(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112528001 B
(45)授权公告日 2023.07.25

(21)申请号 202011538686.1

(22)申请日 2020.12.23

(65)同一申请的已公布的文献号
申请公布号 CN 112528001 A

(43)申请公布日 2021.03.19

(73)专利权人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街10号百度大厦2层

(72)发明人 陈万顺

(74)专利代理机构 北京清亦华知识产权代理事务所(普通合伙) 11201
专利代理师 韩海花

(51)Int.Cl.
G06F 16/332(2019.01)
G06F 16/33(2019.01)
G06F 16/338(2019.01)
G06F 40/216(2020.01)
G06F 40/289(2020.01)
G06F 40/30(2020.01)
G06N 20/00(2019.01)

(56)对比文件
CN 111159409 A,2020.05.15
CN 111708800 A,2020.09.25
CN 109918680 A,2019.06.21
CN 111666372 A,2020.09.15
CN 110991180 A,2020.04.10
CN 108763402 A,2018.11.06
CN 110991185 A,2020.04.10
CN 111027324 A,2020.04.17
KR 20010103151 A,2001.11.23
彭景海.基于问答式语义检索系统中对用户提问处理研究!.电子技术与软件工程.2013,(18),全文.

审查员 赵鹏翔

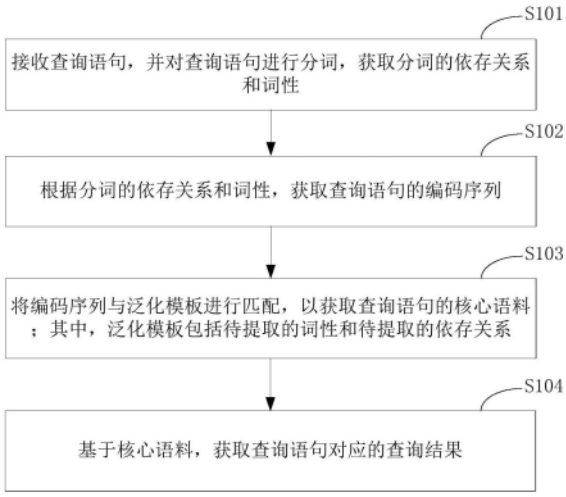
权利要求书2页 说明书12页 附图10页

(54)发明名称

一种信息查询方法、装置及电子设备

(57)摘要

本申请公开了一种信息查询方法、装置及电子设备，涉及深度学习、自然语言处理及人工智能技术领域。该方案为：接收查询语句，并对所述查询语句进行分词，获取所述分词的依存关系和词性；根据所述分词的所述依存关系和所述词性，获取所述查询语句的编码序列；将所述编码序列与泛化模板进行匹配，以获取所述查询语句的核心语料；其中，所述泛化模板包括待提取的词性和待提取的依存关系；基于所述核心语料，获取所述查询语句对应的查询结果。本申请不再依赖海量业务场景数据的积累即可提升泛化能力、确保能够准确、高效地进行信息查询，提高了信息查询过程中的效率和可靠性。同时，能够支持多种不同业务场景下的信息查询，扩展能力强、通用性高。



CN 112528001 B

1.一种信息查询方法,包括:

接收查询语句,并对所述查询语句进行分词,获取所述分词的依存关系和词性;

根据所述分词的所述依存关系和所述词性,获取所述查询语句的编码序列;

将所述编码序列与泛化模板进行匹配,以获取所述查询语句的核心语料;其中,所述泛化模板包括待提取的词性和待提取的依存关系;

基于所述核心语料,获取所述查询语句对应的查询结果;

其中,所述将所述编码序列与泛化模板进行匹配,以获取所述查询语句的核心语料,包括:

根据所述待提取的词性,从所述编码序列中提取与所述待提取词性一致的编码片段;

根据所述待提取的依存关系,确定所述编码片段之间的泛化边界,并按照所述泛化边界,提取所述核心语料。

2.根据权利要求1所述的信息查询方法,其中,所述基于所述核心语料,获取所述查询语句对应的查询结果,包括:

基于所述核心语料,在预先构建的语料数据库中进行检索,以获取所述核心语料对应的目标种子语料;

将所述目标种子语料对应的标签数据,作为所述查询语句对应的查询结果。

3.根据权利要求2所述的信息查询方法,其中,所述基于所述核心语料,在预先构建的语料数据库中进行检索,以获取所述核心语料对应的目标种子语料,包括:

基于所述核心语料,在所述语料数据库中进行检索,以获取所述核心语料对应的至少一个候选种子语料,其中,所述语料数据库存储多个种子语料,其中,所述核心语料为至少一个种子语料的泛化语料;

从所述至少一个候选种子语料中,确定所述目标种子语料。

4.根据权利要求3所述的信息查询方法,其中,所述基于所述核心语料,在所述语料数据库中进行检索,以获取所述核心语料对应的至少一个候选种子语料,包括:

基于倒排索引和语义相似度计算,在所述语料数据库中对所述核心语料进行语料检索,以获取所述至少一个候选种子语料。

5.根据权利要求3所述的信息查询方法,其中,所述从所述至少一个候选种子语料中,确定所述目标种子语料,包括:

获取所述核心语料与每个候选种子语料之间的相似度,选取相似度最高的候选种子语料作为所述目标种子语料。

6.根据权利要求3-5任一项所述的信息查询方法,其特征在于,还包括:

获取所述候选种子语料的数量;

响应于所述候选种子语料的数量大于第一预设数量或者小于第二预设数量,更新所述泛化模板。

7.一种信息查询装置,包括:

第一获取模块,用于接收查询语句,并对所述查询语句进行分词,获取所述分词的依存关系和词性;

第二获取模块,用于根据所述分词的所述依存关系和所述词性,获取所述查询语句的编码序列;

第三获取模块，用于将所述编码序列与泛化模板进行匹配，以获取所述查询语句的核心语料；其中，所述泛化模板包括待提取的词性和待提取的依存关系；

第四获取模块，用于基于所述核心语料，获取所述查询语句对应的查询结果息；

其中，所述第三获取模块，包括：

第一提取子模块，用于根据所述待提取的词性，从所述编码序列中提取与所述待提取词性一致的编码片段；

第二提取子模块，用于根据所述待提取的依存关系，确定所述编码片段之间的泛化边界，并按照所述泛化边界，提取所述核心语料。

8.根据权利要求7所述的信息查询装置，其中，所述第四获取模块，包括：

第一获取子模块，用于基于所述核心语料，在预先构建的语料数据库中进行检索，以获取所述核心语料对应的目标种子语料；

确定子模块，用于将所述目标种子语料对应的标签数据，作为所述查询语句对应的查询结果。

9.根据权利要求8所述的信息查询装置，其中，所述第一获取子模块，包括：

获取单元，用于基于所述核心语料，在所述语料数据库中进行检索，以获取所述核心语料对应的至少一个候选种子语料，其中，所述语料数据库存储多个种子语料，其中，所述核心语料为至少一个种子语料的泛化语料；

确定单元，用于从所述至少一个候选种子语料中，确定所述目标种子语料。

10.根据权利要求9所述的信息查询装置，其中，所述获取单元，包括：

获取子单元，用于基于倒排索引和语义相似度计算，在所述语料数据库中对所述核心语料进行语料检索，以获取所述至少一个候选种子语料。

11.根据权利要求9所述的信息查询装置，其中，所述确定单元，包括：

选取子单元，用于获取所述核心语料与每个候选种子语料之间的相似度，选取相似度最高的候选种子语料作为所述目标种子语料。

12.根据权利要求9-11任一项所述的信息查询装置，其特征在于，还包括：

第五获取模块，用于获取所述候选种子语料的数量；

更新模块，用于响应于所述候选种子语料的数量大于第一预设数量或者小于第二预设数量，更新所述泛化模板。

13. 一种电子设备，其特征在于，包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行权利要求1-6中任一项所述的信息查询方法。

14.一种存储有计算机指令的非瞬时计算机可读存储介质，其特征在于，所述计算机指令用于使所述计算机执行权利要求1-6中任一项所述的信息查询方法。

一种信息查询方法、装置及电子设备

技术领域

[0001] 本申请的实施例总体上涉及数据处理技术领域，并且更具体地涉及深度学习、自然语言处理及人工智能技术领域。

背景技术

[0002] 近年来，AI(Artificial Intelligence，人工智能)技术正在蓬勃发展，随之而来，基于AI技术的智能化信息查询研究，更是受到了越来越多的关注。其中，结构化知识问答正逐渐变成智能化场景中必不可少的一环，针对查询query(口语句子)等口语对话场景则为智能化信息查询方式中极为常见的一种。

[0003] 然而，基于语言表达的丰富性，同一个含义的query通常会有多种不同的表达，由此，根据现有信息查询方法进行查询，为了确保查询结果的准确性，往往存在耗时久、成本高等问题，这样一来，势必导致信息查询过程中存在效率极低的问题，进而导致用户体验极差。因此，如何在确保信息查询结果的准确性，提高信息查询过程中的效率和可靠性，已成为了重要的研究方向之一。

