



(12) 发明专利

(10) 授权公告号 CN 101641694 B

(45) 授权公告日 2013. 05. 22

(21) 申请号 200880005237. 1

(22) 申请日 2008. 02. 12

(30) 优先权数据

11/707, 684 2007. 02. 16 US

(85) PCT申请进入国家阶段日

2009. 08. 17

(86) PCT申请的申请数据

PCT/IB2008/002738 2008. 02. 12

(87) PCT申请的公布数据

W02009/066140 EN 2009. 05. 28

(73) 专利权人 雅虎公司

地址 美国加利福尼亚州

(72) 发明人 雷恩·休

(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 南霆

(51) Int. Cl.

G06F 17/30 (2006. 01)

G06F 17/00 (2006. 01)

(56) 对比文件

US 2003041054 A1, 2003. 02. 27, 摘要, 权利要求 1-6, 说明书第 12-15, 27, 30-36 段, 附图 2-3.

US 2004030690 A1, 2004. 02. 12, 摘要, 权利要求 1-8, 说明书第 11-18, 24-25, 34 段, 附图 1-3, 5-6.

CN 1856791 A, 2006. 11. 01, 全文.

CN 1330330 A, 2002. 01. 09, 全文.

审查员 马晓宇

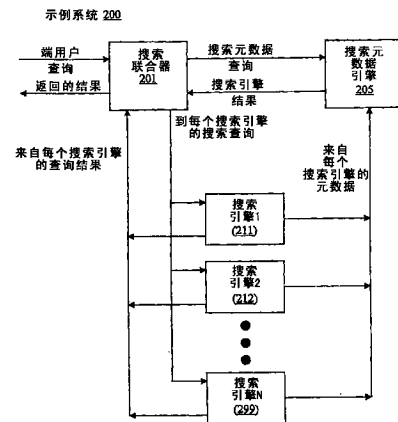
权利要求书3页 说明书11页 附图8页

(54) 发明名称

通过若干搜索引擎实现的联合搜索

(57) 摘要

基于搜索查询和一个或多个标准从若干搜索引擎中选择多个搜索引擎。为了获得针对搜索查询的搜索结果,向所述多个搜索引擎提交一个或多个搜索查询。对于所述多个搜索引擎中的每个搜索引擎,接收针对向该搜索引擎提交的一个或多个搜索查询的某些搜索结果。通过对从所述多个搜索引擎中的每个搜索引擎接收的所述某些搜索结果进行合并来生成合并的搜索结果。



1. 一种用于执行联合搜索的方法,包括如下步骤:  
至少部分地基于某搜索查询来生成生成的查询;  
将所述生成的查询提交给特定搜索引擎;  
响应于将所述生成的查询提交给特定搜索引擎,从所述特定搜索引擎接收指示第一多个搜索引擎的结果;  
从所述第一多个搜索引擎中选择第二多个搜索引擎;  
为了获得针对所述某搜索查询的搜索结果,向所述第二多个搜索引擎提交一个或多个搜索查询;  
对于所述第二多个搜索引擎中的每个搜索引擎,接收针对向该每个搜索引擎提交的所述一个或多个搜索查询的某些搜索结果;以及  
将从所述第二多个搜索引擎中的每个搜索引擎接收的所述某些搜索结果进行合并,以产生合并的搜索结果组;  
将所述合并的搜索结果组呈现给用户,其中,所述合并的搜索结果组中的特定搜索结果基于接收每个所述特定搜索结果的搜索引擎而被可视地分组成搜索结果群组。
2. 如权利要求 1 所述的方法,其中,所述指示第一多个搜索引擎的结果至少部分地基于描述所述第一多个搜索引擎中的每个搜索引擎的信息。
3. 如权利要求 1 所述的方法,其中,向所述第二多个搜索引擎提交一个或多个搜索查询进一步包括:  
至少部分地基于所述某搜索查询来生成特定查询;以及  
将所述特定查询提交给所述第二多个搜索引擎中的至少一个搜索引擎。
4. 如权利要求 1 所述的方法,其中,所述第二多个搜索引擎中的一个或多个搜索引擎包括 web 站点。
5. 如权利要求 1 所述的方法,其中,所述选择第二多个搜索引擎包括选择搜索引擎索引。
6. 如权利要求 1 所述的方法,其中,所述合并包括将从所述第二多个 搜索引擎中的每个搜索引擎返回的结果组合成单组搜索结果,以便呈现给提交所述某搜索查询的端用户。
7. 如权利要求 6 所述的方法,其中,所述单组搜索结果根据相关性而被排名。
8. 如权利要求 1 所述的方法,其中,将所述某些搜索结果进行合并至少部分地基于提交所述某搜索查询的端用户的偏好。
9. 如权利要求 1 所述的方法,  
其中,从所述第一多个搜索引擎中选择所述第二多个搜索引擎至少部分地基于所述第一多个搜索引擎中每个搜索引擎的排名;并且  
其中,所述第二多个搜索引擎包括来自所述第一多个搜索引擎中的最相关的搜索引擎;  
其中,所述最相关的搜索引擎是基于所述第一多个搜索引擎的排名而确定的。
10. 如权利要求 1 所述的方法,其中,所述合并包括将从所述第二多个搜索引擎接收的结果合并成多个搜索结果组,所述搜索结果与所述第二多个搜索引擎的每个搜索引擎相对应,以便呈现给提交所述某搜索查询的端用户。
11. 如权利要求 1 所述的方法,还包括:

确定所述第二多个搜索引擎中的每个搜索引擎的排名；

其中，呈现所述合并的组的搜索结果还包括基于所述第二多个搜索引擎中的每个搜索引擎的排名来对所述搜索结果群组进行排序。

12. 如权利要求 1 所述的方法，还包括：

向所述用户呈现所述第二多个搜索引擎的列表；

其中，所述第二多个搜索引擎的列表可视地独立于所述合并的搜索结果组。

13. 如权利要求 1 所述的方法，其中，所述合并的搜索结果组包括对所述第二多个搜索引擎中的某一搜索引擎的提及，所述方法还包括：

从所述用户接收对所述某一搜索引擎的选择；以及

响应于接收到对所述某一搜索引擎的选择，从所述某一搜索引擎显示 搜索结果。

14. 一种用于执行联合搜索的系统，包括：

用于至少部分地基于某搜索查询来生成生成的查询的装置；

用于将所述生成的查询提交给特定搜索引擎的装置；

用于响应于将所述生成的查询提交给特定搜索引擎，从所述特定搜索引擎接收指示第一多个搜索引擎的结果的装置；

用于从所述第一多个搜索引擎中选择第二多个搜索引擎的装置；

用于为了获得针对所述某搜索查询的搜索结果，向所述第二多个搜索引擎提交一个或多个搜索查询的装置；

用于对于所述第二多个搜索引擎中的每个搜索引擎，接收针对向该每个搜索引擎提交的所述一个或多个搜索查询的某些搜索结果的装置；以及

用于将从所述第二多个搜索引擎中的每个搜索引擎接收的所述某些搜索结果进行合并，以产生合并的搜索结果组的装置；

用于将所述合并的搜索结果组呈现给用户的装置，其中，所述合并的搜索结果组的特定搜索结果基于接收每个所述特定搜索结果的搜索引擎而被可视地分组成搜索结果群组。

15. 如权利要求 14 所述的系统，其中，所述指示第一多个搜索引擎的结果至少部分地基于描述所述第一多个搜索引擎中的每个搜索引擎的信息。

16. 如权利要求 14 所述的系统，其中，用于向所述第二多个搜索引擎提交一个或多个搜索查询的装置进一步包括：

用于至少部分地基于所述某搜索查询来生成特定查询的装置；以及

用于将所述特定查询提交给所述第二多个搜索引擎中的至少一个搜索引擎的装置。

17. 如权利要求 14 所述的系统，其中，所述第二多个搜索引擎中的一个或多个搜索引擎包括 web 站点。

18. 如权利要求 14 所述的系统，其中，用于选择第二多个搜索引擎的装置包括用于选择搜索引擎索引的装置。

19. 如权利要求 14 所述的系统，其中，用于合并的装置包括用于将从 所述第二多个搜索引擎中的每个搜索引擎返回的结果组合成单组搜索结果以便呈现给提交所述某搜索查询的端用户的装置。

