

(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 6 月 30 日 (30.06.2016)



(10) 国际公布号
WO 2016/101737 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2015/095018

(22) 国际申请日: 2015 年 11 月 19 日 (19.11.2015)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201410806927.4 2014 年 12 月 22 日 (22.12.2014) 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: 郭峰 (GUO, Feng); 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。李亚平 (LI, Yaping); 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。彭仁刚 (PENG,

Rengang); 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。秦吉胜 (QIN, Jisheng); 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,

[见续页]

(54) Title: SEARCH QUERY METHOD AND APPARATUS

(54) 发明名称: 搜索查询方法和装置

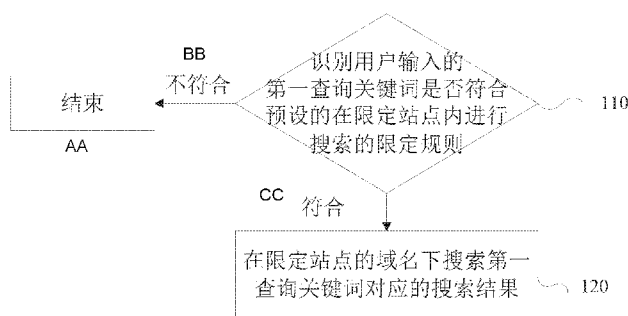


图 1 / FIG.1

110 Recognising whether a first query keyword input by a user conforms to a pre-set limit rule for searching in a limit station
120 Searching under a domain name of the limit station for a search result corresponding to the first query keyword
AA End
BB No
CC Yes

(57) Abstract: Disclosed are a search query method and apparatus, which mainly relate to the technical field of Internet and mainly aim to meet the requirement of a user to acquire resources at a limit station. The method comprises: recognising whether a first query keyword input by a user conforms to a pre-set limit rule for searching in a limit station; and if the first query keyword conforms to the limit rule suitable for searching in the limit station, searching under a domain name of the limit station for a search result corresponding to the first query keyword. According to the present invention, the result of a search query will not contain a result apart from non-limit stations and will not contain home pages of certain stations. Since the search is only performed on the limit station, a calculation amount engendered by the search is small, and interference from a spam page of a non-limit station can be more easily avoided.

(57) 摘要:

[见续页]



WO 2016/101737 A1

CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, **本国际公布:**

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)。

— 包括国际检索报告(条约第 21 条(3))。

本发明公开了一种搜索查询方法和装置，主要涉及互联网技术领域，主要目的在于满足用户在限定站点获取资源的需求。方法包括：识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则；如符合适于在限定站点内进行搜索的限定规则，则在限定站点的域名下搜索第一查询关键词对应的搜索结果。根据本发明，搜索查询的结果既不会包含非限定站点外的结果，也不会包含某些站点的首页；由于仅在限定站点上进行搜索，所以搜索产生的计算量也较小，也更容易避免非限定站点的作弊网页的干扰。

搜索查询方法和装置

技术领域

本发明涉及互联网技术领域，具体而言，涉及一种搜索查询方法和装置。

背景技术

对于目前的搜索引擎，准确理解用户意图，提高搜索引擎结果的质量，改善用户搜索体验是搜索引擎的目标之一。

现有的搜索引擎，通常直接使用用户的查询关键词进行搜索查询，首先对查询关键词进行分词并从中选取核心词，召回含有核心词的网页并进行排序；或者对查询关键词进行转义，如同义词转义，句式转义等，之后该查询关键词被转化成多个查询关键词，然后每个查询关键词会召回的结果进行合并，最后统一进行排序。

现有的搜索引擎，在某些情况下的效果并不理想。用户使用搜索引擎的时候，有时候希望仅仅得到某个网站内的资源页面，所以往往通过在查询关键词中加入站点名称的方式来进行这种查询，例如：“葫芦娃 360 影视”，用户的目的实际上是在“360 影视”的站点上查找“葫芦娃”节目。则现有的搜索引擎会造成：1) 查询结果中含有非用户需要的站点结果，例如非“360 影视”站点上的“葫芦娃”节目；2) 查询结果中，站点首页可能会更靠前，因为首页的权重往往更大，如“360 影视”的站点首页，但这些首页其实不是用户期待的结果；3) 召回结果过多，导致查询计算量较大；4) 有些作弊网站可能会错误的提权，导致放到前面。由此可见，需要一种新的搜索查询方案来满足用户在某个站点获取资源的需求。

发明内容

鉴于上述问题，提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的一种搜索查询方法和装置。

依据本发明的一个方面，提供了一种搜索查询方法，其包括：识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则；如符合适于在限定站点内进行搜索的限定规则，则在所述限定站点的域名下搜索所述第一查询关键词对应的搜索结果。

依据本发明的另一方面，还提供了一种搜索查询装置，其包括：第一查询关键词识别模块，用于识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则；搜索模块，如符合适于在限定站点内进行搜索的限定规则，则在所述限定站点的域名下搜索所述第一查询关键词对应的搜索结果。

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搜索查询方法。

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。

5 根据以上技术方案，可知本发明的搜索查询方法和装置至少具有以下优点：

在接收到用户的查询关键词时，首先分析用户输入查询关键词的目的是不是为了在限定站点搜索资源；在确定用户需求在限定站点搜索资源之后，则可以在该限定站点的域名之下进行搜索查询；因此，搜索查询的结果既不会包含非限定站点外的结果，也不会包含某些站点的首页；由于仅在限定站
10 点上进行搜索，所以搜索产生的计算量也较小，也更容易避免非限定站点的作弊网页的干扰。

上述说明仅是本发明技术方案的概述，为了能够更清楚了解本发明的技术手段，而可依照说明书的内容予以实施，并且为了让本发明的上述和其它
15 目的、特征和优点能够更明显易懂，以下特举本发明的具体实施方式。

附图说明

通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，
20 而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

图 1 示出了根据本发明的一个实施例的搜索查询方法的流程图；
图 2 示出了根据本发明的一个实施例的搜索查询方法的流程图；
图 3 示出了根据本发明的一个实施例的搜索查询方法的流程图；
25 图 4 示出了根据本发明的一个实施例的搜索查询方法的流程图；
图 5 示出了根据本发明的一个实施例的搜索查询方法的流程图；
图 6 示出了根据本发明的一个实施例的搜索查询方法的流程图；
图 7 示出了根据本发明的一个实施例的搜索查询装置的框图；
图 8 示出了根据本发明的一个实施例的搜索查询装置的框图；
30 图 9 示出了根据本发明的一个实施例的搜索查询装置的框图；
图 10 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

图 11 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。

35

具体实施例

下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地

理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

如图 1 所示，本发明的一个实施例中提供了一种搜索查询方法，其包括：

步骤 110，识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。在本实施例中，对限定规则的类型不进行限制，例如，限定规则可以是：对多个站点的名称进行记录，如果第一查询关键词中包含了其中的站点名称，则判断需要在该站点下进行搜索。

步骤 120，如符合适于在限定站点内进行搜索的限定规则，则在限定站点的域名下搜索第一查询关键词对应的搜索结果。根据本实施例的技术方案，搜索查询的结果既不会包含非限定站点外的结果，也不会包含某些站点的首页；由于仅在限定站点上进行搜索，所以搜索产生的计算量也较小，也更容易避免非限定站点的作弊网页的干扰。

根据图 1，假设限定规则包含：如果用户输入的查询关键词中包含“360 影视”，则需要在“360 影视”站点中进行搜索。某用户输入了“葫芦娃 360 影视”进行搜索查询，其中包含了“360 影视”，则搜索引擎选择在“360 影视”的域名“www.360kan.com”下进行搜索查询，得到“360 影视”站点上的“葫芦娃”节目作为搜索结果提供给用户。

