



(12)发明专利申请

(10)申请公布号 CN 109062921 A

(43)申请公布日 2018.12.21

(21)申请号 201810556349.1

(22)申请日 2018.05.31

(71)申请人 武昌船舶重工集团有限公司

地址 430060 湖北省武汉市武昌区张之洞
路2号

(72)发明人 刘文波 谢五一 杨俊

(74)专利代理机构 武汉智权专利代理事务所
(特殊普通合伙) 42225

代理人 张凯

(51)Int.Cl.

G06F 17/30(2006.01)

G06K 9/00(2006.01)

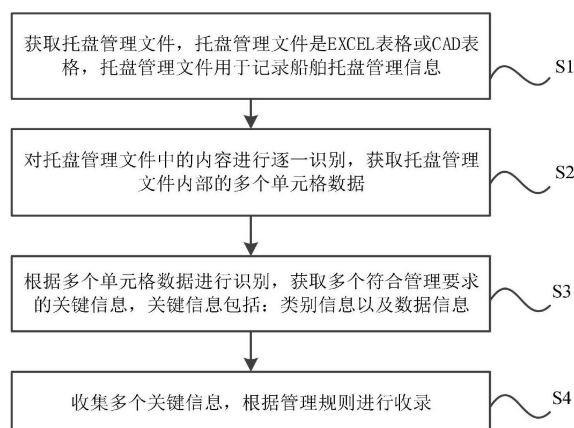
权利要求书2页 说明书9页 附图2页

(54)发明名称

一种提取船舶托盘管理信息的方法及系统

(57)摘要

本发明公开了一种提取船舶托盘管理信息的方法及系统,涉及船舶信息管理技术领域,该方法包括以下步骤:S1、获取托盘管理文件,托盘管理文件是EXCEL表格或CAD表格,托盘管理文件用于记录船舶托盘管理信息;S2、对托盘管理文件中的内容进行逐一识别,获取托盘管理文件内部的多个单元格数据;S3、根据多个单元格数据进行识别,获取多个符合管理要求的关键信息,关键信息包括:类别信息以及数据信息;S4、收集多个关键信息,根据管理规则进行收录。本发明对EXCEL表格或CAD表格形式的托盘管理文件进行筛选,获得进行船舶托盘管理的关键信息,便于进行船舶建造的管理工作,提高工作效率。



1. 一种提取船舶托盘管理信息的方法,其特征在于,所述方法包括以下步骤:

S1、获取托盘管理文件,所述托盘管理文件是EXCEL表格或CAD表格,所述托盘管理文件用于记录船舶托盘管理信息;

S2、对所述托盘管理文件中的内容进行逐一识别,获取托盘管理文件内部的多个单元格数据;

S3、根据多个所述单元格数据进行识别,获取多个符合管理要求的关键信息,所述关键信息包括:类别信息以及数据信息;

S4、收集多个所述关键信息,根据管理规则进行收录。

2. 如权利要求1所述的提取船舶托盘管理信息的方法,其特征在于,当所述托盘管理文件为EXCEL表格时,步骤S2具体包括以下步骤:

S201、识别托盘管理文件中的页面数量,对托盘管理文件中的每一页面进行识别;

S202、逐一识别各页面中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;

S203、将多个所述关键信息进行收录。

3. 如权利要求2所述的提取船舶托盘管理信息的方法,其特征在于,当所述托盘管理文件为EXCEL表格时,步骤S2还包括以下步骤:

识别页面中的单元格的行高,当行高为0时,则跳过该单元格,识别下一单元格。

4. 如权利要求1所述的提取船舶托盘管理信息的方法,其特征在于,当所述托盘管理文件为CAD表格时,步骤S2具体包括以下步骤:

S201、识别CAD表格中的直线以及文本内容;

S202、结合所述直线以及所述文本内容,形成预处理表格;

S203、逐一识别预处理表格中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;

S204、将多个所述关键信息进行收录。

5. 如权利要求4所述的提取船舶托盘管理信息的方法,其特征在于,所述S202具体包括以下步骤:

根据所述CAD表格中的所述直线,识别CAD表格中各直线的交点,并确定各文本内容的位置;

根据各直线以及各交点,形成第一预处理表格;

根据各文本内容的位置以及所述第一预处理表格形成所述预处理表格。

6. 如权利要求2所述的提取船舶托盘管理信息的方法,其特征在于,当所述托盘管理文件为CAD表格时,其特征在于,步骤S2具体包括以下步骤:

S201、手动根据CAD表格中的直线以及文本内容,形成预处理表格;

S202、逐一识别预处理表格中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;

S203、将多个所述关键信息进行收录。

7. 一种提取船舶托盘管理信息的系统,其特征在于,所述系统包括:

托盘管理文件获取模块,其用于获取托盘管理文件,所述托盘管理文件可以是EXCEL表格或CAD表格;

托盘管理文件识别模块,其用于对所述托盘管理文件中的内容进行逐一识别,获取托盘管理文件内部的多个单元格数据;