20. 如权利要求 19 所述的系统，其中，所述单组搜索结果根据相关性而被排名。

21. 如权利要求 14 所述的系统，其中，将所述某些搜索结果进行合并至少部分地基于

提交所述某搜索查询的端用户的偏好。

22. 如权利要求 14 所述的系统,

其中,从所述第一多个搜索引擎中选择第二多个搜索引擎至少部分地基于所述第一多个搜索引擎中每个搜索引擎的排名;并且

其中,所述第二多个搜索引擎包括来自所述第一多个搜索引擎中的最相关的搜索引擎;

其中,所述最相关的搜索引擎是基于所述第一多个搜索引擎的排名而确定的。

23. 如权利要求 14 所述的系统,其中,用于合并的装置包括将从所述第二多个搜索引擎接收的结果合并成多个搜索结果组以便呈现给提交所述某搜索查询的端用户的装置,其中,所述搜索结果与所述第二多个搜索引擎的每个搜索引擎相对应。

24. 如权利要求 14 所述的系统,还包括:

用于确定所述第二多个搜索引擎中的每个搜索引擎的排名的装置;

其中,用于呈现所述合并的组的搜索结果的装置还包括用于基于所述第二多个搜索引擎中的每个搜索引擎的排名来对所述搜索结果群组进行排序的装置。

25. 如权利要求 14 所述的系统,还包括:

用于向所述用户呈现所述第二多个搜索引擎的列表的装置;

其中,所述第二多个搜索引擎的列表可视地独立于所述合并的搜索结果组。

26. 如权利要求 14 所述的系统,其中,所述合并的搜索结果组包括对所述第二多个搜索引擎中的某一搜索引擎的提及,所述系统还包括:

用于从所述用户接收对所述某一搜索引擎的选择的装置;以及

用于响应于接收到对所述某一搜索引擎的选择,从所述某一搜索引擎显示搜索结果的装置。

## 通过若干搜索引擎实现的联合搜索

### 技术领域

[0001] 本发明一般涉及信息检索,并且更具体地涉及通过若干搜索引擎实现的联合搜索。

### 背景技术

[0002] 标准网络和因特网搜索通常覆盖可以经由网络访问的信息的一小部分,或许只有大约百分之十。对于此限制有多种原因。一个原因涉及可扩展性 (scalability)。例如,相当大的花费、努力和基本设施与存储和索引数据、保持信息常新并且可访问相关联。对于相对很少访问的信息,花费相对于明显的需求、信息的有用性等可能似乎过多。对可以经由网络访问的所有信息进行限制可以抑制花费。

[0003] 其次,基于搜索引擎的标准网络被阻止访问网络中的许多信息,这是因为有助于建立搜索引擎用于搜索的索引的标准网络爬虫 (web crawler) 可能缺少获取那些信息的能力。例如,网络爬虫可能不能访问要求诸如口令认证或其它安全措施之类的肯定访问使能 (positive access enablement) 的数据库或网站。

[0004] 最后,标准网络搜索是针对宽范性和大众性而设计的,跨越上层较大众的信息。虽然许多标准网络搜索结果在某种意义上被设计为返回可以最有可能使大多数用户满意的搜索结果,但是大部分时候,对于相当大数量的用户来说可能缺少相关性。

[0005] 为了寻找更加相关的搜索结果,可以使用垂直搜索,垂直搜索比标准搜索更精确,更定向于特定域和主题。然而,垂直搜索的用户针对每个搜索引擎都要或多或少单独地对他们的系统进行配置。对于选择相对较少的搜索引擎,这种实施或多或少是可管理的。然而,对于许多垂直搜索引擎,这可能不久会变得低效并且昂贵。

[0006] 基于前述说明,更加高效地执行垂直搜索将是有益的。

[0007] 在本部分中描述的方法是可以执行的方法,但不一定是之前就已经想到或者已经执行的方法。因此,不应假定本部分中描述的任何方法仅仅因为包括在本部分中就被认为是现有技术,除非另有指定。

### 附图说明

[0008] 在附图的各视图中通过示例而非限制来图示出本发明,其中相似的标号指代相似的元件,并且其中:

[0009] 图 1 示出根据本发明的一个实施例的示例处理;

[0010] 图 2 示出根据本发明的一个实施例的示例系统;

[0011] 图 3A 示出根据本发明的一个实施例的示例搜索结果呈现;

[0012] 图 3B、3C、3D 示出根据本发明的一个实施例的示例截屏;

[0013] 图 4 示出可以基于其实施本发明的实施例的示例计算机平台;以及

[0014] 图 5 示出根据一个实施例的用于对来自若干搜索引擎的结果进行合并的示例处理。

## 具体实施方式

[0015] 这里描述涉及通过若干搜索引擎实现的联合搜索的示例实施例。在下面的描述中,为了说明的目的,大量特定细节被提出以便提供对本发明的彻底理解。然而,很明显,本发明可以在没有这些特定细节的情况下实施。在其它实例中,已知的结构和设备是在没有详尽对待的情况被提及的,以避免不必要地模糊本发明。

### [0016] 概述

[0017] 根据一个实施例,基于搜索查询选择一组垂直搜索引擎。很可能提供与查询比较相关的搜索结果的垂直搜索引擎被选择用于所述组。例如,查询可能是请求“猪肉食谱”。基于在查询中包括这些搜索关键词,针对食谱的搜索引擎被选择。

[0018] 因此,在执行垂直搜索中利用较少的用户配置来访问若干搜索引擎,以利用这些搜索引擎执行查询。此外,高效且低花费的垂直搜索查询通过若干专门的搜索引擎来执行并且是有选择的被执行。

[0019] 对于向其提供搜索查询的每个所选择的搜索引擎,接收搜索结果。通过对从多个搜索引擎的每个搜索引擎接收的某些搜索结果进行合并来生成合并的搜索结果。

[0020] 在一个实施例中,依赖于垂直搜索用户的偏好,对根据跨越若干搜索引擎的查询的结果的合并可以变化。例如,用户可能喜欢被呈现来自若干搜索引擎的组合搜索结果的单个列表。这向端用户呈现出执行查询的搜索引擎系统认为最相关的结果。上述方式以外或者替代上述方式,用户可能喜欢被呈现来自选择的搜索引擎中的每个搜索引擎的结果。在这样的情况中,用户被呈现来自搜索引擎系统认为是最相关的搜索引擎的搜索结果。意图发出将被限于特定域的额外查询的垂直搜索用户因此可以被呈现很可能相关的结果。因此,这里描述的实施例实现了通过若干搜索引擎的联合搜索,并且向搜索用户呈现出相关的搜索结果以及用于此外更多的特定于域的搜索的相关搜索引擎。

### [0021] 用语和注释

[0022] 除非另有特定声明以外,这里所使用一些术语和短语具有本部分所描述的含义。

[0023] 搜索引擎是被配置用于搜索一组信息资源的软件(可执行指令和数据)。执行搜索引擎的计算机生成针对向搜索引擎提交的查询的搜索结果。

[0024] 搜索引擎通常在服务器(这里被称为搜索引擎服务器)上运行。服务器是集成软件组件(包括数据)和诸如在计算机上用于执行集成软件组件的存储器、节点和处理器之类的计算资源的分配的组合,在服务器上,软件和计算资源的组合专用于特别的功能。在搜索引擎服务器的情况中,服务器专用于搜索一组信息资源。

[0025] 搜索引擎在因特网、万维网(www, Web, WWW 等等)以及其它大型内部网络和信息资源网络中被广泛使用。通常,搜索引擎可作为网站在服务器上被访问,所述网站诸如是可利用 Yahoo 和 Google 网页来访问的那些网站,它们分别利用链接(<http://search.yahoo.com/>)和(<http://www.google.com/>)来访问。

[0026] 由搜索引擎搜索出的信息资源的集合在这里称为文档(document)。文档是可通过搜索引擎索引来索引的任何信息单元。搜索引擎索引在下面描述。通常,文档是可以包含无格式的或带格式的文本、行内图形(inlinegraphic)和其它多媒体数据以及对其它文档的超级链接。文档可以是静态或者动态生成的。

