



(12) 发明专利

(10) 授权公告号 CN 109408811 B

(45) 授权公告日 2021. 10. 22

(21) 申请号 201811147486.6

G06F 16/36 (2019.01)

(22) 申请日 2018.09.29

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 104503998 A, 2015.04.08

申请公布号 CN 109408811 A

CN 104035917 A, 2014.09.10

CN 105701253 A, 2016.06.22

(43) 申请公布日 2019.03.01

US 2018150468 A1, 2018.05.31

(73) 专利权人 联想(北京)有限公司

US 2016147737 A1, 2016.05.26

地址 100085 北京市海淀区上地信息产业
基地创业路6号

张志昌. 开放域问答技术研究进展.《电子学
报》.2009,第37卷(第5期),第1058-1069页.

(72) 发明人 金继民 金宝宝 杨帆 张成松

审查员 周永传

(74) 专利代理机构 北京集佳知识产权代理有限
公司 11227

代理人 王宝筠

(51) Int. Cl.

G06F 40/211 (2020.01)

G06F 16/332 (2019.01)

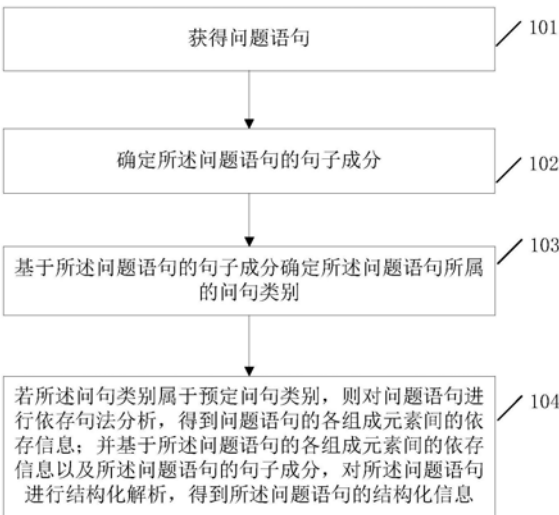
权利要求书3页 说明书34页 附图9页

(54) 发明名称

一种数据处理方法及服务器

(57) 摘要

本申请公开的一种数据处理方法及服务器，在获得问题语句后，基于问题语句的句子成分确定问题语句所属的问句类别，并在问题语句属于预定问句类别的情况下，对问题语句进行依存句法分析，得到问题语句的各组成元素间的依存信息，进而基于问题语句的各组成元素间的依存信息以及所述问题语句的句子成分，对问题语句进行结构化解析，得到问题语句的结构化信息。通过利用问题语句的各组成元素间的依存信息以及问题语句的句子成分，对问题语句进行结构化解析，可达到较好的解析效果，进而可基于解析结果准确地识别用户问句意图。



1. 一种数据处理方法,其特征在于,包括:

获得问题语句;

确定所述问题语句的句子成分;

基于所述问题语句的句子成分确定所述问题语句所属的问句类别;

若所述问句类别属于预定问句类别,则对所述问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息;并基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,得到所述问题语句的结构化信息;

所述问题语句所属的问句类别为简单关系型问题、简单属性型问题及复杂关系属性型问题中的任意一种;所述简单关系型问题中仅包括一条关系线;所述简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问;所述复杂关系属性型问题中包括不少于两条的关系线,或者包括不少于一条的关系线及实体属性;

其中,所述关系线,是指在利用一个谓词将一个实体/实体类别与另一个实体/实体类别进行衔接,使得两个实体/实体类别间产生谓词关系后所得的两个实体/实体类别与谓词间的对应关系。

2. 根据权利要求1所述的方法,其特征在于,所述确定所述问题语句的句子成分,包括:确定所述问题语句的各组成元素中的实体及实体类别。

3. 根据权利要求2所述的方法,其特征在于,所述确定所述问题语句的句子成分,还包括:确定所述问题语句的各组成元素中的属性及疑问词,所述属性包括实体属性和/或实体类别属性;

所述确定所述问题语句的各组成元素中的实体、实体类别、属性及疑问词,具体包括:

利用预定的实体识别方法,识别所述问题语句的各组成元素中的实体;

利用预定的疑问词识别方法,识别所述问题语句的各组成元素中的疑问词;

将所述问题语句中未被识别的各组成元素分别与预定的实体类别库、属性库进行匹配,得到所述问题语句的各组成元素中的实体类别及属性;所述实体类别库及所述属性库为分别预先对知识图谱进行相对应的实体类别信息抽取以及实体属性和/或实体类别属性信息抽取后所得的信息库;所述知识图谱为一预先基于多个语句样本所构建的包括多个节点的有向图,图中的每个节点对应一个实体或实体类别,每个实体对应零个或以上的实体属性、每个实体类别对应零个或以上的实体类别属性,任意两个节点之间的有向边用于表示边所连接的两个节点之间的关系。

4. 根据权利要求3所述的方法,其特征在于,所述若所述问句类别属于预定问句类别,则对所述问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息,包括:

若所述问句类别属于预定问句类别,则基于所述问题语句中各组成元素间的实际依存情况,生成所述问题语句对应的第一依存句法树;

其中,所述第一依存句法树中的每个节点包括所述问题语句的一组成元素,节点间的父子关系对应节点所包括的组成元素间的依存关系。

5. 根据权利要求4所述的方法,其特征在于,所述基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,得到所述问题

语句的结构化信息,包括:

根据所述第一依存句法树中各个节点对应的组成元素所属的句子成分,构建第二依存句法树;所述第二依存句法树中的相应节点至少包括相应组成元素所属的句子成分,节点间的父子关系对应节点所包括的句子成分间的依存关系;

对所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组;每个三元组包括所述问题语句的主体-谓词-宾语信息。

6. 根据权利要求5所述的方法,其特征在于,所述预定问句类别至少包括所述复杂关系属性型问题;

若所述问题语句所属的问句类别为复杂关系属性型问题,所述对所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组,包括:

基于预定的搜索原点确定方式,从所述第二依存句法树中确定出作为搜索原点的目标节点,所述目标节点对应包括的句子成分为实体类别或实体;

搜索所有离所述搜索原点距离最近的实体、实体类别及属性;所述离所述搜索原点距离最近的实体、实体类别或属性为:至所述搜索原点的节点路径上不包括任何其他实体、实体类别及属性的实体、实体类别或属性;

将搜索出的每个实体、实体类别或属性与所述搜索原点所对应的实体类别或实体进行匹配,得到相对应的至少一个二元组,并基于预先制定的关系库为每个二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及相应二元组整合为三元组,得到至少一个三元组;所述关系库为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库,所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系;

若所述第二依存句法树中存在未被搜索的实体、实体类别或属性,则确定未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性间的结构关系,所述结构关系为并列关系或嵌套关系;未被搜索的实体、实体类别或属性的上一级实体、实体类别或属性为:已被搜索的实体、实体类别或属性中与所述未被搜索的实体、实体类别或属性间的节点路径最短的实体、实体类别或属性;

若所述结构关系为并列关系,则将所述未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性对应的三元组中除所述上一级实体、实体类别或属性之外的其他二元组整合为三元组;

若所述结构关系为嵌套关系,则将所述未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性组合为二元组,并基于所述关系库为该二元组匹配出相对应的谓词关系,将匹配出的该谓词关系以及该二元组整合为三元组,直至所述第二依存句法树中不存在未被搜索的实体、实体类别及属性时结束。

7. 根据权利要求5所述的方法,其特征在于,所述预定问句类别还包括所述简单关系型问题及所述简单属性型问题;

若所述问题语句所属的问句类别为简单关系型问题或简单属性型问题,所述对所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组,包括:

基于预定的搜索原点确定方式,从所述第二依存句法树中确定出作为搜索原点的目标节点,所述目标节点对应包括的句子成分为实体类别或实体;

搜索所述第二依存句法树中除所述搜索原点外未被搜索的实体、实体类别或属性;

将搜索出的实体、实体类别或属性与搜索原点所对应的实体类别或实体进行匹配,得到二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及该二元组整合为三元组;所述关系库为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库,所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系。

8. 根据权利要求3所述的方法,其特征在于,所述预定问句类别不包括所述简单关系型问题及所述简单属性型问题,所述方法还包括:

若所述问题语句所属的问句类别为简单关系型问题,则将所述问题语句中已识别出的一个实体或实体类别与另一实体或实体类别组合为二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及该二元组整合为三元组,得到一个三元组;

若所述问题语句所属的问句类别为简单属性型问题,则将所述问题语句中已识别出的实体或实体类别与属性组合为二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及该二元组整合为三元组,得到一个三元组;

所述关系库为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库,所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系。

9. 一种服务器,其特征在于,包括:

存储器,用于至少存储一组指令集;

处理器,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行以下操作:

获得问题语句;

确定所述问题语句的句子成分;

基于所述问题语句的句子成分确定所述问题语句所属的问句类别;

若所述问句类别属于预定问句类别,则对所述问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息;并基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,得到所述问题语句的结构化信息;

所述问题语句所属的问句类别为简单关系型问题、简单属性型问题及复杂关系属性型问题中的任意一种;所述简单关系型问题中仅包括一条关系线;所述简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问;所述复杂关系属性型问题中包括不少于两条的关系线,或者包括不少于一条的关系线及实体属性;

其中,所述关系线,是指在利用一个谓词将一个实体/实体类别与另一个实体/实体类别进行衔接,使得两个实体/实体类别间产生谓词关系后所得的两个实体/实体类别与谓词间的对应关系。

一种数据处理方法及服务器

技术领域

[0001] 本发明属于基于大数据的数据处理技术领域,尤其涉及一种数据处理方法及服务器。

背景技术

[0002] 智能问答系统是在大规模知识处理基础上发展起来的一种面向行业应用的自动服务系统,其为企业与海量用户之间的沟通建立了一种基于自然语言处理的快捷有效的沟通途径。

[0003] 知识图谱问答系统,即为基于知识图谱的智能问答系统,目前,知识图谱问答系统采用的问句解析方法主要分为三类:语义解析法、信息抽取法和向量建模法。但是,这些方法处理复杂的中文问句的效果并不理想,对问句的解析效果较差,相应地会影响系统对用户问句意图的准确识别,进而降低了问答系统的服务质量。

发明内容

[0004] 有鉴于此,本发明的目的在于提供一种数据处理方法及服务器,以能够达到较好的问句句解析效果,进而准确地识别用户问句意图。

[0005] 为此,本发明公开如下技术方案:

[0006] 一种数据处理方法,包括:

[0007] 获得问题语句;

[0008] 确定所述问题语句的句子成分;

[0009] 基于所述问题语句的句子成分确定所述问题语句所属的问句类别;

[0010] 若所述问句类别属于预定问句类别,则对所述问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息;并基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,得到所述问题语句的结构化信息。

[0011] 上述方法,优选的,所述确定所述问题语句的句子成分,包括:

[0012] 确定所述问题语句的各组成元素中的实体及实体类别。

[0013] 上述方法,优选的,所述确定所述问题语句的句子成分,还包括:确定所述问题语句的各组成元素中的属性及疑问词,所述属性包括实体属性和/或实体类别属性;

[0014] 所述确定所述问题语句的各组成元素中的实体、实体类别、属性及疑问词,具体包括:

[0015] 利用预定的实体识别方法,识别所述问题语句的各组成元素中的实体;

[0016] 利用预定的疑问词识别方法,识别所述问题语句的各组成元素中的疑问词;

[0017] 将所述问题语句中未被识别的各组成元素分别与预定的实体类别库、属性库进行匹配,得到所述问题语句的各组成元素中的实体类别及属性;所述实体类别库及所述属性库为分别预先对知识图谱进行相对应的实体类别信息抽取以及实体属性和/或实体类别属

性信息抽取后所得的信息库;所述知识图谱为一预先基于多个语句样本所构建的包括多个节点的有向图,图中的每个节点对应一个实体或实体类别,每个实体对应零个或以上的实体属性、每个实体类别对应零个或以上的实体类别属性,任意两个节点之间的有向边用于表示边所连接的两个节点之间的关系。

[0018] 上述方法,优选的,所述问题语句所属的问句类别为简单关系型问题、简单属性型问题及复杂关系属性型问题中的任意一种,其中:

[0019] 所述简单关系型问题中仅包括一条关系线;

[0020] 所述简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问;

[0021] 所述复杂关系属性型问题中包括不少于两条的关系线,或者包括不少于一条的关系线及实体属性。

[0022] 上述方法,优选的,所述若所述问句类别属于预定问句类别,则对所述问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息,包括:

[0023] 若所述问句类别属于预定问句类别,则基于所述问题语句中各组成元素间的实际依存情况,生成所述问题语句对应的第一依存句法树;

[0024] 其中,所述第一依存句法树中的每个节点包括所述问题语句的一组成元素,节点间的父子关系对应节点所包括的组成元素间的依存关系。

[0025] 上述方法,优选的,所述基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,得到所述问题语句的结构化信息,包括:

[0026] 根据所述第一依存句法树中各个节点对应的组成元素所属的句子成分,构建第二依存句法树;所述第二依存句法树中的相应节点至少包括相应组成元素所属的句子成分,节点间的父子关系对应节点所包括的句子成分间的依存关系;

[0027] 对所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组;每个三元组包括所述问题语句的主体-谓词-宾语信息。

[0028] 上述方法,优选的,所述预定问句类别至少包括所述复杂关系属性型问题;

[0029] 若所述问题语句所属的问句类别为复杂关系属性型问题,所述对所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组,包括:

[0030] 基于预定的搜索原点确定方式,从所述第二依存句法树中确定出作为搜索原点的目标节点,所述目标节点对应包括的句子成分为实体类别或实体;

[0031] 搜索所有离所述搜索原点距离最近的实体、实体类别及属性;所述离所述搜索原点距离最近的实体、实体类别或属性为:至所述搜索原点的节点路径上不包括任何其他实体、实体类别及属性的实体、实体类别或属性;

[0032] 将搜索出的每个实体、实体类别或属性与所述搜索原点所对应的实体类别或实体进行匹配,得到相对应的至少一个二元组,并基于预先制定的关系库为每个二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及相应二元组整合为三元组,得到至少一个三元组;所述关系库为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库,所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系;

[0033] 若所述第二依存句法树中存在未被搜索的实体、实体类别或属性,则确定未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性间的结构关系,所述结构关系

为并列关系或嵌套关系；未被搜索的实体、实体类别或属性的上一级实体、实体类别或属性为：已被搜索的实体、实体类别或属性中与所述未被搜索的实体、实体类别或属性间的节点路径最短的实体、实体类别或属性；

[0034] 若所述结构关系为并列关系，则将所述未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性对应的三元组中除所述上一级实体、实体类别或属性之外的其他二元组整合为三元组；

[0035] 若所述结构关系为嵌套关系，则将所述未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性组合为二元组，并基于所述关系库为该二元组匹配出相对应的谓词关系，将匹配出的该谓词关系以及该二元组整合为三元组，直至所述第二依存句法树中不存在未被搜索的实体、实体类别及属性时结束。

[0036] 上述方法，优选的，所述预定问句类别还包括所述简单关系型问题及所述简单属性型问题；

[0037] 若所述问题语句所属的问句类别为简单关系型问题或简单属性型问题，所述对所述第二依存句法树进行搜索处理，得到所述问题语句的至少一个三元组，包括：

[0038] 基于预定的搜索原点确定方式，从所述第二依存句法树中确定出作为搜索原点的目标节点，所述目标节点对应包括的句子成分为实体类别或实体；

[0039] 搜索所述第二依存句法树中除所述搜索原点外未被搜索的实体、实体类别或属性；

[0040] 将搜索出的实体、实体类别或属性与搜索原点所对应的实体类别或实体进行匹配，得到二元组，并基于预先制定的关系库为该二元组匹配出相对应的谓词关系，将匹配出的谓词关系以及该二元组整合为三元组；所述关系库为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库，所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系。

[0041] 上述方法，优选的，所述预定问句类别不包括所述简单关系型问题及所述简单属性型问题，所述方法还包括：

[0042] 若所述问题语句所属的问句类别为简单关系型问题，则将所述问题语句中已识别出的一个实体或实体类别与另一实体或实体类别组合为二元组，并基于预先制定的关系库为该二元组匹配出相对应的谓词关系，将匹配出的谓词关系以及该二元组整合为三元组，得到一个三元组；

[0043] 若所述问题语句所属的问句类别为简单属性型问题，则将所述问题语句中已识别出的实体或实体类别与属性组合为二元组，并基于预先制定的关系库为该二元组匹配出相对应的谓词关系，将匹配出的谓词关系以及该二元组整合为三元组，得到一个三元组；

[0044] 所述关系库为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库，所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系。

[0045] 一种服务器，包括：

[0046] 存储器，用于至少存储一组指令集；

[0047] 处理器，用于调用并执行所述存储器中的所述指令集，通过执行所述指令集进行以下操作：

[0048] 获得问题语句；

[0049] 确定所述问题语句的句子成分；

[0050] 基于所述问题语句的句子成分确定所述问题语句所属的问句类别；

[0051] 若所述问句类别属于预定问句类别，则对所述问题语句进行依存句法分析，得到所述问题语句的各组成元素间的依存信息；并基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分，对所述问题语句进行结构化解析，得到所述问题语句的结构化信息。

[0052] 根据以上方案可知，本申请提供的数据处理方法、服务器及电子设备，在获得问题语句后，基于问题语句的句子成分确定问题语句所属的问句类别，并在问题语句属于预定问句类别的情况下，对问题语句进行依存句法分析，得到问题语句的各组成元素间的依存信息，进而基于问题语句的各组成元素间的依存信息以及所述问题语句的句子成分，对问题语句进行结构化解析，得到问题语句的结构化信息。通过利用问题语句的各组成元素间的依存信息以及问题语句的句子成分，对问题语句进行结构化解析，可达到较好的解析效果，进而可基于解析结果准确地识别用户问句意图。

附图说明

[0053] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的实施例，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，还可以根据提供的附图获得其他的附图。

