



(12) 发明专利申请

(10) 申请公布号 CN 112632260 A

(43) 申请公布日 2021.04.09

(21) 申请号 202011625530.7

G06F 40/284 (2020.01)

(22) 申请日 2020.12.30

G06F 40/295 (2020.01)

G06F 40/35 (2020.01)

(71) 申请人 平安证券股份有限公司

地址 518000 广东省深圳市福田区福田街
道益田路5023号平安金融中心B座第
22-25层

(72) 发明人 何博

(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

代理人 高杰 于志光

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 16/33 (2019.01)

G06F 16/36 (2019.01)

G06F 40/242 (2020.01)

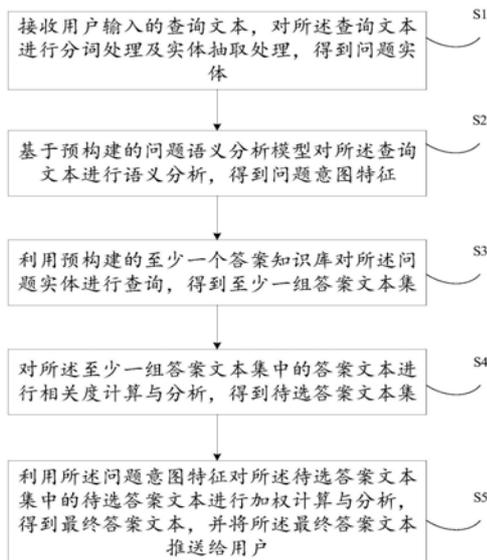
权利要求书3页 说明书12页 附图3页

(54) 发明名称

智能问答方法、装置、电子设备及计算机可
读存储介质

(57) 摘要

本发明涉及大数据技术,揭露一种智能问答方法,包括:接收查询文本,对所述查询文本进行分词处理及实体抽取处理,得到问题实体;对所述查询文本进行语义分析,得到问题意图特征;利用预构建的至少一个答案知识库对所述问题实体进行查询,得到至少一组答案文本集;对所述答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集;利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,并将所述最终答案文本推送给用户。本发明还涉及区块链技术,所述答案知识库可以部署在区块链中。本发明还提出一种智能问答装置、电子设备以及一种计算机可读存储介质。本发明可以提高智能问答的准确率。



1. 一种智能问答方法,其特征在于,所述方法包括:

接收用户输入的查询文本,对所述查询文本进行分词处理及实体抽取处理,得到问题实体;

基于预构建的问题语义分析模型对所述查询文本进行语义分析,得到问题意图特征;

利用预构建的至少一个答案知识库对所述问题实体进行查询,得到至少一组答案文本集;

对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集;

利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,并将所述最终答案文本推送给用户。

2. 如权利要求1所述的智能问答方法,其特征在于,所述对所述查询文本进行分词处理,包括:

字段预设步骤:利用预构建的分词词典对所述查询文本进行分词操作,其中,所述预构建的分词词典中最大词条所含的字符数目为 n 个,若所述查询文本的所有字符数目小于 n 个,则将所述查询文本作为匹配字段,若所述查询文本的所有字符数目等于或大于 n 个,则将所述查询文本的前 n 个字符作为所述匹配字段;

分词匹配步骤:利用所述匹配字段在所述分词词典中进行遍历匹配;若所述分词词典中含有与所述匹配字段相同的词语,确定匹配成功,从所述查询文本中分出所述匹配字段,执行下述顺序匹配步骤;若所述分词词典中不含有与所述匹配字段相同的词语,则执行下述字段更新步骤;

字段更新步骤:若所述匹配字段中包含的字符数目等于1,从所述查询文本中分出所述匹配字段,执行下述顺序匹配步骤;若所述匹配字段中包含的字符数目大于1,将所述匹配字段的最后一个字符剔除,用剔除后的字段替换更新所述匹配字段,返回所述分词匹配步骤;

顺序匹配步骤:若所述查询文本中所述匹配字段后的字符数目小于 n 个,则以所述查询文本中所述匹配字段后的所有字符替换更新所述匹配字段,同时返回所述分词匹配步骤;若所述查询文本中所述匹配字段后的字符数目不小于 n 个,则以所述查询文本匹配字段后的顺序的 n 个字符替换更新所述匹配字段,同时返回所述分词匹配步骤;若所述查询文本匹配字段后的字符数目等于0个,则停止分词处理。

3. 如权利要求1所述的智能问答方法,其特征在于,所述对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集,包括:

将所述答案文本集中的答案文本转化为答案文本向量;

将所述查询文本转化为问题文本向量;

计算所述答案文本向量与所述问题文本向量的相关度值,汇总所有相关度值得到相关度文本集;

将所述相关度文本集中相关度值进行排序,得到最大相关度值;

在所述答案文本集中选取所述最大相关度值对应的答案文本向量,并将选取的答案文本向量对应的答案文本作为该答案文本集对应的待选答案文本;

汇总所述待选答案文本,得到所述待选答案文本集。

4. 如权利要求3所述的智能问答方法,其特征在于,所述利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,包括:

将所述问题意图特征转化为问题意图特征向量;

计算所述问题意图特征向量与每个待选答案文本对应的答案文本向量的相关度值,得到待选答案文本相关度;

计算所述至少一组答案文本集中的答案文本的总数目;

计算所述至少一组答案文本集中每组答案文本集中的答案文本数目与所述答案文本的总数目的比值,得到每组答案文本集权重;

将所述待选答案文本所在的答案文本集对应的答案文本集权重与该待选答案文本对应的待选答案文本相关度相乘,得到待选答案文本权重值;

