



(12) 发明专利

(10) 授权公告号 CN 102640145 B

(45) 授权公告日 2015. 09. 23

(21) 申请号 201080048222. 0

(51) Int. Cl.

(22) 申请日 2010. 08. 26

G06F 17/30(2006. 01)

(30) 优先权数据

(56) 对比文件

61/238, 283 2009. 08. 31 US

CN 101088082 A, 2007. 12. 12,

CN 101191730 A, 2008. 06. 04,

(85) PCT国际申请进入国家阶段日

Arnab Nandi 等. Assisted querying using instant-response interfaces. 《SIGMOD '07 Proceedings of the 2007 ACM SIGMOD international conference on Management of data》. 2007, 1156-1158.

2012. 04. 25

(86) PCT国际申请的申请数据

Holger Bast 等. Output-sensitive autocompletion search. 《Journal Information Retrieval archive》. 2008, 第 11 卷 (第 4 期), 269-286.

PCT/IB2010/002102 2010. 08. 26

(87) PCT国际申请的公布数据

W02011/024064 EN 2011. 03. 03

(73) 专利权人 达索系统公司

地址 法国韦利济-维拉库布莱

(72) 发明人 F·布尔东克勒 F·杜埃托

J·博尔迪耶

审查员 孙国辉

(74) 专利代理机构 永新专利商标代理有限公司

72002

代理人 刘瑜 王英

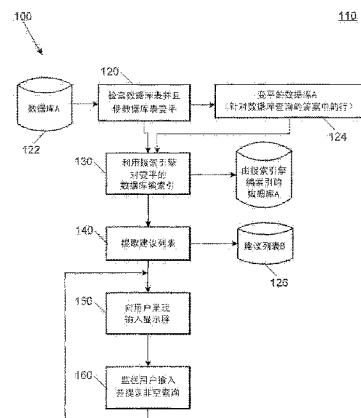
权利要求书5页 说明书13页 附图11页

(54) 发明名称

可信查询系统和方法

(57) 摘要

方法和系统提供了允许用户询问结构化数据库的搜索接口, 并且包括从一个或多个数据库检索数据库条目, 使多个数据库条目变平, 对多个变平的数据数据库条目编索引以形成搜索引擎索引, 以及提示用户录入输入。该系统不断地监视用户输入并且每次用户录入输入时, 该系统响应于用户录入的输入计算非空的部分查询的集合, 将结构化项目与每个非空的部分查询相关联, 以及允许用户选择结构化项目中的一个。如果用户选择了结构化项目中的一个, 则系统使用与所选择的结构化项目相关联的非空的部分查询来替换用户输入。当用户确认输入时, 该系统将输入作为查询来执行。最后, 该系统对应于所执行的查询向用户提供文档。



CN 102640145 B

1. 一种用于基于来自用户的输入在搜索引擎上执行查询的方法,所述方法包括:
从一个或多个数据库检索数据库条目;

使多个所述数据库条目变平;

对所述多个变平的数据库条目编索引以形成搜索引擎索引;

提示所述用户录入输入;

不断地监视所述用户输入;

每次所述用户录入输入时,通过以下各项来处理所述用户输入:

响应于所述用户录入的所述输入,计算非空的部分查询的集合;

将结构化项目与每个非空的部分查询相关联;

允许所述用户选择所述结构化项目中的一个;

如果所述用户选择了所述结构化项目中的一个,则使用与所选择的结构化项目相关联的所述非空的部分查询来替换所述用户输入;

当所述用户确认所述输入时,将所述输入作为查询来执行;以及

对应于所执行的查询向所述用户提供文档;

其中,使所述数据库条目变平生成对应的变平的条目,每个变平的条目将所述数据库条目的每一行复制为文本的单个行,所述文本的单个行具有对应于所述数据库条目分隔行号、列名和列值的定界符;

其中,所确认的查询是在全文搜索时相对于搜索引擎索引运行的。

2. 一种用于基于来自用户的输入在搜索引擎上执行查询的方法,所述方法包括:

从一个或多个数据库检索数据库条目;

使多个所述数据库条目变平;

对所述多个变平的数据库条目编索引以形成搜索引擎索引;

向用户呈现查询框以提示所述用户录入输入;

不断地监视所述用户输入;

每次所述用户录入输入时,通过以下各项来处理所述用户输入:

响应于所述用户录入的所述输入,计算非空的部分查询的集合;

将结构化项目与每个非空的部分查询相关联;

允许所述用户选择所述结构化项目中的一个;

如果所述用户选择了所述结构化项目中的一个,则使用与所选择的结构化项目相关联的所述非空的部分查询来替换所述用户输入;

当所述用户确认所述输入时,将所述输入作为查询来执行;以及

对应于所执行的查询向所述用户提供文档;

其中,使所述数据库条目变平生成对应的变平的条目,每个变平的条目将所述数据库条目的每一行复制为文本的单个行,所述文本的单个行具有对应于所述数据库条目分隔行号、列名和列值的定界符;

其中,所确认的查询是在全文搜索时相对于搜索引擎索引运行的,

其中,响应于所述用户录入的所述输入计算非空的部分查询的集合包括自动完成建议;

其中,在所述自动完成建议中,

如果所述查询框是空,则取出所有的表名和每个表中的行数作为来自所述搜索引擎索引的建议;

如果查询框包含文本,则分析该查询,以分析辨别和识别所述查询框中的表名和列名,并且辨别用户录入的完整的、自由的文本以生成结构化建议;

通过利用 AND 算子将所述查询框中的所有表名和列名连接,并将全文搜索添加在所述自由文本上来构造结构化查询。

3. 如权利要求 1 或 2 所述的方法,其中,所述输入是利用键盘和 / 或定点设备来执行的。

4. 如权利要求 1 或 2 所述的方法,其中,所述输入是利用语音命令来执行的。

5. 如权利要求 3 所述的方法,其中,所述用户录入的所述输入是在每次键击之后被处理的。

6. 如权利要求 1 或 2 所述的方法,其中,所确认的查询被映射至所述一个或多个数据库中的每个数据库的语法,所映射的查询是相对于所述数据库被执行的,并且所执行的数据库查询的结果被合并。

7. 如权利要求 1 或 2 所述的方法,其中,搜索引擎使所述数据库条目变平,所述搜索引擎支持 SPLIT 操作,并且被配置为在包含在所述数据库中的文档的部分内进行搜索。

8. 如权利要求 1 或 2 所述的方法,其中,所述搜索引擎使用转换的文件数据结构。

9. 如权利要求 7 所述的方法,其中,所述搜索引擎执行从由以下各项构成的组中选择的功能:拼写检查功能、词典功能、衍生形态查询功能、词形分析功能、断词功能和标准化功能。

10. 如权利要求 1 所述的方法,其中,响应于所述用户录入的所述输入计算非空的部分查询的集合包括自动完成建议。

11. 如权利要求 1 或 2 所述的方法,其中,每个结构化项目是以对应于所述用户输入高亮显示的方式呈现的。

12. 如权利要求 1 或 2 所述的方法,其中,每个结构化项目是以层次的方式呈现的。

13. 一种用于基于来自用户的输入在搜索引擎上执行查询的设备,包括:

用于从一个或多个数据库检索数据库条目的装置;

用于使多个所述数据库条目变平的装置;

用于对所述多个变平的数据库条目编索引以形成搜索引擎索引的装置;

用于提示所述用户录入输入的装置;

用于不断地监视所述用户输入的装置;

用于每次所述用户录入输入时,通过以下各项来处理所述用户输入的装置:

响应于所述用户录入的所述输入,计算非空的部分查询的集合;

将结构化项目与每个非空的部分查询相关联;

允许所述用户选择所述结构化项目中的一个;

如果所述用户选择了所述结构化项目中的一个,则使用与所选择的结构化项目相关联的所述非空的部分查询来替换所述用户输入;

用于当所述用户确认所述输入时,将所述输入作为查询来执行的装置;以及

用于对应于所执行的查询向所述用户提供文档的装置;

其中,使所述数据库条目变平生成对应的变平的条目,每个变平的条目将所述数据库条目的每一行复制为文本的单个行,所述文本的单个行具有对应于所述数据库条目分隔行号、列名和列值的定界符;

其中,所确认的查询是在全文搜索时相对于搜索引擎索引运行的。

14. 一种用于基于来自用户的输入在搜索引擎上执行查询的设备,包括:

用于从一个或多个数据库检索数据库条目的装置;

用于使多个所述数据库条目变平的装置;

用于对所述多个变平的数据库条目编索引以形成搜索引擎索引的装置;

用于向用户呈现查询框以提示所述用户录入输入的装置;

用于不断地监视所述用户输入的装置;

用于每次所述用户录入输入时,通过以下各项来处理所述用户输入的装置:

响应于所述用户录入的所述输入,计算非空的部分查询的集合;

将结构化项目与每个非空的部分查询相关联;

允许所述用户选择所述结构化项目中的一个;

如果所述用户选择了所述结构化项目中的一个,则使用与所选择的结构化项目相关联的所述非空的部分查询来替换所述用户输入;

用于当所述用户确认所述输入时,将所述输入作为查询来执行的装置;以及

用于对应于所执行的查询向所述用户提供文档的装置;

其中,使所述数据库条目变平生成对应的变平的条目,每个变平的条目将所述数据库条目的每一行复制为文本的单个行,所述文本的单个行具有对应于所述数据库条目分隔行号、列名和列值的定界符;

其中,所确认的查询是在全文搜索时相对于搜索引擎索引运行的;

其中,响应于所述用户录入的所述输入计算非空的部分查询的集合包括自动完成建议;

其中,在所述自动完成建议中,

如果所述查询框是空,则取出所有的表名和每个表中的行数作为来自所述搜索引擎索引的建议;

如果查询框包含文本,则分析该查询,以分析辨别和识别所述查询框中的表名和列名,并且辨别用户录入的完整的、自由的文本以生成结构化建议;

通过利用 AND 算子将所述查询框中的所有表名和列名连接,并将全文搜索添加在所述自由文本上来构造结构化查询。

15. 如权利要求 13 或 14 所述的设备,其中,所述输入是利用键盘和 / 或定点设备来执行的。

16. 如权利要求 13 或 14 所述的设备,其中,所述输入是利用语音命令来执行的。

17. 如权利要求 13 或 14 所述的设备,其中,所述用户录入的所述输入是在每次键击之后被处理的。

18. 如权利要求 13 或 14 所述的设备,其中,所确认的查询被映射至所述一个或多个数据库中的每个数据库的语法,所映射的查询是相对于所述数据库被执行的,并且所执行的数据库查询的结果被合并。

19. 如权利要求 13 或 14 所述的设备,其中,所述搜索引擎使用转换的文件数据结构。

20. 如权利要求 13 或 14 所述的设备,其中,搜索引擎使所述数据库条目变平,所述搜索引擎支持 SPLIT 操作,并且被配置为在包含在所述数据库中的文档的部分内进行搜索。

