



(12) 发明专利申请

(10) 申请公布号 CN 114238595 A

(43) 申请公布日 2022. 03. 25

(21) 申请号 202111459059.3

G06F 40/295 (2020.01)

(22) 申请日 2021.12.01

G06Q 50/04 (2012.01)

(71) 申请人 中冶赛迪重庆信息技术有限公司
地址 401329 重庆市九龙坡区白市驿镇农
科大道66号2幢5-6号

(72) 发明人 姜微 张璟涵 赵丹丹 单凤池
王丹 秦若峰 王明生 杨淋清
祝捷 张伟

(74) 专利代理机构 上海光华专利事务所(普通
合伙) 31219
代理人 李铁

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 16/36 (2019.01)

G06F 40/211 (2020.01)

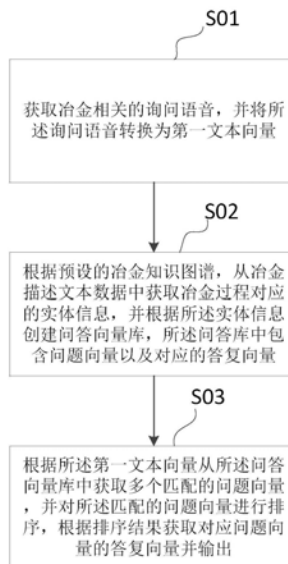
权利要求书2页 说明书5页 附图1页

(54) 发明名称

一种基于知识图谱的冶金知识问答方法及系统

(57) 摘要

本发明提出一种基于知识图谱的冶金知识问答方法及系统,包括:获取冶金相关的询问语音,并将所述询问语音转换为第一文本向量;根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,所述问答库中包含问题向量以及对应的答复向量;根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,并对所述匹配的问题向量进行排序,根据排序结果获取对应问题向量的答复向量并输出;本发明可简化冶金作业人员的检索操作,保障检索响应的快速高效。



1. 一种基于知识图谱的冶金知识问答方法,其特征在于,包括:
获取冶金相关的询问语音,并将所述询问语音转换为第一文本向量;
根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,所述问答库中包含问题向量以及对应的答复向量;
根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,并对所述匹配的问题向量进行排序,根据排序结果获取对应问题向量的答复向量并输出。
2. 根据权利要求1所述的基于知识图谱的冶金知识问答方法,其特征在于,根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,包括:
利用关键词搜索引擎对所述第一文本向量进行分词,获取多个词向量,根据所述词向量从已有的故障报告以及故障文本中获取匹配的关键词,构建关联词向量,并根据所述关联词向量从所述问答向量库中获取多个匹配的问题向量;和/或,
利用向量搜索引擎将所述第一文本向量和所述问答向量库中的问题向量映射到同一个语义向量空间,通过相似度检索算法从所述语义向量空间中获取与所述第一文本向量相似度达到所述设定阈值的问题向量。
3. 根据权利要求2所述的基于知识图谱的冶金知识问答方法,其特征在于,所述关键词搜索引擎包括:ElasticSearch搜索引擎。
4. 根据权利要求2所述的基于知识图谱的冶金知识问答方法,其特征在于,所述向量搜索引擎包括:Faiss引擎和/或Annoy引擎。
5. 根据权利要求1所述的基于知识图谱的冶金知识问答方法,其特征在于,根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,包括:
获取所述冶金知识图谱中的多组实体关系,根据所述实体关系构建第一句子向量;
将所述冶金描述文本数据拆分为多个子句,每个所述子句作为第二句子向量;
将所述第一句子向量和所述第二句子向量输入预训练的句子向量模型进行相似度比对,获取与所述第一句子向量匹配的第二句子向量,作为问题向量;
设置各所述问题向量对应的答复向量,创建所述问答向量库。
6. 根据权利要求1所述的基于知识图谱的冶金知识问答方法,其特征在于,所述冶金描述文本数据包括:冶金故障报告、故障总结。
7. 根据权利要求1所述的基于知识图谱的冶金知识问答方法,其特征在于,对所述匹配的问题向量进行排序,包括:
通过深度学习排序算法获取所述匹配的问题向量与所述第一文本向量的相似度值,根据所述相似度值对所述多个匹配的问题向量进行排序。
8. 根据权利要求7所述的基于知识图谱的冶金知识问答方法,其特征在于,所述深度学习排序算法包括:关键词检索算法、BM25算法和/或余弦相似度算法。
9. 一种基于知识图谱的冶金知识问答系统,其特征在于,包括:
语音识别模块,用于获取冶金相关的询问语音,并将所述询问语音转换为第一文本向量;
问答向量库创建模块,用于根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,所述问答库中包含问题向

