



# [12] 发明专利说明书

专利号 ZL 200480012192.2

[45] 授权公告日 2009年4月22日

[11] 授权公告号 CN 100481070C

[22] 申请日 2004.3.31  
 [21] 申请号 200480012192.2  
 [30] 优先权  
     [32] 2003.5.8   [33] US [31] 10/434,845  
 [86] 国际申请 PCT/US2004/010020 2004.3.31  
 [87] 国际公布 WO2004/102427 英 2004.11.25  
 [85] 进入国家阶段日期 2005.11.7  
 [73] 专利权人 甲骨文国际公司  
     地址 美国加利福尼亚州  
 [72] 发明人 沙米姆·阿尔法 保罗·狄克逊  
 [56] 参考文献  
     US6282538B1 2001.8.20  
     US6032145A 2000.2.29  
     US6324534B1 2001.11.27

Cooperative Answering through Controlled Query Relaxation. Terry Gaasterland. IEEE Expert: Intelligent Systems and Their Applications, Vol. 12 No. Issue 5. 1997

Restricting query relaxation through user constraints. Terry Gaasterland. Intelligent and cooperative information systems. 1993

审查员 张露薇

[74] 专利代理机构 北京康信知识产权代理有限公司  
 代理人 余刚

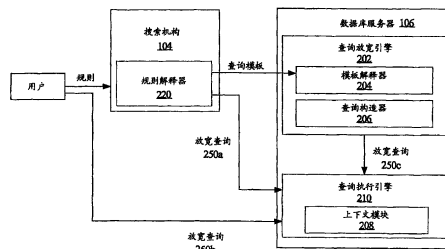
权利要求书4页 说明书20页 附图4页

## [54] 发明名称

搜索条件的逐渐放宽

## [57] 摘要

一种信息搜索机构，用于为请求搜索的用户提供对搜索查询的放宽的控制。通过这种机构，用户可以指定与主搜索条件的变化相关的子查询序列，并且指定执行子查询的进展。因此，用户可以告知其关于在放宽主搜索条件中使用的搜索项变化的优先权，用户还可以告知他们对可以通过特定子查询返回的结果的关联性的想法。如果有，由数据库服务器以基于由用户指定的进展的顺序，接收包括子查询序列的查询并且执行子查询。通过减少数据库服务器的不必要工作并通过减少在客户端和服务器的昂贵通信以及相关处理，来改善响应时间和网络加载。



1. 一种用于管理信息搜索的方法,所述方法包括计算机执行的以下步骤:

接收搜索请求查询,所述搜索请求查询包括子查询序列并指定执行所述子查询的进展,其中,多个所述子查询序列中除第一子查询之外的每个子查询在所述序列中均跟随一个或多个先前子查询;

基于所述进展,执行所述子查询序列中的第一子查询;

以由所述进展指定的顺序执行除所述第一子查询之外的一个或多个所述子查询,其中,只有当先前执行的一个或多个子查询的结果不足以实现搜索请求时,才自动执行除所述第一子查询之外的所述一个或多个子查询中的每个; 以及

提供所述子查询序列中的一个或多个子查询的结果。

2. 根据权利要求1所述的方法,还包括以下计算机执行的步骤:

接收对所述搜索的特定数量结果的请求; 以及

其中,只有当所述一个或多个先前子查询没有实现所述特定数量结果的请求时,才执行除所述第一子查询之外的所述一个或多个子查询中的每个。

3. 根据权利要求1所述的方法,其中,所有所述子查询均基于相应搜索条件从数据库中选择数据,并且其中,所述进展对应于与所述搜索相关的所述搜索条件的放宽。

4. 根据权利要求1所述的方法,其中,所述进展由请求所述搜索的终端用户指定。

5. 根据权利要求3所述的方法,其中,所述进展通过与用户接口的交互来指定。
6. 根据权利要求1所述的方法,其中,所述执行一个或多个子查询的步骤包括以下步骤:
  - 提供所述第一子查询的结果;
  - 接收提供更多结果的请求;
  - 响应于接收到的查询以提供更多结果并基于所述进展,执行所述子查询序列的第二子查询; 以及
  - 提供所述第二子查询的结果。
7. 根据权利要求1所述的方法,其中,所述接收步骤包括接收与所述子查询序列相关的搜索条件以及所述进展作为 XML 文档。
8. 根据权利要求1所述的方法,其中,所述接收步骤包括:接收来自基于一组一个或多个规则构造所述查询的应用程序的所述搜索请求查询,其中,所述组由请求所述信息的终端用户指定。
9. 根据权利要求8所述的方法,其中,所述应用程序是电子搜索引擎。
10. 根据权利要求1所述的方法,其中,所述接收步骤包括接收由终端用户指定的一组搜索条件和基于所述搜索条件构造数据库查询的请求的步骤; 以及其中,所述方法还包括以下计算机执行的步骤:
  - 基于所述搜索条件构造数据库查询。

11. 根据权利要求1所述的方法,其中,所述接收的步骤包括:接收指定与文档的一个或多个部分相关的搜索条件的所述搜索请求查询,并且接收指定对所述文档的一个或多个部分执行与所述搜索条件相关的子查询的顺序的进展。
12. 一种用于管理信息搜索的方法,所述方法包括以下计算机执行的步骤:
  - 指定与正被搜索的所述信息相关的搜索条件;
  - 指定执行与所述搜索条件相关的子查询的顺序; 以及
  - 传输包括所述搜索条件和所述顺序的查询,用于以所指定的顺序通过数据库服务器执行一个或多个所述子查询,其中,多个所述子查询序列中除第一子查询之外的每个子查询在所述序列中均跟随一个或多个先前子查询, 以及
  - 其中,只有当一个或多个先前子查询的结果不足以实现所述搜索时,才自动按所述顺序执行除第一子查询之外的一个或多个子查询中的每个。
13. 根据权利要求12所述的方法,其中,所述指定步骤由请求搜索信息的终端用户启动。
14. 根据权利要求12所述的方法,其中,所述指定搜索条件的步骤包括指定一组信息值,并且其中,所述指定顺序的步骤包括指定与所述组信息值相关的有序的一组一个或多个规则。
15. 根据权利要求14所述的方法,其中,所述有序的一组一个或多个规则包括逻辑运算符。
16. 根据权利要求12所述的方法,其中,所述指定步骤包括以XML文档的形式指定。

17. 根据权利要求 12 所述的方法，其中，所述指定搜索条件的步骤包括指定与文档的一个或多个部分相关的搜索条件，并且其中，所述指定顺序的步骤包括指定对所述文档的所述一个或多个部分执行与所述搜索条件相关的子查询的顺序。
18. 根据权利要求 17 所述的方法，其中，所述指定搜索条件的步骤包括指定与文档的至少两个特定部分相关联的搜索条件，并且其中，所述指定顺序的步骤包括指定以特定顺序对所述文档的所述至少两个特定部分执行与所述搜索条件相关联的子查询。
19. 一种用于管理信息搜索的计算机设备，所述设备包括：

