



(12) 发明专利

(10) 授权公告号 CN 102456054 B

(45) 授权公告日 2016. 05. 18

(21) 申请号 201010528444. 4

US 2007/0266015 A1, 2007. 11. 15,

(22) 申请日 2010. 10. 28

审查员 单娟

(73) 专利权人 深圳市世纪光速信息技术有限公司

地址 518057 广东省深圳市南山区粤海街道
科技中一路腾讯大厦 16 层

(72) 发明人 孙良

(74) 专利代理机构 北京派特恩知识产权代理有限公司 11270

代理人 张颖玲 王黎延

(51) Int. Cl.

G06F 17/30(2006. 01)

(56) 对比文件

- CN 101262494 A, 2008. 09. 10,
- CN 101052181 A, 2007. 10. 10,
- CN 101110077 A, 2008. 01. 23,
- CN 101807213 A, 2010. 08. 18,
- CN 101630330 A, 2010. 01. 20,
- US 2009/0024410 A1, 2009. 01. 22,

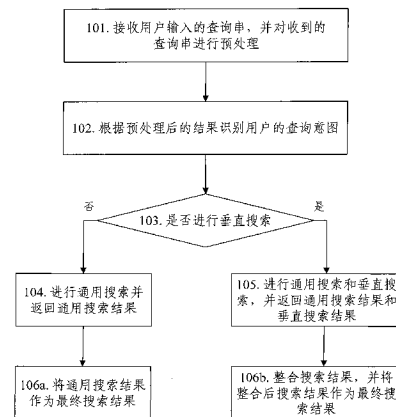
权利要求书1页 说明书5页 附图2页

(54) 发明名称

一种搜索方法及系统

(57) 摘要

本发明公开了一种搜索方法,包括:根据用户输入的查询串识别用户的查询意图,在根据识别出的查询意图确定需要进行垂直搜索时,同时进行通用搜索和垂直搜索,并将垂直搜索结果与通用搜索结果整合,得到最终搜索结果。本发明还同时公开了一种搜索系统,采用本发明能充分理解用户的查询意图,为用户提供更丰富的搜索结果,且实现简单、方便。



1. 一种搜索方法,其特征在于,预先设置查询意图识别模型,所述查询意图识别模型中包括各种查询意图的规则或条件;该方法还包括:

接收用户输入的查询串并进行预处理,结合本次预处理结果和历史搜索记录共同识别出用户的查询意图,根据识别出的查询意图将进行预处理后的查询词与预设的查询意图识别模型中的规则或条件进行匹配,确定是否进行垂直搜索;确定需要进行垂直搜索时,同时进行通用搜索和垂直搜索,并将垂直搜索结果与通用搜索结果整合,得到最终搜索结果;根据所采集的历史数据中的点击量情况,对当前得到的最终搜索结果进行排序。

2. 根据权利要求1所述的方法,其特征在于,该方法还包括:显示所述最终搜索结果。

3. 根据权利要求1或2所述的方法,其特征在于,该方法还包括:根据识别出的查询意图确定不需要进行垂直搜索时,则进行通用搜索;并将通用搜索结果作为最终搜索结果。

4. 根据权利要求1或2所述的方法,其特征在于,该方法进一步包括:识别出查询意图为需要获取文字类信息、以及文字类信息以外的至少一种其他类型信息时,确定需要进行垂直搜索。

5. 一种搜索系统,其特征在于,该系统包括:查询串处理模块、通用搜索引擎、垂直搜索引擎、搜索结果获取模块;其中,

所述查询串处理模块,用于接收用户输入的查询串并进行预处理,结合本次预处理结果和历史搜索记录共同识别出用户的查询意图,根据识别出的查询意图将进行预处理后的查询词与预设的查询意图识别模型中的规则或条件进行匹配,确定是否进行垂直搜索;确定需要进行垂直检索时,向通用搜索引擎和垂直搜索引擎发送查询请求;

所述通用搜索引擎,用于完成通用搜索,并将通用搜索结果发送给搜索结果获取模块;

所述垂直搜索引擎,用于完成垂直搜索,并将垂直搜索结果发送给搜索结果获取模块;

所述搜索结果获取模块,用于对收到的通用搜索结果和垂直搜索结果进行整合,并将整合后的搜索结果作为最终搜索结果;所述最终搜索结果进一步根据所采集的历史数据中的点击量情况进行排序。