如图 2 所示，本发明的一个实施例中提供了一种搜索查询方法，其包括：

步骤 210，从预设的搜索日志中获取 URL 对应的第二查询关键词。在本实施例中，搜索日志可以是根据搜索引擎行为记录的日志，第二查询关键词即为历史查询关键词。

步骤 220，从 URL 对应的第二查询关键词中提取站点名称。在本实施例中，提取的站点名称可以为一个或多个。

步骤 230，根据第二查询关键词和站点名称进行训练得到限定规则，且在第一查询关键词符合限定规则时，得到限定站点的名称。在本实施例中，基于历史数据可以挖掘到限定规则，历史数据反映了用户的历史搜索行为，所以基于历史数据得到的限定规则对用户更加适用。具体地，还可以依据该 URL 是否被点击的数据来进行训练。

步骤 240，识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。

步骤 250，如符合适于在限定站点内进行搜索的限定规则，则按限定站点的名称，确定限定站点的域名，并在限定站点的域名下搜索第一查询关键词对应的搜索结果。在本实施例中，只要确定了站点的名称，那么站点的域名也就能够确定。

根据图 2，在搜索引擎的搜索日志中，假设存在了“www.360kan.com/jqm”的 url，则获取其对应的查询关键词“机器猫 360 影视”，从中提取出站点名称“360 影视”，则基于“机器猫 360 影视”和“360 影视”可以进行限定规则的训练，可以采用决策树方式训练。某用户输入了“葫芦娃 360 影视”进行搜索查询，搜索引擎判断其符合限定规则，

并判断限定站点名称为“360 影视”，则搜索引擎选择在“360 影视”的域名“www.360kan.com”下进行搜索查询，得到“360 影视”站点上的“葫芦娃”节目作为搜索结果提供给用户。

- 5 如图 3 所示，本发明的一个实施例中提供了一种搜索查询方法，其包括：
步骤 310，从预设的搜索日志中获取 URL 对应的第二查询关键词。
步骤 320，从 URL 对应的第二查询关键词中提取站点名称。
步骤 330，根据第二查询关键词和站点名称进行训练得到限定规则，且在第一查询关键词符合限定规则时，得到限定站点的名称。
- 10 步骤 340，从 URL 中提取域名。
步骤 350，建立所提取域名与站点名称之间的对应关系。在本实施例中，一个域名可以与多个站点名称之间建立对应关系。
步骤 360，识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。
- 15 步骤 370，如符合适于在限定站点内进行搜索的限定规则，则根据限定站点的名称和对应关系，查找限定站点的域名，并在限定站点的域名下搜索第一查询关键词对应的搜索结果。根据本实施例的技术方案，通过建立的对应关系，可以快速找到限定站点的域名。
- 根据图 3，在搜索引擎的搜索日志中，假设存在了
20 “www.360kan.com/jqm”的 url，则从中可提取域名“www.360kan.com”，并建立“360 影视”与“www.360kan.com”的对应关系。则某用户输入了“葫芦娃 360 影视”进行搜索查询，搜索引擎判断其符合限定规则，并判断限定站点名称为“360 影视”，则搜索引擎根据对应关系，可知需要在域名“www.360kan.com”下进行搜索查询，得到“360 影视”站点上的“葫芦娃”
25 节目作为搜索结果提供给用户。

- 如图 4 所示，本发明的一个实施例中提供了一种搜索查询方法，其包括：
步骤 410，从预设的搜索日志中获取 URL 对应的第二查询关键词。
步骤 420，从 URL 对应的第二查询关键词中提取站点名称。
- 30 步骤 430，对于每个提取的站点名称，根据每个站点名称出现在所述第二查询关键词中时对应的域名首页的被点击次数高低判断是否对其保留。根据本实施例的技术方案，由于从同一查询关键词提取的站点名称可能有多
个，因为有必要对其进行筛选保留，而对应域名首页点击次数越高则越说明
域名与站点名称的相关度更高，站点名称更有可能是正确的，有必要进行保
35 留。
- 步骤 440，根据第二查询关键词和站点名称进行训练得到限定规则，且在第一查询关键词符合限定规则时，得到限定站点的名称。
- 步骤 450，识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。

步骤 460, 如符合适于在限定站点内进行搜索的限定规则, 则按限定站点的名称, 确定限定站点的域名, 并在限定站点的域名下搜索第一查询关键词对应的搜索结果。

5 根据图 4, 对于从 “www.360kan.com/jqm” 提取的站点名称 “360 影视”, 判断在 “360 影视” 出现在查询关键词时, “www.360kan.com” 的被点击次数较高时则将 “360 影视” 作为 “www.360kan.com” 的站点名称。

如图 5 所示, 本发明的一个实施例中提供了一种搜索查询方法, 其包括:
步骤 510, 从预设的搜索日志中获取 URL 对应的第二查询关键词。

10 步骤 520, 从 URL 对应的第二查询关键词中提取站点名称。

步骤 530, 根据第二查询关键词和站点名称进行训练得到限定规则, 且在第一查询关键词符合限定规则时, 得到限定站点的名称。

步骤 540, 识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。

15 步骤 550, 如符合适于在限定站点内进行搜索的限定规则, 从第一查询关键词中识别出与限定站点的名称的对应内容。在本实施例中, 对应内容可以是与限定站点名称相同的内容, 或者是同义词, 或者是限定站点名称的拼音或英文对照。

步骤 560, 在限定站点的域名下, 按第一查询关键词中除对应内容外的
20 部分进行搜索。根据本实施例的技术方案, 因为用户输入的查询关键词中往往一部分用于限定站点, 而另一部分才用于反映其需求的资源, 所以通过本实施例可以合理将两部分划分以准确地进行搜索。

根据图 5, 某用户输入了 “葫芦娃 360yingshi” 进行搜索查询, 基于限定规则判断用户需求在 “360 影视” 站点上进行搜索, 则识别出与 “360 影
25 视” 对应的 “360yingshi”, 其中, “yingshi” 为 “影视” 的拼音, 则搜索引擎选择在 “360 影视” 的域名 “www.360kan.com” 下以 “葫芦娃” 为新的查询关键词进行搜索。

如图 6 所示, 本发明的一个实施例中提供了一种搜索查询方法, 其包括:

30 步骤 610, 从预设的搜索日志中获取 URL 对应的第二查询关键词。

步骤 620, 从 URL 对应的第二查询关键词中提取站点名称。

步骤 630, 根据第二查询关键词和站点名称进行训练得到限定规则, 且在第一查询关键词符合限定规则时, 得到限定站点的名称。

步骤 640, 识别用户输入的第一查询关键词是否符合预设的在限定站点
35 内进行搜索的限定规则。

步骤 650, 如符合适于在限定站点内进行搜索的限定规则, 对第一查询关键词进行分词得到多个词语, 并分别判断每个词语是否是限定站点名称对应的内容。在本实施例中, 可以采用现有的分词技术进行分词。而在分词之前, 还可以设置黑名单和白名单, 黑名单中可设置一些需要屏蔽掉的词语,

白名单中可设置一些返回固定结果的词语。同时，在分词时，还需要避免对一些保护词进行分词，例如，“钢铁是怎样炼成的”；还可以过滤掉一些词语，例如一些谓语。

步骤 660，在限定站点的域名下，按第一查询关键词中除对应内容外的部分进行搜索。

根据图 6，某用户输入了“葫芦娃 360 影视”进行搜索查询，基于限定规则判断用户需求在“360 影视”站点上进行搜索，对“葫芦娃 360 影视”分词后得到“葫芦娃”和“360 影视”，其中“360 影视”与站点对应，则搜索引擎选择在“360 影视”的域名“www.360kan.com”下以“葫芦娃”为新的查询关键词进行搜索。

如图 7 所示，本发明的一个实施例中提供了一种搜索查询装置，其包括：

第一查询关键词识别模块 710，识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。在本实施例中，对限定规则的类型不进行限制，例如，限定规则可以是：对多个站点的名称进行记录，如果第一查询关键词中包含了其中的站点名称，则判断需要在该站点下进行搜索。

