(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 102687138 B
(45)授权公告日 2015.09.16

(21)申请号 201080059977.0
(22)申请日 2010.12.17
(30)优先权数据
12/648,220 2009.12.28 US
(85)PCT国际申请进入国家阶段日
2012.06.28
(86)PCT国际申请的申请数据
PCT/US2010/061120 2010.12.17
(87)PCT国际申请的公布数据
WO2011/090638 EN 2011.07.28
(73)专利权人 雅虎公司
地址 美国加利福尼亚州
(72)发明人 吉拉德·米思尼 埃帕·詹恩
(74)专利代理机构 北京东方亿思知识产权代理有限责任公司 11258
代理人 宋鹤
(51)Int.Cl.
G06F 17/26(2006.01)
G06F 17/30(2006.01)
G06F 17/10(2006.01)
G06F 15/16(2006.01)
(56)对比文件
US 2009/0171929 A1,2009.07.02,说明书第2页第[0028]段-第7页第[0105]段,图1-10.
CN 1609852 A,2005.04.27,全文.
CN 1684072 A,2005.10.19,全文.
审查员 康厚萍

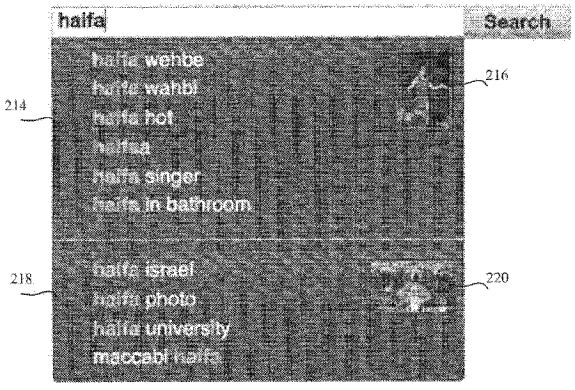
权利要求书4页 说明书14页 附图6页

(54)发明名称

搜索建议聚类和呈现

(57)摘要

本发明公开了用于聚类和呈现搜索建议的方法和装置。通过用户界面的搜索查询区获得文本段,所述文本段是搜索查询的一部分。获得建议集合,该建议集合中的每个建议是与所述文本段相关的所建议的搜索查询。产生两个或两个以上建议群组,所述两个或两个以上建议群组中的每一个包括所述建议集合的不同子集。提供所述两个或两个以上建议群组,使得所述两个或两个以上建议群组中的每一个显示在所述用户界面的搜索辅助段的单独分区中。



CN 102687138 B

1. 一种搜索建议聚类和呈现的方法，包括：

响应于通过用户界面的搜索查询区的文本段的输入来通过所述用户界面的搜索查询区获得所述文本段，所述文本段是用户还未提交的搜索查询的一部分；

获得建议集合，所述建议集合中的每个建议是所述搜索查询的建议完整化以使得该建议包括所述文本段；

产生两个或两个以上建议群组，所述两个或两个以上建议群组中的每一个包括所述建议集合的不同子集，其中所述两个或两个以上建议群组中的每个群组对应于用户还未提交的所述搜索查询的所述部分的不同解释；以及

提供所述两个或两个以上建议群组，使得所述两个或两个以上建议群组中的每一个显示在所述用户界面的搜索辅助段的独立区中；

提供与所述两个或两个以上建议群组中的每一个相关联的标签或图像，使得所述标签或图像被显示在所述用户界面中接近于所述两个或多个建议群组中的相应的建议群组；并且

对于所述两个或两个以上建议群组中的每一个，标识字符序列，该字符序列作为所述建议集合中的相应子集中的每个建议的至少一部分；

其中，提供与所述两个或两个以上建议群组中的每一个相关联的标签或图像包括提供与作为建议的所述相应子集中的每个建议的至少一部分的所述字符序列相关联的标签或图像。

2. 如权利要求 1 所述的方法，所述标签或图像提供，使得所述标签或图像显示为与所述用户界面的搜索辅助段的相应分区相关联。

3. 如权利要求 1 所述的方法，其中产生所述两个或两个以上建议群组包括：

获得与所述建议集合中的每个建议相关联的一个或多个特征的集合；以及

应用与所述建议集合中的每个建议相关联的所述一个或多个特征的集合以产生所述两个或两个以上建议群组。

4. 如权利要求 3 所述的方法，进一步包括：

解析所述建议集合中的每个建议以获得用于相应建议的一个或多个词的集合；

其中与所述建议集合中的每个建议相关联的所述一个或多个特征的集合包括在相应的一个或多个词的集合中的具有代表性的词。

5. 如权利要求 3 所述的方法，进一步包括：

获得与所述建议集合中的每个建议相关联的搜索结果的集合，所述搜索结果的集合中的每个搜索结果包括相应的标题、摘要和全球资源定位器 (URL)；

其中与所述建议集合中的每个建议相关联的所述一个或多个特征的集合包括或者基于在相应的搜索结果的集合中的词的集合。

6. 如权利要求 5 所述的方法，其中相应的所述搜索结果的集合中的所述词的集合包括以下各项中的至少一个：所述搜索结果的集合的至少一部分的标题中的词、所述搜索结果的集合的至少一部分的摘要中的词或所述搜索结果的集合的至少一部分的 URL 中的词。

7. 如权利要求 3 所述的方法，进一步包括：

获得与所述建议集合中的每个建议相关联的搜索结果，搜索结果中的每一个包括全球资源定位器 (URL)；

获得与所述建议集合中的每个建议相关联的点击数据；

其中与所述建议集合中的每个建议相关联的所述一个或多个特征的集合包括所获得的与相应建议相关联的点击数据，其中所述点击数据涉及与所述相应建议相关联的搜索结果的全球资源定位器 (URL)。

8. 如权利要求 1 所述的方法，进一步包括：

确定是否要呈现所述两个或两个以上建议群组；

其中提供所述两个或两个以上建议群组是根据是否确定了呈现所述两个或两个以上建议群组来执行的。

9. 如权利要求 1 所述的方法，进一步包括：

确定所述两个或两个以上建议群组中的每一个中的所述建议集合的子集将被提供的顺序；

其中执行了提供所述两个或两个以上建议群组，使得所述两个或两个以上建议群组中的每一个的建议集合的子集根据所述确定的顺序显示在所述用户界面的搜索辅助段的相应分区中。

10. 如权利要求 1 所述的方法，进一步包括：

在提供所述两个或两个以上建议群组之前确定所述两个或两个以上建议群组将被提供的顺序；

其中执行了提供所述两个或两个以上建议群组，使得所述两个或两个以上建议群组根据所述确定的顺序显示在所述用户界面的搜索辅助段的独立区中。

11. 如权利要求 10 所述的方法，其中确定所述两个或两个以上建议群组将被提供的顺序包括：

应用成本量以从所述两个或两个以上建议群组中产生表示定位建议的预期成本的数值；以及

最小化所述从所述两个或两个以上建议群组中定位建议的所述预期成本。

12. 如权利要求 1 所述的方法，其中所述搜索查询的所述部分是在制定所述搜索查询时由用户键入的字符序列。

13. 如权利要求 12 所述的方法，其中所述搜索查询的所述部分是前缀、中缀或后缀。

14. 如权利要求 1 所述的方法，还包括：

接收对所述两个或两个以上建议群组中的一个中建议的选择；以及

经由一个或多个搜索应用将所选的建议作为搜索查询执行。

15. 一种搜索建议聚类和呈现的装置，包括：

用于响应于通过用户界面的搜索查询区的文本段的输入来通过所述用户界面的搜索查询区获得所述文本段的装置，所述文本段是用户还未提交的搜索查询的一部分；

用于获得建议集合的装置，所述建议集合中的每个建议是所述搜索查询的建议完整化以使得该建议包括所述文本段；

用于从所述建议集合确定用户还未提交的所述搜索查询是模糊查询的装置；

用于产生两个或两个以上建议群组的装置，所述两个或两个以上建议群组中的每一个包括所述建议集合的不同子集，其中所述两个或两个以上建议群组中的每个群组对应于用户还未提交的所述搜索查询的所述部分的不同解释；以及

用于提供所述两个或两个以上建议群组，使得所述两个或两个以上建议群组中的每一个显示在所述用户界面的搜索辅助段的独立区中的装置；

用于提供与所述两个或两个以上建议群组中的每一个相关联的标签或图像，使得所述标签或图像被显示在所述用户界面中接近于所述两个或多个建议群组中的相应的建议群组的装置；并且

用于对于所述两个或两个以上建议群组中的每一个，标识字符序列的装置，该字符序列作为所述建议集合中的相应子集中的每个建议的至少一部分；

其中，提供与所述两个或两个以上建议群组中的每一个相关联的标签或图像包括提供与作为建议的所述相应子集中的每个建议的至少一部分的所述字符序列相关联的标签或图像。

16. 如权利要求 15 所述的装置，与所述两个或两个以上建议群组中的每一个相关联的所述标签或图像被提供在所述用户界面中的搜索辅助段的相应分区中。

17. 如权利要求 16 所述的装置，所述装置还包括：

用于从所述两个或两个以上建议群组中的每一个的建议子集标识具有代表性的建议的装置；以及

用于获得与相应的所述两个或两个以上建议群组之一的具有代表性的建议相关联的具有代表性的标签或图像的装置；

其中提供与所述两个或两个以上建议群组中的每一个相关联的所述标签或图像包括提供与相应的所述两个或两个以上建议群组之一的具有代表性的建议相关联的所述具有代表性的标签或图像。

18. 如权利要求 17 所述的装置，其中所述具有代表性的建议是根据查询日志在所述建议子集中出现频率最高的一个建议。

19. 如权利要求 16 所述的装置，所述装置还包括：

用于对于所述两个或两个以上建议群组中的每一个，获得与所述建议集合的相应子集相关联的搜索结果的集合的装置，所述搜索结果的集合的每个搜索结果包括相应的标题、摘要和全球资源定位器 (URL)；以及

