

(12) 发明专利申请

(10) 申请公布号 CN 103425744 A

(43) 申请公布日 2013. 12. 04

(21) 申请号 201310301230. 7

(22) 申请日 2013. 07. 17

(71) 申请人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街 10 号
百度大厦

(72) 发明人 陈洪亮 呼大为

(74) 专利代理机构 北京汉昊知识产权代理事务
所(普通合伙) 11370
代理人 罗朋 周建华

(51) Int. Cl.
G06F 17/30(2006. 01)

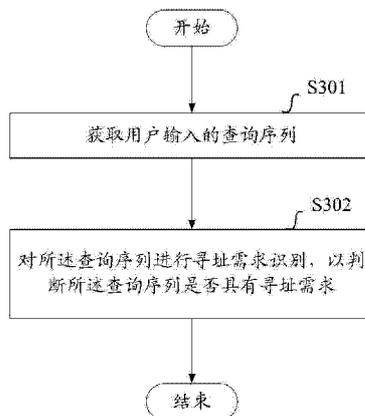
权利要求书3页 说明书12页 附图2页

(54) 发明名称

一种用于识别用户的查询序列中的寻址需求的方法与设备

(57) 摘要

本发明的目的是提供一种用于识别用户的查询序列中的寻址需求的方法与设备;获取用户输入的查询序列;对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。与现有技术相比,本发明获取用户输入的查询序列,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求,准确、高效地识别用户的查询序列中的寻址需求;进一步地,根据识别出的寻址查询序列,将直接满足用户需求的页面进行加权,提高了搜索结果展示效果,提升了用户搜索体验。



1. 一种用于识别用户的查询序列中的寻址需求的方法,其中,该方法包括以下步骤:
 - a 获取用户输入的查询序列;
 - b 对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。
2. 根据权利要求 1 所述的方法,其中,所述步骤 b 包括:
 - 基于寻址词库,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;其中,该方法还包括:
 - 基于查询用户的历史查询记录与查询点击行为,确定具有寻址需求的查询词;
 - 根据所述查询词,建立或更新所述寻址词库。
3. 根据权利要求 2 所述的方法,其中,所述步骤 b 包括:
 - 基于寻址词库,对所述查询序列进行匹配处理,以判断所述查询序列是否具有寻址需求;其中,所述匹配处理的方式包括以下至少任一项:
 - 前缀匹配;
 - 后缀匹配;
 - 中间词匹配。
4. 根据权利要求 1 至 3 中任一项所述的方法,其中,所述步骤 b 包括:
 - b1 基于专名词库,对所述查询序列进行专名识别;
 - b2 当所述查询序列包括专名词组,则判断所述查询序列具有寻址需求。
5. 根据权利要求 4 所述的方法,其中,所述步骤 b2 包括:
 - 当所述查询序列包括专名词组,进一步判断所述查询序列是否包括预定字符串;
 - 当所述查询序列包括所述预定字符串,则判断所述查询序列具有寻址需求。
6. 根据权利要求 1 至 5 中任一项所述的方法,其中,该方法还包括:
 - 基于预置黑名单匹配,对所述查询序列进行过滤处理,以过滤掉范需求查询序列,获得过滤处理后的查询序列;其中,所述步骤 b 包括:
 - 对所述过滤处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。
7. 根据权利要求 1 至 6 中任一项所述的方法,其中,该方法还包括:
 - 对所述查询序列进行预处理,以获得预处理后的查询序列;其中,所述步骤 b 包括:
 - 对所述预处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;
8. 根据权利要求 1 至 7 中任一项所述的方法,其中,该方法还包括:
 - 当所述查询序列具有寻址需求,提高与所述寻址需求相对应的地址查询结果的结果

排序值。

9. 一种用于识别用户的查询序列中的寻址需求的设备,其中,该设备包括:

获取装置,用于获取用户输入的查询序列;

识别装置,用于对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。

10. 根据权利要求9所述的设备,其中,所述识别装置用于:

- 基于寻址词库,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;

其中,该设备还包括更新装置,用于:

- 基于查询用户的历史查询记录与查询点击行为,确定具有寻址需求的查询词;

- 根据所述查询词,建立或更新所述寻址词库。

11. 根据权利要求10所述的设备,其中,所述识别装置用于:

- 基于寻址词库,对所述查询序列进行匹配处理,以判断所述查询序列是否具有寻址需求;

其中,所述匹配处理的方式包括以下至少任一项:

- 前缀匹配;

- 后缀匹配;

- 中间词匹配。

12. 根据权利要求9至11中任一项所述的设备,其中,所述识别装置包括:

识别单元,用于基于专名词库,对所述查询序列进行专名识别;

判断单元,用于当所述查询序列包括专名词组,则判断所述查询序列具有寻址需求。

13. 根据权利要求12所述的设备,其中,所述判断单元用于:

- 当所述查询序列包括专名词组,进一步判断所述查询序列是否包括预定字符串;

- 当所述查询序列包括所述预定字符串,则判断所述查询序列具有寻址需求。

14. 根据权利要求9至13中任一项所述的设备,其中,该设备还包括:

过滤装置,用于基于预置黑名单匹配,对所述查询序列进行过滤处理,以过滤掉范需求查询序列,获得过滤处理后的查询序列;

其中,所述识别装置用于:

- 对所述过滤处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。

15. 根据权利要求9至14中任一项所述的设备,其中,该设备还包括:

预处理装置,用于对所述查询序列进行预处理,以获得预处理后的查询序列;

其中,所述识别装置用于:

- 对所述预处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;

其中,所述预处理包括以下至少任一项:

- 切词处理;

- 无效词过滤处理;

- 同义替换处理。

16. 根据权利要求 9 至 15 中任一项所述的设备,其中,该设备还包括:
排序装置,用于当所述查询序列具有寻址需求,提高与所述寻址需求相对应的地址查询结果的结果排序值。

