



(12)发明专利申请

(10)申请公布号 CN 111522927 A

(43)申请公布日 2020.08.11

(21)申请号 202010296375.2

(22)申请日 2020.04.15

(71)申请人 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 郭辰阳 钱璟 吕继根

(74)专利代理机构 北京英赛嘉华知识产权代理
有限责任公司 11204
代理人 王达佐 马晓亚

(51) Int. Cl.
G06F 16/332(2019.01)
G06F 16/36(2019.01)
G06F 40/289(2020.01)

权利要求书2页 说明书9页 附图3页

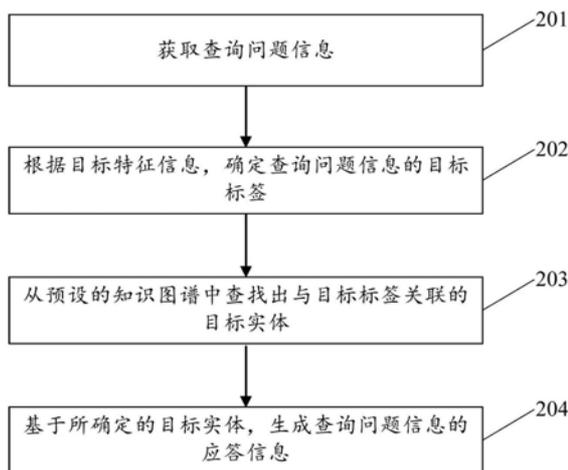
(54)发明名称

基于知识图谱的实体查询方法和装置

(57)摘要

本公开的实施例公开了基于知识图谱的实体查询方法和装置,涉及知识图谱领域。该方法的一具体实施方式包括:获取查询问题信息,根据查询问题信息所包括的用于描述目标关键词的目标特征信息,确定查询问题信息的目标标签,从预设的知识图谱中查找出与目标标签关联的目标实体,基于所确定的目标实体,生成查询问题信息的应答信息。该方法可以根据查询问题信息的标签,快速准确地知识图谱中查找出符合特征信息的实体,并给出对应的应答信息。

200



1. 一种基于知识图谱的实体查询方法,包括:
 - 获取查询问题信息,其中,所述查询问题信息包括目标关键词以及用于描述目标关键词的目标特征信息;
 - 根据所述目标特征信息,确定所述查询问题信息的目标标签;
 - 从预设的知识图谱中查找出与所述目标标签关联的目标实体;
 - 基于所确定的目标实体,生成所述查询问题信息的应答信息。
2. 根据权利要求1所述的方法,其中,所述知识图谱是按照以下步骤生成的:
 - 获取问题信息,所述问题信息包括指定关键词以及用于描述所述指定关键词的特征信息,所述指定关键词包括实体类别;
 - 根据所述特征信息,确定所述问题信息的标签;
 - 获取所述问题信息的搜索结果对应的页面信息,从所述页面信息中抽取属于所述实体类别的实体;
 - 根据所述实体以及所述标签,确定所述标签与所述实体的关联关系;
 - 根据所述关联关系,生成所述知识图谱。
3. 根据权利要求2所述的方法,其中,所述根据所述特征信息,确定所述问题信息的标签,包括:
 - 对所述特征信息进行分词处理,得到分词结果;
 - 根据所述分词结果,确定所述问题信息的标签。
4. 根据权利要求2所述的方法,其中,所述获取所述问题信息的搜索结果对应的页面信息,从所述页面信息中抽取属于所述实体类别的实体,包括:
 - 根据问题信息,获取排序在前预设个数的搜索结果;
 - 从各所述搜索结果对应的页面信息中,抽取属于所述实体类别的实体。
5. 根据权利要求2所述的方法,其中,所述根据所述实体以及所述标签,确定所述标签与所述实体的关联关系,包括:
 - 对所述实体所在的页面信息进行聚合,根据页面信息的聚合结果确定出第一聚合度值;
 - 对包含所述标签的问题信息进行聚合,根据问题信息的聚合结果确定出第二聚合度值;
 - 根据所述第一聚合度值与所述第二聚合度值,确定所述标签与所述实体的关联关系。
6. 根据权利要求2所述的方法,其中,所述方法还包括:
 - 从所述标签中过滤掉预先设置的非法标签表中的非法标签,从所述实体中过滤掉预先设置的非法实体表中的非法实体。
7. 一种基于知识图谱的实体查询装置,包括:
 - 获取单元,被配置成获取查询问题信息,其中,所述查询问题信息包括目标关键词以及用于描述目标关键词的目标特征信息;
 - 确定单元,被配置成根据所述目标特征信息,确定所述查询问题信息的目标标签;
 - 查找单元,被配置成从预设的知识图谱中查找出与所述目标标签关联的目标实体;
 - 输出单元,被配置成基于所确定的目标实体,生成所述查询问题信息的应答信息。
8. 根据权利要求7所述的装置,其中,所述装置还包括生成单元,被配置成通过以下模

