



(12)发明专利

(10)授权公告号 CN 102144228 B

(45)授权公告日 2017.04.26

(21)申请号 200980134320.3

(22)申请日 2009.07.03

(65)同一申请的已公布的文献号  
申请公布号 CN 102144228 A

(43)申请公布日 2011.08.03

(30)优先权数据  
61/078,157 2008.07.03 US  
12/211,712 2008.09.16 US

(85)PCT国际申请进入国家阶段日  
2011.03.03

(86)PCT国际申请的申请数据  
PCT/CN2009/072610 2009.07.03

(87)PCT国际申请的公布数据  
W02010/000208 EN 2010.01.07

(73)专利权人 谷歌股份有限公司  
地址 美国加利福尼亚州

(72)发明人 王咏刚 洪锋 许炜 唐溪柳  
亨利.欧 杨波 张蕾 杨润华  
吴军 姚宝刚

(74)专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 陈炜

(51)Int.Cl.  
G06F 17/28(2006.01)

(56)对比文件  
US 2007/0106653 A1,2007.05.10,说明书  
第[0020]-[0022],[0042]段、附图2-6.

US 2007/0106653 A1,2007.05.10,说明书  
第[0020]-[0022],[0042]段、附图2-6.

US 2005/0086590 A1,2005.04.21,说明书  
第[0080]段.

CN 101203849 A,2008.06.18,全文.

审查员 刘剑

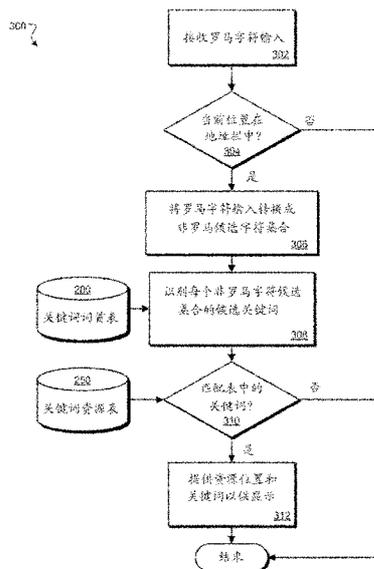
权利要求书3页 说明书8页 附图6页

(54)发明名称

根据输入字符序列的资源定位符建议

(57)摘要

提供了方法、系统和装置以及计算机程序产品,其中,输入法编辑器接收罗马字符输入,识别关于非罗马字符候选集合的关键词,以及识别相关联的资源位置。在识别出相关联的资源位置后,将资源位置与非罗马字符候选集合相关联。



1. 一种计算机实现的方法,包括:
  - 在用户界面的网络地址输入栏接收罗马字符输入;
  - 使用所述罗马字符输入来识别多个非罗马字符候选集合,并且将所述多个非罗马字符候选集合提供给用户界面以供显示;
  - 识别资源位置候选集合,每个资源位置候选集合与相应的非罗马字符候选集合相关联,其中,识别资源位置候选集合包括:对于每个非罗马字符候选集合,
    - 通过将该非罗马字符候选集合用作词首对关键词进行词首搜索来识别关于该非罗马字符候选集合的关键词,所述词首搜索包括利用非罗马字符候选集合搜索词首表以识别关键词,所述词首表将关键词映射到非罗马字符的词首;
    - 对于每个识别出的关键词,通过利用识别出的关键词搜索关键词表以识别资源位置来识别相关联的资源位置,所述关键词表将关键词映射到资源位置;以及
    - 将识别出的资源位置与该非罗马字符候选集合相关联;
    - 接收对所述多个非罗马字符候选集合中的一个的选择;以及
    - 响应于接收所述选择,显示与所选择的非罗马字符候选集合相关联的关键词以及资源位置候选集合,
  - 其中,所述资源位置包括统一资源定位符。
2. 根据权利要求1所述的方法,其中,所述非罗马字符是汉字字符。
3. 根据权利要求1所述的方法,其中:
  - 利用识别出的关键词搜索关键词表以识别资源位置包括:对于每个识别出的关键词,识别映射到识别出的关键词的资源位置;
  - 识别映射到识别出的关键词的每个资源位置的排名;以及
  - 选择具有最高排名的资源位置;并且
  - 将识别出的资源位置与非罗马字符候选集合相关联包括:仅将关于每个关键词的排名最高的资源位置与非罗马字符候选集合相关联。
4. 根据权利要求3所述的方法,进一步包括:
  - 跟踪每个资源位置候选集合中的每个资源位置被选择的次数;以及
  - 基于每个资源位置被选择的次数调整每个资源位置的排名。
5. 根据权利要求1所述的方法,进一步包括:对于每个资源位置候选集合,
  - 跟踪每个资源位置候选集合中的每个资源位置被选择的次数;以及
  - 基于每个资源位置已被选择的次数,将每个所述资源位置候选集合中的资源位置相对于该资源位置候选集合中的其他资源位置进行排名。
6. 根据权利要求1所述的方法,进一步包括:
  - 在所述用户界面上显示每个所述非罗马字符候选集合;
  - 接收对所述非罗马字符候选集合中的一个的选择;以及
  - 仅显示与所述非罗马字符候选集合中的所选择的一个相关联的资源位置候选集合和关键词。
7. 根据权利要求1所述的方法,其中,由用户设备的用户界面执行与所选择的非罗马字符候选集合相关联的关键词以及资源位置候选集合的显示,并且所述方法进一步包括:将与所选择的非罗马字符候选集合相关联的关键词以及资源位置候选集合传送到所述用户

设备。

8. 一种计算机实现的方法,包括:

根据查询日志中的查询识别关键词;

生成关键词到资源位置的映射,每个资源位置是由用户响应于包括所述关键词的搜索而选择的位置,该位置映射到该关键词;

生成词首到关键词的映射;以及

向客户端设备提供所述词首到关键词的映射以及所述关键词到资源位置的映射,以供在所述客户端设备处解析资源位置地址时使用,

其中,所述资源位置包括统一资源定位符。

9. 根据权利要求8所述的方法,其中,生成词首到关键词的映射包括:生成关于关键词的词首树,所述词首树中的每个节点对应于关键词或多个关键词共同的词首。

10. 根据权利要求9所述的方法,其中,每个关键词包括一个或多个汉字字符。

11. 根据权利要求10所述的方法,其中,生成关键词到资源位置的映射包括:基于每个资源位置已被选择的次数,将关于每个关键词的每个资源位置相对于关于该关键词的其他资源位置进行排名。

12. 一种计算机实现的装置,包括:

用于在用户界面的网络地址输入栏接收罗马字符输入的装置;

用于使用所述罗马字符输入来识别多个非罗马字符候选集合,并且将所述多个非罗马字符候选集合提供给用户界面以供显示的装置;

用于识别资源位置候选集合的装置,每个资源位置候选集合与相应的非罗马字符候选集合相关联,其中,识别资源位置候选集合包括:对于每个非罗马字符候选集合,

通过将该非罗马字符候选集合用作词首对关键词进行词首搜索来识别关于该非罗马字符候选集合的关键词,所述词首搜索包括利用非罗马字符候选集合搜索词首表以识别关键词,所述词首表将关键词映射到非罗马字符的词首;