一种用于识别用户的查询序列中的寻址需求的方法与设备

技术领域

[0001] 本发明涉及搜索技术领域,尤其涉及一种用于识别用户的查询序列中的寻址需求的技术。

背景技术

[0002] 如今,搜索引擎已经成为越来越多人获取信息的工具。从技术的角度,如何能快速满足用户的需求成为关注的焦点。想更好的满足用户,首先应该清楚用户的需求。用户的需求大体上可以分为:定向需求和范需求。所谓定向需求,是指用户输入查询序列的目标是想找到特定页面或者是特定页面能够很好的满足用户的需求,比如用户在搜索框中输入“京东商城”,不难发现直接返回“www. jd. com”就能很好的满足用户。所谓范需求,是指用户的需求不是很具体明确,想得到信息比较宽泛,只要与需求相关即可,比如用户输入查询序列“哪个商城好”,用户主要是想比较下商城的优劣,而不是想直接去某个商城。

[0003] 我们将具有定向需求的查询序列称为寻址查询序列,意思是用户输入查询序列的目的是寻找特定网址。寻址查询序列占每日总查询序列的15%,因此,如何准确、高效地识别用户的查询序列中的寻址需求,成为本领域技术人员亟需解决的问题之一。

发明内容

[0004] 本发明的目的是提供一种用于识别用户的查询序列中的寻址需求的方法与设备。

[0005] 根据本发明的一个方面,提供了一种用于识别用户的查询序列中的寻址需求的方法,其中,该方法包括以下步骤:

[0006] a 获取用户输入的查询序列;

[0007] b 对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。

[0008] 根据本发明的另一方面,还提供了一种用于识别用户的查询序列中的寻址需求的设备,其中,该设备包括:

[0009] 获取装置,用于获取用户输入的查询序列;

[0010] 识别装置,用于对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。

[0011] 与现有技术相比,本发明获取用户输入的查询序列,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求,准确、高效地识别用户的查询序列中的寻址需求;进一步地,根据识别出的寻址查询序列,将直接满足用户需求的页面进行加权,提高了搜索结果展示效果,提升了用户搜索体验。

[0012] 进一步地,本发明基于寻址词库与专名词库,通过查询序列分析和查询记录挖掘技术,以及优质资源扩充,从查询序列的语法组成分析用户的寻址意图,同时也从资源满足的维度来反向识别寻址查询序列,从正反两个方向对查询序列进行寻址需求识别,既保证了识别的准确性,同时又增加了召回。

附图说明

[0013] 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,本发明的其它特征、目的和优点将会变得更明显:

[0014] 图 1 示出根据本发明一个方面的用于识别用户的查询序列中的寻址需求的设备示意图;

[0015] 图 2 示出根据本发明一个优选实施例的用于识别用户的查询序列中的寻址需求的设备示意图;

[0016] 图 3 示出根据本发明另一个方面的用于识别用户的查询序列中的寻址需求的方法流程图;

[0017] 图 4 示出根据本发明一个优选实施例的用于识别用户的查询序列中的寻址需求的方法流程图。

[0018] 附图中相同或相似的附图标记代表相同或相似的部件。

具体实施方式

[0019] 下面结合附图对本发明作进一步详细描述。

[0020] 图 1 示出根据本发明一个方面的用于识别用户的查询序列中的寻址需求的设备示意图。设备 1 包括获取装置 101 和识别装置 102。

[0021] 其中,获取装置 101 获取用户输入的查询序列。具体地,获取装置 101 通过与用户的用户设备的交互,例如,通过一次或多次调用该用户设备提供的应用程序接口(API)或其他约定的通信方式,获取该用户输入的查询序列;或者,通过 ASP、JSP 或 PHP 等页面技术,获取用户在搜索框、查询框等输入界面输入的查询序列;又或者,该获取装置 101 通过与搜索引擎等第三方设备的交互,获取该用户输入的查询序列。

[0022] 本领域技术人员应能理解上述获取查询序列的方式仅为举例,其他现有的或今后可能出现的获取查询序列的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0023] 识别装置 102 对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。具体地,识别装置 102 对该查询序列进行寻址需求识别的方式包括但不限于:

[0024] 1) 基于寻址词库,对该获取装置 101 所获取的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;

[0025] 2) 基于寻址词库,对该获取装置 101 所获取的查询序列进行诸如前缀匹配、后缀匹配、中间词匹配等匹配处理,以判断所述查询序列是否具有寻址需求;

[0026] 3) 基于专名词库,对所述查询序列进行专名识别,当所述查询序列包括专名词组,则判断所述查询序列具有寻址需求;

[0027] 4) 基于专名词库,对所述查询序列进行专名识别,当所述查询序列包括专名词组,进一步判断所述查询序列是否包括预定字符串,当所述查询序列包括所述预定字符串,则判断所述查询序列具有寻址需求。

[0028] 本领域技术人员应能理解上述识别寻址需求的方式仅为举例,其他现有的或今后可能出现的识别寻址需求的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0029] 在此,设备 1 获取用户输入的查询序列,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求,准确、高效地识别用户的查询序列中的寻址需求。

[0030] 较佳地,该识别装置 102 基于所述寻址词库与专名词库,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求,其中,所述寻址词库基于查询用户的历史查询记录与查询点击行为建立或更新。

[0031] 在此,设备 1 基于寻址词库与专名词库,通过查询序列分析和查询记录挖掘技术,以及优质资源扩充,从查询序列的语法组成分析用户的寻址意图,同时也从资源满足的维度来反向识别寻址查询序列,从正反两个方向对查询序列进行寻址需求识别,既保证了识别的准确性,同时又增加了召回。

[0032] 优选地,设备 1 的各个装置之间是持续不断工作的。具体地,获取装置 101 获取用户输入的查询序列;识别装置 102 对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。在此,本领域技术人员应理解“持续”是指设备 1 的各装置分别按照设定的或实时调整的工作模式要求进行查询序列的获取及寻址需求的识别,直至该设备 1 在较长时间内停止获取用户输入的查询序列。

