(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 116737915 B
(45)授权公告日 2023.11.21

(21)申请号 202311031334.0
(22)申请日 2023.08.16
(65)同一申请的已公布的文献号
申请公布号 CN 116737915 A
(43)申请公布日 2023.09.12
(73)专利权人 中移信息系统集成有限公司
地址 100071 北京市丰台区东管头1号院3号楼2048-66
专利权人 中移系统集成有限公司
中移雄安信息通信科技有限公司
中国移动通信集团有限公司
(72)发明人 王昀 胡珉 曹植瑞 孙海涛 郭毅峰 许大虎 高有军 于庆军 梅迪菲 陈书钢 陈志刚 张皖哲 郭昱 王学峰 陈仲双 周武爱
(74)专利代理机构 北京国昊天诚知识产权代理有限公司 11315
专利代理师 王思超
(51)Int.Cl.
G06F 16/332(2019.01)
G06F 16/33(2019.01)
G06F 16/36(2019.01)
G06F 40/194(2020.01)
G06F 40/295(2020.01)
G06F 40/30(2020.01)
(56)对比文件
CN 114637760 A,2022.06.17
CN 114817559 A,2022.07.29
CN 105868313 A,2016.08.17
CN 103886099 A,2014.06.25
CN 115329137 A,2022.11.11
CN 114218472 A,2022.03.22
US 2023177363 A1,2023.06.08
US 2020134032 A1,2020.04.30
审查员 王晓燕

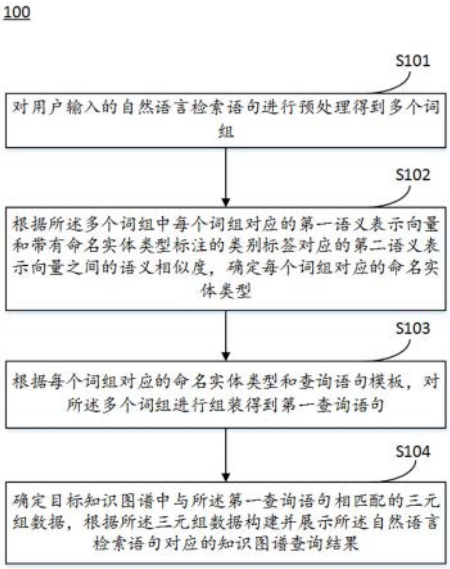
权利要求书2页 说明书11页 附图5页

(54)发明名称

基于知识图谱的语义检索方法、装置、设备及存储介质

(57)摘要

本申请公开一种基于知识图谱的语义检索方法、装置、设备及存储介质，属于自然语言处理技术领域。该方法包括：对用户输入的自然语言检索语句进行预处理得到多个词组；根据多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定每个词组对应的命名实体类型；根据每个词组对应的命名实体类型和查询语句模板，对多个词组进行组装得到第一查询语句；确定目标知识图谱中与第一查询语句相匹配的三元组数据，根据三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。通过该方式，可以实现基于自然语言检索语句进行检索查询，从而能够支撑更加复杂的检索场景。



CN 116737915 B

1.一种基于知识图谱的语义检索方法,其特征在于,包括:

对用户输入的自然语言检索语句进行预处理得到多个词组;

根据所述多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度,确定每个词组对应的命名实体类型;其中,所述命名实体类型包括实体类标签、关系类标签和属性类标签;

根据每个词组对应的命名实体类型和查询语句模板,对所述多个词组进行组装得到第一查询语句;其中,所述查询语句模板包括查询语句的语法规则,以及各查询语句之间的嵌套逻辑;

确定目标知识图谱中与所述第一查询语句相匹配的三元组数据,根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果;

其中,所述根据每个词组对应的命名实体类型和查询语句模板,对所述多个词组进行组装得到第一查询语句,包括:

确定所述多个词组中与类别标签的语义相似度最高的目标词组;

根据所述目标词组的命名实体类型和查询语句模板,确定所述第一查询语句。

2.根据权利要求1所述的方法,其特征在于,所述对用户输入的自然语言检索语句进行预处理得到多个词组,包括:

将所述自然语言检索语句划分成多个分词;

根据所述多个分词中每个分词的词性和不同分词之间的依存关系,确定多个词组。

3.根据权利要求1所述的方法,其特征在于,在所述根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果之后,还包括:

展示所述多个词组中预设数量的目标词组;

响应于所述用户对所述目标词组的选择操作,根据所述目标词组对应的第二命名实体类型和查询语句模板确定第二查询语句;

确定目标知识图谱中与所述第二查询语句相匹配的三元组数据,根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。

4.根据权利要求1所述的方法,其特征在于,在所述对用户输入的自然语言检索语句进行预处理得到多个词组之前,还包括:

获取库表资源,所述库表资源包括目录信息和资源信息;

根据所述目录信息,对所述库表资源按照目录层级进行数据处理,得到中心实体、与所述中心实体具有关联关系的子实体和关联关系;

将所述中心实体、子实体和关联关系分别对应存储在预先构建的知识图谱模型中,得到所述目标知识图谱。

5.根据权利要求4所述的方法,其特征在于,所述根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果,包括:

获取所述用户设置的查询参数;

根据所述查询参数,确定中心实体和子实体的展示位置和展示数量,按照预设的布局规则生成并展示知识图谱,所述知识图谱用于表征所述自然语言检索语句对应的知识图谱查询结果。

6.根据权利要求5所述的方法,其特征在于,在所述根据所述三元组数据构建并展示所

述自然语言检索语句对应的知识图谱查询结果之后，还包括：

响应于所述用户对目标实体的查看操作，展示所述目标实体对应的关联信息，所述目标实体包括中心实体和子实体，所述关联信息包括属性信息、元数据信息和元数据关联字段信息。

7.一种基于知识图谱的语义检索装置，其特征在于，包括：

预处理模块，用于对用户输入的自然语言检索语句进行预处理得到多个词组；

确定模块，用于根据所述多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定所述每个词组对应的命名实体类型；其中，所述命名实体类型包括实体类标签、关系类标签和属性类标签；

组装模块，用于根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行组装得到第一查询语句；其中，所述查询语句模板包括查询语句的语法规则，以及各查询语句之间的嵌套逻辑；

展示模块，用于确定目标知识图谱中与所述第一查询语句相匹配的三元组数据，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果；

其中，组装模块在用于根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行组装得到第一查询语句时，具体用于：

