(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 117931858 B
(45)授权公告日 2024.07.16

(21)申请号 202410325663.4

(22)申请日 2024.03.21

(65)同一申请的已公布的文献号
申请公布号 CN 117931858 A

(43)申请公布日 2024.04.26

(73)专利权人 金蝶软件(中国)有限公司
地址 518000 广东省深圳市南山区科技园科技南十二路2号金蝶软件园A座1-8层

(72)发明人 刘博 宁洪波 赖宇斌 宁义双 宁可

(74)专利代理机构 华进联合专利商标代理有限公司 44224
专利代理师 姚姝娅

(51)Int.Cl.
G06F 16/2453(2019.01)
G06F 16/2455(2019.01)
G06F 16/2458(2019.01)

(56)对比文件
CN 116595026 A,2023.08.15
CN 117591547 A,2024.02.23

审查员 邓丽婉

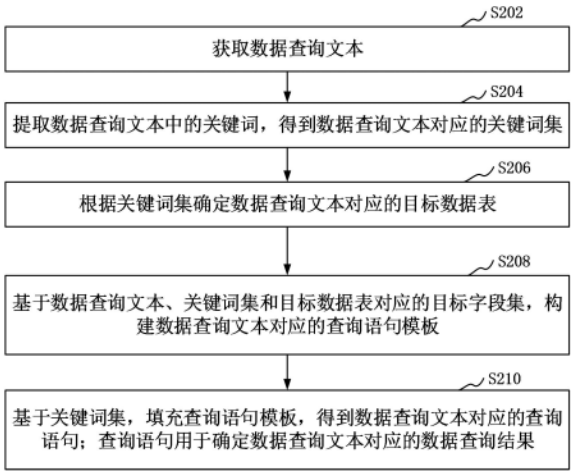
权利要求书3页 说明书13页 附图5页

(54)发明名称

数据查询方法、装置、计算机设备和存储介质

(57)摘要

本申请涉及一种数据查询方法、装置、计算机设备、存储介质和计算机程序产品。所述方法包括:获取数据查询文本;提取数据查询文本中的关键词,得到数据查询文本对应的关键词集;根据关键词集确定数据查询文本对应的目标数据表;基于数据查询文本、关键词集和目标数据表对应的目标字段集,构建数据查询文本对应的查询语句模板;基于关键词集,填充查询语句模板,得到数据查询文本对应的查询语句;查询语句用于确定数据查询文本对应的数据查询结果。采用本方法能够提高数据查询效率。



CN 117931858 B

1.一种数据查询方法,其特征在于,所述方法包括:

获取数据查询文本;

将所述数据查询文本切分为多个文本词;

基于各个所述文本词在所述数据查询文本中的关联关系,提取各个所述文本词分别对应的文本特征,基于所述数据查询文本对应的上下文信息,提取所述数据查询文本对应的文本特征;

将各个所述文本词分别对应的文本特征与所述数据查询文本对应的文本特征进行比对,得到各个所述文本词分别对应的关键指数;所述关键指数用于表示所述文本词为关键词的概率;

基于各个所述文本词分别对应的关键指数,提取所述数据查询文本对应的各个关键词,得到所述数据查询文本对应的关键词集;

根据所述关键词集确定所述数据查询文本对应的目标数据表;

基于所述数据查询文本提取所述关键词集中的各个所述关键词分别对应的属性特征;

基于各个所述关键词分别对应的属性特征,从所述目标字段集包括的各个目标字段中确定所述数据查询文本对应的返回属性字段和条件属性字段;

基于所述返回属性字段和所述条件属性字段,生成所述数据查询文本对应的查询语句模板;

基于所述关键词集,填充所述查询语句模板,得到所述数据查询文本对应的查询语句;所述查询语句用于确定所述数据查询文本对应的数据查询结果。

2.根据权利要求1所述的方法,其特征在于,所述方法还包括:

将所述数据查询文本输入查询语句生成模型,得到所述数据查询文本对应的查询语句;所述查询语句生成模型是基于多个数据查询样本和各个所述数据查询样本分别对应的样本标签集训练得到的;

所述查询语句生成模型包括用于提取所述数据查询文本对应的关键词集的关键词提取分支、用于生成所述数据查询文本对应的查询语句模板的模板生成分支、以及用于根据所述关键词集和所述查询语句模板生成查询语句的语句生成分支。

3.根据权利要求1所述的方法,其特征在于,所述根据所述关键词集确定所述数据查询文本对应的目标数据表,包括:

获取多个候选数据表分别对应的数据表描述信息;

计算所述数据查询文本与各个所述候选数据表分别对应的数据表描述信息之间的文本相似度,得到各个所述候选数据表分别对应的第一匹配度;

比对所述关键词集与各个所述候选数据表分别对应的候选字段集,得到各个所述候选数据表分别对应的第二匹配度;

融合同一候选数据表对应的第一匹配度和第二匹配度,分别得到各个所述候选数据表对应的目标匹配度;

基于各个所述候选数据表分别对应的目标匹配度,确定所述数据查询文本对应的目标数据表。

4.根据权利要求3所述的方法,其特征在于,所述比对所述关键词集与各个所述候选数据表分别对应的候选字段集,得到各个所述候选数据表分别对应的第二匹配度,包括:

提取并融合所述关键词集中各个关键词分别对应的文本特征,得到所述关键词集对应的综合文本特征;

提取并融合同一所述候选字段集中各个候选字段分别对应的文本特征,分别得到各个所述候选字段集对应的综合文本特征;

计算所述关键词集对应的综合文本特征和所述候选字段集对应的综合文本特征之间的相似度,分别得到各个所述候选数据表对应的第二匹配度。

5.根据权利要求1所述的方法,其特征在于,所述基于所述返回属性字段和所述条件属性字段,生成所述数据查询文本对应的查询语句模板,包括:

组合所述返回属性字段对应的语句连接词和所述返回属性字段,得到第一子句;

基于所述数据表标识对应的语句连接词,生成第二子句;

组合所述条件属性字段对应的语句连接词和所述条件属性字段,得到第三子句;

拼接所述第一子句、所述第二子句和所述第三子句,得到所述数据查询文本对应的查询语句模板。

6.根据权利要求1所述的方法,其特征在于,所述基于所述关键词集,填充所述查询语句模板,得到所述数据查询文本对应的查询语句,包括:

从所述关键词集中确定所述数据查询文本对应的条件属性字段所对应的条件关键词;

基于所述数据查询文本和所述条件关键词,提取所述条件关键词对应的词特征;

基于所述条件关键词对应的词特征,确定所述条件关键词对应的目标实体;

将所述目标实体填充至所述查询语句模板中,得到所述数据查询文本对应的查询语句。

7.根据权利要求6所述的方法,其特征在于,所述基于所述条件关键词对应的词特征,确定所述条件关键词对应的目标实体,包括:

基于所述条件关键词对应的词特征,在所述目标数据表中提取所述条件关键词对应的多个候选实体;

从所述目标数据表对应的数据表描述信息中,提取各个所述候选实体分别对应的基础实体特征;

基于各个所述候选实体分别对应的基础实体特征与所述数据查询文本对应的上下文信息之间的匹配度,在各个所述候选实体中确定所述条件关键词对应的目标实体。

8.一种数据查询装置,其特征在于,所述装置包括:

文本获取模块,用于获取数据查询文本;

关键词提取模块,用于将所述数据查询文本切分为多个文本词;基于各个所述文本词在所述数据查询文本中的关联关系,提取各个所述文本词分别对应的文本特征,基于所述数据查询文本对应的上下文信息,提取所述数据查询文本对应的文本特征;将各个所述文本词分别对应的文本特征与所述数据查询文本对应的文本特征进行比对,得到各个所述文本词分别对应的关键指数;所述关键指数用于表示所述文本词为关键词的概率;基于各个所述文本词分别对应的关键指数,提取所述数据查询文本对应的各个关键词,得到所述数据查询文本对应的关键词集;

数据表确定模块,用于根据所述关键词集确定所述数据查询文本对应的目标数据表;