汇总所述待选答案文本权重值得到待选答案文本权重值集;

选取所述待选答案文本权重值集中最大待选答案文本权重值对应的待选答案文本作为所述最终答案文本。

5. 如权利要求3所述的智能问答方法,其特征在于,所述将所述查询文本转化为问题文本向量,包括:

获取预设尺寸的维度窗口;

将所述查询文本输入至所述维度窗口中,生成维度查询文本;

利用预设编码算法对所述维度查询文本进行编码,生成所述问题文本向量。

6. 如权利要求3所述的智能问答方法,其特征在于,所述将所述答案文本集中的答案文本转化为答案文本向量,包括:

利用预先训练的字符向量词典获取所述答案文本集中每个答案文本包含的每个字符的字符向量;

计算每个答案文本中所有字符的字符向量的算术平均值作为每个答案文本对应的答案文本向量。

7. 一种智能问答装置,其特征在于,所述装置包括:

实体抽取模块,用于接收用户输入的查询文本,对所述查询文本进行分词处理及实体抽取处理,得到问题实体;

语义分析模块,用于基于预构建的问题语义分析模型对所述查询文本进行语义分析,得到问题意图特征;

答案筛选模块,用于利用预构建的至少一个答案知识库对所述问题实体进行查询,得到至少一组答案文本集;对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集;利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,并将所述最终答案文本推送给用户。

8. 如权利要求7所述的智能问答装置,其特征在于,所述答案筛选模块通过下述手段对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集,包括:

将所述答案文本集中的答案文本转化为答案文本向量;

将所述查询文本转化为问题文本向量;

计算所述答案文本向量与所述问题文本向量的相关度值,汇总所有相关度值得到相关

度文本集；

将所述相关度文本集中相关度值进行排序,得到最大相关度值；

在所述答案文本集中选取所述最大相关度值对应的答案文本向量,并将选取的答案文本向量对应的答案文本作为该答案文本集对应的待选答案文本；

汇总所述待选答案文本,得到所述待选答案文本集。

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指令,所述指令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6中任一所述的智能问答方法。

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6中任一所述的智能问答方法。

智能问答方法、装置、电子设备及计算机可读存储介质

技术领域

[0001] 本发明涉及大数据领域,尤其涉及一种智能问答的方法、装置、电子设备及计算机可读存储介质。

背景技术

[0002] 随着机器学习和深度学习、自然语言的突飞猛进,智能问答已经成为当今最热门的人工智能的研究和应用方向,例如:苹果公司的Siri,腾讯机器人平台,Amazon的Echo,微软的Luis.AI等,但是目前的智能问答方法,只在单一的业务知识库中查询答案,且未对问题意图进行识别,导致智能问答的准确度低。

发明内容

[0003] 本发明提供一种智能问答方法、装置、电子设备及计算机可读存储介质,其主要目的在于提高了智能问答的准确率。

[0004] 为实现上述目的,本发明提供的一种智能问答方法,包括:

[0005] 接收用户输入的查询文本,对所述查询文本进行分词处理及实体抽取处理,得到问题实体;

[0006] 基于问题语义分析模型对所述查询文本进行语义分析,得到问题意图特征;

[0007] 利用预构建的至少一个答案知识库对所述问题实体进行查询,得到至少一组答案文本集;

[0008] 对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集;

[0009] 利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,并将所述最终答案文本推送给用户。

[0010] 可选地,所述对所述查询文本进行分词处理,包括:

[0011] 字段预设步骤:利用预构建的分词词典对所述查询文本进行分词操作,其中,所述预构建的分词词典中最大词条所含的字符数目为 n 个,若所述查询文本的所有字符数目小于 n 个,则将所述查询文本作为匹配字段,若所述查询文本的所有字符数目等于或大于 n 个,则将所述查询文本的前 n 个字符作为所述匹配字段;

[0012] 分词匹配步骤:利用所述匹配字段在所述分词词典中进行遍历匹配;若所述分词词典中含有与所述匹配字段相同的词语,确定匹配成功,从所述查询文本中分出所述匹配字段,执行下述顺序匹配步骤;若所述分词词典中不含有与所述匹配字段相同的词语,则执行下述字段更新步骤;

[0013] 字段更新步骤:若所述匹配字段中包含的字符数目等于1,从所述查询文本中分出所述匹配字段,执行下述顺序匹配步骤;若所述匹配字段中包含的字符数目大于1,将所述匹配字段的最后一个字符剔除,用剔除后的字段替换更新所述匹配字段,返回所述分词匹配步骤;

[0014] 顺序匹配步骤:若所述查询文本中所述匹配字段后的字符数目小于n个,则以所述查询文本中所述匹配字段后的所有字符替换更新所述匹配字段,同时返回所述分词匹配步骤;若所述查询文本中所述匹配字段后的字符数目不小于n个,则以所述查询文本匹配字段后的顺序的n个字符替换更新所述匹配字段,同时返回所述分词匹配步骤;若所述查询文本匹配字段后的字符数目等于0个,则停止分词处理。

[0015] 可选地,所述对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集,包括:

[0016] 将所述答案文本集中的答案文本转化为答案文本向量;

[0017] 将所述查询文本转化为问题文本向量;

[0018] 计算所述答案文本向量与所述问题文本向量的相关度值,汇总所有相关度值得到相关度文本集;

[0019] 将所述相关度文本集中相关度值进行排序,得到最大相关度值;

[0020] 在所述答案文本集中选取所述最大相关度值对应的答案文本向量,并将选取的答案文本向量对应的答案文本作为该答案文本集对应的待选答案文本;

