

(12) 发明专利申请

(10) 申请公布号 CN 102063453 A

(43) 申请公布日 2011. 05. 18

(21) 申请号 201010187091. 6

(22) 申请日 2010. 05. 31

(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 孙云丰 金海燕 张玉睿 徐波  
李明

(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

(51) Int. Cl.

G06F 17/30 (2006. 01)

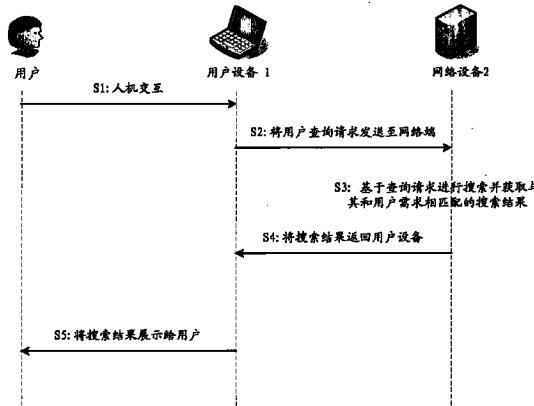
权利要求书 3 页 说明书 11 页 附图 4 页

(54) 发明名称

一种用于基于用户的需求进行搜索的方法和设备

(57) 摘要

本发明提供了一种用于基于用户的需求进行搜索的方法和设备。根据本发明，首先获取该用户的查询请求；随后根据所述查询请求进行匹配搜索，以获得与所述查询请求和用户需求相匹配的一个或多个搜索选项；将所述一个或多个搜索选项提供给所述用户，供其选择。具体地，本发明通过对查询请求进行文义分析获得其中内含的用户实际需求，或进一步结合大样本统计信息、热点信息和 / 或用户的相关信息来更准确判断用户的实际需求。从而，与现有的搜索技术相比，可以准确地获得与查询请求信息以及用户实际需求相匹配的搜索结构。



1. 一种用于基于用户的需求进行搜索的方法,该方法包括以下步骤:
  - a 获取该用户的查询请求;
  - b 根据所述查询请求进行匹配搜索,以获得与所述查询请求和用户需求相匹配的一个或多个搜索选项;
  - c 将所述一个或多个搜索选项提供给所述用户,供其选择。
2. 根据权利要求 1 所述的方法,其中,所述步骤 b 包括:
  - 根据所述查询请求进行匹配搜索,获得多个搜索选项;
  - 根据所述查询请求来判断用户需求;
  - 根据所述用户需求来对所述多个搜索选项进行筛选处理,获得与所述查询请求和所述用户需求相匹配的一个或多个搜索选项。
3. 根据权利要求 1 所述的方法,其中,所述步骤 b 还包括:
  - 根据所述查询请求来判断用户需求;
  - 根据所述查询请求及所述用户需求来进行匹配查询,以获得与所述查询请求和所述用户需求相匹配的一个或多个搜索选项。
4. 根据权利要求 2 或 3 所述的方法,其中,所述判断用户需求的步骤还包括:
  - 对所述查询请求进行文义分析,以确定所述用户的所述用户需求。
5. 根据权利要求 2 至 4 中任一项所述的方法,其中,所述判断用户需求的步骤还包括:
  - 根据所述查询请求来查询预先记录的与所述查询请求相关的大量用户选择的热门搜索项,并根据所述大量用户选择的热门搜索项来确定所述用户的用户需求。
6. 根据权利要求 5 所述的方法,其中,还包括:
  - 进行在线挖掘,以收集并记录热门关键词以及与其相关联的大量用户选择的热门搜索项。
7. 根据权利要求 6 所述的方法,其中,所述进行在线挖掘的步骤还包括:
  - 根据用户点击率来确定热门关键词;
  - 根据以下至少一项确定所述热门关键词的热门搜索项:
    - 所述热门关键词的时效性;
    - 所述热门关键词的主题类别;
    - 所述热门关键词的主题需求类别。
8. 根据权利要求 2 至 7 中任一项所述的方法,其中,所述判断用户需求的步骤还包括:
  - 获取所述用户的用户相关信息;
  - 基于所述用户相关信息来对所述查询请求进行分析,以确定所述用户的所述用户需求。
9. 根据权利要求 8 所述的方法,其中,所述用户相关信息包括以下至少一项:
  - 用户属性;
  - 用户偏好;
  - 用户查询历史记录;
  - 用户 IP 地址。
10. 根据权利要求 9 所述的方法,其中,该方法还包括:
  - d. 获取所述用户在为其提供所述一个或多个搜索选项的页面上的页面操作,并根据所

述页面操作更新所述用户的所述用户相关信息。

11. 根据权利要求 10 所述的方法,其中,所述步骤 d 还包括以下步骤 :

- 当所述页面操作包括翻页、刷新页面和重新搜索其中之一时,跳转至步骤 b。

12. 一种用于基于用户的需求进行搜索的设备,其中,该设备包括 :

获取装置,用于获取该用户的查询请求 ;

查询装置,用于根据所述查询请求进行匹配搜索,以获得与所述查询请求和用户需求相匹配的一个或多个搜索选项 ;

提供装置,用于将所述一个或多个搜索选项提供给所述用户,供其选择。

13. 根据权利要求 12 所述的设备,其中,所述查询装置还包括 :

第一查询装置,用于根据所述查询请求进行匹配搜索,获得多个搜索选项 ;

需求判断装置,用于根据所述查询请求来判断用户需求 ;

筛选装置,用于根据所述用户需求来对所述多个搜索选项进行筛选处理,获得与所述查询请求和所述用户需求相匹配的一个或多个搜索选项。

14. 根据权利要求 12 所述的设备,其中,所述查询装置还包括 :

需求判断装置,用于根据所述查询请求来判断用户需求 ;

第二查询装置,用于根据所述查询请求及所述用户需求来进行匹配查询,以获得与所述查询请求和所述用户需求相匹配的一个或多个搜索选项。

15. 根据权利要求 13 或 14 所述的设备,其中,所述需求判断装置还用于 :

- 对所述查询请求进行文义分析,以确定所述用户的所述用户需求。

16. 根据权利要求 13 至 15 中任一项所述的设备,其中,所述需求判断装置还用于 :

- 根据所述查询请求来查询预先记录的与所述查询请求相关的大量用户选择的热门搜索项,并根据所述大量用户选择的热门搜索项来确定所述用户的用户需求。

17. 根据权利要求 16 所述的设备,其中,还包括 :

在线挖掘装置,用于进行在线挖掘,以收集并记录热门关键词以及与其相关联的大量用户选择的热门搜索项。

18. 根据权利要求 17 所述的设备,其中,所述在线挖掘装置还用于 :

- 根据用户点击率来确定热门关键词 ;

- 根据以下至少一项确定所述热门关键词的热门搜索项 :

- 所述热门关键词的时效性 ;

- 所述热门关键词的主题类别 ;

