可以动态更新再次被搜索到的搜索结果的关联度。由此，可以根据各个搜索结果及补充搜索结果的关联度选取目标搜索结果，进一步提升搜索结果的准确度。

进一步参考图 4，其示出了用于搜索信息的方法的又一个实施例的流程 400。该用于搜索信息的方法的流程 400，包括以下步骤：

步骤 401，利用获取的搜索信息在目标信息库中搜索，得到搜索结果集。

本步骤的具体的执行过程可参考图 2 对应实施例中的步骤 201 的相关说明，在此不再赘述。

步骤 402，确定搜索结果集中的搜索结果的关联度。

本步骤的具体的执行过程可参考图 3 对应实施例中的步骤 302 的相关说明，在此不再赘述。

步骤 403，按照预设的目标数目个关联度区间，将搜索结果集划

分为对应的目标数目个搜索结果子集。

在本实施例中，目标数目可以根据实际应用需求预先设置的。其中，各个关联度区间可以是均匀划分的，也可以是非均匀划分的。举例来说，对于关联度为 0 到 1（包括 0 和 1）之间的情况，可以均匀的划分为如下五个关联度区间：0 到 0.2（包含 0，不包含 0.2）、0.2 到 0.4（包含 0.2，不包含 0.4）、0.4 到 0.6（包含 0.4，不包含 0.6）、0.6 到 0.8（包含 0.6，不包含 0.8）到 0.8-1（包含 0.8 和 1），也可以非均匀的划分为如下五个关联度区间：0 到 0.5（包含 0，不包含 1）、0.5 到 0.7（包含 0.5，不包含 0.7）、0.7 到 0.8（包含 0.7，不包含 0.8）、0.8 到 0.9（包含 0.8，不包含 0.9）、0.9 到 1（包含 0.9 和 1）。

以上述均匀划分五个关联度区间为例，对应地，可以将关联度在 0 到 0.2 之间的搜索结果确定为一个搜索结果子集，将关联度在 0.2 到 0.4 之间的搜索结果确定为一个搜索结果子集，将关联度在 0.4 到 0.6 之间的搜索结果确定为一个搜索结果子集，将关联度在 0.6 到 0.8 之间的搜索结果确定为一个搜索结果子集，将关联度在 0.8 到 1 之间的搜索结果确定为一个搜索结果子集。

步骤 404，从目标数目个搜索结果子集中分别选取搜索结果，得到搜索结果子集，以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息。

在本实施例中，可以从每个搜索结果子集中采样部分搜索结果，并将采样的搜索结果组合得到搜索结果子集。其中，对于每个搜索结果子集所采样的搜索结果的数目可以任意指定，也可以根据采样数目确定规则来确定。例如，采样数目确定规则可以为每个搜索结果子集采样的搜索结果的数目为各搜索结果子集所包含的搜索结果的数目的十分之一。又例如，采样数目也可以根据关联度确定，例如对应的关联度越接近 0.5，对应采样的数目就越大。

可选地，还可以进一步确定目标数目个搜索结果子集中的搜索结果子集的准确度。其中，准确度可以用于表示搜索结果子集中目标搜索结果所占的比例。举例来说，对于包含有 30 条搜索结果的搜索结果子集，其中，有 20 条属于目标搜索结果，那么该搜索结果子集的准确

度即为三分之二。

步骤 405，执行如下搜索步骤：

步骤 4051，将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集。

本步骤 4051 的具体执行过程可参考图 2 对应实施例中的步骤 2031 的相关说明，在此不再赘述。

步骤 4052，针对补充搜索结果集中的补充搜索结果，通过如下方式确定该补充搜索结果的关联度：

步骤 40521，响应于该补充搜索结果存在于搜索结果集中，按照预设的关联度提升算法，更新该补充搜索结果的关联度，以及将更新后的关联度作为该补充搜索结果的关联度。

步骤 40522，响应于该补充搜索结果不存在于搜索结果集中，确定该补充搜索结果的关联度。

上述步骤 40521 和 40522 的具体执行过程可参考图 3 对应实施例中的步骤 30421 和 30422 的相关说明，在此不再赘述。

步骤 4053，根据上述关联度区间，将补充搜索结果集划分为对应的目标数目个补充搜索结果子集。

在本实施例中，在得到补充搜索结果集之后，由于可能包含有新的搜索结果，而且再次被搜索到的搜索结果的关联度有所提升，因此可以将补充搜索结果按照关联度再次进行划分，具体的划分过程可以参考上述步骤 403 中的相关说明。

步骤 4054，确定补充搜索结果集是否满足预设的收敛条件。

本步骤 4054 的具体执行过程可参考图 2 对应实施例中的步骤 2032 的相关说明，在此不再赘述。

步骤 406，响应于确定补充搜索结果集不满足预设的收敛条件，从目标数目个补充搜索结果集中分别选取补充搜索结果，得到补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行上述搜索步骤。

在本实施例中，可以从每个补充搜索结果子集中采样部分补充搜

索结果，并将采样的补充搜索结果组合得到补充搜索结果子集。其中，对于每个补充搜索结果子集所采样的补充搜索结果的数目可以任意指定，也可以根据采样数目确定规则来确定。例如，采样数目确定规则可以为每个补充搜索结果子集采样的补充搜索结果的数目为各补充搜索结果子集所包含的补充搜索结果的数目的十分之一。又例如，采样数目也可以根据关联度确定，例如对应的关联度越接近 0.5，对应采样的数目就越大。

可选地，针对目标数目个补充搜索结果子集中的补充搜索子集，可以先确定该补充搜索结果子集所在的关联度区间对应的搜索结果集的准确度与预设的准确度阈值之间的差值的绝对值；然后，根据确定的绝对值和该补充搜索结果子集包含的补充搜索结果的数目，确定从该补充搜索结果子集中选取的补充搜索结果的数目，其中，确定的绝对值和选取的补充搜索结果的数目成反比，包含的补充搜索结果的数目和选取的补充搜索结果的数目成正比。

其中，预设的准确度阈值可以是由技术人员预先设置的。实践中，可以预先估计准确度的最小值和最大值，并根据预先估计的准确度的最小值和最大值来设置准确度阈值。举例来说，若准确度在 0 到 1 之间，那么可以将准确度阈值设置为 0.5。此时，对应的准确度越接近 0.5 的补充搜索结果子集，对应采样的数目就越多；相反地，对应的准确度越接近 0 和 1 的补充搜索结果子集，对应采样的数目就越少。

举例来说，两个补充搜索结果子集包括的补充搜索结果的数目分别为 200 条和 400 条，那么包含 400 条补充搜索结果的补充搜索结果子集的采样数目可以大于包含 200 条补充搜索结果的补充搜索结果子集的采样数目。