用于使用相应的所述搜索结果的集合标识或产生用于各所述两个或两个以上建议群组中的每一个的标签的装置。

20. 如权利要求 19 所述的装置，其中获得所述搜索结果的集合是通过对所述建议集合的相应子集中的一个或多个建议执行搜索查询而完成的。

21. 如权利要求 19 所述的装置，其中获得所述搜索结果的集合是通过仅使用所述建议集合的相应子集中的部分建议执行搜索查询而完成的，所述建议在所述两个或两个以上建议群组之一中是独特的。

22. 如权利要求 16 所述的装置，所述装置还包括：

用于标识所述两个或两个以上建议群组中的每一个中的所述建议集合的子集所共享的主题或方面的装置，其中所述主题或方面不被所述建议集合中的其他建议子集所共享；以及

用于获得表示所述主题或方面的标签或图像，从而使所述标签或图像能被与所述两个或两个以上建议群组中的相应一个相关联地提供的装置。

23. 如权利要求 16 所述的装置，所述装置还包括：

用于对于所述两个或两个以上建议群组中的每一个确定所述两个或两个以上建议群组之一的关联程度的装置；以及

用于使用所述两个或两个以上建议群组之一以外的信息来根据所述两个或两个以上建议群组之一的关联程度获得与所述两个或两个以上建议群组之一相关联的标签或图像的装置。

搜索建议聚类和呈现

[0001] 本申请要求于2009年12月28日提交的美国专利申请No:12/648,220的权益,为了所有目的,通过引用将其内容结合在此。

技术领域

[0002] 本发明总体上涉及计算机应用搜索和所建议的搜索查询的呈现。

背景技术

[0003] 万维网的用户熟悉网上用于定位感兴趣的内容的各种服务。很多实体提供了搜索引擎,并且搜索能力嵌入到很多网站中。例如,许多网站提供能使用户搜索到该网站的以及因特网上的网站的内容的应用程序。

[0004] 搜索引擎通常提供搜索建议工具,该搜索建议工具通过预测用户将要键入的后面的字符和词来帮助用户更快地完整化他们的查询。例如,当用户开始键入“sacr…”时,下拉式窗口通常出现在搜索框下,提供常用的完整化结果以及相关的建议,诸如“sacramento”、“sacramento airport”和“sacred heart”。用户则可以简单地从列表中选择而无需键入完整的搜索查询。

发明内容

[0005] 本发明公开了用于聚类和呈现所建议的搜索查询(即,搜索建议)的方法和装置。根据一个实施例,通过用户界面的搜索查询区获得文本段,所述文本段是搜索查询的一部分。获得建议集合,该建议集合中的每个建议是与所述文本段相关的或包括所述文本段的所建议的搜索查询。产生两个或两个以上建议群组,所述两个或两个以上建议群组中的每一个包括所述建议集合的不同子集。提供所述两个或两个以上建议群组,使得所述两个或两个以上建议群组中的每一个显示在所述用户界面的搜索辅助段的单独分区中。

[0006] 根据一方面,提供与所述两个或两个以上建议群组中的每一个相关的标签或图像,使得所述标签或图像被显示在所述用户界面中,紧靠所述两个或两个以上建议群组中的相应一个。所述标签或图像可以通过使用所述相应的建议群组之中的信息来获得。可替代地,所述标签或图像可以通过使用所述相应的建议群组以外的信息,加上或代替所述相应的建议群组之中获得的信息来获得。

[0007] 根据又一方面,当根据所提交的搜索查询的一部分确定所述搜索查询是模糊的时,就可以聚类建议集合。例如,如果建议集合的初步聚类产出实质上大小各不相同的建议群组,则搜索查询可以是模糊的。

[0008] 在另一实施例中,本发明涉及一种装置,该装置包括处理器、存储器和显示器。所述处理器和存储器被配置为执行一个或多个以上描述的方法操作。在另一实施例中,本发明涉及一种上面存储有计算机程序指令的计算机可读存储介质,所述计算机程序指令布置为执行一个或多个以上描述的方法操作。

[0009] 在本发明的以下说明书及以示例方式图示了本发明原理的附图中,将对本发明的

这些及其他特征和优点进行更详细的呈现。

附图说明

[0010] 图 1 为示出了实施各种实施例的示例系统的框图。

[0011] 图 2A 为示出了用于呈现搜索查询建议列表的示例图形用户界面的简图。

[0012] 图 2B-C 为图形用户界面，示出了对建议列表执行完聚类之后的示例建议群组。

[0013] 图 3 为示出了根据本发明的各种实施例的用于聚类和呈现建议集合的示例方法的过程流程图。

[0014] 图 4 为示出了示例建议的简图，当用户键入“salsa”到用户界面的搜索查询区中时，该示例建议可被提供给用户。

[0015] 图 5 为示出了可实施各种实施例的示例网络环境的简图。

[0016] 图 6 示出了可实施各种实施例的示例计算机系统。

具体实施方式

[0017] 现在，将对本发明的具体实施例进行详细描述。这些具体实施例的示例在附图中示出。虽然将结合这些具体实施例对本发明进行描述，但是应理解，这并不表示本发明仅限于这些实施例。相反地，由权利要求限定的本发明的精神和范围内的替换、修改和等效内容也试图包括在内。在以下描述中，阐述了许多具体细节以帮助透彻理解本发明。没有这些具体细节中的一些或全部也可以实现本发明。此外，对公知的过程操作不作详细的描述，以避免不必要地模糊本发明。

[0018] 所公开的实施例提供了一种用户界面，该用户界面用于响应于接收部分搜索查询而提供搜索建议。更具体地说，每一个搜索建议可以是与该部分搜索查询相关或包括（例如，完整化或纠正）该部分搜索查询的所建议的搜索查询。因此，术语“建议”、“搜索建议”、“所建议的搜索查询”、“查询完整化结果”、“所建议的搜索查询完整化结果”和“查询完整化建议”可以互换使用。

[0019] 提供给用户的搜索建议可以组织为两个或两个以上群组，可称之为聚类 (cluster) 或分区 (partition)。聚类搜索建议对具有多于一个可能解释的模糊查询特别有用。更具体地说，可以根据已经输入的搜索查询的一部分的不同解释来组织搜索查询。

[0020] 随着用户键入（例如，增加、修改和 / 或删除一个或多个字符），所提供的搜索建议也会变化。同样地，搜索建议的聚类也会随着用户键入搜索查询而动态地执行。因此，随着用户键入查询的部分，建议群组的数量、每一个建议群组中建议的数量以及建议群组的构建方式都会动态变化。换言之，用户修改、增加和 / 或删除至少部分搜索查询都会触发建议的聚类，以下将对此进行更详细的描述。

[0021] 近年来，因特网已经成为百万用户的主要信息来源。这些用户依赖于因特网来给自己搜索感兴趣的信息。用户搜索信息的一个传统方式是通过搜索服务网页开始搜索查询。一般地，用户可以在搜索网页的输入框中输入包括一个或多个搜索项的查询，然后根据所输入的搜索项开始搜索。响应于该查询，网页搜索引擎通常返回搜索结果文档的有序列表。

[0022] 文档可被定义为用于标识文档所位于的位置的统一资源定位符 (URL)。文档可以

位于特定网站上，也可以位于该网站的特定网页上。例如，第一 URL 可以标识文档所位于的网页的位置，而第二 URL 可以标识文档可以位于的网站的位置。

[0023] 图 1 示出了可以实施本发明的各种实施例的示例网络段。如图所示，多个客户端 102a、102b 和 102c 可以通过图形用户界面访问搜索应用（例如，通过网络 104 访问在搜索服务器 106 上的搜索应用）和 / 或访问网络服务（例如，访问在网络服务器 114 上的网络服务），以下将对此进行更详细的描述。网络可以采取任意适当的形式，如广域网或因特网和 / 或一个或多个局域网 (LAN)。网络 104 可以包括任意适当数量和类型的装置，例如，路由器和交换器，用于将搜索或 web 对象请求从每一个客户端转发到搜索或 web 应用并将搜索或网络结果转发回提出请求的客户端。

[0024] 本发明还可以在广泛的网络环境（以网络 104 表示）中实现，包括（例如）基于 TCP/IP 的网络、电信网络、无线网络等。此外，用于实施本发明的实施例的计算机程序指令可存储在任意类型的计算机可读介质中，可以根据多种计算模型在独立的计算装置上执行，这些计算模型包括客户端 / 服务器模型、对等模型；或者根据分布式计算模型执行，在该分布式计算模型中，这里描述的各种功能可以在不同的位置上实现或使用。

[0025] 搜索应用通常允许用户（人类或自动实体）搜索通过网络 104 可访问的并且涉及搜索查询的信息，该搜索查询包括一个或多个搜索项。用户可通过任意方式输入搜索项。例如，图形用户界面可以向客户端呈现输入特征（例如，在客户端的装置上），所以客户端可以输入包括一个或多个搜索项的查询，下文会对图形用户界面进行更详细的描述。在一个具体的实施方式中，图形用户界面呈现输入框（即，搜索查询区），用户可键入包括任意数量的搜索项或其部分的查询。具体地，图形用户界面可以提供用于接收至少部分搜索查询的搜索查询区，以及可提供与搜索查询相关联的所建议的搜索查询（即，搜索建议）的另一部分。用户可以通过图形用户界面选择所建议的搜索查询之一来提交到搜索引擎。

[0026] 然后通过一个或多个搜索应用（例如，与搜索服务器 106 和 / 或网络服务器 114 相关联的）和 / 或一个或多个数据源执行搜索查询。本发明的实施例可以采用任意搜索应用。这些搜索应用可以在任意数量的服务器上实施，但是为了清楚起见，图中仅示出了一个搜索服务器 106。