[0054] 图1是本申请提供的一种数据处理方法实施例一的流程图；

[0055] 图2是本申请提供的一种数据处理方法实施例二的流程图；

[0056] 图3(a)-图3(b)是本申请实施例二提供的一问题语句的第一依存句法树及第二依存句法树的结构示意图；

[0057] 图4(a)-图4(b)是本申请实施例二提供的另一问题语句的第一依存句法树及第二依存句法树的结构示意图；

[0058] 图5是本申请实施例二提供的基于搜索第二依存句法树对复杂关系属性型问题进行三元组提取的流程示意图；

[0059] 图6(a)-图6(b)是本申请实施例二提供的又一问题语句的第一依存句法树及第二依存句法树的结构示意图；

[0060] 图7是本申请实施例三提供的基于搜索第二依存句法树对简单关系型问题/简单属性型问题进行三元组提取的流程示意图；

[0061] 图8是本申请实施例三提供的基于第二依存句法树对复杂关系属性型问题/简单关系型问题/简单属性型问题进行三元组提取的总体逻辑示意图；

[0062] 图9是本申请实施例三提供的通过在句子成分识别的基础上直接匹配关系库对简单关系型问题/简单属性型问题进行三元组提取的流程示意图；

[0063] 图10是本申请提供的一种服务器实施例五的结构示意图；

[0064] 图11是本申请提供的一种电子设备实施例九的结构示意图。

具体实施方式

[0065] 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,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。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前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0066] 为了达到较好的问句句子解析效果,进而准确地识别用户问句意图,本申请提供了一种数据处理方法及服务器,以下将通过多个实施例对本申请的数据处理方法及服务器进行说明。

[0067] 参考图1,是本申请提供的一种数据处理方法实施例一的流程图,该数据处理方法可应用于能够提供数据处理服务的本地/网络端的服务器或服务器集群中,或者还可以应用于智能手机、平板电脑、台式机、笔记本、一体机等各类型终端设备中。如图1所示,本实施例中,所述数据处理方法包括如下步骤:

[0068] 步骤101、获得问题语句。

[0069] 所述问题语句可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句,如用户在其智能手机、笔记本等终端设备中通过智能问答APP(Application,应用程序)的应用界面或网页提交至智能问答系统的问题语句等。

[0070] 相对应地,本步骤中获得问题语句,则可以是本申请方法的执行主体(如用户的终端设备或提供数据处理服务的服务器/服务器集群等)基于智能问答系统获得用户提交的问题语句,如获得用户提交的问题语句“白癜风是什么?”、“治疗感冒的药物有哪些?”、“什么药物可以治疗感冒引起的症状?”等等。其中,若所述执行主体为用户的终端设备本身,则终端设备可直接基于其相应问题输入/提交界面所提供的问题输入/提交功能获得用户所输入或提交的问题语句,若所述执行主体为后台服务器/服务器集群,如本地/网络端的服务器或服务器集群等,则后台服务器/服务器集群可基于局域网、互联网或移动网络等网络的通信功能获得用户在其终端设备所提交的问题语句。

[0071] 需要说明的是,实际应用中,所述用户可以是自然人用户或者基于人工智能的非自然人用户,这里,并不对此进行限定。

[0072] 步骤102、确定所述问题语句的句子成分。

[0073] 本申请中,问题语句的句子成分,是指问题语句的各组成元素所属的实体、实体类别等成分,其中,实体一般是指能够独立存在的、作为属性的基础的具体事物,对于一个句子来说,可以将句子中包括的能够指代具体事物的词语看做属于实体,如“阿莫西林”、“华西医院”、“感冒”等词语都可以认为其属于实体;实体类别则是指具体事物所属的类别,对于一个句子来说,可以将句子中包括的能够指代具体事物所属的类别的词语看做属于实体类别,如“药物”、“医院”、“疾病”等这样的类别词均属于实体类别。

[0074] 相对应地,本步骤中,确定问题语句的句子成分,可以是指确定问题语句的各组成元素(如构成句子的各个分词)中的实体及实体类别。

[0075] 但不限于此,除了实体、实体类别,问题语句的句子成分还可以包括属性、疑问词等成分,其中,所述属性可以是实体属性或实体类别属性,具体地,句子中所包括的关于实体在某一方面的属性的描述词可以认为其对应的句子成分为实体属性,如句子中包括的关于“阿莫西林”的生产日期、功效介绍或使用方法等方面的属性的描述词等;相对应地,句子

中所包括的关于实体类别在某一方面的属性的描述词则可以认为其对应的句子成分为实体类别属性,如句子中包括的关于“医院”的功能定位、性质、位置等方面的属性的描述词等。

[0076] 所述疑问词是指句子中所包括的用于引出提问需求的词语,如“什么”、“哪些”、“哪个”、“哪儿”、“几个”、“如何”等等。

[0077] 从而,本步骤中,确定问题语句的句子成分,还可以是指确定问题语句的各组成元素(如构成句子的各个分词)中的实体、实体类别、属性及疑问词等。

[0078] 具体实施中,在确定问题语句的句子成分时,可以并不以本实施例提供的上述几种具体的句子成分为限,而是可由技术人员根据实际需求来预置具体需确定的句子成分。

[0079] 步骤103、基于所述问题语句的句子成分确定所述问题语句所属的问句类别。

[0080] 问句类别可以包括但不限于简单关系型问题、简单属性型问题及复杂关系属性型问题。

[0081] 其中,所述简单关系型问题中仅包括一条关系线;所述简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问;所述复杂关系属性型问题中包括不少于两条的关系线,或者包括不少于一条的关系线及实体属性。

[0082] 一条关系线,是指在利用一个谓词将一个实体/实体类别与另一个实体/实体类别进行衔接,使得两个实体/实体类别间产生谓词关系后所得的两个实体/实体类别与谓词间的对应关系。

[0083] 例如问题语句“治疗感冒的药物有哪些?”中包括一个实体“感冒”,一个实体类别“药物”,且实体“感冒”与实体类别“药物”通过谓词“治疗”产生谓词关系,从而该问题语句包括一条关系线;问题语句“什么药可以治疗感冒”中同样包括一个实体“感冒”,一个实体类别“药物”(以疑问词“什么(药)”指代的“药物”这一实体类别),且实体“感冒”与实体类别“药物”通过谓词“治疗”产生谓词关系,从而该问题语句同样包括一条关系线。在识别问题语句所属的问句类别时,若通过上述步骤(即识别句子成分的步骤)已确定出句子中共包括两个实体/实体类别,则根据句子的句法特征,一般会认为两个实体/实体类别之间需要通过一谓词关系进行衔接,从而,可确定出该问题语句所属的问句类别为简单关系型问题。

[0084] 若问题语句中不包括关系线,而只对实体或实体类别的属性进行提问,例如问题语句“白癜风是什么?”中不包括关系线,仅对“白癜风”这一实体的“introduction”属性进行提问,“医院在哪里?”中不包括关系线,仅对“医院”这一实体类别的位置属性进行提问,则此类问题语句所属的问句类别为简单属性型问题,具体实施中,对于服务器/服务器集群或终端设备等执行主体来说,可在问题语句中仅包括一个实体/实体类别,以及一个由疑问词指代的属性(如“白癜风是什么”中包括由疑问词“什么”所指代的“白癜风”的“introduction”属性,“医院在哪里?”中包括由疑问词“哪里”所指代的“医院”的位置属性)的情况下,将问题语句的问句类别识别为所述简单属性型问题。

[0085] 一个问题语句,若其不属于上述的简单关系型问题,也不属于上述的简单属性型问题,则将其所属的问句类别确定为复杂关系属性型问题。

[0086] 当然,具体实施中,也可以在识别问题语句中所包括的实体、实体类别、属性及疑问词等句子成分的基础上,同时识别其中的谓词关系,进而结合所识别出的实体、实体类别、属性、疑问词及谓词关系等来确定句子中是否包括关系线、在包括关系线情况下具体包

括的关系线条数以及是否包括对属性的提问等信息,进而确定出问题语句所属的问句类别,本实施例对确定问题语句所属问句类别的实现过程不作限定。

[0087] 步骤104、若所述问句类别属于预定问句类别,则对所述问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息;并基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,得到所述问题语句的结构化信息。

[0088] 所述预定问句类别,可以根据实际实施情况对其进行设定,例如将其设定为包括上述的简单关系型问题、简单属性型问题及复杂关系属性型问题中的全部类别;或者将其设定为仅包括上述三种类别中的其中之一(例如,仅包括所述复杂关系属性型问题)等等,本实施例对此不作限定。

[0089] 本步骤可利用相应依存句法分析算法或依存句法分析器,分析问题语句的依存句法结构信息,以得到问题语句的各组成元素间的依存信息,问题语句的各组成元素间的依存信息能够反映问题语句中各组成元素间的实际依存(依赖)情况。

[0090] 例如问题语句“今天天气怎么样”,通过依存句法分析,可以将其解析成“今天-天气-怎么样”,其中,“今天”依赖于“天气”,“天气”依赖于“怎么样”,而“怎么样”则不依赖于任何词语。

[0091] 在对属于预定问句类别的问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息后,本申请基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,以得到所述问题语句的结构化信息,该对问题语句进行结构化解析,从而得到问题语句的结构化信息的处理过程会在之后的实施例中进行详述,具体请参见后续实施例对该部分的说明。

[0092] 根据以上方案可知,本实施例提供的数据处理方法,在获得问题语句后,基于问题语句的句子成分确定问题语句所属的问句类别,并在问题语句属于预定问句类别的情况下,对问题语句进行依存句法分析,得到问题语句的各组成元素间的依存信息,进而基于问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对问题语句进行结构化解析,得到问题语句的结构化信息。通过利用问题语句的各组成元素间的依存信息以及问题语句的句子成分,对问题语句进行结构化解析,可达到较好的解析效果,进而可基于解析结果准确地识别用户问句意图。

[0093] 以下通过另一实施例继续对上述实施例中的数据处理方法进行进一步详述,参考图2,是本申请提供的一种数据处理方法实施例二的流程图,本实施例二中,所述数据处理方法可以通过如下的处理过程实现:

[0094] 步骤201、获得问题语句。

[0095] 所述问题语句可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句,如用户在其智能手机、笔记本等终端设备中通过智能问答APP的应用界面或网页提交至智能问答系统的问题语句等。

[0096] 相对应地,本步骤中获得问题语句,则可以是本申请方法的执行主体(如用户的终端设备或提供数据处理服务的服务器/服务器集群等)基于智能问答系统获得用户提交的问题语句,如获得用户提交的问题语句“白癜风是什么?”、“治疗感冒的药物有哪些?”、“什么药物可以治疗感冒引起的症状?”等等。其中,若所述执行主体为用户的终端设备本身,则

终端设备可直接基于其相应问题输入/提交界面所提供的问题输入/提交功能获得用户所输入或提交的问题语句,若所述执行主体为后台服务器/服务器集群,如本地/网络端的服务器或服务器集群等,则后台服务器/服务器集群可基于局域网、互联网或移动网络等网络的通信功能获得用户在其终端设备所提交的问题语句。

[0097] 需要说明的是,实际应用中,所述用户可以是自然人用户或者基于人工智能的非自然人用户,这里,并不对此进行限定。

[0098] 步骤202、利用预定的实体识别方法,识别所述问题语句的各组成元素中的实体。

[0099] 可基于相应的实体识别技术,识别问题语句的各组成元素中的实体。

[0100] 具体地,对于实体识别,目前已存在很多可用算法,例如CRF(conditional random field,条件随机场)、HMM(Hidden Markov Model,隐马尔可夫模型)等算法,从而,可选取这些算法中的其中一种对问题语句中的实体进行识别,或者亦可以通过模式匹配的方式进行实体识别,这里不局限于一种技术。

[0101] 步骤203、利用预定的疑问词识别方法,识别所述问题语句的各组成元素中的疑问词。

[0102] 如前文所述,所述疑问词是指句子中所包括的用于引出提问需求的词语,如“什么”、“哪些”、“哪个”、“哪儿”、“几个”、“如何”等等。

[0103] 由于疑问词数量有限,具体实施中可以通过模式匹配的方式对问题语句中的疑问词进行识别。

[0104] 步骤204、将所述问题语句中未被识别的各组成元素分别与预定的实体类别库、属性库进行匹配,得到所述问题语句的各组成元素中的实体类别及属性。

[0105] 所述实体类别库及所述属性库为分别预先对知识图谱进行相对应的实体类别信息抽取以及实体属性和/或实体类别属性信息抽取后所得的信息库;所述知识图谱为一预先基于多个语句样本所构建的包括多个节点的有向图,图中的每个节点对应一个实体或实体类别,每个实体对应零个或以上的实体属性、每个实体类别对应零个或以上的实体类别属性,任意两个节点之间的有向边用于表示边所连接的两个节点之间的关系。

[0106] 具体实施中,可预先采集所需业务领域的一系列语句样本(如问题语句或问答句等)作为语料集,并基于所采集的语料集构建上文所述的知识图谱,其中,所述知识图谱可以是基于某一特定业务领域的语料集所构建的对应于该特定业务领域的知识图谱,也可以是基于多个业务领域的语料集所构建的对应于多个业务领域的知识图谱,本实施例对此不作限定。

[0107] 在构建出所述知识图谱的基础上,可进一步通过对所述知识图谱进行实体类别信息抽取来得到相对应的实体类别库,以及对所述知识图谱进行实体属性和/或实体类别属性信息抽取来得到相对应的属性库。

[0108] 其中,所述实体类别库及属性库,可以是分别通过对对应于某一特定业务领域的知识图谱进行实体类别信息抽取以及属性信息抽取,所获得的对应于该特定业务领域的信息库;或者还可以是分别通过对对应于多个业务领域的知识图谱进行实体类别信息抽取及属性信息抽取,所获得的对应于多个业务领域的信息库;或者,还可以是分别通过对对应于多个业务领域的知识图谱进行某一特定业务领域的实体类别信息抽取及属性信息抽取,所获得的对应于该特定业务领域的信息库,本实施例对此同样不作限定。

[0109] 在构建出知识图谱、并基于对知识图谱进行相应信息抽取构建出实体类别库及属性库的基础上,当对问题语句进行句子成分识别时,具体可通过将问题语句中的各组成元素与所述实体类别库或属性库进行匹配,来识别出问题语句的各组成元素中的实体类别或属性(实体属性和/或实体类别属性)。

[0110] 例如对于问题语句“四川省有哪些医院?”,通过将其各组成元素与实体类别库进行匹配,可识别出其中的组成元素“医院”所属的句子成分为实体类别,对于问句“游乐场在哪个位置?”,通过将其各组成元素与属性库进行匹配,可识别出其中的组成元素“(哪个)位置”所属的句子成分为实体类别属性,且其具体为实体类别的位置属性。

[0111] 步骤205、基于所述问题语句的句子成分确定所述问题语句所属的问句类别。

[0112] 如上文所述,问句类别可以包括但不限于简单关系型问题、简单属性型问题及复杂关系属性型问题。其中,所述简单关系型问题中仅包括一条关系线;所述简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问;所述复杂关系属性型问题中包括不少于两条的关系线,或者包括不少于一条的关系线及实体属性。

[0113] 在已确定出问题语句所包括的实体、实体类别、属性及疑问词等句子成分的基础上,可进一步基于所识别出的句子成分来确定问题语句所属的问句类别具体是上述问句类别中的哪一种。

[0114] 其中,若确定出句子中共包括两个实体/实体类别,则根据句子的句法特征,一般会认为两个实体/实体类别之间需要通过一谓词关系进行衔接,从而,可确定出该问题语句中包括一条关系线,相对应地可确定出该问题语句所属的问句类别为简单关系型问题。如问题语句“治疗感冒的药物有哪些?”中包括一个实体“感冒”以及一个实体类别“药物”,则可认为该实体“感冒”及实体类别“药物”之间需要通过一谓词关系进行衔接,从而该问题语句中包括一条关系线,相对应地,其所属的问句类别为简单关系型问题,这与该问题中“感冒”及“药物”之间通过谓词“治疗”来进行衔接进而构成一条关系线的实际情况也是相符的。

[0115] 若问题语句中仅包括一个实体/实体类别,以及一个疑问词指代的属性,则认为问题语句中不包括关系线,仅包括对实体/实体类别的属性的提问,从而可将该问题语句的问句类别识别为所述简单属性型问题。例如问题语句“白癜风是什么?”中仅包括一个实体“白癜风”以及一个疑问词“什么”指代的“白癜风”的“introduction”属性,本质上属于对“白癜风”这一实体的“introduction”属性进行提问,不包括任何关系线,问题语句“医院在哪里?”仅包括一个实体类别“医院”以及一个疑问词“哪里”指代的“医院”的位置属性,本质上属于对“医院”这一实体类别的位置属性进行提问,不包括任何关系线,则此类问题语句所属的问句类别为简单属性型问题。

[0116] 一个问题语句,若其不属于上述的简单关系型问题,也不属于上述的简单属性型问题,则将其所属的问句类别确定为复杂关系属性型问题。

[0117] 当然,具体实施中,也可以在识别问题语句中所包括的实体、实体类别、属性及疑问词等句子成分的基础上,同时识别其中的谓词关系,进而结合所识别出的实体、实体类别、属性、疑问词及谓词关系等来确定句子中是否包括关系线、在包括关系线情况下具体包括的关系线条数以及是否包括对属性的提问等信息,进而确定出问题语句所属的问句类别,本实施例对确定问题语句所属问句类别的实现过程不作限定。

[0118] 实际应用中,参考表1,所述简单关系型问题可以分为查询型及判断型两种,而所述查询型的简单关系型问题又可以分为列表型及聚合型两种,每种类型的相关示例及对应的句子结构具体可参见以下的表1所示:

[0119] 表1

	分类	关系	问句示例	
[0120]	查询	具体实体+关系->实体类别	列表型	治疗感冒的药物有哪些?
				感冒+治疗->药物
			聚合型	治疗感冒的药物有多少种?
				感冒+治疗->药物种类
	判断	具体实体-关系-具体实体	阿莫西林可以治疗感冒吗?	
			阿莫西林-治疗-感冒	