- 所述热门关键词的主题需求类别。

19. 根据权利要求 13 至 18 中任一项所述的设备,其中,所述需求判断装置还用于 :

- 获取所述用户的用户相关信息 ;

- 基于所述用户相关信息来对所述查询请求进行分析,以确定所述用户的所述用户需求。

20. 根据权利要求 19 所述的设备,其中,所述用户相关信息包括以下至少一项 :

- 用户属性 ;

- 用户偏好 ;

- 用户查询历史记录 ;

- 用户 IP 地址。

21. 根据权利要求 20 所述的设备, 其中, 该设备还包括 :

更新装置, 用于获取所述用户在为其提供所述一个或多个搜索选项的页面上的页面操作, 并根据所述页面操作更新所述用户的所述用户相关信息。

22. 根据权利要求 21 所述的设备, 其中, 所述查询装置还用于 :

- 当所述页面操作包括翻页、刷新页面和重新搜索其中之一时, 用于将所述页面操作作为新的查询请求并根据所述新的查询请求进行匹配搜索, 以获得与所述查询请求和用户需求相匹配的一个或多个搜索选项。

## 一种用于基于用户的需求进行搜索的方法和设备

### 技术领域

[0001] 本发明涉及网络技术领域，尤其涉及基于用户的需求进行搜索的技术。

### 背景技术

[0002] 自从 70、80 年代互联网普及以来，为满足用户在互联网上访问海量信息的要求，出现了许多网络搜索技术，我们简单回顾如下：

[0003] 90 年代早期出现基于人工信息分类的 EINet Galaxy 搜索引擎。随后与 19994 年 Lycos 搜索引擎推出，其包括基于机器人程序 (Robot) 的数据挖掘技术和支持搜索结果相关性排序。

[0004] 1995 年的 AltaVista 推出的搜索引擎第一个支持自然语言搜索的搜索引擎，具备了基于网页内容分析，智能处理的能力，第一个实现高级搜索语法的搜索引擎（如 AND、OR、NOT 等）。

[0005] 1997 年，Northernlight 公司推出的搜索引擎第一个支持对搜索结果进行简单的自动分类，也是当时拥有最大数据库的搜索引擎之一。

[0006] 1998 年，Google 推出基于其 PAGERANK 算法的搜索引擎，其根据网页之间相互的超链接计算的网页排序。

[0007] 2000 年，百度创始人李彦宏推出基于超链分析的百度搜索引擎。

[0008] 现有的搜索引擎的基本工作原理如下：

[0009] 1) 自动搜集信息：搜索引擎的信息搜集基本都是自动的。搜索引擎利用称为网络蜘蛛 (spider) 的自动搜索机器人程序来连上每一个网页上的超链接。机器人程序根据网页链接到其中的超链接，就象日常生活中所说的“一传十，十传百……”一样，从少数几个网页开始，连到数据库上所有到其他网页的链接。理论上，若网页上有适当的超链接，机器人便可以遍历绝大部分网页。

[0010] 2) 建立索引库：搜索引擎整理信息的过程称为“建立索引”。搜索引擎不仅要保存搜集起来的信息，还要将它们按照一定的规则进行编排并保存于索引库中。这样，搜索引擎可以根据查询信息在索引库中迅速找到所要的资料。

[0011] 3) 接受查询：用户向搜索引擎发出查询，搜索引擎接受查询并向用户返回资料。搜索引擎每时每刻都要接到来自大量用户的几乎是同时发出的查询，它按照每个用户的要求检查自己的索引，在极短时间内找到用户需要的资料，并将搜索结果返回给用户，其中搜索结果一般包括网页链接及摘要描述。在有多个搜索结果的情形下，需要利用各种算法进行排序，其中，最著名的搜索排序算法为 Google 的 PageRank 算法（参见美国专利 US6285999），其原理是根据网页之间相互的超链接计算进行网页排序，搜索引擎把从 A 网页到 B 网页的链接解释为 A 网页给 B 网页投票，并根据投票来源（甚至来源的来源，即链接到 A 网页的网页）和投票目标的等级来决定新的等级。简单的说，一个高等级的网页可以使被其链接或引用的其他低等级网页的等级提升。

[0012] 其中，整理信息及接受查询的过程，大量应用了文本信息检索技术，并根据网络超

文本的特点，引入了更多的信息。

[0013] 对于文本信息检索，早期的信息检索系统采用“布尔查询”的方法来进行全文检索。用户必须详细的规划自己的查询，其复杂程度不亚于编程语言。这种检索方式并不提供任何的文档相关性测度，对于文档与查询的评价就只有“匹配”，“不匹配”两种而已。这两点问题决定了布尔查询不能被广泛应用。

[0014] 对于大规模的语料库，任何检索都可能返回数量众多的结果，因此对检索结果进行排序是必须的。因此，一个好的信息检索模型必须提供文档相关性测度。一个好的测度应该使与用户查询需求最相关的那些结果，排在最前面，同时允许尽可能多的，与用户查询有一定关系的结果被包括进来。目前，最为常用的信息检索模型有三种：

[0015] I 向量空间模型 (Vector Space Model, VSM)

[0016] II 概率模型 (Probabilistic Model)

[0017] III 推理网络模型 (Inference Network Model)

[0018] 简而言之，现有的搜索技术基本上均基于用户查询请求信息中所含关键词对索引库进行匹配搜索，根据匹配度或相似度获得多个搜索结果并根据点击率或超链信息进行筛选或排序，也即仅根据一些外在的、或客观信息来进行搜索，因此不能在搜索结果中反映用户的实际需求。

## 发明内容

[0019] 本发明的目的是为克服现有技术的缺陷，提供一种基于用户需求进行搜索的方法和设备。

[0020] 根据本发明的一个方面，还提供了一种用于基于用户的需求进行搜索的方法，该方法包括以下步骤：

[0021] a 获取该用户的查询请求；

[0022] b 根据所述查询请求进行匹配搜索，以获得与所述查询请求和用户需求相匹配的一个或多个搜索选项；

[0023] c 将所述一个或多个搜索选项提供给所述用户，供其选择。

[0024] 根据本发明的另一个方面，提供了一种用于基于用户的需求进行搜索的设备，其中，该设备包括：

[0025] 获取装置，用于获取该用户的查询请求；

[0026] 查询装置，用于根据所述查询请求进行匹配搜索，以获得与所述查询请求和用户需求相匹配的一个或多个搜索选项；

[0027] 提供装置，用于将所述一个或多个搜索选项提供给所述用户，供其选择。

[0028] 与之相比，本发明通过对查询请求进行文义分析获得其中内含的用户实际需求，或进一步结合大样本统计信息、热点信息和 / 或用户的相关信息来更准确地判断用户的实际需求，可以准确地获得与查询请求信息以及用户实际需求相匹配的搜索结构。