[0027] 搜索引擎使用搜索引擎索引（即，一个或多个）来搜索信息，搜索引擎索引在这里也被简单称为索引。搜索引擎索引可以是其中内容或多或少被手动索引的目录，以反映人为观察。更具体地，搜索引擎索引被这里称为爬虫的处理自动创建并维护。爬虫在因特网中探索信息，在该爬虫被配置搜索的位置处本质上连续地寻找它们尽可能多地找到的文档。爬虫可以沿着从一个文档到另一文档的链接而行进。爬虫可以检索（例如，获取、下载）网络资源（例如，页面），保存原始形式的页面或其它资源，识别和提取到其它资源、网页的链接，等等。这使得这些资源 / 页面又能够被随后检索。当检索爬虫得到的页面或其它资源时，可以从爬虫得到的资源 / 页面提取其它形式的信息。

[0028] 从利用爬虫检索的资源 / 页面提取的内容然后可以被索引。在搜索索引中，资源 / 页面的内容被索引（例如，语义索引、概念索引等等）并且被总结在数据库中，此数据库通常具有相当大的大小。响应于搜索查询实际搜索的就是这些索引和数据库。因此，一般而言，用于建立索引的索引处理要使用爬虫获取内容。

[0029] 在一种配置中，用于搜索引擎的软件可以被配置用于搜索一组资源，而在另一种配置中被配置用于搜索另一组资源。这样的配置中的每一种配置被认为是不同的搜索引擎。

[0030] 垂直搜索引擎是使用用于索引被限制于特定域或特定主题的文档的索引的引擎。例如可以通过将爬虫配置为搜索特定位置的方式来限定垂直搜索引擎。例如，针对食谱的用于垂直搜索引擎的爬虫可以被配置为搜索保持有食谱文档的站点和 / 或位置。

[0031] 由搜索引擎生成的搜索结果包括文档的列表并且可以包含关于文档的概要信息。文档的列表可以被排序。为了对文档的列表进行排序，搜索引擎可以为列表中的每个文档分配排名。当列表通过排名而被排序时，具有相对较高排名的文档与具有相对较低排名的文档相比可能被置为离列表的头部更近。搜索引擎可以根据对搜索查询的相关性来对文档排名。相关性是文档的主题与搜索查询关键词匹配得多么近的量度。

[0032] 说明性实施例

[0033] 图 1 示出根据一个实施例的示例处理 100。处理 100 以基于搜索查询和至少一个标准从若干搜索引擎中选择多个搜索引擎而开始。一个或多个搜索查询被提交 102 给多个搜索引擎以便获得搜索结果。对于每个搜索引擎，某一些搜索结果被接收 103 以用于针对提交给每个搜索引擎的每个查询。来自各个搜索引擎的这些结果被合并 104 以生成合并的搜索结果。

[0034] 图 2 示出根据一个实施例的示例系统 200。搜索联合器 201 接收用户查询。诸如联合器 201 之类的联合器包括执行软件的一个或多个计算机处理，其中，软件使得计算机处理接收搜索查询、基于搜索查询选择一个或多个搜索引擎，以及将搜索查询发送给所选择的搜索引擎。联合器还可以合并响应于搜索查询从若干搜索引擎返回的结果。联合器可以是搜索引擎的集成组件，其中分配给搜索引擎的一个或多个处理还操纵此联合器。

[0035] 为了选择向其提交搜索查询的搜索引擎，联合器 201 需要可用来做出选择的信息。在一个实施例中，联合器 201 通过向搜索元数据引擎 205 提交搜索元数据查询来获得这样的信息。

[0036] 搜索元数据引擎是这样的搜索引擎，其中信息资源组是关于其它搜索引擎的信息。此信息包括搜索引擎元数据，搜索引擎元数据是描述搜索引擎和 / 或搜索引擎索引的

信息。元数据可以包括与搜索引擎相关的主题的描述符以及与相关主题相关联的关键字。返回的结果标识一个或多个搜索引擎,例如,网站和 / 或搜索引擎索引的身份。

[0037] 提交给搜索引擎元数据 205 的查询这里被称为搜索元数据查询。在一个实施例中,搜索元数据查询是基于用户的查询(包括用户查询中的关键词)而生成的。在一种实施方式中,转换处理本质上将用户查询转换成适用于查询其它搜索引擎的另一查询。由搜索引擎元数据 205 返回的结果可以基于对用户查询的相关性而被排名。

[0038] 基于从搜索引擎元数据 205 返回的结果,联合器 201 从若干搜索引擎 211 和 212-299 中选择要向其发出搜索查询的搜索引擎,并且然后向所选择的搜索引擎发出查询。所述选择可以选择诸如由搜索引擎服务器寄留的搜索引擎(例如网站)之类的特别的搜索引擎服务器,或者可以是搜索引擎索引。在后一种的情况中,联合器 201 通过向被配置用于搜索索引的搜索引擎提交查询来向所选择的搜索引擎提交查询。

[0039] 然后从每个所选择的搜索引擎返回搜索结果。返回的结果然后被联合器 201 合并和排名。所述排名可以基于多种技术,例如,使用诸如统一资源定位符(URL)之类的文档标识符的频率。此外,在一个实施例中,针对每个搜索引擎计算排名;排名被用于确定如何合并来自所有选择的搜索引擎的搜索结果。可以根据在共同待决的美国专利申请中所描述的方法来计算搜索引擎排名,所述美国专利申请是由 Ryan Sue 于[将被确定的日期]提交的、名称为“对文档进行排名”、序列号为[将被确定]并且转让给本申请的受让者的美国专利申请,该申请通过引用以其全部内容被结合于此,正如其在这里被充分给出一样。在一种实施方式中,按照下面参考图 5 所描述的那样来计算搜索引擎排名。

[0040] 在一个实施例中,可以通过简单地不使用搜索元数据引擎 205 来避免可能与其相关联的执行开销或者其它花费。替代地,搜索查询被简单地提交给若干搜索引擎的组中的所有搜索引擎。然而,通过此措施实现的执行节约也具有必须针对若干搜索引擎中的所有搜索引擎执行端用户查询的额外的花费,其中若干搜索引擎中的一些搜索引擎可能返回缺少与端用户的查询的相关性的内容或其它结果。

[0041] 示例搜索结果呈现

[0042] 这里,搜索联合器 201 向搜索元数据引擎 205 发出对于“鸡肉”的搜索元数据查询,搜索元数据引擎 205 返回标识与对于“鸡肉”的端用户查询相关的搜索引擎的搜索结果。

[0043] 搜索联合器 201 通过由元数据引擎 205 返回的相关搜索引擎发出“鸡肉”查询。从相关搜索引擎“1”和“2”返回的搜索结果被合并成搜索结果组,以便呈现给端用户。从所有相关搜索引擎合并的搜索结果可以以一种或多种方式呈现给端用户。

[0044] 图 3A 示出根据本发明的一个实施例的返回的搜索结果的示例 300。根据一个实施例,当执行用户的查询时,响应于用户偏好,结果呈现 311、结果呈现 312 或者二者可以被呈现给用户。响应于用户的查询,相关搜索引擎“1”和“2”返回很多(例如十个)结果。应当理解,其它数量的搜索引擎可以是相关的并且可以返回任意数量的结果。

[0045] 结果呈现 311 将根据与用户的查询的相关性排名的、来自若干搜索引擎的相互混合的结果呈现为单个列表。因此,一个实施例向用户呈现系统 200(图 2)认为最相关的结果。结果呈现 312 呈现出由选择的若干搜索引擎中的各个搜索引擎分组的结果。按照此格式,用户可以更好地能够可视地观察和识别哪个搜索引擎提供了与查询或用户的希望最相



关的结果。

[0046] 图 3B 示出跨相关搜索引擎合并的搜索结果 311B 的示例截屏。在一种实施方式中, 跨相关搜索引擎合并的搜索结果可以包括对于结果的或多或少“标准的”视图, 这可以根据用户的偏好来改变。

[0047] 图 3C 示出被分组的搜索结果 312C 的示例截屏, 搜索引擎是根据这样的分组返回各组搜索结果的。在一种实施方式中, 被分组 (搜索引擎是根据这样的分组返回各组搜索结果) 的搜索结果可以包括对于结果的或多或少探索性的, 例如“探索 (exploration)”视图, 这可以根据用户的偏好来改变。