模板构建模块,用于基于所述数据查询文本提取所述关键词集中的各个所述关键词分

别对应的属性特征基于各个所述关键词分别对应的属性特征,从所述目标字段集包括的各个目标字段中确定所述数据查询文本对应的返回属性字段和条件属性字段;基于所述返回属性字段和所述条件属性字段,生成所述数据查询文本对应的查询语句模板;

语句确定模块,用于基于所述关键词集,填充所述查询语句模板,得到所述数据查询文本对应的查询语句;所述查询语句用于确定所述数据查询文本对应的数据查询结果。

9.一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述的方法的步骤。

10.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至7中任一项所述的方法的步骤。

11.一种计算机程序产品,包括计算机程序,其特征在于,该计算机程序被处理器执行时实现权利要求1至7中任一项所述的方法的步骤。

数据查询方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及计算机技术领域，特别是涉及一种数据查询方法、装置、计算机设备、存储介质和计算机程序产品。

背景技术

[0002] 随着互联网的快速发展，信息量和数据量不断增长，数据库的重要性日益突出。数据库中存储着海量有价值的结构化数据，技术人员主要通过数据库查询语句来与数据库进行交互，但对于大部分非技术人员来说，往往是直接在信息查询系统输入需要查询的问题，从而实现与数据库之间的交互。

[0003] 传统技术中，在由用户输入的问题转换为相应的数据库查询语言去查询数据的过程中，存在数据查询效率低的问题。

发明内容

[0004] 基于此，有必要针对上述技术问题，提供一种能够提高数据查询效率的数据查询方法、装置、计算机设备、计算机可读存储介质和计算机程序产品。

[0005] 本申请提供了一种数据查询方法。所述方法包括：

[0006] 获取数据查询文本；

[0007] 提取数据查询文本中的关键词，得到数据查询文本对应的关键词集；

[0008] 根据关键词集确定数据查询文本对应的目标数据表；

[0009] 基于数据查询文本、关键词集和目标数据表对应的目标字段集，构建数据查询文本对应的查询语句模板；

[0010] 基于关键词集，填充查询语句模板，得到数据查询文本对应的查询语句；查询语句用于确定数据查询文本对应的数据查询结果。

[0011] 本申请还提供了一种数据查询装置。所述装置包括：

[0012] 文本获取模块，用于获取数据查询文本；

[0013] 关键词提取模块，用于提取数据查询文本中的关键词，得到数据查询文本对应的关键词集；

[0014] 数据表确定模块，用于根据关键词集确定数据查询文本对应的目标数据表；

[0015] 模板构建模块，用于基于数据查询文本、关键词集和目标数据表对应的目标字段集，构建数据查询文本对应的查询语句模板；

[0016] 语句确定模块，用于基于关键词集，填充查询语句模板，得到数据查询文本对应的查询语句；查询语句用于确定数据查询文本对应的数据查询结果。

[0017] 一种计算机设备，包括存储器和处理器，存储器存储有计算机程序，处理器执行计算机程序时实现上述数据查询方法的步骤。

[0018] 一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时实现上述数据查询方法的步骤。

[0019] 一种计算机程序产品，包括计算机程序，计算机程序被处理器执行时实现上述数据查询方法的步骤。

[0020] 上述数据查询方法、装置、计算机设备、存储介质和计算机程序产品，通过获取数据查询文本，从数据查询文本中提取关键词集。根据关键词集确定数据查询文本所针对的目标数据表。基于数据查询文本、关键词集和目标数据表对应的目标字段集，构建数据查询文本对应的查询语句模板。进而根据关键词集填充查询语句模板，得到数据查询文本对应的查询语句。执行查询语句可以得到数据查询文本对应的数据查询结果。这样，在获取到数据查询文本时，基于数据查询文本对应的关键词集确定数据查询文本所查询的目标数据表，进而根据目标数据表和关键词集构建查询语句模板，将关键词集填充至查询语句模板中，能够快速准确地得到数据查询文本对应的查询语句，有效提高查询语句的生成效率，从而提高数据查询的效率。

附图说明

[0021] 图1为一个实施例中数据查询方法的应用环境图；

[0022] 图2为一个实施例中数据查询方法的流程示意图；

[0023] 图3为一个实施例中确定目标数据表的流程示意图；

[0024] 图4为一个实施例中数据查询装置的结构框图；

[0025] 图5为一个实施例中计算机设备的内部结构图；

[0026] 图6为另一个实施例中计算机设备的内部结构图。

具体实施方式

[0027] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0028] 本申请实施例提供的数据查询方法，可以应用于如图1所示的应用环境中。其中，终端102通过网络与服务器104进行通信。数据存储系统可以存储服务器104需要处理的数据。数据存储系统可以集成在服务器104上，也可以放在云上或其他网络服务器上。终端102可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑、物联网设备和便携式可穿戴设备，物联网设备可为智能电视、智能车载设备等。便携式可穿戴设备可为智能手表、智能手环、头戴设备等。服务器104可以用独立的服务器或者是多个服务器组成的服务器集群来实现。终端102以及服务器104可以通过有线或无线通信方式进行直接或间接地连接，本申请在此不做限制。

[0029] 终端和服务器均可单独用于执行本申请实施例中提供的数据查询方法。

[0030] 终端和服务器也可协同用于执行本申请实施例中提供的数据查询方法。

[0031] 例如，终端向服务器发送获取到的数据查询文本。服务器提取数据查询文本中的关键词，得到数据查询文本对应的关键词集。服务器根据关键词集确定数据查询文本对应的目标数据表。服务器基于数据查询文本、关键词集和目标数据表对应的目标字段集，构建数据查询文本对应的查询语句模板。服务器基于关键词集，填充查询语句模板，得到数据查询文本对应的查询语句，查询语句用于确定数据查询文本对应的数据查询结果。服务器可

以将查询语句返回至终端,也可以执行查询语句,得到数据查询文本对应的数据查询结果,将数据查询结果范围值终端。

[0032] 在一个实施例中,如图2所示,提供了一种数据查询方法,以该方法应用于计算机设备为例进行说明,计算机设备可以是终端或服务器,由终端或服务器自身单独执行,也可以通过终端和服务器之间的交互来实现。数据查询方法包括以下步骤:

[0033] 步骤S202,获取数据查询文本。

[0034] 其中,数据查询文本是指用户向系统输入的、用于查询数据的文本。数据查询文本可以是任何形式的自然语言文本,例如问题、命令或对话。举例说明,在利润查询的场景中,用户可以向系统输入问句“A集团21年营收top5的产品是什么?”,该问句即为数据查询文本。

[0035] 示例性地,数据库中存储着海量有价值的结构化数据,技术用户主要通过输入符合数据库操作规则的数据库查询语句来与数据库进行交互,对于大部分不了解数据库知识的非技术用户而言,往往是输入自然语言形式的数据查询文本,计算机设备获取用户输入的数据查询文本,将数据查询文本转换为相应的数据库查询语句,使得非技术用户也能够拥有数据查询的能力。

[0036] 步骤S204,提取数据查询文本中的关键词,得到数据查询文本对应的关键词集。

[0037] 其中,关键词是指数据查询文本包含的用于生成数据查询语句的词汇,即数据查询文本中用于描述数据表、数据表字段、查询条件等信息的特定词汇。关键词集是指由数据查询文本包含的各个关键词组成的集合。

[0038] 示例性地,计算机设备将数据查询文本切分为多个文本词,从各个文本词中提取用于生成数据查询文本的关键词,得到关键词集。具体地,可以从多个业务领域获取数据查询样本,人为标注数据查询样本对应的关键词,得到各个数据查询样本分别对应的标签关键词集,将数据查询样本输入初始的关键词提取模型,得到数据查询样本对应的预测关键词集,基于标签关键词集和预测关键词集之间的差异计算模型损失,基于模型损失调整关键词提取模型的模型参数,直至模型收敛,得到用于关键词提取模型。将数据查询文本输入关键词提取模型,得到数据查询文本对应的关键词集。

[0039] 在一个实施例中,计算机设备也可以提取各个文本词分别对应的语义特征,基于数据查询文本对应的上下文信息提取数据查询文本对应的文本特征,基于上下文信息提取的文本特征能够更加准确性表征数据查询文本对应的文本含义。进而将各个文本词分别对应的文本特征和数据查询文本对应的文本特征进行比对,得到各个文本词分别对应的关键指数,关键指数用于表示文本词是关键词的概率。进而基于各个文本词分别对应的关键指数,提取数据查询文本对应的各个关键词得到关键词集。这样,基于文本词对应的文本特征和数据查询文本对应的文本特征共同来确定关键词,能够提高所确定的关键词的准确性,从而提高生成的查询语句的准确性。