21. 如权利要求 20 所述的设备,其中,所述搜索引擎执行从由以下各项构成的组中选择的功能:拼写检查功能、词典功能、衍生形态查询功能、词形分析功能、断词功能和标准化功能。

22. 如权利要求 13 所述的设备,其中,响应于所述用户录入的所述输入计算非空的部分查询的集合包括自动完成建议。

23. 如权利要求 13 或 14 所述的设备,其中,每个结构化项目是以对应于所述用户输入高亮显示的方式呈现的。

24. 如权利要求 13 或 14 所述的设备,其中,每个结构化项目是以层次的方式呈现的。

25. 一种用于基于来自用户的输入在搜索引擎上执行查询的系统,包括:

处理器,其被配置为从一个或多个数据库检索数据库条目;

数据库变平组件,其被配置为使所述数据库条目变平;

数据库索引组件,其被配置为对所述变平的数据库条目编索引以形成搜索引擎索引;

显示屏,其被配置为提示所述用户录入输入;

所述处理器不断地监视所述用户输入,并且通过下列各项来处理所述用户录入的所述用户输入:

响应于所述用户录入的所述输入,计算非空的部分查询的集合;

将结构化项目与每个非空的部分查询相关联;

允许所述用户选择所述结构化项目中的一个;以及

如果所述用户选择了所述结构化项目中的一个,则使用与所选择的结构化项目相关联的所述非空的部分查询来替换所述用户输入;

当所述用户确认所述输入时,所述处理器将所述输入作为查询来执行,并且对应于所执行的查询向所述用户提供文档;

其中,所述数据库条目变平组件生成对应的变平的条目,每个变平的条目将所述数据库条目的每一行复制为文本的单个行,所述文本的单个行具有对应于所述数据库条目分隔行号、列名和列值的定界符;

其中,所确认的查询是在全文搜索时相对于搜索引擎索引运行的。

26. 一种用于基于来自用户的输入在搜索引擎上执行查询的系统,包括:

处理器,其被配置为从一个或多个数据库检索数据库条目;

数据库变平组件,其被配置为使所述数据库条目变平;

数据库索引组件,其被配置为对所述变平的数据库条目编索引以形成搜索引擎索引;

显示屏,其被配置为向用户呈现查询框以提示所述用户录入输入;

所述处理器不断地监视所述用户输入,并且通过下列各项来处理所述用户录入的所述用户输入:

响应于所述用户录入的所述输入,计算非空的部分查询的集合;

将结构化项目与每个非空的部分查询相关联;

允许所述用户选择所述结构化项目中的一个;以及

如果所述用户选择了所述结构化项目中的一个,则使用与所选择的结构化项目相关联的所述非空的部分查询来替换所述用户输入;

当所述用户确认所述输入时,所述处理器将所述输入作为查询来执行,并且对应于所执行的查询向所述用户提供文档;

其中,所述数据库条目变平组件生成对应的变平的条目,每个变平的条目将所述数据库条目的每一行复制为文本的单个行,所述文本的单个行具有对应于所述数据库条目分隔行号、列名和列值的定界符;

其中,所确认的查询是在全文搜索时相对于搜索引擎索引运行的;

其中,响应于所述用户录入的所述输入计算非空的部分查询的集合包括自动完成建议;

其中,在所述自动完成建议中,

如果所述查询框是空,则取出所有的表名和每个表中的行数作为来自所述搜索引擎索引的建议;

如果查询框包含文本,则分析该查询,以分析辨别和识别所述查询框中的表名和列名,并且辨别用户录入的完整的、自由的文本以生成结构化建议;

通过利用 AND 算子将所述查询框中的所有表名和列名连接,并将全文搜索添加在所述自由文本上来构造结构化查询。

可信查询系统和方法

[0001] 相关申请的交叉参考

[0002] 本申请要求于 2009 年 8 月 31 日提交的标题为可信查询系统和方法 (Trusted Query System and Method) 的临时专利申请 No. 61/238, 283 的优先权权益。在此通过引用的方式将申请 No. 61/238, 283 全部并入本文。

[0003] 版权介绍

[0004] 本专利文档的部分公开内容包含受版权保护的材料。该版权所有人反对如同该专利文档或专利公开中的任何一个在专利和商标局专利文件或记录中所出现的那样对其进行拓制,然而无论如何保留所有的版权。下面的介绍适用于下面所描述的以及在此处的附图中的任何软件和数据 :Copyright © 2010, Exalead, All Right Reserved。

技术领域

[0005] 本公开大体上涉及数据存储和检索。更具体地,本公开涉及允许用户搜索可利用结构化查询语言访问的数据的系统和方法。

背景技术

[0006] 数据库由针对一个或多个使用有组织地收集的数据构成。将数据库分类的一种方法涉及内容的类型,例如,参考书目、全文、数字、图像等。其它的分类方法检查数据库模型或数据库架构,如下面所解释的。

[0007] 由软件根据数据库模型来构造数据库中的数据。最常使用的是关系数据库模型。诸如层次模型和网络模型之类的其它模型使用对关系的更清楚的结构化表示。关系数据库通过使用在数据集中找到的公共特征来匹配数据。以许多人更容易理解的方式来构造得到的数据组。例如,可以通过交易发生的年份、通过交易的销售价格、或者通过购买者的姓氏等等来将包含城镇中所有不动产交易的数据集合分组。这种分组使用关系模型(也称为“模式”)。因此,这种数据库称为“关系数据库”。

[0008] 用来执行这种构造和分组的软件称为关系数据库管理系统 (RDBMS)。术语“关系数据库”常常指这种类型的软件。关系数据库是用于存储财务记录、制造和逻辑信息、人事数据等的当前主流选择。具体地,关系数据库是关系的集合,通常称作表。表由被标记和分类的列中的数据值或关键字的行构成。一些数据库管理系统要求用户在提出查询之前识别其自身,并且取决于为该用户定义的访问权限,表中的一些行或列、或者全部表对特定识别的用户来说可以是或者可以不是可见的。

[0009] 对数据库的查询是用于从数据库请求数据的指定该查询的答案必须满足的条件的公式的示例。结构化查询是根据结构化语法表达的查询。一种这样的语法是在结构化查询语言 (SQL) 中指定的,该结构化查询语言是用于访问关系数据库管理系统 (RDBMS) 中的数据的广泛使用的语言。数据库检索系统是在一个或多个数据库上处理用户查询的软件程序或程序集合。

[0010] 查询处理意味着采用示例用户查询作为输入、分析该查询以理解该查询中指定的

条件、从数据库访问数据以及从数据库返回满足查询中指定的条件的答案。公知的结构化查询是遵守数据库检索系统中实现的语法的查询。

[0011] 示例查询是具有至少一个条件的查询。本地数据库功能是数据库管理系统可以对数据库执行的那些操作,包括创建数据库、修改数据库以及处理对数据库的查询。对数据库查询的搜索结果通常是满足查询中指定的条件的所有数据库元素的列表。

[0012] 搜索引擎是从各种信息源获取通常称为文档的信息段、为该信息编索引并且提供响应于查询使用那些索引来访问某些原始信息表示的方法的软件程序或程序集合。该原始信息或文档可以是文本文档,诸如网页、电子邮件、PDF 文件、图形文件、视频文件、音频文件、来自数据库的行或者经分析的任何其它信息段。

[0013] 可以通过在数据库上运行数据库查询并且将结果列表中的每个项视为文档来为数据库的内容编索引。

[0014] 除了对数据库的内容编索引之外,还可以预先使数据库的内容变平(flatten)。“变平”方法在于将多个表和数据库的多个行组合以创建文档。

[0015] 比起可能具有多个搜索框(可能针对表中的每个字段具有一个搜索框)的数据库接口,诸如 GOOGLE™ 或者 YAHOO!™ 之类的搜索引擎通常具有用于录入用户查询的唯一查询框。搜索引擎通常具有非常简单的输入语法,例如,接收作为输入的单个字并且返回包含该字的所有被编索引的网页。这是自由文本搜索的实例,这是因为该字可以出现在结果文档中的任何位置。如果可以搜索由搜索引擎获取的整个文档,则该搜索引擎实现全文搜索。

[0016] 还可以将文档分为多个部分,这些部分由搜索引擎识别。这些部分包括段落部分、标题部分或者主体部分。一些搜索引擎允许用户将查询限制在特定的部分或字段。搜索引擎查询的搜索结果是与查询匹配的文档的列表。该列表通常根据排序策略来排次序,诸如通过文档长度(通过首先呈现更短的文档)来排序,或者通过将每个文档中的字与查询中的字的密度进行比较来排序。

[0017] 虽然自由文档搜索表达了搜索引擎的有效查询,但是它们通常还提供对更复杂查询的支持。例如,诸如 GOOGLE 之类的流行的搜索引擎通常支持布尔算子(例如,Disney AND world) 或者配置为针对文档的特定部分的算子(例如,“intitle:Disney”,其仅在文档标题中搜索术语“Disney”)。将文档分成部分的许多其它变型也是可能的。

[0018] 搜索引擎通常使用它们从获取的文档中提取出的所有术语的转换的索引。转换的索引指示从哪里、在哪个文档中找到术语。当用户正在将查询键入查询框时,从该索引中提取出的术语的列表允许搜索引擎提议自动完成。自动完成是指示哪个经编索引的术语是用户当前正在键入的字符串的可能完成的机制。自动完成由监视用户正在键入的内容的过程来执行,并且可以在每次键击之后提议可能的完成。自动完成中的每个提议的完成可以是可用作搜索引擎中的查询的经编索引的术语,并且针对该每个提议的完成,搜索引擎知道存在与该术语对应的文档。

[0019] 搜索引擎通常对用户查询执行拼写检查。在拼写检查中,对于具有很少或没有结果的查询,搜索引擎可以从其转换的索引中提议可能是用户打算键入的其它术语。

[0020] 除了拼写检查之外,搜索引擎还可以提供其它的搜索机制,例如,可以对诸如 dogs 之类的搜索术语进行衍生形态查询(stemmed) 或者词形分析(lemmatized),以便其还与术语 dog 匹配。搜索引擎查询句法的另一实例可以是使用星(*) 算子来匹配任意数量的字符,

以使得搜索查询 dog* 还将与转换的索引术语 dog、dogs、dogged、dog-eared 等匹配。该方式的星算子的使用被称为前缀匹配。

[0021] 还可以定义环绕搜索查询的引用的使用,以取消默认的衍生形态查询,从而加强前缀匹配。例如,“dogs”将仅与转换的索引中的术语 dogs 匹配,而不与术语 dog 匹配。搜索引擎还可以使用诸如同义词典之类的词汇语义结构来执行搜索,以使得在假设同义词指示 dog 和 pet 之间的关系并且该关系在查询处理期间被激活的情况下,对字 dog 的搜索还可能检索到包含字 pet 的文档。