[0021] 汇总所述待选答案文本,得到所述待选答案文本集。

[0022] 可选地,所述利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,包括:

[0023] 将所述问题意图特征转化为问题意图特征向量;

[0024] 计算所述问题意图特征向量与每个待选答案文本对应的答案文本向量的相关度值,得到待选答案文本相关度;

[0025] 计算所述至少一组答案文本集中的答案文本的总数目;

[0026] 计算所述至少一组答案文本集中每组答案文本集中的答案文本数目与所述答案文本的总数目的比值,得到每组答案文本集权重;

[0027] 将所述待选答案文本所在的答案文本集对应的答案文本集权重与该待选答案文本对应的待选答案文本相关度相乘,得到待选答案文本权重值;

[0028] 汇总所述待选答案文本权重值得到待选答案文本权重值集;

[0029] 选取所述待选答案文本权重值集中最大待选答案文本权重值对应的待选答案文本作为所述最终答案文本。

[0030] 可选地,所述将所述查询文本转化为问题文本向量,包括:

[0031] 获取预设尺寸的维度窗口;

[0032] 将所述查询文本输入至所述维度窗口中,生成维度查询文本;

[0033] 利用预设编码算法对所述维度查询文本进行编码,生成所述问题文本向量。

[0034] 可选地,所述将所述答案文本集中的答案文本转化为答案文本向量,包括:

[0035] 利用预先训练的字向量词典获取所述答案文本集中每个答案文本包含的每个字符的字向量;

[0036] 计算每个答案文本中所有字符的字向量的算术平均值作为每个答案文本对应的答案文本向量。

[0037] 为了解决上述问题,本发明还提供一种智能问答装置,所述装置包括:

[0038] 实体抽取模块,用于接收用户输入的查询文本,对所述查询文本进行分词处理及

实体抽取处理,得到问题实体;

[0039] 语义分析模块,用于基于预构建的问题语义分析模型对所述查询文本进行语义分析,得到问题意图特征;

[0040] 答案筛选模块,用于利用预构建的至少一个答案知识库对所述问题实体进行查询,得到至少一组答案文本集;对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集;利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,并将所述最终答案文本推送给用户。

[0041] 可选地,所述答案筛选模块对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集,包括:

[0042] 答案文本向量模块:将所述答案文本集中的答案文本转化为答案文本向量;

[0043] 问题文本向量模块:将所述查询文本转化为问题文本向量;

[0044] 相关度文本集模块:计算所述答案文本向量与所述问题文本向量的相关度值,汇总所有相关度值得到相关度文本集;

[0045] 最大相关度值模块:将所述相关度文本集中相关度值进行排序,得到最大相关度值;

[0046] 待选答案文本模块:在所述答案文本集中选取所述最大相关度值对应的答案文本向量,并将选取的答案文本向量对应的答案文本作为该答案文本集对应的待选答案文本;

[0047] 待选答案文本集模块:汇总所述待选答案文本,得到所述待选答案文本集。

[0048] 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:

[0049] 存储器,存储至少一个指令;及

[0050] 处理器,执行所述存储器中存储的指令以实现上述所述的智能问答方法。

[0051] 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,包括存储数据区和存储程序区,存储数据区存储根据区块链节点的使用所创建的数据,存储程序区存储有计算机程序,所述计算机可读存储介质中存储有至少一个指令,所述至少一个指令被电子设备中的处理器执行以实现上述所述的智能问答方法。

[0052] 本发明实施例对所述查询文本进行分词处理及实体抽取处理,剔除了所述查询文本中无关词语对后续答案查询的影响;对所述查询文本进行语义分析,得到问题意图特征,利用问题意图特征可以对答案进行更准确的筛选;利用预构建的至少一个答案知识库对所述问题实体进行查询,使问题查询的更加全面,更加准确;对所述至少一组答案文本集中的答案文本进行相关度计算与分析,利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,提高了答案筛选的准确度;将所述查询文本在不同业务答案知识库中查询,利用问题意图特征对答案进行进一步筛选,提高了智能问答的准确率。

附图说明

[0053] 图1为本发明一实施例提供的智能问答方法的流程示意图;

[0054] 图2为本发明一实施例提供的智能问答方法中分词处理的流程示意图;

[0055] 图3为本发明一实施例提供的智能问答装置的模块示意图;

[0056] 图4为本发明一实施例提供的实现智能问答方法的电子设备的内部结构示意图;

[0057] 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

[0058] 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。

[0059] 本发明提供一种智能问答方法。参照图1所示,为本发明一实施例提供的智能问答方法的流程示意图。该方法可以由一个装置执行,该装置可以由软件和/或硬件实现。

[0060] 在本实施例中,智能问答方法包括:

[0061] S1:接收用户输入的查询文本,对所述查询文本进行分词处理及实体抽取处理,得到问题实体;

[0062] 本发明实施例中所述查询文本为用户输入的需要咨询的问题,例如:“今天天气怎么样?”、“三年期用户存款的利率是多少?”,可通过网页的文本输入模块获取所述查询文本。

[0063] 进一步地,本发明实施例对所述查询文本进行分词处理及实体抽取处理,得到问题实体,其中,所述分词处理包括:

[0064] S10:利用预构建的分词词典对所述查询文本进行分词操作,其中,所述预构建的分词词典中最大词条所含的字符数目为 n 个;

[0065] S11:判断所述查询文本的所有字符数目是否小于 n 个;

[0066] S12:若所述查询文本的所有字符数目小于 n 个,则将所述查询文本作为匹配字段,

[0067] S13:若所述查询文本的所有字符数目等于或大于 n 个,则将所述查询文本的前 n 个字符作为所述匹配字段;