发明内容

[0004] 本申请提供了一种信息查询方法、装置及电子设备。

[0005] 根据第一方面，提供了一种信息查询方法，包括：

[0006] 接收查询语句，并对所述查询语句进行分词，获取所述分词的依存关系和词性；

[0007] 根据所述分词的所述依存关系和所述词性，获取所述查询语句的编码序列；

[0008] 将所述编码序列与泛化模板进行匹配，以获取所述查询语句的核心语料；其中，所述泛化模板包括待提取的词性和待提取的依存关系；

[0009] 基于所述核心语料，获取所述查询语句对应的查询结果。

[0010] 根据第二方面，提供了一种信息查询装置，包括：

[0011] 第一获取模块，用于接收查询语句，并对所述查询语句进行分词，获取所述分词的依存关系和词性；

[0012] 第二获取模块，用于根据所述分词的所述依存关系和所述词性，获取所述查询语句的编码序列；

[0013] 第三获取模块，用于将所述编码序列与泛化模板进行匹配，以获取所述查询语句的核心语料；其中，所述泛化模板包括待提取的词性和待提取的依存关系；

[0014] 第四获取模块，用于基于所述核心语料，获取所述查询语句对应的查询结果息。

[0015] 根据第三方面，提供了一种电子设备，包括：至少一个处理器；以及与所述至少一个处理器通信连接的存储器；其中，所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行本申请第一方面所述的信息查询方法。

[0016] 根据第四方面，提供了一种存储有计算机指令的非瞬时计算机可读存储介质，所

述计算机指令用于使所述计算机执行本申请第一方面所述的信息查询方法。

[0017] 根据第五方面，提供了一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现根据本公开第一方面所述的信息查询方法的步骤。

[0018] 应当理解，本部分所描述的内容并非旨在标识本申请的实施例的关键或重要特征，也不用于限制本申请的范围。本申请的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0019] 附图用于更好地理解本方案，不构成对本申请的限定。其中：

[0020] 图1是根据本申请第一实施例的示意图；

[0021] 图2是一种用户不同提问方式的示意图；

[0022] 图3是一种获取查询语句的编码序列的示意图；

[0023] 图4是一种获取查询语句的核心语料的示意图；

[0024] 图5是一种获取查询语句对应的查询结果的示意图；

[0025] 图6是根据本申请第二实施例的示意图；

[0026] 图7是根据本申请第三实施例的示意图；

[0027] 图8是根据本申请第四实施例的示意图；

[0028] 图9是根据本申请第五实施例的示意图；

[0029] 图10是根据本申请第六实施例的示意图；

[0030] 图11是一种智能语音客户服务交互过程的示意图；

[0031] 图12是用来实现本申请实施例的信息查询方法的信息查询装置的框图；

[0032] 图13是用来实现本申请实施例的信息查询方法的信息查询装置的框图；

[0033] 图14是用来实现本申请实施例的信息查询的电子设备的框图。

具体实施方式

[0034] 以下结合附图对本申请的示范性实施例做出说明，其中包括本申请实施例的各种细节以助于理解，应当将它们认为仅仅是示范性的。因此，本领域普通技术人员应当认识到，可以对这里描述的实施例做出各种改变和修改，而不会背离本申请的范围和精神。同样，为了清楚和简明，以下的描述中省略了对公知功能和结构的描述。

[0035] 以下对本申请的方案涉及的技术领域进行简要说明：

[0036] 数据处理(DataProcessing)，包括对数据的采集、存储、检索、加工、变化和传输等处理，旨从大量的、可能是杂乱无章的、难以理解的数据中抽取并推导出对于某些特定的用户来说有价值、有意义的数据。

[0037] AI(Artificial Intelligence，人工智能)，是研究使计算机来模拟人生的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科，既有硬件层面的技术，也有软件层面的技术。人工智能硬件技术一般包括计算机视觉技术、语音识别技术、自然语言处理技术以及及其学习/深度学习、大数据处理技术、知识图谱技术等几大方面。

[0038] DL(Deep Learning，深度学习)，是ML机器学习(Machine Learning，机器学习)领域中一个新的研究方向，它被引入机器学习使其更接近于最初的目标——人工智能。深度学习是学习样本数据的内在律和表示层次，这些学习过程中获得的信息对诸如文字，图像

和声音等数据的解释有很大的帮助。它的最终目标是让机器能够像人一样具有分析学习能力，能够识别文字、图像和声音等数据。深度学习是一个复杂的机器学习算法，在语音和图像识别方面取得的效果，远远超过先前相关技术。

[0039] NLP(Natural Language Processing，自然语言处理)，是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。因此，这一领域的研究将涉及自然语言，即人们日常使用的语言，所以它与语言学的研究有着密切的联系，但又有重要的区别。

[0040] 需要说明的是，针对结构化知识问答研究，最大的难题即如何解决语义泛化问题。现有技术中，通常基于X-SQL、HydraNet等模型，通过深度学习的方式解决语言表达的泛化问题。

[0041] 然而采用深度学习的方式解决语言表达的泛化问题，往往需要标注海量数据，尽可能地构建一个完善的数据集合，从而为模型提供更多的有效训练样本，以便提升模型的识别能力，这样一来，势必导致现有信息查询过程中存在扩展能力不强，通用性不高、成本高、耗时久、结果可控性差，不易干预等问题。

[0042] 由此，本申请提出的信息查询方式，能够通过基于依存句法分析的泛化匹配技术，在确保查询结果准确性的同时，降低了成本、提高了信息查询过程中的效率和可靠性。

[0043] 下面参考附图描述本申请实施例的信息查询方法、装置及电子设备。

[0044] 图1是根据本申请第一实施例的示意图。其中，需要说明的是，本实施例的信息查询方法的执行主体为服务端。如图1所示，本实施例提出的信息查询方法，包括如下步骤：

[0045] S101、接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性。

[0046] 需要说明的是，基于语言表达的丰富性，同一个含义的query(口语句子)通常会有多种不同的表达。举例而言，获取到的查询语句分别为“车辆甲的发动机马力”与“车辆甲的发动机功率”，此种情况下，“发动机的马力”与“发动机的功率”为“发动机功率”的两种常见的泛化表达。

[0047] 本申请中，为了解决信息查询过程中存在的泛化问题，不再依赖海量标注数据实现查询结果的获取，基于依存句法分析的泛化匹配技术，获取与接收到的查询语句匹配的答案，以实现信息的查询。其中，接收到的查询语句可以为用户所提出的问题。

[0048] 需要说明的是，本申请中，用户可以通过输入文本信息、输入语音信息或者选择文本信息等多种方式进行提问。例如，如图2(a)所示，用户可以于用户终端的输入界面上输入文本，此种情况下，用户的问题即为文本信息1-1；又例如，如图2(b)所示，用户可以于用户终端的输入界面上输入语音，此种情况下，用户的问题即为语音信息1-2；再例如，如图2(c)所示，用户可以于用户终端的输入界面上选择界面提供的文本，此种情况下，用户的问题即为文本信息1-3。

[0049] 本申请实施例中，在用户提出问题之后，可以将文本信息直接作为查询语句，也可以将语音信息转化为文本信息后，将转化得到的文本信息作为查询语句。

[0050] 举例而言，用户以语音输入的方式提出了如下问题：“车辆甲的发动机马力”，此种情况下，在用户提出问题之后，可以将语音信息转化为文本信息：“车辆甲的发动机马力”，并将“车辆甲的发动机马力”作为查询语句。

[0051] 进一步地，在接收到查询语句后，可以对查询语句进行分词处理及词性标注，以获取分词的依存关系和词性。

[0052] 其中，分词处理，指的是基于统计的分词处理，统计的样本内容来自于一些标准的语料库。例如，对“车辆甲的发动机马力”进行分词处理，能够得到车辆甲、的、发动机、马力，共4个分词。

[0053] 其中，依存关系，可以包括：主谓关系(Subject Verb，简称SBV)、动宾关系(Verb Object，简称VOB)、间宾关系(Indirect Object，简称IOB)、定中关系(Attribute，简称ATT)、状中关系(Adverbial，简称ADV)等关系。可选地，依存关系，可以为每两个分词之间的依存关系。例如，发动机与马力之间的依存关系为ATT。

[0054] 其中，词性，指的是以词的特点作为划分词类的根据，可以包括：动词(v)、名词(n)、其他专有名词(nz)、助词(u)等。

[0055] 需要说明的是，本申请中对于进行分词处理的具体方法不做限定，可以根据实际情况进行设定。例如，可以采用正向最大匹配法、反向最大匹配法、最短路径分词法等方法对文本内容进行分词处理。

[0056] 需要说明的是，本申请中对于进行词性标注的具体方法不做限定，可以根据实际情况进行设定。例如，可以将获取到的分词输入马尔可夫模型(Markov Model)中，以获取分词的词性。

[0057] S102、根据分词的依存关系和词性，获取查询语句的编码序列。

[0058] 本申请实施例中，在获取分词的依存关系和词性后，可以将分词的依存关系和词性作为Encode(编码)规则，以获取查询语句的编码序列。

[0059] 举例而言，如图3所示，在接收到查询语句“车辆甲的发动机马力”后，可以对查询语句进行分词处理及词性标注，获取到分词的依存关系分别为：[DE]、[DE]、[ATT]和[SBV]，词性分别为，[nz]、[u]、[n]和[n]。此种情况下，可以将分词的依存关系和词性作为Encode规则，以获取查询语句的编码序列为：[DE][nz][DE][u][ATT][n][SBV][n]。其中，[DE]表示“的字关系”。