用于接收搜索请求查询的数据库服务器，所述搜索请求查询包括子查询序列并指定执行所述子查询的进展，其中，多个所述子查询序列中除第一子查询之外的每个子查询在所述序列中均跟随一个或多个先前子查询；

用于基于所述进展执行所述子查询序列中的第一子查询的装置；以及

用于以由所述进展指定的顺序执行除所述第一子查询之外的一个或多个所述子查询的装置，其中，只有当一个或多个所述先前子查询的结果不足以实现搜索请求时，才自动地执行除所述第一子查询之外的所述一个或多个子查询中的每个。

## 搜索条件的逐渐放宽

### 技术领域

本发明主要涉及信息检索，并且尤其涉及用于指定搜索条件的逐渐放宽的技术。

### 背景技术

服务于客户端或终端用户的应用、查询通常操作，以不仅识别用户指定的查询的准确匹配（如果存在），而且识别最接近的非准确匹配，并且返回给终端用户。在本文中，术语“查询”不限于传统的数据库查询，诸如 SQL（结构化查询语言）查询。通常，查询包括通过任何搜索机构（诸如传统搜索引擎或搜索函数）对信息的任何搜索。典型地，用户的搜索请求最终被转换成结构化数据库查询。

一种使用搜索请求或查询以识别现有准确匹配和非准确匹配的方法，包括：（1）重写或重新构造用户请求以包括原始查询的所有允许变化；（2）通过将查询提交到数据库服务器，检索用于重新构造的连续查询的“命中目录（hit-list）”；以及（3）以基于原始搜索条件的相关性（有时称为“相关性等级”）的顺序对命中目录进行排序。

例如，如果用户在某个形式的信息仓库（诸如，数据库或通过互联网可访问的信息的集合）上启动对关于“cheap pen”的信息的搜索，构造“扩展查询”，以包括原始查询和一个或多个放宽原始

查询要求的子查询。与对“cheap pen”的搜索相关的扩展查询可能包括用于“cheap pen”的其他允许形式的子查询，诸如“cheap NEAR pen”、“cheap AND pen”、“\$cheap AND \$pen”（其中，“\$”代表语法词干操作）、“cheap OR pen”等。基于该组子查询产生命中目录，然后对命中目录进行排序。排序可以基于，例如，产生给定命中和对请求信息的终端用户的推测相关性的特定子查询。

在这样的方法中，因为执行所有的子查询，所以可能执行无用的工作，不管是否必需实际实现用户请求和利益。即，执行的第一子查询可以产生足够数量的命中或充分相关的结果，以满足用户的利益。而且，如果给定子查询是明显非选择性的，那么可以产生比满足用户利益所必需的更多的命中，通过解析查询语句、查询信息仓库、以及产生结果并对结果排序来执行非必要工作。

另一种方法包括：（1）按照优先级的降序，顺序执行与原始查询的允许变化相关的子查询；以及（2）基于某个条件检索命中，直到定位了足够的命中。这种方法包括：除数据库服务器以外的实体，诸如终端用户或搜索机构，基于原始搜索条件，将查询发布到数据库服务器；接收来自查询的结果；将另一查询发布到扩展原始搜索条件的数据库服务器；接收结果；以及继续该迭代过程，直到根据某个定量条件已经满足搜索请求。当由于实体和数据库服务器（其不需要加载系统）之间的多重完全往返通信的可能而使用该方法时，经历查询响应时间和网络执行。在本文中，以及贯穿说明书，完全往返通信指的是在客户端实体和数据库服务器之间的网络通信，以及由数据库服务器执行的处理，其通常包括：（1）解析查询；（2）构造查询执行计划；（3）使查询执行最优化等。次级客户端服务器通信指的是在客户端和服务器应用程序之间的通信，其不会引起与完全往返通信相同的处理开销。

以上两种方法在响应时间、处理、和网络加载方面效率差。而且，这些方法对于搜索应用程序和机制的开发者而言是麻烦的，因为它们需要这样的应用程序和机制以合理地放宽搜索请求，并相对于相关性等级来处理结果。而且，如果有的话，它们提供有限的容量给终端用户，以随后影响在原始搜索条件的放宽的上下文中的搜索项变量的优先级，从而影响相关结果的相关性。

基于以上原因，显然希望提供一种用于提供信息搜索的改进的机构。特别需要提供对请求搜索特定信息的终端用户的更多控制，以提供更有效和更相关的执行。

## 附图说明

通过举例的方式描述本发明，但本发明并不限于此，在附图中，相同的参考标号代表相似元件，其中：

图 1 是示出可以实现实施例的操作环境的简单实例的框图；

图 2 是示出数据库服务器和搜索机构的相关功能组件的框图；

图 3 是示出用于管理信息搜索的处理的流程图；

图 4 是示出用于管理信息搜索的处理的流程图；以及

图 5 是示出可实现本发明的实施例的计算机系统的框图。

## 具体实施方式

本发明描述了一种用于指定搜索条件的逐渐放宽的方法和机构。在以下描述中，为了解释的目的，描述了大量特定细节，以提供本发明的彻底理解。但是，非常明显，在没有这些特定细节的情



况下本发明也可以实现。在其他例子中，以框图形式示出已知结构和装置，以避免不必要地模糊本发明。

## 概述

为了提供更有效的搜索机构，将对搜索查询的放宽的控制提供给请求搜索的用户。查询的逐渐放宽允许在不经过折衷执行的情况下的复杂查询。通常，处理复杂查询，以在不对结果的精确性或质量进行折衷的情况下，提供结果的增加调用或完整性。通过这样的机构，用户可以指定与主搜索项的变化相关的子查询序列，并且指定执行子查询的进展。因此，用户可以告知其关于在放宽主搜索条件中使用的搜索项变化的优先级，其进一步允许用户告知其关于可以由特定子查询返回的结果的相关性的想法。与前述方法不同，当子查询被接收时，他们不通过数据库服务器立即和完全执行。相反地，仅当先前执行的子查询还没有产生满足用户的结果时，才执行每个子查询。

在一个实施例中，子查询基于与被请求的信息相关的搜索条件，从数据库选择数据，其可以包括原始搜索项的变化和进展，根据进展，对应于原始搜索项的逐渐放宽执行子查询。

根据一个方面，由用户指定的一系列子查询，例如在服务器被接收，并且以基于由用户指定的进展的顺序执行。因此，只响应于在用户/客户端和服务器之间的单一完全往返通信，可以由服务器执行多重子查询，并且服务器只需要处理单一查询。从而改善了网络加载和响应时间。