对于每个识别出的关键词,通过利用识别出的关键词搜索关键词表以识别资源位置来识别相关联的资源位置,所述关键词表将关键词映射到资源位置;以及

将识别出的资源位置与该非罗马字符候选集合相关联;

用于接收对所述多个非罗马字符候选集合中的一个的选择的装置;以及

用于响应于接收所述选择,显示与所选择的非罗马字符候选集合相关联的关键词以及资源位置候选集合的装置,

其中,所述资源位置包括统一资源定位符。

13. 根据权利要求12所述的装置,其中,所述非罗马字符是汉字字符。

14. 根据权利要求12所述的装置,其中:

用于识别资源位置候选集合的装置包括:

用于对于每个识别出的关键词识别映射到识别出的关键词的资源位置的装置;

用于识别映射到识别出的关键词的每个资源位置的排名的装置;

用于选择具有最高排名的资源位置的装置;以及

用于仅将关于每个关键词的排名最高的资源位置与所述非罗马字符候选集合相关联的装置。

15. 根据权利要求14所述的装置,进一步包括:  
用于跟踪每个所述资源位置候选集中的每个资源位置被选择的次数的装置;以及  
用于基于每个资源位置被选择的次数调整每个资源位置的排名的装置。

## 根据输入字符序列的资源定位符建议

[0001] 相关申请的交叉引用

[0002] 本申请要求于2008年7月3日提交的序号为No.61/078,157、标题为“Resource Locator Suggestions From Input Character Sequence”(根据输入字符序列的资源定位符建议)的美国临时专利申请和于2008年9月16日提交的序号为No.12/211,712、标题为“Resource Locator Suggestions From Input Character Sequence”(根据输入字符序列的资源定位符建议)的美国专利申请的优先权,其内容通过引用合并于此。

### 技术领域

[0003] 本公开涉及输入法。

### 背景技术

[0004] 在词符文字(logographic script)中,一个或两个字符粗略地对应于一个词或一种含意,使用词符文字的语言所具有的字符多于标准输入设备(如计算机键盘或移动设备键区)上的键。例如,中文语言包含由基本拼音字母和五个音调定义发音的数万个字符。可以通过帮助输入在输入设备上找不到的字符和符号的输入法来实现这些潜在地多对一关联的映射。因此,可以使用罗马字符类型的键盘来输入中文字符。类似地,输入法能够用于使用罗马字符类型的键盘或其他输入设备来输入包括表意文字的许多其他语言,如日语、韩语和其他语言。

[0005] 尽管可以使用输入法编辑器,但因特网站点的地址,例如统一资源定位符(“URL”),通常采用ASCII字符的形式。因而,对于母语是基于词符文字的语言的用户、例如主要讲中文方言的用户来说,即使他们知道在他们的语言中网站的文字名称(literal name),也难以记住网站的URL。因此,这些用户必须采取几个独立的步骤来输入他们期望查看的网站的URL。可以使用用户的语言中的表意字符、以用户的语言来拼写与URL相关联的文字名称。由于表意字符不是使用西式键盘直接输入的,所以使用输入法编辑器来完成从罗马字符音标输入(例如,拼音)到候选表意字符(例如,汉字)的转换。之后,由搜索引擎使用表意字符来找到可能的用户期望的网站,并且如果在产生的搜索结果中出现该网站,则用户选择该网站。

### 发明内容

[0006] 本说明书描述了与输入法编辑器相关的技术。在一些实现方式中,能够直接根据在网页浏览器的地址栏中输入的音标输入(例如,拼音输入)找到资源位置(例如,URL)。在网页浏览器的地址栏中接收罗马字符的用户输入,并且将该用户输入转换成非罗马字符候选集合,例如汉字字符候选集合。将每个非罗马字符的集合用作关键词词首来识别关键词。关键词与URL相关联,并且当选择了候选集合时,为每个相关联的非罗马字符候选集合显示URL和相关联的关键词。

[0007] 一般地,在本说明书中描述的主题的一个方面能够实现为包括以下动作的方法:

在用户界面的网络地址输入栏接收罗马字符输入;使用所述罗马字符输入来识别多个非罗马字符候选集合,并且将所述多个非罗马字符候选集合提供给用户界面以供显示;识别资源位置候选集合,每个资源位置候选集合与相应的非罗马字符候选集合相关联,其中,识别资源位置候选集合包括,对于每个非罗马字符候选集合:识别关于该非罗马字符候选集合的关键词,对于每个识别出的关键词识别相关联的资源位置,以及将所述资源位置与该非罗马字符候选集合相关联;接收对所述多个非罗马字符候选集合中的一个的选择;以及响应于接收所述选择,显示与所选择的非罗马字符候选集合相关联的关键词以及资源位置候选集合。其中,所述资源位置包括统一资源定位符。该方面的其他实施例包括相应的系统、装置和计算机程序产品。

[0008] 在本说明书中描述的主题的另一个方面能够实现为包括以下动作的方法:根据查询日志中的查询识别关键词;生成关键词到资源位置的映射,每个资源位置是由用户响应于包括所述关键词的搜索而选择的位置,其中该位置映射到该关键词;生成词首到关键词的映射;以及向客户端设备提供所述词首到关键词的映射以及所述关键词到资源位置的映射,以供在所述客户端设备处解析资源位置地址时使用。该方面的其他实施例包括相应的系统、装置和计算机程序产品,其中,所述资源位置包括统一资源定位符。

[0009] 在本说明书中描述的主题的再一个方面能够实现为包括以下特征的计算机实施的装置:用于在用户界面的网络地址输入栏接收罗马字符输入的装置;用于使用所述罗马字符输入来识别非罗马字符候选集合的装置;用于识别资源位置候选集合的装置,每个资源位置候选集合与相应的非罗马字符候选集合相关联;以及用于将所述资源位置候选集合和所述非罗马字符候选集合提供给用户界面以供显示的装置,其中,所述资源位置包括统一资源定位符

[0010] 在附图以及下面的描述中阐述了本说明书中描述的主题的一个或多个实施例的细节。所述主题的其他特征、方面和优点从说明书、附图和权利要求中将变得清楚。

## 附图说明

[0011] 图1是图示示例网页浏览器环境的截屏,在该示例网页浏览器环境中从输入法编辑器输入识别资源位置。

[0012] 图2A图示了示例关键词词首(prefix)表。

[0013] 图2B图示了示例关键词资源表。

[0014] 图3是用于提供可选择资源位置的示例过程的流程图。

[0015] 图4是用于解析资源位置地址的示例过程的流程图。

[0016] 图5是能够实现在此描述的方法的示例系统的框图。

[0017] 在各附图中类似的参考数字和标记用于表示相似的元素。

## 具体实施方式

[0018] 图1是图示示例网页浏览器环境104的截屏100,在该示例网页浏览器环境104中,从输入法编辑器输入识别资源位置。地址栏103与IME地址环境101集成。在一些实现方式中,地址栏103可以用作标准URL输入栏,用户能够在其中输入网页地址,例如www.example.com。在一些实现方式中,输入法编辑器输入被提供给IME地址环境101,在所