搜索模块 720，如符合适于在限定站点内进行搜索的限定规则，则在限定站点的域名下搜索第一查询关键词对应的搜索结果。根据本实施例的技术方案，搜索查询的结果既不会包含非限定站点外的结果，也不会包含某些站点的首页；由于仅在限定站点上进行搜索，所以搜索产生的计算量也较小，也更容易避免非限定站点的作弊网页的干扰。

根据图 7，假设限定规则包含：如果用户输入的查询关键词中包含“360 影视”，则需要在“360 影视”站点中进行搜索。某用户输入了“葫芦娃 360 影视”进行搜索查询，其中包含了“360 影视”，则搜索引擎选择在“360 影视”的域名“www.360kan.com”下进行搜索查询，得到“360 影视”站点上的“葫芦娃”节目作为搜索结果提供给用户。

如图 8 所示，本发明的一个实施例中提供了一种搜索查询装置，其包括：

第二查询关键词获取模块 810，从预设的搜索日志中获取 URL 对应的第二查询关键词。在本实施例中，搜索日志可以是根据搜索引擎行为记录的日志，第二查询关键词即为历史查询关键词。

站点名称提取模块 820，从 URL 对应的第二查询关键词中提取站点名称。在本实施例中，提取的站点名称可以为一个或多个。

训练模块 830，根据第二查询关键词和站点名称进行训练得到限定规则，且在第一查询关键词符合限定规则时，得到限定站点的名称。在本实施例中，基于历史数据可以挖掘到限定规则，历史数据反映了用户的历史搜索行为，所以基于历史数据得到的限定规则对用户更加适用。具体地，还可以根据该 URL 是否被点击的数据来进行训练。

第一查询关键词识别模块 840，识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。

搜索模块 850，如符合适于在限定站点内进行搜索的限定规则，则按限定站点的名称，确定限定站点的域名，并在限定站点的域名下搜索第一查询关键词对应的搜索结果。在本实施例中，只要确定了站点的名称，那么站点的域名也就能够确定。

根据图 2，在搜索引擎的搜索日志中，假设存在了“www.360kan.com/jqm”的 url，则获取其对应的查询关键词“机器猫 360 影视”，从中提取出站点名称“360 影视”，则基于“机器猫 360 影视”和“360 影视”可以进行限定规则的训练，可以采用决策树方式训练。某用户输入了“葫芦娃 360 影视”进行搜索查询，搜索引擎判断其符合限定规则，并判断限定站点名称为“360 影视”，则搜索引擎选择在“360 影视”的域名“www.360kan.com”下进行搜索查询，得到“360 影视”站点上的“葫芦娃”节目作为搜索结果提供给用户。

如图 9 所示，本发明的一个实施例中提供了一种搜索查询装置，其包括：第二查询关键词获取模块 910，从预设的搜索日志中获取 URL 对应的第二查询关键词。

站点名称提取模块 920，从 URL 对应的第二查询关键词中提取站点名称。

训练模块 930，根据第二查询关键词和站点名称进行训练得到限定规则，且在第一查询关键词符合限定规则时，得到限定站点的名称。

域名提取模块 940，从 URL 中提取域名。

对应关系建立模块 950，建立所提取域名与站点名称之间的对应关系。

在本实施例中，一个域名可以与多个站点名称之间建立对应关系。

第一查询关键词识别模块 960，识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。

搜索模块 970，如符合适于在限定站点内进行搜索的限定规则，则根据限定站点的名称和对应关系，查找限定站点的域名，并在限定站点的域名下搜索第一查询关键词对应的搜索结果。根据本实施例的技术方案，通过建立的对对应关系，可以快速找到限定站点的域名。

根据图 9，在搜索引擎的搜索日志中，假设存在了“www.360kan.com/jqm”的 url，则从中可提取域名“www.360kan.com”，并建立“360 影视”与“www.360kan.com”的对应关系。则某用户输入了“葫芦娃 360 影视”进行搜索查询，搜索引擎判断其符合限定规则，并判断限定站点名称为“360 影视”，则搜索引擎根据对应关系，可知需要在域名“www.360kan.com”下进行搜索查询，得到“360 影视”站点上的“葫芦娃”节目作为搜索结果提供给用户。

如图 8 所示,本发明的一个实施例中提供了一种搜索查询装置,其包括:
第二查询关键词提取模块 810,从预设的搜索日志中获取 URL 对应的
第二查询关键词。

5 站点名称提取模块 820,从 URL 对应的第二查询关键词中提取站点名称。

站点名称提取模块 820 对于每个提取的站点名称,根据每个站点名称出现在所述第二查询关键词中时对应的域名首页的被点击次数高低判断是否对其保留。根据本实施例的技术方案,由于从同一查询关键词提取的站点名称可能有多个,因为有必要对其进行筛选保留,而对应域名首页点击次数越高则越说明域名与站点名称的相关度更高,站点名称更有可能是正确的,有
10 必要进行保留。

训练模块 830,根据第二查询关键词和站点名称进行训练得到限定规则,且在第一查询关键词符合限定规则时,得到限定站点的名称。

15 第一查询关键词识别模块 840,识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。

搜索模块 850,如符合适于在限定站点内进行搜索的限定规则,则按限定站点的名称,确定限定站点的域名,并在限定站点的域名下搜索第一查询关键词对应的搜索结果。

20 根据图 8,对于从“www.360kan.com/jqm”提取的站点名称“360 影视”,判断在“360 影视”出现在查询关键词时,“www.360kan.com”的被点击次数较高时则将“360 影视”作为“www.360kan.com”的站点名称。

如图 8 所示,本发明的一个实施例中提供了一种搜索查询装置,其包括:

25 第二查询关键词获取模块 810,从预设的搜索日志中获取 URL 对应的第二查询关键词。

站点名称提取模块 820,从 URL 对应的第二查询关键词中提取站点名称。

训练模块 830,根据第二查询关键词和站点名称进行训练得到限定规则,且在第一查询关键词符合限定规则时,得到限定站点的名称。

30 第一查询关键词识别模块 840,识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则。

搜索模块 850,如符合适于在限定站点内进行搜索的限定规则,从第一查询关键词中识别出与限定站点的名称的对应内容。在本实施例中,对应内容可以是与限定站点名称相同的内容,或者是同义词,或者是限定站点名称
35 的拼音或英文对照。

搜索模块 850 在限定站点的域名下,按第一查询关键词中除对应内容外的部分进行搜索。根据本实施例的技术方案,因为用户输入的查询关键词中往往一部分用于限定站点,而另一部分才用于反映其需求的资源,所以通过本实施例可以合理将两部分划分以准确地进行搜索。

根据图 8, 某用户输入了“葫芦娃 360yingshi”进行搜索查询, 基于限定规则判断用户需求在“360 影视”站点上进行搜索, 则识别出与“360 影视”对应的“360yingshi”, 其中, “yingshi”为“影视”的拼音, 则搜索引擎选择在“360 影视”的域名“www.360kan.com”下以“葫芦娃”为新的
5 查询关键词进行搜索。

如图 8 所示, 本发明的一个实施例中提供了一种搜索查询装置, 其包括:
第二查询关键词获取模块 810, 从预设的搜索日志中获取 URL 对应的
第二查询关键词。