6. 根据权利要求5所述的系统,其特征在于,该系统还包括搜索结果显示模块;

所述搜索结果获取模块,还用于将最终搜索结果发送给搜索结果显示模块;

所述搜索结果显示模块,用于将搜索结果获取模块发来的最终搜索结果进行显示。

7. 根据权利要求5或6所述的系统,其特征在于,所述查询串处理模块进一步包括:查询串接收单元、预处理单元、识别及判断单元;其中,

所述查询串接收单元,用于接收用户输入的查询串,并将收到的查询串发送给预处理单元;

所述预处理单元,用于将收到的查询串进行预处理,并将预处理结果发送给识别及判断单元;

所述识别及判断单元,用于根据预处理结果识别用户的查询意图,确定需要进行垂直搜索时,向通用搜索引擎和垂直搜索引擎发送查询请求。

8. 根据权利要求7所述的系统,其特征在于,所述识别及判断单元具体用于:根据预处理结果识别出查询意图为需要获取文字类信息、以及文字类信息以外的至少一种其他类型信息时,确定需要进行垂直搜索,并向通用搜索引擎和垂直搜索引擎发送查询请求。

一种搜索方法及系统

技术领域

[0001] 本发明涉及网页信息检索技术,特别是涉及一种综合各类信息的搜索方法及系统。

背景技术

[0002] 随着网络技术的迅猛发展,通过网络获取所需信息资料已经成为人们必不可少、甚至极其重要的一种信息获取方式。目前,用户需要查询某种或某类信息时,大多是打开浏览器,通过某个检索网站,如百度、Google等等,输入希望查询信息的关键词,之后网页搜索引擎根据用户提交的查询词搜索相关的网页,并将搜索到的网页内容显示给用户,这是通用的网页搜索方式,可简称为通用搜索。

[0003] 但是,由于通用搜索的信息量较大,使得查询不够准确、不够专业、深度也不够,因此,通用搜索渐渐不能很好的满足人们的检索需求,于是出现了垂直搜索,以满足用户对搜索结果专业、精确以及丰富的要求。所谓垂直搜索,是针对某一特定领域、或特定人群、或特定需求提供的专业搜索引擎,是搜索引擎的细分和延伸,是对网页库中某类专门的信息进行一次整合,定向分字段抽取出需要的数据进行处理后,再以某种形式返回给用户。垂直搜索的特点是专、精、深,相对于通用搜索引擎海量信息的无序化,垂直搜索引擎显得更加专注、具体和深入。垂直搜索引擎与通用搜索引擎最大的区别是对网页信息进行了结构化信息抽取,将网页的非结构化数据抽取成特定的结构化信息数据,通用搜索是以网页为最小单位,而垂直搜索是以结构化数据为最小单位。

[0004] 现有技术中,用户输入查询词进行网页搜索时,通常只能提供一些文字信息,如果希望获取视频信息、音频信息、图片信息等等搜索结果,并不能直接得到,还需要进一步的查询或点击相关页面中的其他链接,进入新的网页进行查找或获取,总之需要多次查找才能找到所需的搜索结果,检索过程比较复杂,且不能直接得到所需的搜索结果。并且,目前也没有人提出相关方案将通用搜索和垂直搜索结合起来,以便为用户通过更简单的方式提供更丰富的搜索结果。

[0005] 另外,对于网页搜索而言,要想为用户提供更丰富、更准确的搜索结果,还需要对用户的查询意图进行充分识别,只有准确地了解用户的查询意图,才能得到符合用户需求的搜索结果,而目前也没有类似的技术方案被提出。

发明内容

[0006] 有鉴于此,本发明实施例提供一种搜索方法及系统,能为用户提供更丰富的搜索结果,且实现简单、方便。

[0007] 本发明的技术方案是这样实现的:

[0008] 本发明提供了一种搜索方法,包括:根据用户输入的查询串识别用户的查询意图,根据识别出的查询意图确定需要进行垂直搜索时,同时进行通用搜索和垂直搜索,并将垂直搜索结果与通用搜索结果整合,得到最终搜索结果。