述IME地址环境101中,输入到IME地址栏102中的罗马字符(例如,用于拼音输入的ASCII字符)被转换成非罗马字符(例如,汉字字符)。在一些实现方式中,罗马字符输入代表词符文本的一个或多个候选集合的音标表示,如汉字字符的拼音表示。IME地址环境101能够使用所输入的文本来自动定位统一资源定位符(URL),并且能够建议与罗马字符输入相关的一个或多个用户可选择的URL结果。

[0019] 尽管地址栏103和IME地址栏102被显示为独立的地址栏,但是IME地址栏可以集成到地址栏103中。此外,可以自动地或手动地实例化(instantiate)IME地址环境101。自动实例化可以在用户输入对应于音标输入(例如,“Xin”)并且不对应于标准资源位置地址(例如,www.example.com)的罗马字符时发生。手动实例化可以响应于用户输入的特定命令,例如键盘或菜单命令而发生。

[0020] 对于通常使用汉字或者利用表意字符的其他语言进行读写的用户来说,从罗马字符到非罗马字符的转换以及对URL的识别为他们提供了无需记住URL就能识别URL的便利。IME地址环境101提供能够接收罗马字符输入并且将罗马字符输入转换成非罗马字符的IME处理能力。在一些实现方式中,这些非罗马字符对应于关键词,所述关键词映射到URL。

[0021] 例如,用户可以向IME地址环境101的IME地址栏102输入拼音输入“xin lang”,并且该拼音输入可以被转换成非罗马字符候选集合106。非罗马字符的这些候选集合106在一些实现方式中是以非罗马字符的关键词112的关键词词首。关键词112通过关键词词首106来识别,因而关键词词首106能被用来识别相应的URL 108。

[0022] 因此,对每个非罗马字符候选集合106的选择能够使得相应的关键词112和URL 108被显示,如图1中所图示,图1示出了选择第一候选字符集合“新浪”以及相应的候选URL 108和关键词112。对另一个候选字符集合106的选择将相应地使得另一个URL 108的候选集合以及关键词112被显示。用户可以利用例如键盘或鼠标的输入设备选择URL中的一个(如围绕第二个URL108(“news.sina.com.cn”)的选择框202所例示的)及其相应的关键词。

[0023] 尽管示出了9个非罗马字符候选集合106,但是也可以显示更多或更少的非罗马字符集合。类似地,响应于对非罗马字符候选集合的选择,也可以显示更多或更少的资源位置108的候选集合和关键词112。

[0024] 在一些实现方式中,可以根据例如特定资源位置或关键词的流行度对每个候选集合108中的资源位置进行排名。资源位置108和相关联的关键词112可以例如按照预先的排名顺序来提供,所述预先的排名顺序依照基于点击日志(click log)或查询日志(query log)的选择频率和/或基于在查询日志中的关键词的排名。在一些实现方式中,可以将定义资源位置108、关键词112和相应的排名的数据提供给用户设备,并且可以基于用户设备的用户的选择历史对排名进行修改。例如,IME地址环境101能够跟踪每个资源位置已被用户设备的当前用户选择的次数,并且相应地调整资源位置的排名。例如,可以使用排名来基于资源位置的流行度以特定的优选顺序向用户设备的用户输出资源位置108的列表。以类似方式,可以基于用户设备的用户的选择来调整显示候选字符106的顺序。

[0025] 在一些实现方式中,关键词112到候选字符集合106的映射能够以数据结构来提供,所述数据结构诸如关键词词首表。图2A图示了示例关键词词首表200。在一些实现方式中,表200能够由外部服务器创建并且被提供给客户端设备上的网页浏览器。关键词词首表200包括关键词词首列202、关键词列204和嵌套的关键词列206。罗马字符不包括在关键词

词首列表中,但是在图示中被包括以便于解释。

[0026] 第一行208包括在列202中的关键词词首110,其表示词语“sina”的汉字字符。在示例中,关键词词首110能够被分支成包括“Sina Homepage”(新浪主页)210和“Sina Sports”(新浪体育)212的两个关键词。在一些实现方式中,可以从关键词词首分支出更多的关键词。

[0027] 第二行214包括在列202中的关键词词首216“United”以及表示该关键词词首“United”的汉字字符。在示例中,关键词词首216能够被分支成包括“United Nations”218和“United Airlines”220的两个关键词。“United Nations”关键词218可以额外地被分支成进一步的关键词。例如,关键词218被示为分支成包括“United Nations Charter”222和“Secretary of UN”224的另外两个关键词。

[0028] 表200中的关键词能够映射到资源位置。图2B图示了示例关键词资源表250。关键词资源表250是关键词与资源位置(诸如统一资源定位符)之间的映射,以及依照用户偏好、点击率或其他度量的统一资源定位符的排名。表250能够由外部服务器创建并且被提供给客户端设备上的网页浏览器。在一些实现方式中,表250能够通过确定响应于包括每个关键词的查询而针对该关键词选择每个URL的频率来创建。这些频率能够例如通过挖掘与搜索引擎相关联的查询日志和点击日志来获得。基于这些频率,对于每个关键词候选集合,将统一资源定位符相对于其他统一资源定位符进行排名。

[0029] 表250包括字符列252、URL列254和排名列256。字符列252包括汉字字符。所述字符代表用户输入的关键词。URL列254包括响应于包括列252中示出的关键词汉字字符的查询而已被选择的统一资源定位符的真子集。排名列256包括列254中的每个统一资源位置的排名分值。排名指示关于相应关键词选择每个统一资源位置的相对频率,并且也确定向用户显示关键词和统一资源定位符的顺序。

[0030] 在一个示例中,行258包括第一关键词260、相关的统一资源定位符262、和统一资源定位符的排名264。排名最靠前的统一资源定位符是“www.sina.com”,并且被评分为9.75,而排名较低的统一资源定位符被分别排名为9.22和2.21。因而,对于字符260所代表的关键词,最可能要被选择的统一资源定位符是三个统一资源定位符262。以类似的方式,行266显示第二关键词268、相关的统一资源定位符270和统一资源定位符的排名272。

[0031] 最初可以基于对查询日志和点击日志的挖掘来设置这些排名。在一些实现方式中,可以基于用户输入调整在客户端设备上的特定表250中的排名。

[0032] 网页浏览器能够使用表200和250来在IME地址环境101中选择和显示资源位置108和相应的关键词112。例如,运行实例化IME地址环境101的软件的客户端设备能够访问词首表200,以基于候选字符集合106识别关键词。然后,这些关键词能够被用作关键词位置表250中的索引,以用来识别关于所选择的候选字符集合的最流行的资源位置。例如,如果用户输入了导致识别出关键词词首“United”的拼音,则IME地址环境101将包括建议的关键词“United Nations”和“United Airlines”以及在关键词资源表250中索引的相应的资源位置。此外,IME地址环境101还将建议“United Nations Charter”和“Secretary of UN”以及相应的资源位置。

