作业4-网络爬虫

**题目：**

爬取厦大计算机系网站首页新闻 <https://cs.xmu.edu.cn/>

（1）任务1：爬取首页的新闻标题，保存在文本文件中（每个新闻标题占一行）

（2）任务2：爬取首页的新闻标题，保存在文本文件中，同时，对于每个新闻标题，进入该新闻，爬取新闻详细内容，保存到文本文件中（格式为：新闻标题1，换行，新闻内容1，换行，新闻标题2，换行，新闻内容2……）

代码粘贴在此

1. **import** requests
2. **from** bs4 **import** BeautifulSoup
4. headers = {
5. 'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/67.0.3396.99 Safari/537.36'
6. }
7. url = 'https://cs.xmu.edu.cn/'
9. **try**:
10. response = requests.get(url, headers=headers)
11. response.encoding = 'utf-8'
12. **print**(response.url)
13. **print**(response.status\_code)
15. # 解析HTML
16. soup = BeautifulSoup(response.text, 'lxml')
18. # 找div标签
19. all\_labels = soup.find\_all('div', {'class': 'panel-pane pane-views pane-xinwendongt center-first blankLinks'})
20. with open('output.txt', 'w', encoding='utf-8') as f:
21. # 任务1：
23. hrefs=[]
24. f.write('任务1：爬取首页的新闻标题，保存在文本文件中（每个新闻标题占一行）\n')
25. **print**('任务1')
26. **for** label **in** all\_labels:
27. span\_tags = label.find\_all('span', class\_='field-content')
28. **for** span **in** span\_tags:
29. a\_tags = span.find\_all('a')
30. **for** a **in** a\_tags:
31. # 输出<a>标签的class和内容
32. f.write(f'{a.text}\n')
33. **if** a.get('href'):
34. hrefs.append(f"{url}{a.get('href', [''])}")
35. **print**(a.text)
36. # 任务2：
37. f.write('任务2：爬取首页的新闻标题，保存在文本文件中，同时，对于每个新闻标题，进入该新闻，爬取新闻详细内容，保存到文本文件中（格式为：新闻标题1，换行，新闻内容1，换行，新闻标题2，换行，新闻内容2……）\n')
38. **print**('任务2')
39. **print**(hrefs)
40. **for** href **in** hrefs:
41. response = requests.get(href, headers=headers)
42. response.encoding = 'utf-8'
43. **print**(response.url)
44. **print**(response.status\_code)
45. soup = BeautifulSoup(response.text, 'lxml')
46. title = soup.find('h1',{'class': 'title gutter'})
47. f.write(f'{title.text}\n')
48. **print**(title.text)
49. infos = soup.find('div', {'class': 'Section1'})
50. details=""
51. **for** info **in** infos:
52. details+=info.text
53. **print**(details)
54. f.write(f'{details}\n')
55. **except** Exception as e:
56. **print**(e)

结果：

任务1：爬取首页的新闻标题，保存在文本文件中（每个新闻标题占一行）

计算机科学与技术系教工党支部开展全体党员实践教...

计算机系教工党支部赴南安市水头镇开展党员教学实...

计算机科学与技术系教工党支部与银川能源学院信息...

计算机科学与技术系教工党支部开展师生党员集中培...

计算机科学与技术系教工党支部开展学习《中国共产...

计算机科学与技术系教工党支部开展师生党员集中培...

