



[12] 发明专利申请公布说明书

[21] 申请号 200580015720.4

[43] 公开日 2007年8月29日

[11] 公开号 CN 101027667A

[22] 申请日 2005.3.29
 [21] 申请号 200580015720.4
 [30] 优先权
 [32] 2004.3.31 [33] US [31] 10/813,359
 [86] 国际申请 PCT/US2005/010701 2005.3.29
 [87] 国际公布 WO2005/098680 英 2005.10.20
 [85] 进入国家阶段日期 2006.11.16
 [71] 申请人 GOOGLE 公司
 地址 美国加利福尼亚
 [72] 发明人 卡尔·菲勒格

[74] 专利代理机构 中国国际贸易促进委员会专利商
 标事务所
 代理人 陈 炜

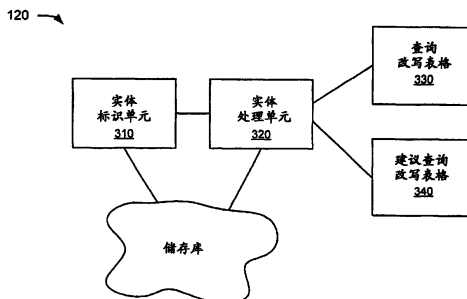
权利要求书 10 页 说明书 20 页 附图 19 页

[54] 发明名称

具有实体检测的查询改写

[57] 摘要

一种系统，确定接收的搜索查询是否包括实体名称，基于与涉及该实体名称的先前搜索有关的信息、确定是否改写所接收的搜索查询，以及当确定应当改写所接收的搜索查询时，改写所接收的搜索查询。该系统还在确定所接收的搜索查询不应当被改写时，基于与涉及该实体名称的先前搜索有关的信息、确定是否提供改写所接收的搜索查询的建议作为改写建议。当确定应该提供改写建议时，该系统还生成该改写建议。该系统基于所接收的搜索查询和所改写的搜索查询之一执行搜索以获得搜索结果，给出该搜索结果，以及当确定应该提供改写建议时，给出该改写建议。



1、一种方法，包含：

接收搜索查询；

确定所接收的搜索查询是否包括实体名称；

基于与涉及该实体名称的先前搜索有关的信息，确定是否改写所接收的搜索查询；

当确定应当改写所接收的搜索查询时，改写所接收的搜索查询；

基于所接收的搜索查询和所改写的搜索查询之一执行搜索，以获得搜索结果；以及

给出该搜索结果。

2、如权利要求 1 所述的方法，还包含：

当基于所改写的搜索查询执行搜索时，提供到所接收的搜索查询的链接。

3、如权利要求 1 所述的方法，还包含：

当确定所接收的搜索查询不应当被改写时，基于与涉及该实体名称的先前搜索有关的信息，确定是否提供改写所接收的搜索查询的建议作为改写建议；以及

当确定应该提供改写建议时，生成该改写建议。

4、如权利要求 3 所述的方法，其中，所述改写建议包括到改写的搜索查询的链接。

5、如权利要求 3 所述的方法，其中，所述给出搜索结果的步骤包含：

当确定应该提供改写建议时，与搜索结果一起给出该改写建议。

6、如权利要求 3 所述的方法，其中，所述确定是否提供改写所接收的搜索查询的建议的步骤包含：

标识与在涉及该实体名称的先前搜索中被选择的文档相关联的实体标识符；

确定每个被标识的实体标识符的选择总数；以及

当与该实体名称相关联的实体标识符接收到的选择总数大于其他所标识的实体标识符时，确定应该提供改写建议。

7、如权利要求 1 所述的方法，其中，所述确定是否改写所接收的搜索查询的步骤包含：

标识与在涉及该实体名称的先前搜索中被选择的文档相关联的实体标识符；

确定每个被标识的实体标识符的选择总数；以及

当与该实体名称相关联的实体标识符接收到的选择总数大于其他所标识的实体标识符时，确定应该改写所接收的搜索查询。

8、如权利要求 7 所述的方法，其中，所述确定是否改写所接收的搜索查询的步骤还包含：

确定与该实体名称相关联的实体标识符的选择总数是否大于阈值；以及

当与该实体名称相关联的实体标识符的选择总数不大于该阈值时，确定不应当改写所接收的搜索查询。

9、如权利要求 1 所述的方法，其中，所述确定是否改写所接收的搜索查询的步骤包含：

标识与在涉及该实体名称的先前搜索中被选择的文档相关联的实体标识符；

确定每个被标识的实体标识符的选择总数的分布；以及

当所述分布指示与该实体名称相关联的实体标识符的选择总数和其它所标识的实体标识符的子集的选择总数相比被峰值化了时，确定应该改写所接收的搜索查询。

10、如权利要求 1 所述的方法，其中，所述改写所接收的搜索查询的步骤包含：

修改所接收的搜索查询，以包括与该实体名称相关联的限制标识符。

11、如权利要求 1 所述的方法，其中，所述基于所接收的搜索查询和所改写的搜索查询之一执行搜索的步骤包含：

当改写了所接收的搜索查询时，使用所改写的搜索查询搜索文档储存库。

12、一种系统，包含：

用于接收搜索查询的装置；

用于确定所接收的搜索查询是否包括实体名称的装置；

用于当所接收的搜索查询包括实体名称时、确定是否改写所接收的搜索查询的装置；

用于当确定应当改写所接收的搜索查询时、改写所接收的搜索查询的装置；

用于基于所接收的搜索查询和所改写的搜索查询之一执行搜索以获得搜索结果的装置；以及

用于提供该搜索结果的装置。

13、一种系统，包含：

存储器，用于存储与先前搜索有关的信息；以及

处理器，用于：

接收搜索查询；

确定所接收的搜索查询是否包括实体名称；

基于该存储器中的信息确定是否改写所接收的搜索查询；

当确定应当改写所接收的搜索查询时，改写所接收的搜索查询；

基于所接收的搜索查询和所改写的搜索查询之一执行搜索，以获得搜索结果；以及

给出该搜索结果。

14、一种方法，包含：

接收搜索查询；

确定所接收的搜索查询是否包括实体名称；

基于与涉及该实体名称的先前搜索有关的信息，确定是否提供改写所接收的搜索查询的建议作为改写建议；

当确定应该提供改写建议时，生成该改写建议；

基于所接收的搜索查询执行搜索，以获得搜索结果；

给出该搜索结果；以及

当应该提供改写建议时，提供该改写建议。

15、如权利要求 14 所述的方法，其中，所述改写建议包括到改写的搜索查询的链接。

16、如权利要求 14 所述的方法，还包含：

接收对该改写建议的选择；

改写所接收的搜索查询；

基于所改写的搜索查询执行搜索，以获得新的搜索结果；以及
给出该新的搜索结果。

17、如权利要求 14 所述的方法，其中，所述确定是否提供改写所接收的搜索查询的建议的步骤包含：

标识与在涉及该实体名称的先前搜索中被选择的文档相关联的实体标识符；

确定每个被标识的实体标识符的选择总数；以及

当与该实体名称相关联的实体标识符接收到的选择总数大于其他所标识的实体标识符时，确定应该提供改写建议。

18、如权利要求 14 所述的方法，还包含：

基于与涉及该实体名称的先前搜索有关的信息，确定是否改写所接收的搜索查询；以及

当确定应当改写所接收的搜索查询时，改写所接收的搜索查询。

19、如权利要求 18 所述的方法，其中，所述确定是否改写所接收的搜索查询的步骤包含：

标识与在涉及该实体名称的先前搜索中被选择的文档相关联的实体标识符；

确定每个被标识的实体标识符的选择总数；以及

当与该实体名称相关联的实体标识符接收到的选择总数大于其他所标识的实体标识符时，确定应该改写所接收的搜索查询。

20、如权利要求 19 所述的方法，其中，所述确定是否改写所接

收的搜索查询的步骤还包含:

确定与该实体名称相关联的实体标识符的选择总数是否大于阈值; 以及

当与该实体名称相关联的实体标识符的选择总数不大于该阈值时, 确定不应当改写所接收的搜索查询。

21、如权利要求 18 所述的方法, 其中, 所述确定是否改写所接收的搜索查询的步骤包含:

标识与在涉及该实体名称的先前搜索中被选择的文档相关联的实体标识符;

确定每个被标识的实体标识符的选择总数的分布; 以及

当所述分布指示与该实体名称相关联的实体标识符的选择总数和其它所标识的实体标识符的子集的选择总数相比被峰值化了时, 确定应该改写所接收的搜索查询。

22、如权利要求 18 所述的方法, 其中, 所述改写所接收的搜索查询的步骤包含:

修改所接收的搜索查询, 以包括与该实体名称相关联的限制标识符。

23、如权利要求 18 所述的方法, 其中, 所述执行搜索的步骤包含:

当改写了所接收的搜索查询时, 使用所改写的搜索查询搜索文档储存库。

24、如权利要求 14 所述的方法, 还包含:

接收对该改写建议的选择;

改写所接收的搜索查询;

基于所改写的搜索查询执行搜索, 以获得结果; 以及
提供该结果以及到所接收的搜索查询的链接。

25、一种系统, 包含:

用于接收搜索查询的装置;

用于确定所接收的搜索查询是否包括实体名称的装置;

用于当所接收的搜索查询包括实体名称时、确定是否提供链接以便改写所接收的搜索查询的装置；

用于当确定应当提供链接时生成该链接的装置；

用于基于所接收的搜索查询执行搜索以获得搜索结果的装置；

用于提供该搜索结果的装置；以及

用于当应该提供链接时提供该链接的装置。

26、一种系统，包含：

存储器，用于存储与先前搜索有关的信息；以及

处理器，用于：

接收搜索查询；

确定所接收的搜索查询是否包括实体名称；

基于存储器中的信息，确定是否提供改写所接收的搜索查询的建议作为改写建议；

当确定应该提供改写建议时，生成该改写建议；

基于所接收的搜索查询执行搜索以获得搜索结果；

给出该搜索结果；以及

当应该提供改写建议时，提供该改写建议。

27、一种方法，包含：

获得实体名称；

标识对应于该实体名称的实体标识符；

基于该实体名字的变体和实体标识符产生候选字符串列表；

基于与先前搜索有关的信息评估每个候选字符串；

确定是否将候选字符串包括在表格中；以及

在基于搜索查询执行搜索之前，使用该表格来改写该搜索查询。

28、如权利要求 27 所述的方法，其中，所述产生候选字符串列表的步骤包含：

通过下列至少其中之一来转换该实体名称：删除修饰词、用连字符或者下划线替换空格、用空格替换连字符或者下划线、删除单引号、用“&”替换“and”、用“and”替换“&”、以及删除“and”与“&”。