[0068] S20:利用所述匹配字段在所述分词词典中进行遍历匹配;

[0069] S21:判断所述分词词典中是否含有与所述匹配字段相同的词语;

[0070] S22:若所述分词词典中含有与所述匹配字段相同的词语,确定匹配成功,从所述查询文本中分出所述匹配字段,执行S40;

[0071] S23:若所述分词词典中不含有与所述匹配字段相同的词语,则执行S30;

[0072] S30:判断所述匹配字段中包含的字符数目是否大于1;

[0073] S31:若所述匹配字段中包含的字符数目等于1,从所述查询文本中分出所述匹配字段,执行S40;

[0074] S32:若所述匹配字段中包含的字符数目大于1,将所述匹配字段的最后一个字符剔除,用剔除后的字段替换更新所述匹配字段,返回S20;

[0075] S40:判断所述查询文本中所述匹配字段后的字符数目是否等于0个;

[0076] S41:若所述查询文本匹配字段后的字符数目不等于0个,判断所述查询文本中所述匹配字段后的字符数目是否小于 n 个;

[0077] S411:若所述查询文本中所述匹配字段后的字符数目小于 n 个,则以所述查询文本中所述匹配字段后的所有字符替换更新所述匹配字段,同时返回S20;

[0078] S412:若所述查询文本中所述匹配字段后的字符数目不小于 n 个,则以所述查询文本匹配字段后的顺序的 n 个字符替换更新所述匹配字段,同时返回S20;

[0079] S42:若所述查询文本匹配字段后的字符数目等于0个,则停止分词处理。

[0080] 通过上述实施方式,以正向切分的方式进行分词处理,分词效果更准确且分词速

度更快。

[0081] 较佳地,本发明实施例中所述实体抽取处理包括:利用命名实体识别技术(Named Entity Recognition,简称NER)对所述分词处理后的查询文本进行识别,得到问题实体。其中,利用命名实体识别技术能够识别出查询文本中具有特定意义的实体(如,人名、地名、机构名、专有名词等)。例如:所述分词处理后的查询文本为“美国总统”“是”“谁”,利用所述命名实体识别技术对所述查询文本进行识别可得到问题实体为“美国总统”。

[0082] 通过所述分词处理及实体抽取处理可以剔除所述查询文本中的无意义词语,避免无意义词语对后续答案查询的影响。

[0083] S2、基于预构建的问题语义分析模型对所述查询文本进行语义分析,得到问题意图特征;

[0084] 本发明实施例中,利用预构建的语义分析模型对所述查询文本进行语义分析,得到所述问题意图特征。例如:查询文本为“三年期用户存款的利率是多少?”,得到所述查询文本的问题意图特征为“存款意图”;查询文本为“哪种基金的年化利率最高”,得到所述查询文本的问题意图特征为“基金购买意图”。

[0085] 较佳地,所述语义分析模型可以利用BERT(Bidirectional Encoder Representations from Transformers,双向编码器)算法进行构建,利用历史用户查询文本集作为训练集,利用标记过的历史用户查询文本集作为标签集,完成所述语义分析模型的训练。其中,所述历史用户查询文本集为多个历史用户查询文本组成的集合。

[0086] 通过对所述查询文本进行语义分析得到问题意图特征,利用问题意图特征可以对答案进行更准确的筛选。

[0087] S3、利用预构建的至少一个答案知识库对所述问题实体进行查询,得到至少一组答案文本集;

[0088] 本发明实施例中,为了保证答案的准确度,预构建不同的答案知识库,如业务答案知识库,其中,所述业务答案知识库是限定业务类别的业务知识组成的数据库。例如:金融行业涵盖多个业务分类,包括:基金业务、存款业务、投资业务等多个业务领域,本发明实施例将基金业务相关的知识汇总存储到数据库得到基金业务答案知识库、将存款相关的知识汇总存储到数据库得到存款业务答案知识库、将投资相关的知识汇总存储到数据库得到投资业务答案知识库。

[0089] 本发明的另一实施例中所述答案知识库为限定领域知识组成,并可以部署在区块链上。

[0090] 进一步地,利用预构建的至少一个答案知识库对所述问题实体进行查询,得到至少一组答案文本集;以业务答案知识库为例,共有基金业务答案知识库、存款业务答案知识库、投资业务答案知识库三种预构建的业务答案知识库,分别在这三种业务知识库中对所述问题实体进行查询得到多组业务答案文本集,在基金业务答案知识库查询得到基金业务答案文本集、在存款业务答案知识库查询得到存款业务答案文本集、在投资业务答案知识库查询得到投资业务答案文本集,共3组业务答案文本集。

[0091] S4、对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集;

[0092] 本发明实施例中,为了筛选出更准确的答案,将所述答案文本集中的答案文本转

化为答案文本向量,将所述查询文本转化为问题文本向量。

[0093] 较佳地,利用预先训练的字向量词典获取所述答案文本集中每个答案文本包含的每个字符的字向量;计算每个答案文本中所有字符的字向量的算术平均值作为每个答案文本对应的答案文本向量。

[0094] 进一步地,所述算术平均值的计算公式如下:

$$[0095] \quad W = \frac{a_1 + a_2 + \dots + a_{n-1} + a_n}{n}$$

[0096] 其中, a_1 至 a_n 表示每个答案文本中每个字符的字向量, n 表示该答案文本中字符的个数, W 表示该答案文本的答案文本向量。

[0097] 进一步地,获取预设尺寸的维度窗口;将所述查询文本输入至所述维度窗口中,生成维度查询文本;利用预设编码算法对所述维度查询文本进行编码,生成问题文本向量。