学习2024年全国两会精神暨人工智能教学沙龙

任务2：爬取首页的新闻标题，保存在文本文件中，同时，对于每个新闻标题，进入该新闻，爬取新闻详细内容，保存到文本文件中（格式为：新闻标题1，换行，新闻内容1，换行，新闻标题2，换行，新闻内容2……）

计算机科学与技术系教工党支部开展全体党员实践教学认定

    9月12日，计算机科学与技术系教工党支部召开党员大会，开展支部全体党员的实践教学认定工作，活动由支部副书记严严同志主持。    会议伊始，严严同志向与会同志介绍了党员实践教学的要求与认定过程。学校要求党支部可根据自身工作实际，结合“三会一课”、“固定党日+”活动组织开展党员实践教学，实践教学是党员教育的一种重要形式，旨在让党员通过实践活动深刻理解和体验党的理论和路线方针政策，增强党性修养和实践能力。随后，党员逐一发言，介绍自己的实践教学情况，并接受支部全体党员的评价和认定。    根据校党委党员实践教学部署安排，计算机科学与技术系教工党支部积极组织、严格督促全体党员开展实践教学。党支部集中安排两次实践活动，个别党员根据兴趣和时间进行了实践教学。支部党员的实践教学包括：前往国家博物馆、哈尔滨工业大学、东北小兴安岭和大庆油田、兰州博物馆、昆明西南联大遗址、青海三江源、南安市水头镇、宁化县革命纪念馆、古田会议旧址、翔安区新圩镇桂林村、华侨博物院、厦门市博物馆、厦门边检总站高崎边检站、厦门罗扬才烈士陵园等地开展实践教学活动。    在集体教学实践中，通过对南安市水头镇和翔安区新圩镇桂林村的走访，同志们感受到了乡村振兴政策、乡村治理和环境保护对地方发展起到的推动作用，感受到改革开放四十年来，人民生活水平大幅提高，经济快速发展，科技水平上到新台阶。在特色教学实践中，通过参观国家博物馆复兴之路专题展，感受到了祖国在中国共产党领导下的艰苦斗争和伟大复兴历程；通过参观哈尔滨工业大学航天馆，党员同志们切身感受了“两弹一星”精神引领下的中国航空航天发展历程；通过参观青海三江源，学习了习近平总书记对生态环境保护的重要指示精神；通过参观昆明西南联大遗址，身临其境回顾起抗战时期师生们的艰苦奋斗岁月；通过对宁化、古田革命遗址的参观，回顾中国共产党在抗战时期开展的艰苦卓绝的斗争和伟大事迹；走访高崎边检站，增强了对深化两岸融合发展的体会和认识；    计算机科学与技术系教工党支部将继续在各级党委的组织和安排下，开展党员实践教学活动，将理论学习与实践学习有机结合，不断增强党性修养和实践能力。计算机科学与技术系教工党支部2024年9月12日

计算机系教工党支部赴南安市水头镇开展党员教学实践活动

        2024年8月12日，计算机系教工党支部赴南安市水头镇开展党员教学实践活动。参加本次活动的人员包括厦门大学信息学院党委薛志平副书记以及计算机系教工党支部曲延云、林子雨、陈龙彪、沈思淇、杨晨晖、黄华勋、吴素贞、郑炜、张丽丽老师等。第一站，支部党员走进五里桥文化公园，听取了该项目介绍，详细了解水头镇在改善五里桥水域环境质量、整合“小田”变“大田”、开发荷花池等农村生态环境整治方面的经验做法。第二站，支部党员前往朴里村“党建+”邻里中心，认真听取了朴一村党委副书记吕炜明的介绍。近年来朴一村在党建引领作用下，走出一条“三步走”的善治之道，尤其是创新“产权证明”无偿拆除房屋、新建房屋退让1.5米写入村规民约，村民积极响应，有效解决村道狭窄、公共空间不足等问题。第三站，支部党员来到全国首家成立党组织的非公有制企业、福建省首个跨地区的非公有制企业党组织——南益集团党委。在南益党建馆，讲解员的生动讲解下南益集团的党建发展历程。1984年，中共南安县南丰针织厂支部委员会成立，成为全国第一家非公有制企业党组织，被誉为“闽南第一枝报春花”。1997年6月，中共南益集团(福建)有限公司委员会成立，成为福建省第一个跨地区的非公有制企业党组织。讲解员通过翔实的文字、图片、实物等资料，讲述企业如何紧随党的步伐，从手抬肩扛到智造升级、从一个支部升格为集团党委的发展印记，回顾了南益集团40年来的党建工作，从彼时的南丰针织厂到如今遍布各行各业的40多家“南”字号企业，南益集团既生动诠释何为“晋江经验”，更为新时代加强非公党建工作提供可遵循、可借鉴的“南益样板”，系统陈述了“党建强、发展强”的南益实践。本次支部党员教学实践活动深入南安水头镇，通过实地参观学习，不仅见证了五里桥文化公园生态治理的成效与智慧，还深刻体会到朴里村“党建+”模式在乡村振兴中的引领作用，尤其是村民自治的创新实践。在南益集团党建馆的参观，则展现了非公企业党建的蓬勃生机与深远影响。此次活动不仅增强了党员们的党性修养，更激发了大家在新时代背景下，积极探索党建引领经济社会发展新路径的热情与责任感。