10 站点名称提取模块 820, 从 URL 对应的第二查询关键词中提取站点名称。

训练模块 830, 根据第二查询关键词和站点名称进行训练得到限定规则, 且在第一查询关键词符合限定规则时, 得到限定站点的名称。

15 第一查询关键词识别模块 840, 识别用户输入的第一查询关键词是否符合预设的在限定点内进行搜索的限定规则。

搜索模块 850, 如符合适于在限定站点内进行搜索的限定规则, 对第一查询关键词进行分词得到多个词语, 并分别判断每个词语是否是限定站点名称对应的内容。在本实施例中, 可以采用现有的分词技术进行分词。而在分词之前, 还可以设置黑名单和白名单, 黑名单中可设置一些需要屏蔽掉的
20 词语, 白名单中可设置一些返回固定结果的词语。同时, 在分词时, 还需要避免对一些保护词进行分词, 例如, “钢铁是怎样炼成的”; 还可以过滤掉一些词语, 例如一些谓语。

搜索模块 850 在限定站点的域名下, 按第一查询关键词中除对应内容外的部分进行搜索。

25 根据图 6, 某用户输入了“葫芦娃 360 影视”进行搜索查询, 基于限定规则判断用户需求在“360 影视”站点上进行搜索, 对“葫芦娃 360 影视”分词后得到“葫芦娃”和“360 影视”, 其中“360 影视”与站点对应, 则搜索引擎选择在“360 影视”的域名“www.360kan.com”下以“葫芦娃”为新的查询关键词进行搜索。

30 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。

35 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。

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明的单独实施例。

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

10 可以把它们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

15 类似目的的替代特征来代替。

此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

20 用。

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搜索查询装置中的一些或者全部部件的一些或者全部功能。

25 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 10 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 1010 和以存储器 1020 形式的计算机程序产品或者计算机可读介质。存储器 1020 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 1020 具有用于执行上述方法中的任何方法步骤的程序代码 1031 的存储空间 1030。例如，用于程序代码的存储空间 1030 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 1031。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 11 所述的便

30

35

携式或者固定存储单元。该存储单元可以具有与图 10 的计算设备中的存储器 1020 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 1031'，即可以由例如诸如 1010 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。

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。

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机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。

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、一种搜索查询方法，其包括：

5 识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则；

如符合适于在限定站点内进行搜索的限定规则，则在所述限定站点的域名下搜索所述第一查询关键词对应的搜索结果。

2、根据权利要求 1 所述的方法，其中，在识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则之前，还包括：

10 从预设的搜索日志中获取 URL 对应的第二查询关键词；

从所述 URL 对应的第二查询关键词中提取站点名称；

根据所述第二查询关键词和所述站点名称进行训练得到所述限定规则，且在所述第一查询关键词符合所述限定规则时，得到所述限定站点的名称；

15 在所述限定站点的域名下搜索所述第一查询关键词对应的搜索结果之前，还包括：

按所述限定站点的名称，确定所述限定站点的域名。

3、根据权利要求 2 所述的方法，其中，在识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则之前，还包括：

从所述 URL 中提取域名；

20 建立所提取域名与所述站点名称之间的对应关系；

按所述限定站点的名称，确定所述限定站点的域名，具体包括：

根据所述限定站点的名称和所述对应关系，查找所述限定站点的域名。

4、根据权利要求 2 所述的方法，其中，从所述 URL 对应的第二查询关键词中提取站点名称，还包括：

25 对于每个提取的站点名称，根据每个站点名称出现在所述第二查询关键词中时对应的域名首页的被点击次数高低判断是否对其保留。

5、根据权利要求 2 所述的方法，其中，在所述限定站点的域名下搜索所述第一查询关键词对应的搜索结果，具体包括：

从所述第一查询关键词中识别出与所述限定站点的名称的对应内容；

30 在所述限定站点的域名下，按所述第一查询关键词中除所述对应内容外的部分进行搜索。

6、根据权利要求 5 所述的方法，其中，从所述第一查询关键词中识别出与所述限定站点的名称的对应内容，具体包括：

对所述第一查询关键词进行分词得到多个词语；

35 分别判断每个词语是否是与所述限定站点名称对应的内容。

7、一种搜索查询装置，其包括：

第一查询关键词识别模块，用于识别用户输入的第一查询关键词是否符合预设的在限定站点内进行搜索的限定规则；

搜索模块,如符合适于在限定站点内进行搜索的限定规则,则在所述限定站点的域名下搜索所述第一查询关键词对应的搜索结果。

8、根据权利要求7所述的装置,其中,还包括:

5 第二查询关键词获取模块,用于从预设的搜索日志中获取 URL 对应的第二查询关键词;

站点名称提取模块,用于从所述 URL 对应的第二查询关键词中提取站点名称;

10 训练模块,用于根据所述第二查询关键词和所述站点名称进行训练得到所述限定规则,且在所述第一查询关键词符合所述限定规则时,得到所述限定站点的名称;

域名确定模块,用于按所述限定站点的名称,确定所述限定站点的域名。

9、根据权利要求8所述的装置,其中,还包括:

域名提取模块,用于从所述 URL 中提取域名;

15 对应关系建立模块,用于建立所提取域名与所述站点名称之间的对应关系;