块生成知识图谱：

获取模块，被配置成获取问题信息，所述问题信息包括指定关键词以及用于描述所述指定关键词的特征信息，所述指定关键词包括实体类别；

第一确定模块，被配置成根据所述特征信息，确定所述问题信息的标签；

抽取模块，被配置成获取所述问题信息的搜索结果对应的页面信息，从所述页面信息中抽取出属于所述实体类别的实体；

第二确定模块，被配置成根据所述实体以及所述标签，确定所述标签与所述实体的关联关系；

生成模块，被配置成根据所述关联关系，生成所述知识图谱。

9. 根据权利要求8所述的装置，其中，所述第一确定模块进一步被配置成：

对所述特征信息进行分词处理，得到分词结果；

根据所述分词结果，确定所述问题信息的标签。

10. 根据权利要求8所述的装置，其中，所述抽取模块进一步被配置成：

根据问题信息，获取排序在前预设个数的搜索结果；

从各所述搜索结果对应的页面信息中，抽取出属于所述实体类别的实体。

11. 根据权利要求8所述的装置，其中，所述第二确定模块进一步被配置成：

对所述实体所在的页面信息进行聚合，根据页面信息的聚合结果确定出第一聚合度值；

对包含所述标签的问题信息进行聚合，根据问题信息的聚合结果确定出第二聚合度值；

根据所述第一聚合度值与所述第二聚合度值，确定所述标签与所述实体的关联关系。

12. 根据权利要求8所述的装置，其中，所述装置还包括：

过滤单元，被配置成从所述标签中过滤掉预先设置的非法标签表中的非法标签，从所述实体中过滤掉预先设置的非法实体表中的非法实体。

13. 一种电子设备，包括：

一个或多个处理器；

存储装置，其上存储有一个或多个程序，

当一个或多个程序被一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-6中任一所述的方法。

14. 一种计算机可读介质，其上存储有计算机程序，其中，该程序被处理器执行时实现如权利要求1-6中任一所述的方法。

基于知识图谱的实体查询方法和装置

技术领域

[0001] 本公开的实施例涉及计算机技术领域,具体涉及基于知识图谱的实体查询方法和装置。

背景技术

[0002] 知识图谱本质上是一种语义网络,其节点代表实体,边代表实体之间的各种语义关系。其强大的语义处理和互联组织能力,为智能化信息应用提供了基础。

[0003] 在基于知识图谱的问答领域,问答形式多种多样,其中,泛需求问答是指基于知识图谱能够确定出多个与泛需求问题对应结果的问答形式。

发明内容

[0004] 本公开的实施例提出了基于知识图谱的实体查询方法和装置。

[0005] 第一方面,本公开的实施例提供了一种基于知识图谱的实体查询方法,该方法包括:获取查询问题信息,其中,查询问题信息包括目标关键词以及用于描述目标关键词的目标特征信息;根据目标特征信息,确定查询问题信息的目标标签;从预设的知识图谱中查找出与目标标签关联的目标实体;基于所确定的目标实体,生成查询问题信息的应答信息。

[0006] 在一些实施例中,上述知识图谱是按照以下步骤生成的:获取问题信息,问题信息包括指定关键词以及用于描述指定关键词的特征信息,指定关键词包括实体类别;根据特征信息,确定问题信息的标签;获取问题信息的搜索结果对应的页面信息,从页面信息中抽取出属于实体类别的实体;根据实体以及标签,确定标签与实体的关联关系;根据关联关系,生成知识图谱。

[0007] 在一些实施例中,根据特征信息,确定问题信息的标签,包括:对特征信息进行分词处理,得到分词结果;根据分词结果,确定问题信息的标签。

[0008] 在一些实施例中,获取问题信息的搜索结果对应的页面信息,从页面信息中抽取出属于实体类别的实体,包括:根据问题信息,获取排序在前预设个数的搜索结果;从各搜索结果对应的页面信息中,抽取出属于实体类别的实体。

[0009] 在一些实施例中,根据实体以及标签,确定标签与实体的关联关系,包括:对实体所在的页面信息进行聚合,根据页面信息的聚合结果确定出第一聚合度值;对包含标签的问题信息进行聚合,根据问题信息的聚合结果确定出第二聚合度值;根据第一聚合度值与第二聚合度值,确定标签与实体的关联关系。

[0010] 在一些实施例中,上述方法还包括:从标签中过滤掉预先设置的非法标签表中的非法标签,从实体中过滤掉预先设置的非法实体表中的非法实体。

[0011] 第二方面,本公开的实施例提供了一种基于知识图谱的实体查询装置,该装置包括:获取单元,被配置成获取查询问题信息,其中,查询问题信息包括目标关键词以及用于描述目标关键词的目标特征信息;确定单元,被配置成根据目标特征信息,确定查询问题信息的目标标签;查找单元,被配置成从预设的知识图谱中查找出与目标标签关联的目标实

体;输出单元,被配置成基于所确定的目标实体,生成查询问题信息的应答信息。

[0012] 在一些实施例中,上述装置还包括生成单元,被配置成通过以下模块生成知识图谱:获取模块,被配置成获取问题信息,问题信息包括指定关键词以及用于描述指定关键词的特征信息,指定关键词包括实体类别;第一确定模块,被配置成根据特征信息,确定问题信息的标签;抽取模块,被配置成获取问题信息的搜索结果对应的页面信息,从页面信息中抽取出属于实体类别的实体;第二确定模块,被配置成根据实体以及标签,确定标签与实体的关联关系;生成模块,被配置成根据关联关系,生成知识图谱。

[0013] 在一些实施例中,第一确定模块进一步被配置成:对特征信息进行分词处理,得到分词结果;根据分词结果,确定问题信息的标签。

[0014] 在一些实施例中,抽取模块进一步被配置成:根据问题信息,获取排序在前预设个数的搜索结果;从各搜索结果对应的页面信息中,抽取出属于实体类别的实体。第二确定模块进一步被配置成