[0040] 步骤S206,根据关键词集确定数据查询文本对应的目标数据表。

[0041] 其中,目标数据表是指数据查询文本所查询的数据表。例如,针对数据查询文本“A集团21年营收top5的产品是什么?”,数据查询文本对应的目标数据表为记录了各个集团分别对应的产品营收信息的利润表。

[0042] 示例性地,计算机设备将关键词集与各个候选数据表进行比对,将相关性最高的

候选数据表作为数据查询文本对应的目标数据表。具体地,可以将数据查询文本和各个候选数据表分别对应的数据表描述信息进行比对,得到各个候选数据表分别对应的第一匹配度,将第一匹配度最大值对应的候选数据表作为目标数据表。也可以将关键词集与各个候选数据表分别对应的候选字段集进行比对,得到各个候选数据表分别对应的第二匹配度,将第二匹配度最大值对应的候选数据表作为目标数据表。也就是基于候选数据表对应的第一匹配度和第二匹配点中的至少一者,计算各个候选数据表分别对应的目标匹配度,将目标匹配度最大值对应的候选数据表作为目标数据表。

[0043] 步骤S208,基于数据查询文本、关键词集和目标数据表对应的目标字段集,构建数据查询文本对应的查询语句模板。

[0044] 其中,目标数据表对应的目标字段集是指由目标数据表中的各个字段组成的集合。查询语句模板是指查询语句的基础结构,包含了将要填充的占位符,这些占位符会在后续的处理中被实际的字符串替换,例如,占位符可以被数据表名、数据表字段名和查询条件值等替换。

[0045] 示例性地,计算机设备基于数据查询文本,提取关键词集中各个关键词分别对应的特征向量,基于各个关键词分别对应的特征向量,在目标数据表中确定关键词对应的数据表字段。通过各个关键词分别对应的数据表字段,构建数据查询文本对应的查询语句模板。例如,针对数据查询文本“A集团21年营收top5的产品是什么?”,关键词集为“产品、A集团、21年、top5”,关键词“产品”对应的是目标数据表中的数据表字段“产品”,关键词“A集团”对应的是目标数据表中的数据表字段“公司名称”,关键词“21年”对应的是目标数据表中的数据表字段“年份”,关键词“top5”对应的是目标数据表中的数据表字段“营业收入”。

[0046] 步骤S210,基于关键词集,填充查询语句模板,得到数据查询文本对应的查询语句;查询语句用于确定数据查询文本对应的数据查询结果。

[0047] 其中,查询语句是指基于数据查询文本生成的数据库查询语句,用于执行针对数据表的查询操作,例如,查询语句可以是SQL(Structured Query Language,结构化查询语言)语句,对Mysql、SQL Server、Hive等数据库进行查询,还可以是非关系型数据库查询语句。

[0048] 示例性地,计算机设备提取关键词集中的关键词在目标数据表中对应的目标实体,将关键词对应的目标实体填充至查询语句模板中,得到数据查询文本对应的查询语句。目标实体是指关键词在目标数据表中对应的字段值。执行查询语句,得到数据查询文本对应的数据查询结果,将数据查询结果返回至数据查询文本对应发送方。

[0049] 上述数据查询方法中,通过获取数据查询文本,从数据查询文本中提取关键词集。根据关键词集确定数据查询文本所针对的目标数据表。基于数据查询文本、关键词集和目标数据表对应的目标字段集,构建数据查询文本对应的查询语句模板。进而根据关键词集填充查询语句模板,得到数据查询文本对应的查询语句。执行查询语句可以得到数据查询文本对应的数据查询结果。这样,在获取到数据查询文本时,基于数据查询文本对应的关键词集确定数据查询文本所查询的目标数据表,进而根据目标数据表和关键词集构建查询语句模板,将关键词集填充至查询语句模板中,能够快速准确地得到数据查询文本对应的查询语句,有效提高查询语句的生成效率,从而提高数据查询的效率。

[0050] 在一个实施例中,数据查询方法还包括:

[0051] 将数据查询文本输入查询语句生成模型,得到数据查询文本对应的查询语句;查询语句生成模型是基于多个数据查询样本和各个数据查询样本分别对应的样本标签集训练得到的;查询语句生成模型包括用于提取数据查询文本对应的关键词集的关键词提取分支、用于生成数据查询文本对应的查询语句模板的模板生成分支、以及用于根据关键词集和查询语句模板生成查询语句的语句生成分支。

[0052] 其中,查询语句生成模型是指用于生成数据查询文本对应的查询语句的模型,查询语句生成模型的输入数据为数据查询文本,输出数据为数据查询文本对应的查询语句。数据查询样本是指从多个业务领域分别对应的数据查询语料库中获取的数据查询文本,数据查询样本对应的样本标签集包括数据查询样本对应的关键词集、查询语句模板和查询语句中的至少一者。

[0053] 示例性地,计算机设备将数据查询文本输入查询语句生成模型,首先由查询语句生成模型中的关键词提取分支,提取数据查询文本对应的关键词集。查询语句生成模型基于关键词集确定数据查询文本对应的目标数据表,进而将数据查询文本、关键词集和目标数据表对应的目标字段集输入模板生成分支,得到数据查询文本对应的查询语句模板。查询语句生成模型提取各个关键词在目标数据表中分别对应的目标实体,将数据查询文本、各个关键词分别对应的目标实体,输入语句生成分支,得到数据查询文本对应的查询语句。

[0054] 上述实施例中,通过从多个不同的业务领域分别对应的数据查询语料库中获取数据查询样本,基于各个数据查询样本训练查询语句生成模型,能够提高查询语句生成模型的泛化性,从而提高所生成的查询语句的准确性。查询语句生成模型首先通过关键词提取分支提取关键词集,进而通过模板生成分支生成查询语句模板,最后通过语句对查询语句模板进行关键词补全,能够快速准确地得到查询语句。

[0055] 在一个实施例中,提取数据查询文本中的关键词,得到数据查询文本对应的关键词集,包括:

[0056] 提取数据查询文本中的多个文本词;

[0057] 基于各个文本词在数据查询文本中的关联关系,提取各个文本词分别对应的文本特征;

[0058] 基于各个文本词分别对应的文本特征,在各个文本词中确定目标词;

[0059] 基于目标词得到数据查询文本对应的关键词集。

[0060] 其中,文本词是指对数据查询文本进行分词处理得到的各个词汇,即组成数据查询文本的词汇。文本特征是指用于表征文本词对应的语义信息和文本词与其他各个文本词之间的关联关系的特征向量。目标词是指数据查询文本包含的关键词。

[0061] 示例性地,计算机设备对数据查询文本进行预处理,去除标点符号等无关信息。从预处理后的数据查询文本中提取多个文本词。进而通过预先训练好的文本特征提取模型,提取各个文本词分别对应的文本特征,文本特征提取模型是基于大量的语料样本训练得到的,能够捕获数据查询文本中的各个文本词之间的语义关系。基于根据各个文本词分别对应的文本特征,将各个文本词分别对应的文本特征与查询文本对应的文本特征进行比对,得到各个文本词分别对应的关键指数,关键指数用于表示文本词是关键词的概率。进而基于各个文本词分别对应的关键指数,提取数据查询文本对应的各个关键词得到关键词集。具体地,可以将关键指数大于预设阈值的文本词作为关键词。

[0062] 上述实施例中,基于文本词对应的文本特征和数据查询文本对应的文本特征共同来确定关键词,综合考虑了文本词的自身的语义信息和文本词在数据查询文本中对应的上下文信息,能够提高所确定的关键词的准确性,从而提高生成的查询语句的准确性。

[0063] 在一个实施例中,如图3所示,根据关键词集确定数据查询文本对应的目标数据表,包括:

