(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115795056 B
(45)授权公告日 2024.08.02

(21)申请号 202310007617.5
(22)申请日 2023.01.04
(65)同一申请的已公布的文献号
申请公布号 CN 115795056 A
(43)申请公布日 2023.03.14
(73)专利权人 中国电子科技集团公司第十五研究所
地址 100083 北京市海淀区北四环中路211号
(72)发明人 嵇晨 张家伟 刘玉龙 于博
(74)专利代理机构 北京惟专知识产权代理事务所(普通合伙) 16074
专利代理师 赵星 霍东岳
(51)Int.Cl.
G06F 16/36(2019.01)
G06F 40/295(2020.01)
审查员 王彩勤

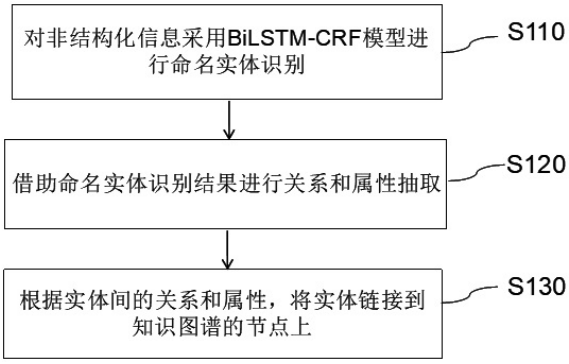
权利要求书2页 说明书7页 附图1页

(54)发明名称

非结构化信息构建知识图谱的方法、服务器及存储介质

(57)摘要

本申请公开了一种非结构化信息构建知识图谱的方法、服务器及存储介质,属于信息处理领域,包括如下步骤:步骤1:对非结构化信息采用BiLSTM-CRF模型进行命名实体识别;步骤2:借助命名实体识别结果进行关系和属性抽取;步骤3:根据实体间的关系和属性,将实体链接到知识图谱的节点上。本发明带来的有益效果是:通过应用BiLSTM-CRF、BiLSTM-Attention、DeepWalk等算法,将命名实体识别、关系抽取、实体链接组合成了一整套系统规范的流程,使得非结构化数据高效、准确地流入知识图谱中;固化了从命名实体识别到关系/属性提取再到实体链接的一整套提取非结构化信息到知识图谱的流程。



CN 115795056 B

1.一种非结构化信息构建知识图谱的方法,其特征在于,包括如下步骤:

步骤1:对非结构化信息采用BiLSTM-CRF模型进行命名实体识别;具体为,

对不同的实体类型定义标签;

根据所述标签对非结构化信息中的实体进行标注;

完成标注后,通过BiLSTM-CRF模型进行训练,完成命名实体识别;

BiLSTM-CRF模型的结构包括输入层、Embedding层、BiLSTM层、CRF层以及输出层,其中:

输入层:每个Xi对应句子中的一个字,每个字进行独热编码;

Embedding层:对字向量embedding化,使用Bert模型或Ernie模型;

BiLSTM层:使用双向的LSTM对embedding化的字向量进行编码,输出每个字的标签预测向量,即维数为标签数;

CRF层:根据BiLSTM层的输出序列化计算所有路径的得分,输出最有可能的路径;

输出层:输出每个字对应的标签,得到最终结果;

步骤2:借助命名实体识别结果进行关系和属性抽取,具体步骤如下:

使用BiLSTM模型进行上下文信息的学习;

用Attention机制对每个位置上的输出进行权重的学习;

将输出的结果进行归一化,得到对于关系的预测;

步骤3:根据实体间的关系和属性,将实体链接到知识图谱的节点上;首先进行候选实体的生成,将知识图谱中节点的单一名称扩充为指向该节点的名词集合,然后对待链接目标的上下文,利用步骤1中的BiLSTM-CRF模型获取上下文信息相关的节点名称,代入向量矩阵中获取上下文信息的低维向量,并通过全局投票评分继续增强待链接目标和图谱中正确节点间的一致性,其评分由待链接目标上下文与图谱中候选节点向量化之间的余弦相似度给出;

候选实体生成过程中实体消歧的具体方法为:给定起始节点,利用DeepWalk在其邻接节点中随机采样,获取一个节点作为下一个访问节点,随后循环此过程直到访问序列长度满足预先设定的值;在采样出足够数量的样本后,使用Skip-gram模型进行向量学习。