[0033] 在一些实现方式中,表200和250可以驻留在服务器设备上,如搜索引擎服务器。在这些实现方式中,服务器设备能够执行IME地址环境101中的所有操作或其一部分。例如,客

户端设备上的浏览器能够执行基本的IME功能,诸如将拼音输入转换为候选字符集合106。候选字符集合106能够被提供给服务器设备,进而服务器设备能够将资源位置108的候选集合和关键词112提供回客户端设备,以在IME地址环境101中进行显示。

[0034] 图3是用于提供可选择的资源位置的示例过程300的流程图。过程300能够例如实现在图5中描述的设备502和/或设备504中。

[0035] 过程300接收罗马字符输入,如ASCII字符串(302)。例如,IME地址环境101包括IME地址栏102,其能够接收罗马字符输入。在一个示例中,用户可以在设备504上显示的地址栏102中输入中文字的拼音表示。例如,用户可以在IME地址栏102中输入词语“ping”。词语“ping”是词语“苹果”的拼音表示。在一些实现方式中,拼音表示“ping”可以是关键词词首。例如,“ping”可以是词语“苹果派”的词首。一般地,例如,能够从诸如西式键盘的键盘或与手写识别引擎一起使用的触笔输入表示“苹果”的罗马字符“ping”。

[0036] 过程300确定是否在地址栏中接收到输入(304)。例如,设备504能够确定罗马字符输入(即,关键词词首)是否放置在URL栏、搜索栏、文本编辑器或其它程序中。如果输入未被输入到URL栏(例如,IME地址栏102)中,则过程300结束并且通过目的程序处理输入的文本。

[0037] 如果关键词词首被输入到URL栏中,则过程300使用罗马字符输入来识别非罗马字符候选集合(306)。例如,过程300能够包括将罗马字符输入从拼音转换成中文(例如,汉字字符)的输入法编辑器(IME)过程。该IME过程能够在服务器设备502或客户端设备504处实现。

[0038] 在将罗马字符转换成非罗马字符后,过程300识别与非罗马字符候选集合相对应的资源位置候选集合(308)。例如,过程300通过识别关键词词首表200中共享相同非罗马关键词词首的关键词来识别关于每个非罗马字符集合的关键词。对关键词的识别可以在服务器设备502处或客户端设备504处实现。

[0039] 过程300识别相关联的资源位置(例如,URL)并且将该资源位置与非罗马字符候选集合进行匹配(310)。例如,过程300将从关键词词首表200中识别出的关键词用作对关键词资源表250的索引,以识别相关联的资源位置。对相关联的资源位置的识别可以在服务器设备502处或客户端设备504处实现。

[0040] 如果未找到资源位置,则过程300结束。如果找到了资源位置,则过程300将资源位置候选集合和非罗马字符候选集合提供给用户界面以供显示(312)。例如,如果在服务器设备502处识别出资源位置,则服务器设备能够将资源位置和相关联的关键词提供给客户端设备504。在接收到相关联的关键词和资源位置后,或者如果在客户端设备504处识别出资源位置和关键词,则关键词和资源位置能够被提供给显示接口,例如网页浏览器界面。之后,过程300响应于对候选字符集合的选择而显示资源位置108和相应的关键词112。

[0041] 在一些实现方式中,可以根据选择频率对资源位置进行排名。在一些实现方式中,可以为每个关键词选择排名最高的URL,并且将排名最高的URL与非罗马字符候选集合相关联。

[0042] 图4是用于解析资源位置地址的示例过程400的流程图。过程400能够在例如图5描述的设备502中实现。

[0043] 过程400根据查询日志401a中的查询以及在点击日志401b中识别出的被选资源位置识别关键词和资源位置。例如,服务器设备502上的数据挖掘引擎可以考虑查询日志401a

和点击日志401b,其中,查询日志401a存储来自许多用户的查询,而点击日志401b存储对资源位置的选择以及为其识别出所述资源位置的相应的查询。

[0044] 过程400能够生成关键词到资源位置的映射(404)。在一些实现方式中,每个资源位置表示用户响应于所输入的包括关键词的搜索查询而选择的位置。映射可以存储在数据库(例如,关键词资源表250)中。在一些实现方式中,映射关键词可以包括将关于每个关键词的每个资源位置相对于关于该关键词的其他资源位置进行排名。在一些实现方式中,排名可以基于各资源位置中的每一个已被选择的次数。也可以使用其他排名度量。然后,可以通过服务器设备502中的数据挖掘引擎生成映射,并且该映射被存储在关键词资源表250中。

[0045] 过程400能够生成词首到关键词的映射(406)。例如,服务器设备502中的数据挖掘引擎能够使用查询日志401生成包括每个关键词或其词首的词首树。词首树中的每个节点可以对应于关键词或多个关键词共同的词首。词首表200中的示例节点包括由列202和204定义的节点。该信息可以存储在关键词词首表200中。在一些实现方式中,可以将来自查询日志401a的关键词的真子集存储在关键词词首表200中,例如仅仅存储那些在某一时间段内使用超过一定次数的关键词。

[0046] 过程400能够向客户端设备提供词首到关键词的映射以及关键词到资源位置的映射(408)。例如,服务器设备502能够向客户端设备502提供关键词词首表200和关键词位置表250。

[0047] 图5是能够实现在此描述的方法的示例系统500的框图。系统500包括服务器处理设备502,其跨越因特网连接506与客户端处理设备504进行数据通信。尽管在图5中仅图示了一个客户端和一个服务器,但是可以包括任意数目的服务器和/或客户端。

[0048] 服务器502包括输入法编辑器代码510,其可由客户端设备504执行并且能够被提供给客户端设备504。在被客户端设备504执行后,输入法编辑器代码510实例化客户端设备504上的IME地址环境101。输入法编辑器代码510能够包括语言模型(未示出),用于处理罗马字符输入(例如,拼音输入)以及将罗马字符输入转换成非罗马字符输入。

[0049] 服务器502还包括如在上面的图2A和2B中描述的关键词词首表200和关键词资源表250。也可以使用其他存储架构。这些表200和250能被提供给客户端设备。

[0050] 服务器502还包括挖掘引擎514,其能够挖掘查询日志401a和点击日志401b,以生成关键词词首表200和关键词资源表250。在一些实现方式中,可以将挖掘引擎514配置为识别超过频率阈值的关键词,例如在一定时间段内使用超过最小次数的关键词,以便排除罕见或很少使用的关键词。该选择过程通过忽略不太可能被用户输入或选择的关键词来降低关键词词首表200的大小。也可以使用其他选择标准。

[0051] 在一些实现方式中,可以将挖掘引擎514进一步配置为识别被选择超过频率阈值的资源,例如网站,例如当在查询中使用关键词时被选择了至少最小次数的资源位置。该选择过程通过忽略不太可能被用户选择的资源位置来降低关键词资源表250的大小。也可以使用其他选择标准。

