

河北工业大学

人工智能与数据科学学院

校内实习报告（三）

报告题目：科技信息智能采编系统-系统实现

学 院：人工智能与数据科学学院

系（专业）：软件工程

班 级：软件 161 班

学 号：164552 164558 164567

学生姓名：董宇航 石凯峰 梁山

实习单位：河北工业大学

2019 年 12 月 5

一、 实现工具

表一 实现工具

编程工具	IDEA、Pycharm、
数据库	MySQL
Python 库	jieba、snowNLP、kashgari
版本控制工具	Git
编程语言	Java、Python、JavaScript 等

二、 系统实现过程

科技信息智能采编系统的开发实现由小组全部三位成员共同参与完成，整个开发过程见图 1，整个过程可分为：

- 1、系统调研及可行性分析。具体包括系统开发背景、意义的调查以及系统技术可行性、经济可行性、运行可行性等一系列分析。
- 2、系统的需求分析。小组成员根据课题要求进行划分用例，分析系统所需要实现的功能，包括用户登录及注册、用户找回密码、用户修改头像或背景图、主题管理功能、任务管理功能、信息管理功能、爬虫策略管理、入口网站管理等一系列功能。
- 3、系统功能的代码实现。小组成员按模块分工，分别实现系统的不同模块功能。
- 4、系统测试。小组成员利用软件工程所学到的知识，用黑盒测试、白盒测试等方法对系统功能的实现情况以及系统对于错误输入的反应情况进行测试。



图 1 系统开发流程图

三、 系统架构

(1) 系统架构图

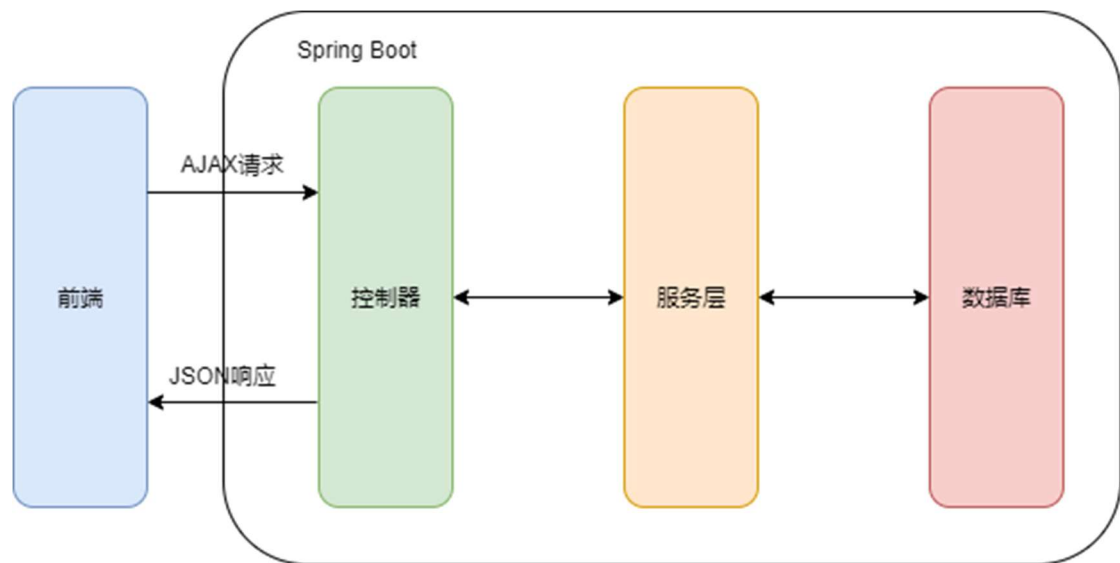


图 2 系统架构图

(2) 详细描述

该智能信息采编系统后台采用 Spring Boot 技术，为前端提供接口，实现前后端分离。使用 MySQL 数据库进行信息存储。使用 JPA，访问数据库。

前端基于 AdminLTE 框架，Vue.JS 实现数据与视图的绑定。前端路由，实现访问控制。

爬虫服务使用 PhantomJs 对动态页面进行爬取。使用 JSoup 对静态页面进行爬取。使用 JSoup、正则表达式对页面元素进行筛选。使用 JPA 对爬取数据进行存储。

系统使用 Python 的 SnowNLP 实现摘要生成。系统使用 Python 的 jiaba 实现关键词提取。系统使用 Python 的 kashgari 实现文章的自动分类。

四、 系统具体实现

（1）注册功能

具体实现如下：



The screenshot shows a mobile application interface for user registration. At the top, there is a dark header with a gear icon and the word "Favorites". Below the header, the text "快速注册新用户" (Quickly register new user) is displayed. The form consists of three input fields: "登录邮箱" (Login email) with the value "1187603543@qq.com", "登录名称" (Login name) with the value "李白", and "密码" (Password) with masked characters. Each field has a corresponding icon (envelope, person, and lock respectively). Below the input fields is a green button labeled "创建账户" (Create account). Underneath the button, the text "已经有账户?" (Already have an account?) is shown, followed by a white button labeled "登录" (Login). At the bottom of the screen, the copyright information "© 0000 - 2019" and the email "youWillGet42@163.com" are displayed.