[0060] S103、将编码序列与泛化模板进行匹配，以获取查询语句的核心语料；其中，泛化模板包括待提取的词性和待提取的依存关系。

[0061] 本申请实施例中，在获取到查询语句的编码序列后，可以将编码序列与泛化模板进行匹配，并将匹配的泛化模板作为Decode(解码)规则，以获取查询语句的核心语料。也就是说，在将编码序列与泛化模板进行匹配后，可以通过转码等方式，获取查询语句的核心语料。其中，泛化模板中存储有待提取的依存关系和待提取的词性。

[0062] 举例而言，如图4所示，获取查询语句的编码序列为[DE][nz][DE][u][ATT][n][SBV][n]，此种情况下，可以将获取到的编码序列与泛化模板进行匹配，并将匹配的泛化模板A作为Decode规则，以获取到查询语句的核心语料为：【车辆甲】、【发动机马力】。

[0063] 需要说明的是，本申请中，为了提升泛化能力、确保能够高效地进行信息查询，进而支持不同的业务场景，可以预先设置多个根据经验配置得到的泛化模板。

[0064] 举例而言，针对业务场景甲，可以预先设置泛化模板B为[DE][nz][DE][u][ATT][n][SBV][n]，此种情况下，针对查询语句“车辆甲的发动机马力”，可以获取到核心语料为：【车辆甲】、【发动机马力】；针对业务场景乙，可以预先设置泛化模板C为[ATT][n][DE][u]

[SBV][n]，此种情况下，针对查询语句“腐烂的西红柿”，可以获取到核心语料为：【烂】、【西红柿】。

[0065] S104、基于核心语料，获取查询语句对应的查询结果。

[0066] 举例而言，如图5所示，获取到查询语句的核心语料5-1为：【车辆甲】、【发动机马力】，此种情况下，可以获取到查询语句对应的查询结果5-2为【车辆甲发动机功率为A瓦特】。

[0067] 根据本申请实施例的信息查询方法，可以通过接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性，并根据分词的依存关系和词性，获取查询语句的编码序列，然后将编码序列与泛化模板进行匹配，以获取查询语句的核心语料，进而基于核心语料，获取查询语句对应的查询结果，以实现信息的查询，不再依赖海量业务场景数据的积累即可提升泛化能力、确保能够准确、高效地进行信息查询，进而提高了信息查询过程中的效率和可靠性。同时，在不增加过多成本的情况下，能够支持多种不同业务场景下的信息查询，扩展能力强、通用性高。

[0068] 图6是根据本申请第二实施例的示意图。如图6所示，在上一实施例的基础上，本实施例提出的信息查询方法，包括如下步骤：

[0069] S601、接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性。

[0070] S602、根据分词的依存关系和词性，获取查询语句的编码序列。

[0071] 该步骤S601～S602与上一实施例中的步骤S101～S102相同，此处不再赘述。

[0072] 上一实施例中的步骤S103具体可包括以下步骤S603～S604。

[0073] S603、根据待提取的词性，从编码信息中提取与待提取词性一致的编码片段。

[0074] 举例而言，预先设置以下泛化模板：[DE][nz][DE][u][ATT][n][SBV][n]、[DE][nz][ATT][n][ATT][n][SBV][n]、[DE][nz][ATT][u][ATT][n][SBV][n]，获取到的待提取的词性为：[nz][u][n][n]。

[0075] 此种情况下，针对“车辆甲的发动机马力”，从编码信息中提取与待提取词性一致的编码片段分别为：[DE][nz][DE][u][ATT][n][SBV][n]和[DE][nz][ATT][u][ATT][n][SBV][n]。

[0076] S604、根据待提取的依存关系，确定编码片段之间的泛化边界，并按照泛化边界，提取核心语料。

[0077] 举例而言，从编码信息中提取与待提取词性一致的编码片段分别为：[DE][nz][DE][u][ATT][n][SBV][n]和[DE][nz][ATT][u][ATT][n][SBV][n]，获取到的待提取的依存关系为：[DE][DE][ATT][SBV]。

[0078] 此种情况下，针对“车辆甲的发动机马力”，根据待提取的依存关系，确定编码片段之间的泛化边界，并按照泛化边界，提取核心语料为与[DE][nz][DE][u][ATT][n][SBV][n]匹配的：【车辆甲】和【发动机马力】。

[0079] S605、基于核心语料，获取查询语句对应的查询结果。

[0080] 该步骤S605与上一实施例中的步骤S104相同，此处不再赘述。

[0081] 根据本申请实施例的信息查询方法，可以根据待提取的词性，提取编码片段，然后根据待提取的依存关系，确定编码片段之间的泛化边界，进而按照泛化边界，提取核心语料，使得能够在解决了泛化问题的同时，缩短了信息查询过程的耗时，进一步提高了信息查

询过程中的效率和可靠性。

[0082] 图7是根据本申请第三实施例的示意图。如图7所示，在上一实施例的基础上，本实施例提出的信息查询方法，包括如下步骤：

[0083] S701、接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性。

[0084] S702、根据分词的依存关系和词性，获取查询语句的编码序列。

[0085] 该步骤S701～S702与上一实施例中的步骤S101～S102相同，此处不再赘述。

[0086] S703、根据待提取的词性，从编码信息中提取与待提取词性一致的编码片段。

[0087] S704、根据待提取的依存关系，确定编码片段之间的泛化边界，并按照泛化边界，提取核心语料。

[0088] 该步骤S703～S704与上一实施例中的步骤S603～S604相同，此处不再赘述。

[0089] 上一实施例中的步骤S104具体可包括以下步骤S705～S706。

[0090] S705、基于核心语料，在预先构建的语料数据库中进行检索，以获取核心语料对应的目标种子语料。

[0091] 作为一种可能的实现方式，如图8所示，在上一实施例的基础上，包括如下步骤：

[0092] S801、基于核心语料，在语料数据库中进行检索，以获取核心语料对应的至少一个候选种子语料，其中，语料数据库存储多个种子语料，其中，核心语料为至少一个种子语料的泛化语料。

[0093] 本申请实施例中，可以基于倒排索引和语义相似度计算，在语料数据库中对核心语料进行语料检索，以通过粗排获取至少一个候选种子语料。

[0094] 其中，语料数据库，为结构化数据存储形式。可选地，可以预先基于用户的表格数据，对表格数据进行内容提取，并利用提取的内容，形成语料数据库。

[0095] 其中，核心语料为至少一个种子语料的泛化语料。例如，【发动机马力】为种子语料【发动机功率】的泛化语料。

[0096] 其中，倒排索引，又称反向索引(Inverted Index)，源于实际应用中需要根据属性的值来查找记录。这种索引表中的每一项都包括一个属性值和具有该属性值的各记录的地址。由于不是由记录来确定属性值，而是由属性值来确定记录的位置，因而称为倒排索引。

[0097] 其中，语义相似度计算，可以采用基于CNN(Convolutional Neural Networks，卷积神经网络)等语义相似度计算方式。

[0098] S802、从至少一个候选种子语料中，确定目标种子语料。

[0099] 本申请实施例中，可以获取核心语料与每个候选种子语料之间的相似度，并选取相似度最高的候选种子语料，以通过精排获取目标种子语料。

[0100] 需要说明的是，本申请对于获取核心语料与每个候选种子语料之间的相似度的具体方式不作限定，可以根据实际情况进行选择。例如，可以通过Simnet(仿真网络)、基于BOW的余弦距离、BM25(Best Matching 25)等方式，保留相似度最高的候选种子语料作为最终的目标种子语料。

[0101] S706、将目标种子语料对应的标签数据，作为查询语句对应的查询结果。

[0102] 举例而言，目标种子语料【车辆甲】和【发动机功率】对应的标签数据分别为【车辆甲】和【发动机功率为A瓦特】，此种情况下，查询语句对应的查询结果为【车辆甲发动机功率为A瓦特】。

[0103] 根据本申请实施例的信息查询方法，可以基于核心语料，在语料数据库中进行检索，以通过粗排获取核心语料对应的至少一个候选种子语料，然后获取核心语料与每个候选种子语料之间的相似度，并选取相似度最高的候选种子语料，以通过精排获取目标种子语料，使得能够通过粗排和精排召回真实的标签数据，进而获取到更加准确地查询结果，进一步提高了信息查询过程中的效率和可靠性。

[0104] 需要说明的是，本申请实施例中，泛化模板可以预先根据领域经验构建。进一步地，为了提升泛化模板的适应性，可以预先针对不同的业务场景，根据核心语料的不同表述方式，配置对应的泛化模板。其中，大部分泛化模板为通用泛化模板，可以适用于多种业务场景，其余泛化模板可以针对特殊领域的需求进行调整，以构建部分定制泛化模板。

