



(12) 发明专利申请

(10) 申请公布号 CN 114881037 A

(43) 申请公布日 2022. 08. 09

(21) 申请号 202210542109.2

(22) 申请日 2022.05.17

(71) 申请人 中国平安财产保险股份有限公司
地址 518000 广东省深圳市福田区益田
路5033号平安金融中心12、13、38、39、
40、62层

(72) 发明人 陈桂华 刘超

(74) 专利代理机构 深圳市沃德知识产权代理事
务所(普通合伙) 44347
专利代理师 高杰 于志光

(51) Int. Cl.
G06F 40/295 (2020.01)
G06F 16/33 (2019.01)
G06F 16/335 (2019.01)

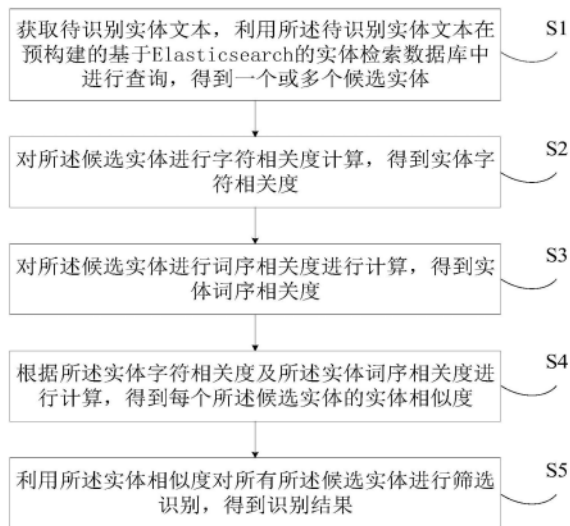
权利要求书2页 说明书9页 附图2页

(54) 发明名称

命名实体识别方法、装置、电子设备及存储
介质

(57) 摘要

本发明涉及人工智能技术,揭露一种命名实体识别方法,包括:获取待识别实体文本,利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询,得到一个或多个候选实体;对所述候选实体进行字符相关度计算,得到实体字符相关度;对所述候选实体进行词序相关度进行计算,得到实体词序相关度;根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果。本发明还涉及一种区块链技术,所述待识别实体文本可以存储在区块链节点中。本发明还提出一种命名实体识别装置、设备以及介质。本发明可以提高命名实体识别的效率。



1. 一种命名实体识别方法,其特征在于,所述方法包括:
 - 获取待识别实体文本,利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询,得到一个或多个候选实体;
 - 对所述候选实体进行字符相关度计算,得到实体字符相关度;
 - 对所述候选实体进行词序相关度进行计算,得到实体词序相关度;
 - 根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;
 - 利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果。
2. 如权利要求1所述的命名实体识别方法,其特征在于,所述对所述候选实体进行词序相关度进行计算,得到实体词序相关度,包括:
 - 计算所述候选实体的折损累计增益,得到候选实体词序相关度;
 - 对所述候选实体词序相关度进行归一化,得到所述实体词序相关度。
3. 如权利要求2所述的命名实体识别方法,其特征在于,所述计算所述候选实体的折损累计增益,得到候选实体词序相关度,包括:
 - 判断所述候选实体中从第一个字符开始每个字符与所述待识别文本中对应顺序的字符是否相同;
 - 根据字符是否相同的判断结果确定所述候选实体中每个字符对应的词序相关系数;
 - 利用所述候选实体中的所有字符的词序相关系数进行计算,得到所述候选实体词序相关度。
4. 如权利要求3所述的命名实体识别方法,其特征在于,所述根据字符是否相同的判断结果确定所述候选实体中每个字符对应的词序相关系数,包括:
 - 当所述候选实体中的字符与所述待识别文本中对应顺序的字符相同,则所述候选实体中该字符对应的词语相关系数为第一预设值;
 - 当所述候选实体中的字符与所述待识别文本中对应顺序的字符不相同,则所述候选实体中该字符对应的词语相关系数为第二预设值,其中所述第一预设值大于第二预设值。
5. 如权利要求4所述的命名实体识别方法,其特征在于,所述对所述候选实体词序相关度进行归一化,得到所述实体词序相关度,包括:
 - 计算所述待识别实体文本的折损累计增益,得到待识别实体词序相关度;
 - 将所述第一预设值确定为所述待识别实体文本中每个字符的词序相关系数;
 - 利用所述待识别实体文本中的所有字符的词序相关系数进行计算,得到待识别实体词序相关度;
 - 根据所述待识别实体词序相关度对所述候选实体词序相关度进行归一化,得到所述实体词序相关度。
6. 如权利要求1所述的命名实体识别方法,其特征在于,所述利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果,包括:
 - 选择所有所述实体相似度中的最大值,得到目标实体相似度;
 - 将所述目标实体相似度对应的所述候选实体确定为所述识别结果。
7. 如权利要求1至6中任意一项所述的命名实体识别方法,其特征在于,所述获取待识别实体文本,包括:

获取问题文本；

对所述待识别文本进行分词，得到多个分词文本；

过滤在预设的非实体文本字典中存在的分词文本，得到所述待识别实体文本。

8. 一种命名实体识别装置，其特征在于，包括：

候选实体查询模块，用于获取待识别实体文本，利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询，得到一个或多个候选实体；

相似度计算模块，用于对所述候选实体进行字符相关度计算，得到实体字符相关度；对所述候选实体进行词序相关度进行计算，得到实体词序相关度；根据所述实体字符相关度及所述实体词序相关度进行计算，得到每个所述候选实体的实体相似度；

实体识别模块，用于利用所述实体相似度对所有所述候选实体进行筛选识别，得到识别结果。

9. 一种电子设备，其特征在于，所述电子设备包括：

至少一个处理器；以及，

与所述至少一个处理器通信连接的存储器；

其中，所述存储器存储有可被所述至少一个处理器执行的计算机程序，所述计算机程序被所述至少一个处理器执行，以使所述至少一个处理器能够执行如权利要求1至7中任一项所述的命名实体识别方法。

10. 一种计算机可读存储介质，存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求1至7中任一项所述的命名实体识别方法。

命名实体识别方法、装置、电子设备及存储介质

技术领域

[0001] 本发明涉及人工智能技术,尤其涉及一种命名实体识别方法、装置、电子设备及存储介质。

背景技术

[0002] 命名实体识别(Named Entity Recognition,NER)是NLP中一项非常基础的任务。NER是信息提取、问答系统、句法分析、机器翻译等众多NLP任务的重要基础工具。

[0003] 但是目前的命名实体识别方法基于规则及字典库进行识别,只能识别已经设置过关联规则的实体,对字典中未设置关联规则的实体无法识别,命名实体识别的效率较低。

发明内容

[0004] 本发明提供一种命名实体识别方法、装置、电子设备及存储介质,其主要目的在于提高了命名实体识别的效率。

[0005] 获取待识别实体文本,利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询,得到一个或多个候选实体;

[0006] 对所述候选实体进行字符相关度计算,得到实体字符相关度;

[0007] 对所述候选实体进行词序相关度进行计算,得到实体词序相关度;

[0008] 根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;

[0009] 利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果。

[0010] 可选地,所述对所述候选实体进行词序相关度进行计算,得到实体词序相关度,包括:

[0011] 计算所述候选实体的折损累计增益,得到候选实体词序相关度;

[0012] 对所述候选实体词序相关度进行归一化,得到所述实体词序相关度。

[0013] 可选地,所述计算所述候选实体的折损累计增益,得到候选实体词序相关度,包括:

[0014] 判断所述候选实体中从第一个字符开始每个字符与所述待识别文本中对应顺序的字符是否相同;

[0015] 根据字符是否相同的判断结果确定所述候选实体中每个字符对应的词序相关系数;

[0016] 利用所述候选实体中的所有字符的词序相关系数进行计算,得到所述候选实体词序相关度。

[0017] 可选地,所述根据字符是否相同的判断结果确定所述候选实体中每个字符对应的词序相关系数,包括:

[0018] 当所述候选实体中的字符与所述待识别文本中对应顺序的字符相同,则所述候选实体中该字符对应的词语相关系数为第一预设值;

[0019] 当所述候选实体中的字符与所述待识别文本中对应顺序的字符不相同,则所述候选实体中该字符对应的词语相关系数为第二预设值,其中所述第一预设值大于第二预设值。

