



(12) 发明专利

(10) 授权公告号 CN 1894689 B

(45) 授权公告日 2010. 11. 03

(21) 申请号 200480029767. 1

代理人 黄威 张金海

(22) 申请日 2004. 08. 27

(51) Int. Cl.

(30) 优先权数据

G06F 17/30(2006. 01)

10/650, 684 2003. 08. 29 US

10/778, 390 2004. 02. 17 US

(56) 对比文件

(85) PCT申请进入国家阶段日

US 2004/0010518 A1, 2004. 01. 15, 说明书
0066-0071 段.

2006. 04. 10

US 6347317 B1, 2002. 02. 12, 全文.

(86) PCT申请的申请数据

US 5826261 A, 1998. 10. 20, 全文.

PCT/CA2004/001540 2004. 08. 27

US 2003/0101413 A1, 2003. 05. 29, 说明书
0045 段.

(87) PCT申请的公布数据

W02005/022402 EN 2005. 03. 10

US 6178419 B1, 2001. 01. 23, 说明书第5栏.

CN 1378674 A, 2002. 11. 06, 说明书第6页,
附图9.

(73) 专利权人 伏泰劳普蒂克斯有限公司

地址 美国内华达州

US 5913208 A, 1999. 06. 15, 全文.

US 2003/0041054 A1, 2003. 02. 27, 说明书
0025-0033 段, 附图1-3.

(72) 发明人 戴维·B·戈斯 容霍·金

贾森·P·尼茨曼 珍妮弗·L·戈斯

季姆·D·法音德尔

贾斯廷·R·尼茨曼

迈克尔·T·温特斯

审查员 吴少鸿

(74) 专利代理机构 北京金信立方知识产权代理

有限公司 11225

权利要求书 3 页 说明书 16 页 附图 9 页

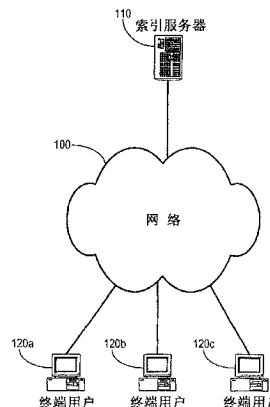
(54) 发明名称

用于查询和显示搜索结果的方法、装置及软件

(57) 摘要

本发明公开了一种用于响应终端用户的查询而显示搜索结果的方法, 装置及软件。搜索结果由来自多个索引的结果组合而成, 每个搜索结果具有相关联的关键字域。对所述多个索引中每个索引的索引条目使用特定索引搜索算法进行查询, 以获得每个索引的匹配搜索结果组, 每个匹配搜索结果具有特定于其索引的匹配质量。确定所述多个索引中每个索引的相对优先级, 并基于该确定的优先级将来自该多个索引的匹配搜索结果组合入经排序的搜索结果合并列表。来自较低优先级索引的搜索结果将被抛弃, 这样有利于任何来自较高优先级索引的匹配搜索结果。

CN 1894689 B



1. 一种计算机执行的响应终端用户查询而显示匹配搜索结果的方法,其中,所述匹配搜索结果是由来自多个索引的结果结合而成,每一个所述匹配搜索结果具有相关联的关键字域,该方法包含:

(i) 使用特定索引搜索算法对所述多个索引中的每一个索引的索引条目进行查询,以获取每个索引的匹配搜索结果组,每个匹配搜索结果具有特定于其索引的匹配质量;

(ii) 确定由网站管理员为所述多个索引中每一个索引分配的相对优先级,其中分配的相对优先级用于从所述多个索引中的被分配了该相对优先级的那个索引得到的所有匹配搜索结果;

(iii) 基于由所述网站管理员分配的所述多个索引中每一个索引的所述相对优先级,将得自所述多个索引的所述匹配搜索结果结合入包括从每个所述多个索引得到的匹配搜索结果的经排序的匹配搜索结果的合并列表,其中任何来自其相关联关键字域与高优先级索引中的匹配搜索结果的相关联关键字域相同的优先级较低的索引的匹配搜索结果都将被抛弃,以利于来自所述高优先级索引的所述匹配搜索结果。

2. 如权利要求 1 所述的方法,进一步包含:确定至少一个分配给所述多个索引中第一索引的每个索引条目的关键字,并将所述至少一个关键字与所述的终端用户查询相匹配,以获得所述第一索引的所述匹配搜索结果组。

3. 如权利要求 2 所述的方法,进一步包含:确定与每个所述至少一个关键字相关联的权重,并通过将与所述终端用户查询相匹配的任何所述关键字的所述权重相加来计算所述匹配质量。

4. 如权利要求 1 所述的方法,进一步包含:确定分配给所述多个索引的第一索引中各个索引条目的关键短语,并将所述关键短语与所述终端用户查询相匹配,以获得所述第一索引的所述匹配搜索结果组。

5. 如权利要求 4 所述的方法,进一步包含:确定所述的终端用户查询的当前时间是否落在分配给所述第一索引的每个所述索引条目的有效时间范围内,以获得所述第一索引的所述匹配搜索结果组。

6. 如权利要求 4 所述的方法,进一步包含:在所述的经排序的匹配搜索结果的所述合并列表中确定分配给所述第一索引的所述匹配搜索结果组中每个匹配搜索结果的相对放置位置,并在所述相对放置位置中放置一个与来自所述第一索引的每个所述匹配搜索结果相关联的值。

7. 如权利要求 6 所述的方法,其中所述的值包含 URL。

8. 如权利要求 4 所述的方法,进一步包含:确定分配给来自所述第一索引的所述匹配搜索结果组中的匹配搜索结果的共享组,每个共享组已在经排序的匹配搜索结果的所述合并列表中被分配了相对放置位置的范围,并在所述相对放置位置范围中放置了与来自所述第一索引的所述匹配搜索结果相关联的值。

9. 如权利要求 8 所述的方法,其中所述的值包含 URL。

10. 如权利要求 8 所述的方法,进一步包含:以连续的顺序显示放置在所述相对放置位置范围中的值。

11. 一种响应终端用户查询而显示匹配搜索结果的计算机装置,其中,所述匹配搜索结果是由结合单元将来自多个索引的结果进行结合而成,且每一个所述匹配搜索结果具有相

关联的关键字域,该计算机装置包括:

计算机可读存储器,其用于存储多个索引及来自该多个索引的结果,每个索引包含多个索引条目;

查询单元,其用于使用特定索引搜索算法对所述计算机可读存储器中存储的所述多个索引中的每一个索引的索引条目进行查询,以获取每个索引的匹配搜索结果组,每个匹配搜索结果具有特定于其索引的匹配质量;

优先级确定单元,其用于确定由网站管理员为所述多个索引中每一个索引分配的相对优先级,其中分配的相对优先级用于从所述多个索引中的被分配了该相对优先级的那个索引得到的所有匹配搜索结果;

所述结合单元,其用于基于由所述网站管理员分配的所述多个索引中每一个索引的所述相对优先级,将得自所述多个索引的匹配搜索结果结合入包括从每个所述多个索引得到的匹配搜索结果的经排序的匹配搜索结果的合并列表,其中任何来自其相关联关键字域与高优先级索引中的匹配搜索结果的相关联关键字域相同的优先级较低的索引的匹配搜索结果都将被抛弃,以利于来自高优先级索引的匹配搜索结果。

12. 如权利要求 11 所述的计算机装置,其中所述结合单元还包括第一匹配单元,其用于确定至少一个分配给所述多个索引中第一索引的每个索引条目的关键字,并将所述至少一个关键字与所述终端用户查询相匹配,以获得所述第一索引的匹配搜索结果组。

13. 如权利要求 12 所述的计算机装置,其中所述结合单元还包括匹配质量计算单元,其用于确定与每个所述至少一个关键字相关联的权重,并通过将与所述终端用户查询相匹配的任何所述关键字的所述权重相加来计算所述匹配质量。

14. 如权利要求 11 所述的计算机装置,其中所述结合单元还包括第二匹配单元,其用于确定分配给所述多个索引中第一索引的每个索引条目的关键短语,并将所述关键短语与所述终端用户查询相匹配,以获得所述第一索引的匹配搜索结果组。

15. 如权利要求 14 所述的计算机装置,其中所述结合单元还包括查询时间确定单元,其用于确定所述终端用户查询的当前时间是否落在分配给所述第一索引的每个所述索引条目的有效时间范围内,以获得所述第一索引的所述匹配搜索结果组。

16. 如权利要求 14 所述的计算机装置,其中所述结合单元还包括第一放置单元,其用于在所述的经排序的匹配搜索结果的所述合并列表中确定分配给所述第一索引的所述匹配结果组中每个匹配搜索结果的相对放置位置,并在所述相对放置位置中放置与来自所述第一索引的每个所述匹配搜索结果相关联的值。

17. 如权利要求 16 所述的计算机装置,其中所述第一放置单元在所述相对放置位置中放置的所述值包含 URL。

18. 如权利要求 14 所述的计算机装置,其中所述结合单元还包括第二放置单元,其用于确定分配给来自所述第一索引的所述匹配搜索结果组中的匹配搜索结果的共享组,每个共享组已在所述经排序的匹配搜索结果的合并列表中分配了相对放置位置的范围,并在所述相对放置位置的范围中放置与来自所述第一索引的所述匹配搜索结果相关联的值。

19. 如权利要求 18 所述的计算机装置,其中所述第二放置单元在所述相对放置位置中放置的所述值包含 URL。

20. 如权利要求 18 所述的计算机装置,其中所述第二放置单元构造为以连续的顺序显

示放置在所述相对放置位置范围中的值。

用于查询和显示搜索结果的方法、装置及软件

[0001] 相关申请

[0002] 本申请是题为“用于查询和显示搜索结果的方法、装置及软件”，序列号为 10/650,684，申请日为 2003 年 8 月 29 日的同时待决的美国专利申请的持续部分申请，其内容在此作为参考。

技术领域

[0003] 本发明涉及搜索引擎及索引，特别涉及一种对从多个索引中获得的搜索结果进行查询和显示的方法、装置及软件。

背景技术

[0004] 因特网和万维网（“web”）的迅速成长导致用于为数亿可用网页编制索引的网络搜索引擎不断增多。众所周知，网络是使用超文本传输协议（“HTTP”）的超文本信息系统及通信系统。HTTP 容许用户访问到这些网页，这些网页可以使用在其它事物中称为超文本链接标示语言（“HTML”）的标准页面描述语言。HTTP 还可用于访问许多不同格式的文件和其他数据，包括文本文件，图像文件，可执行文件，数据文件和其他数据源。典型地，HTTP 通过为公众所熟知的统一资源定位器（“URL”）的寻址模式访问这些网页，文件和数据。

[0005] 通过确定 URL，终端用户能够实质上地访问任何可取得的网络资源（文本文件、图像文件、可执行文件、数据文件和其他数据源），这些资源可得自接入因特网或其它网络的网络服务器。

[0006] 然而，若终端用户对 URL 所知不多，就必须依赖可以搜索网络索引或目录以定位相应网站 URLs 的网络搜索引擎。

[0007] 当某些搜索引擎雄心勃勃地尝试对整个网络的各个重要部分编索引时，其他搜索引擎的注意力可能集中在更为特殊的目标上，如特定的“纵向市场”。针对某个特定“纵向市场”的搜索引擎可能只将该市场中消费者感兴趣的站点挑选出来编为索引。此外，该搜索引擎还可对为将所述搜索引擎主机化的站点准备的内容编索引。最后，该搜索引擎可为其它站点编索引。通过限制编入索引的站点，所述搜索引擎对于对该纵向市场感兴趣的消费者来说具有更大的价值。结果，了解到该搜索引擎可以提供集中的所感兴趣的搜索结果，那么该纵向市场中的消费者可能就会被该站点吸引。这种吸引力当然可被充分利用来产生广告收益及类似的收益。