## 附图说明

[0029] 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：

- [0030] 图 1 为根据本发明一个方面的根据用户需求提供检索结果的方法流程图；
- [0031] 图 2 为根据本发明一个优选实施例的根据用户需求提供检索结果的方法流程图
- [0032] 图 3 为根据本发明一个方面的根据用户需求提供检索结果的系统框图；
- [0033] 图 4 为根据本发明一个优选实施例的根据用户需求提供检索结果的系统框图。
- [0034] 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

- [0035] 下面结合附图对本发明作进一步详细描述。
- [0036] 本发明可以用于网络搜索，也可用于桌面搜索。前者仅需在网络端实现，用户通过常用的用户端与其进行互动，获得网络搜索的搜索选项；而后者仅需在用户端来实现，为用户提供桌面搜索的搜索选项。
- [0037] 图 1 示出根据本发明一个方面的根据用户需求提供检索结果的网络搜索流程图。
- [0038] 如图 1 所示，在步骤 S1 中，用户在利用用户设备 1 进行网络浏览时，可通过诸如键盘、触摸屏、语音输入装置与用户设备 1 进行交互，输入其希望查询的查询信息 (query)，从而发起搜索。用户设备 1 包括桌面搜索引擎，也即在其本地保存有该搜索发起过程可以是用户首先通过用户设备 1 的浏览器软件访问一些搜索网站，如 Yahoo. com, Google. com, Baidu. com 后，在该搜索网站首页的搜索栏中发起搜索；也可通过嵌入浏览器的本地搜索栏发起搜索。本领域技术人员应能理解本发明不限于上述两种搜索发起方式，任何已有的或今后可能出现的搜索发起方式均适用于本发明，应包含在本发明的保护范围以内，并以引用方式而包含于此。
- [0039] 接着，在步骤 S2 中，用户设备 1 将用户所输入的查询信息发送给网络设备 2。在此，用户设备 1 包括，但不限于，计算机、智能手机、PDA 等任何具有上网浏览功能的设备。网络设备 2 包括，但不限于，网络服务器、网络主机、云计算模式下的其他用户设备 1 等可以提供搜索服务的任何设备，其中网络设备 2 包括集中各种功能的单个设备，或由多个分别完成单独功能的物理分离的设备的组合，例如执行搜索的设备与保存索引信息的设备相互分离。此外，用户设备 1 与网络设备 2 之间的通信是基于各种通信协议 (CommunicationsProtocol)，在此“通信协议”指计算机通信的传送协议，如：TCP/IP、UDP、FTP、ICMP、NetBEUI 等，同时还包括存在于计算机中的其他形式通信，例如：面向对象编程里面对对象之间的通信；操作系统内不同程序或计算机不同模块之间的消息传送协议。但本领域技术人员应理解本发明不限于所述几种通信传输协议，任何已有或今后可能出现的通信协议均可适用于本发明，从而以引用方式而包含于此。
- [0040] 随后，在步骤 S3 中，网络设备 2 基于接收到的查询信息在网络索引库中进行搜索。本领域技术人员应能理解，网络索引库可位于网络设备 2 中或与之物理分离。
- [0041] 其中，查询信息包含一个或多个关键词，关键词可以是字，词或者短语。在语音文档检索中，还可以是混淆类、音子、音子串等等单元，或者还包括关键词之间的关联词。如前所述，传统搜索引擎中，信息查询采用的模型，例如向量空间模型，一般通过将查询信息与索引库中的索引信息进行对比以确定两者相似度，从而确定作为搜索结果的一个或多个搜索引擎，其中，索引信息也由一些关键词来描述。
- [0042] 对于所获得的多个搜索结果，传统搜索引擎一般通过其点击率或 Google 的

PageRank 或百度的超链等算法来确定各个搜索结果的排序。但传统搜索引擎的信息查询或排序过程中并未考虑用户输入查询信息的实际需求,事实上,即使用户输入的查询请求中包含相同的关键词,但由于本身内涵文义的区别,反映的用户需求也相应不同,其中查询请求的内涵文义可由关键词之间及关联词的相互顺序,例如当用户分别输入查询请求“川菜最好吃”或“最好吃川菜”,虽然均包含相同关键词,但对于前者,用户更可能希望查询口碑最好的川菜馆,;而对于后者,用户更希望查询最受欢迎的川菜做法,也即,对于前者,用户需求“最好吃的川菜馆”优先级或可能性较高、用户需求“最好吃的川菜做法”优先级或可能性较低;对于后者,正好相反。

[0043] 因此,在步骤 S3 中,网络设备 2 将对查询请求的文义分析与搜索技术相结合,获得一个或多个较符合用户实际需求的搜索结果,而且这些搜索结果可以按与用户需求的匹配度进行排序。步骤 S3 的实现方式包括但不限于,1) 索引库中索引信息应包含用户需求这一项,网络设备 2 可以对查询请求进行文义分析以确定其反映的一个或多个用户需求及其对应的优先级或可能性,随后在基于查询请求对索引库进行查询时,将所确定的用户需求作为信息查询的一个维度,以确定相匹配的一个或多个索引信息作为搜索结果;2) 网络设备 2 可以根据前述已有的各种搜索技术基于查询信息在索引库中进行查询,获得多个搜索结果,随后再基于通过对查询请求文义分析所确定的一个或多个用户需求来对该多个搜索结果进行筛选和排序,以获得最终确定的按用户需求进行排序的搜索结果。本领域技术人员应理解本发明不限于上述两种方式,本领域技术人员根据本发明的构思应能想到的其他实现方式也应包含在本发明的保护范围以内。

[0044] 随后,在步骤 S4 中,网络设备 2 将搜索得到的搜索结果及其排序信息返回给用户设备 1。最后,在步骤 S5 中,用户设备 1 基于排序信息将搜索结果按序排列显示给用户。

[0045] 在一个优选实施例中,在前述步骤 S3 中,网络设备 2 还可根据大样本统计信息来确定关键词与用户需求的关系,例如,对于关键词“Elton John”,网络设备 2 根据大量用户对搜索结构的选择可以确定,通过对大量用户的选择的统计分析可确定其对应的用户需求“Elton John 的音乐”优先级或可能性较高、“Elton John 的新闻”优先级或可能性较低。

[0046] 在另一个优选实施例中,在步骤 S3 中,网络设备 2 还根据所述查询请求来查询预先记录的与所述查询请求相关的大量用户选择的热门搜索项,并根据所述大量用户选择的热门搜索项来确定所述用户的用户需求。例如 2010 年 4 月中国青海省玉树县发生 7 级以上地震,而网络设备 2 中关键词“玉树”与该热门搜索项“玉树地震”相关联,从而当接收到包含关键词“玉树”的查询请求时,可以确定用户希望了解“玉树地震”或“玉树地震赈灾”新闻的需求的优先级或可能性较高,相对的了解“玉树旅游”或“玉树历史”的优先级或可能性较低。