[0020] 可选地,所述对所述候选实体词序相关度进行归一化,得到所述实体词序相关度,包括:

[0021] 计算所述待识别实体文本的折损累计增益,得到待识别实体词序相关度;

[0022] 将所述第一预设值确定为所述待识别实体文本中每个字符的词序相关系数;

[0023] 利用所述待识别实体文本中的所有字符的词序相关系数进行计算,得到待识别实体词序相关度;

[0024] 根据所述待识别实体词序相关度对所述候选实体词序相关度进行归一化,得到所述实体词序相关度。

[0025] 可选地,所述利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果,包括:

[0026] 选择所有所述实体相似度中的最大值,得到目标实体相似度;

[0027] 将所述目标实体相似度对应的所述候选实体确定为所述识别结果。

[0028] 可选地,所述获取待识别实体文本,包括:

[0029] 获取问题文本;

[0030] 对所述待识别文本进行分词,得到多个分词文本;

[0031] 过滤在预设的非实体文本字典中存在的分词文本,得到所述待识别实体文本。

[0032] 为了解决上述问题,本发明还提供一种命名实体识别装置,所述装置包括:

[0033] 候选实体查询模块,用于获取待识别实体文本,利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询,得到一个或多个候选实体;

[0034] 相似度计算模块,用于对所述候选实体进行字符相关度计算,得到实体字符相关度;对所述候选实体进行词序相关度进行计算,得到实体词序相关度;根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;

[0035] 实体识别模块,用于利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果。

[0036] 为了解决上述问题,本发明还提供一种电子设备,所述电子设备包括:

[0037] 存储器,存储至少一个计算机程序;及

[0038] 处理器,执行所述存储器中存储的计算机程序以实现上述所述的命名实体识别方法。

[0039] 为了解决上述问题,本发明还提供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至少一个计算机程序,所述至少一个计算机程序被电子设备中的处理器执行以实现上述所述的命名实体识别方法。

[0040] 本发明实施例对所述候选实体进行字符相关度计算,得到实体字符相关度;对所述候选实体进行词序相关度进行计算,得到实体词序相关度;根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;通过基于Elasticsearch的实体检索数据库查询候选实体,再通过字符相关度与词序相关度对候选实体进行筛选识别,不受数据库中实体关联规则的影响,实体识别的效率更高;因此本发明

实施例提出的命名实体识别方法、装置、电子设备及可读存储介质提高了命名实体识别的效率。

附图说明

[0041] 图1为本发明一实施例提供的命名实体识别方法的流程示意图；

[0042] 图2为本发明一实施例提供的命名实体识别装置的模块示意图；

[0043] 图3为本发明一实施例提供的实现命名实体识别方法的电子设备的内部结构示意图；

[0044] 本发明目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

[0045] 应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

[0046] 本发明实施例提供一种命名实体识别方法。所述命名实体识别方法的执行主体包括但不限于服务端、终端等能够被配置为执行本申请实施例提供的该方法的电子设备中的至少一种。换言之，所述命名实体识别方法可以由安装在终端设备或服务端设备的软件或硬件来执行，所述软件可以是区块链平台。所述服务端包括但不限于：单台服务器、服务器集群、云端服务器或云端服务器集群等，服务器可以是独立的服务器，也可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络(Content Delivery Network, CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0047] 参照图1所示的本发明一实施例提供的命名实体识别方法的流程示意图，在本发明实施例中，所述命名实体识别方法包括：

[0048] S1、获取待识别实体文本，利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询，得到一个或多个候选实体。

[0049] 详细地，本发明实施中所述待识别实体文本为需要进行实体识别的文本。

[0050] 可选地，本发明实施例中获取待识别实体文本，所述方法包括：

[0051] 步骤I：获取问题文本；

[0052] 可选地，本发明实施例中所述问题文本可以为用户输入的问题文本，如：“小米公司是做什么的”，为了更精确的匹配用户输入的问题文本中的答案，需要对待识别文本中的实体进行识别，从而使得后续的问题文本匹配的答案更准确。如：“小米公司”识别对应的实体为“小米科技有限责任公司”，那么“小米公司是做什么的”，就可以完善为“小米科技有限责任公司是做什么的”，问题文本表述更加准确，从而使得对应的答案匹配的也更加准确。