[0121] 复杂关系型问题又可以分为嵌入型、并列型及属性型,每种类型的相关示例及对应的句子结构具体可参见以下的表2所示:

[0122] 表2

	分类	问句示例
[0123]	嵌入型	什么药物可以治疗感冒引起的症状?
		(具体实体+关系->实体)+关系->实体
[0124]	并列型	什么药物可以治疗发烧和头痛?
		(具体实体+关系->实体)^(具体实体+关系->实体)
	属性型	三甲医院能够治疗哪些疾病?
		(实体类别+属性->具体实体)+关系->实体

[0125] 需要说明的是,具体实施中,可以按照上文所述的在针对问题语句识别出实体、实体类别、属性及疑问词等句子成分的基础上(不识别其中的谓词关系),通过推导其关系线情况来获知其所所属的问句类别,或者也可以既识别问题语句的实体、实体类别、属性及疑问词,又识别其中的谓词关系,在此基础上,对其进行句子结构解析,并基于解析出的句子结构识别其对应所属的问句类别,并不局限于其中的一种。

[0126] 步骤206、若所述问句类别属于预定问句类别,则基于所述问题语句中各组成元素间的实际依存情况,生成所述问题语句对应的第一依存句法树;其中,所述第一依存句法树中的每个节点包括所述问题语句的一组成元素,节点间的父子关系对应节点所包括的组成元素间的依存关系。

[0127] 本申请中,所述预定问句类别至少包括所述复杂关系属性型问题,具体地,例如包括上述的简单关系型问题、简单属性型问题及复杂关系属性型问题中的全部类别,或者仅包括所述复杂关系属性型问题等,实际应用中,可根据具体实施情况进行设定。

[0128] 在问题语句的问句类别属于预定问句类别的情况下,本步骤可利用相应依存句法分析算法或依存句法分析器,分析问题语句的依存句法结构信息,以得到问题语句的各组成元素间的依存信息,其中,问题语句的各组成元素间的依存信息能够反映问题语句中各组成元素间的实际依存(依赖)情况,进而基于问题语句的各组成元素间的依存信息,生成所述问题语句对应的第一依存句法树。

[0129] 例如问题语句“今天天气怎么样?”,通过依存句法分析,可以将其解析成“今天-天气-怎么样”,其在句子中的ID(Identity,身份标识号)可按序表示成[1,2,3],其中,“1”表示“今天”,“2”表示“天气”,“3”表示“怎么样”;各成分之间的依存关系可以表示成[2,3,0],即“1”依赖于“2”,“2”依赖于“3”,“3”依赖于“0”,依赖于“0”表示不依赖任何词,也即“天气”支配了“今天”,“怎么样”支配了“天气”。

[0130] 在通过对问题语句进行依存句法信息分析,得到问题语句的各组成元素间的依存信息后,可进一步基于问题语句的各组成元素间的依存信息,生成所述问题语句对应的第一依存句法树,其中,所述第一依存句法树中的每个节点包括所述问题语句的一组成元素,节点间的父子关系对应节点所包括的组成元素间的依存关系。例如,根据问题语句“今天天气怎么样?”的上述依存关系,可知“2”依赖于“3”,则在所构建的该问题语句的第一依存句法树中,“2”为“3”的子节点,即“天气”为“怎么样”的子节点,同理,“1”依赖于“2”,则“1”为子节点,“2”为父节点,也即,“今天”为“天气”的子节点。

[0131] 步骤207、根据所述第一依存句法树中各个节点对应的组成元素所属的句子成分,构建第二依存句法树。所述第二依存句法树中的相应节点至少包括相应组成元素所属的句子成分,节点间的父子关系对应节点所包括的句子成分间的依存关系。

[0132] 本步骤具体可根据问题语句所对应的第一依存句法树,并结合问题语句中各组成元素所属的句子成分(如实体、实体类别、属性或疑问词等),来构建问题语句的第二依存句法树。其中,在基于问题语句的第一依存句法树来构建其第二依存句法树时,对于第一依存句法中的每个节点,可在第二依存句法树中为其构建一相对应的节点,且第二依存句法树中所构建节点的节点值具体为其在第一依存句法中相对应节点的组成元素所属的句子成分(当然其取值也可以同时包括其第一依存句法中相对应节点的组成元素及该组成元素所对应的具体成分,本申请对此不作限制),实际应用中,第一依存句法树中可能存在某些未被识别出所属句子成分的组成元素的节点,对于此类节点,其在第二依存句法树中所对应的节点的节点值则可直接取该组成元素的元素信息,也即维持其第一依存句法树中所对应节点的节点值。

[0133] 参考图3(a),是本实施例提供的问题语句“四川省有哪些医院?”所对应的第一依存句法树的树型结构示意图,在构建其第二依存句法树时,对于“四川省”这一组成元素,可按其在第一依存句法树中的节点位置在第二依存句法树中为其创建一相应节点,且由于“四川省”为地点实体,从而可将创建的该节点的节点值表示为“region”(即地点实体);“医院”为实体类别,则在按“医院”在第一依存句法树中的节点位置在第二依存句法树中为其创建一相应节点后,可将所创建的该节点的节点值表示为“hospital_class”;“哪些”为疑问词,则在按“哪些”在第一依存句法树中的节点位置在第二依存句法树中为其创建一相应节点后,可将所创建的该节点的节点值表示为“what_some”;而在对问题语句中的各组成元素进行实体、实体类别、属性及疑问词的识别时,未识别“有”这一组成元素所属的句子成

分,则在构建第二依存句法树时,该组成元素“有”在第二依存句法树中所对应的节点位置及节点值仍维持不变,即与其在第一依存句法树中所对应的节点位置及节点值一致,对于问题语句“四川省有哪些医院?”,其第二依存句法树的树型结构及树中各节点的取值情况具体可参考图3(b)所示。

[0134] 再比如问题语句“阿莫西林可以治疗发烧和头痛吗?”,其对应的第一依存句法树如图4(a)所示,在构建其所对应的第二依存句法树时,在按“阿莫西林”在其第一依存句法树中所对应的节点位置为其创建一相对应的节点后,由于“阿莫西林”属于药品实体,从而,可将该节点的节点值表示为“medicine”,相对应地,“发烧”和“头痛”为症状实体,可分别将两者在第二依存句法树中所对应的节点的节点值表示为“symptom1”和“symptom2”,“治疗”是药物和症状之间的关系,表示为“indication”,而其他组成元素如“可以”、“吗”由于在句子成分识别时未识别其对应的句子成分,从而,仍维持第一依存句法树中的节点值不变(即仍采用“可以”、“吗”),所创建的第二依存句法树具体如图4(b)所示。

[0135] 步骤208、对所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组;每个三元组包括所述问题语句的主体-谓词-宾语信息。

[0136] 在基于问题语句的各组成元素所属的句子成分及问题语句所对应的第一依存句法树,构建出问题语句所对应的第二依存句法树的基础上,可进一步基于问题语句所对应的第二依存句法树,对问题语句进行结构化解析,以得到问题语句的结构化信息。

[0137] 本实施例具体通过对问题语句的第二依存句法树进行预定的搜索处理,来得到问题语句的至少一个SPO(Subject-Predicate-Object,主体-谓词-宾语)三元组,以此实现对问题语句的结构化解析,并相应地以所得到的至少一个SPO三元组作为问题语句的结构化信息。

[0138] 如上文所述,本实施例中,所述预定问句类别至少包括所述复杂关系属性型问题,从而,当问题语句所属的问句类别为所述复杂关系属性型问题时,则该问题语句所属的问句类别符合所述预定问句类别,从而,可通过对该问题语句的第二依存句法树进行搜索处理,来得到该问题语句的至少一个SPO三元组。

[0139] 其中,参考图5,当问题语句所属的问句类别为所述复杂关系属性型问题时,具体可通过以下的处理过程实现对复杂关系属性型的问题语句进行三元组提取:

[0140] 步骤501、基于预定的搜索原点确定方式,从所述第二依存句法树中确定出作为搜索原点的目标节点,所述目标节点对应包括的句子成分为实体类别或实体。

[0141] 具体地,可根据以下的搜索原点选择规则从问题语句的第二依存句法树中确定出用于作为搜索原点的目标节点:

[0142] a) 若问题语句中只包含一个实体类别,则将该实体类别作为搜索原点;

[0143] b) 若问题语句中包含两个及以上的实体类别,则选择与疑问词最近的一个实体类别作为搜索原点;若没有疑问词,则选择离句末最近的实体类别词作为搜索原点;

[0144] c) 若句子中不包含实体类别,则查看实体个数:其中,若只包含一个具体实体,则将其作为搜索原点;若包含两个及以上个具体实体,则优先选择离根节点最近的左子树(若并非二叉树,则以相对的左右为准)上的节点作为搜索原点;

[0145] d) 若问题语句中既无实体类别也无具体实体,则可能是因为该问题语句本身和该领域的知识图谱无关,也可能是实体识别时未将实体正确识别出来,对于这样的问句,暂

时不处理(也即按结构化解析失败处理)。

[0146] 比如,对于问题语句“阿莫西林可以治疗发烧和头痛吗?”,其对应的第一依存句法树及第二依存句法树分别如图4(a)及图4(b)所示,对于该问题语句,根据上述的搜索原点选择规则,具体可选择图4(b)中“阿莫西林”对应的节点“medicine”作为搜索原点(匹配上述规则中的第c)条规则),也即,以该节点为搜索起始点对第二依存句法树进行后续的搜索处理。再比如,对于问题语句“卖阿莫西林的药厂卖药给哪些医院?”的第一依存句法树及第二依存句法树分别如图6(a)及图6(b)所示,对于该问题语句,根据上述的搜索原点选择规则,可确定出距离疑问词节点“what_some”最近的实体类别节点“hospital_class”为搜索原点。

[0147] 步骤502、搜索所有离所述搜索原点距离最近的实体、实体类别及属性。

[0148] 在确定出搜索原点的基础上,可继续搜索问题语句的第二依存句法树中所有离所述搜索原点距离最近的实体、实体类别及属性,本申请中,所述离所述搜索原点距离最近的实体、实体类别或属性为:至所述搜索原点的节点路径上不包括任何其他实体、实体类别及属性的实体、实体类别或属性。

[0149] 仍以问题语句“阿莫西林可以治疗发烧和头痛吗?”为例,根据图4(b)示出的该问题语句的第二依存句法树,可知,与所述搜索原点“medicine”距离最近的实体、实体类别或属性具体为“发烧”所对应的实体节点“symptom1”,从而可将该节点作为当前待搜索的节点。

[0150] 再以问题语句“卖阿莫西林的药厂卖药给哪些医院?”为例,根据图6(b)示出的该问题语句的第二依存句法树,可知,与所述搜索原点“hospital_class”距离最近的实体、实体类别或属性具体为“药”所对应的实体类别节点“medicine_class”,从而可将该节点作为当前待搜索的节点。

[0151] 步骤503、将搜索出的每个实体、实体类别或属性与所述搜索原点所对应的实体类别或实体进行匹配,得到相对应的至少一个二元组,并基于预先制定的关系库为每个二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及相应二元组整合为三元组,得到至少一个三元组。

[0152] 本实施例中,所述关系库即为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库,所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系。

[0153] 具体地,所述关系库,可以是通过对应于某一特定业务领域的知识图谱进行关系信息抽取,所获得的对应于该特定业务领域的信息库;或者还可以是通过对应于多个业务领域的知识图谱进行关系信息抽取,所获得的对应于多个业务领域的关系库;或者,还可以是通过对应于多个业务领域的知识图谱进行某一特定业务领域的关系信息抽取,所获得的对应于该特定业务领域的关系库,本实施例对此同样不作限定。

[0154] 针对问题语句,在其第二依存句法树中确定出作为搜索原点的目标节点,并搜索所有离所述搜索原点距离最近的实体、实体类别或属性后,可将搜索出的离所述搜索原点距离最近的每个实体、实体类别或属性与所述搜索原点进行匹配,得到相对应的至少一个二元组,并可通过搜索上述的关系库,为每个二元组匹配出相对应的谓词关系,进而将匹配出的谓词关系以及相应二元组整合为三元组,以此得到问题语句的至少一个三元组。

[0155] 比如,针对问题语句“阿莫西林可以治疗发烧和头痛吗?”,如上文所述,通过节点搜索,可确定出与搜索原点“medicine”距离最近的实体、实体类别或属性具体为“发烧”所对应的实体节点“symptom1”,从而可将该节点“symptom1”与搜索原点“medicine”匹配为一二元组(medicine,symptom1),通过对基于知识图谱所提取的上述关系库进行搜索可以确定出两者之间的关系为“indication”(治疗),从而可得到三元组(medicine,indication,symptom1),实际实施中,可继续将该三元组中的实体替换为其在问题语句中所对应的具体组成元素,从而得到三元组(阿莫西林,indication,发烧),以便于智能问题系统基于所提取的三元组对用户意图进行准确识别。

[0156] 相对应地,针对问题语句“卖阿莫西林的药厂卖药给哪些医院?”,通过节点搜索,可确定出与搜索原点“hospital_class”距离最近的实体、实体类别或属性具体为“药”所对应的实体类别节点“medicine_class”以及“factory_class”,从而可将节点“medicine_class”与搜索原点“hospital_class”匹配为一二元组(hospital_class,medicine_class),将节点“factory_class”与搜索原点“hospital_class”匹配为一二元组(factory_class,hospital_class),之后通过对基于知识图谱所提取的上述关系库进行搜索可以确定出前一二元组之间的关系为“buy”(买),从而可得到三元组(hospital_class,buy,medicine_class),后一二元组所对应的关系为“sale to”,从而可得到三元组(factory_class,sale to,hospital_class)。实际应用中,可以将上述三元组中的实体类别替换为其在问题语句中所对应的具体组成元素,从而得到三元组(医院,buy,药)、(药厂,sale to,医院)并输出给智能问答系统。

[0157] 步骤504、若所述第二依存句法树中存在未被搜索的实体、实体类别或属性,则确定未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性间的结构关系,所述结构关系为并列关系或嵌套关系;未被搜索的实体、实体类别或属性的上一级实体、实体类别或属性为:已被搜索的实体、实体类别或属性中与所述未被搜索的实体、实体类别或属性间的节点路径最短的实体、实体类别或属性。

[0158] 具体地,在上述处理的基础上,若第二依存句法树中仅存在一个未被搜索的实体、实体类别或属性,则可直接对该未被搜索的实体、实体类别或属性进行搜索操作,并确定该一个未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性间的结构关系;若第二依存句法树中存在一个以上未被搜索的实体、实体类别或属性,则优先选择未被搜索的各实体、实体类别或属性节点中与已被搜索的实体、实体类别或属性节点的节点距离最近的实体、实体类别或属性节点作为当前待搜索的节点,并确定该前待搜索的节点与其上一级实体、实体类别或属性间的结构关系。这里,某一未被搜索的实体、实体类别或属性节点与已被搜索的实体、实体类别或属性节点的节点距离是指未被搜索的实体、实体类别或属性节点与已被搜索的各实体、实体类别或属性节点的节点距离中的最短距离。

[0159] 比如,针对图4(b)示出的问题语句“阿莫西林可以治疗发烧和头痛吗?”的第二依存句法树,在确定出搜索原点“medicine”,并基于该搜索原点完成对节点“symptom1”的搜索处理后,仍存在一个实体节点“symptom2”未被处理,从而,可确定该实体节点“symptom2”与其上一级节点“symptom1”的结构关系,由于“symptom2”与“symptom1”为并列的两个症状实体,从而两者之间的结构关系为并列关系。

[0160] 步骤505、若所述结构关系为并列关系,则将所述未被搜索的实体、实体类别或属

性与其上一级实体、实体类别或属性对应的三元组中除所述上一级实体、实体类别或属性之外的其他二元组整合为三元组。

[0161] 具体地,例如,问题语句“阿莫西林可以治疗发烧和头痛吗?”中“symptom2”与其上一级节点“symptom1”为并列关系,从而,可将该节点“symptom2”与“symptom1”对应的三元组(medicine,indication,symptom1)中的二元组(medicine,indication)进行整合,得到该节点“symptom2”对应的三元组(medicine,indication,symptom2),至此,问题语句“阿莫西林可以治疗发烧和头痛吗?”的第二依存句法树中已不存在未被搜索的实体、实体类别或属性节点,可结束对该问题语句的第二依存句法树的搜索过程,对于该问题语句,最终可提取出两个三元组:

[0162] (medicine,indication,symptom1);

[0163] (medicine,indication,symptom2)。

[0164] 具体实施中,为便于智能问答系统对用户问句意图进行理解,可将上述三元组中的实体替换为其在问题语句中所对应的具体组成元素,从而得到三元组(阿莫西林,indication,发烧)及(阿莫西林,indication,头痛)。

[0165] 步骤506、若所述结构关系为嵌套关系,则将所述未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性组合为二元组,并基于所述关系库为该二元组匹配出相对应的谓词关系,将匹配出的该谓词关系以及该二元组整合为三元组,直至所述第二依存句法树中不存在未被搜索的实体、实体类别及属性时结束。

[0166] 具体地,例如,问题语句“卖阿莫西林的药厂卖药给哪些医院”中“factory_class”与其上一级节点“medicine_class”为嵌套关系,从而,可将该节点“factory_class”与其上一级节点“medicine_class”组合为二元组(factory_class,medicine_class),通过搜索关系库,可知“factory_class”与 medicine_class”之间的关系为“sale”,从而可得到三元组(factory_class,sale, medicine_class)。

[0167] 在完成对实体类别节点“factory_class”的搜索处理后,根据问题语句“卖阿莫西林的药厂卖药给哪些医院”的第二依存句法树,可知,其中,还存在实体节点“medicine”未被搜索,该节点“medicine”与其上一级节点“factory_class”的关系为嵌套关系,从而可将两者匹配为二元组(factory_class,medicine),通过搜索关系库,可进一步确定出三元组(factory_class,sale,medicine)。