[0052] 在被提供给客户端设备504时关键词资源表250可以包括初始排名。这些初始排名反映了基于查询日志401a和点击日志401b得到的选择概率。在一些实现方式中,IME代码514能够跟踪客户端设备504上的用户选择,并且调整存储在客户端设备504上的关键词资

源表250中的排名。因而,能够基于每个用户的浏览历史为每个用户定制对关键词和资源位置的识别。

[0053] 尽管在特定语言和输入法示例的语境中进行了描述,但是在此描述的系统和方法能够用于利用表意文字书写的任何语言。例如,在此描述的系统和方法能够用于包括表意文字的许多其他语言,例如日语。

[0054] 在本说明书中描述的主题和功能操作的实施例可以以数字电子电路、或计算机软件、固件或硬件(包括在本说明书中公开的结构及其结构等价物)或它们中的一个或多个的组合来实现。可以将在本说明书中描述的主题的实施例实现为一个或多个计算机程序产品,即,编码在有形的程序载体上以供数据处理装置执行或控制数据处理装置的操作的计算机程序指令的一个或多个模块。有形的程序载体可以是传播信号或计算机可读介质。传播信号是人工生成的信号,例如机器生成的电信号、光信号或电磁信号,生成所述传播信号以编码信息,以便传输到适合的接收器装置以供计算机执行。计算机可读介质可以是机器可读存储设备、机器可读存储基片(storage substrate)、存储器设备、影响机器可读传播信号的物质成分、或它们中的一个或多个的组合。

[0055] 术语“处理设备”涵盖用于处理数据的所有装置、设备和机器,例如包括可编程处理器、计算机、或多个处理器或计算机。装置除了硬件外还可以包括创建所讨论的计算机程序的执行环境的代码,例如构成处理器固件、协议栈、数据库管理系统、操作系统或它们中的一个或多个的组的代码。

[0056] 计算机程序(也称为程序、软件、软件应用、脚本或代码)可以以任何形式的编程语言来编写,包括编译语言或解释语言、或者声明性语言或过程语言,并且其可以以任何形式来部署,包括作为单机程序或作为模块、组件、子例程或适合于在计算环境中使用的其它单元。计算机程序不必对应于文件系统中的文件。可以将程序存储在拥有其它程序或数据的文件的一部分中(例如存储在标记语言文档中的一个或多个脚本)、存储在专用于所讨论的程序的单个文件中、或存储在多个协作文件中(例如存储一个或多个模块、子程序或代码部分的文件)。可以将计算机程序部署为在一个计算机或位于一个地点或跨多个地点分布并且通过通信网络互联的多个计算机上执行。

[0057] 作为示例,适合于执行计算机程序的处理器包括通用和专用微处理器两者、以及任何类型的数字计算机的任何一个或多个处理器。通常,处理器将从只读存储器或随机存取存储器或它们两者接收指令和数据。计算机的必要元件是用于执行指令的处理器和用于存储指令和数据的一个或多个存储设备。通常,计算机也将包括一个或多个用于存储数据的大容量存储设备,或者被可操作地耦接到一个或多个大容量存储设备以从其接收数据或者将其传送数据,或者两者均可,所述大容量存储设备例如磁盘、磁光盘或光盘。然而,计算机并非一定要具有这样的设备。此外,计算机可被嵌入另一个设备中,所述另一个设备例如移动电话、个人数字助理(PDA)、移动音频或视频播放器、游戏控制台,等等。

[0058] 适合于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质和存储设备,作为示例包括例如EPROM、EEPROM和闪存设备的半导体存储设备;例如内部硬盘或可移动盘的磁盘;磁光盘;以及CD-ROM和DVD-ROM盘。处理器和存储器可由专用逻辑电路补充或者并入专用逻辑电路中。

[0059] 为了提供与用户的交互,在本说明书中描述的主题的实施例可被实现在具有用于

向用户显示信息的显示设备(例如,CRT(阴极射线管)或LCD(液晶显示器)监视器)和用户能够通过其向计算机提供输入的键盘和指示设备(例如,鼠标或轨迹球)的计算机上。也可以使用其它种类的设备来提供与用户的交互;例如,提供给用户的反馈可以是任意形式的感官反馈,例如视觉反馈、听觉反馈或触觉反馈;并且可以以包括声音、语音或触觉输入的任意形式接收来自用户的输入。

[0060] 在本说明书中描述的主题的实施例能够被实现在计算系统中,所述计算系统包括后端组件,例如作为数据服务器;或包括中间件组件,例如应用服务器;或包括前端组件,例如具有图形用户界面或网页浏览器的客户端计算机,用户能够通过所述图形用户界面或网页浏览器与在本说明书中描述的主题的实现方式进行交互;或者包括一个或多个这样的后端组件、中间件组件或前端组件的任意组合。

[0061] 尽管本说明书包含许多具体的实现方式细节,但是这些不应当被解释为对任何发明的范围或者可能主张的专利权利的范围的限制,相反地,应被解释为是对可能特定于具体发明的具体实施例的特征的描述。在本说明书中,在不同实施例的上下文中描述的某些特征也可在单个实施例中以组合的方式实现。相反地,在单个实施例的上下文中描述的各种特征也可单独地或以任何合适的子组合形式在多个实施例中实现。此外,尽管特征可能在上文被描述为以某种组合形式起作用,甚至最初被主张为以某种组合形式起作用,但是来自所主张的组合的一个或多个特征在一些情况下可从组合中去除,并且所主张的组合可被指向子组合或子组合的变型。

[0062] 类似地,虽然操作在附图中以特定的次序描述,但是这不应当被理解为必须以示出的特定次序或以顺序地次序来执行这些操作,或者必须执行所有图示的操作,才能获得期望的结果。在某些情况中,多任务和并行处理可能是有益的。此外,在上文描述的实施例中的各个系统组件的分割不应当被理解为在所有的实施例中必须是这样的分割,并且应当理解所描述的程序组件和系统通常可以被一起集成在单个软件产品中或打包到多个软件产品中。

[0063] 已描述了在本说明书中描述的主题的特定实施例。其它实施例也在权利要求的范围内。例如,可以以不同的次序执行在权利要求书中叙述的动作而仍然获得期望的结果。作为一个示例,在附图中描述的过程并非必须要求示出的特定次序或者顺序的次序才能获得期望的结果。在某些实现方式中,多任务和并行处理可能是有益的。

