

(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
2019 年 3 月 14 日 (14.03.2019)



(10) 国际公布号
WO 2019/047437 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2017/120062

(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:
201710797899.8 2017 年 9 月 6 日 (06.09.2017) 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: 毕成龙 (**BI, Chenglong**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080

(CN)。潘文彬(**PAN, Wenbin**); 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:** MULTI-DOCUMENT INTERSECTION ACQUISITION METHOD AND DOCUMENT SERVER

(54) 发明名称: 多文档交集的获取方法及文档服务器

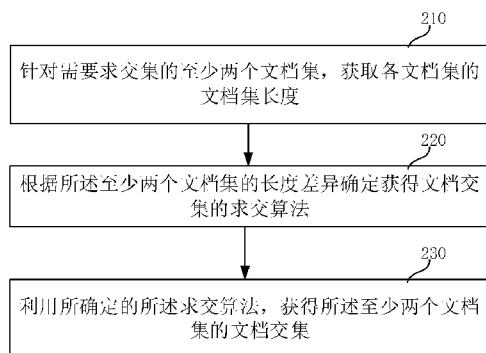


图 2

210 For at least two document sets requiring an intersection, acquire the document set length of each document set
220 Determine an intersection algorithm for obtaining a document intersection according to the length difference of the at least two document sets
230 Use the intersection algorithm to obtain the document intersection of the at least two document sets

(57) **Abstract:** A multi-document intersection acquisition method and device, and a readable storage medium. For at least two document sets requiring an intersection in a search process, the document set length of each document set is acquired (210), and the lengths of the at least two document sets are compared. An intersection algorithm for obtaining a document intersection is determined according to the length difference of the at least two document sets (220). When the document set lengths of the at least two document sets satisfy a preset condition, the elements in the shortest document set can be taken as query elements to sequentially traverse the remaining document sets.

(57) **摘要:** 一种多文档交集获取方法及、装置及可读存储介质。针对搜索过程中需要交集的至少两个文档集，获取各文档集的文档集长度(210)，并对所述至少两个文档集的长度进行比较。根据所述至少两个文档集的长度差异确定获得文档交集的求交算法(220)。在所述至少两个文档集的文档集长度符合预设条件时，可以以长度最短文档集中的元素作为查询元素依次遍历剩余的文档集。

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))。

多文档交集的获取方法及文档服务器

相关申请的交叉引用

5 [001] 本专利申请要求于 2017 年 9 月 6 日提交的、申请号为 201710797899.8、发明名称为“多文档交集获取方法、装置、设备及可读存储介质”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

[002] 本发明实施例涉及搜索引擎技术领域，尤其涉及多文档交集的获取。

背景技术

10 [003] 搜索引擎可能收集了万维网上几千万到几十亿个网页，并对网页中的每一个词进行索引以建立索引数据库。当用户查找某个关键词的时候，页面内容中包含了该关键词的所有网页都将作为搜索结果被搜出来。

[004] 广泛应用的开源搜索引擎、例如 Lucene 采用了一种线性的获取多文档集的交集的方法，即：将每个文档集排序后，从第一个文档集的第一个元素开始遍历其他文档集。这时可能
15 能在当前文档集中找到该元素，则继续遍历下一个文档集。如果没找到，那么以当前文档集中的下一个元素作为查询元素重新遍历其他文档集。如此往复，直到找到一个同时所有文档集中存在的元素、即文档为止。以这种方式，不断重复，直到其中一个文档集遍历结束，则多文档集的求交过程结束。

[005] 其中，在遍历每个文档集时，如果当前比较的元素不符合要求、即不是查询元素，则
20 需要继续比较本文档集的下一个元素。但在不同文档集的长度差距大于某一阈值时，这种查找往往是没有必要的，因为对于多文档集求交而言，只要一个文档中不存在某元素，则该元素就一定不会是交集内的元素。因此，上述文档交集的获取效率可能较低下。

发明内容

[006] 本发明提供一种多文档交集获取方法、装置及可读存储介质，以即使在不同文档集的长度差距大于某一阈值时仍能够以较高效率获取文档交集。
25

[007] 根据本发明的第一方面，提供了一种多文档交集获取方法，包括：针对搜索中需要求

交集的至少两个文档集，获取各所述文档集的文档集长度；根据所述至少两个文档集的长度差异确定获得文档交集的求交算法；以及利用所确定的所述求交算法，获得所述至少两个文档集的文档交集。

[008] 根据本发明的第二方面，提供了一种文档服务器，包括：处理器；以及非临时性计算机可读存储介质，在所述非临时性计算机可读存储介质上存储有能够被所述处理器执行的机器可执行指令。其中，所述处理器被所述机器可执行指令促使：针对搜索中需要求交集的至少两个文档集，获取各所述文档集的文档集长度；根据所述至少两个文档集的长度差异确定获得文档交集的求交算法；以及，利用所确定的所述求交算法，获得所述至少两个文档集的文档交集。

[009] 根据本发明的第三方面，提供了一种非临时性机器可读存储介质，其上存储有能够被处理器执行的机器可执行指令。当所述非临时性机器可读存储介质中的所述机器可执行指令由文档服务器中的处理器执行时，所述文档服务器能够执行上述多文档交集获取方法。

[010] 本发明实施例提供了一种多文档交集获取方法、装置及可读存储介质，针对搜索过程中需要求交集的至少两个文档集，通过在所述至少两个文档集的文档集长度符合预设条件时，以长度最短文档集中的元素作为查询元素依次遍历剩余的文档集，可以有效提高文档求交集效率，加快了搜索引擎针对用户的响应时间。

附图说明

[011] 为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例的描述中所需要使用到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发明的一些实施例，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。

[012] 图 1 是本发明一实施例提供的一种搜索引擎的结构示意图。

[013] 图 2 是本发明一实施例提供的一种多文档交集获取方法的流程图。

[014] 图 3 是本发明另一实施例提供的一种多文档交集获取方法的流程图。

[015] 图 3A 是本发明一实施例提供的一种初级语法树的结构图。

[016] 图 3B 是本发明一实施例提供的一种终级语法树的结构图。

[017] 图 4 是本发明一实施例提供的一种多文档交集获取装置的结构示意图。

[018] 图 5 是本发明另一实施例提供的一种多文档交集获取装置的结构示意图。

具体实施方式

[019] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[020] 本发明实施例提供的一种多文档交集获取方法可应用于搜索引擎技术，其目的是根据用户在搜索引擎界面输入的查询内容，对查询内容进行分词，通过对各个分词进行对应文档的匹配来生成各个分词对应的文档集，通过对所有文档集进行求交集来获得文档交集，并将文档交集返回给用户。

[021] 搜索引擎技术是一种互联网通信技术。在互联网中，服务器侧提供内容，并针对该内容构建索引。当用户使用客户端发送搜索请求至服务器时，服务器可根据搜索请求中的关键词去索引中查找内容，然后将查找到的内容返回给该客户端进行展示。