[0008] 举例说明，某个所述的纵向市场可能是在特定地理区域中提供高尔夫工具和服务的市场。想要吸引终端用户来访问的特定的地区性高尔夫网站可能希望为终端用户提供强大的搜索能力，使得这些终端用户可搜索的不仅是该地区性的高尔夫网站中的内容，还有该地理区域中的其他高尔夫相关站点，以及通常在网络中可得到的高尔夫站点。而且，该地区性高尔夫网站可能还进一步希望从他们的诸如地区性的高尔夫课程及高尔夫工具制造商的广告客户那里得到收益，方法是将他们的网络资源的 URL 列在搜索结果列表中。同时，该地区性高尔夫网站可能希望避免显示任何不适宜的搜索结果，如某个广告客户的主要竞

争对手经营的网站的 URL。可见,在这个示例中,搜索结果是从多个搜索资源中获得的。

[0009] 现有的从多个搜索资源中查询并选择性地显示搜索结果的设计方案已在灵活性方面受到限制,特别是当与搜索资源相应的多个索引具有完全不同的结构和搜索条件时。有必要寻求更为灵活地从多个搜索资源中合并和显示搜索结果的方法。

发明内容

[0010] 本发明提供了一种基于终端用户的特定查询,查询和显示从多个索引中获得的搜索结果的方法、装置及软件。在某个实施例中,通过对每个索引使用特定索引搜索算法而获得搜索结果。这些特定索引搜索算法在某种意义上是模块化的,其中每个搜索算法都可被更改或者个别地替换。根据分配给每个索引的相应优先级,使用具有合并法则的合并算法以便将获取自多个索引的搜索结果进行合并。在合并过程中识别出来的任何同样的搜索结果(也即,具有同样相关联的关键字域的搜索结果)都可通过合并法则进行处理,例如,丢弃产生自优先级较低的索引的搜索结果。

[0011] 举例说明,广告客户索引,本地或私人索引和公共索引的每个都可运用特定索引搜索算法各自进行搜索。使用特定索引搜索算法搜索这些索引获得的搜索结果可随后根据优先级进行合并和排序。例如,可以给所述的广告客户索引分配一个比所述个人索引或公共索引较高的优先级,而给所述个人索引分配一个比所述公共索引较高的优先级。

[0012] 某个来自一个或多个支持明确等级的索引的搜索结果可被置于经合并的搜索结果的顺序中的某个特定的相对放置位置或者等级。在合并的过程中,某个 URL 的明确等级得到保留以便在显示给终端用户的搜索结果列表中,确保该 URL 置于某个相对的放置位置或该位置之上。在一个特定等级中的放置也可以受时间的限制(例如,通过确定终端用户的查询时间是否落在预设的时间范围内,如日期范围)。

[0013] 许多搜索结果可被置于由共享组定义的相对放置位置或者等级的范围内。在任何设定的时间,该共享组中的全体成员可依赖于为广告客户索引中的每个索引条目指定的时间范围(如日期范围)来改变。

[0014] 本发明的一方面提供了一种根据某个终端用户的查询而显示搜索结果的方法,所述搜索结果是由来自多个索引的结果组合而成,每一个搜索结果具有相关联的关键字域,该方法包含:(i) 使用特定索引搜索算法查询每个所述多个索引的索引条目,以获取对于每个索引的一组匹配搜索结果,每个匹配搜索结果具有特定于其索引的匹配质量;(ii) 确定所述多个索引中每一个索引的相对优先级;(iii) 基于所述多个索引中每一个索引的相对优先级,将得自所述多个索引的匹配搜索结果组合入经排序的搜索结果的合并列表中,其中任何来自其关联关键字域与高优先级索引中的匹配搜索结果的关联关键字域相同的优先级较低的索引的搜索结果都将被抛弃,这样有利于来自高优先级索引的匹配搜索结果。

[0015] 本发明的另一方面提供了一种根据终端用户的查询而显示搜索结果的方法,所述搜索结果是由来自至少第一和第二索引的结果组合而成,每一个搜索结果都具有相关联的关键字域,该方法包含:(i) 使用特殊索引算法对第一索引中的索引条目进行查询以获得匹配搜索结果,并确定至少一个与每个匹配搜索结果相关的相对放置位置;(ii) 使用特殊索引算法对第二索引中的索引条目进行查询以获得匹配搜索结果;(iii) 将得自第一和第

二索引的匹配搜索结果组合入经排序合并的搜索结果列表中,其中每个来自第一索引的匹配搜索结果被置于所述相对的放置位置,而且任何来自其关联关键字域与第一索引中的匹配搜索结果的关联关键字域相同的第二索引的搜索结果被抛弃,这样有利于来自第一索引的匹配搜索结果。

[0016] 本发明的再一方面提供了一种计算机装置,其包含处理器和计算机可读存储器,该存储器中存储了多个索引,每个索引包含多个索引条目,该多个索引中的每个索引条目都具有相关联的关键字域,还包含了搜索引擎软件,可使得该装置适于:(i) 使用特定索引搜索算法查询每个所述多个索引的索引条目,以获取对于每个索引的一组匹配搜索结果,每个匹配搜索结果具有特定于其索引的匹配质量;(ii) 确定所述多个索引中每一个索引的相对优先级;(iii) 基于所述多个索引中每一个索引的相对优先级,将得自所述多个索引的匹配搜索结果组合入经排序的搜索结果的合并列表中,其中任何来自其关联关键字域与高优先级索引中的匹配搜索结果的关联关键字域相同的优先级较低的索引的搜索结果都将被抛弃,这样有利于来自高优先级索引的匹配搜索结果。

[0017] 本发明的又一方面提供了一种计算机可读介质,其上存储了计算机可执行指令,当这些指令载入计算机装置(该计算机装置包含处理器和处理器可读存储器,该存储器中存储了多个索引,每个索引包含多个具有相关联的关键字域的索引条目),可使得该装置适于:(i) 使用特定索引搜索算法查询每个所述多个索引的索引条目,以获取对于每个索引的一组匹配搜索结果,每个匹配搜索结果具有特定于其索引的匹配质量;(ii) 确定所述多个索引中每一个索引的相对优先级;(iii) 基于所述多个索引中每一个索引的相对优先级,将得自所述多个索引的匹配搜索结果组合入经排序的搜索结果的合并列表中,其中任何来自其关联关键字域与高优先级索引中的匹配搜索结果的关联关键字域相同的优先级较低的索引的搜索结果都将被抛弃,这样有利于来自高优先级索引的匹配搜索结果。

[0018] 通过查阅下述的结合附图对本发明特定实施例的说明,本技术领域中的普通技术人员可清楚地了解本发明的其他方面及特征。

附图说明

[0019] 下列附图对本发明的典型实施例进行了举例说明。

[0020] 图 1A 为一个与本发明实施例的典型的索引服务器相互连接的典型的数据通信网络的简化示意图,该网络与多个计算机装置进行通信;

[0021] 图 1B 为图 1A 中的索引服务器的硬件结构的简化示意结构图;

[0022] 图 2A 为说明图 1A 和图 1B 中索引服务器的软件和数据组成的逻辑结构图;

[0023] 图 2B 为说明图 2A 中个人索引的典型的数据库图解的示意结构图;

[0024] 图 2C 为与图 2B 中数据库图解对应的关键字/权重数据库表的示例;

[0025] 图 2D 为与图 2B 中数据库图解对应的 URL 数据库表的示例;

[0026] 图 2E 为说明图 2A 中广告客户索引的典型的数据库图解的示意结构图;

[0027] 图 2F 为与图 2E 中数据库图解对应的“合同项目”数据库表的示例;

[0028] 图 2G 为与图 2E 中数据库图解对应的“关键短语”数据库表的示例;

[0029] 图 2H 为与图 2E 中数据库图解对应的“共享组”数据库表的示例;

[0030] 图 3A 是为了在图 2C 和图 2D 的数据库中创建记录而选取与 URLs 相关的关键字以

及为 URLs 分配权重的典型步骤的示意图；

[0031] 图 3B 为索引服务器为了响应某个查询请求而对多个索引进行查询所执行的典型步骤的示意图；

[0032] 图 4 示意性地说明涉及图 2A 的个人索引和公共索引的范例查询所获得的经合并的搜索结果；

[0033] 图 5 示意性地说明了当图 4 中的范例查询进一步包括由图 2A 中广告客户索引得到的搜索结果时，所获得的经合并的搜索结果。

具体实施方式

[0034] 图 1A 所示为与本发明实施例的典型的索引服务器 110 相互连接的典型的数据通信网络 100，该网络与多个计算机装置 120a, 120b 和 120c (单个的和共同的装置 120) 进行通信；

[0035] 计算机装置 120 和索引服务器 110 都是常规的计算机装置，均包括处理器和计算机可读存储器，该可读存储器中存储了操作系统和应用软件以及执行部件。

[0036] 数据通信网络 100 可以为，例如常规局域网，其遵守合适的网络协议如以太网，令牌网 (token ring) 或者类似协议。可选地，该网络协议可以遵从高级协议如网际协议 (IP) , Appletalk 或者 IPX 协议。类似地，网络 100 可以是广域网或者公众网。

[0037] 客户计算机装置 120 为可连接到网络的计算机装置，其为终端用户提供可以查看存储在索引服务器 110 上的信息的接口。计算机装置 120 可以是，例如，常规的基于视窗操作系统的计算机装置，其上可存储以及可执行相应的 HTML 浏览器，如微软的视窗浏览器，Netscape 的 Navigator 浏览器或类似的浏览器。

[0038] 从下文可知，索引服务器 110 存储了网络索引信息，并可以存储使得装置 120 搜索已存储的所述索引信息的软件。

[0039] 图 1B 为示例的索引服务器 110 的优选硬件结构的简化示意图。在图示的实施例中，索引服务器 110 为常规的可连入网络的服务器。索引服务器 110 可以是，例如，基于英特尔 X86 的计算机，该计算机为基于微软 NT, 苹果或 UNIX 操作系统的服务器，工作站，个人计算机等。示例的索引服务器 110 包括处理器 112, 其与计算机存储器 114 进行通信；网络接口 116；输入输出接口 118；以及视频适配器 122。最好是，索引服务器 110 可选择性地包括与适配器 122 相互连接的显示器 124；输入输出装置，如键盘 126, 磁盘驱动器 128, 以及鼠标 130 等。处理器 112 为典型的常规中央处理单元，并可以是，例如英特尔 X86 家族的微处理器。当然，处理器 112 可以是任何其它的为该领域的技术人员所知的适当的处理器。计算机存储器 114 包括了合适的随机存取存储器，只读存储器和磁盘存储器这三者的组合，其由处理器 112 用于存储并执行适于处理器 112 以本发明所示范的方式实现功能的软件程序。磁盘驱动器 128 能将数据写入或读出计算机可读介质 132, 该可读介质可用于存储本发明所示实施例中的可载入存储器 114 的软件和数据。计算机可读介质 132 可以是 CD-ROM, 磁盘, 磁带, ROM- 磁带等等。网络接口 116 为适于将服务器 110 与网络 100 进行物理连接的任意接口。接口 116 可以是例如以太网, ATM, ISDN 接口或者调制解调器等, 可用于使数据出入于网络 100 或者另外的适当通信网络。

[0040] 计算机装置 120 的硬件结构在本质上与索引服务器 110 的相类似，在此不再详述。

[0041] 图 2A 为说明服务器 110 的软件和数据组成的逻辑结构图。如图所示,服务器 110 可以寄存多个索引,每个索引具有与之相关联的特定索引搜索算法。服务器 110 也可寄存带有合并法则的合并算法,其可用于例如,根据分配给每个索引的相应优先级将多个索引中每个索引的搜索结果进行合并。