量以及对应的答复向量；

输出响应模块,用于根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,并对所述匹配的问题向量进行排序,根据排序结果获取对应问题向量的答复向量并输出。

一种基于知识图谱的冶金知识问答方法及系统

技术领域

[0001] 本发明涉及智能冶金领域,尤其涉及一种基于知识图谱的冶金知识问答方法及系统。

背景技术

[0002] 目前知识图谱领域蓬勃发展,并广泛应用于通用领域和垂直领域的知识管理。然而传统冶金行业的生产管理方面也存在着以下问题:

[0003] 1) 缺少冶金行业故障知识的整合库,不利于冶金从业人员快速获取生产故障原因及解决方案;

[0004] 2) 现有的知识库语义检索准确率、高效性不高且智能化程度不高,不能很好的支持近义词、不同表达方式的相同语义识别;

[0005] 3) 多数检索平台仅仅采取文字输入的方式,用户输入文字较为繁琐,便捷性差。

发明内容

[0006] 鉴于以上现有技术存在的问题,本发明提出一种基于知识图谱的冶金知识问答方法及系统,主要解决冶金行业传统检索方式操作繁琐,影响作业效率的问题。

[0007] 为了实现上述目的及其他目的,本发明采用的技术方案如下。

[0008] 一种基于知识图谱的冶金知识问答方法,包括:

[0009] 获取冶金相关的询问语音,并将所述询问语音转换为第一文本向量;

[0010] 根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,所述问答库中包含问题向量以及对应的答复向量;

[0011] 根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,并对所述匹配的问题向量进行排序,根据排序结果获取对应问题向量的答复向量并输出。

[0012] 可选地,根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,包括:

[0013] 利用关键词搜索引擎对所述第一文本向量进行分词,获取多个词向量,根据所述词向量从已有的故障报告以及故障文本中获取匹配的关键词,构建关联词向量,并根据所述关联词向量从所述问答向量库中获取多个匹配的问题向量;和/或,

[0014] 利用向量搜索引擎将所述第一文本向量和所述问答向量库中的问题向量映射到同一个语义向量空间,通过相似度检索算法从所述语义向量空间中获取与所述第一文本向量相似度达到所述设定阈值的问题向量。

[0015] 可选地,所述关键词搜索引擎包括:ElasticSearch搜索引擎。

[0016] 可选地,所述向量搜索引擎包括:Faiss引擎和/或Annoy引擎。

[0017] 可选地,根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,包括:

- [0018] 获取所述冶金知识图谱中的多组实体关系,根据所述实体关系构建第一句子向量;
- [0019] 将所述冶金描述文本数据拆分为多个子句,每个所述子句作为第二句子向量;
- [0020] 将所述第一句子向量和所述第二句子向量输入预训练的句子向量模型进行相似度比对,获取与所述第一句子向量匹配的第二句子向量,作为问题向量;
- [0021] 设置各所述问题向量对应的答复向量,创建所述问答向量库。
- [0022] 可选地,所述冶金描述文本数据包括:冶金故障报告、故障总结。
- [0023] 可选地,对所述匹配的问题向量进行排序,包括:
- [0024] 通过深度学习排序算法获取所述匹配的问题向量与所述第一文本向量的相似度值,根据所述相似度值对所述多个匹配的问题向量进行排序。
- [0025] 可选地,所述深度学习排序算法包括:关键词检索算法、BM25算法和/或余弦相似度算法。
- [0026] 一种基于知识图谱的冶金知识问答系统,包括:
- [0027] 语音识别模块,用于获取冶金相关的询问语音,并将所述询问语音转换为第一文本向量;
- [0028] 问答向量库创建模块,用于根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,所述问答库中包含问题向量以及对应的答复向量;
- [0029] 输出响应模块,用于根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,并对所述匹配的问题向量进行排序,根据排序结果获取对应问题向量的答复向量并输出。
- [0030] 如上所述,本发明一种基于知识图谱的冶金知识问答方法及系统,具有以下有益效果。
- [0031] 利用预设的知识图谱获取冶金行业较为全面的冶金知识数据,通过语音输入快速完成冶金信息的检索匹配,获取需要的数据输出,为作业人员提供便捷的检索方式的同时,提高数据输出的准确性和有效性。

附图说明

- [0032] 图1为本发明一实施例中基于知识图谱的冶金知识问答方法的流程示意图。

具体实施方式

[0033] 以下通过特定的具体实例说明本发明的实施方式,本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用,本说明书中的各项细节也可以基于不同观点与应用,在没有背离本发明的精神下进行各种修饰或改变。需说明的是,在不冲突的情况下,以下实施例及实施例中的特征可以相互组合。