可选地，上述搜索结果子集中的搜索结果的数目和补充搜索结果子集中的搜索结果的数目可以相同。即每次搜索完后，采样的总数目可以是固定的。

可选地，还可以进一步确定目标数目个补充搜索结果集中的补充搜索结果子集的准确度，其中，准确度可以用来表示补充搜索结果子集中的目标搜索结果所占的比例。

可选地，收敛条件可以为目标关联度区间对应的搜索结果子集和补充搜索结果子集的准确度的差值小于预设的差值阈值。

可选地，收敛条件也可以为搜索结果集和补充搜索结果集的差集所包含的搜索结果的数目小于预设的数目差阈值。

步骤 407，响应于确定补充搜索结果集满足上述收敛条件，按照关联度从大到小的顺序，从补充搜索结果集中选取目标数目个补充搜索结果作为目标搜索结果，得到目标搜索结果集合。

本步骤的具体的执行过程可参考图 3 对应实施例中的步骤 306 的相关说明，在此不再赘述。

继续参见图 5，图 5 是根据本实施例的用于搜索信息的方法的应用场景的一个示意图。在图 5 的应用场景中，可以先利用图像 501 在图像库 502 中进行搜索得到搜索结果集 503，其中，搜索结果集 503 包括 100 张图像。之后，可以分别计算搜索结果集 503 中的各个图像与图像 501 的相似度，并将相似度作为各个图像的关联度。之后，可以按照关联度，将搜索结果集划分为三个子集(如图中标号 504 所示)。第一个子集中的图像的关联度在 0 到 0.5 之间，共包含 30 张图像，第二个子集中的图像的关联度在 0.5 到 0.8 之间，共包含 50 张图像，第三个子集中的图像的关联度在 0.8 到 1 之间，共包含 20 张图像。

然后，可以在三个子集中进行采样，并将采样中包含的目标图像组合得到补充搜索图像 506。具体可以对第一个子集采样 10 张图像，对第二个子集采样 10 张图像，对第三个子集采样 3 张图像。具体如图中标号 505 所示，第一个子集采样的图像中包含有 5 张目标图像，第二个子集采样的图像中包含有 8 张目标图像，第三个子集采样的图像中包含有 3 张目标图像，因此，可以得到 16 张目标图像作为补充搜索图像 506。进一步地，还可以得到第一个子集采样的 10 张图像的准确度为二分之一，第二个子集采样的准确度为五分之四，第三个子集采样的准确度为一。

然后，可以将补充搜索图像 506 在图像库 502 中进行搜索，得到本次的搜索结果集 507，其中，本次的搜索结果集 507 中包含 120 条图像。进一步可以将上一次搜索得到搜索结果集 503 和本次搜索的补

充搜索结果集的并集作为补充搜索结果集 508，其中，补充搜索结果集包含 150 张图像。

可见，本次搜索过程中，有 30 张新搜索到的图像，还有 90 张图像再次被搜索到。对于 90 张再次被搜索到的图像的关联度可以进行提升，对于 30 张新搜索到的图像，可以将这些图像分别与图像 501 的相似度作为这些图像的关联度。从而将补充搜索结果集所包含的 150 张图像按照关联度重新分组，得到新的三个子集（如图中 509 所示）。

新得到的关联度在 0 到 0.5 之间的第一个子集包含 45 张图像，新得到的关联度在 0.5 到 0.8 之间的第二个子集包含 80 张图像，新得到的关联度在 0.8 到 1 之间的第三个子集包含 25 张图像。

之后，可以按照上一次计算得到的三个关联度区间对应的子集采样的准确度，对新得到的三个子集再次分别进行采样。上一次三个子集采样的准确度分别为二分之一、五分之四和一，那么本次采样的过程中，可以根据预先设置准确度阈值（如 0.5）来确定在各个子集中的采样数目。

对于新得到的第一个子集，由于该子集的准确度和准确度阈值的差值的绝对值为 0，因此，可以在该子集中多采样一些图像。对于新得到的第二个子集和新得到的第三个子集，由于这两个子集的准确度与准确度阈值的差值的绝对值较大，因此，在这两个子集中可以少采样一些图像。具体地，由于新得到的第二个子集的准确度与准确度阈值的差值的绝对值小于新得到的第三个子集的准确度与准确度阈值的差值的绝对值，因此，对于新得到的第二个子集的采样的数目可以大于新得到的第三个子集的采样的数目。

具体可以在新得到的第一个子集中采样 12 张图像，在新得到的第二个子集中采样 10 张图像，在新得到的第三个子集中采样 1 张图像。具体如图中标号 510 所示，在新得到的第一个子集和第二个子集所采样的图像中均包含有 8 张目标图像，在新得到的第三个子集所采样的图像中包含有 1 张目标图像。进一步还可以计算得到新得到的第一个子集的准确度为三分之二，新得到的第二个子集的准确度为五分之四，新得到的第三个子集的准确度为一。

之后,将新采样的共 17 张图像组成的补充搜索结果集 511 作为补充搜索图像,将补充搜索结果集 508 作为搜索结果集再次执行上述搜索过程。

具体地,先利用补充搜索结果集 511 在图像库 502 中进行搜索,将得到的搜索结果与补充搜索结果集 508 的并集作为新的补充搜索结果集 512。新的补充搜索结果集 512 包含 160 张图像。类似地,可以提升再次被搜索到的图像的关联度,对于新搜索到的图像,可以将新搜索到图像与图像 501 的相似度作为新搜索到的图像的关联度。之后,按照重新确定的关联度对新的补充搜索结果集 512 重新划分为三个子集。具体如图中标号 513 所示,新划分的第一个子集中包含 45 张图像,在新划分的第二个子集中包含 87 张图像,在新划分的第三个子集中包含 28 张图像。

由于上次计算得出三个关联度区间分别对应的准确度为三分之二、五分之四和一。可以再分别从新划分的三个子集中采样图像,并计算本次采样的准确度,其中,可以对于第一个子集多采样一些图像。具体新划分的三个子集分别采样后计算出的准确度分别为二分之一、十分之九和一。由此可以看出新划分的第二子子集和新划分的第三个子集的准确度变化较小,而且本次新的补充搜索结果集 512 和前一次的补充搜索结果集 508 相比,新搜索到的图像也较少,那么可以直接将本次搜索新划分的第二个子集和新划分的第三个子集中的共 115 张图像作为目标图像集 514。