[0022] 构建对使用句法可能性的搜索引擎检索和数据库检索两者的查询对于普通信息搜寻者来说可能是困难的。高级的查询句法难以掌握并且只有很小比例的信息搜寻者能够在没有经过有效训练的情况下掌握该句法。

[0023] 存在指导用户构造数据库查询的商业上可获得的产品。例如,来自 California, Redwood Shores 的 ORACLE™ 的 DISCOVER™ 包括允许用户通过从屏幕上显示的项目下拉列表选择项目来构造查询的查询生成器。这些项目可以代表要从数据库获得的数据,或者可替换地可以代表要对该数据执行的操作。一旦已经选择了项目,那么查询生成器通常以 SQL 生成随后可以被发送到数据库检索系统以用于执行的查询。

[0024] 在 M. Hassan, R. Alhajj 和 M. J. Rodley 的标题为“Combining Free-word Search and Relational Databases”的文章中,作者写道:“结构化查询语言足以允许利用公知的结构和特征来查询关系数据库的内容和结构。然而,给定动态数据库,即,具有变化或未知结构的数据库,使得查询表达过程是非常困难的任务”。上述作者提出了一种通过使用 Java 数据库连接将简单的搜索引擎类查询变换为一系列数据库请求来探查数据库的内容和结构的系统。Java 数据库连接 (JDBC) 是能够使 Java 程序操作存储在数据库中的数据的技术。在“所有等级”模式中,一旦提交了由通过特定布尔连接符连接的一个或两个字构成的查询,则将 JDBC 数据库请求发送到所有的可见数据库,并且显示与查询匹配的任何数据库名、表名、列名或值。

[0025] 商业智能指的是用于对公司的有形和无形数据进行聚集、合并、建模和传送以支持更好的商业决策并为行政管理提供公司活动概况的基于计算机的技术。许多已知的商业智能系统的一个主要缺点源自于它们通常需要能够处理复杂的结构化查询的一个或多个专家的干预。例如,当执行经理具有特定的请求时,他或她将不得不将其表达给数据库专家,该数据库专家的角色是设计复杂的结构化查询以提供满足该请求的合适的答案。该复杂的结构化查询的对话和设计是耗时的。因此,这种要求表现出了商业智能系统发展中的严重约束。因此,存在一种对使用综合的和直观的工具(例如,搜索引擎,其以特定等级的信任执行自由文本搜索)来制定请求的方法和系统的需求。

发明内容

[0026] 本文档描述的实施例在多个方面与上述方法不同。在本发明的一个实施例中,监视每个键击处的用户输入,并且提议该用户输入的所有可能的完成。其它的实施例描述了结构化查询,该结构化查询可以使用搜索引擎在给定数据库上运行,而非运行所有可能的查询。一些实施例还使用数据库中的信息的副本,而非数据库检索系统自身,这基于搜索引擎技术的速度和伸缩性提供了优势。搜索引擎通常非常快并且在亚秒范围内产生结果,而

非在数据检索系统可能要求对查询进行响应的多个秒或者分钟范围内产生结果。一些实施例还通过仅产生具有非空答案的可信查询来帮助用户。其它实施例还向用户提供了提议的可信查询的结果数。

[0027] 在一个实施例中,可信查询系统通过重复地产生可信查询,即在数据库上有效并且已知在数据库中具有匹配的、示例的记录的结构化查询,来为不是数据库构造和操作领域普通技术人员的人提供访问数据库内容的手段。结构化查询将用户查询限制在正被查询的数据库的特定值或特定字段。如本文所提及的,字查询意味着搜索引擎执行的结构化查询。

[0028] 在一个实施例中,一种用于在数据库上执行可信查询的方法包括从一个或多个数据库检索数据库条目,使多个数据库条目变平,对多个变平的数据库条目编索引以形成搜索引擎索引,以及提示用户录入输入。该系统不断地监视用户输入,并且每次用户录入输入时,系统响应于用户录入的输入计算非空的部分查询的集合,将结构化项目与每个非空的部分查询相关联,并且允许用户选择结构化项目中的一个。如果用户选择了结构化项目中的一个,则系统使用与所选择的结构化项目相关联的非空的部分查询来替换用户输入。当用户确认输入时,系统将输入作为查询来执行。最后,系统对应于所执行的查询向用户提供文档。

[0029] 在另一实施例中,一种计算机程序产品包括计算机可使用介质和在计算机可使用介质上实现的计算机可读代码,该计算机可读代码用于基于来自用户的输入在搜索引擎上执行可信查询。计算机可读代码用于由处理器执行,当计算机可读代码被执行时使处理器:从一个或多个数据库检索数据库条目,使多个数据库条目变平,对多个变平的数据库条目编索引以形成搜索引擎索引,以及提示用户录入输入。该计算机程序产品不断地监视用户输入,并且每次用户录入输入时,该处理器响应于用户录入的输入计算非空的部分查询的集合,将结构化项目与每个非空的部分查询相关联,以及允许用户选择结构化项目中的一个。如果用户选择了结构化项目中的一个,则使用与所选择的结构化项目相关联的非空的部分查询来替换用户输入。当用户确认输入时,该处理器将输入作为查询来执行,以及对应于所执行的查询向用户提供文档。

[0030] 在另一实施例中,一种用于基于来自用户的输入在搜索引擎上执行可信查询的系统,包括:处理器,其被配置为从一个或多个数据库检索数据库条目;数据库变平组件,其被配置为使数据库条目变平;数据库索引组件,其被配置为对变平的数据库条目编索引以形成搜索引擎索引;以及显示屏,其被配置为提示用户录入输入。该处理器不断地监视用户输入,并且通过下列各项来处理用户录入的用户输入:响应于用户录入的输入计算非空的部分查询的集合,将结构化项目与每个非空的部分查询相关联,允许用户选择结构化项目中的一个,以及如果用户选择了结构化项目中的一个,则使用与所选择的结构化项目相关联的非空的部分查询来替换用户输入。当用户确认输入时,处理器将输入作为查询来执行并且对应于所执行的查询向用户提供文档。

[0031] 根据本文描述的一些实施例提供和生成可信查询是重要的,这是因为使用自由文本来表达查询本质上是不清楚的,并且这种不清楚在结果集合用于进行决定的环境中(例如,在结果集合中绘制计算特定项目/类别的饼图时的商业智能应用的环境中)时是存在问题的。因此,需要一种可信查询系统和方法,其建议解释,即,预期用户在表达自由文本查

询时头脑中的内容的结构化查询过程。这是重要的,以在使用自由文本表达查询时创建特定等级的“信任”。这在商业智能环境中尤为重要。

附图说明

[0032] 参照附图以及下面的描述可以更好地理解该系统和方法。附图中的组件不一定是按比例,而是将重点置于示出本发明的原理。而且,在附图中,类似的附图标记在不同的视图中指示对应的部件。

[0033] 图 1 是根据本发明的一个实施例的流程图,其示出了可以由可信查询方法和系统采用的步骤;

[0034] 图 2 是示出了根据特定实施例的可以用来使数据库变平的步骤的流程图;

[0035] 图 3 是示出了根据特定实施例的可以用来对变平的数据库编索引的步骤的流程图;

[0036] 图 4 是根据特定实施例的流程图,其示出了可以用来进行用户输入和监视的步骤;

[0037] 图 5 是根据特定实施例的流程图,其示出了在每个用户键击之后如何生成可信查询;

[0038] 图 6 是根据本发明的多个实施例的可以用来执行该方法的计算机系统的框图;

[0039] 图 7 是示出了根据本发明的一个实施例的典型的数据库表的图示;

[0040] 图 8 示出了如何在搜索引擎中表示图 7 的数据库;

[0041] 图 9 和图 10 是示出了根据特定实施例的数据库及其变平的形式示例性实例;以及

[0042] 图 11-21 示出了根据本发明特定实施例的可信查询接口的各种屏幕显示。

具体实施方式

[0043] 图 1 是示出了可信查询系统的操作的流程图,包括可以根据可信查询系统 110 的一个实施例采用的步骤 (100)。作为用户处理的概括,并且根据特定的实施例,向用户呈现诸如 GUI 或图形用户接口之类的接口,以使用搜索引擎查询数据库。然后,用户开始将他或她的查询键入搜索引擎的查询框。在优选的实施例中,在键入每个键击之后,系统显示将从数据库返回非空记录集合的示例和部分示例查询的版本。

[0044] 在任意时刻,用户可以 (i) 点击由系统 110 提议的查询建议, (ii) 在显示的查询框中录入更多的文本,或者 (iii) 向基础搜索引擎或数据库检索系统提交当前查询。可替换地,用户可以使用语音文本接口而非键入作为对系统 110 的数据输入。在下面的描述中,假设用户正在使用本文输入接口。

[0045] 可信查询方法和系统 110 的一个优点在于用户从来不表达将产生 0 个结果的查询。在本文描述的过程中,用户可以录入将导致 0 个结果的文本,但是可信查询方法和系统 110 在用户确认该查询之前首先警告用户该查询将返回 0 个结果。可替换地,例如,系统 110 还可以在该时刻经由拼写检查来进行建议。

[0046] 虽然在用户开始键入之前生成会产生非零结果的所有这种查询在理论上是可能的,但是该组合的大量可能性需要的存储空间为搜索引擎正在查询的原始数据库大小的多

个数量级,因此是不可行的或者在经济上不可实施的。使用本文描述的实施例,不需要明确地生成将在数据库上运行的所有可能的查询,这是因为如下面所解释的,由快速搜索引擎查找实时地或者在空闲时生成可能的非空查询。

[0047] 由于根据一个实施例的可信查询方法和系统 110 使用基础搜索引擎技术而非本地数据库功能,因此系统 110 提供可能不表现在基础数据库系统中的全文搜索和其它搜索引擎能力。例如,可信查询方法和系统 110 提供:

[0048] ●语言检测

[0049] ●断词 (tokenization) 和标准化,具有句子边界识别 (将文本解析为单个的字和句子,应用关于诸如白色空间和标点符号之类的分隔符的语言特定的规则)

[0050] ●衍生形态查询 (共享相同词干的字的标识,例如,“engine”和“engines”)

[0051] ●词形分析 (lemmatization)、形态学和句法处理 (对不仅是基本词干还有更复杂的变型 (例如,“good”和“better”) 的识别,并且应用字和句子结构类型的语言特定的知识)

[0052] 在可信查询方法和系统的特定实施例中,诸如搜索引擎之类的商业上可获得的软件可以用来对被变平的基础数据库编索引。这种商业上可获得的搜索引擎优选地包括“SPLIT”算子能力。首先由搜索引擎对基础数据库 (或者可能多于一个数据库) 完全编索引,以创建搜索引擎索引,然后搜索引擎对搜索引擎索引而非原始数据库进行操作。之后,基于用户查询,搜索引擎返回从搜索引擎索引获取的文档,并且这种结果表示原始数据库的行,如下面解释的。当然,搜索引擎索引可以包括多种结构或子结构,如存储器效率和处理速度所需要的。可替换地,可以用多个“批次”来执行编索引,以最小化存储器使用并促进高效的处理。因此,不需要整个搜索引擎索引同时驻留在存储器中。