[0015] 在一些实施例中,第二确定模块进一步被配置成:对实体所在的页面信息进行聚合,根据页面信息的聚合结果确定出第一聚合度值;对包含标签的问题信息进行聚合,根据问题信息的聚合结果确定出第二聚合度值;根据第一聚合度值与第二聚合度值,确定标签与实体的关联关系。

[0016] 在一些实施例中,上述装置还包括:过滤单元,被配置成从标签中过滤掉预先设置的非法标签表中的非法标签,从实体中过滤掉预先设置的非法实体表中的非法实体。

[0017] 第三方面,本公开的实施例提供了一种电子设备,该电子设备包括:一个或多个处理器;存储装置,其上存储有一个或多个程序;当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

[0018] 第四方面,本公开的实施例提供了一种计算机可读介质,其上存储有计算机程序,该程序被处理器执行时实现如第一方面中任一实现方式描述的方法。

[0019] 本公开的实施例提供的基于知识图谱的实体查询方法和装置,通过获取查询问题信息,然后根据查询问题信息所包括的用于描述目标关键词的目标特征信息,确定查询问题信息的目标标签,从预设的知识图谱中查找出与目标标签关联的目标实体,最后基于所确定的目标实体,生成查询问题信息的应答信息,根据查询问题信息的标签,能够快速准确地知识图谱中查找出符合特征信息的实体。

附图说明

[0020] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本公开的其它特征、目的和优点将会变得更明显:

[0021] 图1是本公开的一个实施例可以应用于其中的示例性系统架构图;

[0022] 图2是根据本公开的基于知识图谱的实体查询方法的一个实施例的流程图;

[0023] 图3是根据本公开的基于知识图谱的实体查询方法的一个应用场景的示意图;

[0024] 图4是根据本公开的基于知识图谱的实体查询方法的一个实施例中生成知识图谱的流程图;

[0025] 图5是根据本公开的基于知识图谱的实体查询装置的一个实施例的结构示意图;

[0026] 图6是适于用来实现本公开的实施例的电子设备的结构示意图。

具体实施方式

[0027] 下面结合附图和实施例对本公开作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释相关发明,而非对该发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与有关发明相关的部分。

[0028] 需要说明的是,在不冲突的情况下,本公开中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本公开。

[0029] 图1示出了可以应用本公开的基于知识图谱的实体查询方法或基于知识图谱的实体查询装置的示例性架构100。

[0030] 如图1所示,系统架构100可以包括终端设备101、102、103,网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0031] 终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用,例如搜索类应用、信息查询类应用、自然语言处理类应用等、即时通信工具、社交平台应用、浏览器类应用等。

[0032] 终端设备101、102、103可以是硬件,也可以是软件。当终端设备101、102、103为硬件时,可以是具有显示屏并且支持互联网访问的各种电子设备,包括但不限于智能手机、平板电脑、笔记本电脑、和台式计算机等等。当终端设备101、102、103为软件时,可以安装在上述所列举的电子设备中。其可以实现成多个软件或软件模块(例如用来提供分布式服务),也可以实现成单个软件或软件模块。在此不做具体限定。

[0033] 服务器105可以是为终端设备101、102、103安装的应用提供支持的服务器。服务器105可以自动处理终端设备101、102、103输入的查询问题信息,查找出与查询问题信息相匹配的应答信息。

[0034] 需要说明的是,本公开的实施例所提供的基于知识图谱的实体查询方法可以通过终端设备101、102、103执行,也可以通过服务器105执行。相应地,基于知识图谱的实体查询装置可以设置于终端设备101、102、103中,也可以设置于服务器105中。本申请对此不做限定。

[0035] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0036] 继续参考图2,示出了根据本公开的基于知识图谱的实体查询方法的一个实施例的流程200。该基于知识图谱的实体查询方法包括以下步骤:

[0037] 步骤201,获取查询问题信息。

[0038] 在本实施例中,上述基于知识图谱的实体查询方法的执行主体(例如图1所示的服务器)可以从搜索类应用、信息查询类应用获取用户输入的查询问题信息。在这里,查询问题信息可以包括目标关键词以及用于描述目标关键词的目标特征信息,目标关键词可以表征用户期待查询的目标对象,目标特征信息可以用来描述目标关键词的属性信息,例如查询问题信息是“表示看的字”,那么目标关键词可以是“字”,目标特征信息可以是“表示看”。

[0039] 步骤202,根据目标特征信息,确定查询问题信息的目标标签。

[0040] 在本实施例中,上述执行主体可以通过多种方式从目标特征信息中确定查询问题信息的目标标签。例如,上述执行主体可以利用基于样本特征信息与样本标签训练出的标

签识别模型,确定出查询问题信息的目标标签。或者,上述执行主体可以根据预先设置的包含标签的标签数据库,从目标特征信息中确定出所包含的标签作为查询问题信息的目标标签。若目标特征信息中不包括标签数据库中的标签,可以直接将目标特征信息作为查询问题信息的目标标签。