[0048] 图 3D 示出包括来自单个搜索引擎的结果的搜索结果 313D 的示例截屏, 所述单个搜索引擎诸如可以是例如用户在考虑了一个或多个探索截屏 (例如搜索结果 312C ; 图 3C)、合并的结果的单个列表 (例如搜索结果 311B ; 图 3B) 或者二者都考虑之后所选择、指定等的搜索引擎。在一种实施方式中, 当用户发现 (例如, 根据先前的执行和 / 或对其搜索结果分组的考虑发现) 特别的搜索引擎返回最相关的结果 (例如, 从该用户的观点来看为最相关的结果) 时, 可以从单个搜索引擎寻找搜索结果。被选择用于呈现结果的搜索引擎可以根据用户的偏好来改变。

[0049] 示例硬件概述

[0050] 图 4 示出可以基于其实施本发明的示例计算机系统 400。计算机系统 400 包括总线 402 或用于传递信息的其它通信机构, 以及与总线耦合用于处理信息的处理器 404。计算机系统 400 还包括耦合到总线 402 用于存储处理器 404 将要执行的信息和指令的主存储器 406, 主存储器 406 诸如是随机存取存储器 (RAM) 或者其它动态存储设备。主存储器 406 还可以用于存储在处理器 404 执行要被执行的指令时的临时变量或其它中间信息。计算机系统 400 还包括耦合到总线 402 用于存储用于处理器 404 的静态信息和指令的只读存储器 (ROM) 408 或其它静态存储设备。诸如磁盘或光盘之类的存储设备 410 被提供并且耦合到总线 402 以存储信息和指令。

[0051] 计算机系统 400 可以经由总线 402 耦合到诸如液晶显示器 (LCD)、阴极射线管 (CRT) 等等用于向计算机用户显示信息的显示器 412。包括字母数字和其它键的输入设备 414 被耦合到总线 402 以向处理器 404 传递信息和指令选择。另一种类型的用户输入设备是诸如鼠标、跟踪球或光标导向键之类用于向处理器 404 传递方向信息和指令选择以及用于控制光标在显示器 412 上的移动的光标控制 416。这种输入设备通常具有两个轴的自由度, 第一轴 (例如 x) 和第二轴 (例如 y), 这使得设备指定平面中的位置。本发明的示例实施例涉及计算机系统 400 用于将对于源数据库系统处的数据块做出的改变复制成备用数据库系统处的数据块的复制品。根据本发明的一个实施例, 将对于源数据块系统处的数据块做出的改变复制成备用数据库系统处的数据块的复制品是由一个或多个实例的计算机系统 400 响应于处理器 404 执行在主存储器 406 中所包含的一个或多个指令而提供的。这样的指令可以从诸如存储设备 410 之类的另一计算机可读介质读入主存储器 406 中。包含在主存储器 406 中的指令序列的执行使得处理器 404 执行这里描述的处理步骤。多处理配置中的一个或多个处理器也可以被采用来执行包含在主存储器 406 中的指令序列。在替代实施例中, 硬线电路可以用来替代软件指令以实施本发明, 或者与软指令组合来实施本发明。因此, 本发明的实施例不限于硬件电路和软件的任何特定组合。

[0052] 这里所使用的术语“计算机可读介质”涉及参与向处理器 404 提供以便执行的指令的任何介质。这样的介质可以采取多种形式,包括但不限于非易失性介质、易失性介质和传输介质。非易失性介质例如包括诸如存储设备 410 之类的光盘或磁盘。易失性介质包括诸如主存储器 406 之类的动态存储器。传输介质包括同轴电缆、铜线和光纤,包括包含总线 402 在内的线缆。传输介质还可以采用诸如在无线电波和红外数据通信期间所生成那些声波或光波的形式。

[0053] 普通形式的计算机可读介质例如包括软盘、柔性盘、硬盘、磁带,或任何其它磁介质、CD-ROM、任何其它光介质、穿孔卡、纸带、具有洞图案的任何其它遗留物或其它物理介质以及其它计算机可读介质、RAM、PROM,和 EPROM、FLASH-EPROM、任何其它存储芯片或存储盒、下文描述的载波,或计算机可读的任何其它介质。

[0054] 在将一个或多个序列的一个或多个指令携带到处理器 404 以便执行的过程中可以涉及多种形式的计算机可读介质。例如,指令可以最初承载在远程计算机的磁盘上。远程计算机可以将指令加载到其动态存储器中并且通过利用调制解调器的电话线来发送指令。计算机系统 400 本地的调制解调器可以接收电话线上的数据并且使用红外发射器将数据转换成红外信号。耦合到总线 402 的红外探测器可以接收红外信号中携带的数据并且将数据置于总线 402 上。总线 402 将数据携带到主存储器 406,处理器 404 从主存储器 406 检索指令并执行指令。由主存储器 406 接收的指令可以可选地在被处理器 404 执行之前或者之后被存储在存储设备 410 上。

[0055] 计算机系统 400 还包括耦合到总线 402 地通信接口 418。通信接口 418 耦合到网络链路 420 的双向数据通信,网络链路 420 被连接到本地网络 422。例如,通信接口 418 可以是综合服务数字网络 (ISDN) 卡、电缆或者数字订户线 (DSL) 调制解调器或用于向相应类型的电话线提供数据通信连接的另一调制解调器。作为另一示例,通信接口 418 可以是用于向兼容 LAN 提供数据通信连接的局域网 (LAN) 卡。还可以实行无线链路。在任何这样的实施方式中,通信接口 418 发送和接收携带有表示各种类型的信息的数字数据流的电信号、光电信号或光信号。

[0056] 网络链路 420 通常通过一个或多个网络向其它数据设备提供数据通信。例如,网络链路 420 可以通过本地网络 422 向主机计算机 424 或者向由因特网服务提供商 (ISP) 426 操纵的数据装备提供连接。ISP 426 进而通过现在通常被称为“因特网”的全世界范围分组数据通信网络 428 提供数据通信服务。本地网络 422 和因特网 428 二者都使用携带有数字数据流的电信号、光电信号或光信号。通过各种网络的信号以及网络链路 420 上或者通过通信接口 418 的信号是传送信息的载波的示例性形式,其中,这些信号携带有到计算机系统 400 或者来自计算机系统 400 的数字数据。

[0057] 计算机系统 400 可以通过网络(一种或多种)、网络链路 420 和通信接口 418 发送消息和接收数据,包括程序代码。在因特网的示例中,服务器 430 可以通过因特网 428、ISP 426、本地网络 422 和通信接口 418 发送应用程序请求的代码。如这里所述,根据本发明,一种这样的下载的应用提供对源数据库系统处的数据块做出的改变到备用数据库系统处的数据块的复制品的复制。

[0058] 接收的代码可以由处理器 404 按照接收那样来执行,和/或被存储在存储设备 410 或者其它非易失性存储装置中以供以后执行。按照这样的方式,计算机系统 400 可以获得

载波形式的应用代码。

[0059] 来自若干搜索引擎的结果的示例合并

[0060] 图 5 示出根据一个实施例的用于合并来自若干搜索引擎的结果的示例处理 500。处理 500 以对多个文档进行排名 501 开始,所述多个文档中的每一个与一个统一资源定位符 (URL) 相关联。所述文档是作为由多个搜索引擎中的每个搜索引擎计算的查询的结果而被返回的。所述多个文档中的至少一个文档是作为来自每个搜索引擎的查询的结果而被返回的。

[0061] 对于与多个文档中的文档相关联的每个 URL, URL 的频率被确定 502。上面 (块 501) 描述的排名包括基于一个或多个因素对多个文档进行排名。这些因素基于针对与多个文档相关联的每个 URL 确定的频率。

[0062] 在一个实施例中,计算与响应于通过多个 (例如一组) 搜索引擎 ‘S’ 执行的查询而返回的多个文档中的文档 ‘i’ 相关联的每个 URL ‘u’ 的频率  $f_{url}(u_i, S)$ , 以进行 URL 频率确定 502。在一个实施例中,根据下面的公式 1 来计 URL 的频率:

$$[0063] \quad f_{url}(u_i, S) = |s_i \supset u_i| / |S|$$

[0064] (公式 1)

[0065] 其中,  $f_{url}$  是 URL 频率,  $|S|$  是执行查询的搜索引擎的总数, 以及  $|s_i \supset u_i|$  是返回包含 URL,  $u_i$  的结果的搜索引擎的总数。