[022] 参照图 1，搜索引擎通常可包括 WEB（World Wide Web，万维网）服务器 110、索引服务器 120 和文档服务器 130。其中，文档服务器 130 可储存有文档的信息。

[023] 当用户通过搜索引擎浏览器 140 输入搜索词后，WEB 服务器 110 会接收到该搜索词并将该搜索词发送至索引服务器 120。接着，索引服务器 120 会将搜索词进行语法分词处理，针对每个分词在索引数据库匹配对应的文档，并将匹配结果发送至文档服务器 130。然后，文档服务器 130 可根据匹配结果建立每个分词对应的文档集，通过对所有分词各自对应的文档集求交集后得到文档交集，并通过 WEB 服务器 110 将该文档交集返回给搜索引擎浏览器 140。这样，搜索引擎浏览器 140 可将文档交集集中的文档展示给用户。

[024] 本发明提供的一种多文档交集获取方法中常见的术语有：

[025] 文档（Document）：一般搜索引擎的处理对象是互联网网页，而文档这个概念要更宽泛些，代表以文本形式存在的存储对象。相比网页来说，文档涵盖更多种形式，比如 Word、PDF、html、XML 等不同格式的文件都可以称之为文档，再比如一封邮件、一条短信、一条微博也可以称之为文档。在本发明中，每个文档对应设置有文档标识，用于识别各个文档。

[026] 文档集（Document Collection）：由若干文档构成的集合称之为文档集。比如海量的互联网网页或者大量的电子邮件都是文档集的具体例子。

[027] 语法树 (Parse Tree) : 语法树是语句结构的图形表示, 它代表了语句的推导结果, 可有利于理解语句的语法结构的层次。简单说, 语法树就是按照某一规则进行推导时所形成的树。

[028] 叶子节点 (leaf node) : 叶子节点为语法树最底层的节点, 叶子节点不再包括更底层的节点。在本发明中, 叶子节点为搜索词的分词。

[029] 参照图 2, 其示出了一种多文档交集获取方法的步骤流程图。

[030] 步骤 210, 针对需要求交集的至少两个文档集, 获取各文档集的文档集长度。

[031] 在索引数据库中, 搜索词的不同分词匹配的文档数量也可能不相同, 更由于不同分词之间存在使用热度的差异, 针对搜索过程中需要求交集的至少两个文档集的长度之间的差值极有可能超过长度阈值。需要说明的是, 文档集的长度是指文档集中包括的文档元素的数量。

[032] 在本发明实施例中, 针对各个分词对应生成的需要求交集的文档集, 获取每个文档集的长度, 具体可获取每个文档集中文档的数量。

[033] 例如, 若用户在搜索引擎浏览器 140 输入搜索词“海底捞火锅”, 则可得到分词 1 “海底捞”和分词 2 “火锅”, 针对分词 1 和分词 2 对应查询得到的文档如下表 1 所示:

海底捞	火锅
文档 1	文档 1
文档 2	文档 3
文档 20	文档 5
文档 85	文档 6
	...
	文档 20
	...
	文档 80

表 1

[034] 因此, 将查询得到的各个分词对应的文档按照文档标识的大小顺排, 可得到表 1。其中, 分词 1 “海底捞”对应的第一文档集中文档元素的个数为 4, 分词 2 “火锅”对应的第二文档集中文档元素的个数为 40 (文档 6 至文档 20 之间和文档 20 至文档 80 之间有省略)。

这样，可获得分词 1 “海底捞”对应的第一文档集长度为 4，分词 2 “火锅”对应的第二文档集的长度为 40。

[035] 步骤 220，对所述至少两个文档集的长度进行比较，以根据所述至少两个文档集的长度差异确定获得文档交集的求交算法。

5 [036] 步骤 230，利用所确定的求交算法，获得所述至少两个文档集的文档交集。

[037] 在本发明实施例中，若至少两个文档集的文档集长度符合预设条件，则可以以最小文档集中的查询元素为遍历起点，查找剩余的文档集中是否有该查询元素。其中，所述最小文档集为所述至少两个文档集中具有最小文档集长度的文档集。预设条件可以为，多个文档集中最长文档集与最短文档集的长度之差大于第一预设阈值。该第一预设阈值可以根据实际情况进行设定，例如可以是通过搜索引擎进行日常搜索试验得到的一个优选值。需要说明的是，预设条件也可以为多个文档集中，最大文档集长度相对于最小文档集长度的比值超过第二预设阈值。

[038] 针对表 1 的数据，若预设条件为多个文档集中最长文档集与最短文档集的长度之差大于 10，则第一文档集和第二文档集的长度之差为 30，符合预设条件。此时，以第一文档集中的文档 1 作为查询元素，进行第二文档集的遍历。发现第二文档集中存在文档 1，并将文档 1 插入文档交集。之后，以第一文档集中的文档 2 作为查询元素进行第二文档集的遍历，并发现第二文档集中不存在文档 2。继续选取第一文档集中的文档 20 作为查询元素，并发现第二文档集中存在文档 20，将文档 20 插入文档交集后，以第一文档集中的文档 85 作为查询元素进行第二文档集的遍历。最后，发现第二文档集中不存在文档 85，并且第一文档集被遍历结束。这样，可终止本次求交过程，并将最终的文档交集[文档 1，文档 20]返回给用户。

[039] 若按照通用的求交算法，则针对表 1，大致的求交过程如下：以第一文档集中的文档 1 作为查询元素进行第二文档集的遍历，并发现第二文档集中存在文档 1，将文档 1 插入文档交集。之后，以第一文档集中的文档 2 作为查询元素进行第二文档集的遍历，并发现第二文档集中不存在文档 2，从而以第二文档集中的文档 3 为查询元素进行第一文档集的遍历。由于第一文档集不存在文档 3，则以第一文档集中的文档 20 为查询元素。发现第二文档集中存在文档 20，则将文档 20 插入文档交集，并将查询元素更换为第一文档集中的文档 85。最后，发现第二文档集中不存在文档 85，并且第一文档集被遍历结束。这样，可终止本次求交过程，并将最终的文档交集[文档 1，文档 20]返回给用户。明显可见，该通用的求交算法造成了以文档 3 进行遍历的浪费。实际上，文档 3 没有在第一文档集中出现，也就是说该文档根本不会被召回，因此，在至少两个文档集的长度之间的差值超过长度阈值的情况下，采用通用求

交算法进行求交的效率较低。

[040] 例如，若用户通过电脑浏览器的搜索引擎查询“海底捞火锅”，则文档 1 和文档 20 对应的网页为用户此次查询得到的结果，并可通过浏览器界面展现给用户文档 1 和文档 20 对应网页的链接。

5 [041] 另外，若用户通过手机应用中的搜索引擎查询“海底捞火锅”，则文档 1 和文档 20 对应的应用页面为用户此次查询得到的结果，并可通过手机界面展现给用户的是文档 1 和文档 20 对应应用页面的链接。