[0042] 作为一个特例,可考虑某特定纵向市场网站的情况,如前文介绍的区域性高尔夫网站。这样一个网站可拥有一个寄存多个索引的服务器 110,包括广告客户索引 211,个人索引 212 和公共索引 213。

[0043] 广告客户索引 211 可包含广告合同的条目。每个广告合同可能涉及特定的广告客户,该客户希望在某个经排序的搜索结果列表中特定的相对放置位置或等级处,放置包含该客户的 URL 的“广告”。该合同可指定必需匹配的特定关键短语,以及该合同的有效日期范围。例如,该广告客户的索引 211 可包含各种区域性高尔夫课程以及高尔夫相关产品制造商的索引条目,以保证在由日期范围(如起始日期与终止日期)界定的特定广告时段内,他们的网站 URL 能够出现在终端用户关于区域性高尔夫网站的查询结果中。

[0044] 个人索引 212 可包含各种对于终端用户搜索所述区域性高尔夫网站可用的 URLs 的条目。例如,个人索引 212 的索引条目可包含所述区域性高尔夫网站内容的 URLs,其它的区域性高尔夫网站的 URLs,以及可能包含一个或多个在广告客户索引 211 中也可找到的 URLs 的条目。

[0045] 公共索引 213 可包含由第三方编为索引的 URLs(如开放的目录数据库 DMOZ- 可用 URL “http://www.dmoz.org” 访问的开放目录项目)。

[0046] 索引服务器 110 可进一步寄存搜索引擎软件 214,终端用户接口 215,以及管理员接口 216。

[0047] 如下文所述,装置 120 的某个终端用户可借助与终端用户接口 215 之间的通信,通过网络 100 访问搜索引擎软件 214。该搜索引擎 214 本身可由一个或多个存储在存储器中并可由索引服务器 110 中的处理器执行的软件模块来体现。例如,终端用户接口 215 可以提供搜索逻辑单元以及接受嵌入在 HTTP GET/POST 请求中表现为“名字=值”对的搜索请求。

[0048] 网站管理员,例如,作为上述示例的区域高尔夫网站的代表,可以通过管理员接口 216 访问广告客户索引 211 和个人索引 212。该管理员可修改索引 211 中的广告客户记录和个人索引 212 中的记录,但是该管理员不能修改公共索引 213 中的记录。

[0049] 搜索引擎软件 214 可以使用搜索算法 221,222 和 223,每个搜索算法对于索引 211,212 和 213 之一都是特定的,并且定义了对于相关联的索引 211,212,213 上的搜索如何被执行。因此,每个索引 211,212,213 是根据它们自己的特定索引搜索算法 221,222,223 进行搜索的。这些搜索算法 221,222,223 在某种程度上是模块化的,其中每个搜索算法 221,222,223 都可被更改或者个别地替换。这种模块性在从可能在结构上有着根本差异的多个索引 211,212,213 中获取期望的搜索结果方面提供了极大的灵活性。在对于纵向市场网站的搜索引擎的情况下,这种灵活性使得网站的管理员可以从多个与纵向市场相关的完全不同的索引中合并及显示搜索结果。

[0050] 如下文所要详述的,合并算法 225 可用于合并对于索引 211,212,213 分别使用特定索引搜索算法 221,222 和 223 得到的搜索结果。

[0051] 需要理解的是本说明书中所用到的术语“管理员”可指任何承担管理一个或多个索引 211, 212, 213, 搜索引擎软件 214, 相关联的搜索算法 221, 222, 223, 或者合并算法 225 的任务的个人。

[0052] 在公开的实施例中, 索引 211, 212 和 213 存储为一个或多个相关的数据库。图 2B 为图 2A 中个人索引 212 的示例的数据库图解 230 的示意结构图。在索引 212 中, 用相关联的一个或多个加权的关键字将网页 URL 编为索引。表 240 的多个记录定义了与网页 URL (HASH_URL) 相关联的关键字 (KEYWORD_HASH) 和权重 (KEYWORD_WEIGHT)。关键字 / 权重表 232 中的 UID 为物理性的主关键字, 用于唯一地识别每个记录。如图解 230 所示, 数据库存储索引 212 可包括关键字 / 权重表 232 和相应的 URL 表 234。

[0053] 图 2C 所示为与数据库图解 230 (图 2B) 对应的关键字 / 权重数据库表 240。更特定地, 表 240 包括多个记录 240a-240c, 每个都包含多个诸如具有特定 URL 的当地的高尔夫课程网站的特定搜索条目的域 232a-232e。为说明起见在此选用任意值, 记录 240a-240c 各自包含: 关键字哈希 (hash) 值“72”, “73”, “74”; 权重 100, 70, 90; 字面的关键字“KW1”, “KW2”, “KW3”; 以及哈希 (hash) URL 值“12”, “12”, “12”。因此, 在这个特例里, 具有哈希 (hash) 值“12”的高尔夫课程 URL 的索引条目包括三个具有相应权重 100, 70 和 90 的关键字“KW1”, “KW2”, “KW3”。

[0054] 图 2D 为与图 2B 中 URL 表 234 对应的索引 212 的数据库表 250 的示例。更特定地, 数据库表 250 包括多个记录 250a-250c, 每个都包含多个字域 234a-234e。表 250 的每个记录都提供了经编索引的 URL 的详细信息。为说明起见在此选用任意值, 数据库 250 的记录 250a-250c 各自包含: 哈希 (hash) URL “12”, “13”, “14”; 对应的各种 URL 地址, 如 www.golf1.foo, www.golf2.foo, www.golf3.foo; 标题“高尔夫一”, “高尔夫二”, “高尔夫三”; 描述“高尔夫一的主页”, “高尔夫二的主页”, “高尔夫三的主页”; 以及相应的日期 / 时间印记。虽然这个简化的例子中的 URL 地址涉及到三个高尔夫相关的网站的主页, 不过一般来说, 须理解这些 URLs 实质上可定位任何类型的网络资源, 该网络资源可类似地在表 250 中编为索引。因此, 更普遍地, 这些 URL 地址或值在后面对图 4 中例子的讨论中被简单地引用为 URL_A, URL_B, URL_C 等等。

[0055] 图 2E 为图 2A 中广告客户索引 211 的示例的数据库图解的示意结构图。更特定地, 在这个说明性的例子里, 数据库图解 260 包含“合同项目”表 262, “关键短语”表 264, “共享组”表 266, “广告客户信息”表 267 以及“网络位置”表 268 (网络位置表 268 在功能上与图 2B 中的 URL 表 234 相对应, 事实上可使用如图 2C 和图 2D 所示的相同的 URL 哈希 (hash) 值, 对这些相同的 URL 进行引用)。需理解数据库图解 260 只是说明性的, 且数据库图解可以许多其他的形式存在。

[0056] 图 2F 为与图 2E 中的“合同项目”表 262 相对应的衍生数据库表 270 的示例。为说明起见, 仅对来自“合同项目”表 262 的属性 262a-262n 图示于数据库表 270: 合同标志符 262a (CONTRACT_ITEM_ID), 搜索结果顺序的等级 262b (RANK), 该等级的标价 262c (PRICE), 起始日期 262d (START_DATE), 终止日期 262e (END_DATE), 可能与该合同项目相关的各种其他日期 (DATE_ENDED, DATE_CREATED, DATE_MODIFIED), “共享组” 262i (GROUP_ID), URL 哈希 (hash) 262j (HASH_URL), 关键短语的关键短语哈希 (hash) 值 262m (HASHED_PHRASE), 以及相应的广告客户信息 262n (AD_SITE_ID)。在该示例里, 所述哈希 (hash) URL 262j 可用于例

如以与上述图 2C 和图 2D 中哈希 (hash)URL232e/234a 实质上类似的方式指向 URL 值。需理解,用于广告合同的 URL 可在实质上指向任何经 URL 就可达到的网络资源。在本示例中,广告客户标志符 262n 识别广告客户(如通过对图 2E 的表 267 的交叉引用),并与识别“广告”URL 的哈希 (hash)URL262j 无关。

[0057] 因此,例如,通过提供必要信息来完成数据库表 270 中所有相关域(如 262a-262n),某区域性高尔夫课程可与所述区域性高尔夫网站的管理人员订立合同并进入广告客户索引 211。该广告合同可指定,例如,该高尔夫课程的 URL 值将在预定的时间范围内(如由起始日期 262d 和终止日期 262e 指定)出现在该区域型高尔夫网站的终端用户使用关键词“高尔夫”所进行的搜索的等级 1 中。

[0058] 在本示例的上下文中,所述“关键词”(对应于关键词哈希 (hash) 值 262m) 包含与终端用户输入的查询相对应的可搜索的字符串或单元。特定的关键词可与广告客户索引 211 的每个索引条目存储在一起,以便为了广告能够以特定的等级显示,与该关键词准确的匹配可以被请求。该关键词的使用允许广告客户为搜索结果顺序中的特定等级而签订合同。

[0059] 起始日期 262d 和终止日期 262e 定义了广告合同的有效时间范围。因而,用于查询广告客户索引 211 的查询表达式应该不仅包含必需匹配的“关键词”,还需包含查询的“当前日期”,该日期必须落在由起始日期 262d 和终止日期 262e 指定的时间限制内。这允许广告客户在特定的广告时间内为了搜索结果顺序中的特定等级或者相对的放置位置而签订合同。

[0060] 例如,在特定的广告时段内,该区域性高尔夫课程可以在对于关键词“高尔夫”的经排序的搜索结果列表中保留为等级 1,同时高尔夫球制造商在对于关键词“高尔夫球”的已排序的搜索结果列表中保留为等级 1。这样,如果终端用户输入查询“高尔夫”,而该查询又进行于该两个广告的广告合同都有效的特定日期,则由于该查询与所述关键词精确地匹配,故该区域性高尔夫课程的广告将以等级 1 出现。该高尔夫球制造商的广告将不会显示出来。相反地,如果终端用户输入查询“高尔夫球”,则有关该高尔夫球制造商的广告将以等级 1 出现,而该区域性高尔夫课程的广告将不会显示出来。

[0061] 可选地,诸如主要高尔夫商品制造商地广告客户,可能希望为许多个不同的关键词(如“高尔夫球”,“高尔夫球袋”,“高尔夫俱乐部”,“高尔夫球鞋”等)预定某一时间段内的等级,以便该广告客户的 URL 能以该等级出现在终端用户对这些关键词中任一查询的结果中。

[0062] 在本实施例中,应用了一个规则,即在任何假定的时刻,与某个假定关键词关联的两个广告不会被分配相同的等级。例如,如果广告客户 A 已经购买的广告合同中的关键词为“高尔夫球”、有效期为 2004 年 1 月 10 日到 2004 年 2 月 15 日,等级为 1,则广告客户 B 可以购买到的广告合同中的关键词为“高尔夫球”、有效期为 2004 年 2 月 16 日到 2004 年 3 月 1 日,等级为 1。但是,广告客户 B 不可以购买其关键词为“高尔夫球”、有效期为 2004 年 2 月 1 日到 2004 年 3 月 1 日,等级为 1 的广告合同,因为该日期范围与在先的广告合同的日期范围相重叠。

[0063] 图 2G 为与图 2E 的数据库图解相对应的“关键词”数据库表 280 的示例,该数据库表可以存储一个或多个这样的关键词。如图所示,关键词哈希 (hash) 值

264a (HASHED_PHRASE) 与合同项目表 270 中的关键短语哈希 (hash) 值 262m 相对应。

[0064] 图 2H 为与图 2E 的数据库图解相对应的“共享组”数据库表 290 的示例。如该名字所显示的,共享组可将两个或多个匹配的搜索结果作为组来进行处理。例如,在本示例中,当希望将与某个关键短语匹配的多个广告放置到为该共享组定义的等级范围内的时候,就可以构建一个共享组。这样可在例如广告客户 A 希望将一组相关的广告显示在经排序的搜索结果列表的特定范围中时较为有利。