[0066] 在一个实施例中,文档的排名是根据其在的一组搜索引擎结果中的罕见度 (rarity) (例如相对于其它 URL 的频率) 来量度的。从搜索引擎返回的搜索结果中的文档的罕见度有效地包括对于文档是多么感兴趣的衡量。在一个实施例中,文档的排名因此与倒数的 URL 频率相对应。根据这样的实施例执行的搜索因此可以返回非常专用的、不明显的、常常被其它搜索引擎忽略的文档。这样的搜索的结果因此很可能返回一般的,例如非专用搜索不能返回的结果。

[0067] 例如,一般的网络搜索通常返回常访问的文档,并且一般的网络搜索是以那些常访问的文档具有相对高的相关性排名而完成的。此外,即使标准网络搜索确实设法返回了相对罕见的文档,其罕见度将很可能使其具有相对低的相关性排名。低的相关性排名很可能将该文档“埋藏”在搜索结果内可能被许多其它搜索结果遮蔽它的地方。这本质上可能迫使用户在搜索结果的页面中找寻它,这可能是单调乏味并且低效的。

[0068] 在文档在搜索结果中的排名基于其相对罕见度的一个实施例中,针对查询返回的排名是正规化的排名 ( $R_{normalized}$ ) 与 URL  $u_i$  的频率  $f_{url}$  的倒数的函数  $f$ , 根据下面的公式 2 来计算:

$$[0069] \quad R_{overall}(u_i, S) = f(R_{normalized}(u_i, S), 1/f_{url}(u_i, S))$$

[0070]

(公式 2)

[0071] 其中,正规化的排名 ( $R_{normalized}$ ) 表示与对组 S 的搜索引擎的本地排名进行组合有关的所有其它计算。在一个实施例中,与文档的排名相关联的其它计算包括但不限于对搜索引擎中完全相同的结果、点击率、用户评价的处理。

[0072] 实施例不限于根据文档在一组搜索引擎结果中的罕见度来对文档进行排名。例如,在一个实施例中,文档的相关性根据该文档在一组搜索引擎结果中多么常见来量度,这可以是非倒数的与 URL 相关联的频率的函数。根据这样的实施例执行的搜索因此可以返回

反映其它搜索者、搜索引擎创建者、管理员和其它参与者常见的看法、多数人的意见等（例如利用垂直搜索）。

[0073] 可以利用诸如已经被之前的用户定制的垂直搜索之类的更专用的搜索的结果。这可以使得根据这样的实施例执行的搜索得到的结果能够反映已经被证明有用或大众的搜索结果，并且因此利用别人的看法、教训和经验，结合诸如集体的知识和智慧之类有用的社会因素，从而促进搜索引擎使用中的高效率和经济实惠。

[0074] 在文档的相关性根据该文档在一组搜索引擎结果的搜索结果中多么常出现来量度一个实施例中，针对查询返回的搜索结果中的文档的总排名是正规化的排名 ( $R_{\text{normalized}}$ ) 与非倒数的 URL  $u_i$  的频率  $f_{\text{url}}$  的函数  $f$ ，根据下面的公式 3 来计算：

$$[0075] \quad R_{\text{overall}}(u_i, S) = f(R_{\text{normalized}}(u_i, S), f_{\text{url}}(u_i, S))$$

[0076] (公式 3)

[0077] 其中，正规化的排名 ( $R_{\text{normalized}}$ ) 表示与对组  $S$  的搜索引擎的本地排名进行组合有关的所有其它计算。在一个实施例中，与正规化的排名相关联的其它计算包括但不限于对搜索引擎中完全相同的结果、点击率、用户评价的处理。

[0078] 对每个 URL 的总排名被分配 503，在一个实施例中，基于简单实施诸如对频率的列表进行分类之类的总排名函数进行分配。

[0079] 例如，对于针对三个单独的搜索引擎 1、2 和 3 执行的示例查询返回的结果（例如 URL）按照下面表 1 所示那样被排名（501）。

[0080]

表 1

<u>搜索引擎 1 结果</u>	<u>搜索引擎 2 结果</u>	<u>搜索引擎 3 结果</u>
A	A	D
B	D	F
[0081] C	E	G

[0082] 响应于查询返回的各个 URL 的频率被确定（502），例如利用上面的公式 1 来计算。在下面的表 2 中示出示例频率的列表。

[0083]

表 2

<u>URL</u>	<u>频率</u>
A	0.66
B	0.33
C	0.33
D	0.66
E	0.33
F	0.33
G	0.33

[0084] 为了对每个 URL 分配 (503) 总排名, 频率列表被分类。当被分类时, 表 2 中所示的示例频率按照下面表 3 中所示那样被排名。可以利用上面的公式 2 和 / 或公式 3 或者根据另一种排名模型来确定排名。

[0085] 表 3

[0086]

<u>URL</u>	<u>排名</u>
A	1
D	2
B	3
C	4
E	5
F	6
G	7

[0087] 这些排名例如可以被呈现给用户和 / 或被呈现用于进一步的处理。

[0088] 在一个实施例中, 使用 URL 频率和 / 或搜索引擎排名来合并 504 从若干搜索引擎返回的结果。例如, 如下面表 4 中所示, 可以将利用各个搜索引擎返回的 URL 的频率进行加和。

[0089]

表 4

搜索引擎 1 结果:  $0.66 + 0.33 + 0.33 = 1.33$

搜索引擎 2 结果:  $0.66 + 0.66 + 0.33 = 1.66$

搜索引擎 3 结果:  $0.66 + 0.33 + 0.33 = 1.33$

[0090] 然后,可以根据搜索引擎结果的加和的 URL 频率对这些搜索引擎结果进行排名。根据表 4 中所示的示例结果进行排名,搜索引擎 2 结果排名第一,搜索引擎 1 结果排名第二,并且搜索引擎 3 结果排名第三(例如,可以使用一种或多种其它标准来在同样得分的结果之间进行区分)。因此,根据分组格式(例如,分组呈现格式 312;图 3A)呈现,示例结果如下面表 5 所示。

[0091] 表 5

[0092]

### 搜索引擎结果 2

A

D

E

### 搜索引擎结果 1

A

B

C

### 搜索引擎结果 3

D

F

G

[0093] 等同物、扩展、替换 & 混合

[0094] 因此,描述了涉及可读物理备用数据库系统的示例实施例。在前面的说明中,已经参考大量特定细节描述了本发明的实施例,所述特定细节可以依赖于实施方式而有所变化。因此,对于本发明是什么以及申请人希望本发明是什么的唯一并且排他的指示是从本申请得出的特定形式的权利要求组(这些权利要求是在特定形式中得出的),包括任何随后的修正。这里对于在这些权利要求中包含的术语明确提出的任何定义应当按照权利要求中所使用那样来决定这些术语的含义。因此,没有在权利要求中明确叙述的限定、元素、特

