(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106033466 A
(43)申请公布日 2016.10.19

(21)申请号 201510123021.7

(22)申请日 2015.03.20

(71)申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 姜南

(74)专利代理机构 北京龙双利达知识产权代理有限公司 11329
代理人 王君 肖鹏

(51)Int.Cl.
G06F 17/30(2006.01)

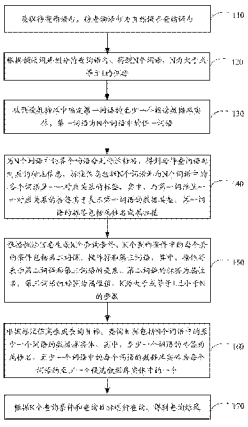
权利要求书5页 说明书21页 附图4页

(54)发明名称

数据库查询的方法和设备

(57)摘要

本发明实施例提供了一种数据库查询的方法和设备，该方法包括：获取待查询语句，待查询语句为自然语言查询语句；根据预设词库划分待查询语句，得到N个词语；从预设数据库中确定第一词语的至少一个候选数据库实体，第一词语为N个词语中的任一词语，为N个词语中的各个词语分别标注标签，得到与待查询语句对应的标注信息；根据标注信息生成K个查询条件，K个查询条件中的每个查询条件包括第二词语、操作符和第三词语；根据标注信息生成查询目标，查询目标包括N个词语中的至少一个词语的数据库实体；根据K个查询条件和查询目标进行查询，得到查询结果。本发明实施例方法能够根据用户请求进行数据库查询，提升用户体验。



CN 106033466 A

1. 一种数据库查询的方法，其特征在于，包括：

获取待查询语句，所述待查询语句为自然语言查询语句；

根据预设词库划分所述待查询语句，得到 N 个词语，N 为大于或等于 1 的整数；

从预设数据库中确定第一词语的至少一个候选数据库实体，所述第一词语为所述 N 个词语中的任一词语；

为所述 N 个词语中的各个词语分别标注标签，得到与所述待查询语句对应的标注信息，所述标注信息包括所述 N 个词语和与所述 N 个词语中的各个词语呈一一对应关系的标签，其中，与所述第一词语呈一一对应关系的标签用于表示所述第一词语的数据类型，所述第一词语的标签包括属性名或属性值；

根据所述标注信息生成 K 个查询条件，所述 K 个查询条件中的每个查询条件包括第二词语、操作符和第三词语，其中，所述操作符表示所述第二词语和所述第三词语的关系，所述第二词语的标签为属性名，所述第三词语的标签为属性值，K 为大于或等于 1 且小于 N 的整数；

根据所述标注信息生成查询目标，所述查询目标包括所述 N 个词语中的至少一个词语的数据库实体，其中，所述至少一个词语的标签为属性名，所述至少一个词语中的每个词语的数据库实体为所述每个词语的至少一个候选数据库实体中的一个；

根据所述 K 个查询条件和所述查询目标进行查询，得到查询结果。

2. 根据权利要求 1 所述的方法，其特征在于，所述根据预设词库划分所述待查询语句，得到 N 个词语，包括：

根据预设词库划分所述待查询语句，得到 N 个初始词语；

根据预设规则，规范化所述 N 个初始词语，得到所述 N 个词语。

3. 根据权利要求 1 或 2 所述的方法，其特征在于，所述从预设数据库中确定第一词语的至少一个候选数据库实体，包括：

从预设数据库中确定所述第一词语的 n 个初始候选数据库实体，n 为大于或等于 1 的整数；

当 n 大于 1 时，确定所述 n 个初始候选数据库实体中每个初始候选数据库实体与所述第一词语的相关度，将所述 n 个初始候选数据库实体中相关度高于预设阈值的初始候选数据库实体确定为所述第一词语的至少一个候选数据库实体，

或者，当 n 等于 1 时，将所述第一词语的 n 个初始候选数据库实体确定为所述第一词语的至少一个候选数据库实体。

4. 根据权利要求 3 所述的方法，其特征在于，所述确定所述 n 个初始候选数据库实体中每个初始候选数据库实体与所述第一词语的相关度，包括：

根据以下方法中的至少一种方法确定所述 n 个初始候选数据库实体中每个初始候选数据库实体与所述第一词语的相关度：

命中率、向量空间余弦和编辑距离。

5. 根据权利要求 1 至 4 中任一项所述的方法，其特征在于，在根据所述标注信息生成 K 个查询条件之前，还包括：

根据所述标注信息中的词语的候选数据库实体，合并所述标注信息中连续标签为属性名的词语，得到第一合并词语，所述第一合并词语为所述标注信息中连续标签为属性名的

词语的候选数据库实体的交集，使用所述第一合并词语替换所述标注信息中所述连续标签为属性名的词语，以对所述标注信息进行更新，

和 / 或

根据所述标注信息中的词语的候选数据库实体，合并所述标注信息中连续标签为属性值的词语，得到第二合并词语，所述第二合并词语为所述标注信息中连续标签为属性值的词语的候选数据库实体的交集，使用所述第二合并词语替换所述标注信息中所述连续标签为属性值的词语，以对所述标注信息进行更新，

其中，所述根据所述标注信息生成 K 个查询条件，包括根据更新后的标注信息生成所述 K 个查询条件，

所述根据所述标注信息生成查询目标，包括根据更新后的标注信息生成所述查询目标。

6. 根据权利要求 1 至 5 中任一项所述的方法，其特征在于，所述根据所述标注信息生成 K 个查询条件，包括 ：

根据所述标注信息生成 M 个候选查询条件，所述 M 个候选查询条件中的每个候选查询条件包括第一候选词语、操作符和第二候选词语的对应关系，其中第一候选词语的标签为属性名，第二候选词语的标签为属性值，M 为大于或等于 K 的整数 ；

确定所述每个候选查询条件的第一候选词语和所述第二候选词语的匹配指数 ；

将所述 M 个候选查询条件中的匹配指数大于预设阈值的 K 个候选查询条件确定为所述 K 个查询条件。

7. 根据权利要求 6 所述的方法，其特征在于，所述根据所述标注信息生成 M 个候选查询条件，包括 ：

根据所述标注信息生成 M 个初始候选查询条件 ；

根据用户信息，对所述 M 个初始候选查询条件进行消歧处理，得到所述 M 个候选查询条件，所述消歧处理包括根据用户信息消除所述 M 个初始候选查询条件中存在歧义的初始候选查询条件中的歧义，其中，所述用户信息包括终端设备的硬件信息、终端系统的软件信息、保存在终端内存或者存储设备上的用户数据、用户的历史操作和用户的设定中的至少一种。

8. 根据权利要求 6 或 7 所述的方法，其特征在于，所述确定所述每个候选查询条件的第一候选词语和所述第二候选词语的匹配指数，包括 ：

根据所述第一候选词语和所述第二候选词语的配对概率、序列距离、数据库数据类型匹配度和语言习惯约束中的至少一种确定所述匹配指数。

9. 根据权利要求 8 所述的方法，其特征在于，所述配对概率由所述第一候选词语所对应的数据库实体与所述第二候选词语所对应的数据库实体之间的交集决定，其中，所述第一候选词语所对应的数据库实体与所述第二候选词语所对应的数据库实体之间的交集越少，所述配对概率越大，所述匹配指数越大。

10. 根据权利要求 8 或 9 所述的方法，其特征在于，所述序列距离由所述第一候选词语和所述第二候选词语在所述标注信息或所述查询语句中的距离决定，其中，所述第一候选词语和所述第二候选词语在所述标注信息或所述查询语句中的距离越大，所述序列距离越大，所述匹配指数越小，所述标注信息或所述查询语句中所述第一候选词语和所述第二候

选词语之间的词语的多少，表示所述距离的大小。

11. 根据权利要求 8 至 10 中任一项所述的方法，其特征在于，所述数据库数据类型匹配度由所述第一候选词语和所述第二候选词语的数据库数据类型是否一致决定，其中，所述第一候选词语和所述第二候选词语的数据类型一致时的数据库数据类型匹配度大于所述第一候选词语和所述第二候选词语的数据类型不一致时的数据库数据类型匹配度，所述匹配指数与所述数据库类型匹配度正相关。

12. 根据权利要求 8 至 11 中任一项所述的方法，其特征在于，所述语言习惯约束由所述第一候选词语和所述第二候选词语是否符合数据库或语言习惯决定，其中，所述第一候选词语和所述第二候选词语符合数据库或语言习惯时的语言习惯约束小于所述第一候选词语和所述第二候选词语不符合数据库或语言习惯时的语言习惯约束，所述匹配指数与所述语言习惯约束负相关。

13. 根据权利要求 1 至 12 中任一项所述的方法，其特征在于，所述根据所述标注信息生成查询目标，包括：

确定所述标注信息中的标签为属性名的词语满足预设条件和 / 或为孤点词语，其中所述孤点词语没有对应的标签为属性值的词语；

将所述标注信息中的标签为属性名的词语的属性名作为所述查询目标。

14. 一种数据库查询的设备，其特征在于，包括：

获取单元，用于获取待查询语句，所述待查询语句为自然语言查询语句；

划分单元，用于根据预设词库划分所述待查询语句，得到 N 个词语，N 为大于或等于 1 的整数；

确定单元，用于从预设数据库中确定第一词语的至少一个候选数据库实体，所述第一词语为所述 N 个词语中的任一词语；

标注单元，用于为所述 N 个词语中的各个词语分别标注标签，得到与所述待查询语句对应的标注信息，所述标注信息包括所述 N 个词语和与所述 N 个词语中的各个词语呈一一对应关系的标签，其中，与所述第一词语呈一一对应关系的标签用于表示所述第一词语的数据类型，所述第一词语的标签包括属性名或属性值；

第一生成单元，用于根据所述标注信息生成 K 个查询条件所述 K 个查询条件中的每个查询条件包括第二词语、操作符和第三词语，其中，所述操作符表示所述第二词语和所述第三词语的关系，所述第二词语的标签为属性名，所述第三词语的标签为属性值，K 为大于或等于 1 且小于 N 的整数；第二生成单元，用于根据所述标注信息生成查询目标，所述查询目标包括所述 N 个词语中的至少一个词语的数据库实体，其中，所述至少一个词语的标签为属性名，所述至少一个词语中的每个词语的数据库实体为所述每个词语的至少一个候选数据库实体中的一个；

查询单元，用于根据所述 K 个查询条件和所述查询目标进行查询，得到查询结果。

15. 根据权利要求 14 所述的设备，其特征在于，所述划分单元根据预设词库划分所述待查询语句，得到 N 个初始词语；根据预设规则，规范化所述 N 个初始词语，得到所述 N 个词语。

16. 根据权利要求 14 或 15 所述的设备，其特征在于，所述确定单元从预设数据库中确定所述第一词语的 n 个初始候选数据库实体，n 为大于或等于 1 的整数；当 n 大于 1 时，确

定所述 n 个初始候选数据库实体中每个初始候选数据库实体与所述第一词语的相关度，将所述 n 个初始候选数据库实体中相关度高于预设阈值的初始候选数据库实体确定为所述第一词语的至少一个候选数据库实体，或者，当 n 等于 1 时，将所述第一词语的 n 个初始候选数据库实体确定为所述第一词语的至少一个候选数据库实体。

17. 根据权利要求 16 所述的设备，其特征在于，所述确定单元根据以下方法中的至少一种方法确定所述 n 个初始候选数据库实体中每个初始候选数据库实体与所述第一词语的相关度：

命中率、向量空间余弦和编辑距离。

18. 根据权利要求 14 至 17 中任一项所述的设备，其特征在于，还包括：合并单元，用于在第一生成单元根据所述标注信息生成 K 个查询条件之前，根据所述标注信息中的词语的候选数据库实体，合并所述标注信息中连续标签为属性名的词语，得到第一合并词语，所述第一合并词语为所述标注信息中连续标签为属性名的词语的候选数据库实体的交集，使用所述第一合并词语替换所述标注信息中所述连续标签为属性名的词语，以对所述标注信息进行更新，和 / 或根据所述标注信息中的词语的候选数据库实体，合并所述标注信息中连续标签为属性值的词语，得到第二合并词语，所述第二合并词语为所述标注信息中连续标签为属性值的词语的候选数据库实体的交集，使用所述第二合并词语替换所述标注信息中所述连续标签为属性值的词语，以对所述标注信息进行更新，

