

PYSE练习题

2016年4月16日

乙醇



天气抓取

- 假如你是某公司的hr，请从中国天气网(<http://www.weather.com.cn>)或者百度天气获取当地明天的天气情况，如果明天下雨，请发送邮件通知全体同事带伞，如果明天气温低于10度，请邮件提醒同事注意保暖，如果气温高于30度则提醒同事注意高温。假设存在发送邮件的方法
`self.send_email(email_content)`

微博热门话题

- 抓取微博24小时热门话题的前15个
(http://d.weibo.com/100803?cfs=&Pl_Discover_Pt6Rank_5_filter=hothtlist_type%3D1#_0), 注意需要翻页, 抓取的内容
请保存至txt文件中, 需要抓取阅读数

热门微博

- 抓取24小时热门微博前十条
(http://d.weibo.com/102803?from=unlogin_home&mod=pindao&type=hotweibo#_rnd1456632403002)保存成txt格式，需要抓取发帖人，正文的html，转发量，阅读量和赞的数量

知乎热点

- 抓取知乎今日最热及本月最热的问题和首个回答，保存至html文件，该html文件的文件名应该是20160228_zhihu_today_hot.html，也就是日期+zhihu_today_hot.html

登录什么值得买

- 注册（这个不需要自动化）并登录<http://www.smzdm.com/>，点击签到领积分按钮，获取每日积分

豆瓣电影

- 抓取豆瓣电影
(<http://movie.douban.com/nowplaying/shenzhen/>)中的正在热映前12部电影，并按照评分排序，保存至txt文件

豆瓣读书

- 抓取豆瓣读书中的(<http://book.douban.com/>)最受关注图书，按照评分排序，并保存至txt文件中，需要抓取书籍的名称，作者，评分，体裁和一句话评论

PM2.5

- 从<http://www.pm25.com/shenzhen.html>抓取北京，深圳，上海，广州的pm2.5指数，并按照空气质量从优到差排序，保存在txt文档里



抓取infoq里的测试新闻

- 描述

- 打开infoq页面，抓取最新的一些测试文章
- 需要抓取文章的标题和内容
- 如果你有个人blog的话，可以将这些文章转载到自己的blog

- 要求

- 不要在新窗口打开文章
- 自行了解最新的测试思潮与实践

简单的开发者头条客户端

▫ 描述

- 去开发者头条抓取本日的top 10内容，也就是排在前面的10个，需要抓取标题和url
- 将这些内容保存在数据库，推荐使用sqlite
- 写个简单的客户端，要求可以展示每日的内容，点击标题后可以打开浏览器浏览该内容

▫ 要求

- 用同一种语言实现
- 使用phantomjs去抓取内容