[0041] 步骤203,从预设的知识图谱中查找出与目标标签关联的目标实体。

[0042] 在本实施例中,上述执行主体可以从预设的知识图谱中定位出目标标签,并查找出与目标标签关联的目标实体。在这里,目标实体可以是符合目标特征信息的实体。上述知识图谱可以是基于实体和实体属性之间的关联关系预先进行构建的,可以利用关系数据库或图数据库存储知识图谱中的实体、实体之间的关联关系。知识图谱中可以包括标签与实体的关联对,例如标签为“很美很冷门”,与标签关联的实体为“浮云在空碧,来往议阴晴”,那么“很美很冷门”与“浮云在空碧,来往议阴晴”为标签与实体的关联对。

[0043] 步骤204,基于所确定的目标实体,生成查询问题信息的应答信息。

[0044] 在本实施例中,上述执行主体可以基于步骤203中查找到的目标实体,可以直接将目标实体作为查询问题信息的应答信息返回给用户。例如,查询问题信息为“表示思乡之情的诗”,目标实体为“《静夜思》”,则查询问题信息的应答信息可以为“《静夜思》”。或者,上述执行主体也可以基于查找出的实体,筛选出属于目标关键词所对应的实体类别的实体作为应答信息。例如,查询问题信息为“表示看的词语”,则目标实体为“瞧”、“俯视”,则查询问题信息的应答信息可以为“俯视”。在这里,应答信息可以为针对查询问题信息返回的、向用户呈现的答案,可以是将目标实体输入至预先设置的答案模板中生成的信息。

[0045] 继续参考图3,图3是根据本公开的基于知识图谱的实体查询方法的一个应用场景的示意图。在图3的应用场景中,执行主体可以为服务器302,用户从终端301安装的搜索类应用中输入查询问题信息,服务器302根据查询问题信息所包括的用于描述目标关键词的目标特征信息,确定出查询问题信息的目标标签,并从知识图谱303中定位出目标标签,确定出与目标标签关联的目标实体(实体A、实体B),再根据目标实体(实体A、实体B)输出应答信息,服务器302将包含目标实体(实体A、实体B)的应答信息返回给终端301。

[0046] 本公开的上述实施例提供的方法通过获取查询问题信息,然后根据查询问题信息所包括的用于描述目标关键词的目标特征信息,确定查询问题信息的目标标签,从预设的知识图谱中查找出与目标标签关联的目标实体,最后基于所确定的目标实体,生成查询问题信息的应答信息,根据查询问题信息的标签,能够快速准确地在知识图谱中查找出符合特征信息的实体。

[0047] 继续参考图4,示出了根据本公开的基于知识图谱的实体查询方法的一个实施例中生成知识图谱的流程图400。该知识图谱的生成方法包括以下步骤:

[0048] 步骤401,获取问题信息。

[0049] 在本实施例中,上述执行主体可以获取用户在搜索类应用、搜索平台中输入的问题信息,具体地,可以设置指定关键词筛选出指定关键词对应的垂类问题信息。在这里,问题信息可以包括指定关键词以及用于描述指定关键词的特征信息,指定关键词可以包括实体类别。例如问题信息是“很美很冷门的诗句”,那么指定关键词可以是“诗句”,表示问题信息都是属于用于描述诗句实体类别的问题,用于描述指定关键词的特征信息可以是“很美很冷门”。

[0050] 步骤402,根据特征信息,确定问题信息的标签。

[0051] 在本实施例中,上述执行主体可以从问题信息中所包括特征信息的内容中确定出问题信息的标签。具体地,上述执行主体可以从预先设置的包括多个标签词的标签数据库中,确定出特征信息中所包含的标签作为问题信息的标签。

[0052] 在本实施的一些可选的实现方式中,根据特征信息,确定问题信息的标签,包括:对特征信息进行分词处理,得到分词结果;根据分词结果,确定问题信息的标签。

[0053] 在该可选实现方式中,上述执行主体可以利用分词工具对特征信息进行分词处理,从得到的分词结果中获取问题信息的标签。例如,特征信息是“很美很冷门的”,那么分词结果可以包括“很美”、“很冷门”、“很美很冷门”,可以直接将“很美”、“很冷门”、“很美很冷门”其中任一个作为问题信息的标签。可选地,上述执行主体可以将分词结果中最大粒度的词作为标签,即将“很美很冷门”作为问题信息的标签,这样可以更加精准地表示指定关键词的特征。

[0054] 步骤403,获取问题信息的搜索结果对应的页面信息,从页面信息中抽取出属于实体类别的实体。

[0055] 在本实施例中,上述执行主体可以从搜索类应用、搜索平台获取问题信息的搜索结果,根据搜索结果对应的链接获取搜索结果对应的页面信息,从页面信息中抽取出属于实体类别的实体。具体地,上述执行主体还可以根据页面信息中包含的内容预先设置实体抽取模式,利用实体抽取模式确定出页面信息中属于实体类别的部分,例如页面信息以诗词、诗名、诗词作者的内容排布,利用实体抽取模式定位出页面信息中的诗词部分,从诗词部分抽取出属于实体类别的实体。

[0056] 在本实施的一些可选的实现方式中,获取问题信息的搜索结果对应的页面信息,从页面信息中抽取出属于实体类别的实体,包括:根据问题信息,获取排序在前预设个数的搜索结果;从各搜索结果对应的页面信息中,抽取出属于实体类别的实体。

[0057] 在该可选实现方式中,上述执行主体可以从根据问题信息,选取出排序在前预设个数的搜索结果,进行实体的抽取。这样可以与问题信息关联程度越强的搜索结果中获取置信度较高的实体。