其中，所述第一生成单元根据更新后的标注信息生成所述 K 个查询条件，所述第二生成单元根据更新后的标注信息生成所述查询目标。

19. 根据权利要求 14 至 18 中任一项所述的设备，其特征在于，所述第一生成单元根据所述标注信息生成 M 个候选查询条件，所述 M 个候选查询条件中的每个候选查询条件包括第一候选词语、操作符和第二候选词语的对应关系，其中第一候选词语的标签为属性名，第二候选词语的标签为属性值，M 为大于或等于 K 的整数；确定所述每个候选查询条件的第一候选词语和所述第二候选词语的匹配指数；将所述 M 个候选查询条件中的匹配指数大于预设阈值的 K 个候选查询条件确定为所述 K 个查询条件。

20. 根据权利要求 19 所述的设备，其特征在于，所述第一生成单元，根据所述标注信息生成 M 个初始候选查询条件；根据用户信息，对所述 M 个初始候选查询条件进行消歧处理，得到所述 M 个候选查询条件，所述消歧处理包括根据用户信息消除所述 M 个初始候选查询条件中存在歧义的初始候选查询条件中的歧义，其中，所述用户信息包括终端设备的硬件信息、终端系统的软件信息、保存在终端内存或者存储设备上的用户数据、用户的历史操作和用户的设定中的至少一种。

21. 根据权利要求 19 或 20 所述的设备，其特征在于，所述第一生成单元根据所述第一候选词语和所述第二候选词语的配对概率、序列距离、数据库数据类型匹配度和语言习惯约束中的至少一种确定所述匹配指数。

22. 根据权利要求 21 所述的设备，其特征在于，所述配对概率由所述第一候选词语所对应的数据库实体与所述第二候选词语所对应的数据库实体之间的交集决定，其中，所述第一候选词语所对应的数据库实体与所述第二候选词语所对应的数据库实体之间的交集越少，所述配对概率越大，所述匹配指数越大。

23. 根据权利要求 21 或 22 所述的设备，其特征在于，所述序列距离由所述第一候选词

语和所述第二候选词语在所述标注信息或所述查询语句中的距离决定，其中，所述第一候选词语和所述第二候选词语在所述标注信息或所述查询语句中的距离越大，所述序列距离越大，所述匹配指数越小，所述标注信息或所述查询语句中所述第一候选词语和所述第二候选词语之间的词语的多少，表示所述距离的大小。

24. 根据权利要求 21 至 23 中任一项所述的设备，其特征在于，所述数据库数据类型匹配度由所述第一候选词语和所述第二候选词语的数据库数据类型是否一致决定，其中，所述第一候选词语和所述第二候选词语的数据类型一致时的数据库数据类型匹配度大于所述第一候选词语和所述第二候选词语的数据类型不一致时的数据库数据类型匹配度，所述匹配指数与所述数据库类型匹配度正相关。

25. 根据权利要求 21 至 24 中任一项所述的设备，其特征在于，所述语言习惯约束由所述第一候选词语和所述第二候选词语是否符合数据库或语言习惯决定，其中，所述第一候选词语和所述第二候选词语符合数据库或语言习惯时的语言习惯约束小于所述第一候选词语和所述第二候选词语不符合数据库或语言习惯时的语言习惯约束，所述匹配指数与所述语言习惯约束负相关。

26. 根据权利要求 14 至 25 中任一项所述的设备，其特征在于，所述第二生成单元确定所述标注信息中的标签为属性名的词语满足预设条件和 / 或为孤点词语，其中，所述孤点词语没有对应的标签为属性值的词语；将所述标注信息中的标签为属性名的词语的属性名作为所述查询目标。

数据库查询的方法和设备

技术领域

[0001] 本发明涉及通信领域，特别涉及一种数据库查询的方法和设备。

背景技术

[0002] 对于传统的数据库查询，当前仍然需要专业人员深入理解数据库内部的结构信息，并且构建适当的结构化查询语言(Structured Query Language，SQL)查询语句，对于非专业人员来说，如果不具备数据库的专业知识，对于数据库操作将比较困难。而随着互联网搜索引擎技术的不断发展，人们逐渐习惯了在搜索框中输入自然语言搜索结果，同样希望通过自然语言查询数据库。

[0003] 由于普通用户不了解数据库中的结构、数据库字段名／值，同时在描述查询请求的时候会省略上下文信息，因此现有技术存在诸多问题，例如，用户请求中的描述无法完全和数据库字段名／值一一对应，而对于SQL，如果描述的请求与数据库字段名／值对应不上可能查询不到结果；用户请求中可能包含歧义的信息，即用户查询语句中包含的一个或者多个词语可能包含不只一种数据库对象（表、字段），导致无法得到查询结果，用户体验差。

[0004] 因此，希望提供一种技术，能够根据用户请求进行数据库查询。

发明内容

[0005] 本发明实施例提供了一种数据库查询的方法和设备，该方法能够根据用户请求进行数据库查询，提升用户体验。

[0006] 第一方面，提供了一种数据库查询的方法，包括：获取待查询语句，该待查询语句为自然语言查询语句；根据预设词库划分该待查询语句，得到N个词语，N为大于或等于1的整数；从预设数据库中确定第一词语的至少一个候选数据库实体，该第一词语为该N个词语中的任一词语；为该N个词语中的各个词语分别标注标签，得到与该待查询语句对应的标注信息，该标注信息包括该N个词语和与该N个词语中的各个词语呈一一对应关系的标签，其中，与该第一词语呈一一对应关系的标签用于表示该第一词语的数据类型，该第一词语的标签包括属性名或属性值；根据该标注信息生成K个查询条件，该K个查询条件中的每个查询条件包括第二词语、操作符和第三词语，其中，该操作符表示该第二词语和该第三词语的关系，该第二词语的标签为属性名，该第三词语的标签为属性值，K为大于或等于1且小于N的整数；根据该标注信息生成查询目标，该查询目标包括该N个词语中的至少一个词语的数据库实体，其中，该至少一个词语的标签为属性名，该至少一个词语中的每个词语的数据库实体为该每个词语的至少一个候选数据库实体中的一个；根据该K个查询条件和该查询目标进行查询，得到查询结果。

[0007] 结合第一方面，在第一种可能的实现方式中，该根据预设词库划分该待查询语句，得到N个词语，包括：根据预设词库划分该待查询语句，得到N个初始词语；根据预设规则，规范化该N个初始词语，得到该N个词语。

[0008] 结合第一方面或第一种可能的实现方式，在第二种可能的实现方式中，该从预设

数据库中确定第一词语的至少一个候选数据库实体，包括：从预设数据库中确定该第一词语的 n 个初始候选数据库实体，n 为大于或等于 1 的整数；当 n 大于 1 时，确定该 n 个初始候选数据库实体中每个初始候选数据库实体与该第一词语的相关度，将该 n 个初始候选数据库实体中相关度高于预设阈值的初始候选数据库实体确定为该第一词语的至少一个候选数据库实体，或者，当 n 等于 1 时，将该第一词语的 n 个初始候选数据库实体确定为该第一词语的至少一个候选数据库实体。

[0009] 结合第二种可能的实现方式，在第三种可能的实现方式中，该确定该 n 个初始候选数据库实体中每个初始候选数据库实体与该第一词语的相关度，包括：根据以下方法中的至少一种方法确定该 n 个初始候选数据库实体中每个初始候选数据库实体与该第一词语的相关度：命中率、向量空间余弦和编辑距离。

[0010] 结合第一方面、第一至第三种可能的实现方式中的任一种可能的实现方式，在第四种可能的实现方式中，在根据该标注信息生成 K 个查询条件之前，还包括：根据该标注信息中的词语的候选数据库实体，合并该标注信息中连续标签为属性名的词语，得到第一合并词语，该第一合并词语为该标注信息中连续标签为属性名的词语的候选数据库实体的交集，使用该第一合并词语替换该标注信息中该连续标签为属性名的词语，以对该标注信息进行更新，和 / 或根据该标注信息中的词语的候选数据库实体，合并该标注信息中连续标签为属性值的词语，得到第二合并词语，该第二合并词语为该标注信息中连续标签为属性值的词语的候选数据库实体的交集，使用该第二合并词语替换该标注信息中该连续标签为属性值的词语，以对该标注信息进行更新，其中，该根据该标注信息生成 K 个查询条件，包括根据更新后的标注信息生成该 K 个查询条件，该根据该标注信息生成查询目标，包括根据更新后的标注信息生成该查询目标。

[0011] 结合第一方面、第一至第四种可能的实现方式中的任一种可能的实现方式，在第五种可能的实现方式中，该根据该标注信息生成 K 个查询条件，包括：根据该标注信息生成 M 个候选查询条件，该 M 个候选查询条件中的每个候选查询条件包括第一候选词语、操作符和第二候选词语的对应关系，其中第一候选词语的标签为属性名，第二候选词语的标签为属性值，M 为大于或等于 K 的整数；确定该每个候选查询条件的第一候选词语和该第二候选词语的匹配指数；将该 M 个候选查询条件中的匹配指数大于预设阈值的 K 个候选查询条件确定为该 K 个查询条件。

[0012] 结合第五种可能的实现方式，在第六种可能的实现方式中，该根据该标注信息生成 M 个候选查询条件，包括：根据该标注信息生成 M 个初始候选查询条件；根据用户信息，对该 M 个初始候选查询条件进行消歧处理，得到该 M 个候选查询条件，该消歧处理包括根据用户信息消除该 M 个初始候选查询条件中存在歧义的初始候选查询条件中的歧义，其中，该用户信息包括终端设备的硬件信息、终端系统的软件信息、保存在终端内存或者存储设备上的用户数据、用户的历史操作和用户的设定中的至少一种。

[0013] 结合第五种或第六种可能的实现方式，在第七种可能的实现方式中，该确定该每个候选查询条件的第一候选词语和该第二候选词语的匹配指数，包括：根据该第一候选词语和该第二候选词语的配对概率、序列距离、数据库数据类型匹配度和语言习惯约束中的至少一种确定该匹配指数。

[0014] 结合第七种可能的实现方式，在第八种可能的实现方式中，该配对概率由该第一

候选词语所对应的数据库实体与该第二候选词语所对应的数据库实体之间的交集决定，其中，该第一候选词语所对应的数据库实体与该第二候选词语所对应的数据库实体之间的交集越少，该配对概率越大，该匹配指数越大。

[0015] 结合第七种或第八种可能的实现方式，在第九种可能的实现方式中，该序列距离由该第一候选词语和该第二候选词语在该标注信息或该查询语句中的距离决定，其中，该第一候选词语和该第二候选词语在该标注信息或该查询语句中的距离越大，该序列距离越大，该匹配指数越小，该标注信息或该查询语句中该第一候选词语和该第二候选词语之间的词语的多少，表示该距离的大小。

[0016] 结合第七至第九种可能的实现方式中的任一种可能的实现方式，在第十种可能的实现方式中，该数据库数据类型匹配度由该第一候选词语和该第二候选词语的数据库数据类型是否一致决定，其中，该第一候选词语和该第二候选词语的数据类型一致时的数据库数据类型匹配度大于该第一候选词语和该第二候选词语的数据类型不一致时的数据库数据类型匹配度，该匹配指数与该数据库类型匹配度正相关。

[0017] 结合第七至第十种可能的实现方式中的任一种可能的实现方式，在第十一种可能的实现方式中，该语言习惯约束由该第一候选词语和该第二候选词语是否符合数据库或语言习惯决定，其中，该第一候选词语和该第二候选词语符合数据库或语言习惯时的语言习惯约束小于该第一候选词语和该第二候选词语不符合数据库或语言习惯时的语言习惯约束，该匹配指数与该语言习惯约束负相关。

[0018] 结合第一方面、第一至第十一种可能的实现方式中的任一种可能的实现方式，在第十二种可能的实现方式中，该根据该标注信息生成查询目标，包括：确定该标注信息中的标签为属性名的词语满足预设条件和/或为孤点词语，其中该孤点词语没有对应的标签为属性值的词语；将该标注信息中的标签为属性名的词语的属性名作为该查询目标。