10 [042] 综上所述，本发明实施例提供了一种多文档交集获取方法，针对搜索过程中需要求交集的至少两个文档集，通过在两个文档集的文档集长度符合预设条件时，以长度最短文档集中的元素作为查询元素依次遍历剩余的文档集，可以有效提高文档求交集效率，加快了搜索引擎针对用户的响应时间。

[043] 参照图 3，其示出了一种多文档交集获取方法的具体步骤流程图。

[044] 步骤 310，接收搜索词。

15 [045] 在本发明实施例中，搜索引擎可以接收用户输入的搜索词，并对搜索词进行语法树构建。

[046] 步骤 320，根据接收到的搜索词构建搜索语法树。其中，所述语法树的叶子节点为搜索词的分词。

20 [047] 在本发明实施例中，可通过语法解析用户输入的搜索词并构建语法树。例如，若设定规则：在通过语法解析发现有空格时，判定空格的前后两个词为“与（and）”的关系。假设用户输入了“北京市 全时便利店”，则会将空格的前后两个词“北京市”和“全时便利店”保存，并将解析结果构造成如图 3A 所示的初级语法树。

25 [048] 之后，在构造最终语法树节点时，会根据图 3A 所示的初级语法树的结构进行同步构造。这个过程中，系统会判别初级语法树的节点是否为文本。如果是文本，系统会对其进行再次分词，例如“北京市”会分为“北京”和“市”，“全时便利店”会分为“全时”和“便利”和“店”。可根据分词后的结果对初级语法树重建，并将分词后的“与”节点添加到初级语法树中，形成如图 3B 所示的终级语法树，以供搜索引擎对终极语法树做各个分词的求交算法。

[049] 步骤 330，针对需要求交集的至少两个文档集，获取各文档集的文档集长度。

[050] 该步骤的基本实现可以参照上述步骤 210，此处不再赘述。

[051] 此外，针对步骤 320 中构建的搜索语法树，可从所述搜索语法树中还未进行求交计算的最低层级的求交节点开始，根据所述求交节点的各子节点确定需要求交集的所述至少两个文档集，并确定所述至少两个文档集各自的文档集长度。

5 [052] 在本发明实施例中，可从根据搜索词构建的语法树中的最底层求交节点开始进行求交计算。当最底层的求交节点完成求交计算后，根据得到的底层的文档交集，再对次底层的求交节点进行求交计算，直到得到最顶层的文档交集，并将其返回给用户。

10 [053] 例如，图 3B 示出的针对“北京市 全时便利店”的终级语法树中，求交节点 10、20 为底层节点，求交节点 30 为顶层节点。此时，根据底层节点 10 和底层节点 20，可确定顶层节点 30 需要求交集的文档集为“北京”和“市”对应的文档集与“全时”、“便利”和“店”对应的文档集。

[054] 接着，可分别得到“北京”和“市”各自对应的文档集中文档元素的数量，以及“全时”、“便利”和“店”各自对应的文档集中文档元素的数量。

15 [055] 步骤 340，对所述至少两个文档集的长度进行比较，以根据所述至少两个文档集的长度差异确定获得文档交集的求交算法。

20 [056] 步骤 350，在所述至少两个文档集的长度差异符合预设条件的情况下，以最小文档集中的查询元素为遍历起点，查找剩余的文档集中是否有所述查询元素。如果在每个所述剩余的文档集中均找到了匹配所述当前排序序号的查询元素的元素，则将所述查询元素作为所述文档交集的元素。其中，所述最小文档集为所述至少两个文档集中具有最小文档集长度的文档集。

[057] 其中，所述预设条件可包括：所述至少两个文档集的文档集长度中，最大文档集长度与最小文档集长度之间的差值超过第一预设阈值；或者，所述至少两个文档集中，最大文档集长度相对于最小文档集长度的比值超过第二预设阈值。该第一预设阈值与第二预设阈值均可以根据实际情况进行设定。例如，可通过对搜索引擎进行日常搜索试验得到的一个优选值。

25 [058] 可对所述最小文档集中的当前排序序号的查询元素进行与每个剩余文档集中的元素的匹配。如果在至少一个所述剩余文档集中未找到匹配所述当前排序序号的查询元素的元素，则对所述最小文档集中的下一排序序号的查询元素进行与每个剩余文档集中的元素的匹配。如果在所有所述剩余文档集中均找到了匹配所述当前排序序号的查询元素的元素，则将所述查询元素作为所述文档交集的元素，并对所述最小文档集中的下一排序序号的查询元素进行

与剩余的文档集中的元素的匹配。以此循环，直至所述最小文档集中的所有元素被遍历。

[059] 例如，针对图 3B 示出的“北京市 全时便利店”的终级语法树中的求交节点 10，假设分词“北京”的第一文档集中文档元素的数量为 4，分词“市”对应的第二文档集中文档元素的数量为 40，第一预设阈值为 10，并且分词“北京”和“市”分别对应的文档元素如下

5 表 2 所示。

[060] 由于第一文档集和第二文档集的长度之差 30 大于第一预设阈值 10，则此时以第一文档集中的文档 1 作为查询元素，进行第二文档集的遍历。

北京	市
文档 1	文档 1
文档 2	文档 3
文档 20	文档 5
文档 85	文档 6
	...
	文档 20
	...
	文档 80

表 2

[061] 可选的，文档集中的元素已按标识大小顺排或者倒排。

北京	市
文档 85	文档 80
文档 20	...
文档 2	文档 20
文档 1	...
	文档 6
	文档 5

	文档 3
	文档 1

表 3

5 [062] 在本发明实施例中，文档集中元素的标识即为表 2 中文档后面对应的数字，该标识用于识别文档元素，表 2 是由文档集中的元素以按标识大小顺排得到的。若将文档集中的元素以按标识大小倒排，则可得到表 3，此时可以将第一文档集中的第一个元素文档 85 作为查询元素进行遍历，本发明实施例对此不作限定。

[063] 当以第一文档集中的文档 1 作为查询元素进行第二文档集的遍历，并发现第二文档集中存在文档 1，则将文档 1 插入文档交集。之后，以第一文档集中的文档 2 作为查询元素进行第二文档集的遍历，并发现第二文档集中不存在文档 2。这时，将选取第一文档集中的文档 20 作为查询元素，并发现第二文档集中存在文档 20，则将文档 20 插入文档交集。最后，再以第一文档集中的最后一个文档 85 作为查询元素进行第二文档集的遍历，并发现第二文档集中不存在文档 85，则停止遍历。接着，可将求交节点 10 得到的文档交集[文档 1，文档 20]返回给图 3B 示出的顶层求交节点 30。

15 [064] 步骤 360，在所述至少两个文档集的长度差异不符合预设条件的情况下，则利用通用的求交算法来获取所述文档交集。例如，可以以所述至少两个文档集中的关注文档集内的查询元素为遍历起点，查找每个剩余文档集中是否有所述查询元素。当每个所述剩余文档集中均有所述查询元素，则将所述查询元素作为所述文档交集的元素。其中，前一查询元素出自所述关注文档集且所述前一查询元素被确定为所述文档交集的元素，或者，所述关注文档集被第一个确定不具有所述前一查询元素；所述查询元素的排序序号为所述前一查询元素的下一排序序号。