[0047] 优选地,在步骤 S6(未示出)中,网络设备 2 可以自行经由网络在线搜索热门搜索项,包括但不限于以下方式:

[0048] - 首先,通过大样本统计信息,也即通过统计用户对网页的点击率,并对点击率排序较高的网页进行信息抓取,确定相应的关键词;

[0049] - 接着,根据以下至少一项确定所述热门关键词的热门搜索项:

[0050] - 所述热门关键词的时效性;

[0051] - 所述热门关键词的主题类别；

[0052] - 所述热门关键词的主题需求类别。

[0053] 其中，主题类别是指关键词的所述类别，例如关键词“苹果”的主题类别上可以是水果、计算机、电影。而主题需求类别是指相对于关键词的用户需求的类别，例如当用户输入关键词“苹果”，如果其主题类型是属于计算机类别的，则主题需求类别可能是寻求苹果公司信息，或是寻求苹果电脑的报价。

[0054] 在另一个优选实施例中，考虑由于用户的个体差异，使得当不同用户输入包含相同关键词时，其实际需求也是因人而异，因此在步骤例如，对（成年）男性、（成年）女性、儿童、老年人等，当输入查询请求中包含“\*\*百货公司”时，其实际需求区别较大，其中，男性用户可能对运动产品、男性服饰、金融商业书籍和重金属摇滚乐 CD 等更感兴趣、女性用户可能对护肤 / 化妆品、女性服饰、浪漫言情书籍和轻音乐 CD 更感兴趣，儿童用户对儿童服饰、儿童读物和卡通 DVD 等更感兴趣，而老年用户对老年服饰、保养品、文史类书籍和古典音乐 CD 更感兴趣。当然，用户的个人背景，例如教育背景、宗教背景等也会成为其实际需求的影响要素，例如理工科毕业的用户可能对于技术类书籍更感兴趣，而经济类专业的用户对于金融或商战类书籍更感兴趣。另外，例如，对于旅游网站时，当用户处于不同地域时其所感兴趣的信息通常也是不同的，例如，对于处于北京地域的用户，当其输入的查询请求中包括关键词“旅游”时，其实际需求可能是北京周边区域（例如，北戴河、河北）等自驾游和旅店住宿信息，当其查询请求包括关键词“机票酒店”时，可能对由北京出发的特价机票更感兴趣。此外，用户以往的搜索历史记录也可以成为确定其实际需求的指示信息，例如，网络设备 2 通过查询一名用户的历史记录信息，发现其以往在访问红酒购物网站时更多地关注梅乐（MERLOT）红酒，则可当其输入的查询请求包括“\*\* 红酒网站”时，其实际需求可能是与梅乐红酒有关的信息。

[0055] 基于以上理解，在步骤 S7（未示出）中，在网络设备 2 接收到来自用户设备 1 的查询请求后，会查询该用户的用户相关信息，并在步骤 S3 中基于上述文义分析（或者进一步结合热点搜索项）并结合用户的相关信息来判断用户的实际需求，以获得与用户实际需求相匹配的搜索结果。如前所述，其中用户相关信息包括以下各项中至少一项：

[0056] - 用户属性，其包含，但不限于，用户预先输入的个人背景信息，如性别、年龄、宗教、国籍、教育程度等；

[0057] - 用户偏好，其包含，但不限于，用户预先设置的一些偏好选择，如喜爱的运动类型、购物方式、红酒品种、音乐类型等；

[0058] - 用户查询历史记录，其由系统来记录和统计用户先前的搜索和点击浏览历史，从而可判断或预测用户的需求。其中，当用户为登录用户时，网络设备 2 可通过该用户的 ID 在网络端保存有该用户相关信息的用户特征库中进行查询；当用户为非登录用户时，可通过检测该用户使用本网络设备 2 的搜索服务过程中的动作，或通过向用户设备 1 询问并获取该用户的 COOKIE 信息来获得，优选地，可赋予该用户一个临时 ID 并为其建立一个用户特征库，将此次搜索过程中用户的历史记录保存与该临时用户特征库中；

[0059] - 用户设备 1 的 IP 地址，其可以反映该用户当前所处地域，可由网络设备 2 实时检测用户设备 1 的 IP 地址来获取。

[0060] 优选地，在步骤 S8（未示出）中，网络设备 2 根据与用户设备 1 及用户的交互来记

录和更新所述用户历史记录信息,具体地,包括但不限于以下方式:1) 网络设备 2 可在用户一次搜索过程结束后记录用户在此搜索过程中的所有动作,例如点击和浏览;2) 网络设备 2 在用户一次搜索过程中记录用户每次动作并实时对网络端该用户的特征库进行更新,从而可在用户在此次搜索过程中点击其他搜索结果时提供最新的用户历史记录信息。

[0061] 在另外一个优选实施例中,在步骤 S5 用户设备 1 向用户展示获得的搜索结果之后,当用户对所显示的包含搜索结果的页面进行页面操作,例如包括但不限于,点击其中的某一搜索结果项所含链接、或点击下一页按钮以显示随后的搜索结果、或点击刷新页面按钮、或点击搜索按钮进行重新搜索时,该新交互信息经由用户设备 1 重新发送至网络设备 2,网络设备 2 将根据该新交互信息作为用户历史记录更新该用户的用户相关信息。优选地,网络设备 2 还会将该新交互信息作为新的查询请求,重复上述步骤 S3 以及 S6-S8,并将进一步的搜索结果返回用户设备 1。

[0062] 图 2 为根据本发明一个优选实施例的根据用户需求提供检索结果的方法流程图。

[0063] 除了以上参照图 1 描述的网络搜索方式,本发明同样适用于桌面端的搜索方式,也即搜索引擎位于用户设备 1。其主要对用户设备的本地信息进行搜索,包括用户设备 1 中保存的全部可用信息,包括网页浏览器历史,电子邮件档案,字处理器文档等等。为实现本地搜索的合理性能,用户设备 1 通常也需要建立和维护一个索引数据库,对应的索引信息包括但不限于以下至少一项:

[0064] 1) 文件和目录名;

[0065] 2) 元数据,例如标题,作者,注释;

[0066] 3) 所支持文档的内容。

[0067] 如图 2 所示,在步骤 s1 中,用户与用户设备 1 进行交互输入查询请求,具体过程与前面参照图 1 描述的步骤 S1 相同,为简明起见,将步骤 S1 的描述以引用方式包含在此,而不做赘述。

[0068] 随后,在步骤 s2 中,用户设备 1 基于查询请求对于本地索引库中保存的索引信息进行查询,获得与用户需求相匹配的索引信息作为搜索结果,该信息包括但不限于文档、图片、音乐等,具体过程与前面参照图 1 描述的步骤 S3 相同,为简明起见,将步骤 S3 的描述以引用方式包含在此,而不做赘述。