[0168] 从而,对于问题语句“卖阿莫西林的药厂卖药给哪些医院?”,最终可提取出四个三元组:

[0169] (hospital_class,buy,medicine_class);

[0170] (factory_class,sale to,hospital_class);

[0171] (factory_class,sale,medicine_class);

[0172] (factory_class,sale,medicine)。

[0173] 具体实施中,为便于智能问答系统对用户问句意图进行理解,可将上述三元组中的实体/实体类别替换为其在问题语句中所对应的具体组成元素,从而可得到三元组:(医院,buy,药),(药厂,sale to,医院)(药厂,sale,药),(药厂,sale,阿莫西林)。

[0174] 需要说明的是,实际应用中,包括两个实体类别的三元组的实用性较低,如上述的(factory_class,sale to,hospital_class)及(factory_class,sale, medicine_class),

从而可考虑将其过滤掉。

[0175] 还需要说明的是,本实施例中,所述附图4(b)及6(b)中节点与节点间的边(非树的自身的边,在附图中未带箭头)上的关系,如图4(b)中边上的两个“indication”,图6(b)中边上的两个“sale”、一个“sale to”及一个“buy”,均是通过关系库为边的两端节点的词所匹配出来的。

[0176] 本实施例通过搜索复杂关系型问题的第二依存句法树,并基于关系库为搜索过程中所确定出的实体/实体类别/属性与实体/实体类别/属性的二元组匹配相对应的谓词关系,得到复杂关系型问题语句的三元组,实现了利用问题语句的各组成元素间的依存信息以及问题语句的句子成分,对复杂关系型问题的问题语句进行结构化解析,可达到较好的解析效果,进而可基于解析结果准确地识别用户问句意图。由于在进行问句结构解析时,不需依赖于句子的语法规则,且不要求句子必须遵循某一特定的表达形式(如不要求句子必须是非口语化的句子),从而对于不符合标准语法规则和/或采用口语化表达形式的复杂句子,仍能达到较好的解析效果(对于口语化句子可在构建知识图谱时,同时搜集口语化表达形式的语料,使其能够适应该类问题语句)。

[0177] 在接下来的实施例三中,所述预定问句类别除了包括所述复杂关系属性型问题,还包括简单关系型问题及所述简单属性型问题。

[0178] 基于预定问句类别的上述内容,若问题语句所属的问句类别为简单关系型问题或简单属性型问题,则问题语句所属的问句类别符合所述预定问句类别。此种情况下,即问题语句所属的问句类别为简单关系型问题或简单属性型问题的情况下,参考图7,所述步骤208对所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组,可以通过以下的处理过程实现:

[0179] 步骤701、基于预定的搜索原点确定方式,从所述第二依存句法树中确定出作为搜索原点的目标节点,所述目标节点对应包括的句子成分为实体类别或实体。

[0180] 上文中提供的针对复杂关系属性型问题的搜索原点选择规则对于简单关系型问题或简单属性型问题同样适用,具体可参考上文中针对复杂关系属性型问题所提供的搜索原点选择规则,这里,不再对该规则进行详述。

[0181] 比如针对简单关系型问题“四川省有哪些医院?”,其对应的第一依存句法树及第二依存句法树具体可参考图3(a)及图3(b)所示,基于上文的搜索原点选择规则,可将“hospital_class”对应的实体类别节点作为搜索原点。

[0182] 步骤702、搜索所述第二依存句法树中除所述搜索原点外未被搜索的实体、实体类别或属性。

[0183] 在确定出搜索原点的基础上,可从第二依存句法树中继续搜索除所述搜索原点外未被搜索的实体、实体类别或属性,例如,针对图3(b),可继续搜索图中的实体节点“region”。

[0184] 步骤703、将搜索出的实体、实体类别或属性与搜索原点所对应的实体类别或实体进行匹配,得到二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及该二元组整合为三元组;所述关系库为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库,所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系。

[0185] 对于搜索出的实体、实体类别或属性,可将其与搜索原点所对应的实体类别或实体进行匹配,得到二元组,例如,对于图3(b)中的实体节点“region”可将其与搜索原点“hospital_class”匹配为二元组(region,hospital_class),通过进一步查询关系库,可确定出两者间的关系,本实施例中将该关系表示为“local”,从而可得到简单关系型问题“四川省有哪些医院?”的三元组 (region,local,hospital_class),为便于问答系统准确理解用户问句意图,具体实施中,可将该三元组中的实体/实体类别替换为其在问题语句中对应的组成元素,从而可得到最终的三元组(四川省,local,医院)。

[0186] 简单关系型问题中仅包括一条关系线,简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问,对于这两种类型的问题语句,仅能提取出一组三元组,从而,在提取出一组三元组的基础上,即可结束对其对应的第二依存句法树的搜索过程(事实上而言,在提取出一组三元组的基础上,其第二依存句法树也不存在未被搜索的节点)。

[0187] 在上一实施例及本实施例中分别针对复杂关系属性型问题及简单关系型问题/简单属性型问题,描述了通过对其第二依存句法树进行搜索处理,提取其三元组的过程,该过程的逻辑示意图具体可参考图8所示。

[0188] 本实施例通过搜索简单关系型问题/简单属性型问题的第二依存句法树,并基于关系库为搜索过程中所确定出的实体/实体类别/属性与实体/实体类别/ 属性的二元组匹配相应的谓词关系,得到简单关系型问题/简单属性型问题语句的三元组,实现了利用问题语句的各组成元素间的依存信息以及问题语句的句子成分,对简单关系型问题/简单属性型问题的问题语句进行结构化解析,可达到较好的解析效果,进而可基于解析结果准确地识别用户问句意图。

[0189] 在接下来的实施例四中,所述预定问句类别不包括所述简单关系型问题及所述简单属性型问题,也即,所述预定问句类别仅包括所述复杂关系属性型问题。

[0190] 基于预定问句类别的上述内容,若问题语句所属的问句类别为简单关系型问题或简单属性型问题,则问题语句所属的问句类别不符合所述预定问句类别。此种情况下,参考图8示出的数据处理方法流程图,所述数据处理方法还可以包括以下步骤:

[0191] 步骤901、若所述问题语句所属的问句类别为简单关系型问题,则将所述问题语句中已识别出的一个实体或实体类别与另一实体或实体类别组合为二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及该二元组整合为三元组,得到一个三元组。

[0192] 简单关系型问题中仅包括一条关系线,在对其进行实体、实体类别、属性及疑问词等句子成分识别时,一般仅可从中识别出两个实体/实体类别(而不会识别出更多数量的实体/实体类别),例如简单关系型问题语句“四川省有哪些医院?”,从中仅可识别出两个实体/实体类别:实体“四川省”及实体类别“医院”。

[0193] 从而对于所识别出的两个实体/实体类别,可直接将其匹配为一二元组,进而通过搜索所构建的关系库,来为该二元组匹配一谓词关系,例如对于上述的简单关系型问题语句“四川省有哪些医院?”,可将其实体“四川省”与实体类别“医院”直接匹配为二元组(四川省,医院),并基于关系库为其匹配一关系“local”,从而最终可得到该问题语句对应的三元组(四川省,local,医院)。

[0194] 步骤902、若所述问题语句所属的问句类别为简单属性型问题,则将所述问题语句

中已识别出的实体或实体类别与属性组合为二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及该二元组整合为三元组,得到一个三元组。

[0195] 简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问,在对其进行实体、实体类别、属性及疑问词等句子成分识别时,一般仅可从中识别出一个实体/实体类别以及一个由疑问词指代的属性,如上文所述,对于问题语句“白癜风是什么”可识别出实体“白癜风”,以及由疑问词“什么”指代的“introduction”属性,对于问题语句“医院在哪里?”可识别出实体类别“医院”以及由疑问词“哪里”指代的位置属性。

[0196] 对于所识别出的一个实体/实体类别以及一个由疑问词指代的属性,同样可直接将其匹配为一二元组,进而通过搜索所构建的关系库,来为该二元组匹配一谓词关系,例如对于上述的简单关系型问句“白癜风是什么”,可将其实体“白癜风”与疑问词指代的“introduction”属性直接匹配为二元组(白癜风,introduction),并基于关系库为其匹配一关系“attribute”,从而最终可得到该问题语句对应的三元组(白癜风,attribute,introduction);对于上述的简单关系型问句“医院在哪里?”可将其实体“医院”与疑问词指代的位置属性“location”直接匹配为二元组(医院,location),并基于关系库为其匹配一关系“attribute”,从而最终可得到该问题语句对应的三元组(医院,attribute,location)。

[0197] 本实施例根据简单关系型问题/简单属性问题的问句结构较为简单的特点,直接基于其句子成分并通过匹配关系库对其进行结构化解析,同样可具备较好的解析效果,且相比于上一实施例,同时还可以具备较高的解析效率。

[0198] 与上述的数据处理方法相对应,本申请还提供一种服务器,所述服务器可以是本地/网络端单独部署的服务器,或者本地/网络端的服务器集群中的服务器,参考图10,是本申请提供的一种服务器实施例五的结构示意图,所述服务器包括:

[0199] 存储器1001,用于至少存储一组指令集。

[0200] 所述存储器1001可以是具备数据存储功能的各类型存储器件,如ROM (Read Only Memory,只读存储器),FLASH,光盘,硬盘等,所存储的所述指令集用于指示处理器1002执行如下文所述的相应数据处理操作,该指令集可以以程序形式存储于所述存储器1001中。

[0201] 处理器1002,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行以下操作:

[0202] 获得问题语句;

[0203] 确定所述问题语句的句子成分;

[0204] 基于所述问题语句的句子成分确定所述问题语句所属的问句类别;

[0205] 若所述问句类别属于预定问句类别,则对所述问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息;并基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,得到所述问题语句的结构化信息。

[0206] 所述问题语句可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句,如用户在其智能手机、笔记本等终端设备中通过智能问答APP (Application,应用程序)的应用界面或网页提交至智能问答系统的问题语句等。

[0207] 相对应地,所述获得问题语句,则可以是本申请方法的执行主体(如服务器/服务器集群等)基于智能问答系统获得用户提交的问题语句,如获得用户提交的问题语句“白癜风是什么?”、“治疗感冒的药物有哪些?”、“什么药物可以治疗感冒引起的症状?”等等。其中,作为执行主体的服务器/服务器集群,如本地/网络端的服务器或服务器集群等,可基于局域网、互联网或移动网络等网络的通信功能获得用户在其终端设备所提交的问题语句。

[0208] 需要说明的是,实际应用中,所述用户可以是自然人用户或者基于人工智能的非自然人用户,这里,并不对此进行限定。

[0209] 本申请中,问题语句的句子成分,是指问题语句的各组成元素所属的实体、实体类别等成分,其中,实体一般是指能够独立存在的、作为属性的基础的具体事物,对于一个句子来说,可以将句子中包括的能够指代具体事物的词语看做属于实体,如“阿莫西林”、“华西医院”、“感冒”等词语都可以认为其属于实体;实体类别则是指具体事物所属的类别,对于一个句子来说,可以将句子中包括的能够指代具体事物所属的类别的词语看做属于实体类别,如“药物”、“医院”、“疾病”等这样的类别词均属于实体类别。

[0210] 相对应地,确定问题语句的句子成分,可以是指确定问题语句的各组成元素(如构成句子的各个分词)中的实体及实体类别。

[0211] 但不限于此,除了实体、实体类别,问题语句的句子成分还可以包括属性、疑问词等成分,其中,所述属性可以是实体属性或实体类别属性,具体地,句子中所包括的关于实体在某一方面的属性的描述词可以认为其对应的句子成分为实体属性,如句子中包括的关于“阿莫西林”的生产日期、功效介绍或使用方法等方面的属性的描述词等;相对应地,句子中所包括的关于实体类别在某一方面的属性的描述词则可以认为其对应的句子成分为实体类别属性,如句子中包括的关于“医院”的功能定位、性质、位置等方面的属性的描述词等。

[0212] 所述疑问词是指句子中所包括的用于引出提问需求的词语,如“什么”、“哪些”、“哪个”、“哪儿”、“几个”、“如何”等等。

[0213] 从而,确定问题语句的句子成分,还可以是指确定问题语句的各组成元素(如构成句子的各个分词)中的实体、实体类别、属性及疑问词等。

[0214] 具体实施中,在确定问题语句的句子成分时,可以并不以本实施例提供的上述几种具体的句子成分为限,而是可由技术人员根据实际需求来预置具体需确定的句子成分。

[0215] 问句类别可以包括但不限于简单关系型问题、简单属性型问题及复杂关系属性型问题。

[0216] 其中,所述简单关系型问题中仅包括一条关系线;所述简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问;所述复杂关系属性型问题中包括不少于两条的关系线,或者包括不少于一条的关系线及实体属性。

[0217] 一条关系线,是指在利用一个谓词将一个实体/实体类别与另一个实体/实体类别进行衔接,使得两个实体/实体类别间产生谓词关系后所得的两个实体/实体类别与谓词间的对应关系。

[0218] 例如问题语句“治疗感冒的药物有哪些?”中包括一个实体“感冒”,一个实体类别“药物”,且实体“感冒”与实体类别“药物”通过谓词“治疗”产生谓词关系,从而该问题语句包括一条关系线;问题语句“什么药可以治疗感冒”中同样包括一个实体“感冒”,一个实体

类别“药物”(以疑问词“什么(药)”指代的“药物”这一实体类别),且实体“感冒”与实体类别“药物”通过谓词“治疗”产生谓词关系,从而该问题语句同样包括一条关系线。在识别问题语句所属的问句类别时,若通过上述处理已确定出句子中共包括两个实体/实体类别,则根据句子的句法特征,一般会认为两个实体/实体类别之间需要通过一谓词关系进行衔接,从而,可确定出该问题语句所属的问句类别为简单关系型问题。

[0219] 若问题语句中不包括关系线,而只对实体或实体类别的属性进行提问,例如问题语句“白癜风是什么?”中不包括关系线,仅对“白癜风”这一实体的“introduction”属性进行提问,“医院在哪里?”中不包括关系线,仅对“医院”这一实体类别的位置属性进行提问,则此类问题语句所属的问句类别为简单属性型问题,具体实施中,对于服务器/服务器集群或终端设备等执行主体来说,可在问题语句中仅包括一个实体/实体类别,以及一个由疑问词指代的属性(如“白癜风是什么”中包括由疑问词“什么”所指代的“白癜风”的“introduction”属性,“医院在哪里?”中包括由疑问词“哪里”所指代的“医院”的位置属性)的情况下,将问题语句的问句类别识别为所述简单属性型问题。

[0220] 一个问题语句,若其不属于上述的简单关系型问题,也不属于上述的简单属性型问题,则将其所属的问句类别确定为复杂关系属性型问题。

[0221] 当然,具体实施中,也可以在识别问题语句中所包括的实体、实体类别、属性及疑问词等句子成分的基础上,同时识别其中的谓词关系,进而结合所识别出的实体、实体类别、属性、疑问词及谓词关系等来确定句子中是否包括关系线、在包括关系线情况下具体包括的关系线条数以及是否包括对属性的提问等信息,进而确定出问题语句所属的问句类别,本实施例对确定问题语句所属问句类别的实现过程不作限定。

[0222] 所述预定问句类别,可以根据实际实施情况对其进行设定,例如将其设定为包括上述的简单关系型问题、简单属性型问题及复杂关系属性型问题中的全部类别;或者将其设定为仅包括上述三种类别中的其中之一(例如,仅包括所述复杂关系属性型问题)等等,本实施例对此不作限定。

[0223] 可利用相应依存句法分析算法或依存句法分析器,分析问题语句的依存句法结构信息,以得到问题语句的各组成元素间的依存信息,问题语句的各组成元素间的依存信息能够反映问题语句中各组成元素间的实际依存(依赖)情况。

[0224] 例如问题语句“今天天气怎么样”,通过依存句法分析,可以将其解析成“今天-天气-怎么样”,其中,“今天”依赖于“天气”,“天气”依赖于“怎么样”,而“怎么样”则不依赖于任何词语。

[0225] 在对属于预定问句类别的问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息后,本申请基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,以得到所述问题语句的结构化信息,该对问题语句进行结构化解析,从而得到问题语句的结构化信息的处理过程会在之后的实施例中进行详述,具体请参见后续实施例对该部分的说明。

[0226] 根据以上方案可知,本实施例提供的服务器,在获得问题语句后,基于问题语句的句子成分确定问题语句所属的问句类别,并在问题语句属于预定问句类别的情况下,对问题语句进行依存句法分析,得到问题语句的各组成元素间的依存信息,进而基于问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对问题语句进行结构化解析,

得到问题语句的结构化信息。通过利用问题语句的各组成元素间的依存信息以及问题语句的句子成分,对问题语句进行结构化解析,可达到较好的解析效果,进而可基于解析结果准确地识别用户问句意图。

[0227] 在接下来的实施例六中,继续对上述服务器中处理器1002的数据处理功能进行进一步详述。本实施例中,所述处理器1002具体可通过执行以下处理实现其数据处理功能:

[0228] 获得问题语句;

[0229] 利用预定的实体识别方法,识别所述问题语句的各组成元素中的实体;

[0230] 利用预定的疑问词识别方法,识别所述问题语句的各组成元素中的疑问词;

[0231] 将所述问题语句中未被识别的各组成元素分别与预定的实体类别库、属性库进行匹配,得到所述问题语句的各组成元素中的实体类别及属性;

[0232] 基于所述问题语句的句子成分确定所述问题语句所属的问句类别;

[0233] 若所述问句类别属于预定问句类别,则基于所述问题语句中各组成元素间的实际依存情况,生成所述问题语句对应的第一依存句法树;其中,所述第一依存句法树中的每个节点包括所述问题语句的一组成元素,节点间的父子关系对应节点所包括的组成元素间的依存关系;

[0234] 根据所述第一依存句法树中各个节点对应的组成元素所属的句子成分,构建第二依存句法树。所述第二依存句法树中的相应节点至少包括相应组成元素所属的句子成分,节点间的父子关系对应节点所包括的句子成分间的依存关系;