确定所述多个词组中与类别标签的语义相似度最高的目标词组；根据所述目标词组的命名实体类型和查询语句模板，确定所述第一查询语句。

8.一种电子设备，其特征在于，所述电子设备包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如权利要求1至6任一项所述的方法的步骤。

9.一种可读存储介质，其特征在于，所述可读存储介质上存储程序或指令，所述程序或指令被处理器执行时实现如权利要求1至6任一项所述的方法的步骤。

基于知识图谱的语义检索方法、装置、设备及存储介质

技术领域

[0001] 本申请实施例涉及自然语言处理技术领域，特别涉及一种基于知识图谱的语义检索方法、装置、设备及存储介质。

背景技术

[0002] 数据共享交换平台是数字政府类项目需要建设的一个基础平台。政务数据通过平台实现交换共享，旨在打破政务数据“各自为政、信息孤岛”的局面。

[0003] 现有的数据共享交换平台的管理方法是通过为部门之间数据的共享交换提供通道，以通过检索目录名称、资源名称等实现单个资源的申请，但是该方法只是对不同来源、不同形式的资源，按目录划分进行汇聚整合，而没有打通政务数据之间的关系，本质上这些数据之间仍然是互相独立的，碎片化的，没有形成知识，并且检索方式较为单一，只能按照目录名称、资源名称和关键词等检索资源，无法支撑复杂的检索场景。

发明内容

[0004] 本申请实施例提供了基于知识图谱的语义检索方法、装置、设备及存储介质，以至少解决现有的数据共享交换平台只能按照目录名称、资源名称和关键词等检索资源，其检索方式单一，无法支撑复杂的检索场景的问题。

[0005] 为了解决上述技术问题，本申请是这样实现的：

[0006] 第一方面，本申请实施例提供了一种基于知识图谱的语义检索方法，包括：

[0007] 对用户输入的自然语言检索语句进行预处理得到多个词组；

[0008] 根据所述多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定每个词组对应的命名实体类型；

[0009] 根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行组装得到第一查询语句；

[0010] 确定目标知识图谱中与所述第一查询语句相匹配的三元组数据，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。

[0011] 第二方面，本申请实施例提供了一种基于知识图谱的语义检索装置，包括：

[0012] 预处理模块，用于对用户输入的自然语言检索语句进行预处理得到多个词组；

[0013] 确定模块，用于根据所述多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定所述每个词组对应的命名实体类型；

[0014] 组装模块，用于根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行组装得到第一查询语句；

[0015] 展示模块，用于确定目标知识图谱中与所述第一查询语句相匹配的三元组数据，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。

[0016] 第三方面，本申请实施例提供了一种电子设备，包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如上述第一方面所述的方法的步骤。

[0017] 第四方面，本申请实施例提供了一种可读存储介质，所述可读存储介质上存储程序或指令，所述程序或指令被处理器执行时实现如上述第一方面所述的方法的步骤。

[0018] 本申请实施例提供的基于知识图谱的语义检索方法，对用户输入的自然语言检索语句进行预处理得到多个词组；根据多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定每个词组对应的命名实体类型；根据每个词组对应的命名实体类型和查询语句模板，对多个词组进行组装得到第一查询语句；确定目标知识图谱中与第一查询语句相匹配的三元组数据，根据所述三元组数据构建并展示自然语言检索语句对应的查询结果。

[0019] 通过上述方式，由于确定自然语言检索语句的多个词组对应的命名实体类型，根据命名实体类型对多个词组进行组装得到第一查询语句，通过该第一查询语句查询目标知识图谱，确定对应的查询结果，因此，相对于现有的按照目录名称、资源名称和关键词等检索资源的方式，本申请实施例能够实现基于自然语言检索语句进行检索查询，从而可以支撑更加复杂的检索场景，同时，自然语言检索语句中包含更加丰富的语义信息，基于自然语言检索语句进行检索查询，能够得到更为准确且全面的检索结果。

[0020] 应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本申请。

附图说明

[0021] 此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本申请的实施例，并与说明书一起用于解释本申请的原理。

[0022] 图1示出了本申请实施例提供的基于知识图谱的语义检索方法的流程示意图；

[0023] 图2示出了本申请实施例提供的一种可能的应用场景示意图；

[0024] 图3示出了本申请实施例提供的基于知识图谱的语义检索装置的结构示意图之一；

[0025] 图4示出了本申请实施例提供的本申请实施例提供的基于知识图谱的语义检索装置的结构示意图之二；

[0026] 图5示出了本申请实施例提供的电子设备的结构示意图。

具体实施方式

[0027] 这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本申请的一些方面相一致的装置和方法的例子。

[0028] 数据共享交换平台是数字政府类项目需要建设的一个基础平台，政务数据通过平台实现交换共享。目录和资源是平台的核心，目录实现对资源的层级和分类划分，资源包括库表、文件和接口等，每个资源都归属在一个目录下，各政府主体可在平台上进行资源申

请。

[0029] 现有技术中，数据共享交换平台的检索方式单一，通常按照目录名称、资源名称和关键词等检索资源，无法适用于复杂的检索场景。本申请实施例提供了一种基于知识图谱的语义检索方法，能够实现基于自然语言检索语句进行检索查询，相对于目录名称、资源名称、关键词等，自然语言检索语句中包含更加丰富的语义信息，基于自然语言检索语句进行检索查询，能够得到更为准确且全面的检索结果。

[0030] 图1示出了本申请实施例提供的基于知识图谱的语义检索方法的流程示意图，该方法的执行主体可以为终端设备或服务器，其中，该终端设备可以如个人计算机等设备，也可以如手机、平板电脑等移动终端设备，该终端设备可以为用户使用的终端设备。该服务器可以是独立的服务器，也可以是由多个服务器组成的服务器集群，而且，该服务器可以是独立的服务器，也可以是由多个服务器组成的服务器集群。该方法可以应用于数据共享交换平台，实现基于自然语言检索语句进行检索查询。本申请实施例中以执行主体为服务器为例进行说明，对于终端设备的情况，可以根据下述相关内容处理，在此不再赘述。如图中所示，该基于知识图谱的语义检索方法100可以包括以下步骤：

[0031] S101：对用户输入的自然语言检索语句进行预处理得到多个词组。

[0032] 在具体实施中，用户通过终端设备的检索页面上输入自然语言检索语句，获取该自然语言检索语句，并进行如分词、词性标注、依存分析等预处理得到多个词组。举例而言，用户输入的自然语言检索语句为“苹果的颜色是红色的”，对自然语言检索语句进行分词处理得到“苹果/的/颜色/是/红色/的”，根据各分词的词性和不同分词之间的依存关系，得到多个词组“苹果”、“颜色”、“红色”。