[0065] 在该公开的实施例中,数据库表 270 中的等级 262b (RANK) 仅应用在未对给定合同的共享组 262i (GROUP_ID) 区段作定义时。如果该共享组 262i 已经定义,则查询时间内的合同的实际等级也得到确定。如果该共享组 262i 未经定义,则由等级 262b 确定放置位置。

[0066] 例如,如果存在某个关键短语的共享组 Y,且该共享组中最低等级 266b 为 10,最高等级 266c 为 5 (其中在经排序的列表中最高可能的等级为 1),在合同被生成时,不放在共享组 Y 里的新的合同项目不能被赋予任何介于 5-10 之间的等级。更确切地说,为了能以等级 5-10 之一出现,该合同项目必须置于共享组 Y 中。本公开的实施例中,在查询时,所述广告客户的索引搜索算法 221 将随机地为共享组 Y 中的每个合同项目分配等级。因此,对于共享组中的合同,没有使用等级 262b 域。

[0067] 在该公开的实施例中,多个广告客户也可可作为某个共享组的成员而共享一段等级范围。例如,共享组 Z 可为关键短语“高尔夫球”定义范围 11-15。下述合同可以是共享组 Z 的成员:有效期为 2004 年 1 月 1 日到 2004 年 2 月 1 日的合同 A;有效期为 2004 年 2 月 1 日到 2004 年 3 月 1 日的合同 B;有效期为 2004 年 1 月 15 日到 2004 年 2 月 15 日的合同 C。如果某个执行于 2004 年 2 月 1 日的查询使用了关键短语“高尔夫球”,则该三个合同 A, B 和 C 都可匹配,而相关联的广告的等级就被分在 11-15 的范围中。另一方面,如果查询执行于 2004 年 1 月 15 日,仅有合同 A 和 C 可匹配,而相关联的广告的等级就被分在 11-15 的范围中。注意该时间范围与每个合同有关,而与该共享组无关。因而,在任何假定的日期,一个查询会引起同一共享组中不同合同的组合。

[0068] 如图,为说明起见,共享组数据库表 290 可以有許多定义属性 266a-266d,包括标志符 266a (GROUP_ID),保证最低等级 266b (LOWEST_RANK),默认最高等级 266c (HIGHEST_RANK),以及与关键短语表 264 的哈希 (hash) 短语 264a 和合同项目表 262 的哈希 (hash) 短语 262m 相对应的哈希 (hash) 短语 266d (HASHED_PHRASE)。在本示例里,对于共享组 A,最低等级 266b 其值为 3,而最小的最高等级 266c 取值为 2 (这里,为说明起见,约定 3 为比 2 “低”的等级),表明了共享组中的广告保证被置于等级 2 或 3 中。

[0069] 最高等级 266c 被称为“最小的”最高等级,是由于有可能共享组 A 的最高等级实际上是作为甚至更高的最高等级显示给终端用户的。例如,可考虑当终端用户的查询结果中没有广告在等级 1 里,同时等级 2 和等级 3 又被与共享组 A 相关联的广告占用的情况。这样,虽然共享组 A 的保证最低等级为 3,同时共享组 A 的最小的最高等级为 2,但是在显示给终端用户的时候,由于等级 1 没有被占用,故共享组 A 的实际等级将为 1 或 2。这里,需要意识到的是能放置在所希望的放置位置是相对的,同时当显示给终端用户时的实际的放置位置可根据是否所有的放置位置都被占用而变化。

[0070] 需要了解的是,对于假定的共享组,上述保证最低等级 266b 和最高等级 266c 可定义任何合适的等级范围。这样,该共享组定义了一个连续的等级范围,以便任何置于共享组

中的广告都能连续地显示给终端用户。

[0071] 虽然没有给出特定的例子,图 2E 中的广告客户信息表 267 可以包含各种属性 267a-267h,其中包括特定广告客户的地址,帐单和状态信息,以及图 2E 中的“网络位置”表 268 可包含各种属性 268a-268g,其定义了包含“广告”的广告客户的 URL。可理解这些属性 268a-268g 仅用作说明。

[0072] 图 3A 是为了在个人索引 212(图 2A)的表 240 和 250 中创建记录而为 URLs 选择关键字以及分配权重的典型步骤 S300A 的流程图。对于那些该技术领域的技术人员,需了解到,步骤 S300A 可通过本发明典型实施例的计算机软件而体现,包括用合适的计算机语言编写的可读代码。因此,步骤 300A 可在本发明典型实施例中的软件的控制下通过服务器 200 而执行。

[0073] 如图示,在步骤 S302 中,从管理员处获取将要编索引的 URL。在步骤 S304 中,该 URL 已被获取。在步骤 S306 中,为了确定可用于对该网络资源编索引的可能的关键字,该 URL 的内容被分解与分析。例如,关键字可以根据其在网页中出现的频率而被确定,使用元标签(meta-tags)或任何其他的为那些普通技术人员所知的方法关键字。在某个实施例中,多达 20 个最相关的关键字(如在步骤 S306 中所识别出的)其每个都被分配了一个与它们的可察觉的相关性对应的数字权重。在步骤 S307 中,该关键字和权重列表通过管理员的接口(如图 2A 中的管理员接口 216)呈现给管理员。可选地,管理员可以通过管理员接口 216 改变所呈现的关键字和/或权重,其原因显而易见。在步骤 S308 中,一经编辑,管理员可提交包括关键字和 URL 列表的索引条目关键字,将其存储为索引 212 的表 240 和表 250 中的纪录。每个关键字用于衍生表 240 的一行。

[0074] 重复进行步骤 S300A 使得管理员建立已编索引的 URLs 的集合,每个 URL 在个人索引 212 中包含索引条目。在所述区域性高尔夫球网站例子的上下文中,已编索引的站点的集合可能包含区域性高尔夫球课程列表的 URLs。显而易见地,通过为已编索引的 URLs 分配期望的关键字和权重,管理员可以有效地“调整”或组织从任何由搜索算法 222 执行的搜索中所获取的搜索结果。

[0075] 通过获取置于广告客户索引 211 中的每个广告合同的必要的细节,管理员可以用类似的方式收集广告客户索引 211 中的纪录。在所公开的实施例中,这样的信息至少包括广告客户网页 URL,与该 URL 相关联的关键短语,该广告运行的日期范围,以及其在共享组中的明确的等级或放置位置等。其他信息也可提供作为每个记录的一部分。

[0076] 通过合同,公共索引 213 包含并非由管理员收集的索引信息,而是可能由第三方索引提供者制作,例如 DMOZ 数据库。在本公开的实施例中,索引 213 存储在与存储索引 212 的数据库具有同样格式的数据库中。当然,索引 213 也可以具有与索引 212 完全不同的数据结构,如同广告客户索引 211 那样。

[0077] 在本发明的典型实施例的方法中,管理员可以在个人索引 212 中对已经在公共索引 213 中编过索引的 URL 编索引。而且,在公共索引 213 或个人索引 212 中编过索引的 URL 也可在广告客户索引 211 中编索引,作为某个广告客户的特定广告合同的一部分。

[0078] 如下文所要解释的,具有了多个索引 211,212,213 的特定索引搜索算法 221,222,223,并进一步具有带合并法则的合并算法 225 以便对获取自每个特定索引搜索算法 221,222,223 的搜索结果加以合并,给网站管理员提供了高度灵活的查询和显示搜索结果的方法。

法。这种灵活性可包括对特定索引搜索算法 221, 222, 223 中任意一个或者合并算法 225 的修改, 以便对实质上实时显示给终端用户的搜索结果进行定形或组织进行修改。相比可能具有日、周、或更长时间周期的传统搜索引擎, 可觉察到这种类型的灵活性正是纵向市场搜索引擎最需要的特性。

[0079] 通过举例的方式, 图 3B 所示为索引服务器 110 为了响应一个包括了访问该区域性高尔夫网站的终端用户输入的一个或多个关键字的查询请求, 而对广告客户索引 211, 个人索引 212 和公共索引 213 进行查询所执行的典型步骤的示意流程图。本技术领域中的技术人员可了解到, 步骤 S300B 可由本发明典型实施例中的计算机软件实现, 包括以合适的计算机语言编写的可读代码。

[0080] 如图所示, 终端用户输入索引服务器 110 的查询在步骤 S314 中被接收。在步骤 S315 中, 终端用户在步骤 S314 中输入的查询用于产生相应的关键短语 (如合同项目表 262 中的关键短语 262k) 或者一个或多个关键字 (如表 240 中的关键字 232d), 根据具体情况而定, 可用于搜索索引 211, 212, 213。

[0081] 在步骤 S316 中, 在步骤 S315 中产生的关键字或关键短语被特定索引搜索算法 221, 222 和 223 用于分别对索引 211, 212 和 213 进行查询。每个特定索引搜索算法 221, 222, 223 使用从终端用户查询中产生的合适的搜索信息, 并对结果进行排序或放置。例如, 特定索引搜索算法 221 可用该查询的关键字和“当前日期”来定位广告客户索引 211 中的匹配条目, 并根据特定等级或共享组对搜索结果进行排序, 如前文所讨论。特定索引搜索算法 222 可用由终端用户查询所产生的关键字在个人索引 212 中搜索匹配条目, 并根据关键字权重 (参考下文图 4 的详细说明中有详细解释) 对搜索结果进行排序。类似地, 特定索引搜索算法 223 可用由终端用户查询所产生的关键字在公共索引 213 中搜索匹配条目, 并根据另一组关键字权重 (同样参考下文图 4 的详细说明中有详细解释) 对搜索结果进行排序。

[0082] 在步骤 S318 中, 为每个匹配记录计算匹配指示符的质量。在该公开的实施例中, 索引 211 的匹配指示符的质量是通过确定关键短语是否与查询精确匹配, 以及“当前日期”是否落在规定的日期范围中来进行计算。在这个方面, 广告客户索引 211 中的匹配索引条目的匹配质量可说是 100%。同样, 在本公开的实施例中, 每个索引 212 和 213 的匹配指示符质量是通过将与终端用户查询匹配的每个关键字的权重 (如表 240 的域 232c 中所包含的) 相加来计算得到的。

[0083] 便利地, 特定索引搜索算法 221, 222, 223 (图 2A) 可以基于完全不同的规则为每个获取自广告客户索引 211, 个人索引 212 和公共索引 213 的搜索结果分别计算不同的匹配质量。

[0084] 在区域性高尔夫网站例子的内容中, 这给网站的管理员在对特定的终端客户查询的查询和显示搜索结果的方面提供了极大的灵活性。例如, 管理员可以用一组权重可分配的关键字搜索个人索引 212, 用一组预先已分配权重的关键字搜索公共索引 213, 以及用关键短语和特定日期范围的精确匹配广告客户索引 211。

[0085] 在步骤 S319 中, 获取自步骤 S316 并具有在步骤 S318 中计算出的匹配质量的搜索结果由诸如图 2A 的合并算法 225 得到。

[0086] 来自广告客户索引 211, 个人索引 212 和公共索引 213 的匹配搜索结果可随后在步

骤 S320 中合并。根据由合并算法 225 执行的预定的合并法则,该结果可用多种方法合并。例如,来自个人索引和公共索引 212,213 的索引条目可基于为每个索引条目计算的匹配质量而整体排序。匹配质量较高的索引条目可显示在匹配质量较低的索引条目的前面。可选地,所有来自个人索引 212 的匹配条目可显示于公共索引 312 的条目的前面。此外,广告客户索引 211 的匹配搜索结果可被赋予最高的优先级以便广告客户的网页 URL 可放置在如前文所讨论的特定的等级,或者规定的等级范围内。