所述域名确定模块根据所述限定站点的名称和所述对应关系,查找所述限定站点的域名。

10、根据权利要求8所述的装置,其中,

20 所述站点名称提取模块对于每个提取的站点名称,根据每个站点名称出现在所述第二查询关键词中时对应的域名首页的被点击次数高低判断是否对其保留。

11、根据权利要求8所述的装置,其中,

25 所述搜索模块从所述第一查询关键词中识别出与所述限定站点的名称的对应内容,并在所述限定站点的域名下,按所述第一查询关键词中除所述对应内容外的部分进行搜索。

12、根据权利要求11所述的装置,其中,

所述搜索模块对所述第一查询关键词进行分词得到多个词语,并分别判断每个词语是否是与所述限定站点名称对应的内容。

30 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中的任一项所述的搜索查询方法。

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

1/9

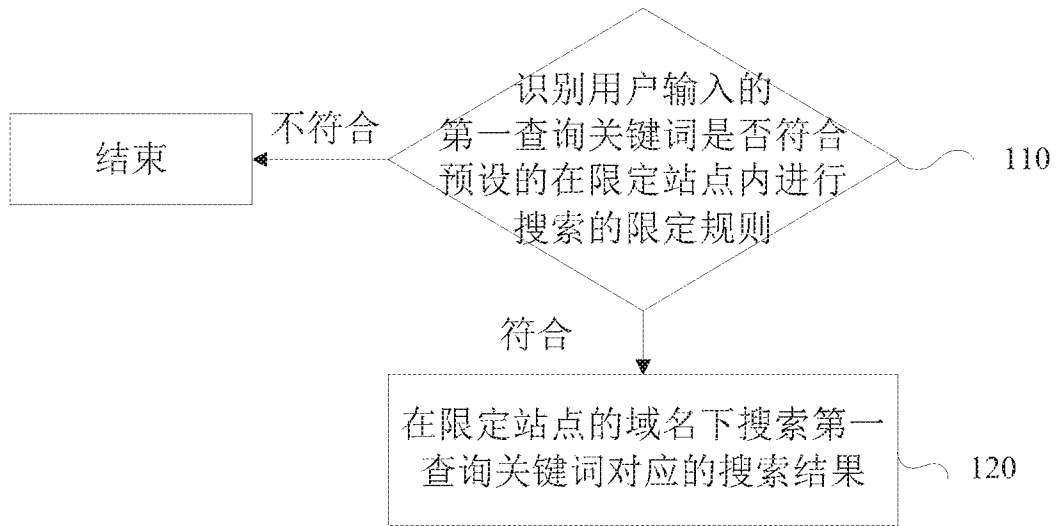


图 1

2/9

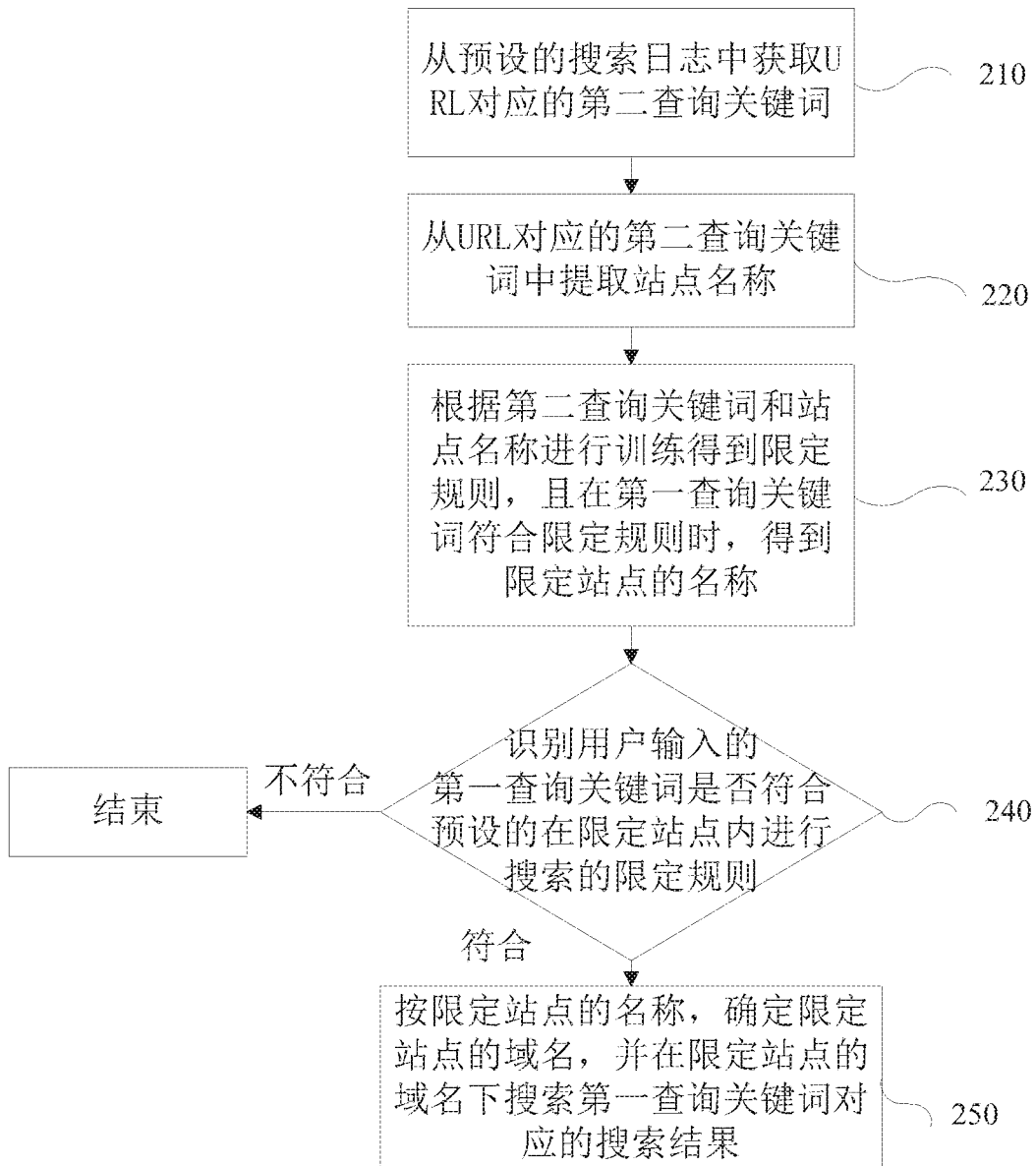


图 2

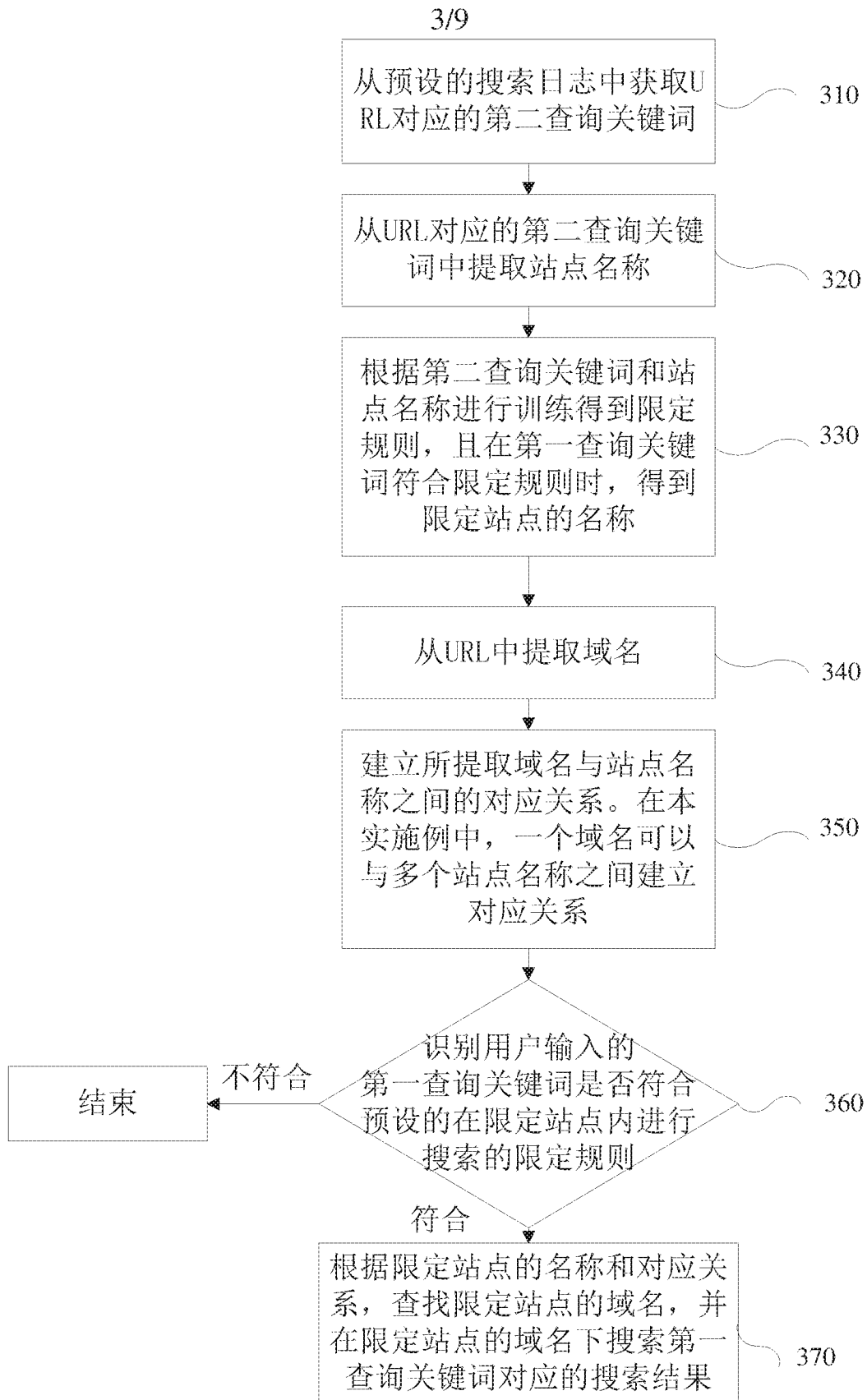


图 3

4/9

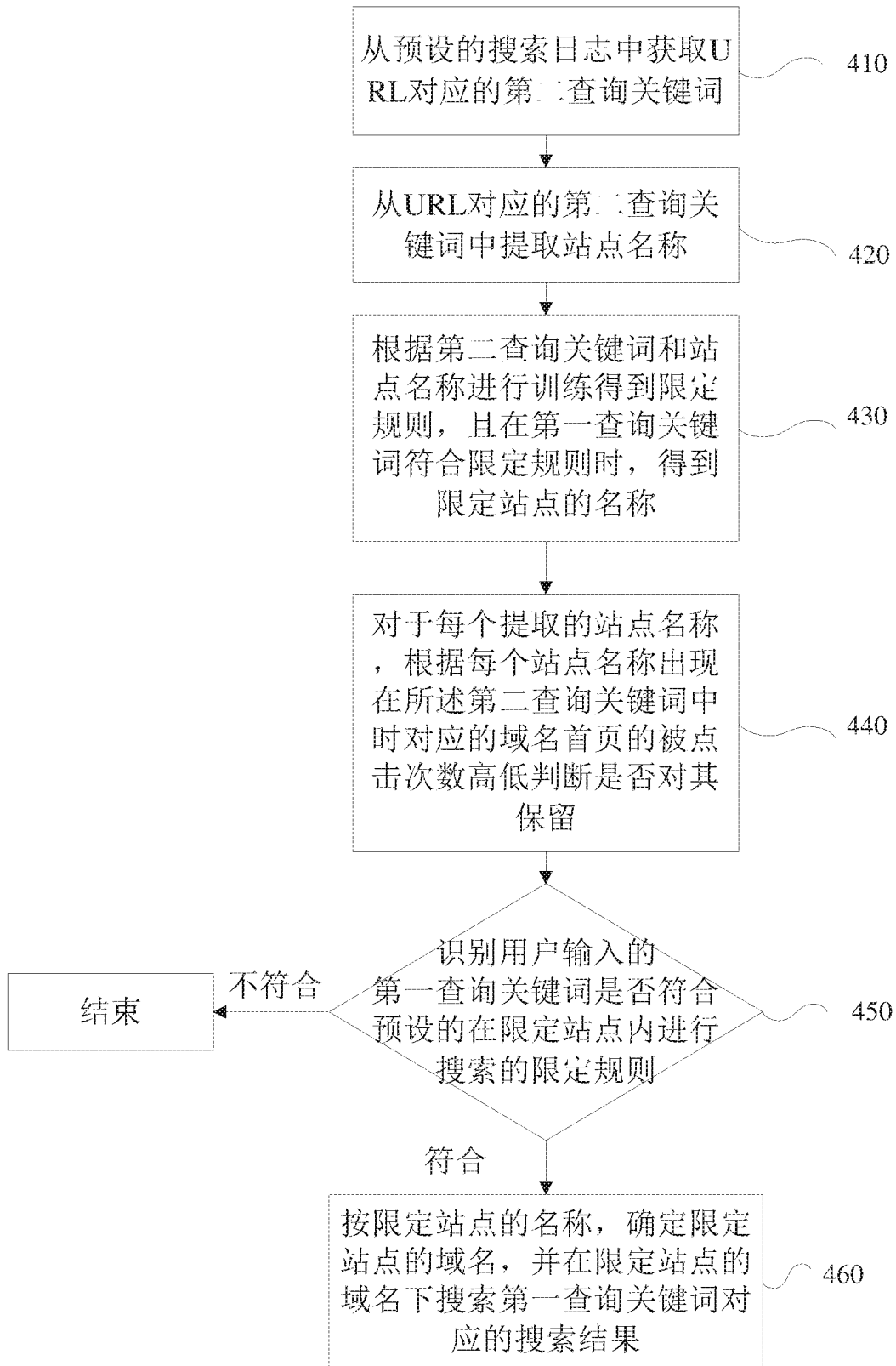


图 4

5/9

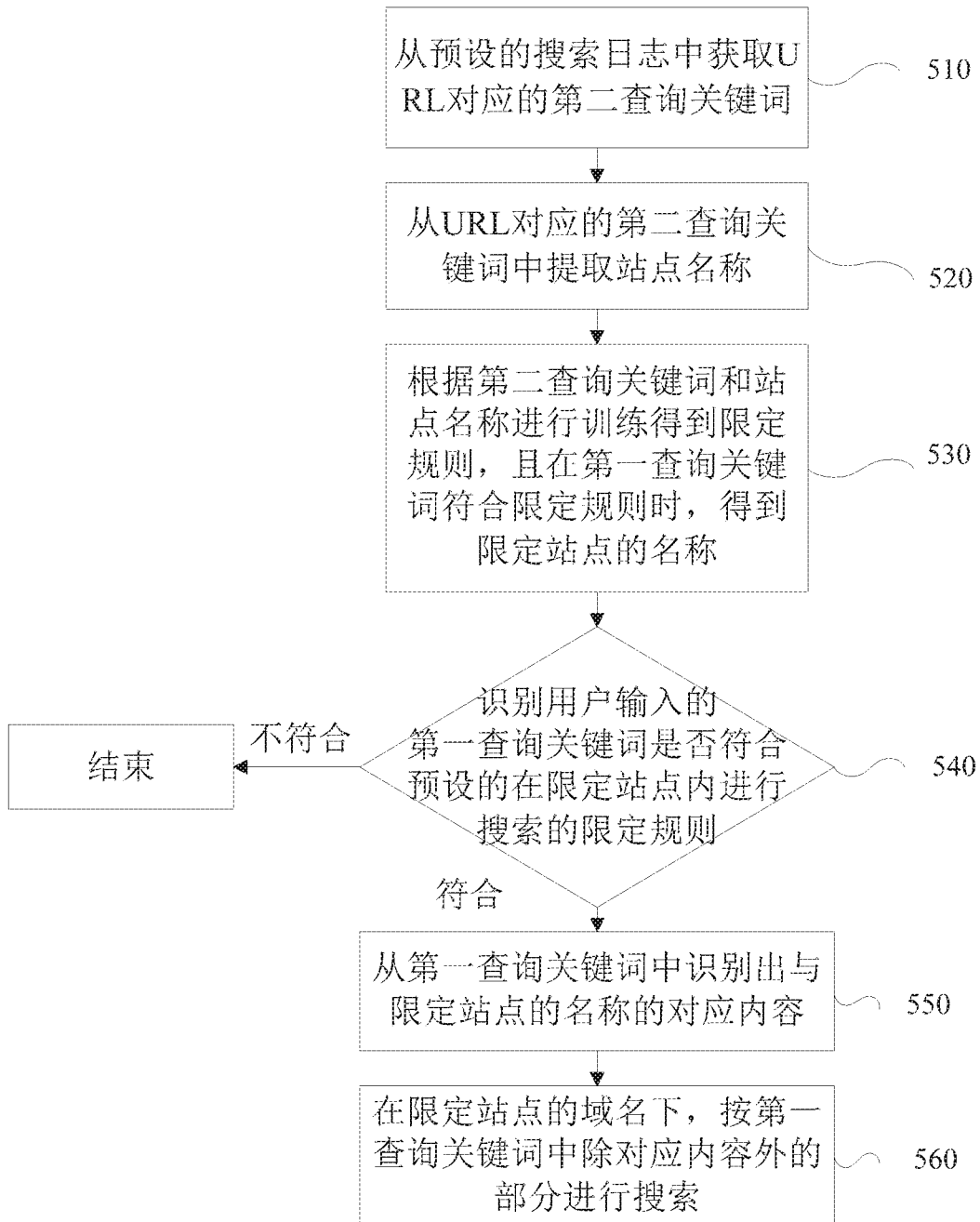


图 5

6/9

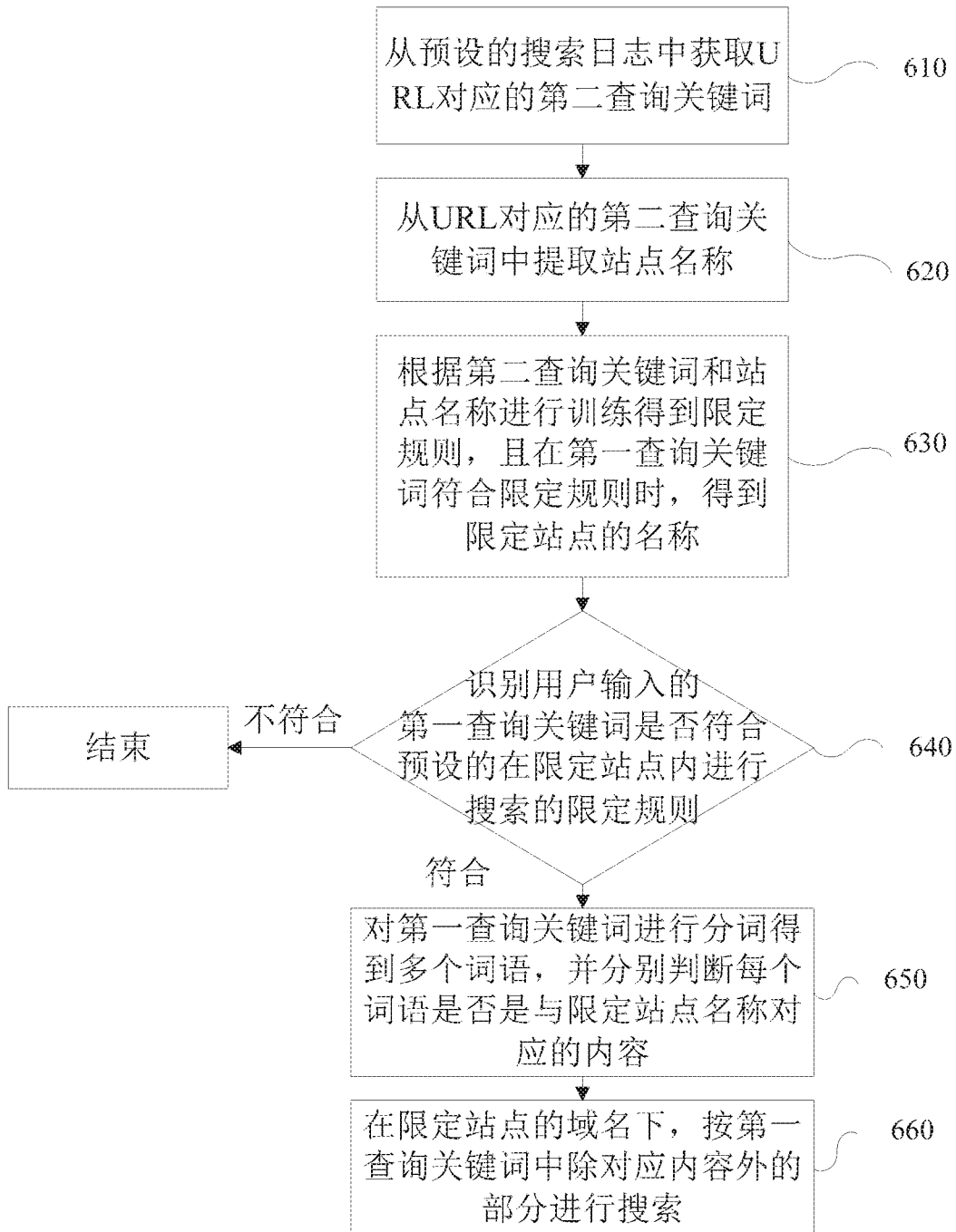


图 6

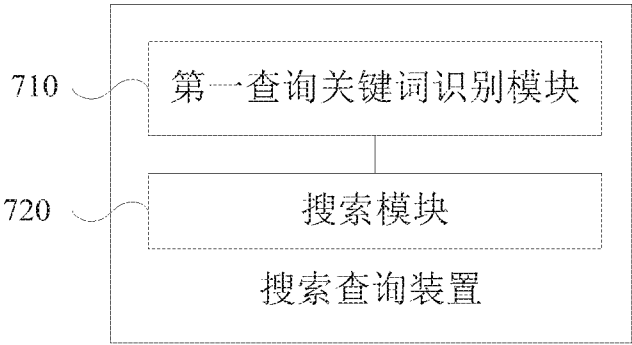


图 7

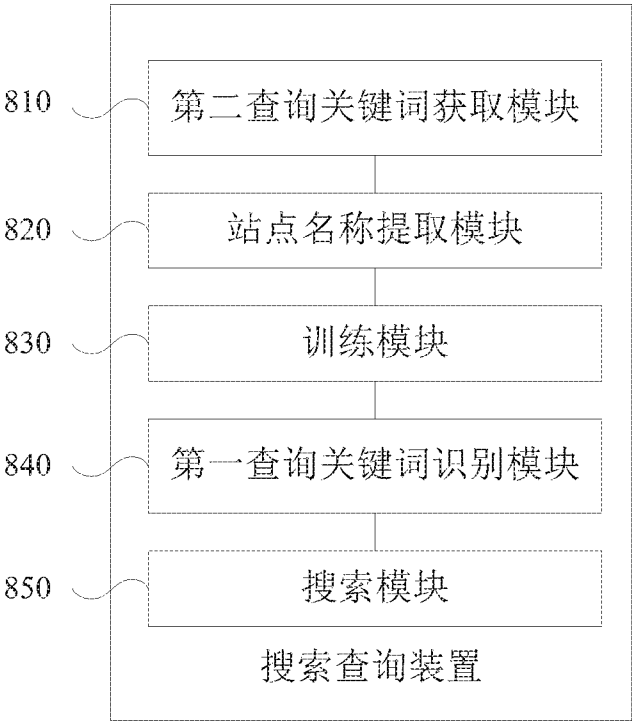


图 8

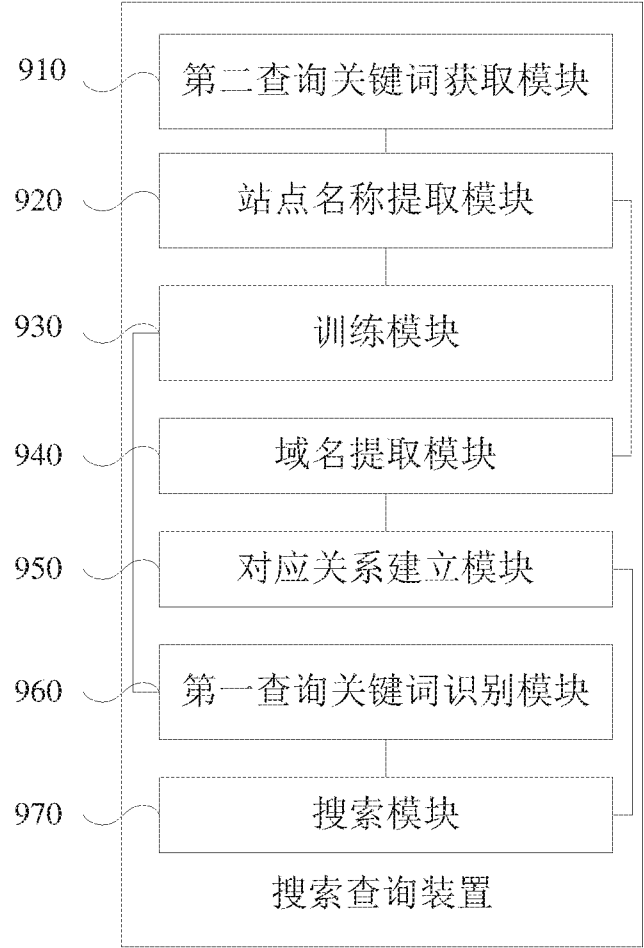


图 9

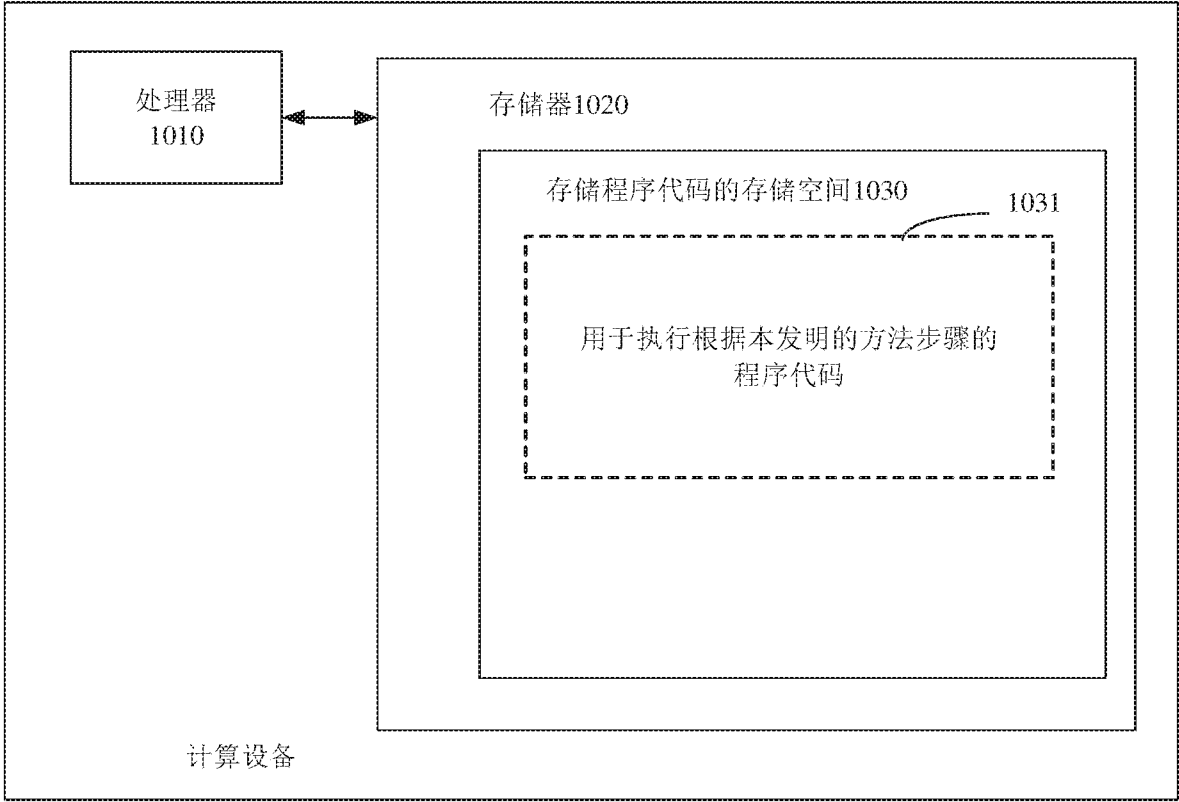


图10

