



(12) 发明专利申请

(10) 申请公布号 CN 117195855 A

(43) 申请公布日 2023. 12. 08

(21) 申请号 202210582273.6

G06F 40/30 (2020.01)

(22) 申请日 2022.05.26

(71) 申请人 华为云计算技术有限公司

地址 550025 贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心

(72) 发明人 蔡蔓菁 曹高雄 陆明哲 李明磊

(74) 专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285

专利代理师 陈松浩

(51) Int. Cl.

G06F 40/194 (2020.01)

G06F 16/2458 (2019.01)

G06F 16/215 (2019.01)

G06F 16/22 (2019.01)

G06F 18/22 (2023.01)

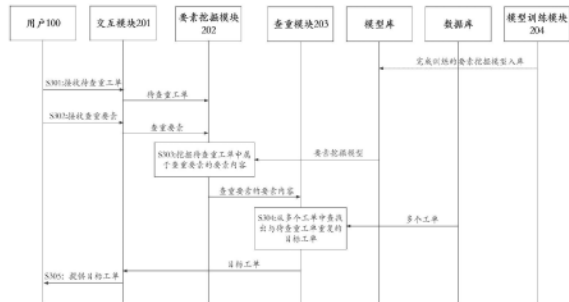
权利要求书3页 说明书11页 附图5页

(54) 发明名称

工单查重方法、装置及相关设备

(57) 摘要

本申请提供了一种工单查重方法,具体的,工单查重装置获取待查重工单,并获取用户指定的查重要素,然后,工单查重装置利用模型库中的要素挖掘模型,挖掘出待查重工单中属于该查重要素的要素内容,并从数据库存储的多个工单中查找出目标工单,所查找出的目标工单中属于查重要素的要素内容的语义与待查重工单中属于查重要素的要素内容的语义相匹配,从而工单查重装置提供用于呈现给用户的目标工单。如此,即使两个工单中表达语义的关键词不同,工单查重装置也能将这两个工单识别为重复工单,从而可以有效提高工单查重的准确率、提高工单查重效果。此外,本申请还提供了对应的装置及相关设备。



1. 一种工单查重方法,其特征在于,所述方法包括:
 - 获取待查重工单;
 - 获取用户指定的查重要素;
 - 利用模型库中的要素挖掘模型,挖掘出所述待查重工单中属于所述查重要素的要素内容;
 - 从数据库存储的多个工单中查找出目标工单,所述目标工单中属于所述查重要素的要素内容的语义与所述待查重工单中属于所述查重要素的要素内容的语义相匹配;
 - 提供所述目标工单,所述目标工单用于呈现给所述用户。
2. 根据权利要求1所述的方法,其特征在于,所述查重要素包括结构化要素和非结构化要素,所述从数据库存储的多个工单中查找出目标工单,包括:
 - 根据所述待查重工单中属于所述结构化要素的要素内容,对数据库存储的多个工单进行筛选,得到候选工单集合,所述候选工单集合包括至少一个候选工单,所述候选工单中属于所述结构化要素的要素内容的语义与所述待查重工单中属于所述结构化要素的要素内容的语义相匹配;
 - 分别计算所述待查重工单中属于所述非结构化要素的要素内容的语义与各个所述候选工单中属于所述非结构化要素的要素内容的语义的相似度;
 - 从所述至少一个候选工单中确定出所述目标工单,所述目标工单中属于所述非结构化要素的要素内容的语义与所述待查重工单中属于所述非结构化要素的要素内容的语义的相似度大于预设阈值。
3. 根据权利要求2所述的方法,其特征在于,所述获取用户指定的查重要素,包括:
 - 生成配置界面,所述配置界面用于将多个查重要素呈现给所述用户;
 - 响应于用户在所述配置界面针对所述多个查重要素的选择操作,确定所述用户选择的所述结构化要素以及所述非结构化要素。
4. 根据权利要求3所述的方法,其特征在于,所述方法还包括:
 - 响应于所述用户在所述配置界面针对所述非结构化要素的配置操作,确定所述预设阈值;
 - 响应于所述用户在所述配置界面针对多种计算方式的选择操作,确定所述相似度对应的计算方式。
5. 根据权利要求2至4任一项所述的方法,其特征在于,所述非结构化要素包括文本向量、图片向量以及视频向量中的至少一种。
6. 根据权利要求1至5任一项所述的方法,其特征在于,所述方法还包括:
 - 生成校验界面,所述校验界面用于将所述目标工单以及所述待查重工单呈现给所述用户;
 - 响应于所述用户的确认操作,合并所述待查重工单与所述目标工单,以使得所述待查重工单与所述目标工单基于同一操作被处理。
7. 一种工单查重装置,其特征在于,所述工单查重装置包括:
 - 交互模块,用于获取待查重工单以及用户指定的查重要素;
 - 要素挖掘模块,用于利用模型库中的要素挖掘模型,挖掘出所述待查重工单中属于所述查重要素的要素内容;

查重模块,用于从数据库存储的多个工单中查找出目标工单,所述目标工单中属于所述查重要素的要素内容的语义与所述待查重工单中属于所述查重要素的要素内容的语义相匹配;

所述交互模块,还用于提供所述目标工单,所述目标工单用于呈现给所述用户。

8.根据权利要求7所述的工单查重装置,其特征在于,所述查重要素包括结构化要素和非结构化要素,所述查重模块,用于:

根据所述待查重工单中属于所述结构化要素的要素内容,对数据库存储的多个工单进行筛选,得到候选工单集合,所述候选工单集合包括至少一个候选工单,所述候选工单中属于所述结构化要素的要素内容的语义与所述待查重工单中属于所述结构化要素的要素内容的语义相匹配;

分别计算所述待查重工单中属于所述非结构化要素的要素内容的语义与各个所述候选工单中属于所述非结构化要素的要素内容的语义的相似度;

从所述至少一个候选工单中确定出所述目标工单,所述目标工单中属于所述非结构化要素的要素内容的语义与所述待查重工单中属于所述非结构化要素的要素内容的语义的相似度大于预设阈值。

9.根据权利要求8所述的工单查重装置,其特征在于,所述交互模块,用于:

生成配置界面,所述配置界面用于将多个查重要素呈现给所述用户;

响应于用户在所述配置界面针对所述多个查重要素的选择操作,确定所述用户选择的所述结构化要素以及所述非结构化要素。

10.根据权利要求9所述的工单查重装置,其特征在于,所述交互模块,还用于:

响应于所述用户在所述配置界面针对所述非结构化要素的配置操作,确定所述预设阈值;

响应于所述用户在所述配置界面针对多种计算方式的选择操作,确定所述相似度对应的计算方式。

11.根据权利要求8至10任一项所述的工单查重装置,其特征在于,所述非结构化要素包括文本向量、图片向量以及视频向量中的至少一种。

12.根据权利要求7至11任一项所述的工单查重装置,其特征在于,所述交互模块,还用于:

生成校验界面,所述校验界面用于将所述目标工单以及所述待查重工单呈现给所述用户;

响应于所述用户的确认操作,合并所述待查重工单与所述目标工单,以使得所述待查重工单与所述目标工单基于同一操作被处理。

13.一种计算设备集群,其特征在于,所述计算设备集群包括至少一个计算设备,每个计算设备包括处理器以及存储器:

所述存储器用于存储指令;