[0064] 步骤S302,获取多个候选数据表分别对应的数据表描述信息。

[0065] 步骤S304,计算数据查询文本与各个候选数据表分别对应的数据表描述信息之间的文本相似度,得到各个候选数据表分别对应的第一匹配度。

[0066] 步骤S306,比对关键词集与各个候选数据表分别对应的候选字段集,得到各个候选数据表分别对应的第二匹配度。

[0067] 步骤S308,融合同一候选数据表对应的第一匹配度和第二匹配度,分别得到各个候选数据表对应的目标匹配度。

[0068] 步骤S310,基于各个候选数据表分别对应的目标匹配度,确定数据查询文本对应的目标数据表。

[0069] 其中,数据表描述信息是指用自然语言(如中文、英文等)对数据表的内容、结构、字段含义等信息进行解释和描述的信息,通过易于理解的文字来阐述数据表的基本信息和用途。候选数据表是指数据库中存储的数据表。候选字段集是指包含候选数据表中的各个数据表字段的集合。

[0070] 第一匹配度是指数据查询文本和候选数据表之间的匹配程度。第二匹配度是指关键词集与候选数据表对应的候选字段集之间的匹配程度。目标匹配度是指融合第一匹配度和第二匹配度得到的候选数据表与数据查询文本之间的匹配度。匹配度用于指示候选数据表是数据查询文本所查询的数据表的概率。

[0071] 示例性地,计算机设备获取各个候选数据表分别对应的数据表描述信息。提取各个数据表描述信息分别对应的文本特征,以及数据查询文本对应的文本特征。计算数据表描述信息对应的文本特征与数据查询文本之间的文本相似度,分别得到各个候选数据表与数据查询文本之间的第一匹配度。比对关键词集与候选数据表对应的候选字段集,得到各个候选数据表分别对应的第一匹配度。具体地,可以提取并融合关键词集中各个关键词分别对应的文本特征,得到关键词集对应的综合文本特征,提取并融合候选字段集中各个候选字段分别对应的文本特征,得到候选字段集对应的综合文本特征。计算关键词集对应的综合文本特征和候选字段集对应的综合文本特征之间的相似度,得到候选数据表对应的第二匹配度。融合同一候选数据表对应的第一匹配度和第二匹配度,分别得到各个候选数据表对应的目标匹配度。进而将目标匹配度最大值对应的候选数据表作为数据查询文本对应的目标数据表。

[0072] 上述实施例中,基于数据查询文本和候选数据表对应的数据表描述信息之间的相似度,关键词集和候选数据表对应的候选字段集之间的相似度,共同确定目标数据表,能够提高所确定的目标数据表的准确性,从而提高生成的查询语句的准确性。

[0073] 在一个实施例中,基于数据查询文本、关键词集和目标数据表对应的目标字段集,构建数据查询文本对应的查询语句模板,包括:

[0074] 基于数据查询文本提取关键词集中的各个关键词分别对应的属性特征;

[0075] 基于各个关键词分别对应的属性特征,从目标字段集包括的各个目标字段中确定数据查询文本对应的返回属性字段和条件属性字段;

[0076] 基于返回属性字段和条件属性字段,生成数据查询文本对应的查询语句模板。

[0077] 其中,属性特征用于指示关键词分别与查询语句模板对应的返回属性和条件属性之间的匹配程度。返回属性在查询语句中用于指示在数据表中查询的字段。条件属性在查询语句中用于指示查询条件。返回属性字段是指在数据表中确定的返回属性对应的数据表字段。条件属性字段是指在数据表中确定的条件属性对应的字段。例如,当数据查询文本为“A集团21年营收top5的产品是什么?”时,数据查询文本对应的关键词集为“产品、A集团、21年、top5”,查询语句模板为SELECT 产品 FROM 利润表 WHERE 公司名称 = “候选词” and 年份 in “候选词” order by 营业收入 limit “候选词”;”其中,SELECT后跟随的第一个字符串“产品”即为返回属性字段,FROM后跟随的第一个字符串“利润表”即为数据表标识,WHERE后跟随的“公司名称”、“年份”和“营业收入”即为条件属性字段,“候选词”为占位符,可以被实际的查询条件值替换。

[0078] 示例性地,计算机设备基于数据查询文本提取关键词集中各个关键词在数据查询文本中对应的属性特征。基于各个关键词分别对应的属性特征,确定返回属性对应的关键词和条件属性对应的关键词。进而在目标数据表对应的目标字段集中,获取返回属性对应的关键词所对应的目标字段作为返回属性字段,以及获取条件属性对应的关键词所对应的目标字段作为条件属性字段。例如,当数据查询文本为“A集团21年营收top5的产品是什么?”时,数据查询文本对应的关键词集为“产品、A集团、21年、top5”,“产品”为返回属性的关键词,关键词“产品”在目标数据表中对应的返回属性字段为“产品”,“A集团”、“21年”、“top5”均为条件属性的关键词,关键词“A集团”在目标数据表中对应的条件属性字段为“公司名称”。组合返回属性字段对应的语句连接词和返回属性字段,得到第一子句,例如,在SQL语句中,返回属性字段对应的连接词为SELECT。基于数据表标识对应的语句连接词生成第二子句。组合条件属性字段对应的语句连接词和条件属性字段,得到第三子句。拼接第一子句、第二子句和第三子句,得到数据查询文本对应的查询语句模板。

[0079] 上述实施例中,首先提取组成查询语句模板的返回属性字段,条件属性字段,进而基于相应的语句连接词组合各个属性字段和数据表标识分别对应的子句,最后拼接各个子句,能够快速准确地生成数据查询文本对应的查询语句模板,从而提高数据查询的效率和准确性。

[0080] 在一个实施例中,基于关键词集,填充查询语句模板,得到数据查询文本对应的查询语句,包括:

[0081] 从关键词集中确定数据查询文本对应的条件属性字段所对应的条件关键词;

[0082] 基于数据查询文本和条件关键词,提取条件关键词对应的词特征;

[0083] 基于条件关键词对应的词特征,确定条件关键词对应的目标实体;

[0084] 将目标实体填充至查询语句模板中,得到数据查询文本对应的查询语句。

[0085] 其中,条件关键词是指关键词集中用于指示条件属性字段对应的字段值(即查询条件值)的关键词。条件关键词对应的词特征是指用于表征条件关键词在数据查询文本中的语义信息的特征向量。目标实体是指将条件关键词在数据库中对应的字段值。例如,当数据查询文本为“A集团21年营收top5的产品是什么?”时,查询语句模板为“SELECT 产品

FROM 利润表 WHERE 公司名称 = “候选词”and 年份 in “候选词” order by 营业收入 limit “候选词”;”,“A集团”、“21年”、“top5”即为数据查询文本对应的多个条件关键词,“A集团”对应的目标实体为“Axx Group”,“21年”对应的目标实体为“2021”,“top5”对应的目标实体为“5”,将目标实体填充至查询语句模板中得到的查询语句为“SELECT 产品 FROM 利润表 WHERE 公司名称 = ″ Axx Group″ and 年份 in (″2021″) order by 营业收入 limit 5;”。

[0086] 示例性地,计算机设备从关键词集中确定数据查询文本对应的条件属性字段所对应条件关键词,即属于条件属性对应的关键词。条件属性字段可以有一个或多个,确定各个条件属性字段分别对应的条件关键词。进而提取各个条件关键词分别对应的词特征,基于各个条件关键词分别对应的词特征,确定各个条件关键词分别对应的目标实体。

[0087] 计算机设备将每个条件关键词对应的目标实体和目标数据表对应的数据表标识填充至查询语句模板中,得到数据查询文本对应的查询语句。数据表标识是指数据表的表名。具体地,查询语句模板中,数据表标识对应一个占位符,每个条件属性字段均对应一个占位符,基于数据表标识替换数据表标识对应的占位符,基于条件属性字段对应的条件关键词,替换该条件属性字段对应的占位符,直至每个条件属性字段分别对应的占位符均被相应的条件关键词替换,得到数据查询文本对应的查询语句。