[0098] 例如,维度窗口为 $k*k$ 的维度窗口,其中, k 为大于或者等于1的正整数。

[0099] 较佳地,本实施例中,所述预设编码算法可以为哈夫曼编码算法。

[0100] 进一步地,计算所述答案文本向量与所述问题文本向量的相关度值,汇总所有相关度值得到相关度文本集;将所述相关度文本集中相关度值进行排序,得到最大相关度值;在所述答案文本集中选取所述最大相关度值对应的答案文本向量,并将选取的答案文本向量对应的答案文本作为该答案文本集对应的待选答案文本。

[0101] 详细地,所述相关度值计算可用如下公式:

$$[0102] \quad sim(x_i, y) = \frac{\sum x_i y}{\sqrt{\sum x_i^2} \sqrt{\sum y}}$$

[0103] 其中, x_i 为所述答案文本向量, i 为正整数, y 为所述问题文本向量, $sim(x_i, y)$ 为所述答案文本向量与所述问题文本向量的相关度值。

[0104] 进一步地,汇总所述待选答案文本,得到所述待选答案文本集。

[0105] S5、利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,并将所述最终答案文本推送给用户。

[0106] 详细地,本发明实施例为了进一步对答案进行筛选,保证答案的准确度,本发明实施例利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,包括:

[0107] S51、将所述问题意图特征转化为问题意图特征向量;

[0108] S52、计算所述问题意图特征向量与每个待选答案文本对应的答案文本向量的相关度值,得到待选答案文本相关度;

[0109] S53、计算所述至少一组答案文本集中的答案文本的总数目;

[0110] S54、计算所述至少一组答案文本集中每组答案文本集中的答案文本数目与所述答案文本的总数目的比值,得到每组答案文本集权重;

[0111] S55、将所述待选答案文本所在的答案文本集对应的答案文本集权重与该待选答案文本对应的待选答案文本相关度相乘,得到待选答案文本权重值,汇总所述待选答案文本权重值得到待选答案文本权重值集;

[0112] S56、选取所述待选答案文本权重值集中最大待选答案文本权重值对应的待选答案文本作为所述最终答案文本。

[0113] 进一步地,本发明实施例将最终答案文本推送给用户,同时获取用户对所述最终答案文本是否认可的反馈信息,并基于所述反馈信息对所述知识库中的数据进行对应更新,实现对所述知识库的纠错处理。

[0114] 本发明实施例对所述查询文本进行分词处理及实体抽取处理,剔除了所述查询文本中无关词语对后续答案查询的影响;对所述查询文本进行语义分析,得到问题意图特征,利用问题意图特征可以对答案进行更准确的筛选;利用预构建的至少一个答案知识库对所述问题实体进行查询,使问题查询的更加全面,更加准确;对所述至少一组答案文本集中的答案文本进行相关度计算与分析,利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,提高了答案筛选的准确度;将所述查询文本在不同业务答案知识库中查询,利用问题意图特征对答案进行进一步筛选,提高了智能问答的准确率。

[0115] 如图2所示,是本发明智能问答装置的功能模块图。

[0116] 本发明所述智能问答装置100可以安装于电子设备中。根据实现的功能,所述智能问答装置可以包括实体抽取模块101、语义分析模块102、答案筛选模块103。本发明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。

[0117] 在本实施例中,关于各模块/单元的功能如下:

[0118] 所述实体抽取模块101用于接收用户输入的查询文本,对所述查询文本进行分词处理及实体抽取处理,得到问题实体。

[0119] 本发明实施例中所述查询文本为用户输入的需要咨询的问题,例如:“今天天气怎么样?”、“三年期用户存款的利率是多少?”,可通过网页的文本输入模块获取所述查询文本。

[0120] 进一步地,本发明实施例所述实体抽取模块101对所述查询文本进行分词处理及实体抽取处理,得到问题实体,其中,所述实体抽取模块101通过下述手段进行分词处理:

[0121] 字段预设步骤:利用预构建的分词词典对所述查询文本进行分词操作,其中,所述预构建的分词词典中最大词条所含的字符数目为 n 个,若所述查询文本的所有字符数目小于 n 个,则将所述查询文本作为匹配字段,若所述查询文本的所有字符数目等于或大于 n 个,则将所述查询文本的前 n 个字符作为所述匹配字段;

[0122] 分词匹配步骤:利用所述匹配字段在所述分词词典中进行遍历匹配;若所述分词词典中含有与所述匹配字段相同的词语,确定匹配成功,从所述查询文本中分出所述匹配字段,执行下述顺序匹配步骤;若所述分词词典中不含有与所述匹配字段相同的词语,则执行下述字段更新步骤;

[0123] 字段更新步骤:若所述匹配字段中包含的字符数目等于1,从所述查询文本中分出所述匹配字段,执行下述顺序匹配步骤;若所述匹配字段中包含的字符数目大于1,将所述匹配字段的最后一个字符剔除,用剔除后的字段替换更新所述匹配字段,返回所述分词匹配步骤;

[0124] 顺序匹配步骤:若所述查询文本中所述匹配字段后的字符数目小于 n 个,则以所述

查询文本中所述匹配字段后的所有字符替换更新所述匹配字段,同时返回所述分词匹配步骤;若所述查询文本中所述匹配字段后的字符数目不小于n个,则以所述查询文本匹配字段后的顺序的n个字符替换更新所述匹配字段,同时返回所述分词匹配步骤;若所述查询文本匹配字段后的字符数目等于0个,则停止分词处理。