图片 1 注册功能截图

功能描述：用户输入邮箱、用户名、密码等信息，点击创建账户按钮完成注册。当邮箱、用户名已被使用时，系统会做出相应提示。密码应在 6 到 20 位之间。密码格式错误时，系统将做出相应提示。

（2）系统异常处理提示功能

异常处理。例如，当邮箱已被注册时，系统作出相应的提示。截图如下。

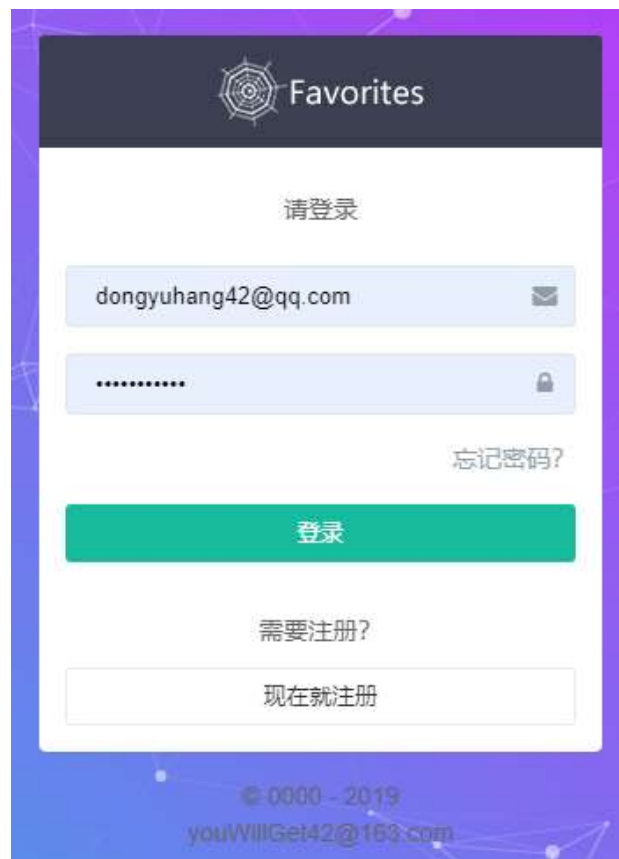


图片 2 邮箱已注册截图

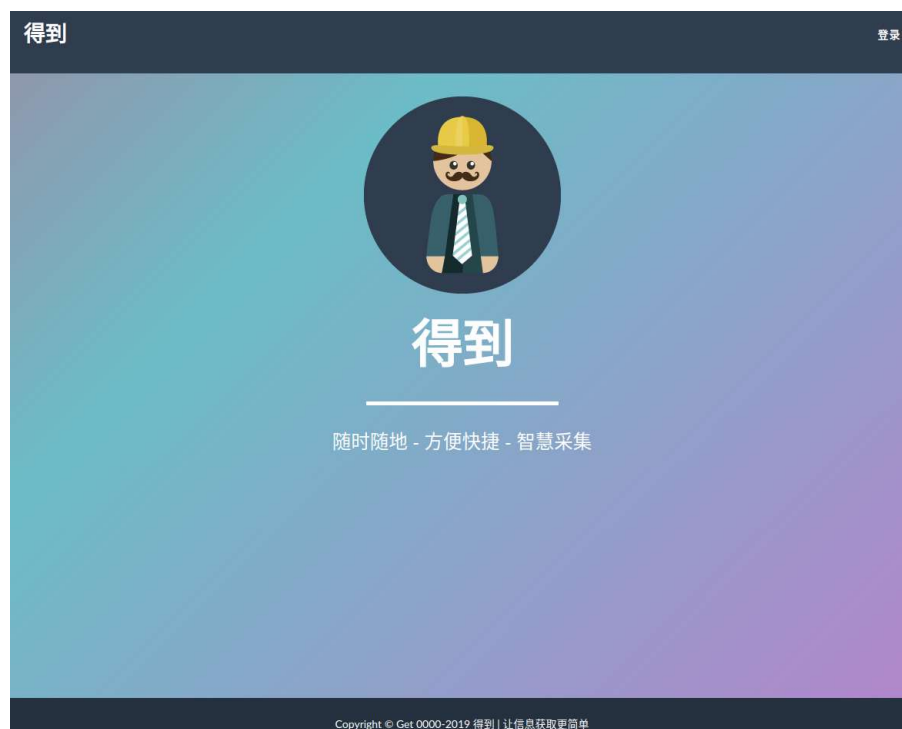
功能描述：用户在使用该系统时，可能进行误操作或者输入错误，系统会对异常情况进行检查，并在明显位置提示用户错误原因或者处理方式。从而，提升用户使用该系统的用户体验。

（3）用户登录功能

用户打开首页，点击登录按钮，系统跳转到登录页面。具体实现如下：



图片 3 登录功能截图



图片 4 系统首页

功能描述：用户打开登录界面，输入注册邮箱、密码，点击登录按钮，完成登录。当登录失败时，系统会作出相应提示，引导用户修改正确。若登录成功，系统会跳转到系统首页，用户点击用户名，完成登录。