[0033] 优选地,所述识别装置 102 基于寻址词库,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;其中,该设备还包括更新装置(未示出),该更新装置基于查询用户的历史查询记录与查询点击行为,确定具有寻址需求的查询词;根据所述查询词,建立或更新所述寻址词库。具体地,对于网络中的查询用户,其历史输入了查询序列,将该历史输入的查询序列记为该查询用户的历史查询记录;对于每个查询序列,搜索引擎都会返回一定的查询结果,查询用户对其中的某个查询结果进行点击查看,将该点击查看的查询结果记为该查询用户的查询点击行为。例如,查询用户历史输入了查询序列“京东”,并且,在返回的多个查询结果中,点击进入了网址 [http://www. jd. com/](http://www.jd.com/),则该设备 1 记录该查询用户的历史查询记录与查询点击行为。特别地,对于多个查询用户的历史查询记录与查询点击行为,更新装置可确定具有寻址需求的查询词。例如,假设在预定时间段内,有超过预定比例阈值的查询用户在输入查询序列“京东”的情况下,点击进入了网址 <http://www. jd. com/>,则该更新装置将“京东”确定为具有寻址需求的查询词。进而,该更新装置通过将该确定的具有寻址需求的查询词存入寻址词库,建立或更新该寻址词库。基于该寻址词库,识别装置 102 对该获取装置 101 所获取的查询序列进行寻址需求识别,以判断该查询序列是否具有寻址需求。例如,当该查询序列命中该寻址词库时,判断该查询序列具有寻址需求。

[0034] 更优选地,所述识别装置 102 基于寻址词库,对所述查询序列进行匹配处理,以判断所述查询序列是否具有寻址需求;其中,所述匹配处理的方式包括以下至少任一项:

[0035] - 前缀匹配;

[0036] - 后缀匹配;

[0037] - 中间词匹配。

[0038] 具体地,识别装置 102 基于寻址词库,对获取装置 101 所获取的查询序列进行前缀匹配、后缀匹配、中间词匹配等匹配处理,以判断该查询序列是否具有寻址需求,例如,当该查询序列的前缀部分、后缀部分或中间词部分命中该寻址词库时,判断该查询序列具有寻址需求。在此,前缀匹配为该用户输入的查询序列的前缀命中该寻址词库;后缀匹配为该查

询序列的后缀部分命中该寻址词库；同理，中间词匹配为该查询序列的中间词部分命中该寻址词库。

[0039] 较佳地，该识别装置 102 基于寻址词库，对该查询序列进行匹配处理之后，还可进一步对该查询序列的剩余部分进行识别处理，以判断该查询序列是否具有寻址需求。例如，对于用户输入的查询序列“京东商城官网”，识别装置 102 基于寻址词库，对该查询序列进行前缀匹配，判断该查询序列的前缀部分“京东商城”命中了该寻址词库，对于该查询序列的剩余部分“官网”，该识别装置 102 进一步对其进行识别处理，识别其同样具有寻址需求，据此，该识别装置 102 判断该查询序列“京东商城官网”具有寻址需求。而对于用户输入的查询序列“京东商城好不好”，虽然识别装置 102 基于寻址词库，判断该查询序列的前缀部分“京东商城”命中了该寻址词库，然而，对于该查询序列的剩余部分“好不好”，该识别装置 102 进一步对其进行识别处理，未识别出其寻址需求，据此，该识别装置 102 最终判断该查询序列“京东商城好不好”不具有寻址需求。

[0040] 更佳地，该识别装置 102 可对该查询序列的不同部分设置对应的权重，例如，通过对该查询序列进行切词处理，得到切词处理后的词组；进而，分别基于寻址词库，对该切词处理后的词组进行匹配处理；接着，根据匹配结果，并结合对应权重值，通过加权计算，得到查询序列的寻址需求值；最后，基于预定阈值，判断该查询序列是否具有寻址需求。

[0041] 本领域技术人员应能理解上述匹配处理的方式仅为举例，其他现有的或今后可能出现的匹配处理的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0042] 优选地，所述识别装置 102 包括识别单元（未示出）和判断单元（未示出）。该识别单元基于专名词库，对所述查询序列进行专名识别；判断单元当所述查询序列包括专名词组，则判断所述查询序列具有寻址需求。具体地，对于获取装置 101 所获取的查询序列，识别单元基于专名词库，对该查询序列进行专名识别，例如，判断该查询序列是否命中该专名词库中所包括的专名词组；当该查询序列包括专名词组，即，该查询序列命中该专名词库所存储的专名词组时，判断单元判断该查询序列具有寻址需求。

[0043] 在此，该专名词库中存储有诸如机构名、网站名、软件名、品牌名、人名或地名等专名词组，其可位于该设备 1 中，也可位于与该设备 1 通过网络相连接的第三方设备中。

[0044] 例如，对于用户输入的查询序列“清华大学”，识别单元基于专名词库，对该查询序列进行专名识别，假设该专名词库中存储有专名词组“清华大学”，则该查询序列命中该专名词库，判断单元判断该查询序列具有寻址需求。

[0045] 更优选地，所述判断单元当所述查询序列包括专名词组，进一步判断所述查询序列是否包括预定字符串；当所述查询序列包括所述预定字符串，则判断所述查询序列具有寻址需求。具体地，当该识别单元识别出该查询序列包括专名词组时，判断单元进一步判断该查询序列是否包括预定字符串，在此，该预定字符串例如“官网”、“主页”、“.com”等字符串，当该查询序列包括预定字符串时，判断单元判断该查询序列具有寻址需求。

[0046] 例如，对于用户输入的查询序列“清华大学官网”，识别单元基于专名词库，对该查询序列进行专名识别，假设该专名词库中存储有专名词组“清华大学”，则该查询序列命中该专名词库；判断单元进一步判断该查询序列是否包括预定字符串，由于该查询序列包括预定字符串“官网”，则该判断单元判断该查询序列具有寻址需求。

[0047] 较佳地,对于不同类型的专名词组,识别装置 102 根据专名识别结果制定识别策略,例如,对于机构名、网站名、软件名或人名,用户的查询序列包含这些专名词组的情况,与查询序列既包含这些专名词组又包含预定字符串的情况,具有同样强烈的寻址需求。例如,假设识别装置 102 通过对查询序列进行打分的方式,确定其寻址需求值,则在该两种情况下,查询序列的寻址需求值是相同的。而对于品牌名或地名等专名词组,查询序列既包含这些专名词组又包含预定字符串的情况,比查询序列仅包含专名词组的情况具有更强烈的寻址需求,即,前者具有更大的寻址需求值。