- [0009] 上述方案中,该方法还包括:显示所述最终搜索结果。
- [0010] 上述方案中,所述根据用户输入的查询串识别用户的查询意图为:接收用户输入的查询串,并对收到的查询串进行预处理;根据预处理结果识别用户的查询意图。
- [0011] 上述方案中,该方法还包括:根据识别出的查询意图确定不需要进行垂直搜索时,则进行通用搜索;并将通用搜索结果作为最终搜索结果。
- [0012] 上述方案中,该方法进一步包括:识别出查询意图为需要获取文字类信息、以及文字类信息以外的至少一种其他类型信息时,确定需要进行垂直搜索。
- [0013] 上述方案中,该方法进一步包括:对所述最终搜索结果根据点击量情况进行排序。
- [0014] 本发明还提供了一种搜索系统,包括:查询串处理模块、通用搜索引擎、垂直搜索引擎、搜索结果获取模块;其中,
- [0015] 所述查询串处理模块,用于根据用户输入的查询串识别用户的查询意图,根据识别出的查询意图确定需要进行垂直检索时,向通用搜索引擎和垂直搜索引擎发送查询请求;
- [0016] 所述通用搜索引擎,用于完成通用搜索,并将通用搜索结果发送给搜索结果获取模块;
- [0017] 所述垂直搜索引擎,用于完成垂直搜索,并将垂直搜索结果发送给搜索结果获取模块;
- [0018] 所述搜索结果获取模块,用于对收到的通用搜索结果和垂直搜索结果进行整合,并将整合后的搜索结果作为最终搜索结果。
- [0019] 上述方案中,该系统还包括搜索结果显示模块;所述搜索结果获取模块,还用于将最终搜索结果发送给搜索结果显示模块;所述搜索结果显示模块,用于将搜索结果获取模块发来的最终搜索结果进行显示。
- [0020] 上述方案中,所述查询串处理模块进一步包括:查询串接收单元、预处理单元、识别及判断单元;其中,所述查询串接收单元,用于接收用户输入的查询串,并将收到的查询串发送给预处理单元;所述预处理单元,用于将收到的查询串进行预处理,并将预处理结果发送给识别及判断单元;所述识别及判断单元,用于根据预处理结果识别用户的查询意图,确定需要进行垂直搜索时,向通用搜索引擎和垂直搜索引擎发送查询请求。
- [0021] 其中,所述识别及判断单元具体用于:根据预处理结果识别出查询意图为需要获取文字类信息、以及文字类信息以外的至少一种其他类型信息时,确定需要进行垂直搜索,并向通用搜索引擎和垂直搜索引擎发送查询请求。
- [0022] 本发明提供的搜索方法及系统,根据用户输入的查询串识别用户的查询意图,在根据识别出的查询意图确定需要进行垂直搜索时,同时进行通用搜索和垂直搜索,并将垂直搜索的搜索结果与通用搜索的搜索结果整合在一起提供给用户。如此,能为用户提供多种多样符合用户需求的搜索结果,如新闻、视频、图片、博客、论坛等等。
- [0023] 本发明根据用户输入的查询串识别用户查询意图,不仅能充分理解用户的查询意图,而且实现方式简单、方便;本发明将通用搜索和垂直搜索结合在一起,能为用户提供更专业、精度更高、更丰富的搜索结果。

附图说明

[0024] 图1为本发明搜索方法的实现流程示意图；

[0025] 图2为本发明搜索系统的组成结构示意图。

具体实施方式

[0026] 本发明实施例中,根据用户输入的查询串识别用户的查询意图,在根据识别出的查询意图确定需要进行垂直搜索时,同时进行通用搜索和垂直搜索,并将垂直搜索结果与通用搜索结果整合,整合后的搜索结果作为最终搜索结果。

[0027] 这里,如果仅有通用搜索结果,则直接将通用搜索结果作为最终搜索结果;也就是说,如果只有通用搜索结果,按现有技术的实现方案处理即可。

[0028] 进一步的,得到最终搜索结果后,可将得到的最终搜索结果显示给用户。

[0029] 下面结合附图对本发明的实现进行详细地说明。

[0030] 本发明的搜索方法如图1所示,包括以下步骤:

[0031] 步骤101:接收用户输入的查询串,并对收到的查询串进行预处理;

[0032] 这里,所述预处理主要是指对收到的查询串进行分词处理,所述分词处理就是按语音学上的含义识别收到的查询串,并将查询串切分为多个单词,所述单词可分别作为查询词;例如,收到的查询串为:26届大运会,该查询串可被切分为:26、届、大运会三个单词,那么,26、届、大运会可作为三个查询词。具体如何对查询串进行预处理属于现有技术,这里不再详述。

