(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101523338 B
(45)授权公告日 2016.10.19

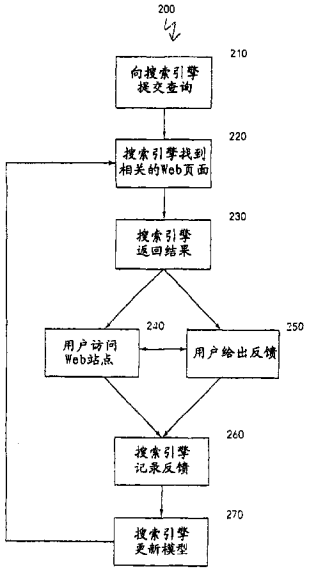
(21)申请号 200680017270.7
(22)申请日 2006.03.17
(65)同一申请的已公布的文献号
申请公布号 CN 101523338 A
(43)申请公布日 2009.09.02
(30)优先权数据
60/663,361 2005.03.18 US
(85)PCT国际申请进入国家阶段日
2007.11.19
(86)PCT国际申请的申请数据
PCT/US2006/009800 2006.03.17
(87)PCT国际申请的公布数据
WO2006/102122 EN 2006.09.28
(73)专利权人 搜索引擎科技有限责任公司
地址 美国加利福尼亚州
(72)发明人 Y·卢 G·P·里奥斯 M·坦纳
(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 王岳 王小衡
(51)Int.Cl.
G06F 7/00(2006.01)
G06F 17/30(2006.01)
(56)对比文件
US 2004/0024755 A1,2004.02.05,
US 6327590 B1,2001.12.04,
Vladimir Eske.User Profile Management in a Web Search Engine.《http://domino.mpi-inf.mpg.de/intranet/ag5/ag5publ.nsf/951b6516df8639d3c1256464004a33d0/983fc84653920a97c1256ede00394b65/$FILE/Eske.pdf》.2004,9-10、19-29、33、42-55、79、83.
审查员 董洪梅
权利要求书1页 说明书12页 附图5页

(54)发明名称

应用来自用户的反馈来改进搜索结果的搜索引擎

(57)摘要

本发明针对用于对由搜索引擎返回的结果进行排序的方法和系统。根据本发明的方法包括:确定具有变量和参数的公式,其中所述公式用于计算文献与搜索查询的相关性分数,并且基于所述相关性分数对所述文献进行排序。优选地,确定所述公式包括基于用户输入调整所述参数。优选地,使用机器学习技术来确定所述参数,比如包括某种形式的统计分类的机器学习技术。



CN 101523338 B

1.一种响应于搜索查询而对结果列表中返回的文献进行排序的计算机实施的方法，包括：

接收并存储来自用户的关于文献与该用户的相关性的反馈，所述用户反馈用于对所述结果列表中的文献进行排序；

确定具有变量和参数的相关性公式，所述相关性公式用于计算文献与所述搜索查询的相关性分数，其中所述变量包括第一特征和第二特征，所述第一特征包括：标签、从搜索结果列表访问文献的次数、用户表示、用户输入的时间、阻绝、用户标识符或用户对文献的评分，且所述参数包括对应于所述第一特征和所述第二特征的第一系数和第二系数；以及，其中确定所述相关性公式包括利用用户反馈来通过更新所述参数而更新所述相关性公式，其中所述用户反馈是以下各项的其中之一：给文献加标签、响应于搜索查询而对文献的相关性进行评分、以及阻绝文献；

基于所述相关性分数对所述文献进行排序，其中所述相关性分数取决于与所述文献和所述搜索查询相关联的用户反馈；以及

将文献的排序的结果列表展现给所述用户。

2.权利要求1所述的方法，还包括：响应于所述搜索查询而返回结果列表，其中该结果列表包含到基于所述相关性分数而被排序的所述文献的链接。

3.权利要求1所述的方法，其中，所述第二特征是从包括以下各项的组中选择的：文献的结构、文献的长度、文献在搜索结果列表中的位置、从搜索结果列表访问文献的次数、项目分数、链接结构、文献内的项目、到文献的链接、锚文本概要以及文献在搜索结果列表中的位置。

4.权利要求1所述的方法，其中，所述第二特征是从包括以下各项的组中选择的：用户阻绝文献、文献的用户标识符、用户保存到文献的链接、用户为文献加书签、用户为文献加标签、用户保存文献以及文献的摘要内的项目的出现频率。

5.权利要求1所述的方法，其中，所述相关性公式对应于用户模型和组模型，其中该用户模型用于确定一个用户的文献与搜索查询的相关性分数，以及该组模型用于确定一组用户的文献与搜索查询的相关性分数。

6.权利要求5所述的方法，还包括：把所述用户模型与所述组模型进行比较，以便确定对所述文献的偏向。

7.权利要求1所述的方法，其中，利用机器学习技术来确定所述参数。

8.权利要求7所述的方法，其中，所述机器学习技术包括统计分类。

9.权利要求8所述的方法，其中，所述统计分类是以下各项当中的任何一个：逻辑斯谛回归分析、支撑矢量机、神经网络、提升树、随机森林、朴素贝叶斯、以及图形模型。

10.权利要求7所述的方法，其中，所述机器学习技术使用共轭梯度下降。

11.权利要求1所述的方法，其中，所述相关性分数对应于在搜索引擎站点上注册的用户，并且该相关性分数被用来为没有在该搜索引擎站点上注册的用户确定文献与搜索查询的相关性分数。

12.权利要求1所述的方法，其中，利用对于多个结果的每个结果的用户反馈来更新所述参数。

应用来自用户的反馈来改进搜索结果的搜索引擎

[0001] 相关申请

[0002] 本申请根据35 U.S.C.§119(e)要求2005年3月18日提交的、顺序号为60/663,361、以及标题为“Search Engine that Applies Feedback from Users to Improve Search Results”的同时待审的美国临时申请的优先权，该申请被结合于此以作参考。

发明领域

[0003] 本发明涉及一种用于搜索因特网的应用，更具体而言，涉及这样一种用于搜索因特网的应用，其应用来自使用该搜索应用的人的反馈来提高搜索结果的质量。

[0004] 发明背景

[0005] 因特网搜索引擎被设计成从在因特网中包含的大量信息中定位所期望的信息。用户通过输入包含搜索项目的查询来描述他们所寻找的信息。所述搜索引擎利用多种相关性计算把所述搜索项目与Web页面的索引相匹配，其目的是识别出最有可能与用户所寻求的信息相关的那些Web页面。随后，所述搜索引擎返回到这些Web页面的超链接的有序列表，其中到被认为是最相关的那些页面的链接更靠近该列表的顶部。

[0006] 搜索引擎的目的是对于给定查询提供最相关的Web页面。早期的搜索引擎利用在每个页面内包含的信息来确定Web页面的相关性，所述信息比如是所述搜索项目在文献内的存在、密度和近似。更高级的搜索引擎在确定相关性时考虑与Web页面之间的链接相关的信息。

[0007] 确定哪些Web页面最相关的过程是非常困难的，因为因特网上的Web页面的数量非常巨大而且不断增长，并且常常有大量名义上满足用户的查询的Web页面。对于相同的或相似的项目可能会提到许多不同的概念。大多数用户并不精于创建及输入合适的查询的过程，因此在其正在寻求的信息类型方面存在模糊性。

[0008] 可以对由搜索引擎返回的结果进行操纵。Web站点运营商可以在其Web页面中添加来自其他Web站点的内容或元数据或超链接，其目的 是让搜索引擎返回在所述有序结果列表中排序高的到其Web页面的链接。这样做的结果是不包含用户所寻求的信息的一些Web页面在该结果列表中的排序高，从而降低了搜索结果的总体感知精度。这种实践常常被称作搜索引擎优化(或者“SEO”)。搜索引擎必须花费不断增加的努力来仅仅在SEO方面保持一致的相关性水平。期望的将是搜索引擎能够直接从用户收集反馈，以便确定哪些Web页面满足或者不满足其查询，从而为由相同或不同用户进行的后续查询提供更相关的结果。

[0009] 搜索引擎的运营商随着时间对用来确定相关性的方法进行调节，并且对应用于每种所述方法的权重进行调节，以便保持或者提高其搜索引擎的精度。这种过程通常涉及到进行实验，比如让测试用户对于由所述搜索引擎为不同查询提供的结果进行打分。可以对所述分数进行编辑和分析，以便判定将使用什么方法或者将应用什么权重。这一过程耗时、不精确、没有代表性并且不灵活。期望的将是使用反馈机制，所述反馈机制直接从真实用户取得输入并且调节所述搜索相关性方法和权重，以便提高搜索引擎的精度。

[0010] 另外，用户在其输入查询时具有不同的意图，这是因为他们对于搜索项目可能具

有不同的理解，具有不同的品位和兴趣，或者处在不同的搜索“模式”下。例如，如果三个不同的用户输入查询“iPod”，第一个用户可能是想要购买iPod，第二个用户可能是想要搜索关于iPod的新闻，而第三个用户可能正在搜索关于iPod的信息或评论。用户在进行搜索时可以给出关于其兴趣和目的的某种指示。因此，期望的将是开发出这样一种搜索引擎，其能够在确定哪些结果与用户查询相关时考虑不同的搜索模式以及用户兴趣之间的差异。