计算机科学与技术系教工党支部与银川能源学院信息传媒学院教工党支部开展联合主题党日活动

    6月12日，计算机科学与技术系教工党支部与来访的银川能源学院信息传媒学院教工党支部开展联合主题党日活动。活动在翔安西部片区三号楼3304会议室进行，支部书记曲延云同志主持会议，计算机科学与技术系主任刘向荣教授参加此次活动。  会议伊始，与会同志共同学习了习近平总书记在二十届中央纪委三次全会上发表的重要讲话精神——深入推进党的自我革命、坚决打赢反腐败斗争攻坚战持久战。随后，双方支部负责同志分别介绍了各自支部开展推进全面从严治党的学习教育情况，并针对全面从严治党这一主题进行了热烈的讨论。  在交流环节，与会双方针对专业以及课程建设、科研项目、师资培养、本科毕业设计等内容进行了分享与交流。曲延云同志介绍了计算机科学与技术系教工党支部的基本情况。刘向荣教授以“入厦正当时”为题介绍了计算机科学与技术系的基本情况。李杰院长介绍了信息传媒学院的基本情况。随后，双方针对闽宁协作、学生学业培养、教师队伍建设、科研算力建设等方面开展了深入的交流与探讨。  最后，与会双方共同参观了信息学院西部片区新院址以及部分实验室，并合影留念。计算机科学与技术系教工党支部2024年06月12日

计算机科学与技术系教工党支部开展师生党员集中培训第二课学习交流

    5月13日，计算机科学与技术系教工党支部开展专题组织生活会，学习交流校师生党员集中培训第二课。会议在翔安西部片区一号楼108会议室以线上线下相结合的方式展开，支部委员黄晓阳同志主持本次会议。    会议主题是开展校党委书记张荣为全校师生作的第二次集中培训——“中国式现代化与世界一流大学建设”的学习讨论。在培训中，张荣书记勉励全体师生要在思根淬炼中提升能力本领、在专业训练中提升能力本领、在实践锻炼中提升能力本领。在新时代新征程上奋力谱写与时俱进建设世界一流大学的崭新篇章，为以中国式现代化全面推进强国建设、民族复兴伟业作出新的更大贡献。    会后，支部党员进行了充分且热烈的学习心得交流与讨论。林子雨同志认为，作为党员，深感中国式现代化对于一流大学建设的重要性。通过学习，认识到现代化不仅是经济技术的飞跃，更是文化自信的体现。一流大学应立足中国实际，传承中华文化，培养具有国际视野的创新人才。将以更加坚定的信念，积极投身到大学建设中，为实现中国式现代化贡献自己的力量。    陈龙彪同志说，在今天的党课中，张荣书记深入阐释了厦门大学在中国式现代化进程中的关键角色。他强调了根据国家战略培养人才、加强科研以及慕课系统在疫情期间的重要作用。通过这次学习，我更加明白了一流大学的使命不仅限于教育和研究，还包括文化传承与创新。作为一名党员，我深切体会到要在自己的岗位上，以不懈的努力和坚定的信念，为推动国家的现代化贡献自己的力量。这是一场思想的洗礼，也是对未来方向的明确指引。    陈中贵同志谈到，听完“中国式现代化与一流大学建设”主题报告，深感启发。报告逻辑清晰，从中国现代化发展脉络入手，详实阐述一流大学建设的重要性。案例丰富，以厦大的发展为主线，展示中国高等教育的多样性与活力。数据支撑充足，各项指标实现跃升，凸显中国大学的国际竞争力。这样的报告不仅让我对中国高等教育发展有了更深入的了解，也激励我们在自己的领域里追求卓越。    支部新进党员黄华勋同志说，今天张荣书记的讲话加深了我对厦门大学在我国现代化中扮演的关键角色的认识。他详细讲解了按照国家战略培养人才、强化科研活动，以及疫情期间慕课系统的重要性，使我深刻感受到作为大学教师及党员的重要职责。我们的角色不仅限于知识传播，还包括文化的继承与创新。这次党课不仅是对思想的一次净化，也清晰指出了我们的奋斗目标。我将坚持不懈地努力，在我的岗位上为国家的现代化进程贡献力量。计算机科学与技术系教工党支部2024年05月13日