[0088] 在一些实施例中,将数据查询文本和条件关键词输入实体预测模型,实体预测模型首先提取条件关键词在数据查询文本中对应的词特征,进而基于条件关键词对应的词特征,预测条件关键词对应的目标实体。实体预测模型可以是一个单独的神经网络模型,也可以是查询语言生成模型中包含的模型分支。实体预测模型是基于语料库中的多个数据查询样本和数据查询样本对应的样本标签集训练得到的,样本标签集包括数据查询样本对应的关键词集、关键词集中各个关键词分别目标实体、查询语句模板和查询语句。

[0089] 在一些实施例中,基于条件关键词对应的词特征,确定条件关键词与目标数据表中的各个字段值之间的语义相似度,将语义相似度最大值的字段值作为条件关键词对应的目标实体。这样,在目标数据表包含的各个字段值中确定条件关键词对应的目标实体,能够提高所确定的目标实体的准确性。

[0090] 上述实施例中,确定各个条件属性字段分别对应的条件关键词后,进一步基于条件关键词和数据查询文本,提取条件关键词对应的词特征。进而基于关键词对应的词特征,确定关键词对应的目标实体,最后将目标实体填充至查询语句模板中,能够消除自然语言的多样性所带来的歧义,这样得到的查询语句更加准确,能够提高数据查询的准确性。

[0091] 在一个实施例中,基于条件关键词对应的词特征,确定条件关键词对应的目标实体,包括:

[0092] 基于条件关键词对应的词特征,在目标数据表中提取条件关键词对应的多个候选实体;

[0093] 从目标数据表对应的数据表描述信息中,提取各个候选实体分别对应的基础实体特征;

[0094] 基于各个候选实体分别对应的基础实体特征与数据查询文本对应的上下文信息之间的匹配度,在各个候选实体中确定条件关键词对应的目标实体。

[0095] 其中,候选实体是指将条件关键词映射至目标数据表中得到的候选字段值。基础

实体特征是指对候选实体进行特征提取得到的特征向量,包含了候选实体对应的语义信息和候选实体与其他各个候选实体之间的语义关系等信息。数据查询文本对应的上下文信息包括数据查询文本对应的用户查询历史和用户查询状态,是用于辅助理解数据查询文本的重要依据。

[0096] 示例性地,计算机设备通过多路召回策略,在目标数据表包含的各个字段值中,根据条件关键词对应的词特征,提取条件关键词对应的多个候选实体。例如,可以通过关键词匹配、向量检索、同义词命中、主题识别等不同的召回算法,快速地从目标数据表包含的大量数据中检索到每个条件关键词分别对应的多个候选实体。进而,从目标数据表对应的数据表描述信息中,提取每个候选实体分别对应的基础实体特征。进而从数据查询文本对应的上下文信息中,提取数据查询文本对应的上下文特征。基于每个候选实体对应的基础实体特征和上下文特征,确定每个候选实体分别与数据查询文本对应的上下文信息之间的匹配度,将匹配度最高的候选实体作为条件关键词对应的目标实体。

[0097] 上述实施例中,首先通过多路召回策略,在目标数据表中快速确定条件关键词对应的少量候选实体,大大缩小了目标实体的检索范围,能够提高确定目标实体的效率。进而基于候选实体对应的基础特征特征和数据查询文本对应上下文信息之间的匹配度,能够快速准确地在各个候选实体中确定目标实体,提高数据查询的效率和准确性。

[0098] 在一个具体的实施例中,本申请提出的数据查询方法可以应用于信息查询系统。数据查询方法包括以下步骤:

[0099] 1、构建语料库

[0100] 信息查询系统提取用户在信息查询系统中查询数据库数据时输入的各种数据查询样本,并通过SQL专家对数据查询样本进行标注,得到各个数据查询样本分别对应的样本标签。基于各个数据查询样本和数据查询样本对应的样本标签,得到通用领域的语料库。例如,语料库规模可以为10000个数据查询样本,其中8000条作为训练集,2000条作为验证集。训练样本的数据格式包括用户的问题(即数据查询样本)和样本标签,样本标签包括问题所针对的数据表的名称、数据表的字段模式和SQL模板等信息。例如,训练样本可以为以下形式:

[0101] {表格名称:“利润表”,

[0102] 表格字段: “营业收入,营业支出,产品,公司名称,年份,营业收入等”,

[0103] 问题: “A集团21年营收top5的产品是什么?”,

[0104] 关键词:“产品,A集团,21年,top5”,

[0105] SQL模板: SELECT 产品 FROM 利润表 WHERE 公司名称 = “候选词” and 年份 in “候选词” order by “营业收入” limit “候选词”;”}

[0106] 2、模型构建

[0107] 信息查询系统基于构建的语料库,训练SQL语句生成模型。SQL语句生成模型可以是大语言模型,即一种基于海量文本数据训练的深度学习模型。SQL语句生成模型包括关键词识别和SQL模板生成两个分支。每个分支的训练过程主要包括数据输入、编码器工作和结果输出三个步骤。在数据输入步骤中,SQL语句生成模型的输入数据可以是任何形式的自然语言文本,例如,问题、命令或对话。输入数据被模型转化为“token”格式,即模型能够理解和处理的数据格式。编码器包含多层自注意力机制(Self-Attention Mechanism),能够捕

获文本中的各种复杂关系,并生成一个上下文相关的词向量表示。经过编码器处理后,生成一个预测序列,预测序列由一系列概率最高的token组成,将预测序列转化为自然语言文本,得到输出数据。对于关键词识别模型,预测序列即为各个关键词,对于SQL模板生成模型,预测序列即为SQL模板。在获取到数据查询文本后,将数据查询文本输入SQL语句生成模型,得到对应的关键词集和SQL模板,对关键词集进行实体对齐,映射为数据库中的标准术语(即目标实体),最后将各个关键词分别对应的目标实体填入SQL模板中,得到最终的SQL语句。

[0108] 上述实施例中,通过SQL模板生成技术、实体对齐技术和SQL语句补全技术,能够为快速将数据查询文本转化为相应的SQL语句。解决了用户通过自然语言,直接在各类信息查询系统中进行查询的问题,提升了用户查询的效率。并先通过先生成较为通用的SQL草稿(即SQL模板),再结合大模型出色的泛化能力对SQL草稿进行关键词补全,从而提高NL2SQL(Natural Language to SQL,即将用户的自然语言转换为可执行的SQL语句的技术)技术在通用领域的泛化能力。

[0109] 应该理解的是,虽然如上的各实施例所涉及的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,如上的各实施例所涉及的流程图中的至少一部分步骤可以包括多个步骤或者多个阶段,这些步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0110] 基于同样的发明构思,本申请实施例还提供了一种用于实现上述所涉及的数据查询方法的数据查询装置。该装置所提供的解决问题的实现方案与上述方法中所记载的实现方案相似,故下面所提供的一个或多个数据查询装置实施例中的具体限定可以参见上文中对于数据查询方法的限定,在此不再赘述。

[0111] 在一个实施例中,如图4所示,提供了一种数据查询装置,包括:文本获取模块402、关键词提取模块404、数据表确定模块406、模板构建模块408和语句确定模块410,其中:

[0112] 文本获取模块402,用于获取数据查询文本。

[0113] 关键词提取模块404,用于提取数据查询文本中的关键词,得到数据查询文本对应的关键词集。

[0114] 数据表确定模块406,用于根据关键词集确定数据查询文本对应的目标数据表。

[0115] 模板构建模块408,用于基于数据查询文本、关键词集和目标数据表对应的目标字段集,构建数据查询文本对应的查询语句模板。

[0116] 语句确定模块410,用于基于关键词集,填充查询语句模板,得到数据查询文本对应的查询语句;查询语句用于确定数据查询文本对应的数据查询结果。

[0117] 在一个实施例中,数据查询装置还包括模型处理模块,模型处理模块用于将数据查询文本输入查询语句生成模型,得到数据查询文本对应的查询语句;查询语句生成模型是基于多个数据查询样本和各个数据查询样本分别对应的样本标签集训练得到的;查询语句生成模型包括用于提取数据查询文本对应的关键词集的关键词提取分支、用于生成数据查询文本对应的查询语句模板的模板生成分支、以及用于根据关键词集和查询语句模板生

成查询语句的语句生成分支。

[0118] 在一个实施例中,关键词提取模块404还用于:

[0119] 提取数据查询文本中的多个文本词;基于各个文本词在数据查询文本中的关联关系,提取各个文本词分别对应的文本特征;基于各个文本词分别对应的文本特征,在各个文本词中确定目标词基于目标词得到数据查询文本对应的关键词集。

[0120] 在一个实施例中,数据表确定模块406还用于:

[0121] 获取多个候选数据表分别对应的数据表描述信息;计算数据查询文本与各个候选数据表分别对应的数据表描述信息之间的文本相似度,得到各个候选数据表分别对应的第一匹配度;比对关键词集与各个候选数据表分别对应的候选字段集,得到各个候选数据表分别对应的第二匹配度;融合同一候选数据表对应的第一匹配度和第二匹配度,分别得到各个候选数据表对应的目标匹配度;基于各个候选数据表分别对应的目标匹配度,确定数据查询文本对应的目标数据表。

[0122] 在一个实施例中,模板构建模块408还用于:

[0123] 基于数据查询文本提取关键词集中的各个关键词分别对应的属性特征;基于各个关键词分别对应的属性特征,从目标字段集包括的各个目标字段中确定数据查询文本对应的返回属性字段和条件属性字段;基于返回属性字段和条件属性字段,生成数据查询文本对应的查询语句模板。

[0124] 在一个实施例中,语句确定模块410还用于:

[0125] 从关键词集中确定数据查询文本对应的条件属性字段所对应的条件关键词;基于数据查询文本和条件关键词,提取条件关键词对应的词特征;基于条件关键词对应的词特征,确定条件关键词对应的目标实体;将目标实体填充至查询语句模板中,得到数据查询文本对应的查询语句。

[0126] 在一个实施例中,语句确定模块410还用于:

[0127] 基于条件关键词对应的词特征,在目标数据表中提取条件关键词对应的多个候选实体;从目标数据表对应的数据表描述信息中,提取各个候选实体分别对应的基础实体特征;基于各个候选实体分别对应的基础实体特征与数据查询文本对应的上下文信息之间的匹配度,在各个候选实体中确定条件关键词对应的目标实体。

[0128] 上述数据查询装置,通过在获取到数据查询文本时,基于数据查询文本对应的关键词集确定数据查询文本所查询的目标数据表,进而根据目标数据表和关键词集构建查询语句模板,将关键词集填充至查询语句模板中,能够快速准确地得到数据查询文本对应的查询语句,有效提高查询语句的生成效率,从而提高数据查询的效率。

[0129] 上述数据查询装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0130] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是服务器,其内部结构图可以如图5所示。该计算机设备包括处理器、存储器、输入/输出接口(Input/Output,简称I/O)和通信接口。其中,处理器、存储器和输入/输出接口通过系统总线连接,通信接口通过输入/输出接口连接到系统总线。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存储器。该非易失性存储介质存储

有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储关键词集、目标数据表等数据。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种数据查询方法。

[0131] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是终端,其内部结构图可以如图6所示。该计算机设备包括处理器、存储器、输入/输出接口、通信接口、显示单元和输入装置。其中,处理器、存储器和输入/输出接口通过系统总线连接,通信接口、显示单元和输入装置通过输入/输出接口连接到系统总线。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端进行有线或无线方式的通信,无线方式可通过WIFI、移动蜂窝网络、NFC(近场通信)或其他技术实现。该计算机程序被处理器执行时以实现一种数据查询方法。该计算机设备的显示单元用于形成视觉可见的画面,可以是显示屏、投影装置或虚拟现实成像装置。显示屏可以是液晶显示屏或者电子墨水显示屏,该计算机设备的输入装置可以是显示屏上覆盖的触摸层,也可以是计算机设备外壳上设置的按键、轨迹球或触控板,还可以是外接的键盘、触控板或鼠标等。

[0132] 本领域技术人员可以理解,图5、6中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0133] 在一个实施例中,提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现上述各方法实施例中的步骤。

[0134] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现上述各方法实施例中的步骤。

[0135] 在一个实施例中,提供了一种计算机程序产品或计算机程序,该计算机产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述各方法实施例中的步骤。

[0136] 需要说明的是,本申请所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等),均为经用户授权或者经过各方充分授权的信息和数据,且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。

[0137] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory,ROM)、磁带、软盘、闪存、光存储器、高密度嵌入式非易失性存储器、阻变存储器

(ReRAM)、磁变存储器(Magnetoresistive Random Access Memory,MRAM)、铁电存储器(Ferroelectric Random Access Memory,FRAM)、相变存储器(Phase Change Memory,PCM)、石墨烯存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器等。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(Static Random Access Memory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等,不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等,不限于此。

[0138] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0139] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对本申请专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请的保护范围应以所附权利要求为准。