从图 4 中可以看出,与图 2 和图 3 对应的实施例相比,本实施例中的用于搜索信息的方法的流程突出了在每次搜索完并确定了各搜索结果关联度之后,还可以按照预设的关联度区间,将搜索结果划分为对应的多个子集,并在各个子集中分别采样进行补充搜索,从而使得补充搜索过程可以覆盖各个关联度区间,进一步提升搜索结果的覆盖率。在此基础上,在对各个子集分别采样后,还可以进一步根据采样结果确定各个子集对应的准确度,并根据准确度确定下一次该关联度区间的采样数目,从而实现有针对性地进行补充搜索,进一步提升搜索结果的准确度和搜索效率。

进一步参考图 6，作为对上述各图所示方法的实现，本申请提供了用于搜索信息的装置的一个实施例，该装置实施例与图 2 所示的方法实施例相对应，该装置具体可以应用于各种电子设备中。

如图 6 所示，本实施例提供的用于搜索信息的装置 600 包括搜索单元 601、选取单元 602、补充搜索单元 603 和确定单元 604。其中，搜索单元 601，被配置成利用获取的搜索信息在目标信息库中搜索，得到搜索结果集；选取单元 602，被配置成从搜索结果集中选取搜索结果子集，以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息；补充搜索单元 603，被配置成执行如下搜索步骤：将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集；确定补充搜索结果集是否满足预设的收敛条件；确定单元 604，被配置成响应于确定补充搜索结果集不满足预设的收敛条件，从补充搜索结果集中选取补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行上述搜索步骤。

在本实施例中，用于搜索信息的装置 600 中：搜索单元 601、选取单元 602、补充搜索单元 603 和确定单元 604 的具体处理及其所带来的技术效果可分别参考图 2 对应实施例中的步骤 201、步骤 202、步骤 203 和步骤 204 的相关说明，在此不再赘述。

在本实施例的一些可选的实现方式中，上述补充搜索单元 603 进一步被配置成：响应于确定补充搜索结果集满足上述收敛条件，从补充搜索结果集中确定目标搜索结果集。

在本实施例的一些可选的实现方式中，上述选取单元 602 进一步被配置成：获取搜索结果子集中的搜索结果的标注信息，其中，标注信息用于表示搜索结果是否是目标搜索结果；根据标注信息，确定搜索结果子集中属于目标搜索结果的搜索结果。

在本实施例的一些可选的实现方式中，该用于搜索信息的装置 600 还包括：关联度确定单元（图中未示出）被配置成确定搜索结果集中的搜索结果的关联度，其中，关联度用于表示搜索结果与搜索信

息的关联程度。

在本实施例的一些可选的实现方式中，上述关联度确定单元进一步被配置成：确定该搜索结果与搜索信息的相似度；根据相似度，确定该搜索结果的关联度，其中，相似度与关联度成正比。

在本实施例的一些可选的实现方式中，上述关联度确定单元进一步被配置成：针对补充搜索结果集中的补充搜索结果，响应于该补充搜索结果存在于搜索结果集中，按照预设的关联度提升算法，更新该补充搜索结果的关联度，以及将更新后的关联度作为该补充搜索结果的关联度；响应于该补充搜索结果不存在于搜索结果集中，确定该补充搜索结果的关联度。

在本实施例的一些可选的实现方式中，上述关联度确定单元进一步被配置成：确定该补充搜索结果与搜索信息的相似度；根据相似度，确定该补充搜索结果的关联度，其中，相似度与关联度成正比。

在本实施例的一些可选的实现方式中，上述补充搜索单元 603 进一步被配置成：按照关联度从大到小的顺序，从补充搜索结果集中选取目标数目个补充搜索结果作为目标搜索结果，得到目标搜索结果集。

在本实施例的一些可选的实现方式中，该用于搜索信息的装置 600 还包括划分单元（图中未示出）被配置成：按照预设的目标数目个关联度区间，将搜索结果集划分为对应的目标数目个搜索结果子集；以及上述选取单元，进一步被配置成：从目标数目个搜索结果子集中分别选取搜索结果，得到搜索结果子集。

在本实施例的一些可选的实现方式中，该用于搜索信息的装置 600 还包括：准确度确定单元（图中未示出）被配置成确定目标数目个搜索结果子集中的搜索结果子集的准确度，其中，准确度用于表示搜索结果子集中目标搜索结果所占的比例。

在本实施例的一些可选的实现方式中，上述划分单元进一步被配置成：按照上述关联度区间，将补充搜索结果集划分为对应的目标数目个补充搜索结果子集；以及上述确定单元进一步被配置成：从目标数目个补充搜索结果子集中分别选取补充搜索结果，得到补充搜索结果子集。

在本实施例的一些可选的实现方式中，准确度确定单元进一步被配置成：确定目标数目个补充搜索结果集中的补充搜索结果子集的准确度，其中，准确度用于表示补充搜索结果子集中的目标搜索结果所占的比例。

在本实施例的一些可选的实现方式中，上述确定单元进一步被配置成：针对目标数目个补充搜索结果子集中的补充搜索子集，确定该补充搜索结果子集所在的关联度区间对应的搜索结果集的准确度与预设的准确度阈值之间的差值的绝对值；根据绝对值和该补充搜索结果子集包含的补充搜索结果的数目，确定从该补充搜索结果子集中选取的补充搜索结果的数目，其中，绝对值和选取的补充搜索结果的数目成反比，包含的补充搜索结果的数目和选取的补充搜索结果的数目成正比。

在本实施例的一些可选的实现方式中，搜索结果集中的搜索结果的数目和补充搜索结果子集中的补充搜索结果的数目相同。

在本实施例的一些可选的实现方式中，收敛条件包括：目标关联度区间对应的搜索结果子集和补充搜索结果子集的准确度的差值小于预设的差值阈值。

在本实施例的一些可选的实现方式中，收敛条件包括：搜索结果集和补充搜索结果集的差集所包含的搜索结果的数目小于预设的数目差阈值。

本申请的上述实施例提供的装置，通过搜索单元直接得到搜索结果之后，可以通过选取单元采样部分搜索结果并从中选取属于目标搜索结果的搜索结果由补充搜索单元再次进行搜索，之后，确定单元根据再次搜索的搜索结果判断是否继续采样进行搜索，从而实现了增大搜索结果的数目，提升搜索结果的覆盖率。

下面参考图 7，其示出了适于用来实现本申请实施例的电子设备的计算机系统 700 的结构示意图。图 7 示出的电子设备仅仅是一个示例，不应对本申请实施例的功能和使用范围带来任何限制。