（4）找回密码功能

若用户忘记密码，可以在登录界面，点击忘记密码按钮，系统会跳转到找回密码页面。具体实现如下。



图片 5 重置密码页面



图片 6 发送重置邮件成功截图



图片 7 用户收到重置密码邮件截图

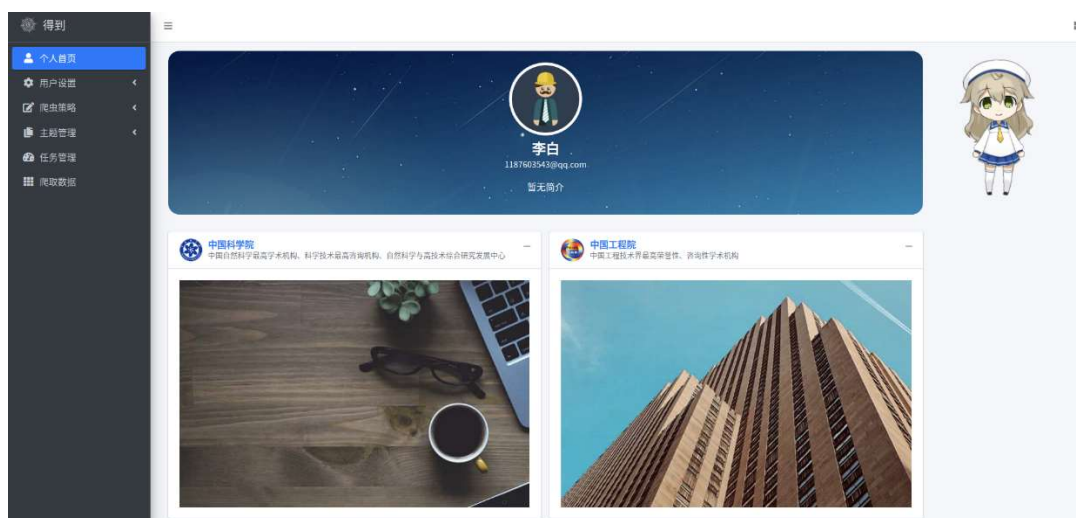


图片 8 设置新密码截图

功能描述：用户输入邮箱、密码完成登录。若用户忘记密码，可在登录界面点击忘记密码按钮，进入找回密码功能。用户输入注册邮箱，点击发送重置按钮，系统会发送邮件到用户的注册邮箱。用户登录自己的注册邮箱，找到密码重置邮件，点击重置密码链接。之后，用户进入设置新密码界面，用户输入新密码，完成密码重置。

重置链接 30 分钟内有效。系统会通过对链接进行参数检查，判断是否重置链接是否过期、失效。

(5) 个人主页功能



功能描述：左侧为系统功能栏，用户通过点击按钮，可进入相应功能界面。中间上半部分显示用户头像、用户背景图、用户邮箱、用户简介等内容。中间下半部分为系统当前支持爬取的网站列表，头像为相应网站的 LOGO，点击网站名称，可打开相应网页。右侧为 Live 2D 动态人物形象。

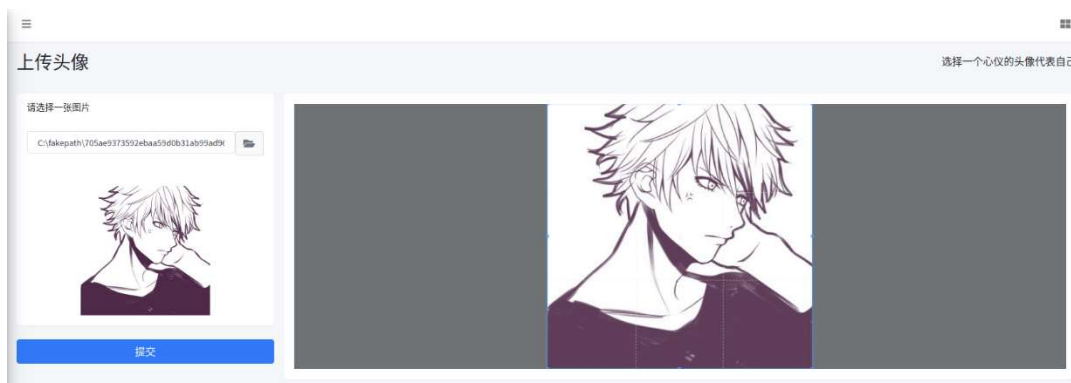
(6) 修改密码功能

The screenshot shows a '修改密码' (Change Password) form. At the top is the title '修改密码'. Below it is the instruction '请输入原密码和新密码'. The form contains three input fields: '原密码' (Original Password), '新密码' (New Password), and '确认新密码' (Confirm New Password), each with a password icon on the right. At the bottom is a blue button labeled '修改密码'.