[0069] 然后,在步骤 s3 中,用户设备 1 基于排序信息将搜索结果按序排列显示给用户。

[0070] 同样,在优选实施例中,当用户设备 1 与网络相连时,在步骤 s4(未示出)中,用户设备 1 优选地还可自行经由网络在线搜索热门搜索项,具体过程与前面参照图 1 描述的步骤 S6 相同,为简明起见,将步骤 S6 描述以引用方式包含在此,而不做赘述。

[0071] 同样,在另一优选实施例中,在步骤 s5 中,当用户设备 1 收到来自用户的查询请求后,会查询该用户的用户相关信息,并在步骤 s2 中基于上述文义分析(或者进一步结合热点搜索项)并结合用户的相关信息来判断用户的实际需求,以获得与用户实际需求相匹配的搜索结果。步骤 s5 的具体过程与与前面参照图 1 描述的步骤 S7 相同,为简明起见,将步骤 S7 描述以引用方式包含在此,而不做赘述。

[0072] 在另一个优选实施例中,在步骤 s6(未示出)中,用户设备 1 会根据与用户的交互来记录和更新所述用户历史记录信息,具体过程与与前面参照图 1 描述的步骤 S8 相同,为简明起见,将步骤 S8 描述以引用方式包含在此,而不做赘述。

[0073] 在另外一个优选实施例中，在步骤 s3 用户设备 1 向用户展示获得的搜索结果之后，当用户对所显示的包含搜索结果的页面进行页面操作，例如包括但不限于，点击其中的某一搜索结果项所含链接、或点击下一页按钮以显示随后的搜索结果、或点击刷新页面按钮、或点击搜索按钮进行重新搜索时，用户设备 1 将根据该新交互信息作为用户历史记录更新该用户的用户相关信息。优选地，用户设备 1 还会将该新交互信息将作为新的查询请求将重复上述步骤 s2 以及 s4-s6，并将进一步的搜索结果返回用户设备 1。

[0074] 图 3 为根据本发明一个方面的根据用户需求提供检索结果的系统框图。

[0075] 如图 3 所示，用户在利用用户设备 1 进行网络浏览时，可通过诸如键盘、触摸屏、语音输入装置与用户设备 1 进行交互，输入其希望查询的查询信息 (query)，从而发起搜索。该搜索发起过程可以是用户首先通过用户设备 1 的浏览器软件访问一些搜索网站，如 Yahoo. com, Google. com, Baidu. com 后，在该搜索网站首页的搜索栏中发起搜索；也可通过嵌入浏览器的本地搜索栏发起搜索。本领域技术人员应能理解本发明不限于上述两种搜索发起方式，任何已有的或今后可能出现的搜索发起方式均适用于本发明，应包含在本发明的保护范围以内，并以引用方式而包含于此。

[0076] 接着，用户设备 1 将用户所输入的查询信息发送给网络设备 2 的获取装置 21。其中，在此，用户设备 1 包括，但不限于，计算机、智能手机、PDA 等任何具有上网浏览功能的设备。网络设备 2 包括，但不限于，网络服务器、网络主机、云计算模式下的其他用户设备 1 等可以提供搜索服务的任何设备，其中网络设备 2 包括集中各种功能的单个设备，或由多个分别完成单独功能的物理分离的设备的组合，例如执行搜索的设备与保存索引信息的设备相互分离。此外，用户设备 1 与网络设备 2 之间的通信是基于各种通信协议 (CommunicationsProtocol)，在此“通信协议”指计算机通信的传送协议，如：TCP/IP、UDP、FTP、ICMP、NetBEUI 等，同时还包括存在于计算机中的其他形式通信，例如：面向对象编程里面对对象之间的通信；操作系统内不同程序或计算机不同模块之间的消息传送协议。但本领域技术人员应理解本发明不限于所述几种通信传输协议，任何已有或今后可能出现的通信协议均可适用于本发明，从而以引用方式而包含于此。

[0077] 随后，网络设备 2 的查询装置 22 基于获取装置 21 提供的查询信息在网络索引库中进行搜索。本领域技术人员应能理解，网络索引库可位于网络设备 2 中或与之物理分离。

[0078] 其中，查询信息包含一个或多个关键词，关键词可以是字，词或者短语。在语音文档检索中，还可以是混淆类、音子、音子串等等单元，或者还包括关键词之间的关联词。如前所述，传统搜索引擎中，信息查询采用的模型，例如向量空间模型，一般通过将查询信息与索引库中的索引信息进行对比以确定两者相似度，从而确定作为搜索结果的一个或多个搜索引擎，其中，索引信息也由一些关键词来描述。

[0079] 对于所获得的多个搜索结果，传统搜索引擎一般通过其点击率或 Google 的 PageRank 或百度的超链等算法来确定各个搜索结果的排序。但传统搜索引擎的信息查询或排序过程中并未考虑用户输入查询信息的实际需求，事实上，即使用户输入的查询请求中包含相同的关键词，但由于本身内涵文义的区别，反映的用户需求也相应不同，其中查询请求的内涵文义可由关键词之间及关联词的相互顺序，例如当用户分别输入查询请求“川菜最好吃”或“最好吃川菜”，虽然均包含相同关键词，但对于前者，用户更可能希望查询口碑最好的川菜馆，；而对于后者，用户更希望查询最受欢迎的川菜做法，也即，对于前者，用户

需求“最好吃的川菜馆”优先级或可能性较高、用户需求“最好吃的川菜做法”优先级或可能性较低；对于后者，正好相反。

[0080] 因此，查询装置 2 将对查询请求的文义分析与搜索技术相结合，获得一个或多个较符合用户实际需求的搜索结果，而且这些搜索结果可以按与用户需求的匹配度进行排序。

[0081] 在一个实施方式中，索引库中索引信息应包含用户需求这一项，查询装置 22 包括第一查询装置（未示出）、需求判断装置（未示出）、筛选装置（未示出），其中第一查询装置根据前述已有的各种搜索技术基于查询信息在索引库中进行查询，获得多个搜索结果；需求判断装置对查询请求进行文义分析以确定其反映的一个或多个用户需求及其对应的优先级或可能性；筛选装置基于通过对查询请求文义分析所确定的一个或多个用户需求来对该多个搜索结果进行筛选和排序，以获得最终确定的按用户需求进行排序的搜索结果。

[0082] 在另一实施方式中，查询装置 22 包括需求判断装置（未示出）、第二查询装置（未示出）、需求判断装置对查询请求进行文义分析以确定其反映的一个或多个用户需求及其对应的优先级或可能性；第二查询装置在基于查询请求对索引库进行查询时，将所确定的用户需求作为信息查询的一个维度，以确定相匹配的一个或多个索引信息作为搜索结果。

[0083] 本领域技术人员应理解本发明不限于上述两种实施方式，本领域技术人员根据本发明的构思应能想到的其他实施方式也应包含在本发明的保护范围以内。

[0084] 随后，网络设备 2 的提供装置 23 将查询装置 22 搜索得到的搜索结果及其排序信息返回给用户设备 1。最后，用户设备 1 基于排序信息将搜索结果按序排列显示给用户。

