



(12) 发明专利

(10) 授权公告号 CN 112800769 B

(45) 授权公告日 2024.06.14

(21) 申请号 202110195296.7

G06F 16/31 (2019.01)

(22) 申请日 2021.02.20

G06F 16/33 (2019.01)

G06F 16/903 (2019.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 112800769 A

(56) 对比文件

CN 111159535 A, 2020.05.15

(43) 申请公布日 2021.05.14

审查员 张池

(73) 专利权人 深圳追一科技有限公司

地址 518057 广东省深圳市南山区粤海街道科技园社区科苑路8号讯美科技广场3号楼23A、23B

(72) 发明人 崔健

(74) 专利代理机构 华进联合专利商标代理有限公司 44224

专利代理师 纪婷婧

(51) Int. Cl.

G06F 40/295 (2020.01)

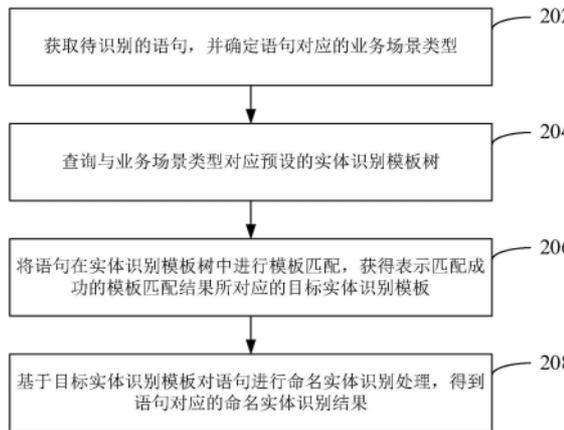
权利要求书2页 说明书16页 附图3页

(54) 发明名称

命名实体识别方法、装置、计算机设备和存储介质

(57) 摘要

本申请涉及一种命名实体识别方法、装置、计算机设备和存储介质。所述方法包括：获取待识别的语句，并确定语句对应的业务场景类型；查询与业务场景类型对应预设的实体识别模板树；将语句在实体识别模板树中进行模板匹配，获得表示匹配成功的模板匹配结果所对应的目标实体识别模板；基于目标实体识别模板对语句进行命名实体识别处理，得到语句对应的命名实体识别结果。采用本方法能够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:
获取待识别的语句,并确定所述语句对应的业务场景类型;
查询与所述业务场景类型对应预设的实体识别模板树;
对所述语句中各字符依次进行编码,得到所述语句对应的字符编码序列;
基于所述字符编码序列在所述实体识别模板树中进行字符状态转移,得到状态转移结果;
根据所述状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板;
基于所述目标实体识别模板对所述语句进行命名实体识别处理,得到所述语句对应的命名实体识别结果。
2. 根据权利要求1所述的方法,其特征在于,所述确定匹配成功的模板匹配结果所对应的目标实体识别模板,包括:
若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定所述业务场景类型对应预设的模板权重映射关系;
基于所述模板权重映射关系,从各所述候选实体识别模板中确定目标实体识别模板。
3. 根据权利要求2所述的方法,其特征在于,所述基于所述模板权重映射关系,从各所述候选实体识别模板中确定目标实体识别模板,包括:
根据所述模板权重映射关系,确定各所述候选实体识别模板分别对应的模板权重;
将所述模板权重中最大的模板权重所对应的候选实体识别模板,确定为目标实体识别模板。
4. 根据权利要求1所述的方法,其特征在于,在所述查询与所述业务场景类型对应预设的实体识别模板树之前,还包括:
基于所述业务场景类型对应的各业务流程,确定由各模板字符按序组成的模板字符序列;
按照各所述业务流程对应的模板特征类型,对各所述模板字符序列进行模板特征化处理,得到至少包括一个所述模板字符的模板特征字段;
基于所述模板特征字段得到实体识别模板,并根据所述实体识别模板构建所述业务场景类型对应的实体识别模板树。
5. 根据权利要求4所述的方法,其特征在于,所述模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。
6. 根据权利要求1至5任意一项所述的方法,其特征在于,所述方法还包括:
确定所述业务场景类型中业务流程对应的待填充词槽;
将所述命名实体识别结果填充至所述待填充词槽中,得到所述业务场景类型对应的业务词槽;
若所述业务词槽满足流程执行条件,则基于所述业务词槽执行所述业务流程对应的业务。
7. 一种命名实体识别装置,其特征在于,所述装置包括:
语句获取模块,用于待识别的语句,并确定所述语句对应的业务场景类型;
模板树查询模块,用于查询与所述业务场景类型对应预设的实体识别模板树;

模板匹配模块,用于对所述语句中各字符依次进行编码,得到所述语句对应的字符编码序列;基于所述字符编码序列在所述实体识别模板树中进行字符状态转移,得到状态转移结果;根据所述状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板;

实体识别模块,用于基于所述目标实体识别模板对所述语句进行命名实体识别处理,得到所述语句对应的命名实体识别结果。

8. 根据权利要求7所述的装置,其特征在于,

所述模板匹配模块,还用于若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定所述业务场景类型对应预设的模板权重映射关系;基于所述模板权重映射关系,从各所述候选实体识别模板中确定目标实体识别模板。

9. 一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至6中任一项所述的方法的步骤。

10. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至6中任一项所述的方法的步骤。

命名实体识别方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及人工智能技术领域,特别是涉及一种命名实体识别方法、装置、计算机设备和存储介质。

背景技术

[0002] 随着人工智能(Artificial Intelligence, AI)技术的发展,自然语言处理(Natural Language Processing, NLP)技术在语音识别、语音翻译、理解完整的句子、理解匹配词的同义词,以及生成语法正确完整句子和段落等方面得到广泛的应用。作为自然语言处理的一个基础任务,命名实体识别(Named Entities Recognition, NER)的目的在于识别语料中人名、地名、组织机构名等有具体含义的实体,如从语句中识别出人名、地名、机构名、时间、日期等命名实体。

[0003] 目前,在对语句进行命名实体识别时,一般对语句进行文本匹配,随着目标词汇的不断扩充,语句需要进行文本匹配的范围急剧增大,导致针对语句的命名实体识别的处理效率较低。

发明内容

[0004] 基于此,有必要针对上述技术问题,提供一种能够提高命名实体识别效率的命名实体识别方法、装置、计算机设备和存储介质。

[0005] 一种命名实体识别方法,所述方法包括:

[0006] 获取待识别的语句,并确定语句对应的业务场景类型;

[0007] 查询与业务场景类型对应预设的实体识别模板树;

[0008] 将语句在实体识别模板树中进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板;

[0009] 基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。

[0010] 在其中一个实施例中,将语句在实体识别模板树进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板,包括:

[0011] 对语句中各字符依次进行编码,得到语句对应的字符编码序列;

[0012] 基于字符编码序列在实体识别模板树中进行字符状态转移,得到状态转移结果;

[0013] 根据状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板。

[0014] 在其中一个实施例中,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板,包括:

[0015] 获得模板匹配的模板匹配结果;

[0016] 若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定业务场景类型对应预设的模板权重映射关系;

[0017] 基于模板权重映射关系,从各候选实体识别模板中确定目标实体识别模板。

[0018] 在其中一个实施例中,基于模板权重映射关系,各候选实体识别模板中确定目标实体识别模板,包括:

[0019] 根据模板权重映射关系,确定各候选实体识别模板分别对应的模板权重;

[0020] 将模板权重中最大的模板权重所对应的候选实体识别模板,确定为目标实体识别模板。

[0021] 在其中一个实施例中,在查询与业务场景类型对应预设的实体识别模板树之前,还包括:

[0022] 基于业务场景类型对应的各业务流程,确定由各模板字符按序组成的模板字符序列;

[0023] 按照各业务流程对应的模板特征类型,对各模板字符序列进行模板特征化处理,得到至少包括一个模板字符的模板特征字段;

[0024] 基于模板特征字段得到实体识别模板,并根据实体识别模板构建业务场景类型对应的实体识别模板树。

[0025] 在其中一个实施例中,模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。

[0026] 在其中一个实施例中,命名实体识别方法还包括:

[0027] 确定业务场景类型中业务流程对应的待填充词槽;

[0028] 将命名实体识别结果填充至待填充词槽中,得到业务场景类型对应的业务词槽;

[0029] 若业务词槽满足流程执行条件,则基于业务词槽执行业务流程对应的业务。

[0030] 一种命名实体识别装置,所述装置包括:

[0031] 语句获取模块,用于待识别的语句,并确定语句对应的业务场景类型;

[0032] 模板树查询模块,用于查询与业务场景类型对应预设的实体识别模板树;

[0033] 模板匹配模块,用于将语句在实体识别模板树中进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板;

[0034] 实体识别模块,用于基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。

[0035] 在其中一个实施例中,模板匹配模块包括字符编码模块、状态转移模块和模板确定模块;其中:字符编码模块,用于对语句中各字符依次进行编码,得到语句对应的字符编码序列;状态转移模块,用于基于字符编码序列在实体识别模板树中进行字符状态转移,得到状态转移结果;模板确定模块,用于根据状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板。

[0036] 在其中一个实施例中,模板匹配模块包括匹配结果获得模块、权重关系确定模块和权重关系处理模块;其中:匹配结果获得模块,用于获得模板匹配的模板匹配结果;权重关系确定模块,用于若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定业务场景类型对应预设的模板权重映射关系;权重关系处理模块,用于基于模板权重映射关系,从各候选实体识别模板中确定目标实体识别模板。

[0037] 在其中一个实施例中,权重关系处理模块包括模板权重确定模块和目标模板选定模块;其中:模板权重确定模块,用于根据模板权重映射关系,确定各候选实体识别模板分

别对应的模板权重；目标模板选定模块，用于将模板权重中最大的模板权重所对应的候选实体识别模板，确定为目标实体识别模板。

[0038] 在其中一个实施例中，还包括字符序列确定模块、特征字段确定模块和模板树构建模块；其中：字符序列确定模块，用于基于业务场景类型对应的各业务流程，确定由各模板字符按序组成的模板字符序列；特征字段确定模块，用于按照各业务流程对应的模板特征类型，对各模板字符序列进行模板特征化处理，得到至少包括一个模板字符的模板特征字段；模板树构建模块，用于基于模板特征字段得到实体识别模板，并根据实体识别模板构建业务场景类型对应的实体识别模板树。

[0039] 在其中一个实施例中，模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。

[0040] 在其中一个实施例中，还包括待填充词槽确定模块、业务词槽确定模块和业务执行模块；其中：待填充词槽确定模块，用于确定业务场景类型中业务流程对应的待填充词槽；业务词槽确定模块，用于将命名实体识别结果填充至待填充词槽中，得到业务场景类型对应的业务词槽；业务执行模块，用于若业务词槽满足流程执行条件，则基于业务词槽执行业务流程对应的业务。

[0041] 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现以下步骤：

[0042] 获取待识别的语句，并确定语句对应的业务场景类型；

[0043] 查询与业务场景类型对应预设的实体识别模板树；

[0044] 将语句在实体识别模板树中进行模板匹配，获得表示匹配成功的模板匹配结果所对应的目标实体识别模板；

[0045] 基于目标实体识别模板对语句进行命名实体识别处理，得到语句对应的命名实体识别结果。

[0046] 在其中一个实施例中，处理器执行计算机程序时还实现以下步骤：对语句中各字符依次进行编码，得到语句对应的字符编码序列；基于字符编码序列在实体识别模板树中进行字符状态转移，得到状态转移结果；根据状态转移结果得到模板匹配结果，并确定匹配成功的模板匹配结果所对应的目标实体识别模板。

[0047] 在其中一个实施例中，处理器执行计算机程序时还实现以下步骤：获得模板匹配的模板匹配结果；若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一，则确定业务场景类型对应预设的模板权重映射关系；基于模板权重映射关系，从各候选实体识别模板中确定目标实体识别模板。

[0048] 在其中一个实施例中，处理器执行计算机程序时还实现以下步骤：根据模板权重映射关系，确定各候选实体识别模板分别对应的模板权重；将模板权重中最大的模板权重所对应的候选实体识别模板，确定为目标实体识别模板。

[0049] 在其中一个实施例中，处理器执行计算机程序时还实现以下步骤：基于业务场景类型对应的各业务流程，确定由各模板字符按序组成的模板字符序列；按照各业务流程对应的模板特征类型，对各模板字符序列进行模板特征化处理，得到至少包括一个模板字符的模板特征字段；基于模板特征字段得到实体识别模板，并根据实体识别模板构建业务场景类型对应的实体识别模板树。

[0050] 在其中一个实施例中,模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。

[0051] 在其中一个实施例中,处理器执行计算机程序时还实现以下步骤:确定业务场景类型中业务流程对应的待填充词槽;将命名实体识别结果填充至待填充词槽中,得到业务场景类型对应的业务词槽;若业务词槽满足流程执行条件,则基于业务词槽执行业务流程对应的业务。

[0052] 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现以下步骤:

[0053] 获取待识别的语句,并确定语句对应的业务场景类型;

[0054] 查询与业务场景类型对应预设的实体识别模板树;

[0055] 将语句在实体识别模板树中进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板;

[0056] 基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。

[0057] 在其中一个实施例中,计算机程序被处理器执行时还实现以下步骤:对语句中各字符依次进行编码,得到语句对应的字符编码序列;基于字符编码序列在实体识别模板树中进行字符状态转移,得到状态转移结果;根据状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板。

[0058] 在其中一个实施例中,计算机程序被处理器执行时还实现以下步骤:获得模板匹配的模板匹配结果;若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定业务场景类型对应预设的模板权重映射关系;基于模板权重映射关系,从各候选实体识别模板中确定目标实体识别模板。

[0059] 在其中一个实施例中,计算机程序被处理器执行时还实现以下步骤:根据模板权重映射关系,确定各候选实体识别模板分别对应的模板权重;将模板权重中最大的模板权重所对应的候选实体识别模板,确定为目标实体识别模板。

[0060] 在其中一个实施例中,计算机程序被处理器执行时还实现以下步骤:基于业务场景类型对应的各业务流程,确定由各模板字符按序组成的模板字符序列;按照各业务流程对应的模板特征类型,对各模板字符序列进行模板特征化处理,得到至少包括一个模板字符的模板特征字段;基于模板特征字段得到实体识别模板,并根据实体识别模板构建业务场景类型对应的实体识别模板树。

[0061] 在其中一个实施例中,模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。

[0062] 在其中一个实施例中,计算机程序被处理器执行时还实现以下步骤:确定业务场景类型中业务流程对应的待填充词槽;将命名实体识别结果填充至待填充词槽中,得到业务场景类型对应的业务词槽;若业务词槽满足流程执行条件,则基于业务词槽执行业务流程对应的业务。

[0063] 上述命名实体识别方法、装置、计算机设备和存储介质,确定与待识别的语句对应的业务场景类型,将语句在业务场景类型对应预设的实体识别模板树中进行模板匹配,得到匹配成功的目标实体识别模板,并基于目标实体识别模板对语句进行命名实体识别处