[0105] 进一步地，在信息查询的过程中，可能会出现欠召回或者扩召回现象，本申请提出的信息查询方法，可以针对欠召回、扩召回现象，更新泛化模板。

[0106] 作为一种可能的实现方式，如图9所示，在上述实施例的基础上，包括以下步骤：

[0107] S901、获取候选种子语料的数量。

[0108] S902、响应于候选种子语料的数量大于第一预设数量或者小于第二预设数量，更新泛化模板。

[0109] 其中，第一预设数量和第二预设数量可以根据实际情况进行设定。

[0110] 举例而言，候选种子语料的数量为k、第一预设数量为k_1、第二预设数量k_2，此种情况下，若$k>k_1$，说明当前存在扩召回现象，则可以对泛化模板进行更新；若$k<k_2$，说明当前存在欠召回现象，则可以对泛化模板进行更新。

[0111] 根据本申请实施例的信息查询方法，可以获取候选种子语料的数量，并响应于候选种子语料的数量大于第一预设数量或者小于第二预设数量，更新泛化模板，使得能够在出现欠召回或者扩召回现象时，随时更新泛化模板，立即对错误进行干预，确保了获取到的查询结果更加准确，进一步提高了信息查询过程中的效率和可靠性，增强了信息查询过程中的可控性及稳定性。

[0112] 图10是根据本申请第六实施例的示意图。如图10所示，在上述实施例的基础上，本实施例提出的信息查询方法，包括如下步骤：

[0113] S1001、接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性。

[0114] S1002、根据分词的依存关系和词性，获取查询语句的编码序列。

[0115] S1003、根据待提取的词性，从编码信息中提取与待提取词性一致的编码片段。

[0116] S1004、根据待提取的依存关系，确定编码片段之间的泛化边界，并按照泛化边界，提取核心语料。

[0117] S1005、基于核心语料，在语料数据库中进行检索，以获取核心语料对应的至少一个候选种子语料，其中，语料数据库存储多个种子语料，其中，核心语料为至少一个种子语料的泛化语料。

[0118] S1006、从至少一个候选种子语料中，确定目标种子语料。

[0119] S1007、将目标种子语料对应的标签数据，作为查询语句对应的查询结果。

[0120] 需要说明的是，关于步骤S1001～S1007的介绍可参见上述实施例中的相关记载，此处不再赘述。

[0121] 需要说明的是，本申请提出的信息查询方法，可以运用于多种场景中。

[0122] 针对智能语音客户服务等人机交互应用场景，可以基于NLP技术，将用户输入的语音转化为文本信息，并作为查询语句，对查询语句进行分词，获取分词的依存关系和词性，然后可以基于DL以及AI技术，根据分词的依存关系和词性，获取查询语句的编码序列，然后将编码序列与泛化模板进行匹配，以获取查询语句的核心语料，进而基于核心语料，获取查询语句对应的查询结果，以实现信息的查询。

[0123] 举例而言，如图11所示，用户试图通过输入语音“车辆甲的发动机马力”，以进行提问。可选地，通过将语音转化为文本信息“车辆甲的发动机马力”作为查询语句，对查询语句进行分词处理，可以获取分词的依存关系和词性，然后可以基于DL以及AI技术，根据分词的依存关系和词性，获取查询语句的编码序列为[DE][nz][DE][u][ATT][n][SBV][n]。进一步地，可以将编码序列与泛化模板进行匹配，以获取查询语句的核心语料为【车辆甲】、【发动机马力】，进而基于核心语料，获取查询语句对应的查询结果为【车辆甲发动机功率为A瓦特】，以实现信息的查询。

[0124] 根据本申请实施例的信息查询方法，可以通过接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性，并根据分词的依存关系和词性，获取查询语句的编码序列，然后将编码序列与泛化模板进行匹配，以获取查询语句的核心语料，进而基于核心语料，获取查询语句对应的查询结果，以实现信息的查询，不再依赖海量业务场景数据的积累即可提升泛化能力、确保能够准确、高效地进行信息查询，进而提高了信息查询过程中的效率和可靠性。同时，在不增加过多成本的情况下，能够支持多种不同业务场景下的信息查询，扩展能力强、通用性高。

[0125] 与上述几种实施例提供的信息查询方法相对应，本申请的一个实施例还提供一种信息查询装置，由于本申请实施例提供的信息查询装置与上述几种实施例提供的信息查询方法相对应，因此在信息查询方法的实施方式也适用于本实施例提供的信息查询装置，在本实施例中不再详细描述。

[0126] 图12是根据本申请一个实施例的信息查询装置的结构示意图。

[0127] 如图12所示，该信息查询装置1200，包括：第一获取模块1210、第二获取模块1220、第三获取模块1230和第四获取模块1240。其中：

[0128] 第一获取模块1210，用于接收查询语句，并对所述查询语句进行分词，获取所述分词的依存关系和词性；

[0129] 第二获取模块1220，用于根据所述分词的所述依存关系和所述词性，获取所述查询语句的编码序列；

[0130] 第三获取模块1230，用于将所述编码序列与泛化模板进行匹配，以获取所述查询语句的核心语料；其中，所述泛化模板包括待提取的词性和待提取的依存关系；

[0131] 第四获取模块1240，用于基于所述核心语料，获取所述查询语句对应的查询结果息。

[0132] 图13是根据本申请另一个实施例的信息查询装置的结构示意图。

[0133] 如图13所示，该信息查询装置1300，包括：第一获取模块1310、第二获取模块1320、第三获取模块1330和第四获取模块1340。其中：

[0134] 第三获取模块1330，包括：

[0135] 第一提取子模块1331，用于根据所述待提取的词性，从所述编码信息中提取与所

述待提取词性一致的编码片段；

[0136] 第二提取子模块1332，用于根据所述待提取的依存关系，确定所述编码片段之间的泛化边界，并按照所述泛化边界，提取所述核心语料。

[0137] 第四获取模块1340，包括：

[0138] 第一获取子模块1341，用于基于所述核心语料，在预先构建的语料数据库中进行检索，以获取所述核心语料对应的目标种子语料；

[0139] 确定子模块1342，用于将所述目标种子语料对应的标签数据，作为所述查询语句对应的查询结果。

[0140] 第一获取子模块1341，包括：

[0141] 获取单元13411，用于基于所述核心语料，在所述语料数据库中进行检索，以获取所述核心语料对应的至少一个候选种子语料，其中，所述语料数据库存储多个种子语料，其中，所述核心语料为至少一个种子语料的泛化语料；

[0142] 确定单元13412，用于从所述至少一个候选种子语料中，确定所述目标种子语料。

[0143] 获取单元13411，包括：

[0144] 获取子单元134111，用于基于倒排索引和语义相似度计算，在所述语料数据库中对所述核心语料进行语料检索，以获取所述至少一个候选种子语料。

[0145] 确定单元13412，包括：

[0146] 选取子单元134121，用于获取所述核心语料与每个候选种子语料之间的相似度，选取相似度最高的候选种子语料作为所述目标种子语料。

[0147] 该信息查询装置1300，还包括：

[0148] 第五获取模块1350，用于获取所述候选种子语料的数量；

[0149] 更新模块1360，用于响应于所述候选种子语料的数量大于第一预设数量或者小于所述第二预设数量，更新所述泛化模板。

[0150] 需要说明的是，第二获取模块1320与第二获取模块1220具有相同功能和结构。

[0151] 根据本申请实施例的信息查询装置，可以通过接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性，并根据分词的依存关系和词性，获取查询语句的编码序列，然后将编码序列与泛化模板进行匹配，以获取查询语句的核心语料，进而基于核心语料，获取查询语句对应的查询结果，以实现信息的查询，不再依赖海量业务场景数据的积累即可提升泛化能力、确保能够准确、高效地进行信息查询，进而提高了信息查询过程中的效率和可靠性。同时，在不增加过多成本的情况下，能够支持多种不同业务场景下的信息查询，扩展能力强、通用性高。

[0152] 根据本申请的实施例，本申请还提供了一种电子设备和一种可读存储介质。

[0153] 如图14所示，是根据本申请实施例的信息查询的电子设备的框图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本申请的实现。

[0154] 如图14所示，该电子设备包括：一个或多个处理器1410、存储器1420，以及用于连

接各部件的接口，包括高速接口和低速接口。各个部件利用不同的总线互相连接，并且可以被安装在公共主板上或者根据需要以其它方式安装。处理器可以对在电子设备内执行的指令进行处理，包括存储在存储器中或者存储器上以在外部输入/输出装置(诸如，耦合至接口的显示设备)上显示GUI的图形信息的指令。在其它实施方式中，若需要，可以将多个处理器和/或多条总线与多个存储器和多个存储器一起使用。同样，可以连接多个电子设备，各个设备提供部分必要的操作(例如，作为服务器阵列、一组刀片式服务器、或者多处理器系统)。图14中以一个处理器1410为例。

[0155] 存储器1420即为本申请所提供的非瞬时计算机可读存储介质。其中，所述存储器存储有可由至少一个处理器执行的指令，以使所述至少一个处理器执行本申请所提供的信息查询方法。本申请的非瞬时计算机可读存储介质存储计算机指令，该计算机指令用于使计算机执行本申请所提供的信息查询方法。