发明概要

[0011] 本发明针对用于预测文献与搜索查询的相关性的方法和系统，由此向执行所述搜索的用户返回更相关的结果。在一个优选实施例中，根据本发明的方法使用公式来预测多个文献与搜索查询的相关性，基于每个文献的相关性对所述文献进行排序，并且响应于搜索查询而向用户返回该有序列表。优选地，使用用户输入来调整所述公式的参数，以便提高所返回的文献与所述搜索查询相关的可能性。

[0012] 在本发明的第一方面中，一种响应于搜索查询而对文献进行排序的 方法包括：(a)确定具有变量和参数的公式，其中所述公式用于计算文献和搜索查询的相关性分数；并且(b)基于所述相关性分数对所述文献进行排序。优选地，该方法还包括响应于所述搜索查询而返回结果列表。该结果列表包含到所述文献的链接，所述文献基于所述相关性分数被排序在该结果列表内。

[0013] 在一个实施例中，确定所述公式包括基于用户输入调整所述参数。所述公式是从以下的任何一个或多个特征导出的，比如标签、文献内的项目、项目在文献内的位置、文献的结构、到文献的链接、文献在搜索结果列表中的位置、从搜索结果列表访问文献的次数、项目分数、段信息、链接结构、锚文本以及概要。可选择地或另外，所述特征包括用户表示、用户输入的时间、阻绝、用户标识符或者用户对文献的评分。

[0014] 在一个实施例中，所述公式对应于用户模型和组模型。该用户模型用于确定一个用户的文献与搜索查询的相关性分数。该组模型用于确定一组用户的文献与搜索查询的相关性分数。所述方法还包括把该用户模型与该组模型进行比较，以便确定对所述文献的偏向(bias)。

[0015] 优选地，利用机器学习技术来确定所述参数。在一个实施例中，所述机器学习技术包括某种形式的统计分类，比如逻辑斯谛回归分析、支撑矢量机、神经网络、提升树(boosted tree)、随机森林、朴素贝叶斯、或者图形模型。在另一个实施例中，所述机器学习技术使用共轭梯度下降。

[0016] 在另一个实施例中，从用户输入生成所述公式的一个或多个变量。所述用户输入是给文献加标签、对文献的相关性进行评分、阻绝文献、或者这些的任何组合。

[0017] 在一个实施例中，所述相关性分数对应于在搜索引擎站点上注册的用户，并且该相关性分数被用来为没有在该搜索引擎站点上注册的用户确定文献与搜索查询的相关性分数。

[0018] 在本发明的第二方面中，一种组织搜索结果列表的方法包括：(a)更新包括变量和参数的相关性公式，其中该相关性公式用于为响应于搜索查询而返回的多个结果当中的每一个确定相关性分数；并且(b)响应于所述搜索查询而返回包含所述多个结果的搜索结果列表，其中利用对应于所述多个结果当中的每一个的相关性分数对所述多个结果进行排

序。在一个实施例中，更新所述相关性公式包括更新所述参数。优选 地，利用用户输入来更新所述参数，比如通过给文献加标签、响应于搜索查询而对文献的相关性进行评分、以及阻绝文献。

[0019] 优选地，利用机器学习技术来导出所述相关性公式，比如包括某种形式的统计分类的技术。优选地，所述统计分类是逻辑斯谛回归分析、支撑矢量机、神经网络、提升树、随机森林、朴素贝叶斯、或者图形模型。在另一个实施例中，所述机器学习技术使用共轭梯度下降。

[0020] 在本发明的第三方面中，一种对由第一搜索引擎返回的结果进行排序的方法包括：响应于搜索查询而接收包含由第一搜索引擎排序的文献的第一结果列表；响应于该搜索查询而接收包含由第二搜索引擎排序的文献的第二结果列表；确定具有变量和参数的公式，该公式用于响应于该搜索查询而为文献确定相关性分数；并且响应于该搜索查询而返回第三结果列表，该第三结果列表包含利用对应于每个文献的相关性分数进行了排序的第一结果列表和第二结果列表中的文献。因此，本发明的实施例能够利用由一个或多个搜索引擎返回的结果来运行。优选地，利用用户输入来确定所述公式。

[0021] 在一个实施例中，利用所述文献在第一结果列表中的顺序以及所述文献在第二结果列表中的顺序来确定所述公式。利用机器学习技术来确定所述参数，比如包括某种形式的统计分类的技术。在可选实施例中，所述统计分类是逻辑斯谛回归分析、支撑矢量机、神经网络、提升树、随机森林、朴素贝叶斯、或者图形模型。

[0022] 在本发明的第四方面中，一种用于响应于搜索查询而返回有序结果列表的系统包括耦合到相关性模型的第一数据库。该第一数据库用于存储用户输入，以用于响应于第一搜索查询而对所述多个结果当中的结果进行排序。该相关性模型用于使用所述用户输入为对应于第二搜索查询的多个结果当中的每个结果计算相关性分数。

[0023] 在一个实施例中，所述系统还包括耦合到第一数据库的搜索引擎。该搜索引擎用于接收搜索查询，基于对应于每个结果的相关性分数对所述多个结果进行排序，以及返回包含所述经过排序的多个结果的搜索结果列表。优选地，所述相关性模型被配置成利用用户输入来更新。在另一个实施例中，该相关性模型被配置成使用机器学习技术，比如包括某种形式的统计分类的技术。在可选实施例中，所述统计分类是逻辑斯谛回归分析、支撑矢量机、神经网络、提升树、随机森林、朴素贝叶斯、 或者图形模型。在另一个实施例中，所述机器学习技术使用共轭梯度下降。

[0024] 在一个实施例中，所述相关性模型用于确定一组特定于用户的参数，以便为特定用户确定搜索查询与文献的相关性分数。在另一个实施例中，该相关性模型还用于确定组模型的参数，以便为一组用户确定搜索查询与文献的相关性分数。

[0025] 在一个实施例中，所述系统还包括第二数据库，其用于存储用于更新所述相关性模型的一个或多个特征。所述特征包括以下各项当中的任何一个或多个：标签、文献内的项目、项目在文献内的位置、文献的结构、到文献的链接、文献在搜索结果列表中的位置、从搜索结果列表访问文献的次数、项目分数、段信息、链接结构、锚文本、概要、用户表示、用户输入的时间、阻绝、用户标识符以及用户对文献的评分。

[0026] 在本发明的第五方面中，一种用于在搜索结果列表中组织多个结果的系统包括耦合到搜索引擎的相关性模型以及耦合到该搜索引擎和该相关性模型的数据库。该相关性模

型用于使用用户输入来确定文献与搜索查询的相关性分数。该搜索引擎用于接收搜索查询并且返回搜索结果列表，所述搜索结果列表包含根据每个文献与所述搜索查询的相关性分数进行了排序的结果。该数据库用于存储一组特征，该组特征被该相关性模型使用来确定文献与搜索查询的相关性分数。

[0027] 在一个实施例中，所述数据库包括耦合到所述搜索引擎的用户数据库以及耦合到所述相关性模型的用户输入数据库。该用户数据库用于存储关于所述搜索结果列表的用户输入，并且所述用户输入数据库用于存储所述一组特征。该组特征包括以下各项当中的任何一个或多个：标签、文献内的项目、项目在文献内的位置、文献的结构、到文献的链接、文献在搜索结果列表中的位置、从搜索结果列表访问文献的次数、项目分数、段信息、链接结构、锚文本、概要、用户表示、用户输入的时间、阻绝、用户标识符以及用户对文献的评分。

[0028] 优选地，所述系统还包括耦合到所述数据库的Web服务器以及耦合到所述相关性模型和所述搜索引擎的文献索引。

[0029] 优选地，所述相关性模型被配置成使用机器学习技术，比如包括某种形式的统计分类的技术。在可选实施例中，所述统计分类是逻辑斯谛回归分析、支撑矢量机、神经网络、提升树、随机森林、朴素贝叶斯、或者图形模型。在另一个实施例中，所述机器学习技术使用共轭梯度下降。

[0030] 附图简述

[0031] 图1是显示搜索结果列表的示例性图形用户界面的示意图，其中结合了用户反馈以及用于使用户提供反馈的机制。

[0032] 图2是说明能够对于排序文献的过程应用用户反馈的示例性因特网搜索应用的操作的流程图。

[0033] 图3是说明根据本发明的示例性因特网搜索应用流程图的组件的示意图。

[0034] 图4是说明根据本发明利用用户反馈数据计算在结果列表中包含的结果的步骤的流程图。

[0035] 图5是说明根据本发明的示例性因特网搜索应用的组件的硬件图。

[0036] 优选实施例的详细描述

[0037] 与传统的搜索引擎不同，本发明的实施例利用用户反馈来为正在搜索因特网的用户提供更相关的信息。例如，根据本发明，执行搜索的第一用户能够对该搜索的结果进行评分。第一用户可以基于所述结果与他的搜索的相关性、在结果列表中返回的特定Web页面中所包含的信息的丰富程度或者任何其他标准来为所述结果进行评分。于是，执行类似的或相关的搜索的第二用户能够观看基于第一用户的评分或者受到第一用户的评分的影响、从而显示出更有可能与第二用户最相关的搜索结果的结果列表。该第二用户也能够对搜索结果进行评分。因此，一个用户团体能够提供反馈，所述反馈帮助用户接收到与他们所寻求的信息尽可能相关的搜索结果。在顺序号为11/364,617的美国专利申请中描述了使用用户反馈的系统和方法，该申请的标题为“Methods of and Systems for Searching by Incorporating User-Entered Information”，并且于2006年2月27日提交，这里合并该申请以作参考。