关键信息提取模块,其用于根据多个所述单元格数据进行识别,获取多个符合管理要求的关键信息,所述关键信息包括:类别信息以及数据信息;

关键信息收录模块,其用于收集多个所述关键信息,根据管理规则进行收录。

8.如权利要求7所述的提取船舶托盘管理信息的系统,其特征在于,当所述托盘管理文件为EXCEL表格时,托盘管理文件识别模块的工作流程具体包括以下步骤:

A1、识别托盘管理文件中的页面数量,对托盘管理文件中的每一页面进行识别;

A2、逐一识别各页面中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;

A3、将多个所述关键信息进行收录。

9.如权利要求7所述的提取船舶托盘管理信息的系统,其特征在于,当所述托盘管理文件为EXCEL表格时,托盘管理文件识别模块的工作流程具体包括以下步骤:

B1、识别CAD表格中的直线以及文本内容;

B2、结合所述直线以及所述文本内容,形成预处理表格;

B3、逐一识别预处理表格中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;

B4、将多个所述关键信息进行收录。

10.如权利要求7所述的提取船舶托盘管理信息的系统,其特征在于,托盘管理文件识别模块的进行步骤B2操作时,具体包括以下步骤:

根据所述CAD表格中的所述直线,识别CAD表格中各直线的交点,并确定各文本内容的位置;

根据各直线以及各交点,形成第一预处理表格;

根据各文本内容的位置以及所述第一预处理表格形成所述预处理表格。

一种提取船舶托盘管理信息的方法及系统

技术领域

[0001] 本发明涉及船舶信息管理技术领域,具体涉及一种提取船舶托盘管理信息的方法及系统。

背景技术

[0002] 随着信息化建设的不断深入,管理部门对生产设计信息的需求与日俱增。

[0003] 在生产设计过程中,产生大量的产品数据,大量数据保存在数据库中,设计人员在抽取数据库数据形成安装托盘表后,托盘表生成过程中可能会手工修改一些数据,导致托盘表数据与数据库数据不一致,因此,急需一种能够保证托盘表数据的正确性与一致性,利于后期按分段、区域或全船进行材料汇总的方法和系统,从而提高工作效率。

发明内容

[0004] 针对现有技术中存在的缺陷,本发明的目的在于提供一种提取船舶托盘管理信息的方法及系统,对EXCEL表格或CAD表格形式的托盘管理文件进行筛选,获得进行船舶托盘管理的关键信息,便于进行船舶建造的管理工作,提高工作效率。

[0005] 为达到以上目的,本发明采取的技术方案是:

[0006] 一种提取船舶托盘管理信息的方法,所述方法包括以下步骤:

[0007] S1、获取托盘管理文件,所述托盘管理文件是EXCEL表格或CAD表格,所述托盘管理文件用于记录船舶托盘管理信息;

[0008] S2、对所述托盘管理文件中的内容进行逐一识别,获取托盘管理文件内部的多个单元格数据;

[0009] S3、根据多个所述单元格数据进行识别,获取多个符合管理要求的关键信息,所述关键信息包括:类别信息以及数据信息;

[0010] S4、收集多个所述关键信息,根据管理规则进行收录。

[0011] 在上述技术方案的基础上,当所述托盘管理文件为EXCEL表格时,步骤S2具体包括以下步骤:

[0012] S201、识别托盘管理文件中的页面数量,对托盘管理文件中的每一页面进行识别;

[0013] S202、逐一识别各页面中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;

[0014] S203、将多个所述关键信息进行收录。

[0015] 在上述技术方案的基础上,当所述托盘管理文件为EXCEL表格时,步骤S2还包括以下步骤:

[0016] 识别页面中的单元格的行高,当行高为0时,则跳过该单元格,识别下一单元格。

[0017] 在上述技术方案的基础上,当所述托盘管理文件为CAD表格时,步骤S2具体包括以下步骤:

[0018] S201、识别CAD表格中的直线以及文本内容;