在一个实施例中，由用户指定的一系列子查询在单一往返通信中在数据库服务器被接收。然后，服务器以基于由用户指定的进展的顺序执行子查询，然而，只有当先前执行的子查询的结果不足以

实现搜索请求时，才执行每个子查询。例如，如果用户请求特定数量的结果命中，则顺序执行子查询，直到已经产生请求命中的数量，但是不执行比所需的更多的子查询。从而，例如，如果第一子查询的执行提供了足够的命中或足够的相关命中来满足用户请求，那么不执行其他子查询，并且随后保存计算资源。在这样的实施例中，用户在提交搜索前，可以通过与用户接口的交互来指定“停止条件”（例如，最小数量的命中、时间量、数据量、或这些条件的结合）。一旦搜索结果满足停止条件，数据库服务器就不执行任何进一步的子查询。因此，这个用户指定的停止条件方法将只要求在客户端和服务器之间的一个完全往返通信，并且服务器只需要处理单一查询。

可选择地，为了满足用户利益的需要，用户可以通过与用户接口的交互，反复请求更多命中。在这种情况下，用户接口不发送新查询，而是指示服务器进行在原始接收的查询中的下一子查询。类似于用户指定的停止条件的方法，只请求一个完全往返通信和相关处理。这种方案可以要求在客户端和服务器的多个次级通信，因为服务器在公式化查询执行计划之前不知道停止条件。但是，这些次级客户端-服务器通信不是要求有效的查询处理开销的、计算上昂贵的完全往返通信。

由在此描述的技术提供的一些好处的非限制实例为：（1）由于减少了在客户端和服务器的计算上昂贵的完全往返通信，因此减少了网络负载；（2）由于执行不是实际上要求满足终端用户需要的查询，因此减少了数据库服务器上的不必要负载；（3）消除或减少了关于相关等级处理的搜索引擎或中间件应用程序的工作；以及（4）通过将任务从终端用户转移到数据库服务器，使关于基于原始搜索项的查询结构的使用容易，和/或通过将任务从搜索引擎或中间件应用程序转移到数据库服务器，使关于基于原始搜索项的查询结构的应用程序开发容易。

## 操作环境实例

图1是示出可实现本发明的实施例的操作环境的简单实例的框图。

实施例包括用于管理信息搜索的技术，其包括管理数据库查询。因此，操作环境的实例包括客户端**102**、搜索机构**104**、数据库服务器**106**和数据库**108**。

客户端**102**是在计算平台（诸如台式计算机或膝上型计算机）上执行的客户端计算机软件应用程序，与服务器计算机软件应用程序（诸如搜索机构**104**或数据库服务器**106**）通信。客户端**102**有助于终端用户（诸如请求搜索信息的人）与搜索机构**104**通信以请求数据或信息的处理。例如，客户端**102**可以是便于基于网络（诸如互联网）的通信的传统Web浏览器。因此，客户端**102**可以在显示终端上显示多种网页、数据输入帧、查询结果等。客户端**102**显示由搜索机构**104**提供的信息或页面。

搜索机构**104**是典型的计算机软件应用程序和在其上执行应用程序的计算机硬件（诸如，服务器计算机）的结合。搜索机构**104**使用客户端**102**将用户接口和搜索功能提供给终端用户。对于非限制实例，搜索机构**104**可以是通常所称的搜索引擎，用于在互联网上搜索网页中包含的信息，或可以是到更特殊的信息源的接口，诸如关于电子商务或公司网页的搜索功能，或可以是应用程序服务器。

搜索机构**104**起到客户端**102**和数据库服务器**106**之间的接口的作用，并且可以被认为是数据库服务器**106**的客户端。搜索机构**104**和数据库服务器**106**可以一起设置在单一计算平台上，并且可

以设置为提供集成功能给客户端 **102** 的相关软件模块。因此，不要要求搜索机构 **104** 和数据库服务器 **106** 与图 1 所示的实体分开。

数据库服务器 **106** 提供由应用程序服务器(诸如搜索机构 **104**) 代表客户端 **102** 请求的数据。数据库服务器执行未被搜索机构 **104** 或客户端 **102** 的用户执行的所有剩余查询处理。因此，数据库服务器 **106** 是到数据仓库(诸如数据库 **108**) 的高级功能接口，用于管理在多用户环境下的大量数据。在更宽的条件下，数据库服务器 **106** 访问与请求相符的数据库 **108** 中的数据。

数据库 **108** 是被作为一个单元的数据的集合。通常，数据库 **108** 用于存储和恢复相关信息。数据库 **108** 包括一些类型的数据存储单元或数据容器，诸如具有相关行和列的数据表，或具有相关对象的对象类。数据库 **108** 典型地包括用于存储、访问和操作数据的多个其他结构，例如，关于表的索引。实施例不要求用于数据库 **108** 的任何特定逻辑或物理结构。

## 文本查询

根据一个实施例，CONTAINS 运算符用于实现在此描述的逐渐放宽机构。CONTAINS 运算符用于 SELECT 语句的 WHERE 子句中，以指定用于文本查询的查询表达式。CONTAINS 运算符为所选择的每行返回关联值，其可用 SCORE 运算符获得。用于 CONTAINS 运算符的一种语法称为 CONTEXT。另一种可应用的语法为 CTXCAT。包括 CONTAINS 运算符的查询的一个例子如下：

```
CONTAINS (  
    [schema.]column  
    text_query  VARCHAR2  
    [,label    NUMBER])  
RETURN NUMBER;
```

其中，[schema.]列指定将被搜索的文本列，其优选具有与其相关的文本索引。text\_query 参数指定（1）查询表达式，其限定在列中搜索，或（2）标记字符串，其基于 CTXCAT 语法指定查询。在基于 CTXCAT 语法指定查询时，通过查询模板，查询字符串使用以下标记：

<query></query>，其用信号通知该查询被解释为查询模板；

<textquery></textquery>，其指定查询字符串；

grammar=，其指定查询的语法；

<score></score>，其指定值的优先权；以及

datatype=，其指定作为值返回的数的类型。

对于选择的每一行，CONTAINS 返回 0 和 100 之间的数，指示文档行对于查询的相关程度，0 指示在给定行中没有找到匹配。

CONTEXT 语法是用于 CONTAINS 运算符的默认语法，并允许在查询表达式中使用查询运算符。例如，逻辑运算符 AND 允许搜索包括两个不同词的所有文档；ABOUT 运算符允许搜索概念；WITHIN 运算符允许部分搜索，诸如在 XML 或 HTML 文档的部分中搜索；以及 NEAR 运算符允许邻近搜索、词干搜索、模糊搜索，以及相近（thesaural）运算符用于扩展查询表达式，换句话说，用于放宽与查询相关的搜索条件。另外，如果 CONTEXT 语法希望被用于下一搜索，存储文本的表的索引典型的是索引类型“上下文”。