20 [065] 在本发明实施例中，可以以所述至少两个文档集中的第一个文档集中的最小排序序号的查询元素作为最初的遍历起点，查找每个剩余文档集中是否有所述查询元素。如果在当前被查找的剩余文档集中没有找到匹配所述当前排序序号的查询元素的元素，则以该剩余文档集中的下一排序序号的元素作为新的查询元素。如果在所有剩余文档集中均找到了匹配所述当前排序序号的查询元素的元素，则将所述查询元素作为所述文档交集的元素，并以选出所述查询元素的文档集中的下一排序序号的元素作为新的查询元素。以此循环，直至所述至少

25 两个文档集中有一个文档集的所有元素被遍历。

[066] 在本发明实施例中，针对图 3B 示出的针对“北京市 全时便利店”的终级语法树中

的求交节点 20，假设分词“全时”对应的第三文档集中文档元素的数量为 2，分词“便利”对应的第四文档集中文档元素的数量为 3，分词“店”对应的第五文档集中文档元素的数量为 4，第一预设阈值为 10，分词“全时”、“便利”、“店”各自对应的文档元素可如下表 4 所示。

全时	便利	店
文档 1	文档 2	文档 20
文档 20	文档 20	文档 40
	文档 21	文档 50
		文档 60

表 4

[067] 此时，由于最长的第五文档集与最短的第三文档集之间的文档集长度差值 2 小于第一预设阈值 10，可通过通用求交算法进行第三文档集、第四文档集和第五文档集之间的求交计算，具体如下：以第三文档集中的文档 1 作为查询元素遍历第四文档集，并发现第四文档集中不存在文档 1，于是将查询元素换为第四文档集中的文档 2。发现第五文档集中不存在文档 2，于是将查询元素换成第五文档集中的文档 20，并发现第三文档集和第四文档集中均存在文档 20，则将文档 20 插入求交节点 20 对应的文档交集。当发现第三文档集和第四文档集中存在文档 20 之后，再以第五文档集中的文档 40 作为查询元素，并发现在第三文档集不存在文档 40 且第三文档集被遍历结束。此时，查询停止，并将求交节点 20 对应的文档交集[文档 20]上传至上层的求交节点 30。在各个分词对应的文档集之间的长度差值较小的情况下，采用通用求交算法会可具有更快的查询效率。

[068] 基于以上处理，在图 3B 示出的针对“北京市 全时便利店”的终级语法树中，求交节点 30 对应的文档集分别为求交节点 10 对应的文档交集[文档 1，文档 20]和求交节点 20 对应的文档交集[文档 20]。此时，可对求交节点 30 进行求交，得到最终的文档交集[文档 20]并返回给用户。

[069] 针对本发明提供的上述多文档交集获取方法，通过测试系统的测试，在关于响应延时的 TP90，TP99，TP999 指标方面均有提升，并且提升幅度在 10%以上。其中，TP90 时间为满足 90%请求的最低耗时；TP99 时间为满足 99%请求的最低耗时；TP999 时间为满足 99.9%请求的最低耗时。

[070] 综上所述,本发明实施例提供了的上述多文档交集获取方法,通过在搜索过程中需要求交集的至少两个文档集的文档集长度符合预设条件时,以长度最短文档集中的元素作为查询元素依次遍历剩余的文档集,可以有效提高文档求交集效率,加快了搜索引擎针对用户的响应时间。

5 [071] 参照图 4,其示出了一种多文档交集获取装置的结构图。

[072] 如图 4 所示,该多文档交集获取装置可包括:获取模块 401,用于针对搜索过程中需要求交集的至少两个文档集,获取各文档集的文档集长度;长度比较模块 402,用于对所述至少两个文档集的长度进行比较,以确定获得文档交集的求交算法;以及求交模块 403,用于根据所确定的求交算法获得所述至少两个文档的交集。

10 [073] 综上所述,本发明实施例提供了一种多文档交集获取装置,通过在需要求交集的至少两个文档集的文档集长度符合预设条件时,以长度最短文档集中的元素作为查询元素依次遍历剩余的文档集,可以有效提高文档求交集效率,加快了搜索引擎针对用户的响应时间。

[074] 参照图 5,其示出了一种多文档交集获取装置的具体结构图。

[075] 如图 5 所示,该多文档交集获取装置可包括:

15 [076] 接收模块 501,用于接收搜索词。

[077] 语法树构建模块 502,用于根据所述搜索词构建搜索语法树,所述语法树的叶子节点为所述搜索词的分词。

[078] 获取模块 503,用于针对搜索过程中需要求交集的至少两个文档集,获取各文档集的文档集长度;

20 [079] 长度比较模块 504,用于对所述至少两个文档集的长度进行比较,以确定获得文档交集的求交算法。

[080] 第一求交模块 505,用于在所述至少两个文档集的长度差异符合预设条件的情况下,以最小文档集中的查询元素为遍历起点,查找每个剩余文档集中是否有所述查询元素;以及,当每个剩余文档集中均有所述查询元素,则将所述查询元素作为所述文档交集的元素。其中,25 所述最小文档集为所述至少两个文档集中具有最小文档集长度的文档集。

[081] 第二求交模块 506,用于在所述至少两个文档集的长度差异不符合预设条件的情况下,则利用通用的求交算法来获得所述文档交集。例如,可以以所述至少两个文档集中的关注文档集内的查询元素为遍历起点,查找每个剩余文档集中是否有所述查询元素;以及,当每个

剩余文档集中均有所述查询元素，则将所述查询元素作为所述文档交集的元素。其中，前一查询元素出自所述关注文档集且所述前一查询元素被确定为所述文档交集的元素，或者，所述关注文档集被第一个确定不具有所述前一查询元素；所述查询元素的排序序号为所述前一查询元素的下一排序序号。

- 5 [082] 其中，获取模块 503 可具体用于：从所述搜索语法还未进行求交计算的求交节点中的最低层级的第一求交节点开始，根据所述第一求交节点的子节点确定需要求交集的所述至少两个文档集；以及，获取每个文档集的文档集长度。

[083] 第一求交模块 505 可具体用于：对所述最小文档集中的当前排序序号的查询元素进行与剩余文档集中的元素的匹配；如果在至少一个所述剩余文档集中未找到匹配所述当前排序
10 序号的查询元素的元素，则对所述最小文档集中的下一排序序号的查询元素进行与剩余的文档集中的元素的匹配；如果在所有所述剩余文档集中均找到了匹配所述当前排序序号的查询元素的元素，则将所述查询元素作为所述文档交集的元素，并对所述最小文档集中的下一排序序号的查询元素进行与剩余文档集中的元素的匹配。以此循环，直至所述最小文档集中的所有元素被遍历。