29、如权利要求 27 所述的方法，其中，所述产生候选字符串列表的步骤包含：

通过下列至少其中之一来转换该实体标识符：用“&”替换“and”、用“and”替换“&”、删除“and”与“&”、删除“www.”、删除“.com”、以及将两边没有空格的句点视为空格或者删除该句点。

30、如权利要求 27 所述的方法，其中，所述确定是否将候选字符串包括在表格中的步骤包含：

标识与在涉及候选字符串中的一个候选字符串的先前搜索中被选择的文档相关联的实体标识符；

确定每个被标识的实体标识符的选择总数；以及

当与所述一个候选字符串相关联的实体标识符接收到的选择总数大于其他所标识的实体标识符时，确定应该将所述一个候选字符串包括在该表格中。

31、如权利要求 30 所述的方法，其中，所述确定是否将每个候选字符串包括在表格中的步骤还包含：

确定与所述一个候选字符串相关联的实体标识符的选择总数是否大于阈值；以及

当与所述一个候选字符串相关联的实体标识符的选择总数不大于该阈值时，确定所述一个候选字符串不应当被包括在该表格中。

32、如权利要求 27 所述的方法，其中，所述确定是否将候选字符串包括在表格中的步骤包含：

标识与在涉及候选字符串中的一个候选字符串的先前搜索中被选择的文档相关联的实体标识符；

确定每个被标识的实体标识符的选择总数的分布；以及

当所述分布指示与所述一个候选字符串相关联的实体标识符的选择总数与其它所标识的实体标识符的子集的选择总数相比被峰值化了时，确定应该将所述一个候选字符串包括在该表格中。

33、如权利要求 27 所述的方法，其中，所述使用该表格来改写搜索查询的步骤包含：

接收包括该实体名称的变体或者实体标识符的搜索查询；

确定该实体名称的变体或者该实体标识符是否被包括在该表格中；以及

当该实体名称的变体或者该实体标识符被包括在该表格中时，改写该搜索查询。

34、如权利要求 33 所述的方法，其中，所述改写该搜索查询的步骤包含：

修改该搜索查询，以包括与该实体名称相关联的限制标识符。

35、一种系统，包含：

用于获得实体名称的装置；

用于基于该实体名字的变体生成候选字符串列表的装置；

用于基于与先前搜索有关的信息评估每个候选字符串的装置；

用于确定是否将候选字符串包括在表格中的装置；以及

用于在基于搜索查询执行搜索之前，使用该表格来改写该搜索查询的装置。

36、一种系统，包含：

存储器，用于存储与先前搜索有关的信息；以及

处理器，用于：

标识实体名称；

生成该实体名称的变体；

基于与先前搜索有关的信息评估该实体名称的每个变体，以确定是否将该实体名称的变体包括到表格中；以及

在基于搜索查询执行搜索之前，使用该表格来改写该搜索查询。

37、一种方法，包含：

获得实体名称；

生成该实体名称的变体；

基于与先前搜索有关的信息评估该实体名称的每个变体；

确定是否将该实体名称的变体包括在表格中；

接收包括该实体名称的变体之一的搜索查询；
确定该实体名称的所述一个变体是否被包括在该表格中；以及
当该实体名称的所述一个变体被包括在该表格中时，为所接收的
搜索查询建议不同的搜索查询。

38、一种方法，包含：

接收搜索查询；
确定所接收的搜索查询是否包括实体名称；
基于与涉及该实体名称的先前搜索有关的信息，确定是否改写所
接收的搜索查询；

当确定应当改写所接收的搜索查询时，改写所接收的搜索查询；
当确定所接收的搜索查询不应当被改写时，基于与涉及该实体名
称的先前搜索有关的信息，确定是否提供改写所接收的搜索查询的建
议作为改写建议；

当确定应该提供改写建议时，生成该改写建议；
基于所接收的搜索查询和所改写的搜索查询之一执行搜索，以获
得搜索结果；

给出该搜索结果；以及
当确定应该提供改写建议时，给出该改写建议。

39、一种方法，包含：

接收搜索查询；
确定所接收的搜索查询是否包括商店名称；
基于与涉及该商店名称的先前搜索有关的信息，确定是否改写所
接收的搜索查询；

当确定应该改写所接收的搜索查询时，改写所接收的搜索查询，
以包括与该商店名称相关联的限制标识符；

基于所接收的搜索查询和所改写的搜索查询之一执行搜索，以获
得搜索结果；以及

给出该搜索结果。

40、一种方法，包含：

接收搜索查询;

确定所接收的搜索查询是否包括新闻来源名称;

基于与涉及该新闻来源名称的先前搜索有关的信息,确定是否改写所接收的搜索查询;

当确定应该改写所接收的搜索查询时,改写所接收的搜索查询,以包括与该新闻来源名称相关联的限制标识符;

基于所接收的搜索查询和所改写的搜索查询之一执行搜索,以获得搜索结果; 以及

给出该搜索结果。

具有实体检测的查询改写

技术领域

根据本发明原理的系统和方法通常涉及信息检索，并且尤其涉及基于对搜索查询中某些实体名称的检测而改写该搜索查询。

背景技术

万维网（“web”）包含巨量的信息。搜索引擎通过分类 web 文档来帮助用户定位这些信息的期望部分。通常，响应于用户的请求，搜索引擎返回到与该请求相关的文档的链接。

搜索引擎可以基于由用户提供的搜索项（被称为搜索查询）确定用户的兴趣。搜索引擎的目标是标识到基于该搜索查询的相关结果的链接。通常，搜索引擎通过将搜索查询中的项目（term）与预先存储的 web 文档的文集（corpus）进行匹配来实现它。包含用户搜索项的 web 文档被认为是“命中”，并且将其返回到用户。

一些搜索引擎通过在搜索查询中包括特殊字符或者项目、而允许用户将搜索限制到诸如与相同网站相关联的文档之类的相关文档集合。然而，时常地，用户忘记了包括这些特殊字符/项目或者不知道它们。

发明内容

根据按照本发明原理的一个方面，一种方法可以包括：接收搜索查询；确定所接收的搜索查询是否包括实体名称；基于与涉及该实体名称的先前搜索有关的信息、确定是否要改写所接收的搜索查询；以及当确定应当改写所接收的搜索查询时，改写所接收的搜索查询。该方法还可以包括：基于所接收的搜索查询和改写的搜索查询之一执行搜索以获得搜索结果，以及给出该搜索结果。

根据又一个方面，一个系统可以包括存储器和处理器。该存储器可以存储与先前搜索有关的信息。该处理器可以接收搜索查询，确定所接收的搜索查询是否包括实体名称，基于存储器中的信息、确定是否要提供改写所接收的搜索查询的建议作为改写建议，以及当确定应该提供改写建议时生成该改写建议。该处理器还可以基于所接收的搜索查询执行搜索以获得搜索结果，给出该搜索结果，并且当应该提供改写建议时提供该改写建议。

根据又一个方面，一个系统可以包括存储器和处理器。该存储器可以存储与先前搜索有关的信息。该处理器可以标识实体名称，生成该实体名称的变体，以及基于与先前搜索有关的信息评估该实体名称的每个变体、以确定在表格中是否包括该实体名称的变体。该处理器可以使用该表格来在基于搜索查询执行搜索之前改写该搜索查询。

根据进一步的方面，一种方法可以包括：获得实体名称，生成该实体名称的变体，基于与先前搜索有关的信息评估该实体名称的每个变体，以及确定在表格中是否包括该实体名称的变体。该方法还可以包括：接收包括该实体名称的变体之一的搜索查询，确定该实体名称的一个变体是否被包括在该表格中，以及当该实体名称的一个变体被包括在该表格中时，为所接收的搜索查询建议不同的搜索查询。

根据另一方面，一种方法可以包括：接收搜索查询，并且确定所接收的搜索查询是否包括实体名称。该方法还可以包括：基于与涉及该实体名称的先前搜索有关的信息、确定是否要改写所接收的搜索查询，以及当确定应当改写所接收的搜索查询时，改写所接收的搜索查询。该方法还可以包括：当确定所接收的搜索查询不应当被改写时，基于与涉及该实体名称的先前搜索有关的信息、确定是否要提供改写所接收的搜索查询的建议作为改写建议；以及当确定应该提供该改写建议时，生成该改写建议。该方法还可以包括：基于所接收的搜索查询和所改写的搜索查询之一执行搜索，以获得搜索结果；给出该搜索结果；以及当确定应该提供该改写建议时，给出该改写建议。

附图说明

并入这个说明书中并且构成其一部分的附图，举例说明了本发明的实施例，而且它连同该描述一起来阐述本发明。在附图中，

图 1 是其中可以实现根据本发明原理的系统和方法的示范性网络的图示；

图 2 是根据按照本发明原理的实现方式的、图 1 中的客户端和/或服务器的示范性图示；

图 3 是根据按照本发明原理的实现方式的、图 1 中的服务器的一部分的示范性功能框图；

图 4 是根据按照本发明原理的实现方式的、查询日志的示范性图示；

图 5 是根据按照本发明原理的实现方式的、候选字符串列表的示范性图示；

图 6 是根据按照本发明原理的实现方式的、图 3 中的查询改写表格的示范性图示；

图 7 是根据按照本发明原理的实现方式的、图 3 中的提议查询改写表格的示范性图示；

图 8 是根据按照本发明原理的实现方式的、用于生成查询改写表格和提议查询改写表格的示范性处理的流程图；

图 9 是根据按照本发明原理的实现方式的、有选择地改写查询的示范性处理的流程图；

图 10 和 11 是根据按照本发明的原理的实现方式的、在商店环境中的自动查询改写示例的图示；

图 12 - 14 是根据按照本发明原理的实现方式的、在商店环境中的查询改写建议示例的图示；

图 15 和 16 是根据按照本发明的原理的实现方式的、在新闻环境中的自动查询改写示例的图示；以及

图 17 - 19 是根据按照本发明原理的实现方式的、在新闻环境中的查询改写建议示例的图示。

具体实施方式

下面对本发明的详细说明需要引用附图。在不同附图中的相同参考数字可以标识相同或者类似的单元。此外，以下的详细说明不对本发明构成限制。

概述

根据本发明原理的系统和方法可以在检测到某些实体的名称时，改写搜索查询或者生成到改写的搜索查询的建议链接。此处使用的“实体”可以指任何可以被标记为与某些文档相关联的事物。实体的示例可以包括诸如在线商店之类的商店、新闻来源、产品类别、商标或者生产商、特定产品模型、状态（例如，新的、用过的、整修的等）、作者、艺术家、人、地点、和组织。