性、特征、优点或属性不应当以任何方式限制这些权利要求的范围。说明书和附图因此被认为是说明性的,而非具有限制意义。

## 示例处理 100

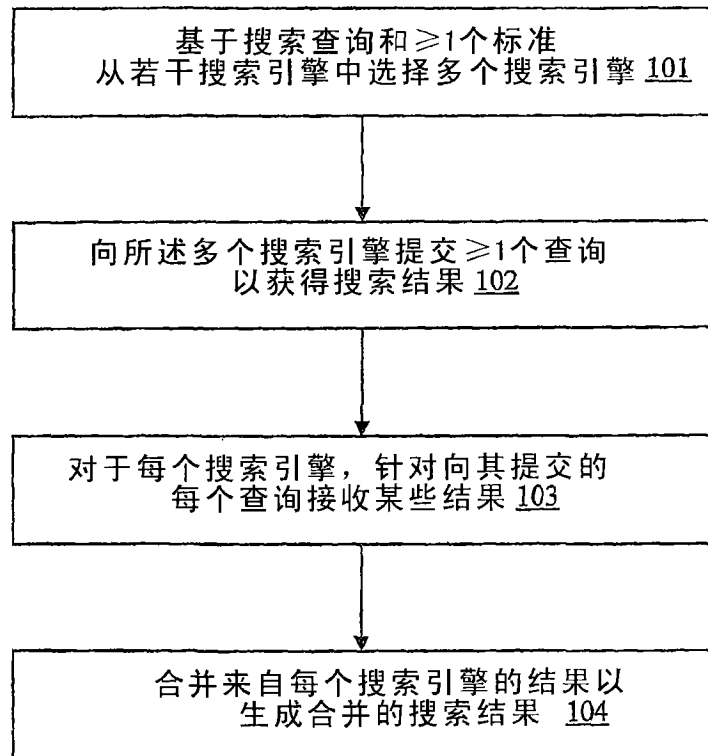


图 1



示例系统 200

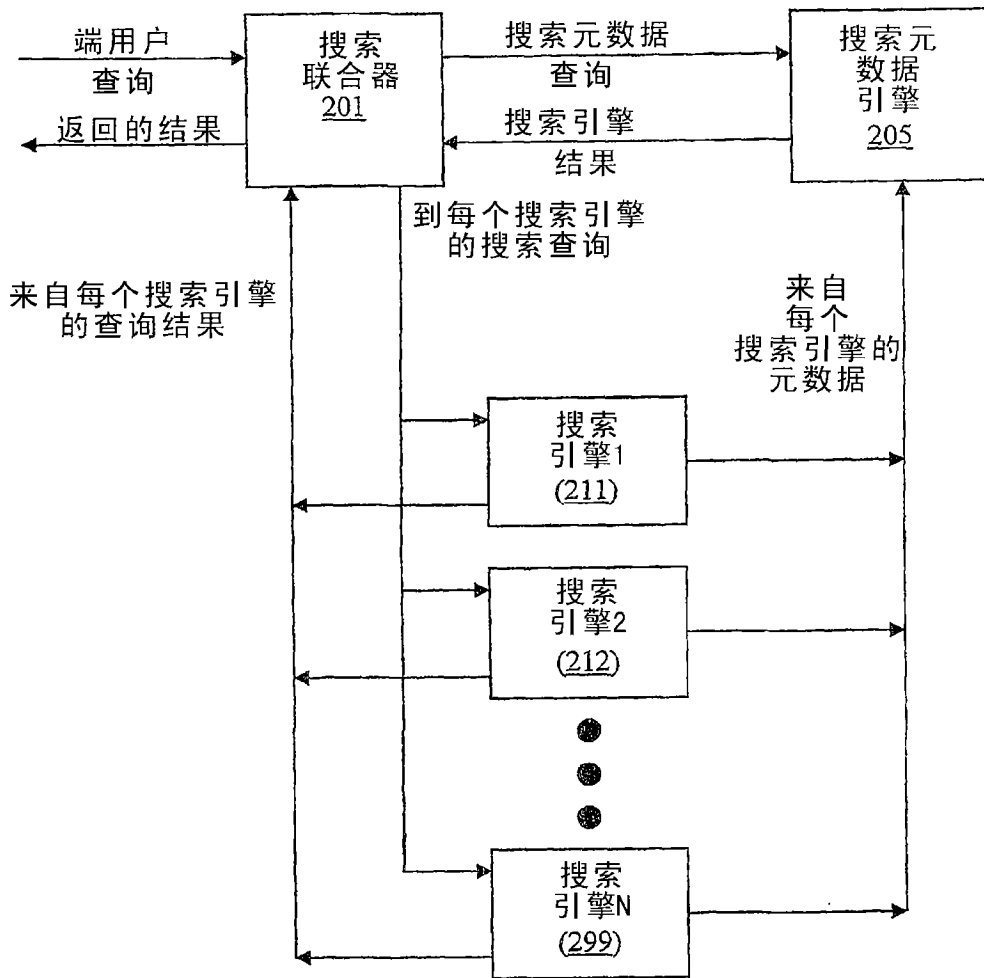


图 2