[0019] 第二方面，提供了一种数据库查询的设备，包括：获取单元，用于获取待查询语句，该待查询语句为自然语言查询语句；划分单元，用于根据预设词库划分该待查询语句，得到N个词语，N为大于或等于1的整数；确定单元，用于从预设数据库中确定第一词语的至少一个候选数据库实体，该第一词语为该N个词语中的任一词语；标注单元，用于为该N个词语中的各个词语分别标注标签，得到与该待查询语句对应的标注信息，该标注信息包括该N个词语和与该N个词语中的各个词语呈一一对应关系的标签，其中，与该第一词语呈一一对应关系的标签用于表示该第一词语的数据类型，该第一词语的标签包括属性名或属性值；第一生成单元，用于根据该标注信息生成K个查询条件，该K个查询条件中的每个查询条件包括第二词语、操作符和第三词语，其中，该操作符表示该第二词语和该第三词语的关系，该第二词语的标签为属性名，该第三词语的标签为属性值，K为大于或等于1且小于N的整数；第二生成单元，用于根据该标注信息生成查询目标，该查询目标包括该N个词语中的至少一个词语的数据库实体，其中，该至少一个词语的标签为属性名，该至少一个词语中的每个词语的数据库实体为该每个词语的至少一个候选数据库实体中的一个；查询单元，用于根据该K个查询条件和该查询目标进行查询，得到查询结果。

[0020] 结合第二方面，在第一种可能的实现方式中，该划分单元根据预设词库划分该待查询语句，得到N个初始词语；根据预设规则，规范化该N个初始词语，得到该N个词语。

[0021] 结合第二方面或第二方面的第一种可能的实现方式，在第二种可能的实现方式

中，该确定单元从预设数据库中确定该第一词语的 n 个初始候选数据库实体，n 为大于或等于 1 的整数；当 n 大于 1 时，确定该 n 个初始候选数据库实体中每个初始候选数据库实体与该第一词语的相关度，将该 n 个初始候选数据库实体中相关度高于预设阈值的初始候选数据库实体确定为该第一词语的至少一个候选数据库实体，或者，当 n 等于 1 时，将该第一词语的 n 个初始候选数据库实体确定为该第一词语的至少一个候选数据库实体。

[0022] 结合第二方面的第二种可能的实现方式，在第三种可能的实现方式中，该确定单元根据以下方法中的至少一种方法确定该 n 个初始候选数据库实体中每个初始候选数据库实体与该第一词语的相关度：命中率、向量空间余弦和编辑距离。

[0023] 结合第二方面、第二方面的第一至第三种可能的实现方式中的任一种可能的实现方式，在第四种可能的实现方式中，还包括：合并单元，用于在第一生成单元根据该标注信息生成 K 个查询条件之前，根据该标注信息中的词语的候选数据库实体，合并该标注信息中连续标签为属性名的词语，得到第一合并词语，该第一合并词语为该标注信息中连续标签为属性名的词语的候选数据库实体的交集，使用该第一合并词语替换该标注信息中该连续标签为属性名的词语，以对该标注信息进行更新，和 / 或根据该标注信息中的词语的候选数据库实体，合并该标注信息中连续标签为属性值的词语，得到第二合并词语，该第二合并词语为该标注信息中连续标签为属性值的词语的候选数据库实体的交集，使用该第二合并词语替换该标注信息中该连续标签为属性值的词语，以对该标注信息进行更新，其中，该第一生成单元根据更新后的标注信息生成该 K 个查询条件，该第二生成单元根据更新后的标注信息生成该查询目标。

[0024] 结合第二方面、第二方面的第一至第四种可能的实现方式中的任一种可能的实现方式，在第五种可能的实现方式中，该第一生成单元根据该标注信息生成 M 个候选查询条件，该 M 个候选查询条件中的每个候选查询条件包括第一候选词语、操作符和第二候选词语的对应关系，其中第一候选词语的标签为属性名，第二候选词语的标签为属性值，M 为大于或等于 K 的整数；确定该每个候选查询条件的第一候选词语和该第二候选词语的匹配指数；将该 M 个候选查询条件中的匹配指数大于预设阈值的 K 个候选查询条件确定为该 K 个查询条件。

[0025] 结合第二方面的第五种可能的实现方式，在第六种可能的实现方式中，该第一生成单元，根据该标注信息生成 M 个初始候选查询条件；根据用户信息，对该 M 个初始候选查询条件进行消歧处理，得到该 M 个候选查询条件，该消歧处理包括根据用户信息消除该 M 个初始候选查询条件中存在歧义的初始候选查询条件中的歧义，其中，该用户信息包括终端设备的硬件信息、终端系统的软件信息、保存在终端内存或者存储设备上的用户数据、用户的历史操作和用户的设定中的至少一种。

[0026] 结合第二方面的第五种或第六种可能的实现方式，在第七种可能的实现方式中，该第一生成单元根据该第一候选词语和该第二候选词语的配对概率、序列距离、数据库数据类型匹配度和语言习惯约束中的至少一种确定该匹配指数。

[0027] 结合第二方面的第七种可能的实现方式，在第八种可能的实现方式中，该配对概率由该第一候选词语所对应的数据库实体与该第二候选词语所对应的数据库实体之间的交集决定，其中，该第一候选词语所对应的数据库实体与该第二候选词语所对应的数据库实体之间的交集越少，该配对概率越大，该匹配指数越大。

[0028] 结合第二方面的第七种或第八种可能的实现方式，在第九种可能的实现方式中，该序列距离由该第一候选词语和该第二候选词语在该标注信息或该查询语句中的距离决定，其中，该第一候选词语和该第二候选词语在该标注信息或该查询语句中的距离越大，该序列距离越大，该匹配指数越小，该标注信息或该查询语句中该第一候选词语和该第二候选词语之间的词语的多少，表示该距离的大小。

[0029] 结合第二方面的第七至第九种可能的实现方式中的任一种可能的实现方式，在第十种可能的实现方式中，该数据库数据类型匹配度由该第一候选词语和该第二候选词语的数据库数据类型是否一致决定，其中，该第一候选词语和该第二候选词语的数据类型一致时的数据库数据类型匹配度大于该第一候选词语和该第二候选词语的数据类型不一致时的数据库数据类型匹配度，该匹配指数与该数据库类型匹配度正相关。

[0030] 结合第二方面的第七至第十种可能的实现方式中的任一种可能的实现方式，在第十一种可能的实现方式中，该语言习惯约束由该第一候选词语和该第二候选词语是否符合数据库或语言习惯决定，其中，该第一候选词语和该第二候选词语符合数据库或语言习惯时的语言习惯约束小于该第一候选词语和该第二候选词语不符合数据库或语言习惯时的语言习惯约束，该匹配指数与该语言习惯约束负相关。

[0031] 结合第二方面、第一至第十一种可能的实现方式中的任一种可能的实现方式，在第十二种可能的实现方式中，该第二生成单元确定该标注信息中的标签为属性名的词语满足预设条件和 / 或为孤点词语，其中，该孤点词语没有对应的标签为属性值的词语 ；将该标注信息中的标签为属性名的词语的属性名作为该查询目标。

[0032] 基于上述技术方案，本发明实施例通过将为自然语言查询语句的待查询语句生成查询目标和查询条件，根据查询目标和查询条件进行查询，进而得到查询结果，能够根据用户请求进行数据库查询。本发明实施例无需用户熟悉数据库查询语言，提升用户体验。

附图说明

[0033] 为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例中所需要使用的附图作简单地介绍，显而易见地，下面所描述的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0034] 图 1 是根据本发明一个实施例的数据库查询的方法的示意性流程图。

[0035] 图 2 是根据本发明另一实施例的数据库查询的方法的示意性流程图。

[0036] 图 3 是根据本发明一个实施例的数据库查询的设备的示意框图。

[0037] 图 4 是根据本发明另一实施例的数据库查询的设备的示意框图。

具体实施方式

[0038] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明的一部分实施例，而不是全部实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都应属于本发明保护的范围。

[0039] 应理解，在本发明实施例中，用户设备 (UE,User Equipment) 包括但不限于移动台

(MS，Mobile Station)、移动终端 (Mobile Terminal)、移动电话 (Mobile Telephone)、手机 (handset) 及便携设备 (portable equipment) 等，该用户设备可以经无线接入网 (RAN，Radio Access Network) 与一个或多个核心网进行通信，例如，用户设备可以是移动电话（或称为“蜂窝”电话）、具有无线通信功能的计算机等，用户设备还可以是计算机、Pad、便携式、袖珍式、手持式、计算机内置的或者车载的移动装置。

[0040] 图 1 是根据本发明一个实施例的数据库查询的方法的示意性流程图。图 1 所示的方法可以由数据库查询的设备执行，具体地，如图 1 所示的方法包括：

[0041] 110，获取待查询语句，待查询语句为自然语言查询语句。

[0042] 120，根据预设词库划分待查询语句，得到 N 个词语，N 为大于或等于 1 的整数。

[0043] 130，从预设数据库中确定第一词语的至少一个候选数据库实体，第一词语为 N 个词语中的任一词语。

[0044] 140 为 N 个词语中的各个词语分别标注标签，得到与待查询语句对应的标注信息，标注信息包括 N 个词语和与 N 个词语中的各个词语呈一一对应关系的标签，其中，与第一词语呈一一对应关系的标签用于表示第一词语的数据类型，第一词语的标签包括属性名或属性值。

[0045] 150，根据标注信息生成 K 个查询条件，K 个查询条件中的每个查询条件包括第二词语、操作符和第三词语，其中，操作符表示第二词语和第三词语的关系，第二词语的标签为属性名，第三词语的标签为属性值，K 为大于或等于 1 且小于 N 的整数。

[0046] 160，根据标注信息生成查询目标，查询目标包括 N 个词语中的至少一个词语的数据库实体，其中，至少一个词语的标签为属性名，至少一个词语中的每个词语的数据库实体为每个词语的至少一个候选数据库实体中的一个。

[0047] 170，根据 K 个查询条件和查询目标进行查询，得到查询结果。

[0048] 因此，本发明实施例通过根据将为自然语言查询语句的待查询语句生成查询目标和查询条件，根据查询目标和查询条件进行查询，进而得到查询结果，能够根据用户请求进行数据库查询。本发明实施例无需用户熟悉数据库查询语言，提升用户体验。

[0049] 应理解，N 个词语可以为待查询语句中的 Y 个词语中的 N 个具有实际意义的词语，例如，对于查询语句“大于 30 岁的人数”，可以划分为 Y ＝ 4 个词语“大于”“30 岁”“的”和“人数”，其中，N 个词语为 4 个词语中的 2 个词语，即 N ＝ 2，该 2 个词语为“30 岁”和“人数”。换句话说，N 个词语中的每一个词语都存在候选数据库实体，也就是说 N 个词语可以为 Y 个词语中具有候选数据库实体的词语。N 可以为大于或等于 1 的整数。还应理解，数据库实体为数据库中的属性名或属性值，数据库实体也可以为具有实际意义的词，例如可以为实词等。

[0050] 应理解，操作符可以包括多种符号，例如可以为≥、≤、＝、<、> 等。其中，可以通过预定义规则的方式，识别查询语句中包含的操作符。例如预定义操作符与规则对为“<: 在 ** 以下 | 小于”，那么对于“年龄在 30 岁以下”，识别出查询条件（年龄，操作符，30），“在 ** 以下”根据预定义规则为操作符“<”，那么完整的查询条件为（年龄，<，30）。

[0051] 应理解，本发明实施例中的标注信息也可以表述为标注序列或标注序列信息。

[0052] 应注意，在 150 中，第二词语和第三词语中的至少一个为 N 个词语的候选数据库实体中的数据库实体。第二词语也可以称为第二数据库实体，第三词语也可以称为第三数据

库实体；换句话说，在150中，根据标注信息生成K个查询条件，K个查询条件中的每个查询条件包括第二数据库实体、操作符和第三数据库实体，其中，操作符表示第二数据库实体和第三数据库实体的关系，第二数据库实体的标签为属性名，第三数据库实体的标签为属性值。其中第二数据库实体和第三数据库实体中的至少一个为N个词语的候选数据库实体中的数据库实体，1 ≤ K<N。

[0053] 可选地，在170中，可以根据K个查询条件和查询目标生成目标查询语句，目标查询语句为数据库查询语言，执行目标查询语句，得到查询结果。

[0054] 例如，用户输入查询语句（待查询语句）“年龄小于30岁的高级工程师的姓名”，经过上述过程可以得到查询条件为：“age<30岁”和“Job ＝高级工程师”，查询目标为“姓名”(name)，则生成的sql语句（目标查询语句）为：select name from view where age<30and job ＝‘高级工程师’。