理,得到语句对应的命名实体识别结果。在命名实体识别处理过程中,根据与待识别语句的业务场景类型对应的实体识别模板树进行模板匹配,可以使实体识别模板树适应于语句对应的业务场景,提高命名实体识别的针对性,进一步通过实体识别模板树进行模板匹配,可以提高模板匹配的处理效率,从而提高了命名实体识别的处理效率。

附图说明

- [0064] 图1为一个实施例中命名实体识别方法的应用环境图;
- [0065] 图2为一个实施例中命名实体识别方法的流程示意图;
- [0066] 图3为一个实施例中确定目标实体识别模板的流程示意图;
- [0067] 图4为另一个实施例中trie树的数据结构示意图;
- [0068] 图5为一个实施例中命名实体识别装置的结构框图;
- [0069] 图6为一个实施例中计算机设备的内部结构图。

具体实施方式

[0070] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0071] 本申请提供的命名实体识别方法,可以应用于如图1所示的应用环境中。其中,终端102通过网络与服务器104进行通信。用户在终端102输入语音数据或文本数据,终端102获得用户输入的待识别的语句,并将待识别的语句发送至服务器104,服务器104接收到待识别的语句后,确定与待识别的语句对应的业务场景类型,将语句在业务场景类型对应预设的实体识别模板树中进行模板匹配,得到匹配成功的目标实体识别模板,并基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果,服务器104可以将命名实体识别结果反馈至终端102。其中,终端102可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备,服务器104可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

[0072] 在一个实施例中,如图2所示,提供了一种命名实体识别方法,以该方法应用于图1中的服务器为例进行说明,包括以下步骤:

[0073] 步骤202,获取待识别的语句,并确定语句对应的业务场景类型。

[0074] 其中,语句是需要进行命名实体识别的语料,如可以为文本语句,语句由一个或多个字符组成。例如,对于语句“今天上完班,我就想休假去北京玩几天了。”语句还可以是基于语音识别技术获得的语音文本,如终端可以通过麦克风采集用户的语音数据,基于语音识别技术对采集的语音数据进行自然语言理解处理后得到文本语句。命名实体是指语句中具有特定意义的实体,如人名、地名、组织机构、日期、时间、百分数或货币等,命名实体识别即需要从语句中识别出命名实体,如从待识别的语句中识别出人名、地名、机构名、专有名词、时间、日期等。业务场景类型根据业务场景预先进行划分确定,在不同的业务场景中,会有不同的业务需求,不同的业务需求对应于不同的业务流程和语句,即在不同的业务场景中,用户通过不同的语句表达相应的业务需求。例如,在出租车预定的业务场景中,用户输入的待识别语句与出租车预定业务相关。在不同的业务场景,不同的业务需求中,需要识别不同

的命名实体,及有不同的命名实体识别任务,如在机票预定的业务场景中,命名实体识别需从语句中识别出发地点、到达地点和出发时间等;而在预订酒店的业务场景中,命名实体识别需从语句中识别出入住时间、离店时间、酒店名称等。

[0075] 具体地,在触发进行命名实体识别处理时,如接收到用户通过终端发送与业务流程相关的语句时,获取待识别的语句,并确定语句对应的业务场景类型。在具体应用时,可以对待识别的语句进行意图识别,如通过词表穷举法、规则模板解析法、深度学习意图识别模型等,对待识别的语句进行意图识别,根据意图识别结果确定语句对应的业务场景,从确定对应的业务场景类型。在具体实现时,若待识别的语句并非业务场景中触发业务流程的语句,即待识别的语句为在业务场景中触发业务流程后的交互语句,如用户在触发预订机票的业务流程后,待识别的语句为用户在人机交互过程中关于出发时间的描述时,可以根据待识别的语句的上下文确定待识别的语句对应的业务场景类型。

[0076] 步骤204,查询与业务场景类型对应预设的实体识别模板树。

[0077] 其中,实体识别模板树由实体识别模板构建,实体识别模板为预先设定的用于命名实体识别的语句模板,在满足语句与实体识别模板匹配时,可以用匹配的实体识别模板对语句进行识别,从而实现对语句的命名实体识别处理。实体识别模板树与业务场景类型对应设置,即不同业务场景类型可以对应于不同的实体识别模板树,从而提高实体识别模板树与语句的针对性,通过对应业务场景类型的实体识别模板树进行模板匹配,可以利用业务场景类型对各种类型的模板进行过滤,有效提高模板匹配的针对性,提高模板匹配的处理效率。

[0078] 在具体应用时,实体识别模板树可以为基于实体识别模板构建的双数组树,具体可以为双数组trie树。双数组Trie树(Double-array Trie,DAT)是一种Trie树的高效实现,其兼顾了查询效率与空间存储,极大地节省了内存占用。双数组Trie树是一种空间复杂度低的Trie树,应用于字符区间大的语言(如中文、日文等)分词领域。双数组的原理是,将原来需要多个数组才能表示的Trie树,使用两个数据就可以存储下来,可以极大的减小空间复杂度。具体来说:使用两个数组base和check来维护Trie树,base负责记录状态,check负责检查各个字符串是否是从同一个状态转移而来,当check[i]为负值时,表示此状态为字符串的结束。

[0079] 具体地,在确定待识别的语句对应的业务场景类型后,服务器可以查询与业务场景类型对应预设的实体识别模板树。具体应用时,可以针对不同的业务场景类型预先构建相应的实体识别模板树,根据各业务场景类型对应的实体识别模板树得到模板树库。服务器在确定语句对应的业务场景类型后,可以查询模板树库,根据业务场景类型的类型标识从模板树库中查询得到业务场景类型对应的实体识别模板树。

[0080] 步骤206,将语句在实体识别模板树中进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板。

[0081] 其中,实体识别模板为预先设定的用于命名实体识别的语句模板,目标实体识别模板为从实体识别模板树中包括的各实体识别模板中,确定的用于对待识别的语句进行命名实体识别处理的实体识别模板。

[0082] 具体地,服务器查询得到与业务场景类型对应预设的实体识别模板树后,将语句在实体识别模板树中进行模板匹配,从而从实体识别模板树中确定用于对语句进行命名实

体识别处理的目标实体识别模板。在具体实现时,服务器可以将语句与实体识别模板树中各节点进行匹配,将模板匹配结果为匹配成功的实体识别模板确定为目标实体识别模板。例如,实体识别模板树为双数组trie树时,可以对语句的各字段进行拆分,基于双数组trie树算法确定语句中的状态转移信息,根据状态转移信息,利用base和check数组在实体识别模板树中进行状态转移,从而跟状态转移结果确定目标实体识别模板。

[0083] 步骤208,基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。

[0084] 具体地,在得到目标实体识别模板后,服务器基于该目标实体识别模板对语句进行命名实体识别处理,从而从语句中识别出相应的命名实体,得到语句对应的命名实体识别结果。在具体应用中,可以通过目标实体识别模板将语句进行拆分,并将相应的字段填充到目标实体识别模板对应的词槽中,得到语句对应的命名实体识别结果。

[0085] 上述命名实体识别方法中,确定与待识别的语句对应的业务场景类型,将语句在业务场景类型对应预设的实体识别模板树中进行模板匹配,得到匹配成功的目标实体识别模板,并基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。在命名实体识别处理过程中,根据与待识别语句的业务场景类型对应的实体识别模板树进行模板匹配,可以使实体识别模板树适应于语句对应的业务场景,提高命名实体识别的针对性,进一步通过实体识别模板树进行模板匹配,可以提高模板匹配的处理效率,从而提高了命名实体识别的处理效率。