一些实体名称是明确的，并且唯一地标识特定实体。然而，大量的名称多少有一些是不明确或者通用的，这使得当将它们包括在用户的搜索查询中时，更加难以标识它们想要对应的实体。根据本发明原理的系统和方法提供了用于确定实体名称所对应的实体、以及当确定该名称对应于实体时改写用户的搜索查询或者提议改写用户的搜索查询的机制。因此，可以将用户的搜索查询限制为对与用户在该搜索中想要的实体相关联的文档（一个或多个）的搜索。

示范性网络配置

图1是其中可以实现根据本发明原理的系统和方法的网络100的示范性图示。网络100可以包括经由网络150连接到多个服务器120-140的多个客户端110。网络150可以包括局域网（LAN）、广域网（WAN）、诸如公共交换电话网（PSTN）之类的电话网络、企业内部网、Internet、存储器设备、另一类型的网络、或者网络的组合。为了简化起见，已经示出了有两个客户端110和三个服务器120-140连接到网络150。实际上，可以有更多或者更少的客户端和服务器。此外，在有些情况下，客户端可以执行服务器的功能，而服务器可以执行客户端的功能。

客户端 110 可以包括客户端部件。部件可以被定义为诸如无线电话、个人计算机、个人数字助理 (PDA)、膝上型或者另一类型的计算或者通讯设备之类的设备,在这些设备之一上运行的线程或者处理,和/或可由这些设备之一执行的对象。服务器 120 - 140 可以包括服务器部件,其以根据本发明原理的方式收集、处理、搜索、和/或维护文档。客户端 110 和服务器 120 - 140 可以经由有线、无线、和/或光学连接而连接到网络 150。

在根据本发明原理的实现方式中,服务器 120 可以包括可由客户端 110 使用的搜索引擎 125。服务器 120 可以在文档(例如,网页)的文集 (corpus) 上爬行 (crawl), 索引该文档, 并且在所爬行文档的储存库 (repository) 中存储与这些文档相关联的信息。服务器 130 和 140 可以存储或者维护可由服务器 120 爬行的文档。虽然服务器 120 - 140 被示为是单独的实体, 但是有可能让服务器 120 - 140 中的一个或者多个执行其它一个或者多个服务器 120 - 140 的一个或者多个功能。例如, 服务器 120 - 140 中的两个或多个有可能被实现为单个服务器。还有可能将服务器 120 - 140 中的单个服务器实现为两个或多个单独的 (并且有可能是分布式的) 设备。

此处使用的术语“文档”被广泛地解释为包括任何机器可读的和机器可存储的工作成果。文档可以包括电子邮件、网站、文件、文件的组合、具有到其它文件的嵌入式链接的一个或者多个文件、新闻组发布 (posting)、博客、网络广告等。在 Internet 的环境中, 常用的文档是网页。网页经常包括文本信息, 而且可以包括嵌入的信息 (诸如元信息、图像、超链接等) 和/或嵌入的指令 (诸如 Javascript (Java 脚本) 等)。

示范性客户端/服务器结构

图 2 是根据按照本发明原理的实现方式、可以对应于客户端 110 和服务器 120 - 140 中的一个或多个的客户端或者服务器部件 (在下文中被称为“客户端/服务器部件”) 的示范性框图。该客户端/服务器部件可以包括总线 210、处理器 220、主存储器 230、只读存储器 (ROM)

240、存储设备 250、输入设备 260、输出设备 270、和通信接口 280。总线 210 可以包括允许在该客户端/服务器部件的单元当中进行通信的导线集合。

处理器 220 可以包括用于解释并且执行指令的传统处理器或者微处理器。主存储器 230 可以包括随机存取存储器 (RAM) 或者另一类型的动态存储设备, 其存储信息以及由处理器 220 执行的指令。ROM 240 可以包括传统的 ROM 设备或者另一类型的静态存储设备, 其存储用于由处理器 220 使用的静态信息和指令。存储设备 250 可以包括磁和/或光记录介质以及其相应的驱动器。

输入设备 260 可以包括诸如键盘、鼠标、笔、语音识别和/或生物机构等之类的、允许操作者将信息输入到该客户端/服务器部件中的传统机构。输出设备 270 可以包括一个包括显示器、打印机、扬声器等在内的、将信息输出到操作者的传统机构。通信接口 280 可以包括任何类似于收发信机的机构, 其允许该客户端/服务器部件与其它设备和/或系统进行通信。例如, 通信接口 280 可以包括用于经由诸如网络 150 之类的网络与另一设备或者系统进行通信的机构。

如下面详细描述的那样, 根据本发明原理的客户端/服务器部件可以执行某些与搜索相关的操作。该客户端/服务器部件可以响应于处理器 220 执行包含在诸如存储器 230 之类的计算机可读介质中的软件指令, 而执行这些操作。计算机可读介质可被定义为物理或者逻辑存储器设备和/或载波。

软件指令可以从诸如数据存储设备 250 之类的另一计算机可读介质中、或者经由通信接口 280 从另一设备中读入到存储器 230 中。包含在存储器 230 中的软件指令可以导致处理器 220 执行稍后所述的处理。作为选择, 可以使用硬布线的电路来代替软件指令或者与软件指令相结合以实现根据本发明原理的处理。因此, 根据本发明原理的实现方式不局限于任何硬件电路和软件的特定组合。

示范性服务器

图 3 是根据按照本发明原理的实现方式的、服务器 120 的一部分

的示范性功能框图。根据一个实现方式，如下所述的一个或多个功能可以由搜索引擎 125 执行。根据另一实现方式，这些功能中的一个或多个可以由诸如与服务器 120 相关联的计算机或者服务器 130 和 140 之一之类的、在服务器 120 外部的部件执行。

服务器 120 可以包括连接到储存库的实体标识单元 310 和实体处理单元 320。储存库可以包括与先前例如由服务器 120 爬行和存储的文档相关联的信息。储存库还可以存储与关于先前执行的搜索的用户行为相关联的信息。例如，储存库可以存储一个查询日志，其描述了用户过去已经进行的搜索、他们选择（例如，通过点击来选择）的结果，并且还可能描述了与这个搜索处理相关的其它用户行为。

图 4 是根据按照本发明原理的实现方式的、可以存储在该储存库中的查询日志 400 的示范性框图。查询日志 400 可以包括多个与先前执行的搜索相关联的条目。查询日志 400 中的条目可以包括查询字段 410、选中实体标识符 (ID) 字段 420，并且可能还包括用户行为信息字段 430。查询字段 410 可以存储与包括在由用户输入的搜索查询中的项目（一个或多个）相关的信息。选中实体 ID 字段 420 可以存储诸如域（或者主机名或者统一资源定位符 (URL)）、类别标识符、产品标识符等之类的、可以唯一地标识特定实体的信息，该特定实体与由用户从使用该搜索查询执行的搜索的结果中选择的文档相关联。用户行为信息字段 430 可以存储关于用户对与该实体相关联的文档的访问的信息，诸如可以趋向于显示用户对该实体的兴趣的信息。因为用户可以选择超过一个的与搜索查询相关联的结果，所以相同的搜索查询可能在该查询日志 400 中的多个条目中出现（例如，查询 2）。

查询日志 400 中的条目的示例可以包括用于查询字段 410 的“harry potter barnes and noble”，和用于选中实体 ID 字段 420 的“barnesandnoble.com”。用户行为信息字段 430 可以存储与用户访问 barnesandnoble.com 域所花费的时间量相关的信息、是否出现了转化 (conversion)（即，是否出现了在 barnesandnoble.com 上的后续购买）、是否出现了部分转化（即，用户是否只是在 barnesandnoble.com

域上将一件物品增加到购物车中)、和/或其它可以趋向于示出用户对 **barnesandnoble.com** 域的兴趣的信息。

返回图 3, 实体标识单元 310 可以生成实体名称列表。实体标识单元 310 可以获得在特定环境中的实体的实体名称的初始集合(例如, 在商店环境中的商店名称或者在新闻来源环境中的新闻来源名称)。有很多种实体标识单元 310 可以获得特定环境中的实体名称的初始集合的方法。例如, 实体标识单元 310 可以通过分析文档文集等从在线目录、列表、组发布中获得实体名称。

对于这些名称中的每一个, 实体标识单元 310 还可以标识与该名称相关联的、诸如主页域名或者类别标识符之类的实体标识符。例如, 如果名称是 **Barnes & Noble**, 则相关联的实体标识符可以是 **barnesandnoble.com**。实体标识单元 310 可以根据例如对储存库中的文档信息的分析, 标识出相关联的实体标识符。

实体标识单元 310 然后可以处理该实体名称以生成该名称的变体列表。实体标识单元 310 可以向该名称和其标识符应用几个转换, 诸如: 按原样使用实体名称; 按原样使用实体标识符; 从实体名称中删除诸如“一”、“该”、“公司”、“股份有限公司”之类的修饰词; 在实体名称内用连字符或者下划线替换空格, 或者反过来用空格代替连字符或者下划线; 删除实体名称中的单引号; 在实体名称和/或标识符中互换“and”和“&”; 删除实体名称和/或标识符中的“and”和“&”; 删除标识符中的开头的“www.”和/或结尾的“.com”; 和/或将标识符中两边没有空格的句点视为空格或者删除这些句点。还可以使用其它或者不同的转换。

实体标识单元 310 可以将这些名称变体形成为候选字符串列表。图 5 是根据按照本发明原理的实现方式的、候选字符串列表 500 的示范性图示。候选字符串列表 500 可以包括与实体名称的各个版本相关联的许多条目(候选字符串)和它们的相关联的实体标识符。列表 500 中的条目可以包括实体名称字段 510 和实体 ID 字段 520。实体名称字段 510 可以包括实体名称的变体或者其相关联的实体标识符。实体 ID

字段 520 可以包括诸如域、URL、或者类别标识符之类的、唯一地标 识与实体名称字段 510 中的实体名称相对应的实体的信息。用于 Barnes & Noble 的条目的示例可以在实体名称字段 510 中包括“barnes & noble”，并且在实体 ID 字段 520 中包括“www.barnesandnoble.com”。

返回图 3，实体处理单元 320 可以处理该候选字符串列表以生成 两个表格：查询改写表格 330 和建议查询改写表格 340。如下面更详 细描述的那样，查询改写表格 330 可以指示是否应该自动地改写搜索 查询，而建议查询改写表格 340 可以指示是否应该建议改写查询。