15 [084] 第二求交模块 506 可具体用于：以所述至少两个文档集中的第一个文档集中的最小排序序号的查询元素作为最初的遍历起点，查找每个剩余文档集中是否有所述查询元素。如果在当前被查找的剩余文档集中没有找到匹配当前排序序号的查询元素的元素，则以该剩余文档集中的下一排序序号的元素作为新的查询元素。如果在所有剩余文档集中均找到了匹配所述当前排序序号的查询元素的元素，则将所述查询元素作为所述文档交集的元素，并以选出
20 所述查询元素的文档集中的下一排序序号的元素作为新的查询元素。以此循环，直至所述至少两个文档集中有一个文档集的所有元素被遍历。

[085] 综上所述，本发明实施例提供了一种多文档交集获取装置，通过在需要求交集的至少两个文档集的文档集长度符合预设条件时，以长度最短文档集中的元素作为查询元素依次遍历剩余的文档集，可以有效提高了文档集长度差值较大情况下的文档求交集效率，加快了搜索
25 引擎针对用户的响应时间。

[086] 本发明实施例还提供了一种文档服务器，包括：处理器；以及非临时性计算机可读存储介质，在所述非临时性计算机可读存储介质上存储有能够被所述处理器执行的机器可执行指令。其中，所述处理器被所述机器可执行指令促使执行前述实施例的多文档交集获取方法中的步骤，例如：针对搜索中需要求交集的至少两个文档集，获取各所述文档集的文档集长度；根据所述至少两个文档集的长度差异确定获得文档交集的求交算法；以及，利用所确定
30

的所述求交算法，获得所述至少两个文档集的文档交集。

[087] 本发明实施例还提供了一种非临时性机器可读存储介质，当所述存储介质中的指令由文档服务器的处理器执行时，使得文档服务器能够执行前述实施例的多文档交集获取方法。

5 [088] 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。

10 [089] 在此提供的算法和显示不与任何特定计算机、虚拟系统或者其它设备固有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述，构造这类系统所要求的结构是显而易见的。此外，本发明也不针对任何特定编程语言。应当明白，可以利用各种编程语言实现在此描述的本发明的内容，并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[090] 在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

15 [091] 类似地，应当理解，为了精简本公开并帮助理解各个发明方面中的一个或多个，在上面对本发明的示例性实施例的描述中，本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释成反映如下意图：即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如下面的权利要求书所反映的那样，发明方面在于少于前面公开的单个实施例的所有特征。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本发
20 明的单独实施例。

[092] 本领域那些技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可以把它分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

25

[093] 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支付信息处理设备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。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载得到，或者在载体信号上提供，或者以任何其他形式提供。

[094] 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“包含”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是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。

权利要求书

1、一种多文档交集获取方法，包括：

针对搜索中需要求交集的至少两个文档集，获取各所述文档集的文档集长度；

根据所述至少两个文档集的长度差异确定获得文档交集的求交算法；以及

5 利用所确定的所述求交算法，获得所述至少两个文档集的文档交集。

2、如权利要求 1 所述的方法，其特征在于，利用所确定的所述求交算法，获得所述至少两个文档集的文档交集，包括：

在所述至少两个文档集的长度差异符合预设条件的情况下，以最小文档集中的查询元素为遍历起点，查找每个剩余文档集中是否有所述查询元素，其中，所述最小文档集为所述至少两个文档集中具有最小文档集长度的文档集；

当每个所述剩余文档集中均有所述查询元素，将所述查询元素作为所述文档交集的元素。

3、根据权利要求 2 所述的方法，其特征在于，所述预设条件包括以下任意一个或多个：

所述至少两个文档集中，最大文档集长度与最小文档集长度之间的差值超过第一预设阈值；以及

15 所述至少两个文档集中，最大文档集长度相对于最小文档集长度的比值超过第二预设阈值。

4、根据权利要求 2 所述的方法，其特征在于，以所述最小文档集中的查询元素为遍历起点，查找每个剩余文档集中是否有所述查询元素，包括：

对所述最小文档集中的当前排序序号的查询元素进行与剩余文档集中的元素的匹配；

20 如果在至少一个剩余文档集中未找到匹配所述查询元素的元素，则对所述最小文档集中的下一排序序号的查询元素进行与剩余文档集中的元素的匹配；

如果在所有剩余文档集中均找到了匹配所述查询元素的元素，则将所述查询元素作为所述文档交集的元素，并对所述最小文档集中的下一排序序号的查询元素进行与剩余文档集中的元素的匹配。

25 5、根据权利要求 1 所述的方法，其特征在于，所述文档集中的元素按标识大小顺排或者倒排而具有所述排序序号。

6、根据权利要求 1 所述的方法，其特征在于，还包括：

接收用于触发所述搜索的搜索词；

根据所述搜索词构建搜索语法树，其中，所述语法树的叶子节点为所述搜索词的分词。

30 7、根据权利要求 6 所述的方法，其特征在于，针对所述搜索中需要求交集的所述至少两个文档集，获取各所述文档集的文档集长度，包括：

从所述搜索语法树中还未进行求交计算的最低层级的求交节点开始，根据所述求交节点
的各子节点确定需要求交集的所述至少两个文档集；

获取每个所述文档集的文档集长度。

8、根据权利要求 1-7 中任一所述的方法，其特征在于，利用所确定的所述求交算法，获
5 得所述至少两个文档集的文档交集，包括：

在所述至少两个文档集的长度差异不符合预设条件的情况下，则以所述至少两个文档集
中的关注文档集内的查询元素为遍历起点，查找每个剩余文档集中是否有所述查询元素；

当每个所述剩余文档集中均有所述查询元素，则将所述查询元素作为所述文档交集的元
素，

10 其中，前一查询元素出自所述关注文档集且所述前一查询元素被确定为所述文档交集的
元素，或者，所述关注文档集被第一个确定不具有所述前一查询元素；所述查询元素的排序
序号为所述前一查询元素的下一排序序号。

9、一种文档服务器，包括：

处理器；以及

15 非临时性计算机可读存储介质，在所述非临时性计算机可读存储介质上存储有能够被所
述处理器执行的机器可执行指令，

其中，所述处理器被所述机器可执行指令促使：

针对搜索中需要求交集的至少两个文档集，获取各所述文档集的文档集长度；

根据所述至少两个文档集的长度差异确定获得文档交集的求交算法；以及

20 利用所确定的所述求交算法，获得所述至少两个文档集的文档交集。

10、如权利要求 9 所述的装置，其特征在于，在利用所确定的所述求交算法，获得所述
至少两个文档集的文档交集时，所述处理器被所述机器可执行指令促使：

在所述至少两个文档集的长度差异符合预设条件的情况下，以最小文档集中的查询元素
为遍历起点，查找每个剩余文档集中是否有所述查询元素，其中，所述最小文档集为所述至

25 少两个文档集中具有最小文档集长度的文档集；

当每个所述剩余文档集中均有所述查询元素，将所述查询元素作为所述文档交集的元素。

11、根据权利要求 10 所述的装置，其特征在于，所述预设条件包括以下任意一个或多个：

所述至少两个文档集中，最大文档集长度与最小文档集长度之间的差值超过第一预设阈
值；以及

30 所述至少两个文档集中，最大文档集长度相对于最小文档集长度的比值超过第二预设阈
值。

12、根据权利要求 10 所述的装置，其特征在于，在以所述最小文档集中的查询元素为遍历起点，查找每个剩余文档集中是否有所述查询元素时，所述处理器被所述机器可执行指令促使：