[0038] 根据本发明的其他实施例，存在几种用于对搜索结果进行评分的方法，其中包括但不限于：(1)用于提供关于所述结果列表中的链接的相关性的反馈的机制；(2)用于保存

可以被显示在个人搜索页面上的链接或者用于针对相关链接进行投票的机制;以及(3)用于“阻绝” 到与所述搜索结果无关或者本质上是冒犯性的Web页面的链接的机制。其他实施例包括显示并且链接到相关的搜索项目和赞助商链接。

[0039] 根据本发明的其他实施例,搜索结果页面还可以包括用于显示的所选项目,其中包括但不限于:(1)来自与所述搜索相关的Web页面的文本;(2)对于与所述查询项目相关的不同概念的描述;(3)所建议的查询项目;(4)到其他概念的“See also(另见)”链接;以及(5)赞助商链接。

[0040] 在下面的整个描述中,术语“搜索引擎”被用来指代一种取得作为输入的查询并且返回到电子文献或Web页面的超链接的结果列表的设备(或者在通用计算机上运行的程序)。该搜索引擎包括文献本体的索引、确定每个文献的相关性的代码和算法、以及向用户提供结果列表的图形用户界面。

[0041] 在下面的整个描述中,术语“查询”指代被提交给搜索引擎的一组项目,而不管其是被键入的、说出的、通过已经嵌入一组搜索项目的“链接”提交的还是通过任何其他接口提交的。一个查询可以包括单个单词、多个单词、或者短语。所述查询可以被措辞为一个问题(例如“自然语言”查询)、一组松散项目、或者一个结构化布尔表达式。实际上,一个查询可以包括由搜索引擎使用来搜索包含搜索字符或者与搜索字符相关的电子文献或Web页面的符号或任何其他字符。

[0042] 在下面的整个描述中,术语“Web站点”被用来指代被链接在一起的并且可以在万维网上获得的Web页面的集合。术语“Web页面”指代可以通过万维网从任何数目的主机访问并且包括但不限于文本、视频、图像、音乐和图形的公开物。

[0043] 在下面的整个描述中,术语“结果列表”指代一个超链接的列表,所述超链接索引可以利用超文本传输协议(HTTP)或用于访问Web页面或其他电子文献的任何其他协议来访问的文献或Web页面、以及对应于每个链接的其他相关联的信息,其中包括但不限于文献的标题、文献的概要、到文献的高速缓存的拷贝的链接、最后一次索引或者最后一次修改文献的日期、与文献相关联或者位于文献内的图像、以及从文献提取的信息。

[0044] 这里所使用的术语“文献”的定义很广泛,并且除了其普通的含义之外,还包括计算机文件和Web页面,而不管这些页面是被实际存储的 还是响应于显示请求而被动态生成的。术语“文献”不限于包含文本的计算机文件,而且还包括含有图形、音频、视频和其他多媒体数据的计算机文件。

[0045] 这里使用的术语“机器学习”指代软件系统可以通过其来使它的行为适于通过观察某些事件或者分析特定信息来产生最佳结果的过程和算法。

[0046] 这里使用的术语“统计模型”指代在给定一组输入的情况下计算分数的数学公式(例如数值的或解析的)。所述公式的参数可以利用机器学习过程来获得。用在本发明中的统计模型可以基于用户反馈、来自搜索事件的其他信息、或者二者的组合,并且可以利用多种数学技术当中的任何一种来生成。

[0047] 如下面将更详细地描述的那样,搜索引擎取得由用户输入的查询,并且利用多种相关性计算把所述搜索项目与Web页面的索引相匹配,其目的是识别出最有可能与用户所寻求的信息相关的Web页面。搜索引擎随后返回到这些Web页面的有序超链接列表,其中被认为最相关的链接更靠近所述列表的顶部。根据本发明,搜索引擎基于用户输入返回结果

列表，并且用户能够对所述结果进行评分，以便例如影响在该结果列表中列出的文献或链接的顺序。根据本发明，能够响应于用户对他们认为相关的站点加标签来对搜索结果进行排序，即使当他们访问在搜索的上下文之外的站点或者利用不同于被用来生成所述搜索结果的项目对所述站点加标签时也是如此。

[0048] 如下面将更详细地描述的那样，搜索引擎取得由用户输入的查询，并且利用多种相关性计算把所述搜索项目与Web页面的索引相匹配，其目的是识别出最有可能与用户所寻求的信息相关的Web页面。搜索引擎随后返回到这些Web页面的有序超链接列表，其中被认为最相关的链接更靠近所述列表的顶部。在典型的搜索中，所述搜索引擎主要返回结果列表，并且用户不能向所述系统输入信息。

[0049] 根据本发明，当向用户提供包含结果列表的页面时，他可以选择关于该页面上的结果提供反馈，所述反馈将被提交给一个模型，该模型分析所述反馈并且调节所述相关性方法和权重，以便提高被提供给随后通过输入相同的或不同的查询来访问该搜索引擎的用户的结果的相关性。

[0050] 图1是图形用户界面(GUI)的截屏图，其显示根据本发明响应于 查询而返回的结果页面100。该GUI允许用户对单独的搜索结果进行评分、阻绝单独的搜索结果或者保存单独的搜索结果。另外，用户可以添加、编辑和观看关于查询项目的一个或多个概念的描述，以及添加、编辑或观看关于如何搜索与所述概念相关的信息的建议。

[0051] 结果页面100包括用于插入查询项目的框110以及包含由搜索引擎返回的结果列表的区域160。该区域160还包含用于输入用户反馈的机制170以及用于保存与由所述搜索引擎返回的每个结果相关联的链接的机制190。该结果页面100还包括用于显示对于与所述查询项目相关的概念的描述的区域120、包含对于与该查询项目相关的不同概念的描述的区域130、包含到与其他查询项目相关的概念的“See also(另见)”链接的区域140、包含将导致执行相关的查询项目的链接列表的区域150、以及赞助商链接的区域180。如下面将更详细地描述的那样，在一个优选实施例中，可以基于对链接进行评分或阻绝170或者保存链接190以用于随后显示来对于其他用户的后续查询修改区域160中的结果。

[0052] 如图1的例子中所示，当用户在框110中输入查询项目“U2”并且请求搜索时，结果页面100被返回给该用户。区域120显示对于与查询项目“U2”相关的一个概念的描述，在这里是对乐队“U2”的描述，如用户所输入的那样。区域130显示对应于查询“U2”的不同概念的描述，在这里是U2侦察机，如用户所输入的那样。区域150显示对应于相关的搜索的查询项目，用户在执行搜索引擎时可能还对所述相关的搜索感兴趣，比如“U2音乐会门票”或“U2 iPod”，如用户所输入的或者通过算法所导出的那样。区域140包含到与其他查询项目相关的概念的“See also(另见)”超链接，如用户所输入的或者通过算法所导出的那样，比如对应于“U2乐队”的概念的“Bono”或者对应于“U2侦察机”的概念的“Dragon Lady”。

[0053] 区域160包含搜索结果以及用户反馈机制170。利用该用户反馈机制170，用户可以针对相应的Web页面与他所寻找的内容的匹配程度来进行评分。换句话说，如果在区域160中列出的第一Web页面包含用户所寻求的关于摇滚乐队U2的相关信息，则该用户可以使用用户反馈机制170来给该链接评高分(例如五星)。专用于称为“U2”的时装品牌的名称的第二Web页面与用户所寻求的概念无关但是也被列在区域160 中，可以给该第二Web页面评低分(例如1星)。根据本发明，当同样对“U2”乐队感兴趣的稍后的用户利用查询“U2”进行搜索

时，被返回给他的结果列表将包含更靠近该结果列表的顶部的第一Web页面(被评为5星)以及更靠近该结果列表的底部或者甚至根本没有被列出的第二Web页面(被评为1星)。这样，将向用户呈现首先仅列出最相关的结果的结果列表。顺序地访问该结果列表中的站点的用户有更大的机会看到与其所寻求的概念最相关的站点。因此，除了元数据和用户没有输入的其他信息之外，结果列表中的各项目的顺序还基于用户反馈。