[0048] 据此,设备 1 还可对该查询序列进行纵向分档,例如,根据查询序列的寻址需求的强烈程度,如根据其寻址需求值,将查询序列分为强寻址需求、弱寻址需求、范寻址需求、非寻址需求等。例如,按照寻址需求值,对于包含品牌名的查询序列“欧莱雅官网”即具有强寻址需求;“欧莱雅”则具有弱寻址需求。范需求,是指用户的需求不是很具体明确,想得到信息比较宽泛,只要与需求相关即可,比如用户输入查询序列“哪个商城好”,用户主要是想比较下商城的优劣,而不是想直接去某个商城,该查询序列“哪个商城好”即具有范寻址需求。而诸如查询序列“为什么黄金价格下跌”等用户不需要特定页面作为结果页面的查询序列,则具有非寻址需求。

[0049] 在此,该设备 1 还可对查询序列进行横向分档,将该查询序列按照站点名称类型、品牌机构类型等进行分类。

[0050] 图 2 示出根据本发明一个优选实施例的用于识别用户的查询序列中的寻址需求的设备示意图。该设备 1 还包括过滤装置 203。以下参照图 2 对该优选实施例进行详细描述:获取装置 201 获取用户输入的查询序列;过滤装置 203 基于预置黑名单匹配,对所述查询序列进行过滤处理,以过滤掉范需求查询序列,获得过滤处理后的查询序列;识别装置 202 对所述过滤处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。其中,获取装置 201 与图 1 所示对应装置相同,故此处不再赘述,并通过引用的方式包含于此。

[0051] 其中,过滤装置 203 基于预置黑名单匹配,对所述查询序列进行过滤处理,以过滤掉范需求查询序列,获得过滤处理后的查询序列。具体地,过滤装置 203 基于预置黑名单匹配,对获取装置 201 所获取的查询序列进行过滤处理,过滤掉用户需求不明确的范需求查询序列,获得过滤处理后的查询序列,以供识别装置 202 进行寻址需求识别处理。

[0052] 在此,该黑名单中存储有范需求查询序列,其可位于该设备 1 中,也可位于与该设备 1 通过网络相连接的第三方设备中。

[0053] 随后,所述识别装置 202 对所述过滤处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。例如,对于过滤处理后的查询序列,识别装置 202 基于寻址词库,对该过滤处理后的查询序列进行寻址需求识别;或者,基于寻址词库,进一步对该过滤处理后的查询序列进行诸如前缀匹配、后缀匹配、中间词匹配等匹配处理,以做寻址需求识别;又或者,基于专名词库,对该查询序列进行专名识别,或进一步判断该查询序列是否包括预定字符串,对该过滤处理后的查询序列进行寻址需求识别,判断该查询序列是否具有寻址需求。

[0054] 较佳地,该过滤装置 203 还可基于预置白名单匹配,对该查询序列进行过滤处理,从中过滤出明确具有寻址需求的查询序列,进而,识别装置 202 对剩下的查询序列进行寻

址需求识别,进一步从中识别出具有寻址需求的查询序列。

[0055] 在此,该白名单中存储有具有明确寻址需求的查询序列,其可位于该设备 1 中,也可位于与该设备 1 通过网络相连接的第三方设备中。

[0056] 优选地,该设备 1 还包括预处理装置(未示出)。该预处理装置对所述查询序列进行预处理,以获得预处理后的查询序列;其中,所述识别装置 202 对所述预处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;其中,所述预处理包括以下至少任一项:

[0057] - 切词处理;

[0058] - 无效词过滤处理;

[0059] - 同义替换处理。

[0060] 具体地,对于获取装置 201 所获取的查询序列,预处理装置对其进行切词处理、无效词过滤处理或同义替换处理等预处理操作,获得预处理后的查询序列;随后,该识别装置 202 对该预处理后的查询序列进行寻址需求识别,例如,基于寻址词库,对该预处理后的查询序列进行寻址需求识别;或者,基于寻址词库,进一步对该预处理后的查询序列进行诸如前缀匹配、后缀匹配、中间词匹配等匹配处理,以做寻址需求识别;又或者,基于专名词库,对该查询序列进行专名识别,或进一步判断该查询序列是否包括预定字符串,对该预处理后的查询序列进行寻址需求识别,判断该查询序列是否具有寻址需求。

[0061] 例如,预处理装置按照自然语言处理,对该获取装置 201 获取到的查询序列进行切词处理,获得切词处理后的词组,以供识别装置 202 进行寻址需求识别。

[0062] 又如,该预处理装置基于无效词库,对该获取装置 201 获取到的查询序列,或该切词处理后的词组进行过滤处理,过滤掉其中无实际意义的功能字词,如过滤掉该无效词库中所包括的“的、了、得、在”等字词,获得过滤处理后的查询序列,以供识别装置 202 进行寻址需求识别。

[0063] 在此,该无效词库中存储有无实际意义的功能字词,可以是预设的无效词库,也可通过用户反馈建立或更新,该无效词库可位于该设备 1 中,也可位于与该设备 1 通过网络相连接的第三方设备中。

[0064] 再如,该预处理装置基于同义词表或者其他存储有同义词映射关系的词表或词库中进行匹配查询,确定与该查询序列同义的同义词组,或者,确定与包含于该查询序列中的词组同义的同义词组,进而,对该查询序列或其中包含的词组进行同义替换处理,以获得同义替换后的查询序列,供识别装置 202 进行寻址需求识别。

[0065] 本领域技术人员应能理解上述预处理的方式仅为举例,其他现有的或今后可能出现的预处理的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0066] 优选地,该设备 1 还包括排序装置(未示出)。该排序装置当所述查询序列具有寻址需求,提高与所述寻址需求相对应的地址查询结果的结果排序值。具体地,当识别装置 202 识别出该查询序列具有寻址需求,该排序装置相应地提高与该寻址需求对应的地址查询结果的结果排序值,如 pagerank 值,以使该地址查询结果能够在较前的排序位置提供给用户。

[0067] 在此,设备 1 获取用户输入的查询序列,对所述查询序列进行寻址需求识别,以判

断所述查询序列是否具有寻址需求,准确、高效地识别用户的查询序列中的寻址需求;进一步地,根据识别出的寻址查询序列,将直接满足用户需求的页面进行加权,提高了搜索结果展示效果,提升了用户搜索体验。