所述处理器用于根据所述指令,以使得所述计算设备集群执行权利要求1-6中任一项所述的方法。

14.一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指令,当其在计算设备上运行时,使得所述计算设备执行如权利要求1至6任一项所述的方法。

15. 一种包含指令的计算机程序产品,当其在计算设备上运行时,使得所述计算设备执行如权利要求1至6任一项所述的方法。

工单查重方法、装置及相关设备

技术领域

[0001] 本申请涉及数据处理技术领域,尤其涉及一种工单查重方法、装置及相关设备。

背景技术

[0002] 工单,具体可以称为工作单据,通常是客户或企业内部员工针对产品或服务问题向服务商、企业部门提交的一项任务、工作或请求,以便服务商或企业部门根据该工单执行相应的处理流程。

[0003] 随着线上业务的发展(如线上热线的普及等),工单的数量不断的累积。相比于工单的数量,工单所反映的案件类型的数量通常有限,这就意味着新产生的工单可能会与历史工单(即在过去时间段内产生的工单)重复,从而可以通过将新产生的工作与重复的历史工单进行并案处理,以降低处理工单所需的人力成本。

[0004] 目前,通常是通过关键词匹配的方式,确定新产生的工单是否与历史工单重复。但是,这种工单查重方式的准确率较低、查重效果较差。

发明内容

[0005] 本申请提供了一种工单查重方法,用于通过语义匹配的方式进行工单查重,以此提供工单查重的准确率、提高查重效果。此外,本申请还提供了一种工单查重装置、计算设备、计算机可读存储介质以及计算机程序产品。

[0006] 第一方面,本申请提供了一种工单查重方法,该工单查重方法可以由相应的工单查重装置执行,具体的,工单查重装置获取待查重工单,该待查重工单例如可以是由用户通过信息输入界面提供给工单查重装置等,并且,工单查重装置还获取用户指定的查重要素,例如用户可以通过配置界面指定工单查重时所使用的一个或者多个要素等,然后,工单查重装置利用模型库中的要素挖掘模型,挖掘出待查重工单中属于该查重要素的要素内容,并从数据库存储的多个工单中查找出目标工单,所查找出的目标工单中属于查重要素的要素内容的语义与待查重工单中属于查重要素的要素内容的语义相匹配,从而工单查重装置提供用于呈现给用户的目标工单,例如,工单查重装置可以通过对外提供的客户端将该目标工单呈现给用户等。

[0007] 如此,工单查重装置可以根据查重要素的要素内容的语义是否匹配的方式,判定多个工单中是否存在与待查重工单重复的工单,这使得两个相同语义的工单中,即使表达该语义的关键词不同,工单查重装置也能将这两个工单识别为重复工单,如此,可以有效提高工单查重的准确率、提高工单查重效果。并且,当用户指定的查重要素包括多个要素时,基于多个查重要素进行工单查询,不仅可以不同维度的特征信息确定待查重工单是否为重复工单,以此使得工单查重的准确率达到较高水平,而且,可以实现针对不同场景选择不同组合的查重要素进行工单查重,以此提高工单查重的普适性。

[0008] 在一种可能的实施方式中,用户所指定的查重要素包括结构化要素以及非结构化要素,从而工单查重装置在查找目标工单的过程中,具体可以先根据待查重工单中属于

结构化要素的要素内容,对数据库存储的多个工单进行筛选,得到候选工单集合,该候选工单集合包括至少一个候选工单,该候选工单中属于结构化要素的要素内容的语义与该待查重工单中属于该结构化要素的要素内容的语义相匹配,从而工单查重装置可以从至少一个候选工单中确定出目标工单,该目标工单中属于非结构化要素的要素内容的语义与待查重工单中属于非结构化要素的要素内容的语义的相似度大于预设阈值。如此,工单查重装置可以利用二层的匹配过程,从多个工单中确定出与待查重工单重复的目标工单。

[0009] 相应的,当数据库中不存在工单的属于结构化要素的要素内容的语义与待查重工单中属于结构化要素的要素内容的语义相匹配,或者候选工单中不存在非结构化要素的要素内容的语义与待查重工单中属于非结构化要素的要素内容的语义的相似度大于预设阈值,则工单查重装置可以确定数据库中不存在与待查重工单重复的工单。

[0010] 在一种可能的实施方式中,工单查重装置在获取用户指定的查重要素时,具体可以是生成配置界面,该配置界面用于将多个查重要素呈现给用户,例如可以是通过对外提供的客户端呈现给用户等,从而工单查重装置可以响应于用户在该配置界面针对该多个查重要素的选择操作,确定用户选择的结构化要素以及非结构化要素。如此,可以通过人机交互的方式,实现用户对于工单查重所需的查重要素进行自定义,以此在提高用户体验的同时,也能满足用户针对不同场景中工单选择不同组合的查重要素的需求,提高工单查重的灵活性。

[0011] 在一种可能的实施方式中,针对非结构化要素,工单查重装置还响应于用户在该配置界面针对非结构化要素的配置操作,确定用于衡量两个工单在非结构化要素的要素内容的语义上的相似度大小的预设阈值,并且,工单查重装置还可以响应于用户在该配置界面针对多种计算方式的选择操作,确定计算该相似度所对应的计算方式。如此,工单查重装置可以支持用户对于预设阈值以及计算方式的自定义,以此可以进一步提高工单查重的灵活性。

[0012] 在一种可能的实施方式中,非结构化要素包括文本向量、图片向量以及视频向量中的至少一种。如此,工单查重装置可以深层次的理解待查重工单中的文本信息、图片信息、或视频信息,从而有助于提高工单查重的准确性。

[0013] 在一种可能的实施方式中,工单查重装置还可以生成校验界面,该校验界面用于将目标工单以及待查重工单呈现给用户,并响应于用户的确认操作,合并待查重工单与目标工单,以使得该目标工单与待查重工单基于同一操作被处理。如此,可以无需重新设计针对重复的工单的处理操作,从而可以提高工单的处理效率。

[0014] 其中,目标工单可以是与待查重工单重复并且尚未完成处理的工单,从而通过工单合并,可以实现对多个重复的工单进行统一处理,以此提高工单的处理效率。或者,目标工单可以是历史工单,即在过去时间段内已经被处理完成的工单,从而通过工单合并,可以基于处理目标工单的操作,对待查重工单进行相应处理,从而无需再对待查重工单的处理操作进行重新设计。

[0015] 第二方面提供的工单查重装置,对应于第一方面提供的工单查重方法,故第二方面以及第二方面任一种可能实现方式中的工单查重装置所具有的技术效果,可以参照第一方面以及第一方面中相应实现方式所具有的技术效果,在此不做赘述。

[0016] 第三方面,本申请提供一种计算设备集群,所述计算设备集群包括至少一个计算

设备,每个计算设备包括处理器和存储器;该存储器用于存储指令,当该计算设备集群运行时,每个计算设备中的处理器执行存储器存储的指令,以使该计算设备集群执行上述第一方面或第一方面任一种可能实现方式中的工单查重方法。需要说明的是,该存储器可以集成于处理器中,也可以是独立于处理器之外。每个计算设备还可以包括总线。其中,处理器通过总线连接存储器。其中,存储器可以包括可读存储器以及随机存取存储器。