[0058] 步骤404,根据实体以及标签,确定标签与实体的关联关系。

[0059] 在本实施例中,上述执行主体可以将从问题信息中抽取的标签,与问题信息的搜索结果对应的页面信息中抽取的实体,确定为具有关联关系的标签实体对。

[0060] 在本实施的一些可选的实现方式中,根据实体以及标签,确定标签与实体的关联关系,包括:对实体所在的页面信息进行聚合,根据页面信息的聚合结果确定出第一聚合度值;对包含标签的问题信息进行聚合,根据问题信息的聚合结果确定出第二聚合度值;根据第一聚合度值与第二聚合度值,确定标签与实体的关联关系。

[0061] 在该可选实现方式中,上述执行主体可以对实体所在的页面信息进行聚合,生成页面信息的集合,根据集合中所包括的页面信息的个数确定出第一聚合度值,对包含标签的问题信息进行聚合,生成问题信息的集合,根据集合中包括的问题信息的个数确定出第二聚合度值,可以设置第一聚合度阈值与第二聚合度阈值,筛选出有效实体与有效标签作为具有关联关系的有效标签实体对。在这里,无论实体在同一页面信息中出现一次或者多次,将实体在该页面信息的个数都记为一次。第一聚合度值表示签与实体的关联程度以

及实体的可信度,页面信息的个数越多,第一聚合度值越大,说明标签与实体的关联程度越高,实体的可信度越高。第二聚合度值表示标签的热门程度以及置信度,问题信息的个数越多,第二聚合度值越大,说明标签的热门程度越高,标签的置信度越高。例如上述执行主体可以将第一聚合度阈值设置为3,将第二聚合度阈值设置为3,即页面信息的个数超过3个对应的实体作为有效实体,问题信息的个数超过3个对应的标签作为有效标签,将有效实体与有效标签关联。

[0062] 步骤405,根据关联关系,生成知识图谱。

[0063] 在本实施例中,上述执行主体可以基于标签与实体的关联关系构建知识图谱。知识图谱可以包括表征实体和标签的节点、以及表征实体与标签之间关联关系的连接边。具体地,上述执行主体首先可以根据实体与实体之间的关系创建各个实体节点,构建出初始知识图谱,然后将与实体相关联的标签添加至初始知识图谱中,具体地,可以利用边将实体节点、以及与实体具有关联关系的标签连接,得到知识图谱。

[0064] 在本实施的一些可选的实现方式中,上述方法还包括:从标签中过滤掉预先设置的非法标签表中的非法标签,从实体中过滤掉预先设置的非法实体表中的非法实体。

[0065] 在该可选实现方式中,上述执行主体可以预先从反动词库中获取反动词,从色情词库中获取色情词,利用反动词和色情词构建非法标签表、非法实体表,然后从步骤402获取的标签中过滤掉预先设置的非法标签表中的非法标签,从步骤404获取的实体中过滤掉预先设置的非法实体表中的非法实体。通过该实现方式可以过滤掉标签和实体中的敏感词,避免用户收到包含敏感词的应答信息,防止不良信息的传播。

[0066] 本实施例中生成知识图谱的流程图400,利用问题信息与问题信息的搜索结果对应的页面信息之间的关系,将问题信息中抽取的标签、与搜索结果对应的页面信息中抽取的实体关联,基于标签与实体的关联关系添生成知识图谱,丰富实体所关联的标签以及丰富知识图谱的实体属性特征,进而可以将丰富后的知识图谱应用至泛需求问答中,根据泛需求问答中的查询问题信息所包含的目标标签,从知识图谱中筛选出与目标标签关联的实体作为应答信息,提高了应答信息的丰富度及准确性,满足了泛需求问答的需求。

[0067] 进一步参考图5,作为对上述各图所示方法的实现,本公开提供了基于知识图谱的实体查询装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0068] 如图5所示,本实施例提供的基于知识图谱的实体查询装置500包括获取单元501、确定单元502、查找单元503以及输出单元505。其中,获取单元501,被配置成获取查询问题信息,其中,查询问题信息包括目标关键词以及用于描述目标关键词的目标特征信息;确定单元502,被配置成根据目标特征信息,确定查询问题信息的目标标签;查找单元503,被配置成从预设的知识图谱中查找出与目标标签关联的目标实体;输出单元504,被配置成基于所确定的目标实体,生成查询问题信息的应答信息。

[0069] 在本实施例中,基于知识图谱的实体查询装置500中:获取单元501、确定单元502、查找单元503以及输出单元505的具体处理及其所带来的技术效果可分别参考图2对应实施例中的步骤201、步骤202、步骤203和步骤204的相关说明,在此不再赘述。

[0070] 在本实施例的一些可选的实现方式中,上述装置还包括生成单元(图中未示出),被配置成通过以下模块生成知识图谱:获取模块,被配置成获取问题信息,问题信息包括指

定关键词以及用于描述指定关键词的特征信息,指定关键词包括实体类别;第一确定模块,被配置成根据特征信息,确定问题信息的标签;抽取模块,被配置成获取问题信息的搜索结果对应的页面信息,从页面信息中抽取出属于实体类别的实体;第二确定模块,被配置成根据实体以及标签,确定标签与实体的关联关系;生成模块,被配置成根据关联关系,生成知识图谱。

