



(12) 发明专利

(10) 授权公告号 CN 114691265 B

(45) 授权公告日 2024. 11. 15

(21) 申请号 202210224918.9

G06F 8/38 (2018.01)

(22) 申请日 2022.03.07

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 113015956 A, 2021.06.22

申请公布号 CN 114691265 A

CN 111460259 A, 2020.07.28

(43) 申请公布日 2022.07.01

审查员 刘冰心

(73) 专利权人 达观数据有限公司

地址 201203 上海市浦东新区博霞路66号

501、502、503室

(72) 发明人 张涛 何双钻 刘迪康 金克

邵万骏 纪达麒 陈运文

(74) 专利代理机构 上海智力专利商标事务所

(普通合伙) 31105

专利代理师 周涛

(51) Int. Cl.

G06F 9/451 (2018.01)

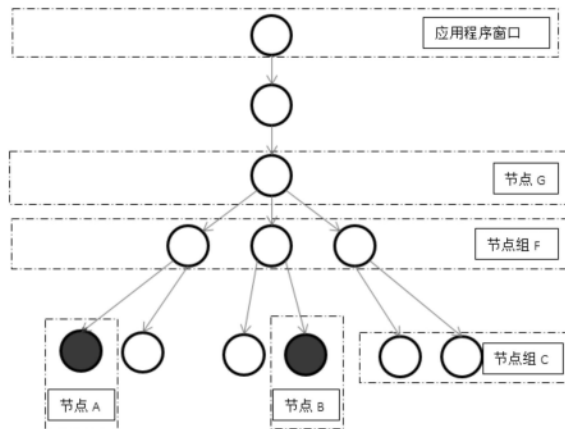
权利要求书1页 说明书4页 附图3页

(54) 发明名称

一种软件界面结构化元素的批量抓取方法

(57) 摘要

本发明涉及一种软件界面结构化元素的批量抓取方法,包括开发流程阶段和运行流程阶段,在开发流程阶段获取三个选择器,三个选择器分别对应祖先节点、孩子节点组和孩子节点组中的目标节点;运行流程阶段RPA机器人通过三个选择器提取结构化数据,获得由目标节点组成的节点组,RPA机器人获取节点组内目标节点的属性并按照节点组和孩子节点组生成结构化数据的二维数组。本发明解决现有技术中RPA机器人获取结构化数据时遇到的系统软件不提供API、数据分页、数据结构不明显的问题,简化了RPA机器人获取结果化数据的过程以及提高RPA机器人获取结果化数据的效率。



1. 一种软件界面结构化元素的批量抓取方法,其特征在于,包括开发流程阶段和运行流程阶段,在开发流程阶段获取三个选择器,三个选择器分别对应祖先节点、孩子节点组和孩子节点组中的目标节点,所述选择器保存了目标节点、孩子节点组以及祖先节点中的元素属性及层级关系;运行流程阶段RPA机器人通过三个选择器提取结构化数据,获得由目标节点组成的节点组,RPA机器人获取节点组内目标节点的属性并按照节点组和孩子节点组生成结构化数据的二维数组;

开发流程阶段包括以下步骤:

步骤一、人工选取结构化数据中的两个控件类型相同的界面元素;

步骤二、借助分析模块分析两个界面元素的层级和位置关系,找到两个所述界面元素的共同祖先节点,将祖先节点的选择器作为第一个选择器;

步骤三、获取祖先节点的所有孩子节点组,将从祖先节点查找孩子节点组的选择器作为第二个选择器;

步骤四、将孩子节点组中的查找控件类型相同的节点的选择器作为第三个选择器;若步骤1中的两个界面元素在孩子节点组层级,则第三个选择器为空;

在运行流程阶段,RPA机器人通过三个选择器提取结构化数据时包括以下步骤:

步骤1、RPA机器人根据第一个选择器找到祖先节点;

步骤2、RPA机器人根据第二个选择器找到孩子节点组;

步骤3、RPA机器人根据第三个选择器,分别从孩子节点组的每个孩子节点出发查找孩子节点下对应的目标节点并形成目标节点组;