[0053] 现在参照图 1,以高级形式概括了根据一个实施例用于执行可信查询方法的步骤 110。下面将更详细地描述在本文章中描述的所选择的高级步骤。在第一步骤 120 中,从数据库 (“A”) 122 中获取数据库表,将表变平并且将表存储在变平的数据库存储设备 124 中。图 3 中更详细地示出了该过程。接下来,使用搜索引擎或者其它合适的工具对变平的数据库 124 编索引 130。之后提取建议列表 140 并将其存储在建议列表 (“B”) 126 中。之后向用户呈现输入显示屏 150 以允许用户录入各种搜索术语或者查询。然后,处理继续监视用户输入并向用户提议会导致对查询的非空响应的查询 160。重复地循环该步骤以向用户提供可信查询处理。

[0054] 参照图 1 的变平步骤 120,访问图 7 中示出的数据库表 700,其是典型数据库表的简化表示。来自数据库的数据库表 700 是具有名 D(720) 的类型 T(710) 的,并且由被标记为 C1、C2、... CN 的多个数据库表列 730 以及示例性具有值 T11、T12、... T1N、T21、T22、... T2N 直至 TM1、... TMN 的数据库表行 740 组成。在本描述中要注意,将可交换地使用术语“列”和“字段”。

[0055] 在图 1 的步骤 120 中,可以使用已知的数据库变平技术来将数据库 A 变平,并且可以将其变平为其中每行包含标识数据库表类型、数据库表名、数据库列名 (字段名) 以及数据库表行中的示例性值的信息的特定格式。可以使用任意合适的搜索引擎来将数据库变平,诸如可以从法国的 Exalead S. A 获得的 CLOUDVIEW™ 搜索引擎。在步骤 120 期间,可以将多个数据库以及每个数据库中的多个表变平并且对其编索引,但是仅为了示例的目的,图 7

中示出了一个这样的表。

[0056] 图 2 更详细地示出了用来使数据库变平的步骤,其对应于图 1 中的步骤 120。例如,示出了将图 7 中的数据库表转换为图 8 中所示的变平的形式。图 8 中所示的变平的形式在由冒号 810、逗号 820 以及分号分隔的单个文本行上复制了每个数据库表行,并且示出了指示行号、列名和列值的位置。例如,图 7 的第二行第二列的值 T22 在图 8 中表现为行:

[0057] -ROW-2 ;T/D/C1, T21 ;T/D/C2, T22 ;... ;T/D/CN, T2N

[0058] 在图 8 中所示的该实例中,第一字段 ROW-2840(由分号分隔的)指示该行对应于表的第二行。该分号后面是 N 个字段。这些字段中的第二字段包含由逗号分隔的两个项,即 T/D/C2 和 T22。由于该字段是冒号之后的第二字段,因此其对应于表中的第二列。该字段中的第一项给出了数据库类型 T(可选值)、数据库表名 D 以及字段或列名 C2 的层次表示。注意,这些分离转换可以由任意合适的或等价的分离方案来替代,可以在下列 URL :http://en.wikipedia.org/wiki/Flat_file_database处找到这种合适的分离方案的描述。

[0059] 如图 2 中所示,使用搜索引擎对数据库执行查询 202,并且将数据库查询结果的下一行读入或输入系统 204。如果处理没有到数据库查询结果结束处 210,则对于每个列 216,进行检查以确定是否已经到了行的结束处 226。如果已经到了行的结束处 226,则在该行的结束处写入“行结束”分隔符符号,并且处理转回到步骤 204 以读取下一行。

[0060] 如果还没有到行结束处 226,则将后面跟随列名分隔符符号的列名写入文件 250。接下来,将后面跟随列分隔符符号的列值写入文件 260。然后进行检查以确定文件中是否存在更多的列 270。如果存在更多的列,则处理转回到步骤 226 以完全地处理行。如果不存在额外的列,则处理转回到退出点 280。在步骤 210 中,如果到了文件结束处,则认为该例程结束 290,处理转到退出点 280。变平过程允许搜索引擎立即在所有的列名和字段值中进行搜索,因此允许对一个有效查询中的所有该数据构建建议列表。

[0061] 重新参照图 1,在步骤 130 中,对数据库 A 的变平的版本以及在步骤 120 中已经变平的所有其它数据库编索引。可以使用允许在文档的部分内搜索的商业上可获得的或标准的搜索引擎来执行使数据库变平。这种标准搜索引擎接收文档作为输入并且以转换的文件数据结构来放置文档中的所有的字,其可以随后被处理以将用户查询与包含该用户查询术语的文档相匹配。在图 8 中给出的变平的数据库实例中,文档的部分是由分号分离的。

[0062] 图 3 示出了与图 1 中的步骤 130 中示出的索引步骤有关的额外的细节。首先,进行检查以确定商业上可获得的搜索引擎是否提供了 SPLIT 算子能力 310。SPLIT 算子允许用户在文档的部分内进行搜索。如果没有 SPLIT 算子可用 310,则处理退出 314。如果 SPLIT 算子可用,则对变平的数据库表编索引 320,然后处理退出 314。

[0063] 例如,在该实例中,下面的搜索 (T/D/C2 AND T22) SPLIT “;” 将返回文档 ROW-2。图 9 示出了从客户关系数据库提取的称为“A11 leads”900 的表的实例,而图 10 示出了其变平的版本 1000。图 9 和图 10 中的称为“A11_Leads”的数据库表以及在此没有示出的经类似处理的数据库表将用于示出下面将更详细描述的可信查询系统和方法的机能。

[0064] 接下来,重新参照图 1,如步骤 140 中所示的,将在索引步骤 (130) 期间被编索引的所有术语的列表写入到字列表“B”中。该步骤可以是可选的。字列表 B 可以与索引步骤 130 同时地创建,或者可以在步骤 130 之后创建,或者与步骤 130 的时间无关。字列表 B 可以用于拼写检查目的以及其它的自然语言处理,诸如语音搜索。

[0065] 然后,向用户呈现显示器(150)或者其它接口,其中,用户可以使用诸如键盘、定点设备、触摸屏、语音输入之类的输入设备来将文本输入到查询框中和/或点击各种选项。可以使用适当的商业上可获得的语音识别软件和/或硬件来实现语音识别并处理用户发出的语音命令。优选地,在呈现给用户的初始显示中,每行一个地显示所有被编索引的数据库表,其中具有每个表的被编索引的行数。在图 11 中示出了该呈现 1100 的实例,其中,对三个表“All-Leads”1100、“All-Contacts”1120 和“All-Accounts”1130 编索引,其分别具有被编索引的 97、73 和 21 个行。

[0066] 在图 1 中所示的步骤 160 中,录入到查询框中的任何用户输入激发可信查询建议生成的重复过程。可信查询是满足下列条件的查询:

[0067] (a) 该查询是针对搜索引擎的被较好构造的查询;以及

[0068] (b) 当在搜索引擎上执行时,该查询返回非空的多个答案。可选地,如果需要用户许可来访问数据库内容,那么当可信查询由识别的终端用户执行时,该查询将针对所识别的终端用户来返回非空的多个答案。

[0069] 在步骤 160 中,在用户使用输入设备输入每个字母或者音素之后,对应于用户录入的文本,生成查询并通过搜索引擎索引来执行查询。步骤 150 和 160 中示出的上述过程重复,以执行可信数据库查询过程 100。

[0070] 图 4 示出了图 1 中的步骤 150 和 160 的额外细节。首先,向用户 150 呈现出出查询框的用户接口,以使得用户可以录入查询。为了连贯性,在该图中示出了步骤 150。如上所述,可以使用任何适当的数据录入或请求方法。处理不断地循环 404 以等待用户输入。在用户已经输入了她或他的数据或查询之后,处理确定是否按下了“录入(enter)”键 408。如果已经按下了录入键 408,则向搜索引擎 410 提交用户录入(确认)的查询,处理结果页以用于向用户 416 呈现,并且处理分支回到显示用户接口 150。在另一实施例中,将用户录入的查询映射到一个或多个原始数据库 122 的语法,在数据库上执行所映射的查询,并且向用户 416 呈现数据库搜索结果。

[0071] 如果用户没有按下录入键 408,则处理确定用户是否点击了所提议的部分查询中的一个 420。如果用户没有点击所提议的部分查询中的一个 420,则处理确定用户是否录入了键击 430。如果用户没有录入键击 430,则处理确定用户是否点击了外部建议 440。如果用户点击了外部建议 440,则建议可以对用户隐藏 446,并且处理分支返回到显示用户接口 150。在一个实施例中,建议可以被隐藏,这是因为用户已经点击了所提供的建议外部的区域,并且因此假设用户没有关注于搜索字段或者对搜索字段不感兴趣。

[0072] 如果用户已经点击了所提议的部分查询中的一个 420,则将所选择的查询显示在查询输入框的输入字段中 460。接下来,从可信查询处理获得自动完成建议 470,这将在下面参照图 5 更详细地解释。在从可信查询处理获得自动完成建议 470 之后,处理该自动完成建议以显示给用户 480,并且处理分支返回到显示用户接口 150。

[0073] 在步骤 430 中,如果用户已经录入了键击,则从可信查询处理获得自动完成建议 470。如果处理确定用户没有点击外部建议 440,则处理分支返回到显示用户接口 150。

[0074] 在图 5 中更详细地示出了步骤 470 中示出的获得自动完成建议的过程,并且所显示的该自动完成建议对应于在被编索引的变平的数据库中找到的术语的可能有效完成。自动完成过程涉及预测用户可能想要录入的字或短语,但是不需要用户实际完整地键入该字

或短语。

[0075] 当基于用户已经键入的字或短语易于预测正在键入的字或短语时,诸如当存在有限数量的可能或常用的字(如可以是具有电子邮件程序、web 浏览器或命令行解释的情况)时,或者当编辑以高结构性、易于预测的语言(诸如在源代码编辑器中)写入的文本时,参照步骤 470 示出的根据一个实施例的自动完成特征是尤其有效的。自动完成加速了人与计算机的交互并且提高了用户满意度。

[0076] 可信查询系统 110 的一个实施例中的自动完成允许用户自动完成 SQL 语句中的表名以及 SQL 语句中引用的表的列名。当文本被键入编辑器中时,SQL 语句中的光标的内容提供了关于用户是否需要表完成或表列完成的指示。该表完成提供了在用户连接的数据库服务器中可用的表的列表。列完成提供了仅在 SQL 语句中引用的表的列的列表。