[0055] 应理解，数据库查询语言可以是SQL语言，也可以为NO-SQL语言，本发明实施例并不对此做限定。

[0056] 可选地，作为另一实施例，在120中，根据预设词库划分待查询语句，得到N个初始词语；根据预设规则，规范化N个初始词语，得到N个词语。

[0057] 应理解，在本发明实施例中的词语可以为词组或者短语等。

[0058] 具体而言，可以根据从自然语言的词语、词组或者短语的概念、关系、属性等方面，解析待查询语句，例如，可以根据词语、词组或者短语的概念、关系、属性等对用户查询语句（待查询语句）进行分词，即将待查询语句切分成N个词语、词组或者短语（初始词语）。

[0059] 根据词语、词组或者短语的概念、关系、属性等对用户查询语句进行命名实体识别，即标识用户查询语句中的特定词语、词组或者短语的实体名称、类别。例如用户查询语句“销售部过去三年的业绩”，命名实体的结果可以为“销售部－机构名”，“过去三年－时间”等。同时还可以将其中的特定词语、词组、短语规范化成特定的词语，例如“过去三年”可以规范化成距离当前时间前三年的日期时间，最终得到N个词语。

[0060] 根据本发明实施例，还可以从自然语言的句法方面，解析用户查询语句，包括但不限于：根据词法分析结果，以及自然语言的句法结果，为其中的每个词标注词性，划分包含多个词、词组的短句，并且生成句法结构图便于后续生成查询条件。

[0061] 应理解，词库保存了特定词语、词组、短语和指示其概念、属性、关系的实体之间的关联。词库还可以保存词语的同义词、近义词等。词库可以但不限于保存在文件或者数据库中。

[0062] 可选地，作为另一实施例，在130中，可以根据N个词语从预设数据库中确定N个词语中的第一词语的n个初始候选数据库实体；，n为大于或等于1的整数；当n大于1时，确定n个初始候选数据库实体中每个初始候选数据库实体与第一词语的相关度，将n个初始候选数据库实体中相关度高于预设阈值的初始候选数据库实体确定为第一词语的至少一个候选数据库实体，或者，当n等于1时，将第一词语的n个初始候选数据库实体确定为第一词语的至少一个候选数据库实体。

[0063] 应理解，第一词语可以为N个词语中的任意一个词语。

[0064] 进一步地，作为另一实施例，确定n个初始候选数据库实体中每个初始候选数据库实体与每个词语的相关度，包括：根据以下方法中的至少一种方法确定n个初始候选数

据库实体中每个初始候选数据库实体与第一词语的相关度：命中率、向量空间余弦和编辑距离等。

[0065] 具体地，相关度也可以称为相似度，例如，可以根据命中率、向量空间余弦，编辑距离等确定至少一个初始候选数据库实体中每个初始候选数据库实体与每个词语的相关度，并对至少一个初始候选数据库实体的实体进行排序或过滤。假设以编辑距离作为相似度的计算方式，关键词“北京大学”的候选数据库实体有｛属性值 1——北京大学，属性值 2——北京大学深圳分院｝，对于属性值 1 的编辑距离为 0，属性值 2 的编辑距离为 4，属性值比属性值 2 的要小，则认为属性值 1 更相似。假如设定编辑距离过滤阈值为 1，那么属性值 2 将被过滤掉。

[0066] 应理解，预定阈值为已确定的值，可以认为预先已设定好的值，也可以认为是在之前的预测过程中得到的值，较优的，本发明实施例中的预定阈值可以直接使用，不需要计算或通过其他求解即可获得。

[0067] 可选地，作为另一实施例，在 140 中，可以对每一个待识别的实体检索数据库实体库，得到至少一个候选数据库实体。检索的方式可以是直接使用待识别实体本身或其数据类型。待识别实体假如是时间 / 日期型或者数值型，默认是待确定的属性值。例如用户查询语句“2013 年毕业于北京大学的有多少人”，经过步骤 120 后，换句话说预处理后，输出若干关键词序列 (2013 年 /Date，毕业，北京大学)，那么对于“2013 年”是时间 / 日期类型，则检索其相同数据类型的属性名，比如可能的候选数据库实体｛属性名 1——销售时间；属性名 2——入职时间；属性名 3——离职时间……｝，而对于“毕业”可能的候选数据库实体｛属性名 1——毕业时间；属性名 2——毕业学校；属性名 3——毕业证书｝，对于“北京大学”可能为｛属性值 1——北京大学，属性值 2——北京大学深圳分院｝。从上面可见“2013 年”是默认的待确定属性值，标注为 value(属性值)，“毕业”的候选数据实体都是属性名，可标注为 field(属性名)，“北京大学”的候选数据库实体都是属性值，可标注为 value，那么输出的标注信息为 (2013 年 /value，毕业 /field，北京大学 /value)。

[0068] 可选地，作为另一实施例，在 150 之前，本发明实施例方法还包括：根据标注信息中的词语的候选数据库实体，合并标注信息中连续标签为属性名的词语，得到第一合并词语，第一合并词语为标注信息中连续标签为属性名的词语的候选数据库实体的交集，使用第一合并词语替换标注信息中连续标签为属性名的词语，以对标注信息进行更新，和 / 或根据标注信息中的词语的候选数据库实体，合并标注信息中连续标签为属性值的词语，得到第二合并词语，第二合并词语为标注信息中连续标签为属性值的词语的候选数据库实体的交集，使用第二合并词语替换标注信息中连续标签为属性值的词语，以对标注信息进行更新，其中，在 150 中，根据更新后的标注信息生成 K 个查询条件。在 160 中，根据更新后的标注信息生成 KG 个查询条件

[0069] 具体而言，合并标注信息中连续标签为属性名或属性值的词语，包括合并计算 P(Field|field_1,field_2…field_n) 或 P(Value|value_1,value_2…value_n)；具体地，当标注信息出现连续的 field 或 value 标签时，以贪心的方式尝试合并 field_1,field_2…field_n 或 value_1,value_2…value_n，计算减少原候选数据库实体数量的概率。例如用户查询语句“张三所属岗位的职责”，其中关键词“岗位”的候选数据库实体可能有｛岗位名称，岗位职责，岗位类型……｝，关键词“职责”可能有｛职位职责，岗位职责……｝，用户查询

语句对应的标注信息（张三 /value，岗位 /field，职责 /field），其中“岗位”与“职责”出现连续 field，那么尝试合并“岗位”和“职责”，主要通过对两者的候选数据库实体求交集进行判定是否最终合并，如果交集中候选数据库实体数量减少了（不为 0），证明 P(Field| 岗位，职责）比 P(Field| 岗位）和 P(Field| 职责）要大，那么直接合并，继续尝试合并下一个，直到 P(Field|field_1,field_2…field_n) 或 P(Value|value_1,value_2…value_n) 出现最大值，更新标注信息，比如当前查询语句合并后，更新标注信息为（张三 /value，岗位职责 /field）。

[0070] 可选地，作为另一实施例，在 150 中，根据标注信息生成 M 个候选查询条件，M 个候选查询条件中的每个候选查询条件包括第一候选词语、操作符和第二候选词语的对应关系，其中第一候选词语的标签为属性名，第二候选词语的标签为属性值，M 为大于或等于 K 的整数；

[0071] 确定每个候选查询条件的第一候选词语和第二候选词语的匹配指数；

[0072] 将 M 个候选查询条件中的匹配指数大于预设阈值的 K 个候选查询条件确定为 K 个查询条件。

[0073] 根据标注信息生成 M 个候选查询条件；

[0074] 换句话说，根据 M 个候选查询条件得到第一候选查询条件，第一候选查询条件包括第一候选词语、操作符和第二候选词语的对应关系，其中第一候选词语的标签为属性名，第二候选词语的标签为属性值；其中第一候选词语和第二候选词语中的至少一个为 N 个词语中的词语；确定第一候选词语和第二候选词语的匹配指数；在匹配指数大于预设参数阈值时，将第一候选查询条件确定为第一查询条件，其中，第一候选词语作为第一词语，第二候选词语作为为第二词语。

[0075] 具体而言，可以扫描标注信息，配对 field 和 value，或者根据隐指的 Field，生成候选查询条件。例如用户查询语句“年龄小于 30 岁的高级工程师”，其标注信息为（年龄 /field，小于，30 岁 /value，高级工程师 /value），其中“年龄”对应属性名“Age”，“30 岁”隐指“Age”的属性值，“高级工程师”隐指属性名“Job”的属性值，假如没有存在歧义或多个候选数据库实体，则可配对 field 与 value。没配对的“高级工程师 /value”，使用其隐指的 field，生成候选查询条件 (age，操作符，30) 和“(Job，操作符，高级工程师）”。

[0076] 进一步地，作为另一实施例，根据标注信息生成 M 个候选查询条件，包括：根据标注信息生成 M 个初始候选查询条件；根据用户信息，对 M 个初始候选查询条件进行消歧处理，得到 M 个候选查询条件，消歧处理包括根据用户信息消除 M 个初始候选查询条件中存在歧义的初始候选查询条件中的歧义，其中，用户信息包括终端设备的硬件信息、终端系统的软件信息、保存在终端内存或者存储设备上的用户数据、用户的历史操作和用户的设定中的至少一种。

[0077] 具体而言，可以根据用户个人信息，消除用户查询语句中的歧义。例如在企业 HR(Human Resource，人力资源）数据库搜索系统中，用户查询“部门任职高级工程师的有多少人”，其中“部门”是存在歧义的实体，不知道其指的是某个或某几个部门，但是从查询用户的个人信息，如工号、姓名、所在部门等信息，可以确认查询语句中的“部门”隐性表示是用户的所在部门，根据用户信息对其进行消歧处理，得到查询条件。

[0078] 应理解，用户个人信息包括用户个人信息数据包括但不限于：终端设备的硬件信

息，包括但不限于日期和时钟信息（例如但不限于当前日期、时间、时区等），位置信息（例如但不限于GPS、国家、城市），通过传感器产生的信息（例如但不限于加速度、磁力、方向、陀螺仪、光线感应、压力、温度、脸部感应、重力、旋转矢量等信息），或者上述方式的混合。终端系统的软件信息，包括但不限于操作系统和运行的软件、进程、服务的状态、事件和提供的数据。保存在终端内存或者存储设备上的用户数据，包括但不限于短文本，通讯录，备忘录，提醒事项，照片，应用，视频，音频，邮件，书签，网页浏览记录，商品/服务的购买记录，酒店预订记录，机票购买记录。用户的历史操作，包括但不限于用户历史查询语句。用户的设定，包括但不限于用户信息（例如姓名、电话号、地址、账户等等），用户偏好设置。

[0079] 可选地，作为另一实施例，确定每个候选查询条件的第一候选词语和第二候选词语的匹配指数，包括：

[0080] 根据第一候选词语和第二候选词语的配对概率、序列距离、数据库数据类型匹配度和语言习惯约束中的至少一种确定匹配指数。

[0081] 其中，匹配指数与配对概率、序列距离和语言习惯约束成负相关。匹配指数与数据库数据类型匹配度成正相关。配对概率、序列距离、数据库数据类型匹配度和语言习惯约束的定义如下，配对概率指第一候选词语所对应的数据库实体与第二候选词语所对应的数据库实体之间的交集的多少，当第一候选词语所对应的数据库实体与第二候选词语所对应的数据库实体之间的交集越少，配对概率越大；序列距离也可以称为语句距离，指在标注信息或查询语句中第一候选词语和第二候选词语之间的词语或字数的多少，当查询语句中第一候选词语和第二候选词语之间的词语或字数越多时，序列距离越大；数据库数据类型匹配度指第一候选词语和第二候选词语的数据库数据类型是否匹配（一致），第一候选词语和第二候选词语的数据类型匹配时的数据库数据类型匹配度大于第一候选词语和第二候选词语的数据类型不匹配时的数据库数据类型匹配度；语言习惯约束指第一候选词语和第二候选词语是否符合数据库或语言习惯决定，第一候选词语和第二候选词语符合数据库或语言习惯时的语言习惯约束小于第一候选词语和第二候选词语不符合数据库或语言习惯时的语言习惯约束。