- [0019] S202、结合所述直线以及所述文本内容,形成预处理表格;
- [0020] S203、逐一识别预处理表格中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;
- [0021] S204、将多个所述关键信息进行收录。
- [0022] 在上述技术方案的基础上,所述S202具体包括以下步骤:
- [0023] 根据所述CAD表格中的所述直线,识别CAD表格中各直线的交点,并确定各文本内容的位置;
- [0024] 根据各直线以及各交点,形成第一预处理表格;
- [0025] 根据各文本内容的位置以及所述第一预处理表格形成所述预处理表格。
- [0026] 在上述技术方案的基础上,当所述托盘管理文件为CAD表格时,其特征在于,步骤S2具体包括以下步骤:
- [0027] S201、手动根据CAD表格中的直线以及文本内容,形成预处理表格;
- [0028] S202、逐一识别预处理表格中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;
- [0029] S203、将多个所述关键信息进行收录。
- [0030] 本发明还公开一种提取船舶托盘管理信息的系统,所述系统包括:
- [0031] 托盘管理文件获取模块,其用于获取托盘管理文件,所述托盘管理文件可以是EXCEL表格或CAD表格;
- [0032] 托盘管理文件识别模块,其用于对所述托盘管理文件中的内容进行逐一识别,获取托盘管理文件内部的多个单元格数据;
- [0033] 关键信息提取模块,其用于根据多个所述单元格数据进行识别,获取多个符合管理要求的关键信息,所述关键信息包括:类别信息以及数据信息;
- [0034] 关键信息收录模块,其用于收集多个所述关键信息,根据管理规则进行收录。
- [0035] 在上述技术方案的基础上,当所述托盘管理文件为EXCEL表格时,托盘管理文件识别模块的工作流程具体包括以下步骤:
- [0036] A1、识别托盘管理文件中的页面数量,对托盘管理文件中的每一页面进行识别;
- [0037] A2、逐一识别各页面中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;
- [0038] A3、将多个所述关键信息进行收录。
- [0039] 在上述技术方案的基础上,当所述托盘管理文件为EXCEL表格时,托盘管理文件识别模块的工作流程具体包括以下步骤:
- [0040] B1、识别CAD表格中的直线以及文本内容;
- [0041] B2、结合所述直线以及所述文本内容,形成预处理表格;
- [0042] B3、逐一识别预处理表格中的单元格,将各单元格中的类别信息与所述管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成所述关键信息;
- [0043] B4、将多个所述关键信息进行收录。
- [0044] 在上述技术方案的基础上,托盘管理文件识别模块的进行步骤B2操作时,具体包括以下步骤:
- [0045] 根据所述CAD表格中的所述直线,识别CAD表格中各直线的交点,并确定各文本内

容的位置；

[0046] 根据各直线以及各交点，形成第一预处理表格；

[0047] 根据各文本内容的位置以及所述第一预处理表格形成所述预处理表格。

[0048] 与现有技术相比，本发明的优点在于：

[0049] (1) 本发明对EXCEL表格或CAD表格形式的托盘管理文件进行筛选，获得进行船舶托盘管理的关键信息，便于进行船舶建造的管理工作，提高工作效率。

附图说明

[0050] 图1为本发明实施例1中提取船舶托盘管理信息的方法的步骤图；

[0051] 图2为本发明实施例2中提取船舶托盘管理信息的方法的步骤图；

[0052] 图3为本发明实施例3中提取船舶托盘管理信息的方法中步骤S2的步骤流程图；

[0053] 图4为本发明实施例4中提取船舶托盘管理信息的方法中步骤S2的步骤流程图；

具体实施方式

[0054] 以下结合附图对本发明的实施例作进一步详细说明。

[0055] 实施例1

[0056] 参见图1所示，本发明实施例1提供一种提取船舶托盘管理信息的方法，该方法包括以下步骤：

[0057] S1、获取托盘管理文件，托盘管理文件是EXCEL表格或CAD表格，托盘管理文件用于记录船舶托盘管理信息；

[0058] S2、对托盘管理文件中的内容进行逐一识别，获取托盘管理文件内部的多个单元格数据；

[0059] S3、根据多个单元格数据进行识别，获取多个符合管理要求的关键信息，关键信息包括：类别信息以及数据信息；

[0060] S4、收集多个关键信息，根据管理规则进行收录。

[0061] 本发明中，首先获取存储了船舶托盘管理信息的托盘管理文件，而托盘管理文件可以是EXCEL表格或CAD表格，具体根据数据采集设备等因素进行选择；

[0062] 而后对托盘管理文件中的内容进行逐一识别，获取托盘管理文件内部的多个单元格数据，由于托盘管理文件可以是EXCEL表格或CAD表格，因此托盘管理文件内可以包含有表格数据以及其他数据，而船舶托盘管理信息都是根据存储规则存储于托盘管理文件的单元格中，因此需要进行逐一识别，从而获取托盘管理文件内部的多个单元格数据，以便进一步进行数据的筛选；

[0063] 根据多个单元格数据进行识别，获取多个符合管理要求的关键信息，关键信息包括：类别信息以及数据信息，具体识别方式可以是识别关键字，从而对类别信息进行识别，从而寻找到需要获取数据的单元格，从而获取单元格的数据，即关键信息。

[0064] 最终，收集多个关键信息，根据管理规则进行收录。

[0065] 本发明对EXCEL表格或CAD表格形式的托盘管理文件进行筛选，获得进行船舶托盘管理的关键信息，便于进行船舶建造的管理工作，提高工作效率。

[0066] 实施例2

[0067] 参见图2所示,本发明实施例2提供一种提取船舶托盘管理信息的方法,在实施例1的基础上:

[0068] 当托盘管理文件为EXCEL表格时,步骤S2具体包括以下步骤:

[0069] S201、识别托盘管理文件中的页面数量,对托盘管理文件中的每一页面进行识别;

[0070] S202、逐一识别各页面中的单元格,将各单元格中的类别信息与管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成关键信息;