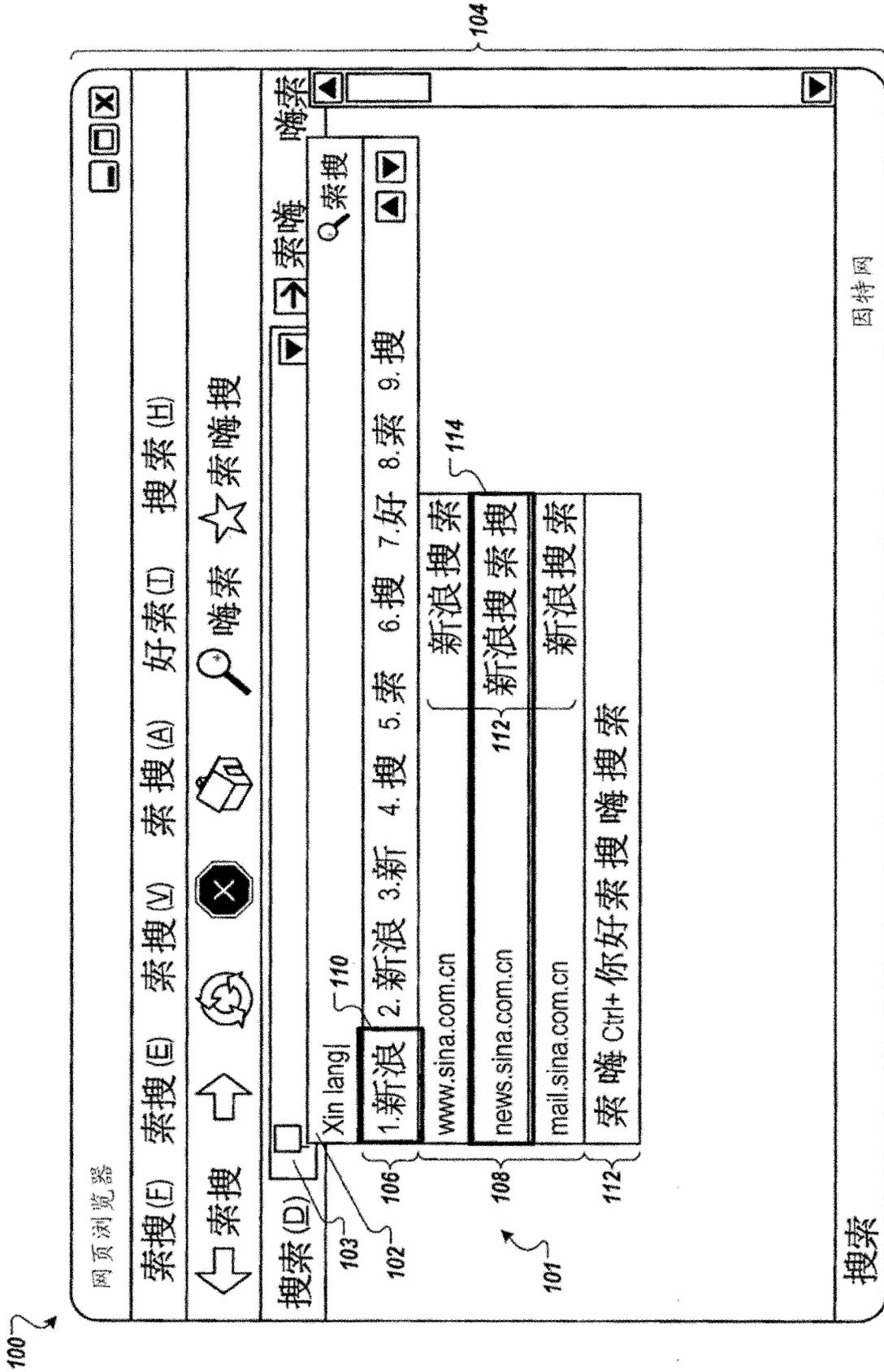


图1

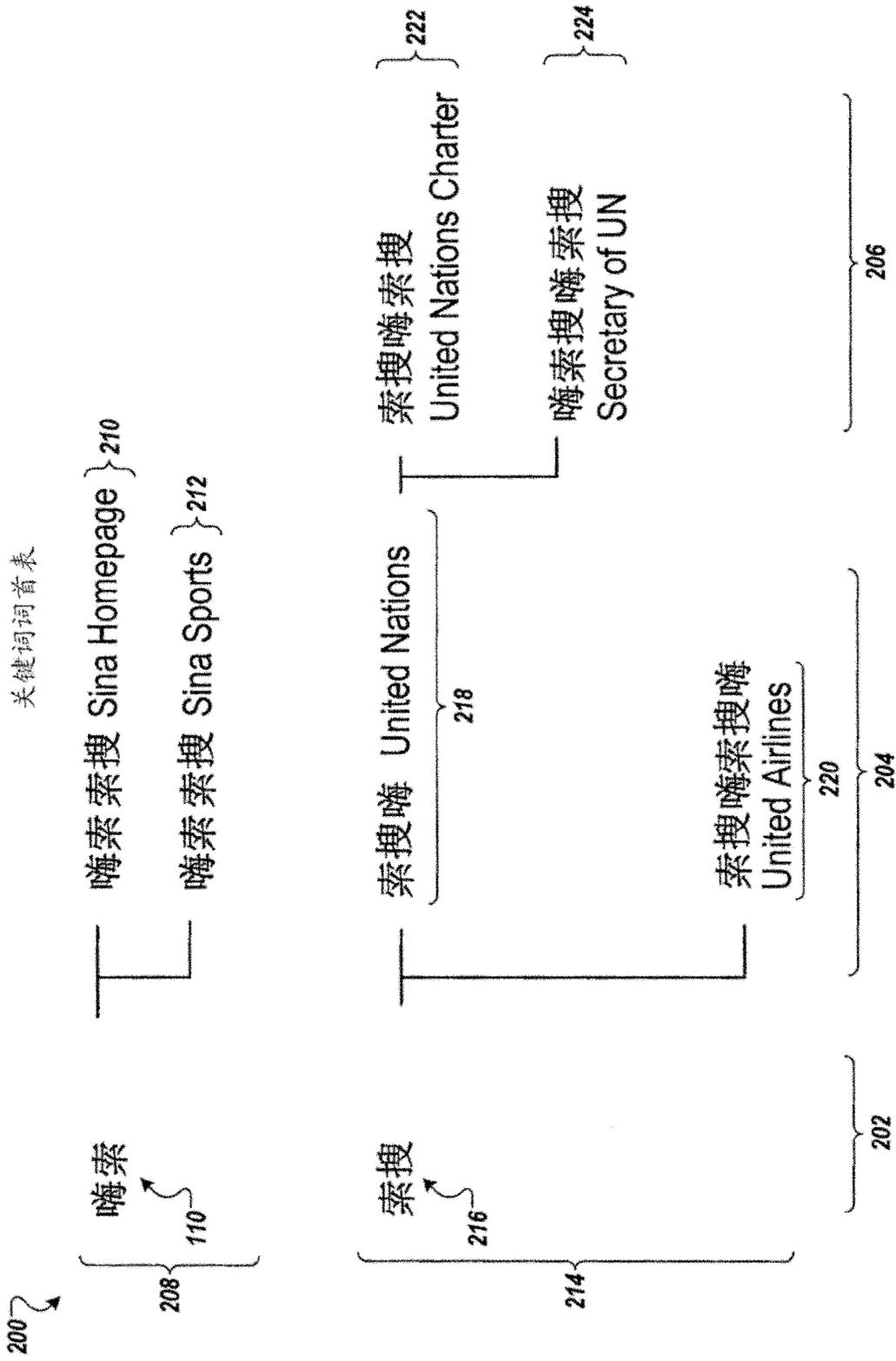


图2A