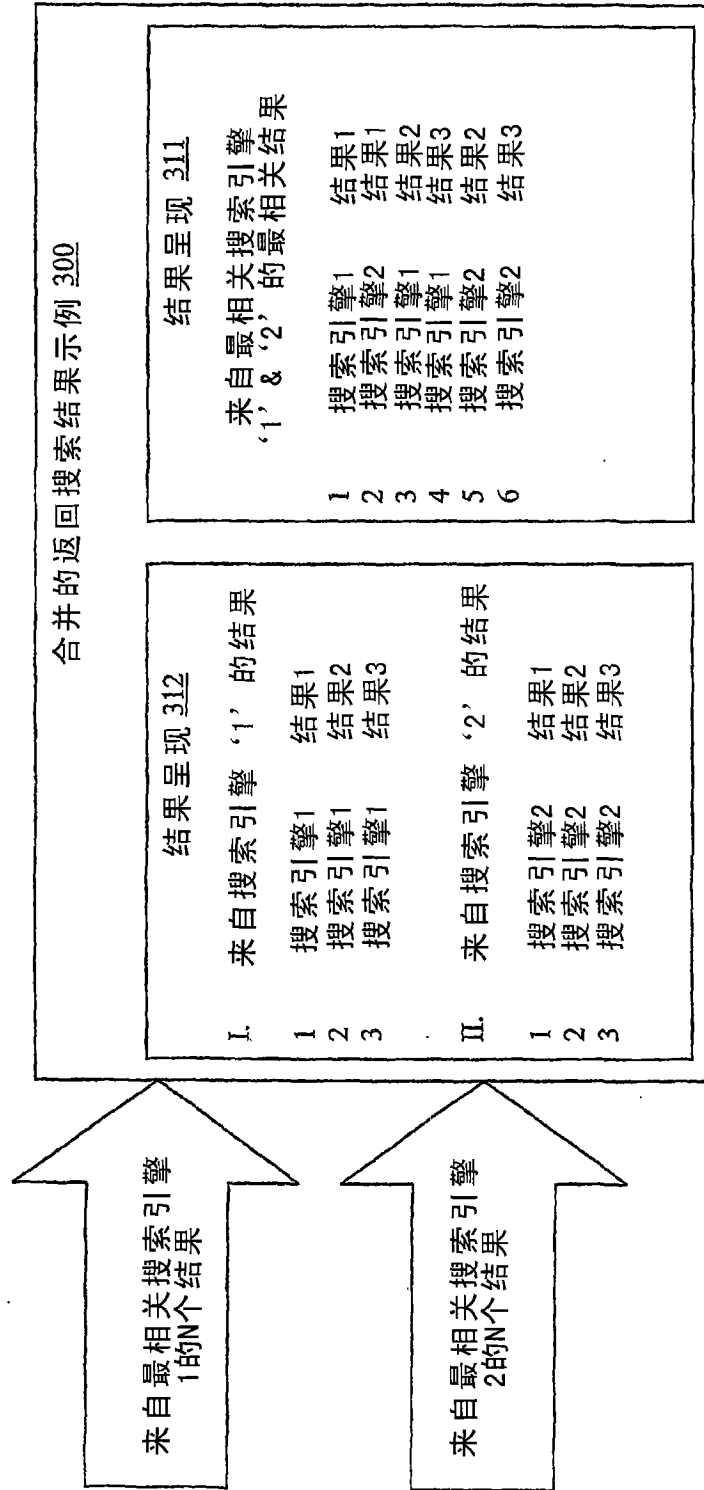


图 3A

311B

多个相关搜索引擎的合并的搜索结果（标准视图）

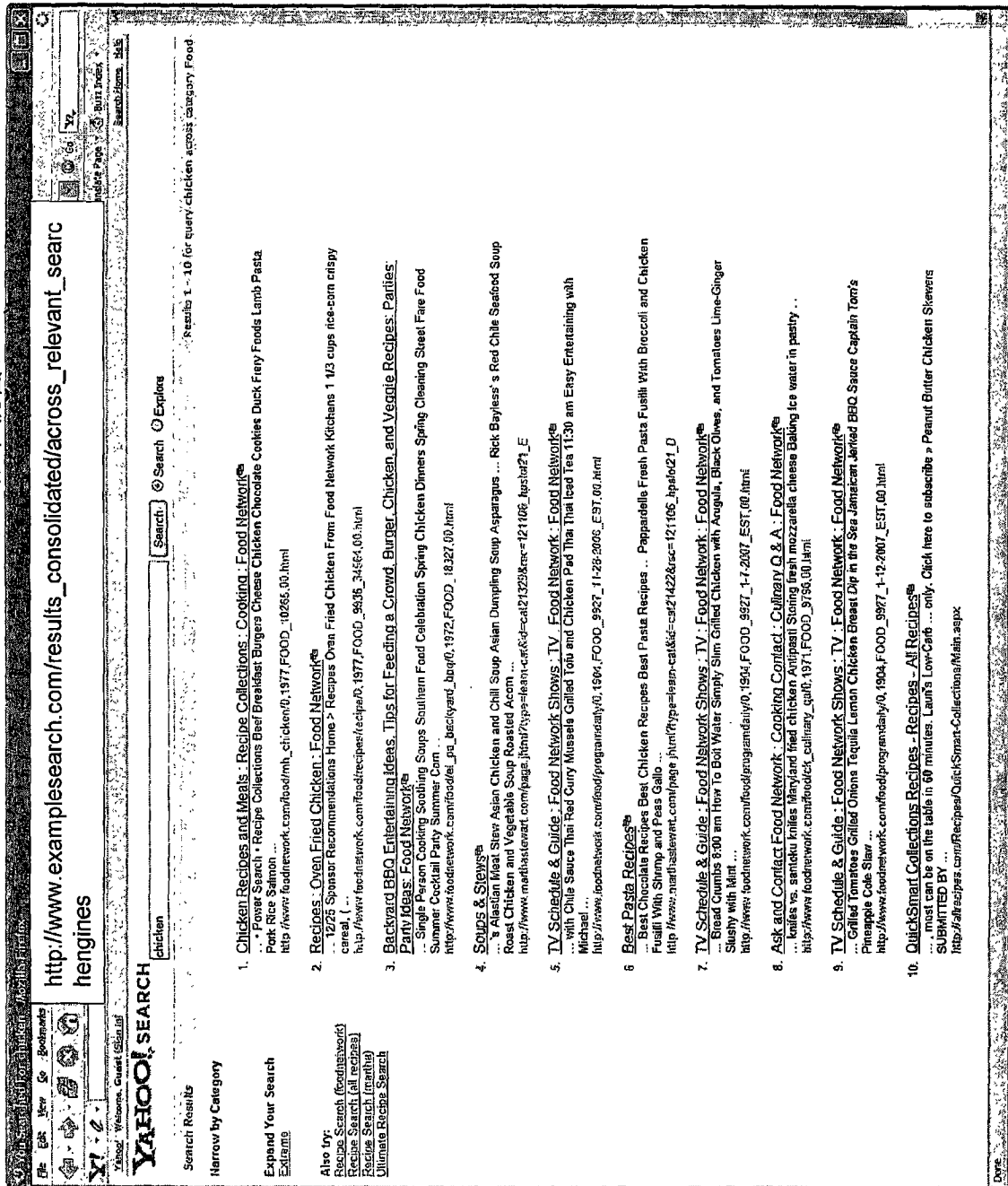


图 3B

由搜索引擎分组的搜索结果（探索视图）

312C

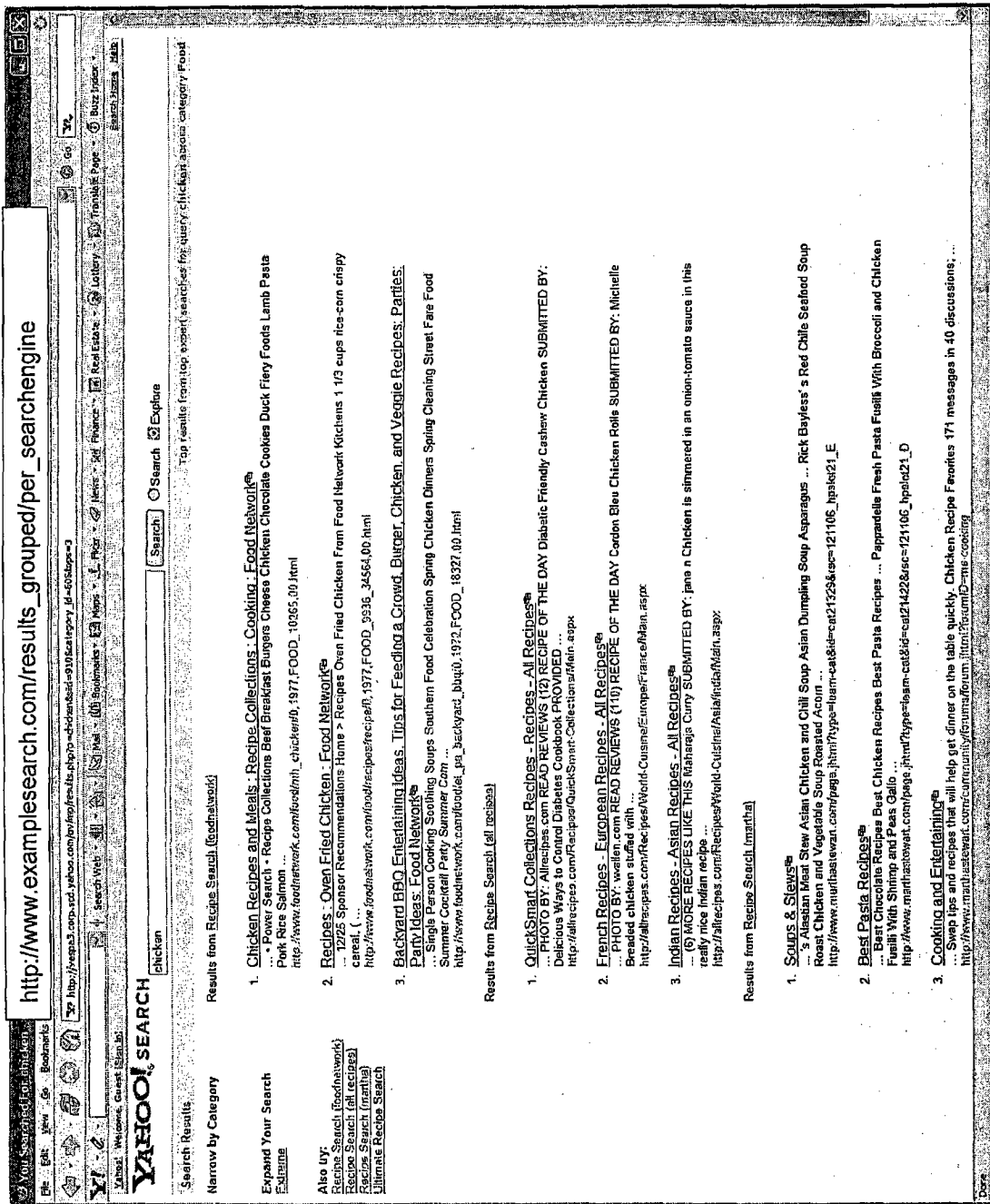


图 3C