[0086] 在一个实施例中,将语句在实体识别模板树进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板,包括:对语句中各字符依次进行编码,得到语句对应的字符编码序列;基于字符编码序列在实体识别模板树中进行字符状态转移,得到状态转移结果;根据状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板。

[0087] 其中,语句包括不同的字段,语句中的各字段由相应字符组成。字符编码序列为依次对语句中的各字符进行编码后得到的字符编码按序组成的序列。不同的字符编码对应于不同的字符,但同一字符可以对应于多个不同的字符编码。在具体实现时,对语句中各字符的编码算法与对实体识别模板树中各实体识别模板包括字符的编码算法相同,从而将语句中各字符映射到与各实体识别模板的字符相同的编码空间,确保模板匹配的准确性。字符状态转移,是指将语句对应的字符编码序列中各个字符编码按序进行状态转移处理,状态转移时,每个字符编码认为为一种状态,而字符编码之间的先后关系即为字符编码的状态转移。

[0088] 具体地,在确定语句的业务场景类型对应的实体识别模板树后,服务器可对语句中各字符依次进行编码,得到语句中各字符分别对应字符编码,服务器将字符编码按照字符在语句中的排序依次进行组合,得到语句对应的字符编码序列。在具体应用时,可以查询构建实体识别模板树时对应的字符映射规则,服务器按照该映射规则对语句中各字符依次进行编码,得到语句对应的字符编码序列。服务器将获得的字符编码序列在实体识别模板树中进行字符状态转移,具体可以将字符编码序列中各个字符对应的字符编码认为为一种字符状态,字符编码的先后联系,为字符状态的转移。例如,对于语句A“王四还欠张三20块钱。”其中,可以将“王”“四”“还”“欠”“张”“三”“20”“块”分别对应的字符编码看作为一种字

符状态,而该语句A正是由“王”“四”“还”“欠”“张”“三”“20”“块”的字符状态,按照顺序进行状态转移得到的。通过将字符编码序列在实体识别模板树中进行字符状态转移,可以确定字符编码序列是否与实体识别模板树中某一实体识别模板的字符状态转移过程匹配,若匹配,则表明待识别的语句命中该实体识别模板,即可以通过该实体识别模板对语句进行命名实体识别处理。服务器得到字符编码序列在实体识别模板树中进行字符状态转移的状态转移结果后,服务器根据状态转移结果得到模板匹配结果,并从模板匹配结果中确定匹配成功的模板匹配结果所对应的目标实体识别模板。具体实现时,服务器可以直接将状态转移结果确定为模板匹配结果,服务器将实体识别模板树中与字符编码序列的字符状态转移过程匹配的实体识别模板,确定其对应的模板匹配结果为匹配成功,即将其作为目标实体识别模板。

[0089] 本实施例中,通过将语句各字符分别编码后获得的字符编码序列,在实体识别模板树中进行字符状态转移,根据状态转移结果确定目标实体识别模板,从而通过状态转移方式实现模板匹配,可以提高模板匹配的处理效率,从而提高命名实体识别的处理效率。

[0090] 在一个实施例中,如图3所示,目标实体识别模板确定的步骤,即获得表示匹配成功的模板匹配结果所对应的目标实体识别模板,包括:

[0091] 步骤302,获得模板匹配的模板匹配结果。

[0092] 具体地,服务器在将语句在实体识别模板树中进行模板匹配后,获得模板匹配的模板匹配结果,模板匹配结果反映了语句在实体识别模板树的匹配信息。若语句在实体识别模板树中匹配成功,则表明语句可以命中实体识别模板树中的实体识别模板,即语句可以通过命中的实体识别模板进行准确命名实体识别处理。若语句在实体识别模板树中无法命中任一实体识别模板,则表明语句与任一实体识别模板不匹配,则服务器不可通过预设的实体识别模板进行命名实体识别处理,服务器可以直接对语句进行命名实体识别处理,如可以基于词性分析进行命名实体识别,也可以基于预训练的神经网络模型对语句进行命名实体识别处理。

[0093] 步骤304,若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定业务场景类型对应预设的模板权重映射关系。

[0094] 得到模板匹配的模板匹配结果后,服务器从各模板匹配结果中,确定表示匹配成功的模板匹配结果对应的候选实体识别模板。在候选实体识别模板的数量大于一,即表示匹配成功的模板匹配结果的数量大于一,语句在实体识别模板树中命中多个实体识别模板,则服务器确定业务场景类型对应预设的模板权重映射关系。其中,模板权重映射关系包括实体识别模板树中各实体识别模板对应预设的模板权重,模板权重反映了实体识别模板在业务场景中的重要程度和优先级别。模板权重越大的实体识别模板,在业务场景的优先级别高,重要程度高,则可以进行优先命中。模板权重映射关系中各实体识别模板的模板权重可以根据对应业务场景类型的实际需求进行灵活设置,以使模板匹配与业务场景对应,确保模板匹配的准确性。

[0095] 步骤306,基于模板权重映射关系,从各候选实体识别模板中确定目标实体识别模板。

[0096] 得到业务场景类型对应预设的模板权重映射关系后,服务器基于该模板权重映射关系,从各候选实体识别模板中确定目标实体识别模板。具体地,服务器可以将模板权重最

大的候选实体识别模板确定为目标实体识别模板,从而确保目标实体识别模板的优先级别。

[0097] 本实施例中,在匹配成功的候选实体识别模板数量超过1,即待识别的语句在实体识别模板树中命中多个实体识别模板时,服务器基于业务场景类型对应预设的模板权重映射关系,从各个候选实体识别模板中确定目标实体识别模板,可以确保目标实体识别模板的优先级别,从而确保模板匹配的准确性,提高了基于目标实体识别模板进行命名实体识别处理时的处理效率。

[0098] 在一个实施例中,基于模板权重映射关系,各候选实体识别模板中确定目标实体识别模板,包括:根据模板权重映射关系,确定各候选实体识别模板分别对应的模板权重;将模板权重中最大的模板权重所对应的候选实体识别模板,确定为目标实体识别模板。

[0099] 具体地,服务器得到模板权重映射关系后,从模板权重映射关系中,确定各候选实体识别模板分别对应的模板权重。具体实现时,服务器可以确定各候选实体识别模板的模板标识,如模板编号、ID或名称等,将模板标识在模板权重映射关系中进行标识匹配,根据标识匹配结果确定候选实体识别模板对应的模板权重。得到各候选实体识别模板分别对应的模板权重后,服务器比较各候选实体识别模板分别对应的模板权重的数值大小,确定模板权重中最大的模板权重所对应的候选实体识别模板,将其确定为目标实体识别模板。

[0100] 本实施例中,在匹配成功的候选实体识别模板数量超过1,即待识别的语句在实体识别模板树中命中多个实体识别模板时,服务器基于业务场景类型对应预设的模板权重映射关系,确定各个候选实体识别模板分别对应的模板权重,并将模板权重最大的候选实体识别模板确定为目标实体识别模板,可以确保目标实体识别模板的优先级别,从而确保模板匹配的准确性,提高了基于目标实体识别模板进行命名实体识别处理时的处理效率。

[0101] 在一个实施例中,在查询与业务场景类型对应预设的实体识别模板树之前,还包括:基于业务场景类型对应的各业务流程,确定由各模板字符按序组成的模板字符序列;按照各业务流程对应的模板特征类型,对各模板字符序列进行模板特征化处理,得到至少包括一个模板字符的模板特征字段;基于模板特征字段得到实体识别模板,并根据实体识别模板构建业务场景类型对应的实体识别模板树。