如图 7 所示，计算机系统 700 包括中央处理单元（CPU）701，其

可以根据存储在只读存储器 (ROM) 702 中的程序或者从存储部分 708 加载到随机访问存储器 (RAM) 703 中的程序而执行各种适当的动作和处理。在 RAM 703 中, 还存储有系统 700 操作所需的各种程序和数据。CPU 701、ROM 702 以及 RAM 703 通过总线 704 彼此相连。输入/输出 (I/O) 接口 705 也连接至总线 704。

以下部件连接至 I/O 接口 705: 包括键盘、鼠标等的输入部分 706; 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 707; 包括硬盘等的存储部分 708; 以及包括诸如 LAN 卡、调制解调器等网络接口卡的通信部分 709。通信部分 709 经由诸如因特网的网络执行通信处理。驱动器 710 也根据需要连接至 I/O 接口 705。可拆卸介质 711, 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 710 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 708。

特别地, 根据本公开的实施例, 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 709 从网络上被下载和安装, 和/或从可拆卸介质 711 被安装。在该计算机程序被中央处理单元 (CPU) 701 执行时, 执行本申请的方法中限定的上述功能。

需要说明的是, 本申请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。在本申请中, 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。而在本申请中，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。

附图中的流程图和框图，图示了按照本申请各种实施例的系统、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，该模块、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。也要注意的，框图和/或流程图中的每个方框、以及框图和/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。

描述于本申请实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器，包括搜索单元、选取单元、补充搜索单元和确定单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，搜索单元还可以被描述为“利用获取的搜索信息在目标信息库中搜索，得到搜索结果集的单元”。

作为另一方面，本申请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电子设备中所包含的；也可以是单独存在，而未装配入该电子设备中。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该电子设备执行时，使

得该电子设备：利用获取的搜索信息在目标信息库中搜索，得到搜索结果集；从搜索结果集中选取搜索结果子集，以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息；执行如下搜索步骤：将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集；确定补充搜索结果集是否满足预设的收敛条件；响应于确定补充搜索结果集不满足预设的收敛条件，从补充搜索结果集中选取补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行上述搜索步骤。

以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

权 利 要 求 书

1、一种用于搜索信息的方法，包括：

利用获取的搜索信息在目标信息库中搜索，得到搜索结果集；

从搜索结果集中选取搜索结果子集，以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息；

执行如下搜索步骤：将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集；确定补充搜索结果集是否满足预设的收敛条件；

响应于确定补充搜索结果集不满足预设的收敛条件，从补充搜索结果集中选取补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行所述搜索步骤。

2、根据权利要求 1 所述的方法，其中，所述搜索步骤还包括：

响应于确定补充搜索结果集满足所述收敛条件，从补充搜索结果集中确定目标搜索结果集。

3、根据权利要求 1 所述的方法，其中，所述将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息，包括：

获取搜索结果子集中的搜索结果的标注信息，其中，标注信息用于表示搜索结果是否是目标搜索结果；

根据标注信息，确定搜索结果子集中属于目标搜索结果的搜索结果。

4、根据权利要求 2 所述的方法，其中，在所述利用搜索信息在目标信息库中搜索，得到搜索结果集之后，还包括：

确定搜索结果集中的搜索结果的关联度，其中，关联度用于表示搜索结果与搜索信息的关联程度。

5、根据权利要求 4 所述的方法，其中，针对搜索结果集中的搜索结果，该搜索结果的关联度通过如下步骤确定：

确定该搜索结果与搜索信息的相似度；

根据相似度，确定该搜索结果的关联度，其中，相似度与关联度成正比。

6、根据权利要求 4 所述的方法，其中，在所述将补充搜索信息在目标信息库中的搜索结果与搜索结果集的并集确定为补充搜索结果集之后，还包括：

针对补充搜索结果集中的补充搜索结果，响应于该补充搜索结果存在于搜索结果集中，按照预设的关联度提升算法，更新该补充搜索结果的关联度，以及将更新后的关联度作为该补充搜索结果的关联度；响应于该补充搜索结果不存在于搜索结果集中，确定该补充搜索结果的关联度。

7、根据权利要求 6 所述的方法，其中，所述确定该补充搜索结果的关联度包括：

确定该补充搜索结果与搜索信息的相似度；

根据相似度，确定该补充搜索结果的关联度，其中，相似度与关联度成正比。

8、根据权利要求 6 所述的方法，其中，所述从补充搜索结果集中确定目标搜索结果集，包括：

按照关联度从大到小的顺序，从补充搜索结果集中选取目标数目个补充搜索结果作为目标搜索结果，得到目标搜索结果集。

9、根据权利要求 4 所述的方法，其中，所述方法还包括：

按照预设的目标数目个关联度区间，将搜索结果集划分为对应的目标数目个搜索结果子集；以及

所述从搜索结果集中选取搜索结果子集，包括：

从目标数目个搜索结果子集中分别选取搜索结果，得到搜索结果子集。

10、根据权利要求 9 所述的方法，其中，所述方法还包括：

确定目标数目个搜索结果子集中的搜索结果子集的准确度，其中，准确度用于表示搜索结果子集中目标搜索结果所占的比例。

11、根据权利要求 10 所述的方法，其中，在所述将补充搜索信息在目标信息库中的搜索结果与搜索结果集的并集确定为补充搜索结果集之后，还包括：

按照所述关联度区间，将补充搜索结果集划分为对应的目标数目个补充搜索结果子集；以及

所述从补充搜索结果集中选取补充搜索结果子集，包括：

从目标数目个补充搜索结果子集中分别选取补充搜索结果，得到补充搜索结果子集。

12、根据权利要求 11 所述的方法，其中，所述方法还包括：

确定目标数目个补充搜索结果集中的补充搜索结果子集的准确度，其中，准确度用于表示补充搜索结果子集中的目标搜索结果所占的比例。

13、根据权利要求 12 所述的方法，其中，所述从目标数目个补充搜索结果子集中分别选取补充搜索结果，包括：

针对目标数目个补充搜索结果子集中的补充搜索子集，确定该补充搜索结果子集所在的关联度区间对应的搜索结果集的准确度与预设的准确度阈值之间的差值的绝对值；根据所述绝对值和该补充搜索结果子集包含的补充搜索结果的数目，确定从该补充搜索结果子集中选取的补充搜索结果的数目，其中，所述绝对值和选取的补充搜索结果的数目成反比，包含的补充搜索结果的数目和选取的补充搜索结果的数目成正比。