[0082] 本发明实施例中，可以根据用户查询语句上下文，对序列中存在歧义或有多个候选数据库实体的待识别实体计算上述特征值（配对概率、序列距离、数据库数据类型匹配度和语言习惯约束）。

[0083] 具体而言，配对概率由第一候选词语所对应的数据库实体与第二候选词语所对应的数据库实体之间的交集决定，其中，第一候选词语所对应的数据库实体与第二候选词语所对应的数据库实体之间的交集越少，配对概率越大，匹配指数越大。

[0084] 配对概率：P(Field-Value|field,value) 表示序列中 field 与 value 配对，生成查询条件 (Field,操作符，Value) 的概率，主要方式是根据两者的候选数据库实体是否存在交集，交集的元素的多少决定。例如用户查询语句用户查询语句“去年毕业的研究生有多少人”，假设“去年”的候选数据库实体有｛毕业时间，入职时间，离职时间……｝，“毕业”的候选数据库实体有｛毕业学校，毕业证书，毕业时间……｝，其标注信息为（去年/value，毕业/field，研究生/value），计算P(Field-Value| 毕业，去年）时，两者存在交集｛毕业时间｝，可认为P(Field-Value| 毕业，去年）＝s(s>0)，即生成查询条件（毕业时间，操作符，去年）的概率为s。假如交集中存在m元素，P(Field-Value| 毕业，去年）＝s/m。而对于

P(Field-Value| 毕业，研究生），因不存在交集，则为 0。

[0085] 具体而言，序列距离由第一候选词语和第二候选词语在标注信息或查询语句中的距离决定，其中，第一候选词语和第二候选词语在标注信息或查询语句中的距离越大，序列距离越大，匹配指数越小，标注信息或查询语句中第一候选词语和第二候选词语之间的词语的多少，表示距离的大小。

[0086] 序列距离：L(Field-Value|field,value) 表示序列中 field 与 value 配对，生成查询条件 (Field,操作符,Value) 时 field 与 value 之间的距离。距离越小，生成查询条件的概率越大。主要的计算方式是根据两者在标注信息或查询语句中的距离，例如（年龄 /field，小于，30 岁 /value，职级 /field，大于，18/value），其中“年龄”与“30 岁”在序列中相隔“小于”，即 L(Field-Value| 年龄，30 岁）为 2；而 L(Field-Value| 年龄，18) 则为 8。

[0087] 具体而言，数据库数据类型匹配度由第一候选词语和第二候选词语的数据库数据类型是否一致决定，其中，第一候选词语和第二候选词语的数据类型一致时的数据库数据类型匹配度大于第一候选词语和第二候选词语的数据类型不一致时的数据库数据类型匹配度，匹配指数与数据库类型匹配度正相关。

[0088] 数据库数据类型匹配度：Type(Field-Value|field,value) 表示序列中 field 的数据库数据类型与 value 的数据类型是否一致。若一致，则配对生成查询条件的可能性更大。例如“年龄 /field”的数据库数据类型是数值型，因此与数值型的“18/value”的 Type(Field-Value| 年龄，18) ＝ 1，对于字符型“中国 /value”的 Type(Field-Value| 年龄，中国）＝ 0。

[0089] 具体而言，语言习惯约束由第一候选词语和第二候选词语是否符合数据库或语言习惯决定，其中，第一候选词语和第二候选词语符合数据库或语言习惯时的语言习惯约束小于第一候选词语和第二候选词语不符合数据库或语言习惯时的语言习惯约束，匹配指数与语言习惯约束负相关。

[0090] 语言习惯约束：C(Field-Value|field,value) 表示序列中 field 与 value 配对，value 是否符合 field 在数据库或语言习惯约束。若符合，则配对生成查询条件的可能性更大，这里的约束一般指量词与数值范围约束。例如（年龄 /field，小于，30 岁 /value，职级 /field，大于，25/value），其中“职级 /field”和“30 岁 /value”，因量词“岁”不符合“职级”的量词约束，那么 C(Field-Value| 职级，30 岁）为 0。假设“职级 /field”在数据库中数值范围的约束是 13 ～ 21，那么对于“职级 /field”和“25/value”，因 value 不符合该约束，则 C(Field-Value| 职级，25) 为 0。

[0091] 经过以上处理，field 和 value 配对生成查询条件 (Field,操作符,Value) 的匹配指数可以为上述特征值的线性加权值。例如，

[0092] 匹配指数 Score ＝ z1*P+z2*L+z3*Type+z4*C。其中 z1、z2、z3 和 z4 为预先确定的权重值。

[0093] 最后通过设定预设阈值（过滤规则），筛选输出查询条件。

[0094] 可选地，作为另一实施例，在 160 中，可以确定标注信息中的标签为属性名的词语满足预设条件和 / 或为孤点词语，其中，孤点词语没有对应的标签为属性值的词语和隐性标签为属性值的词语；将标注信息中的标签为属性名的词语的属性名作为查询目标。

[0095] 具体地，预设条件可以包括通过句法或者预定义规则的方式，换句话说可以通过

句法或者预定义规则的方式识别用户查询语句或标注信息中的查询目标。例如预设条件包括标签为属性名的词语之前具有“的”字，例如，预设条件可以为“.*的field1和field2”表示查询目标是field1和field2，当用户输入查询语句类似“张三的工号和部门”时，标注信息为（张三/value，的，工号/field，和，部门/field），符合该预定义规则，“工号”和“部门”即为查询目标；类似的，预设条件可以为“.*的field”。

[0096] 本发明实施例中，也可以将孤点词语作为查询目标，例如，如果出现没有value与其配对的field，则忽略或者加入到查询目标当中；如果出现没有field与其配对的value，且value的候选数据库实体拥有同一个隐性field，则使用隐性的field与其配对生成查询条件，否则忽略。例如用户查询语句“年龄张三的部门”，其中“年龄/field”，但是没有value与其配对，且不是查询目标，则忽略或者加入查询目标当中。例如用户查询语句“销售部过去三年的业绩”，其中“销售部/value”的候选数据库实体｛属性值1——手机销售部，属性值2——服务器销售部｝，所有的候选数据库实体拥有同一个隐性的field——“部门”，则生成查询条件（部门，操作符，手机销售部）和（部门，操作符，服务器销售部）。

[0097] 上文中结合图1详细描述了本发明实施例的数据库查询的方法，下面将结合图2具体地例子，更加详细的描述本发明实施例的数据库查询的方法。应注意，图2的例子是为了帮助本领域技术人员更好地理解本发明实施例，而非要限制本发明实施例的范围。本领域技术人员根据所给出的图2的例子，显然可以进行各种等价的修改或变化，这样的修改或变化也落入本发明实施例的范围内。

[0098] 应理解，上述各过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

[0099] 图2是根据本发明另一实施例的数据库查询的方法的示意性流程图。如图2所示的方法包括：

[0100] 201，获取查询语句。

[0101] 具体地，接收用户输入的自然语言查询语句，例如查询语句可以为“去年我部门毕业于北大，年龄小于30任职大于18级人员所在的岗位的名称”。

[0102] 202，预处理。

[0103] 具体地，预处理过程包括对查询语句进行分句、分词、词性标注、命名实体识别、句法分析等。同时进行规范化，例如查询语句中的“去年”规范化成2013年（假设当前时间是2014年），并且关联实体“时间”，“北大”关联实体“机构名”，“30”及“18级”关联为数量词等等。识别谓语（动词）“毕业”的直接宾语“北大”等。

[0104] 203，获取候选数据库实体。

[0105] 具体地，根据预处理的结果，对每一个待识别的实体检索数据库实体库，返回一个或多个候选数据库实体——属性名(field)或者属性值(value)。对于时间/日期、数字型等待识别实体，获取数据库中相同数据类型的属性名作为其候选数据库实体。其余字符型关键词，获取属性名/属性值中包含该关键词或者同义词的属性名/属性值作为候选数据库实体，假如通过先验知识知道待识别实体是数据库实体的别名，应使用数据库实体的正式名称去获取相关的候选数据库实体，例如查询语句中“毕业”的候选数据实体可能是｛毕业时间，毕业学校，毕业证书……｝；而对于“北大”，它是北京大学的别名，应该以“北京大学”这个正式的数据库实体去获取其他相关的候选数据库实体，比如｛北京大学，北京大学

研究生院，北京大学深圳研究院……}，不应该包含“北京理工大学”等只命中关键词的数据库实体。最终输出与用户查询语句对应的标注信息 (2013 年 /value，我部门，毕业 /field，北京大学 /value，年龄 /field，小于，30/value，任职 /field，大于，18 级 /value，人员，所在，的，岗位 /field，的，名称 /field)。

[0106] 204，相似度计算。

[0107] 具体地，计算待识别实体或数据实体正式名与候选数据库实体之间的相似度（相关度）。可以根据命中率、向量空间余弦和编辑距离中的至少一种确定相似度，例如以命中率和覆盖率的线性加权计算相似度。命中率＝｛关键词或数据库实体正式名与候选数据库实体的交集的权重和｝/｛关键词的权重和｝，比如查询语句中“毕业”与候选数据库实体“毕业时间”的交集是｛毕业｝，其权重为 w1，那么关键词“毕业”与候选数据库实体“毕业时间”的命中＝ w1/w1 ＝ 1.0 ；覆盖率＝｛关键词或数据库实体正式名与候选数据库实体的交集的权重和｝/｛候选数据库实体的权重和｝，比如查询语句中“毕业”与候选数据库实体“毕业时间”的交集是｛毕业｝，其权重为 w1，“毕业时间”包含两个词“毕业”与“时间”，假设“时间”的权重为 w2，那么“毕业时间”的权重和＝ w1+w2，关键词“毕业”与候选数据库实体“毕业时间”的覆盖率＝ w1/(w1+w2)。最后关键词“毕业”与候选数据库实体“毕业时间”的相似度＝ a1* 命中率 +a2* 覆盖率，其中 a1 与 a2 分别为命中率与覆盖率的权重，a1 和 a2 可以为预设值。

[0108] 205，合并计算。

[0109] 具体地，根据标注信息中的词语的候选数据库实体合并标注信息中连续标签为属性名或属性值的词语，得到合并词语，合并词语为标注信息中连续标签为属性名或属性值的词语的候选数据库实体的交集；使用合并词语替换标注信息中连续标签为属性名或属性值的词语，以对标注信息进行更新。

[0110] 换句话说，根据标注信息中的词语的候选数据库实体，合并标注信息中连续标签为属性名的词语，得到第一合并词语，第一合并词语为标注信息中连续标签为属性名的词语的候选数据库实体的交集，使用第一合并词语替换标注信息中连续标签为属性名的词语，以对标注信息进行更新，和 / 或根据标注信息中的词语的候选数据库实体，合并标注信息中连续标签为属性值的词语，得到第二合并词语，第二合并词语为标注信息中连续标签为属性值的词语的候选数据库实体的交集，使用第二合并词语替换标注信息中连续标签为属性值的词语，以对标注信息进行更新，

[0111] 具体地，扫描输出序列（标注信息），发现“岗位”和“名称”是连续 field，“岗位”的候选数据库实体有｛岗位职责，岗位名称，岗位等级｝，“名称”的候选数据库实体有｛职位名称，岗位名称｝，尝试合并，两者候选数据库实体交集｛岗位名称｝，元素个数为 1，数量比原来要小，根新标注信息为 (2013 年 /value，我部门，毕业 /field，北京大学 /value，年龄 /field，小于，30/value，任职 /field，大于，18 级 /value，人员，所在，的，岗位名称 /field)。

[0112] 206，查询目标识别。

[0113] 具体地，通过句法或者预定义规则的方式识别用户查询语句中的查询目标。例如预定义规则“.* 的 field”表示查询目标是 field。当前查询语句符合该规则，生成查询目标——“岗位名称”。

[0114] 207，查询条件识别。

[0115] 具体地，扫描标注信息，配对 field 和 value，或者根据隐指的 Field，生成候选查询条件。由于序列中多个待识别实体包含多个候选数据库实体，所以判断存在歧义，需要消歧。