[0027] 搜索服务器 106（或多个服务器）可以访问一个或多个查询日志 110，搜索信息保存在该查询日志中。例如，查询日志 110 可以保存在耦接到搜索服务器 106 的一个或多个存储器中。每次用户对一个或多个搜索项执行搜索时，关于该搜索的信息可以保存在查询日志 110 中。例如，用户的搜索请求可以包括任意数量的参数，如用户或浏览器身份和搜索项，这些都可以保存在查询日志 110 中。与搜索有关的额外信息（如时间戳）也可以和搜索请求参数一起保存在查询日志 110 中。当根据所输入的搜索项向用户呈现结果时，该搜索结果的参数也可以保存在查询日志 110 中。例如，具体的搜索结果，如网站、搜索结果呈现的顺序、每一个搜索结果是赞助搜索结果还是算法搜索结果、每一个搜索结果的所有者（如，网站）、每一个搜索结果是否是由用户（如果有的话）选择（即，点击）的和 / 或时间戳，也可以保存在查询日志 110 中。

[0028] 在接收到搜索查询之后，搜索服务器 106 可以标识并且呈现与该查询相关的适当的网页。例如，搜索服务器 106 可以标识并呈现多个超文本链接（该超文本链接标识与搜索查询有关的内容），以及呈现与多个超文本链接相关联的总结或摘要。

[0029] 这里公开的实施例可以通过搜索服务器（或其他服务器）106 和 / 或客户端 102a、102b 和 102c 实施。例如，各种特征可以通过客户端 102a、102b 和 102c 上的网络浏览器和 / 或应用来实施。所公开的实施例可以通过软件和 / 或硬件来实施。

[0030] 搜索引擎不断探寻着减少执行搜索相关任务的用户工作的方法。这些工作导致了广泛使用的自动完整化机制，当用户制定查询时，该自动完整化机制自动建议搜索查询的可能的完整化结果。然而，常规的自动完整化机制有可能提供让用户感到混乱的搜索建议，特别是当完整化结果的集合是由以交错方式显示的查询的不同解释构成时。

[0031] 图 2A 是示出了示例图形用户界面的简图，该图形用户界面示出了通过常规的自动完整化机制提供的所建议的搜索查询。考虑这样的情况 :用户通过键入字符序列 haifa 到与搜索引擎相关联的图形用户界面的搜索查询区 202 开始搜索引擎查询。当用户在图形用户界面的搜索查询区 202 键入内容时，用户的输入可以当作搜索查询的一部分。该部分搜索查询可包括一个或多个字符，以及一个或多个词或其部分。在此示例中，该部分搜索查询被当作搜索查询前缀。对于该搜索查询前缀而言，主搜索引擎呈现的建议集合可包括有序列表，如图中 204 所示。

[0032] 根据过去用户行为的、诸如点击行为、查询频率或查询重制的各种因素可以确定由搜索引擎提供的所建议的搜索查询完整化结果（即搜索建议）的集合。所公开的实施例可以通过组织根据主题自动完整化的建议来拓展当前查询完整化方法。

[0033] 如图 2A 所示，查询完整化建议可对应于不完全相同的现实世界中的实体、方面或主题。例如，在位置 1、2 和 5 的建议对应于流行艺人，而在位置 3 和 6 的建议对应于城市。此外，与类似或相同方面或主题相关联的查询建议不会被划分在同一组群中，因此从局部视角来看，这些建议通常呈现为无序列表。

[0034] 如图 2A 所示，当搜索查询（或其部分）具有不同的可能的含义时，这些建议可涉及已经提供的搜索查询的部分的不同解释。此外，这些建议通常只根据流行度来分类，从而导致对应于不同解释的建议以交错方式被提供。所公开的实施例实现了为所建议的搜索查询进行分组，从而允许用户容易地标识包括最相关搜索建议的搜索查询群组。

[0035] 图 2B-C 是示出了示例图形用户界面的屏幕截图，该图形用户界面可根据各种实施例呈现。如图 2B-C 所示，所公开的实施例能使所建议的查询的集合被聚类，并且通过图形用户界面呈现。此外，每一个聚类可以通过图形用户界面中的标签或图像标识，分别如图 2B 和 2C 所示。

[0036] 如图 2B 和 2C 所示，当用户输入搜索查询部分“Haifa”到图形用户界面的搜索查询区 202 中时，可以获得包括搜索查询部分的建议集合（例如，通过一个或多个搜索查询日志）。建议集合可以通过使用一个或多个聚类方法根据现实世界中的实体、方面、主题或其他准则被聚类为两个或两个以上建议群组。例如，每一个建议群组可呈现在图形用户界面的单独分区或分段中。

[0037] 此外，标识每一个建议群组的合适的标签或图像可被确定并且被与建议群组相关联地提供，以便帮助用户在相应的建议群组之间进行区分。如图 2B-C 所示，标签或图像可被与所述两个或两个以上建议群组中的每一个相关联地显示，使得该标签或图像被与相应的群组相关联地提供。例如，标签或图像可以显示在用户界面中，紧靠所述两个或两个以上建议群组中的相应一个。

[0038] 如图 2B 所示，不同的标签可被与每一个建议群组或图形用户界面的对应分区相关联地显示。例如，可以为显示在分区 206 中的第一建议群组指定一个标签，即 208 处显示的“Haifa(Singer(歌手))”，而为显示在分区 210 中的第二建议群组指定一个标签，即 212 处显示的“Haifa(City(城市))”。

[0039] 同样地，如图 2C 所示，不同的图像可被与每一个建议群组或图形用户界面的分区相关联地显示。例如，显示在分区 214 中的第一建议群组可以由 216 处显示的图像来标识，而显示在分区 218 中的第二建议群组可以由 220 处显示的图像来标识。

[0040] 当用户选择建议群组之一中的建议之一时，与所选择的建议相关联的搜索结果可被获得和提供。以此方式，所公开的实施例可给用户搜索过程带来方便。

[0041] 图 3 为示出了根据各种实施例的执行搜索建议聚类的示例方法的过程流程图。通过用户界面的搜索查询区可以在 302 处获得文本段，其中文本段是搜索查询的一部分。更具体地说，搜索查询的该部分可以是搜索查询的第一部分，其可被称为搜索查询的“前缀”或“查询前缀”。例如，查询前缀可以是制订搜索查询时由用户键入的字符序列。可替代地，搜索查询的该部分可以在预期搜索查询的中间或末端，其可被分别称为“中缀”或“后缀”。

[0042] 可在 304 处获得建议集合，其中在建议集合中的每一个建议是包括文本段的所建议的搜索查询。可以通过针对包括用户输入文本（例如，查询前缀）的查询来搜索搜索查询数据库，从而获得建议集合。搜索查询数据库可以与用户相关联，或者可以是存储用于多个用户的数据的全局数据库。一般而言，建议是根据相应的搜索查询的流行度来排序的。

[0043] 在一个实施例中，可以根据建议集合确定搜索查询是否是模糊查询。当之前输入的搜索查询的部分有多于一个的可能的解释时，搜索查询可以被确定为模糊的。例如，在初步聚类建议集合之后，可根据每一个建议群组中的建议的数量确定查询是模糊的。更具体地说，当建议落入两个或两个以上群组中时，则查询可以确定为是模糊的。然而，如果一个群组相比于另一个群组存在非常少的建议，这表示查询不是模糊的。如果搜索查询是模糊查询，就可以聚类建议集合，如以下参照方框 306 和 308 所述。

[0044] 可在 306 处产生两个或两个以上建议群组，其中所述两个或两个以上建议群组中的每一个包括建议集合的不同子集。更具体地说，可以获得与建议集合中的每一个建议相关联的一个或多个特征的集合。然后，与建议集合中的每一个建议相关联的一个或多个特征的集合可用于产生两个或两个以上建议群组。特征可以从建议和 / 或使用建议执行搜索查询时所获得的搜索结果的至少一部分获得。例如，特定建议的特征可以包括建议中的一个或多个词的集合和 / 或搜索结果中的一个或多个词的集合。特定建议的搜索结果中的词可包括在一个或多个文档的标题、摘要和 / 或统一资源定位符 (URL) 中发现的词。特定建议的特征还可以包括与该建议相关联的点击数据。以下将对用于获得和使用各种特征的各种机制进行更详细的描述。

[0045] 在产生建议群组之后，可取的是对建议群组重新分组。例如，当特定建议群组中的建议的数量明显小于另一个建议群组中的建议的数量时，重新分组是可取的。

[0046] 可取的是，确定是否呈现两个或两个以上建议群组。例如，可以确定查询不是模糊的。如果查询确定为不是模糊的，就可以不提供（例如，显示）两个或两个以上建议群组。

[0047] 当查询是模糊的时，可在 308 处提供两个或两个以上建议群组，使得所述两个或两个以上建议群组中的每一个被显示在用户界面的搜索辅助段的单独分区中。例如，这些

分区可以在用户界面的搜索辅助段中按顺序呈现。以下将对在搜索辅助段中排序建议群组的多种方法进行更详细的描述。

[0048] 特定建议群组中的建议可以根据各种方法进行排序。例如，特定建议群组中的建议可以按作为搜索查询的建议的执行流行度或选择流行度的顺序来显示。作为搜索查询的特定建议的流行度可以使用输入当前搜索查询的用户的查询日志数据来确定。可替代地，作为搜索查询的特定建议的流行度可以使用多个用户的查询日志数据来确定。

[0049] 此外，可以提供标识所述两个或两个以上建议群组中的每一个的标签或图像，使得该标签或图像被与相应的建议群组相关联地显示。例如，标签或图像可被与搜索辅助段的相应分区相关联地显示。更具体地说，与所述两个或两个以上建议群组中的每一个相关联的标签或图像可提供在与用户界面的搜索辅助段的相应段中。以下将对用于标识或产生将为特定建议群组呈现的标签或图像的各种方法进行更详细的描述。