[0033] 步骤102:根据预处理结果,识别用户的查询意图;

[0034] 这里,根据预处理后得到的查询词,可以分析出用户当前的查询意图,例如,如果当前的查询词为移动通信,那么,查询意图为查询文字相关信息,包括新闻、名词介绍、相关技术介绍、相关产品介绍等等;如果查询词为刘德华,那么,用户期望查询的应该包括新闻信息、视频信息、图片信息、博客信息等等,是多种类型的组合。

[0035] 由于每个用户只要之前检索过,都会存在历史搜索记录,本步骤中,进一步的,可以结合本次预处理结果和历史搜索记录共同识别用户的查询意图。比如:用户第一次输入某个人名查询时,可能只提供了新闻信息、文字介绍信息等,但用户通过进一步搜索,又获取了该人物的视频信息、博客信息、图片信息,那么,输入人名后需要获取新闻信息、文字介绍信息、视频信息、博客信息、以及图片信息就成为历史搜索记录;当该用户再次检索输入查询串后,如果确认是人名,则可以同时提供新闻信息、文字介绍信息、视频信息、博客信息、图片信息等多种类型的信息,以满足用户的需求。

[0036] 步骤103~105:根据识别出的用户查询意图确定是否需要进行垂直搜索,如果需要进行垂直搜索,则搜索引擎在完成通用搜索的同时,进行垂直搜索,之后,搜索引擎分别返回通用搜索结果和垂直搜索结果,执行步骤106b;否则,搜索引擎仅完成通用搜索,并返回通用搜索结果,执行步骤106a;

[0037] 例如,如果根据预处理后的查询词确认仅需要获取文字类的信息,就确定不需要进行垂直搜索;如果根据预处理后的查询词确认除了需要获取文字类信息以外,还需要获取视频、和/或图片、和/或博客等其他类型的信息,则确定需要进行垂直搜索。

[0038] 本步骤中,确定查询意图是否包括垂直搜索后,再向相应的搜索引擎发送查询请求,具体的,确定不需要进行垂直搜索时,仅向通用搜索引擎发送查询请求,请求中携带查

询串;确定需要进行垂直搜索时,同时向通用搜索引擎和垂直搜索引擎发送查询请求,请求中携带查询串。

[0039] 对于确定是否进行垂直搜索,还可以预先设置查询意图识别模型,在查询意图识别模型中给出各种查询意图的规则或条件,那么,根据预处理结果、或根据预处理结果和历史搜索记录分析符合哪种规则就完成哪种检索,比如:符合规则A、C的仅需要通用搜索的信息,符合规则B、D的同时还需要垂直搜索的信息;相应的,将进行预处理后的查询词与查询意图识别模型中的规则或条件进行匹配,以便识别出用户当前的查询意图,进而确定是否进行垂直搜索。

[0040] 本申请中采用垂直搜索,相对来说数据量较小,几乎能做到很快更新,这样,还可以直接提升用户的使用体验。

[0041] 步骤106a:将通用搜索结果作为最终搜索结果,结束当前处理流程;

[0042] 步骤106b:将通用搜索结果和垂直搜索结果进行整合,并将整合后的搜索结果作为最终搜索结果;

[0043] 这里,对于只有通用搜索结果的情况,直接将通用搜索结果作为最终搜索结果;对于同时返回通用搜索结果和垂直搜索结果的情况,需要将两种搜索结果进行整合,也就是将两个搜索结果的集合进行合并求并集,并将整合后的搜索结果作为最终搜索结果。

[0044] 其中,对于最终搜索结果,还可以进一步根据所采集的历史数据中的点击量情况,对当前得到的最终搜索结果进行排序,将点击最多的搜索结果排在最前面,其他搜索结果可依次按点击量从多到少顺序排列;并且,由于点击量情况是实时变化的,所以对每次得到的最终搜索结果都可以对搜索结果的排序进行实时调整。如:得到的最终搜索结果包括七条搜索结果A1~A7,根据历史的点击量,可将七条搜索结果排列为A3、A5、A1、A2、A4、A6、A7;A3可以为视频搜索结果,A5可以为音频搜索结果,A1为图片搜索结果,A2可以为新闻搜索结果,A4可以为博客搜索结果,A6、A7可以为文字搜索结果。