图片 9 修改密码截图

功能描述：用户选择修改密码连接，进入修改密码页面。用户输入原密码、新密码、确认新密码，点击修改密码按钮，完成密码修改。

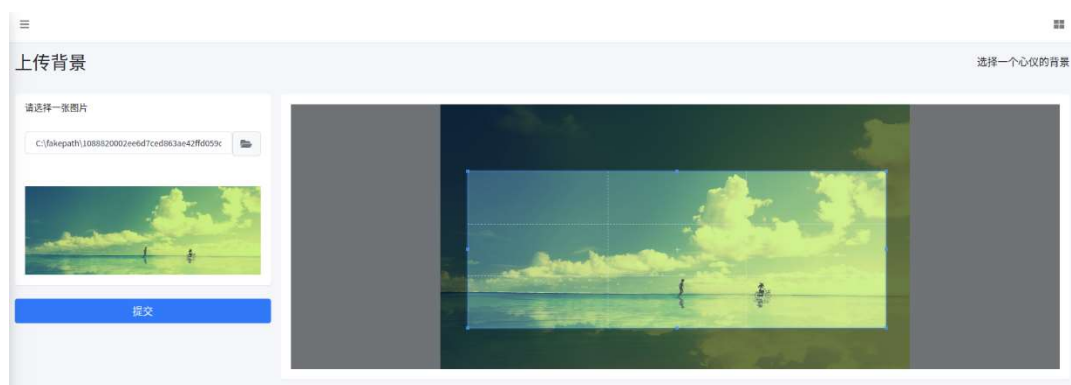
(7) 上传头像功能



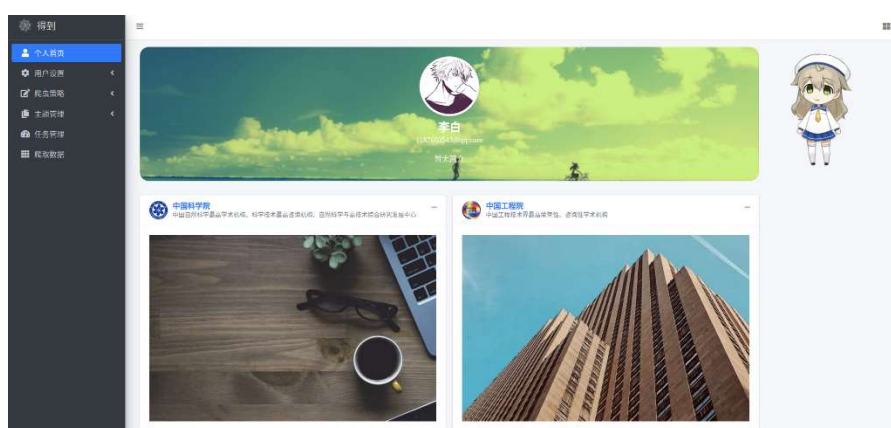
图片 10 上传头像功能

功能描述：用户可以选择上传本地图片，进行区域选择，点击提交，完成用户头像的修改。

(8) 上传背景图功能



图片 11 修改背景图



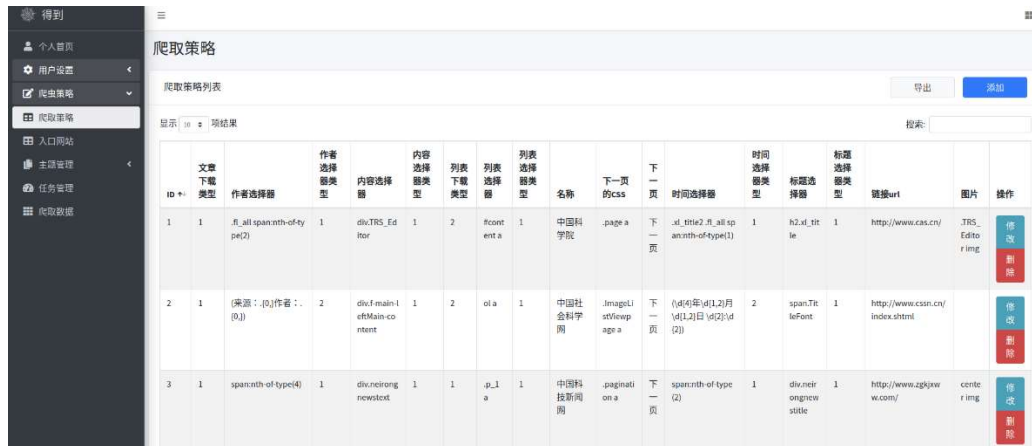
图片 12 修改后的个人首页

功能描述：用户可以选择上传本地图片，进行区域选择，点击提交，完成用户背景图的修改。

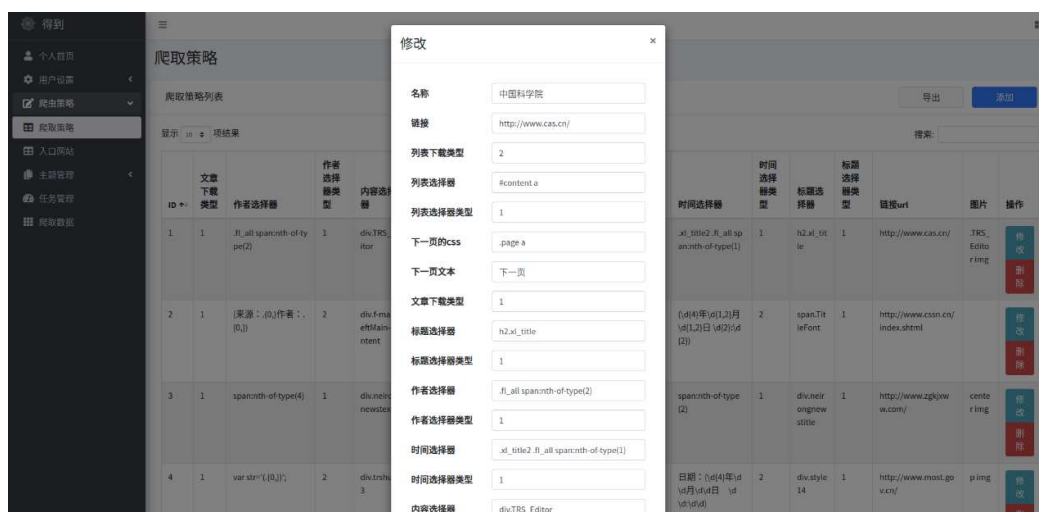
(9) 爬取策略管理



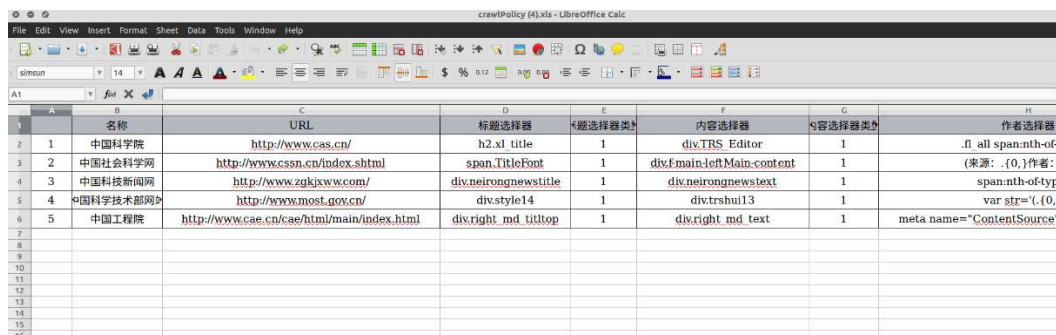
图片 13 系统内置爬取网站列表



图片 14 系统内置的爬取策略



图片 15 爬取策略修改



图片 16 爬取策略导出

功能描述：系统内置了五个可以爬取的网站策略模板。管理员可以对爬虫策略进行管理，点击导出按钮，可以完成爬虫策略的导出。

(10) 入口网站管理

入口网站				
入口网站列表				
显示 50 项结果			搜索:	
ID	名称	网站ID	链接url	操作
1	政策解读	1	http://www.cas.ac.cn/zcjd/	修改 删除
2	科研进展	1	http://www.cas.ac.cn/syky/	修改 删除
3	传媒扫描	1	http://www.cas.ac.cn/cm/	修改 删除