实体处理单元 320 可以使用查询日志 400（图 4）评估每个候 选字符串（即，实体名称和/或实体标识符的每个版本）。例如，实体处 理单元 320 可以检查查询日志 400 以标识包括候选字符串的全部搜索 查询。实体处理单元 320 然后可以检查与所标识的查询相关联的点击。 例如，实体处理单元 320 可以与所标识的查询相关联地分析与由用户 （例如，通过点击）选择的文档相关联的实体标识符。

实体处理单元 320 然后可以分析点击的分布。例如，实体处理单 元 320 可以确定被最频繁点击的实体标识符是否为被认为对应于这个 候选字符串的实体标识符。如果不是的话，则实体处理单元 320 可以 忽略这个候选字符串，并且选择下一个候选字符串来评估。

如果最频繁点击的实体标识符是被认为对应于这个候选字符串 的实体标识符，则实体处理单元 320 可以确定到正确的、最频繁点击 的实体标识符的点击总数是否超过了一个小的阈值（例如，5）。使用 这个阈值确保了与一个用户点击一次的意外不同、这个候选字符串常 用到足以证明改写它是对的。如果点击总数低于该阈值，则实体处理 单元 320 可以忽略这个候选字符串，并且选择下一个候选字符串来评 估。

如果该正确的、最频繁点击的实体标识符具有高于该阈值的点击 总数，则实体处理单元 320 可以分析有多少点击与每个实体标识符相 关联的分布情况。如果在统计意义上该分布是足够“峰值化的

(**peaked**)”(例如,大多数点击是对正确实体标识符的点击,而非常少的点击是对其它实体标识符的点击),则实体处理单元 320 可以将该候选字符串添加到查询改写表格 330 中。相反,如果该分布不是峰值化的,而是下落得非常缓慢(例如,其它实体标识符获得几乎一样多的点击),则实体处理单元 320 可以将该候选字符串添加到建议查询改写表格 340 中。

可以使用许多可能的对峰度 (**peakedness**) 的检测,其中包括像分布(被视作多项式分布)熵值这样的信息理论度量。为了高效和简单起见,可以使用简单的测试,诸如,如果对正确的、最频繁点击的实体标识符的点击数目超过对次于最频繁点击的 N 个实体标识符(例如, $N=4$) 的点击数目组合,则可以认为它足够峰值化到可以添加到查询改写表格 330 中。

除了点击总数之外,还可以使用查询日志 400 中的、诸如用户行为信息字段 430 中的信息之类的其它数据来确定峰度。例如,与用户访问与该实体相关联的文档所花费的时间量相关的信息、是否出现全部或者部分转化、和/或可以趋向于示出用户对该实体的兴趣的其它信息可以是确定峰度的因素。趋向于指示期望结果的任何数据可以与连续的权重相关联。例如,可以一起使用这些数据中的多个数据,并且向它们分配权重(例如,转化与部分转化相比可以被更多地加权,而部分转化与点击相比可以被更多地加权)。作为选择,取决于用户访问与该实体相关联的文档所花费的时间量、或者与用户访问与该实体相关联的文档所花费的时间量相关的某些类别,可以将这些点击用可变的量加权。然后,峰度可以基于总体加权的度量。

这个技术的优点是它直接获取当用户提供搜索查询时用户想要的内容。例如, **Amazon.com** 是常用的商店,但是它也是南美洲的河流。如果当大多数用户将它包括在搜索查询中时他们仅仅意指商店,则将它添加到查询改写表格 330 中。如果显著一部分的用户意指河流或者别的东西,则他们将会点击其它结果,而且该分布将不会足够峰值化到可以将 **Amazon** 添加到查询改写表格 330 中。在这种情况下,

可以将 Amazon 添加到建议查询改写表格 340 中。

查询改写表格 330 可以存储与其名称明确的实体（即，相当确信它们的名称对应于该实体而不是别的东西，诸如“barnes and noble”）相关联的信息。图 6 是根据按照本发明原理的实现方式的查询改写表格 330 的示范性图示。查询改写表格 330 可以包括多个与名称字符串相关联的条目，如果这些名称字符串出现在用户搜索查询中，则可以自动地将它们改写为特定的实体标识符（例如，域、URL、或者类别标识符）。查询改写表格 330 中的条目可以包括字符串字段 610 和“改写为（rewrite to）”字段 620。字符串字段 610 可以存储如上所述的候选字符串。“改写为”字段 620 可以存储可以将该候选字符串改写成的实体标识符。例如，如果用户搜索查询包括“barnes & noble”，则基于查询改写表格 330，可以将该查询改写为“store:barnesandnoble”，以指示该搜索将只限于与商店 Barnes & Noble 相关联的实体标识符（域）。“store:”可以对应于商店环境中的限制标识符，其指示该搜索应该被限于遵循该限制标识符的商店。在其它环境中可以使用类似的限制标识符。

建议查询改写表格 330 可以存储与其名称不明确的实体相关联的信息（即，该名称可以包括可具有其它意思的通用字，诸如除了是指商店 Coach.com 之外还具有几个用法的“coach”）。图 7 是根据按照本发明原理的实现方式的、建议查询改写表格 340 的示范性图示。建议查询改写表格 340 可以包括多个与名称字符串相关联的条目，如果这些名称字符串出现在用户搜索查询中，则可以建议到特定实体标识符（例如，域、URL、或者类别标识符）的链接。建议查询改写表格 340 中的条目可以包括字符串字段 710 和“建议改写为（suggest rewrite to）”字段 720。字符串字段 710 可以存储如上所述的候选字符串。

“建议改写为”字段 720 可以存储可以为候选字符串建议到其的链接的实体标识符。例如，如果用户搜索查询包括“coach”，则基于建议查询生成表格 340，可以生成一个链接，以便将该搜索查询改写为“store:coach”，以指示该搜索将只限于与商店 Coach 相关联的实体标识符

(域)。

在一个实现方式中，可以在生成查询改写表格 330 和建议查询改写表格 340 后人工评估它们。这可以有助于将新的实体放入正确的表格中。此外，可以定期更新表格 330 和 340。当更新表格 330 和 340 时不同地加权用户行为（例如，点击）可能是所期望的，这是因为，如下面详细描述的那样，用户的行为可能受到自动查询改写的影响。

示范性处理

图 8 是根据按照本发明原理的实现方式的、用于生成表格 330 和 340 (图 3) 的示范性处理的流程图。处理可以从获得用于特定环境的实体名称列表开始 (动作 810)。对于每个实体名称，还可以标识出相应的实体标识符 (动作 820)。存在几种用于为该列表标识出实体名称和/或实体标识符的技术。例如，可以通过分析文档文集等，从在线目录、列表、组发布中标识出实体名称和/或实体标识符。

然后，可以通过转换实体名称和/或实体标识符生成候选字符串列表 (动作 830)。例如，用于特定实体名称及其相关联的实体标识符的候选字符串列表可以包括：按照原样的实体名称，按照原样的实体标识符，没有修饰词 (例如，“一”、“该”、“股份有限公司”、“公司”) 的实体名称，用连字符或者下划线代替空格以及反过来用空格代替连字符或者下划线的实体名称，没有单引号的实体名称，用“and”代替“&”以及反过来用“&”代替“and”的实体名称和/或实体标识符，没有“and”和“&”的实体名称和/或实体标识符，没有开头的“www.”和/或结尾的“.com”的实体标识符，以及用空格代替在两边都没有空格的句点或者删除了该句点的实体标识符。还可以使用其它或者不同的转换。

然后可以评估候选字符串 (即，实体名称和/或实体标识符的每个版本)，并且将其用于形成查询改写表格 330 和建议查询改写表格 340 (动作 840 和 850)。例如，可以搜索查询日志 400，以标识出所有在它们的查询字段 410 中包括特定候选字符串的条目。然后，可以检查所标识出的条目，以确定为相应的搜索查询点击了哪些实体标识符 (例如，域、URL、或者类别标识符)。可以分析该点击的分布，

以便例如确定被最频繁点击的实体标识符是否是被认为对应于这个候选字符串的实体标识符。如果不是的话，则可以忽略该候选字符串，并且选择下一个候选字符串用于评估。

如果被最频繁点击的实体标识符是被认为对应于这个候选字符串的实体标识符，则可以确定对正确的、最频繁点击的实体标识符的点击总数是否超过了一个小的阈值（例如，5）。如果点击总数低于该阈值，则可以忽略这个候选字符串，并且选择下一个候选字符串来评估。

如果该正确的、最频繁点击的实体标识符具有高于该阈值的点击总数，则可以确定有多少点击与每个实体标识符相关联的分布情况。如果该分布是足够“峰值化的”（例如，大多数点击是对正确实体标识符的点击，而比较少的点击是对其它实体标识符的点击），则可以将该候选字符串添加到查询改写表格 330 中。如果该分布不是峰值化的，而是下落得非常缓慢（例如，其它实体标识符获得几乎一样多的点击），则可以将该候选字符串添加到建议查询改写表格 340 中。上面已经描述了用于确定峰度的一些可能测试。

图 9 是根据按照本发明原理的实现方式的、用于有选择地改写搜索查询的示范性处理的流程图。处理可以从接收来自用户的搜索查询开始（动作 910）。该搜索查询可包含一个或多个项目，这些项目可以包括或者可以不包括实体名称。

可以使用查询改写表格 330 和/或建议查询改写表格 340 评估该搜索查询（动作 920）。例如，可以搜索查询改写表格 330 中的条目，以确定该搜索查询中的项目（或者项目的组合）与字符串字段 610（图 6）中的名称字符串相匹配。如果搜索查询中的一个项目（或者多个项目）与名称字符串之一相匹配，则可以自动地改写该查询（动作 930 和 940）。例如，如果原有的查询包括“harry potter barnes & noble”，则可以将该查询改写为“harry potter store: barnesandnoble”，以指示用户想要在 Barnes & Noble 网站上查找有关 Harry Potter 的产品或者信息，而不是简单地搜索包含项目“harry”、“potter”、“barnes”、“&”、

和“noble”的文档。

可选地，可以生成查询链接以向用户提供原有查询，用于询问用户是否想要不同于自动改写的查询的搜索查询（动作 950）。在这种情况下，该查询链接可以向用户提供原有的查询“harry potter barnes & noble”。