[0156] 存储器1420作为一种非瞬时计算机可读存储介质，可用于存储非瞬时软件程序、非瞬时计算机可执行程序以及模块，如本申请实施例中的信息查询方法对应的程序指令/模块(例如，附图12所示的第一获取模块1210、第二获取模块1220、第三获取模块1230和第四获取模块1240)。处理器1310通过运行存储在存储器1420中的非瞬时软件程序、指令以及模块，从而执行服务器的各种功能应用以及数据处理，即实现上述方法实施例中的信息查询方法。

[0157] 存储器1420可以包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需要的应用程序；存储数据区可存储根据定位电子设备的使用所创建的数据等。此外，存储器1420可以包括高速随机存取存储器，还可以包括非瞬时存储器，例如至少一个磁盘存储器件、闪存器件、或其他非瞬时固态存储器件。在一些实施例中，存储器1420可选包括相对于处理器1410远程设置的存储器，这些远程存储器可以通过网络连接至定位电子设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0158] 信息查询的电子设备还可以包括：输入装置1430和输出装置1440。处理器1410、存储器1420、输入装置1430和输出装置1440可以通过总线或者其他方式连接，图14中以通过总线连接为例。

[0159] 输入装置1430可接收输入的数字或字符信息，以及产生与定位电子设备的用户设置以及功能控制有关的键信号输入，例如触摸屏、小键盘、鼠标、轨迹板、触摸板、指示杆、一个或者多个鼠标按钮、轨迹球、操纵杆等输入装置。输出装置1440可以包括显示设备、辅助照明装置(例如，LED)和触觉反馈装置(例如，振动电机)等。该显示设备可以包括但不限于，液晶显示器(LCD)、发光二极管(LED)显示器和等离子体显示器。在一些实施方式中，显示设备可以是触摸屏。

[0160] 此处描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、专用ASIC(专用集成电路)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出

装置。

[0161] 这些计算程序(也称作程序、软件、软件应用、或者代码)包括可编程处理器的机器指令,并且可以利用高级过程和/或面向对象的编程语言、和/或汇编/机器语言来实施这些计算程序。如本文使用的,术语“机器可读介质”和“计算机可读介质”指的是用于将机器指令和/或数据提供给可编程处理器的任何计算机程序产品、设备、和/或装置(例如,磁盘、光盘、存储器、可编程逻辑装置(PLD)),包括,接收作为机器可读信号的机器指令的机器可读介质。术语“机器可读信号”指的是用于将机器指令和/或数据提供给可编程处理器的任何信号。

[0162] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0163] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)、互联网以及区块链网络。

[0164] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务端可以是云服务器,又称为云计算服务器或云主机,是云计算服务体系中的一项主机产品,以解决了传统物理主机与VPS服务(“Virtual Private Server”,或简称“VPS”)中,存在的管理难度大,业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0165] 本申请还提供一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时,实现根据本申请第一方面实施例所述的信息查询方法。

[0166] 根据本申请实施例的信息查询方法,可以通过接收查询语句,并对查询语句进行分词,获取分词的依存关系和词性,并根据分词的依存关系和词性,获取查询语句的编码序列,然后将编码序列与泛化模板进行匹配,以获取查询语句的核心语料,进而基于核心语料,获取查询语句对应的查询结果,以实现信息的查询,不再依赖海量业务场景数据的积累即可提升泛化能力、确保能够准确、高效地进行信息查询,进而提高了信息查询过程中的效率和可靠性。同时,在不增加过多成本的情况下,能够支持多种不同业务场景下的信息查询,扩展能力强、通用性高。

[0167] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发申请中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本申请公开的技术方案所期望的结果,本文在此不进行限制。

[0168] 上述具体实施方式，并不构成对本申请保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本申请的精神和原则之内所作的修改、等同替换和改进等，均应包含在本申请保护范围之内。

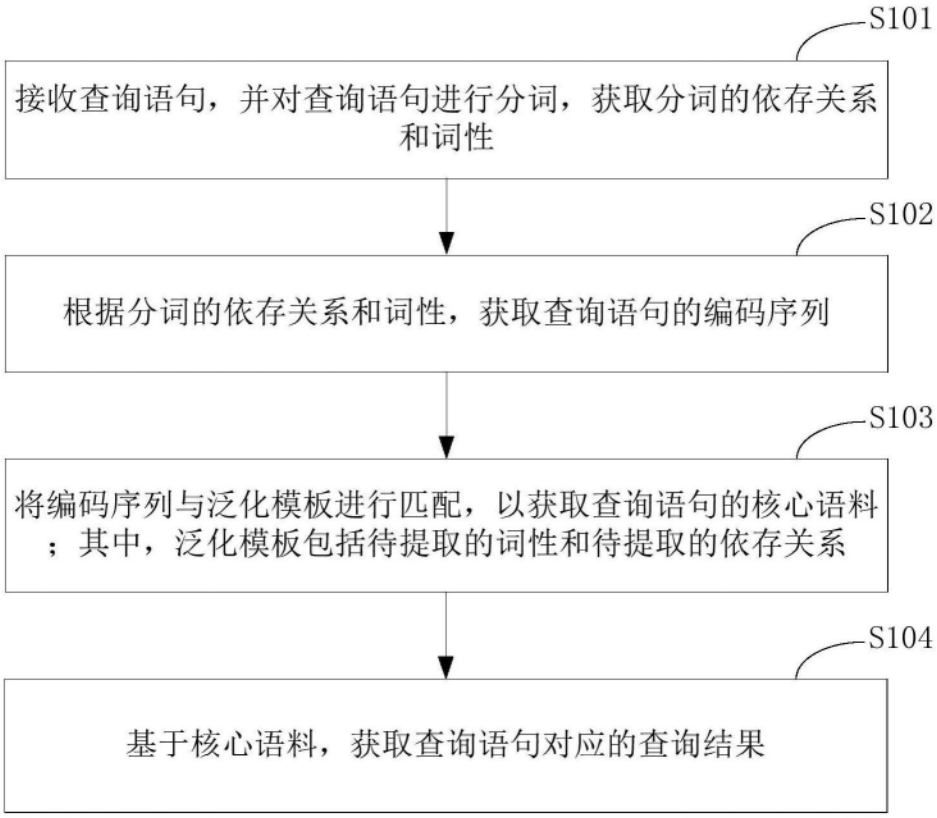

S101
接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性
S102
根据分词的依存关系和词性，获取查询语句的编码序列
S103
将编码序列与泛化模板进行匹配，以获取查询语句的核心语料；其中，泛化模板包括待提取的词性和待提取的依存关系
S104
基于核心语料，获取查询语句对应的查询结果

图1

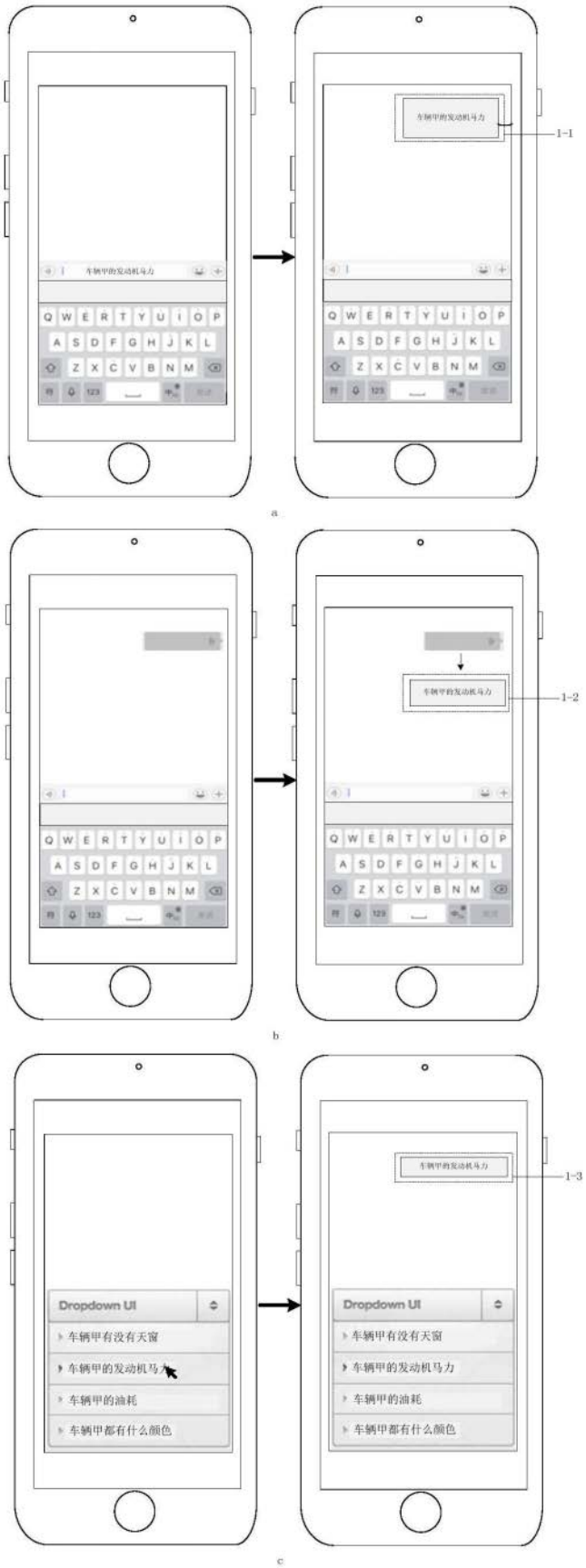

车辆甲的发动机马力
1-1
a
车辆甲的发动机马力
1-2
b
Dropdown UI
车辆甲有没有天窗
车辆甲的发动机马力
车辆甲的油耗
车辆甲都有什么颜色
车辆甲的发动机马力
1-3
c