[0102] 其中,不同的业务场景类型对应于不同的业务场景,不同的业务场景下存在不同的业务流程,业务流程可以为通过语句执行的业务相应的流程,如用户通过语句进行出租车预订处理时,业务流程即为预定出租车的处理流程。一般地,不同的业务场景对应于不同的业务流程,在不同的业务流程的交互过程中,涉及不同的语句。业务流程涉及不同的语句句式模板,例如出租车预定中,一般包括“[出发时间]从[出发地点]到[目的地]”的句式模板,即业务流程对应的实体识别模板具有一定的格式。对于业务场景类型中的各个业务流程,可以生成不同的实体识别模板,实体识别模板由不同的模板字符有序组合得到。模板字符为组成实体识别模板的字符元素,模板字符序列由模板字符按照与业务流程对应的顺序组成得到。模板特征类型与业务流程对应,模板特征类型反映了业务流程的业务特点。模板特征字段根据对模板字符序列进行特征化处理得到,不同的模板特征字段描述了实体识别模板不同的特征,如对于省略词特征字段,描述了实体识别模板中可以省略的字段,对于省略的字段,不用将其进行匹配,可以直接省略,而对其他字符进行模板匹配。

[0103] 具体地,服务器预先构建与各种业务场景类型对应的实体识别模板树,服务器可

以确定业务场景类型对应的各业务流程,基于各业务流程,确定对应于每个业务流程构建实体识别模板的模板字符,即业务流程中需要采用的特征字符,如“[出发时间]从[出发地点]到[目的地]”的实体识别模板中,“从”和“到”可以为该业务流程中的模板字符。服务器将各模板字符按照业务流程的需要按序进行组合,得到模板字符序列。进一步地,服务器确定各业务流程对应的模板特征类型,模板特征类型描述了模板字符序列中各模板字符的特征类型,不同特征类型的模板字符对实体识别模板的构建起不同的限定作用。服务器按照各业务流程对应的模板特征类型,对各模板字符序列进行模板特征化处理,具体可以对各模板字符序列中的各模板字符分别进行模板特征化处理,如可以对各模板字符进行编码或格式转换等,得到至少包括一个模板字符的模板特征字段。服务器根据模板特征字段构建实体识别模板,并基于实体识别模板构建业务场景类型对应的实体识别模板树。具体实现时,可以基于双数组Trie树算法,将实体识别模板构建为实体识别模板树,即得到实体识别模板的双数组Trie树。

[0104] 本实施例中,预先根据业务场景类型中的各业务流程确定模板字符序列,并按照业务流程对应的模板特征类型对模板字符序列进行模板特征化处理,基于获得的模板特征字段得到实体识别模板,并根据实体识别模板构建业务场景类型对应的实体识别模板树,从而确保实体识别模板树与业务场景类型中的各业务流程对应,且能够准确反映各业务流程的模板特征,从而确保实体识别模板树的准确性。

[0105] 在一个实施例中,模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。

[0106] 其中,省略词特征字段表示对应位置字段可以省略字符,即在模板匹配时可以跳过一部分字符进行匹配,具体格式可以为“[w:2-4]”,表示可以省略2-4各字符。词槽特征字段表示需要进行词槽填充的字段,具体格式为“[name]”“[money]”,分别表示需要填充姓名和金额的词槽。关键词特征字段表示需要进行匹配的关键字段,匹配成功才可以命中相应模板,如可以为“的天气”,表示语句需要命中“的天气”这三个字符,才可以命中该实体识别模板。符合特征字段可以将各个模板特征字段进行连接,可以对各个模板特征字段的范围和联系进行限定,如可以通过“()”限定模板特征字段的范围,通过“|”来表示模板特征字段之间的或关系,即命中存在或关系的任一模板特征字段,则可以命中对应的实体识别模板。

[0107] 具体地,模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。通过配置的多种类型的模板特征字段,可以对实体识别模板进行灵活配置,能满足更多的自然语言处理场景,使运行效率高,从而提高命名实体识别的处理效率。

[0108] 在一个实施例中,命名实体识别方法还包括:确定业务场景类型中业务流程对应的待填充词槽;将命名实体识别结果填充至待填充词槽中,得到业务场景类型对应的业务词槽;若业务词槽满足流程执行条件,则基于业务词槽执行业务流程对应的业务。

[0109] 其中,待填充词槽将用户意图转化为明确的业务流程指令所需要补充的信息,如预订机票的业务流程中,待填充词槽可以包括出发地点、出发时间和到达地点。业务词槽为根据命名实体识别结果对待填充词槽进行填充后得到的信息。流程执行条件用于触发业务流程的执行,如预订机票的业务流程中,出发地点、出发时间和到达地点等三个词槽均填充

完成,且不冲突时,如出发地点与到达地点不同,此时可以执行预订机票的业务流程,则满足流程执行条件,触发业务流程的执行,进行机票预订处理。

[0110] 具体地,得到语句的命名实体识别结果后,服务器确定业务场景类型中业务流程对应的待填充词槽,不同的业务流程对应于不同的待填充词槽,对业务流程对应的待填充词槽进行填充后,满足流程执行条件时,可以触发业务流程的执行。具体由服务器将命名实体识别结果填充至待填充词槽中,得到业务流程对应的业务词槽,服务器根据业务词槽确定是否满足业务流程对应的流程执行条件,若是,则触发业务流程的执行,服务器基于业务词槽执行业务流程。在具体应用时,不同的业务流程对应于不同的待填充词槽,具有不同的流程执行条件,根据实际需要进行预先配置。

[0111] 本实施例中,根据语句的命名实体识别结果对业务流程对应的待填充词槽进行填充,从而执行相应的业务流程,可以通过识别到的准确的命名实体识别结果进行词槽填充,能够确保业务流程执行的准确性。

[0112] 在一个实施例中,提供了一种命名实体识别方法,该方法中实体识别模板树基于双数组trie树算法构建。具体地,Trie树是一种树形结构,是一种哈希树的变种。其优点是可以利用字符串的公共前缀来减少查询时间,最大限度地减少无谓的字符串比较,能在常数时间 $O(len)$ 内实现插入和查询操作,是一种以空间换取时间的数据结构,广泛用于词频统计和输入统计领域。如图4所示,假设有一批数据,里面包含5个字符串,它们分别是:Code,Cook,Five,File,Fat,将图4所示的形式存储在树形结构里,如查询数据中是否有Cook时,从第一个根节点(/)出发,按照左左右右可以直接找出查询到该单词,效率为 $O(len)$,即查询时间与查询数据的字符长度对应。

[0113] 而对于双数组trie树是在Trie数实现过程中,发现每个节点均需要一个数组来存储next节点,非常占用存储空间,空间复杂度大,双数组Trie树正是解决该问题。双数组Trie树是一种空间复杂度低的Trie树,应用于字符区间大的语言(如中文、日文等)分词领域。双数组的原理是,将原来需要多个数组才能表示的Trie树,使用两个数据就可以存储下来,可以极大的减小空间复杂度。具体来说,使用两个数组base和check来维护Trie树,base负责记录状态,check负责检查各个字符串是否是从同一个状态转移而来,当 $check[i]$ 为负值时,表示此状态为字符串的结束。例如,假定两个单词ta,tb,其base和check的值会满足条件: $base[t]+a.code=base[ta]$ 、 $base[t]+b.code=base[tb]$ 和 $check[ta]=check[tb]$ 。在每个节点插入的过程中会修改这两个数组,具体说来,首先初始化root节点 $base[0]=1$; $check[0]=0$;对于每一群兄弟节点(属于同一层级的节点),寻找一个begin值使得 $check[begin+a1\cdots an]=0$,也就是找到了n个空闲空间, $a1\cdots an$ 是siblings中的n个节点对应的code;然后将这群兄弟节点的check设为 $check[begin+a1\cdots an]=begin$;接着对每个兄弟节点,如果它没有孩子(即节点的子节点),令其base为负值;否则为该节点的子节点的插入位置(也就是begin值),同时插入子节点,返回重新寻找一个begin值使得 $check[begin+a1\cdots an]=0$ 的处理。基于双数组trie树构建实体识别模板树,可以有效提高模板匹配的处理效率。