[0050] 1. 聚类建议

[0051] 产生两个或两个以上建议群组以使得在建议群组之间分配建议集合，这可被定义为一个数学问题。

[0052] 问题：假设一部分查询（例如，前缀 p）和建议集合（例如有序的建议集合），S = {s1，s2，…，sn}，我们可以将 S 划分为 k 个不相交的分区（例如，有序分区），P = {P1，P2，…，Pk}，使得每个 si 属于恰好一个 Pj，并且每个 Pj 中的成员都是主题相关的（即，指查询 q 的单一主题或方面）。在将 S 划分之后，我们可以给每个分区指定不同的标签 L（和 / 或图像 I），使得 L(Pj) 或 I(Pj) 向用户表示或描述由分区 P(j) 中的项共享的主题或方面，而非由 S 中的剩余元素共享的主题或方面。更具体地说，我们可以标识由分区 P(j) 中的成员共享的主题或方面，然后获得表示所标识的主题或方面的标签或图像。我们还可以对分区 P(j) 和 / 或每个分区 P(j) 中的建议进行排序，使得集合 S 的效用对用户最大化。

[0053] 各种聚类机制可用于根据查询的一部分（例如，查询前缀）将建议集合划分为两个或两个以上建议群组。以下将描述 3 种不同的聚类机制。在下列描述中，假设建议集合中的建议所共享的部分查询是查询前缀。然而，重要的是，注意共享的查询的部分可在查询的不同位置发生。

[0054] 聚类任务可以简化为找出任意两个正被聚类的元素（如，建议）之间的相似性（或距离）的任务。以下描述的 3 种示例聚类机制提供了估计提供用于部分查询的建议集合中的两个建议之间的相似性的不同方法。

[0055] 1.1 中心词聚类

[0056] 用户键入搜索查询时所提供的许多建议是完整化结果，其中用户输入作为前缀。有时候，用户输入被当作后缀或中缀。结果，集合 S 在词汇水平上可能已经非常相似了。总之，建议 si 可以看作 si = p ∪ ci，其中 p 是用户提供的查询前缀，ci 是添加到特定建议 si 中的附加上下文（例如，一个或多个字符）。如果用户已经输入的查询的部分是查询前缀，附加上下文 ci 可以是查询前缀之后出现的一个或多个字符。可替代地，附加上下文 ci 可以包括在查询的部分之前出现的一个或多个字符和 / 或查询的部分之后出现的一个或多个字符。已经输入的查询的部分之前和 / 或之后的一个或多个字符可以包括一个或多个词或其部分。

[0057] 图 4 为示出了示例建议的简图，当用户键入“salsa”到用户界面的搜索查询区时，

这些示例建议可提供给用户。如在此示例中所示，建议已经共享了前缀 p。可用于标识建议 si 所属的聚类的术语最可能在该建议 si 的附加上下文中。

[0058] 在一个实施例中，我们可以从每个建议 si 中选择单个术语，其中该单个术语是最具代表性的术语，即，使建议 si 最区分于剩余建议的术语。然后，可以使用这些术语对 S 执行聚类。在图 4 所示的示例中，区别性的术语是“recipes（食谱）”、“dancing（舞蹈）”、“dance（跳舞）”、“music（音乐）”、“singer（歌手）”、“homemade（自制）”、“lessons（课程）”和“classes（班）”。这些术语中的每一个都可以称作相应建议 si 的“中心词”。

[0059] 可以解析在建议集合 S 中的每一个建议 si 以获得一个或多个词的集合。然后，可以对每个建议 si 标识一个或多个词的集合中的“中心词”（例如，具有代表性的词）。因此，与建议集合中的每个建议 si 相关联的特征的集合可以包括该建议的中心词。

[0060] 可以使用用于估计语义或主题词水平的相似度的各种方法来确定建议的中心词之间的相似度，从而确定建议 si 之间的相似度。通常使用的方法包括根据大语料库或词汇资源（如词汇网络）中的词语上下文的方法。例如，使用信息检索 (IR) 的点间互信息 (PMI) (PMI-IR) 是可用于确定两个词 {wi，wj} 之间的相似度的简单共现技术。两个词 {wi，wj} 之间的相似度可被定义为两个词之间的点间互信息，其中单个词的几率 P(wi) 以及联合概率 P(wi,wj) 是使用语料库中的最大出现可能性估计的。特别地，在此情况下两个词之间的相似度测量可被定义为：

[0061] $$\mathrm{Sim(wi,wj)} = \log \frac{\frac{|counts(wi) \cap counts(wj)|}{n}}{\frac{|counts(wi)|}{n} \cdot \frac{|counts(wj)|}{n}}$$

[0062] 其中，计数 (x) 是包含 x 的文档的集合，n 是语料库尺寸（例如，搜索结果的数量）。两个建议之间的相似度可以是中心词之间的相似度。

[0063] 中心词选择

[0064] 由于 web 查询的平均长度较短，附加上下文 ci 经常包括单个术语。因此，此单个术语可以用作建议 si 的中心词。然而，存在附加上下文 ci 包括两个或两个以上词的情况。因此，对于特定建议 si，可以使用用于选择中心词的各种方法来从这些词中选出中心词。以下描述了几个示例方法。

[0065] 首词：选择附加上下文 ci 中最左边的词。例如，当建议为“salsa singer cruz”时，附加上下文为“singer cruz”，首词为“singer”。

[0066] 尾词：选择附加上下文 ci 中最右边的词（例如，建议“salsa singer cruz”中的 cruz）。

[0067] 频率：对附加上下文 ci 中的每个词，计算其术语频率 (tf) 值和逆向文档频率 (idf) 值的积：tf·idf，其中可用于计算 tf 的“文档”可以包括正被聚类的建议集合 S 中的所有词，并且对用户输入的所有建议的集合 S 计算 idf：

[0068] $$\mathrm{tf(w)} = \frac{\sum_{s \in S} count_w(s)}{\sum_{s \in S} |s|}$$

[0069] $$\mathrm{idf(w)} = \log \frac{|S|}{|\{s \mid w \in s\}|}$$

[0070] 特定建议 si 的中心词可以通过选择具有最高的 tf·idf 值的词来选择。

[0071] 1.2 结果集合聚类

[0072] 为了确定两个查询建议之间的相似度，与查询建议中的每一个相关联的搜索结果可被利用。对于相应查询建议的排名前 N 个搜索结果（例如，文档）中找出的术语，建议查询中的每一个可以使用相应的 tf·idf 值来表示。因此，与建议集合中的每一个建议相关联的特征的集合可以包括或者基于相应的搜索结果的集合中的词的集合。

[0073] 假设查询建议 si，我们可以获得搜索引擎返回的建议 si 的前 N 个文档的搜索结果 R(si) 的集合。每一个文档 $d \in R(si)$ 可包括标题、摘要和统一资源定位符 (URL)。摘要可以是示出给用户的文档 d 的一部分，包括查询中的术语和术语周围的少量上下文。因此，对于排名前 N 个的搜索结果中的每一个的标题 t(d)、摘要 a(d) 和 / 或 URL u(d) 中的一个或多个词，确定其 tf·idf 值。

[0074] 在一个实施例中，每一个文档组成部分（标题、摘要和 / 或 URL）可以表示为出现在其中的术语的 tf·idf 矢量，即，每个位置存储一个词的 tf·idf 值的矢量。可以针对前 N 个的文档中的每一个来确定文档组成部分的矢量。结果集合 R(s) 的文档组成部分的矢量可以通过对特定建议 si 的所有文档获得每一个组成部分矢量的形心（例如，平均矢量）来获得。例如，结果集合 R(si) 的矢量标题 (si) 可通过获得限定结果集合 R(si) 的文档的排名前 N 个的标题的矢量标题 (d) 的形心来获得。特定建议 si 的单个矢量 vs 可以通过将与该建议 si 的结果集合 R(si) 相对应的矢量标题 (d)、摘要 (d) 和 / 或 url(d) 连接起来而获得。可对每个建议 si 执行该过程。诸如余弦相似性函数的相似性函数可用于确定两个不同的形心矢量 vs 之间的相似性，因此，两个对应的建议 si 之间的相似性是它们的点积。

[0075] $Sim(s_i, s_j) = v_{si} \cdot v_{sj}$

[0076] 1.3 基于点击的聚类

[0077] 搜索引擎保持的点击数据可被用于将建议集合 S 划分为两个或两个以上群组。点击数据可包括关于用户所点击的 URL 的信息，这些 URL 来自呈现给一个或多个用户的搜索结果。例如，对于多个用户的特定查询建议“pineapple salsa”，搜索日志可包括 3 个不同点击的 URL。

[0078] URL1:www.allrecipes.com/pineapple-salsa/detail.aspx

[0079] URL2:www.cooks.com/rec/pineapple_salsa.html

[0080] URL3:www.blogchef.net/pineapple-salsa-recipe/

[0081] 使用特定查询建议 si 的点击数据，我们可以将部分搜索查询（例如，查询前缀）的每一个建议 si 特征化为与该建议相关联的点击的 URL 的集合。具有类似用户点击行为的建议可被一起分组在相同的群组中。更具体地说，产生对一个或多个相同的 URL 的点击的不同查询可以捕捉类似的用户意图。例如，查询建议“pineapple salsa for fish”可产生对以上 URL 之一的点击，从而表明两个建议是类似的。

[0082] 使用点击的 URL 可导致被证明为过于受限的具体表示，因为网站趋向于依据概念给出网页。因此，我们可以使用来自点击数据的基础 URL，而不是具体点击的 URL。例如，URL1 可以一般化为 www.allrecipes.com。因此，可以使用与网站相关联的 URL，而不是与具体网页相关联的 URL。

[0083] 此外，诸如 www.wikipedia.org的信息网站或百科全书式网站将会引入非期望的

偏差，并且导致不相似的概念放置在相同的聚类中。同样地，诸如www.youtube.com的其他网站也可以引入这种偏差。为了解决这个问题，我们将每个建议当作一个文档，并且计算每个基础URL的逆文档频率，并且当产生表现时，使用其作为权值，以下将对此进行更详细的描述。可替代地，我们可以根据他们的逆文档频率排除一个或多个URL。更具体地说，逆文档频率可以表示建议在查询日志中的出现频率的逆。