## 功能组件

图 2 是示出数据库服务器 106 和搜索机构 104 的相关功能组件的框图。

搜索机构 104 包括规则解释器 220。规则解释器 220 操作，解释和翻译与信息搜索相关的规则。例如，在以下更加详细描述的一

个实施例中，客户端 **102**（图 1）的用户以规则的形式指定搜索项或值，称为搜索条件，其中，规则由规则解释器 **220** 解释并翻译成另外一种格式。例如，用户可以提供用于指定搜索条件（即，原始形式的原始搜索项，搜索项的连接形式；搜索项的分离形式）的一组搜索规则（诸如，搜索项、AND、OR）和执行相关子查询的顺序（即，首先执行原始形式，接下来执行连接形式，然后执行分离形式）。

当接收到一组规则时，规则解释器 **220** 基于该组规则构造查询模板。如上所述，根据一个实施例，查询模板为<text\_query>模板的形式。根据另外一个实施例，查询模板为在下文中描述的“重写模板”形式。在两个实施例中，查询模板代表搜索条件和与搜索条件相关的子查询执行的顺序。

在另外一个实施例中，规则解释器 **220** 基于接收到的规则产生放宽查询 **250a**，用于提交给数据库服务器 **106**。放宽查询 **250a** 包括搜索条件和与搜索条件相关的构成（constituent）子查询执行的顺序。

规则解释器 **220** 或在搜索机构 **104** 中的其他模块可以被设置成提供其他功能，以使用户与数据库服务器 **106** 友好地进行交互。

数据库服务器 **106** 在功能上包括：查询放宽引擎 **202** 和查询执行引擎 **210**。查询放宽引擎 **202** 包括模板解释器 **204** 和查询构造器 **206**，查询执行引擎 **210** 包括上下文模块 **208**。

模板解释器 **204** 被设置用于解释不同形式的搜索条件（诸如查询模板），搜索条件用于指定或表示子查询序列和执行子查询的顺序。模板解释器 **204** 将信息传送到查询构造器 **206**，查询构造器将

接收到的信息处理成查询。执行查询以从数据库 108 (图 1) 获取信息, 可能与上下文对象 208 结合。

模板解释器 204 被设置用于解释查询模板, 诸如, 可能从规则解释器 220 或另外的源输出什么。例如, 规则解释器 220 从终端用户接收一组规则, 并且基于这些规则产生查询模板, 规则作为请求, 以基于查询模板构造查询。依次, 模板解释器 204 接收并解释查询模板, 并将信息传送到根据数据库服务器 106 支持的查询语言 (诸如, SQL 或 PL/SQL) 构造查询所必需的查询构造器 206。在一个实施例中, 重写模板从规则解释器 220 被接收并包括有效地请求基于该组规则构造查询的语句。关于对 “cheap pen” 的搜索的重写模板的一个实例如下:

```
<query>
  <textquery>
    'cheap pen'
    <PROGRESSION>
      <seq><REWRITE> TRANSFORM(TOKENS, ' ',';','')
      </REWRITE></seq>
      <seq><REWRITE> TRANSFORM(TOKENS, ' ',';','&')
      </REWRITE></seq>
      <seq><REWRITE> TRANSFORM(TOKENS, ' ','$','&')
      </REWRITE></seq>
    </PROGRESSION>
  </textquery>
</query>;
```

其中, TRANSFORM 函数的自变量 (arguments) 包括与搜索项关联的前缀、与搜索项关联的后缀、和与搜索项关联的运算符。从而, 除与搜索项关联的相关运算符 (诸如表示逻辑运算符 “AND” 的 “&” 或者表示逻辑运算符 “OR” 的 “/”), 可以与搜索项关联

地指定前缀或后缀。前缀和后缀字段提供对诸如词干和通配符搜索的变量的支持。因此，与规则解释器 220 结合的该实施例的使用提供了对终端用户的用户友好的能力，以简单地提供一组规则作为用于搜索的输入，从而通过一系列处理，为数据库服务器 106 的执行构造相关查询，以从数据库 108（图 1）获取信息。

基于前述的重写模板，模板解释器 204 将搜索词组“cheap pen”解析成多个词“cheap”和“pen”，并且每个词都根据指定的规则被转换并被传送到用于产生对数据库 108（图 1）的查询的查询构造器 206。例如，关于搜索项“cheap pen”，函数 TRANSFORM（TOKENS, ‘,’,\$’,’&’）的处理用于以连接形式基于“cheap”和“pen”的后缀词干构造子查询。因此，基于 REWRITE 运算符和 TRANSFORM 函数构造子查询，其搜索（1）“cheaper”和“pen”；以及（2）“cheapest”和“pen”。

前面的模板解释器 204 和查询构造器 206 之间的功能性描述不是限制性的，因为这些功能可以由集成的解释器 204/构造器 206 软件模块执行。

查询构造器 206 接收限定搜索的多种形式的信息，诸如，文本字符串、文本查询模板、重写模板、和非文本参数。然后，查询构造器 206 从接收的信息（诸如，放宽查询 250c）构造查询。这样的查询依次被提交给查询执行引擎 210，以对数据库 108（图 1）进行查询。

查询执行引擎 210 与诸如放宽查询 250a、250b、250c 的查询进行交互，并对数据库 108（图 1）执行该查询。查询执行引擎 210 包括上下文模块 208，当搜索文本时，上下文模块实现 CONTEXT 语法的功能。因此，如果 CONTAINS 运算符在执行 SQL 查询的处理中被识别，CONTAINS 自变量文本字符串或模板被传送到上下文



模块 **208**，可能经由模板解释器 **204**，用于根据 CONTEXT 语法的功能进行处理。此外，查询执行引擎 **210** 支持直接从终端用户接收到的放宽查询 **250b** 的执行。

### 用户指定的查询请求的逐渐放宽

根据本发明的一个方面，在数据库服务器应用程序中提供用于数据库查询的逐渐放宽的固有支持。因此，应用程序和终端用户能够指定与数据搜索相关的子查询序列和执行子查询以产生希望结果的顺序。从而，用户可以影响执行搜索的方式和范围，以增强搜索的性能。另外，用户可以影响相关结果或命中目录的相关性等级的质量，而不明显降低查询响应时间或通过量。这样的技术通过减少不必要的查询/搜索项扩大和在客户端和服务器应用程序之间的计算上昂贵的完全往返通信，改善了查询响应时间和通过量。