[0033] S102：根据所述多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定所述每个词组对应的命名实体类型。

[0034] 其中，命名实体类型包括实体类标签、关系类标签和属性类标签。

[0035] 在具体实施中，对上述S101中获取的多个词组进行向量表示，得到每个词组对应的第一语义表示向量，例如，对词组“苹果”进行向量表示，得到“苹果”的第一语义表示向量。具体可以使用预训练词向量对所述多个词组中的每个词组进行语义表示，得到每个词组对应的第一语义表示向量。由于预训练词向量包含大量的语义信息，能够更好地反映词语之间的相似关系，因此利用预先训练好的大规模语料库得到的词向量，来计算两个词之间的语义相似度。例如，预训练词向量可以让“猫”和“狗”这两个词在向量空间中比较接近，因为它们通常都被用来描述宠物动物，而“猫”和“桌子”之间的距离更远，因为它们通常没有太多联系。可以利用词向量之间的相似度来度量词语之间的语义相似度，从而实现自然语言处理中的各种任务，比如文本分类、信息检索等。

[0036] 获取带有命名实体类型标注的类别标签对应的第二语义表示向量，类别标签如“水果”、“植物”等。进而，根据第一语义表示向量与第二语义表示向量之间的语义相似度，确定多个词组对应的命名实体类型，例如，“苹果”属于“水果”，“水果”对应的命名实体类别为实体类标签，则“苹果”为实体类标签，通过同样的方式，可以确定“颜色”为关系类标签，“红色”为属性类标签。

[0037] S103：根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行

组装得到第一查询语句。

[0038] 在具体实施中，根据支持的match、lookup、fetch等查询语句，设计一条通用的查询语句模板，根据自然语言检索语句对应的多个词组和每个词组对应的命名实体类型，按照该查询语句模板进行组装形成第一查询语句，通过第一查询语句进行目标知识图谱的查询，得到相应的查询结果。

[0039] 这里，查询语句模板包括查询语句的语法规则，以及各查询语句之间的嵌套逻辑。

[0040] S104：确定目标知识图谱中与所述第一查询语句相匹配的三元组数据，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。

[0041] 在具体实施中，通过上述的第一查询语句查询目标知识图谱，得到与第一查询语句相匹配的三元组数据，也即通过第一查询语句能够进行图数据库检索。根据三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。这里，目标知识图谱可以为数据共享交换平台获取的政务知识图谱，也可以是其他类别的知识图谱。

[0042] 本申请实施例提供了一种基于知识图谱的语义检索方法，由于确定自然语言检索语句的多个词组对应的命名实体类型，根据命名实体类型对多个词组进行组装得到第一查询语句，通过该第一查询语句查询目标知识图谱，确定对应的查询结果，因此，相对于现有的按照目录名称、资源名称和关键词等检索资源的方式，能够实现基于自然语言检索语句进行检索查询，从而可以支撑更加复杂的检索场景，同时，自然语言检索语句中包含更加丰富的语义信息，基于自然语言检索语句进行检索查询，能够得到更为准确且全面的检索结果。

[0043] 可选地，在上述步骤S101中，对用户输入的自然语言检索语句进行预处理得到多个词组，包括：

[0044] 将所述自然语言检索语句划分成多个分词；根据所述多个分词中每个分词的词性和不同分词之间的依存关系，确定多个词组。

[0045] 其中，依存分析是一种句法分析方法，它分析句子中不同分词之间的依存关系，以此来确定句子的结构。它将句子中的每个分词都视为一个节点，然后通过分析它们之间的关系(如主谓、动宾等)，建立一棵树形结构来表示整个句子的结构。举例而言，对于句子“我爱你”，依存分析会将“我”和“爱”之间建立一个主语关系，将“你”和“爱”之间建立一个宾语关系，从而建立一棵包含“爱”作为根节点，以“我”和“你”作为子节点的树形结构。这样，有助于识别句子的语法结构，进而实现自然语言处理等任务。

[0046] 进而，根据多个分词中每个分词的词性和不同分词之间的依存关系，确定多个词组，可以抽取出自然语言检索语句中的能够表达句子语义的词组，从而能够使查询结果更加符合用户需求，提高用户的满意度。

[0047] 可选地，在上述步骤103中，根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行组装得到第一查询语句，包括：

[0048] 确定所述多个词组中与类别标签的语义相似度最高的目标词组；根据所述目标词组的命名实体类型和查询语句模板，确定所述第一查询语句。

[0049] 在具体实施中，以词组为“苹果”为例，候选类别标签可以是“水果”、“植物”等，假设“苹果”与“水果”之间的语义相似度为0.9，“苹果”与“植物” 之间的语义相似度为0.6，则根据候选标签中语义相似度最高的类型标签“水果”对应的目标命名实体类型，确定词组

对应的命名实体类型。在词组为多个的情况下，语义相似度可以是各词组的第一语义表示向量与候选类别标签的第二语义表示向量之间的相似度的和值，也可以是各词组的第一语义表示向量与候选类别标签的第二语义表示向量之间的相似度的平均值或者加权平均值。在确定第一查询语句时，可以根据多个词组中与类别标签的语义相似度最高的目标词组的命名实体类型和查询语句模板，确定第一查询语句。

[0050] 可选地，在上述步骤S104之后，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果之后，还包括：

[0051] 展示所述多个词组中预设数量的目标词组；响应于所述用户对所述目标词组的选择操作，根据所述目标词组对应的第二命名实体类型和查询语句模板确定第二查询语句；确定目标知识图谱中与所述第二查询语句相匹配的三元组数据，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。

[0052] 在具体实施中，为了提高用户查询的便捷性，在显示知识图谱查询结果之后，展示多个词组中预设数量的目标词组，当用户选择目标词组之后，重新根据目标词组的第二命名实体类型和查询语句模板，确定第二查询语句，进而通过第二查询语句查询目标知识图谱，确定自然语言检索语句对应的知识图谱查询结果。