[0235] 对所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组;每个三元组包括所述问题语句的主体-谓词-宾语信息。

[0236] 具体地,所述问题语句可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句,如用户在其智能手机、笔记本等终端设备中通过智能问答APP的应用界面或网页提交至智能问答系统的问题语句等。

[0237] 相对应地,所述获得问题语句,则可以是本申请方法的执行主体(如能够提供数据处理服务的服务器/服务器集群等)基于智能问答系统获得用户提交的问题语句,如获得用户提交的问题语句“白癜风是什么?”、“治疗感冒的药物有哪些?”、“什么药物可以治疗感冒引起的症状?”等等。其中,作为执行主体的服务器/服务器集群,如本地/网络端的服务器或服务器集群等,可基于局域网、互联网或移动网络等网络的通信功能获得用户在其终端设备所提交的问题语句。

[0238] 需要说明的是,实际应用中,所述用户可以是自然人用户或者基于人工智能的非自然人用户,这里,并不对此进行限定。

[0239] 可基于相应的实体识别技术,识别问题语句的各组成元素中的实体。

[0240] 具体地,对于实体识别,目前已存在很多可用算法,例如CRF、HMM等算法,从而,可选取这些算法中的其中一种对问题语句中的实体进行识别,或者亦可以通过模式匹配的方式进行实体识别,这里不局限于一种技术。

[0241] 如前文所述,所述疑问词是指句子中所包括的用于引出提问需求的词语,如“什么”、“哪些”、“哪个”、“哪儿”、“几个”、“如何”等等。

[0242] 由于疑问词数量有限,具体实施中可以通过模式匹配的方式对问题语句中的疑问词进行识别。

[0243] 所述实体类别库及所述属性库为分别预先对知识图谱进行相对应的实体类别信息抽取以及实体属性和/或实体类别属性信息抽取后所得的信息库;所述知识图谱为一预先基于多个语句样本所构建的包括多个节点的有向图,图中的每个节点对应一个实体或实体类别,每个实体对应零个或以上的实体属性、每个实体类别对应零个或以上的实体类别属性,任意两个节点之间的有向边用于表示边所连接的两个节点之间的关系。

[0244] 具体实施中,可预先采集所需业务领域的一系列语句样本(如问题语句或问答句等)作为语料集,并基于所采集的语料集构建上文所述的知识图谱,其中,所述知识图谱可以是基于某一特定业务领域的语料集所构建的对应于该特定业务领域的知识图谱,也可以是基于多个业务领域的语料集所构建的对应于多个业务领域的知识图谱,本实施例对此不作限定。

[0245] 在构建出所述知识图谱的基础上,可进一步通过对所述知识图谱进行实体类别信息抽取来得到相对应的实体类别库,以及对所述知识图谱进行实体属性和/或实体类别属性信息抽取来得到相对应的属性库。

[0246] 其中,所述实体类别库及属性库,可以是分别通过对对应于某一特定业务领域的知识图谱进行实体类别信息抽取以及属性信息抽取,所获得的对应于该特定业务领域的信息库;或者还可以是分别通过对对应于多个业务领域的知识图谱进行实体类别信息抽取及属性信息抽取,所获得的对应于多个业务领域的信息库;或者,还可以是分别通过对对应于多个业务领域的知识图谱进行某一特定业务领域的实体类别信息抽取及属性信息抽取,所获得的对应于该特定业务领域的信息库,本实施例对此同样不作限定。

[0247] 在构建出知识图谱、并基于对知识图谱进行相应信息抽取构建出实体类别库及属性库的基础上,当对问题语句进行句子成分识别时,具体可通过将问题语句中的各组成元素与所述实体类别库或属性库进行匹配,来识别出问题语句的各组成元素中的实体类别或属性(实体属性和/或实体类别属性)。

[0248] 例如对于问题语句“四川省有哪些医院?”,通过将其各组成元素与实体类别库进行匹配,可识别出其中的组成元素“医院”所属的句子成分为实体类别,对于问句“游乐场在哪个位置?”,通过将其各组成元素与属性库进行匹配,可识别出其中的组成元素“(哪个)位置”所属的句子成分为实体类别属性,且其具体为实体类别的位置属性。

[0249] 如上文所述,问句类别可以包括但不限于简单关系型问题、简单属性型问题及复杂关系属性型问题。其中,所述简单关系型问题中仅包括一条关系线;所述简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问;所述复杂关系属性型问题中包括不少于两条的关系线,或者包括不少于一条的关系线及实体属性。

[0250] 在已确定出问题语句所包括的实体、实体类别、属性及疑问词等句子成分的基础上,可进一步基于所识别出的句子成分来确定问题语句所属的问句类别具体是上述问句类别中的哪一种。

[0251] 其中,若确定出句子中共包括两个实体/实体类别,则根据句子的句法特征,一般会认为两个实体/实体类别之间需要通过一谓词关系进行衔接,从而,可确定出该问题语句中包括一条关系线,相对应地可确定出该问题语句所属的问句类别为简单关系型问题。如问题语句“治疗感冒的药物有哪些?”中包括一个实体“感冒”以及一个实体类别“药物”,则可认为该实体“感冒”及实体类别“药物”之间需要通过一谓词关系进行衔接,从而该问题语

句中包括一条关系线,相对应地,其所属的问句类别为简单关系型问题,这与该问题中“感冒”及“药物”之间通过谓词“治疗”来进行衔接进而构成一条关系线的实际情况也是相符的。

[0252] 若问题语句中仅包括一个实体/实体类别,以及一个疑问词指代的属性,则认为问题语句中不包括关系线,仅包括对实体/实体类别的属性的提问,从而可将该问题语句的问句类别识别为所述简单属性型问题。例如问题语句“白癜风是什么?”中仅包括一个实体“白癜风”以及一个疑问词“什么”指代的“白癜风”的“introduction”属性,本质上属于对“白癜风”这一实体的“introduction”属性进行提问,不包括任何关系线,问题语句“医院在哪里?”仅包括一个实体类别“医院”以及一个疑问词“哪里”指代的“医院”的位置属性,本质上属于对“医院”这一实体类别的位置属性进行提问,不包括任何关系线,则此类问题语句所属的问句类别为简单属性型问题。

[0253] 一个问题语句,若其不属于上述的简单关系型问题,也不属于上述的简单属性型问题,则将其所属的问句类别确定为复杂关系属性型问题。

[0254] 当然,具体实施中,也可以在识别问题语句中所包括的实体、实体类别、属性及疑问词等句子成分的基础上,同时识别其中的谓词关系,进而结合所识别出的实体、实体类别、属性、疑问词及谓词关系等来确定句子中是否包括关系线、在包括关系线情况下具体包括的关系线条数以及是否包括对属性的提问等信息,进而确定出问题语句所属的问句类别,本实施例对确定问题语句所属问句类别的实现过程不作限定。

[0255] 实际应用中,参考表1,所述简单关系型问题可以分为查询型及判断型两种,而所述查询型的简单关系型问题又可以分为列表型及聚合型两种,每种类型的相关示例及对应的句子结构具体可参见以下的表1所示:

[0256] 表1

[0257]	分类	关系	问句示例	
	查询	具体实体+关系->实体类别	列表型	治疗感冒的药物有哪些?
				感冒+治疗->药物
			聚合型	治疗感冒的药物有多少种?
				感冒+治疗->药物种类
	判断	具体实体-关系-具体实体	阿莫西林可以治疗感冒吗?	
			阿莫西林-治疗-感冒	

[0258] 复杂关系型问题又可以分为嵌入型、并列型及属性型,每种类型的相关示例及对应的句子结构具体可参见以下的表2所示:

[0259] 表2

[0260]	分类	问句示例
	嵌入型	什么药物可以治疗感冒引起的症状?
		(具体实体+关系->实体)+关系->实体
	并列型	什么药物可以治疗发烧和头痛?
		(具体实体+关系->实体)^(具体实体+关系->实体)
	属性型	三甲医院能够治疗哪些疾病?
		(实体类别+属性->具体实体)+关系->实体

[0261] 需要说明的是,具体实施中,可以按照上文所述的在针对问题语句识别出实体、实体类别、属性及疑问词等句子成分的基础上(不识别其中的谓词关系),通过推导其关系线情况来获知其所属的问句类别,或者也可以既识别问题语句的实体、实体类别、属性及疑问词,又识别其中的谓词关系,在此基础上,对其进行句子结构解析,并基于解析出的句子结构识别其对应所属的问句类别,并不局限于其中的一种。

[0262] 本申请中,所述预定问句类别至少包括所述复杂关系属性型问题,具体地,例如包括上述的简单关系型问题、简单属性型问题及复杂关系属性型问题中的全部类别,或者仅包括所述复杂关系属性型问题等,实际应用中,可根据具体实施情况进行设定。

[0263] 在问题语句的问句类别属于预定问句类别的情况下,处理器1002可利用相应依存句法分析算法或依存句法分析器,分析问题语句的依存句法结构信息,以得到问题语句的各组成元素间的依存信息,其中,问题语句的各组成元素间的依存信息能够反映问题语句中各组成元素间的实际依存(依赖)情况,进而基于问题语句的各组成元素间的依存信息,生成所述问题语句对应的第一依存句法树。

[0264] 例如问题语句“今天天气怎么样?”,通过依存句法分析,可以将其解析成“今天-天气-怎么样”,其在句子中的ID(Identity,身份标识号)可按序表示成[1,2,3],其中,“1”表示“今天”,“2”表示“天气”,“3”表示“怎么样”;各成分之间的依存关系可以表示成[2,3,0],即“1”依赖于“2”,“2”依赖于“3”,“3”依赖于“0”,依赖于“0”表示不依赖任何词,也即“天气”支配了“今天”,“怎么样”支配了“天气”。

[0265] 在通过对问题语句进行依存句法信息分析,得到问题语句的各组成元素间的依存信息后,可进一步基于问题语句的各组成元素间的依存信息,生成所述问题语句对应的第一依存句法树,其中,所述第一依存句法树中的每个节点包括所述问题语句的一组成元素,节点间的父子关系对应节点所包括的组成元素间的依存关系。例如,根据问题语句“今天天气怎么样?”的上述依存关系,可知“2”依赖于“3”,则在所构建的该问题语句的第一依存句法树中,“2”为“3”的子节点,即“天气”为“怎么样”的子节点,同理,“1”依赖于“2”,则“1”为子节点,“2”为父节点,也即,“今天”为“天气”的子节点。

[0266] 处理器1002具体可根据问题语句所对应的第一依存句法树,并结合问题语句中各组成元素所属的句子成分(如实体、实体类别、属性或疑问词等),来构建问题语句的第二依存句法树。其中,在基于问题语句的第一依存句法树来构建其第二依存句法树时,对于第一依存句法树中的每个节点,可在第二依存句法树中为其构建一相对应的节点,且第二依存句法树中所构建节点的节点值具体为其在第一依存句法中相对应节点的组成元素所属的句

子成分（当然其取值也可以同时包括其在第一依存句法中相对应节点的组成元素及该组成元素所对应的具体成分，本申请对此不作限制），实际应用中，第一依存句法树中可能存在某些未被识别出所属句子成分的组成元素的节点，对于此类节点，其在第二依存句法树中所对应的节点的节点值则可直接取该组成元素的元素信息，也即维持其在第一依存句法树中所对应节点的节点值。

[0267] 参考图3(a)，是本实施例提供的问题语句“四川省有哪些医院？”所对应的第一依存句法树的树型结构示意图，在构建其第二依存句法树时，对于“四川省”这一组成元素，可按其在第一依存句法树中的节点位置在第二依存句法树中为其创建一相应节点，且由于“四川省”为地点实体，从而可将创建的该节点的节点值表示为“region”（即地点实体）；“医院”为实体类别，则在按“医院”在第一依存句法树中的节点位置在第二依存句法树中为其创建一相应节点后，可将所创建的该节点的节点值表示为“hospital_class”；“哪些”为疑问词，则在按“哪些”在第一依存句法树中的节点位置在第二依存句法树中为其创建一相应节点后，可将所创建的该节点的节点值表示为“what_some”；而在对问题语句中的各组成元素进行实体、实体类别、属性及疑问词的识别时，未识别“有”这一组成元素所属的句子成分，则在构建第二依存句法树时，该组成元素“有”在第二依存句法树中所对应的节点位置及节点值仍维持不变，即与其在第一依存句法树中所对应的节点位置及节点值一致，对于问题语句“四川省有哪些医院？”，其第二依存句法树的树型结构及树中各节点的取值情况具体可参考图3(b)所示。

[0268] 再比如问题语句“阿莫西林可以治疗发烧和头痛吗？”，其对应的第一依存句法树如图4(a)所示，在构建其所对应的第二依存句法树时，在按“阿莫西林”在其第一依存句法树中所对应的节点位置为其创建一相对应的节点后，由于“阿莫西林”属于药品实体，从而，可将该节点的节点值表示为“medicine”，相对应地，“发烧”和“头痛”为症状实体，可分别将两者在第二依存句法树中所对应的节点的节点值表示为“symptom1”和“symptom2”，“治疗”是药物和症状之间的关系，表示为“indication”，而其他组成元素如“可以”、“吗”由于在句子成分识别时未识别其对应的句子成分，从而，仍维持第一依存句法树中的节点值不变（即仍采用“可以”、“吗”），所创建的第二依存句法树具体如图4(b)所示。

[0269] 在基于问题语句的各组成元素所属的句子成分及问题语句所对应的第一依存句法树，构建出问题语句所对应的第二依存句法树的基础上，可进一步基于问题语句所对应的第二依存句法树，对问题语句进行结构化解析，以得到问题语句的结构化信息。

[0270] 本实施例具体通过对问题语句的第二依存句法树进行预定的搜索处理，来得到问题语句的至少一个SP0 (Subject-Predicate-Object, 主体-谓词-宾语) 三元组，以此实现对问题语句的结构化解析，并相应地以所得到的至少一个 SP0三元组作为问题语句的结构化信息。

[0271] 如上文所述，本实施例中，所述预定问句类别至少包括所述复杂关系属性型问题，从而，当问题语句所属的问句类别为所述复杂关系属性型问题时，则该问题语句所属的问句类别符合所述预定问句类别，从而，可通过对该问题语句的第二依存句法树进行搜索处理，来得到该问题语句的至少一个SP0 三元组。

[0272] 其中，参考图5，当问题语句所属的问句类别为所述复杂关系属性型问题时，处理器1002具体可通过以下的处理过程实现对复杂关系属性型的问题语句进行三元组提取：

[0273] 步骤501、基于预定的搜索原点确定方式,从所述第二依存句法树中确定出作为搜索原点的目标节点,所述目标节点对应包括的句子成分为实体类别或实体。

[0274] 具体地,可根据以下的搜索原点选择规则从问题语句的第二依存句法树中确定出用于作为搜索原点的目标节点:

[0275] a) 若问题语句中只包含一个实体类别,则将该实体类别作为搜索原点;

[0276] b) 若问题语句中包含两个及以上的实体类别,则选择与疑问词最近的一个实体类别作为搜索原点;若没有疑问词,则选择离句末最近的实体类别词作为搜索原点;

[0277] c) 若句子中不包含实体类别,则查看实体个数:其中,若只包含一个具体实体,则将其作为搜索原点;若包含两个及以上个具体实体,则优先选择离根节点最近的左子树(若并非二叉树,则以相对的左右为准)上的节点作为搜索原点;

[0278] d) 若问题语句中既无实体类别也无具体实体,则可能是因为该问题语句本身和该领域的知识图谱无关,也可能是实体识别时未将实体正确识别出来,对于这样的问句,暂时不处理(也即按结构化解析失败处理)。

[0279] 比如,对于问题语句“阿莫西林可以治疗发烧和头痛吗?”,其对应的第一依存句法树及第二依存句法树分别如图4(a)及图4(b)所示,对于该问题语句,根据上述的搜索原点选择规则,具体可选择图4(b)中“阿莫西林”对应的节点“medicine”作为搜索原点(匹配上述规则中的第c)条规则),也即,以该节点为搜索起始点对第二依存句法树进行后续的搜索处理。再比如,对于问题语句“卖阿莫西林的药厂卖药给哪些医院?”的第一依存句法树及第二依存句法树分别如图6(a)及图6(b)所示,对于该问题语句,根据上述的搜索原点选择规则,可确定出距离疑问词节点“what_some”最近的实体类别节点“hospital_class”为搜索原点。

[0280] 步骤502、搜索所有离所述搜索原点距离最近的实体、实体类别及属性。

[0281] 在确定出搜索原点的基础上,可继续搜索问题语句的第二依存句法树中所有离所述搜索原点距离最近的实体、实体类别及属性,本申请中,所述离所述搜索原点距离最近的实体、实体类别或属性为:至所述搜索原点的节点路径上不包括任何其他实体、实体类别及属性的实体、实体类别或属性。

[0282] 仍以问题语句“阿莫西林可以治疗发烧和头痛吗?”为例,根据图4(b)示出的该问题语句的第二依存句法树,可知,与所述搜索原点“medicine”距离最近的实体、实体类别或属性具体为“发烧”所对应的实体节点“symptom1”,从而可将该节点作为当前待搜索的节点。

[0283] 再以问题语句“卖阿莫西林的药厂卖药给哪些医院?”为例,根据图6(b)示出的该问题语句的第二依存句法树,可知,与所述搜索原点“hospital_class”距离最近的实体、实体类别或属性具体为“药”所对应的实体类别节点“medicine_class”,从而可将该节点作为当前待搜索的节点。

[0284] 步骤503、将搜索出的每个实体、实体类别或属性与所述搜索原点所对应的实体类别或实体进行匹配,得到相对应的至少一个二元组,并基于预先制定的关系库为每个二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及相应二元组整合为三元组,得到至少一个三元组。