在一个实施例中，使用户具有相对于文档（诸如，HTML 和 XML 文档）的不同部分指定待执行的子查询的进展序列的能力。例如，用户可以指定首先在一个或多个文档的标题中被搜索、接着在一个或多个文档的主体中被搜索、然后在一个或多个文档中出现的图标对象的描述中被搜索的项。因此，提供了精确搜索能力。

在一个实施例中，包括 CONTAINS 子句的查询可以发布给数据库，其中，搜索字符串不要求遵守传统的文本查询语法。相反，如以下举例说明的，搜索字符串可以符合文本查询模板格式，其中，查询模板被用作用于 CONTAINS 运算符的查询表达式自变量。下面提供了用于搜索存储在用上下文索引类型进行索引的数据表中的文本的示例性查询：

```
SELECT * from tablename  
WHERE CONTAINS (columnname, 'search criteria',1) >0.
```

从而，根据该实施例，查询模板被提供作为 CONTAINS 子句的自变量，其在前述的查询中描述为“搜索条件”。在这样的实施例中，以下示例性的标记文档或查询模板可以用作 CONTAINS 子句的自变量。即，查询模板用于表示搜索条件。查询模板指定子查询序列（例如，到一个或多个表的指针）和对表的指定列执行子查询的进展顺序。

```
<query>
  <textquery>
    <PROGRESSION>
      <seq>cheap pen</seq>
      <seq>cheap & pen</seq>
      <seq>cheap/pen</seq>
    </PROGRESSION>
  </textquery>
</query>
```

前述查询和模板示出了对文本的搜索，但是，存储在一个或多个数据库中的任何数据值和数据类型都可以通过应用在此所述的技术来搜索。而且，数据库可以被分布并与多个分布服务器相关联，例如在搜索互联网的上下文中。

在可选的实施例中，进展由 SQL 或 PL/SQL 扩展名运算符实现，而不是标记模板。例如，PROGRESSION 运算符被定义为对现有 SQL 运算符组的扩展名，从而提供对根据在此描述的技术进展的指定的固有数据库支持。

在执行包括文本的子句（诸如 CONTAINS 子句）或非文本数据的类似子句的查询期间，根据<query>标记的识别，执行引擎将不会将查询的相关部分处理为典型查询。相反，诸如数据库服务器 106（图 1）的数据库服务器提取在<textquery>标记之间的信息，诸

如通过<PROGRASSION>标记来确定进展是否被指定，并以进展指定的顺序执行相关子查询。

在以上实例中，子查询将首先被执行，以搜索包括词组“cheap pen”的文档。接下来，可以执行子查询，以搜索包括词“cheap”和“pen”的文档。最后，可以执行子查询，以搜索包括词“cheap”或“pen”的文档。

在一个实施例中，三个子查询中的每个都可以根据指定进展连续自动运行，并对结果排序，子查询根据排序的结果产生给定结果。例如，从第一子查询返回的命中被产生或显示，具有比从第二和第三子查询返回的命中更高的相关性等级。

相关性等级呈现的方式是没有限制的。例如，结果的相关性等级可以简单地通过呈现给搜索请求者的顺序来识别，或可以分配给数字等级、百分比等。这样的情况消除了在客户端和服务端之间的多重往返通信，必然消除了网络开销。

在可选实施例中，只有当先前执行的子查询不能够完成信息搜索请求时，才执行每个子查询。例如，如果成百上千个包含词组“cheap pen”的文档通过只执行第一子查询被定位，然后中止或完全放弃执行剩余子查询。因此，这种情况消除了客户端和服务端之间的多重完全往返通信，必然消除了查询处理和网络开销，以及不必要的检查。但是，如果用户特别地要求附加结果，那么当必须完成搜索请求的特殊要求时，然后执行第二和/或第三查询。在这样的情况下，随后的客户端服务器通信不是上面所说的完全往返通信，但是被认为是次级通信，因为他们要求很少或不要求查询的诸如检查、优化等的“开销”处理。

## 用于管理信息搜索客户端的方法

图3是示出用于管理信息搜索的方法的流程图。根据本发明的一个方面，该方法通过请求信息搜索的终端用户、典型地经由用户接口来执行。

在块**302**，指定关于正被搜索的信息的一个或多个信息值。在块**304**，指定执行与一个或多个值相关的子查询的顺序。例如，用户可以提交如上所述的查询模板或如上所述的构造查询的一组规则。块**302**和**304**优选同时执行，但是也可以顺序执行该步骤。在一个实施例中，经由用户接口指定值和顺序，并且在另一个实施例中，被指定为XML文档形式。

在块**306**，一个或多个信息值和顺序被提交用于以指定顺序执行一个或多个子查询。该信息可以被直接提交到诸如数据库服务器**106**（图1）的数据库服务器，或直接通过诸如搜索引擎或函数的中间件应用程序。

在一个实施例中，在块**302**指定与文档的一个或多个特定部分相关的信息值。在相关实施例中，在块**302**指定与文档的至少两个特定部分相关的信息值，在块**304**指定的顺序包括对于文档的至少两个特定部分执行与信息值相关的子查询的顺序，如前所述。

## 用于管理信息搜索服务器的方法

图4是示出用于管理信息搜索的方法的流程图。根据本发明的一个方面，该方法通过数据库服务器执行。

在块**402**，接收包括子查询序列和执行子查询的进展的指定的查询。例如，如前所述，包括用于查询文本的CONSTAINS子句的

查询被接收，其包括作为子句的自变量的查询模板。在一个实例中，接收 XML 文档形式的查询。接收的查询可以，例如，包括与文档的特定部分相关的子查询。

在块 404，以由进展指定的顺序执行一个或多个子查询。在一个实施例中，该方法包括从搜索接收对特定数量结果的请求，例如，还有在块 402 接收或反复地从终端用户接收的查询等。例如迭代方法，终端用户可以通过连续通信一次请求十个查询结果。从而，响应于这些请求，数据库服务器或搜索机构连续提供结果，即，一次十个。而且，仅在先前执行的子查询还没有产生足够的结果以实现指定数量的结果的请求时，才执行一个或多个子查询的每个子查询。例如，如果第一执行的子查询产生三十个结果，并且用户反复请求一次十个命中，那么第二子查询不执行，直到用户请求第三十一个至第四十个命中。

与使用连接查询的现有方法对比，不必要执行所有子查询。而且，与顺序地将每个子查询提交给数据库服务器的现有方法相比，本实施例可以只使用在请求搜索的终端用户（诸如客户端 102（图 1））和搜索信息的服务器（诸如数据库服务器 106（图 1））之间的一个完全往返通信，将多个子查询提供给数据库服务器。以任何顺序实际执行子查询的控制由终端用户提供，而不是由软件应用程序和/或搜索引擎或函数假定，并且由数据库服务器应用。