[0034] 需要说明的是,以下实施例中所提供的图示仅以示意方式说明本发明的基本构想,遂图式中仅显示与本发明中有关的组件而非按照实际实施时的组件数目、形状及尺寸绘制,其实际实施时各组件的型态、数量及比例可为一种随意的改变,且其组件布局型态也

可能更为复杂。

[0035] 请参阅图1,本发明提供基于知识图谱的冶金知识问答方法,包括以下步骤。

[0036] 步骤S01,获取冶金相关的询问语音,并将所述询问语音转换为第一文本向量;

[0037] 步骤S02,根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,所述问答库中包含问题向量以及对应的答复向量;

[0038] 步骤S03,根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,并对所述匹配的问题向量进行排序,根据排序结果获取对应问题向量的答复向量并输出。

[0039] 在一实施例中,语音录入功能可采用了现有的语音实时转写工具。前端调用对应的语音转写api,将语音流实时转换为文字信息,输入到问答对话框中实现了语音智能输入,具体的语音转写工具可根据实际应用需求进行选择,这里不作限制。

[0040] 在一实施例中,精准识别用户语音输入的冶金会话后,可利用自然语言处理技术HanLP对冶金会话进行预处理并转换成语义向量(即第一文本向量)。具体地,利用HanLP工具及收集的冶金行业专用名词词典对冶金会话进行分词、词性标注、实体识别、纠错及语义召回等预处理,并进一步利用句子向量模型Sentence BERT将问题语句转换成词向量,得到对应的第一文本向量。

[0041] 在一实施例中,根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,包括:

[0042] 获取所述冶金知识图谱中的多组实体关系,根据所述实体关系构建第一句子向量;

[0043] 将所述冶金描述文本数据拆分为多个子句,每个所述子句作为第二句子向量;

[0044] 将所述第一句子向量和所述第二句子向量输入预训练的句子向量模型进行相似度比对,获取与所述第一句子向量匹配的第二句子向量,作为问题向量;

[0045] 设置各所述问题向量对应的答复向量,创建所述问答向量库。

[0046] 具体地,可收集冶金行业的相关资料如故障描述文本、故障汇总解决方案等,通过提取实体-关系-实体或实体-关系-属性三元组,生成对应的知识图谱。具体的知识图谱构建过程可采用现有技术,这里不再赘述。基于预先构建的底层知识图谱中的实体关系,通过搜集冶金故障报告和故障总结,抽取出故障知识,并利用句子向量模型Sentence BERT构建故障问答向量库。

[0047] 在一实施例中,根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,包括:

[0048] 利用关键词搜索引擎对所述第一文本向量进行分词,获取多个词向量,根据所述词向量从已有的故障报告以及故障文本中获取匹配的关键词,构建关联词向量,并根据所述关联词向量从所述问答向量库中获取多个匹配的问题向量;和/或,

[0049] 利用向量搜索引擎将所述第一文本向量和所述问答向量库中的问题向量映射到同一个语义向量空间,通过相似度检索算法从所述语义向量空间中获取与所述第一文本向量相似度达到所述设定阈值的问题向量。

[0050] 在一实施例中,利用关键词搜索引擎进行关键词召回,关键词搜索引擎可采用ElasticSearch搜索引擎,ElasticSearch将第一文本向量进行自动分词处理,并在非结构

化故障报告及故障文本中进行全文索引和相关性搜索,快速获取高匹配度返回结果。

[0051] 在一实施例中,利用向量搜索引擎进行语义召回,向量搜索引擎可采用Faiss或Annoy向量索引算法,Faiss或Annoy索引算法为问答向量库中的问题向量构建一个索引,用户输入问题和问答向量库中的问题被映射到同一个语义向量空间,通过ANN算法快速获得最相似的K个问题向量。

[0052] 在一实施例中,对所述匹配的问题向量进行排序,包括:

[0053] 通过深度学习排序算法获取所述匹配的问题向量与所述第一文本向量的相似度值,根据所述相似度值对所述多个匹配的问题向量进行排序。

[0054] 具体地,所述深度学习排序算法,包括:结合传统的关键词检索、BM25、Cosine相似度算法以及一些基于神经网络训练的Word2vec模型或者卷积匹配模型等共同对topK问题进行排序。