[0285] 本实施例中,所述关系库即为预先对所述知识图谱进行相对应的关系信息抽取后

所得的信息库,所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系。

[0286] 具体地,所述关系库,可以是通过对对应于某一特定业务领域的知识图谱进行关系信息抽取,所获得的对应于该特定业务领域的信息库;或者还可以是通过对对应于多个业务领域的知识图谱进行关系信息抽取,所获得的对应于多个业务领域的关系库;或者,还可以是通过对对应于多个业务领域的知识图谱进行某一特定业务领域的关系信息抽取,所获得的对应于该特定业务领域的关系库,本实施例对此同样不作限定。

[0287] 针对问题语句,在其第二依存句法树中确定出作为搜索原点的目标节点,并搜索所有离所述搜索原点距离最近的实体、实体类别或属性后,可将搜索出的离所述搜索原点距离最近的每个实体、实体类别或属性与所述搜索原点进行匹配,得到相对应的至少一个二元组,并可通过搜索上述的关系库,为每个二元组匹配出相对应的谓词关系,进而将匹配出的谓词关系以及相应二元组整合为三元组,以此得到问题语句的至少一个三元组。

[0288] 比如,针对问题语句“阿莫西林可以治疗发烧和头痛吗?”,如上文所述,通过节点搜索,可确定出与搜索原点“medicine”距离最近的实体、实体类别或属性具体为“发烧”所对应的实体节点“symptom1”,从而可将该节点“symptom1”与搜索原点“medicine”匹配为一二元组(medicine,symptom1),通过对基于知识图谱所提取的上述关系库进行搜索可以确定出两者之间的关系为“indication”(治疗),从而可得到三元组(medicine,indication,symptom1),实际实施中,可继续将该三元组中的实体替换为其在问题语句中所对应的具体组成元素,从而得到三元组(阿莫西林,indication,发烧),以便于智能问题系统基于所提取的三元组对用户意图进行准确识别。

[0289] 相应地,针对问题语句“卖阿莫西林的药厂卖药给哪些医院?”,通过节点搜索,可确定出与搜索原点“hospital_class”距离最近的实体、实体类别或属性具体为“药”所对应的实体类别节点“medicine_class”以及“factory_class”,从而可将节点“medicine_class”与搜索原点“hospital_class”匹配为一二元组(hospital_class,medicine_class),将节点“factory_class”与搜索原点“hospital_class”匹配为一二元组(factory_class,hospital_class),之后通过对基于知识图谱所提取的上述关系库进行搜索可以确定出前一二元组之间的关系为“buy”(买),从而可得到三元组(hospital_class,buy,medicine_class),后一二元组所对应的关系为“sale to”,从而可得到三元组(factory_class,sale to,hospital_class)。实际应用中,可以将上述三元组中的实体类别替换为其在问题语句中所对应的具体组成元素,从而得到三元组(医院,buy,药)、(药厂,sale to,医院)并输出给智能问答系统。

[0290] 步骤504、若所述第二依存句法树中存在未被搜索的实体、实体类别或属性,则确定未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性间的结构关系,所述结构关系为并列关系或嵌套关系;未被搜索的实体、实体类别或属性的上一级实体、实体类别或属性为:已被搜索的实体、实体类别或属性中与所述未被搜索的实体、实体类别或属性间的节点路径最短的实体、实体类别或属性。

[0291] 具体地,在上述处理的基础上,若第二依存句法树中仅存在一个未被搜索的实体、实体类别或属性,则可直接对该未被搜索的实体、实体类别或属性进行搜索操作,并确定该一个未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性间的结构关系;若

第二依存句法树中存在一个以上未被搜索的实体、实体类别或属性,则优先选择未被搜索的各实体、实体类别或属性节点中与已被搜索的实体、实体类别或属性节点的节点距离最近的实体、实体类别或属性节点作为当前待搜索的节点,并确定该前待搜索的节点与其上一级实体、实体类别或属性间的结构关系。这里,某一未被搜索的实体、实体类别或属性节点与已被搜索的实体、实体类别或属性节点的节点距离是指未被搜索的实体、实体类别或属性节点与已被搜索的各实体、实体类别或属性节点的节点距离中的最短距离。

[0292] 比如,针对图4(b)示出的问题语句“阿莫西林可以治疗发烧和头痛吗?”的第二依存句法树,在确定出搜索原点“medicine”,并基于该搜索原点完成对节点“symptom1”的搜索处理后,仍存在一个实体节点“symptom2”未被处理,从而,可确定该实体节点“symptom2”与其上一级节点“symptom1”的结构关系,由于“symptom2”与“symptom1”为并列的两个症状实体,从而两者之间的结构关系为并列关系。

[0293] 步骤505、若所述结构关系为并列关系,则将所述未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性对应的三元组中除所述上一级实体、实体类别或属性之外的其他二元组整合为三元组。

[0294] 具体地,例如,问题语句“阿莫西林可以治疗发烧和头痛吗?”中“symptom2”与其上一级节点“symptom1”为并列关系,从而,可将该节点“symptom2”与“symptom1”对应的三元组(medicine,indication,symptom1)中的二元组(medicine,indication)进行整合,得到该节点“symptom2”对应的三元组(medicine,indication,symptom2),至此,问题语句“阿莫西林可以治疗发烧和头痛吗?”的第二依存句法树中已不存在未被搜索的实体、实体类别或属性节点,可结束对该问题语句的第二依存句法树的搜索过程,对于该问题语句,最终可提取出两个三元组:

[0295] (medicine,indication,symptom1);

[0296] (medicine,indication,symptom2)。

[0297] 具体实施中,为便于智能问答系统对用户问句意图进行理解,可将上述三元组中的实体替换为其在问题语句中所对应的具体组成元素,从而得到三元组(阿莫西林,indication,发烧)及(阿莫西林,indication,头痛)。

[0298] 步骤506、若所述结构关系为嵌套关系,则将所述未被搜索的实体、实体类别或属性与其上一级实体、实体类别或属性组合为二元组,并基于所述关系库为该二元组匹配出相对应的谓词关系,将匹配出的该谓词关系以及该二元组整合为三元组,直至所述第二依存句法树中不存在未被搜索的实体、实体类别及属性时结束。

[0299] 具体地,例如,问题语句“卖阿莫西林的药厂卖药给哪些医院”中“factory_class”与其上一级节点“medicine_class”为嵌套关系,从而,可将该节点“factory_class”与其上一级节点“medicine_class”组合为二元组(factory_class,medicine_class),通过搜索关系库,可知“factory_class”与“medicine_class”之间的关系为“sale”,从而可得到三元组(factory_class,sale,medicine_class)。

[0300] 在完成对实体类别节点“factory_class”的搜索处理后,根据问题语句“卖阿莫西林的药厂卖药给哪些医院”的第二依存句法树,可知,其中,还存在实体节点“medicine”未被搜索,该节点“medicine”与其上一级节点“factory_class”的关系为嵌套关系,从而可将两者匹配为二元组(factory_class,medicine),通过搜索关系库,可进一步确定出三元组

(factory_class,sale,medicine)。

[0301] 从而,对于问题语句“卖阿莫西林的药厂卖药给哪些医院?”,最终可提取出四个三元组:

[0302] (hospital_class,buy,medicine_class);

[0303] (factory_class,sale to,hospital_class);

[0304] (factory_class,sale,medicine_class);

[0305] (factory_class,sale,medicine)。

[0306] 具体实施中,为便于智能问答系统对用户问句意图进行理解,可将上述三元组中的实体/实体类别替换为其在问题语句中所对应的具体组成元素,从而可得到三元组:(医院,buy,药),(药厂,sale to,医院)(药厂,sale,药),(药厂,sale,阿莫西林)。

[0307] 需要说明的是,实际应用中,包括两个实体类别的三元组的实用性较低,如上述的(factory_class,sale to,hospital_class)及(factory_class,sale, medicine_class),从而可考虑将其过滤掉。

[0308] 还需要说明的是,本实施例中,所述附图4(b)及6(b)中节点与节点间的边(非树的自身的边,在附图中未带箭头)上的关系,如图4(b)中边上的两个“indication”,图6(b)中边上的两个“sale”、一个“sale to”及一个“buy”,均是通过关系库为边的两端节点的词所匹配出来的。

[0309] 本实施例通过搜索复杂关系型问题的第二依存句法树,并基于关系库为搜索过程中所确定出的实体/实体类别/属性与实体/实体类别/属性的二元组匹配相对应的谓词关系,得到复杂关系型问题语句的三元组,实现了利用问题语句的各组成元素间的依存信息以及问题语句的句子成分,对复杂关系型问题的问题语句进行结构化解析,可达到较好的解析效果,进而可基于解析结果准确地识别用户问句意图。由于在进行问句结构解析时,不需依赖于句子的语法规则,且不要求句子必须遵循某一特定的表达形式(如不要求句子必须是非口语化的句子),从而对于不符合标准语法规则和/或采用口语化表达形式的复杂句子,仍能达到较好的解析效果(对于口语化句子可在构建知识图谱时,同时搜集口语化表达形式的语料,使其能够适应该类问题语句)。

[0310] 在接下来的实施七中,所述预定问句类别除了包括所述复杂关系属性型问题,还包括简单关系型问题及所述简单属性型问题。

[0311] 基于预定问句类别的上述内容,若问题语句所属的问句类别为简单关系型问题或简单属性型问题,则问题语句所属的问句类别符合所述预定问句类别。此种情况下,即问题语句所属的问句类别为简单关系型问题或简单属性型问题的情况下,所述服务器中的处理器1002所述第二依存句法树进行搜索处理,得到所述问题语句的至少一个三元组,可以通过以下的处理过程实现:

[0312] 基于预定的搜索原点确定方式,从所述第二依存句法树中确定出作为搜索原点的目标节点,所述目标节点对应包括的句子成分为实体类别或实体;

[0313] 搜索所述第二依存句法树中除所述搜索原点外未被搜索的实体、实体类别或属性;

[0314] 将搜索出的实体、实体类别或属性与搜索原点所对应的实体类别或实体进行匹配,得到二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出

的谓词关系以及该二元组整合为三元组;所述关系库为预先对所述知识图谱进行相对应的关系信息抽取后所得的信息库,所述关系信息包括实体或实体类别与实体或实体类别间的谓词关系的对应关系。

[0315] 上文中提供的针对复杂关系属性型问题的搜索原点选择规则对于简单关系型问题或简单属性型问题同样适用,具体可参考上文中针对复杂关系属性型问题所提供的搜索原点选择规则,这里,不再对该规则进行详述。

[0316] 比如针对简单关系型问题“四川省有哪些医院?”,其对应的第一依存句法树及第二依存句法树具体可参考图3(a)及图3(b)所示,基于上文的搜索原点选择规则,可将“hospital_class”对应的实体类别节点作为搜索原点。

[0317] 在确定出搜索原点的基础上,可从第二依存句法树中继续搜索除所述搜索原点外未被搜索的实体、实体类别或属性,例如,针对图3(b),可继续搜索图中的实体节点“region”。

[0318] 对于搜索出的实体、实体类别或属性,可将其与搜索原点所对应的实体类别或实体进行匹配,得到二元组,例如,对于图3(b)中的实体节点“region”可将其与搜索原点“hospital_class”匹配为二元组(region,hospital_class),通过进一步查询关系库,可确定出两者间的关系,本实施例中将该关系表示为“local”,从而可得到简单关系型问题“四川省有哪些医院?”的三元组 (region,local,hospital_class),为便于问答系统准确理解用户问句意图,具体实施中,可将该三元组中的实体/实体类别替换为其在问题语句中对应的组成元素,从而可得到最终的三元组(四川省,local,医院)。

[0319] 简单关系型问题中仅包括一条关系线,简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问,对于这两种类型的问题语句,仅能提取出一组三元组,从而,在提取出一组三元组的基础上,即可结束对其对应的第二依存句法树的搜索过程(事实上而言,在提取出一组三元组的基础上,其第二依存句法树也不存在未被搜索的节点)。

[0320] 在上一实施例及本实施例中分别针对复杂关系属性型问题及简单关系型问题/简单属性型问题,描述了通过对其第二依存句法树进行搜索处理,提取其三元组的过程,该过程的逻辑示意图具体可参考图8所示。

[0321] 本实施例通过搜索简单关系型问题/简单属性型问题的第二依存句法树,并基于关系库为搜索过程中所确定出的实体/实体类别/属性与实体/实体类别/属性的二元组匹配相应的谓词关系,得到简单关系型问题/简单属性型问题语句的三元组,实现了利用问题语句的各组成元素间的依存信息以及问题语句的句子成分,对简单关系型问题/简单属性型问题的问题语句进行结构化解析,可达到较好的解析效果,进而可基于解析结果准确地识别用户问句意图。

[0322] 在接下来的实施八中,所述预定问句类别不包括所述简单关系型问题及所述简单属性型问题,也即,所述预定问句类别仅包括所述复杂关系属性型问题。

[0323] 基于预定问句类别的上述内容,若问题语句所属的问句类别为简单关系型问题或简单属性型问题,则问题语句所属的问句类别不符合所述预定问句类别。此种情况下,服务器中的处理器1002还可以通过以下处理实现对简单关系型问题或简单属性型问题的进行三元组提取:

[0324] 若所述问题语句所属的问句类别为简单关系型问题,则将所述问题语句中已识别

出的一个实体或实体类别与另一实体或实体类别组合为二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及该二元组整合为三元组,得到一个三元组;

[0325] 若所述问题语句所属的问句类别为简单属性型问题,则将所述问题语句中已识别出的实体或实体类别与属性组合为二元组,并基于预先制定的关系库为该二元组匹配出相对应的谓词关系,将匹配出的谓词关系以及该二元组整合为三元组,得到一个三元组。

[0326] 其中,简单关系型问题中仅包括一条关系线,在对其进行实体、实体类别、属性及疑问词等句子成分识别时,一般仅可从中识别出两个实体/实体类别(而不会识别出更多数量的实体/实体类别),例如简单关系型问题语句“四川省有哪些医院?”,从中仅可识别出两个实体/实体类别:实体“四川省”及实体类别“医院”。

[0327] 从而对于所识别出的两个实体/实体类别,可直接将其匹配为一二元组,进而通过搜索所构建的关系库,来为该二元组匹配一谓词关系,例如对于上述的简单关系型问题语句“四川省有哪些医院?”,可将其实体“四川省”与实体类别“医院”直接匹配为二元组(四川省,医院),并基于关系库为其匹配一关系“local”,从而最终可得到该问题语句对应的三元组(四川省,local,医院)。

[0328] 简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问,在对其进行实体、实体类别、属性及疑问词等句子成分识别时,一般仅可从中识别出一个实体/实体类别以及一个由疑问词指代的属性,如上文所述,对于问题语句“白癜风是什么”可识别出实体“白癜风”,以及由疑问词“什么”指代的“introduction”属性,对于问题语句“医院在哪里?”可识别出实体类别“医院”以及由疑问词“哪里”指代的位置属性。

[0329] 对于所识别出的一个实体/实体类别以及一个由疑问词指代的属性,同样可直接将其匹配为一二元组,进而通过搜索所构建的关系库,来为该二元组匹配一谓词关系,例如对于上述的简单关系型问句“白癜风是什么”,可将其实体“白癜风”与疑问词指代的“introduction”属性直接匹配为二元组(白癜风,introduction),并基于关系库为其匹配一关系“attribute”,从而最终可得到该问题语句对应的三元组(白癜风,attribute,introduction);对于上述的简单关系型问句“医院在哪里?”可将其实体“医院”与疑问词指代的位置属性“location”直接匹配为二元组(医院,location),并基于关系库为其匹配一关系“attribute”,从而最终可得到该问题语句对应的三元组(医院,attribute,location)。

[0330] 本实施例根据简单关系型问题/简单属性问题的问句结构较为简单的特点,直接基于其句子成分并通过匹配关系库对其进行结构化解析,同样可具备较好的解析效果,且相比于上一实施例,同时还可以具备较高的解析效率。

[0331] 与上述的数据处理方法相对应,本申请还提供一种电子设备,所述电子设备可以是但不限于智能手机、平板电脑、台式机、笔记本、一体机等各类型终端设备,参考图11,是本申请提供的一种电子设备实施例九的结构示意图,所述电子设备包括:

[0332] 存储器1101,用于至少存储一组指令集。

[0333] 所述存储器1101可以是具备数据存储功能的各类型存储器件,如ROM (Read Only Memory,只读存储器),FLASH,光盘,硬盘等,所存储的所述指令集用于指示处理器1102执行如下文所述的相应数据处理操作,该指令集可以以程序形式存储于所述存储器1101中。

[0334] 处理器1102,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行以下操作:

[0335] 获得问题语句;

[0336] 确定所述问题语句的句子成分;

[0337] 基于所述问题语句的句子成分确定所述问题语句所属的问句类别;

[0338] 若所述问句类别属于预定问句类别,则对所述问题语句进行依存句法分析,得到所述问题语句的各组成元素间的依存信息;并基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分,对所述问题语句进行结构化解析,得到所述问题语句的结构化信息。

[0339] 所述问题语句可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句,如用户在其智能手机、笔记本等终端设备中通过智能问答APP(Application,应用程序)的应用界面或网页提交至智能问答系统的问题语句等。

[0340] 相对应地,所述获得问题语句,则可以是本申请方法的执行主体(如如用户的终端设备)基于智能问答系统获得用户提交的问题语句,如获得用户提交的问题语句“白癜风是什么?”、“治疗感冒的药物有哪些?”、“什么药物可以治疗感冒引起的症状?”等等。其中,作为执行主体的终端设备,可直接基于其相应问题输入/提交界面所提供的问题输入/提交功能获得用户所输入或提交的问题语句。

[0341] 需要说明的是,实际应用中,所述用户可以是自然人用户或者基于人工智能的非自然人用户,这里,并不对此进行限定。

[0342] 本申请中,问题语句的句子成分,是指问题语句的各组成元素所属的实体、实体类别等成分,其中,实体一般是指能够独立存在的、作为属性的基础的具体事物,对于一个句子来说,可以将句子中包括的能够指代具体事物的词语看做属于实体,如“阿莫西林”、“华西医院”、“感冒”等词语都可以认为其属于实体;实体类别则是指具体事物所属的类别,对于一个句子来说,可以将句子中包括的能够指代具体事物所属的类别的词语看做属于实体类别,如“药物”、“医院”、“疾病”等这样的类别词均属于实体类别。