[0071] 在本实施例的一些可选的实现方式中,第一确定模块进一步被配置成:对特征信息进行分词处理,得到分词结果;根据分词结果,确定问题信息的标签。

[0072] 在本实施例的一些可选的实现方式中,抽取模块进一步被配置成:根据问题信息,获取排序在前预设个数的搜索结果;从各搜索结果对应的页面信息中,抽取出属于实体类别的实体。

[0073] 在本实施例的一些可选的实现方式中,第二确定模块进一步被配置成:对实体所在的页面信息进行聚合,根据页面信息的聚合结果确定出第一聚合度值;对包含标签的问题信息进行聚合,根据问题信息的聚合结果确定出第二聚合度值;根据第一聚合度值与第二聚合度值,确定标签与实体的关联关系。

[0074] 在本实施例的一些可选的实现方式中,上述装置还包括:过滤单元(图中未示出),被配置成从标签中过滤掉预先设置的非法标签表中的非法标签,从实体中过滤掉预先设置的非法实体表中的非法实体。

[0075] 本公开的上述实施例提供的装置,通过获取单元501获取查询问题信息,然后确定单元502根据查询问题信息所包括的用于描述目标关键词的目标特征信息,确定查询问题信息的目标标签,查找单元503从预设的知识图谱中查找出与目标标签关联的目标实体,最后输出单元504基于所确定的目标实体,生成查询问题信息的应答信息,实现了根据筛选查询问题信息的标签,快速准确地知识图谱中查找出符合特征信息的实体。

[0076] 下面参考图6,下面参考图6,其示出了适于用来实现本公开的实施例的电子设备(例如图1中的服务器)600的结构示意图。图6示出的服务器仅仅是一个示例,不应对本公开的实施的功能和使用范围带来任何限制。

[0077] 如图6所示,电子设备600可以包括处理装置(例如中央处理器、图形处理器等)601,其可以根据存储在只读存储器(ROM)602中的程序或者从存储装置608加载到随机访问存储器(RAM)603中的程序而执行各种适当的动作和处理。在RAM 603中,还存储有电子设备600操作所需的各种程序和数据。处理装置601、ROM、602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口606也连接至总线604。

[0078] 通常,以下装置可以连接至I/O接口605:包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置606;包括例如液晶显示器(LCD,Liquid Crystal Display)、扬声器、振动器等的输出装置607;包括例如磁带、硬盘等的存储装置608;以及通信装置609。通信装置609可以允许电子设备600与其他设备进行无线或有线通信以交换数据。虽然图6示出了具有各种装置的电子设备600,但是应理解的是,并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。图6中示出的每个方框可以代表一个装置,也可以根据需要代表多个装置。

[0079] 特别地,根据本公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质

上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信装置609从网络上被下载和安装,或者从存储装置608被安装,或者从ROM 602被安装。在该计算机程序被处理装置601执行时,执行本公开的实施例的方法中限定的上述功能。

[0080] 需要说明的是,本公开的实施例的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPR0M或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开的实施例中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开的实施例中,计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:电线、光缆、RF(Radio Frequency,射频)等等,或者上述的任意合适的组合。

[0081] 上述计算机可读介质可以是上述电子设备中所包含的;也可以是单独存在,而未装配入该电子设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被该电子设备执行时,使得该电子设备:获取查询问题信息,其中,查询问题信息包括目标关键词以及用于描述目标关键词的目标特征信息;根据目标特征信息,确定查询问题信息的目标标签;从预设的知识图谱中查找出与目标标签关联的目标实体;基于所确定的目标实体,生成查询问题信息的应答信息。

[0082] 可以以一种或多种程序设计语言或其组合来编写用于执行本公开的实施例的操作的计算机程序代码,程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0083] 附图中的流程图和框图,图示了按照本公开的各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要

注意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0084] 描述于本公开的实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中,例如,可以描述为:一种处理器,包括获取单元、确定单元、查找单元以及输出单元。其中,这些单元的名称在某种情况下并不构成对该单元本身的限定,例如,获取单元还可以被描述为“获取查询问题信息的单元”。

[0085] 上描述仅为本公开的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本公开的实施例中所涉及的发明范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开的实施例中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