[0085] 在一个优选实施例中，查询装置 22 中的需求判断装置还可根据大样本统计信息来确定关键词与用户需求的关系，例如，对于关键词“Elton John”，网络设备 2 根据大量用户对搜索结构的选择可以确定，通过对大量用户的选择的统计分析可确定其对应的用户需求“Elton John 的音乐”优先级或可能性较高、“Elton John 的新闻”优先级或可能性较低。

[0086] 在另一个优选实施例中，网络设备 2 还包括在线挖掘装置（未示出），其根据所述查询请求来查询预先记录的与所述查询请求相关的大量用户选择的热门搜索项，并根据所述大量用户选择的热门搜索项来确定所述用户的用户需求。例如 2010 年 4 月中国青海省玉树县发生 7 级以上地震，而网络设备 2 中关键词“玉树”与该热门搜索项“玉树地震”相关联，从而当接收到包含关键词“玉树”的查询请求时，可以确定用户希望了解“玉树地震”或“玉树地震赈灾”新闻的需求的优先级或可能性较高，相对的了解“玉树旅游”或“玉树历史”的优先级或可能性较低。

[0087] 优选地，在线挖掘装置可以自行经由网络在线搜索热门搜索项，包括但不限于以下方式：

[0088] - 首先，在线挖掘装置通过大样本统计信息，也即通过统计用户对网页的点击率，并对点击率排序较高的网页进行信息抓取，确定相应的关键词；

[0089] - 接着，在线挖掘装置根据以下至少一项确定所述热门关键词的热门搜索项：

[0090] - 所述热门关键词的时效性；

[0091] - 所述热门关键词的主题类别；

[0092] - 所述热门关键词的主题需求类别。

[0093] 其中，主题类别是指关键词的所述类别，例如关键词“苹果”的主题类别上可以是

水果、计算机、电影。而主题需求类别是指相对于关键词的用户需求的类别，例如当用户输入关键词“苹果”，如果其主题类型是属于计算机类别的，则主题需求类别可能是寻求苹果公司信息，或是寻求苹果电脑的报价。

[0094] 在另一个优选实施例中，考虑由于用户的个体差异，使得当不同用户输入包含相同关键词时，其实际需求也是因人而异，因此在步骤例如，对（成年）男性、（成年）女性、儿童、老年人等，当输入查询请求中包含“\*\* 百货公司”时，其实际需求区别较大，其中，男性用户可能对运动产品、男性服饰、金融商业书籍和重金属摇滚乐 CD 等更感兴趣、女性用户可能对护肤 / 化妆品、女性服饰、浪漫言情书籍和轻音乐 CD 更感兴趣，儿童用户对儿童服饰、儿童读物和卡通 DVD 等更感兴趣，而老年用户对老年服饰、保养品、文史类书籍和古典音乐 CD 更感兴趣。当然，用户的个人背景，例如教育背景、宗教背景等也会成为其实际需求的影响要素，例如理工科毕业的用户可能对于技术类书籍更感兴趣，而经济类专业的用户对于金融或商战类书籍更感兴趣。另外，例如，对于旅游网站时，当用户处于不同地域时其所感兴趣的信息通常也是不同的，例如，对于处于北京地域的用户，当其输入的查询请求中包括关键词“旅游”时，其实际需求可能是北京周边区域（例如，北戴河、河北）等自驾游和旅店住宿信息，当其查询请求包括关键词“机票酒店”时，可能对由北京出发的特价机票更感兴趣。此外，用户以往的搜索历史记录也可以成为确定其实际需求的指示信息，例如，网络设备 2 通过查询一名用户的历史记录信息，发现其以往在访问红酒购物网站时更多地关注梅乐（MERLOT）红酒，则可当其输入的查询请求包括“\*\* 红酒网站”时，其实际需求可能是与梅乐红酒有关的信息。

[0095] 基于以上理解，网络设备 2 的查询装置 22 接收到获取装置 21 提供的用户设备 1 的查询请求后，其中的需求判断装置会查询该用户的用户相关信息，基于上述文义分析（或者进一步结合热点搜索项）并结合用户的相关信息来判断用户的实际需求，如前所述，其中用户相关信息包括以下各项中至少一项：

[0096] - 用户属性，其包含，但不限于，用户预先输入的个人背景信息，如性别、年龄、宗教、国籍、教育程度等；

[0097] - 用户偏好，其包含，但不限于，用户预先设置的一些偏好选择，如喜爱的运动类型、购物方式、红酒品种、音乐类型等；

[0098] - 用户查询历史记录，其由系统来记录和统计用户先前的搜索和点击浏览历史，从而可判断或预测用户的需求。其中，当用户为登录用户时，需求判断装置可通过该用户的 ID 在网络端保存有该用户相关信息的用户特征库中进行查询；当用户为非登录用户时，可通过检测该用户使用本网络设备 2 的搜索服务过程中的动作，或通过向用户设备 1 询问并获取该用户的 Cookie 信息来获得，优选地，可赋予该用户一个临时 ID 并为其建立一个用户特征库，将此次搜索过程中用户的历史记录保存与该临时用户特征库中；

[0099] - 用户设备 1 的 IP 地址，其可以反映该用户当前所处地域，可由网络设备 2 实时检测用户设备 1 的 IP 地址来获取。

[0100] 优选地，网络设备 2 还包括更新装置（未示出），其根据与用户设备 1 及用户的交互来记录和更新所述用户历史记录信息，具体地，包括但不限于以下方式：1) 更新装置可在用户一次搜索过程结束后记录用户在此搜索过程中的所有动作，例如点击和浏览；2) 更新装置在用户一次搜索过程中记录用户每次动作并实时对网络端该用户的特征库进行更

新,从而可在用户在此次搜索过程中点击其他搜索结果时提供最新的用户历史记录信息。

[0101] 在另外一个优选实施例中,当用户设备 1 向用户展示获得的搜索结果之后,当用户对所显示的包含搜索结果的页面进行页面操作,例如包括但不限于,点击其中的某一搜索结果项所含链接、或点击下一页按钮以显示随后的搜索结果、或点击刷新页面按钮、或点击搜索按钮进行重新搜索时,该新交互信息经由用户设备 1 重新发送至网络设备 2,网络设备 2 的更新装置将根据该新交互信息作为用户历史记录更新该用户的用户相关信息。优选地,查询装置还会将该新交互信息将作为新的查询请求将重复上述查询过程,并将进一步的搜索结果返回用户设备 1。