14、根据权利要求 1-13 之一所述的方法，其中，搜索结果集中的搜索结果的数目和补充搜索结果子集中的补充搜索结果的数目相同。

15、根据权利要求 13 所述的方法，其中，所述收敛条件包括：
目标关联度区间对应的搜索结果子集和补充搜索结果子集的准确度的差值小于预设的差值阈值。

16、根据权利要求 1 所述的方法，其中，所述收敛条件包括：
搜索结果集和补充搜索结果集的差集所包含的搜索结果的数目小于预设的数目差阈值。

17、一种用于搜索信息的装置，包括：
搜索单元，被配置成利用获取的搜索信息在目标信息库中搜索，得到搜索结果集；
选取单元，被配置成从搜索结果集中选取搜索结果子集，以及将搜索结果子集中属于目标搜索结果的搜索结果确定为补充搜索信息；
补充搜索单元，被配置成执行如下搜索步骤：将补充搜索信息在目标信息库中搜索的搜索结果与搜索结果集的并集确定为补充搜索结果集；确定补充搜索结果集是否满足预设的收敛条件；
确定单元，被配置成响应于确定补充搜索结果集不满足预设的收敛条件，从补充搜索结果集中选取补充搜索结果子集，以及将补充搜索结果子集中属于目标搜索结果的补充搜索结果确定为补充搜索信息，将补充搜索结果集确定为搜索结果集，继续执行所述搜索步骤。

18、根据权利要求 17 所述的装置，其中，所述补充搜索单元进一步被配置成：

响应于确定补充搜索结果集满足所述收敛条件，从补充搜索结果集中确定目标搜索结果集。

19、根据权利要求 17 所述的装置，其中，所述选取单元进一步被配置成：

获取搜索结果子集中的搜索结果的标注信息，其中，标注信息用于表示搜索结果是否是目标搜索结果；

根据标注信息，确定搜索结果子集中属于目标搜索结果的搜索结果。

20、根据权利要求 18 所述的装置，其中，所述装置还包括：

关联度确定单元，被配置成确定搜索结果集中的搜索结果的关联度，其中，关联度用于表示搜索结果与搜索信息的关联程度。

21、根据权利要求 20 所述的装置，其中，所述关联度确定单元进一步被配置成：

确定该搜索结果与搜索信息的相似度；

根据相似度，确定该搜索结果的关联度，其中，相似度与关联度成正比。

22、根据权利要求 20 所述的装置，其中，所述关联度确定单元进一步被配置成：

针对补充搜索结果集中的补充搜索结果，响应于该补充搜索结果存在于搜索结果集中，按照预设的关联度提升算法，更新该补充搜索结果的关联度，以及将更新后的关联度作为该补充搜索结果的关联度；响应于该补充搜索结果不存在于搜索结果集中，确定该补充搜索结果的关联度。

23、根据权利要求 22 所述的装置，其中，所述关联度确定单元进一步被配置成：

确定该补充搜索结果与搜索信息的相似度；

根据相似度，确定该补充搜索结果的关联度，其中，相似度与关联度成正比。

24、根据权利要求 22 所述的装置，其中，所述补充搜索单元，进一步被配置成：

按照关联度从大到小的顺序，从补充搜索结果集中选取目标数目个补充搜索结果作为目标搜索结果，得到目标搜索结果集。

25、根据权利要求 20 所述的装置，其中，所述装置还包括划分单元，被配置成：

按照预设的目标数目个关联度区间，将搜索结果集划分为对应的目标数目个搜索结果子集；以及

所述选取单元，进一步被配置成：

从目标数目个搜索结果子集中分别选取搜索结果，得到搜索结果子集。

26、根据权利要求 25 所述的装置，其中，所述装置还包括：

准确度确定单元，被配置成确定目标数目个搜索结果子集中的搜索结果子集的准确度，其中，准确度用于表示搜索结果子集中目标搜索结果所占的比例。

27、根据权利要求 26 所述的装置，其中，所述划分单元进一步被配置成：

按照所述关联度区间，将补充搜索结果集划分为对应的目标数目个补充搜索结果子集；以及

所述确定单元进一步被配置成：

从目标数目个补充搜索结果子集中分别选取补充搜索结果，得到补充搜索结果子集。

28、根据权利要求 27 所述的装置，其中，所述准确度确定单元进一步被配置成：

确定目标数目个补充搜索结果集中的补充搜索结果子集的准确

度，其中，准确度用于表示补充搜索结果子集中的目标搜索结果所占的比例。

29、根据权利要求 28 所述的装置，其中，所述确定单元进一步被配置成：

针对目标数目个补充搜索结果子集中的补充搜索子集，确定该补充搜索结果子集所在的关联度区间对应的搜索结果集的准确度与预设的准确度阈值之间的差值的绝对值；根据所述绝对值和该补充搜索结果子集包含的补充搜索结果的数目，确定从该补充搜索结果子集中选取的补充搜索结果的数目，其中，所述绝对值和选取的补充搜索结果的数目成反比，包含的补充搜索结果的数目和选取的补充搜索结果的数目成正比。