[0045] 进一步的,该方法还包括:将所得到的最终搜索结果显示给用户;其中,所述最终搜索结果可以是经过排序的搜索结果。

[0046] 为实现上述方法,本发明还提供了一种搜索系统,如图2所示,包括:查询串处理模块、通用搜索引擎、垂直搜索引擎、搜索结果获取模块;其中,

[0047] 查询串处理模块,用于根据用户输入的查询串识别用户的查询意图,根据识别出的查询意图确定需要进行垂直检索时,向通用搜索引擎和垂直搜索引擎发送查询请求;其中,查询请求中携带有查询串。

[0048] 通用搜索引擎,用于完成通用搜索,并将通用搜索结果发送给搜索结果获取模块;

[0049] 垂直搜索引擎,用于完成垂直搜索,并将垂直搜索结果发送给搜索结果获取模块;

[0050] 搜索结果获取模块,用于对收到的通用搜索结果和垂直搜索结果进行整合,并将整合后的搜索结果作为最终搜索结果。

[0051] 进一步的,如果查询串处理模块根据识别出的查询意图,确定不需要进行垂直检索,则查询串处理模块仅向通用搜索引擎发送查询请求;相应的,仅有通用搜索引擎向搜索结果获取模块发送通用搜索结果;搜索结果获取模块仅收到通用搜索结果时,将通用搜索结果作为最终搜索结果。

[0052] 上述查询串处理模块进一步包括:查询串接收单元、预处理单元、识别及判断单

元;其中,

[0053] 查询串接收单元,用于接收用户输入的查询串,并将收到的查询串发送给预处理单元;预处理单元,用于将收到的查询串进行预处理,并将预处理结果发送给识别及判断单元;

[0054] 这里,预处理单元可以是分词单元,所述分词单元可将收到的查询串按照语言学含义划分为一个或多个单词,之后将分词结果发送给识别及判断单元;

[0055] 识别及判断单元,用于根据预处理结果识别用户的查询意图,确定需要进行垂直搜索时,向通用搜索引擎和垂直搜索引擎发送查询请求。

[0056] 该识别及判断单元具体用于:根据预处理结果识别出查询意图为需要获取文字类信息、以及文字类信息以外的至少一种其他类型信息时,确定需要进行垂直搜索,并向通用搜索引擎和垂直搜索引擎发送查询请求。

[0057] 另外,该识别及判断单元在确定不需要进行垂直搜索时,仅向通用搜索引擎发送查询请求。

[0058] 该系统还包括搜索结果显示模块;所述搜索结果获取模块,还用于将最终搜索结果发送给搜索结果显示模块,所述搜索结果显示模块,则用于将搜索结果获取模块发来的最终搜索结果进行显示。

[0059] 以上所述,仅为本发明的较佳实施例而已,并非用于限定本发明的保护范围,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