[0084] 可以使用点击数据表示查询建议。更具体地说，假设前缀p和与其相关联的建议集合S，我们可以定义p的点击图。点击图可被定义为包括两类节点的二分图：建议节点（s节点）和基础URL节点（u节点），以及定向边缘的集合E。建议集合S中的每一个建议可以表示为s节点。为了产生u节点，我们可以取与每一个建议相关联的基本URL的集合的并集，并且依据不同的基本URL产生一个节点。建议节点s和URL节点u之间的边缘s->u表示当s提交为查询时，点击了URL u。可以为每一个边缘指定一个权值，该权值是当s提交为查询时URL u被点击的次数。

[0085] 使用点击图，对于该图中的每个建议，我们可以产生L2标准化特征矢量，该标准化特征矢量的大小等于图中URL节点的数量，其中矢量中的每个维表示图中的一个URL。与URL j相关联的维的值可以计算为：

[0086] $f_j \frac{w_{sj}}{\sqrt{\sum_i^{|u|} w_{si}^2}}$ 如果建议s和j之间存在边缘的话；

[0087] 否则为0。

[0088] 其中U是点击图中URL的集合，w_{sj}是在点击图中与边缘s->j相关联的权值。为了计算前缀p的两个建议之间的相似度，我们使用诸如余弦相似度函数的相似度函数来产生如下的相似度量：

[0089] $$Sim(x,y) = \sum_i^{|U|} \frac{x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$$

[0090] 1.4 聚类算法

[0091] 一旦使用以上所述的3种方法之一限定给定查询前缀的S中的任意一对建议之间的相似度时，就可以将该相似度用作用于聚类的相似度量。然后，可以使用聚类算法对使用相应的相似度量的建议来分组，使得相似的建议划分在一起。更具体地说，一旦排除了两个不同的建议之间的相似度，就会使用诸如层次聚合聚类方法的无监督聚类算法来将建议划分为两个或两个以上聚类。

[0092] 2. 标记聚类

[0093] 一旦建议集合S被划分为两个或两个以上群组，不同的标签或图像可以指定给每一个建议群组，并且被与相应的建议群组相关联地显示。以此方式，可以提供视觉线索来指示相应建议群组的主题。以下会对将标签或图像指定给建议群组的各种方法进行详细的描述。

[0094] 2.1 最高频的建议（MFS）

[0095] 选择用于查询建议聚类的标签（或图像）的一种方法是选择聚类中最具代表性的建议。由于聚类中的每个建议都是查询，选择最具代表性的建议的一种方法是选择已被呈现和/或被用户点击过的（例如，根据查询日志）最高频的建议。更具体地说，MFS指定给

特定建议聚类 S 的标签是：

[0096] $MFS(S) = s_i : s_i \in S, \forall_{s_j \in S} \; Freq(s_j) \leq Freq(s_i)$

[0097] 其中，Freq(x) 是查询日志中观察到 x 的次数。

[0098] 一旦标识出建议群组中最具代表性的建议，就可以获得并提供（例如，显示）与具有代表性的建议相关联的标签和 / 或图像。例如，该标签可以简单地是具有代表性的建议（例如，“Nursing（护理）”）。作为另一个示例，可提供护士的图像，而不是标签“nursing”。

[0099] 2.2 最长公共子序列 (LCS)

[0100] 通常，聚类中的建议共享字符序列，而不与其他聚类中的建议共享该字符序列。例如，用户提交的部分查询“us a”可以完整化为“us airways”和“us airways flights”（都在一个聚类中），以及“us army”和“us army jobs”（在不同的聚类中）。可取的是，使用该建议的最长公共子序列（或选择图像）作为查询建议聚类的标签。建议集合 S 的 LCS 可以表示如下：

[0101] $LCS(S) =$

[0102] $li : li \in Q(S), \forall_{l_j \in Q(S)} \; Length(lj) \leq Length(li)$

[0103] 其中，Q(S) 是任意建议 $s \in S$ 的子序列的集合。例如，LCS 方法指定给搜索查询建议集合的标签是“nursing home”，而该搜索查询建议集合包括“nursing home”、“nursing home compare”和“nursing home costs”。因此，一旦标识出两个或两个以上建议的群组所共有的字符序列，就可以提供（例如，显示）与该建议群组所共有的该字符序列相关联（例如，标识）的标签或图像。

[0104] 2.3 最高频的结果集合 (MFRS)

[0105] MFS 和 LCS 方法都具有的缺点是，他们从属于聚类的建议中产生用于该聚类的标签。然而，对于某些建议聚类而言，有意义的标签是不能单独从聚类的建议中确定的。在这些情况下，用于聚类的标签可以使用该聚类以外的资源获得。例如，对于包括建议“los angeles daily news”、“los angeles times”和“los angeles times newspaper”的聚类，有用的标签可以是“los angeles newspapers”-与该聚类中的所有建议仅部分重叠的标签。

[0106] 就执行聚类自身而言，我们可以使用每个建议（当其作为提交给搜索引擎的查询时）的排名靠前的文档的集合作为该外部知识。更具体地说，每一个搜索查询建议可以通过搜索引擎执行为搜索查询，以获得相应的文档集合。通过将特定聚类中的建议集合转换为文档集合，我们可以使用为标记文档（而非查询）开发的各种方法。

[0107] 标记文档聚类的一个标准方法是从文档中获取词 n- 元，并且选择最高频的 n- 元。n- 元是 n 个词的连续序列。设 R(s) 为建议 s 的排名靠前的搜索结果的集合；设 $R(S) = \cup_{S_i \in S} R(S_i)$；设 NG(d) 为文档 d 中包含的词 n- 元的集合；并且设 NG(R(S)) 为建议聚类的所有排名靠前的文档中的所有 n- 元的集合，$NG(R(S)) = \cup_{d \in R(S)} NG(d)$。则 MFRS 方法指定给建议集合 S 的标签是

[0108] $MFRS(S) = li : li \in NG(R(S)), \forall_{lj \in NG(R(S))} Count(l_j, R(S)) \leq Count(l_i, R(S)).$

[0109] 例如，MFRS 方法可以将标签“news”指定给包括建议“los angeles daily news”、“los angeles times”和“los angeles times newspaper”的建议聚类。

[0110] 根据一个实施例，对于每一个建议群组，可以获得与相应建议集合相关联的搜索结果（例如，文档）的集合，其中每一个搜索结果包括相应的标题、摘要和统一资源定位符(URL)。然后，使用相应的搜索结果的集合，可以标识或产生每一个建议群组的标签（或图像）。

[0111] 2.4 最高频的修改结果集合 (MFRS*)

[0112] 作为一批用于聚类的实体，搜索建议的独特之处在于他们具有高词汇重叠度。在具有长公共子序列的聚类中，我们感兴趣的用于标记的元素有时候在不为该聚类的所有元素所共享的那部分建议中被最好表示。因此，可以使用附加标记机制 MFRS*。MFRS*类似于 MFRS，但是为了获得排名靠前的文档的集合而执行的查询可以通过仅执行在聚类中独特的部分建议来获得（而非执行他们全部的搜索建议）。例如，对于包括建议“los angeles public library”、“los angeles police department”和“los angeles unified school district”的建议聚类，可以执行搜索查询“public library”和“police department”和“unified school district”。MFRS*机制可被定义为如下。

[0113] 设 s_i^* 是除去了建议集合 S 的最长公共子序列的建议 s_i，$s_i^* = s_i - LCS(S)$，并且设 S^* 是从所有建议中除去了最长公共子序列的建议集合 S，$S^* = \cup_i si^*$，则 MFRS* 指定给 S 的标签是：

[0114] $MFRS^*(S) = MFRS(S^*)$.

[0115] 例如，MFRS*方法可以将标签“services”指定给包括建议“los angeles public library”、“los angeles police department” 和“los angeles unified school district”的建议聚类。

[0116] 2.5 组合标记策略

[0117] 如以上所述的一个或多个标记机制可以单独地或者彼此结合地使用，以将标签（或图像）指定给各种建议群组。建议聚类可以具有不同的特点，因此可通过不同的标记方法获得益处。因此，选择并且应用的一个或多个标记机制可以根据他们所应用于的系统而变化。此外，选择并且应用的标记机制可根据聚类的聚类特点而变化。

[0118] 可以单独地使用来自聚类中的信息（例如，建议）来将标签（或图像）指定给聚类。例如，可以使用诸如 MFS 或 LCS 的机制来指定标签（或图像）。可替代地，可以使用该聚类以外的信息（例如，搜索结果）加上或代替该聚类中的信息来将标签（或图像）指定给聚类。例如，可以使用诸如 MFRS 或 MFRS* 的机制来指定标签（或图像）。

[0119] 在一个实施例中，可以检查聚类以确定聚类的聚类关联程度。换言之，可以检查聚类以确定聚类的元素（例如，建议）的相似程度。聚类越紧凑（例如，聚类的元素越相似），就越可能在聚类的成员中找出合适的标签，而非从外部将其找出。建议集合 S 的关联程度可以使用聚类 S 的元素之间的平均距离来测量。当聚类 S 的关联程度小于阈值量，可以应用诸如 MFRS 或 MFRS* 的、使用聚类以外的信息的机制，在其他情况下，可以应用诸如 MFS 或 LCS 的、使用聚类中的信息的机制。

[0120] 3. 排序建议聚类

[0121] 所公开的实施例可以应用于呈现用于使查询完整化的建议集合，以减少在建议集合中定位期望的建议的用户花费。建议集合所分组的方式可以减少用户花费量。同样地，建议群组呈现的顺序以及特定建议群组中的建议呈现的顺序也可以影响用户花费量，其中，

用户花费是指在所呈现的建议集合中定位所期望的建议。

[0122] 根据一方面，将要提供两个或两个以上建议群组的顺序可以在提供用于显示的两个或两个以上建议群组之前确定。然后，可以提供所述两个或两个以上建议群组，使得所述两个或两个以上建议群组根据所确定的顺序显示在用户界面的搜索辅助段的单独分区中。

[0123] 成本度量可被应用以描述当在建议聚类集合中定位建议时所付出的用户花费的特征。更具体地说，成本度量可以产生表示从两个或两个以上建议的群组中定位建议的预期成本的数值。然后，可以应用算法来最小化从建议聚类集合中定位建议的预期成本。