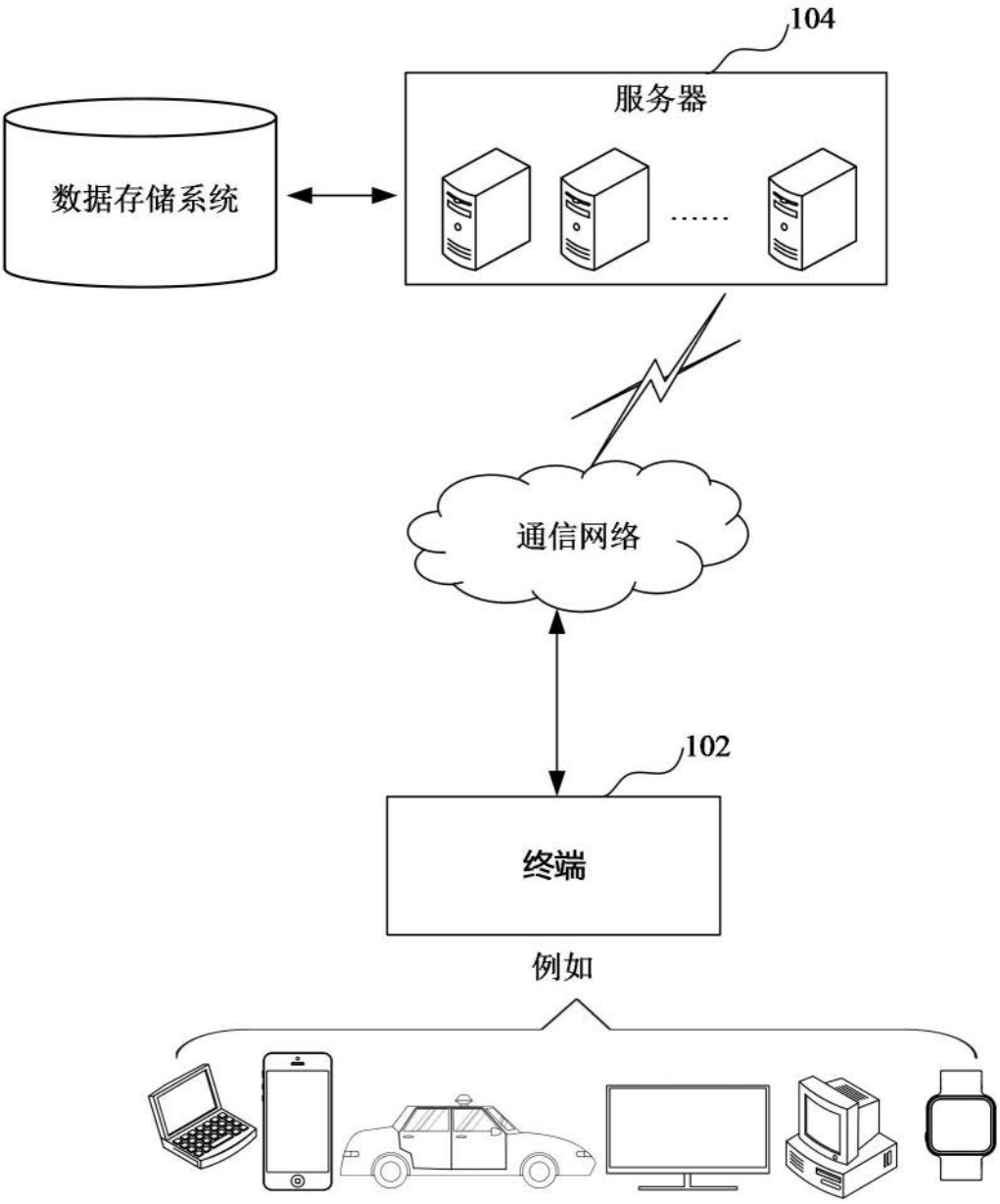


图 1

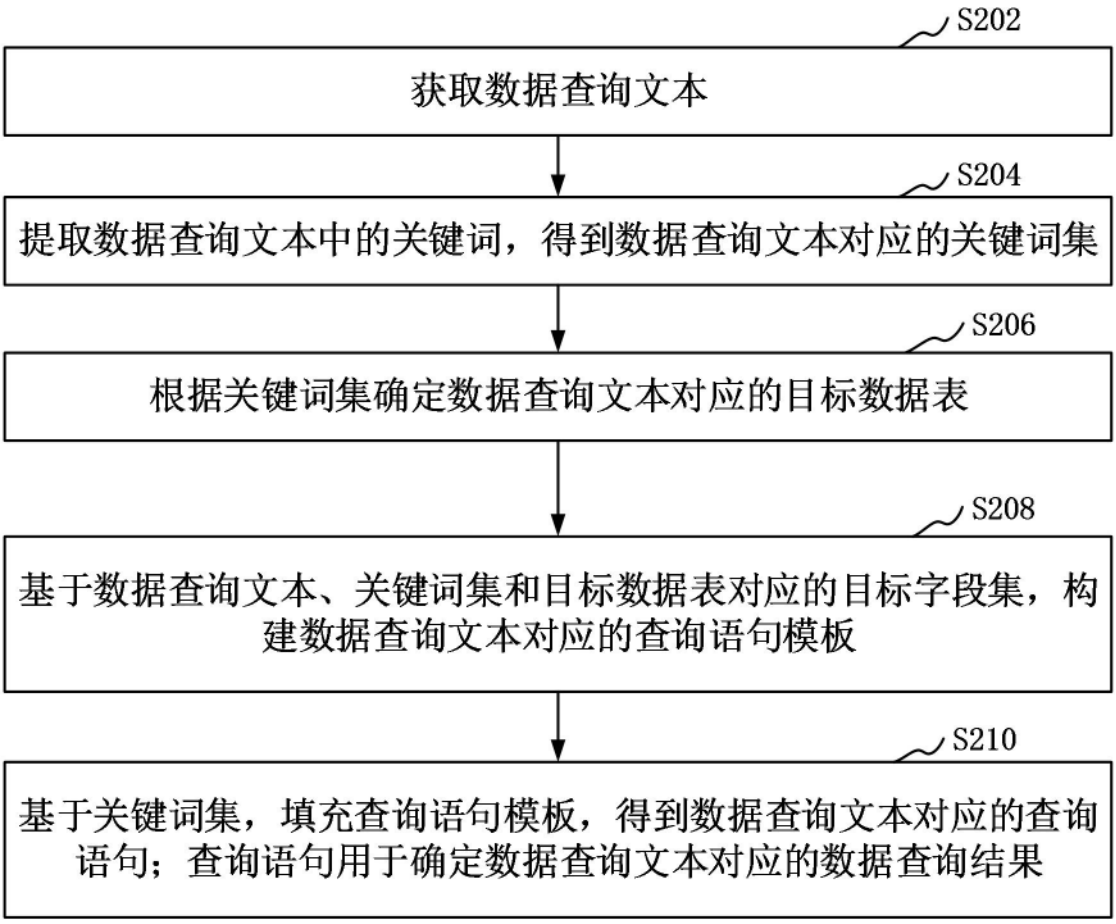

S202
获取数据查询文本
S204
提取数据查询文本中的关键词，得到数据查询文本对应的关键词集
S206
根据关键词集确定数据查询文本对应的目标数据表
S208
基于数据查询文本、关键词集和目标数据表对应的目标字段集，构建数据查询文本对应的查询语句模板
S210
基于关键词集，填充查询语句模板，得到数据查询文本对应的查询语句；查询语句用于确定数据查询文本对应的数据查询结果

图 2

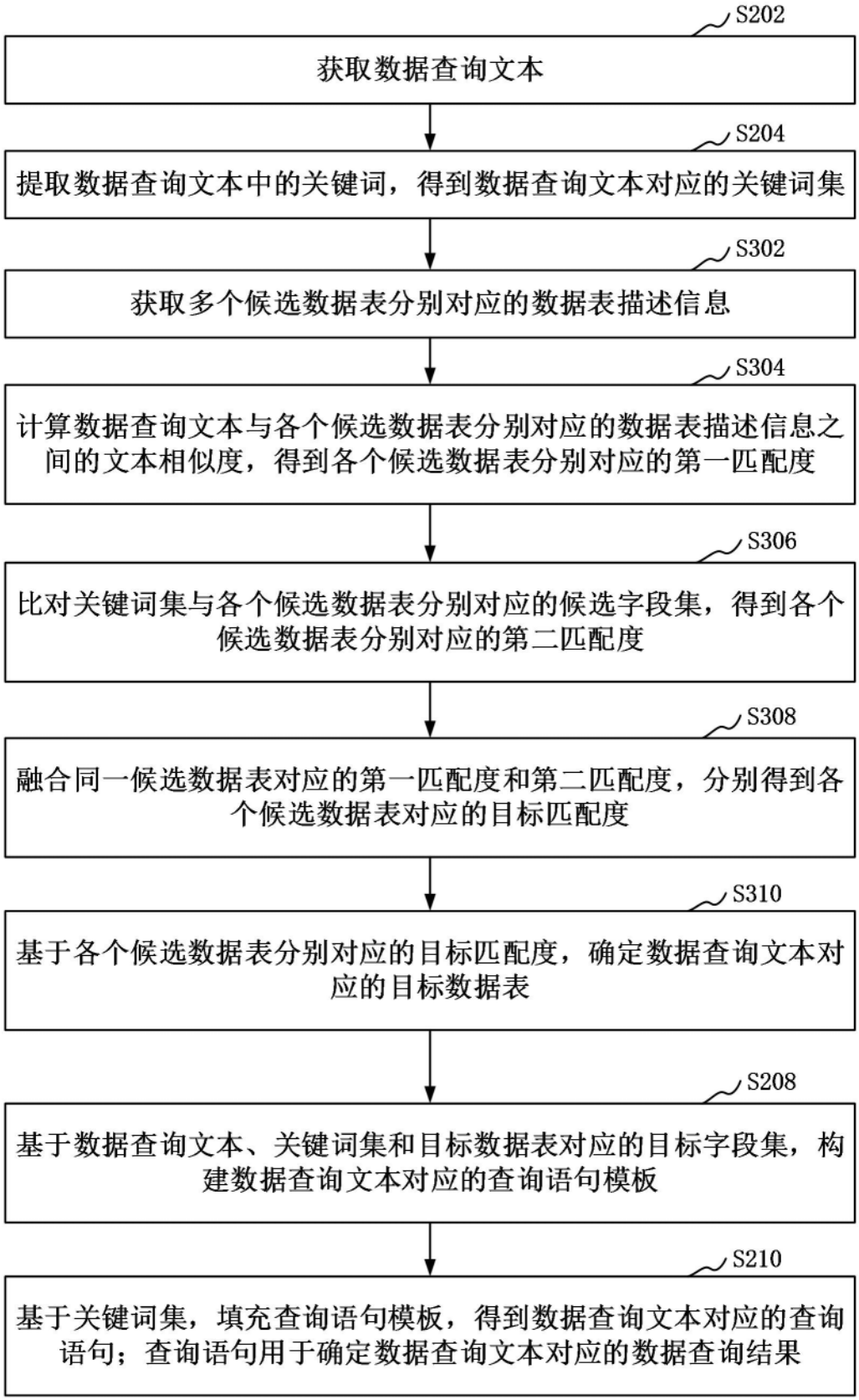

S202
获取数据查询文本
S204
提取数据查询文本中的关键词，得到数据查询文本对应的关键词集
S302
获取多个候选数据表分别对应的数据表描述信息
S304
计算数据查询文本与各个候选数据表分别对应的数据表描述信息之间的文本相似度，得到各个候选数据表分别对应的第一匹配度
S306
比对关键词集与各个候选数据表分别对应的候选字段集，得到各个候选数据表分别对应的第二匹配度
S308
融合同一候选数据表对应的第一匹配度和第二匹配度，分别得到各个候选数据表对应的目标匹配度
S310
基于各个候选数据表分别对应的目标匹配度，确定数据查询文本对应的目标数据表
S208
基于数据查询文本、关键词集和目标数据表对应的目标字段集，构建数据查询文本对应的查询语句模板
S210
基于关键词集，填充查询语句模板，得到数据查询文本对应的查询语句；查询语句用于确定数据查询文本对应的数据查询结果