图2

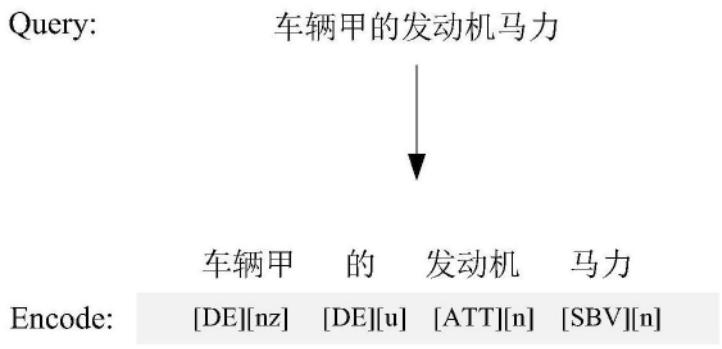

Query:
车辆甲的发动机马力
车辆甲 的 发动机 马力
Encode:
[DE][nz] [DE][u] [ATT][n] [SBV][n]

图3

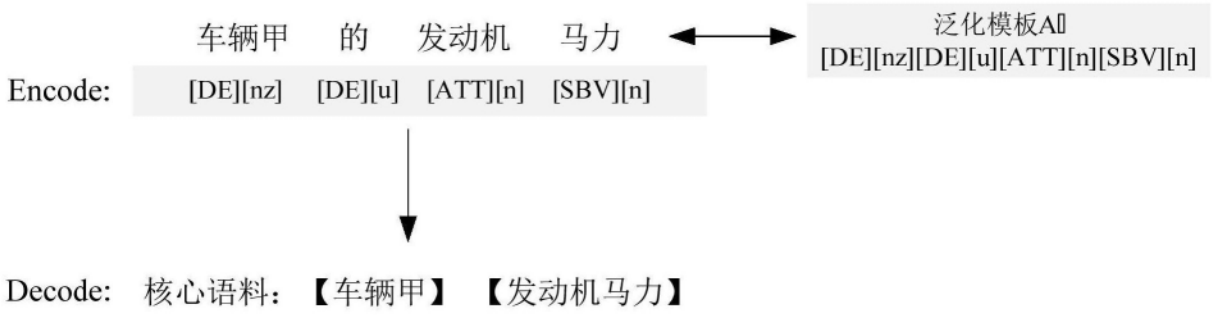

车辆甲 的 发动机 马力
泛化模板A
[DE][nz][DE][u][ATT][n][SBV][n]
Encode:
[DE][nz] [DE][u] [ATT][n] [SBV][n]
Decode: 核心语料：【车辆甲】 【发动机马力】

图4

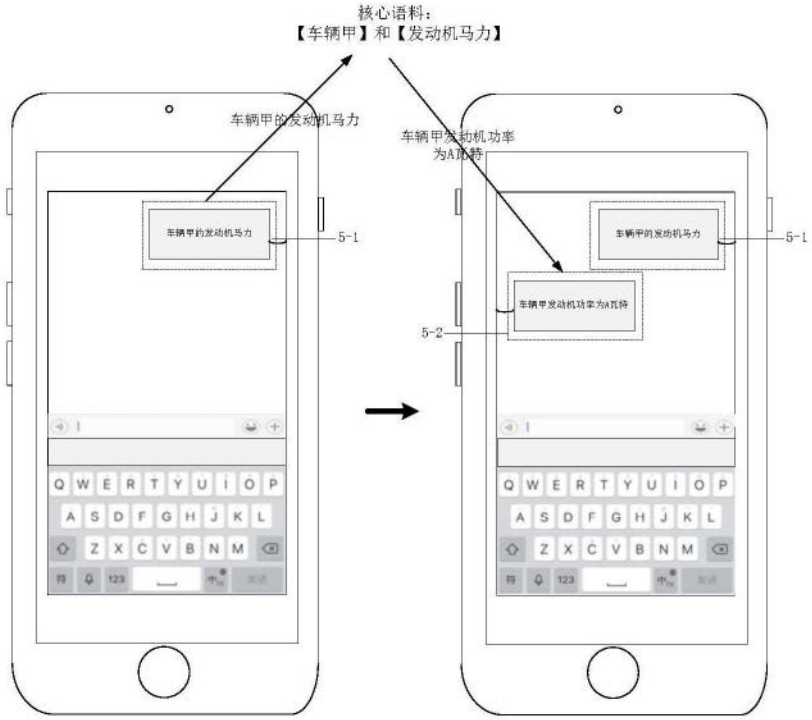

核心语料：
【车辆甲】和【发动机马力】
车辆甲的发动机马力
车辆甲发动机功率为A瓦特
5-1
5-2

图5

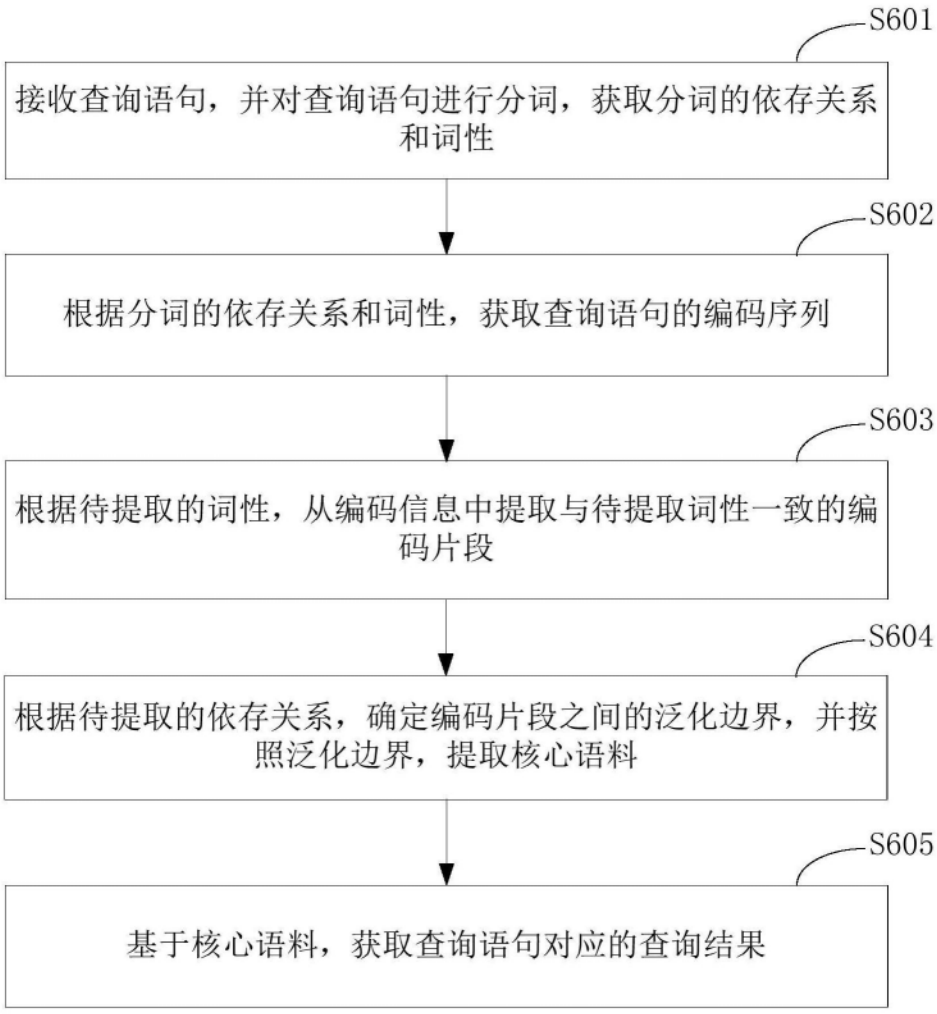

S601
接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性
S602
根据分词的依存关系和词性，获取查询语句的编码序列
S603
根据待提取的词性，从编码信息中提取与待提取词性一致的编码片段
S604
根据待提取的依存关系，确定编码片段之间的泛化边界，并按照泛化边界，提取核心语料
S605
基于核心语料，获取查询语句对应的查询结果