4	一线动态	1	http://www.cas.ac.cn/yx/	修改 删除
5	工作进展	1	http://www.cas.cn/sygz/	修改 删除
6	访谈·视点	1	http://www.cas.cn/zjs/	修改 删除

图片 17 入口网站管理

修改

名称

政策解读

网站ID

1

链接url

http://www.cas.ac.cn/zcjd/

取消

确认修改

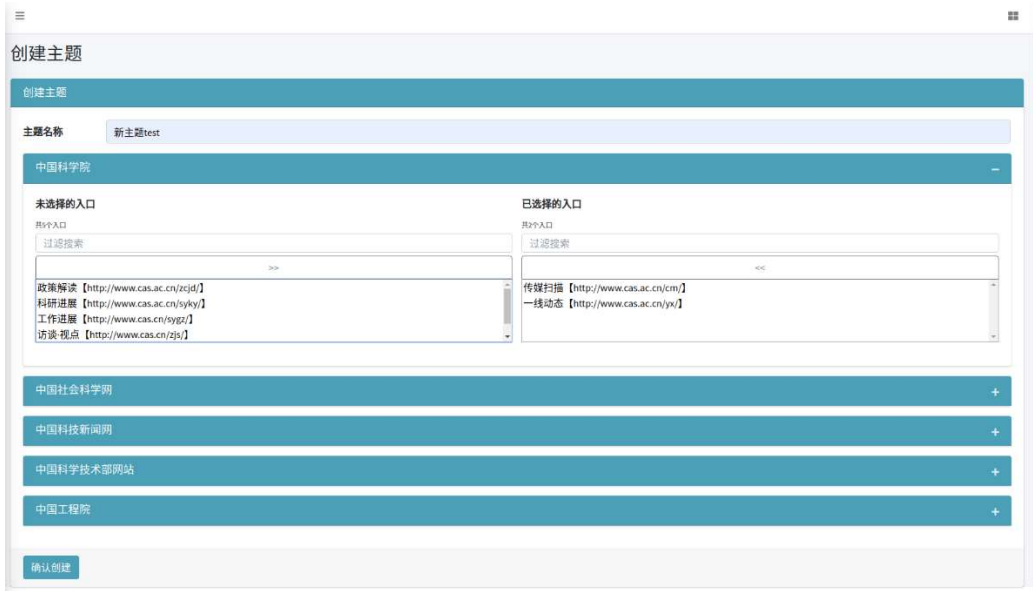
图片 18 入口网站的修改

	A	B	C	D	
1	编号	名称	URL	策略编号	
2	1	政策解读	http://www.cas.ac.cn/zcjd/	1	
3	2	科研进展	http://www.cas.ac.cn/syky/	1	
4	3	传媒扫描	http://www.cas.ac.cn/cm/	1	
5	4	一线动态	http://www.cas.ac.cn/yx/	1	
6	5	工作进展	http://www.cas.cn/sygz/	1	
7	6	访谈·视点	http://www.cas.cn/zjs/	1	
8	7	院内要闻	http://www.cas.ac.cn/yw/	1	
9	8	今日要闻	http://orig.cssn.cn/sf/	2	
10	9	军事技术	http://mil.cssn.cn/jsx/jszk_jsx/jsjs_jsx/	2	
11	10	国际经济	http://intl.cssn.cn/qj/gj_gwshkx/qj_jj/	2	
12	11	学术视角	http://news.cssn.cn/zx/xshshj/	2	
13	12	社会万象	http://news.cssn.cn/zx/shwx/	2	