100

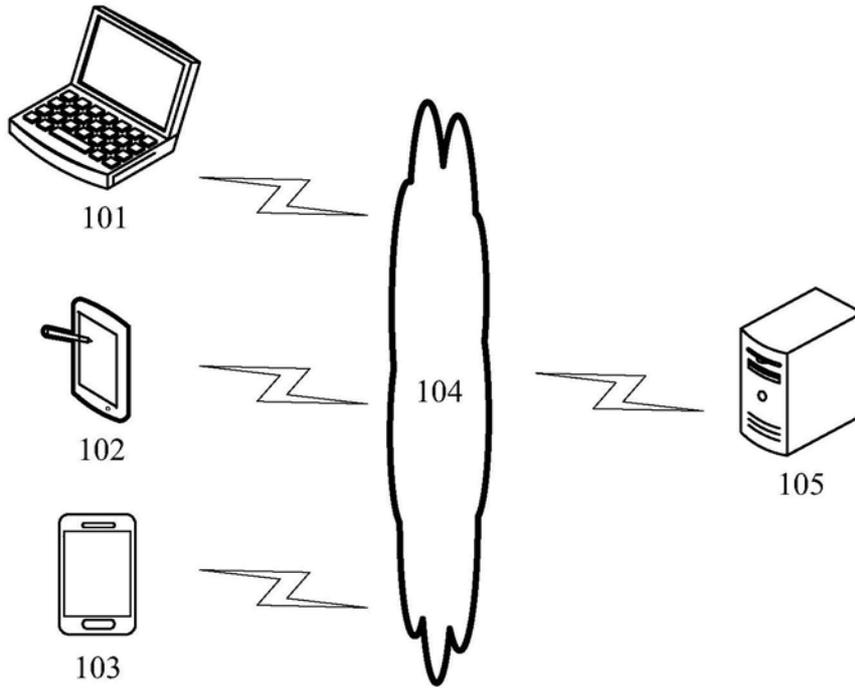


图1

200

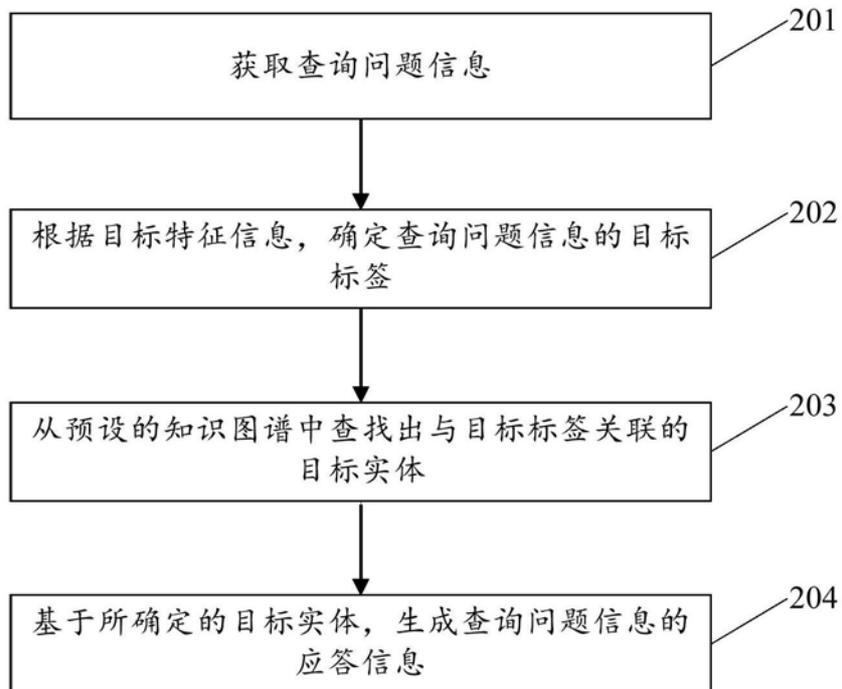


图2

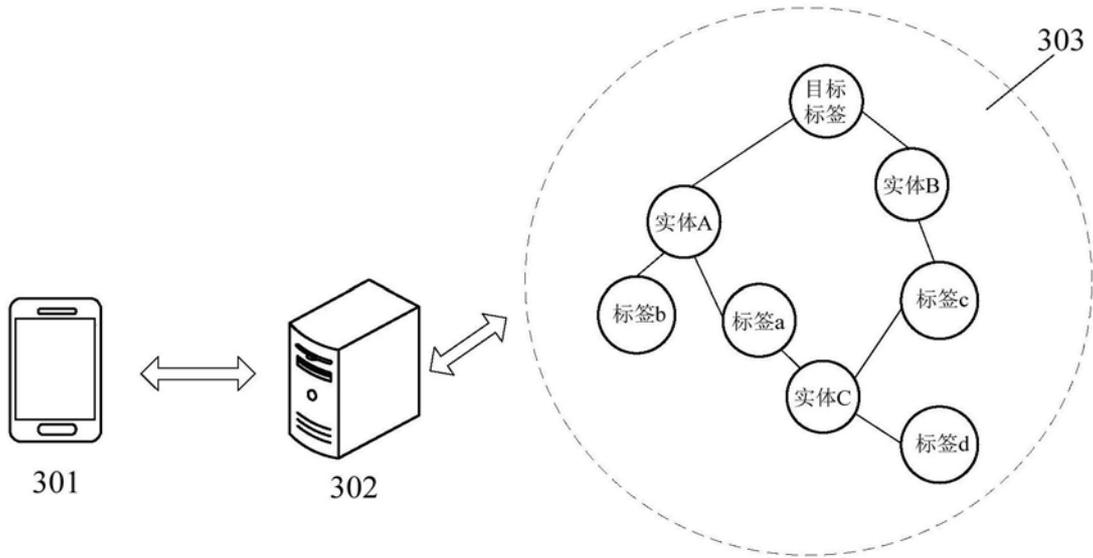