[0053] 步骤II：对所述待识别文本进行分词，得到多个分词文本；

[0054] 可选地，由于所述待识别文本中并不是所有词语都需要实体识别，所以需要对待识别文本进行分词，从而过滤掉所述待识别文本中不需要实体识别的普通词语。

[0055] 步骤III：过滤在预设的非实体文本字典中存在的分词文本，得到待识别实体文本。

[0056] 具体地，本发明实施例中所述非实体文本字典为不需要进行实体识别的非实体文本构建的字典数据库。

[0057] 例如:共有三个分词文本:分词文本A、分词文本B、分词文本C,分词文本C在所述非实体文本字体中,那么分词文本A与分词文本B都为待识别实体文本。

[0058] 进一步地,本发明实施例中所述实体检索数据库为基于elastic search引擎构建的Elasticsearch数据库,相较于普通的数据库检索速度更快,同时检索的结果不需要必须包含被检索字段,检索的结果更加全面。

[0059] 可选地,本发明实施例中所述实体检索数据库中定义实体索引结构可以根据需求定义,通常至少有三项entityName (实体标准名)、entityType (实体分类)及entityAlias (实体别名)。

[0060] 本发明另一实施例中所述待识别实体可以存储在区块链节点中,利用区块链节点高吞吐的特性,提高数据的取用效率。

[0061] S2、对所述候选实体进行字符相关度计算,得到实体字符相关度。

[0062] 可选地,本发明实施例所述字体字符相关度利用bm25算法计算所述候选实体在所述实体检索数据库中的评分。

[0063] 本发明另一实施例中,所述实体字符相关度还可以为所述候选实体在所述实体检索数据库中的TF-IDF值。

[0064] S3、对所述候选实体进行词序相关度进行计算,得到实体词序相关度;

[0065] 可选地,本发明实施例中对所述候选实体进行词序相关度进行计算,得到实体词序相关度,包括:

[0066] 步骤A:计算所述候选实体的折损累计增益,得到候选实体词序相关度;

[0067] 可选地,本发明实施例中所述字符相似度只是衡量了候选实体与所述待识别实体文本的字符相关性,并未考虑本发明实施例通过候选实体与所述待识别实体文本的词序相关度,因此,本发明实施例计算所述候选实体的折损累计增益作为词序相关度的评价标准。

[0068] 详细地,本发明实施例中计算所述候选实体的折损累计增益,得到候选实体词序相关度,包括:

[0069] 判断所述候选实体中从第一个字符开始每个字符与所述待识别文本中对应顺序的字符是否相同;

[0070] 根据字符是否相同的判断结果确定所述候选实体中每个字符对应的词序相关系数;

[0071] 利用所述候选实体中的所有字符的词序相关系数进行计算,得到所述候选实体词序相关度。

[0072] 具体地,本发明实施例中根据字符是否相同的判断结果确定所述候选实体中每个字符对应的词序相关系数,包括:

[0073] 当所述候选实体中的字符与所述待识别文本中对应顺序的字符相同,则所述候选实体中该字符对应的词语相关系数为第一预设值;

[0074] 当所述候选实体中的字符与所述待识别文本中对应顺序的字符不相同,则所述候选实体中该字符对应的词语相关系数为第二预设值,其中所述第一预设值大于第二预设值。

[0075] 可选地,当所述候选实体中的字符与所述待识别文本中对应顺序的字符相同,则所述候选实体中该字符对应的词语相关系数为1,当所述候选实体中的字符与所述待识别

文本中对应顺序的字符不相同,则所述候选实体中该字符对应的词语相关系数为第二预设值为0。

[0076] 例如:所述第一预设值为1,第二预设值为0,所述待识别实体文本为“甲乙”,所述候选实体为“丙乙”,那么所述候选实体中字符“丙”对应的词序相关系数为0,所述候选实体中字符“乙”对应的词序相关系数为1。

[0077] 本发明另一实施例中计算所述候选实体的折损累计增益,得到候选实体词序相关度,包括:

[0078] 判断所述候选实体中从倒数第一个字符开始每个字符与所述待识别实体文本中对应顺序的字符是否相同;

[0079] 根据字符是否相同的判断结果确定所述候选实体中每个字符对应的词序相关系数;

[0080] 利用所述候选实体中的所有字符的词序相关系数进行计算,得到所述候选实体词序相关度。

[0081] 本发明另一实施例中计算所述候选实体的折损累计增益,得到候选实体词序相关度,包括:

[0082] 判断所述候选实体中每个字符在所述待识别实体文本中是否存在;

[0083] 根据字符是否存在的判断结果确定所述候选实体中每个字符对应的词序相关系数;

[0084] 利用所述候选实体中的所有字符的词序相关系数进行计算,得到所述候选实体词序相关度。

[0085] 具体地,本发明实施例中根据字符是否存在的判断结果确定所述候选实体中每个字符对应的词序相关系数,包括:

[0086] 当所述候选实体中每个字符在所述待识别实体文本中存在,则所述候选实体中该字符对应的词语相关系数为第一预设值;

[0087] 当所述候选实体中每个字符在所述待识别实体文本中不存在,则所述候选实体中该字符对应的词语相关系数为第二预设值,其中所述第一预设值大于第二预设值。

[0088] 可选地,本发明实施例中利用如下公式计算所述候选实体的候选实体词序相关度:

$$[0089] \quad DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

[0090] 其中,i为所述候选实体中字符的顺序,rel_i为所述候选实体中顺序为i的字符的词序相关系数,p为所述候选实体中字符的个数,DCG_p为所述候选实体的候选实体词序相关度。

[0091] 步骤B:对所述候选实体词序相关度进行归一化,得到所述实体词序相关度。

[0092] 可选地,本发明实施例中为了降低数据计算资源,对所述候选实体词序相关度进行归一化,得到所述实体词序相关度。

[0093] 具体地,本发明实施例中对所述候选实体词序相关度进行归一化,得到所述实体词序相关度,包括:

[0094] 计算所述待识别实体文本的折损累计增益,得到待识别实体词序相关度;

[0095] 将所述第一预设值确定为所述待识别实体文本中每个字符的词序相关系数;

[0096] 利用所述待识别实体文本中的所有字符的词序相关系数进行计算,得到所述待识别实体词序相关度。

[0097] 可选地,本发明实施例利用如下公式计算所述待识别实体词序相关度:

$$[0098] \quad IDCG_n = \sum_{j=1}^n \frac{2^{rel} - 1}{\log_2(j + 1)}$$

[0099] 其中,j为所述待识别实体文本中字符的顺序,rel为所述待识别实体文本中任意字符的词序相关系数,n为所述待识别实体文本中字符的个数,IDCG_n为所述候选实体词序相关度。

[0100] 根据所述待识别实体词序相关度对所述候选实体词序相关度进行归一化,得到所述实体词序相关度。

[0101] 可选地,本发明实施例利用如下公式进行归一化:

$$[0102] \quad X = \frac{DCG_p}{IDCG_n}$$

[0103] 其中X为所述所述实体词序相关度。

[0104] S4、根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;

[0105] 可选地,本发明实施例将所述实体字符相关度及所述实体词序相关度进行乘法计算,得到所述实体相关度。

[0106] 例如:候选实体A对应的实体字符相关度为0.8,候选实体A对应的实体词序相关度为0.5,那么候选实体A对应的实体相关度为0.8*0.5=0.4。

[0107] S5、利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果。

[0108] 详细地,本发明实施例中利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果,包括:

[0109] 选择所有所述实体相似度中的最大值,得到目标实体相似度;

[0110] 将所述目标实体相似度对应的所述候选实体确定为所述识别结果。

[0111] 如图2所示,是本发明命名实体识别装置的功能模块图。

[0112] 本发明所述命名实体识别装置100可以安装于电子设备中。根据实现的功能,所述命名实体识别装置可以包括候选实体查询模块101、相似度计算模块102、实体识别模块103,本发所述模块也可以称之为单元,是指一种能够被电子设备处理器所执行,并且能够完成固定功能的一系列计算机程序段,其存储在电子设备的存储器中。