步骤4、RPA机器人获取目标节点的指定属性,并按照目标节点组和孩子节点组形成结构化数据的二维数组。

2. 根据权利要求1所述的一种软件界面结构化元素的批量抓取方法,其特征在于,运行流程阶段,当结构数据处于不同页面时,RPA机器人进行翻页操作,获取每页的结构化数据并将获取的结构化数据拼接在一起。

3. 根据权利要求1所述的一种软件界面结构化元素的批量抓取方法,其特征在于,所述分析模块为语法分析算法。

一种软件界面结构化元素的批量抓取方法

技术领域

[0001] 本发明涉及RPA机器人流程自动化领域,具体涉及一种软件界面结构化元素的批量抓取方法。

背景技术

[0002] 机器人流程自动化(Robotic Process Automation,后文简称RPA)是运行在个人计算机或服务上的软件,主要功能是模拟人操控各种系统软件、自动执行业务流程,适用于大批量、高重复的业务场景。RPA机器人的主要功能之一就是获取系统软件界面上的数据,其中就包括结构化数据(结构化数据指的是在一个记录文件里面以固定格式存在的数据,例如表格、目录等)。

[0003] 目前获取结构化数据的主要方法是调用系统软件提供的应用程序编程接口(Application Programming Interface,后文简称API),该方法存在以下问题:

[0004] 1、对于不提供API的系统软件无效;

[0005] 2、对于分页显示的结构化数据无效;

[0006] 3、对于结构不明显的的数据无效。

[0007] 为了解决上述技术问题,现有技术提出的解决方案是分别获取结构化中的每个数据,然后按照发现的规律拼接起来,然而这种解决方案存在以下问题:

[0008] 1、发现结构化数据的规律困难;

[0009] 2、数据量大的时候,单独获取每个数据耗费时间;

[0010] 3、将获取到的数据按照特定规律组合效率低,容易出错。

发明内容

[0011] 为了解决现有技术中RPA机器人获取结构化数据时遇到的系统软件不提供API、数据分页、数据结构不明显的问题,简化RPA机器人获取结构化数据的过程以及提高RPA机器人获取结构化数据的效率,本发明提供了一种软件界面结构化元素的批量抓取方法。

[0012] 本发明的技术目的是通过以下技术方案实现的:

[0013] 一种软件界面结构化元素的批量抓取方法,包括开发流程阶段和运行流程阶段,在开发流程阶段获取三个选择器,三个选择器分别对应祖先节点、孩子节点组和孩子节点组中的目标节点;运行流程阶段RPA机器人通过三个选择器提取结构化数据,获得由目标节点组成的节点组,RPA机器人获取节点组内目标节点的属性并按照节点组和孩子节点组生成结构化数据的二维数组。

[0014] 进一步地,运行流程阶段,当结构数据处于不同页面时,RPA机器人进行翻页操作,获取每页的结构化数据并将获取的结构化数据拼接在一起。

[0015] 进一步地,选择器保存了目标节点、孩子节点组以及祖先节点中的元素属性及层级关系。

[0016] 进一步地,开发流程阶段包括以下步骤:

- [0017] 步骤一、人工选取结构化数据中的两个控件类型相同的界面元素；
- [0018] 步骤二、借助分析模块分析两个界面元素的层级和位置关系,找到两个所述界面元素的共同祖先节点,将祖先节点的选择器作为第一个选择器；
- [0019] 步骤三、获取祖先节点的所有孩子节点组,将从祖先节点查找孩子节点组的选择器作为第二个选择器；
- [0020] 步骤四、将孩子节点组中的查找控件类型相同的节点的选择器作为第三个选择器；若步骤1中的两个界面元素在孩子节点组层级,则第三个选择器为空。
- [0021] 进一步地,在运行流程阶段,RPA机器人通过三个选择器提取结构化数据时包括以下步骤：
- [0022] 步骤1、RPA机器人根据第一个选择器找到祖先节点；
- [0023] 步骤2、RPA机器人根据第二个选择器找到孩子节点组；
- [0024] 步骤3、RPA机器人根据第二个选择器,分别从孩子节点组的每个孩子节点出发查找孩子节点下对应的目标节点并形成目标节点组；
- [0025] 步骤4、RPA机器人获取目标节点的指定属性,并按照目标节点组和孩子节点组形成结构化数据的二维数组。
- [0026] 进一步地,分析模块为语法分析算法。
- [0027] 相比现有技术,本发明的有益效果在于：
- [0028] 1、本发明的软件界面结构化元素的批量抓取方法支持不提供提取结构化数据API的系统软件；
- [0029] 2、通过RPA机器人自动翻页支持获取分页现实的结构化数据；
- [0030] 3、本发明的方法支持提取结构不明显的的数据；
- [0031] 4、本发明的方法操作简单,不需要理解数据间的结构规律；
- [0032] 5、生成结果为三个元素选择器,直观展示了数据的结果关系,方便理解和修改；
- [0033] 6、通过本发明的方法大幅提高获取结构化数据的效率。