图6

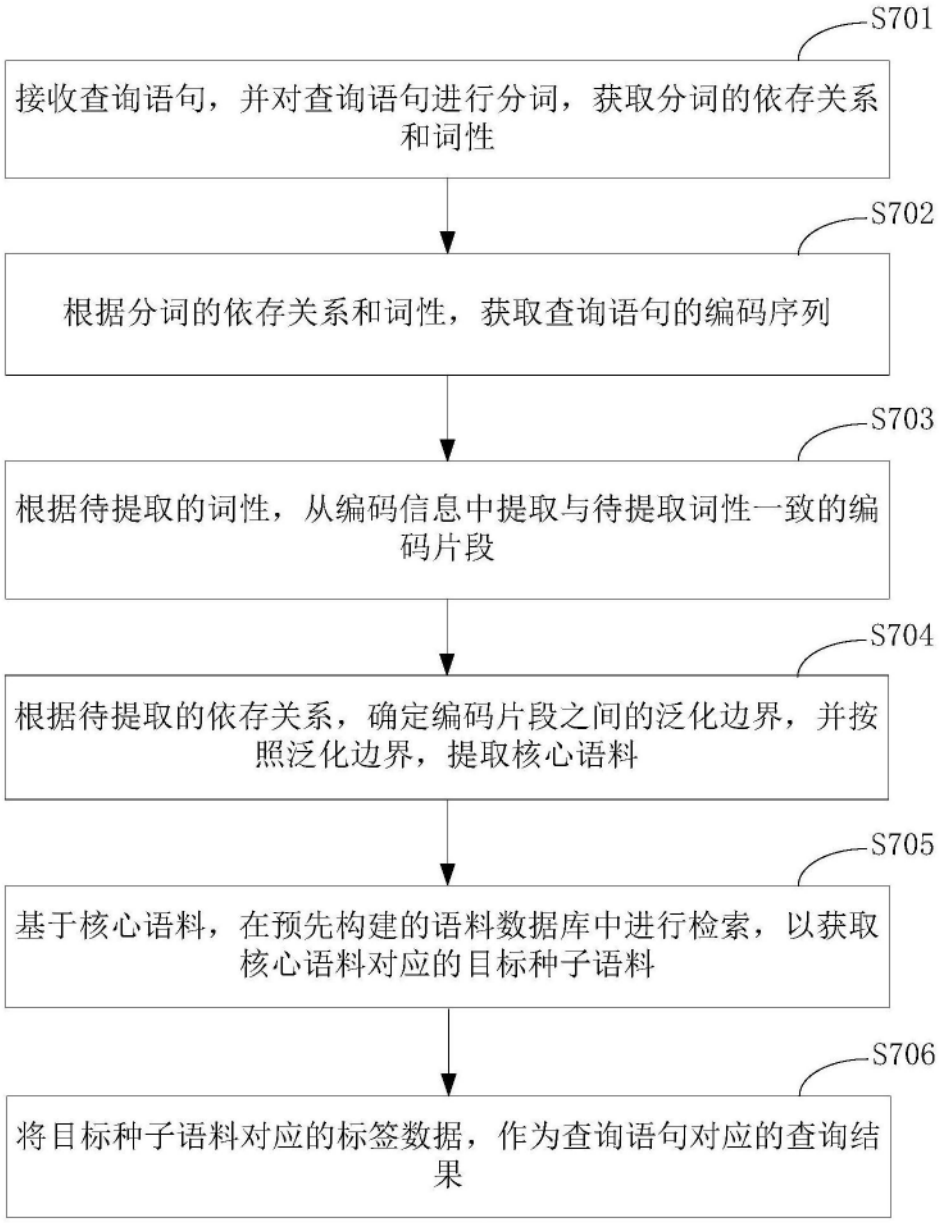

S701
接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性
S702
根据分词的依存关系和词性，获取查询语句的编码序列
S703
根据待提取的词性，从编码信息中提取与待提取词性一致的编码片段
S704
根据待提取的依存关系，确定编码片段之间的泛化边界，并按照泛化边界，提取核心语料
S705
基于核心语料，在预先构建的语料数据库中进行检索，以获取核心语料对应的目标种子语料
S706
将目标种子语料对应的标签数据，作为查询语句对应的查询结果

图7

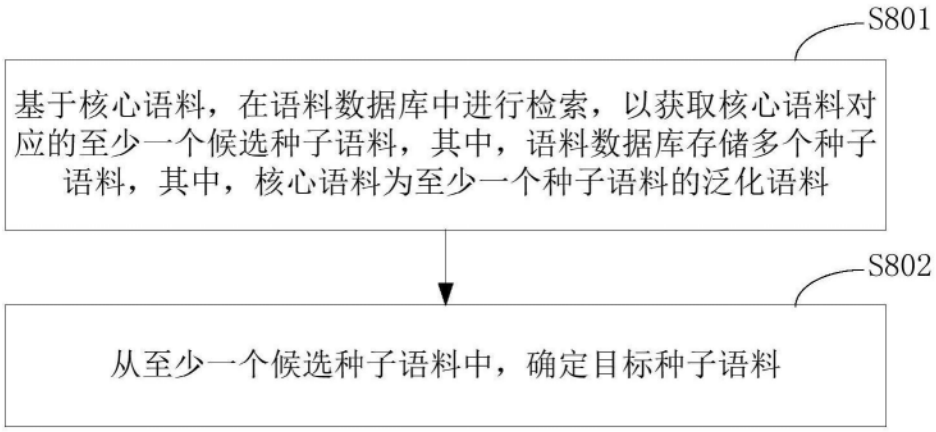

S801
基于核心语料，在语料数据库中进行检索，以获取核心语料对应的至少一个候选种子语料，其中，语料数据库存储多个种子语料，其中，核心语料为至少一个种子语料的泛化语料
S802
从至少一个候选种子语料中，确定目标种子语料

图8

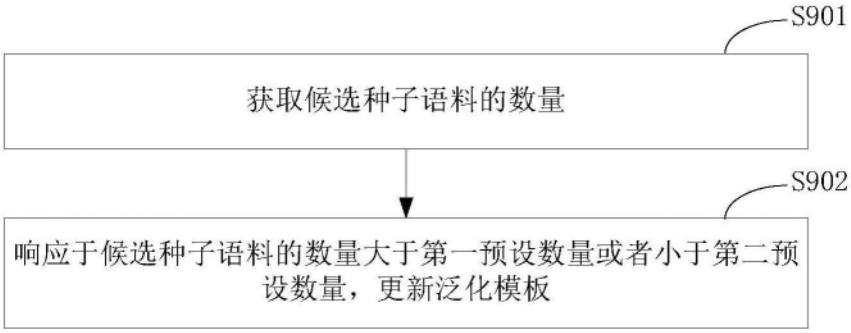

S901
获取候选种子语料的数量
S902
响应于候选种子语料的数量大于第一预设数量或者小于第二预设数量，更新泛化模板

图9

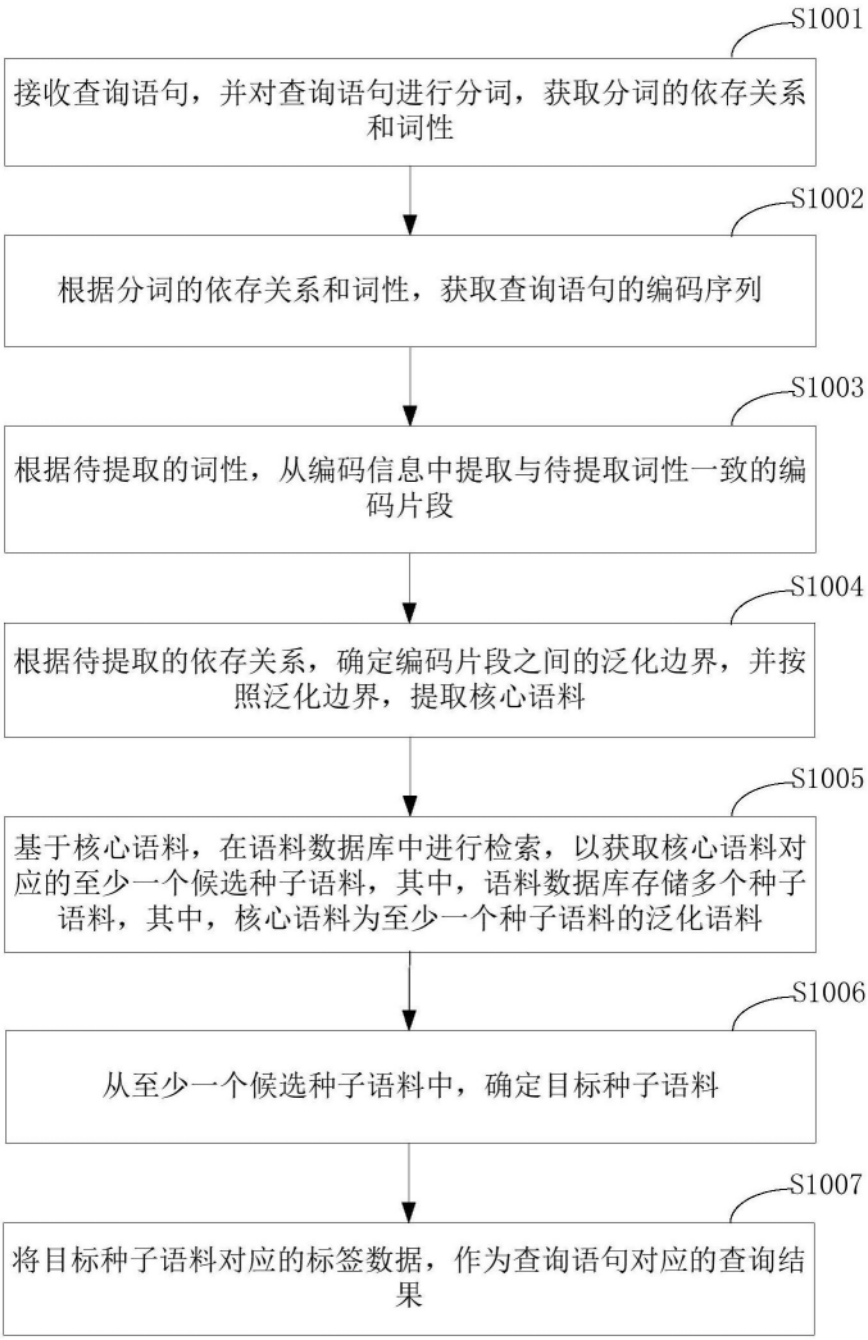

S1001
接收查询语句，并对查询语句进行分词，获取分词的依存关系和词性
S1002
根据分词的依存关系和词性，获取查询语句的编码序列
S1003
根据待提取的词性，从编码信息中提取与待提取词性一致的编码片段
S1004
根据待提取的依存关系，确定编码片段之间的泛化边界，并按照泛化边界，提取核心语料
S1005
基于核心语料，在语料数据库中进行检索，以获取核心语料对应的至少一个候选种子语料，其中，语料数据库存储多个种子语料，其中，核心语料为至少一个种子语料的泛化语料
S1006
从至少一个候选种子语料中，确定目标种子语料
S1007
将目标种子语料对应的标签数据，作为查询语句对应的查询结果