2.根据权利要求1所述的非结构化信息构建知识图谱的方法,其特征在于,BiLSTM-Attention模型的结构包括:

输入层:在输入的句子中直接对词进行独热编码,识别出的属于所需关系类型的两个实体单独作为词,其他部分利用结巴分词工具分好;

Embedding层:词向量embedding化,使用Bert模型或Ernie模型;

BiLSTM层:进行每个词的编码表示;

Attention层:使用Soft-Attention机制,在初始化时随机生成权值,使用上述权值的向量与BiLSTM层的输出进行匹配获取输出,再将每个词上的输出向量进行Softmax函数计算;

输出层:输出关系数维数的预测向量。

3.根据权利要求1所述的非结构化信息构建知识图谱的方法,其特征在于,候选实体的生成采用创建实体词典的方式,所述实体词典的来源至少包括百科、搜索引擎、网页爬取和人工标注,辅以文本的相似度计算增加匹配的概率,并进行人工校验后再加入实体词典中。

4.一种服务器,其特征在于,包括:存储器和至少一个处理器;

所述存储器存储计算机程序,所述至少一个处理器执行所述存储器存储的计算机程序,以实现权利要求1至3中任一项所述的非结构化信息构建知识图谱的方法。

5.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有计算机程序,所述计算机程序被执行时实现权利要求1至3中任一项所述的非结构化信息构建知识图谱的方法。

非结构化信息构建知识图谱的方法、服务器及存储介质

技术领域

[0001] 本申请属于信息处理领域,特别涉及一种非结构化信息构建知识图谱的方法、服务器及存储介质。

背景技术

[0002] 知识图谱(Knowledge Graph),在图书情报界称为知识域可视化或知识领域映射地图,是显示知识发展进程与结构关系的一系列各种不同的图形,用可视化技术描述知识资源及其载体,挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计量学引文分析、共现分析等方法结合,并利用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构达到多学科融合目的的现代理论。

[0003] 在知识图谱构建的过程中,会有许多的非结构化信息,而非结构化信息的形式相对不固定,常常是各种格式的文件。非结构化信息在构建知识图谱时,会出现信息提取难、关联图谱中实体难、关联信息放入图谱中难的问题,因此现有技术中通常会将非结构化信息进行命名实体识别、关系抽取、实体链接这三方面的处理,但是现有技术对于命名实体识别、关系抽取、实体链接这三方面的技术在知识图谱构建的领域内各自为战,没有串联成一个整体;关系抽取的成果难以直接提取到知识图谱中;实体链接的相关技术只是利用上下文进行链指,没有提取上下文中富余的相关信息。

[0004] 因此,需要一种针对非结构化信息构建知识图谱的方法,能够解决上述问题。

发明内容

[0005] 为了解决所述现有技术的不足,本申请提供了一种非结构化信息构建知识图谱的方法、服务器及存储介质,形成一整套提取非结构化信息到知识图谱的流程,包括实体识别、关系/属性抽取、实体链接三个环节,通过各个环节的技术手段以及各个环节间的紧密联系,将非结构化数据中蕴含的实体名称、实体属性、不同实体间的关系都提取出来放入知识图谱中。

[0006] 本申请所要达到的技术效果通过以下方案实现:

[0007] 根据本发明的第一方面,提供了一种非结构化信息构建知识图谱的方法,包括如下步骤:

[0008] 步骤1:对非结构化信息采用BiLSTM-CRF模型进行命名实体识别;

[0009] 步骤2:借助命名实体识别结果进行关系和属性抽取;

[0010] 步骤3:根据实体间的关系和属性,将实体链接到知识图谱的节点上。

[0011] 优选地,在步骤1中,具体步骤如下:

[0012] 对不同的实体类型定义标签;

[0013] 根据所述标签对非结构化信息中的实体进行标注;

[0014] 完成标注后,通过BiLSTM-CRF模型进行训练,完成命名实体识别。

[0015] 优选地,BiLSTM-CRF模型的结构包括输入层、Embedding层、BiLSTM层、CRF层以及输出层,其中:

[0016] 输入层:每个Xi对应句子中的一个字,每个字进行独热编码;

[0017] Embedding层:对字向量embedding化,使用Bert模型或Ernie模型;