[0068] 图 3 示出根据本发明另一个方面的用于识别用户的查询序列中的寻址需求的方法流程图。

[0069] 在步骤 S301 中,设备 1 获取用户输入的查询序列。具体地,在步骤 S301 中,设备 1 通过与用户的用户设备的交互,例如,通过一次或多次调用该用户设备提供的应用程序接口(API)或其他约定的通信方式,获取该用户输入的查询序列;或者,通过 ASP、JSP 或 PHP 等页面技术,获取用户在搜索框、查询框等输入界面输入的查询序列;又或者,在步骤 S301 中,设备 1 通过与搜索引擎等第三方设备的交互,获取该用户输入的查询序列。

[0070] 本领域技术人员应能理解上述获取查询序列的方式仅为举例,其他现有的或今后可能出现的获取查询序列的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0071] 在步骤 S302 中,设备 1 对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。具体地,在步骤 S302 中,设备 1 对该查询序列进行寻址需求识别的方式包括但不限于:

[0072] 1) 基于寻址词库,对在步骤 S301 中所获取的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;

[0073] 2) 基于寻址词库,对在步骤 S301 中所获取的查询序列进行诸如前缀匹配、后缀匹配、中间词匹配等匹配处理,以判断所述查询序列是否具有寻址需求;

[0074] 3) 基于专名词库,对所述查询序列进行专名识别,当所述查询序列包括专名词组,则判断所述查询序列具有寻址需求;

[0075] 4) 基于专名词库,对所述查询序列进行专名识别,当所述查询序列包括专名词组,进一步判断所述查询序列是否包括预定字符串,当所述查询序列包括所述预定字符串,则判断所述查询序列具有寻址需求。

[0076] 本领域技术人员应能理解上述识别寻址需求的方式仅为举例,其他现有的或今后可能出现的识别寻址需求的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0077] 在此,设备 1 获取用户输入的查询序列,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求,准确、高效地识别用户的查询序列中的寻址需求。

[0078] 较佳地,在步骤 S302 中,设备 1 基于所述寻址词库与专名词库,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求,其中,所述寻址词库基于查询用户的历史查询记录与查询点击行为建立或更新。

[0079] 在此,设备 1 基于寻址词库与专名词库,通过查询序列分析和查询记录挖掘技术,以及优质资源扩充,从查询序列的语法组成分析用户的寻址意图,同时也从资源满足的维度来反向识别寻址查询序列,从正反两个方向对查询序列进行寻址需求识别,既保证了识别的准确性,同时又增加了召回。

[0080] 优选地,设备 1 的各个步骤之间是持续不断工作的。具体地,在步骤 S301 中,设备 1 获取用户输入的查询序列;在步骤 S302 中,设备 1 对所述查询序列进行寻址需求识别,以

判断所述查询序列是否具有寻址需求。在此,本领域技术人员应理解“持续”是指设备 1 的各步骤分别按照设定的或实时调整的工作模式要求进行查询序列的获取及寻址需求的识别,直至该设备 1 在较长时间内停止获取用户输入的查询序列。

[0081] 优选地,在步骤 S302 中,设备 1 基于寻址词库,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;其中,该方法还包括步骤 S304 (未示出),在步骤 S304 中,设备 1 基于查询用户的历史查询记录与查询点击行为,确定具有寻址需求的查询词;根据所述查询词,建立或更新所述寻址词库。具体地,对于网络中的查询用户,其历史输入了查询序列,将该历史输入的查询序列记为该查询用户的历史查询记录;对于每个查询序列,搜索引擎都会返回一定的查询结果,查询用户对其中的某个查询结果进行点击查看,将该点击查看的查询结果记为该查询用户的查询点击行为。例如,查询用户历史输入了查询序列“京东”,并且,在返回的多个查询结果中,点击进入了网址 [http://www. jd. com/](http://www.jd.com/),则该设备 1 记录该查询用户的历史查询记录与查询点击行为。特别地,对于多个查询用户的历史查询记录与查询点击行为,在步骤 S304 中,设备 1 可确定具有寻址需求的查询词。例如,假设在预定时间段内,有超过预定比例阈值的查询用户在输入查询序列“京东”的情况下,点击进入了网址 <http://www. jd. com/>,则在步骤 S304 中,设备 1 将“京东”确定为具有寻址需求的查询词。进而,在步骤 S304 中,设备 1 通过将该确定的具有寻址需求的查询词存入寻址词库,建立或更新该寻址词库。基于该寻址词库,在步骤 S302 中,设备 1 对在步骤 S301 中所获取的查询序列进行寻址需求识别,以判断该查询序列是否具有寻址需求。例如,当该查询序列命中该寻址词库时,判断该查询序列具有寻址需求。

[0082] 更优选地,在步骤 S302 中,设备 1 基于寻址词库,对所述查询序列进行匹配处理,以判断所述查询序列是否具有寻址需求;其中,所述匹配处理的方式包括以下至少任一项:

[0083] - 前缀匹配;

[0084] - 后缀匹配;

[0085] - 中间词匹配。

[0086] 具体地,在步骤 S302 中,设备 1 基于寻址词库,对在步骤 S301 中所获取的查询序列进行前缀匹配、后缀匹配、中间词匹配等匹配处理,以判断该查询序列是否具有寻址需求,例如,当该查询序列的前缀部分、后缀部分或中间词部分命中该寻址词库时,判断该查询序列具有寻址需求。在此,前缀匹配为该用户输入的查询序列的前缀命中该寻址词库;后缀匹配为该查询序列的后缀部分命中该寻址词库;同理,中间词匹配为该查询序列的中间词部分命中该寻址词库。

[0087] 较佳地,在步骤 S302 中,设备 1 基于寻址词库,对该查询序列进行匹配处理之后,还可进一步对该查询序列的剩余部分进行识别处理,以判断该查询序列是否具有寻址需求。例如,对于用户输入的查询序列“京东商城官网”,在步骤 S302 中,设备 1 基于寻址词库,对该查询序列进行前缀匹配,判断该查询序列的前缀部分“京东商城”命中了该寻址词库,对于该查询序列的剩余部分“官网”,在步骤 S302 中,设备 1 进一步对其进行识别处理,识别其同样具有寻址需求,据此,在步骤 S302 中,设备 1 判断该查询序列“京东商城官网”具有寻址需求。而对于用户输入的查询序列“京东商城好不好”,虽然在步骤 S302 中,设备 1 基于寻址词库,判断该查询序列的前缀部分“京东商城”命中了该寻址词库,然而,对于该

