



网站首页

活动介绍

最新动态

组织机构

往届赛题

活动专题

就业通道

English

第七届赛题

网店工商信息图片文字提取

2018-02-08 15:32:30 来源: 作者: 【大 中 小】 浏览:23862次


第七届大赛--A组赛题

本赛题于2018.5.11修改了内容，请大家留意蓝字部分！

赛题名称: 网店工商信息图片文字提取											
组类：A 本科及以上											
赛题简介：介绍整个赛题的思路和整体要求	根据国家工商总局《网络交易管理办法》要求对网店营业执照信息进行公示，天猫网店经营者营业执照信息会在天猫店铺上以图片形式进行公示，但图片信息不能进行结构化处理，需要提取出图片中的企业注册号、企业名称形成结构化文档。										
赛题业务场景：描述赛题相关的真实企业业务背景。从真实场景中，适当简化或者提炼出适合比赛的赛题场景	如果需要统计某家企业在天猫平台的网店经营情况，就需要先把天猫平台的企业信息采集下来进行结构化处理。参赛者可以用附件提供的天猫平台上采集的50家企业图片作为测试数据，提取出文字信息后汇总进Excel作为交付文件。										
功能性需求	<div>1、程序能够自动读取企业工商信息图片所在的文件夹路径。</div> <div>2、从图片文件夹路径中顺序取出图片进行识别。</div> <div>3、因为天猫平台公示的图片内容没有固定格式，所以需要程序能匹配不同格式的图片内容提取信息。</div> <div>4、能够提取出图片中的企业注册号、企业名称数据项，并保存进Excel中，因为企业注册号、企业名称数据项要进行分析处理，所以需要保证提取信息的准确性，识别准确率需要保证在95%以上。</div> <div>5、最终的识别结果以一份汇总的Excel交付，格式如下：</div> <table><tr><th>A</th><th>B</th></tr><tr><td>企业名称</td><td>企业注册号</td></tr><tr><td>福娃集团有限公司</td><td>914210236162401254</td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr></table>	A	B	企业名称	企业注册号	福娃集团有限公司	914210236162401254				
A	B										
企业名称	企业注册号										
福娃集团有限公司	914210236162401254										
非功能性需求	识别速度保持在60秒识别50张图片										
其他限制条件：开发环境、实验平台、开发语言、数据库、编译器等限制条件（请尽量明确）	<div>开发环境操作系统：Windows 和 Linux均可</div> <div>开发语言：C++、JAVA、Python</div>										

测试数据或平台：提供给参赛者的测试环境和测试数据。 (可提供电子档)	<div>网页原始图片信息：<div>企业注册号：914210236162401254 企业名称：福娃集团有限公司 类 型：有限责任公司(自然人投资或控股) 住 所：监利县新沟经济开发区内 法定代表人：谢松柏 成立时间：1997-12-03 注册资本：5000万人民币元 营业期限：2014-03-11至 经营范围：粮食收购、仓储服务；大米、小麦粉、食用植物油、农副产品产品、膨化米果、面包、糕点、糖果、饼干、饮料、营养品、方便米制品、罐头、粥类、软饮料、乳制品、维生素糖果、袋装茶等预包装食品、散装食品批发兼零售（有效期至2018年4月14日）；玩具的加工、批发、零售；服装鞋帽、皮革制品、办公文具销售；经营本企业和本企业成员企业自产产品及相关技术的进出口业务（出口商品目录见附件，国家规定公司经营或禁止出口的商品除外）；经营本企业和本企业成员企业生产、科研所需的原辅材料、机械设备、仪器仪表、零配件及相关技术的进口业务（国家规定公司经营或禁止进口的商品除外）；经营本企业的进料加工和“三来一补”业务；普通货运（有效期至2018年7月31日）；物流服务；稻虾共育、水产养殖销售及技术服务。</div><div>登记机关：监利县工商行政管理局</div></div> <div>识别后数据信息：<table><thead><tr><th>A</th><th>B</th></tr></thead><tbody><tr><td>企业名称</td><td>企业注册号</td></tr><tr><td>福娃集团有限公司</td><td>914210236162401254</td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr></tbody></table></div>	A	B	企业名称	企业注册号	福娃集团有限公司	914210236162401254				
A	B										
企业名称	企业注册号										
福娃集团有限公司	914210236162401254										
开发所需设备及设备指标需求说明	暂无										
其他要求	<div>1.参赛人员与企业共享知识产权和成果代码</div> <div>2.不能使用百度等在线api接口服务</div> <div>3.各种语言的开源工具包可以使用，最好能有自己对工具包代码改造的东西</div> <div>4.需要识别的图片格式比较混乱，程序要有容错机制，很难识别的程序自动放弃</div>										

测试数据：

A5测试数据.rar

word版赛题下载

网店工商信息图片文字提取.doc

答疑老师联系方式：

李永伟 liywei@inspur.com

关键字：

责任编辑：cnsoft

【大 中 小】【打印】【繁体】【投稿】【收藏】【推荐】【举报】【评论】【关闭】【返回顶部】

分享到：

上一篇：智能问答系统设计与开发

下一篇：基于企业进销项发票数据的异常企..

主办单位：工业和信息化部
教育部
江苏省人民政府

承办单位：江苏省经济和信息化委员会
江苏省教育厅
中国电子信息产业发展研究院
教育部高等学校计算机类专业教学指导委员会

执行单位：中国信息化周报
江苏产业技术国际研究院
江苏软件产业人才发展基金会

客服电话：010-88559646 邮编：100048

联系地址：北京市海淀区紫竹院路66号赛迪大厦18层

网站备案/许可证号：京ICP备05039896号-10  京公网安备 11010802020860号