[0053] 可选地，在所述对用户输入的自然语言检索语句进行预处理得到多个词组之前，还包括：

[0054] 获取库表资源，所述库表资源包括目录信息和资源信息；

[0055] 根据所述目录信息，对所述库表资源按照目录层级进行数据处理，得到中心实体、与所述中心实体具有关联关系的子实体和关联关系；

[0056] 将所述中心实体、子实体和关联关系分别对应存储在预先构建的知识图谱模型中，得到所述目标知识图谱。

[0057] 在具体实施中，在对用户输入的自然语言检索语句进行预处理得到多个词组之前，构建目标知识图谱，这里的目标知识图谱可以为政务知识图谱，政务知识图谱的构建过程可以包括以下步骤：

[0058] (1)数据采集：依托于外部大数据治理平台，定时采集数据共享交换平台库表资源，接入数据仓库。同时在数据仓库的数据表中新增字段，将资源目录信息(目录层级信息，目录名称)和资源信息(资源名称，资源id)以自定义格式json字符串形式，放入新增字段中；

[0059] (2)实体、关系、属性挖掘：挖掘的原则是将所有采集的库表资源和资源中的信息项视为平级，逐个进行分析挖掘，通过分析数据表的字段内容、主键、外键等信息识别出实体和关系，进而抽取出实体的概念、属性，从描述关系的表中抽取出概念间的关系，以定义其要作为实体，关系或属性中的哪类元素。同时将整个知识图谱划分为多个体系，共同组成知识图谱，体系之间也存在交叉关系。知识图谱的建设基于数据共享交换平台的政务数据，政务数据是在共享交换平台上挂接的政务数据资源，这些政务数据资源是对自然人、法人、政府机构、公司组织和物品等全方位、全生命周期的重点信息描述，通过知识图谱的建设，打破政务数据之间的壁垒，多场景、多维度实现对政务知识的全方位展示，实现政务知识可视化。政务数据包括但不限于以下三个体系：

[0060] ①个人画像体系：

[0061] 个人画像体系以自然人为核心，可以将身份证号码作为自然人实体唯一id(即中心实体)，将姓名、性别、年龄和籍贯等基本信息作为自然人实体属性，将来自于不同部门的学校、地址、职业、疾病、医保、社保、信用、资质等信息实体(即子实体)与自然人产生关系，关系则可按毕业院校、住址、上班地址、职业、疾病等进行命名。自然人之间可存在婚姻关系、亲情关系等。

[0062] ②组织画像体系：

[0063] 组织画像体系类同个人画像体系，组织包括公司、政府机构和事业单位等，可将统一社会信用代码作为组织实体唯一id(即中心实体)，将组织名称、组织类型、经营者、登记状态等基本信息作为组织实体属性，将黑名单、红名单、信用情况等信息实体(即子实体)与组织实体产生关系。同时组织之间可存在子母关系，自然人和组织可存在法人关系，雇佣关系等。地址实体只有一份，自然人和组织都可与其产生关系，只不过关系可能各种各样，这是体系之间关系交叉的一个例子。

[0064] ③事项体系：

[0065] 事项是指国家政务服务事项，事项包括事项名称、办理地址、办理条件、办理所需材料、办理前置环节等信息。同时与事项体系有关系的还包括自然人、组织、办件和好差评。自然人和组织可办理办件，并且可以对办件进行评价，同时办件又属于具体事项。

[0066] (3)模型类设计：在实体、关系、属性挖掘的基础上，进行实体类和关系类模型设计。模型设计的原则是实体和关系分开放入不同的表，可能存在实体和关系都在同一张表的情况，这种情况还是要将实体和关系分开，方便后续维护。同时，在模型设计的时候，必须包含库表资源元数据信息。实体的每个属性都要后缀元数据信息，并且给每个属性编号，同时要在实体增加一个元数据关系属性，该属性通过列表的形式，以[[编号,关联字段=关联字段,编号],[...]]的格式记录实体属性之间的关系，如果实体和属性在一个原始表里，则以[编号,编号]的格式记录关系。关系也要增加一个元数据关系属性，以[[vid1,关联字段=关联字段,vid2],[...]]格式记录实体之间的关系，如果实体和实体在一个原始表里，则以[vid,vid]格式记录实体之间的关系，此处暂时不考虑关系的属性关联关系；

[0067] (4)数据开发：数据开发基于设计好的模型进行实体关系表开发。基于数据共享交换平台库表资源的政务知识图谱，由于库表资源数据量大、种类多，知识图谱的实体关系数量也会同时很大，每次全量导入图数据库可能存在效率问题，因此采用增量方式。基于hive分区，进行前后分区实体和关系去重，每次只增量导入新增的实体和关系。但同时由于库表资源动态变化的特点，整体资源包含的信息项也在不断变化，因此会存在阶段性模型修改，因此政务知识图谱的建设维护是一项持续性工作；

[0068] (5)数据导入：基于图数据库引擎功能，在图数据库中构建政务知识图谱模型，添加实体类和关系类，并基于数据源管理功能进行实体关系数据导入。导入后需整体观察知识图谱情况，观察有无明显错误，如果把不可能有关系的实体连在了一起，实体之间存在多条相同的关系等。同时基于模型数据，抽样进行知识图谱实体关系明细校对，以确保建模，开发和数据导入的准确无误。

[0069] 这样，通过将采集的库表资源中的中心实体、子实体和关联关系分别对应存储在预先构建的知识图谱模型中，具体可以将关联数据的实体作为点存储，关系作为边存储，形成政务知识图谱，可以打通政务数据之间的关系。

[0070] 可选地，所述根据所述三元组数据构建并展示所述自然语言检索语句对应的查询结果，包括：

[0071] 获取所述用户设置的查询参数；根据所述查询参数，确定中心实体和子实体的展示位置和展示数量，按照预设的布局规则生成并展示知识图谱，所述知识图谱用于表征所述自然语言检索语句对应的知识图谱查询结果。

[0072] 其中，查询参数包括实体之间关联关系的方向，如流入、流出、双向等，搜寻步数和搜寻范围，展示实体数量。

[0073] 可选地，在所述根据所述三元组数据构建并展示所述自然语言检索语句对应的查询结果之后，还包括：

[0074] 响应于所述用户对目标实体的查看操作，展示所述目标实体对应的关联信息，所述目标实体包括中心实体和子实体，所述关联信息包括属性信息、元数据信息和元数据关联字段信息。

[0075] 在具体实施中，解析目标实体携带的库表资源元数据信息，可以通过视图切换，展示实体、关系、属性的来源信息，即来自于共享交换平台的哪个资源，信息包括目录信息、资源信息和资源id，同时可以展示实体和属性，关系和属性之间关联字段信息。