附图说明

- [0034] 图1是本发明中系统应用界面元素节点示意图。
- [0035] 图2是实施例1中的组织构架图。
- [0036] 图3是实施例2中的组织构架图。
- [0037] 图4是实施例3中的组织构架图。

具体实施方式

- [0038] 下面结合具体实施方式对本发明的技术方案进行进一步描述：
- [0039] 一种软件界面结构化元素的批量抓取方法,如图1-4所示,包括开发流程阶段和运行流程阶段,选择器是一种选择元素的模式,以文本的形式存储和传递,如CSS选择器,选择器的作用是定位界面元素,选择器中保存了目标界面元素的一些属性和这些元素之间的层级关系,根据选择器中的属性,RPA机器人可以定位到符合选择器条件的一个或多个界面元素。
- [0040] 在开发流程阶段,捕获结构化数据中两个界面元素,分析获取三个选择器：

[0041] 步骤一、人工选取结构化数据中的两个界面元素(简称节点A和节点B),如果这两个界面元素的控件类型(控件类型称为控件类型C)不同,则提示重新选取。

[0042] 步骤二、通过分析模块分析这两个数据的层级和位置关系,分析模块采用通用的语法分析算法,找到这两个元素的共同祖先节点(简称节点G),将节点G的选择器作为第一个选择器。语法分析算法的作用是对输入的两个选择器文本进行分析,根据特定规则生成输出的三个选择器文本,通过对文本的分析处理,间接实现对选择条件的分析处理。

[0043] 步骤三、获取节点G的所有孩子节点组(简称节点组F),将从节点G查找节点组F的选择器作为第二个选择器;

[0044] 步骤四、将从节点组F中的某个节点查找控件类型C的节点的选择器作为第三个选择器;如果节点A和节点B处于节点组F的层级,也即节点A和节点B直接属于节点G,则第三个选择器为空。

[0045] 在运行流程阶段,RPA机器人根据三个选择器提取结构化数据:

[0046] 步骤1、RPA机器人根据第一个选择器找到节点G;

[0047] 步骤2、RPA机器人根据第二个选择器找到节点组F;

[0048] 步骤3、RPA机器人根据第二个选择器,分别从节点组F的每个节点出发,查找所有的目标节点并形成目标节点组(节点组C);

[0049] 步骤4、RPA机器人获取目标节点的指定属性,并按照节点组C和节点组F形成结构化数据的二维数组;

[0050] 步骤5、如果需要翻页,RPA机器人自动点击翻页按钮翻页,新一页从第一步开始获取结构化数据,并将之前获得的数据拼接在一起。

[0051] 下面结合具体实施例进行说明:

[0052] 以公司为例,将员工看作元素,员工姓名和性别看作元素属性,公司的组织架构对应元素的层级关系。

[0053] 实施例1

[0054] A男为研发部门开发组组长,B男为研发部门测试组组长;

[0055] 分析A男和B男的共同点,生成三个选择器,三个选择器分别对应研发部门、研发部门下的组长、每个组长管理的男性组员;

[0056] 选择器1-选择研发部门;

[0057] 选择器2-选择所有组长;

[0058] 选择器3-选择所有男性组员;

[0059] RPA机器人通过三个选择器提取结构化数据,可以得到一张表,表的行和列分别对应研发部门组长和对应研发部门组长管理的男性组员。

[0060] 实施例2

[0061] 男C为研发部门开发组组长,女D为研发部门测试组组长;分析男C与女D的共同点,生成三个选择器,三个选择器分别对应研发部门、研发部门下的组长、每个组长管理下的组员;