[0054] 用户可以添加关于与查询项目相关的一个或多个概念的描述120和130，从而提供关于该查询所提到的概念的一些背景信息，或者提供关于如何搜索关于该概念的信息的建议。用户还可以修改、增强或者除去由他们自己或者其他用户先前添加或修改的关于与查询项目相关的概念的描述。用户可以添加、修改或删除链接到与不同查询项目相关的概念的超链接或“See also(另见)”索引140。用户可以添加针对某一概念的所建议的查询150，当其被点击时，导致把所述查询提交给搜索引擎。该搜索引擎还可以利用计算机算法来生成所建议的查询项目。

[0055] 用户能够添加或保存到他们认为与所述概念高度相关的文献的链接。用户可以保存到他们认为与所述概念高度相关的文献的链接。这可以通过点击被标记为“Save(保存)”的超链接或图标190来实现，或者可以由诸如“Bookmark(书签)”、“Tag(标签)”或“Add to Favorites(添加到收藏夹)”之类的其他项目来指代。由于不同的用户对于哪些站点是最相关的将具有不同的想法，因此根据本发明的算法确定所列出的站点的顺序。在一个实施例中，所述算法使用民主过程，从而使得接收到最多“投票”(例如被数目最多的用户“保存”)的文献被放置在结果列表中的更高位置处。

[0056] 如果到被“保存”的文献的链接还出现在由搜索引擎生成的结果列表中，则可以使用图标165来表明该链接还是已经由用户投了票的链接。此外，在每个搜索结果下面有一个“By(由)”条目167，其示出添加该链接的用户的姓名，从而其可以作为该结果列表的一部分而被返回，并且在每个搜索结果下面还有一个“Tags(标签)”条目168，其列出用户用来对所述链接加标签的项目或者由先前搜索生成的项目。

[0057] 根据本发明，到Web站点的链接可以按照两种方式被列出，或者是 两个单独的列表：结果列表(算法的)和用户输入的链接，或者是被集成到一个列表中，其中如上所述地用图标来标记用户输入的链接。

[0058] 将会认识到，可以根据本发明进行许多修改。例如，可以从文件中读取用户生成的反馈，而不是由用户直接从终端输入。此外，虽然结果页面100示出了诸如“See also(另见)”链接140之类的区域，但是将会认识到，根据本发明，可以利用任何区域组合来显示包含用户输入的信息的结果页面，其中包括图1中示出的组合或者作为对图1中示出的组合的补充。该信息被用来使得搜索结果更加全面、精确以及有意义。

[0059] 图2是说明根据本发明的因特网搜索应用200的操作的流程图。该因特网搜索应用200能够让用户向系统提供反馈，从而基于对所述用户反馈的分析来允许其他用户接收到更相关的搜索结果。所述信息被用来调节由所述搜索引擎使用来在响应于特定查询而生成的结果列表中对文献进行排序的方法和权重。因此可以响应于用户反馈而对搜索引擎进行“调整”，以便返回更相关的结果。

[0060] 在步骤210中，用户向搜索引擎提交查询。该过程随后继续到步骤220，在该步骤中，该搜索引擎对该查询进行匹配，以便组合出最相关结果的列表。步骤220继续到步骤

230，在该步骤中，向用户发送结果页面（例如图1中的100）。步骤230继续到步骤240或步骤250。

[0061] 在步骤240中，用户遵循一个或多个所述链接以访问结果列表中的各Web页面。可选择地，在步骤250中，用户能够与反馈机制（例如图1的区域170）交互，以便向搜索引擎提供反馈。在步骤250中，用户能够点击用来保存链接的机制（例如图1的区域190），以便用搜索引擎来记录链接。用户可以从访问Web站点的步骤240继续到步骤250以给出反馈，以及可选择地，用户可以从步骤250继续到步骤240。步骤240和250都通向步骤260，在该步骤中，搜索引擎记录来自用户的反馈。步骤260通向步骤270，在该步骤中，更新用于搜索相关性的模型以用在后续搜索中。步骤270循环回到步骤220，在该步骤中，搜索引擎利用由更新后的模型提供的值来确定哪些链接对于后续查询是相关的。

[0062] 将会认识到，本发明的实施例能够用于一个或多于一个的搜索引擎。作为一个例子，第一搜索引擎根据由该第一搜索引擎采用的相关性因素返回文献列表。根据本发明，第二搜索引擎随后能够单独对上述结 果进行排序，或者能够把上述结果与由该第二搜索引擎生成的结果相组合地进行排序。根据本发明，随后能够使用所有的或者任何的结果组合来更新相关性模型。因此，根据本发明的搜索能够运行在元搜索引擎上。

[0063] 根据一个实施例的系统是基于具有注册用户和未注册访客的Web搜索引擎入口站点。与所有搜索引擎一样，对于每个用户查询，从最高相关性分数到最低相关性分数对文献的统一资源定位符（urls）进行排序，并且按照该顺序将其呈现回给用户。所述相关性分数是文献d关于在X中包含的给定特征（或所观察到的证据）与查询q相关的概率$p(R|X)$，其中为了清楚起见在这里和后面略去了下标。通过函数$\mu(X)=1/[1+e^{-\theta*X}]$来近似$p(R|X)$，其中X是特征的矢量，以及θ是参数的矢量，其包括乘以1.0的常数输入的截距项；取负的θ，从而使得正系数表示肯定的相关性证据。假设使用逻辑斯谛回归来对该概率进行建模，X由特定于文献的特征（例如文献长度等等）以及查询-文献特征（比如文献标题中的查询出现）构成，文献从搜索引擎E等等得到排序K，并且随后将被U增强，其中U是一组稀疏指标变量，其对应于查询与文献间的用户评分。

[0064] 在所述引擎上观察到查询的结果之后，用户被允许按照某种序数形式关于与该查询对应的文献给出反馈，其中最低等级（例如一星）意味着完全谬误，而最高等级（例如五星）表示完全相关或者接近完全相关。随后作为新的查询-文献特征来记录反馈$(q,d,u)=s$，其中s是用户u所分配的星数。在任何给定时刻，该值就是用户u对于文献d与查询q的当前评分。重要的是，该特征不取决于用户执行该查询的次数，而仅仅存储他或她的后来的评分。通过收集对应于给定的查询-文献对的所有用户反馈，如果用户u对于给定查询q为文献d分配了5星，则获得新的一组稀疏指标变量$U(q,d)_{u,s}=1$：随后利用U中的特征来增强该特征组，并且利用矢量θ来对完整的模型进行参数化。在收集用户反馈之前，不知道或者不清楚所述星数，从而表明用户尚未明确评估文献的相关性。这是重要的，因为将纯粹由已经被用户评过分的文献形成数据集。

[0065] 所述系统采用用户反馈来执行三个基本功能。首先，所述系统使用反馈来按照对θ中的全局模型参数的添加的形式开发特定于用户的参数矢量ψ。其次，所述系统能够在用户ψ矢量的总体空间上对用户进行聚类。可以使用诸如K均值聚类之类的简单技术，可以预期所述技术能够 良好地工作，这是因为各维度是具有非常纯的数值属性（比如完全相同的尺

度、独立性等等)的回归系数。这可以被用来建立一组特定于群集的参数矢量,比如可以按照与特定于用户的ψ类似的方式被添加到矢量θ的γ。第三,所述系统能够在建模过程中作为数据点结合用户反馈事件以便估计θ,以及结合用户的ψ矢量的属性以便对特定事件的重要性进行加权。

[0066] 当用户对于某一查询-文献对公布一个星数时,在特定于用户的数据集中创建一个数据点。把值1.0与高度相关的评分相关联,并且把0.0与谬误的或者不相关的评分相关联。在五星系统中,五星被用于高度相关,而一星被用于不相关。所述系统使用其他评分作为对应于其他用户参数和全局参数的特征。在技术上,如果期望对完全分布进行建模,则将需要把输出变量作为从多项式分布中抽取的来对待。随后把所述数据点添加到更新数据库,该更新数据库由在所述系统中追溯到某一确立时间段或者到某一最大数据集尺寸的所有评分事件数据点构成。可以在数百万个数据点上执行稀疏逻辑斯谛回归。在到达该极限之后,针对最后N个数据点执行所述回归,其中可以预期N是以百万计。对于任何给定的查询-文献对,在每用户评分事件的基础上如下指定所估计的模型:

[0067] $$\mu(X) = 1/[1 + e^{-(\theta+\psi_u)*X}]$$ 等式(1)

[0068] 其中θ被取为固定的,并且在ψ中的用户参数上执行所述回归。U中的用户评分指标也被结合,从而ψ_u潜在地包含对应于V中的其他用户的指标$U(q,d)_{v,s}=1$的系数。例如,如果用户v对于查询q为某一文献评了五星,则这可以具有一个正系数,从而表明其是该文献对于用户u高度相关的肯定证据。所述增强的特征组包括对应于高度相关的自指标变量$U(q,d)_{u,5}$,因为这是通过其把该实例包括在所述数据集中的机制。这样包括非常稀疏的自指标将允许所述回归移动经过潜在的矛盾并且在所述模型中前进,以便确保所述文献对于该查询和用户将得到高分。另一方面,如果用户u正在寻找大体效果,则该稀疏自指标将不被赋予重的权重。在所述过程的结尾,提交了评分事件的每个用户u将具有可以被检查、被用于聚类等等的唯一的ψ_u,其为该用户个人化所述相关性打分模型。