[0076] 此外，还可以将知识图谱查询结果导出为图片形式；将视图切换后的库表资源元数据信息导出为图片形式；将视图切换后的库表资源元数据信息导出为excel表格形式；针对知识图谱中涉及的库表资源，可以展示资源列表，且可以个性化选择批量申请。

[0077] 知识图谱侧重于关系，在知识图谱建设过程中，无法将库表资源中的所有信息放入知识图谱中，所以在展示知识图谱的同时，可以展示库表资源元数据信息，及库表资源之间的关联字段信息，且可以进行库表资源批量申请，这样用户在得到图谱知识的同时，还可以获取到原始数据及其关联字段，自行进行sql关联计算，得到更为详尽的知识，通过该方式，可以提高资源信息的获取效率。

[0078] 图2示出了本申请实施例提供的一种可能的应用场景示意图，如图中所示，本申请实施例提供的共享交换平台200，包括：知识图谱检索模块210、语义处理引擎220、图数据库引擎230和库表资源240；其中，

[0079] 知识图谱检索模块210，用于获取用户输入的自然语言检索语句和查询步数等查询参数，将自然语言检索语句传入语义处理引擎220，查询参数透传至图数据库引擎230，并根据图数据库引擎230返回的查询结果进行展示，同时提供知识图谱申请下载能力和资源申请下载能力；

[0080] 语义处理引擎220，用于处理知识图谱检索模块210输入的自然语言检索语句，对自然语言检索语句进行预处理得到多个词组；根据多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定每个词组对应的命名实体类型，将词组和对应的命名实体类型传入图数据库引擎230；

[0081] 图数据库引擎230，用于根据每个词组对应的命名实体类型和查询语句模板，对多个词组进行组装得到第一查询语句；确定目标知识图谱中与第一查询语句相匹配的三元组数据，将根据三元组数据确定的查询结果返回至知识图谱检索模块210；

[0082] 图数据库引擎230，还用于：获取库表资源；根据目录信息，对库表资源按照目录层级进行数据处理，得到中心实体、与所述中心实体具有关联关系的子实体和关联关系；将中

心实体、子实体和关联关系分别对应存储在预先构建的知识图谱模型中，得到目标知识图谱。

[0083] 在具体实施中，图数据库引擎230可以采用neo4j、nebulagraph等图数据库，具备词库管理、图谱模板、数据源管理和图谱实体功能，具体为：

[0084] 词库管理：新增词条，编辑词条，删除词条，词条导出，词条导入；

[0085] 图谱模板：实体类创建，关系类创建，类删除，属性添加，属性删除；

[0086] 数据源管理：模板下载，mysql源导入，hive源导入；

[0087] 图谱实体：实体关系展示，查询，批量删除，搜寻步数设置，关系方向设置。

[0088] 图数据库引擎230分前端和后端，一个后端支撑两个前端，其中的一个前端用以支撑上述的功能，主要用以知识图谱建设和验证，另一个前端集成在共享交换平台知识图谱检索功能页面，用于语义检索。

[0089] 图3示出了本申请实施例提供的基于知识图谱的语义检索装置的结构示意图之一，如图中所示，该语义检索装置300，包括：

[0090] 预处理模块310，用于对用户输入的自然语言检索语句进行预处理得到多个词组；

[0091] 确定模块320，用于根据所述多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定所述每个词组对应的命名实体类型；

[0092] 组装模块330，用于根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行组装得到第一查询语句；

[0093] 展示模块340，用于确定目标知识图谱中与所述第一查询语句相匹配的三元组数据，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。

[0094] 其中，预处理模块310在用于对用户输入的自然语言检索语句进行预处理得到多个词组时，具体用于：

[0095] 将所述自然语言检索语句划分成多个分词；根据所述多个分词中每个分词的词性和不同分词之间的依存关系，确定多个词组。

[0096] 可选地，组装模块330在用于根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行组装得到第一查询语句时，具体用于：

[0097] 确定所述多个词组中与类别标签的语义相似度最高的目标词组；根据所述目标词组的命名实体类型和查询语句模板，确定所述第一查询语句。

[0098] 可选地，展示模块340在用于根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果之后，还用于：

[0099] 展示所述多个词组中预设数量的目标词组；响应于所述用户对所述目标词组的选择操作，根据所述目标词组对应的第二命名实体类型和查询语句模板确定第二查询语句；确定目标知识图谱中与所述第二查询语句相匹配的三元组数据，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。

[0100] 图4示出了本申请实施例提供的本申请实施例提供的基于知识图谱的语义检索装置的结构示意图之二，如图中所示，语义检索装置300，还包括：

[0101] 图谱获取模块350，用于获取库表资源，所述库表资源包括目录信息和资源信息；根据所述目录信息，对所述库表资源按照目录层级进行数据处理，得到中心实体、与所述中

心实体具有关联关系的子实体和关联关系;将所述中心实体、子实体和关联关系分别对应存储在预先构建的知识图谱模型中,得到所述目标知识图谱。

[0102] 其中,展示模块340在用于根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果时,具体用于:

[0103] 获取所述用户设置的查询参数;

[0104] 根据所述查询参数,确定中心实体和子实体的展示位置和展示数量,按照预设的布局规则生成知识图谱;

[0105] 在交互界面上展示所述知识图谱,所述知识图谱用于表征所述自然语言检索语句对应的知识图谱查询结果。

[0106] 展示模块340在用于根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果之后,还用于:

[0107] 响应于所述用户对目标实体的查看操作,展示所述目标实体对应的关联信息,所述目标实体包括中心实体和子实体,所述关联信息包括属性信息、元数据信息和元数据关联字段信息。

[0108] 本申请实施例提供了一种基于知识图谱的语义检索装置,包括预处理模块、确定模块、组装模块和展示模块,通过预处理模块对用户输入的自然语言检索语句进行预处理得到多个词组;确定模块根据所述多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度,确定所述每个词组对应的命名实体类型;组装模块根据每个词组对应的命名实体类型和查询语句模板,对所述多个词组进行组装得到第一查询语句;展示模块确定目标知识图谱中与所述第一查询语句相匹配的三元组数据,根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果。相对于现有的按照目录名称、资源名称和关键词等检索资源的方式,本申请实施例能够实现基于自然语言检索语句进行检索查询,从而可以支撑更加复杂的检索场景,同时,自然语言检索语句中包含更加丰富的语义信息,基于自然语言检索语句进行检索查询,能够得到更为准确且全面的检索结果。