对所述最小文档集中的当前排序序号的查询元素进行与剩余文档集中的元素的匹配；

5 如果在至少一个剩余文档集中未找到匹配所述查询元素的元素，则对所述最小文档集中的下一排序序号的查询元素进行与剩余文档集中的元素的匹配；

如果在所有剩余文档集中均找到了匹配所述查询元素的元素，则将所述查询元素作为所述文档交集的元素，并对所述最小文档集中的下一排序序号的查询元素进行与剩余文档集中的元素的匹配。

10 13、根据权利要求 9 所述的装置，其特征在于，所述文档集中的元素按标识大小顺排或者倒排而具有所述排序序号。

14、根据权利要求 9 所述的装置，其特征在于，所述处理器还被所述机器可执行指令促使：

接收用于触发所述搜索的搜索词；

15 根据所述搜索词构建搜索语法树，其中，所述语法树的叶子节点为所述搜索词的分词。

15、根据权利要求 14 所述的装置，其特征在于，在针对所述搜索中需要交集的所述至少两个文档集，获取各所述文档集的文档集长度时，所述处理器被所述机器可执行指令促使：

从所述搜索语法树中还未进行求交计算的最低层级的求交节点开始，根据所述求交节点的各子节点确定需要交集的所述至少两个文档集；

20 获取每个所述文档集的文档集长度。

16、根据权利要求 9-15 中任一所述的装置，其特征在于，在利用所确定的所述求交算法，获得所述至少两个文档集的文档交集时，所述处理器被所述机器可执行指令促使：

在所述至少两个文档集的长度差异不符合预设条件的情况下，则以所述至少两个文档集中的关注文档集中的查询元素为遍历起点，查找每个剩余文档集中是否有所述查询元素；

25 当每个所述剩余文档集中均有所述查询元素，将所述查询元素作为所述文档交集的元素，其中，前一查询元素出自所述关注文档集且所述前一查询元素被确定为所述文档交集的元素，或者，所述关注文档集被第一个确定不具有所述前一查询元素；所述查询元素的排序序号为所述前一查询元素的下一排序序号。

17、一种非临时性机器可读存储介质，其上存储有能够被处理器执行的机器可执行指令，
30 当所述非临时性机器可读存储介质中的所述机器可执行指令由文档服务器中的处理器执行时，所述文档服务器能够执行如权利要求 1-8 任一所述的多文档交集获取方法。