[0069] 一旦在所述系统中有了对应于每个所识别的用户的ψ矢量之后,就可以把所述矢量组织成一个坐标空间并且执行聚类。这样做可能是有帮助的,因为可能期望隔离将具有古怪的参数以反映其对于自身内容或相关内容的任意偏向的搜索引擎优化器或者兜售器(spammer)。另外,所识别的用户的硬群集能够被用来为未识别的用户(即未注册的"访客")提供更好的默认参数。例如,基于发出查询q的未识别的用户,可以通过如下计算p(c|q)来形成该用户由群集c表示的概率:

[0070] $p(c|q)=p(q|c)p(c)/[\Sigma_{c\in C}p(q|c)p(c)]$

[0071] 对于所述该组用户群集C使用贝叶斯法则。随后所述相关性分数p(R|X)将等于每个群集成员概率乘以相关性分数或者对应于给定群集成员的概率$p_c(R|X)$的加权和:

[0072] $p(R|X)=\Sigma_{c\in C}p(c|q)p_c(R|X)$ 等式(2)

[0073] 其中,对应于每个群集γ_c的参数将被计算为对于全局参数θ的添加。以下描述这样的过程。

[0074] 在估计了特定于用户的参数γ_c之后,对于以θ参数化的原始模型执行更新。这是通过使用被开发来估计ψ_u的相同的用户评分数据集合来实现的,但是这次ψ参数是固定的。因此,尝试以交替的方式从所有用户的累积体验估计大体效果——第一用户特定效果以及

随后的全局效果。在对于每个用户估计ψ参数(通过单个大尺度回归实现)的过程中所学习到的东西是适当地对于每种情况有区别地过加权及欠加权。例如,假设ψ中的系数的独立性,则能够利用在由用户u提交的数据上的内核函数把每个数据点的值计算为所述用户与用户群体的平均值的相似性:

[0075] $$K(u) = e^{-\alpha\|\psi_u - \psi_{avg}\|^2}$$

[0076] 该函数将取得最大值1.0,并且对于其ψ矢量更加远离所有ψ矢量的形心的用户将以速率α衰减。利用该方法,所述回归的形式将是在给定θ的情况下最大化数据的对数似然的和:

[0077] $\max\ l(\theta) = \Sigma\ \omega_i[y_i\log(\mu(\theta+\psi_{u(i)})_i)+(1-y_i)\log(1-\mu(\theta+\psi_{u(i)})_i)$等式(3)

[0078] 其中,在θ上取得所述最大值,并且求和是从i=1到N,其中N是数据集D中的数据点的数目,如果数据点被评为高度相关,则y_i是1.0,以及如果数据点被评为不相关,则y_i是0.0,ω_i是基于提交者的ψ_u参数的属性的数据集中的每次观测i的权重,并且为了清楚起见略去对于数据D的依赖性。

[0079] 可选择地,通过使用前面讨论的硬聚类步骤的结果,参数γ_c能够是适合的,其根据哪个群集被分配给提交所述评分数据点的用户来修改θ。硬聚类指的是把每个用户分配给单个群集,其与软聚类相对,软聚类是把用户分配给多个群集,在每个群集中具有某种程度的成员性。在这种情况下,利用等式(3)执行回归,但是在θ以及其中指定了μ的γ_c上联合执行优化:

[0080] $\mu(X;\theta,\gamma_c) = 1/[1+e^{-(\theta+\gamma c+\psi u)*X}]$

[0081] 基于提交特定数据点的用户的群集成员性来分配γ_c,所述特定数据点被用来生成特征X以及计算$\mu(\theta,\gamma)_i$。如早前对于特定于用户的ψ回归所示出的那样,对于在γ_c中包含的每个参数,在评分行为中的任何冲突都可以被吸收到特定于群集的修改器中。

[0082] 作为使用所述γ的一个具体例子,假设由200个用户在从几个到数百个评分事件中提交的数据集,其中对于每个用户从群集组C中进行群集分配c。当估计θ(比如θ($EngineRank_{E,4}$))时,还估计相应的γ_c,比如γ_k($Enginerank_{E,4}$),其中假设所述用户在群集k中。在另一个数据点中估计θ($EngineRank_{E,4}$),但是这一次对于群集m确定γ_m($Enginerank_{E,4}$),其中假设文献对于所述查询具有$EngineRank_{E,4}$,并且所述两个用户分别来自群集k和m。结果,现在θ($EngineRank_{E,4}$)是所述变量的平均效果,并且γ($Enginerank_{E,4}$)包含所述变量的特定于群集的效果,例如一个SEO群集可能具有负系数,这是因为它们将很有可能没有平均到良好的变量那么敏感,但是它们的有害影响被有效地从全局模型中除去。在所述模型拟合过程内,从根据用户的群集分配来填充特征矢量那时以来就都是相同的。

[0083] 在对于全局更新所描述的两种方案中,用户的评分确定每个用户对全局参数的拟合的影响。在每种情况下,基于用户的评分行为的用户输入的效果被自动限制。

[0084] 通过对于所述用户优化第一组参数以及随后对于所述全局和/或群集效果优化另一组参数,所述拟合过程的焦点在两个不同的分析级别之间交替。并不完全清楚应当在建模过程中顺序地执行所述优化还是交替优化的单独迭代。

[0085] 图3说明根据本发明的系统300的组件。该系统300包括用户客户端305,其连接到

Web服务器310。Web服务器310被耦合到搜索引擎320、用户数据库330和反馈数据库340。搜索引擎320被耦合到包含文献索引的数据存储库350。用户数据库330也被耦合到搜索引擎320。反馈数据库340被耦合到机器学习模型360以便计算新的相关性因子。机器学习模型360也被耦合到包含文献索引的数据存储库350，该数据存储库350又被耦合到索引器370。索引器370被耦合到Web内容数据库380，其被耦合到Web爬虫(crawler)390。Web爬虫390通过因特网395被耦合到一个或多个Web站点399。

[0086] 在操作中，web爬虫390在因特网395上进行导航，从而访问Web站点399，并且利用其所访问的Web页面的内容来填充Web内容数据库380。索引器370使用Web内容数据库380来创建文献索引350。当用户在用户客户端305上生成查询时，Web服务器310把该搜索请求传送到搜索引擎320。搜索引擎320使用相关性算法以及从上述用户反馈导出的因子来确定哪些Web页面可能与该查询最相关，并且创建结果列表，该结果列表被发送到Web服务器310。Web服务器310随后把结果页面提供给用户客户端以进行显示。

[0087] 此外，当用户实施搜索时，他利用用户客户端305输入查询，该查询被提交给Web服务器310。Web服务器310把该查询提交给搜索引擎320，该搜索引擎320把该查询与文献索引350进行匹配以便确定最相关的文献，并且把结果列表返回到Web服务器310。此外，响应于该查询，用户数据库330记录关于该用户的搜索的信息，比如使用所述保存链接机制(例如图1中的区域190)保存的链接、后面的链接(例如图1中的区域160)、以及利用所述反馈机制(例如图1中的区域170)给出的反馈。该信息被Web服务器310和搜索引擎320使用来定制对应于 该用户的后续搜索结果。此外，响应于查询，还在反馈数据库340中记录来自所述反馈机制(例如图1中的区域170)的反馈。在本发明的一个实施例中，在用户数据库330和反馈数据库340中存储的反馈信息可以被实施为两个单独的数据库，或者它们可以被实施在同一个数据库内。

[0088] 在适时的基础上(但是不必在执行查询时)，在反馈数据库340中包含的反馈信息被发送到机器学习模型360，在那里所述反馈信息被处理以便生成由搜索引擎使用来确定对应于查询的最相关的Web页面的方法和权重。机器学习模型360在文献索引350中记录该反馈信息，以便在后续搜索中使用。

[0089] 反馈数据库340把各特征发送到机器学习模型360，其中包括但不限于查询项目、用户标识符、文献ID、文献链接、结果列表中的位置、用户评分、以及用户点击。机器学习模型360还可以对于给定文献查找其他特征，其中包括但不限于项目分数、段信息、链接结构、锚文本概要、标签、文献内的项目、项目在文献内的位置、文献的结构、从搜索结果列表访问文献的次数、项目分数、段信息、链接结构、用户表示、用户输入的时间、阻绝。

[0090] 机器学习模型360使用这些特征开发出具有对全局模型参数的添加的形式的特定于用户的参数组。这些参数被如下导出：每个反馈事件构成一个数据点，该数据点被添加到反馈数据库340中的所有数据点的数据库。使用一个模型在每个用户评分事件上对于任何给定的查询-文献对进行估计，如上面的等式(1)所给出的那样。

[0091] 如前面所提到的那样，对输入数据执行回归。一旦找到了最佳参数，这些参数就被使用来更新全局模型和特定于用户的模型。一般来说，所述最佳参数将具有最佳的预测能力，这转换为关于未见过的数据的更好的结果。在本发明的一个实施例中，执行使用共轭梯度下降的逻辑斯谛回归，以作为所述建模过程的一部分。将会认识到，可以利用其他形式的