[0109] 图5示出执行本申请实施例提供的电子设备的硬件结构示意图,参考该图,在硬件层面,电子设备包括处理器,可选地,包括内部总线、网络接口、存储器。其中,存储器可能包含内存,例如高速随机存取存储器(Random-Access Memory,RAM),也可能还包括非易失性存储器(non-volatile memory),例如至少1个磁盘存储器等。当然,该计算机设备还可能包括其他业务所需要的硬件。

[0110] 处理器、网络接口和存储器可以通过内部总线相互连接,该内部总线可以是工业标准体系结构(Industry Standard Architecture,ISA)总线、外设部件互连标准(Peripheral Component Interconnect,PCI)总线或扩展工业标准结构(Extended Industry Standard Architecture,EISA)总线等。所述总线可以分为地址总线、数据总线、控制总线等。为便于表示,该图中仅用一个双向箭头表示,但并不表示仅有一根总线或一种类型的总线。

[0111] 存储器,存放程序。具体地,程序可以包括程序代码,所述程序代码包括计算机操作指令。存储器可以包括内存和非易失性存储器,并向处理器提供指令和数据。

[0112] 处理器从非易失性存储器中读取对应的计算机程序到内存中然后运行,在逻辑层

面上形成定位目标用户的装置。处理器，执行存储器所存放的程序，并具体执行：图1所示实施例揭示的方法并实现前文方法实施例中所述的各方法的功能和有益效果，在此不再赘述。

[0113] 上述如本申请图1所示实施例揭示的方法可以应处理器中，或者由处理器实现。处理器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。上述的处理器可以是通用处理器，包括中央处理器(Central Processing Unit，CPU)、网络处理器(Network Processor，NP)等；还可以是数字信号处理器(Digital Signal Processor，DSP)、专用集成电路(Application Specific Integrated Circuit，ASIC)、现场可编程门阵列(Field－Programmable Gate Array，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申请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。结合本申请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储器，闪存、只读存储器，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，处理器读取存储器中的信息，结合其硬件完成上述方法的步骤。

[0114] 该计算机设备还可执行前文方法实施例中所述的各方法，并实现前文方法实施例中所述的各方法的功能和有益效果，在此不再赘述。

[0115] 当然，除了软件实现方式之外，本申请的电子设备并不排除其他实现方式，比如逻辑器件抑或软硬件结合的方式等等，也就是说以下处理流程的执行主体并不限定于各个逻辑单元，也可以是硬件或逻辑器件。

[0116] 本申请实施例还提出了一种计算机可读存储介质，所述计算机可读介质存储一个或多个程序，所述一个或多个程序当被包括多个应用程序的电子设备执行时，使得所述电子设备执行图1所示实施例揭示的方法并实现前文方法实施例中所述的各方法的功能和有益效果，在此不再赘述。

[0117] 其中，所述的计算机可读存储介质包括只读存储器(Read-Only Memory，简称ROM)、随机存取存储器(Random Access Memory，简称RAM)、磁碟或者光盘等。

[0118] 进一步地，本申请实施例还提供了一种计算机程序产品，所述计算机程序产品包括存储在非暂态计算机可读存储介质上的计算机程序，所述计算机程序包括程序指令，当所述程序指令被计算机执行时，实现以下流程：图1所示实施例揭示的方法并实现前文方法实施例中所述的各方法的功能和有益效果，在此不再赘述。

[0119] 总之，以上所述仅为本申请的较佳实施例，并非限定本申请的保护范围。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

[0120] 上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的，计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[0121] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体(transitory media)，如调制的数据信号和载波。

[0122] 还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0123] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