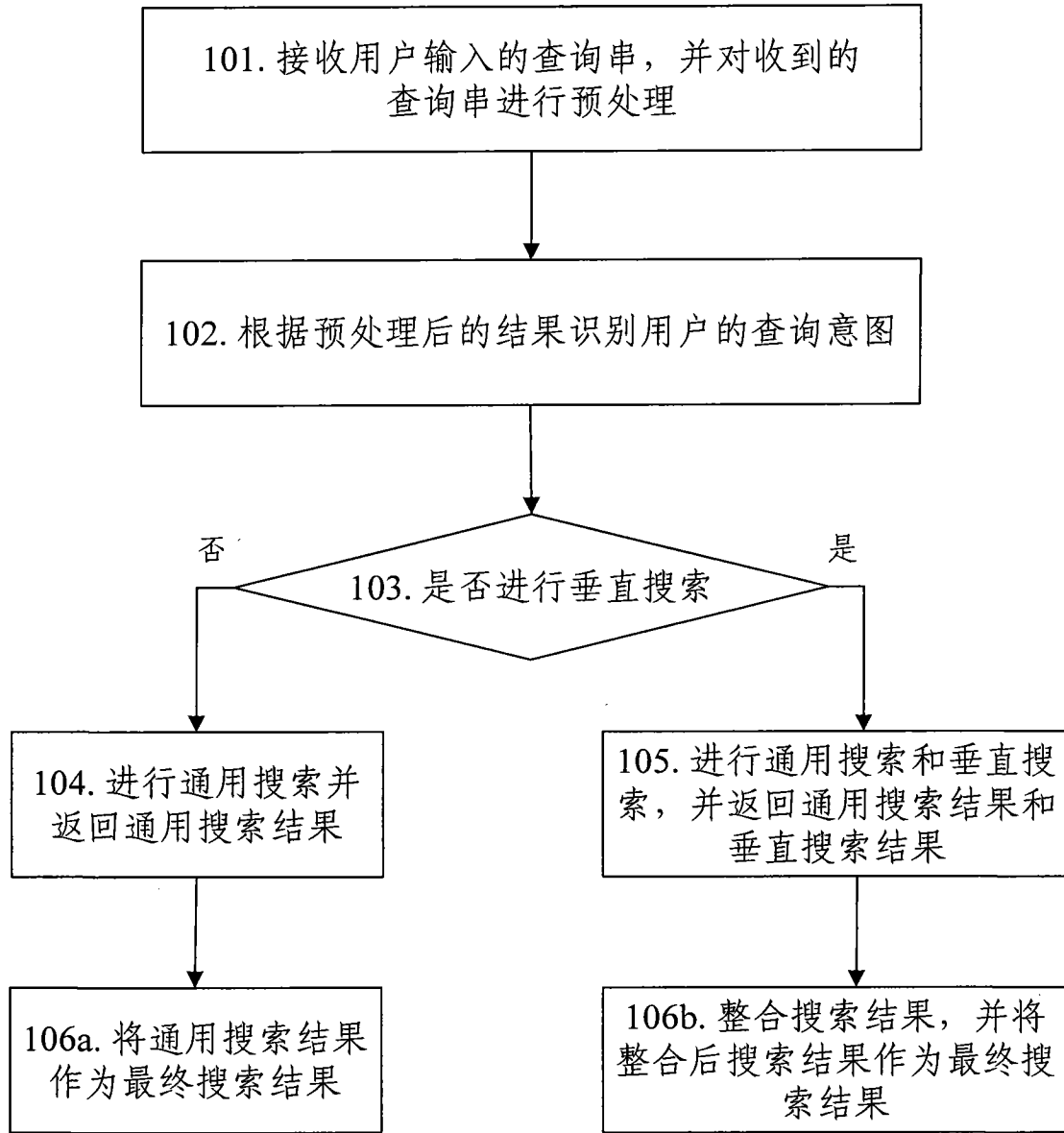


图1

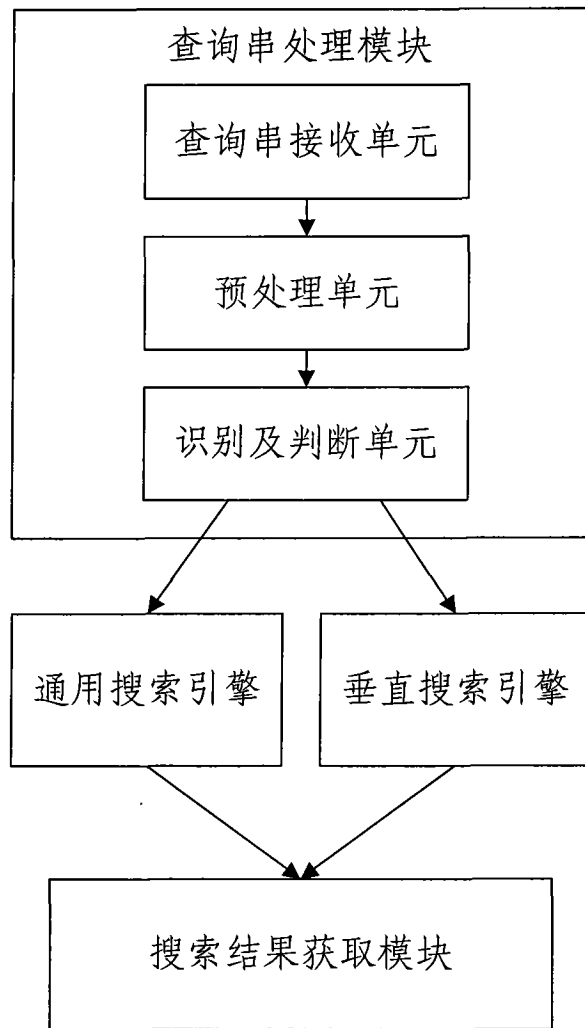


图2