[0124] 通过聚类（和标记）将与用户已经输入的搜索查询的一部分相关联地呈现的建议集合，我们能让用户在聚类之间跳过，然后在标识相关的聚类之后，用户可以在该聚类中细看以定位期望的建议。因此，标识期望的建议的成本可被定义为：

[0125] 读取聚类标签的时间：用户可通过读取相应的标签（或图像）来浏览建议的聚类。在每一个聚类 C 处，根据标签是否捕捉用户感兴趣的领域，用户可以决定应该跳过还是细看该聚类。我们可以将读取聚类标签的成本表示为 $T_{lb}(C)$。

[0126] 细看聚类的时间：一旦标识出包含期望的建议 s 的聚类 C，用户就可以细看聚类 C 中的建议，直到期望的建议被定位为止。我们可以将细看聚类中的每一个建议的成本表示为 $T_{sc}(s)$。

[0127] 考虑这样的情况：用户已经输入查询前缀 p 并且有兴趣从聚类 C1、C2....、Cn 的集合中定位建议 s，设 Cm 为包含建议 s1、s2...、sj 的聚类，使得 sk = s。换言之，建议 s 定位在聚类 Cm 中的位置 k。用户定位建议 s 的成本（可表示为 T(s)）可被定义为 $\sum_{i=1}^{m} T_{lb}(Ci)+\sum_{j=1}^{k} T_{sc}(S_j)$。为了简化，我们可以假设读取任意聚类标签的成本对于所有聚类都是相同的，即 T_{lb}。同样地，我们可以假设，无论怎样的建议，在聚类 id 中看完建议的成本 Tsc 都是相同的。聚类 m 中在位置 k 处的建议 s 的 T(s) 则变成了 T(s) = m •Tlb+k •Tsc。

[0128] 对于输入了前缀 p 的用户，在建议中定位感兴趣的建议的预期成本 T(p) 可被定义为：

[0129] $$Tp(R)=\sum_{\forall s} T(s)\cdot P(s\mid p),$$

[0130] 其中，P(s|p) 表示输入前缀后用户倾向于建议 s 的几率，Tp 是建议 s 的排名 R 的函数。当输入前缀 p 时，可以基于观察用户的偏好根据查询日志来估计 P(s|p)。更具体地说，可以标识该用户（或通常为多个用户）已经提交或选择的包括前缀 p 的查询。然后，可以根据标识出的查询来确定已经提交或选择的查询 s 的次数相对于包括前缀 s 的查询的总数。具体地，如果 f(p) 是一个用户（或多个用户）输入前缀的次数（例如，一个用户或多个用户曾提交包括前缀的查询的次数），并且 f(s) 是建议 s 曾被作为用户查询提交的次数，则

[0131] $$P(s\mid p)=\frac{f(s)}{f(p)}$$

[0132] 注意，$\sum_{\forall s} P(s\mid p)$，将通常小于 1，因为用户可输入了不在建议集合中的查询。我们可以假设对建议集合中不存在的建议感兴趣的用户的成本并不取决于呈现的建议集合的排名。

[0133] 可以使用排名算法对聚类以及聚类中的建议进行排序，以最小化 Tp(R)。在一个实

施例中，排名算法可以按频率 f(s) 的非递增顺序（例如，递减顺序）将聚类中的建议排序。为了对建议聚类排序，可以为每一个聚类 S 指定总频率 F(C)，该总频率 F(C) 等于聚类 C 中所有建议的频率的总和。因此，排名算法可以按总频率 F(C) 的非递增顺序（例如，递减顺序）将建议聚类排序。

[0134] 根据另一方面，可以对每一个建议群组的建议排序。更具体地说，可以确定两个或两个以上建议群组中的每一个中的建议集合的子集将被提供的顺序。例如，该顺序可指示根据查询日志的建议的流行度。然后，两个或两个以上建议群组中的每一个的建议可以根据所确定的顺序被显示在用户界面的搜索辅助段的相应分区中。

[0135] 使用本发明的实施例，可以通过图形用户界面执行搜索，同时可以使用同一图形用户界面提供搜索建议。所公开的实施例可以实施在任意各种计算上下文中。例如，如图 5 所示，设想了这样的实施方式，其中用户通过任意类型的计算机（例如，台式计算机、笔记本计算机、平板计算机等）1102、媒体计算平台 1103（例如，有线电视和卫星电视机顶盒和数字视频录像机）、手持计算装置（例如，个人数字助理）1104、蜂窝电话 1106 或任意类型的计算或通信平台与多种多样的网络环境交互。

[0136] 而且根据各种实施例，根据本发明进行的输入可以使用各种技术来获得。例如，可以通过图形用户界面从用户与本地应用、网站或基于 web 的应用或服务的交互来获得搜索查询，并且可以使用用于从用户获得信息的任意多种已知的机制来完成。然而，应当理解的是，从用户获得输入的这些方法仅为示例，搜索查询还可以通过多种其他方式获得。

[0137] 根据所公开的实施例可按照某些集中方式聚类和呈现搜索建议。这在图 5 中表示为服务器 1108 和数据存储装置 1110，将会理解的是，服务器和数据存储装置可对应于多个分布式装置和数据存储装置。本发明还可以在各种网络环境（以网络 1112 表示）中实现，例如包括基于 TCP/IP 的网络、电信网络、无线网络等。此外，实施本发明的实施例的计算机程序指令可存储在任意类型的计算机可读介质中，可以根据包括客户端 / 服务器模型、对等模型在内的各种计算模型在独立的计算装置上执行，或者根据分布式计算模型执行，在该分布式计算模型中，这里描述的各种功能可以在不同的位置上实现或使用。

[0138] 本发明所公开的技术可以软件和 / 或硬件系统的任意适当组合的方式实施，诸如基于 web 的服务器或台式计算机系统。此外，实施本发明的各种实施例的系统可以是便携式装置，诸如笔记本计算机或蜂窝电话。本发明的搜索装置和 / 或网络浏览器可以专门构造用于所需的目的，或者也可以是由计算机程序和 / 或存储在计算机中的数据结构选择性地激活或者重新配置的通用计算机。这里表示的处理不是固有地与任意特定计算机或其他装置有关。特别地，可以使用根据这里的教导写有程序的各种通用机器，或者构造更加专业的装置来执行所需的方法步骤将会更加方便。

[0139] 不管系统配置，可以使用一种或多种存储器或存储模块，这些存储器或存储模块配置为存储用于通用目的处理操作和 / 或这里描述的本发明的技术的数据及程序指令。例如，程序指令可以控制操作系统和 / 或一个或多个应用程序的操作。一个或多个存储器也可以配置为存储用于执行所公开的方法的指令、以及查询日志、标签、图像、搜索结果等。

[0140] 因为这种信息和程序指令可以用于实施这里描述的系统 / 方法，所以本发明涉及机器可读介质，该机器可读介质包括用于执行这里描述的各种操作的程序指令、状态信息等。机器可读介质的示例包括（但不限于）：诸如硬盘的磁性介质、软盘和磁带；诸如

CD-ROM 盘的光学介质 ;诸如软式光盘的磁光介质 ;以及专门配置为存储和执行程序指令的硬件装置,如只读存储器装置 (ROM) 和随机存取存储器 (RAM)。程序指令的示例包括机器代码 (诸如通过编译器生成的) 和包含高级代码的文件,计算机可以使用解释器执行该文件。

[0141] 图 6 示出了典型的计算系统,当适当地配置或设计时可以用作本发明的系统。计算系统 1200 包括任意数量的处理器 1202(也称作中央处理器或 CPU),该处理器耦接至存储装置,该存储装置包括主存储器 1206(通常为随机存取存储器或 RAM) 和主存储器 1204(通常为只读存储器或 ROM)。CPU 1202 可以是各种类型,包括微控制器和微处理器,如可编程装置 (如 CPLD 和 FPGA) 和诸如如门阵列 ASIC 或通用微处理器的非可编程装置。本领域已知的是,主存储器 1204 起着单向传递数据和指令到 CPU 的作用,主存储器 1206 通常用于以双向的方式传递数据和指令到 CPU。这些主存储装置都可以包括如上所述的任意适当的计算机可读介质。大容量存储装置 1208 也双向耦接至 CPU 1202,并且提供额外的数据存储能力,并且可以包括以上所述的任意计算机可读介质。大容量存储装置 1208 可用于存储程序、数据等,并且通常为诸如硬盘的次级存储介质。将会认识到,在适当的情况下,保存在大容量存储装置 1208 中的信息可以标准方式纳入到作为虚拟内存的部分主存储器 1206 中。诸如 CD-ROM 1214 的特定大容量存储装置也可以单向传递数据到 CPU。

[0142] CPU 1202 也可以耦接至接口 1210,该接口连接至一个或多个输入 / 输出装置,诸如视频监视器、轨迹球、鼠标、键盘、麦克风、触摸式显示屏、传感器智能卡阅读器、磁性或纸带阅读机、平板计算机、光笔、语音或手写识别器或其他公知的输入装置,如 (当然) 其他计算机。最后,CPU 1202 可选择地使用一般如 1212 所示的外部连接耦接至诸如数据库或计算机或电信网络的外部装置。由于具有这种连接,设想 CPU 在执行这里描述的方法步骤的过程中可以接收来自网络的信息,或可输出信息到网络。

[0143] 虽然前文出于清楚理解的目的在一些细节方面对本发明进行了描述,然而应清楚的是,可以在所附权利要求的范围内进行某些改变和修改。因此,本实施例将被认为是示例性的而非限制性的,并且本发明并不仅限于这里给出的细节,也可在所附权利要求的范围和等效内容内进行修改。