如果该搜索查询中没有项目与查询改写表格 330 中的名称字符串相匹配，则可以搜索建议查询改写表格 340 中的条目，以确定该查询中的项目（或者项目的组合）是否与字符串字段 710（图 7）中的名称字符串相匹配。如果搜索查询中的一个项目（或者多个项目）与名称字符串之一相匹配，则可以生成查询链接（动作 960 和 970）。该查询链接可以向用户提供查询改写，以询问用户是否想要不同的搜索查询。例如，如果原有的查询包括“purse coach”，则该查询链接可以提供改写的查询“purse store: coach”（在商店环境中），以指示该用户想要在 Coach 网站上查找有关 purse 的产品或者信息，而不是仅仅搜索包含项目“purse”和“coach”的文档。

然后，如果可应用改写的查询，则基于改写的查询执行搜索，或者如果可应用原有查询，则基于原有查询执行搜索，以标识出与改写/原有的查询相关的文档。例如，可以搜索文档储存库以标识出包括该查询中的一个或多个项目的文档。所获得的文档可以形成可向用户给出的搜索结果（动作 980）。在一个实现方式中，该搜索结果可以采用到文档的链接的形式。

自动查询改写示例 - 商店环境

图 10 和 11 是根据按照本发明的原理的实现方式的、在商店环境中的自动查询改写示例的图示。如图 10 所示，用户可以经由与诸如搜索引擎 125（图 1）之类的搜索引擎相关联的图形用户接口输入搜索查询。在这个示例中，用户输入搜索查询“harry potter amazon”。假定项目“amazon”清楚地标识了商店 Amazon.com，并因此作为名称字符串被包括在查询改写表格 330（图 3）中。

搜索引擎 125 可以基于查询改写表格 330 和建议查询改写表格

340 评估该搜索查询中的项目。在这种情况下，搜索引擎 125 发现项目“amazon”出现在查询改写表格 330 中。如图 11 所示，搜索引擎 125 然后将该搜索查询改写为“harry potter store: amazon”。

搜索引擎 125 在储存库中搜索与改写的查询相关的、与商店 amazon.com 相关联的文档（例如，产品）。有许多种确定文档相关性的方法。例如，可以将包含该改写的查询中的一个或多个搜索项的文档标识为是相关的。包括更多数量的搜索项的文档可以被标识为比包含较少数量的搜索项的文档更相关。

搜索引擎 125 然后将相关的文档作为搜索结果向用户给出。如图 11 所示，每个搜索结果可以包括到相应文档的链接 1110、与该文档相关联的价格和商店标识符 1120、以及相应文档的简短描述 1130。如果可用的话，该搜索结果还可以包括与该文档相关联的图像 1140。可选地，搜索引擎 125 还可以提供到由用户输入的原有查询的链接 1150。在这种情况下，链接 1150 可以对应于与对搜索项“harry”、搜索项“potter”、和/或搜索项“amazon”的搜索相关联的查询。

建议查询改写示例 - 商店环境

图 12 - 14 是根据按照本发明原理的实现方式的、在商店环境中的查询改写建议示例的图示。如图 12 所示，用户可以经由与诸如搜索引擎 125(图 1)之类的搜索引擎相关联的图形用户接口输入搜索查询。在这个示例中，用户输入搜索查询“purse coach”。假定项目“coach”是通用项目，并且不明确地标识商店 Coach.com。进一步假定“coach”作为名称字符串被包括在建议查询改写表格 340（图 3）中。

搜索引擎 125 可以基于查询改写表格 330 和建议查询改写表格 340 评估该搜索查询中的项目。在这种情况下，搜索引擎 125 发现搜索项都没有出现在查询改写表格 330 中，但是项目“coach”出现在建议查询改写表格 340 中。如图 13 所示，搜索引擎 125 然后将该搜索查询改写为“purse store: coach”，并且生成到所改写的查询的链接 1310（“Search Merchant Coach for Purse（在商店 Coach 中搜索 Purse）”）。

搜索引擎 125 在储存库中搜索与原有搜索查询相关的文档(例如, 产品)。如上所述, 有许多种确定文档相关性的方法。例如, 可以将包含该原有查询中的一个或多个搜索项的文档标识为是相关的。包括更多数量的搜索项的文档可以被标识为比包含较少数量的搜索项的文档更相关。在这种情况下, 搜索引擎 125 搜索包括搜索项“purse”和/或搜索项“coach”的文档。

搜索引擎 125 然后将相关的文档作为搜索结果向用户给出。如图 13 所示, 每个搜索结果可以包括到相应文档的链接 1320、与该文档相关联的价格和商店标识符 1330、以及相应文档的简短描述 1340。因为该搜索不局限于商店 Coach, 所以搜索结果与多个不同的商店(例如, Yahoo! Auctions 和 Vendio)相关联。该搜索结果还可以包括链接 1350, 其是到与由商店标识符 1330 所标识的商店相关联的附加文档的链接。如果可用的话, 该搜索结果还可以包括与该文档相关联的图像 1360。

如果用户选择了与改写的查询相关联的链接 1310, 则搜索引擎 125 在储存库中搜索与该改写的查询相关的、与商店 Coach.com 相关联的文档(例如, 产品)。搜索引擎 125 然后将相关的文档作为搜索结果向用户给出。如图 14 所示, 每个搜索结果可以包括到相应文档的链接 1410、与该文档相关联的价格和商店标识符 1420、以及相应文档的简短描述 1430。如果可用的话, 该搜索结果还可以包括与该文档相关联的图像 1440。可选地, 搜索引擎 125 还可以提供到由用户输入的原有查询的链接 1450。在这种情况下, 链接 1450 可以对应于与对搜索项“purse”和/或搜索项“coach”的搜索相关联的查询。

自动查询改写示例 - 新闻环境

图 15 和 16 是根据按照本发明的原理的实现方式的、在新闻环境中的自动查询改写示例的图示。如图 15 所示, 用户可以经由与诸如搜索引擎 125(图 1)之类的搜索引擎相关联的图形用户接口输入搜索查询。在这个示例中, 用户输入搜索查询“george bush washingtonpost”。假定项目“washingtonpost”明确地标识新闻来源 washingtonpost.com,

并因此作为名称字符串被包括在查询改写表格 330 (图 3) 中。

搜索引擎 125 可以基于查询改写表格 330 和建议查询改写表格 340 评估该搜索查询中的项目。在这种情况下, 搜索引擎 125 发现项目“washingtonpost”出现在查询改写表格 330 中。如图 16 所示, 搜索引擎 125 然后将该搜索查询改写为“george bush source: washington_post”。

搜索引擎 125 在储存库中搜索与改写的查询相关的、与来源 washingtonpost.com 相关联的文档(例如, 新闻文档)。有许多种确定文档相关性的方法。例如, 可以将包含该改写的查询中的一个或多个搜索项的文档标识为是相关的。包括更多数量的搜索项的文档可以被标识为比包含较少数量的搜索项的文档更相关。

搜索引擎 125 然后将相关的文档作为搜索结果向用户给出。如图 16 所示, 每个搜索结果可以包括到相应文档的链接 1610、新闻来源标识符以及文档创建时间的指示符 1620、以及相应文档的简短描述 1630。该搜索结果还可以包括链接 1640, 其是到和与链接 1610 相关联的文档的主题相关的其它文档的链接。可选地, 搜索引擎 125 还可以提供到由用户输入的原有查询的链接 1650。在这种情况下, 链接 1650 可以对应于与对搜索项“george”、搜索项“bush”、和/或搜索项“washingtonpost”的搜索相关联的查询。

提议查询改写示例 - 新闻环境

图 17 - 19 是根据按照本发明原理的实现方式的、在新闻环境中的查询改写建议示例的图示。如图 17 所示, 用户可以经由与诸如搜索引擎 125(图 1)之类的搜索引擎相关联的图形用户接口输入搜索查询。在这个示例中, 用户输入搜索查询“mutual funds business week”。假定项目“business week”是通用项目, 并且不明确地标识新闻来源 businessweek.com。进一步假定“business week”作为名称字符串被包括在建议查询改写表格 340 (图 3) 中。

搜索引擎 125 可以基于查询改写表格 330 和建议查询改写表格 340 评估该搜索查询中的项目。在这种情况下, 搜索引擎 125 发现搜

索项都没有出现在查询改写表格 330 中，但是项目“business week”出现在建议查询改写表格 340 中。如图 18 所示，搜索引擎 125 然后将该搜索查询改写为“mutual funds source: businessweek”，并且生成到所改写的查询的链接 1810（“Search News Source Business Week for Mutual Funds(在新闻来源 Business Week 中搜索 Mutual Funds)”）。

搜索引擎 125 在储存库中搜索与原有搜索查询相关的文档(例如，新闻文档)。如上所述，有许多确定文档相关性的方法。例如，可以将包含该改写的查询中的一个或多个搜索项的文档标识为是相关的。包括更多数量的搜索项的文档可以被标识为比包含较少数量的搜索项的文档更相关。在这种情况下，搜索引擎 125 搜索包括搜索项“mutual”、“funds”、“business”和/或“week”的文档。

搜索引擎 125 然后将相关的文档作为搜索结果向用户给出。如图 18 所示，每个搜索结果可以包括到相应文档的链接 1820、新闻来源标识符以及文档创建时间的指示符 1830、以及相应文档的简短描述 1840。因为该搜索不局限于新闻来源 Business Week (商业周刊)，所以搜索结果与多个不同的新闻来源(例如，San Antonio Express (圣安东尼奥快报)、Baltimore Sun (巴尔的摩太阳报)、和 Arizona Business Gazette (亚利桑那州商业公报))相关联。

如果用户选择了与改写的查询相关联的链接 1810，则搜索引擎 125 在储存库中搜索与该改写的查询相关的、与新闻来源 businessweek.com 相关联的文档(例如，新闻文档)。搜索引擎 125 然后将相关的文档作为搜索结果向用户给出。如图 19 所示，每个搜索结果可以包括到相应文档的链接 1910、新闻来源标识符以及与该文档创建日期相对应的日期指示符 1920、以及相应文档的简短描述 1930。可选地，搜索引擎 125 还可以提供到由用户输入的原有查询的链接 1950。在这种情况下，链接 1950 可以对应于与对搜索项“mutual”、搜索项“funds”、搜索项“business”、和/或搜索项“week”的搜索相关联的查询。

结论

根据本发明原理的系统和方法可以在检测到某些实体的名称时，

有选择地改写搜索查询。

对本发明优选实施例的上述描述提供了举例说明和描述，但是不是穷举的，并且也不打算将本发明限制为所公开的准确形式。根据上述示教，许多修改和变化是可能的，或者可以从本发明的实践中获取这些修改和变化。