[0017] 第四方面,本申请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指令,当其在计算设备集群(该计算设备集群包括至少一个计算设备)上运行时,使得计算设备集群执行上述第一方面或第一方面的任一种实现方式所述的方法。

[0018] 第五方面,本申请提供了一种包含指令的计算机程序产品,当其在计算设备集群(该计算设备集群包括至少一个计算设备)上运行时,使得计算设备集群执行上述第一方面或第一方面的任一种实现方式所述的方法。

[0019] 本申请在上述各方面提供的实现方式的基础上,还可以进行进一步组合以提供更多实现方式。

附图说明

[0020] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例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,还可以根据这些附图获得其它的附图。

[0021] 图1为本申请实施例提供的一示例性应用场景示意图;

[0022] 图2为本申请实施例提供的另一示例性应用场景示意图;

[0023] 图3为本申请实施例提供的一种工单查重方法的流程示意图;

[0024] 图4为本申请实施例提供的一种信息输入界面的示意图;

[0025] 图5为本申请实施例提供的一种配置界面的示意图;

[0026] 图6为本申请实施例提供的一种校验界面的示意图;

[0027] 图7为本申请实施例提供的一种计算设备的结构示意图;

[0028] 图8为本申请实施例提供的一种计算设备集群的结构示意图。

具体实施方式

[0029] 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,这仅仅是描述本申请的实施例中具有相同属性的对象在描述时所采用的区分方式。

[0030] 参见图1,为一示例性应用场景示意图。在图1所示的应用场景中,用户100可以(通过客户端或者用户终端等)向工单查重装置200提供一个或者多个工单,以请求对该工单进行查重。示例性地,用户100提供的工单,可以是与企业或者其他机构相关的工单,如企业内容用于工作安排所生成的工单、公共机构基于市民线上投诉内容所生成的工单等。为便于区分,以下将用户100提供的工单称为待查重工单。

[0031] 此时,如果工单查重装置200通过关键词匹配的方式,从多个工单中查找出与待查重工单具有相同关键词的工单,以此实现工单查重,则工单查重效果可能较差。比如,假设

工单查重装置200的本地数据库中存储有包括关键词“举报XXX”的工单,而待查重工单包括关键词“投诉XXX”,此时,工单查重装置200通过关键词匹配的方式进行查重,则因为关键词“举报XXX”与关键词“投诉XXX”并不相同,则工单查重装置200很可能会将这两个工单判定为不重复的工单(实际上应该为重复的工单),从而导致工单查重装置200的查重准确率较低。

[0032] 为此,本实施例中,工单查重装置200可以从语义匹配的角度,为待查重工单进行查重处理。具体地,工单查重装置200不仅获得待查重工单,还获取用户100所指定的查重要素,从而利用模型库中的要素挖掘模型可以挖掘出待查重工单中属于该查重要素的要素内容,其中,查重要素用于表达工单中数据的语义内容,并且,不同的查重要素可以表达不同方面的语义。比如,对于文本“A投诉B”,其中,工单查询装置200确定“A”对应的查重要素为投诉人、“投诉”对应的查重要素为工单类型、“B”对应的查重要素为涉事责任主体。然后,工单查重装置200从数据库存储的多个工单中查找是否存在目标工单,该目标工单中属于查重要素的要素内容的语义与待查重工单中属于该查重要素的要素内容的语义相匹配,从而在查找出目标工单后,工单查重装置200向用户100提供该目标工单,例如可以将该目标工单(以及待查重工单)一并呈现给用户100。

[0033] 如此,工单查重装置200可以根据查重要素的要素内容的语义是否匹配的方式,判定多个工单中是否存在与待查重工单重复的工单,这使得两个相同语义的工单中,即使表达该语义的关键词不同,工单查重装置200也能将这两个工单识别为重复工单,如此,可以有效提高工单查重的准确率、提高工单查重效果。并且,当用户100指定的查重要素包括多个要素时,基于多个查重要素进行工单查询,不仅可以不同维度的特征信息确定待查重工单是否为重复工单,以此使得工单查重的准确率达到较高水平,而且,可以实现针对不同场景选择不同组合的查重要素进行工单查重,以此提高工单查重的普适性。

[0034] 仍以查重工单为包括关键词“投诉XXX”的工单为例,假设工单查重装置200的本地数据库中存储有包括关键词“举报XXX”的工单,工单查重装置200根据属于查重要素的要素内容进行匹配,可以确定这两个工单中的属于工单类型要素的要素内容“投诉”、“举报”属于同一语义,属于涉事责任主体要素的要素内容“XXX”、“XXX”也属于同一语义,因此,工单查重装置200可以将这两个工单判定为重复的工单,从而提高了工单查重的准确率。

[0035] 示例性地,工单查重装置200可以通过硬件实现,或者可以通过软件实现。其中,当通过软件实现时,工单查重装置200可以是运行在计算设备上的应用程序,如计算引擎、虚拟机等。当通过硬件实现时,工单查重装置200中可以包括至少一个计算设备,如服务器等。或者,工单查重装置200也可以是利用专用集成电路(application-specific integrated circuit,ASIC)实现、或可编程逻辑器件(programmable logic device,PLD)实现的设备等。其中,上述PLD可以是复杂程序逻辑器件(complex programmable logical device,CPLD)、现场可编程门阵列(field-programmable gate array,FPGA)、通用阵列逻辑(generic array logic,GAL)或其任意组合实现。

[0036] 实际部署时,工单查重装置200可以部署在用户侧,即可以作为本地设备为用户100提供本地的工单查重服务;或者,工单查重装置200可以部署在云端,如公有云、边缘云或者分布式云等,用于为用户100提供云端的工单查重服务等,本申请对于工单查重装置200的部署方式并不进行限定。

[0037] 值得注意的是,图1所示的应用场景仅作为一种实现示例,实际应用时,工单查重装置200也可以应用于其他可适用的场景中。例如,在图2所示的应用场景中,工单查重装置200用于为用户100提供云端的工单查重服务,并且,工单查重装置200可以对外提供客户端300,从而用户100可以通过客户端300请求工单查重装置200进行工单查重。

[0038] 为便于理解,下面结合附图,对工单查重装置200为用户100提供工单查重服务的过程进行详细介绍。

[0039] 参见图3,示出了本申请实施例提供的一种工单查重方法的流程示意图,该方法可以应用于图1或者图2所示的工单查重装置200。为便于描述,下面以应用于图1所示的场景为例进行说明。其中,工单查重装置200可以包括交互模块201、要素挖掘模块202、查重模块203,并且,工单查重装置200本地可以部署有模型库以及数据库;进一步地,工单查重装置200还可以包括模型训练模块204。工单查重装置200中的各个模块的功能,具体可参见图3所示实施例的相关之处描述。如图3所示,该方法具体可以包括:

[0040] S301:交互模块201获取待查重工单。

[0041] 作为一种实现示例,交互模块201可以向用户100提供信息输入界面,以使用户100通过该信息输入界面将一个或者多个待查重工单提供给工单查重装置200,并请求为该待查重工单进行查重处理。例如,交互模块201可以提供如图4所示的信息输入界面,该信息输入界面上可以呈现有提示信息“请导入需要查重的工单!”,并配置有“导入”控件,从而用户100在触发该“导入”控件后,将一个或者多个待查重工单的文件通过信息输入界面导入工单查重装置200。