图10

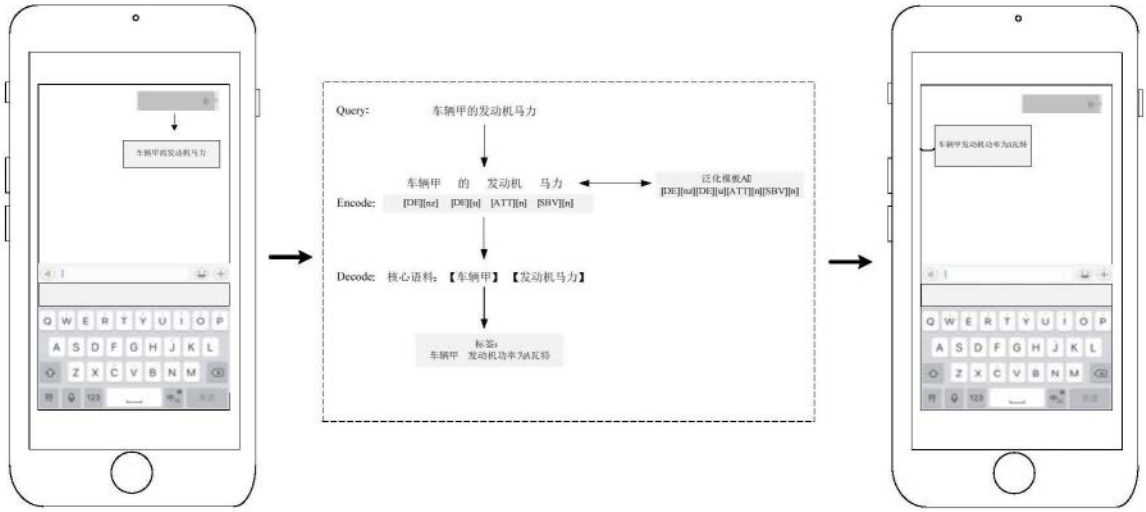

Query:
车辆甲的发动机马力
车辆甲 的 发动机 马力
Encode:
[DE][nz] [DE][u] [ATT][n] [SBV][n]
泛化模板A2
[DE][nz][DE][u][ATT][n][SBV][n]
Decode: 核心语料：【车辆甲】【发动机马力】
标签1
车辆甲 发动机功率为6瓦特

图11

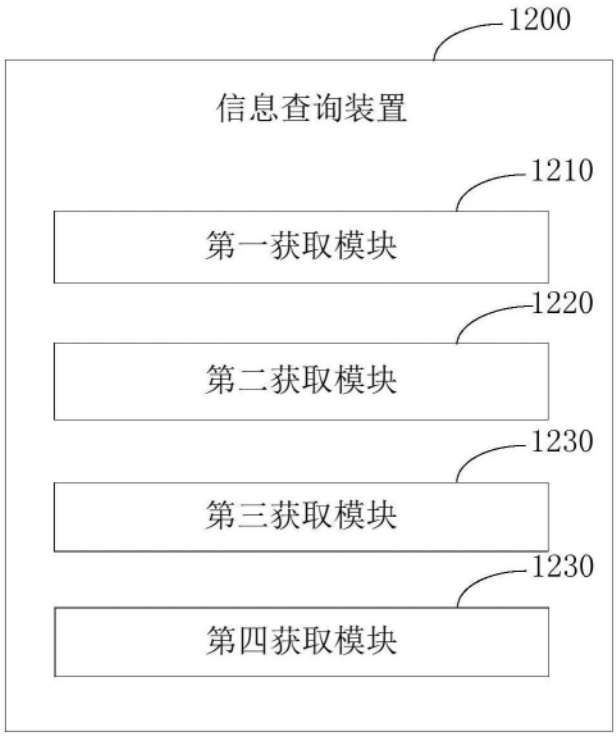

1200
信息查询装置
1210
第一获取模块
1220
第二获取模块
1230
第三获取模块
1230
第四获取模块

图12

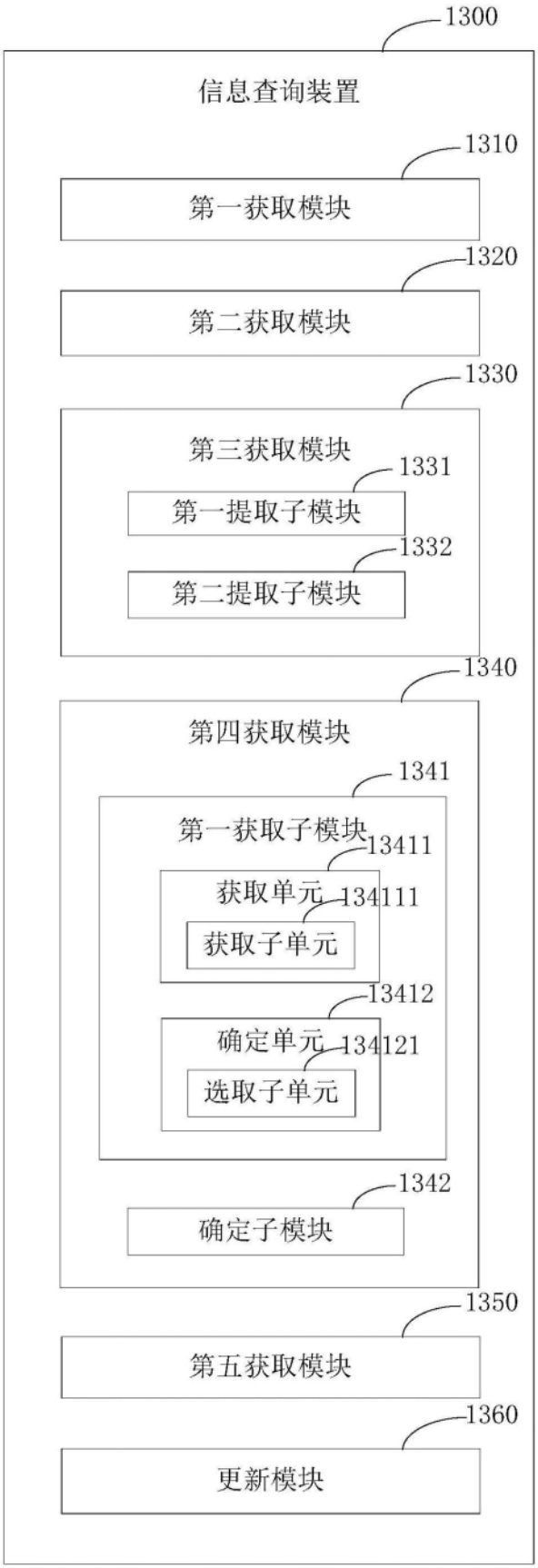

1300
信息查询装置
1310
第一获取模块
1320
第二获取模块
1330
第三获取模块
1331
第一提取子模块
1332
第二提取子模块
1340
第四获取模块
1341
第一获取子模块
13411
获取单元
134111
获取子单元
13412
确定单元
134121
选取子单元
1342
确定子模块
1350
第五获取模块
1360
更新模块

图13

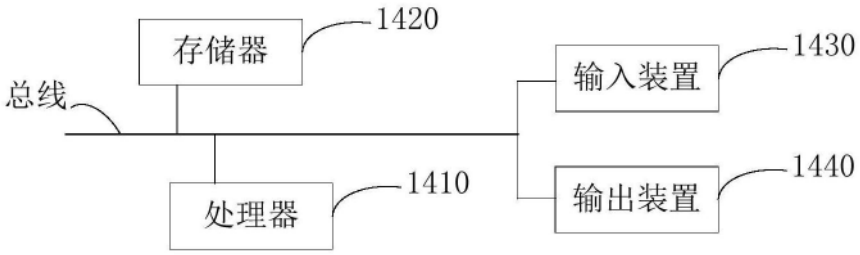


图14