[0077] 可信查询系统 110 的一个实施例中的自动完成处理可以与商业上可得到的软件程序类似,商业上可得到的软件程序例如来自 AquaFold 公司的 Aqua Data Studio, release 7.5, 其除了提供 SQL 编辑器以外还提供了用于数据库中的各种查询的自动完成工具。在许多字处理程序中,自动完成减少了键入重复字和短语所花费的时间量。可以从用户正在操作的当前文档或者从用户定义的常用字的列表来收集用于自动完成的源材料。

[0078] 当前,例如,California, Redwood City 的 Oracle 公司的 OpenOffice.org、Sun Microsystems 公司的 StarOffice、Microsoft 公司的 Microsoft Office 和 KDE 公司的 KOffice 包括上述对于自动完成的支持,如诸如 Emacs 和 Vim 之类的前进文本编辑器(advanced text editor)。在命令行解释中,诸如 Unix 的 sh 或 bash,或者 Microsoft Window 的 cmd.exe 或 PowerShell,或者在类似的命令行接口中,可以通过跟踪用户可能访问的事物的所有可能的名称来完成命令名和文件名的自动完成。

[0079] 在一些程序中,在键入了字的前几个字母之后,可以通过按下 Tab ⇌ 键来执行自动完成。例如,如果当前目录中以 x 开始的文件仅仅是 xLongFileName,则用户可以优选地键入 x,并且针对完整的名称使用自动完成。如果在相同范围内存在以 x 开始的另一文件名或命令,则用户将键入额外的字母或重复地按下 Tab 键以选择合适的文本。在可信数据库查询的一些实施例中,有效的完成可以是原始数据库中找到的任何列名,或者原始数据库中的行值的任何部分匹配。在本发明的一些实施例中,列名和行值均已经被编索引。

[0080] 图 5 更详细地示出了图 4 中的步骤 470 的自动完成建议过程。首先,分析查询框的内容 506。如果查询框是空的 510,则取出所有的表名和每个表中的行数作为来自搜索引擎索引的建议,并且将其格式化 530。

[0081] 如果查询框包含文本 510,则分析该查询 534。分析辨别和识别查询框中的表名和列名(结构化块),并且辨别用户录入的完整的、自由的文本以生成结构化建议。

[0082] 在分析 534 之后,处理确定完整的、自由的文本是否已经被识别为分析的结果 538。如果已经分析了完整的、自由的文本 538,则处理分支到确定表名是否出现在查询框中 544。在该点,处理确定新的文本是否跟随表名。如果表名出现在了查询框中 544,则获取特定表的列名并将其与所有的表和列名匹配 546,然后构建结构化查询 564。

[0083] 关于已经录入到查询框中的所有字和 / 或字段名生成对变平的数据库的查询。步骤 544 确定是否在先前的结构化块中指定了表名。如果是,则执行全文与表列名的匹配(546)。如果没有在先前的结构化块中指定表名(544),则执行针对所有的表名和所有的表

列名的匹配 548。

[0084] 如果不存在指定的表名 544,则将完整的、自由的文本与所有的表名及其列名匹配 548,以生成建议。无论是找到 546 还是没有找到 548 表名,均构建结构化查询 564。可以通过利用 AND 算子将查询框中的所有结构化块(表名和列名)连接,并将全文搜索添加在检测到的自由文本上来构造该结构化查询。

[0085] 当该查询部分地或者完全地匹配变平的数据库表的列名的前缀时(见步骤 548),在用户接口的窗口中显示匹配的列的表名,并且还显示部分或完全匹配的列,其中以粗体或高亮显示匹配的部分 576。当该查询与数据库表列的值部分或完全匹配时,在用户接口的窗口中显示找到该值的数据库表名。还显示找到该值的列名以及该值,其中以粗体或高亮显示其匹配的部分 576。还将对应于匹配的列名或具有其列名的匹配的列值的示例行的数量显示为例如括号中的数目。

[0086] 然后,将如上所述构造的结构化查询发送到搜索引擎 570。如果该结构化查询在搜索引擎中具有匹配的结果(命中)572,则从这些结果中提取建议 574。所提取的结果中的每一个可以包含表名、列名以及值。所提取的建议对应于针对给定表和给定列的值,其中这些值通过完全匹配或者与出现在特定表的特定列的值中的字的前缀匹配,而对应于自由文本。

[0087] 接下来,聚集在步骤 546、548 和 574 中创建的所有建议,并且匹配的部分是高亮的 576。然后,格式化高亮的建议 530。在附图中,为了示例的目的,还以增加的字形大小来显示高亮的部分。然而,可以使用任何形式的文本强调来容易地指出向用户提议的查询。如果该结构化查询在搜索引擎中不具有匹配结果(命中),则没有要提取的建议,并且执行高亮 576。由于高亮对于每个查询命中是重复的过程,因此如果没有命中,则没有高亮要执行。

[0088] 格式化过程 530 对其接收的建议进行重复并且根据每个建议的建议类型,相应地格式化显示。如果建议类型是表名 578,则处理所有的表名以用于显示 580。如果建议类型是列名 578,则处理所有的列名以用于显示 582。类似地,如果建议类型是值 578,则处理所有的列名和值以用于显示 584。在处理完建议类型以用于显示 580、582、584 之后,返回处理建议 586,并且使处理建议可用于显示。然后,如图 4 的步骤 470 所示,处理返回到其子例程调用点。至少在图 11、12、16 和 17 中示出了这些格式化显示的实例。

[0089] 返回到步骤 538,如果还没有分析完整的、自由的文本,则处理确定是否指定了表名 590。如果还没有指定表名 590,则处理分支以格式化建议 530。如果已经指定了表名 590,则返回指定的表的列名 592,并且处理分支以格式化建议 530。注意,如果完整的、自由的文本是不可得的,则将对应于特定表的列名获得描述表名的“结构化块”。

[0090] 在优选的实施例中,将变平的数据库表的每一行编索引为单独的文档,在该单独的文档中,出现了以下列方式(例如,图 10 中所示的)格式化的表名、每个列的名、以及每个列的值;使得可以辨别数据库表名、列名和列值,并且可以将列值与其列名相关联。可以使用保留数据库表名和列名以及列值之间的区别的对文档编索引的任何其它合适的方法。例如,可以在一个搜索引擎索引中对所有的数据库表名和表列名独立地编索引,并且在另一搜索引擎索引中对所有的数据库值独立地编索引。

[0091] 在一个实施例中,诸如本体论(ontology)之类的简单的或结构化的词典或词汇语义结构可以用于在编索引时或者在查询时将变平的数据库中的实际值映射到可替换值

的集合。本体论是域中的一组概念和这些概念之间的关系的正式表示。本体论可以用来推断域的特性,并且可以用于定义域。理论上,本体论是“共享概念化的正式的、清楚的说明”。

[0092] 本体论提供共享的词汇表,该词汇表可以用于对域进行建模,即,存在的对象和/或概念的类型及其特性和关系。本体论作为关于世界或其某些部分的知识表示的形式用于人工智能、语义网、软件工程、生物医学信息学、图书馆学以及信息架构中。大多数本体论描述了个体(实例)、类(概念)、属性和关系。(额外的细节见 [http://en.wikipedia.org/wiki/Ontology \(computer science\)](http://en.wikipedia.org/wiki/Ontology_(computer_science)))。

[0093] 例如,本体论可能指定“dog”具有可替换的值“poodle”。在该情况中,用户查询可以相对于列值匹配本体论,并且在用户录入他或她的查询时可以显示替换的本体论或者原始查询值,或者这两者。

[0094] 在另一可替换的实施例中,当查询匹配列名,并且该列名仅包含原始数据库中的数字值时,可以在用户接口中显示列名,其中具有额外的菜单显示符号指示该系统将显示与查询匹配的所有列值的总和、或者平均值、或者计数等,而不是个体值自身。

[0095] 图 12 示出了用户在图 11 中所示的查询框中录入了三个字母“nam”时,这些过程的结果的实例。这些字母与“All-Leads”表中的列名(“name”)部分匹配(以粗体示出)。它们还与“All-Leads”表中的多个行的内容部分匹配,与八行的“name”列中的“Robert Namais”部分匹配,以及与“All-Leads”表中的 18 行中出现的“Namibia”部分匹配。类似地,录入的字符串“nam”还与变平的“All-Contacts”表中的列名(“name”)和行值匹配。如图 12 所示,可以以可见层次的方式来显示结构化项目,诸如用户看到“Namibia”是“All-Leads”表的“country”字段中的值,或者“name”是该“All-Leads”表中的另一字段。所呈现的结构化项目帮助用户解释相关联的非空部分查询。每个结构化项目提供了对用户搜索意图的明确解疑,并因此向用户提供关于其搜索被正确解释和起作用的特定信任等级。可以根据任何适当的排序标准排序的层次的形式,例如,以字母表顺序、基于最常用的、或者基于找到的出现的次数,来呈现这种结构化项目。

[0096] 一旦用户已经录入了一些文本,如图 12 中所示,则用户可以决定下列:

[0097] 1. 输入更多的文本(产生诸如图 12 中所见的输出,其具有对应于录入的新字符串的值);

[0098] 2. 按下返回键,并因此验证该查询,其发送当前文本框的内容作为查询(见图 13);或者

[0099] 3. 点击显示器上的示例字段中的一个。

[0100] (A) 如果示例字段在内容部分包含部分匹配,则该匹配替换查询框中的当前文本。例如,如果用户点击了图 12 中的包含“Robert Namias”的框,则将向用户显示图 14。在该图中,我们看到,由于在“Robert Namias”上点击的字段在“All-leads”表的“name”字段中出现,因此替换用户录入的文本。这对应于图 4 中示出的过程 460(“在输入字段中显示所选择的查询”)。

[0101] (B) 如果用户点击对应于字段名的部分匹配,诸如图 12 的“All-contacts”部分中的字段名“name”,则该列名出现在查询框中,如图 15 中所示。

[0102] (C) 大体上,从自由文本部分移除并且由结构化部分替换查询框的自由部分中的所有字(跟随包含表名和列名或值的结构化部分)、与表名、列名或者选择点击的建议的

值相匹配的所有的这种剩余的字。

[0103] 如图 11 所示,如果查询框为空,并且用户点击表名,则在接口上显示表的列名,如图 16 中所示,其在用户点击图 11 中的表名“All-contacts”的情况下出现。

[0104] 在该点,用户可以录入自由文本,其对应的自动完成建议受限于该数据库表,如图 17 所示。可替换地,用户可以点击字段名,这会导致该字段名出现在查询框中,如图 18 中所示,并且因此,进一步的文本录入过程可信查询受限于该字段,如图 19 中所示。

[0105] 在图 20 中,示出了当用户已经选择了上述建议中的一个时的结果。在图 21 中,示出了将文本键入图 18 的可信查询中的结果的实例,其文本与图 18 的可信查询结果的一些其它列中的行内容匹配。用户还可以按下返回键以将当前的可信查询发送到基础搜索引擎,如图 22 中所示。