查询序列的剩余部分“好不好”，在步骤 S302 中，设备 1 进一步对其进行识别处理，未识别出其寻址需求，据此，在步骤 S302 中，设备 1 最终判断该查询序列“京东商城好不好”不具有寻址需求。

[0088] 更佳地，在步骤 S302 中，设备 1 可对该查询序列的不同部分设置对应的权重，例如，通过对该查询序列进行切词处理，得到切词处理后的词组；进而，分别基于寻址词库，对该切词处理后的词组进行匹配处理；接着，根据匹配结果，并结合对应权重值，通过加权计算，得到查询序列的寻址需求值；最后，基于预定阈值，判断该查询序列是否具有寻址需求。

[0089] 本领域技术人员应能理解上述匹配处理的方式仅为举例，其他现有的或今后可能出现的匹配处理的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0090] 优选地，步骤 S302 包括子步骤 S302a (未示出)和子步骤 S302b (未示出)。在子步骤 S302a 中，设备 1 基于专名词库，对所述查询序列进行专名识别；当所述查询序列包括专名词组，在子步骤 S302b 中，设备 1 则判断所述查询序列具有寻址需求。具体地，对于在步骤 S301 中所获取的查询序列，在子步骤 S302a 中，设备 1 基于专名词库，对该查询序列进行专名识别，例如，判断该查询序列是否命中该专名词库中所包括的专名词组；当该查询序列包括专名词组，即，该查询序列命中该专名词库所存储的专名词组时，在子步骤 S302b 中，设备 1 判断该查询序列具有寻址需求。

[0091] 在此，该专名词库中存储有诸如机构名、网站名、软件名、品牌名、人名或地名等专名词组，其可位于该设备 1 中，也可位于与该设备 1 通过网络相连接的第三方设备中。

[0092] 例如，对于用户输入的查询序列“清华大学”，在子步骤 S302a 中，设备 1 基于专名词库，对该查询序列进行专名识别，假设该专名词库中存储有专名词组“清华大学”，则该查询序列命中该专名词库，在子步骤 S302b 中，设备 1 判断该查询序列具有寻址需求。

[0093] 更优选地，当所述查询序列包括专名词组，在子步骤 S302b 中，设备 1 进一步判断所述查询序列是否包括预定字符串；当所述查询序列包括所述预定字符串，则判断所述查询序列具有寻址需求。具体地，当在子步骤 S302a 中，设备 1 识别出该查询序列包括专名词组时，在子步骤 S302b 中，设备 1 进一步判断该查询序列是否包括预定字符串，在此，该预定字符串例如“官网”、“主页”、“. com”等字符串，当该查询序列包括预定字符串时，在子步骤 S302b 中，设备 1 判断该查询序列具有寻址需求。

[0094] 例如，对于用户输入的查询序列“清华大学官网”，在子步骤 S302a 中，设备 1 基于专名词库，对该查询序列进行专名识别，假设该专名词库中存储有专名词组“清华大学”，则该查询序列命中该专名词库；在子步骤 S302b 中，设备 1 进一步判断该查询序列是否包括预定字符串，由于该查询序列包括预定字符串“官网”，则设备 1 判断该查询序列具有寻址需求。

[0095] 较佳地，对于不同类型的专名词组，在步骤 S302 中，设备 1 根据专名识别结果制定识别策略，例如，对于机构名、网站名、软件名或人名，用户的查询序列包含这些专名词组的情况，与查询序列既包含这些专名词组又包含预定字符串的情况，具有同样强烈的寻址需求。例如，假设在步骤 S302 中，设备 1 通过对查询序列进行打分的方式，确定其寻址需求值，则在该两种情况下，查询序列的寻址需求值是相同的。而对于品牌名或地名等专名词组，查询序列既包含这些专名词组又包含预定字符串的情况，比查询序列仅包含专名词组的情况

具有更强烈的寻址需求,即,前者具有更大的寻址需求值。

[0096] 据此,设备 1 还可对该查询序列进行纵向分档,例如,根据查询序列的寻址需求的强烈程度,如根据其寻址需求值,将查询序列分为强寻址需求、弱寻址需求、范寻址需求、非寻址需求等。例如,按照寻址需求值,对于包含品牌名的查询序列“欧莱雅官网”即具有强寻址需求;“欧莱雅”则具有弱寻址需求。范需求,是指用户的需求不是很具体明确,想得到信息比较宽泛,只要与需求相关即可,比如用户输入查询序列“哪个商城好”,用户主要是想比较下商城的优劣,而不是想直接去某个商城,该查询序列“哪个商城好”即具有范寻址需求。而诸如查询序列“为什么黄金价格下跌”等用户不需要特定页面作为结果页面的查询序列,则具有非寻址需求。

[0097] 在此,该设备 1 还可对查询序列进行横向分档,将该查询序列按照站点名称类型、品牌机构类型等进行分类。

[0098] 图 4 示出根据本发明一个优选实施例的用于识别用户的查询序列中的寻址需求的方法流程图。以下参照图 4 对该优选实施例进行详细描述:在步骤 S401 中,设备 1 获取用户输入的查询序列;在步骤 S403 中,设备 1 基于预置黑名单匹配,对所述查询序列进行过滤处理,以过滤掉范需求查询序列,获得过滤处理后的查询序列;在步骤 S402 中,设备 1 对所述过滤处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。其中,步骤 S401 与图 3 所示对应步骤相同,故此处不再赘述,并通过引用的方式包含于此。

[0099] 其中,在步骤 S403 中,设备 1 基于预置黑名单匹配,对所述查询序列进行过滤处理,以过滤掉范需求查询序列,获得过滤处理后的查询序列。具体地,在步骤 S403 中,设备 1 基于预置黑名单匹配,对在步骤 S401 中所获取的查询序列进行过滤处理,过滤掉用户需求不明确的范需求查询序列,获得过滤处理后的查询序列,以供设备 1 在步骤 S401 中进行寻址需求识别处理。