图片 19 入口网站的导出

功能描述：管理员可以对入口网站进行管理，点击导出按钮，可以进行入口网站的导出。

(11) 主题管理功能



图片 20 主题管理的创建



图片 21 主题的修改

主题列表			
主题列表			
显示 10 项结果		搜索:	
ID	主题名称	所选入口	操作
1	新主题test	传媒扫描 【http://www.cas.ac.cn/cm/】 一线动态 【http://www.cas.ac.cn/yx/】 工程院图片新闻 【http://www.cae.cn/cae/html/main/column_3_1.html】	修改 删除
ID	主题名称	所选入口	操作
显示第 1 至 1 项结果，共 1 项			
		上一页	下一页

图片 22 查看列表

功能描述：用户可以创建主题，查看主题列表，对于主题进行管理。

(12) 任务管理功能

任务列表

任务列表

一个任务也没有，快去创建一个吧

创建任务

任务名称

新任务test

选择主题

未选择的主题

列表为空

过滤搜索

>>>

已选择的主题

共1个主题

过滤搜索

<<<

新主题test

确认创建

图片 23 任务列表界面

任务列表

任务列表

☐ 新任务test 所选主题:新主题test ① 未开始

创建任务

任务名称

新任务test

选择主题

未选择的主题

共2个主题

过滤搜索

>>>

新主题test

已选择的主题

列表为空

过滤搜索

<<<

确认创建

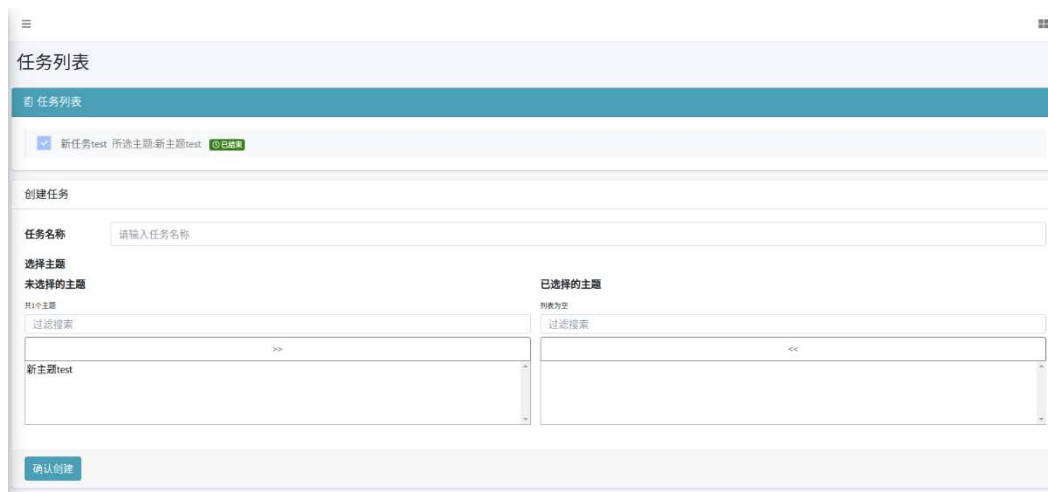
图片 24 任务列表查看



图片 25 启动任务



图片 26 查看任务启动情况



图片 27 查看任务状态

功能描述：用户可以创建、修改、删除爬虫任务，查看已创建的任务列表。启动爬虫任务。

(13) 爬取数据列表

爬取数据列表									
数据列表						导出数据	生成摘要	热词提取	
任务过滤 <input checked="" type="checkbox"/> 中国科学院爬虫任务1 <input type="checkbox"/> 中国科学技术部爬虫任务1									
显示 10 项结果						搜索: <input type="text"/>			
<input type="checkbox"/>	所属任务	爬取时间	标题	关键字	发布时间	分类	操作		
<input type="checkbox"/>	中国科学院爬虫任务1	2019-12-31	【中国科学院】中科院与河北省签署合作协议	创新,建设,新区,发展,研究院	2019-12-30	时政	查看		
<input type="checkbox"/>	中国科学院爬虫任务1	2019-12-31	中央第十五巡视组巡视中国科学院党组工作动员会召开	巡视,工作,中央,政治,巡视组	2019-09-12	时政	查看		
<input type="checkbox"/>	中国科学院爬虫任务1	2019-12-31	中科院科创中心暨北京分院召开2019年终院务会议	分院,工作,中心,北京,会议	2019-12-30	时政	查看		
<input type="checkbox"/>	中国科学院爬虫任务1	2019-12-31	【科技日报】2019年国内国际十大科技新闻揭晓	基因,全球,人类,编辑,科技	2019-12-30	科技	查看		
<input type="checkbox"/>	中国科学院爬虫任务1	2019-12-31	新时代南海科技发展高峰论坛暨南沙综合科考学术研讨会在广州召开	海洋,南海,科技,领域,建设	2019-12-27	时政	查看		
<input type="checkbox"/>	中国科学院爬虫任务1	2019-12-31	中科院在国家高端智库综合评估中成绩优异	智库,咨询,战略,国家,科技	2019-12-30	科技	查看		

图片 28 爬取数据列表

2019年(第六届)新能源材料高峰论坛在北京举办

三局科技合作处

原文:http://www.cae.cn/cae/html/main/col1/2019-12/10/20191210094546279291170_1.html