例如，已经描述了生成查询改写表格 330 和建议查询改写表格 340，然后使用它们用于帮助自动改写搜索查询或者生成到改写的搜索查询的建议链接。在另一实现方式中，可以在接收到来自用户的搜索查询时即时地执行可生成表格 330 和 340 的分析。在这种情况下，可以使用搜索查询中的项目（一个或多个）来搜索查询日志 400（图 4）。然后，可以用与如上所述的方式类似的方式确定点击和峰度的分布。可以使用与上述分析类似的、用于在查询改写表格 330 或者建议查询改写表格 340 中存储搜索项（一个或多个）的分析，来确定是否自动地改写该查询或者生成到改写的查询的建议链接。

此外，可以使用不同的技术来确定怎样以及是否改写搜索查询中的项目。一种技术是手动创建表格 330 和 340。另一类常用技术是基于大量文本文集（有可能是被搜索的文档的储存库）进行对任何潜在名称的统计分析，并且基于该名称出现于其中的环境的易变性，或者基于该名称在已知适当字段（例如，商店名称字段）而不是在其它更通用字段的一般文本（诸如文章的描述或者正文）中出现的相对次数，来确定该名称是否为通用项目。这些方法的组合也是可能的，诸如，使用自动化方法，并继之以人工评估和少量编辑所生成的表格条目。

此外，虽然已经就图 8 和 9 描述了一系列动作，但是在根据本发明原理的其它实现方式中，可以修改这些动作的次序。此外，不相关的动作可以并行执行。

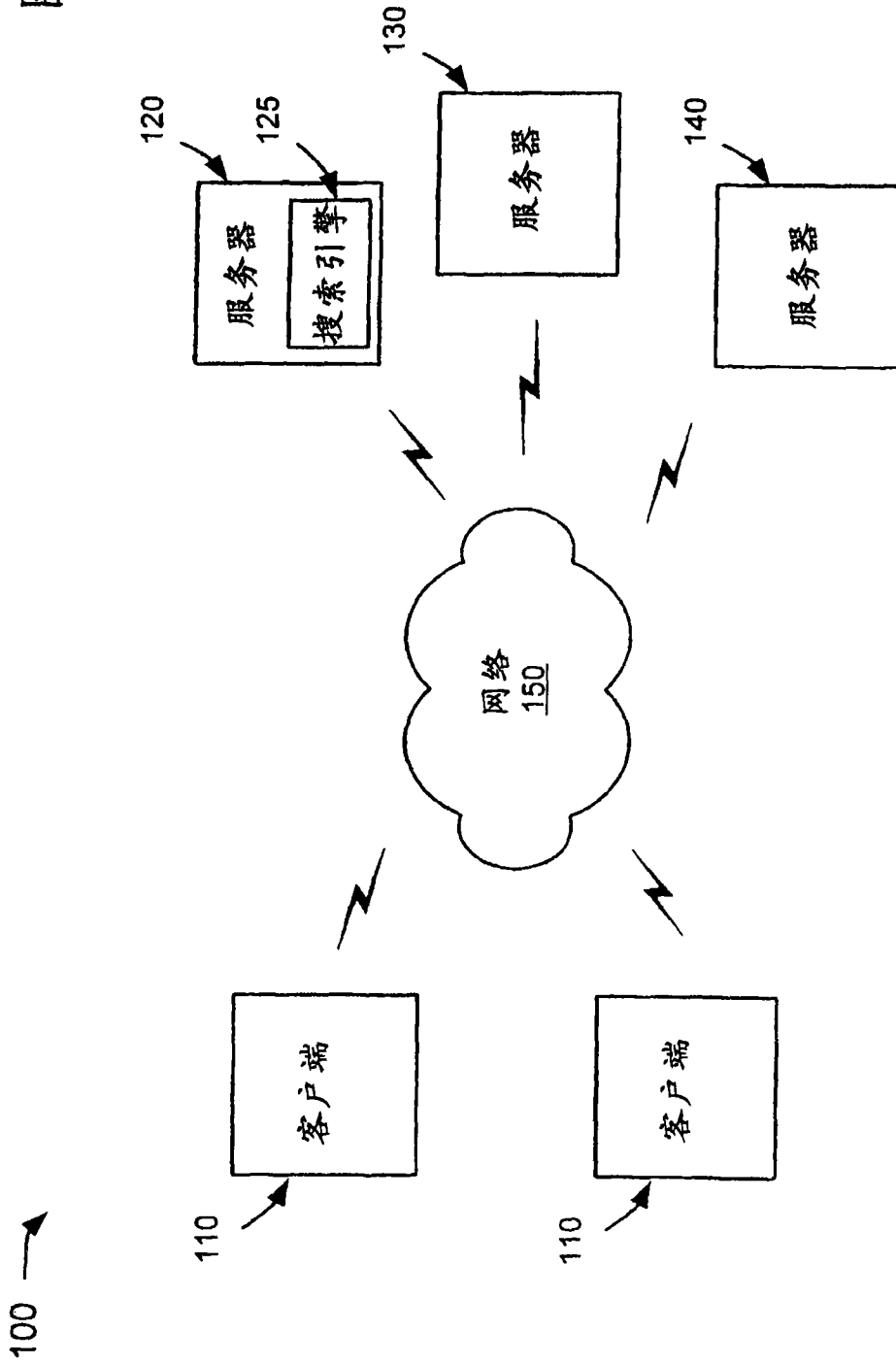
在一个实现方式中，服务器 120 可以执行大多数（如果不是全部的话）的就图 8 和/或 9 中的处理所描述的动作。在按照本发明原理的另一个实现方式中，这些动作中的一个或者多个、或者其全部都可以由诸如另一个服务器 130 和/或 140 或者客户端 110 之类的另一部件所

执行。

本领域的普通技术人员还应当明白，在附图所说明的实现方式中，如上所述的本发明的各个方面可以用软件、固件、和硬件的许多不同形式实现。用于实现根据本发明原理的各个方面的实际软件代码或者专用控制硬件不是对本发明的限制。因此，在没有参考特定的软件代码的情况下描述了各个方面的操作和动作——应当理解，基于此处的描述，本领域的普通技术人员将能够设计出软件和控制硬件以实现这些方面。

除非被明确描述为这样，否则在当前申请中所使用的要素、动作、或者指令都不应该被看作是本发明关键和必要的。此外，此处使用的冠词“一”意图包括一或多项。在意图表示仅仅一项的地方，使用了术语“一个”或者类似的语言。此外，除非明确申明不是这样，否则短语“基于”意指“至少部分地基于”。

图 1



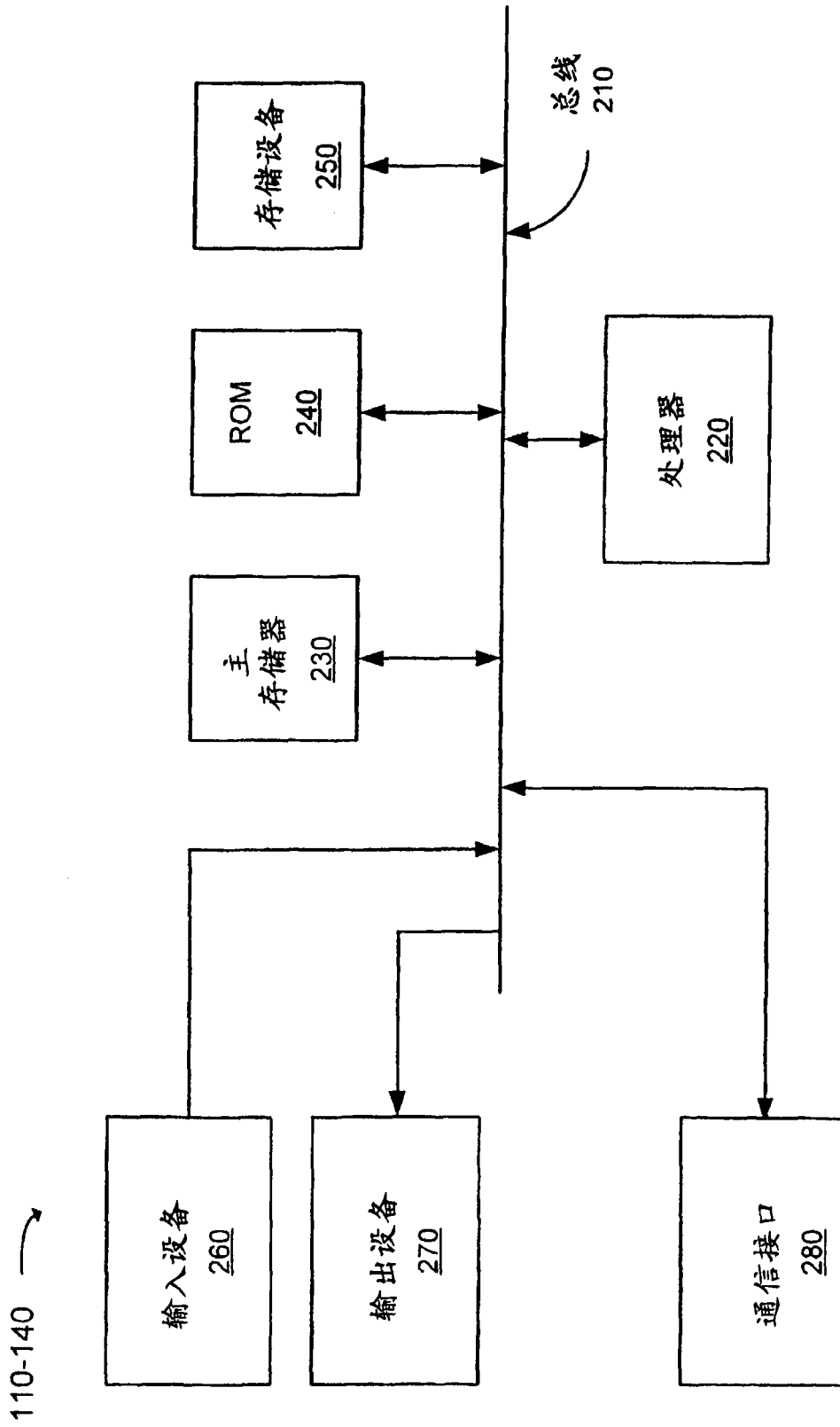
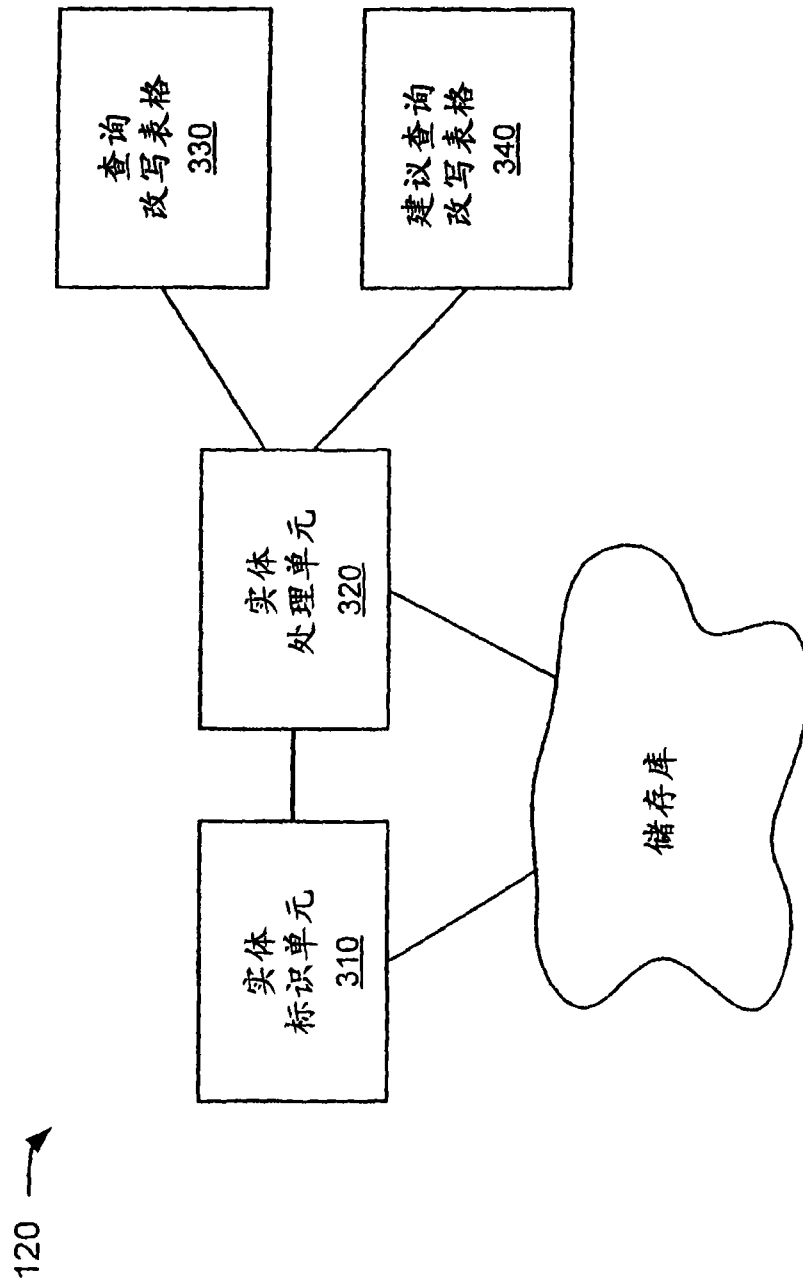