[0125] 通过上述实施方式,所述实体抽取模块101以正向切分的方式进行分词处理,分词效果更准确且分词速度更快。

[0126] 较佳地,本发明实施例中所述实体抽取模块101进行实体抽取处理包括:利用命名实体识别技术(Named Entity Recognition,简称NER)对所述分词处理后的查询文本进行识别,得到问题实体。其中,利用命名实体识别技术能够识别出查询文本中具有特定意义的实体(如,人名、地名、机构名、专有名词等)。例如:所述分词处理后的查询文本为“美国总统”“是”“谁”,利用所述命名实体识别技术对所述查询文本进行识别可得到问题实体为“美国总统”。

[0127] 通过所述分词处理及实体抽取处理可以剔除所述查询文本中无意义词语,避免无意义词语对后续答案查询的影响。

[0128] 所述语义分析模块102用于基于预构建的问题语义分析模型对所述查询文本进行语义分析,得到问题意图特征。

[0129] 本发明实施例中,所述语义分析模块102利用预构建的语义分析模型对所述查询文本进行语义分析,得到所述问题意图特征。例如:查询文本为“三年期用户存款的利率是多少?”,得到所述查询文本的问题意图特征为“存款意图”;查询文本为“哪种基金的年化利率最高”,得到所述查询文本的问题意图特征为“基金购买意图”。

[0130] 较佳地,所述语义分析模型可以利用BERT(Bidirectional Encoder Representations from Transformers,双向编码器)算法进行构建,利用历史用户查询文本集作为训练集,利用标记过的历史用户查询文本集作为标签集,完成所述语义分析模型的训练。其中,所述历史用户查询文本集为多个历史用户查询文本组成的集合。

[0131] 通过对所述查询文本进行语义分析得到问题意图特征,利用问题意图特征可以对答案进行更准确的筛选。

[0132] 所述答案筛选模块103用于利用预构建的至少一个答案知识库对所述问题实体进行查询,得到至少一组答案文本集;对所述至少一组答案文本集中的答案文本进行相关度计算与分析,得到待选答案文本集;利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,并将所述最终答案文本推送给用户。

[0133] 本发明实施例中,为了保证答案的准确度,所述答案筛选模块103预构建不同的答案知识库,如业务答案知识库,其中,所述业务答案知识库是限定业务类别的业务知识组成的数据库。例如:金融行业涵盖多个业务分类,包括:基金业务、存款业务、投资业务等多个业务领域,本发明实施例将基金业务相关的知识汇总存储到数据库得到基金业务答案知识库、将存款相关的知识汇总存储到数据库得到存款业务答案知识库、将投资相关的知识汇总存储到数据库得到投资业务答案知识库。

[0134] 本发明的另一实施例中所述答案知识库为限定领域知识组成,并可以部署在区块链上。

[0135] 进一步地,所述答案筛选模块103利用预构建的至少一个答案知识库对所述问题

实体进行查询,得到至少一组答案文本集;以业务答案知识库为例,共有基金业务答案知识库、存款业务答案知识库、投资业务答案知识库三种预构建的业务答案知识库,分别在这三种业务知识库中对所述问题实体进行查询得到多组业务答案文本集,在基金业务答案知识库查询得到基金业务答案文本集、在存款业务答案知识库查询得到存款业务答案文本集、在投资业务答案知识库查询得到投资业务答案文本集,共3组业务答案文本集。

[0136] 本发明实施例中,为了筛选出更准确的答案,所述答案筛选模块103将所述答案文本集中的答案文本转化为答案文本向量,将所述查询文本转化为问题文本向量。

[0137] 较佳地,所述答案筛选模块103通过下述手段得到答案文本向量:

[0138] 利用预先训练的字向量词典获取所述答案文本集中每个答案文本包含的每个字符的字向量;

[0139] 计算每个答案文本中所有字符的字向量的算术平均值作为每个答案文本对应的答案文本向量。

[0140] 进一步地,所述算术平均值的计算公式如下:

$$[0141] \quad W = \frac{a_1 + a_2 + \dots + a_{n-1} + a_n}{n}$$

[0142] 其中, a_1 至 a_n 表示每个答案文本中每个字符的字向量, n 表示该答案文本中字符的个数, W 表示该答案文本的答案文本向量。

[0143] 进一步地,所述答案筛选模块103通过下述手段得到问题文本向量:

[0144] 获取预设尺寸的维度窗口;

[0145] 将所述查询文本输入至所述维度窗口中,生成维度查询文本;

[0146] 利用预设编码算法对所述维度查询文本进行编码,生成问题文本向量。

[0147] 例如,维度窗口为 $k*k$ 的维度窗口,其中, k 为大于或者等于1的正整数。

[0148] 较佳地,本实施例中,所述预设编码算法可以为哈夫曼编码算法。

[0149] 进一步地,所述答案筛选模块103通过下述手段获得待选答案文本:

[0150] 计算所述答案文本向量与所述问题文本向量的相关度值,汇总所有相关度值得到相关度文本集;

[0151] 将所述相关度文本集中相关度值进行排序,得到最大相关度值;

[0152] 在所述答案文本集中选取所述最大相关度值对应的答案文本向量,并将选取的答案文本向量对应的答案文本作为该答案文本集对应的待选答案文本。

[0153] 详细地,所述相关度值计算可用如下公式:

$$[0154] \quad \text{sim}(x_i, y) = \frac{\sum x_i y}{\sqrt{\sum x_i^2} \sqrt{\sum y}}$$

[0155] 其中, x_i 为所述答案文本向量, i 为正整数, y 为所述问题文本向量, $\text{sim}(x_i, y)$ 为所述答案文本向量与所述问题文本向量的相关度值。