[0102] 图 4 为根据本发明一个优选实施例的根据用户需求提供检索结果的系统框图。

[0103] 除了以上参照图 3 描述的网络搜索方式,本发明同样适用于桌面端的搜索方式,也即搜索引擎位于用户设备 1 用,其主要对用户设备的本地信息进行搜索,包括用户设备 1 中保存的全部可用信息,包括网页浏览器历史,电子邮件档案,字处理器文档等等。为实现本地搜索的合理性能,用户设备 1 通常也需要建立和维护一个索引数据库,对应的索引信息包括但不限于以下至少一项:

[0104] 1) 文件和目录名;

[0105] 2) 元数据,例如标题,作者,注释;

[0106] 3) 所支持文档的内容。

[0107] 如图 4 所示,用户设备 1 包括获取装置 11、查询装置 12、提供装置 13,其所执行的功能分别于前面参照图 3 中描述的网络设备 2 的获取装置 21、查询装置 22、提供装置 23 相同,为简明起见,将前面针对获取装置 21、查询装置 22、提供装置 23 的描述以引用方式包含在此,而不做赘述。

[0108] 同样,在优选实施例中,当用户设备 1 与网络相连时,用户设备 1 还包括在线挖掘装置(未示出),其可以自行经由网络在线搜索热门搜索项,具体过程与前面参照图 3 描述的网络设备 2 的在线挖掘装置相同,为简明起见,将该描述以引用方式包含在此,而不做赘述。

[0109] 同样,在另一优选实施例中,当用户设备 1 收到来自用户的查询请求后,会查询该用户的用户相关信息,用户设备 1 中查询装置 12 的需求判断装置(未示出)基于上述文义分析(或者进一步结合热点搜索项)并结合用户的相关信息来判断用户的实际需求,以获得与用户实际需求相匹配的搜索结果,具体过程与前面参照图 3 描述的网络设备 2 中查询装置 22 的需求判断装置相同,为简明起见,将该描述以引用方式包含在此,而不做赘述。

[0110] 在另一个优选实施例中,用户设备 1 还包括更新装置(未示出)会根据与用户的交互来记录和更新所述用户历史记录信息,具体过程与与前面参照图 3 描述的网络装置 2 中的更新装置相同,为简明起见,将该步骤描述以引用方式包含在此,而不做赘述。

[0111] 在另外一个优选实施例中,在用户设备 1 的提供装置 13 向用户展示获得的搜索结果之后,当用户对所显示的包含搜索结果的页面进行页面操作,例如包括但不限于,点击其中的某一搜索结果项所含链接、或点击下一页按钮以显示随后的搜索结果、或点击刷新页面按钮、或点击搜索按钮进行重新搜索时,更新装置将根据该新交互信息作为用户历史记录更新该用户的用户相关信息。优选地,查询装置 12 还会将所述页面操作作为新的查询请求并根据所述新的查询请求进行前述匹配搜索,以获得与所述查询请求和用户需求相匹配