30、根据权利要求 17-29 之一所述的装置，其中，搜索结果集中的搜索结果的数目和补充搜索结果子集中的补充搜索结果的数目相同。

31、根据权利要求 29 所述的装置，其中，所述收敛条件包括：

目标关联度区间对应的搜索结果子集和补充搜索结果子集的准确度的差值小于预设的差值阈值。

32、根据权利要求 17 所述的装置，其中，所述收敛条件包括：

搜索结果集和补充搜索结果集的差集所包含的搜索结果的数目小于预设的数目差阈值。

33、一种电子设备，包括：

一个或多个处理器；

存储装置，其上存储有一个或多个程序；

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 1-16 中任一所述的方法。

34、一种计算机可读介质，其上存储有计算机程序，其中，该程序被处理器执行时实现如权利要求 1-16 中任一所述的方法。

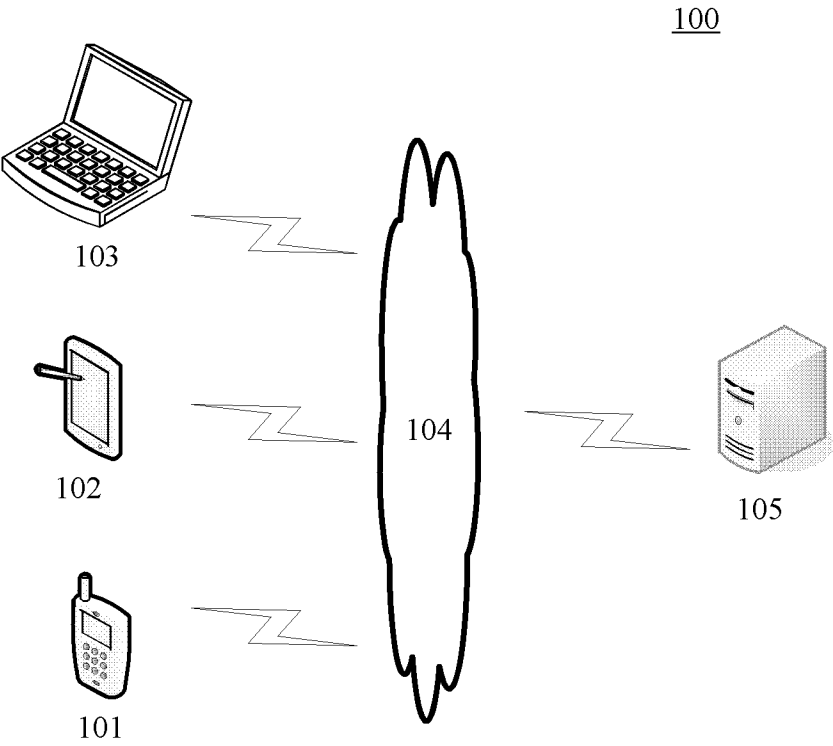


图 1

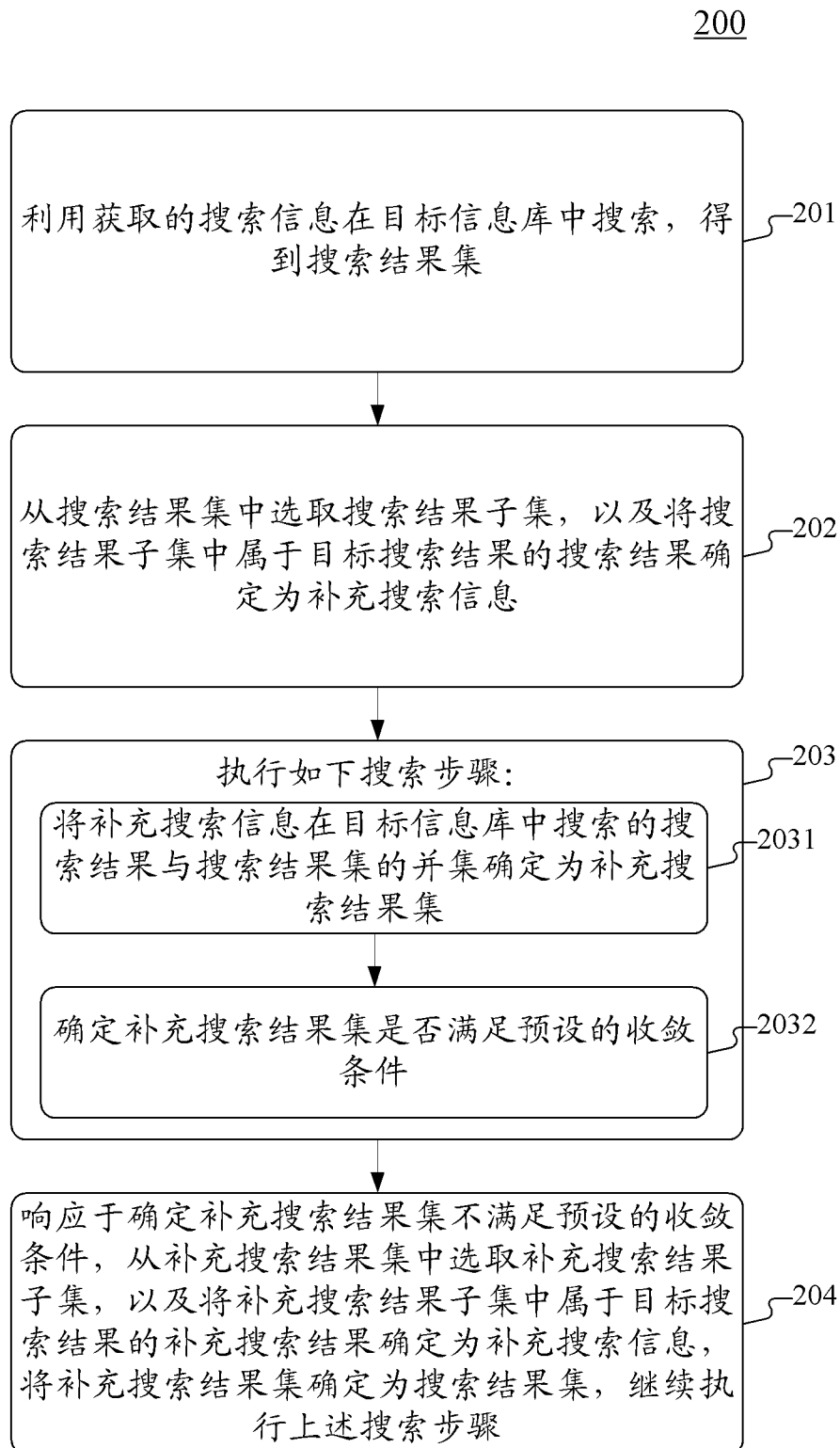


图 2

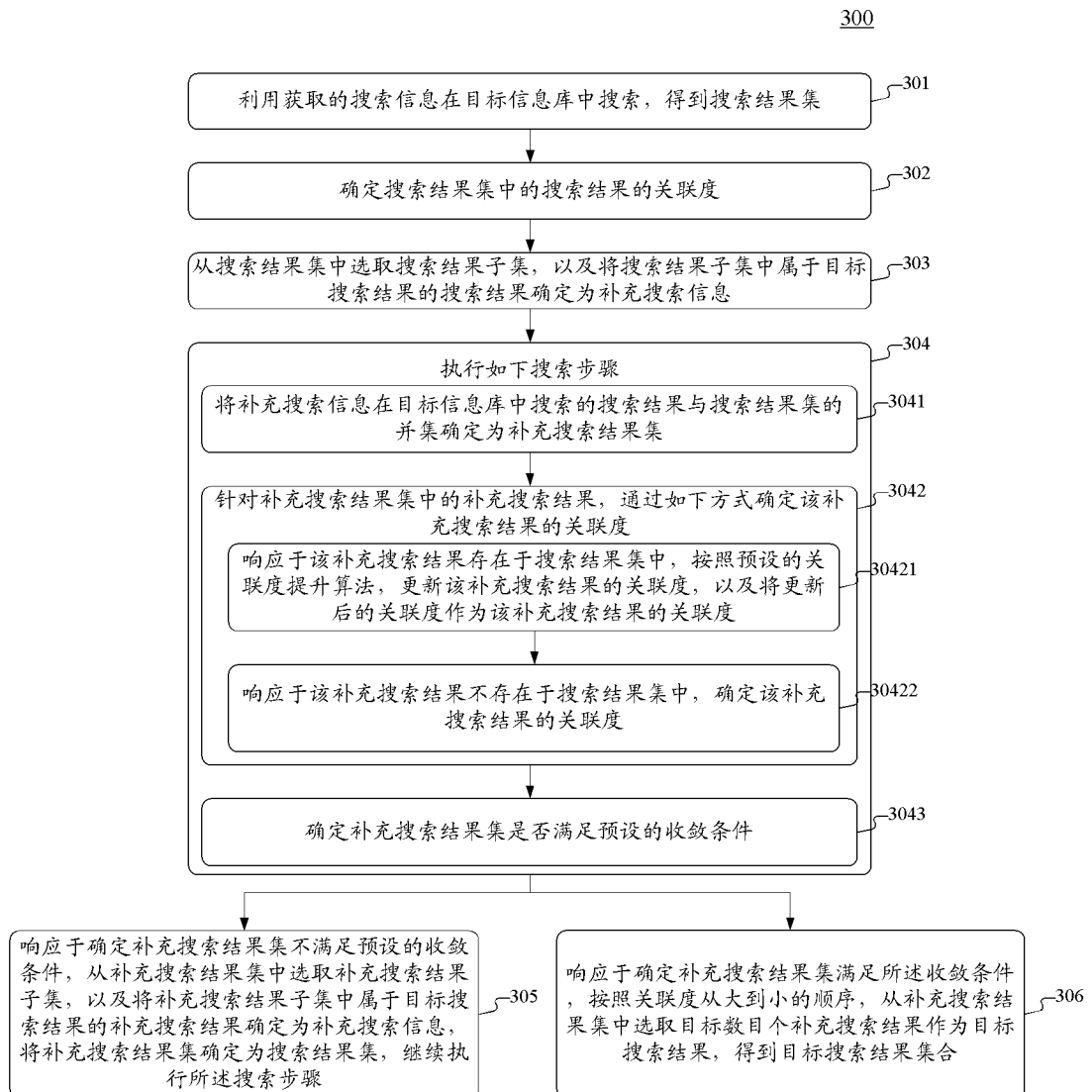


图 3

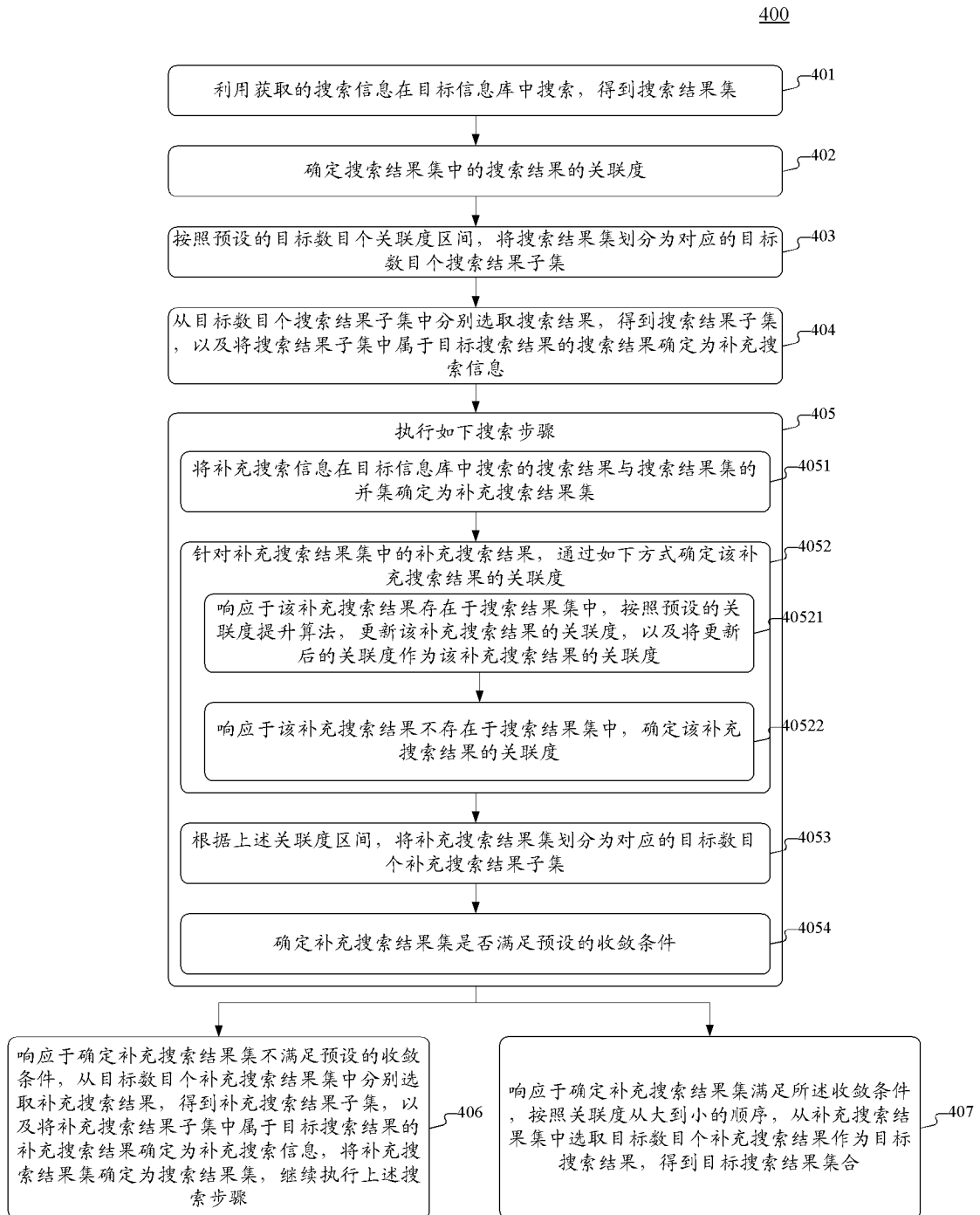


图 4

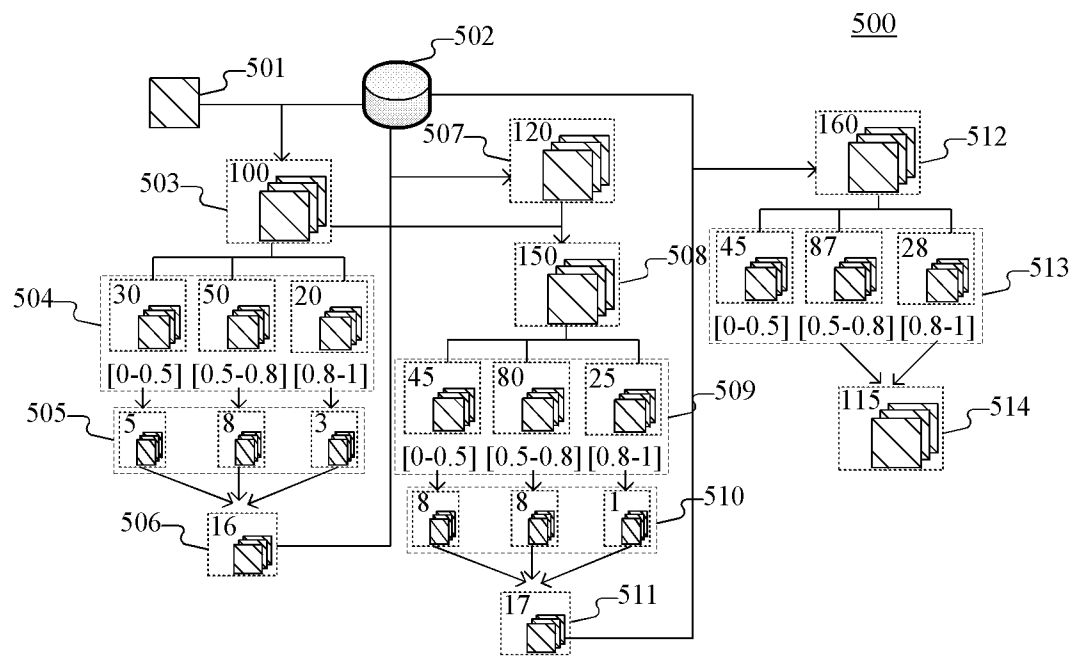


图 5

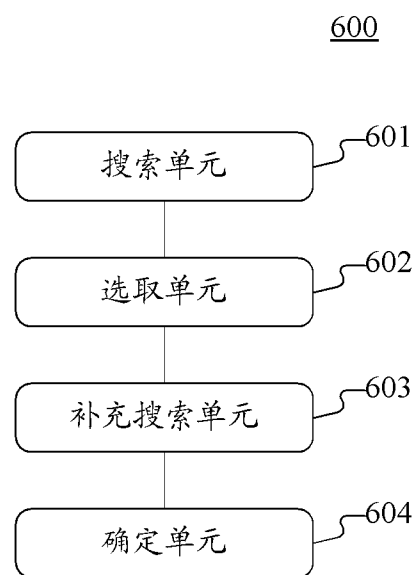


图 6

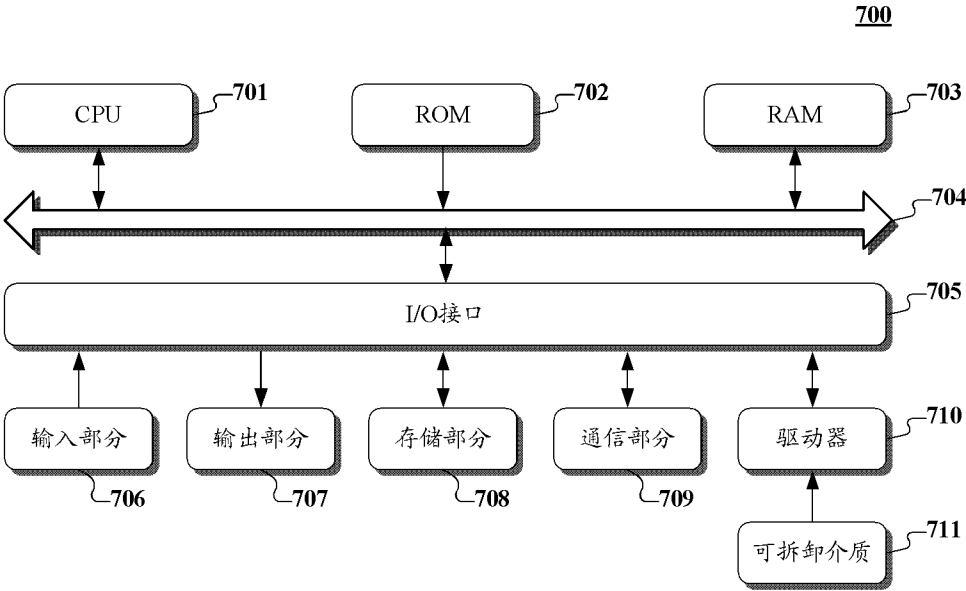


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/116342

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/242(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)

CNKI, CNPAT, WPI, EPODOC, IEEE: 搜索, 检索, 并集, 子集, 补充, 增大, 结果, 合并, search, set, subset, result, increase