[0114] 进一步地,实体识别模板树中的各实体识别模板,根据业务场景类型中的各业务流程对应配置。在配置时,还可以配置各个实体识别模板对应的模板权重,以同步至模板引擎中。实体识别模板的配置方式可以为,默认所有的字符都是有序序列,例如,用户问句“你

好”可以命中模板“你好”,用户问句“好你”不会命中“你好”;“*”或“[w:2-4]”代表省略词片段,“*”代表固定省略0-10个字符,“[w:2-4]”表示这个位置可以省略固定字数,其中“2”表示至少有2个字符,“4”表示最多有四个字符;“[Slot_Key]”代表词槽片段,“[Slot_Key]”表示这个位置可以由词槽键(key)所示词槽中的所有实例或别名,如模板“给[name]转[money]”可以识别到“给张三转1000元”或者“给李四转个800块吧”。

[0115] 具体地,对于“[Slot_Key]”,表示词槽,即表示该位置可以是该词槽键(key)所示词槽中的所有实例或别名,如“给张三转1000元”会匹配到模板“给[name]转[money]”,而所有的文字都是有序序列,顺序必须对应才能匹配。对于“[w:2-4]”,表示省略字符,该位置可以省略固定字数2-4个字,如“我决定给张三转1000元”会匹配到模板“[w:2-4]给[name]转[money]”,且“[w:2-4]”不能连续相连多个省略片段,比如“[w:2-4][w:3-5]”为不合法。对于“()”,用于划定“或”关系的范围,如“我决定给张三现金1000元”或“我决定给张三转账1000元”都可以匹配到模板“[w:2-4]给[name](现金|转账)[money]”,且“()”需要配合“|”使用,两者是强关联,“()”单独使用没有意义,“(给)”与“给”等同,但不能多个“()”进行嵌套,例如:“(给[name](现金)|(转账)[money])”不合法,但是“()”中可以包括表示词槽的“[]”,例如:“([city]|[countryregion])”表示城市或乡村的词槽。对于“|”,表示“或”关系:命中“|”两边表达式中的一个即为命中,如“我决定给张三现金1000元”或“我决定给张三转账1000元”或“我决定给张三支付宝1000元”都可以匹配到模板“[w:2-4]给[name](现金|转账|支付宝)[money]”,表示“或”关系的范围的两边需要用小括号“()”括起来,且支持多个“|”连用。

[0116] 具体地,对于实体识别模板“[w:2-4]([city]|[countryregion])的(降雨|风向)”,其中包括类型词片段“[city]”为词槽特征字段,表示该位置可以由该类型的所有表达中的一项构成,类型词片段的识别功能,可以通过提供词典“或”要素/实体识别来实现。省略词特征字段“[w:2-4]”,表示这个位置可以省略几个字,其中“2”表示至少有2个字符,“4”表示最多有四个字符,但多个省略片段不能相连,如“[w:2-4][w:3-5]”为不合法,语句中每一个字母、数字、标点、汉字,都算作一个字符。字序列片段“的天气”是关键词特征字段,表示由字序列片段中包含的有序词构成,每一实体识别模板可以由以上三种片段中的一项或多项组合构成。对于小括号“()”为符合特征字段,其为了将一部分内容限定在一起,括号内是一个子模板,如“[city]的天气”中,包含4个子模板:“[city]”、“的”、“天”、“气”;“[city]的(天气)”中,则包含3个子模板:“[city]”、“”的和“天气”;“([city])”和“[city]”等效,“的”和“(的)”等效。对于或关系符号“|”为符合特征字段,表示几个表达式是或的关系,命中一个即为命中。例如,“([city]|[countryregion])”表示这部分可能是城市也可能是农村地区。“降雨|风向”表示这部分可能是“降雨”,也可能是“风向”。“([city]的天气|[countryregion]的美食)”表示可能是“某个城市的天气”或“某个乡村的美食”。但各种括号之前不能交叉重叠,括号不支持嵌套,或关系必须用小括号括起来。

[0117] 在具体应用时,将语句query在实体识别模板树中进行模板匹配时,需要只有语句query和模板完全匹配时,则命中该模板。具体地,模板表达式匹配命中规则为当query和该模板完全匹配时,则命中该模板。如模板“[w:0-4][city]的天气”,query为“查下北京的天气”,则命中;若query为“查下北京的天气怎么样”,则不命中。

[0118] 本实施例中,使用实体识别模板构建模板树,模板树为双数组trie树,文本查询

时,使用文本在检索树中查找,匹配类型名词典、类型词词典、省略字的通配符,如果能匹配成功,将命中的实体识别模板返回,并基于命中的目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。实体识别模板的配置灵活,能满足更多的自然语言处理场景,运行效率高,在大数据量和大量模板情况下尤为显著,提高了命名实体识别的处理效率。

[0119] 应该理解的是,虽然图2-3的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,图2-3中的至少一部分步骤可以包括多个步骤或者多个阶段,这些步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0120] 在一个实施例中,如图5所示,提供了一种命名实体识别装置500,包括:语句获取模块502、模板树查询模块504、模板匹配模块506和实体识别模块508,其中:

[0121] 语句获取模块502,用于待识别的语句,并确定语句对应的业务场景类型;

[0122] 模板树查询模块504,用于查询与业务场景类型对应预设的实体识别模板树;

[0123] 模板匹配模块506,用于将语句在实体识别模板树中进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板;

[0124] 实体识别模块508,用于基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。

[0125] 在一个实施例中,模板匹配模块506包括字符编码模块、状态转移模块和模板确定模块;其中:字符编码模块,用于对语句中各字符依次进行编码,得到语句对应的字符编码序列;状态转移模块,用于基于字符编码序列在实体识别模板树中进行字符状态转移,得到状态转移结果;模板确定模块,用于根据状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板。

[0126] 在一个实施例中,模板匹配模块506包括匹配结果获得模块、权重关系确定模块和权重关系处理模块;其中:匹配结果获得模块,用于获得模板匹配的模板匹配结果;权重关系确定模块,用于若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定业务场景类型对应预设的模板权重映射关系;权重关系处理模块,用于基于模板权重映射关系,从各候选实体识别模板中确定目标实体识别模板。

[0127] 在一个实施例中,权重关系处理模块包括模板权重确定模块和目标模板选定模块;其中:模板权重确定模块,用于根据模板权重映射关系,确定各候选实体识别模板分别对应的模板权重;目标模板选定模块,用于将模板权重中最大的模板权重所对应的候选实体识别模板,确定为目标实体识别模板。

[0128] 在一个实施例中,还包括字符序列确定模块、特征字段确定模块和模板树构建模块;其中:字符序列确定模块,用于基于业务场景类型对应的各业务流程,确定由各模板字符按序组成的模板字符序列;特征字段确定模块,用于按照各业务流程对应的模板特征类型,对各模板字符序列进行模板特征化处理,得到至少包括一个模板字符的模板特征字段;模板树构建模块,用于基于模板特征字段得到实体识别模板,并根据实体识别模板构建业务场景类型对应的实体识别模板树。