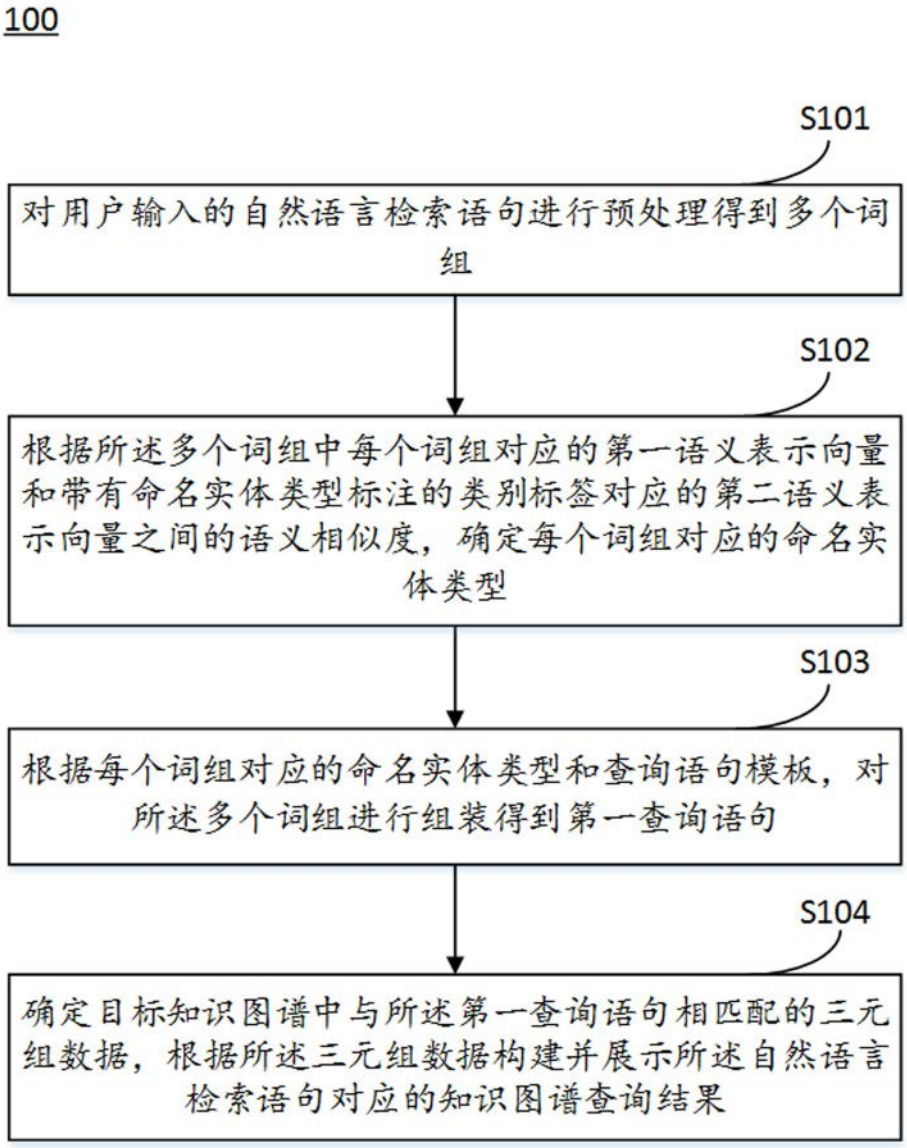

100
S101
对用户输入的自然语言检索语句进行预处理得到多个词组
S102
根据所述多个词组中每个词组对应的第一语义表示向量和带有命名实体类型标注的类别标签对应的第二语义表示向量之间的语义相似度，确定每个词组对应的命名实体类型
S103
根据每个词组对应的命名实体类型和查询语句模板，对所述多个词组进行组装得到第一查询语句
S104
确定目标知识图谱中与所述第一查询语句相匹配的三元组数据，根据所述三元组数据构建并展示所述自然语言检索语句对应的知识图谱查询结果