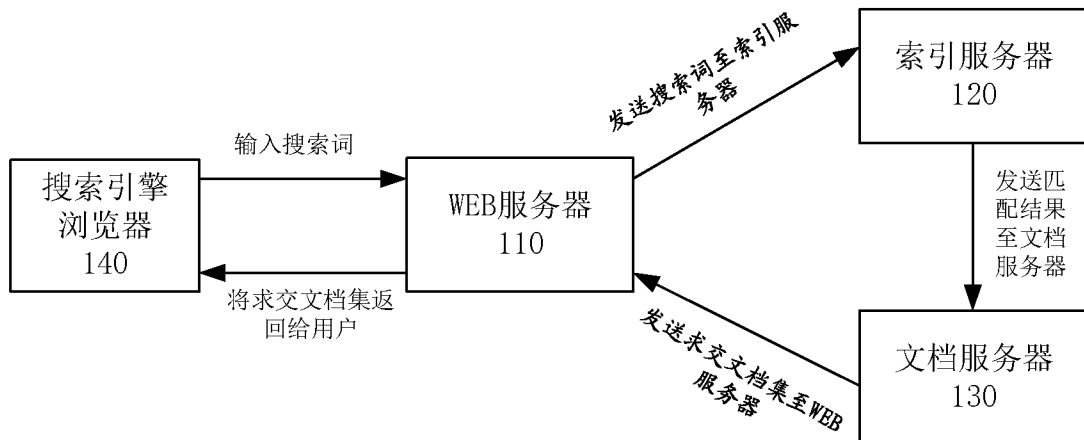


图 1

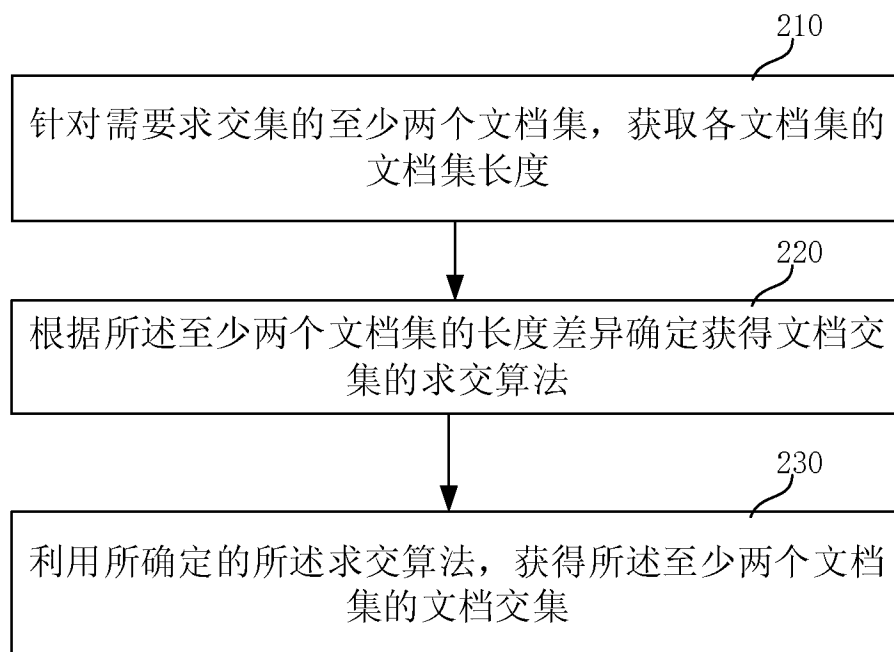


图 2

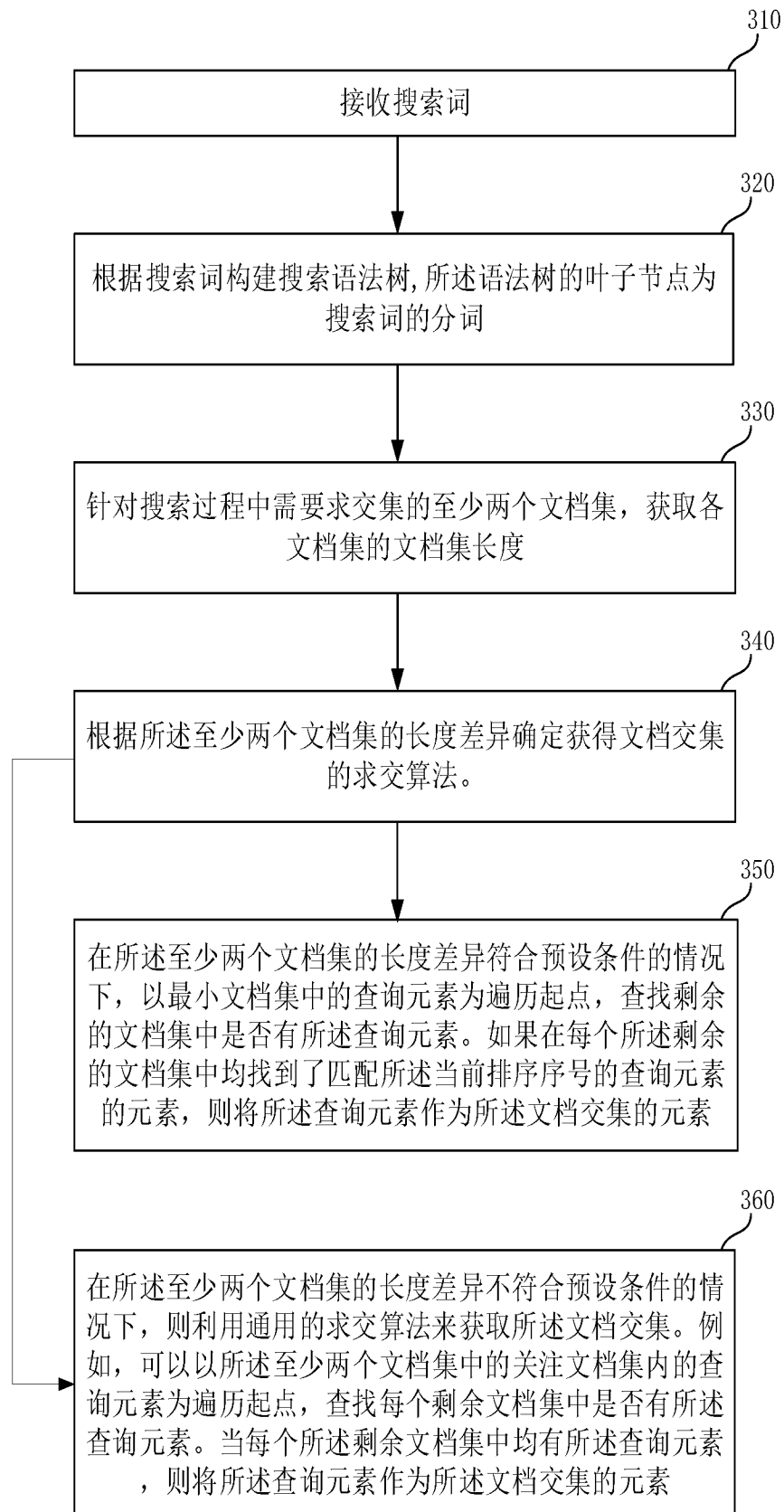


图 3

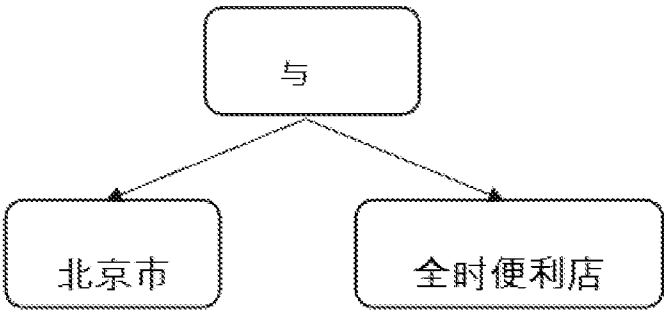


图 3A

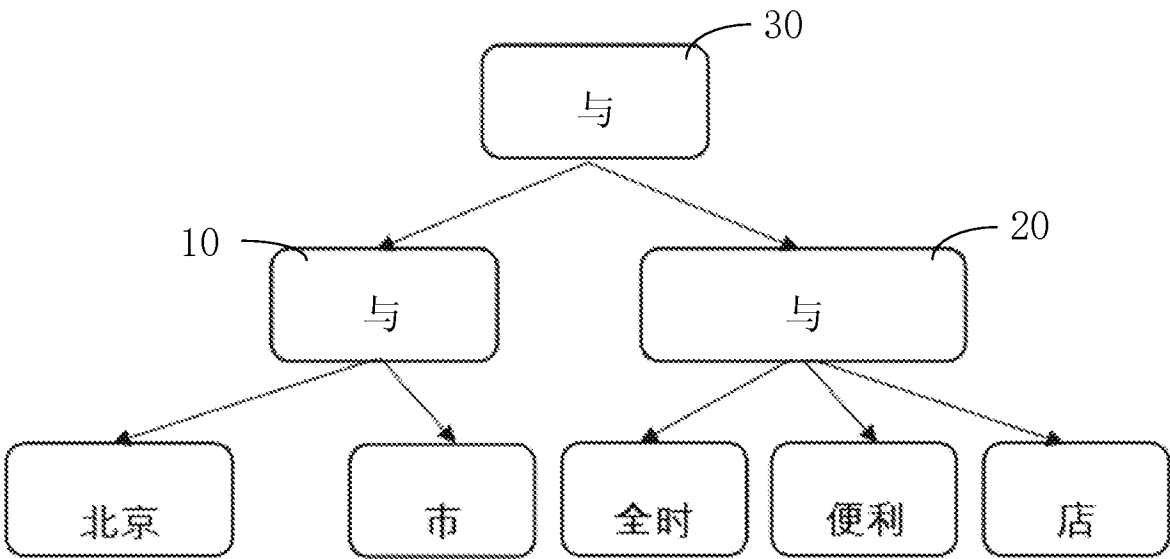


图 3B

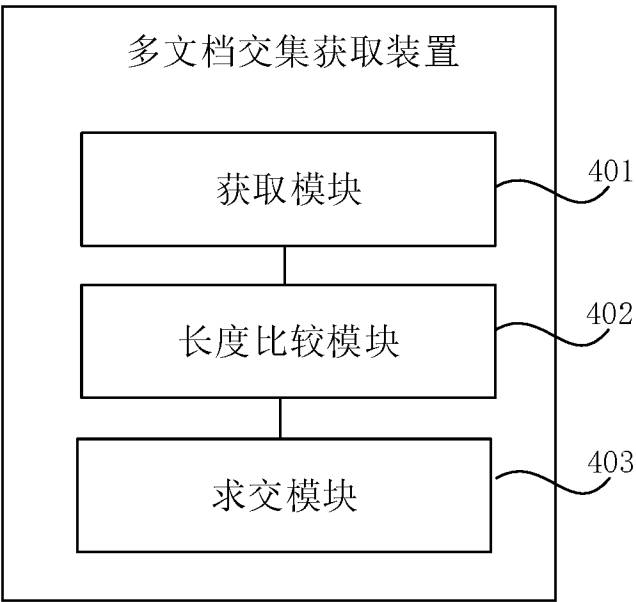


图 4

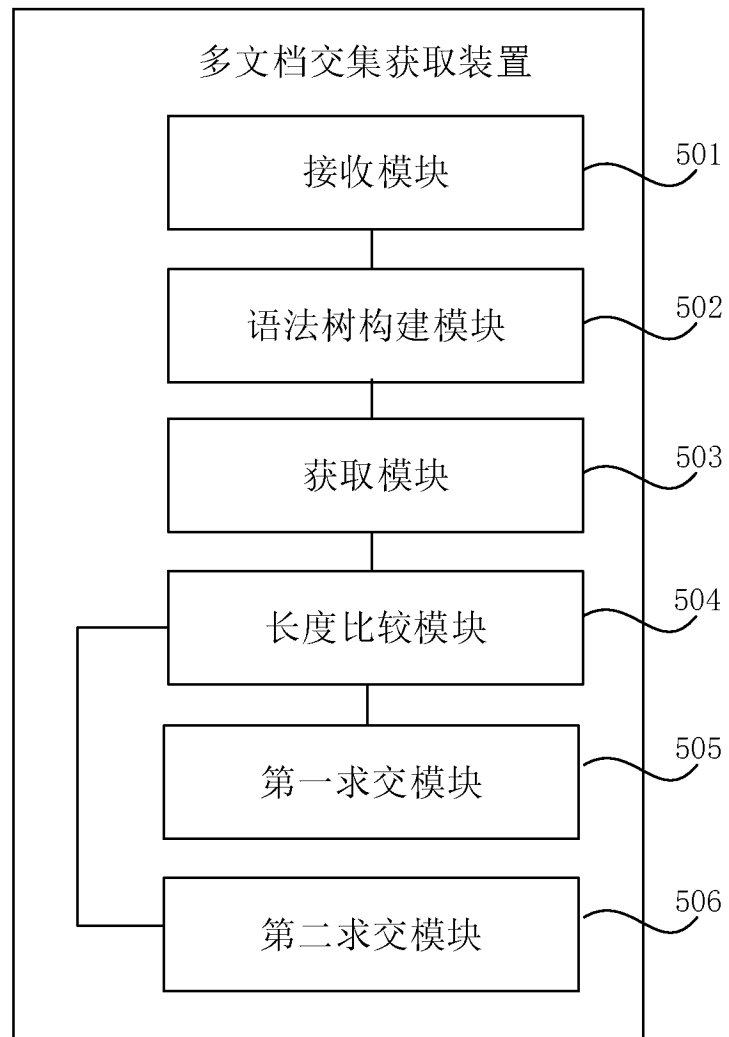


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/120062

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)

WPI, EPODOC, CNKI, CNPAT, IEEE: 文档, 文档集, 集合, 求交, 交集, 搜索, 排序, 引擎, 长度, 数量, 最小文档集, 遍历, 匹配, 预设条件, document, set, and, search, small, match, preset, number

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102810096 A (ALIBABA GROUP HOLDING LTD.) 05 December 2012 (05.12.2012), description, paragraphs [0032]-[0103], and figures 3-11	1, 5-9, 13-17
A	CN 1858737 A (HUAWEI TECHNOLOGIES CO., LTD.) 08 November 2006 (08.11.2006), entire document	1-17
A	CN 102201007 A (YOUYI BEIJING ADVERTISING CO., LTD.) 28 September 2011 (28.09.2011), entire document	1-17
A	CN 102750393 A (CTRIIP COMPUTER TECHNOLOGY (SHANGHAI) CO., LTD.) 24 October 2012 (24.10.2012), entire document	1-17
A	WO 2016173366 A1 (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 03 November 2016 (03.11.2016), entire document	1-17

☐ 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 22 May 2018	Date of mailing of the international search report 05 June 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 ZHANG, Qianzhen Telephone No. (86-10) 53961401

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/120062

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102810096 A	05 December 2012	US 9311389 B2	12 April 2016
		CN 102810096 B	16 March 2016
		US 2012310915 A1	06 December 2012
		HK 1173819 A1	17 February 2017
		JP 2014519123 A	07 August 2014
		WO 2012166960 A1	06 December 2012
		TW 201250502 A	16 December 2012
		JP 5746426 B2	08 July 2015