计算机科学与技术系教工党支部开展学习《中国共产党纪律处分条例》专题组织生活会

    5月8日，计算机科学与技术系教工党支部开展专题组织生活会，学习何丽新教授在信息学院关于解读《中国共产党纪律处分条例》的讲座。会议在翔安西部片区四号楼106会议室以线上线下相结合的方式展开，活动由支部委员陈胤燃同志主持。    会议伊始，主持人院党委副书记龙坚毅向与会同志介绍了组织生活会的主题——学习交流厦门大学法学院党委书记何丽新教授为信息学院师生所作的关于解读《中国共产党纪律处分条例》的讲座。在讲座中，何丽新教授结合多个具体的案例，详细介绍和深入解读了2023年新修订的条例，并指导学院师生学习和领会条例精神。    随后，与会党员进行了充分且热烈的学习心得交流与讨论。谢怡同志谈到，讲座中提到的条例修改情况和具体案例令自己获益匪浅，需要在今后的工作生活中认真学习各个条目，并以条例为指导把握好工作方向。    郑炜同志说，通过此次讲座，对党的纪律处分条例的权威性有了更深的认识。作为一名党员教师，应当在今后的教学工作中加强对学生的引导和教育，以新的条例为指导，做到为人师表。    新进青年教师沈耀斌同志谈到，条例的修改体现了党纪的与时俱进，对于党的团结和统一是十分重要的，作为一名刚工作的党员教师，要始终把党纪国法牢记在心中。    支部书记曲延云同志做了总结发言，她谈到新修订的《中国共产党纪律处分条例》犹如一把悬在头上的利剑，时刻提醒着每一位党员同志在思想和行动上要坚决维护党章和其他党内法规，坚决维护习近平总书记党中央的核心、全党的核心地位，坚决维护以习近平同志为核心的党中央权威和集中统一领导。要时刻牢记和深入领会条例的警示作用，在条例的监督指导下做好各项工作。  计算机科学与技术系教工党支部2024年05月08日