2019年12月5日,由中国工程院指导、中国有色金属工业协会主办的2019年(第六届)新能源材料高峰论坛在北京举行。钟志华副院长出席并致辞,干勇院士作专题报告,何季麟院士出席。

钟志华指出,大力发展新能源材料,是科技界和企业界必须携手推进的紧迫任务。我国新材料产业发展取得了长足进步,产业规模持续扩大,整体实力大幅提升,但与国际先进相比,我国新能源与新材料产业的发展还有很大的提升空间,新材料产业的突破与升级已经迫在眉睫。国家新材料产业相关政策的出台,将有力推动包括新能源材料在内的先进基础材料、关键战略材料和前沿新材料的研发和应用,从而带动新能源产业和新材料产业的快速发展。

干勇作了题为《制造业强国新材料发展战略》的专题报告,分析了材料技术发展趋势,重点阐述了新材料产业发展战略,并提出制造业技术创新体系建设。

多位专家围绕新能源材料的若干问题作了专题报告,把脉行业企业发展,聚焦新能源、新材料、新能源汽车的应用、投融资、环保、竞争技术路线等热点话题,深入研究产业发展方向,共商新能源和新材料产业发展大计,力争打造为引领全球新能源新材料产业发展的风向标。(赵文成 文/图)

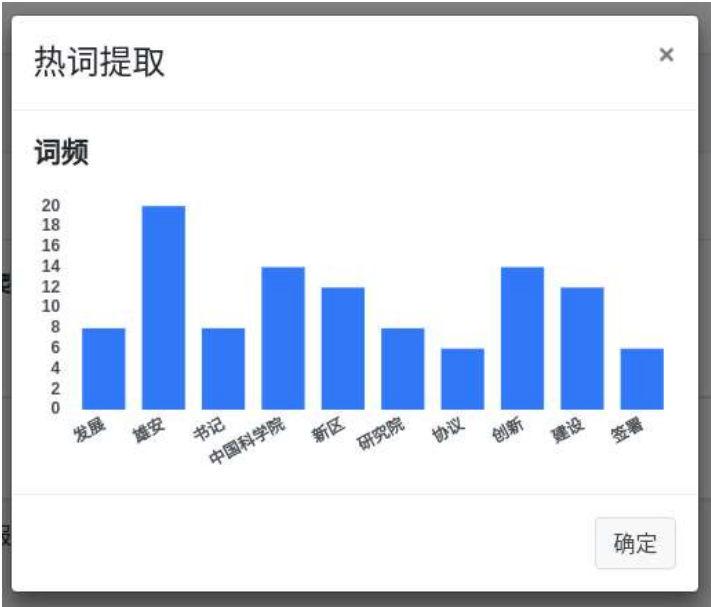


确定

图片 29 查看详细文章

功能描述：用户可以根据任务爬取数据进行管理。击查看按钮看文章详细内容。输入关键词，对于内容进行检索。

(14) 热词提取功能



图片 30 热词提取截图

功能描述：用户选择爬取文章数据，点击热词提取按钮，系统显示选中文章的热词统计图。

(15) 数据导出功能

A	B	C	D	E	F
编号	标题	作者	发布时间	分类	关键词
1	【中国科学报】中科院与河北省签署合作协议	中国科学报 高雅丽	2019-12-30	8	[创新, 建设, 新区, 发展, 研究院]
2	中央第十五巡视组巡视中国科学院党组工作动员会召开	中央纪委国家监委网站	2019-09-12	8	[巡视, 工作, 中央, 政治, 巡视组]
3	中科院科创中心暨北京分院召开2019年终院务会议	科技创新发展中心 (北京分院)	2019-12-30	8	[分院, 工作, 中心, 北京, 会议]
4	【科技日报】2019年国内国际十大科技新闻揭晓	科技日报 高博 陆成宽	2019-12-30	6	[基因, 全球, 人类, 编辑, 科技]
5	新时代南海科技发展高端论坛暨南沙综合科考学术研讨会在广州召开	重大科技任务局	2019-12-27	8	[海洋, 南海, 科技, 领域, 建设]
6	中科院在国家高端智库综合评估中成绩优异	科技战略咨询研究院	2019-12-30	6	[智库, 咨询, 战略, 国家, 科技]
7	【中国科学报】怀柔科学城26个科学设施平台全部开建	中国科学报 郑金武	2019-12-30	6	[平台, 科学, 创新, 研究, 设施]
8	中科院党组2019年冬季扩大会议召开	科学传播局	2019-12-22	8	[工作, 发展, 会议, 建设, 精神]
9	中国科技云2.0发布会在京举行	计算机网络信息中心	2019-12-27	6	[科技, 中国, 服务, 科学, 数据]
10	【中国科学报】中科院科技创新年度巡展2019在京开幕	中国科学报 程唯唯	2019-12-30	6	[科学, 科技, 巡展, 成果, 创新]
11	中科院与生态环境部举行深化生态保护监管领域合作工作座谈会	科技促进发展局	2019-12-20	8	[生态, 合作, 保护, 监管, 科技]
12	中科院《科研组织知识产权管理规范》贯标工作总结交流会在京召开	科技促进发展局	2019-12-26	6	[知识产权, 贯标, 工作, 研究所, 认证]
13	中科院与天津市举行科技合作座谈	科技促进发展局	2019-12-20	8	[科技, 合作, 天津市, 中心, 创新]
14	中科院召开2019年部门决算及财务报告编制视频会	条件保障与财务局	2019-12-25	8	[工作, 财务, 单位, 部门, 财务报告]
15	【科技日报】世界科学前沿的方向标	科技日报 张梦然	2019-12-30	6	[细胞, 量子, 超导, 黑洞, 材料]
16	“中美物理合作：机遇与挑战”圆桌论坛举行	国际合作局	2019-12-20	6	[物理, 科学, 挑战, 美国, 论坛]
17	中科院党组传达学习中央经济工作会议精神	办公厅	2019-12-13	8	[工作, 经济, 会议, 目标, 精神]
18	【科技日报】中国创新路上的里程碑	科技日报 陆成宽 高博	2019-12-30	6	[中国, 黑洞, 病毒, 密钥, 量子]