[0087] 如果一个网站在索引 211,212 和 213 的一个以上中被编为索引,那么来自较高优先级索引的索引条目可在先于来自较低优先级索引的索引条目。即,并非是来自个人索引 212 和公共索引 213 的索引条目都包括在内,例如,而是仅可能显示来自个人索引 212 的索引条目。

[0088] 在区域性高尔夫网站例子的内容中,网站管理员可把索引 211,212,213 区分优先次序,这样来自广告客户索引 211 的搜索结果在先于任何列于个人索引 212 或者公共索引 213 中的同样的 URL,以及来自个人索引 212 的搜索结果依次在先于任何列于公共索引 213 中的同样的 URL。

[0089] 使用合并算法 225 的合并策略,在步骤 S320 中结合的记录中相应的域在步骤 S322 中被排序,且在步骤 S324 中,相应的域以合并的、经排序列表的形式显示给终端客户。如上所述,用于在步骤 S322 中合并的具有特定等级的 URLs 事实上在步骤 S324 中显示给终端用户时可具有更高的等级。

[0090] 在本公开的实施例中,图 2B 中的 URL 域 234b 或者图 2E 中的 URL 域 268a 可在步骤 S324 中显示给终端用户。另外的域如标题域 234c,268e,说明域 234d,268f 以及日期印记/更改域 234e,268d 也可显示出来。步骤 S300B 随之结束。

[0091] 图 4 示意性地说明了服务器 110 执行的示例查询 400 的结果。暂时只考虑个人索引 212(如索引 412 所体现)和公共索引 213(如索引 432 所体现)。

[0092] 索引 412(具有个人索引 212 的结构)编辑 URL 列表于块 414a。表 240 和 250 适宜地衍生。在图 4 中,块 414a 中任意的示例 URLs 标注为“URL_A”,“URL_B”,“URL_C”和“URL_D”。块 414b 中所示的“URL_E”在下文将详细描述。假定在当前例子中,这些 URL 每个都与一个高尔夫网站相关。

[0093] 对于块 414a 中的每个 URL,块 424a 图示了一个用于对该 URL 编索引的相关的关键词列表。块 424a 中的每个关键词都分配了如括号中所示的权重。块 414a 中的 URLs 由例如网站“客户 1”(如所述区域性高尔夫网站)的管理员加以控制。因此,424a 中的关键词和权重可以很容易地由网站“客户 1”的管理员加以更改。例如,管理员可用图 3 的示例方法 S300A 关联关键词并为“URL_A”至“URL_D”的各个 URL 分配关键词权重。

[0094] 在该示例中,权重为“80”的第一关键词“KW1”和权重为“100”的第二关键词“KW2”都与“URL_A”相关联。在另一个例子中,具有不同的权重“70”的同一个第一关键词“KW1”与具有不同的权重“90”的同一个第二关键词“KW2”都与“URL_B”相关联。为说明起见,此权重范围 0-100 为任意选择。

[0095] 图 4 进一步示意性地图示了公共索引 213 的表格的公共索引 432 中的条目,其展示了块 434 中多个已编索引的 URLs,即,“URL_E”到“URL_H”。对于块 434 中的每个 URL,都有一个或多个相关联的关键词,并在括号中示有权重,如块 442 中所示。

[0096] 范例公共索引 432 可由可连于网页以产生索引的软件自动产生。对于每个网页，该软件识别出一列有效的相关关键字。除了自动产生关键字以外，权重可分配给每个与给定网页相关联的关键字。例如，权重可由给定的关键字在网页中出现的频率来推导出，或者该关键字是否出现在该网页的特殊区域例如标题或描述。这里所示的固定权重范围 0.00-0.99 是为说明起见而任选的。

[0097] 无论如何，网站“客户 1”的管理员最初没有有效能力来编辑索引 432 中的索引条目（即，块 434 中“URL_E”至“URL_H”或者任何示于块 442 中的关键字权重）。

[0098] 然而，如之前参考图 3A 的图示和说明，与个人索引 412 的给定 URL 相关联的给定关键字的相应权重可由网站“客户 1”的管理员轻易地改变。例如，对于与记录“URL_A”相关联的“KW1”，当前权重“80”可通过分配新的权重任意升高或降低。这样，当某个查询包括关键字“KW1”时，“URL_A”的匹配质量可直接控制以便 URL_A 可在搜索结果列表中出现在或更高或更低的位置。

[0099] 较为有利的是，对于个人索引 412 中的 URLs，任何关键字可与给定 URL 相关联，即使该关键字不是自动产生的，甚至该关键字没有出现在该目标网页上。换言之，关键字可以任意分配给 URL，仅仅为了当查询中使用该关键字时，可使得该 URL 出现或者不出现在查询结果中。例如，如果希望当终端用户的查询中输入了关键字“KW9”（未图示）时显示“URL_D”，可将关键字“KW9”简单地与“URL_D”相关联，以及将合适的权重分配于“URL_D”的 KW9，用以确保一旦使用关键字“KW9”就会出现“URL_D”。

[0100] 便利地，对于个人索引 412，管理员可简单地通过调节该关键字的已编入索引的 URL 的相应权重，来调整或组织任何使用了该关键字的搜索结果的顺序。然而，需要理解地是，如果在搜索查询中使用了一个以上的关键字，分配较高或较低关键字权重的效果可能被其他关键字权重的取均值或取总值的效果冲淡或抵消。因此，例如，除非给所有分配给 URL 的关键字都分配较高的权重，否则不太可能确保较高的等级。

[0101] 为了在任何已成形的搜索中都把公共索引 432 中的 URL 包括进来，如图 4 所示，公共索引 432 中的一个或多个 URL 可由个人索引 412 中的管理员选择性地编为索引。特别是，在本示例中，“URL_E”已经在个人索引 412 中编为索引。这是考虑到例如所述区域性高尔夫网站希望把“URL_E”排除在搜索结果之外的情况，原因是该 URL 属于直接竞争对手的网站或者广告客户。

[0102] 显然，个人索引 412 中的索引法使得管理员可影响到搜索结果中“URL_E”的显示。因此，对“URL_E”的控制级别变为与对块 414a 中的其他 URL 的控制级别相同。换言之，可以任意将“URL_E”与关键字相关联，同时可由网站机构的管理员任意给那些关键字分配权重。

[0103] 在本例中，“URL_E”已经与关键字“KW1”和“KW2”相关联，其中“KW1”和“KW2”各被分配了权重“0”或权重“空”。该空权重可在例如当输入“KW1”和“KW2”中任一个时，不希望在合并的搜索结果中包括该记录的情况下予以分配。

[0104] 为了进一步说明，块 450 描述了响应使用关键字“KW1”和“KW2”的搜索的搜索结果，其合并了获取自个人索引 412 和公共索引 432（如图中的块 452 所示）的 URL。这里，关键字“KW1”和“KW2”已经由某终端用户在搜索查询中输入，如图中块 454a 所示。如图中 456 所示，公共索引 432 中任何匹配关键字“KW1”和“KW2”的 URL 的结果可由个人索引 412（例

如,与公共数据库 432 的块 434 中的“URL_E”相应的索引结果可由个人索引 412 的块 414b 中的相应索引条目“URL_E”先占)中的相应 URLs 先占。“URL_E”的“KW1”和“KW2”的权重都为“空”,示于块 460 的删除线区域。

[0105] 因此,如 460a 的经排序的列表所示,匹配“KW1”和“KW2”的个人索引 412 的 URLs 列表基于关键字权重的总和而排序。匹配“KW1”或“KW2”的公共索引 432 的 URLs 列表也遵循关键字权重总和的次序。在本示例中,个人索引 412 的 URLs 显示于公共索引 432 的 URLs 之前。这里体现的要求是赋予个人索引 432(即与索引 212 相应)中索引条目的优先级要优先于公共索引 434(即与索引 212 相应)中的索引条目。

[0106] 假定个人索引 412 和公共索引 434 的 URLs 的排序如上所述,相应记录的相关域就可以同样的顺序显示给终端用户。例如,URL 域 234b(图 2D)和其他相关域都可显示给终端用户。但是,在该示例,即使“URL_E”在该已排序的列表中,由于“URL_E”的“KW1”和“KW2”都被赋予了“空”权重,所以“URL_E”不会显示给终端用户。

[0107] 因此,获取自公共索引 432 的不需要的 URL 可被有效地排除在显示给终端用户的经合并的搜索结果列表之外。例如,由计算匹配某个搜索请求的关键字的权重总和和得到的匹配质量的预定值可使得相应的记录以特定的方式进行处理。例如,该经相加总和的关键字权重的空权重可用于指示所述相关联的 URL(本例中的 URL_E)被排除在给终端用户的显示以外。可选地,如果希望促使特定的 URL 从公共索引 432 中更突出地显示出来(例如“URL_E”),管理员也可以给与“URL_E”相关联的关键字分配合适的更高权重,以使得“URL_E”可以突出地显示在合并的搜索结果中。

[0108] 可见,公共索引 434 中的索引条目的先占或者丢弃由个人索引 412 和公共索引 434 两者中关键字域中的普通值触发。在本公开的实施例,该关键字域通过关系型数据库中常见的连接机制与 URL 域 234b 相连接(图 3B),例如通过示于图示(图 2C 和 2D)表 240 和表 250 中的 HASH_URL 域 232e/234a。当相同的 URL 从个人索引 412 和公共索引 434 中取回时,所述先占或者丢弃随之触发。当然,可了解到另一个合适的域也可使用。

[0109] 给定的上述图 4 的示例了合并获取自两个索引的搜索结果,也就是个人索引 212/412 和公共索引 312/432,现在考虑有第三个索引的例子,比如图 2A 中的广告客户索引 211。为了说明该示例,考虑合并算法 225(图 2A)的一个合并法则,其中该算法赋予了广告客户索引 211 比个人索引 212/412 更高的优先级,并顺次赋予个人索引 212/412 比公共索引 213/432 更高的优先级。

[0110] 同样,作为示例,考虑四个广告的情况,分别以合同项目标志符“CID1”,合同项目标志符“CID2”,合同项目标志符“CID3”以及合同项目标志符“CID4”作为标志。这些合同项目标志符在合同项目表 270(见图 2F)中定义为记录 270a-270d。假设为了本例子,图 2E 的“网络位置”表 268 和图 2B 的“URL 表”234 实际上使用同样的“哈希(hash)URL”以指向某个网页 URL。这样,CID1 可由哈希(hash)URL262j 和带有网址 URL_A 的表 268 进行关联。类似地,CID2 可由哈希(hash)URL262j 和带有网址 URL_B 的表 268 进行关联。同样,CID3 可由哈希(hash)URL262j 和带有网址 URL_C 的表 268 进行关联。最后,CID4 可由哈希(hash)URL262j 和带有网址 URL_D 的表 268 进行关联。

[0111] 同样,如数据库表 270 的等级 262b 域所示,CID1 指定了“等级 1”的放置,CID4 指定了“等级 4”的放置。但是,对 CID2 和 CID3 的定义在等级 262b 域并没有作出,而是分配

了共享组 262i 的共享组“A”中。CID1, CID2 和 CID3 都有相同的起始日期 262d, 终止日期 262e, 以及哈希 (hash) 关键短语 262m(带有如图示的关键短语哈希 (hash) 值“34”)。通过对关键短语数据库表 280 的记录 280c 的交叉参考, 可见关键短语哈希 (hash) 值“34”与关键短语“KW1KW2”相对应。