[0071] S203、将多个关键信息进行收录。

[0072] 需要说明的是,以下提到的EXCEL表格的页面,均为EXCEL表格的sheet。

[0073] 本实施例中,当托盘管理文件为EXCEL表格时,步骤S2还包括以下步骤:

[0074] 识别页面中的单元格的行高,当行高为0时,则跳过该单元格,识别下一单元格。

[0075] 在处理EXCEL表格时,存在以下几个特殊操作

[0076] 1、面对合并单元格时,对于合并单元格,数据存放在这些单元格中行数和列数均为最小的单元格中,而因为格式要求,每一行的单元格的合并形式应该是一致的,因此,所有的数据与其相对应的数据项仍然在同一列;

[0077] 亦即是说,通过引入整型数组resPos[]记录一行中不同数据项所在的列数,只取获得了数据项的列进行数据提取,便可排除其它无数据的单元格。因此,本文的数据提取方式能够应对普遍存在的合并单元格的情况而不受影响;

[0078] 2、面对单元格的不同数据类型:

[0079] 单元格的数据类型存在多种,有字符串列、数值列还有公式列;

[0080] 引入一个获取单元格的值的方法,若单元格中的数据为字符串,则返回该字符串;

[0081] 若单元格中的数据为数值,则返回该数值转换为的字符串,由于空白单元格不需要提取数据,所以该方法返回空字符串;

[0082] 而在数据有效性的判别中,需要依照所提取的单元格中的数据所对应的数据项的格式要求来进行甄选,此时只需要该数据的值判定为非空字符串即可剔除掉空白行、空白列以及公式列;

[0083] 3、面对隐藏行或列:

[0084] 数据中存在隐藏行或隐藏列,这些行或列中的数据为暂时不用的数据,是不需要进行数据提取的;

[0085] 根据隐藏行的行高以及隐藏列的列宽为0的特性可以在程序中加入判断从而达到忽略隐藏行或列的数据的目的;

[0086] 一旦这些行或列取消隐藏,它们的行高或列宽便会发生变化,于是这些行或列中的数据便能通过程序进行提取了。

[0087] 本实施例中,面对EXCEL表格时,在数据提取过程存在以下几个优势:

[0088] 1、准确地找到数据所在的位置

[0089] 由于数据所在的页面以及数据的起始行或起始列不确定,所以程序必须能够准确地找到数据所在的位置

[0090] EXCEL托盘表中的数据是以一个个数据表单的形式存放的,这些表单可能会在同一个页面中。

[0091] 本发明的数据提取方式是通过设立关键字来定位数据所在的位置的,关键字分为

类别关键字和属性关键字,其中类别关键字对应表单的类别,属性关键字则对应该类表单具有的数据项;

[0092] 该数据提取方法的主要特点是通过这两类关键字进行匹配查找,可以实现对EXCEL表内容的全文检索,只要内部存在特定的信息并满足提取条件(根据船舶物料描述的有关信息要求,即描述属性的数目来衡量是数据的有效性),不论在文件中的任何地方,都能获取到信息;

[0093] 通过对所有sheet的每一行的单元格进行遍历,分别找到包含类别关键字的单元格所在的行数以及包含属性关键字的单元格所在的列数,从而获取到数据所在的位置。同时,通过对找到的不同关键字的判别,还能确认数据所属的表单以及数据项,从而确保了数据提取的准确性;

[0094] 本数据提取方式可在不破坏表单的完整性以及应有格式的前提下无论怎么改变数据所在的sheet、行数以及列数程序都能准确地找到数据所在的位置。

[0095] 2、能够对图号进行提取:

[0096] 托盘表内大多存在图号信息,图号中有产品代码、专业及分段或区域码,这些信息对于材料按产品、阶段统计是非常关键的。

[0097] 图号提取过程如下:

[0098] 由于封面所在的页面不定,因此需要先获取托盘表中页面的数目,然后依次对每个页面进行遍历,

[0099] 获取页面包含的行数,然后循环取得一行,判断是否是隐藏行,如果是跳过执行下一行,若不是则获取其所包含的列数,循环取出一列,判断是否是隐藏列,如果是则跳过执行下一列,若不是,则获取其值,比较其内容是否是“生产设计”,若是则记下所在的列,往后循环五行(主要依据图表模板的规定的样式,图号信息所处的列与“生产设计”列相同,其信息在其行号后面五行之内,所以为了提高分析数据的效率,超过五行后的数据就不进行分析),取指定的列,取出其列中包含的值,判断其是否包含图号取名规则所用的字符串,若包含则,该值为图号,若循环完都没有找到图号则该文件格式不符合托盘表格式规范;

[0100] 3、托盘信息的提取

[0101] 托盘信息的数据提取过程如下:

[0102] 1) 定义关键字

[0103] 在系统的配置文件中定义关键字,有利于程序的维护,即当关键字发生变化时只需要修改相应的配置文件,而不用修改程序代码,

[0104] 类别关键字定义了EXCEL托盘表中可能出现的不同表单的表名,属性关键字则定义了这些表单中可能出现的数据项。

[0105] 2) 获取所有的页面

[0106] 由于数据所在的页面不定,可能是第一个页面也可能是其它页面,因此需要先获取托盘表中页面的数目,然后依次对每个页面进行遍历。

[0107] 3) 获取页面中的行

[0108] 在数据所在的页面中,逐行遍历,可以依次找到类别关键字所对应的表名和属性关键字所对应的数据项,利用这些关键字可以准确地提取EXCEL托盘表中不同表单的数据并存入相应的数据库表中。

[0109] 具体过程如下:

[0110] 获取EXCEL表格页面包含的行数,依次对行进行遍历,遍历过程中获取行高,当行高等于0时,可以判断此行为设计人员隐藏的行,对于这类行中的数据不需要进行提取,程序转到执行下一行数据的提取;

[0111] 对于每一行,获取该行所包含的单元格数目,依次对行中的各单元格进行遍历,首先获取单元格类型,根据类型对值进行转换,尤其对于类型是公式的单元格,需按字符串的方式提取,再转化为数字;对于单元格的内容包含类别关键字所对应的字符串,则用整型变量currentType记录该类别关键字的序号,用以区分不同的数据表单;如若该值包含属性关键字所对应的字符串,则用整型数组的一个元素resPos[i]来记录该属性关键字所在的列数,该整型数组可记录遍历该行所有单元格获取的每个属性关键字所在的列数;

[0112] 整型变量currentType的值在遍历单元格的过程中再次找到包含类别关键字(表名)的单元格时会发生改变,亦即是说currentType的值决定了提取的数据属于哪个表单;

[0113] 整型数组的元素resPos[i],在遍历某一行单元格的过程中找到包含下标i对应的属性关键字(数据项)的值时会进行记录;

[0114] 由于托盘表中各个表单中的数据项一般大于3个(即船舶零件的属性条目的数量),所以遍历一行所得的属性关键字的个数大于3,则表明下述内容为船舶零部件托盘清单信息,而对于关键字个数少于或等于3的数据是需要过滤的;

[0115] 通过上述判断可以分析出托盘信息表的数据结构,依据此结构对信息进行全文检索,通过循环取得整型数组resPos[]每一个元素resPos[i]的值,获得其值对应的表格中的列数据。

[0116] 采用本方法对EXCEL中的数据进行全文检索,只需要保证设计人员在设计船舶托盘表时采用定义的属性关键字,在托盘信息所在的SHEET,零件属性所在的列和顺序不确定的情况下,依然可以准确地提取到所需的信息。已在公司产品信息管理系统中进行了部署与应用,实现了设计部门与制造管理部门间的信息共享,提高了物资采购、集配的效率。

[0117] 实施例3

[0118] 参见图3所示,本发明实施例2提供一种提取船舶托盘管理信息的方法,在实施例1的基础上,

[0119] 当托盘管理文件为CAD表格时,步骤S2具体包括以下步骤:

[0120] S201、识别CAD表格中的直线以及文本内容;

[0121] S202、结合直线以及文本内容,形成预处理表格;

[0122] S203、逐一识别预处理表格中的单元格,将各单元格中的类别信息与管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成关键信息;

[0123] S204、将多个关键信息进行收录。

[0124] 本实施例中,S202具体包括以下步骤:

[0125] 根据CAD表格中的直线,识别CAD表格中各直线的交点,并确定各文本内容的位置;

[0126] 根据各直线以及各交点,形成第一预处理表格;

[0127] 根据各文本内容的位置以及第一预处理表格形成预处理表格。

[0128] 在形成预处理表格时,需要进行以下几个步骤:

[0129] 1、提取直线与文本

[0130] 直线和文本是构成表格和数据的基本元素,由于存在设计人员将表格合成“块”的情况,因此必须首先将图中的“块”炸开,便于提取“直线”与“文本”实体,

[0131] 提取“直线”与“文本”有两种方式,一种由程序自动提取,通过遍历文件中的所有元素,依次判断其类别与位置属性;另一种方式是由设计人员手工框选,然后由程序对选择的元素进行遍历,再判断其类别与位置属性,分别将“水平线”、“垂直线”与“文本”归入水平线集合、垂直线集合与文本集合。

[0132] 2、对文本排序

[0133] 对文本集合内的文本按其插入点的位置按从上到下、从左到右排列,文本排序的目的是为了后期与单元格进行关联。

[0134] 进一步,再对单元格计算及数据提取

[0135] 1) 直线交点计算及排序

[0136] 取出所有的竖直线和水平线,在定义某一点及其坐标后,再判断是否满足此点既在取出的竖直线起点与终点之间又在取出的水平线起点与终点之间的条件,若条件满足即可判定为交点。这样就得到了所有交点,并为后续单元格的形成提供数据,

[0137] 对交点集合中的数据应按交点从上到下、从左到右进行排列,便于计算单元格。

[0138] 2) 单元格计算及排序:

[0139] 单元格的计算依据对集合内交点的处理,每个单元格存在四个交点,取左下点和右上点作为单元格的特征点;

[0140] 单元格中交点的计算顺序如下:

[0141] (1) 循环从交点集合中取出第一点;

[0142] (2) 循环从交点集合中取出第一点之后的另一个交点,作为第二点;

[0143] (3) 第一点与第二点进行比较计算,当第一点的Y值等于第二点的Y值,并且其X不等则该交点满足组成单元格第二点的要求,否则执行第二步;

[0144] (4) 循环从交点集合中取出第二点之后的交点;

[0145] (5) 第二点与第三点进行比较计算,当第二点的X值等于第三点的X值,并且其Y值不等则该交点满足第三点的要求,否则执行第四步;

[0146] (6) 循环从交点集合中取出第三点之后的另一个交点,作为第四点;

[0147] (7) 当第四点的Y值等于第三点的Y值,并且第四点的X值等于第一点的X值,则该点满足第四点的要求,否则执行第六步;

[0148] (8) 当求得的四点满足条件时,创建单元格对象,单元格的左下点和右上点作为该单元格的特征点,并将单元格存入单元格集合;

[0149] (9) 对单元格按从上到下,从左到右进行排列。

[0150] 3) 单元格与文本关联:

[0151] 单元格与文本关联是通过判断文本的插入点是否在单元格内,当插入点在单元格内,将该文本赋值给该单元格,执行过程如下:

[0152] (1) 循环从单元格集合中取出一个单元格;

[0153] (2) 循环从文本集合中取出一个文本;

[0154] (3) 当文本插入点的X值、Y值分别大于等于单元格的左下点的X值、Y值,并且分别小于单元格右上点的X值、Y值时满足条件进行关联,执行第一步;否则执行第二步取另一个

文本进行判断。

[0155] 4) 提取表格信息:

[0156] 单元格与文本关联后,则对单元格进行分析,将首尾相连的单元格组成单元行,行的特征点为第一列的左下点及最后一列的右上点。将单元行存放到行集合中,然后对集合内的单元行进行排序,排序的依据是单元行的左下点,按从上到下,从左到右的方式排列;

[0157] 语义关系是指关键字与信息内容的位置关系,通常存在上下关系和左右关系,上图中关键字与信息内容是上下关系,并且它们所在表的列序号相同。

[0158] 关键字是进行数据提取的基础,包含表格信息的图表文件一般都会含有关键字,对单元行集合进行遍历,检索关键字,根据关键字、语义关系等进行判断,当连续单元行关键字的数目达到指定的要求,下面的单元行就是提取信息开始的地方。

[0159] 实施例4

[0160] 参见图4所示,本发明实施例提供一种提取船舶托盘管理信息的方法,该方法在实施例1的基础上,当托盘管理文件为CAD表格时,步骤S2具体包括以下步骤:

[0161] S201、手动根据CAD表格中的直线以及文本内容,形成预处理表格;

[0162] S202、逐一识别预处理表格中的单元格,将各单元格中的类别信息与管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成关键信息;

[0163] S203、将多个关键信息进行收录。

[0164] 相对于EXCEL表格,AutoCAD格式的托盘表也可以看作是一个“表格”,尽管不是真正意义上的“表格”,它可能是“块”也可能是多个直线与文本组成,直线和文本之间除了位置关系之外,没有其它的任何联系,所以为提取数据信息,先提取所有直线与文本信息,同时按位置关系对直线与文本进行排序,便于形成单元格,以及单元格与文本进行关联。

[0165] 实施例5

[0166] 本发明实施例提供一种提取船舶托盘管理信息的系统,该系统包括:

[0167] 托盘管理文件获取模块,其用于获取托盘管理文件,托盘管理文件可以是EXCEL表格或CAD表格;

[0168] 托盘管理文件识别模块,其用于对托盘管理文件中的内容进行逐一识别,获取托盘管理文件内部的多个单元格数据;

[0169] 关键信息提取模块,其用于根据多个单元格数据进行识别,获取多个符合管理要求的关键信息,关键信息包括:类别信息以及数据信息;

[0170] 关键信息收录模块,其用于收集多个关键信息,根据管理规则进行收录。

[0171] 本发明中,首先获取存储了船舶托盘管理信息的托盘管理文件,而托盘管理文件可以是EXCEL表格或CAD表格,具体根据数据采集设备等因素进行选择;

[0172] 而后对托盘管理文件中的内容进行逐一识别,获取托盘管理文件内部的多个单元格数据,由于托盘管理文件可以是EXCEL表格或CAD表格,因此托盘管理文件内可以包含有表格数据以及其他数据,而船舶托盘管理信息都是根据存储规则存储于托盘管理文件的单元格中,因此需要进行逐一识别,从而获取托盘管理文件内部的多个单元格数据,以便进一步进行数据的筛选;