[0042] 本实施例中,信息输入界面(以及下文提及的各种交互界面)可以通过配置在交互模块201中的显示器直接呈现给用户100,或者可以通过用户侧的客户端或者用户终端呈现给用户100等,本实施例对此并不进行限定。

[0043] 然后,交互模块201可以将接收到的待查重工单转发给要素挖掘模块202。

[0044] S302:交互模块201获取用户指定的查重要素。

[0045] 待查重工单中可以包括属于一个或者多个要素的要素内容。比如,针对投诉类的工单,该工单中可以包括属于投诉人、投诉电话、投诉地址、涉事责任主体、情感、话题、工单类别、文本表征向量、图片表征向量、视频表征向量等要素的要素内容。

[0046] 本实施例中,工单查重装置200可以支持用户100指定查重工单时所使用的要素。例如,在处理投诉工单的场景中,用户100主要处理相同区域内的不同人员在三个月内投诉工单,从而可以将针对同一件事的重复工单进行去重处理,此时,用户可以指定利用工单类型、涉事责任主体、投诉地址、工单生成时间、文本向量等要素作为进行工单查重所使用的要素。

[0047] 作为一种实现示例,工单查重装置200可以通过交互模块201生成如图5所示的配置界面,该配置界面用于将多个查重要素呈现给用户100,其中,所呈现的多个查重要素可以包括结构化要素以及非结构化要素。这样,用户100可以在呈现的配置界面上对多个查重要素进行选择,例如用户100可以在配置界面上对部分或者全部查重要素的选择按钮进行点击等。相应地,交互模块201可以响应用户100的选择操作,确定用户所选择的查重要素。其中,用户100指定的查重要素的数量可以是一个或者多个。实际应用时,用户100可以结合实际应用场景生成的工单中的数据特点,灵活选择相应的查重要素的组合进行工单查重,

以此可以支持多种场景的工单查重需求,提高了工单查重的灵活性。

[0048] 进一步地,用户100所指定的查重要素,可以是结构化要素,或者可以是非结构化要素。其中,结构化要素,是指其要素内容能够从工单中直接匹配,如投诉人、投诉电话、投诉地址、涉事责任主体、情感、话题、工单类别等。非结构化要素,是指其要素内容需要通过计算才能得到,如文本表征向量、图片表征向量、视频表征向量等。

[0049] 然后,交互模块201可以将获取的查重要素转发给要素挖掘模块202。

[0050] S303:要素挖掘模块202利用模型库中的要素挖掘模型,挖掘出交互模块201所获取的待查重工单中属于用户100指定的查重要素的要素内容。

[0051] 本实施例中,工单查重装置200可以根基于表达工单语义的查重要素,对工单进行查重处理。在一种可能的实现方式中,工单查重装置200中的要素挖掘模块202可以从模型库中获取一个或者多个要素挖掘模型,不同要素挖掘模型用于挖掘工单中属于不同查重要素的要素内容。其中,模型库可以部署于工单查重装置200中(如图1所示),从而要素挖掘模块202可以通过读取本地的模型库,获取要素挖掘模型;或者,模型库独立于工单查重装置200部署,从而要素挖掘模块202可以通过交互模块201与存储模型库的装置进行交互,获得要素挖掘模型。然后,要素挖掘模块202利用获取到的一个或者多个要素挖掘模型挖掘出待查重工单中属于该查重要素的要素内容。在其它实施方式中,要素挖掘模块202也可以利用一个要素挖掘模型挖掘出工单中属于多个查重要素的要素内容等。

[0052] 其中,在工单查重装置200提供工单查重服务之前,模型库中的各个要素挖掘模型,可以预先通过至少一组训练样本完成训练,该训练样本例如可以是携带有要素标签的历史工单,并且,该要素挖掘模型可以由工单查重装置200完成训练,也可以是由其它装置进行训练。以工单查重装置200训练用于挖掘话题要素的模型为例,工单查重装置200中的模型训练模块204可以获取训练样本,该训练样本包括多个携带有话题标签的历史工单,从而模型训练模块204可以利用该训练样本对预先构建的模型进行训练,以此训练得到用于挖掘话题要素的模型。示例性地,要素挖掘模型,例如可以是基于深度学习网络进行构建,例如可以是bert模型、T5模型等。这样,基于要素挖掘模型可以深层次的理解待查重工单中的文本信息、图片信息、或视频信息,从而有助于提高工单查重的准确性。实际应用是,要素挖掘模型也可以通过其它方式完成构建,如通过规则匹配方式进行构建等,本实施例对此并不进行限定。

[0053] 然后,要素挖掘模块202可以将待查重工单以及挖掘出的属于用户100指定的查重要素的要素内容转发给查重模块203。

[0054] S304:查重模块203从数据库存储的多个工单中查找出目标工单,该目标工单中属于该查重要素的要素内容的语义与待查重工单中属于该查重要素的要素内容的语义相匹配。

[0055] 具体实现时,工单查重装置200可以在本地部署有数据库,并且,该数据库在存储工单的同时,还可以存储该工单的非结构化要素的要素内容以及标识该工单中属于各个结构化要素的要素内容,从而查重模块203可以对数据库中的多个工单进行遍历,确定具有查重要素的工单。这样,查重模块203通过进一步比对该工单与待查重工单在该查重要素的要素内容上的语义,确定这两个工单在属于查重要素的要素内容上的语义是否匹配。并且,当其匹配时,查重模块203将该工单确定为与待查重工单重复的目标工单;而当其不匹配时,

查重模块203可以确定数据库中不存在与待查重工单重复的工单。其中,查重模块203可以基于语义分析技术(latent semantic analysis,LSA)确定两个工单在查重要素的要素内容上的语义是否匹配,或者可以通过其他技术实现,如通过语义识别模型等方式进行判断等。

[0056] 作为一种实现示例,用户100指定的查重要素可以包括结构化要素以及非结构化要素。则,查重模块203可以先根据待查重工单中属于该结构化要素的要素内容,对数据库存储的多个工单进行筛选,得到候选工单集合,该候选工单集合包括至少一个候选工单,也即数据库中被筛选出的工单,该候选工单中属于该结构化要素的要素内容的语义与待查重工单中属于该结构化要素的要素内容的语义相匹配,如两个工单中关于该结构化要素的文本相同或者相似等。然后,查重模块203可以分别计算待查重工单中属于该非结构化要素的要素内容的语义与各个候选工单中属于该非结构化要素(如图片表征向量等)的要素内容的语义的相似度,从而可以从至少一个候选工单中确定出相似度大于预设阈值(如0.85等)的候选工单,并且该候选工单确定为与待查重工单重复的目标工单。

[0057] 示例性地,当非结构化要素包括多个要素时,如包括文本表征向量以及图片表征向量等,查重模块203可以通过加权求和的方式计算两个工单在非结构化要素的要素内容的语义上的相似度。比如,查重模块203可以基于下述公式(1)计算相似度:

$$[0058] \quad W = \alpha * A + \beta * B \quad (1)$$

[0059] 其中,W为两个工单在非结构化要素的要素内容的语义上的整体相似度大小,A为两个工单在文本表征向量上的相似度,B为两个工单在图片表征向量上的相似度, α 以及 β 为权重值,并且, α 与 β 的和为1。实际应用时,查重模块203也可以采用其它计算方式计算两个工单在非结构化要素上的相似度。如此,对于包括图片以及视频的待查重工单,工单查重装置200也能对待查重工单进行准确查重。