[0116] 208，是否存在歧义。

[0117] 具体地，如果存在歧义则执行步骤 209，如果不存在歧义，则执行步骤 211。

[0118] 209，用户信息消歧。

[0119] 具体地，通过用户的个人信息和预定义规则的方式对查询语句进行消歧。例如在用户登录的情况下，输入查询语句，默认情况下或针对某类型关键词增加某类查询条件，对于标注信息中的“我部门”等此类关键词，结合用户信息，在查询条件中增加（部门，操作符，用户所在部门）进行消歧。

[0120] 应理解，用户个人信息包括用户个人信息数据包括但不限于：终端设备的硬件信息，包括但不限于日期和时钟信息（例如但不限于当前日期、时间、时区等），位置信息（例如但不限于 GPS、国家、城市），通过传感器产生的信息（例如但不限于加速度、磁力、方向、陀螺仪、光线感应、压力、温度、脸部感应、重力、旋转矢量等信息），或者上述方式的混合。终端系统的软件信息，包括但不限于操作系统和运行的软件、进程、服务的状态、事件和提供的数据。保存在终端内存或者存储设备上的用户数据，包括但不限于短文本，通讯录，备忘录，提醒事项，照片，应用，视频，音频，邮件，书签，网页浏览记录，商品 / 服务的购买记录，酒店预订记录，机票购买记录。用户的历史操作，包括但不限于用户历史查询语句。用户的设定，包括但不限于用户信息（例如姓名、电话号、地址、账户等等），用户偏好设置。

[0121] 210，上下文消歧。

[0122] 具体地，根据用户查询语句上下文，对序列中存在歧义或多个候选数据库实体的待识别实体计算以下特征值，假设“年龄”的候选数据库实体有｛年龄｝，“30”按数据类型可能获得的候选数据库实体有｛年龄，任职等级，试用期天数……｝，“18 级”按按数据类型可能的候选数据库实体有｛年龄，任职等级，试用期天数……｝，下面举例以“年龄 /field”和“30/value”与“18 级 /value”配对时的计算过程：

[0123] 具体地，可以根据第一候选词语和第二候选词语的配对概率 P、序列距离 L、数据库数据类型匹配度 Type 和语言习惯约束 C 中的至少一种确定匹配指数。

[0124] 其中，P(Field-Value|field,value) 表示序列中 field 与 value 配对，生成查询条件 (Field，操作符，Value) 的概率。主要方式是根据两者的候选数据库实体是否存在交集，交集的元素的多少决定。对于标注信息，计算 P(Field-Value| 年龄，30) 时，两者存在交集｛年龄｝且元素个数为 1，可认为 P(Field-Value| 年龄，30) ＝ s(s>0)，生成查询条件（毕业时间，操作符，去年）的概率为 s。同理 P(Field-Value| 年龄，18 级）＝ s。

[0125] L(Field-Value|field,value) 表示序列中 field 与 value 配对，生成查询条件 (Field，操作符，Value) 时，field 与 value 之间的距离。距离越小，生成查询条件的概率越大。主要的计算方式是根据两者在标注信息或查询语句中的距离。对于标注信息中 L(Field-Value| 年龄，30) 为 2；而 L(Field-Value| 年龄，18 级）则为 8。

[0126] Type(Field-Value|field,value) 表示序列中 field 的数据库数据类型与 value 的数据类型是否一致。若一致，则配对生成查询条件的可能性更大。对于标注信息中 Type(Field-Value| 年龄，30) ＝ 1，Type(Field-Value| 年龄，18 级）＝ 1。

[0127] C(Field-Value|field,value) 表示序列中 field 与 value 配对，value 是否符合 field 在数据库或语言习惯约束。若符合，则配对生成查询条件的可能性更大，这里的约束一般指量词与数值范围约束。对于标注信息中 C(Field-Value| 年龄 ,30) = 1，C(Field-Value| 年龄 ,18 级) = 0。

[0128] 经过以上处理，年龄和 30 的匹配指数为：

[0129] Score1 = z1*P(Field-Value| 年 龄 ,30)+z2*L(Field-Value| 年龄 ,30)+z3*Type(Field-Value| 年 龄 ,30)+z4*C(Field-Value| 年 龄 ,30) = z1*s+z2*2+z3*1+z4*1 = z1*s+z2*2+z3+z4

[0130] 年龄和 18 级的匹配指数为：

[0131] Score2 = z1*P(Field-Value| 年 龄 ,18 级)+z2*L(Field-Value| 年 龄 ,18 级)+z3*Type(Field-Value| 年 龄 ,18 级)+z4*C(Field-Value| 年 龄 ,18 级) = z1*s+z2*2+z3*1+z4*0 = z1*s+z2*8+z3

[0132] 其中 z1、z2、z3 和 z4 是线下通过机器学习的方式生成的权重值，换句话说，z1、z2、z3 和 z4 为预先确定的值，存放在语义消歧模型中。从上述特征的设计上看，特征 1)、3)、4) 为正向特征，则 z1，z3, 和 z4 为正数，而 z2 为负向特征，其值为负数，可知 Score1 要比 Score2 大。最后通过设定阈值或过滤规则，筛选查询条件，比如 C(Field-Value|field,value) 为 0 的查询条件忽略，那么查询条件 (年龄，操作符，18 级) 就被忽略掉了。

[0133] 211，孤点处理。

[0134] 具体地，假如出现没有 value 与其配对的 field，则忽略或者加入到查询目标当中；假如出现没有 field 与其配对的 value，且 value 的候选数据库实体拥有同一个隐性 field，则使用隐性的 field 与其配对生成查询条件，否则忽略。按上述计算，当前标注信息不存在孤点。

[0135] 212，操作符处理。

[0136] 换句话说，识别操作符，具体地，通过预定义规则的方式，识别查询语句中包含的操作符。例如默认操作符为“=”，预定义其他操作符与规则对为“<: 在 ** 以下 | 小于”，那么对于查询条件 (年龄，操作符，30)，其在查询语句或序列中，(年龄 /field，小于，30/value) 符合预定义规则，那么完整的查询条件为 (年龄，<，30)。最后输出的查询目标——岗位名称，查询条件为 (毕业时间，=，2013 年)、(毕业学校，=，北京大学)、(年龄，<，30)、(任职等级，=，18 级) 以及 (部门，=，用户所在部门)。

[0137] 213，数据库查询语句生成。

[0138] 具体地，根据上述模块输出查询条件与目标，生成数据库查询语句，如 SQL，那么对于当前查询语句生成的数据库查询语句为——select 岗位名称 from view where 毕业时间 = 2013and 毕业学校 = 北京大学 and 年龄 <30and 任职等级 = 18and 部门 = 用户所在部门，对数据库进行检索。

[0139] 214，输出结果。

[0140] 具体地，执行数据库查询语句，把检索结果返回给用户。

[0141] 因此，本发明实施例通过根据将为自然语言查询语句的待查询语句生成查询目标和查询条件，根据查询目标和查询条件进行查询，进而得到查询结果，能够根据用户请求进

行数据库查询。本发明实施例无需用户熟悉数据库查询语言，提升用户体验。

[0142] 上文中结合图 1 至图 2，详细描述了根据本发明实施例的数据库查询的方法，下面结合图 3 至图 4 详细描述根据本发明实施例的数据库查询的设备。

[0143] 图 3 是根据本发明一个实施例的数据库查询的设备的示意框图。数据库查询的设备可以为用户设备或数据库服务器等，如图 3 所示 3 的设备 300 包括：获取单元 310、划分单元 320、确定单元 330、标注单元 340、第一生成单元 350、第二生成单元 360 和查询单元 370。

[0144] 具体地，获取单元 310 用于获取待查询语句，待查询语句为自然语言查询语句；划分单元 320 用于根据预设词库划分待查询语句，得到 N 个词语；确定单元 330 用于从预设数据库中确定第一词语的至少一个候选数据库实体，第一词语为 N 个词语中的任一词语；标注单元 340 用于为 N 个词语中的各个词语分别标注标签，得到与待查询语句对应的标注信息，标注信息包括 N 个词语和与 N 个词语中的各个词语呈一一对应关系的标签，其中，与第一词语呈一一对应关系的标签用于表示第一词语的数据类型，第一词语的标签包括属性名或属性值；第一生成单元 350 用于根据标注信息生成 K 个查询条件，K 个查询条件中的每个查询条件包括第二词语、操作符和第三词语，其中，操作符表示第二词语和第三词语的关系，第二词语的标签为属性名，第三词语的标签为属性值；第二生成单元 360 用于根据标注信息生成查询目标，查询目标包括 N 个词语中的至少一个词语的数据库实体，其中，至少一个词语的标签为属性名，至少一个词语中的每个词语的数据库实体为每个词语的至少一个候选数据库实体中的一个；查询单元 370 用于根据 K 个查询条件和查询目标进行查询，得到查询结果。

[0145] 因此，本发明实施例通过将为自然语言查询语句的待查询语句生成查询目标和查询条件，根据查询目标和查询条件进行查询，进而得到查询结果，能够根据用户请求进行数据库查询。本发明实施例无需用户熟悉数据库查询语言，提升用户体验。

[0146] 可选地，作为另一实施例，划分单元 320 根据预设词库划分待查询语句，得到 N 个初始词语；根据预设规则，规范化 N 个初始词语，得到 N 个词语。

[0147] 可选地，作为另一实施例，确定单元 330 从预设数据库中确定第一词语的 n 个初始候选数据库实体，n 为大于或等于 1 的整数；当 n 大于 1 时，确定 n 个初始候选数据库实体中每个初始候选数据库实体与第一词语的相关度，将 n 个初始候选数据库实体中相关度高于预设阈值的初始候选数据库实体确定为第一词语的至少一个候选数据库实体，或者，当 n 等于 1 时，将第一词语的 n 个初始候选数据库实体确定为第一词语的至少一个候选数据库实体。

[0148] 进一步地，作为另一实施例，确定单元 330 根据以下方法中的至少一种方法确定 n 个初始候选数据库实体中每个初始候选数据库实体与第一词语的相关度：命中率、向量空间余弦和编辑距离。

[0149] 可选地，作为另一实施例，设备 300 还包括：合并单元。具体地，合并单元用于在第一生成单元 350 根据标注信息生成 K 个查询条件之前，根据标注信息中的词语的候选数据库实体，合并标注信息中连续标签为属性名的词语，得到第一合并词语，第一合并词语为标注信息中连续标签为属性名的词语的候选数据库实体的交集，使用第一合并词语替换标注信息中连续标签为属性名的词语，以对标注信息进行更新，和 / 或根据标注信息中的词语的候选数据库实体，合并标注信息中连续标签为属性值的词语，得到第二合并词语，第二合

并词语为标注信息中连续标签为属性值的词语的候选数据库实体的交集，使用第二合并词语替换标注信息中连续标签为属性值的词语，以对标注信息进行更新，其中，第一生成单元350根据更新后的标注信息生成K个查询条件，第二生成单元360根据更新后的标注信息生成查询目标。

[0150] 可选地，作为另一实施例，第一生成单元350根据标注信息生成M个候选查询条件，M个候选查询条件中的每个候选查询条件包括第一候选词语、操作符和第二候选词语的对应关系，其中第一候选词语的标签为属性名，第二候选词语的标签为属性值；确定每个候选查询条件的第一候选词语和第二候选词语的匹配指数；将M个候选查询条件中的匹配指数大于预设阈值的K个候选查询条件确定为K个查询条件。

[0151] 进一步地，作为另一实施例，第一生成单元350根据标注信息生成M个初始候选查询条件；根据用户信息，对M个初始候选查询条件进行消歧处理，得到M个候选查询条件，消歧处理包括根据用户信息消除M个初始候选查询条件中存在歧义的初始候选查询条件中的歧义，其中，用户信息包括终端设备的硬件信息、终端系统的软件信息、保存在终端内存或者存储设备上的用户数据、用户的历史操作和用户的设定中的至少一种。

[0152] 进一步地，作为另一实施例，第一生成单元350根据第一候选词语和第二候选词语的配对概率、序列距离、数据库数据类型匹配度和语言习惯约束中的至少一种确定匹配指数。

[0153] 具体地，作为另一实施例，配对概率由第一候选词语所对应的数据库实体与第二候选词语所对应的数据库实体之间的交集决定，其中，第一候选词语所对应的数据库实体与第二候选词语所对应的数据库实体之间的交集越少，配对概率越大，匹配指数越大。