[0018] BiLSTM层:使用双向的LSTM对embedding化的字向量进行编码,输出每个字的标签预测向量,即维数为标签数;

[0019] CRF层:根据BiLSTM层的输出序列化计算所有路径的得分,输出最有可能的路径;

[0020] 输出层:输出每个字对应的标签,得到最终结果。

[0021] 优选地,在步骤2中,使用BiLSTM-Attention模型进行关系和属性抽取,具体步骤如下:

[0022] 使用BiLSTM模型进行上下文信息的学习;

[0023] 用Attention机制对每个位置上的输出进行权重的学习;

[0024] 将输出的结果进行归一化,得到对于关系的预测。

[0025] 优选地,BiLSTM-Attention模型的结构包括:

[0026] 输入层:在输入的句子中直接对词进行独热编码,识别出的属于所需关系类型的两个实体单独作为词,其他部分利用结巴分词工具分好;

[0027] Embedding层:词向量embedding化,使用Bert模型或Ernie模型;

[0028] BiLSTM层:进行每个词的编码表示;

[0029] Attention层:使用Soft-Attention机制,在初始化时随机生成权值,使用该权值的向量与BiLSTM层的输出进行匹配获取输出,再将每个词上的输出向量进行Softmax函数计算;

[0030] 输出层:输出关系数维数的预测向量。

[0031] 优选地,在步骤3中,实体链接到知识图谱的节点上具体为:首先进行候选实体的生成,将知识图谱中节点的单一名称扩充为指向该节点的名词集合,然后对待链接目标的上下文,利用命名实体识别中的BiLSTM-CRF模型获取上下文信息相关的节点名称,代入向量矩阵中获取上下文信息的低维向量,并通过全局投票评分继续增强待链接目标和图谱中正确节点间的一致性,其评分由待链接目标上下文与图谱中候选节点向量化之间的余弦相似度给出。

[0032] 优选地,候选实体的生成采用创建实体词典的方式,所述实体词典的来源至少包括百科、搜索引擎、网页爬取和人工标注,辅以文本的相似度计算增加匹配的概率,并进行人工校验后再加入实体词典中。

[0033] 优选地,候选实体生成过程中实体消歧的具体方法为:给定起始节点,利用DeepWalk在其邻接节点中随机采样,获取一个节点作为下一个访问节点,随后循环此过程直到访问序列长度满足预先设定的值;在采样出足够数量的样本后,使用Skip-gram模型进行向量学习。

[0034] 根据本发明的第二方面,提供了一种服务器,包括:存储器和至少一个处理器;

[0035] 所述存储器存储计算机程序,所述至少一个处理器执行所述存储器存储的计算机程序,以实现上述任一项所述的非结构化信息构建知识图谱的方法。

[0036] 根据本发明的第三方面,提供了一种计算机可读存储介质,所述计算机可读存储

介质中存储有计算机程序,所述计算机程序被执行时实现上述任一项所述的非结构化信息构建知识图谱的方法。

[0037] 根据本发明的实施例,本发明带来的有益效果是:通过应用BiLSTM-CRF、BiLSTM-Attention、DeepWalk等算法,将命名实体识别、关系抽取、实体链接组合成了一整套系统规范的流程,使得非结构化数据高效、准确地流入知识图谱中;固化了从命名实体识别到关系/属性提取再到实体链接的一整套提取非结构化信息到知识图谱的流程;

[0038] 通过先进行命名实体识别对实体进行标注的方式,定位非结构化数据中的实体项,从而使得关系抽取的结果能够直接对应上其主体和客体;

[0039] 在关系抽取这一环节同步进行属性抽取,使得非结构化数据中的对象链接到图谱实体中后,可以对实体的相关属性和关联关系进行进一步完善。

附图说明

[0040] 为了更清楚地说明本申请实施例或现有的技术方案,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申请中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0041] 图1为本申请一实施例中一种非结构化信息构建知识图谱方法的流程图;

[0042] 图2为本申请一实施例中一种服务器的结构示意图。

具体实施方式

[0043] 为使本申请的目的、技术方案和优点更加清楚,下面将结合具体实施例及相应的附图对本申请的技术方案进行清楚、完整地描述。显然,所描述的实施例仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,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,都属于本申请保护的范围。

[0044] 如图1所示,本申请一实施例中的非结构化信息构建知识图谱的方法,包括如下步骤:

[0045] S110:对非结构化信息采用BiLSTM-CRF模型进行命名实体识别;

[0046] 在该步骤具体方法如下:

[0047] S111:对不同的实体类型定义标签;

[0048] S112:根据标签对非结构化信息中的实体进行标注;

[0049] S113:完成标注后,通过BiLSTM-CRF模型进行训练,完成命名实体识别。

[0050] BiLSTM能够学习到上下文的信息,根据整条句子判断该位置的词性;在BiLSTM上再加一层CRF,则可以从训练数据中学到更多约束,从而确保最终预测标签序列的有效性。

[0051] 在本申请一实施例中,BiLSTM-CRF模型的结构包括输入层、Embedding层、BiLSTM层、CRF层以及输出层,其中:

[0052] 输入层:每个Xi对应句子中的一个字,每个字进行独热编码;

[0053] Embedding层:对字向量embedding化,使用Bert模型或Ernie模型;

[0054] BiLSTM层:使用双向的LSTM对embedding化的字向量进行编码,输出每个字的标签预测向量,即维数为标签数;

[0055] CRF层:BiLSTM层输出的是每个字对应各个标签的概率向量,直接将其作为结果,可能会产生‘B-Person’标签对应的字的下一个字对应的标签是‘I-Loc’这样的无效输出序列,因此加入了条件随机场这一层;根据BiLSTM层的输出序列化计算所有路径的得分,输出最有可能的路径;

[0056] 输出层:输出每个字对应的标签,得到最终结果。

[0057] 在该步骤中,并没有使用分词工具预先分词后以词向量的形式作为输入,因为考虑到错误的分词会对结果产生影响,故而使用字向量将分词也作为模型学习的一部分。

[0058] 命名实体识别可以剔除掉一些没有信息的句子,但是光是命名实体识别并不能对知识图谱的构建起到太多作用,后续还需要从句子中进行关系抽取。

[0059] 在本申请一个具体的例子中,以非公经济领域知识图谱为例,根据实体类型定义的标签如表1:

[0060] 表1根据实体类型定义标签

[0061]

标签名称	标签含义(B为词中首字,I为剩下的字)
O	无用字
B-Person、I-Person	人员标签
B-Company、I-Company	企业标签
B-GSL、I-GSL	工商联标签
B-Court、I-Court	法院标签
B-Club、I-Club	商会标签
B-Loc、I-Loc	地址标签

[0062] 依据这些标签对句子进行标注,例如“马某作为某公司股东代表出席了位于北京的全国工商联举办的民营企业家发展大会”即{‘马’,‘某’,‘作’,‘为’,‘某’,‘公’,‘司’,‘股’,‘东’,‘代’,‘表’,‘出’,‘席’,‘了’,‘位’,‘于’,‘北’,‘京’,‘的’,‘全’,‘国’,‘工’,‘商’,‘联’,‘举’,‘办’,‘的’,‘民’,‘营’,‘企’,‘业’,‘家’,‘发’,‘展’,‘大’,‘会’}应该被标注为{‘B-Person’,‘I-Person’,‘O’,‘O’,‘B-Company’,‘I-Company’,‘I-Company’,‘I-Company’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘B-Loc’,‘I-Loc’,‘O’,‘B-GSL’,‘I-GSL’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’,‘O’}。

[0063] 完成标注后,通过BiLSTM-CRF的模型结构进行训练。

[0064] S120:借助命名实体识别结果进行关系和属性抽取;

[0065] 在该步骤中,使用BiLSTM-Attention模型进行关系和属性抽取,具体步骤如下:

[0066] S121:使用BiLSTM模型进行上下文信息的学习;

[0067] S122:用Attention机制对每个位置上的输出进行权重的学习;

[0068] S123:将输出的结果进行归一化,得到对于关系的预测。

[0069] 其中,BiLSTM-Attention模型的结构包括:

[0070] 输入层:在输入的句子中直接对词进行独热编码,识别出的属于该关系类型的两个实体单独作为词,其他部分利用结巴分词工具分好;

[0071] Embedding层:词向量embedding化,使用Bert模型或Ernie模型;

[0072] BiLSTM层:进行每个词的编码表示;

[0073] Attention层:使用Soft-Attention机制,在初始化时随机生成权值,使用该权值向量与BiLSTM层的输出进行匹配获取输出,再将每个词上的输出向量进行Softmax函数计算;