[0055] 完成排序后,可将排序结果反馈给用户,每个问题向量对应一个索引信息,用户可通过语音输入匹配对应的索引信息,进行问题选择。根据用户选择输出对应问题向量对应的答复向量。答复向量对应具体问题的解决方案,可在通过知识图谱从各非结构化冶金描述文本中获取问题向量后,基于获取的问题向量配置对应的解决方案作为问题向量对应的答复向量,存储在问答向量库中。

[0056] 在一实施例中,问答流程包括:

[0057] 步骤1,利用Sentence BERT语意向量模型将提出的问题转换成语义向量,同时基于底层知识图谱抽取构建故障问答向量库;

[0058] 步骤2,结合Elasticsearch关键词召回及Faiss、Annoy向量召回技术,将问题向量与问答向量库训练的语义向量进行相似度对比,获取与提出的问题相似的topK问题;

[0059] 步骤3,利用排序算法对topK问题进行排序,从而关联出输入问题的最佳答案。

[0060] 本实施例中还提供了一种基于知识图谱的冶金知识问答系统,用于执行前述方法实施例中所述的基于知识图谱的冶金知识问答方法。由于系统实施例的技术原理与前述方法实施例的技术原理相似,因而不再对同样的技术细节做重复性赘述。

[0061] 在一实施例中,基于知识图谱的冶金知识问答系统,包括:

[0062] 语音识别模块,用于获取冶金相关的询问语音,并将所述询问语音转换为第一文本向量;

[0063] 问答向量库创建模块,用于根据预设的冶金知识图谱,从冶金描述文本数据中获取冶金过程对应的实体信息,并根据所述实体信息创建问答向量库,所述问答库中包含问题向量以及对应的答复向量;

[0064] 输出响应模块,用于根据所述第一文本向量从所述问答向量库中获取多个匹配的问题向量,并对所述匹配的问题向量进行排序,根据排序结果获取对应问题向量的答复向量并输出。

[0065] 综上所述,本发明一种基于知识图谱的冶金知识问答方法及系统,支持问答会话语音输入,不需要手动输入,极大地节省了冶金一线从业人员的时间;借助底层知识图谱搭建了一个较为全面的冶金故障知识库,帮助冶金从业人员快速获取生产故障原因及解决方案;基于关键词搜索引擎和向量搜索引擎结合的搜索匹配索引技术大大提高了搜索的准确率,并且支持语义理解实现了问答的智能化;能够达到毫秒级响应的高校检索,更适合钢铁

生产场景。所以,本发明有效克服了现有技术中的种种缺点而具高度产业利用价值。

[0066] 上述实施例仅例示性说明本发明的原理及其功效,而非用于限制本发明。任何熟悉此技术的人士皆可在不违背本发明的精神及范畴下,对上述实施例进行修饰或改变。因此,举凡所属技术领域中具有通常知识者在未脱离本发明所揭示的精神与技术思想下所完成的一切等效修饰或改变,仍应由本发明的权利要求所涵盖。

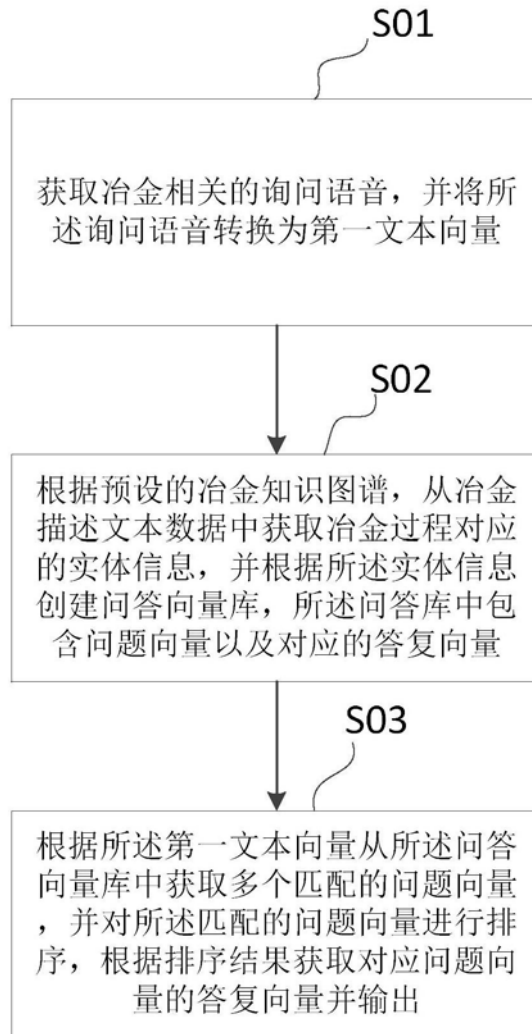


图1