计算机科学与技术系教工党支部开展师生党员集中培训第一课学习研讨

    4月19日，计算机科学与技术系教工党支部开展专题组织生活会，学习交流校师生党员集中培训第一课。会议在翔安西部片区三号楼302会议室以线上线下相结合的方式展开，支部副书记严严同志主持本次会议。    会议伊始，严严同志向与会同志介绍了中共中央党校（国家行政学院）分管日常工作的副校长（副院长）、中国现代史学会会长谢春涛在厦门大学作的首场专题辅导报告“中国共产党与中华民族伟大复兴”。    随后，与会党员进行了充分且热烈的学习心得交流与讨论。曹刘娟同志谈到，中国共产党经历过百年的奋斗和建设，在新百年的开端面临着中西方的激烈技术竞争，作为一名科技工作者应该更加努力工作，为国家的科学技术发展贡献自己的力量。    范晓亮同志认为，在中国共产党的坚强带领下，我们国家发展蒸蒸日上，在新百年的征途上，应该更加努力地工作，为新百年的建设添砖加瓦。    唐璐同志说，报告中提到的中国共产党艰苦奋斗的历程和精神深深感染着所有师生，作为一名人民教师，不仅要把这种艰苦奋斗的精神延续下去，还要把这种精神传递给学生。    新进青年教师黄华勋同志谈到，作为一名新进教师，为了党的新百年建设，应该努力提升自己的综合业务能力和整体素质，同时提升科研和教学水平，做好教书育人工作。    支部书记曲延云同志总结道，谢春涛同志通过第一课的讲座带领我们重温了党的奋斗史。结合当今的百年未有之大变局，作为一名共产党员，作为一个党支部，做到统一思想、统一意志、统一行动是非常重要的。另外，在发展新质生产力的要求下，作为高校教师应更加努力地开展教学科研工作，让学术成果更好地转化，为突破卡脖子技术、发展新质生产力做出更好地贡献。    计算机科学与技术系教工党支部将根据上级党委的统一安排部署，认真学习领会和深入贯彻落实党的二十大精神，巩固拓展学习贯彻习近平新时代中国特色社会主义思想主题教育成果，在新时代精神的指引下让支部事业更上一层楼。计算机科学与技术系教工党支部2024年04月19日

学习2024年全国两会精神暨人工智能教学沙龙

----信息学院与教师发展中心、教务处联合开展“固定党日+”活动为深入学习贯彻习近平新时代中国特色社会主义思想和党的二十大精神，贯彻落实全国两会精神，大力弘扬教育家精神，提高教师教学能力，推动营造追求卓越教学文化，信息学院计算机科学与技术系教工党支部、通信系教工党支部、教师发展中心党支部、教务处现教中心党支部联合开展“固定党日+”和人工智能+教学教师沙龙活动，活动在翔安校区西部片区信息学院一号楼108举办。活动由计算机系教工党支部书记曲延云主持。活动第一项内容为传达学习全国两会精神。大家集体观看我校党委书记张荣畅谈拔尖创新人才具备什么样的核心素养、应该怎么培养的视频。随后，由美育与通识教育中心副主任、教务处副处长黄艳萍，教务处副处长陈雪芬，教师发展中心副主任薛成龙，信息学院党委副书记龙坚毅，计算机系副主任郑炜，计算机系教工支部书记曲延云分别进行全国政府工作报告的导读，并分享学习感悟。信息学院副书记龙坚毅最后总结表示，学院要和机关部门群策群力培育一流的教育生态，才能培养出真正一流的人才，满足国家和社会的需求，并希望今后有更多促进党建和业务发展深度融合的活动方式。活动第二项邀请北京市青年教学名师、清华大学融合式教学指导专家组副组长、清华大学计算机科学与技术系教授马昱春作主题为“人工智能大模型在编程类课程中的应用”的教学沙龙。马老师通过清华大学的课程实践分享了人工智能大模型在教学中的应用，并实际操作展示了“教学资源+AI编程教学助手+课堂讲授”有机结合的教学新范式，为在座的教师提供了教育技术应用的新思路。八十余名教师通过线上线下相结合的方式参会，与会教师纷纷表示收获颇多，现场互动氛围热烈。今后，计算机科学与技术系教工党支部，继续深入推进“党建+学科”融合发展模式，通过支部共建模式，围绕课程思政、AI赋能教学创新、产教融合等话题继续举办系列活动，以实际行动推动厦门大学的事业发展。（计算机系教工党支部、通信工程系教工党支部、教师发展中心党支部、教务处现教中心联合党支部、）                                                                 2024年3月15日

截图：