图片 31 数据导出功能

功能描述：用户选择要导出的数据列表，点击导出数据按钮，完成数据的导出。

(16) 摘要生成功能

<p>标题:【中国科学报】中科院与河北省签署合作协议</p> <p>作者:中国科学报 高雅丽</p> <p>URL:http://www.cas.ac.cn/cn/201912/20191230_4729650.shtml</p> <p>摘要:雄安新区党工委书记、管委会主任陈刚代表双方签署了《中国科学院河北省人民政府共建中国科学院雄安创新研究院框架协议》。雄安新区党工委书记、管委会主任陈刚代表双方签署了《中国科学院河北省人民政府共建中国科学院雄安创新研究院框架协议》。中国科学院院长、党组书记白春礼。中国科学院院长、党组书记白春礼。建设中科院支持和参与雄安新区规划建设发展的综合性集成创新平台</p> <p>标题:在中科院与生态环境部举行深化生态保护监管领域合作工作座谈会</p> <p>作者:科技促进发展局</p> <p>URL:http://www.cas.ac.cn/yw/201912/20191220_4728444.shtml</p> <p>摘要:中国科学院与生态环境部在北京举行深化生态保护监管领域合作工作座谈会并签署深化生态保护监管领域合作备忘录。中国科学院与生态环境部在北京举行深化生态保护监管领域合作工作座谈会并签署深化生态保护监管领域合作备忘录。李树深与黄润秋代表院部双方签署了《深化生态保护监管领域合作备忘录》。李树深与黄润秋代表院部双方签署了《深化生态保护监管领域合作备忘录》。座谈现场签署《深化生态保护监管领域合作备忘录》签署《国家生态状况监测网络建设合作协议》 12月19日上午</p> <p>标题:在中科院与天津市举行科技合作座谈</p> <p>作者:科技促进发展局</p> <p>URL:http://www.cas.ac.cn/yw/201912/20191220_4728413.shtml</p> <p>摘要:他希望进一步加强与中科院在技术创新中心建设、人才引进、成果转化、体制机制改革等方面的合作。他希望进一步加强与中科院在技术创新中心建设、人才引进、成果转化、体制机制改革等方面的合作。连茂君介绍了中科院与天津市科技合作的进展情况。连茂君介绍了中科院与天津市科技合作的进展情况。共同建设国家合成生物技术创新中心</p> <p>标题:在中科院与甘肃省举行科技合作座谈会</p> <p>作者:科技促进发展局</p> <p>URL:http://www.cas.ac.cn/yw/201912/20191213_4727530.shtml</p> <p>摘要:座谈现场 12月12日下午。张世珍简要汇报了改善优化甘肃科技生态环境、加强院省合作等相关情况以及下一阶段的重点合作考虑。张世珍简要汇报了改善优化甘肃科技生态环境、加强院省合作等相关情况以及下一阶段的重点合作考虑。进一步深化院省科技合作机制。进一步深化院省科技合作机制</p> <p>标题:在中科院与大连市举行科技合作座谈会</p> <p>作者:重大科技任务局</p>	<p>URL:http://www.cas.ac.cn/yw/201912/20191213_4727432.shtml</p> <p>摘要:会议听取了大连市副市长张志宏和中科院洁净能源创新研究院院长刘中民关于洁净能源创新院、大连先进光源等工作的汇报。会议听取了大连市副市长张志宏和中科院洁净能源创新研究院院长刘中民关于洁净能源创新院、大连先进光源等工作的汇报。中科院重大科技任务局、洁净能源创新院和大连市有关部门相关人员参加座谈。中科院重大科技任务局、洁净能源创新院和大连市有关部门相关人员参加座谈。大连市科技创新工作一直以来都得到了中科院大力支持</p> <p>总摘要:中国科学院与生态环境部在北京举行深化生态保护监管领域合作工作座谈会并签署深化生态保护监管领域合作备忘录。中国科学院与生态环境部在北京举行深化生态保护监管领域合作工作座谈会并签署深化生态保护监管领域合作备忘录。座谈现场签署《深化生态保护监管领域合作备忘录》签署《国家生态状况监测网络建设合作协议》 12月19日。座谈现场签署《深化生态保护监管领域合作备忘录》签署《国家生态状况监测网络建设合作协议》 12月19日上午。他希望进一步加强与中科院在技术创新中心建设、人才引进、成果转化、体制机制改革等方面的合作</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

图片 32 摘要导出 word 文件例子

功能描述：用户点击选择文章数据，点击摘要生成按钮，完成摘要的生成、导出。