[0112] 给定的上述对 CID1, CID2, CID3 和 CID4 的说明性信息, 以及前文图 4 中的例子, 所述广告客户搜索算法 221 可使用关键短语“KW1KW2”在广告客户索引 211 中寻找匹配记录。因此, 在该示例中, CID1, CID2, CID3 和 CID4 都通过其哈希 (hash) 值“34”以及它们各自的, 包括终端用户查询的“当前”日期的日期范围来与该关键短语“KW1KW2”相匹配。

[0113] 给定的当前关于合并算法 225 的合并法则的假设, 任何来自索引 211 的搜索结果都将先占在个人索引 212(如图 4 的索引 412 所体现) 或者公共索引 213(如图 4 的索引 432 所体现) 中任何同样的匹配搜索结果。如同公共索引 434 中的索引条目的先占或者丢弃的情况, 个人索引 412 或公共索引 434 中索引条目被来自广告客户索引 411 的结果先占可由个人索引 412, 公共索引 432 和广告客户索引 211 中关键字域的普通值进行触发。例如, 如图 2F 所示的广告客户索引 211 中的每个匹配索引条目的哈希 (hash) URL262j 可与如图 2C 所示的个人索引 212/412 和公共索引 213/432 的哈希 (hash) URL232e 相比较。

[0114] 图 5 示意性地说明了, 当图 4 中的示例查询进一步包括获自广告客户索引 511(与图 2A 中的广告客户索引 211 相应) 的搜索结果时所获的搜索结果的例子。为避免混乱, 图 4 中的个人索引 412 和公共索引 432 未在图 5 中进行显示。但是, 可理解到广告客户索引 511, 个人索引 412 和公共索引 432 中的每个都可由前文所述的特定索引搜索算法进行搜索。

[0115] 如图 5 所示, 假设“当前”日期落在图 2F 中的纵列 262d 和 262e 为各个合同项目 CID1, CID2, CID3 和 CID4 所定义的范围之内, 并使用块 516 所示的终端用户指定的关键短语来匹配关键短语 512(与图 2F 中的哈希 (hash) 短语 262m 相对应), 则可获得块 520 中的经合并和排序的搜索结果。

[0116] 更为明确地, 使用上表 270 定义的说明性的值以及期望的放置位置 513, 对于某个特定的落在由起始日期 262d 和终止日期 262e 指定的范围内的日子, URL_A 根据其被分配的等级被置于等级 1 中。类似地, URL_D 也根据其被分配的等级被置于等级 4 中。在这个方面, 虽然关键字 KW1 和 KW2 相应的权重对于索引 212 和 213 来说非常重要, 但它们与索引 211 不相关。相反, 给定了匹配关键短语和日期范围, 如果图 2F 的纵列 262b 中所分配的等级存在, 则决定了 URL 所放置的等级。

[0117] 对于共享组“A”, 等级范围由表 290 中的最高等级 266c 和最低等级 266b 进行设定(在本示例中分别为 2 和 3)。也有关键短语哈希 (hash) 值 266d 与共享组 A(本例中为“34”) 相关联。基于共享组 A 的成员资格, URL_B 和 URL_C 都放置于为共享组 A 所定义的等级范围中。在本发明的实施例中, URL_B 和 URL_C 可以随机方式放置在该范围中。因此, 不能保证为等级范围中的某个特定的等级。但是, 在另一个实施例中, 可了解到 URL_B 和 URL_C 可设的等级范围由基于一些其他子规则(未图示) 的共享组进行定义。

[0118] 在等级 1-4 之后, 随后获取自个人索引 212/412 的搜索结果以先前图 4(即, 通过总计关键字权重的方式) 所述的方式进行排序。注意到 URL_E 可能因为对其关键字分配了零权重而被删除或者被禁制, 如前文图 4 所述。这可在例如假如 URL_E 为 URL_A, URL_B,

URL_C或者URL_D的竞争对手,并且该广告客户不希望该竞争对手的URL显示在搜索结果列表中时非常有利。最后,来自公共索引 212/412 的余下的搜索结果以如前文图 4 所述的方式跟随在来自个人索引 212/412 的搜索结果之后。

[0119] 但是,可以观察到由于所有 URL_A, URL_B 和 URL_C 现在都被放置于等级 1-3 作为由来自广告客户索引 211 的匹配搜索结果确定的广告,相应的来自如前文图 4 所图示的个人索引 213/432 的搜索结果已经被先占(即,URL_A, URL_B 和 URL_C 在块 520 中未被复制)。显示于等级 4 中的 URL_J 不具有相应的来自个人索引 213/432 的搜索结果,所以也不会占先于另一个搜索结果。当然,如果使用另一个关键短语,一组完全不同的广告可能放置于等级 1-4 中,而且这样可以进一步影响其它获取自个人索引 212/412 和公共索引 213/432 的搜索结果的先占或删除。

[0120] 可了解到,在上文所述的区域性高尔夫网站的例子里,管理员在使用特定索引搜索算法查询和显示获取自多个索引 211, 212, 213 的搜索结果的方面拥有极大的灵活性。得自特定索引搜索算法的结果随后可根据合并算法中灵活的合并法则进行合并。该合并法则可以,例如,基于相关联关键字域的匹配条目,把多个索引区分优先次序并且用来自较高优先级索引的索引条目先占于来自较低优先级索引的索引条目。因此,例如该区域性高尔夫网站的广告客户可确保得到搜索结果等级中的特殊位置,同时来自竞争网站的 URL 可有效地被先占。

[0121] 更普遍地说,上述说明性例子公开了一种将与终端客户查询相应的搜索结果进行显示的方法。如图示,搜索结果由来自多个索引的结果组合而成,每个搜索结果具有一个相关联的关键字域。该方法包含了使用特定索引搜索算法对所述多个索引中每个索引的索引条目进行查询,以便从每个索引中获取一组匹配搜索结果,每个匹配搜索结果具有特定于其索引的匹配质量。该方法进一步包含确定所述多个索引中每个索引相对的优先级以及基于该多个索引的每个索引的相对优先级,将得自所述多个索引的匹配搜索结果组合进入经排序的搜索结果的合并列表。如图示,任何来自其相关关键字域与高优先级索引中的匹配搜索结果的相关关键字域相同的优先级较低的索引的搜索结果都将被抛弃,这样有利于来自优先级较高的索引的匹配搜索结果。

[0122] 如同样所说明的,该方法可进一步包含确定至少一个分配给所述多个索引中至少一个索引的每个索引条目的关键字,并将至少一个关键字与终端用户查询相匹配,以获取匹配的搜索结果。该方法可进一步包含确定与每个关键字相关联的权重,并通过将匹配终端用户查询的任何关键字的权重进行相加而计算匹配质量。

[0123] 如进一步所说明的,该方法可进一步包含确定分配给至少一个索引中每个索引条目的关键短语,并将该关键短语与终端用户查询相匹配,以便获取匹配搜索结果。该方法可进一步包含确定终端用户查询的当前时间是否落在分配给每个索引条目的有效时间范围之内,以便获取匹配搜索结果。

[0124] 可理解到,以上所讨论的示例的硬件结构,软件,以及数据组成可恰当地适用于实践该阐明的方法。

[0125] 本发明的实施例可类似地包括多于三个的索引,每个索引分配一个相应的优先级。如果共享一个同样的关键字域的多个索引条目在响应某搜索中被取回,则来自较低优先级索引的结果将被来自较高优先级索引的结果先占。因此,仅有来自最高优先级索引的

匹配结果可被包括在任何被显示的结果列表中。有利的是,每个索引可通过某个仅与该索引相关联的特定索引搜索算法(如算法 211,222 或 223)进行搜索。当增加索引时,搜索引擎 214 中也会增加模块化搜索算法。

[0126] 当然,上述实施例仅用于说明,并非限制性的。所述实行本发明的实施例可在形式上,各部分的安排上,细节上以及操作顺序上作多种更改。更确切地说,本发明意在包括所有这些属于由其权利要求所定义的范围内的更改。