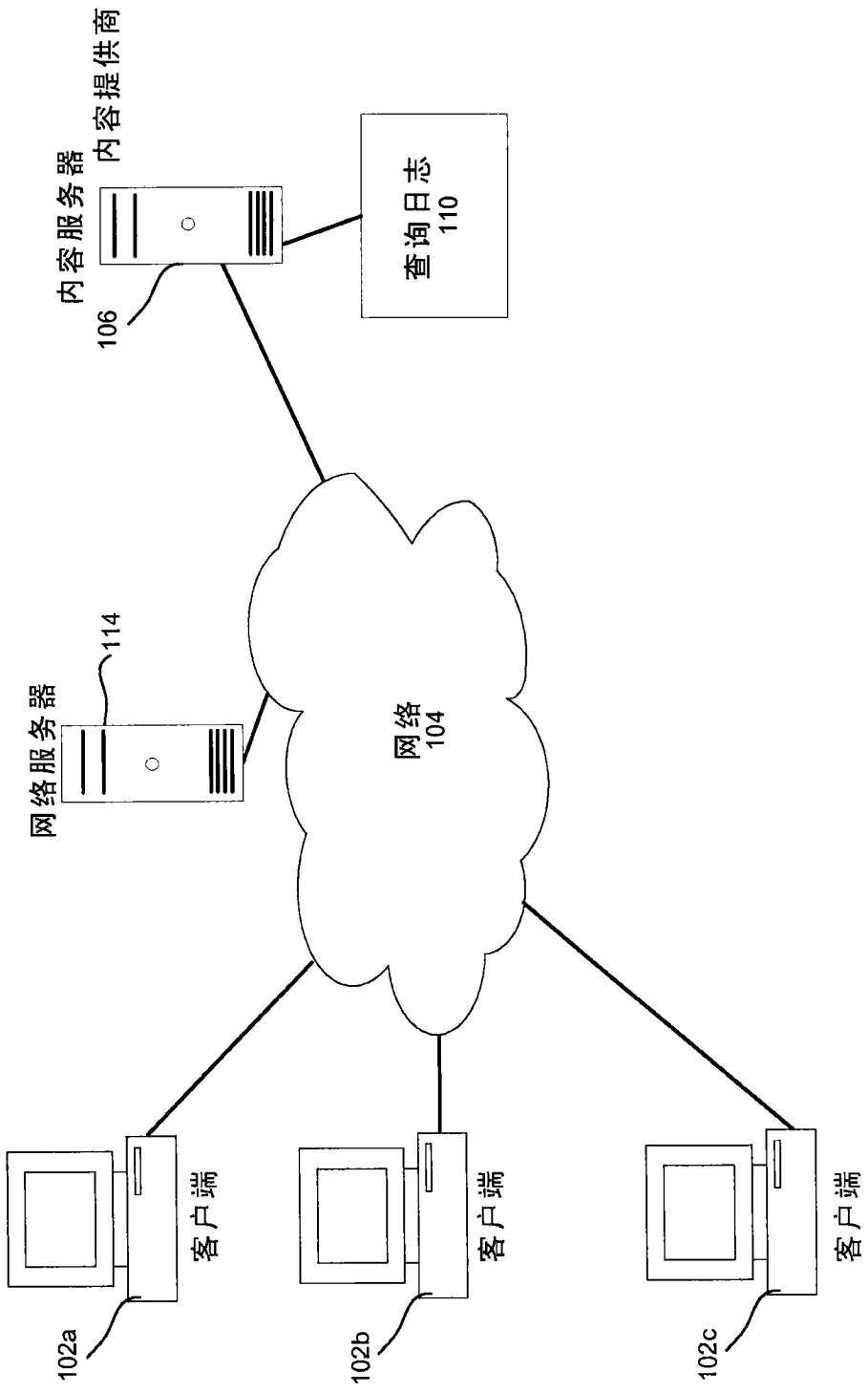

内容服务器
内容提供商
106
查询日志
110
网络服务器
114
网络
104
客户端
102a
客户端
102b
客户端
102c

图 1

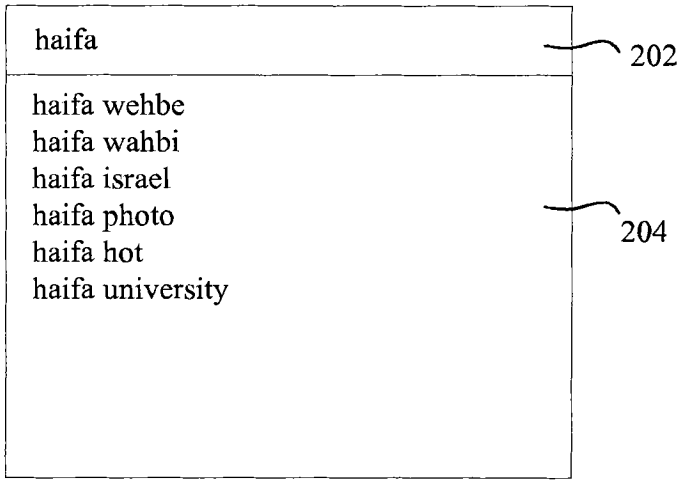

haifa
202
haifa wehbe
haifa wahbi
haifa israel
haifa photo
haifa hot
haifa university
204

图 2A

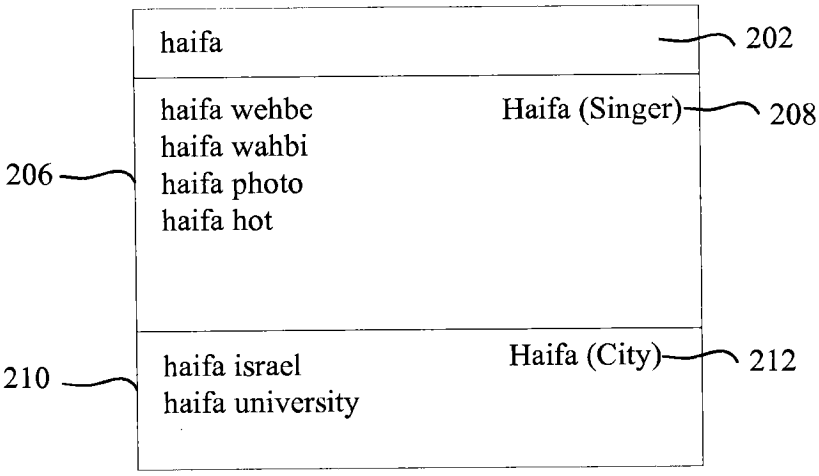

haifa
202
haifa wehbe
haifa wahbi
haifa photo
haifa hot
206
Haifa (Singer)
208
haifa israel
haifa university
210
Haifa (City)
212

图 2B

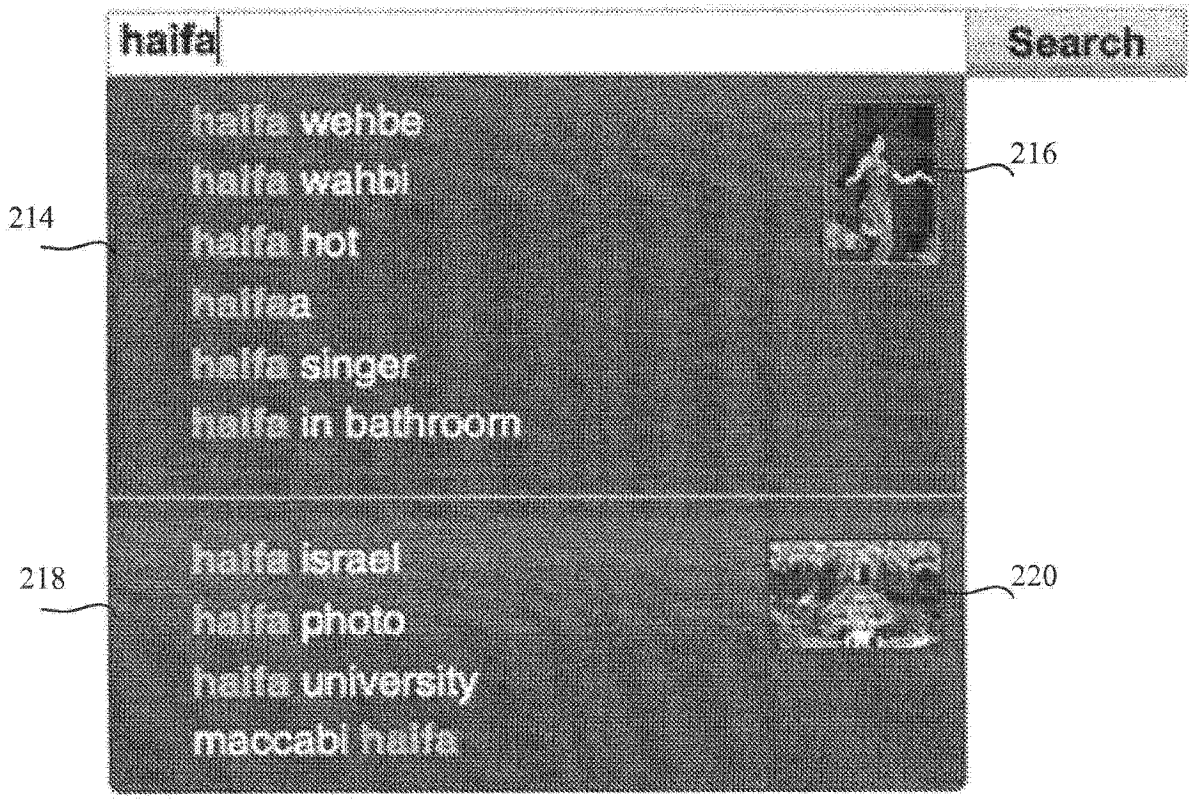

haifa
Search
haifa wehbe
haifa wahbi
haifa hot
haifaa
haifa singer
haifa in bathroom
haifa israel
haifa photo
haifa university
maccabi haifa
214
216
218
220