[0343] 相对应地,确定问题语句的句子成分,可以是指确定问题语句的各组成元素(如构成句子的各个分词)中的实体及实体类别。

[0344] 但不限于此,除了实体、实体类别,问题语句的句子成分还可以包括属性、疑问词等成分,其中,所述属性可以是实体属性或实体类别属性,具体地,句子中所包括的关于实体在某一方面的属性的描述词可以认为其对应的句子成分为实体属性,如句子中包括的关于“阿莫西林”的生产日期、功效介绍或使用方法等方面的属性的描述词等;相对应地,句子中所包括的关于实体类别在某一方面的属性的描述词则可以认为其对应的句子成分为实体类别属性,如句子中包括的关于“医院”的功能定位、性质、位置等方面的属性的描述词等。

[0345] 所述疑问词是指句子中所包括的用于引出提问需求的词语,如“什么”、“哪些”、“哪个”、“哪儿”、“几个”、“如何”等等。

[0346] 从而,确定问题语句的句子成分,还可以是指确定问题语句的各组成元素(如构成句子的各个分词)中的实体、实体类别、属性及疑问词等。

[0347] 具体实施中,在确定问题语句的句子成分时,可以并不以本实施例提供的上述几

种具体的句子成分为限,而是可由技术人员根据实际需求来预置具体需确定的句子成分。

[0348] 问句类别可以包括但不限于简单关系型问题、简单属性型问题及复杂关系属性型问题。

[0349] 其中,所述简单关系型问题中仅包括一条关系线;所述简单属性型问题中不包括关系线,只对实体或实体类别的属性进行提问;所述复杂关系属性型问题中包括不少于两条的关系线,或者包括不少于一条的关系线及实体属性。

[0350] 一条关系线,是指在利用一个谓词将一个实体/实体类别与另一个实体/ 实体类别进行衔接,使得两个实体/实体类别间产生谓词关系后所得的两个实体/实体类别与谓词间的对应关系。

[0351] 例如问题语句“治疗感冒的药物有哪些?”中包括一个实体“感冒”,一个实体类别“药物”,且实体“感冒”与实体类别“药物”通过谓词“治疗”产生谓词关系,从而该问题语句包括一条关系线;问题语句“什么药可以治疗感冒”中同样包括一个实体“感冒”,一个实体类别“药物”(以疑问词“什么(药)”指代的“药物”这一实体类别),且实体“感冒”与实体类别“药物”通过谓词“治疗”产生谓词关系,从而该问题语句同样包括一条关系线。在识别问题语句所属的问句类别时,若通过上述处理已确定出句子中共包括两个实体/实体类别,则根据句子的句法特征,一般会认为两个实体/实体类别之间需要通过一谓词关系进行衔接,从而,可确定出该问题语句所属的问句类别为简单关系型问题。

[0352] 若问题语句中不包括关系线,而只对实体或实体类别的属性进行提问,例如问题语句“白癜风是什么?”中不包括关系线,仅对“白癜风”这一实体的“introduction”属性进行提问,“医院在哪里?”中不包括关系线,仅对“医院”这一实体类别的位置属性进行提问,则此类问题语句所属的问句类别为简单属性型问题,具体实施中,对于服务器/服务器集群或终端设备等执行主体来说,可在问题语句中仅包括一个实体/实体类别,以及一个由疑问词指代的属性(如“白癜风是什么”中包括由疑问词“什么”所指代的“白癜风”的“introduction”属性,“医院在哪里?”中包括由疑问词“哪里”所指代的“医院”的位置属性)的情况下,将问题语句的问句类别识别为所述简单属性型问题。

[0353] 一个问题语句,若其不属于上述的简单关系型问题,也不属于上述的简单属性型问题,则将其所属的问句类别确定为复杂关系属性型问题。

[0354] 当然,具体实施中,也可以在识别问题语句中所包括的实体、实体类别、属性及疑问词等句子成分的基础上,同时识别其中的谓词关系,进而结合所识别出的实体、实体类别、属性、疑问词及谓词关系等来确定句子中是否包括关系线、在包括关系线情况下具体包括的关系线条数以及是否包括对属性的提问等信息,进而确定出问题语句所属的问句类别,本实施例对确定问题语句所属问句类别的实现过程不作限定。

[0355] 所述预定问句类别,可以根据实际实施情况对其进行设定,例如将其设定为包括上述的简单关系型问题、简单属性型问题及复杂关系属性型问题中的全部类别;或者将其设定为仅包括上述三种类别中的其中之一(例如,仅包括所述复杂关系属性型问题)等等,本实施例对此不作限定。

[0356] 可利用相应依存句法分析算法或依存句法分析器,分析问题语句的依存句法结构信息,以得到问题语句的各组成元素间的依存信息,问题语句的各组成元素间的依存信息能够反映问题语句中各组成元素间的实际依存(依赖)情况。

[0357] 例如问题语句“今天天气怎么样”，通过依存句法分析，可以将其解析成“今天-天气-怎么样”，其中，“今天”依赖于“天气”，“天气”依赖于“怎么样”，而“怎么样”则不依赖于任何词语。

[0358] 在对属于预定问句类别的问题语句进行依存句法分析，得到所述问题语句的各组成元素间的依存信息后，本申请基于所述问题语句的各组成元素间的依存信息以及所述问题语句的句子成分，对所述问题语句进行结构化解析，以得到所述问题语句的结构化信息，该对问题语句进行结构化解析，从而得到问题语句的结构化信息的处理过程会在之后的实施例中进行详述，具体请参见后续实施例对该部分的说明。

[0359] 根据以上方案可知，本实施例提供的服务器，在获得问题语句后，基于问题语句的句子成分确定问题语句所属的问句类别，并在问题语句属于预定问句类别的情况下，对问题语句进行依存句法分析，得到问题语句的各组成元素间的依存信息，进而基于问题语句的各组成元素间的依存信息以及所述问题语句的句子成分，对问题语句进行结构化解析，得到问题语句的结构化信息。通过利用问题语句的各组成元素间的依存信息以及问题语句的句子成分，对问题语句进行结构化解析，可达到较好的解析效果，进而可基于解析结果准确地识别用户问句意图。

[0360] 需要说明的是，本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

[0361] 为了描述的方便，描述以上系统或装置时以功能分为各种模块或单元分别描述。当然，在实施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0362] 通过以上的实施方式的描述可知，本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在存储介质中，如ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等等）执行本申请各个实施例或者实施例的某些部分所述的方法。

[0363] 最后，还需要说明的是，在本文中，诸如第一、第二、第三和第四等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，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、物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、物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、物品或者设备中还存在另外的相同要素。

[0364] 以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

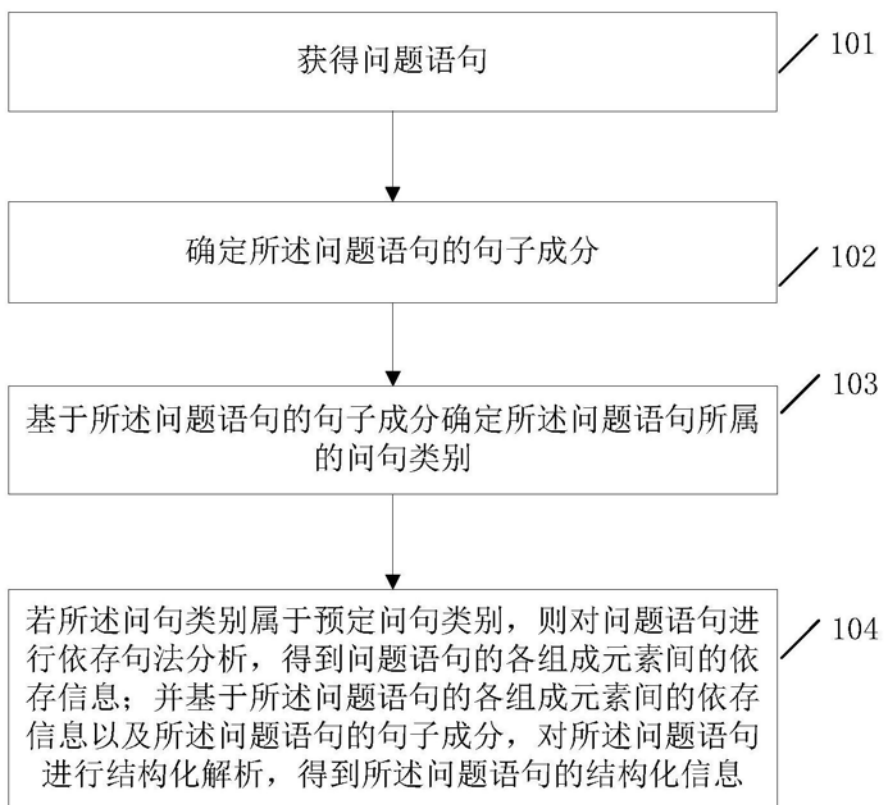


图1

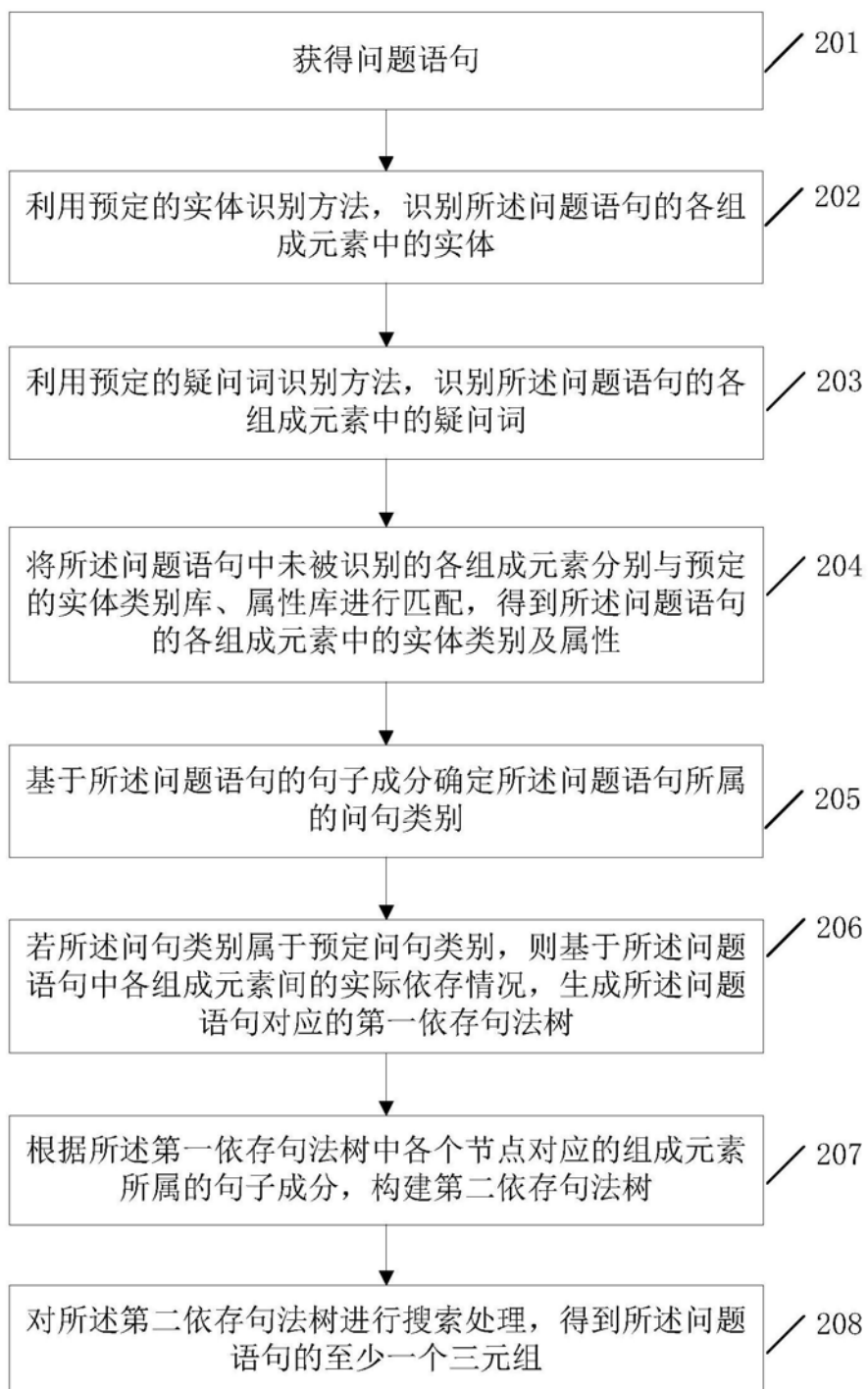


图2

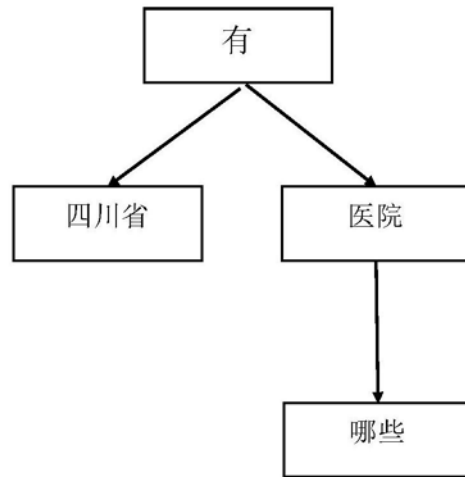


图3 (a)

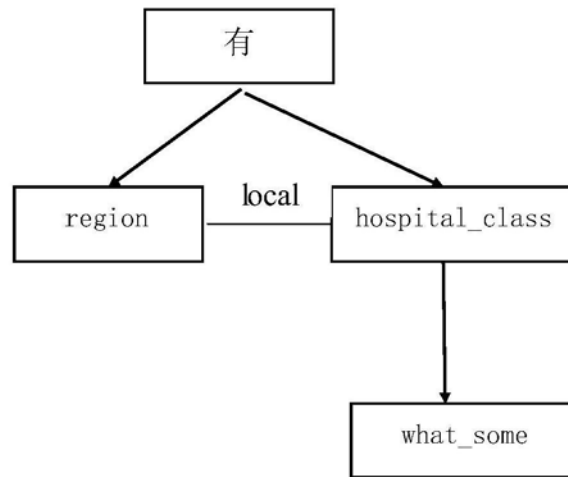


图3 (b)

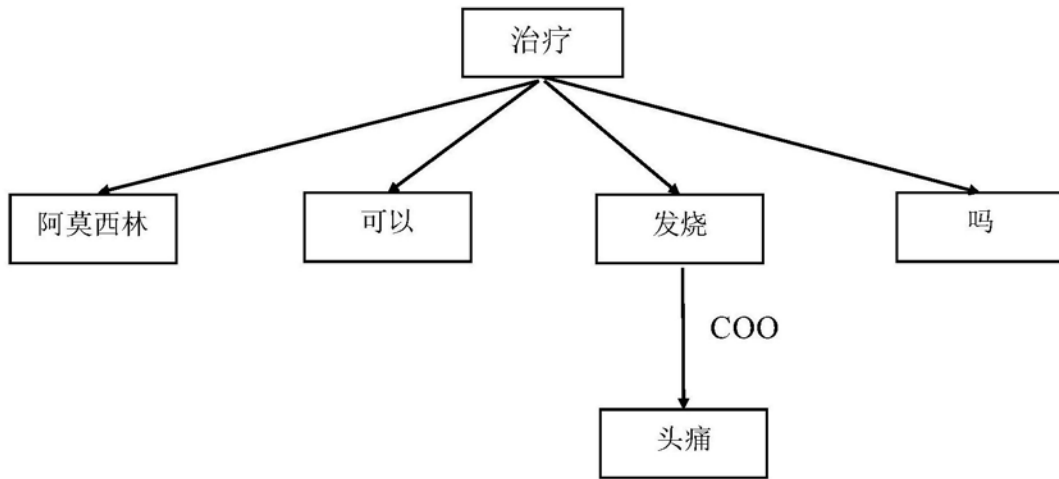


图4 (a)

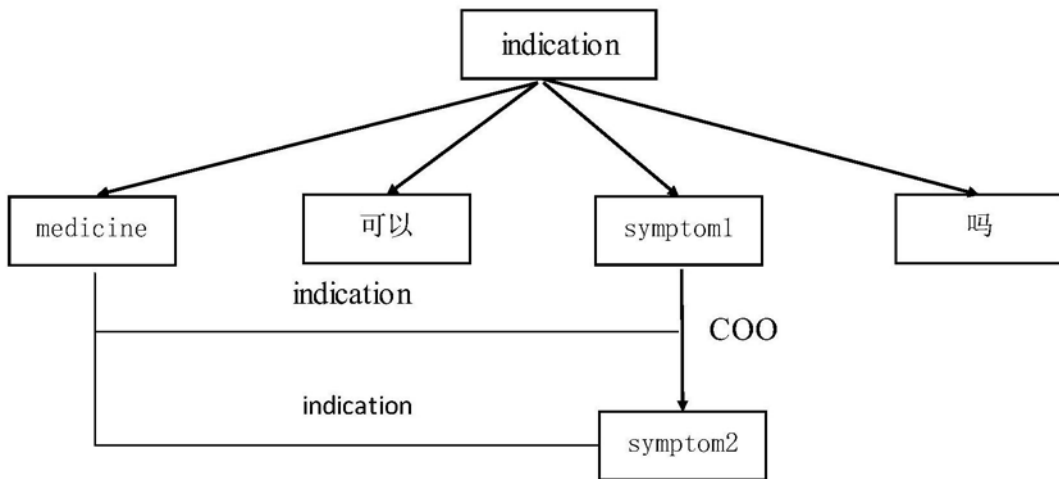


图4 (b)