在一个实施例中，所有子查询均基于相应的特定组值（即，搜索条件）以及与搜索条件的放宽对应的进展从数据库选择数据。例如，使用特定组值（诸如（1）“cheap”和“pen”，（2）词干“cheap”和“pen”，以及（3）“cheap”或“pen”）为每个原始搜索项和搜索项“cheap pen”的变形构造子查询。因此，搜索条件“cheap pen”通过根据指定的进展执行相关子查询（其扩大搜索）被逐渐放宽。

## 实施机构硬件概述

图 5 是示出可以实施本发明的实施例的计算机系统 500 的框图。计算机系统 500 包括：总线 502 或用于传输信息的其他通信机构，和与总线 502 连接、用于处理信息的处理器 504。计算机系统 500 还包括主存储器 506，诸如随机存取存储器（RAM）或其他动态存储装置，与总线 502 连接，用于存储信息和将由处理器 504 执行的指令。主存储器 506 还可以用于在执行将由处理器 504 执行的指令期间，存储临时变量或其他中间信息。计算机系统 500 还包括只读存储器（ROM）508 或其他静态存储器，连接到总线 502，用于存储静态信息和处理器 504 的指令。提供存储装置 510，诸如磁盘、光盘、或磁光盘，连接到总线 502，用于存储信息和指令。

计算机系统 500 可以经由总线 502 连接到显示器 512，诸如阴极射线管（CRT）或液晶显示器（LCD），用于将信息显示给计算机用户。包括字母数字键或其他键的输入装置 514 连接到总线 502，用于将信息和命令选择传输到处理器 504。另一种用户输入装置是光标控制器 516，诸如鼠标，跟踪球、或光标方向键，用于将方向信息和命令选择传送到处理器 504，并用于控制显示器 512 上的光标移动。该输入装置通常在两个轴上（第一轴（例如，x）和第二轴（例如，y））具有两个自由度，允许装置指定在平面中的位置。

本发明涉及计算机系统 500 的使用，用于实现在此描述的技术的。根据本发明的一个实施例，由计算机系统 500 响应于执行包括在主存储器 506 中的一个或多个指令的一个或多个序列的处理器 504，来执行这些技术。这样的指令可以从另一计算机可读介质（诸如存储装置 510）读入主存储器 506。通过执行包括在主存储器 506 中的指令序列，使处理器 504 执行在此描述的处理步骤。在可选实施例中，硬接线电路可以替换软件指令或与软件指令结合以实现本

发明。从而，本发明的实施例不限于硬件电路和软件的任何特定结合。

在此使用的术语“计算机可读介质”指的是参与将指令提供给用于执行的处理器 **504** 的任何介质。这样的介质可以有多种形式，包括但不限于非易失性介质、易失性介质、和传输介质。非易失性介质包括，例如，光盘、磁盘、或磁光盘，诸如存储装置 **510**。易失性介质包括动态存储器，诸如主存储器 **506**。传输介质包括同轴电缆、铜线、和光纤，包括构成总线 **502** 的导线。传输介质还可以是声波或光波形式，诸如那些在无线电波或红外线数据通信期间产生的声波或光波。

普通形式的计算机可读介质包括，例如，软盘、柔性盘、硬盘、磁带、或任何其他磁介质、CD-ROM、DVD、任何其他光学或磁光介质、打孔纸、纸带、具有孔图样的任何其他物理介质、RAM、PROM、EPROM、FLASH-EPROM、任何其他储存芯片或盒式磁带、以下描述的载波、或计算机可读取的任何其他介质。

多种形式的计算机可读介质可以参与承载一个或多个指令的一个或多个序列到处理器 **504** 用于执行。例如，指令最初可以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，并使用调制解调器通过电话线发送指令。计算机系统 **500** 本地的调制解调器可以接收在电话线上的数据，并使用红外发射器将数据转换成红外信号。红外监测器可以接收在红外信号中承载的数据，并使用红外发射器将数据转换成红外信号。红外探测器可以接收红外信号承载的数据，合适的电路可以将数据放在总线 **502** 上。总线 **502** 将数据传送到主存储器 **506**，处理器 **504** 从存储器恢复并执行指令。在被处理器 **504** 执行之前或之后，主存储器 **506** 接收的指令可以选择性地存储在存储装置 **510** 中。

计算机系统 **500** 还包括与总线 **502** 连接的通信接口 **518**。通信接口 **518** 提供双向数据通信，连接到与局部网络 **522** 连接的网络链接 **520**。例如，通信接口 **518** 可以是综合业务数字网 (ISDN) 卡或调制解调器，以提供到相应类型的电话线的数据通信连接。作为另一实例，通信接口 **518** 可以是局域网 (LAN) 卡，以提供到兼容 LAN 的数据通信连接。也可以实现无线链接。在任何这样的实现中，通信接口 **518** 均发送和接收承载表示各种信息的数字数据流的电信号、电磁信号、或光信号。

网络链接 **520** 通常通过一个或多个网络提供到其他数据装置的数据通信。例如，网络链接 **520** 可以提供通过局部网络 **522** 到主机 **524** 或到由互联网服务提供者 (ISP) **526** 操作的数据装置的连接。ISP **526** 又通过现在通常称为“互联网” **528** 的全球分组数据通信网络提供数据通信服务。局部网络 **522** 和互联网 **528** 都使用承载数字数据流的电信号、电磁信号、或光信号。通过各种网络上的信号，网络链路 **520** 上的信号，通过通信接口 **518** 上的信号（都传送数字数据给计算机系统 **500** 或者传送来自计算机系统 **500** 的数字数据）是传输信息的载波的示范性形式。

计算机系统 **500** 可以通过网络、网络链路 **520**、和通信接口 **518** 发送信息和接收数据（包括程序代码）。在互联网的实例种，服务器 **530** 可以通过互联网 **528**、ISP **526**、局部网络 **522**、和通信接口 **518** 传输所请求的用于应用程序的代码。

接收到的代码可以在接收到时由处理器 **504** 执行，和/或存储在存储装置 **510**、或其他非易失性存储器中用于随后处理。通过这种方式，计算机系统 **500** 可以获取载波形式的应用程序代码。



## 扩展和选择

通过以上描述和最便于理解实施例的上下文的位置描述了本发明的可选实施例。而且，还参考其特定实施例描述本发明。但是，非常明显，在不脱离本发明的更宽的精神和范围下，可以对本发明作出多种修改和变化。例如，实施例可以在数据库查询的上下文中描述；但是，通常，所描述的技术可以应用到查询信息的电子搜索。再例如，实施例可以在文本搜索的上下文中描述；但是，该技术可以应用到除文本外的多种类型信息的搜索。从而，说明书和附图应该被认为是说明性的而不是限制性的。