图 1

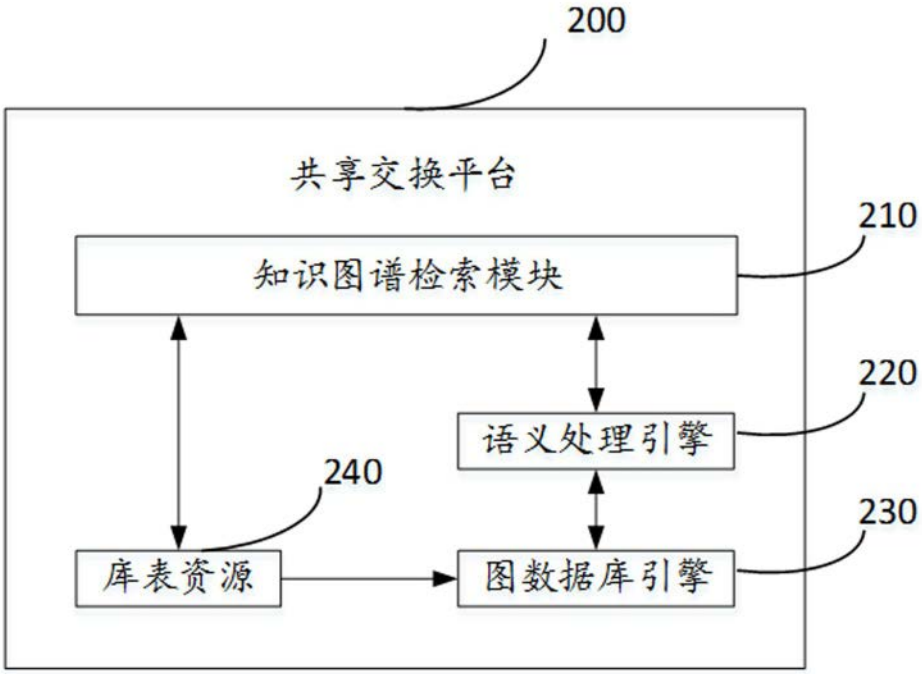


图 2

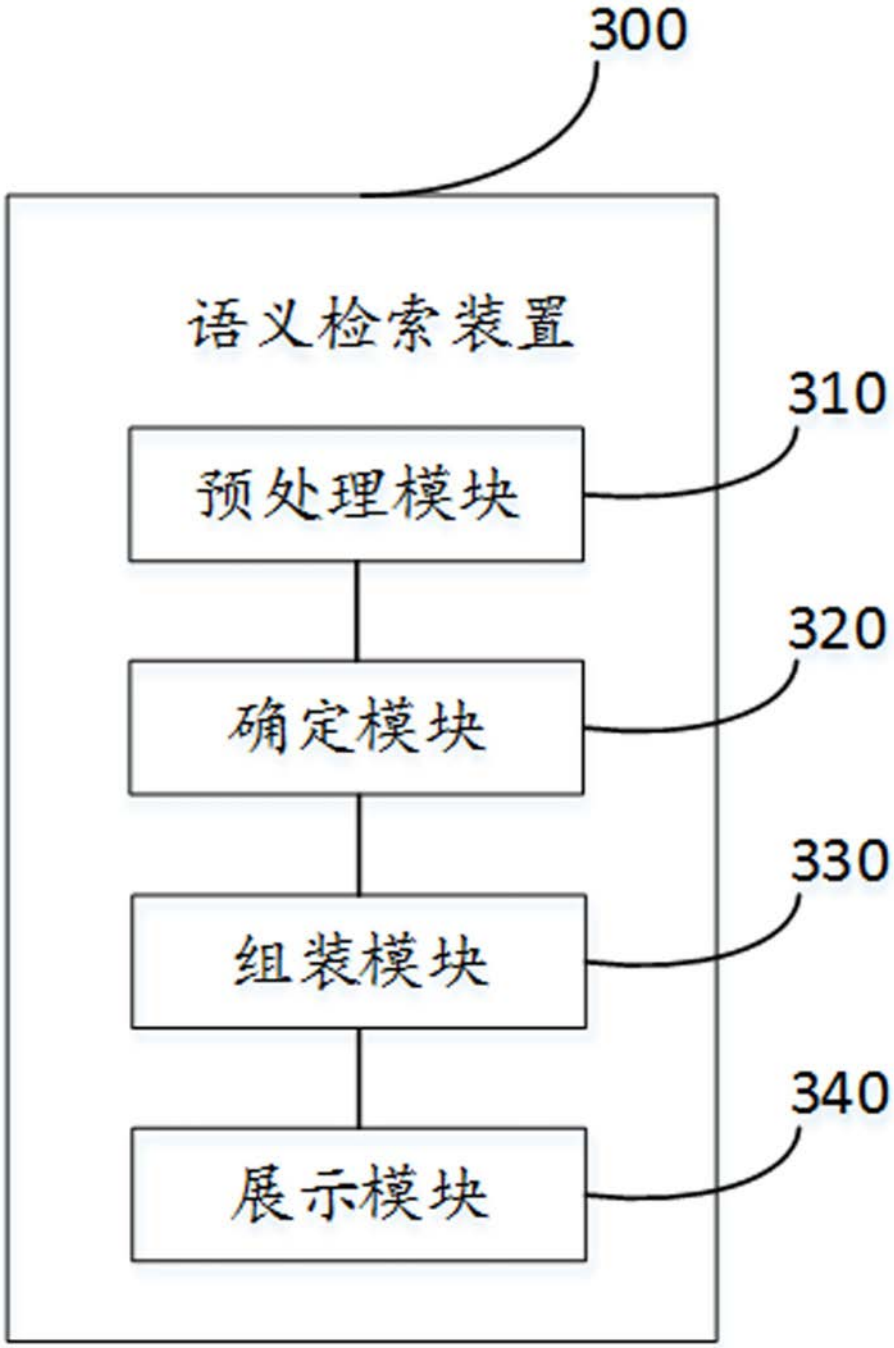


图 3

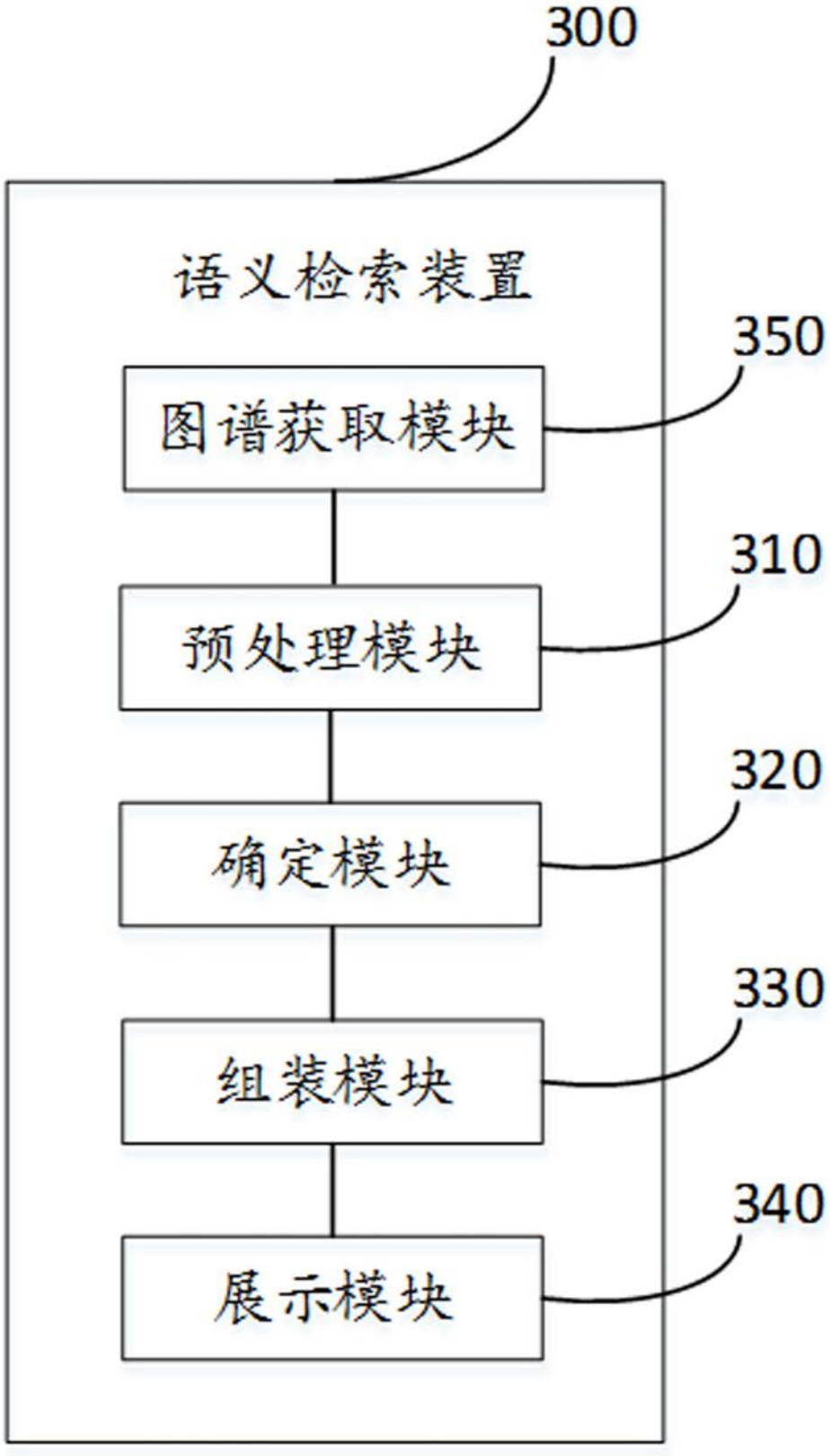


图 4

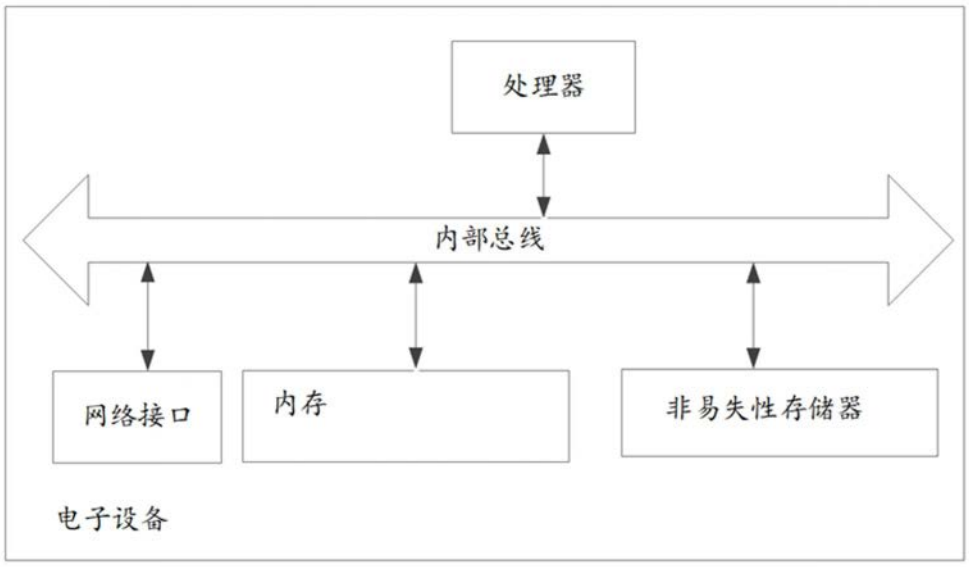


图 5