[0100] 在此,该黑名单中存储有范需求查询序列,其可位于该设备 1 中,也可位于与该设备 1 通过网络相连接的第三方设备中。

[0101] 随后,在步骤 S402 中,设备 1 对所述过滤处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求。例如,对于过滤处理后的查询序列,在步骤 S402 中,设备 1 基于寻址词库,对该过滤处理后的查询序列进行寻址需求识别;或者,基于寻址词库,进一步对该过滤处理后的查询序列进行诸如前缀匹配、后缀匹配、中间词匹配等匹配处理,以做寻址需求识别;又或者,基于专名词库,对该查询序列进行专名识别,或进一步判断该查询序列是否包括预定字符串,对该过滤处理后的查询序列进行寻址需求识别,判断该查询序列是否具有寻址需求。

[0102] 较佳地,在步骤 S403 中,设备 1 还可基于预置白名单匹配,对该查询序列进行过滤处理,从中过滤出明确具有寻址需求的查询序列,进而,在步骤 S403 中,设备 1 对剩下的查询序列进行寻址需求识别,进一步从中识别出具有寻址需求的查询序列。

[0103] 在此,该白名单中存储有具有明确寻址需求的查询序列,其可位于该设备 1 中,也可位于与该设备 1 通过网络相连接的第三方设备中。

[0104] 优选地,该方法还包括步骤 S405(未示出)。在步骤 S405 中,设备 1 对所述查询序列进行预处理,以获得预处理后的查询序列;其中,在步骤 S402 中,设备 1 对所述预处理后的查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求;其中,所述预处理

包括以下至少任一项：

- [0105] - 切词处理；
- [0106] - 无效词过滤处理；
- [0107] - 同义替换处理。

[0108] 具体地,对于在步骤 S401 中所获取的查询序列,在步骤 S405 中,设备 1 对其进行切词处理、无效词过滤处理或同义替换处理等预处理操作,获得预处理后的查询序列;随后,在步骤 S402 中,设备 1 对该预处理后的查询序列进行寻址需求识别,例如,基于寻址词库,对该预处理后的查询序列进行寻址需求识别;或者,基于寻址词库,进一步对该预处理后的查询序列进行诸如前缀匹配、后缀匹配、中间词匹配等匹配处理,以做寻址需求识别;又或者,基于专名词库,对该查询序列进行专名识别,或进一步判断该查询序列是否包括预定字符串,对该预处理后的查询序列进行寻址需求识别,判断该查询序列是否具有寻址需求。

[0109] 例如,在步骤 S405 中,设备 1 按照自然语言处理,对在步骤 S401 中获取到的查询序列进行切词处理,获得切词处理后的词组,以供设备 1 在步骤 S402 中进行寻址需求识别。

[0110] 又如,在步骤 S405 中,设备 1 基于无效词库,对在步骤 S401 中获取到的查询序列,或该切词处理后的词组进行过滤处理,过滤掉其中无实际意义的功能字词,如过滤掉该无效词库中所包括的“的、了、得、在”等字词,获得过滤处理后的查询序列,以供设备 1 在步骤 S402 中进行寻址需求识别。

[0111] 在此,该无效词库中存储有无实际意义的功能字词,可以是预设的无效词库,也可通过用户反馈建立或更新,该无效词库可位于该设备 1 中,也可位于与该设备 1 通过网络相连接的第三方设备中。

[0112] 再如,在步骤 S405 中,设备 1 基于同义词表或者其他存储有同义词映射关系的词表或词库中进行匹配查询,确定与该查询序列同义的同义词组,或者,确定与包含于该查询序列中的词组同义的同义词组,进而,对该查询序列或其中包含的词组进行同义替换处理,以获得同义替换后的查询序列,供设备 1 在步骤 S402 中进行寻址需求识别。

[0113] 本领域技术人员应能理解上述预处理的方式仅为举例,其他现有的或今后可能出现的预处理的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0114] 优选地,该方法还包括步骤 S406 (未示出)。当所述查询序列具有寻址需求,在步骤 S406 中,设备 1 提高与所述寻址需求相对应的地址查询结果的结果排序值。具体地,当在步骤 S402 中,设备 1 识别出该查询序列具有寻址需求,在步骤 S406 中,设备 1 相应地提高与该寻址需求对应的地址查询结果的结果排序值,如 pagerank 值,以使该地址查询结果能够在较前的排序位置提供给用户。

[0115] 在此,设备 1 获取用户输入的查询序列,对所述查询序列进行寻址需求识别,以判断所述查询序列是否具有寻址需求,准确、高效地识别用户的查询序列中的寻址需求;进一步地,根据识别出的寻址查询序列,将直接满足用户需求的页面进行加权,提高了搜索结果展示效果,提升了用户搜索体验。

[0116] 需要注意的是,本发明可在软件和 / 或软件与硬件的组合体中被实施,例如,可采用专用集成电路(ASIC)、通用目的计算机或任何其他类似硬件设备来实现。在一个实施例

中,本发明的软件程序可以通过处理器执行以实现上文所述步骤或功能。同样地,本发明的软件程序(包括相关的数据结构)可以被存储到计算机可读记录介质中,例如, RAM 存储器,磁或光驱动器或软磁盘及类似设备。另外,本发明的一些步骤或功能可采用硬件来实现,例如,作为与处理器配合从而执行各个步骤或功能的电路。

[0117] 另外,本发明的一部分可被应用为计算机程序产品,例如计算机程序指令,当其被计算机执行时,通过该计算机的操作,可以调用或提供根据本发明的方法和/或技术方案。而调用本发明的方法的程序指令,可能被存储在固定的或可移动的记录介质中,和/或通过广播或其他信号承载媒体中的数据流而被传输,和/或被存储在根据所述程序指令运行的计算机设备的工作存储器中。在此,根据本发明的一个实施例包括一个装置,该装置包括用于存储计算机程序指令的存储器和用于执行程序指令的处理器,其中,当该计算机程序指令被该处理器执行时,触发该装置运行基于前述根据本发明的多个实施例的方法和/或技术方案。

[0118] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。装置权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一,第二等词语用来表示名称,而并不表示任何特定的顺序。