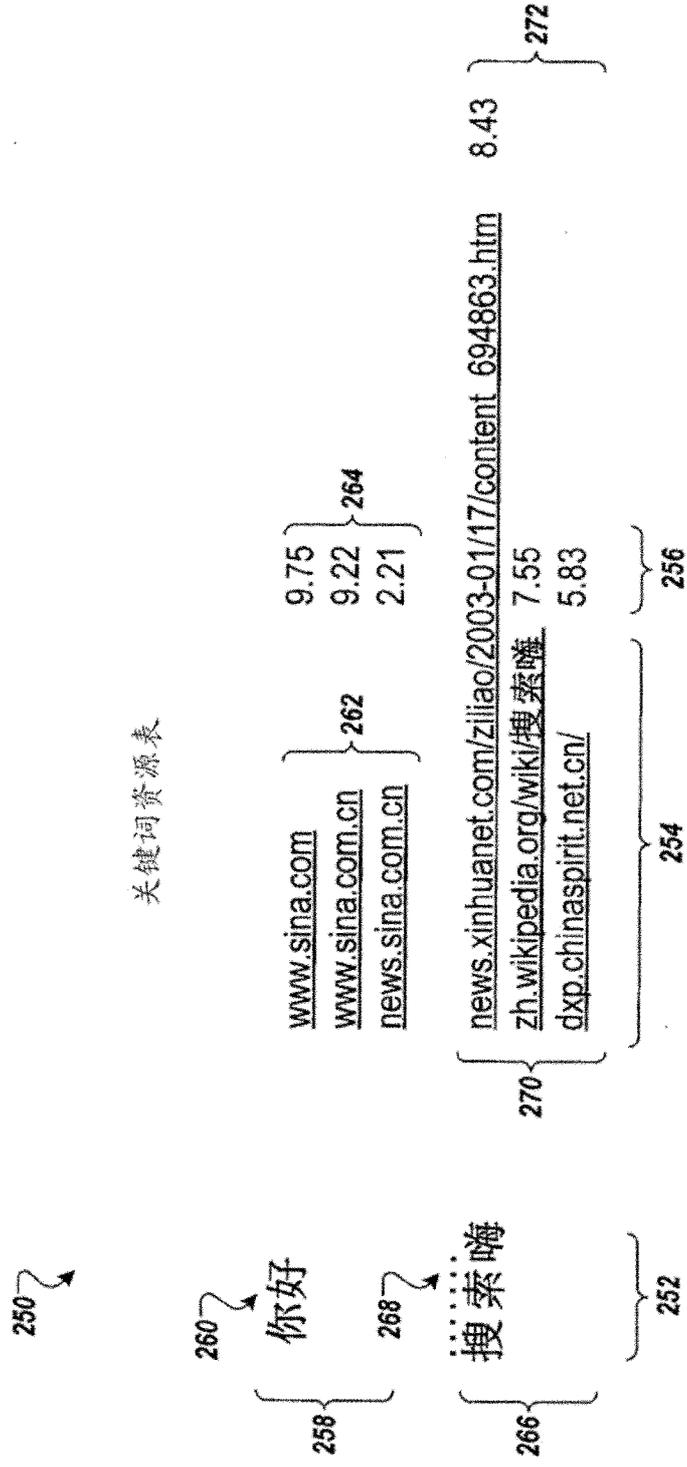


图2B

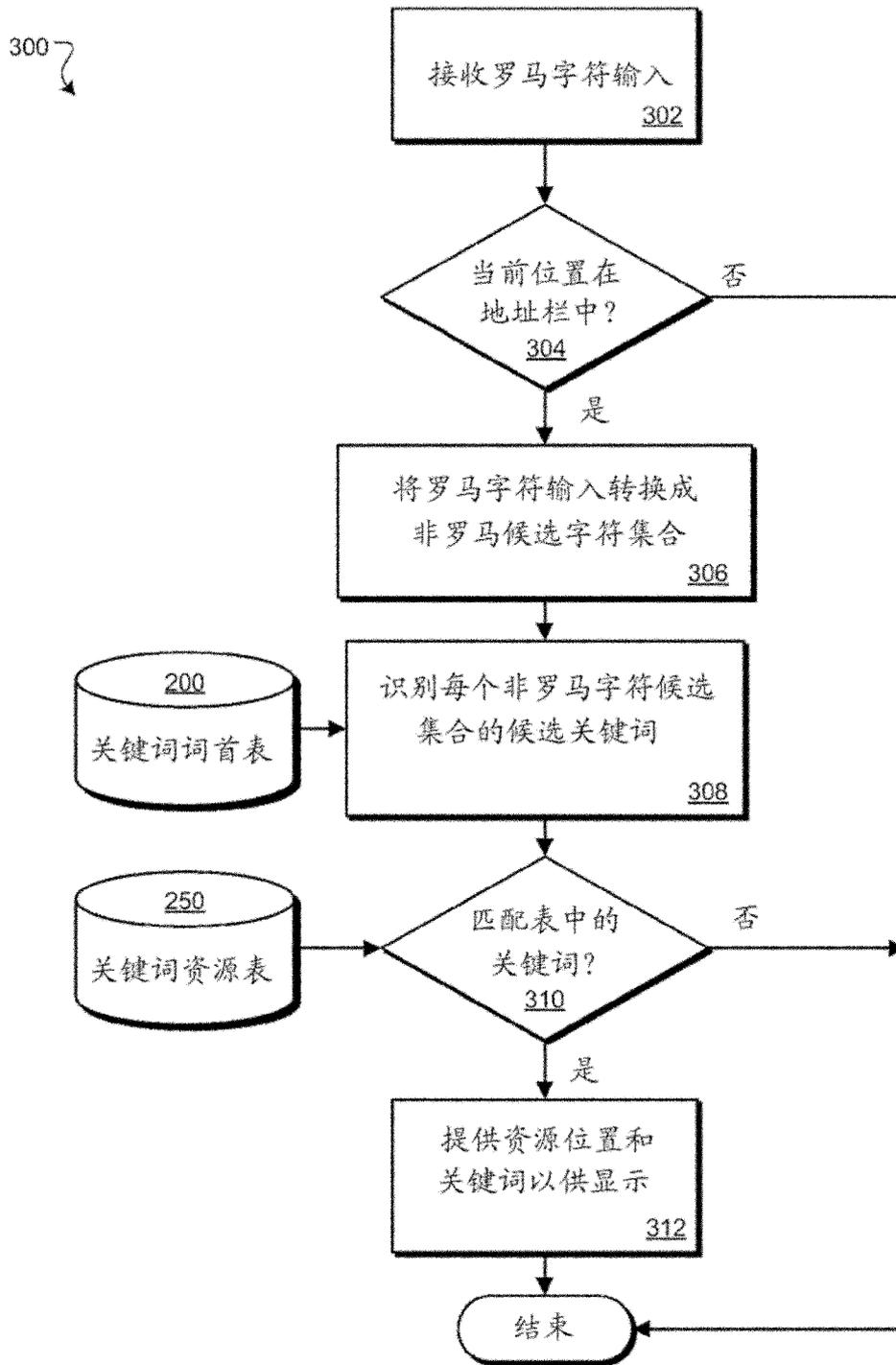