另外，在该描述中以特定顺序描述了特定处理步骤，字母和数字标记可以用于识别特定步骤。除非在说明书中特别说明，本发明的实施例将不限于执行这些步骤的任何特定顺序。尤其是，标记仅用于方便识别步骤，而不用于指定或要求执行这些步骤的特定顺序。

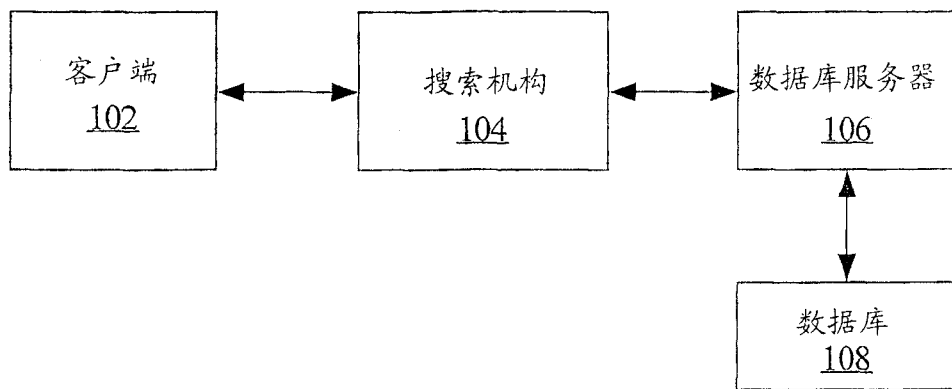


图 1