[0074] 输出层:一层分类器,加入了L2正则化项,损失函数为交叉熵,输出关系数维数的预测向量。

[0075] 具体的,非结构化数据的关系抽取可以看作是SPO三元组(主谓宾三元组)的抽取任务,即三元组【S,P,O】,S和O分别为主客实体,P为主实体指向客实体的关系。例如当输入文本是“张某入股了位于北京朝阳区的某团公司,该公司的董事长是王某”,那么基于字面意思,至少能够提取出以下三个SPO三元组:【张某,控股,某团】,【某团,位于,北京朝阳区】,【某团,董事长,王某】。

[0076] 通过对非结构化文本的观察,可以发现很多句子中都能提取出不止一条关系,句子提取出的关系主要有以下几种情况,如表2:

[0077] 表2:句子中S、P、O结构类型

[0078]

1	只有一条【S，P，O】	“王某是某团的董事长”，抽取出的关系应该为【王某，董事长，某团】
2	一个S对应多个【P，O】	“王某既是某团的董事长，又入股了某点”，抽取出的关系应该为【王某，董事长，某团】和【王某，控股，某点】
3	多个S对应一个【P，O】	“王某和张某都入股了某团”，抽取出的关系应该为【王某，控股，某团】和【张某，控股，某团】
4	一个【S，O】对应多个P	“王某是某团的董事长兼法人”，抽取出的关系应该为【王某，董事长，某团】和【王某，法人，某团】
5	一个【S，P】对应多个O	“王某入股了某团和某点”，抽取出的关系应该为【王某，控股，某团】和【王某，控股，某点】
6	多个【S，P】对应一个O	“王某和张某分别是某团的董事长和法人”，抽取出的关系应该为【王某，董事长，某团】和【张某，法人，某团】
7	多个【S，O】对应一个P	“王某和张某分别是某团和某点的董事长”，抽取出的关系应该为【王某，董事长，某团】和【张某，董事长，某点】
8	S，P，O都为多个	“张某入股了位于北京朝阳区的某团公司，该公司的董事长是王某”，抽取出的关系应该为【张某，控股，某团】，【某团，位于，北京朝阳区】，【某团，董事长，王某】

[0079] 在该表格中,由于关系是实体间的关系,可以将关系也表示为标签,再标注出足够的数据进行训练,就能够在识别实体的同时抽取出关系。

[0080] 关系对应的节点类型基本上是固定的,因此实际的关系三元组的类别只有68种左右,既然每种关系对应的实体类型是固定的,那么问题就变成了从句子中抽取确定类型的两个实体之间的关系,而关系的总数量是确定的。

[0081] 所以在命名实体识别的基础中,关系抽取可以被定义为输入为包含实体类型确定的句子,输出为预测关系的问题。为了减少标注的数量节省人工成本,对包含不同实体类型的句子分别进行训练,这同时也解决了单一模型下一条句子对应超过两个实体,从而含有多条关系的问题(只需要把这些句子分别放入对应实体类型的多个模型里)。

[0082] S130:根据实体间的关系和属性,将实体链接到知识图谱的节点上。

[0083] 在该步骤中,实体链接到知识图谱的节点上具体为:首先进行候选实体的生成,将知识图谱中节点的单一名称扩充为指向该节点的名词集合,然后对待链接目标的上下文,利用命名实体识别中的BiLSTM-CRF模型获取上下文信息相关的节点名称,代入向量矩阵中获取上下文信息的低维向量,并通过全局投票评分继续增强待链接目标和图谱中正确节点间的一致性,其评分由待链接目标上下文与图谱中候选节点向量化之间的余弦相似度给出。

[0084] 候选实体的生成采用创建实体词典的方式,实体词典的来源至少包括百科、搜索引擎、网页爬取和人工标注,辅以文本的相似度计算增加匹配的概率,并进行人工校验后再加入实体词典中。此外,人工还用于处理没匹配到或相似度过低的样例,并继续做词典的完善。

[0085] 实体消歧的具体方法为:给定起始节点,利用DeepWalk在其邻接节点中随机采样,获取一个节点作为下一个访问节点,随后循环此过程直到访问序列长度满足预先设定的值;在采样出足够数量的样本后,使用Skip-gram模型进行向量学习。