		EP 2715572 A1	09 April 2014
CN 1858737 A	08 November 2006	WO 2007085187 A1	02 August 2007
		CN 1858737 B	02 June 2010
CN 102201007 A	28 September 2011	None	
CN 102750393 A	24 October 2012	CN 102750393 B	16 July 2014
WO 2016173366 A1	03 November 2016	CN 106156000 A	23 November 2016
		US 2017193081 A1	06 July 2017

国际检索报告

国际申请号

PCT/CN2017/120062

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))

WPI, EPODOC, CNKI, CNPAT, IEEE: 文档, 文档集, 集合, 求交, 交集, 搜索, 排序, 引擎, 长度, 数量, 最小文档集, 遍历, 匹配, 预设条件, document, set, and, search, small, match, preset, number

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
X	CN 102810096 A (阿里巴巴集团控股有限公司) 2012年 12月 5日 (2012 - 12 - 05) 说明书第[0032]-[0103] 段, 及附图3-11	1, 5-9, 13-17
A	CN 1858737 A (华为技术有限公司) 2006年 11月 8日 (2006 - 11 - 08) 全文	1-17
A	CN 102201007 A (悠易互通北京广告有限公司) 2011年 9月 28日 (2011 - 09 - 28) 全文	1-17
A	CN 102750393 A (携程计算机技术上海有限公司) 2012年 10月 24日 (2012 - 10 - 24) 全文	1-17
A	WO 2016173366 A1 (腾讯科技深圳有限公司) 2016年 11月 3日 (2016 - 11 - 03) 全文	1-17

☐ 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 5月 22日

国际检索报告邮寄日期

2018年 6月 5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

受权官员

张乾桢

电话号码 86-(10)-53961401

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/120062

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	102810096	A	2012年 12月 5日	US	9311389	B2	2016年 4月 12日
				CN	102810096	B	2016年 3月 16日
				US	2012310915	A1	2012年 12月 6日
				HK	1173819	A1	2017年 2月 17日
				JP	2014519123	A	2014年 8月 7日
				WO	2012166960	A1	2012年 12月 6日
				TW	201250502	A	2012年 12月 16日
				JP	5746426	B2	2015年 7月 8日
				EP	2715572	A1	2014年 4月 9日
CN	1858737	A	2006年 11月 8日	WO	2007085187	A1	2007年 8月 2日
				CN	1858737	B	2010年 6月 2日
CN	102201007	A	2011年 9月 28日	无			
CN	102750393	A	2012年 10月 24日	CN	102750393	B	2014年 7月 16日
WO	2016173366	A1	2016年 11月 3日	CN	106156000	A	2016年 11月 23日
				US	2017193081	A1	2017年 7月 6日

表 PCT/ISA/210 (同族专利附件) (2015年1月)