回归以及其他方法来执行根据本发明的建模过程。

[0092] 将会认识到，能够根据本发明使用许多类型的机器学习技术，其中包括使用某种形式的统计分类的那些技术。所述统计分类包括但不限于逻辑斯谛回归分析、支撑矢量机、神经网络、提升树、随机森林、朴素贝叶斯、图形模型、以及最大后验。在其他实施例中，所述机器学习技 术使用共轭梯度下降。

[0093] 图4是说明根据本发明一个实施例的利用用户反馈数据来计算结果的步骤400的流程图。在图4以及附随文字中提到的等式号指代所述等式。

[0094] 参考图4，在步骤410中，用户输入反馈数据，以及在步骤420中，该输入数据被规范化，这包括提取全局和每个用户的数据并且对其进行规范化。在步骤430中，使用等式(1)为各数据点打分，以及在步骤440中，利用上面的等式(3)计算目的函数。等式(1)可以包括许多形式，其中包括利用聚类来进行打分。在步骤450中，计算共轭梯度，以及在步骤460中，使用新的梯度来更新参数。在步骤470中，确定是否满足一个或多个标准。如果尚未满足所述一个或多个标准，则所述过程循环回到步骤430；否则，该过程继续到步骤480，在该步骤中更新全局和用户模型。

[0095] 任何数目和类型的标准都可以在步骤470中被用作停止标准。例如，所述停止标准可以包括但不限于：执行了预定的最大次数的迭代；(2)交叉验证失败(例如测试数据不同于试验数据)；以及(3)解收敛，也就是，在前一次迭代与下一次迭代中的参数之间的差小于预定值。

[0096] 图5说明根据本发明的供用户510使用的因特网搜索应用系统500的硬件组件。该系统500包括通过因特网530耦合到Web服务器540的客户端设备520。客户端设备520是用来访问Web服务器540的任何设备，其被配置成利用因特网协议进行通信，所述协议包括但不限于http(超文本传输协议)和WAP(无线应用协议)。优选地，客户端设备520是个人计算机，但是它也可以是另一设备，其中包括但不限于诸如蜂窝电话或个人数字助理(PDA)之类的手持设备，并且它能够使用诸如HTML(超文本标记语言)、HDML(手持设备标记语言)、WML(无线标记语言)等等之类的标准来呈现信息。

[0097] Web服务器540被耦合到搜索服务器550和反馈数据存储装置560。该反馈数据存储装置560被耦合到机器学习服务器570，并且搜索服务器550被耦合到索引数据存储装置580。另外，机器学习服务器570被耦合到索引数据存储装置580。