[0129] 在一个实施例中,模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。

[0130] 在一个实施例中,还包括待填充词槽确定模块、业务词槽确定模块和业务执行模块;其中:待填充词槽确定模块,用于确定业务场景类型中业务流程对应的待填充词槽;业务词槽确定模块,用于将命名实体识别结果填充至待填充词槽中,得到业务场景类型对应的业务词槽;业务执行模块,用于若业务词槽满足流程执行条件,则基于业务词槽执行业务流程对应的业务。

[0131] 关于命名实体识别装置的具体限定可以参见上文中对于命名实体识别方法的限定,在此不再赘述。上述命名实体识别装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0132] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是服务器,其内部结构图可以如图6所示。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储实体识别模板树数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种命名实体识别方法。

[0133] 本领域技术人员可以理解,图6中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0134] 在一个实施例中,提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现以下步骤:

[0135] 获取待识别的语句,并确定语句对应的业务场景类型;

[0136] 查询与业务场景类型对应预设的实体识别模板树;

[0137] 将语句在实体识别模板树中进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板;

[0138] 基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。

[0139] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:对语句中各字符依次进行编码,得到语句对应的字符编码序列;基于字符编码序列在实体识别模板树中进行字符状态转移,得到状态转移结果;根据状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板。

[0140] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:获得模板匹配的模板匹配结果;若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定业务场景类型对应预设的模板权重映射关系;基于模板权重映射关系,从各候选实体识别模板中确定目标实体识别模板。

[0141] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:根据模板权重映射

关系,确定各候选实体识别模板分别对应的模板权重;将模板权重中最大的模板权重所对应的候选实体识别模板,确定为目标实体识别模板。

[0142] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:基于业务场景类型对应的各业务流程,确定由各模板字符按序组成的模板字符序列;按照各业务流程对应的模板特征类型,对各模板字符序列进行模板特征化处理,得到至少包括一个模板字符的模板特征字段;基于模板特征字段得到实体识别模板,并根据实体识别模板构建业务场景类型对应的实体识别模板树。

[0143] 在一个实施例中,模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。

[0144] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:确定业务场景类型中业务流程对应的待填充词槽;将命名实体识别结果填充至待填充词槽中,得到业务场景类型对应的业务词槽;若业务词槽满足流程执行条件,则基于业务词槽执行业务流程对应的业务。

[0145] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现以下步骤:

[0146] 获取待识别的语句,并确定语句对应的业务场景类型;

[0147] 查询与业务场景类型对应预设的实体识别模板树;

[0148] 将语句在实体识别模板树中进行模板匹配,获得表示匹配成功的模板匹配结果所对应的目标实体识别模板;

[0149] 基于目标实体识别模板对语句进行命名实体识别处理,得到语句对应的命名实体识别结果。

[0150] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:对语句中各字符依次进行编码,得到语句对应的字符编码序列;基于字符编码序列在实体识别模板树中进行字符状态转移,得到状态转移结果;根据状态转移结果得到模板匹配结果,并确定匹配成功的模板匹配结果所对应的目标实体识别模板。

[0151] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:获得模板匹配的模板匹配结果;若表示匹配成功的模板匹配结果所对应的候选实体识别模板的数量大于一,则确定业务场景类型对应预设的模板权重映射关系;基于模板权重映射关系,从各候选实体识别模板中确定目标实体识别模板。

[0152] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:根据模板权重映射关系,确定各候选实体识别模板分别对应的模板权重;将模板权重中最大的模板权重所对应的候选实体识别模板,确定为目标实体识别模板。

[0153] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:基于业务场景类型对应的各业务流程,确定由各模板字符按序组成的模板字符序列;按照各业务流程对应的模板特征类型,对各模板字符序列进行模板特征化处理,得到至少包括一个模板字符的模板特征字段;基于模板特征字段得到实体识别模板,并根据实体识别模板构建业务场景类型对应的实体识别模板树。

[0154] 在一个实施例中,模板特征字段包括省略词特征字段、词槽特征字段、关键词特征字段和符合特征字段中的至少一种。

[0155] 在一个实施例中, 计算机程序被处理器执行时还实现以下步骤: 确定业务场景类型中业务流程对应的待填充词槽; 将命名实体识别结果填充至待填充词槽中, 得到业务场景类型对应的业务词槽; 若业务词槽满足流程执行条件, 则基于业务词槽执行业务流程对应的业务。

[0156] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程, 是可以通过计算机程序来指令相关的硬件来完成, 所述的计算机程序可存储于一非易失性计算机可读取存储介质中, 该计算机程序在执行时, 可包括如上述各方法的实施例的流程。其中, 本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用, 均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器 (Read-Only Memory, ROM)、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器 (Random Access Memory, RAM) 或外部高速缓冲存储器。作为说明而非局限, RAM 可以是多种形式, 比如静态随机存取存储器 (Static Random Access Memory, SRAM) 或动态随机存取存储器 (Dynamic Random Access Memory, DRAM) 等。

[0157] 以上实施例的各技术特征可以进行任意的组合, 为使描述简洁, 未对上述实施例中的各个技术特征所有可能的组合都进行描述, 然而, 只要这些技术特征的组合不存在矛盾, 都应当认为是本说明书记载的范围。

[0158] 以上所述实施例仅表达了本申请的几种实施方式, 其描述较为具体和详细, 但并不能因此而理解为对发明专利范围的限制。应当指出的是, 对于本领域的普通技术人员来说, 在不脱离本申请构思的前提下, 还可以做出若干变形和改进, 这些都属于本申请的保护范围。因此, 本申请专利的保护范围应以所附权利要求为准。