[0086] 随后,对于候选实体的上下文,利用命名实体识别中的Bi-LSTM+CRF模型获取上下文信息相关的节点名称,代入向量矩阵中获取上下文信息的低维向量,并通过全局投票评分继续增强链接实体和消歧实体间的一致性。其评分由候选实体与上下文信息的标准化平均值之间的余弦相似度给出。

[0087] 最后的输出是候选实体和文本相关信息的相似度,即对应候选实体为正确链接对象的概率。

[0088] 在上述步骤中,DeepWalk算法使用图中包含的节点、关系构成序列,拆出其中的一个节点作为Skip-gram模型的输出,剩下的作为输入,来学习节点的低维向量表示。获取序列样本的方法是随机游走(Random Walk),该策略作为一种深度优先遍历(Deep First Search,DFS)的方法,沿途访问并提取序列,并且可以重复访问已访问的节点。考虑到实体链接中候选实体与具体节点间的协同信息往往不止一跳,例如(马某)->[控股]->(某集团控股有限公司)->[组织]->(某公司20周年晚会)的具体链接和“马某身着个性服装亮相某公司20周年晚会现场并上台表演了节目”,采用该方法对节点进行学习表示能够更好地学习到节点周围的相关信息。

[0089] 该方法适用于非公经济领域的实体消歧,解决候选实体都从事非公经济领域的工作导致的可用信息大大减少的问题,能够保证较高的准确率。

[0090] 通过以上方法,将所在领域的不同类别的相关实体定义成不同种类的标签,进行标注后应用BiLSTM-CRF解决了命名实体识别问题;

[0091] 利用BiLSTM-Attention对标注好实体的文本进行关系和属性的提取;

[0092] 利用DeepWalk在图谱上进行采样,形成实体的向量表示,与利用上下文信息得到的向量进行相似度计算,将对象与实体进行链接;

[0093] 将以上三个环节串联起来形成了一个整体,固化了从命名实体识别到关系/属性提取再到实体链接的一整套提取非结构化信息到知识图谱的流程。

[0094] 根据本发明的第二方面,如图2所示,提供了一种服务器,包括:存储器201和至少一个处理器202;

[0095] 存储器201存储计算机程序,至少一个处理器202执行存储器201存储的计算机程序,以实现上述任一项的非结构化信息构建知识图谱的方法。

[0096] 根据本发明的第三方面,提供了一种计算机可读存储介质,计算机可读存储介质中存储有计算机程序,计算机程序被执行时实现上述任一项的非结构化信息构建知识图谱的方法。

[0097] 应该指出,上述详细说明都是示例性的,旨在对本申请提供进一步的说明。除非另有指明,本文使用的所有技术和科学术语均具有与本申请所属技术领域的普通技术人员的通常理解所相同的含义。

[0098] 需要注意的是,这里所使用的术语仅是为了描述具体实施方式,而非意图限制根据本申请的示例性实施方式。如在这里所使用的,除非上下文另外明确指出,否则单数形式也意图包括复数形式。此外,还应当理解的是,当在本说明书中使用术语“包含”和/或“包括”时,其指明存在特征、步骤、操作、器件、组件和/或它们的组合。

[0099] 需要说明的是,本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的术语在适当情况下可以互换,以便这里描述的本申请的实施方式能够以除了在这里图示或描述的那些以外的顺序实施。

[0100] 此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含。例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0101] 为了便于描述,在这里可以使用空间相对术语,如“在……之上”、“在……上方”、“在……上表面”、“上面的”等,用来描述如在图中所示的一个器件或特征与其他器件或特征的空间位置关系。应当理解的是,空间相对术语旨在包含除了器件在图中所描述的方位之外的在使用或操作中的不同方位。例如,如果附图中的器件被倒置,则描述为“在其他器件或构造上方”或“在其他器件或构造之上”的器件之后将被定位为“在其他器件或构造下方”或“在其他器件或构造之下”。因而,示例性术语“在……上方”可以包括“在……上方”和“在……下方”两种方位。该器件也可以其他不同方式定位,如旋转90度或处于其他方位,并且对这里所使用的空间相对描述作出相应解释。

