

Level 2 实验报告

PB20111686 黄瑞轩

选题：Level 2.13【计算机】

基于 Python，使用浏览器引擎 Selenium 爬取科大首页中的科大要闻第一条新闻，要求爬取结果至少有新闻标题、时间、以及新闻内容。

思路：Selenium 使用 Chrome 来做 (chromedriver)，通过 Selenium 的 `find_element_by_xpath` 方法找到对应的内容。

关键代码：

```
browser = webdriver.Chrome()
news = []

browser.get("http://ustc.edu.cn/")
browser.find_element_by_xpath("/html/body/main/div[1]/section[1]/ul/li[1]/div[2]/h4[1]/a").click()

news.append(browser.find_element_by_xpath("/html/body/main/div/div/div[1]/form/div[1]").text)
news.append(browser.find_element_by_xpath("/html/body/main/div/div/div[1]/form/div[2]").text)
news.append(browser.find_element_by_xpath('//*[@id="vsb_content_2"]/div').text)
```

爬取结果：

['2022年05月21日', '中国科大举办2022科技活动周开幕式暨全国科普教育基地揭牌仪式', '5月21日上午，中国科大2022科技活动周开幕式暨全国科普教育基地揭牌仪式在东校区水上报告厅举行。党委书记舒歌群、校长包信和、党委常务书记蒋一出席仪式，党委常委、副校长罗喜胜主持仪式。相关职能部门、院系负责人、部分感兴趣师生参加仪式。\\n罗喜胜介绍了学校2022科技活动周的组织筹备情况。他表示，为加快推进科技规划政策扎实落地，广泛宣传科技创新成果，开展科学普及惠民活动，根据科技部、中宣部、中国科协及中科院总体部署，学校于5月21日举办科技活动周暨第十八届公众科学日活动。本次活动采用线上线下相结合、线上即时互动的方式，让更多观众享受这一科学盛宴。\\n日前中国科协公布《2021-2025年第一批全国科普教育基地名单》，我校国家同步辐射实验室和火灾科学国家重点实验室入选。舒歌群、包信和、国家同步辐射实验室书记李良彬、火灾科学国家重点实验室主任刘乃安共同为全国科普教育基地揭牌。\\n舒歌群在致辞中强调，值此第二个百年奋斗目标新征程扬帆起航之际，要深入学习领会习近平总书记关于科技创新系列重要指示精神，坚持“四个面向”，认真贯彻落实习近平总书记“科技创新与科学普及两翼齐飞”的重要论述精神，深刻把握全民科学素质建设的责任使命，推动在全社会形成讲科学、爱科学、学科学、用科学的良好氛围。他表示，学校将办好此次科技活动周作为深入贯彻《科技进步法》《科普法》、坚决落实党中央关于科技自立自强的战略部署、推进科技强国建设的一项重要任务抓好抓实，精心组织筹办27个科普点共105个方案和九场大型科普报告，科普点数、活动方案数均达到历史新高。舒歌群指出，要以举办此次科技活动周为契机，进一步充分发挥我校在教学、科研、人才、基地方面的综合优势，大力培育青少年尊崇科学的人生价值观，激发青少年热爱科学探索的兴趣爱好，引导青少年树立实现高水平科技自立自强的理想志向，在青春的赛道上奋力奔跑，跑出当代青年的最好成绩。\\n\\n(党委宣传部)']

存在 `news` 列表中，`news[0]` 为时间，`news[1]` 为标题，`news[2]` 为内容。

用法：

```
$ python3 main.py
```

注：selenium 需要为特定的 Chrome 版本匹配特定的 chromedriver.exe 版本，测试时请助教前往 <http://chromedriver.storage.googleapis.com/index.html> 下载适合测试环境 Chrome 版本的 chromedriver.exe 并于 main.py 放置于同一目录下。