[0156] 进一步地,所述答案筛选模块103汇总所述待选答案文本,得到所述待选答案文本集。

[0157] 详细地,本发明实施例为了进一步对答案进行筛选,保证答案的准确度,本发明实

施例所述答案筛选模块103通过下述手段利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,包括:

[0158] 将所述问题意图特征转化为问题意图特征向量;

[0159] 计算所述问题意图特征向量与每个待选答案文本对应的答案文本向量的相关度值,得到待选答案文本相关度;

[0160] 计算所述至少一组答案文本集中的答案文本的总数目;

[0161] 计算所述至少一组答案文本集中每组答案文本集中的答案文本数目与所述答案文本的总数目的比值,得到每组答案文本集权重;

[0162] 将所述待选答案文本所在的答案文本集对应的答案文本集权重与该待选答案文本对应的待选答案文本相关度相乘,得到待选答案文本权重值,汇总所述待选答案文本权重值得到待选答案文本权重值集;

[0163] 选取所述待选答案文本权重值集中最大待选答案文本权重值对应的待选答案文本作为所述最终答案文本。

[0164] 进一步地,本发明实施例所述答案筛选模块103将最终答案文本推送给用户,同时获取用户对所述最终答案文本是否认可的反馈信息,并基于所述反馈信息对所述知识库中的数据进行对应更新,实现对所述知识库的纠错处理。

[0165] 本发明实施例所述实体抽取模块101对所述查询文本进行分词处理及实体抽取处理,剔除了所述查询文本中无关词语对后续答案查询的影响;所述语义分析模块102对所述查询文本进行语义分析,得到问题意图特征,利用问题意图特征可以对答案进行更准确的筛选;所述答案筛选模块103利用预构建的至少一个答案知识库对所述问题实体进行查询,使问题查询的更加全面,更加准确;所述答案筛选模块103对所述至少一组答案文本集中的答案文本进行相关度计算与分析,所述答案筛选模块103利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算与分析,得到最终答案文本,提高了答案筛选的准确度;将所述查询文本在不同业务答案知识库中查询,利用问题意图特征对答案进行进一步筛选,提高了智能问答的准确率。

[0166] 如图3所示,是本发明实现智能问答方法的电子设备的结构示意图。

[0167] 所述电子设备1可以包括处理器10、存储器11和总线,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,如智能问答程序。

[0168] 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1的内部存储单元,例如该电子设备1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1的外部存储设备,例如电子设备1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1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1的应用软件及各类数据,例如智能问答程序的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

[0169] 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 (例如智能问答程序等), 以及调用存储在所述存储器11内的数据, 以执行电子设备1的各种功能和处理数据。

[0170] 所述总线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。

[0171] 图3仅示出了具有部件的电子设备的, 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1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。

[0172] 例如, 尽管未示出, 所述电子设备1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、电源故障检测电路、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1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。

[0173] 进一步地, 所述电子设备1还可以包括网络接口, 可选地, 所述网络接口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1与其他电子设备之间建立通信连接。

[0174] 可选地, 该电子设备1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1中处理的信息以及用于显示可视化的用户界面。

[0175] 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。

[0176] 所述电子设备1中的所述存储器11存储的智能问答程序12是多个指令的组合, 在所述处理器10中运行时, 可以实现:

[0177] 接收用户输入的查询文本, 对所述查询文本进行分词处理及实体抽取处理, 得到问题实体;

[0178] 基于预构建的问题语义分析模型对所述查询文本进行语义分析, 得到问题意图特征;

[0179] 利用预构建的至少一个答案知识库对所述问题实体进行查询, 得到至少一组答案文本集;

[0180] 对所述至少一组答案文本集中的答案文本进行相关度计算与分析, 得到待选答案文本集;