[0060] 进一步地,当基于非结构化要素的要素内容进行工单查重时,工单查重装置200还可以支持用户100对上述确定目标工单的预设阈值以及相似度的计算方式进行配置。比如,在图5所示的配置界面中,当用户所指定的查重要素包括非结构化要素时,用户100可以在该配置界面上配置用于衡量两个工单在非结构化要素的要素内容的语义上的相似度大小所采用的阈值,并且,配置界面还可以呈现用于计算相似度的多种计算方式,以使用户100从多种计算方式中选择其中一种计算方式进行工单查重。

[0061] 另外,在进行工单查重过程中,若查重模块203根据待查重工单中属于结构化要素的要素内容,未从数据库中查找出在属于结构化要素的要素内容的语义上相匹配的工单,或者待查重工单与各个候选工单之间在非结构化要素的要素内容的语义上的相似度均小于或等于预设阈值,则查重模块203可以确定数据库中不存在与待查重工单重复的目标工单。实际应用时,对于不存在重复工单的待查重工单,工单查重装置200还可以将该待查重工单对应的一个或者多个要素添加至数据库中,以丰富数据库,以便基于更新后的数据库指导后续其他工单的查重过程。

[0062] 值得注意的是,上述确定目标工单的方式仅作为一种示例性说明,实际应用时,查重模块203也可以通过其他方式确定目标工单,比如,查重模块203在基于上述公式(1)计算出各个候选工单与待查重工单在非结构化要素的要素内容的语义上的相似度后,可以对各个候选工单分别对应的相似度大小进行排序,并将前N个相似度最大的候选工单均确定为

目标工单(N为正整数,如为1等)。

[0063] 本实施例中,数据库中的工单,可以是未被处理的其它待查重工单,如可以是24小时内接收到但尚未处理的工单等。这样,在工单查重装置200对这些工单进行查重处理后,可以删除重复的工单,或者对重复的工单进行同一操作的批量处理等,以此减少用户100处理工单的效率、降低处理工单的成本。

[0064] 又比如,数据库中的工单,例如可以是历史工单,如预先可以将企业过去所处理的工单作为历史工单存储于数据库中。这样,用户100在从数据库中查找出与待查重工单存在重复的历史工单后,可以基于企业处理历史工单的操作对待查重工单进行相应处理;或者,可以基于企业对历史工单的处理过程,分析历史处置无法彻底解决问题的原因。

[0065] 示例性地,当数据库中存储的工单为历史工单时,工单查重装置200在进行工单查重之前,还可以利用模型库中的要素挖掘模型对历史工单进行要素挖掘,以此得到各个历史工单分别对应的要素(包括结构化要素以及非结构化要素),从而可以将历史工单以及要素对应保存在数据库。这样,查重模块203在进行工单查重时,即可从数据库中确定各个历史工单所具有的要素以及属于各个要素的要素内容。实际应用时,若数据库中预先对应存储历史工单以及该历史工单的部分要素,则工单查重装置200可以利用训练要素挖掘模型挖掘数据库中的历史工单的要素后,对数据库中存储的历史工单对应的要素进行补充。

[0066] 基于上述过程,查重模块203可以从数据库中查找出目标工单,并将该目标工单发送给交互模块201。

[0067] S305:交互模块提供该目标工单,该目标工单用于呈现给用户100。

[0068] 在一种可能的实施方式中,当工单查重装置200确定出数据库中与待查重工单重复的目标工单时,工单查重装置200还可以进一步将目标工单与待查重工单进行捆绑输出,并将该目标工单的相关信息呈现给用户100(或者技术人员),如呈现目标工单的名称、工单内容、待查重工单与目标工单之间在非结构化要素上的语义的相似度、工单发现时间、针对目标工单的历史处置内容等,以便通过人工判断这两个工单是否重复的方式,进一步提高工单查重的准确性。

[0069] 例如,交互模块201可以生成用于将待查重工单以及目标工单呈现给用户100的校验界面,该校验界面例如可以如图6所示,包括目标工单以及待查重工单的相关信息。这样,用户100在查看到可能与待查重工单存在重复的目标工单后,可以通过呈现的这两个工单的相关信息,人工判断这两个工单是否重复。并且,如图6所示,该配置界面中还可以配置有“并单”控件,从而用户100在确定目标工单与待查重工单重复时,可以点击该“并单”控件。相应的,交互模块201基于用户100的确认操作(即点击“并单”控件的操作),合并待查重工单与目标工单。这样,后续可以利用处理目标工单的操作方式对待查重工单进行相应的处理,以此提高处理工单的效率、降低处理工单的成本,或者可以借此分析出现重复工单的原因,如分析历史处置无法彻底解决问题的根因等。同时,该配置界面中还可以配置有“不并单”控件,用户100在确定目标与待查重工单不重复时,可以点击该“不并单”控件,以便后续单独处理该待查重工单。

[0070] 进一步地,在工单查重装置200完成对于待查重工单的查重过程后,工单查重装置200可以将该待查重工单作为新的训练样本,对模型库中的要素挖掘模型进行更新,以便通过对该要素挖掘模型的动态更新,保持要素挖掘模型的活力、使得要素挖掘模型的推理精

度保持在较高水平。

[0071] 值得注意的是,本实施例中,是以针对一个待查重工单为例介绍工单查重装置200所执行的工单查重过程,实际应用时,用户100所提供的待查重工单的数量可以是多个,从而针对用户100提供的每个工单,工单查重装置200均可以采用上述类似方式为其进行查重处理。

[0072] 上述各实施例中,针对工单查重过程中所涉及到的工单查重装置200可以是配置于计算设备或者计算设备集群上的软件,并且,通过在计算设备或者计算设备集群上运行该软件,可以使得计算设备或者计算设备集群实现上述工单查重装置200所具有的功能。下面,基于硬件设备实现的角度,对工单查重过程中所涉及到的工单查重装置200进行详细介绍。

[0073] 图7示出了一种计算设备的结构示意图,上述工单查重装置200可以部署在该计算设备上,该计算设备可以是云环境中的计算设备(如服务器),或边缘环境中的计算设备,或终端设备等具体可以用于实现上述图3所示实施例中工单查重装置200中交互模块201、要素挖掘模块202、查重模块203以及模型训练模块204的功能。

[0074] 如图7所示,计算设备600包括处理器610、存储器620、通信接口630和总线640。处理器610、存储器620和通信接口630之间通过总线640通信。总线640可以是外设部件互连标准(peripheral component interconnect,PCI)总线或扩展工业标准结构(extended industry standard architecture,EISA)总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示,图7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。通信接口630用于与外部通信,例如接收用户提供的待查重工单,向用户呈现目标工单等。

[0075] 其中,处理器610可以为中央处理器(central processing unit,CPU)、专用集成电路(application specific integrated circuit,ASIC)、图形处理器(graphics processing unit,GPU)或者一个或多个集成电路。处理器610还可以是一种集成电路芯片,具有信号的处理能力。在实现过程中,工单查重装置200中各个模块的功能可以通过处理器610中的硬件的集成逻辑电路或者软件形式的指令完成。处理器610还可以是通用处理器、数据信号处理器(digital signal process,DSP)、现场可编程逻辑门阵列(fieldprogrammable gate array,FPGA)或者其他可编程逻辑器件,分立门或者晶体管逻辑器件,分立硬件组件,可以实现或者执行本申请实施例中公开的方法、步骤及逻辑框图。其中,通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等,结合本申请实施例所公开的方法可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器、闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器620,处理器610读取存储器620中的信息,结合其硬件完成工单查重装置200中的部分或全部功能。