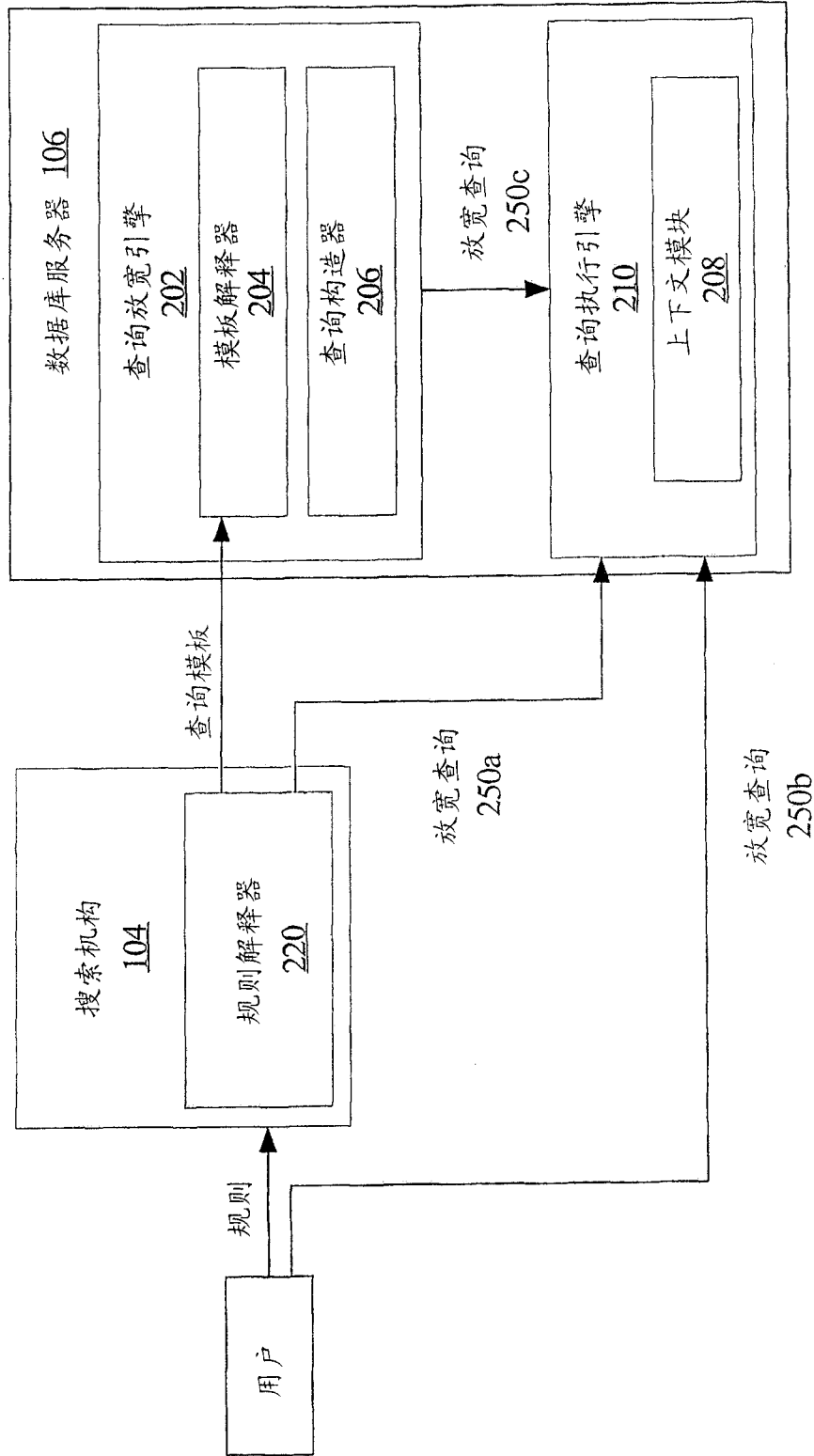


图 2

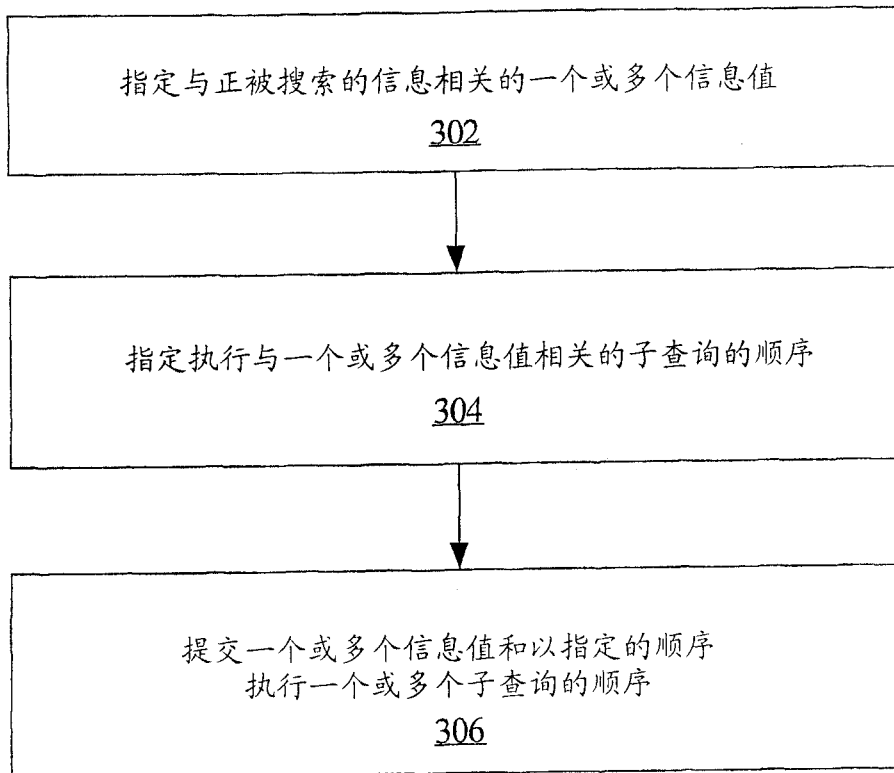


图 3

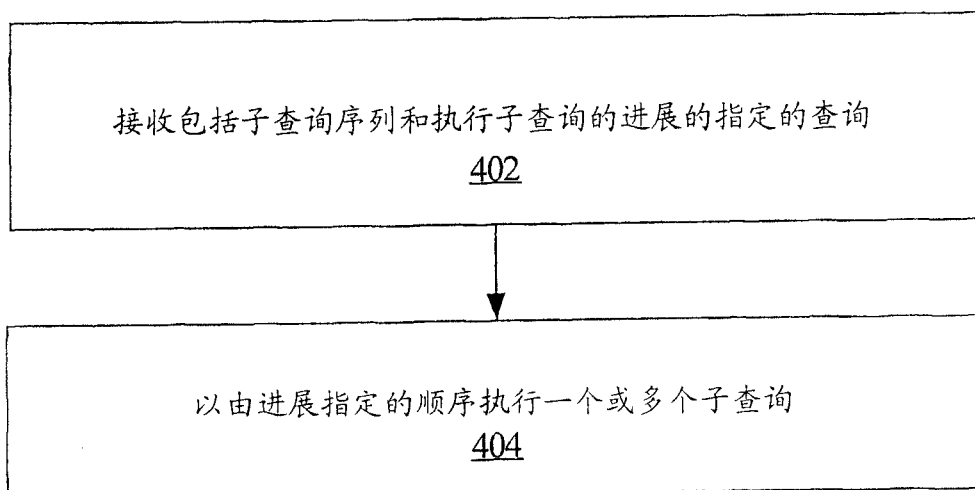


图 4

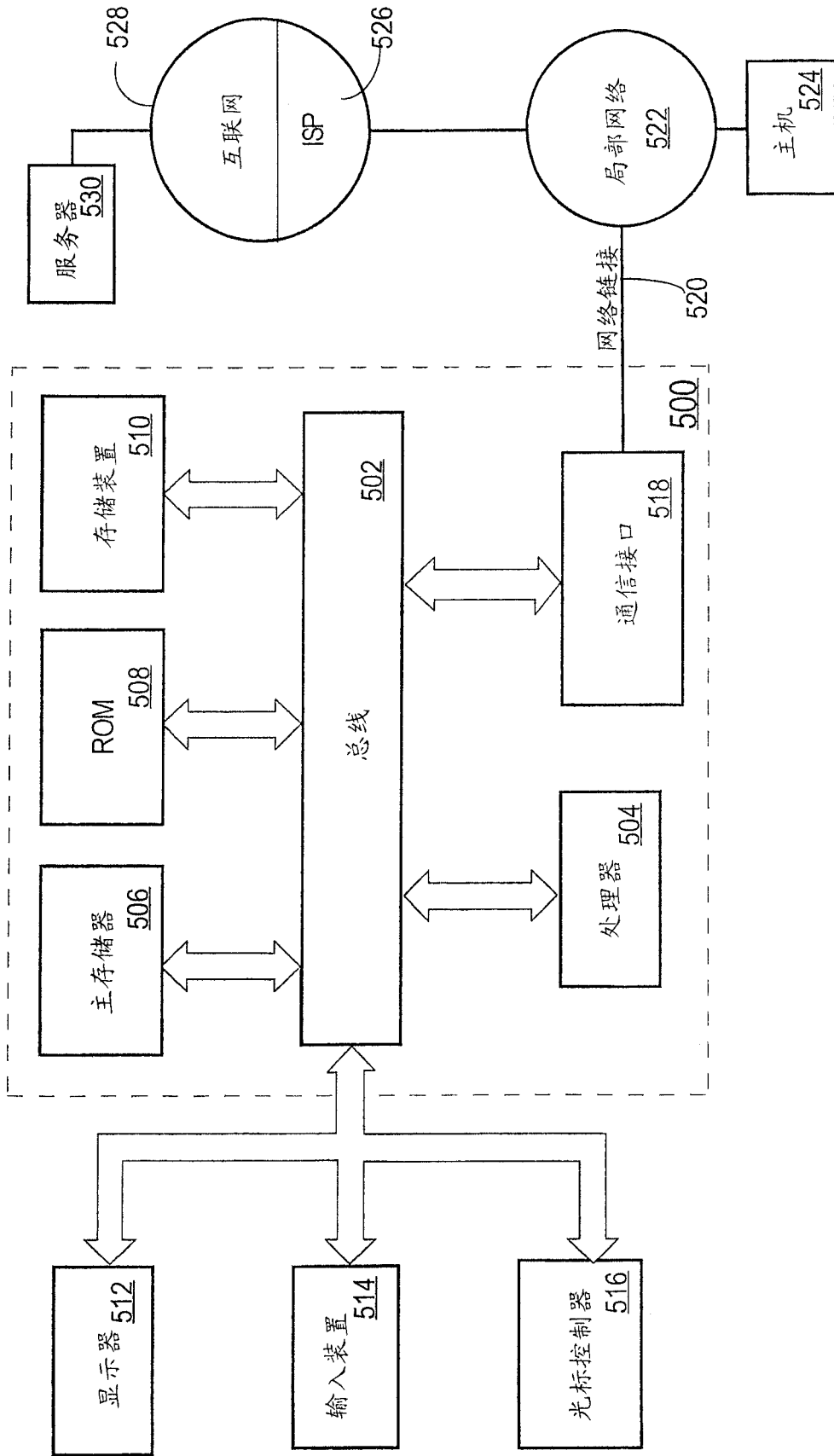


图 5