[0106] 现在参照图 6,示出了用于执行可信查询搜索的系统的一个实施例的高级硬件框图。该可信查询系统可以实现为与计算机硬件组件协作的系统 and / 或计算机实现的方法。该可信查询系统 110 可以包括多个软件模块或子系统。可以在硬件、软件、固件或硬件、软件和固件的任意组合中实现该模块或子系统,并且该模块或子系统可以或者可以不驻留在单个的物理或逻辑空间中。例如,本文档中所指的并且可能或者可能不在附图中示出的模块或子系统可以彼此远离,并且可以通过通信网络耦合。

[0107] 此外,图 6 是可以用于执行软件或逻辑以实现可信查询处理的计算机系统 600 的高级硬件框图。计算机 600 可以是个人计算机,并且可以包括各种硬件组件,诸如 RAM 614、ROM 616、硬盘存储 618、高速缓存存储器 620、数据库存储 622 等(也被称为“存储器子系统 627”)。计算机 600 可以包括任何适当的处理设备 628,诸如计算机、微处理器、RISC 处理器(精简指令集计算机)、CISC 处理器(复杂指令集计算机)、主帧计算机、工作站、单片计算机、分布式处理器、服务器、控制器、微控制器、分立逻辑计算机等,如本领域所知的。例如,处理设备 628 可以是 Intel Pentium®微处理器、x86 兼容的微处理器或者等价设备,并且可以被合并到服务器、个人计算机或任何合适的计算平台中。

[0108] 存储器子系统 626 可以包括任何适当的存储组件,诸如 RAM、EPROM(电可编程 ROM)、闪存、动态存储器、静态存储器、FIFO(先进先出)存储器、LIFO(后进先出)存储器、循环存储器、半导体存储器、泡存储器、缓冲存储器、盘存储器、光存储器、高速缓存存储器等。可以使用任何适当形式的存储器,无论是磁介质上的固定存储、半导体设备中的存储或者通过通信链路可访问的远程存储。用户或系统接口 630 可以耦合至计算机 600,并且可以包括各种输入设备 636,诸如可由系统管理器和 / 或键盘选择的开关。用户接口还可以包括适当的输出设备 640,诸如 LCD 显示器、CRT、各种 LED 指示器、打印机和 / 或语音输出设备,如本领域中所知的。

[0109] 为了促进计算机 600 和外部源的通信,通信接口 642 可以可操作地耦合至计算机系统。通信接口 642 可以是例如局域网、诸如以太网、企业内部互联网、因特网或其它适当的网络 544。通信接口 642 还可以连接至公共交换电话网 (PSTN) 646 或者 POTS(简单的旧的电话系统),这可以促进经由因特网 644 的通信。可以使用任何适当的商业上可获得的通信设备或网络。

[0110] 上述逻辑、电路和处理可以作为例如用于由处理器、控制器或其它处理设备执行的指令而被编码在或者存储在机器可读或计算机可读介质中,诸如光盘只读存储器

(CDROM)、磁或光盘、闪存、随机存取存储器 (RAM) 或只读存储器 (ROM)、可擦除可编程只读存储器 (EPROM) 或其它机器可读介质。

[0111] 可以将该介质实现为包含、存储、传送、传播或传输可执行指令以用于由指令可执行系统、装置或设备使用,或者连接至指令可执行系统、装置或设备的任何设备。可替换地或者额外地,可以使用诸如一个或多个集成电路或者一个或多个执行指令的处理器之类的硬件来将逻辑实现为模拟或数字逻辑、或者在应用程序接口 (API) 或动态链接库 (DLL) 中 (其中功能在共享存储器中可用或者被定义为本地或远程处理调用) 的软件中实现该逻辑;或者将逻辑实现为硬件和软件的组合。

[0112] 在其它的实现中,可以用信号或者传播信号的介质来表示逻辑。例如,实现任何给定程序的逻辑的指令可以采用电、磁、光、电磁、红外或其它类型的信号的形式。上述系统可以在诸如光纤接口、天线或其它模拟或数字信号接口之类的通信接口处接收这种信号,从信号恢复指令、将其存储在机器可读存储器中和 / 或利用处理器执行它们。

[0113] 该系统可以包括额外的或不同的逻辑,并且可以以多种不同的方式实现。可以将处理器实现为控制器、微处理器、微控制器、专用集成电路 (ASIC)、分立逻辑或其它类型的电路或逻辑的组合。类似地,存储器可以是 DRAM、SRAM、闪存或其它类型的存储器。参数 (例如,条件和阈值) 以及其它数据结构可以被分别存储和管理,可以被合并到单个存储器或数据库中,或者可以以多种不同的方式逻辑地和物理地组织。程序和指令可以是单个程序的部分、分别的程序或者分布在多个存储器和处理器中。

[0114] 已经描述了本发明的各种实施例,本领域普通技术人员将理解,在本发明的范围内,更多的实施例和实现也是可能的。因此,除了考虑所附权利要求及其等价物,本发明不受其它限制。