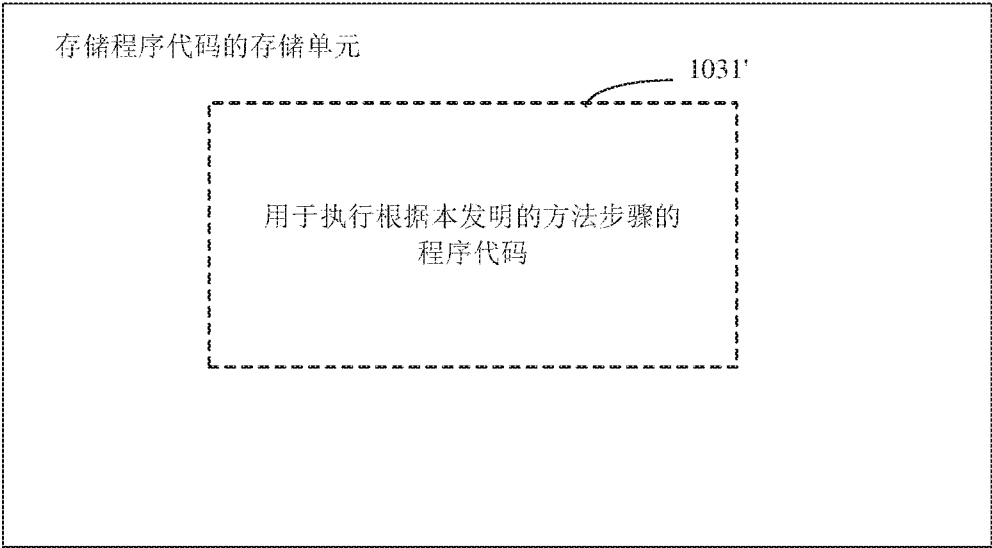


图11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/095018**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)

VEN, CNABS, CNTXT, CNKI: site name, name, URL, rule, decision tree, log, history, input, search, query, keyword, domain name, web site, site, station, train

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102982150 A (PAN, Yanhui et al.), 20 March 2013 (20.03.2013), description, paragraphs [0003]-[0024]	1, 7, 13, 14
X	CN 102947824 A (NITIN, D.A.), 27 February 2013 (27.02.2013), description, paragraphs [0018]-[0020] and [0028]-[0077], and figures 1-4	1, 7, 13, 14
PX	CN 104462519 A (BEIJING QIHOO TECHNOLOGY CO., LTD. et al.), 25 March 2015 (25.03.2015), claims 1-12, and description, paragraph [0125]	1-14
A	US 2005060296 A1 (WHITBY, D.S. et al.), 17 March 2005 (17.03.2005), the whole document	1-14