[0098] 本领域技术人员将容易显而易见的是，在不背离由所附权利要求书限定的本发明的精神和范围的情况下可以对实施例做出其他修改。

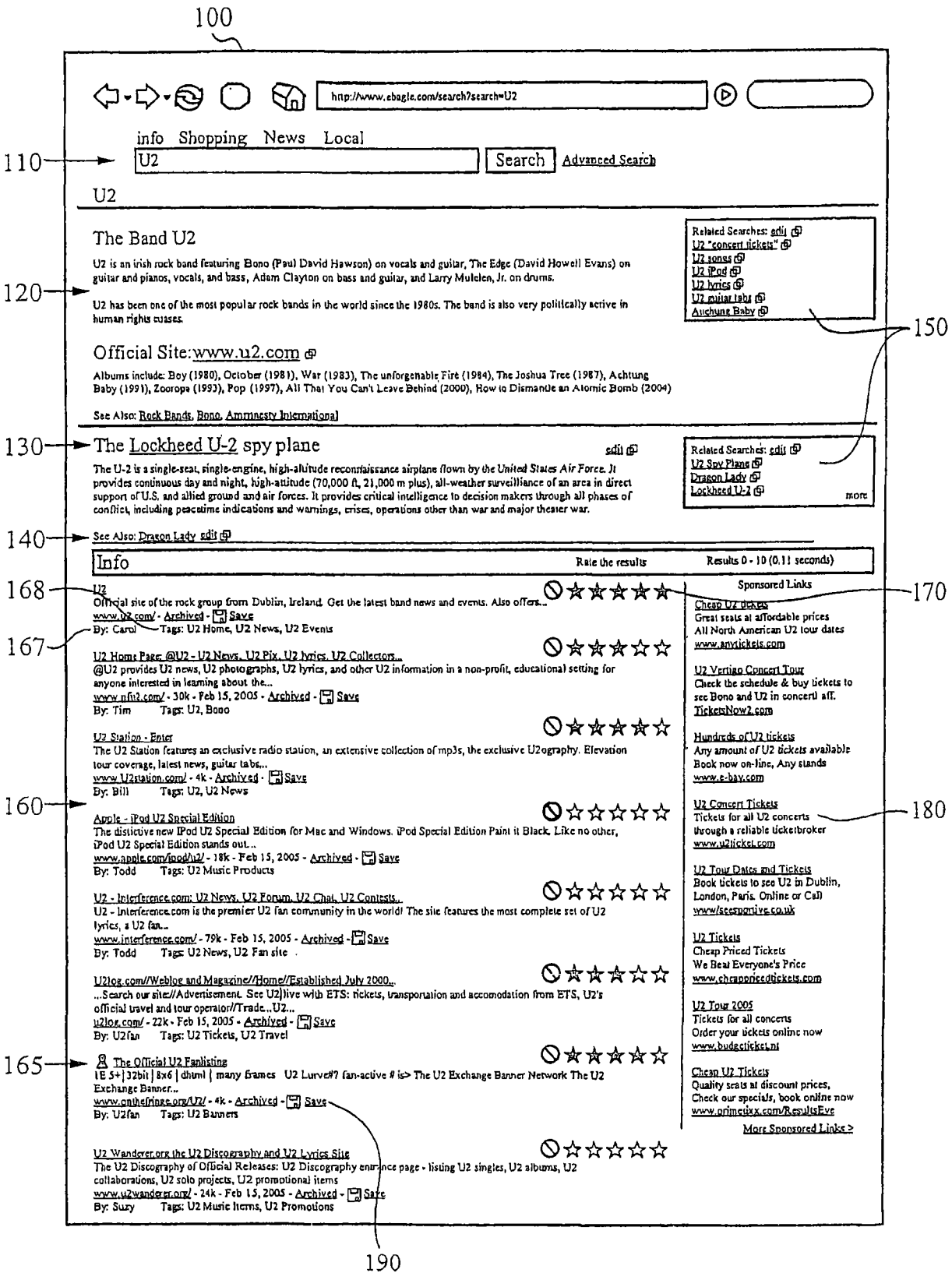

100
http://www.ebagle.com/search?search=U2
info Shopping News Local
110
U2
Search
Advanced Search
U2
The Band U2
U2 is an irish rock band featuring Bono (Paul David Hawson) on vocals and guitar, The Edge (David Howell Evans) on guitar and pianos, vocals, and bass, Adam Clayton on bass and guitar, and Larry Mulelen, Jr. on drums.
120
U2 has been one of the most popular rock bands in the world since the 1980s. The band is also very politically active in human rights cuases.
Related Searches: edit
U2 "concert tickets"
U2 songs
U2 iPod
U2 lyrics
U2 guitar tabs
Auchtung Baby
150
Official Site:www.u2.com
Albums include: Boy (1980), October (1981), War (1983), The unforgenable Fire (1984), The Joshua Tree (1987), Achtung Baby (1991), Zooropa (1993), Pop (1997), All That You Can't Leave Behind (2000), How to Dismantle an Atomic Bomb (2004)
See Also: Rock Bands, Bono, Ammnesty International
130
The Lockheed U-2 spy plane
edit
The U-2 is a single-seat, single-engine, high-altitude reconnaissance airplane flown by the United States Air Force. It provides continuous day and night, high-attitude (70,000 ft, 21,000 m plus), all-weather surveilliance of an area in direct support of U.S. and allied ground and air forces. It provides critical intelligence to decision makers through all phases of conflict, including peacetime indications and warnings, crises, operations other than war and major theater war.
Related Searches: edit
U2 Spy Plane
Dragon Lady
Lockheed U-2
more
140
See Also: Dragon Lady edit
Info
Rate the results
Results 0 - 10 (0.11 seconds)
Sponsored Links
168
U2
Official site of the rock group from Dublin, Ireland. Get the latest band news and events. Also offers...
www.u2.com/ - Archived - Save
By: Carol Tags: U2 Home, U2 News, U2 Events
167
U2 Home Page: @U2 - U2 News, U2 Pix, U2 lyrics, U2 Collectors...
@U2 provides U2 news, U2 photographs, U2 lyrics, and other U2 information in a non-profit, educational setting for anyone interested in learning about the...
www.atu2.com/ - 30k - Feb 15, 2005 - Archived - Save
By: Tim Tags: U2, Bono
U2 Station - Enter
The U2 Station features an exclusive radio station, an extensive collection of mp3s, the exclusive U2ography. Elevation tour coverage, latest news, guitar tabs...
www.U2station.com/ - 4k - Archived - Save
By: Bill Tags: U2, U2 News
160
Apple - iPod U2 Special Edition
The distictive new IPod U2 Special Edition for Mac and Windows. iPod Special Edition Paint it Black. Like no other, iPod U2 Special Edition stands out...
www.apple.com/ipod/u2/ - 18k - Feb 15, 2005 - Archived - Save
By: Todd Tags: U2 Music Products
U2 - Interference.com: U2 News, U2 Forum, U2 Chat, U2 Contests...
U2 - Interference.com is the premier U2 fan community in the world! The site features the most complete set of U2 lyrics, a U2 fan...
www.interference.com/ - 79k - Feb 15, 2005 - Archived - Save
By: Todd Tags: U2 News, U2 Fan site
U2log.com//Weblog and Magazine//Home//Established July 2000...
...Search our site//Advertisement. See U2|live with ETS: tickets, transportation and accomodation from ETS, U2's official travel and tour operator//Trade...U2...
u2log.com/ - 22k - Feb 15, 2005 - Archived - Save
By: U2fan Tags: U2 Tickets, U2 Travel
165
The Official U2 Fanlisting
IE 5+ | 32bit | 8x6 | dhtml | many frames U2 Lurve!? fan-active # is> The U2 Exchange Banner Network The U2 Exchange Banner...
www.onthefringe.org/U2/ - 4k - Archived - Save
By: U2fan Tags: U2 Banners
U2 Wanderer.org the U2 Discography and U2 Lyrics Site
The U2 Discography of Official Releases: U2 Discography entrance page - listing U2 singles, U2 albums, U2 collaborations, U2 solo projects, U2 promotional items
www.u2wanderer.org/ - 24k - Feb 15, 2005 - Archived - Save
By: Suzy Tags: U2 Music Items, U2 Promotions
190
Cheap U2 tickets
Great seats at affordable prices
All North American U2 tour dates
www.anytickets.com
U2 Vertigo Concert Tour
Check the schedule & buy tickets to see Bono and U2 in concert! aff.
TicketsNow2.com
Hundreds of U2 tickets
Any amount of U2 tickets available
Book now on-line, Any stands
www.e-bay.com
U2 Concert Tickets
Tickets for all U2 concerts through a reliable ticketbroker
www.u2ticket.com
180
U2 Tour Dates and Tickets
Book tickets to see U2 in Dublin, London, Paris. Online or Call
www/seesportive.co.uk
U2 Tickets
Cheap Priced Tickets
We Beat Everyone's Price
www.cheappricedtickets.com
U2 Tour 2005
Tickets for all concerts
Order your tickets online now
www.budgetticket.ni
Cheap U2 Tickets
Quality seats at discount prices, Check our specials, book online now
www.primeuxx.com/ResultsEve
More Sponsored Links >

图 1

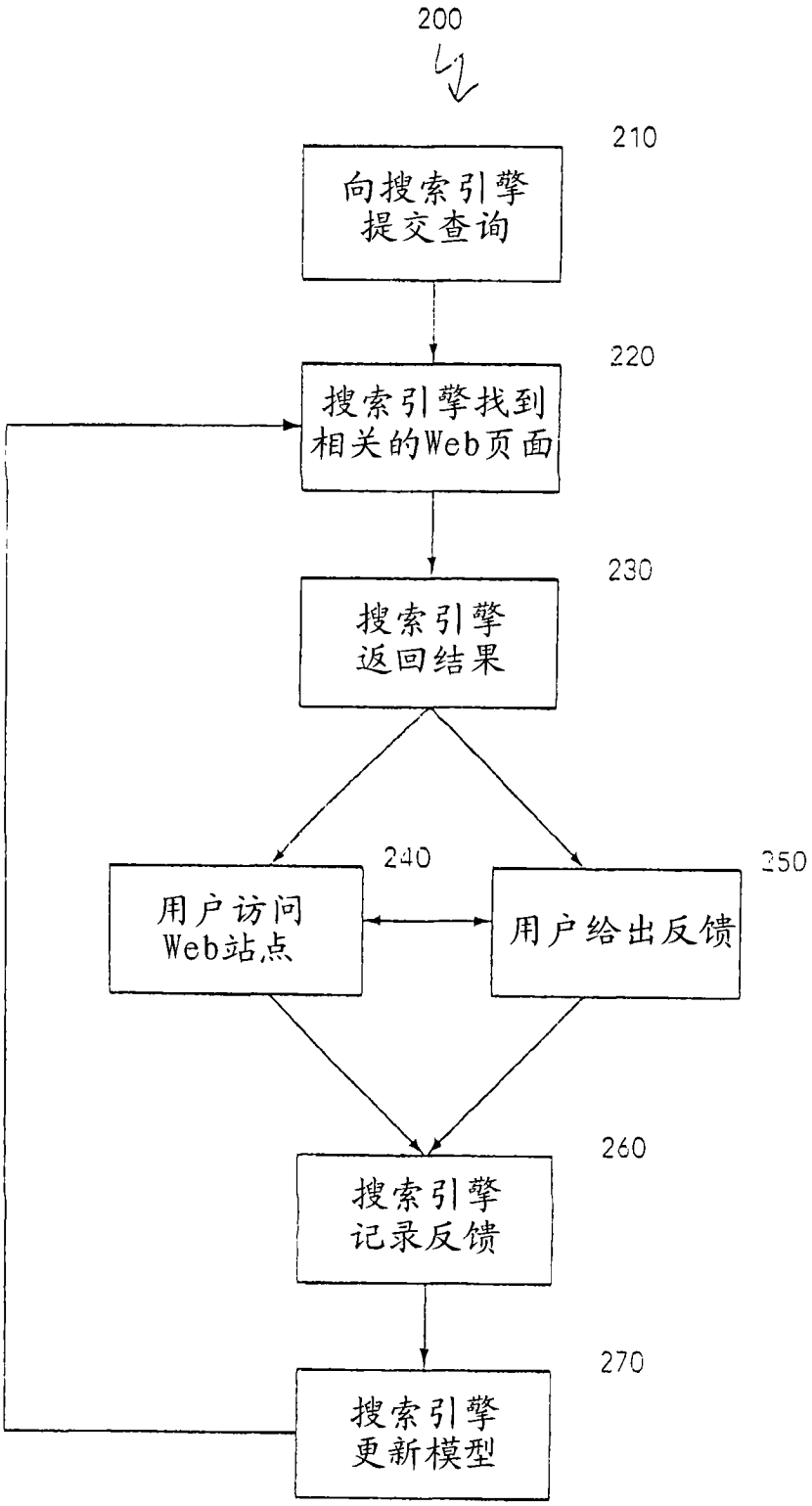

200
210
向搜索引擎提交查询
220
搜索引擎找到相关的Web页面
230
搜索引擎返回结果
240
用户访问Web站点
250
用户给出反馈
260
搜索引擎记录反馈
270
搜索引擎更新模型