[0113] 在本实施例中,关于各模块/单元的功能如下:

[0114] 所述候选实体查询模块101用于获取待识别实体文本,利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询,得到一个或多个候选实体;

[0115] 所述相似度计算模块102用于对所述候选实体进行字符相关度计算,得到实体字符相关度;对所述候选实体进行词序相关度进行计算,得到实体词序相关度;根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;

[0116] 所述实体识别模块103用于利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果。

[0117] 详细地,本发明实施例中所述命名实体识别装置100中所述的各模块在使用时采用与上述图1中所述的命名实体识别方法一样的技术手段,并能够产生相同的技术效果,这里不再赘述。

[0118] 如图3所示,是本发明实现命名实体识别方法的电子设备的结构示意图。

[0119] 所述电子设备可以包括处理器10、存储器11、通信总线12和通信接口13,还可以包括存储在所述存储器11中并可在所述处理器10上运行的计算机程序,如命名实体识别程序。

[0120] 其中,所述存储器11至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、移动硬盘、多媒体卡、卡型存储器(例如:SD或DX存储器等)、磁性存储器、磁盘、光盘等。所述存储器11在一些实施例中可以是电子设备的内部存储单元,例如该电子设备的移动硬盘。所述存储器11在另一些实施例中也可以是电子设备的外部存储设备,例如电子设备上配备的插接式移动硬盘、智能存储卡(Smart Media Card,SMC)、安全数字(Secure Digital,SD)卡、闪存卡(Flash Card)等。进一步地,所述存储器11还可以既包括电子设备的内部存储单元也包括外部存储设备。所述存储器11不仅可以用于存储安装于电子设备的应用软件及各类数据,例如命名实体识别程序的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

[0121] 所述处理器10在一些实施例中可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing unit,CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。所述处理器10是所述电子设备的控制核心(Control Unit),利用各种接口和线路连接整个电子设备的各个部件,通过运行或执行存储在所述存储器11内的程序或者模块(例如命名实体识别程序等),以及调用存储在所述存储器11内的数据,以执行电子设备的各种功能和处理数据。

[0122] 所述通信总线12可以是外设部件互连标准(peripheral component interconnect,简称PCI)总线或扩展工业标准结构(extended industry standard architecture,简称EISA)总线等。该总线可以分为地址总线、数据总线、控制总线等。所述通信总线12总线被设置为实现所述存储器11以及至少一个处理器10等之间的连接通信。为便于表示,图中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0123] 图3仅示出了具有部件的电子设备,本领域技术人员可以理解的是,图3示出的结构并不构成对所述电子设备的限定,可以包括比图示更少或者更多的部件,或者组合某些部件,或者不同的部件布置。

[0124] 例如,尽管未示出,所述电子设备还可以包括给各个部件供电的电源(比如电池),优选地,电源可以通过电源管理装置与所述至少一个处理器10逻辑相连,从而通过电源管理装置实现充电管理、放电管理、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障分类电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述电子设备还可以包括多种传感器、蓝牙模块、Wi-Fi模块等,在此不再赘述。

[0125] 可选地,所述通信接口13可以包括有线接口和/或无线接口(如WI-FI接口、蓝牙接口等),通常用于在该电子设备与其他电子设备之间建立通信连接。

[0126] 可选地,所述通信接口13还可以包括用户接口,用户接口可以是显示器(Display)、输入单元(比如键盘(Keyboard)),可选地,用户接口还可以是标准的有线接口、无线接口。可选地,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备中处理的信息以及用于显示可视化的用户界面。

[0127] 应该了解,所述实施例仅为说明之用,在专利申请范围上并不受此结构的限制。

[0128] 所述电子设备中的所述存储器11存储的命名实体识别程序是多个计算机程序的组合,在所述处理器10中运行时,可以实现:

[0129] 获取待识别实体文本,利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询,得到一个或多个候选实体;

[0130] 对所述候选实体进行字符相关度计算,得到实体字符相关度;

[0131] 对所述候选实体进行词序相关度进行计算,得到实体词序相关度;

[0132] 根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;

[0133] 利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果。

[0134] 具体地,所述处理器10对上述计算机程序的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

[0135] 进一步地,所述电子设备集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。所述计算机可读取介质可以是非易失性的,也可以是易失性的。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)。