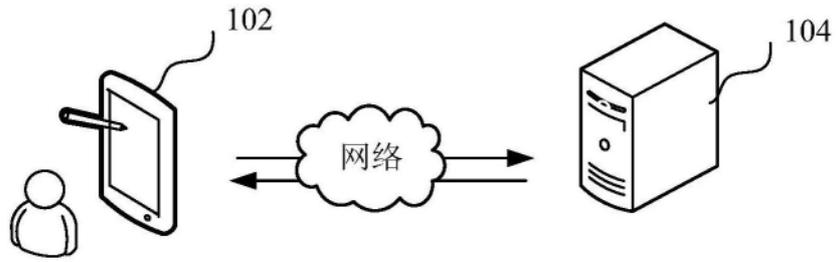


图1

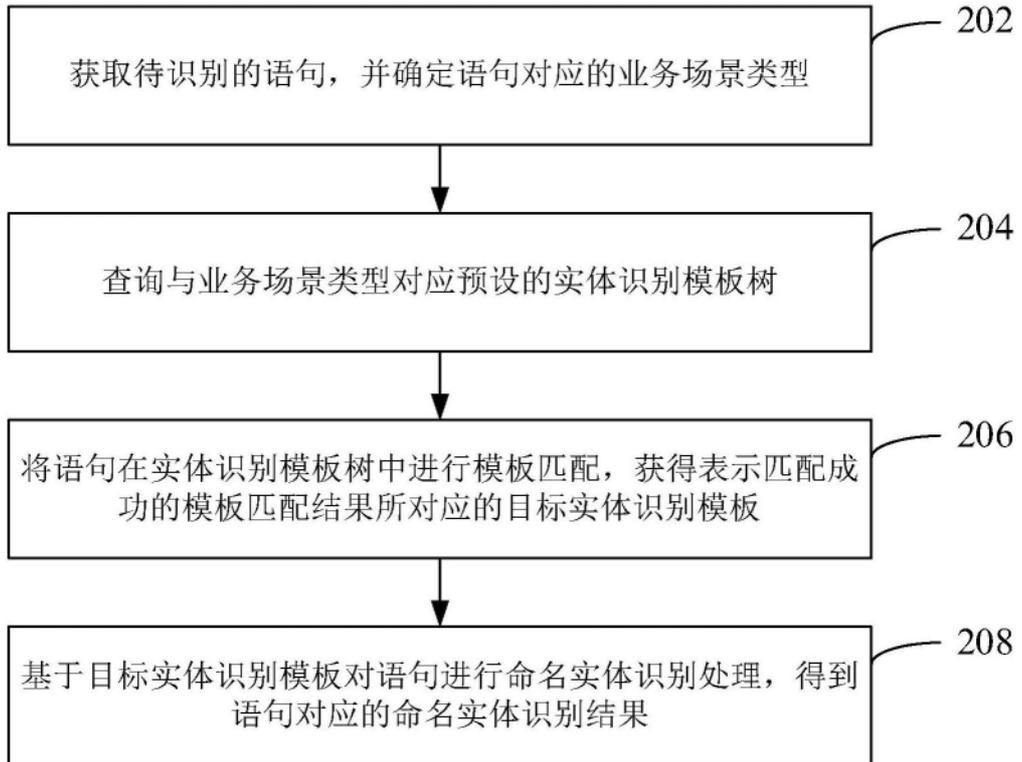


图2

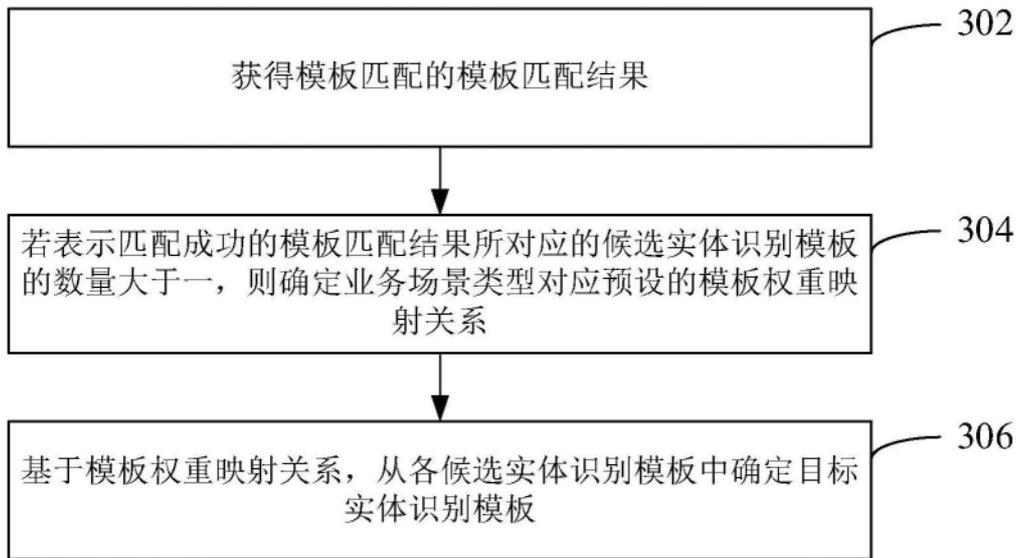


图3

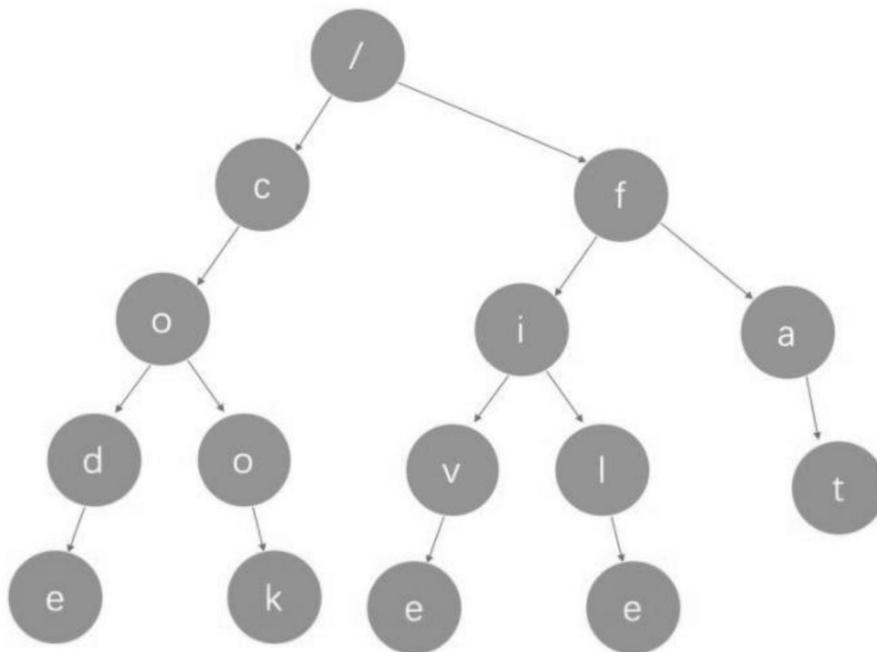


图4

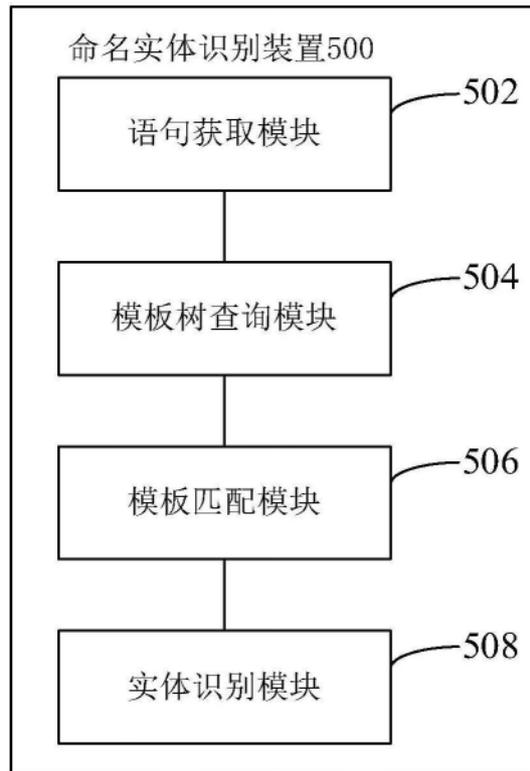


图5

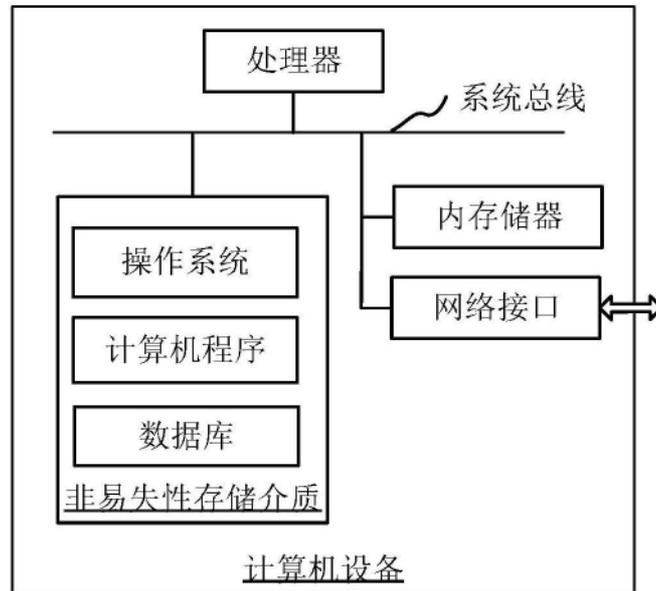


图6