[0076] 存储器620可以包括易失性存储器(volatile memory),例如随机存取存储器(random access memory,RAM)。存储器620还可以包括非易失性存储器(non-volatile memory),例如只读存储器(read-only memory,ROM),快闪存储器,HDD或SSD。

[0077] 存储器620中存储有可执行代码,处理器610执行该可执行代码以执行前述工单查

重装置200所执行的方法。

[0078] 具体地,在实现图3所示实施例的情况下,且图3所示实施例中所描述的工单查重装置200为通过软件实现的情况下,执行图3中的工单查重装置200的功能所需的软件或程序代码存储在存储器620中,工单查重装置200与其它设备的交互通过通信接口630实现,处理器用于执行存储器620中的指令,实现工单查重装置200所执行的方法。

[0079] 图8示出的一种计算设备集群的结构示意图。其中,图8所示的计算设备集群70包括多个计算设备,上述工单查重装置200可以分布式地部署在该计算设备集群70中的多个计算设备上。如图8所示,计算设备集群70包括多个计算设备700,每个计算设备700包括存储器720、处理器710、通信接口730以及总线740,其中,存储器720、处理器710、通信接口730通过总线740实现彼此之间的通信连接。

[0080] 处理器710可以采用CPU、GPU、ASIC或者一个或多个集成电路。处理器710还可以是一种集成电路芯片,具有信号的处理能力。在实现过程中,工单查重装置200的部分功能可用通过处理器710中的硬件的集成逻辑电路或者软件形式的指令完成。处理器710还可以是DSP、FPGA、通用处理器、其他可编程逻辑器件,分立门或者晶体管逻辑器件、分立硬件组件,可以实现或者执行本申请实施例中公开的部分方法、步骤及逻辑框图。其中,通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等,结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器、闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器720,在每个计算设备700中,处理器710读取存储器720中的信息,结合其硬件可以完成工单查重装置200的部分功能。

[0081] 存储器720可以包括ROM、RAM、静态存储设备、动态存储设备、硬盘(例如SSD、HDD)等。存储器720可以存储程序代码,例如,用于实现交互模块201的部分或者全部程序代码、用于实现要素挖掘模块202的部分或者全部程序代码、用于实现查重模块203的部分或者全部程序代码、用于实现模型训练模块204的部分或全部程序代码等。针对每个计算设备700,当存储器720中存储的程序代码被处理器710执行时,处理器710基于通信接口730执行工单查重装置200所执行的部分方法,如其中一部分计算设备700可以用于执行上述交互模块201和要素挖掘模块202所执行的方法,另一部分计算设备700用于执行上述查重模块203执行的方法,还有一部分计算设备700用于执行上述模型训练模块204所执行的方法。存储器720还可以存储数据,例如:处理器710在执行过程中产生的中间数据或结果数据,例如,属于查重要素的要素内容、目标工单等。

[0082] 每个计算设备700中的通信接口703用于与外部通信,例如与其它计算设备700进行交互等。

[0083] 总线740可以是外设部件互连标准总线或扩展工业标准结构总线等。为便于表示,图8中每个计算设备700内的总线740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。

[0084] 上述多个计算设备700之间通过通信网络建立通信通路,以实现工单查重装置200的功能。任一计算设备可以是云环境中的计算设备(例如,服务器),或边缘环境中的计算设备,或终端设备。

[0085] 本申请实施例还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可以是计算设备能够存储的任何可用介质或者是包含一个或多个可用介质的数据中心等数据存储设备。所述可用介质可以是磁性介质, (例如, 软盘、硬盘、磁带)、光介质 (例如, DVD)、或者半导体介质 (例如固态硬盘) 等。该计算机可读存储介质包括指令, 所述指令指示计算设备执行上述应用于工单查重装置200用于执行工单查重方法。

[0086] 本申请实施例还提供了一种包含指令的计算机程序产品。所述计算机程序产品可以是包含指令的, 能够运行在计算设备上或被储存在任何可用介质中的软件或程序产品。当所述计算机程序产品在至少一个计算机设备上运行时, 使得至少一个计算设备执行上述工单查重方法。另外需说明的是, 以上所描述的装置实施例仅仅是示意性的, 其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的, 作为单元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。另外, 本申请提供的装置实施例附图中, 模块之间的连接关系表示它们之间具有通信连接, 具体可以实现为一条或多条通信总线或信号线。

[0087] 通过以上的实施方式的描述, 所属领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件的方式来实现, 当然也可以通过专用硬件包括专用集成电路、专用CPU、专用存储器、专用元器件等来实现。一般情况下, 凡由计算机程序完成的功能都可以很容易地用相应的硬件来实现, 而且, 用来实现同一功能的具体硬件结构也可以是多种多样的, 例如模拟电路、数字电路或专用电路等。但是, 对本申请而言更多情况下软件程序实现是更佳实施方式。基于这样的理解, 本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来, 该计算机软件产品存储在可读取的存储介质中, 如计算机的软盘、U盘、移动硬盘、ROM、RAM、磁碟或者光盘等, 包括若干指令用以使得一台计算设备 (可以是个人计算机, 训练设备, 或者网络设备等) 执行本申请各个实施例所述的方法。

[0088] 在上述实施例中, 可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时, 可以全部或部分地以计算机程序产品的形式实现。

[0089] 所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时, 全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中, 或者从一个计算机可读存储介质向另一计算机可读存储介质传输, 例如, 所述计算机指令可以从一个网站站点、计算机、训练设备或数据中心通过有线 (例如同轴电缆、光纤、数字用户线 (DSL)) 或无线 (例如红外、无线、微波等) 方式向另一个网站站点、计算机、训练设备或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存储的任何可用介质或者是包含一个或多个可用介质集成的训练设备、数据中心等数据存储设备。所述可用介质可以是磁性介质, (例如, 软盘、硬盘、磁带)、光介质 (例如, DVD)、或者半导体介质 (例如固态硬盘 (Solid State Disk, SSD)) 等。