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106156109 A (ALIBABA GROUP HOLDING LIMITED) 23 November 2016 (2016-11-23) description, paragraphs [0050]-[0091]	1-34
A	CN 102999556 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 27 March 2013 (2013-03-27) entire document	1-34
A	CN 1609859 A (SUN, BIN) 27 April 2005 (2005-04-27) entire document	1-34
A	CN 103136213 A (ALIBABA GROUP HOLDING LIMITED) 05 June 2013 (2013-06-05) entire document	1-34



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

23 May 2019

Date of mailing of the international search report

12 June 2019

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/CN2018/116342

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106156109	A	23 November 2016	None			
CN	102999556	A	27 March 2013	None			
CN	1609859	A	27 April 2005	US	2006117002	A1	01 June 2006
CN	103136213	A	05 June 2013	HK	1181487	A1	09 March 2018

国际检索报告

国际申请号

PCT/CN2018/116342

A. 主题的分类

G06F 16/242(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))

CNKI, CNPAT, WPI, EPDOC, IEEE: 搜索, 检索, 并集, 子集, 补充, 增大, 结果, 合并, search, set, subset, result, increase

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
A	CN 106156109 A (阿里巴巴集团控股有限公司) 2016年 11月 23日 (2016 - 11 - 23) 说明书50-91段	1-34
A	CN 102999556 A (百度在线网络技术北京有限公司) 2013年 3月 27日 (2013 - 03 - 27) 全文	1-34
A	CN 1609859 A (孙斌) 2005年 4月 27日 (2005 - 04 - 27) 全文	1-34
A	CN 103136213 A (阿里巴巴集团控股有限公司) 2013年 6月 5日 (2013 - 06 - 05) 全文	1-34

☐ 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

2019年 5月 23日

国际检索报告邮寄日期

2019年 6月 12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

赵天奇

电话号码 86-(10)-53961320

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/116342

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	106156109	A	2016年 11月 23日	无	
CN	102999556	A	2013年 3月 27日	无	
CN	1609859	A	2005年 4月 27日	US 2006117002 A1	2006年 6月 1日
CN	103136213	A	2013年 6月 5日	HK 1181487 A1	2018年 3月 9日