[0173] 根据多个单元格数据进行识别,获取多个符合管理要求的关键信息,关键信息包括:类别信息以及数据信息,具体识别方式可以是识别关键字,从而对类别信息进行识别,

从而寻找到需要获取数据的单元格,从而获取单元格的数据,即关键信息。

[0174] 最终,收集多个关键信息,根据管理规则进行收录。

[0175] 本发明对EXCEL表格或CAD表格形式的托盘管理文件进行筛选,获得进行船舶托盘管理的关键信息,便于进行船舶建造的管理工作,提高工作效率。

[0176] 实施例6

[0177] 本发明实施例提供一种提取船舶托盘管理信息的系统,在实施例5的基础上:

[0178] 当托盘管理文件为EXCEL表格时,托盘管理文件识别模块的工作流程具体包括以下步骤:

[0179] A1、识别托盘管理文件中的页面数量,对托盘管理文件中的每一页面进行识别;

[0180] A2、逐一识别各页面中的单元格,将各单元格中的类别信息与管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成关键信息;

[0181] A3、将多个关键信息进行收录。

[0182] 本实施例中,当托盘管理文件为EXCEL表格时,托盘管理文件识别模块的工作流程具体包括以下步骤:

[0183] B1、识别CAD表格中的直线以及文本内容;

[0184] B2、结合直线以及文本内容,形成预处理表格;

[0185] B3、逐一识别预处理表格中的单元格,将各单元格中的类别信息与管理要求进行比对,将符合管理要求的单元格的类别信息以及数据信息组合成关键信息;

[0186] B4、将多个关键信息进行收录。

[0187] 本实施例中,托盘管理文件识别模块的进行步骤B2操作时,具体包括以下步骤:

[0188] 根据CAD表格中的直线,识别CAD表格中各直线的交点,并确定各文本内容的位置;

[0189] 根据各直线以及各交点,形成第一预处理表格;

[0190] 根据各文本内容的位置以及第一预处理表格形成预处理表格。

[0191] 本发明不仅局限于上述最佳实施方式,任何人在本发明的启示下都可得出其他各种形式的产品,但不论在其形状或结构上作任何变化,凡是具有与本发明相同或相近似的技术方案,均在其保护范围之内。