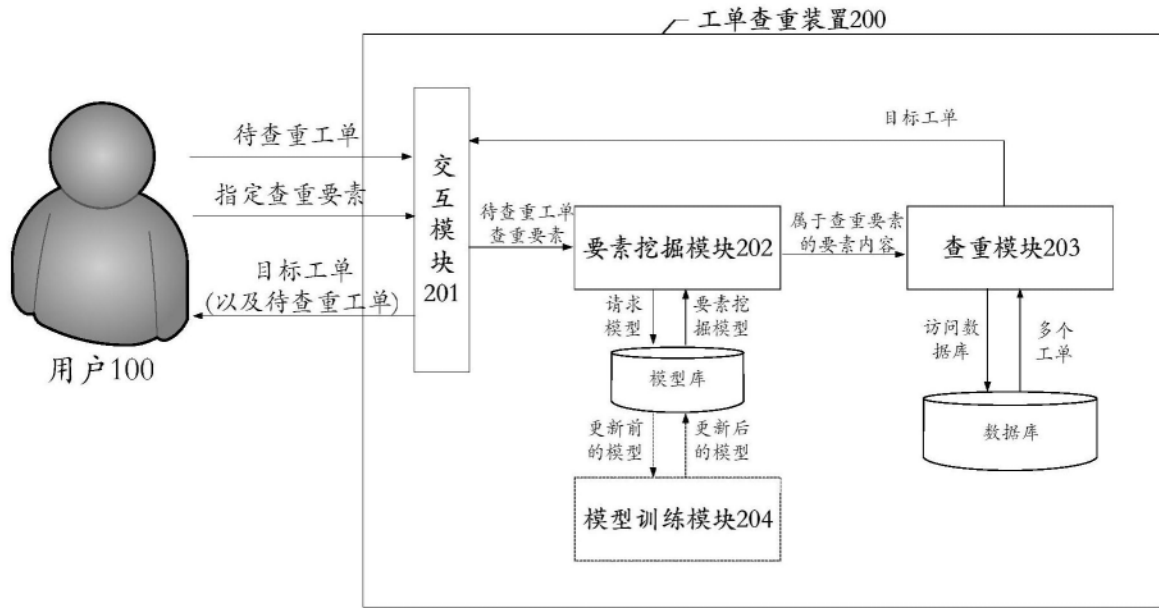


图1

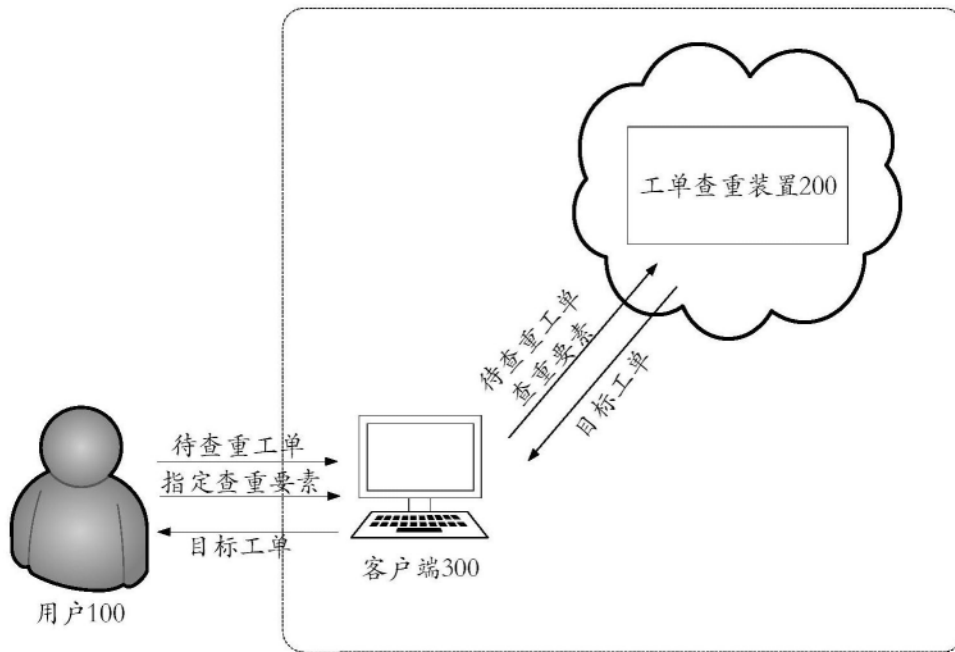


图2

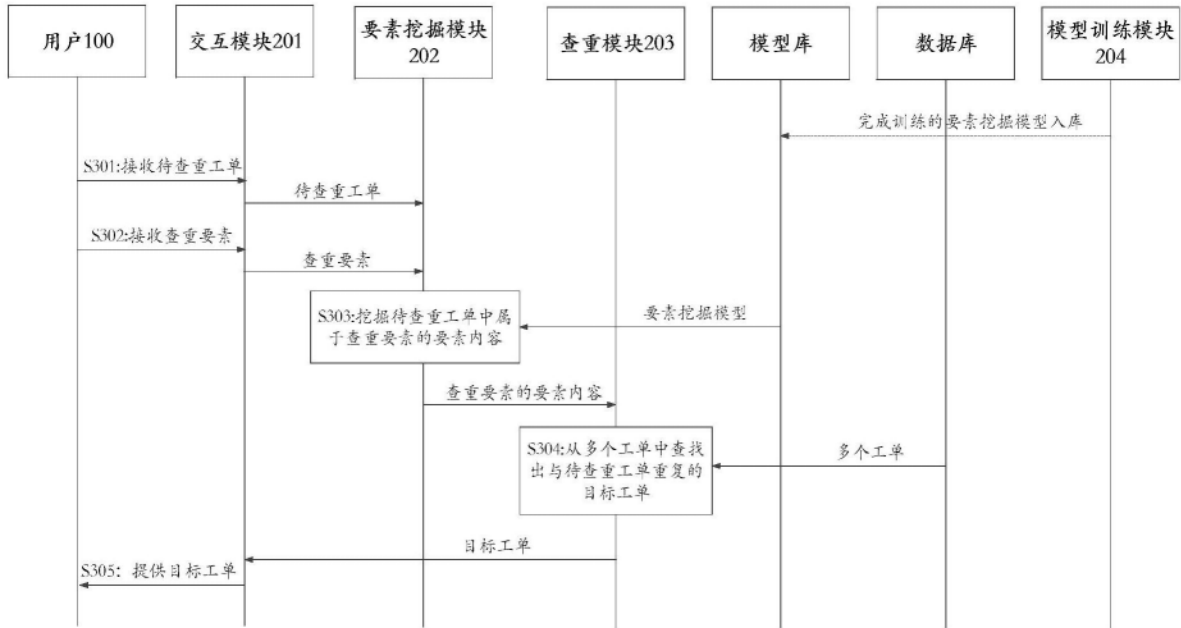


图3

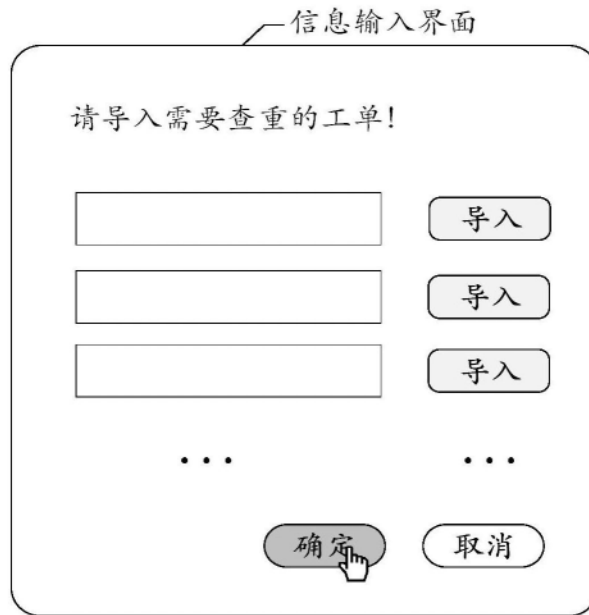


图4

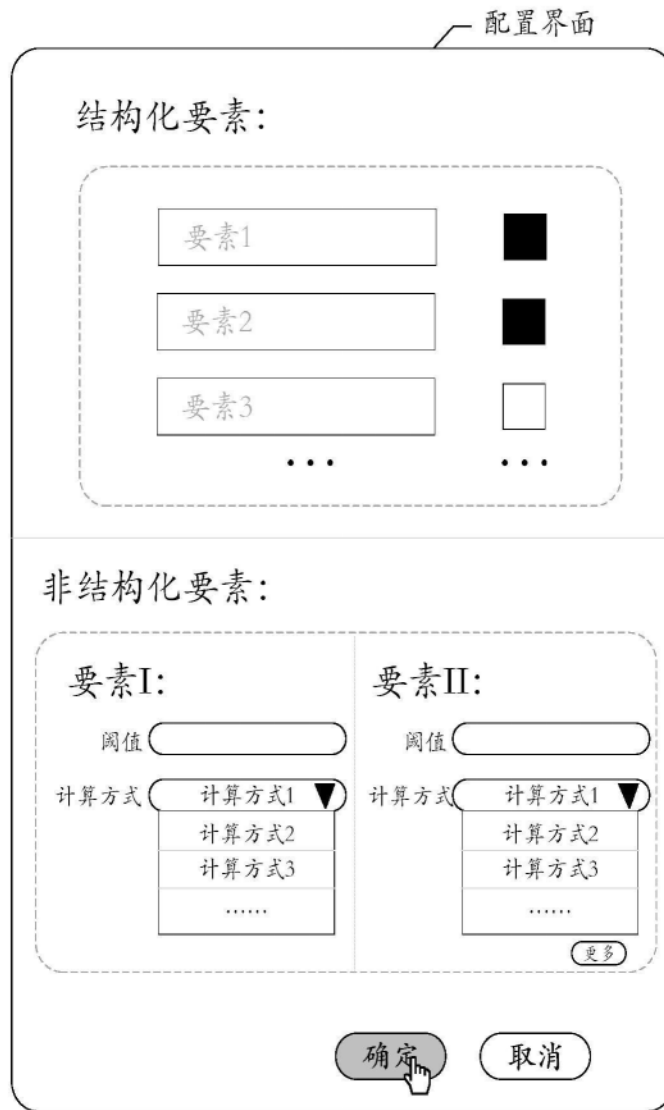


图5

校验界面

序号	工单标识	工单内容	相似度	时间	工单状态	操作