图 2

图 3



400 →	410	查询	选中实体ID	用户行为信息
		查询 1	实体 ID 12	行为信息 1
		查询 2	实体 ID 3	行为信息 2
		查询 2	实体 ID 87	行为信息 3
		查询 2	实体 ID 123	行为信息 4
		查询 2	实体 ID 47	行为信息 5
		查询 3	实体 ID 89	行为信息 6
		查询 3	实体 ID 873	行为信息 7
		HARRY POTTER BARNES AND NOBLE	BARNES&NOBLE.COM	有关访问、 兴趣等的信息
			● ● ●	

图 4

500 →

510 实体名称	520 实体 ID
名字1版本1	实体 ID 1
名字1版本2	实体 ID 1
名字1版本3	实体 ID 1
名字2版本1	实体 ID 2
BARNES&NOBLE	WWW.BARNESANDNOBLE.COM
● ● ●	

图 5

330 →

字符串	改写为
名称字符串 1	实体 ID 1
名称字符串 2	实体 ID 1
名称字符串 3	实体 ID 1
名称字符串 4	实体 ID 2
BARNES & NOBLE	BARNESANDNOBLE.COM
● ● ●	

图 6

340 →

字符串	建议改写为
名称字符串 1	实体 ID 1
名称字符串 2	实体 ID 1
名称字符串 3	实体 ID 1
名称字符串 4	实体 ID 2
COACH	COACH.COM
● ● ●	

图 7

图 8

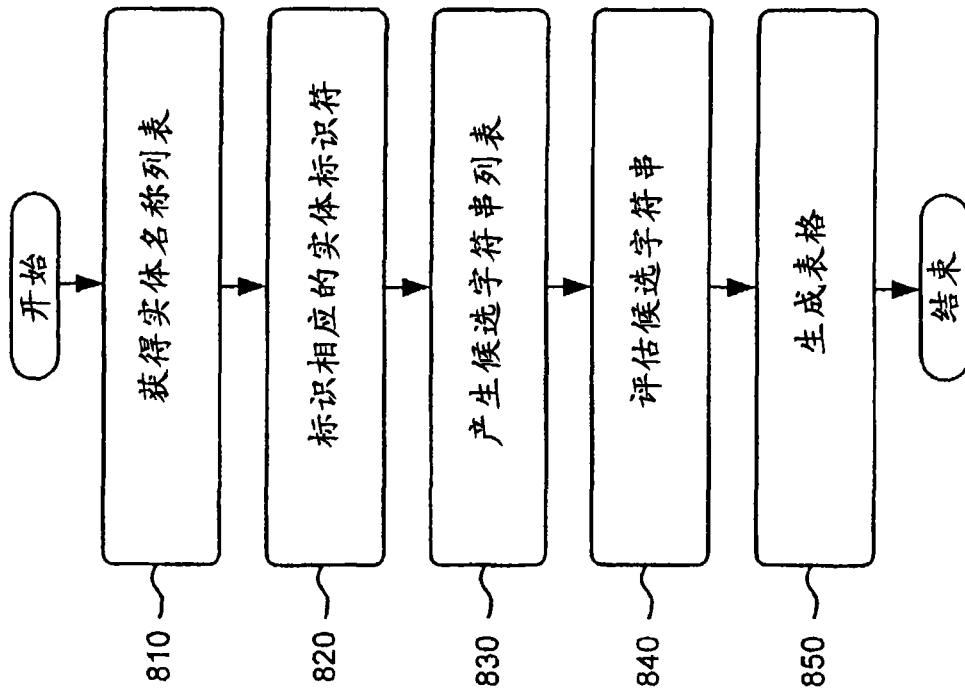
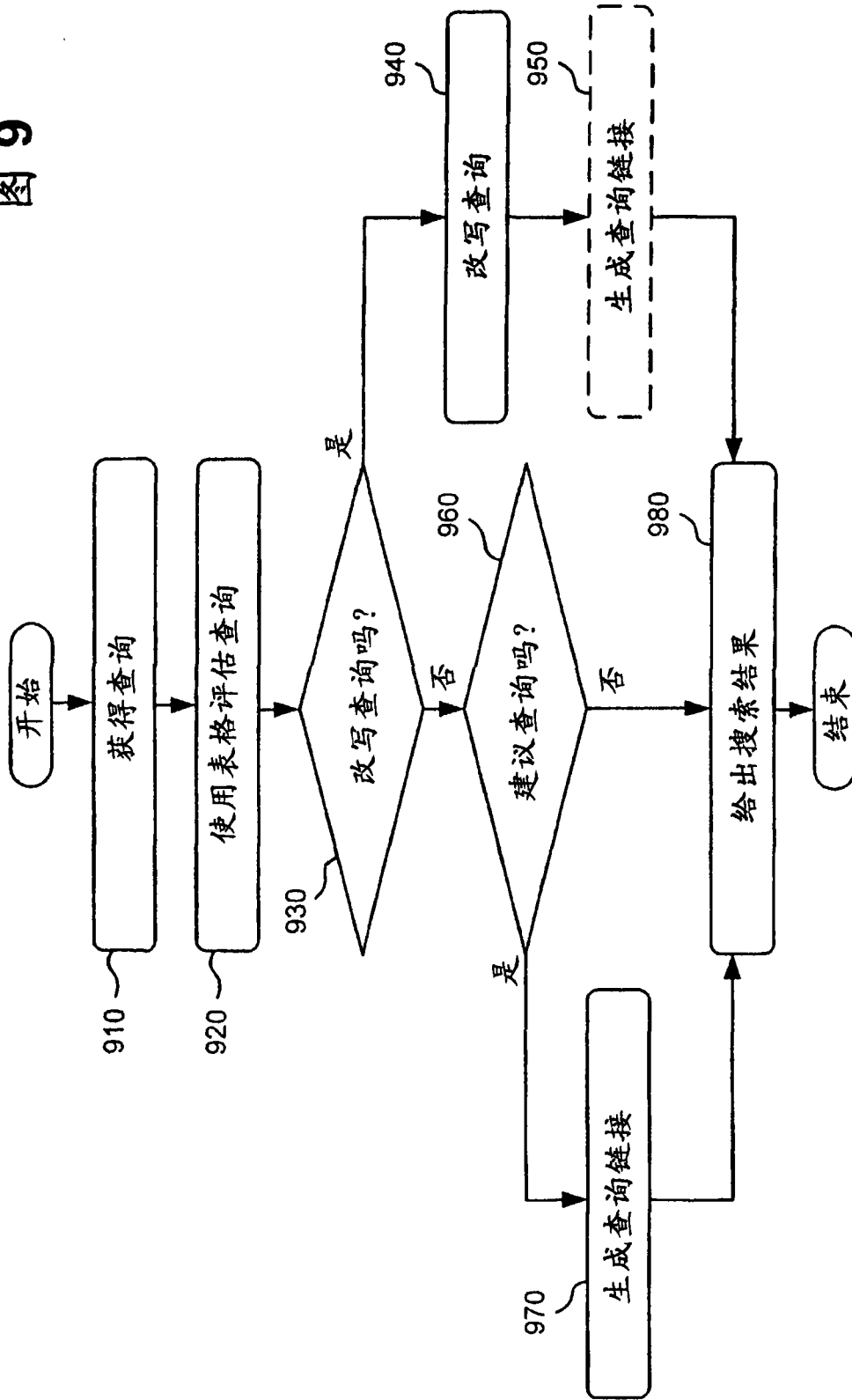


图 9



Froogle™
BETA

harry potter amazon |

SEARCH FROOGLE

图 10

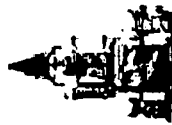
Froogle[™] BETA

harry potter store:amazon

SEARCH FROOGLE

SEARCHED PRODUCTS FOR HARRY POTTER AT AMAZON.COM

SEARCH PRODUCTS FOR HARRY POTTER AMAZON ~ 1150



1140

LEGO Harry Potter: Dumbledore's Office (4729)

\$39.98 - Amazon.com - Blocks & Construction
... Connect Dumbledore's Office to your Hogwarts
Castle set for the complete Harry Potter experience!
Includes Professor Dumbledore, Professor ...