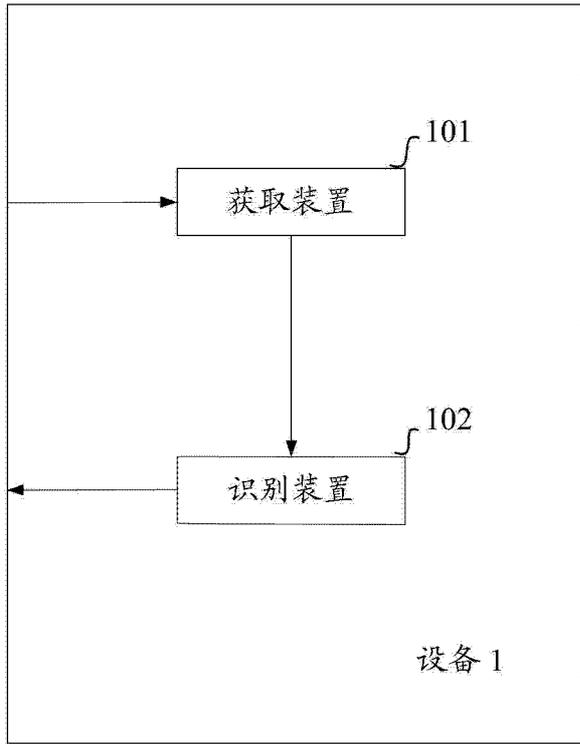


图 1

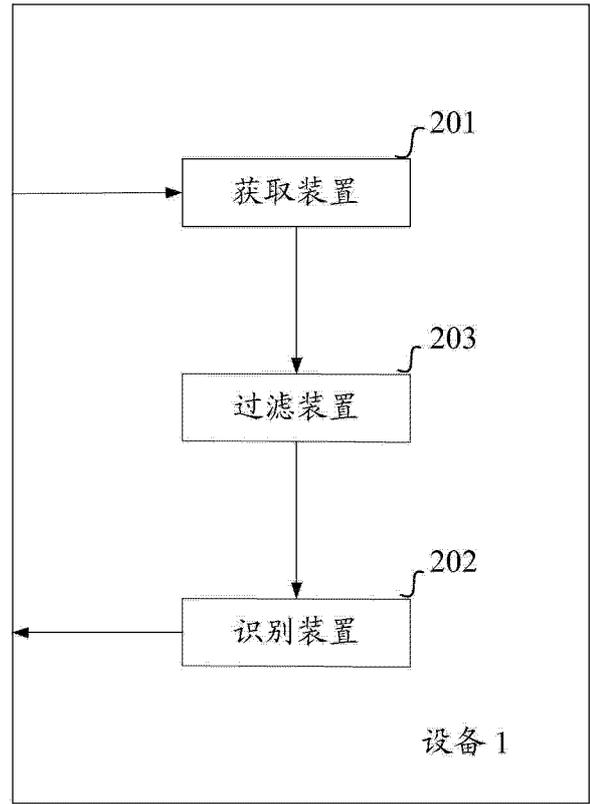


图 2

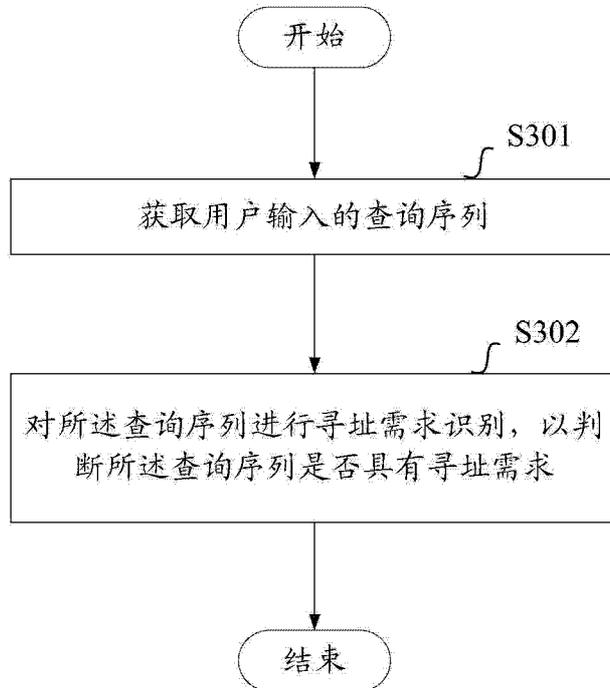


图 3

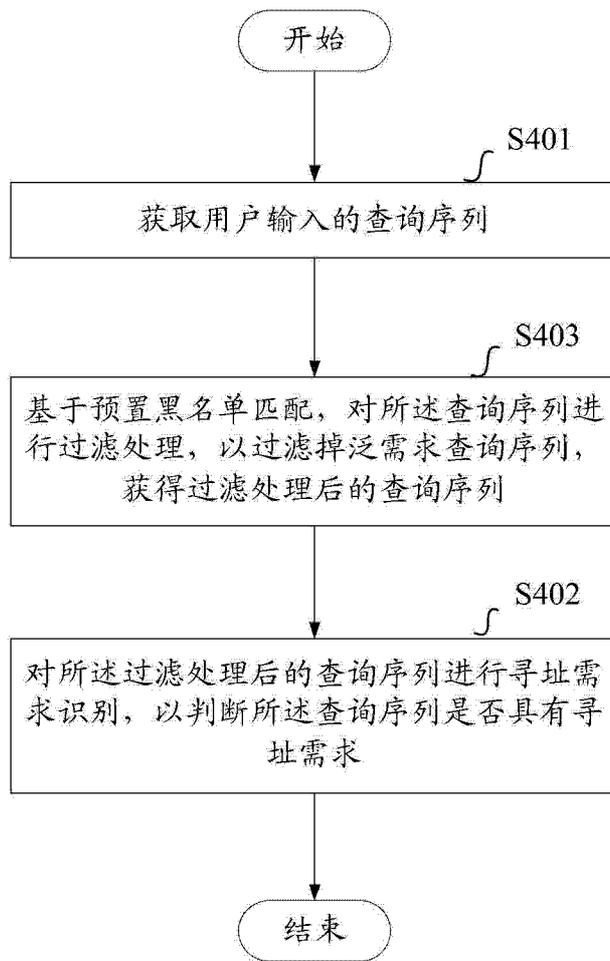


图 4