来自单个引擎的搜索结果（用户在查看联合结果之后所选择的）

313D

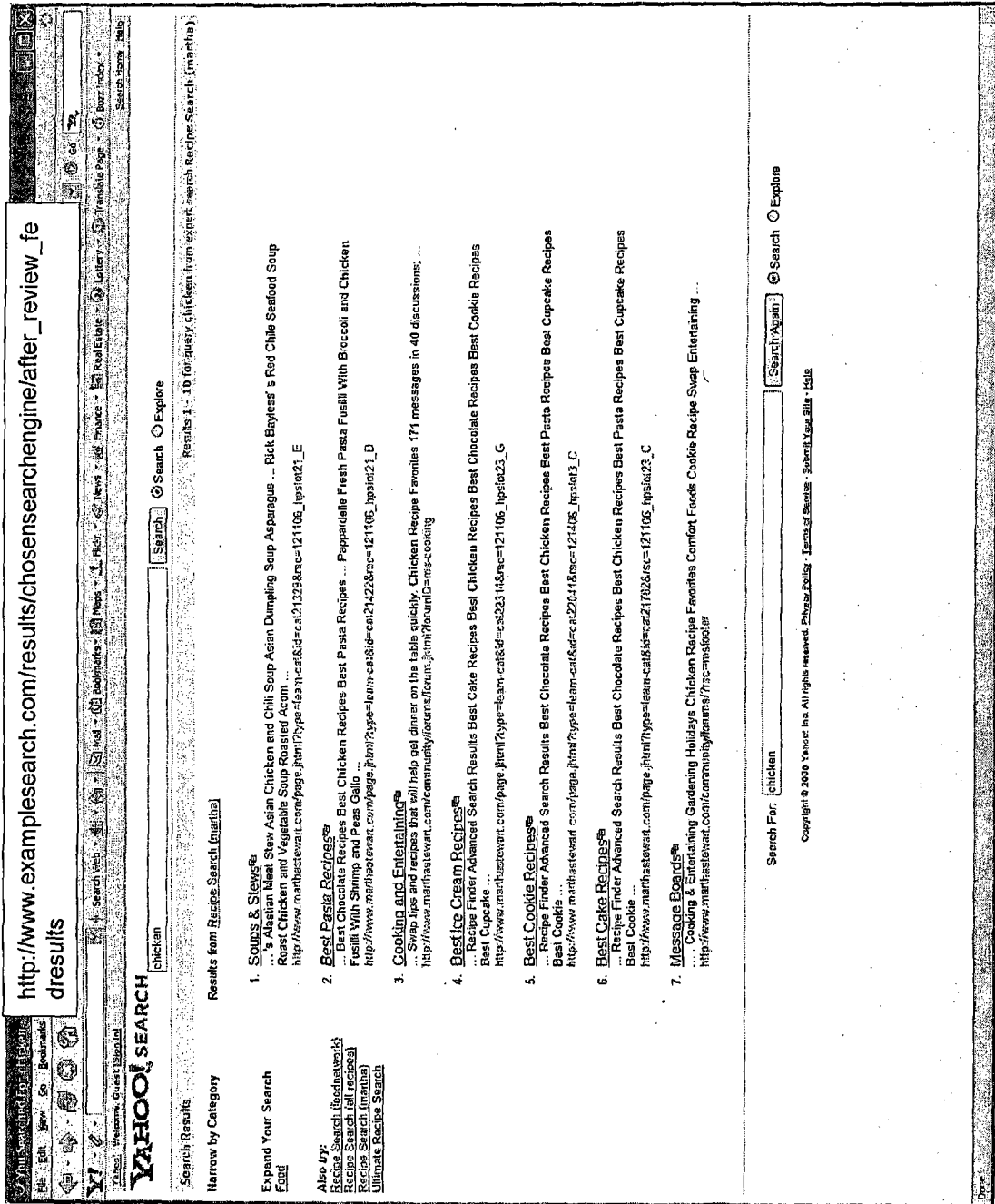


图 3D

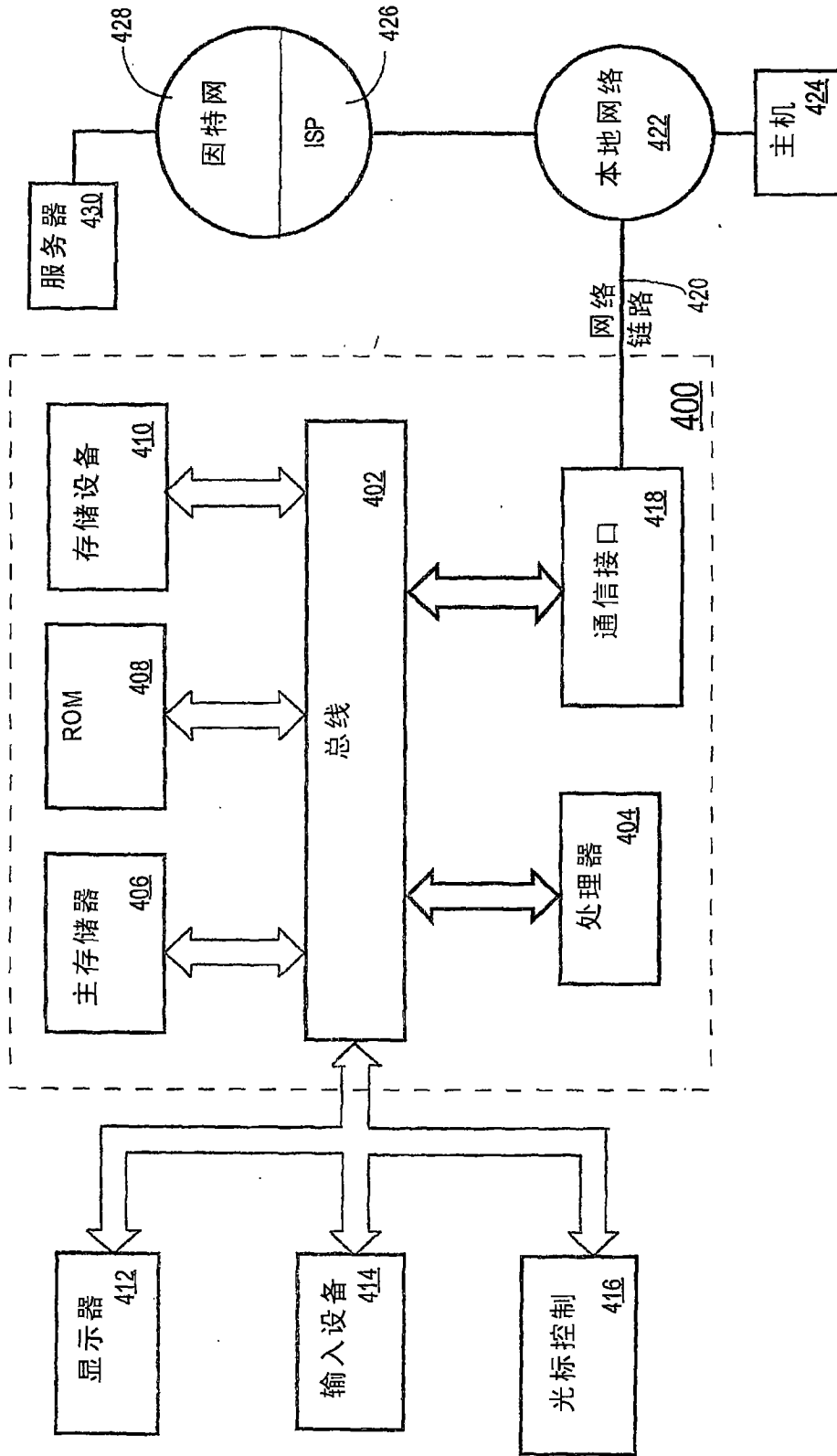


图 4

## 用于合并来自若干搜索引擎的结果的示例处理500

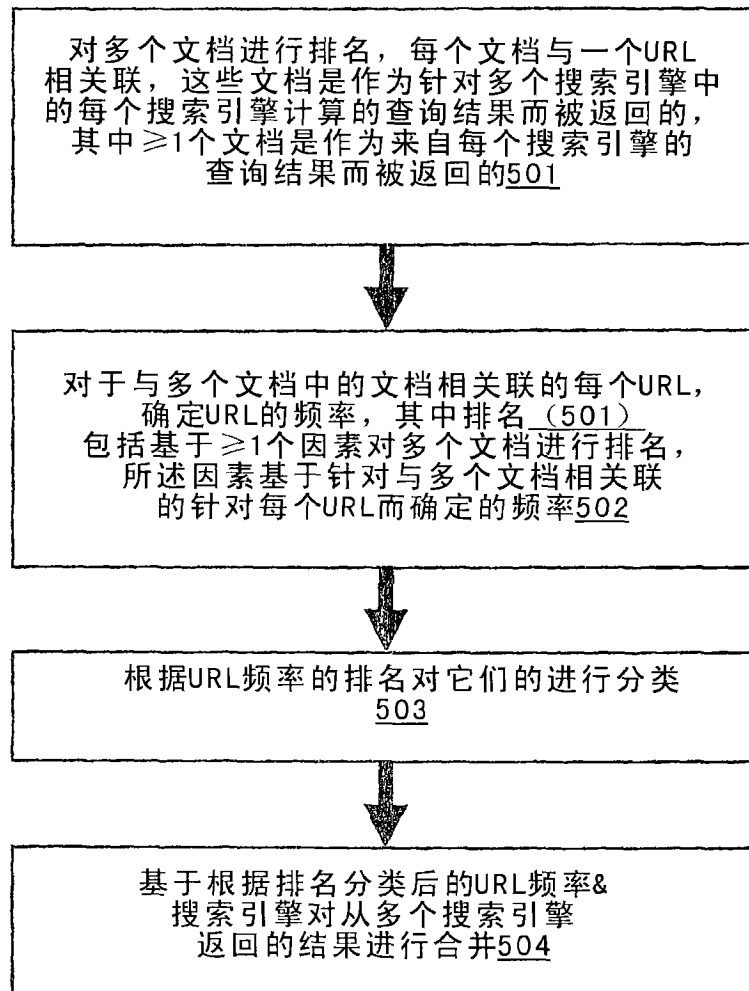


图 5