[0154] 具体地，作为另一实施例，序列距离由第一候选词语和第二候选词语在标注信息或查询语句中的距离决定，其中，第一候选词语和第二候选词语在标注信息或查询语句中的距离越大，序列距离越大，匹配指数越小，标注信息或查询语句中第一候选词语和第二候选词语之间的词语的多少，表示距离的大小。

[0155] 具体地，作为另一实施例，数据库数据类型匹配度由第一候选词语和第二候选词语的数据库数据类型是否一致决定，其中，第一候选词语和第二候选词语的数据类型一致时的数据库数据类型匹配度大于第一候选词语和第二候选词语的数据类型不一致时的数据库数据类型匹配度，匹配指数与数据库类型匹配度正相关。

[0156] 具体地，作为另一实施例，语言习惯约束由第一候选词语和第二候选词语是否符合数据库或语言习惯决定，其中，第一候选词语和第二候选词语符合数据库或语言习惯时的语言习惯约束小于第一候选词语和第二候选词语不符合数据库或语言习惯时的语言习惯约束，匹配指数与语言习惯约束负相关。

[0157] 可选地，作为另一实施例，第二生成单元360确定标注信息中的标签为属性名的词语满足预设条件和/或为孤点词语，其中，孤点词语没有对应的标签为属性值的词语；将标注信息中的标签为属性名的词语的属性名作为查询目标。

[0158] 应注意，图3所示的数据库查询的设备能够实现图1-图2的方法实施例中由数据库查询的设备完成的各个过程。数据库查询的设备300的其他功能和操作可以参考图1和图2的方法实施例中涉及数据库查询的设备的各个过程。为避免重复，此处不再详述。

[0159] 图4是根据本发明另一实施例的数据库查询的设备的示意框图。如图4所示的设

备 400 包括 :处理器 410、存储器 420 和总线系统 430。

[0160] 具体地,处理器 410 通过总线系统 430 调用存储在存储器 420 中的代码,获取待查询语句,待查询语句为自然语言查询语句 ;根据预设词库划分待查询语句,得到 N 个词语 ;从预设数据库中确定第一词语的至少一个候选数据库实体,第一词语为 N 个词语中的任一词语 ;为 N 个词语中的各个词语分别标注标签,得到与待查询语句对应的标注信息,标注信息包括 N 个词语和与 N 个词语中的各个词语呈一一对应关系的标签,其中,与第一词语呈一一对应关系的标签用于表示第一词语的数据类型,第一词语的标签包括属性名或属性值 ;根据标注信息生成 K 个查询条件,K 个查询条件中的每个查询条件包括第二词语、操作符和第三词语,其中,操作符表示第二词语和第三词语的关系,第二词语的标签为属性名,第三词语的标签为属性值 ;根据标注信息生成查询目标,查询目标包括 N 个词语中的至少一个词语的数据库实体,其中,至少一个词语的标签为属性名,至少一个词语中的每个词语的数据库实体为每个词语的至少一个候选数据库实体中的一个 ;根据 K 个查询条件和查询目标进行查询,得到查询结果。

[0161] 因此,本发明实施例通过根据将为自然语言查询语句的待查询语句生成查询目标和查询条件,根据查询目标和查询条件进行查询,进而得到查询结果,能够根据用户请求进行数据库查询。本发明实施例无需用户熟悉数据库查询语言,提升用户体验。

[0162] 上述本发明实施例揭示的方法可以应用于处理器 410 中,或者由处理器 410 实现。处理器 410 可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过处理器 410 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 410 可以是通用处理器、数字信号处理器 (英文 Digital Signal Processor,简称 DSP)、专用集成电路 (英文 Application Specific Integrated Circuit,简称 ASIC)、现成可编程门阵列 (英文 Field Programmable Gate Array,简称 FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存取存储器 (英文 Random Access Memory,简称 RAM)、闪存、只读存储器 (英文 Read-Only Memory,简称 ROM)、可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 420,处理器 410 读取存储器 420 中的信息,结合其硬件完成上述方法的步骤,该总线系统 430 除包括数据总线之外,还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见,在图中将各种总线都标为总线系统 430。

[0163] 可选地,作为另一实施例,处理器 410 根据预设词库划分待查询语句,得到 N 个初始词语 ;根据预设规则,规范化 N 个初始词语,得到 N 个词语。

[0164] 可选地,作为另一实施例,处理器 410 从预设数据库中确定第一词语的 n 个初始候选数据库实体,n 为大于或等于 1 的整数 ;当 n 大于 1 时,确定 n 个初始候选数据库实体中每个初始候选数据库实体与第一词语的相关度,将 n 个初始候选数据库实体中相关度高于预设阈值的初始候选数据库实体确定为第一词语的至少一个候选数据库实体,或者,当 n 等于 1 时,将第一词语的 n 个初始候选数据库实体确定为第一词语的至少一个候选数据库实体。

[0165] 进一步地，作为另一实施例，处理器410根据以下方法中的至少一种方法确定n个初始候选数据库实体中每个初始候选数据库实体与第一词语的相关度：命中率、向量空间余弦和编辑距离。

[0166] 可选地，作为另一实施例，处理器410在根据标注信息生成K个查询条件之前，根据标注信息中的词语的候选数据库实体，合并标注信息中连续标签为属性名的词语，得到第一合并词语，第一合并词语为标注信息中连续标签为属性名的词语的候选数据库实体的交集，使用第一合并词语替换标注信息中连续标签为属性名的词语，以对标注信息进行更新，和/或根据标注信息中的词语的候选数据库实体，合并标注信息中连续标签为属性值的词语，得到第二合并词语，第二合并词语为标注信息中连续标签为属性值的词语的候选数据库实体的交集，使用第二合并词语替换标注信息中连续标签为属性值的词语，以对标注信息进行更新，其中，处理器410根据更新后的标注信息生成K个查询条件，包括根据更新后的标注信息生成查询目标。

[0167] 可选地，作为另一实施例，处理器410根据标注信息生成M个候选查询条件，M个候选查询条件中的每个候选查询条件包括第一候选词语、操作符和第二候选词语的对应关系，其中第一候选词语的标签为属性名，第二候选词语的标签为属性值；确定每个候选查询条件的第一候选词语和第二候选词语的匹配指数；将M个候选查询条件中的匹配指数大于预设阈值的K个候选查询条件确定为K个查询条件。

[0168] 进一步地，作为另一实施例，处理器410根据标注信息生成M个初始候选查询条件；根据用户信息，对M个初始候选查询条件进行消歧处理，得到M个候选查询条件，消歧处理包括根据用户信息消除M个初始候选查询条件中存在歧义的初始候选查询条件中的歧义，其中，用户信息包括终端设备的硬件信息、终端系统的软件信息、保存在终端内存或者存储设备上的用户数据、用户的历史操作和用户的设定中的至少一种。

[0169] 进一步地，作为另一实施例，处理器410根据第一候选词语和第二候选词语的配对概率、序列距离、数据库数据类型匹配度和语言习惯约束中的至少一种确定匹配指数。

[0170] 具体地，作为另一实施例，配对概率由第一候选词语所对应的数据库实体与第二候选词语所对应的数据库实体之间的交集决定，其中，第一候选词语所对应的数据库实体与第二候选词语所对应的数据库实体之间的交集越少，配对概率越大，匹配指数越大。

[0171] 具体地，作为另一实施例，序列距离由第一候选词语和第二候选词语在标注信息或查询语句中的距离决定，其中，第一候选词语和第二候选词语在标注信息或查询语句中的距离越大，序列距离越大，匹配指数越小，标注信息或查询语句中第一候选词语和第二候选词语之间的词语的多少，表示距离的大小。

[0172] 具体地，作为另一实施例，数据库数据类型匹配度由第一候选词语和第二候选词语的数据库数据类型是否一致决定，其中，第一候选词语和第二候选词语的数据类型一致时的数据库数据类型匹配度大于第一候选词语和第二候选词语的数据类型不一致时的数据库数据类型匹配度，匹配指数与数据库类型匹配度正相关。

[0173] 具体地，作为另一实施例，语言习惯约束由第一候选词语和第二候选词语是否符合数据库或语言习惯决定，其中，第一候选词语和第二候选词语符合数据库或语言习惯时的语言习惯约束小于第一候选词语和第二候选词语不符合数据库或语言习惯时的语言习惯约束，匹配指数与语言习惯约束负相关。

[0174] 可选地，作为另一实施例，处理器 410 确定标注信息中的标签为属性名的词语满足预设条件和 / 或为孤点词语，其中，孤点词语没有对应的标签为属性值的词语；将标注信息中的标签为属性名的词语的属性名作为查询目标。

[0175] 应注意，图 4 所示的数据库查询的设备 400 与图 3 所示的数据库查询的设备 300 相对应，能够实现图 1- 图 2 的方法实施例中由数据库查询的设备完成的各个过程。数据库查询的设备 400 的其他功能和操作可以参考图 1 和图 2 的方法实施例中涉及数据库查询的设备的各个过程。为避免重复，此处不再详述。

[0176] 应理解，说明书通篇中提到的“一个实施例”或“一实施例”意味着与实施例有关的特定特征、结构或特性包括在本发明的至少一个实施例中。因此，在整个说明书各处出现的“在一个实施例中”或“在一实施例中”未必一定指相同的实施例。此外，这些特定的特征、结构或特性可以任意适合的方式结合在一个或多个实施例中。应理解，在本发明的各种实施例中，上述各过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

[0177] 另外，本文中术语“系统”和“网络”在本文中常被可互换使用。本文中术语“和 / 或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和 / 或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

[0178] 应理解，在本发明实施例中，“与 A 相应的 B”表示 B 与 A 相关联，根据 A 可以确定 B。但还应理解，根据 A 确定 B 并不意味着仅仅根据 A 确定 B，还可以根据 A 和 / 或其它信息确定 B。

[0179] 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

[0180] 所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0181] 在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另外，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口、装置或单元的间接耦合或通信连接，也可以是电的，机械的或其它的形式连接。

[0182] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本发明实施例方案的目的。

[0183] 另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以

是各个单元单独物理存在，也可以是两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

[0184] 通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到本发明可以用硬件实现，或固件实现，或它们的组合方式来实现。当使用软件实现时，可以将上述功能存储在计算机可读介质中或作为计算机可读介质上的一个或多个指令或代码进行传输。计算机可读介质包括计算机存储介质和通信介质，其中通信介质包括便于从一个地方向另一个地方传送计算机程序的任何介质。存储介质可以是计算机能够存取的任何可用介质。以此为例但不限于：计算机可读介质可以包括 RAM、ROM、EEPROM、CD-ROM 或其他光盘存储、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质。此外。任何连接可以适当的成为计算机可读介质。例如，如果软件是使用同轴电缆、光纤光缆、双绞线、数字用户线 (DSL) 或者诸如红外线、无线电和微波之类的无线技术从网站、服务器或者其他远程源传输的，那么同轴电缆、光纤光缆、双绞线、DSL 或者诸如红外线、无线和微波之类的无线技术包括在所属介质的定影中。如本发明所使用的，盘 (Disk) 和碟 (disc) 包括压缩光碟 (CD)、激光碟、光碟、数字通用光碟 (DVD)、软盘和蓝光光碟，其中盘通常磁性的复制数据，而碟则用激光来光学的复制数据。上面的组合也应当包括在计算机可读介质的保护范围之内。

[0185] 总之，以上所述仅为本发明技术方案的较佳实施例而已，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