[0062] 选择器1-选择研发部门;

[0063] 选择器2-选择所有组长;

[0064] 选择器3-选择到组长自身,选择器3为空;

[0065] RPA机器人通过三个选择器提取结构化数据,如可以得到一张单列表,包含研发部门所有的组长。

[0066] 实施例3

[0067] 男E为研发部门开发组组长,男F为研发部门开发组员;分析男E和男F的共同点,发现男E和男F相对于公共节点研发部门的位置不同,处于不同层级,生成选择器失败,需要重新选取。

[0068] 本实施例只是对本发明的进一步解释,并不是对本发明的限制,本领域技术人员在阅读完本说明书后可以根据需要对本实施例做出没有创造性的修改,但是只要在本发明的权利要求范围内都受到专利法的保护。

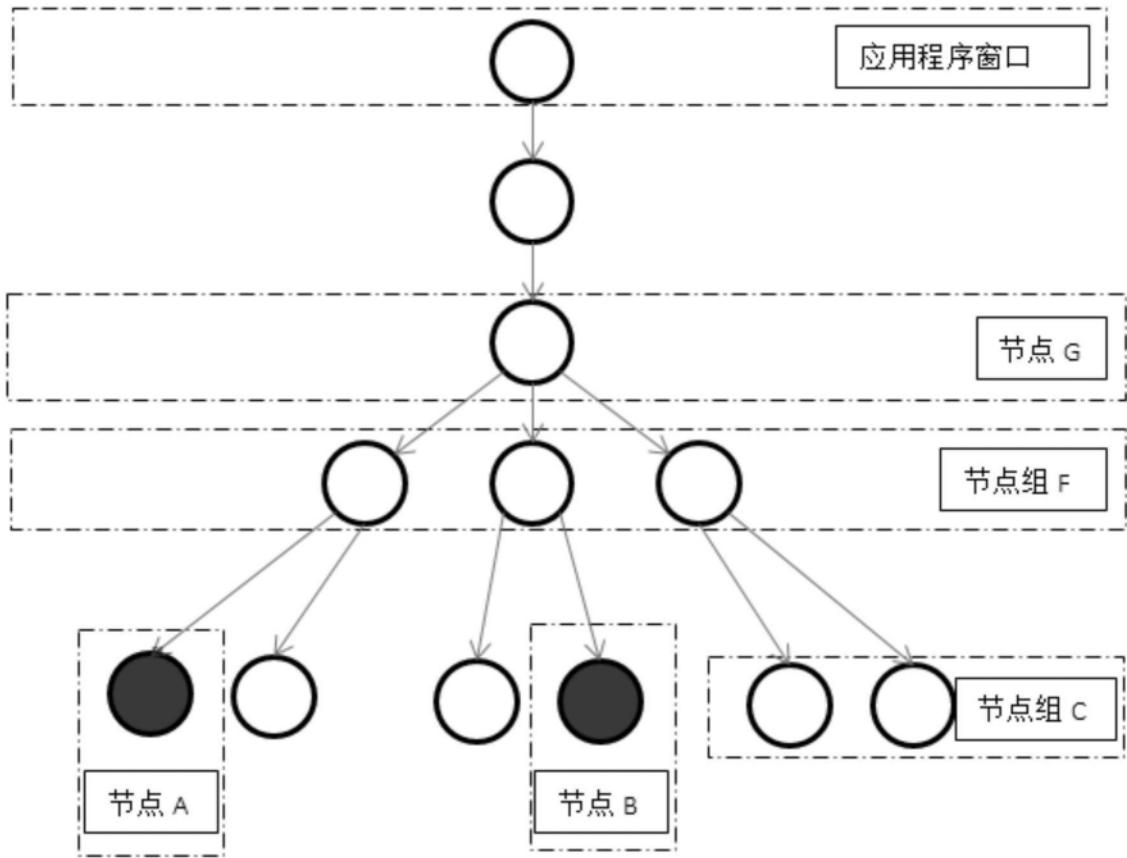


图1

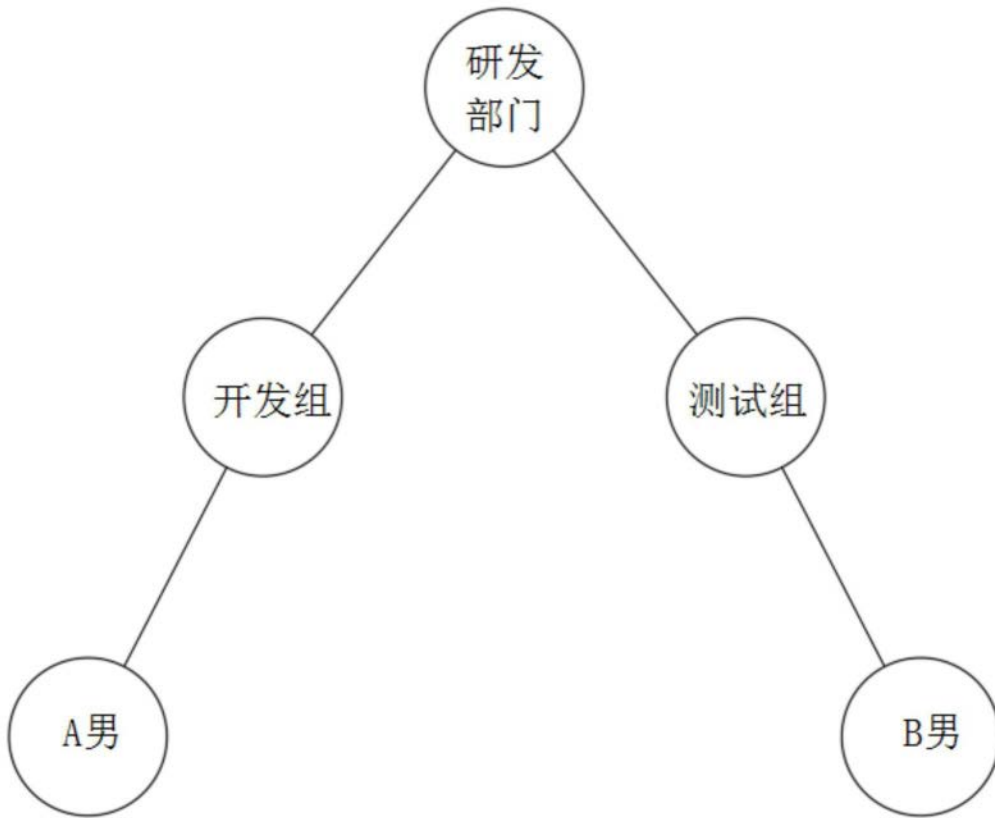


图2

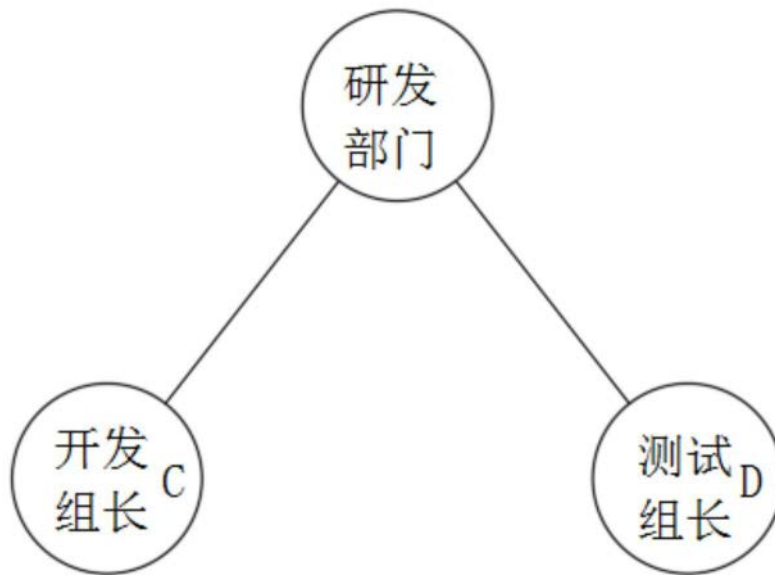


图3



图4