的一个或多个搜索选项。

[0112] 以上对本发明的具体实施例进行了描述。需要理解的是，本发明并不局限于上述特定实施方式，本领域技术人员可以在所附权利要求的范围内做出各种变形或修改。

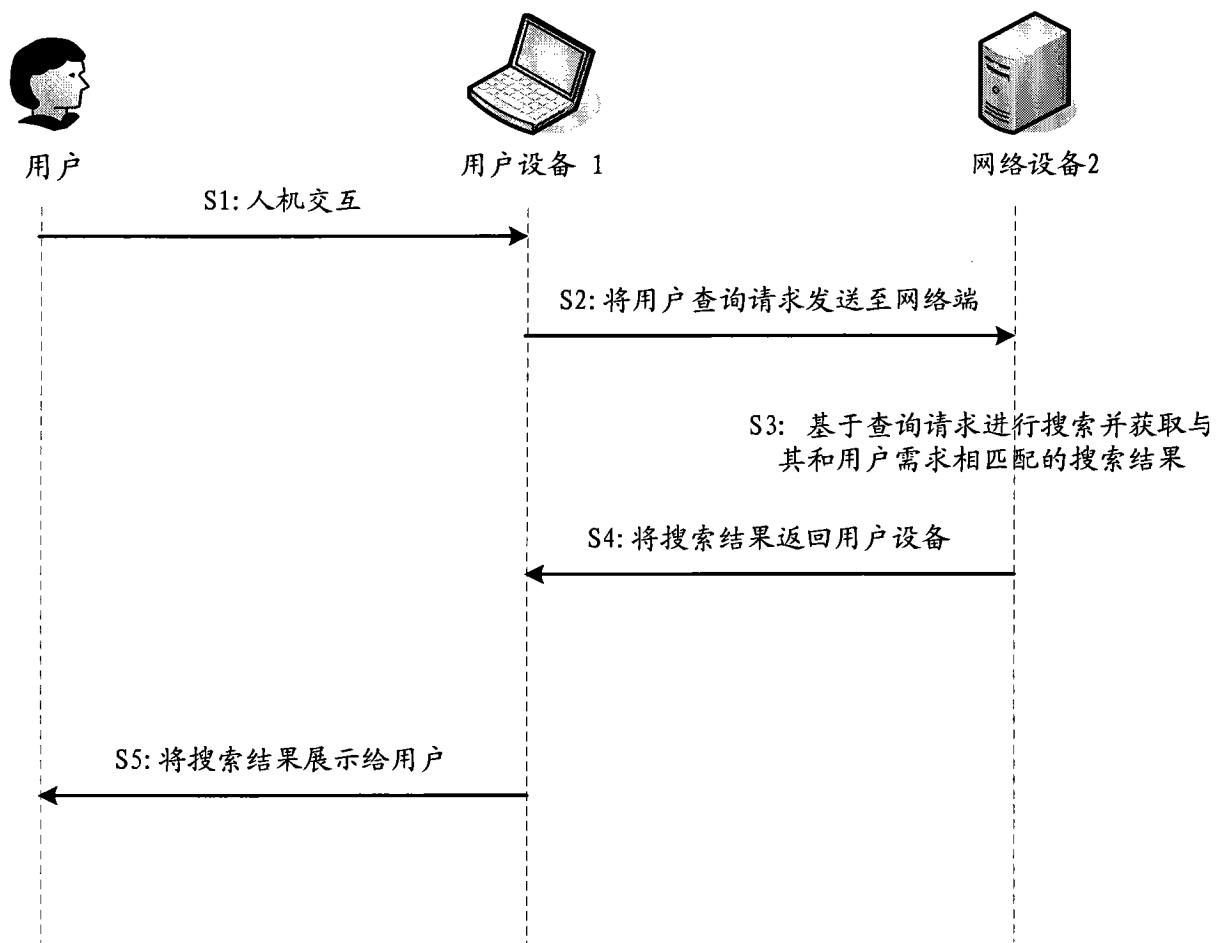


图 1

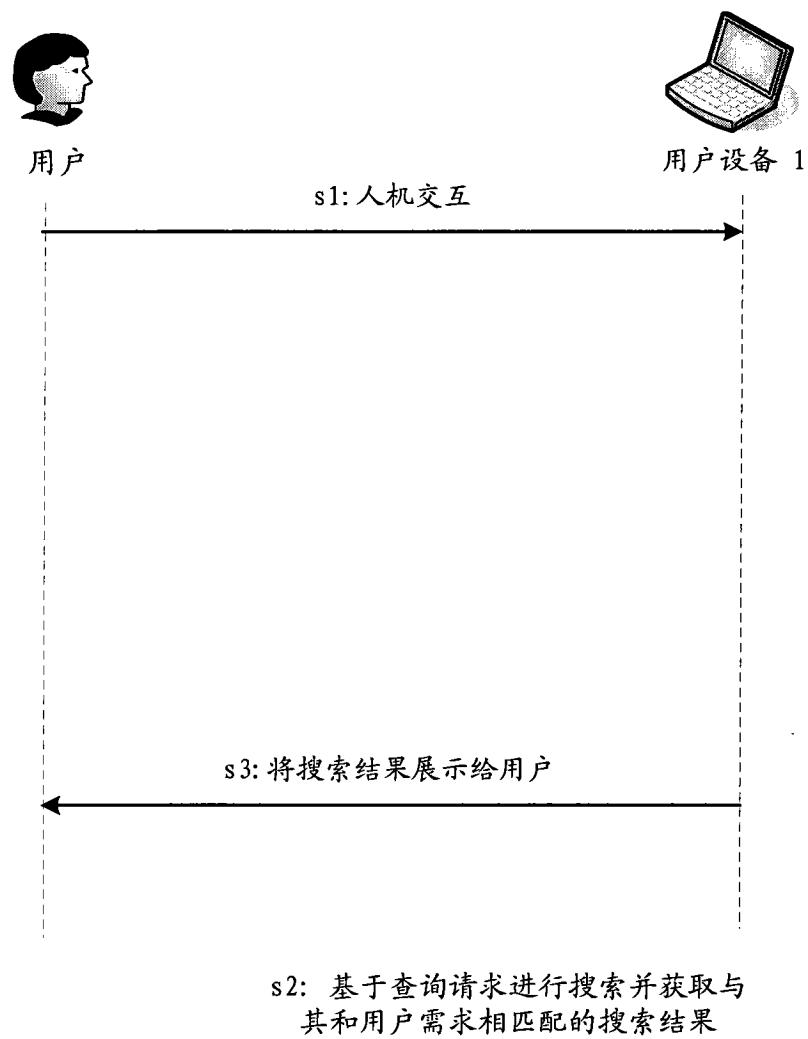


图 2

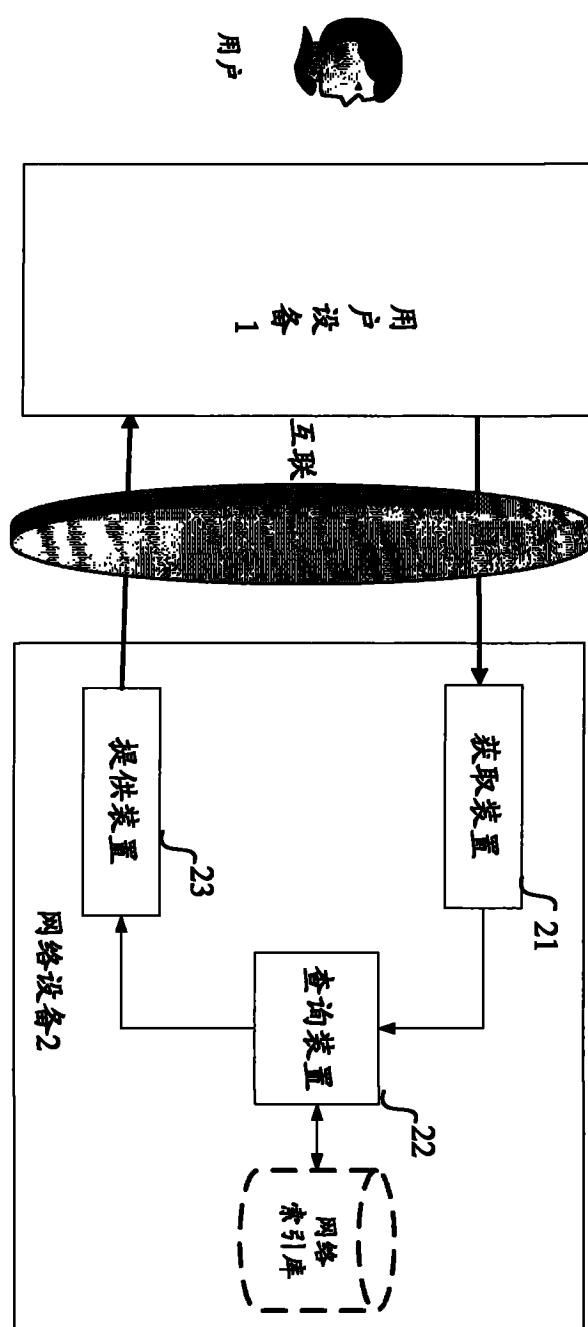


图 3

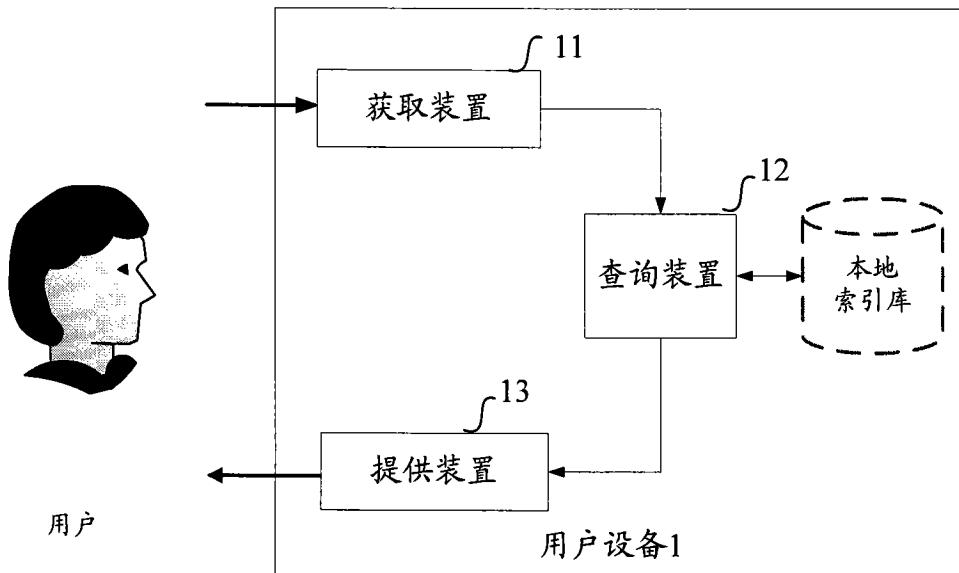


图 4