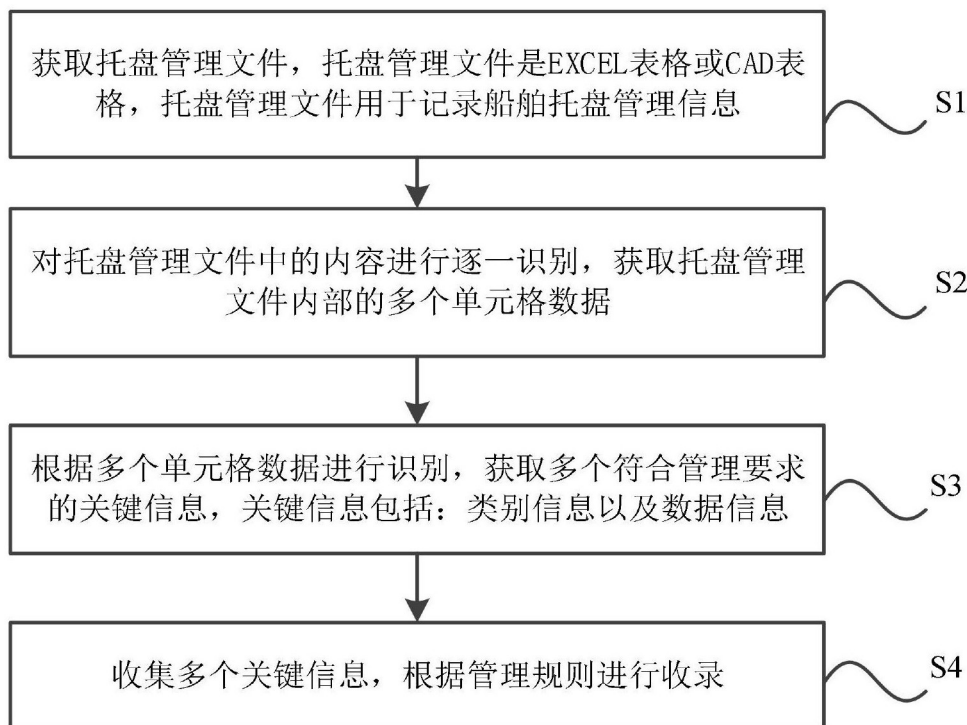


图1

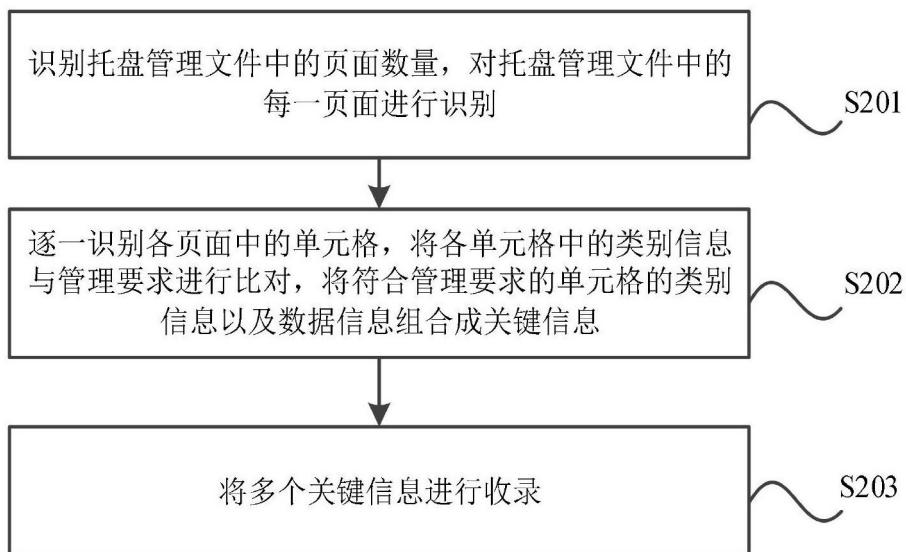


图2

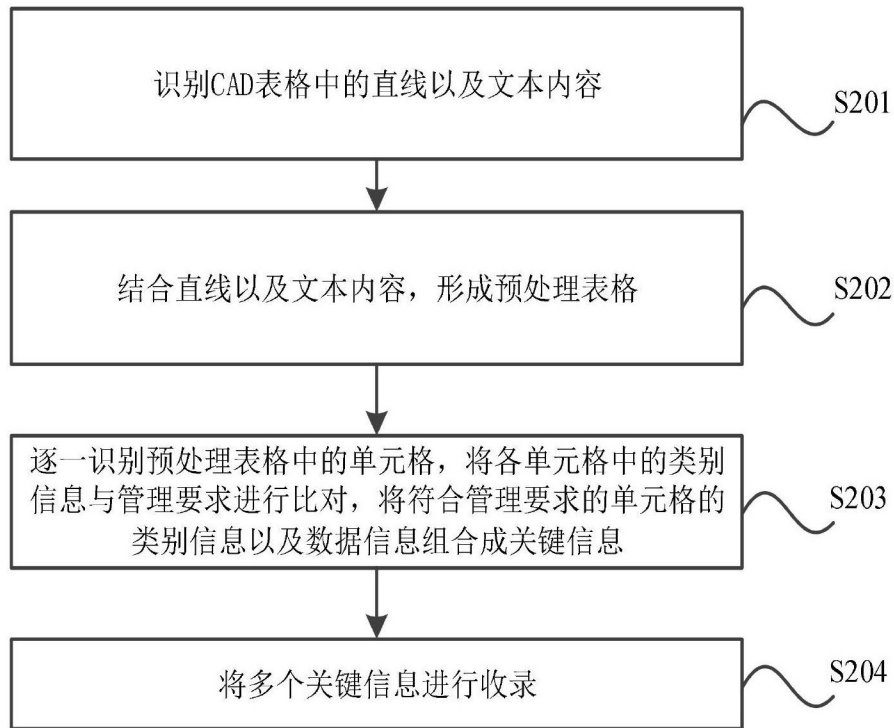


图3

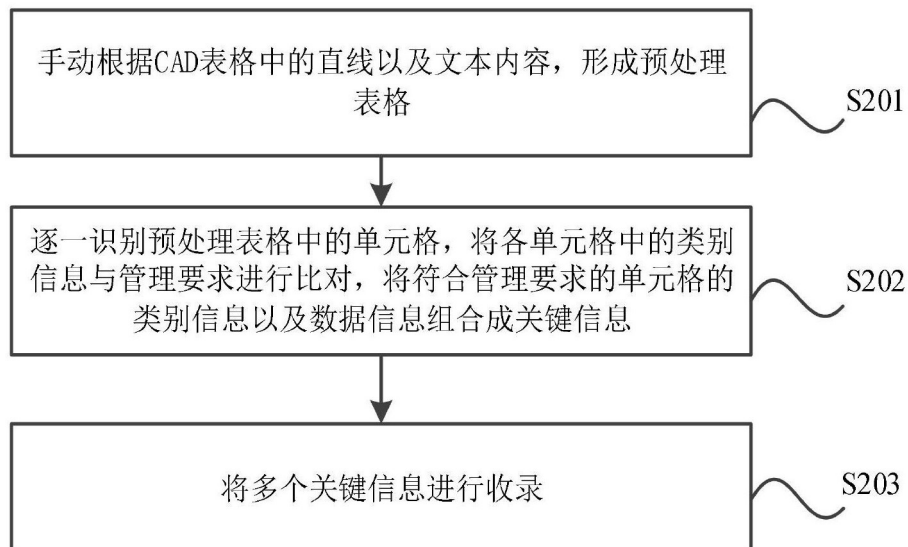


图4