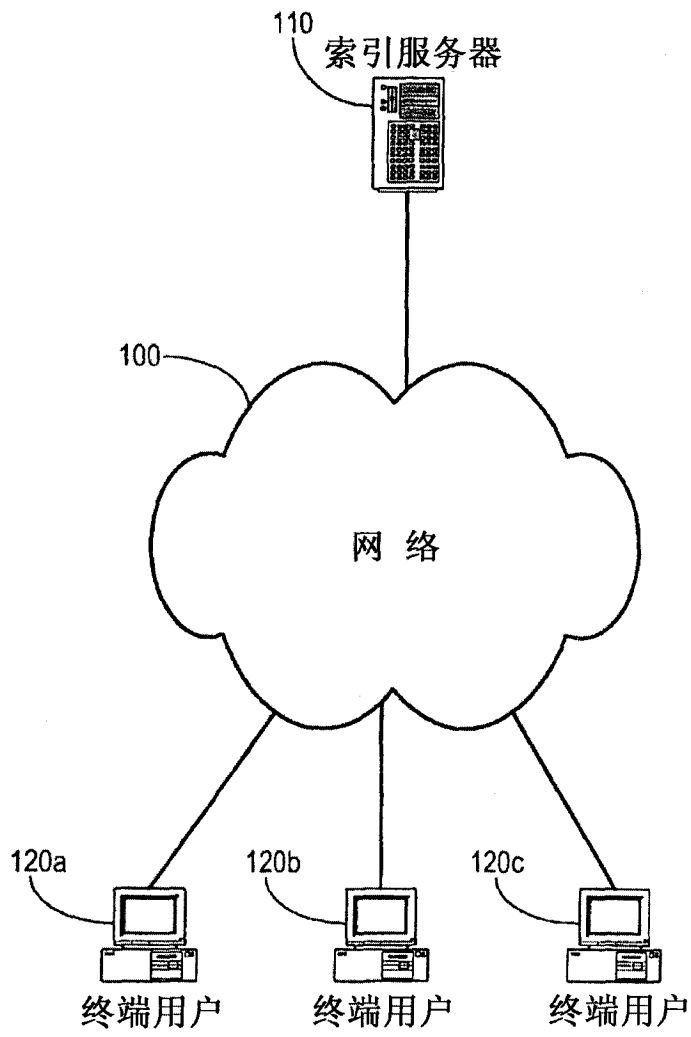


图 1A

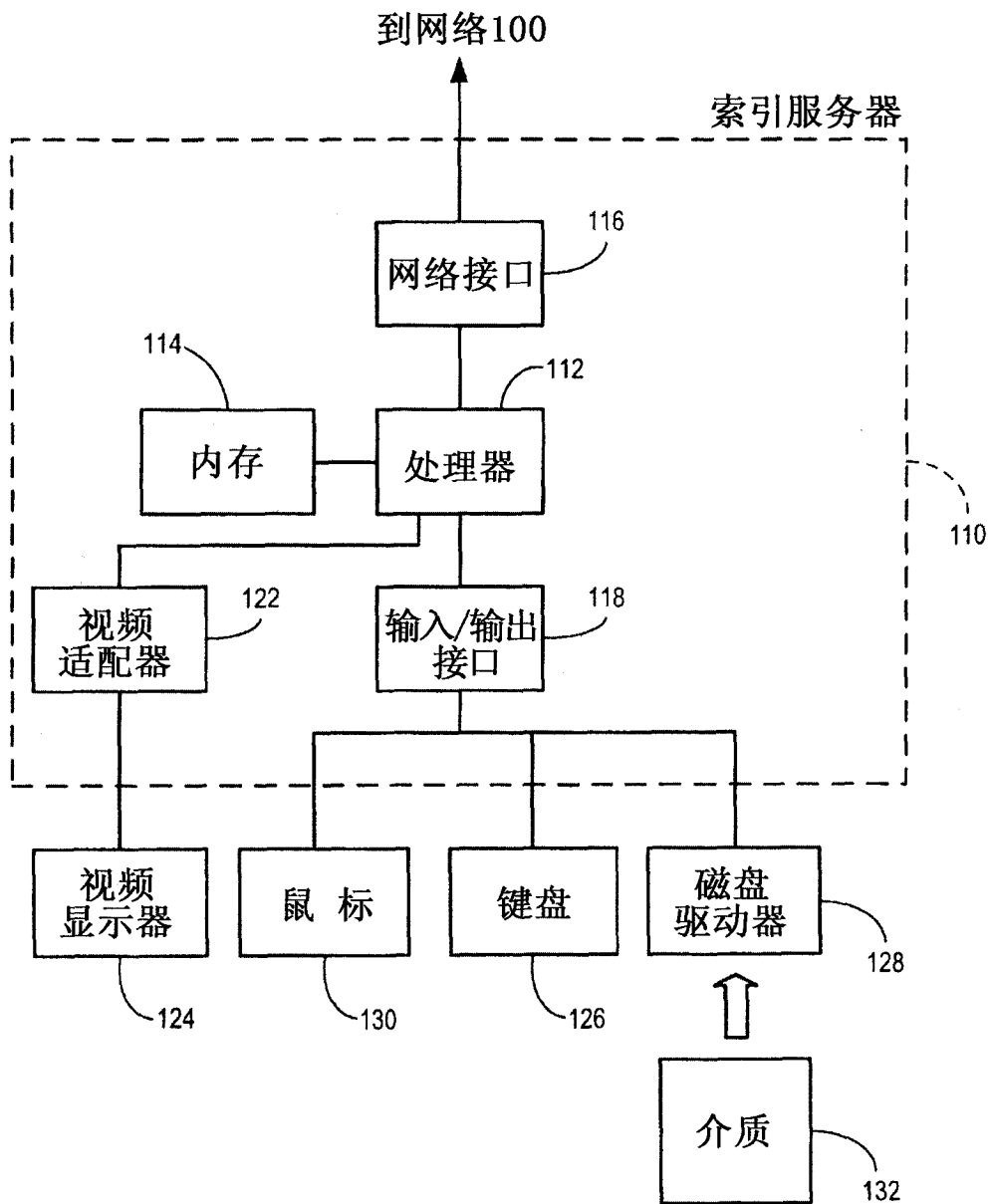


图 1B

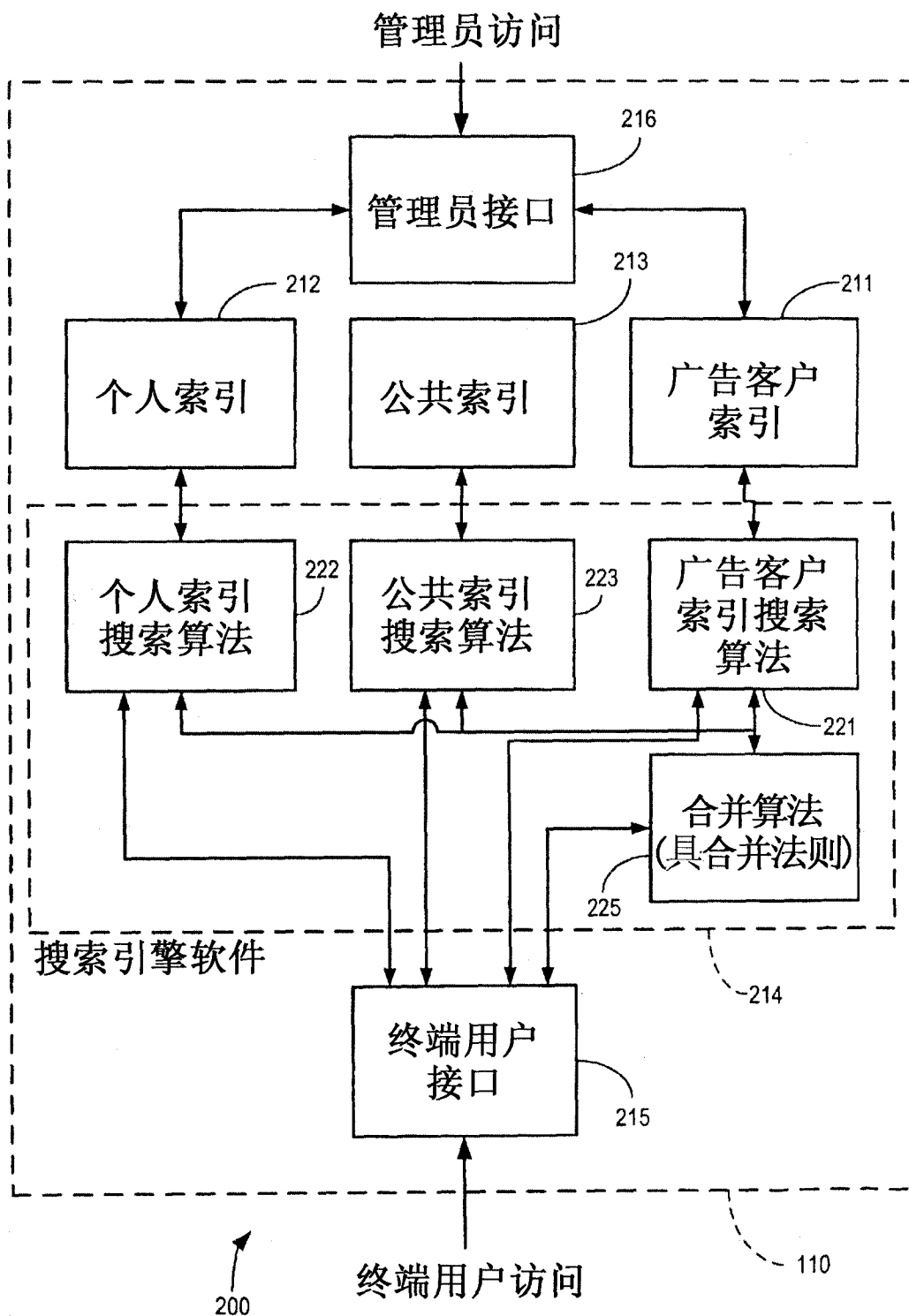


图 2A

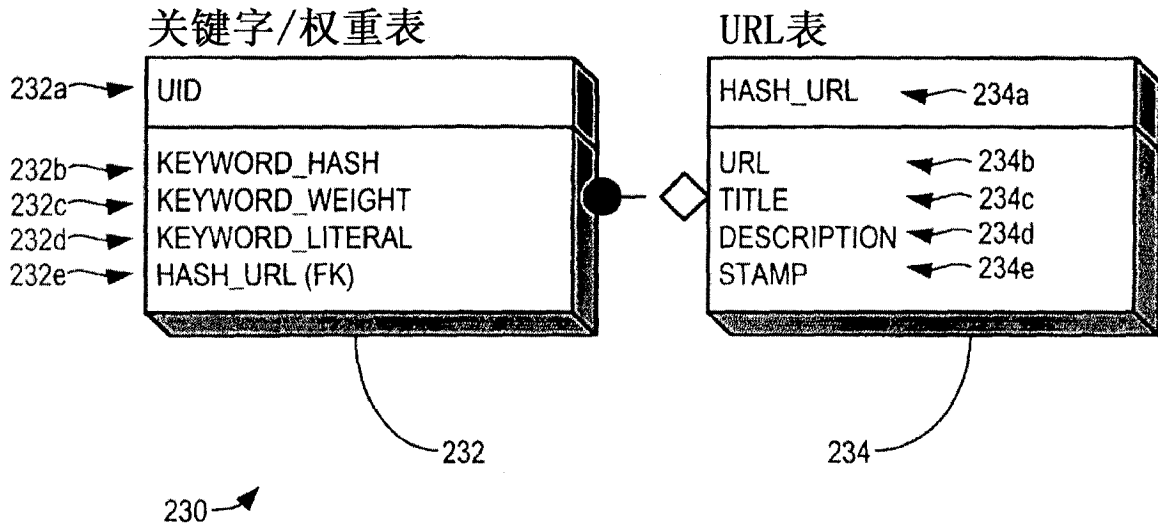


图 2B

UID	KEYWORD_HASH	KEYWORD_WEIGHT	KEYWORD_LITERAL	HASH_URL
UID:1	72	100	KW1	12
UID:2	73	70	KW2	12
UID:3	74	90	KW3	12

图 2C

HASH_URL	URL	标题	描述	印记
12	www.golf1.foo	高尔夫一	高尔夫一的主页	yy-mm-dd-hh:mm
13	www.golf2.foo	高尔夫二	高尔夫二的主页	yy-mm-dd-hh:mm
14	www.golf3.foo	高尔夫三	高尔夫三的主页	yy-mm-dd-hh:mm

图 2D

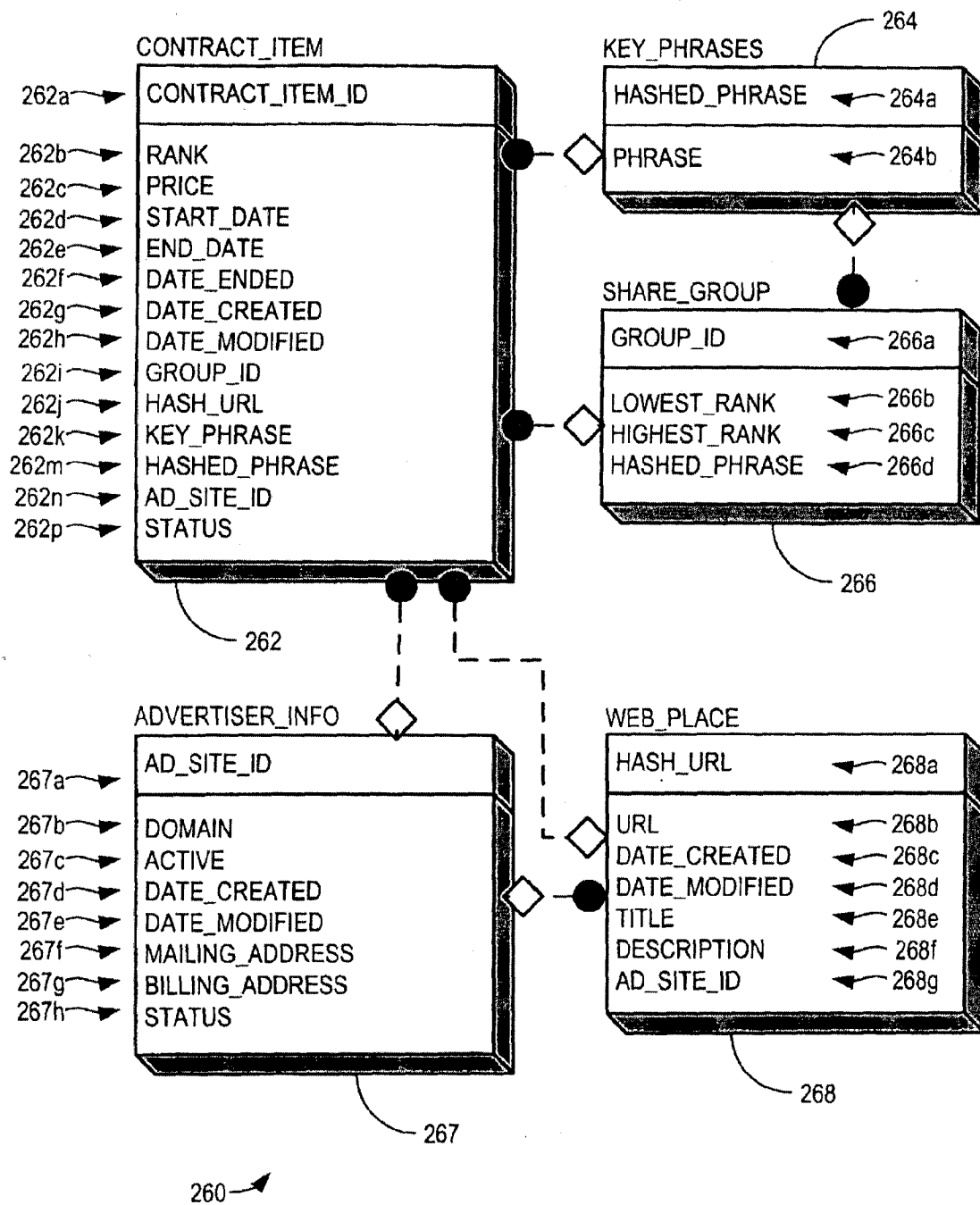


图 2E

	262a	262b	262c	262d	262e	262i	262j	262m	262n
	CONTRACT ITEM_ID	RANK	PRICE	START DATE	END DATE	GROUP ID	URL HASH	HASHED PHRASE	AD_ID
270a →	CID1	1	\$300	01-FEB-04	29-FEB-04		12	34	AD_A
270b →	CID2		\$150	01-FEB-04	29-FEB-04	GRP_A	13	34	AD_B
270c →	CID3		\$150	01-FEB-04	29-FEB-04	GRP_A	14	34	AD_C