[0102] 在上面详细的说明中,参考了附图,附图形成本文的一部分。在附图中,类似的符号典型地确定类似的部件,除非上下文以其他方式指明。在详细的说明书、附图及权利要求书中所描述的图示说明的实施方案不意味是限制性的。在不脱离本文所呈现的主题的精神或范围下,其他实施方案可以被使用,并且可以作其他改变。

[0103] 以上仅为本发明的优选实施例而已,并不用于限制本发明,对于本领域的技术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

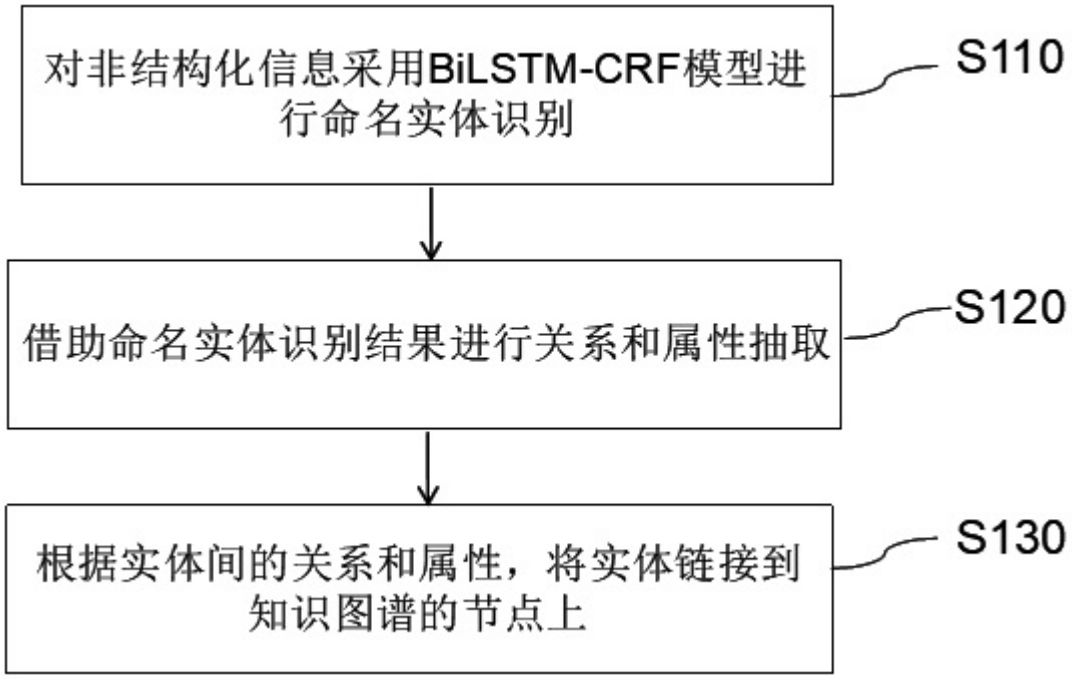

对非结构化信息采用BiLSTM-CRF模型进行命名实体识别
S110
借助命名实体识别结果进行关系和属性抽取
S120
根据实体间的关系和属性，将实体链接到知识图谱的节点上
S130

图1

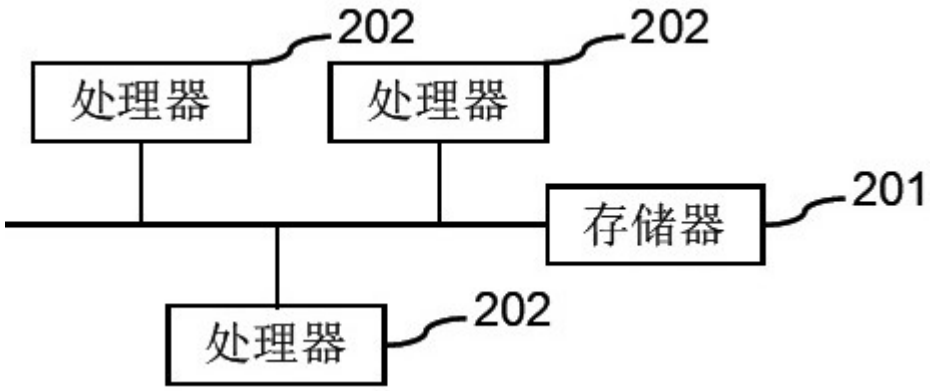

202
处理器
202
处理器
存储器
201
处理器
202

图2