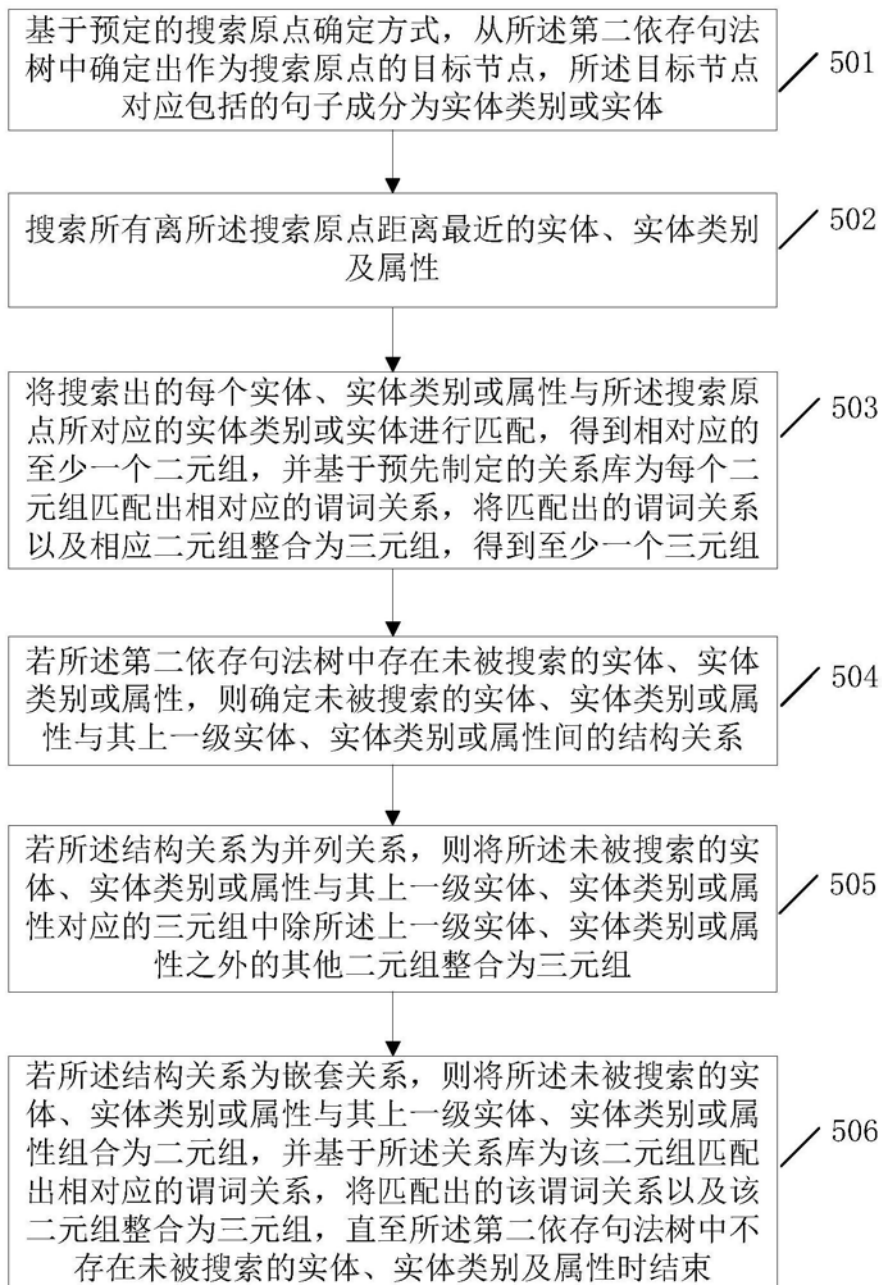


图5

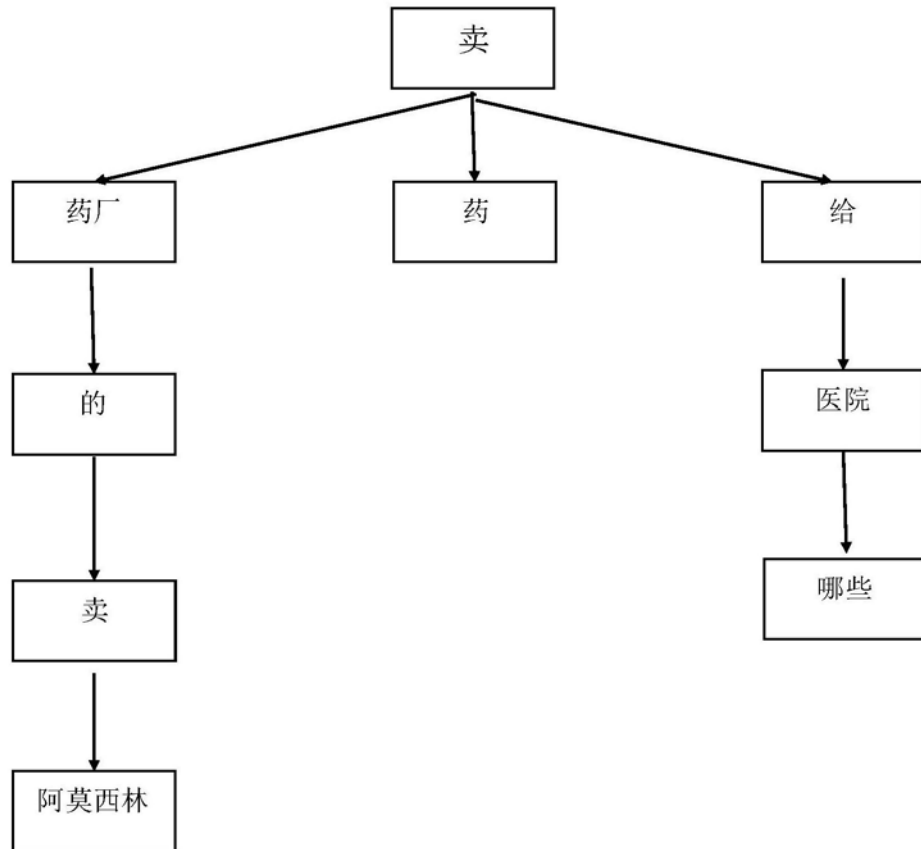


图6(a)

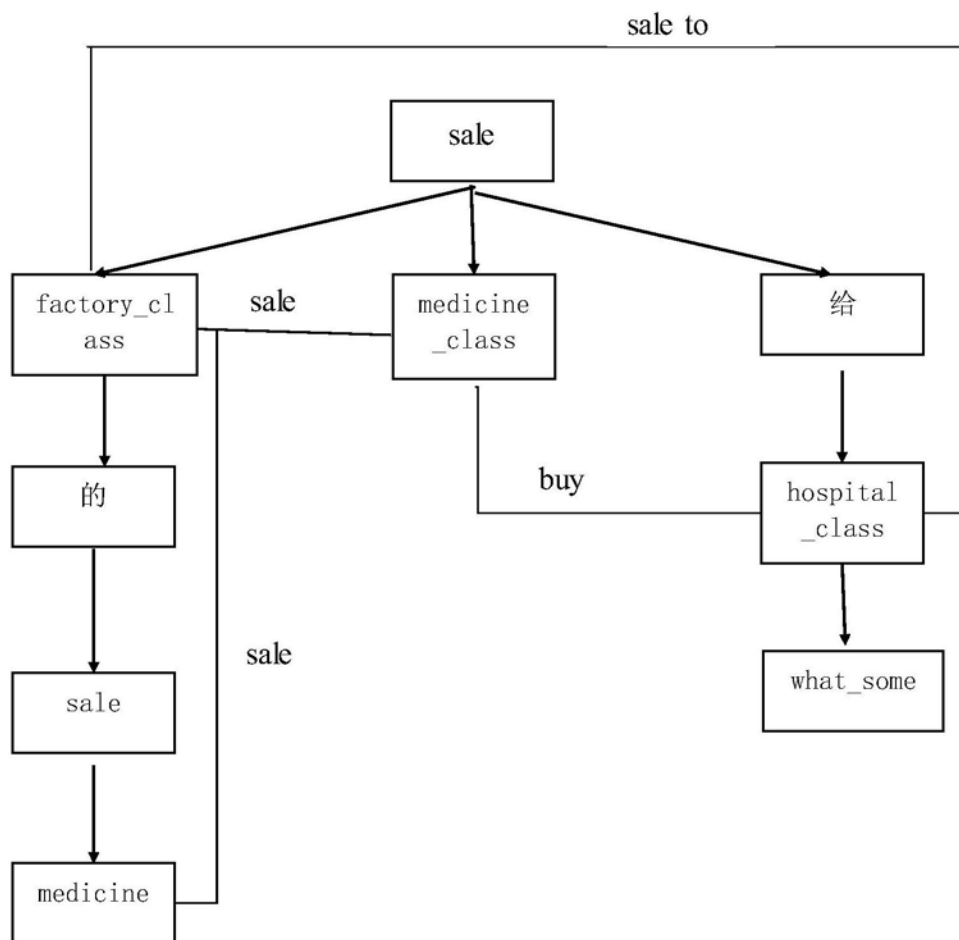


图6 (b)

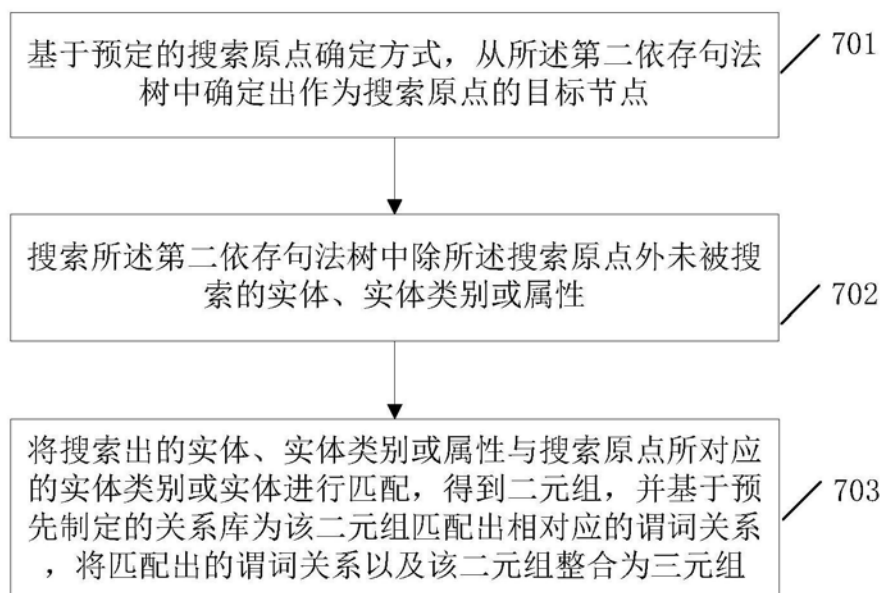


图7

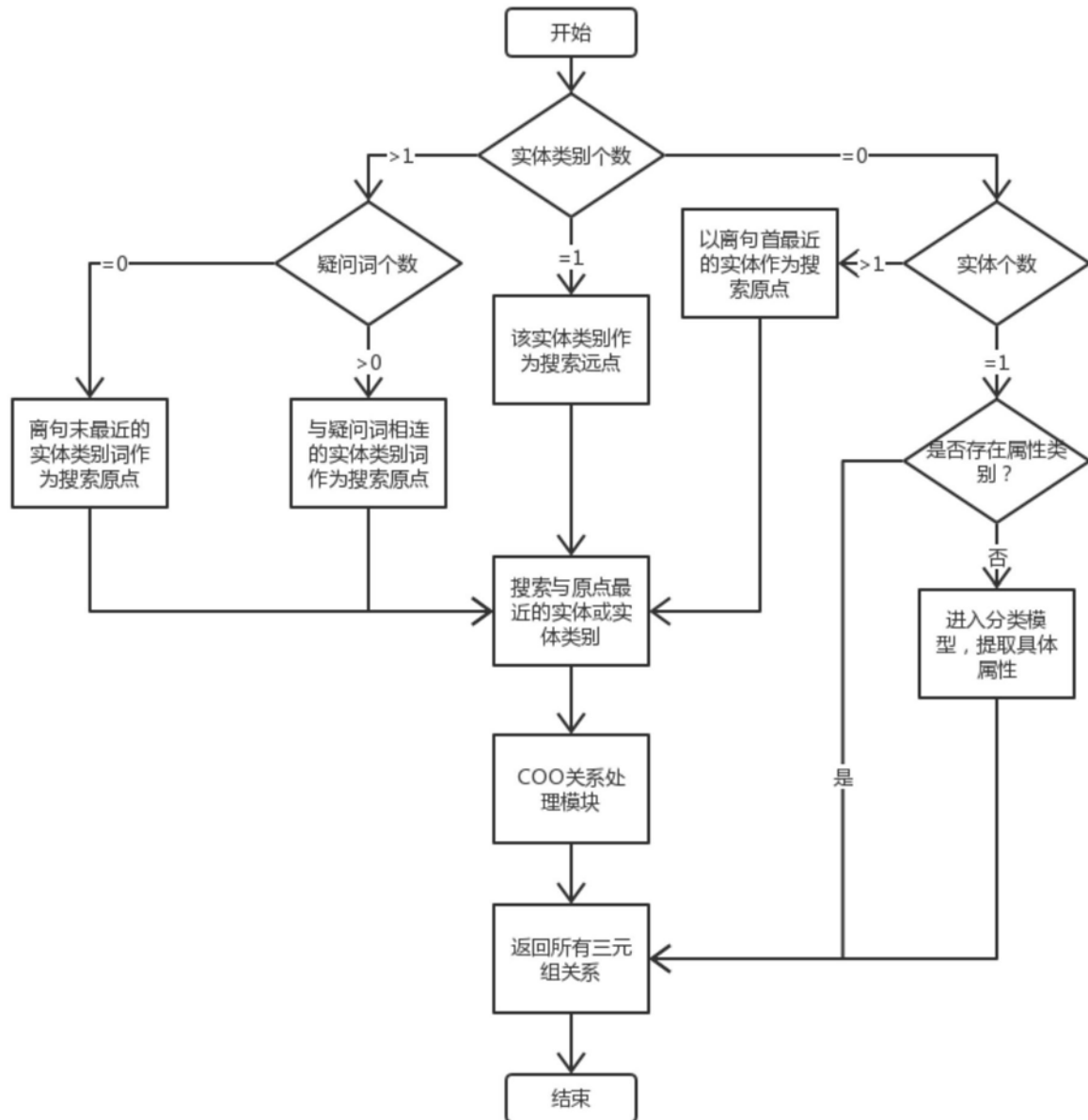


图8

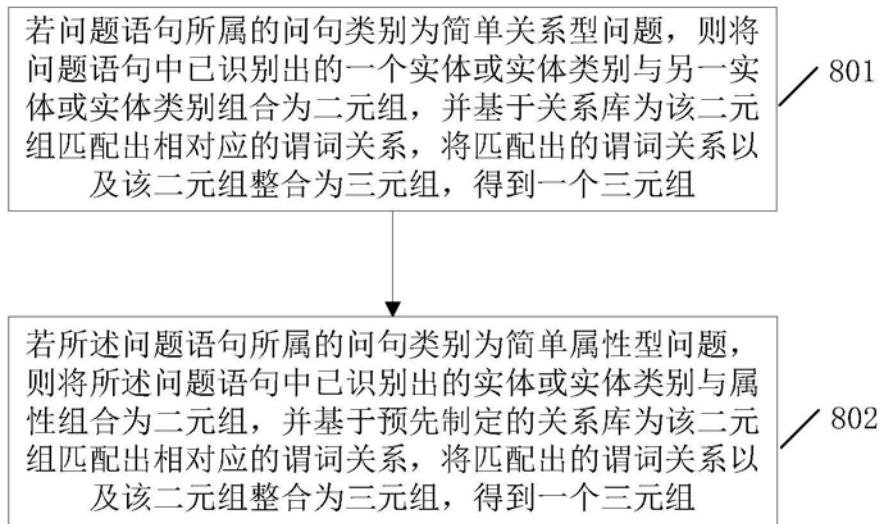


图9

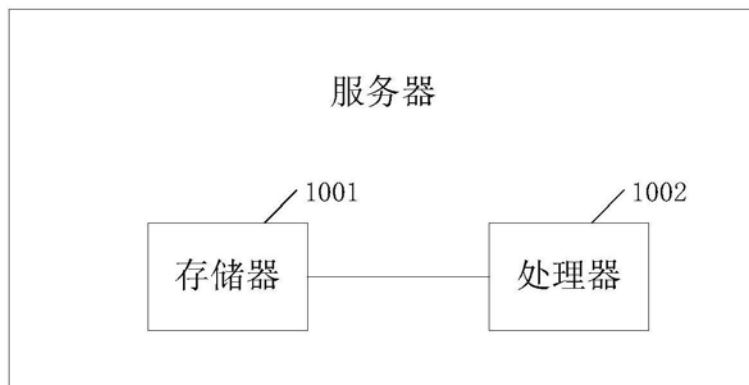


图10

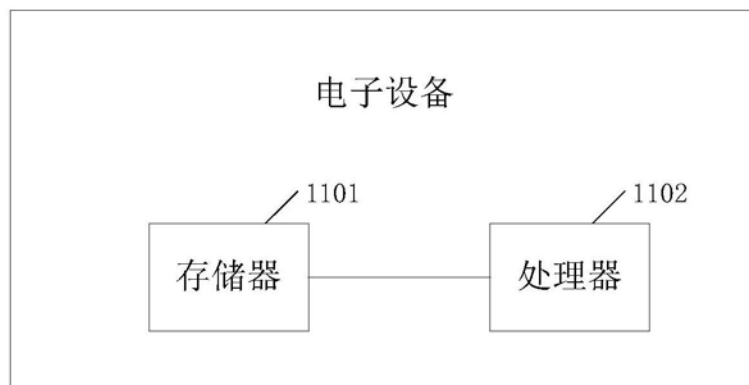


图11