图 2C

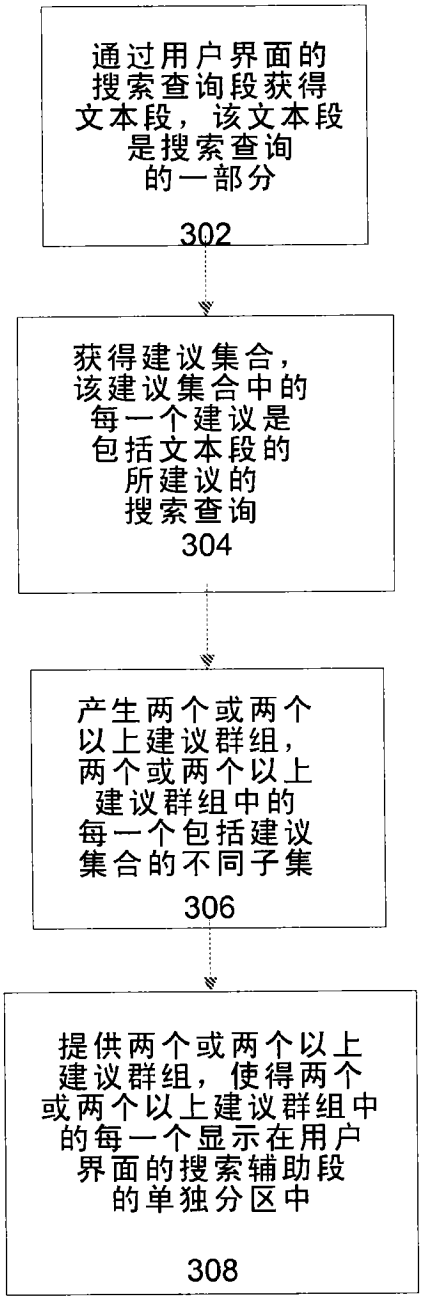

通过用户界面的搜索查询段获得文本段，该文本段是搜索查询的一部分
302
获得建议集合，该建议集合中的每一个建议是包括文本段的所建议的搜索查询
304
产生两个或两个以上建议群组，两个或两个以上建议群组中的每一个包括建议集合的不同子集
306
提供两个或两个以上建议群组，使得两个或两个以上建议群组中的每一个显示在用户界面的搜索辅助段的单独分区中
308

图 3

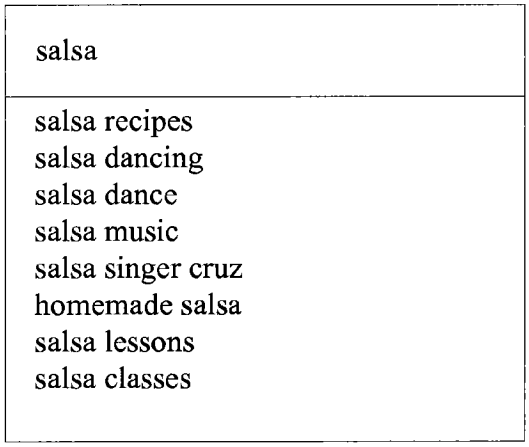

salsa
salsa recipes
salsa dancing
salsa dance
salsa music
salsa singer cruz
homemade salsa
salsa lessons
salsa classes

图 4

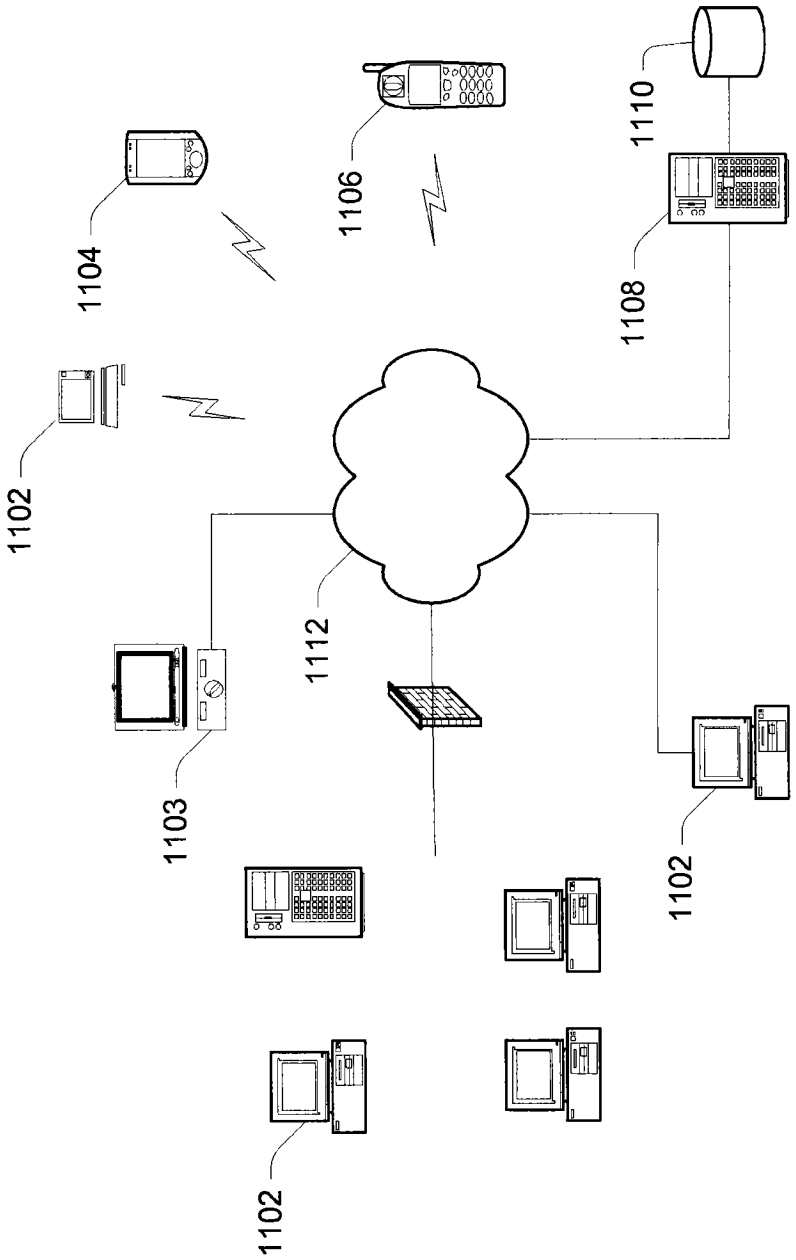


图 5

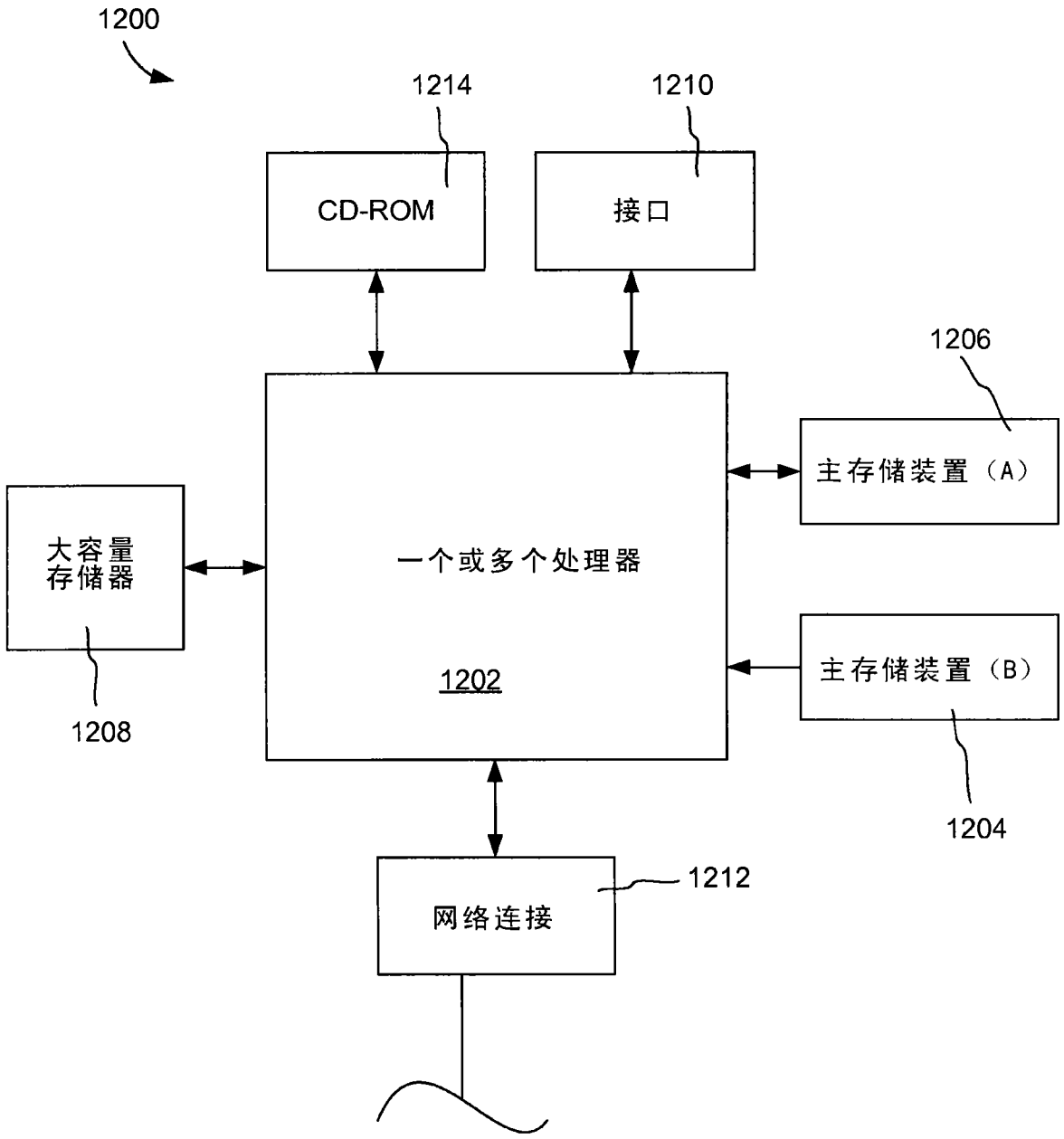

1200
1214
1210
CD-ROM
接口
1206
主存储装置（A）
大容量
存储器
一个或多个处理器
1202
主存储装置（B）
1208
1204
1212
网络连接

图 6