A	CN 104123366 A (XIE, Jianping), 29 October 2014 (29.10.2014), 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 12 January 2016 (12.01.2016)	Date of mailing of the international search report 22 January 2016 (22.01.2016)
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 WANG, Yue Telephone No.: (86-10) 62089109

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/095018

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102982150 A	20 March 2013	None	
CN 102947824 A	27 February 2013	US 2011307484 A1	15 December 2011
		EP 2616964 A2	24 July 2013
		WO 2011154974 A2	15 December 2011
		WO 2011154974 A3	02 February 2012
CN 104462519 A	25 March 2015	None	
US 2005060296 A1	17 March 2005	None	
CN 104123366 A	29 October 2014	None	

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)) VEN, CNABS, CNTXT, CNKI: 搜索, 查询, 检索, 关键词, 域名, 网站, 站名, 名称, URL, 规则, 训练, 决策树, 日志, 历史, 输入, search, query, keyword, domain name, web site, site, station, train		
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 102982150 A (潘燕辉 等) 2013年 3月 20日 (2013 - 03 - 20) 说明书第[0003]-[0024]段	1, 7, 13, 14
X	CN 102947824 A (迪内希 阿南德 尼丁) 2013年 2月 27日 (2013 - 02 - 27) 说明书第[0018]-[0020], [0028]-[0077]段、图1-4	1, 7, 13, 14
PX	CN 104462519 A (北京奇虎科技有限公司 等) 2015年 3月 25日 (2015 - 03 - 25) 权利要求1-12、说明书第[0125]段	1-14
A	US 2005060296 A1 (WHITBY D S 等) 2005年 3月 17日 (2005 - 03 - 17) 全文	1-14
A	CN 104123366 A (谢建平) 2014年 10月 29日 (2014 - 10 - 29) 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 1月 12日		国际检索报告邮寄日期 2016年 1月 22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 王越 电话号码 (86-10)62089109

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/095018

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	102982150	A	2013年 3月 20日	无			
CN	102947824	A	2013年 2月 27日	US	2011307484	A1	2011年 12月 15日
				EP	2616964	A2	2013年 7月 24日
				WO	2011154974	A2	2011年 12月 15日
				WO	2011154974	A3	2012年 2月 2日
CN	104462519	A	2015年 3月 25日	无			
US	2005060296	A1	2005年 3月 17日	无			
CN	104123366	A	2014年 10月 29日	无			