[0136] 本发明实施例还可以提供一种计算机可读存储介质,所述可读存储介质存储有计算机程序,所述计算机程序在被电子设备的处理器所执行时,可以实现:

[0137] 获取待识别实体文本,利用所述待识别实体文本在预构建的基于Elasticsearch的实体检索数据库中进行查询,得到一个或多个候选实体;

[0138] 对所述候选实体进行字符相关度计算,得到实体字符相关度;

[0139] 对所述候选实体进行词序相关度进行计算,得到实体词序相关度;

[0140] 根据所述实体字符相关度及所述实体词序相关度进行计算,得到每个所述候选实体的实体相似度;

[0141] 利用所述实体相似度对所有所述候选实体进行筛选识别,得到识别结果。

[0142] 进一步地,所述计算机可用存储介质可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序等;存储数据区可存储根据区块链节点的使用所创建的数据等。

[0143] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的

划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0144] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0145] 本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中,人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0146] 另外,在本发明各个实施例中的各功能模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0147] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。

[0148] 因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0149] 本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一批次网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

[0150] 此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第二等词语用来表示名称,而并不表示任何特定的顺序。

[0151] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

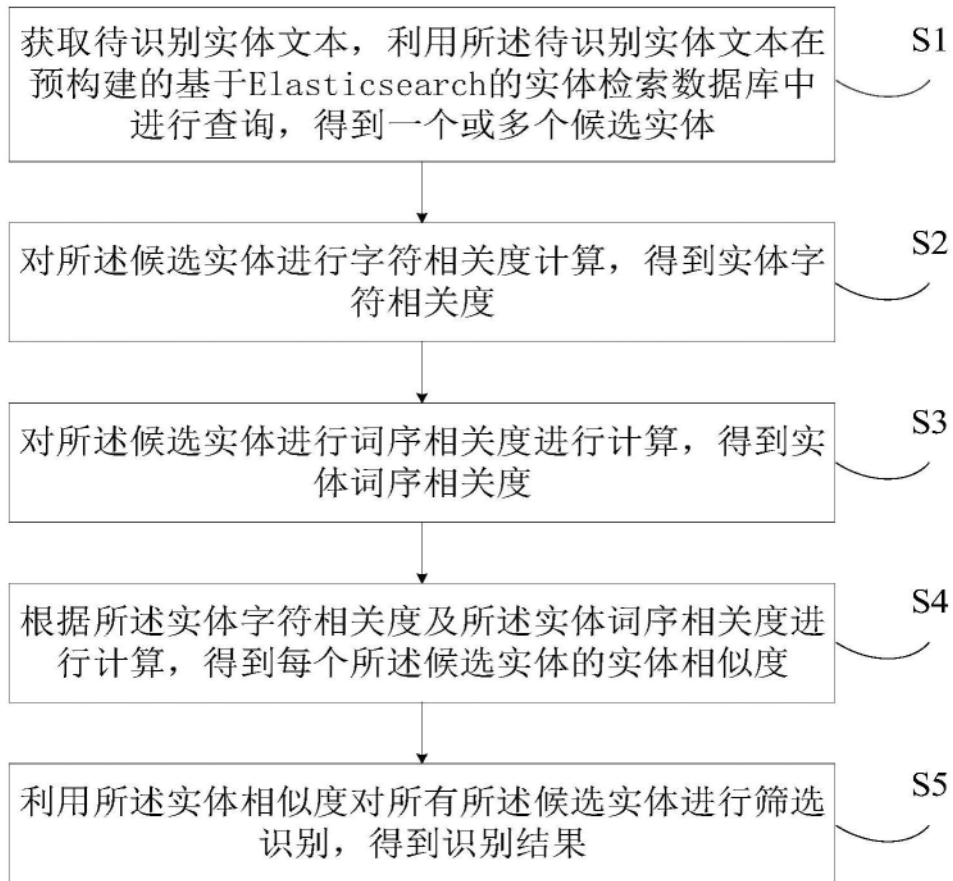


图1

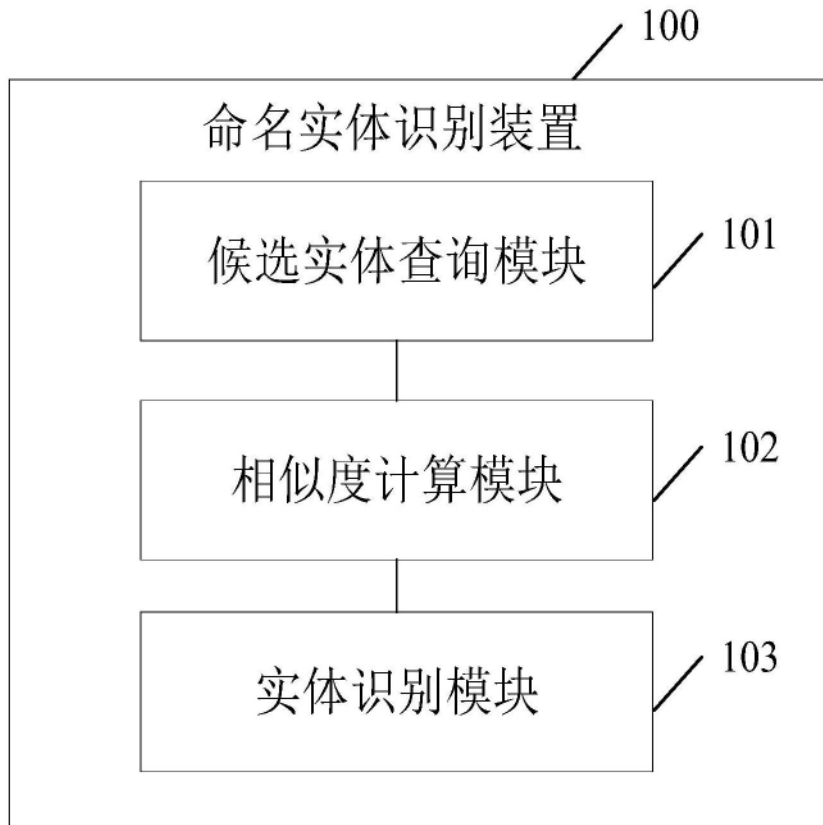


图2

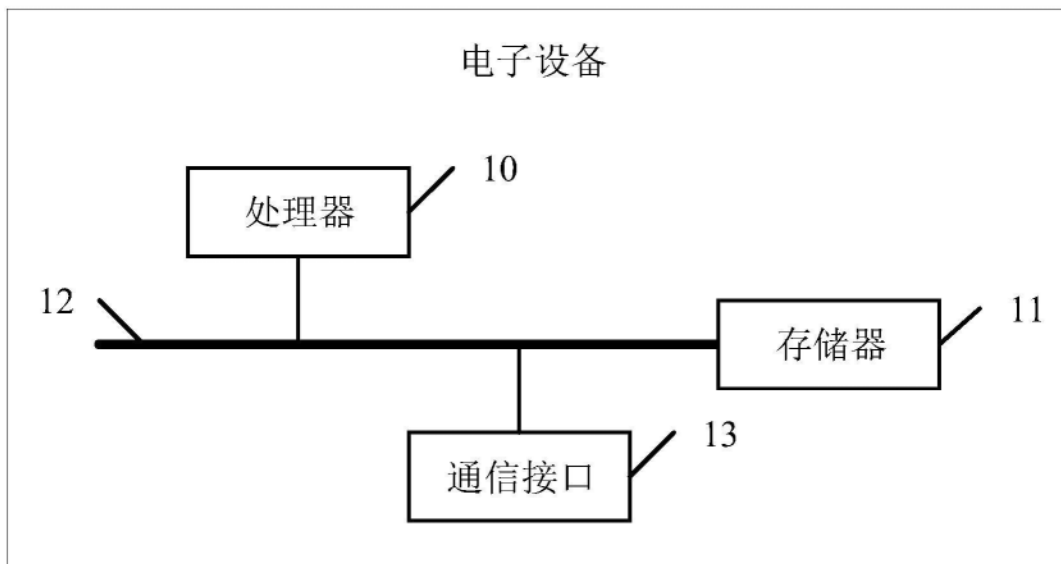


图3