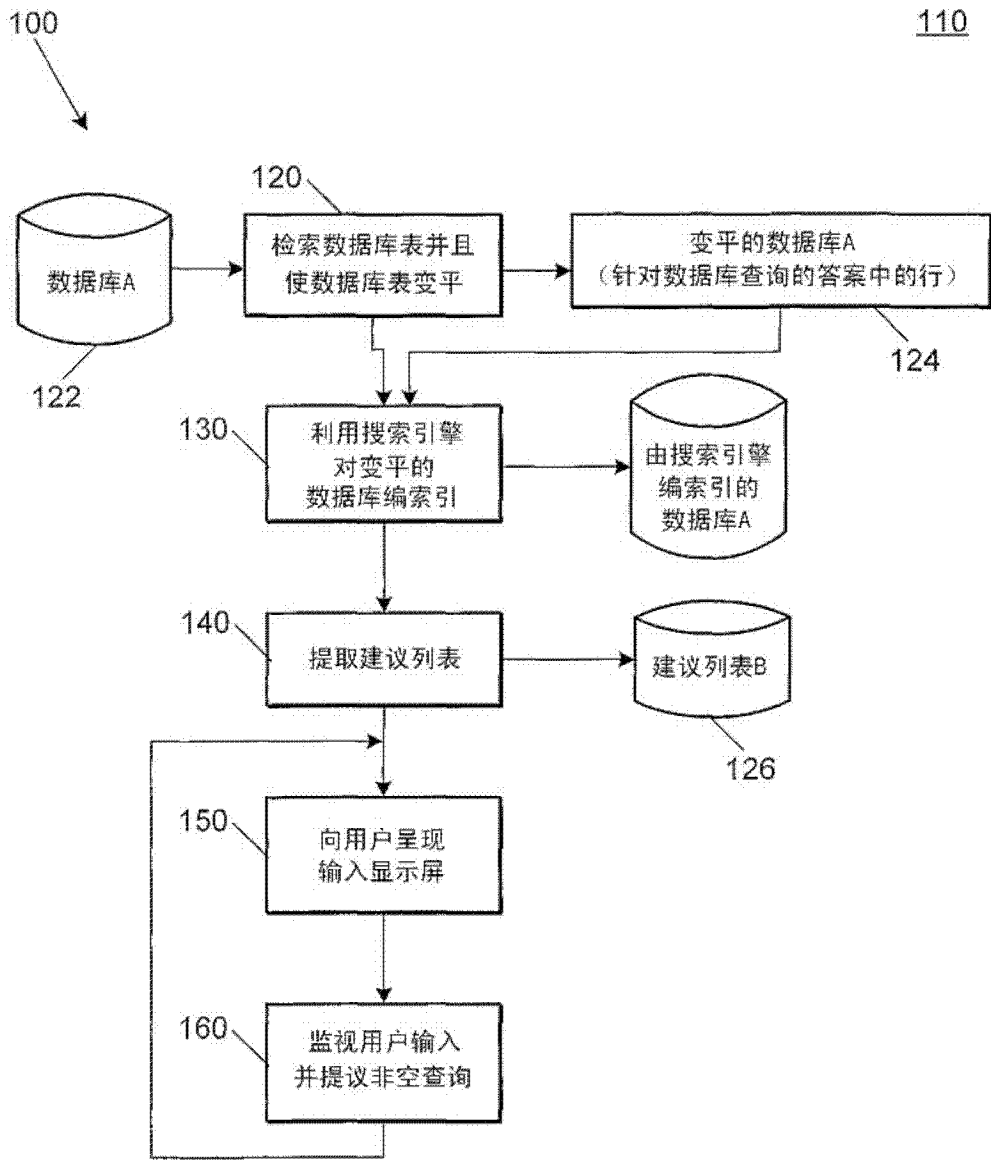


图 1

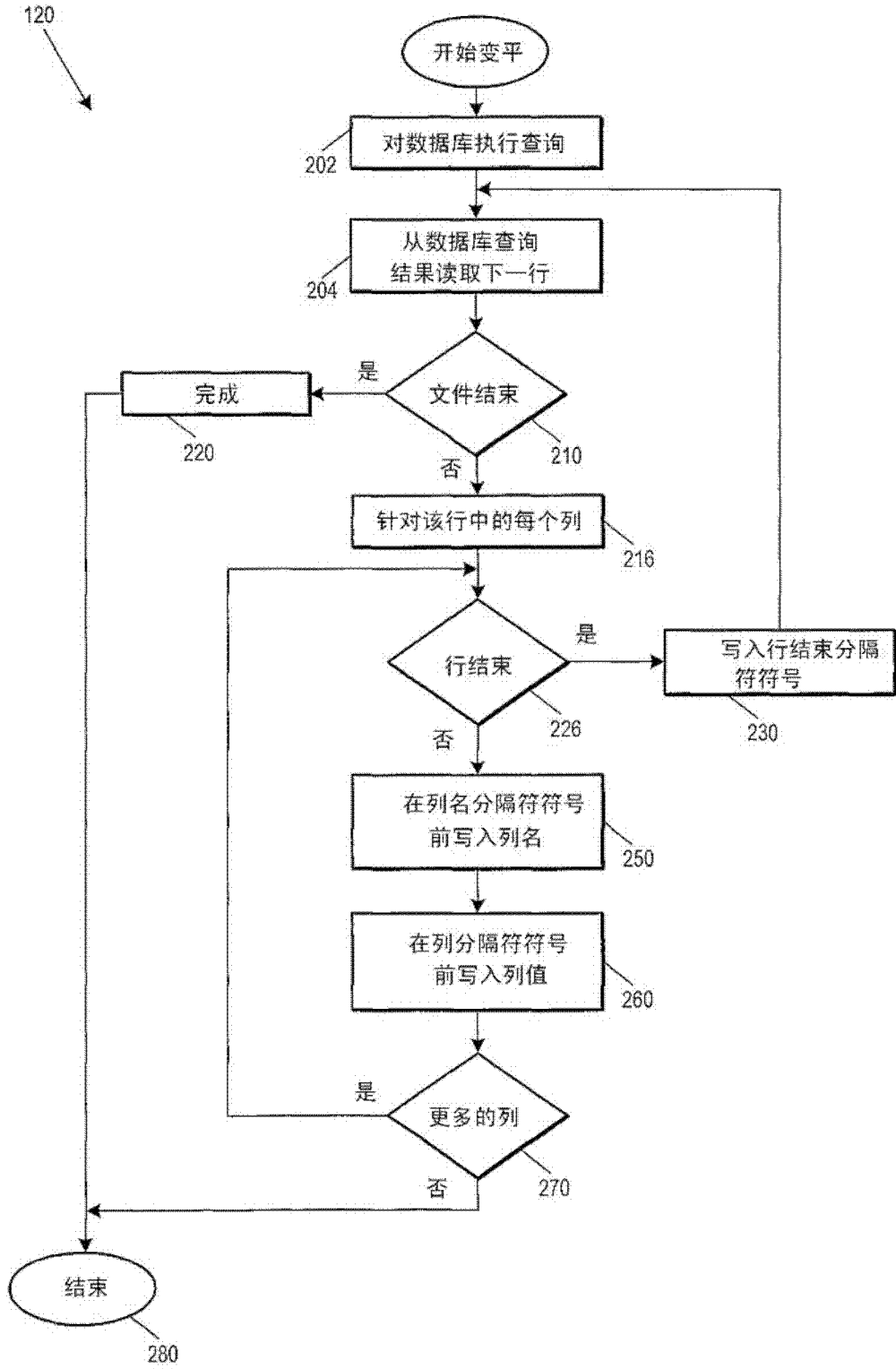


图 2

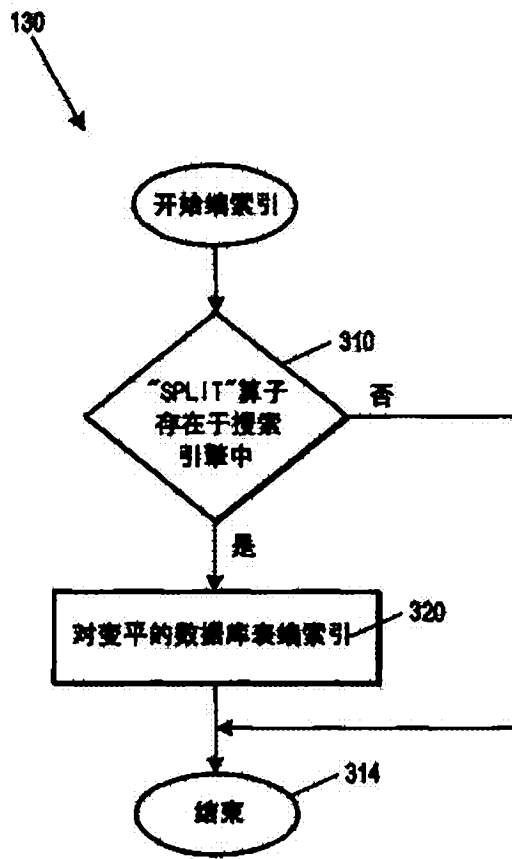


图 3

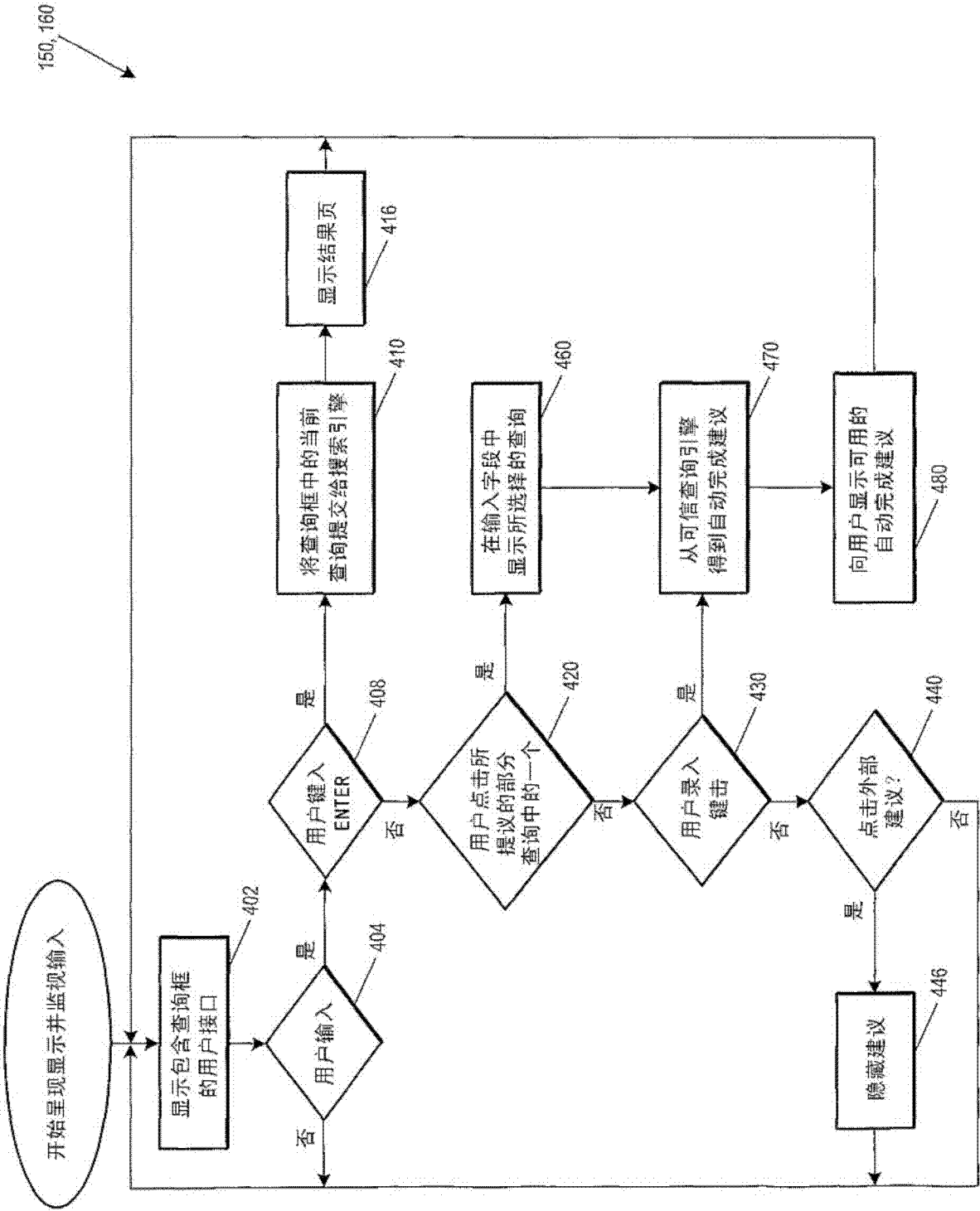


图 4

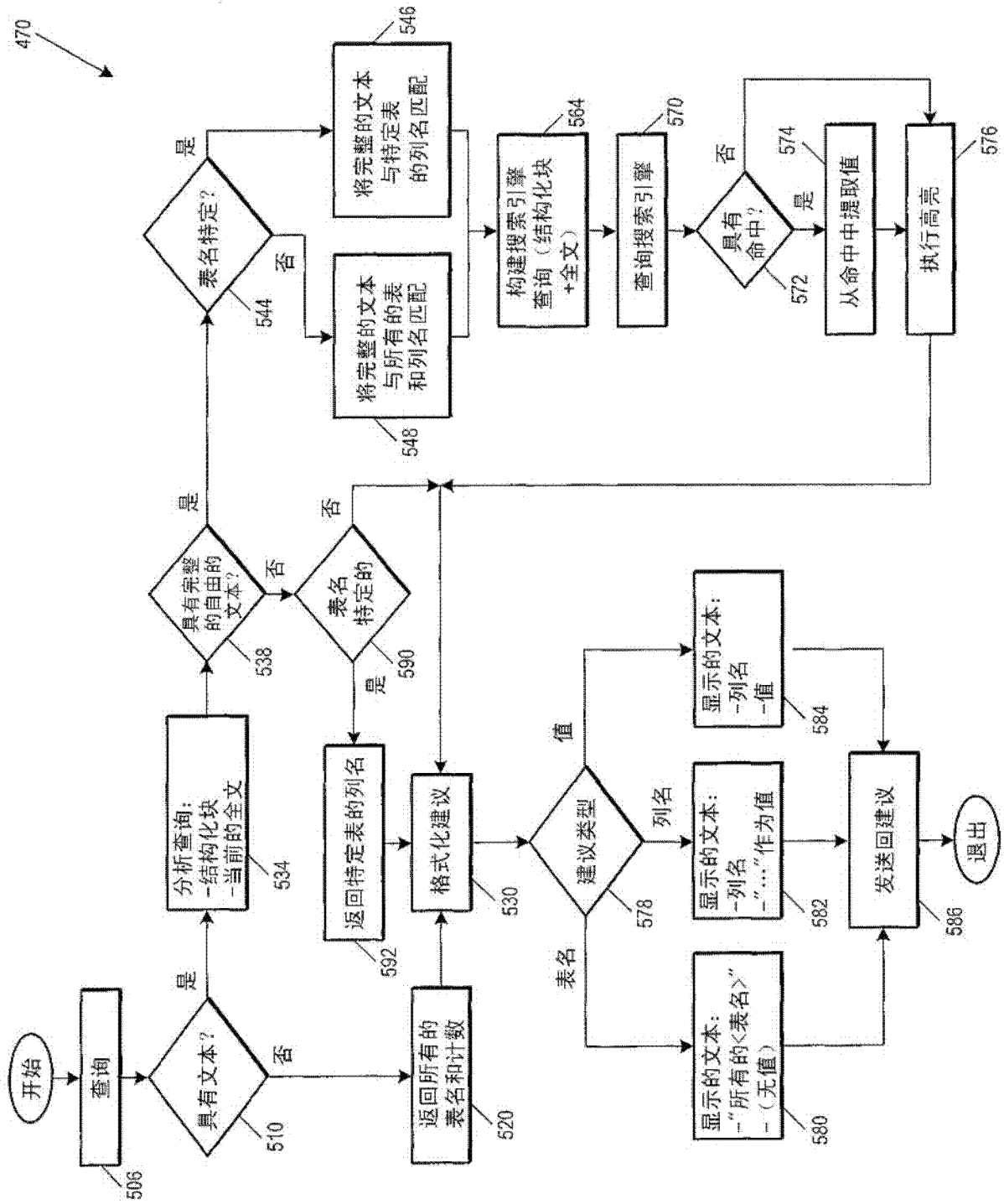


图 5

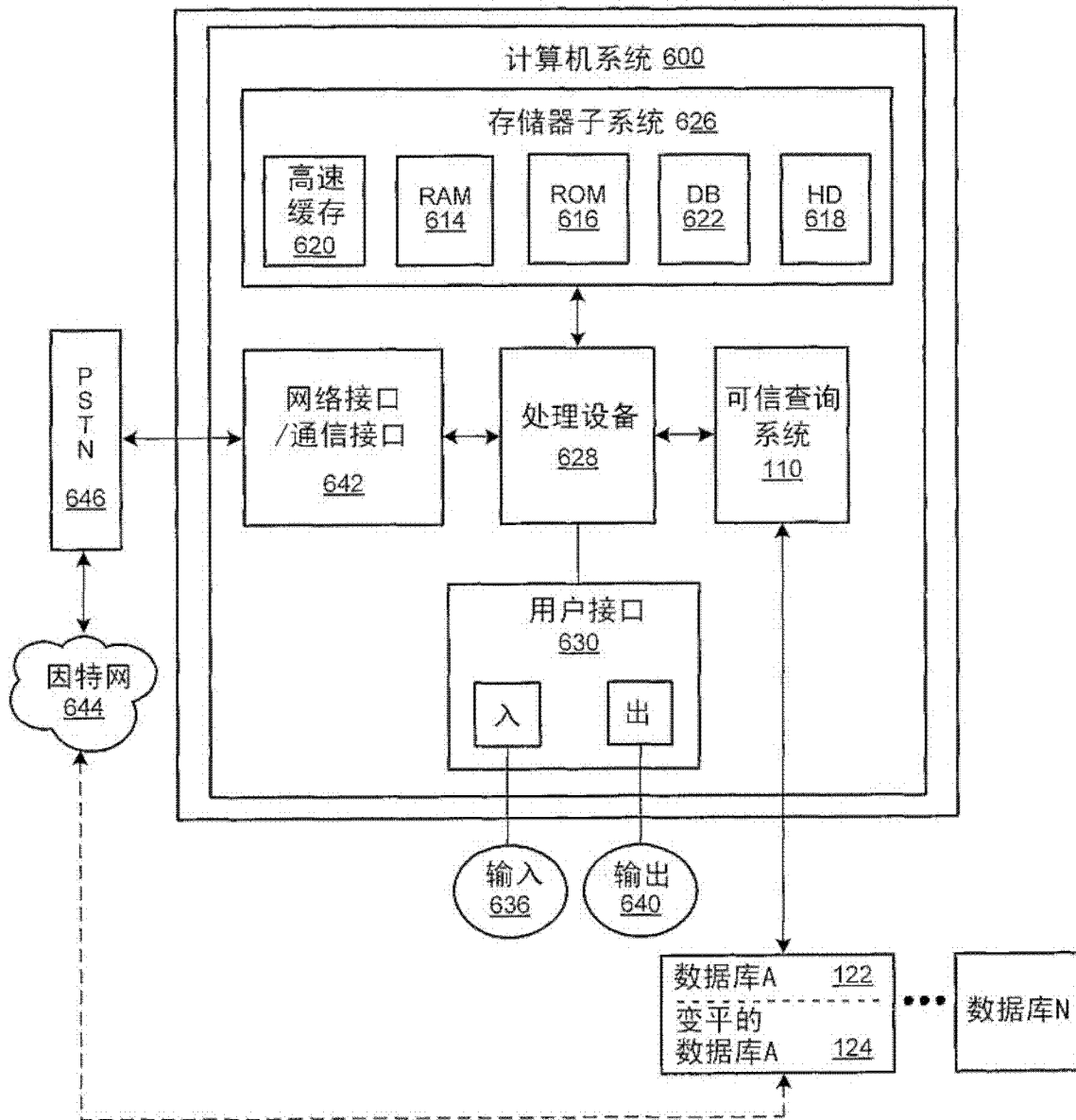


图 6

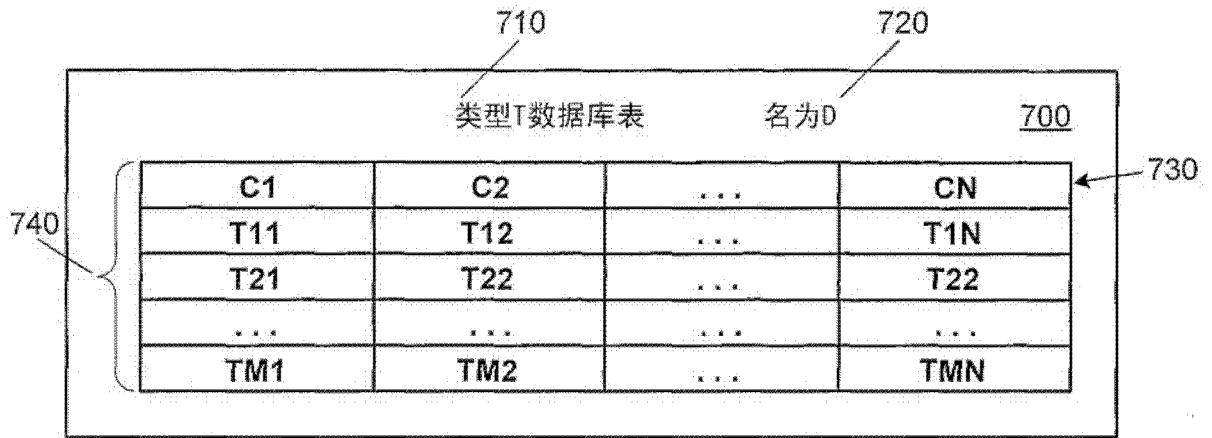


图 7

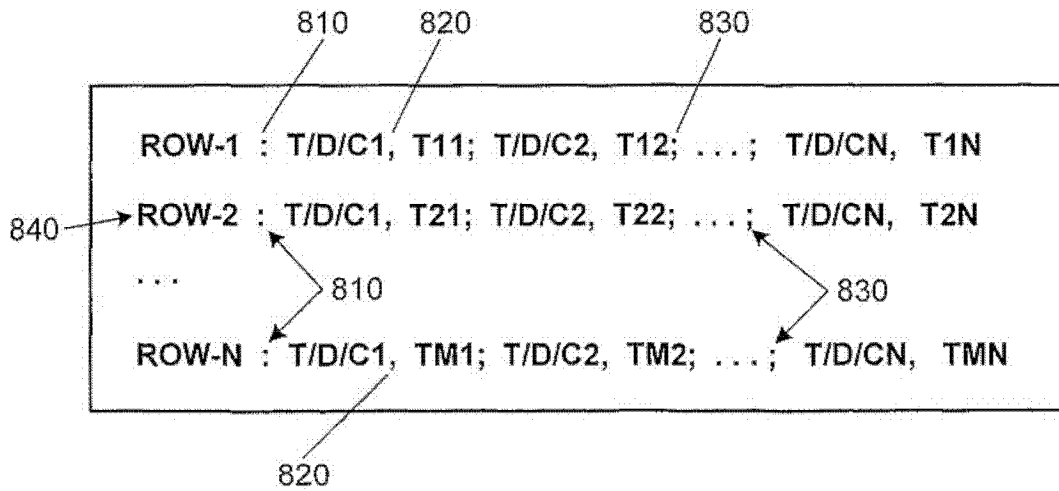


图 8

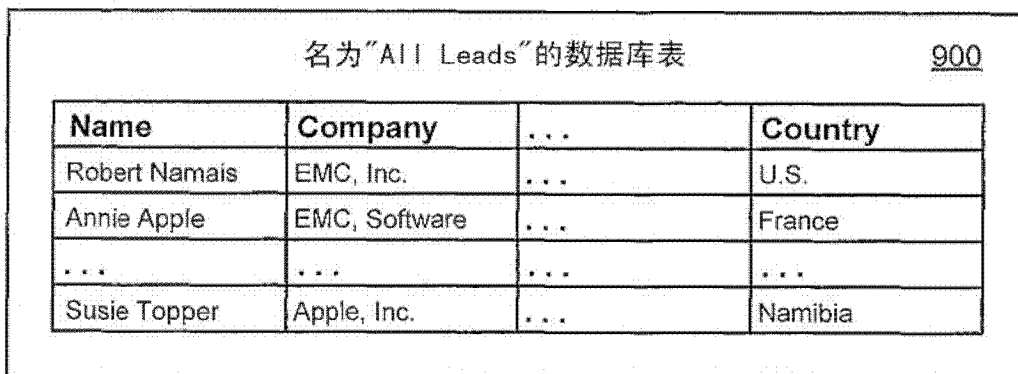


图 9

<u>1000</u>
ROW-1 : All-Leads/Name, Robert Namais; All-Leads/Company, EMC Inc.; . . . ; All-Leads/Country, US;
ROW-2 : All-Leads/Name, Annie Apple; All-Leads/Company, EMC Softwrae; . . . ; All-Leads/Country, France;
. . .
ROW-N : All-Leads/Name, Susie Topper; All-Leads/Company, Apple Inc.; . . . ; All-Leads/Country, Namibia;

图 10




<u>1100</u>		
	All leads — 1110	(97)
	All contacts — 1120	(73)
	All accounts — 1130	(21)

图 11


nam					
\$	All leads				
	<table border="1"> <tr> <td>name</td> <td>...</td> </tr> <tr> <td></td> <td>Robert Namias (8)</td> </tr> </table>	name	...		Robert Namias (8)
	name	...			
	Robert Namias (8)				
country	Namibia (13)				
	All contacts				
	<table border="1"> <tr> <td>name</td> <td>...</td> </tr> <tr> <td>title</td> <td>Naming Specialist (7)</td> </tr> </table>	name	...	title	Naming Specialist (7)
	name	...			
title	Naming Specialist (7)				

图 12

nam



Nam Inc
Country: India, CIO: Raman Tata



Eric **Nam**
Company: EMC, Title: CIO



Track & Trace
Company: **Nam** Inc, Amount: \$200,000

图 13

All leads

Name: Robert Namias

图 14

All contacts	name	
--------------	------	--

图 15

All contacts	
name	...
title	...
department	...
company	...

图 16

All contacts	nam
name	...
title	Naming Specialist (7)

图 17

All contacts	name	
--------------	------	--

图 18

All contacts	name	bro
Robert B rown		(2)
Richard B rownson		(1)
B roderick Ernst		(4)

图 19

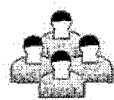
All contacts	Name: Robert Brown
--------------	--------------------

图 20

All contacts	Name: Robert Brown	ital
country	Italy	(1)

图 21

All contacts	name: Robert Brown	country: Italy
--------------	--------------------	----------------



Robert Brown
 Company: EMC, Title: CIO, **Country: Italy**

图 22