1	▼ ID1	XX投诉XX	--	xx年xx月xx日	待处理	不并单
-	id1	XX投诉XX	--	xx年xx月xx日	待处理	并单
-	id2	XX举报XX	--	xx年xx月xx日	待处理	并单
2	▶ ID2	XX举报XX	--	xx年xx月xx日	待处理	不并单
3	▶ ID3	XX反映XX	--	xx年xx月xx日	待处理	不并单

图6

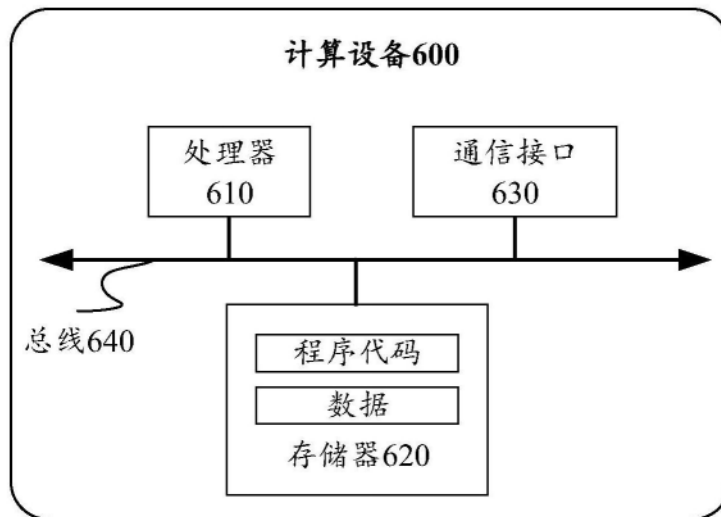


图7

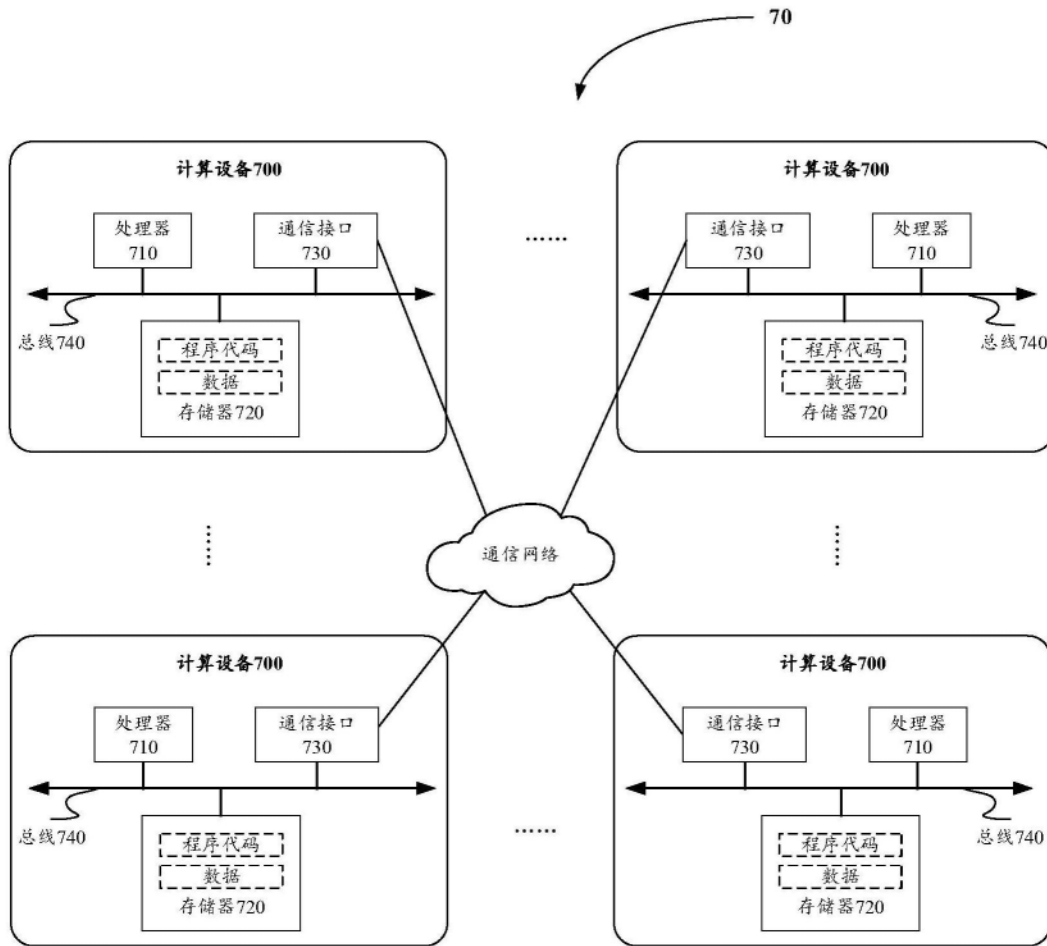


图8