图3

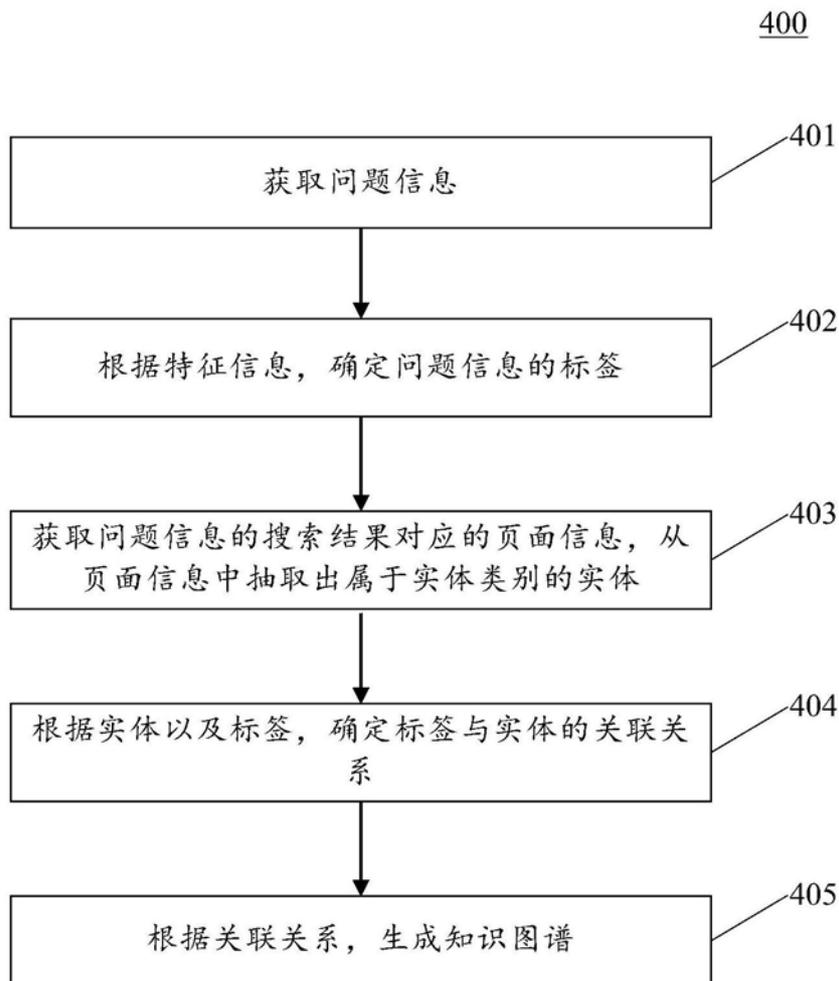


图4

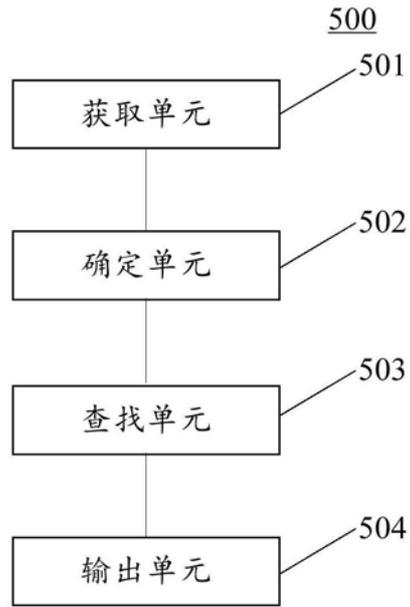


图5

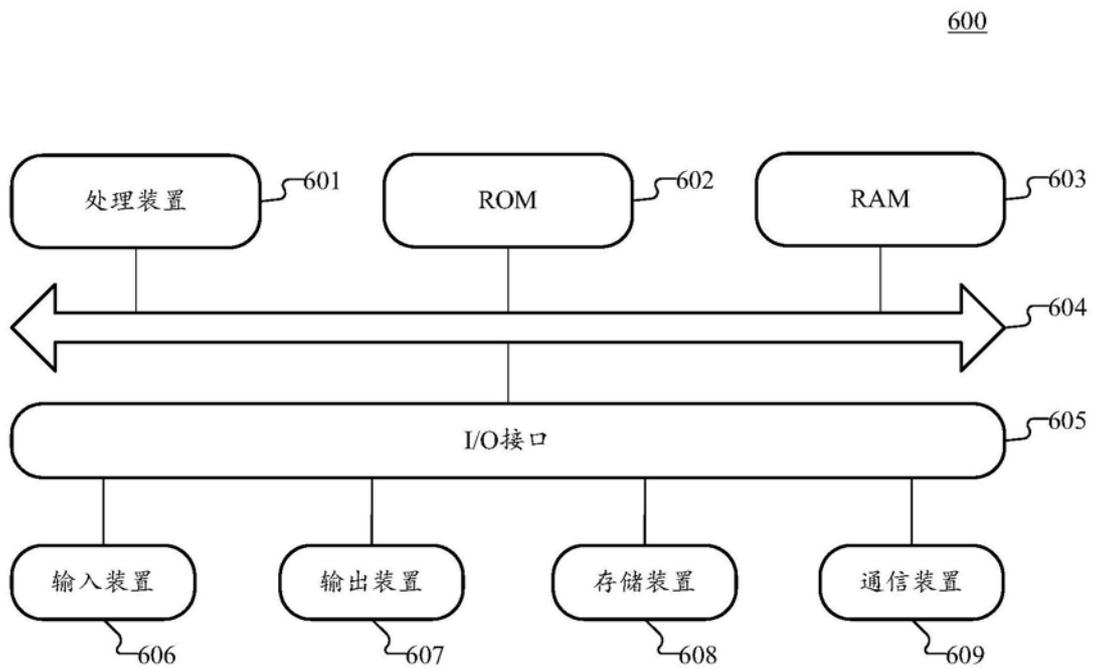


图6