270 →

图 2F

	264a	264b
	HASHED PHRASE	短语
280a →	32	"KW1"
280b →	33	"KW2"
280c →	34	"KW1KW2"

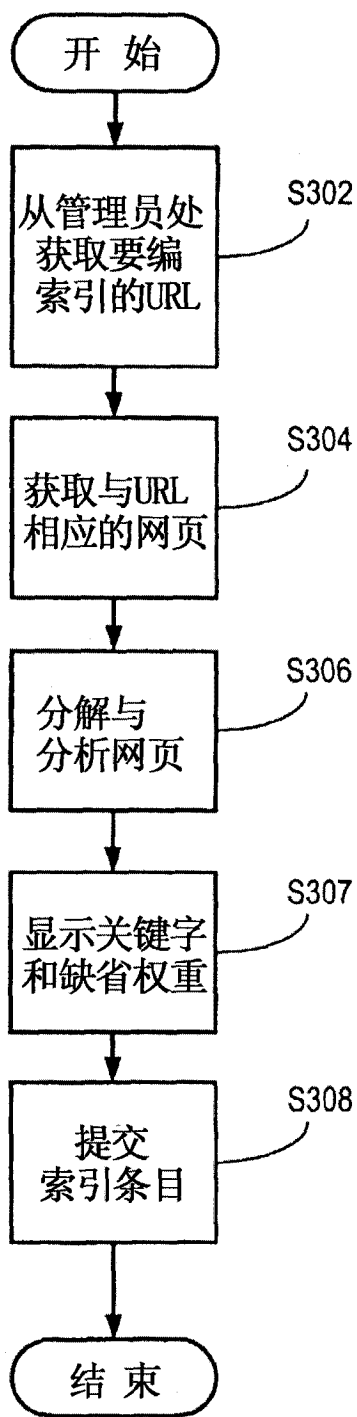
280 →

图 2G

	266a	266b	266c	266d
	GROUP ID	LOWEST RANK	HIGHEST RANK	HASHED PHRASE
290a →	GRP_A	3	2	34

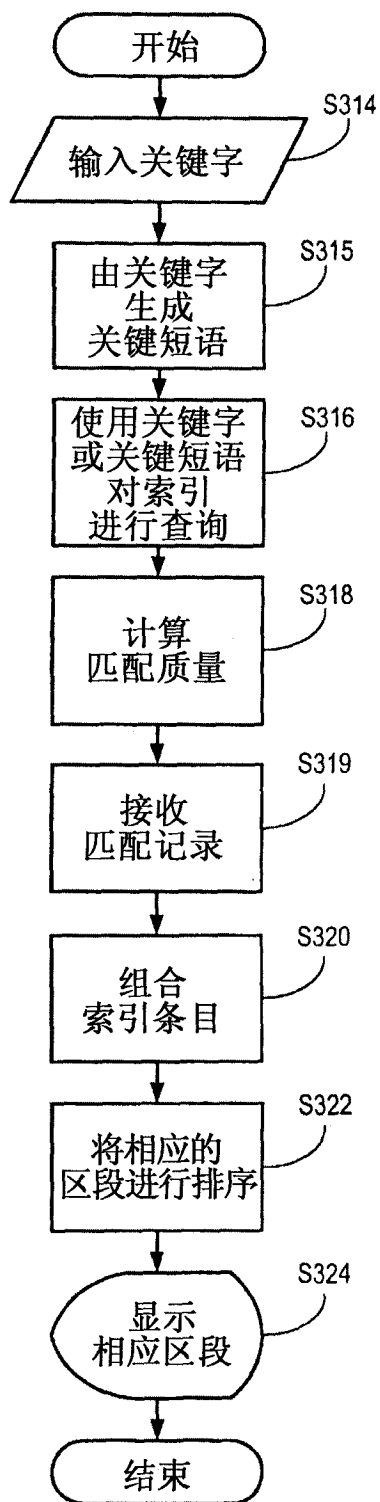
290 →

图 2H



S300A

图 3A



S300B

图 3B

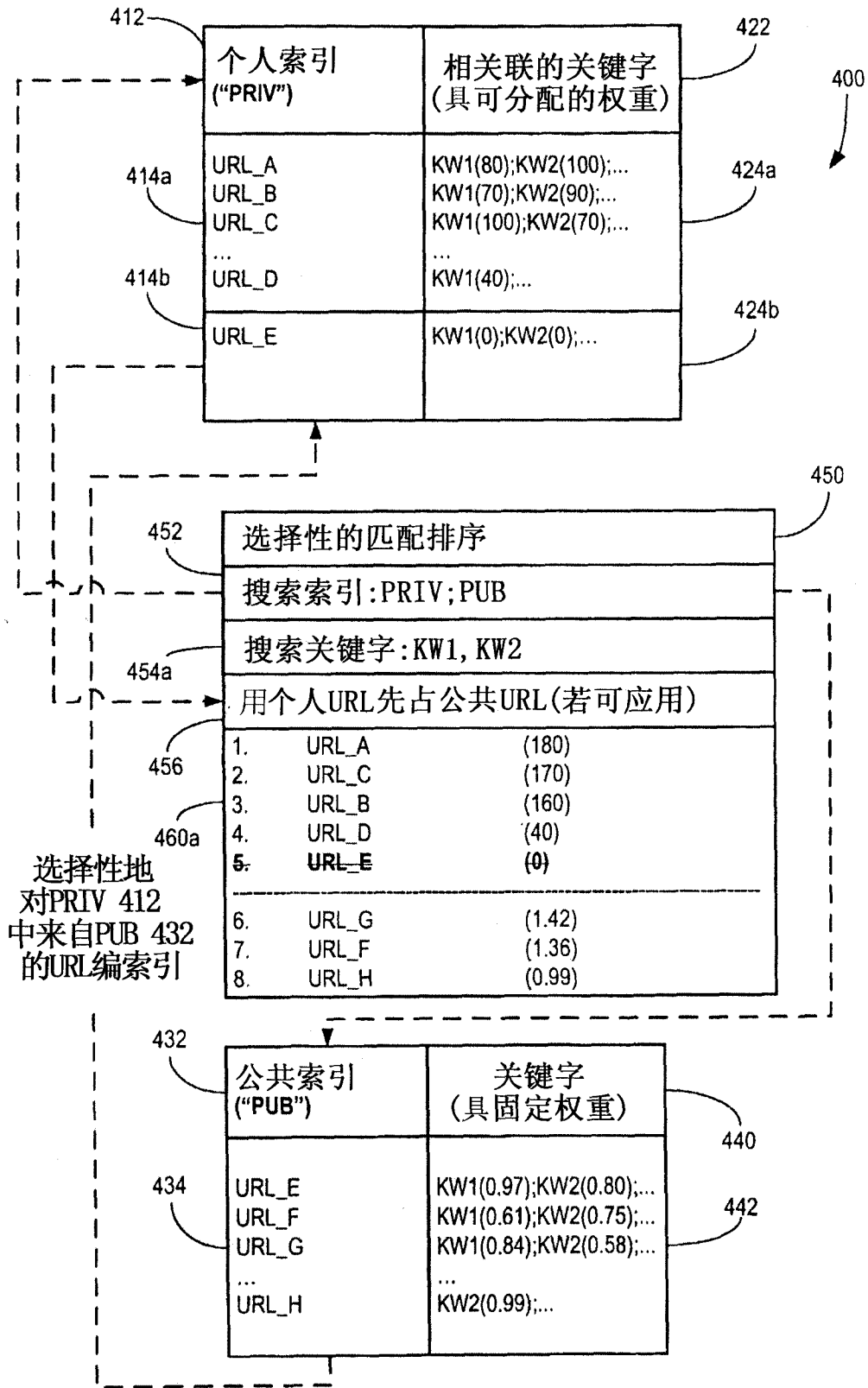


图 4

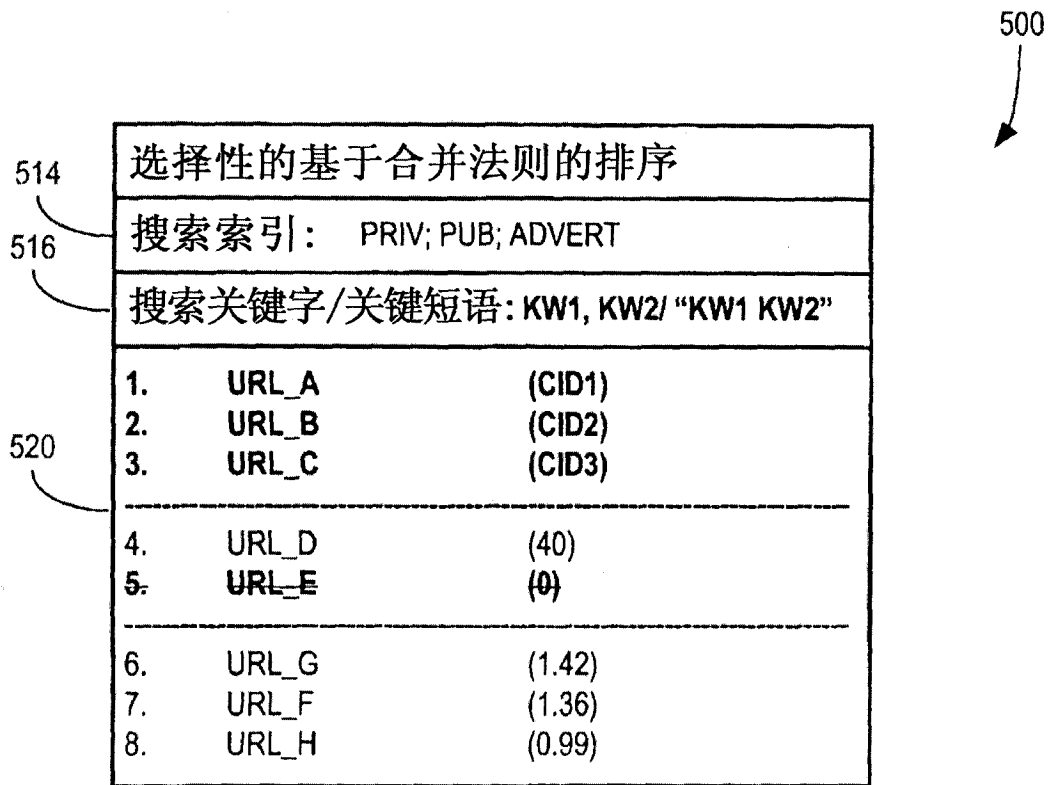


图 5