图 2

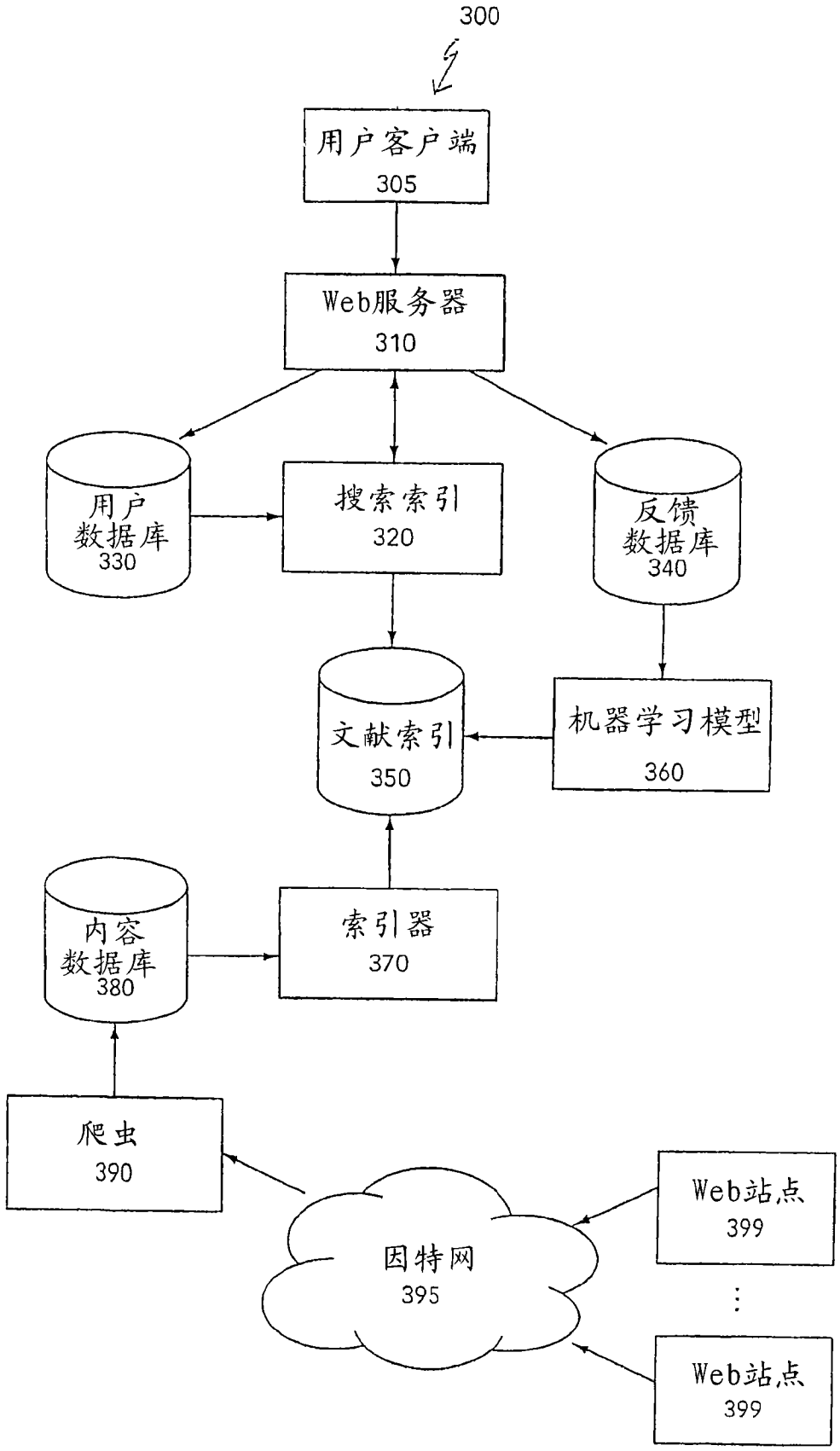

300
用户客户端
305
Web服务器
310
用户
数据库
330
搜索索引
320
反馈
数据库
340
文献索引
350
机器学习模型
360
内容
数据库
380
索引器
370
爬虫
390
因特网
395
Web站点
399
Web站点
399

图 3

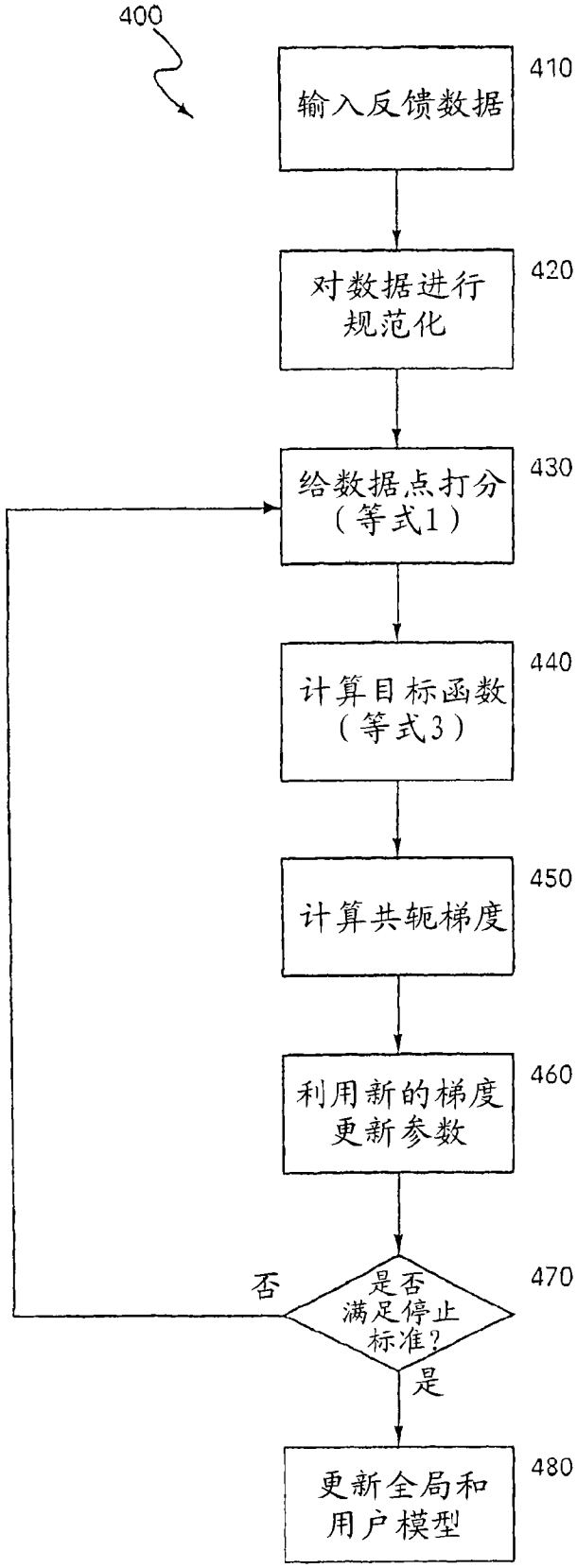

400
输入反馈数据
410
对数据进行规范化
420
给数据点打分（等式1）
430
计算目标函数（等式3）
440
计算共轭梯度
450
利用新的梯度更新参数
460
是否满足停止标准？
470
否
是
更新全局和用户模型
480

图 4

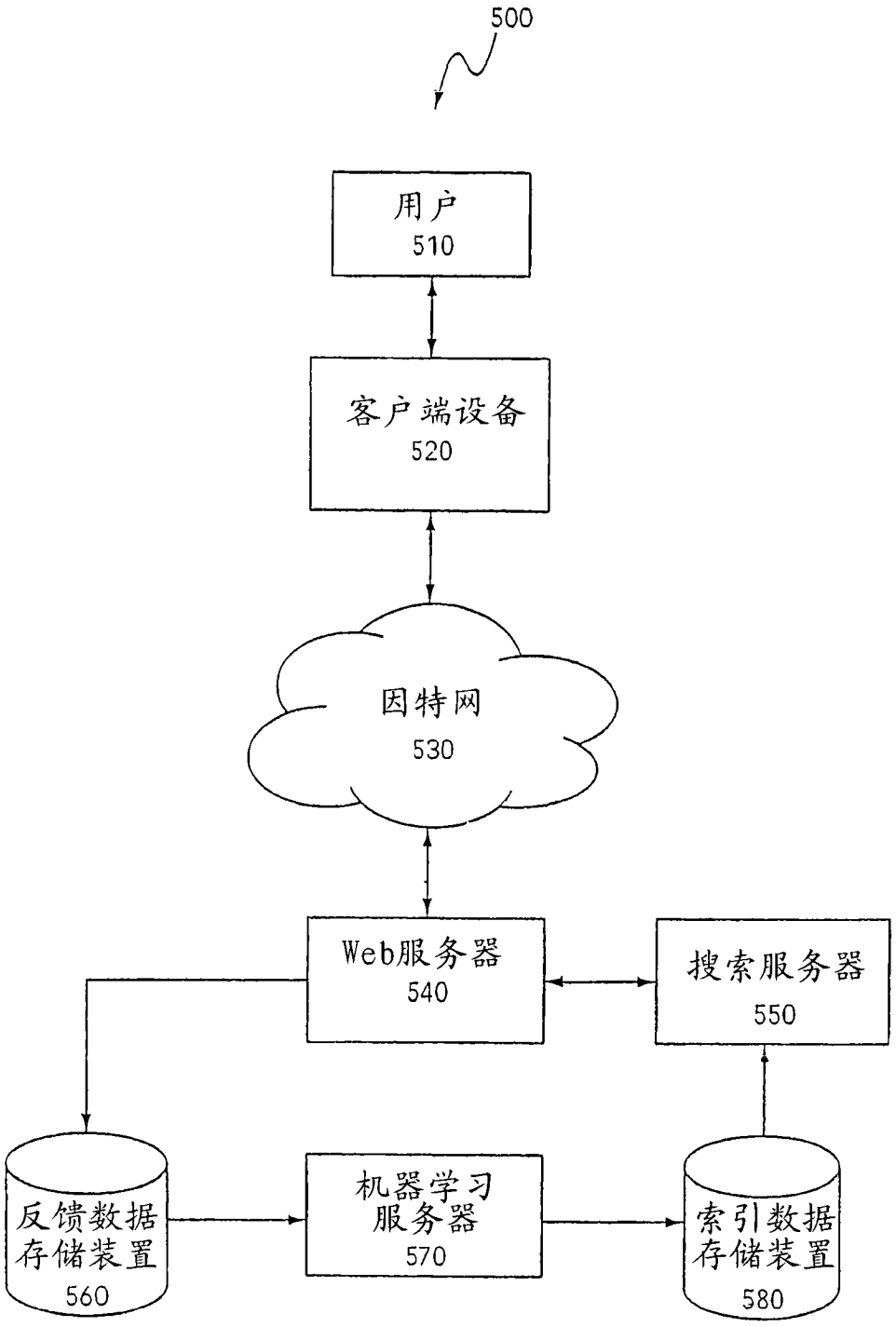

500
用户
510
客户端设备
520
因特网
530
Web服务器
540
搜索服务器
550
反馈数据
存储装置
560
机器学习
服务器
570
索引数据
存储装置
580

图 5