[0181] 利用所述问题意图特征对所述待选答案文本集中的待选答案文本进行加权计算

与分析,得到最终答案文本,并将所述最终答案文本推送给用户。

[0182] 具体地,所述处理器10对上述指令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。

[0183] 进一步地,所述电子设备1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)。

[0184] 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。

[0185] 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。

[0186] 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。

[0187] 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。

[0188] 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。

[0189] 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。

[0190] 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。

[0191] 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。

[0192] 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。

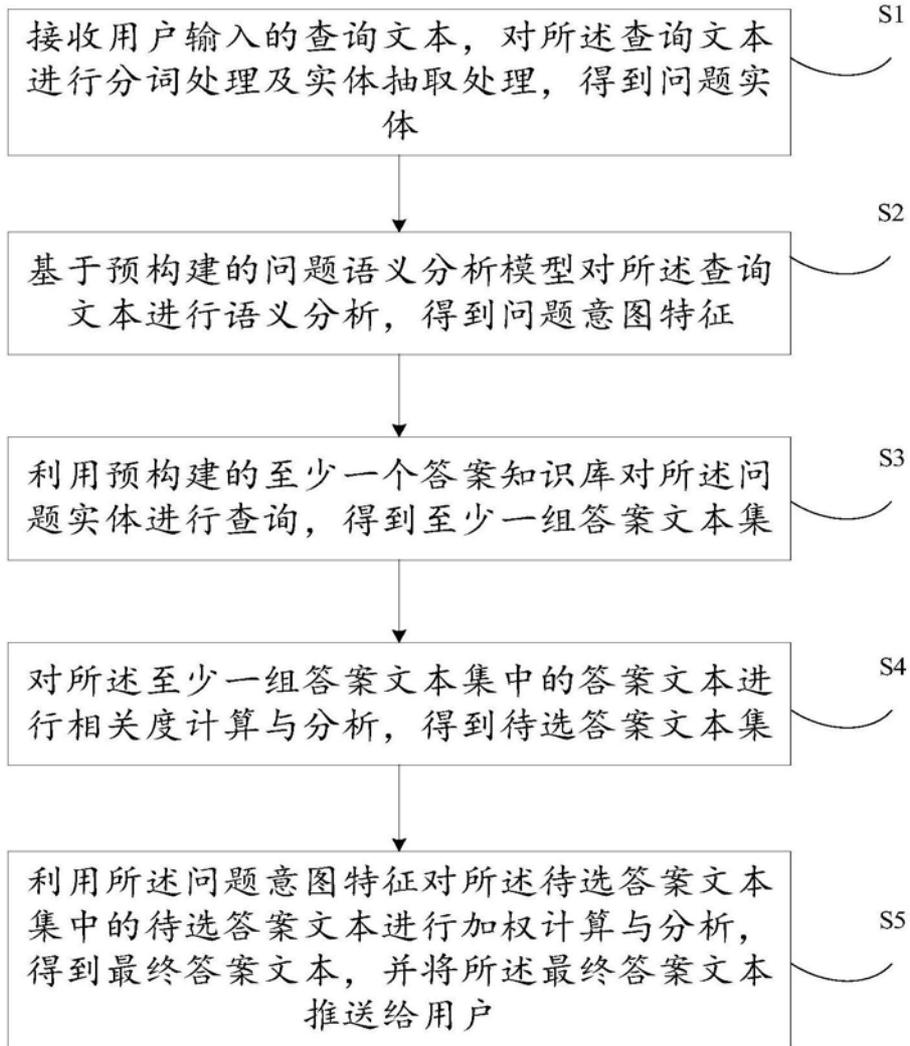


图1

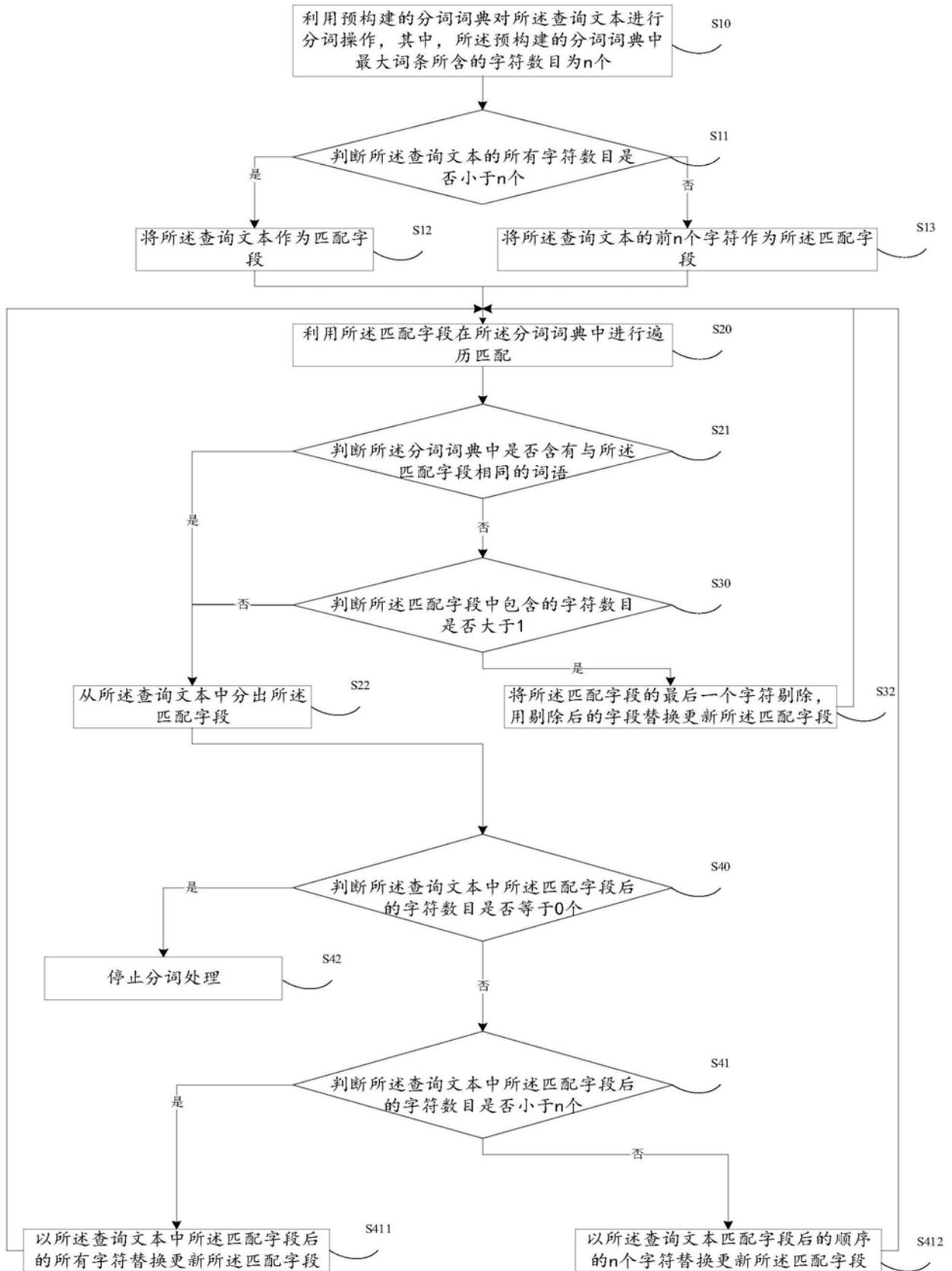


图2



图3



图4