图 3

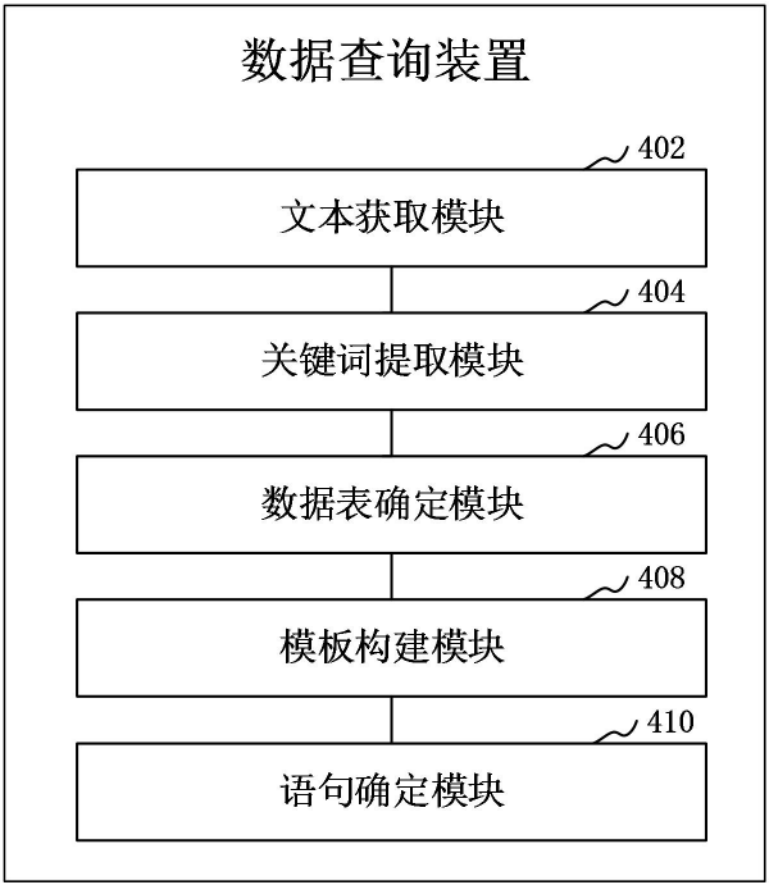

数据查询装置
402
文本获取模块
404
关键词提取模块
406
数据表确定模块
408
模板构建模块
410
语句确定模块

图 4

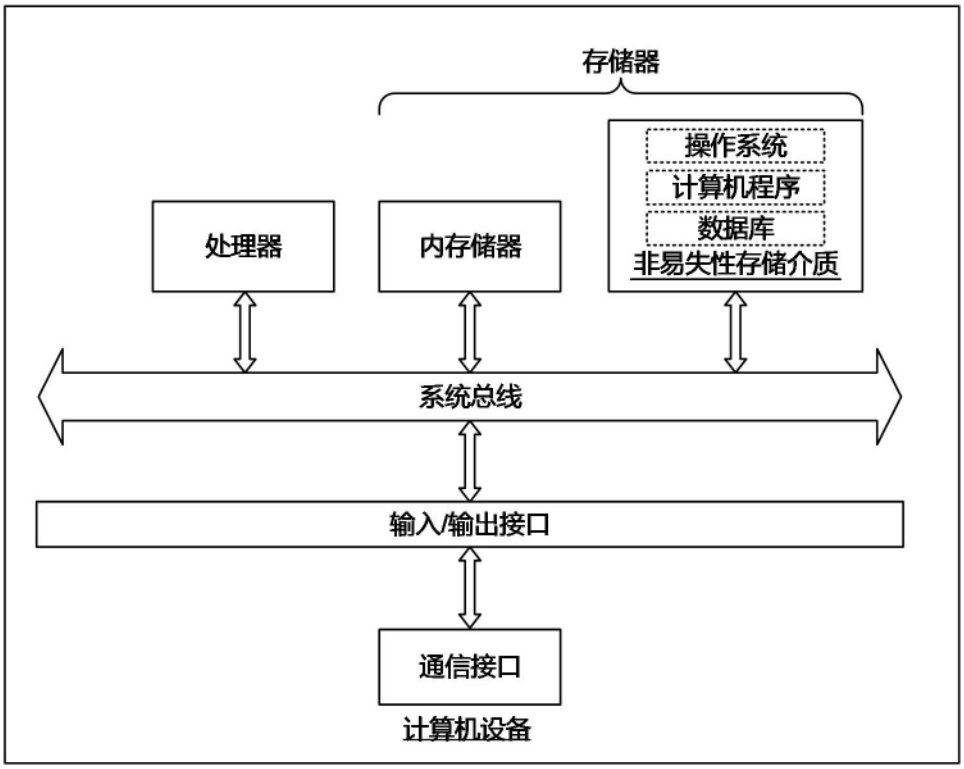

存储器
操作系统
计算机程序
数据库
非易失性存储介质
处理器
内存储器
系统总线
输入/输出接口
通信接口
计算机设备

图 5

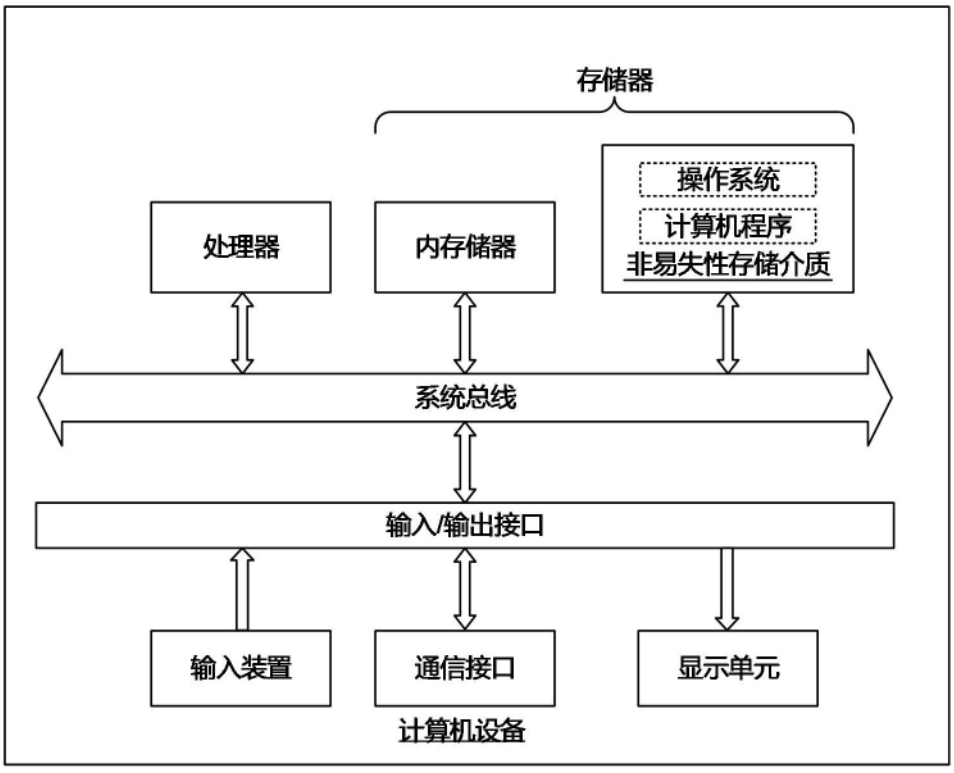

存储器
操作系统
计算机程序
非易失性存储介质
处理器
内存储器
系统总线
输入/输出接口
输入装置
通信接口
显示单元
计算机设备

图 6