1110

1120

1130



LEGO Harry Potter Escape from Private Drive

\$18.74 - Amazon.com - Blocks & Construction
... Attach the string on the rear of the car to the window
and pull it free so Harry can escape. Includes Harry
Potter, Ron Weasley, and Uncle Vernon Dursley.

图 11

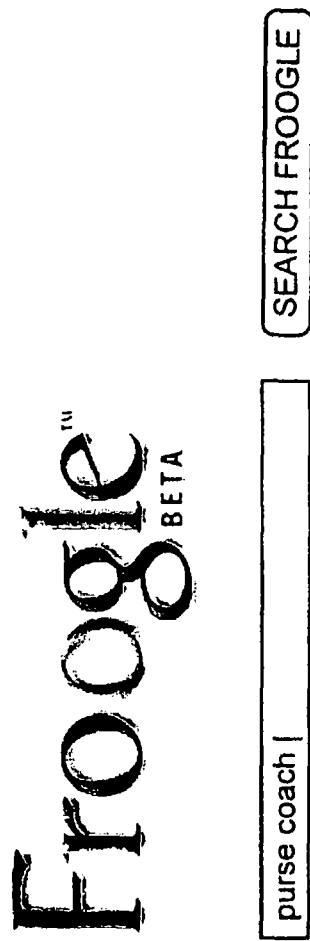



图 12

Froogle[™] BETA

SEARCHED PRODUCTS FOR PURSE COACH

SEARCH MERCHANT COACH FOR PURSE  1310



1140

[Coach Mini Signature Bucket Tote Purse Khaki](#)

\$44.99 - Yahoo! Auctions - Women's Accessories

Brand New COACH mini Signature Bucket Tote Purse in Khaki color. The size of this gorgeous purse is 10.5" x 8" x 5", and the adjustable ...

[\[More from Yahoo! Auctions\]](#)  1350



1320



1330



1340



[Coach black signature demi purse](#)

\$50.00 - Vendio - Women's Accessories

This is a black Coach signature demi purse. There are no guarantees on this item -- I never went and checked to see if it really ...

[\[More from Vendio\]](#)

图 13

Froogle

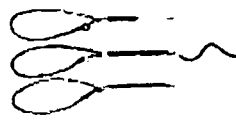
BETA

purse store:coach

SEARCH FROOGLE

SEARCHED PRODUCTS FOR PURSE AT COACH

SEARCH PRODUCTS FOR PURSE.COACH ~ 1450

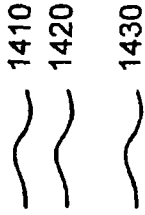


1440

Coach Purse Pen

\$38.00 - Coach - Women's Accessories

This colorful enamel pen comes with a leather strap that can be clipped into a handbag. 4" enamel pen. 10 1/4" calfskin wrist strap with dogleash closure.



1410

1420

1430

1440



COACH Women's Wallet Polished Calfskin Elongated French Purse

\$198.00 - Coach - Women's Accessories

A chic addition to your handbag. Ten credit card pockets. Three multi-function pockets. Outside gusseted zip coin compartment. Double bill compartments.

图 14

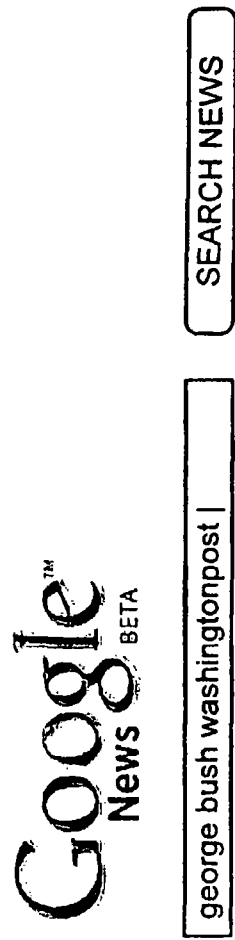


图 15



george bush source:washington_post

SEARCH NEWS

SEARCHED NEWS SOURCE WASHINGTON POST FOR GEORGE BUSH

SEARCH NEWS GEORGE BUSH WASHINGTONPOST — 1650

Bush: Hussein Eluded Containment Efforts

Washington Post, DC - 10 hours ago

In defending his decision to go to war in Iraq, President Bush suggested yesterday a belief that UN inspections and sanctions were of limited utility in . . .

1610

1620

1630

Bush Offers Shifting Rationale for Iraq War - Washington Post

President Revises Rationale for War - Washington Post

and more >>

1640

Bush Offers Defense of Iraq War Rationale

Washington Post, DC - 23 hours ago

. . . Bush said he decided to go to war based on the intelligence he had at hand about Saddam, but said CIA Director George Tenet's job is not in jeopardy. . . .

Bush, Aides Ignored CIA Caveats - Washington Post

Imminent Danger at the CIA - Washington Post

and more >>

图 16

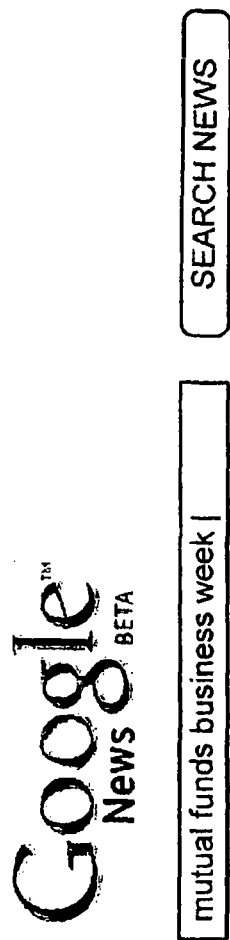


图 17

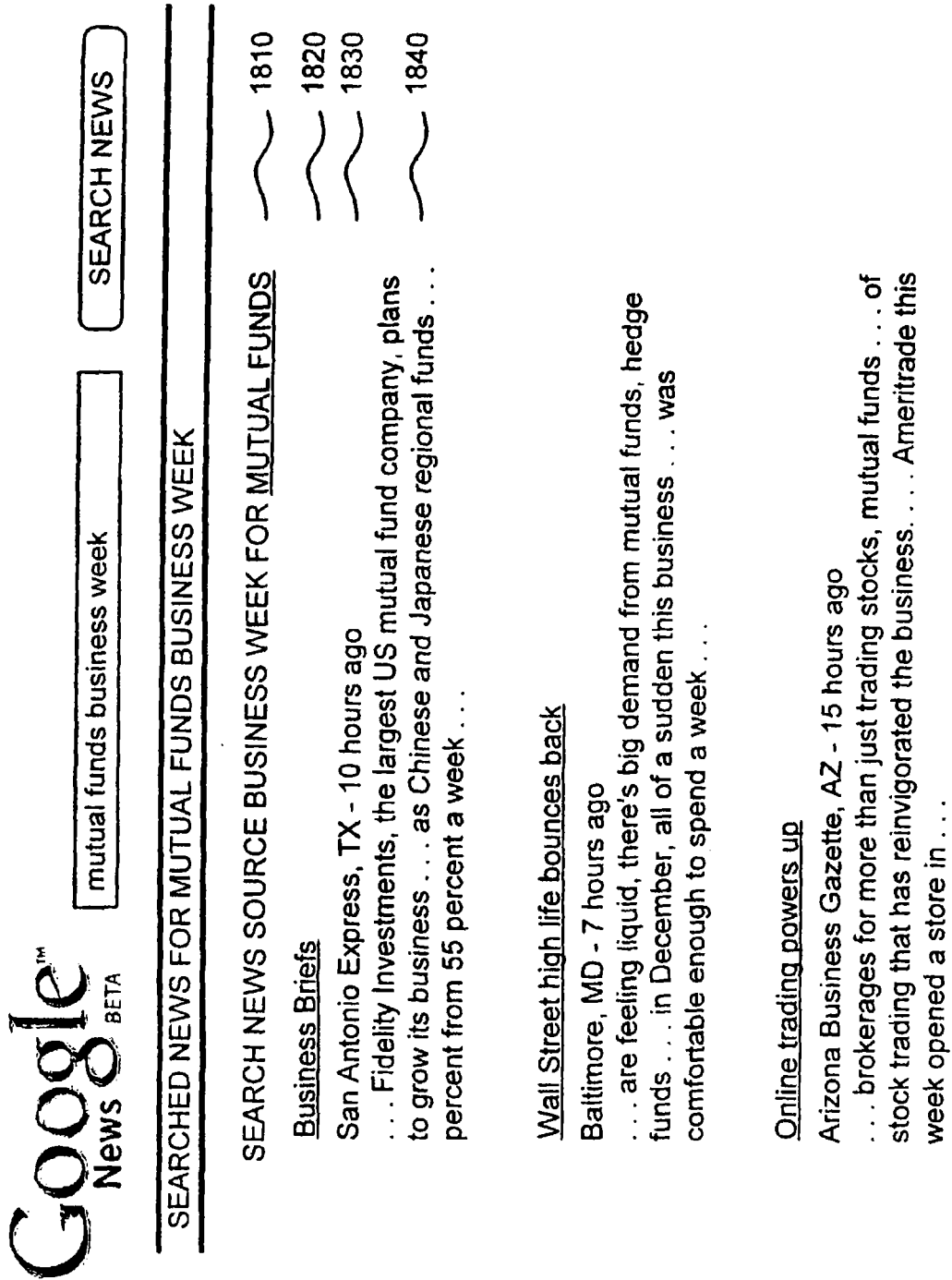


图 18



mutual funds source:businessweek

SEARCH NEWS

SEARCHED NEWS SOURCE BUSINESS WEEK FOR MUTUAL FUNDS

SEARCH NEWS FOR MUTUAL FUNDS BUSINESS WEEK ~ 1950

Dynasty In Distress

BusinessWeek - Feb 1, 2004

... by New York Attorney General Eliot Spitzer for "fraudulent" late trading and market timing of mutual funds. He continues to meet with Spitzer, says a source ...

1910
1920
1930

Rescuing Reuters

BusinessWeek - Feb 1, 2004

... capitalization, then moved on into the stock positions of hedge and mutual funds. ... that their product and service are compelling," says a Goldman source. ...

图 19