图3

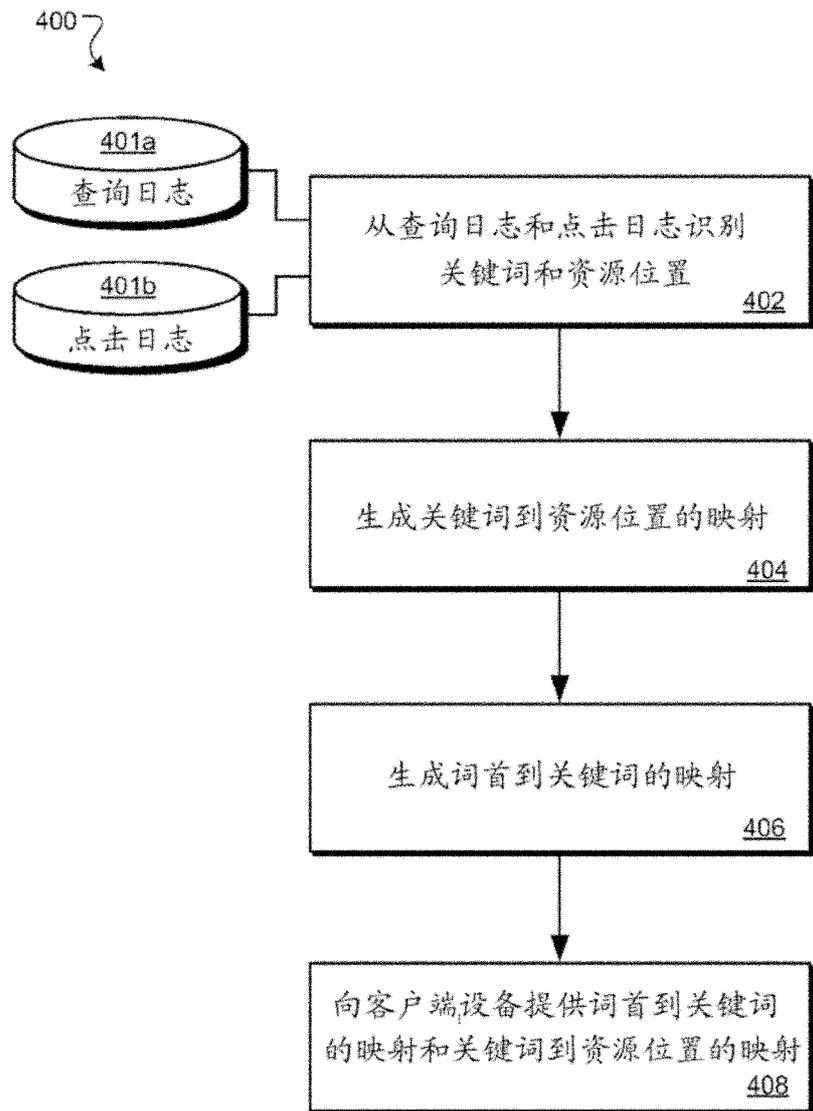


图4

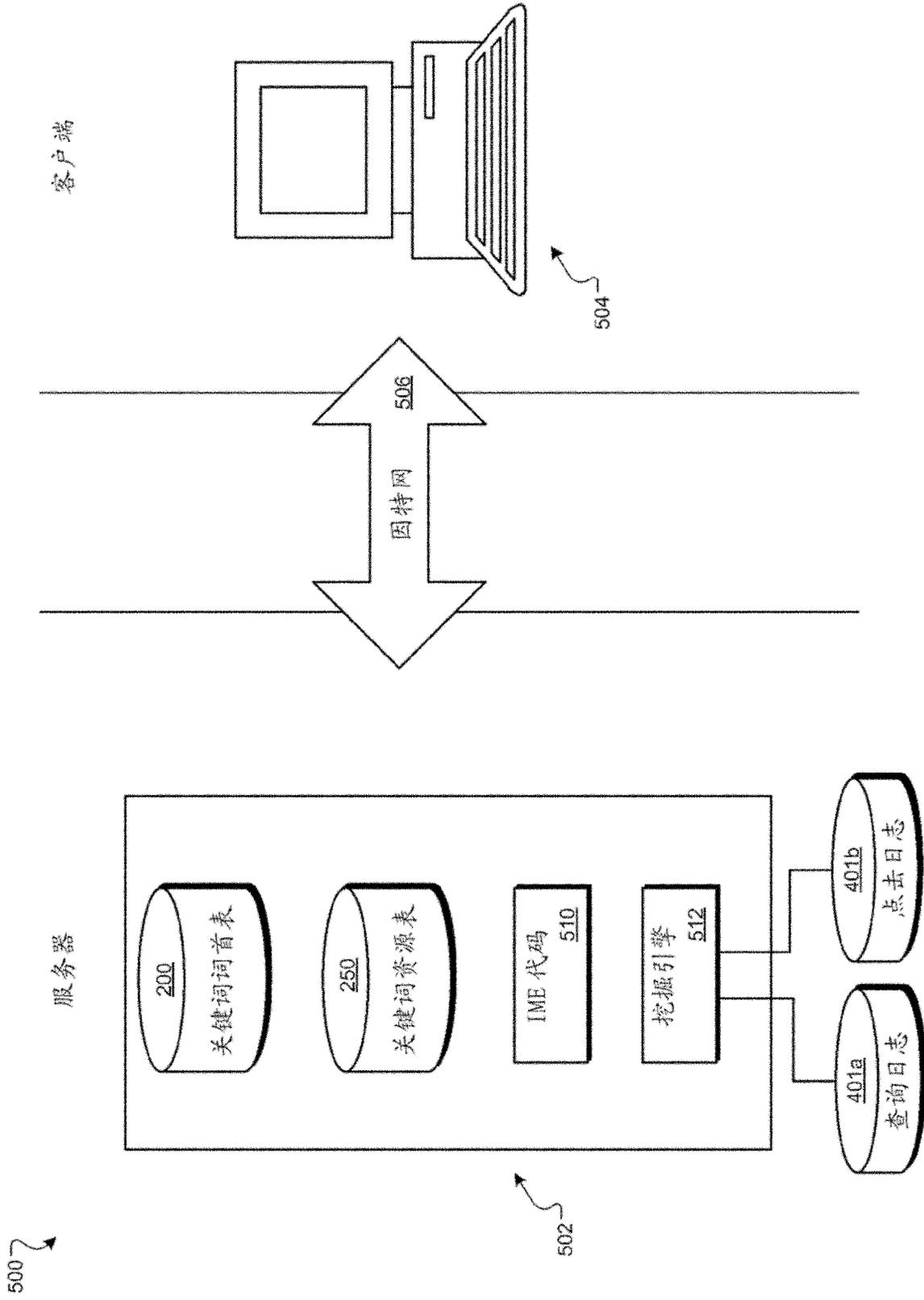


图5