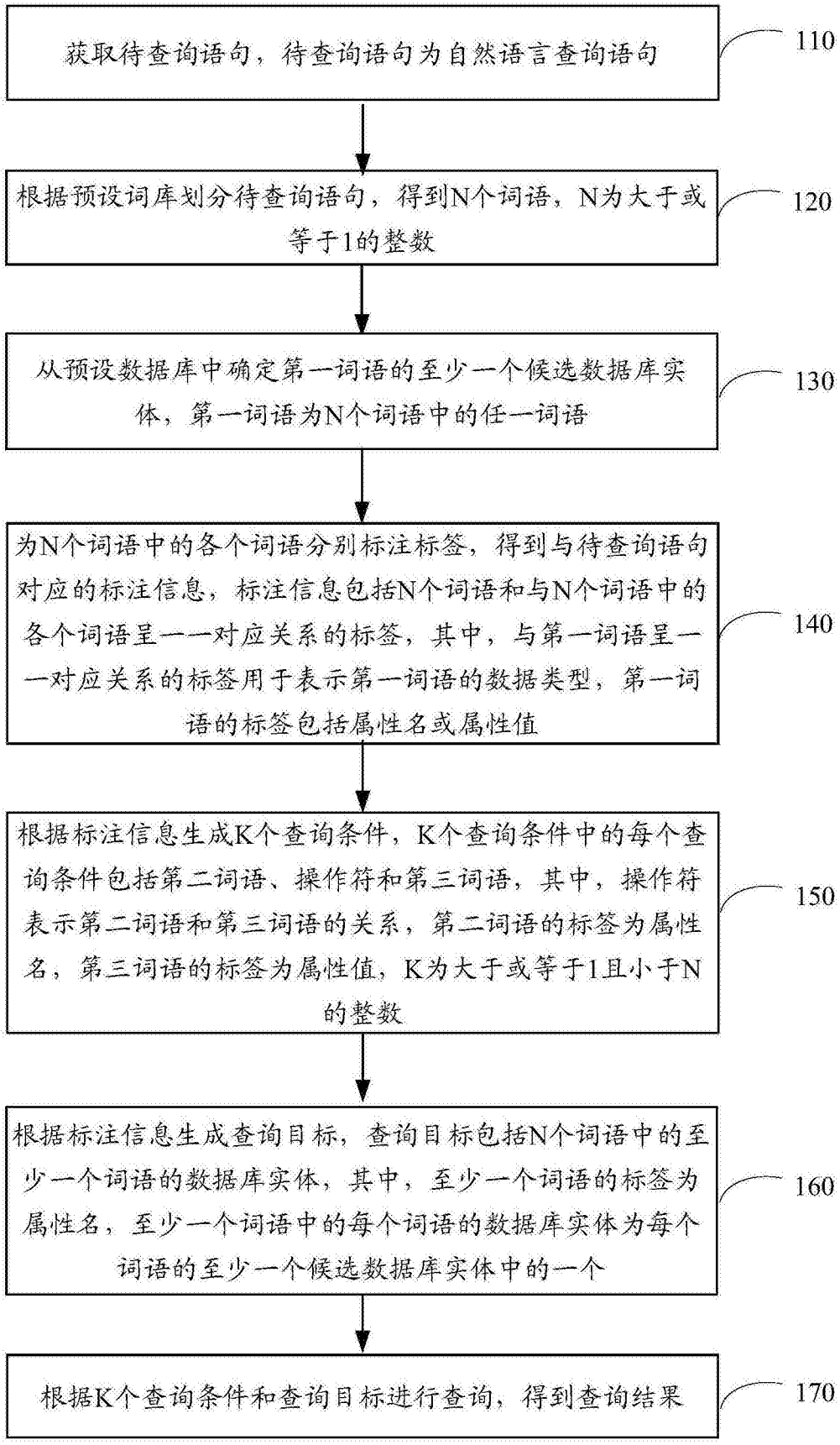

获取待查询语句，待查询语句为自然语言查询语句
110
根据预设词库划分待查询语句，得到N个词语，N为大于或等于1的整数
120
从预设数据库中确定第一词语的至少一个候选数据库实体，第一词语为N个词语中的任一词语
130
为N个词语中的各个词语分别标注标签，得到与待查询语句对应的标注信息，标注信息包括N个词语和与N个词语中的各个词语呈一一对应关系的标签，其中，与第一词语呈一一对应关系的标签用于表示第一词语的数据类型，第一词语的标签包括属性名或属性值
140
根据标注信息生成K个查询条件，K个查询条件中的每个查询条件包括第二词语、操作符和第三词语，其中，操作符表示第二词语和第三词语的关系，第二词语的标签为属性名，第三词语的标签为属性值，K为大于或等于1且小于N的整数
150
根据标注信息生成查询目标，查询目标包括N个词语中的至少一个词语的数据库实体，其中，至少一个词语的标签为属性名，至少一个词语中的每个词语的数据库实体为每个词语的至少一个候选数据库实体中的一个
160
根据K个查询条件和查询目标进行查询，得到查询结果
170

图 1

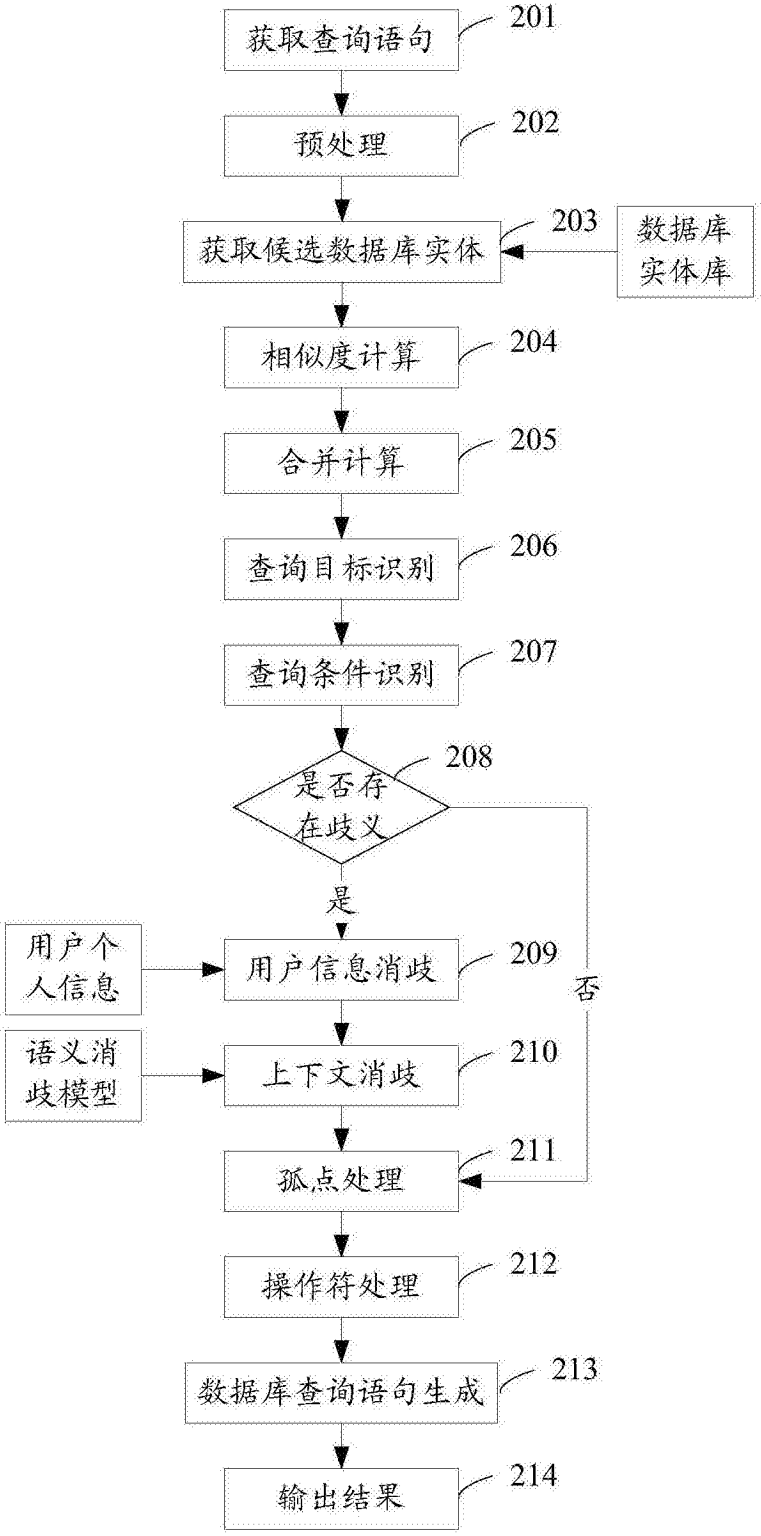

获取查询语句
201
预处理
202
获取候选数据库实体
203
数据库实体库
相似度计算
204
合并计算
205
查询目标识别
206
查询条件识别
207
是否存在歧义
208
是
否
用户个人信息
用户信息消歧
209
语义消歧模型
上下文消歧
210
孤点处理
211
操作符处理
212
数据库查询语句生成
213
输出结果
214

图 2

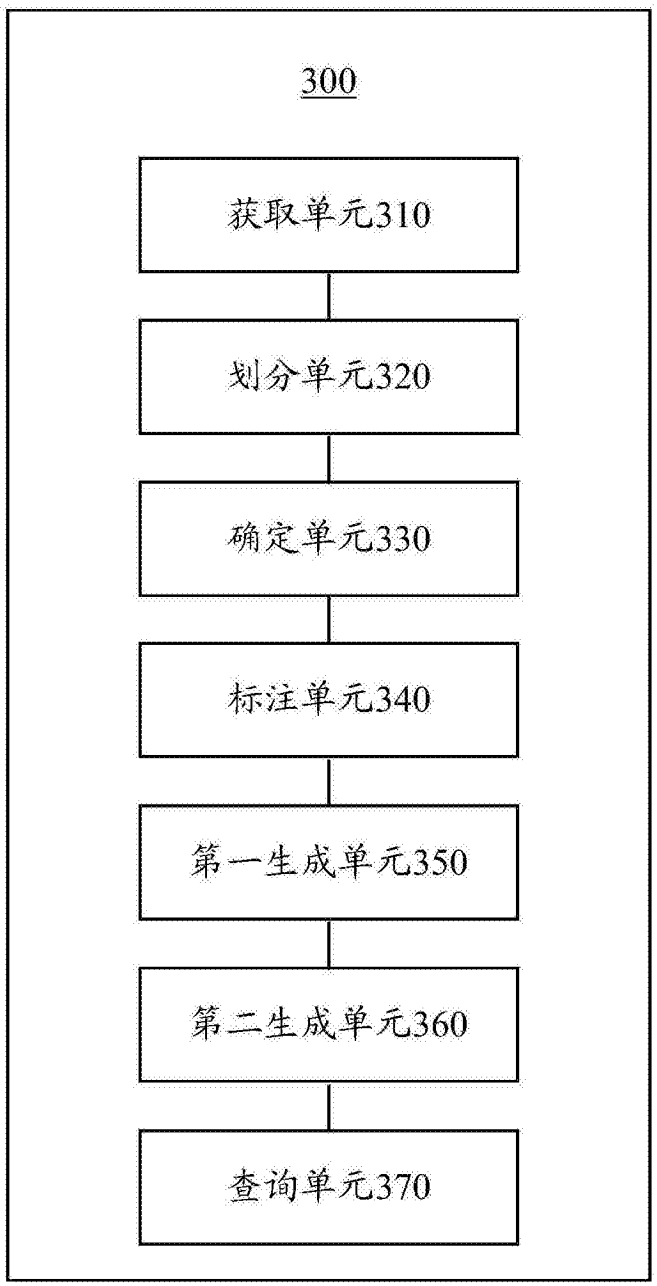

300
获取单元310
划分单元320
确定单元330
标注单元340
第一生成单元350
第二生成单元360
查询单元370

图 3

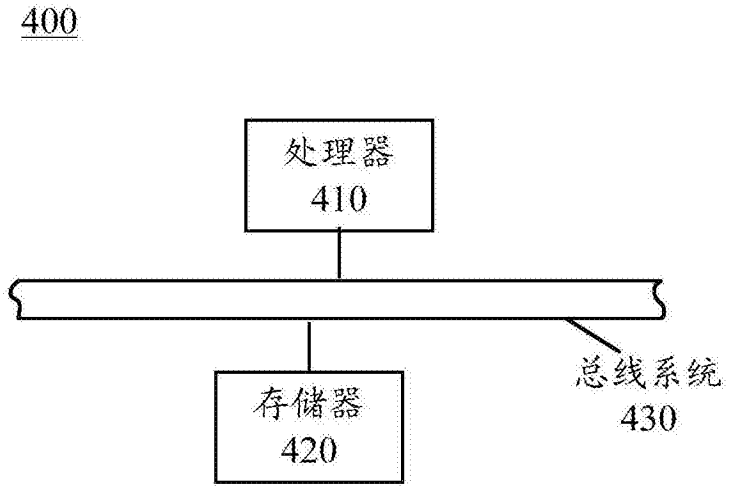


图 4