湖南石油化工职业技术学院
Hunan Petrochemical Vocational Technology College

课程标准

课程名称：网络爬虫

课程代码：31091430

适用专业：大数据技术与应用

制订时间：2020 年 3 月

湖南石油化工职业技术学院

目　　录

《网络爬虫》课程标准

课程名称：网络爬虫

课程代码：31091430

总学时数：48 学时（理论课学时数：20； 实践课学时数：28）

适用专业：大数据技术

1 课程概述

1.1 课程的性质

《网络爬虫》是大数据技术与应用专业的一门专业核心课（技术技能课），旨在对学生的程序设计思想和技能进行强化，培养学生利用主流 scrapy 框架进行爬虫项目的设计和开发的能力。

1.2 课程定位

本课程以爬虫工程师岗位的基本要求为指导，依据该岗位真实业务内容与流程选取课程内容、构建学习单元，将目前爬虫程序必备功能组件如网页数据下载、数据分析、数据存储、网页地柜爬取等技术作为项目中的系列任务。课程内容编排符合循序渐进的认知规律，培养学生的网页爬虫实际应用能力。

先导课程： web 应用开发技术、python 基础、数据库基础。

1.3 课程设计思路

本课程依据网页爬虫开发岗位的 PGSD 能力要求而设置，主要工作时根据需求进行数据采集，获得有效数据，课程对应的职业能力分析具体如表 1-1 所示。

表 1-1“python 爬虫”课程对应 PGSD 能力要求。

能力类别	编号	内容
	P-A1.1	明确爬虫的目的、主题
	P-A1.2	明确爬虫的数据，含字段、时间范围等
	P-A1.3	明确爬虫的步骤、工具
	P-A2.1	能够使用 ullib 实现网页下载

	P-A2.2	能够使用正则表达式获取网页数据
职业能力	P-A2.3	能够使用 beautifulsoup 工具选择数据
	P-A2.4	能够使用 xpath 、css 选择数据
	P-A2.5	能够使用 scrapy 编写网页爬虫程序
	P-A2.6	能够使用 item 、pipeline 实现数据序列化与存储
	P-A2.7	能够使用 scrapy 实现网页递归爬取
	P-A2.8	能够进行网站爬虫程序综合开发
	G-A1	具备信息收集和信息处理能力
通用能力	G-A2 G-A3	具备结构化思维和数据化思维能力 具备一定的互联网和网页知识
	G-A4	具备一定的数学素养

2. 课程基本目标

2.1 素质目标

（1）培养学生良好的职业道德

（2）培养按时、守时的软件交付观念

（3）培养阅读设计文档、编写程序设计的能力

（4）培养学生的团队协作精神

（5）培养学生分析问题、解决问题的能力

（6）培养学生勇于创新、敬业乐观的工作作风

（7）培养学生自主、开放的学习能力

2.2 知识目标

（1）掌握爬虫程序设计理念

（2）掌握数据提取与存储思想

（3）掌握 scrapy 爬虫框架设计思想

（4）熟练掌握 ullib 网页下载方法

（5）熟练掌握正则表达式选取数据的规则

（6）熟练掌握 beautifulsoup 工具选取数据的方法

（7）熟练掌握 xpath、css 选择数据的方法

（8）熟练掌握 scrapy 网页爬取的工作流程

（9）熟练掌握 scrapy 中 item 、pipeline 数据的序列化输出方法

（10）熟练掌握 scrapy 中 spider 的网页递归爬取技术

2.3 技能目标

（1）熟练掌握 scrapy 中间件的使用方法

（2）能够完成真实业务逻辑向代码的转化

（3）能够独立分析解决技术问题

（4）自学能力强，能够快速准确地查找参考资料

（5）能够安好规范编写技术文档

（6）沟通能力强，能够与小组其他人通力合作

2.4 职业证书融通要求

3、课程教学内容及学时安排

3.1 课程主要内容说明

本课程重点是培养学生的网络爬虫使用的基本技能。要求学生掌握爬虫概述、Ullib 实现网站下载、使用正则表达式获取网页数据、使用 beautifulsoup 工具选择数据、使用 scrapy 编写网页爬虫程序、使用 item、pipeline 实现数据序列化与存储等基础知识。学生首先了解网络爬虫的特点、发展及推荐学习方法，然后学习使用 scrapy 实现网页递归爬取、第三方库相关知识等。

课程根据“理论实践一体化教学”模式，按照 Python 的有关知识由浅入深、从易到难进行教学，课后布置实训与习题练习，实现“教、学、做”一体，从而切实提高学生的持续发展能力。

3.2 课程组织安排说明

本课程主要使用集“教、学、做”于一体，采用案例演示法、项目教学法等教学方法，在电脑上理论结合实际，采用理实一体化教学模式完成课程组织和教学。

3.3 课程教学内容及要求

序号	教学单元（或者模块）	素质内容及要求	知识内容及要求	技能内容及要求	参考学时
1	爬虫概述	培养学生的团队精神和服务意识 培养学生的自主学习能力	● 能够初步了解爬虫的概念，了解爬虫的历史、发展、功能等 ●了解现有的爬虫工具， 使用爬虫工具爬取一次数据 ● 具备信息收集和信息处理能力 ● 具备自学能力，能适应行业的不断变革发展 ● 具备一定的设计素养	爬虫工具的使用 简单的数据爬取	4
2	前置技能准备	培养学生的团队精神和服务意识 培养学生的自主学习能力	● Python 语言回顾 ● web 开发基础回顾 ●具备一定的互联网和网页知识	使用 python 编写一个程序，使用 web 开发一个网页	4
3	Ullib 实现网站下载	培养学生的团队精神和服务意识 培养学生的自主学习能力	● 搭建前端开发环境 ● 搭建后端静态网页 ● 利用 urllib 下载后端网页 ● 编写程序实现编码（GBK,UTF -8 ）的自动识别与转换 ● 存储网页到文件或数据库	能够通 ullib 网页下载函数方法下载网页，实现编码的转换	4
4	使用正则表达式获取网页数据	培养学生的团队精神和服务意识 培养学生的自主学习能力	●搭建前端开发环境 ●搭建后端静态网页 ●利用 urllib 下载后端网页 ●使用正则表达式匹配并提取网页数据	能够根据功能组件的不同实现需求，使用正则表达式匹配并提取网页中的数据	6
5	使用 beautifulso	培养学生的团队精	●搭建前端开发环境	能够使用 beautifulsoup	4

	up 工具选择数据	神和服务意识 培养学生的自主学习能力	●搭建后端静态网页 ●利用 urllib 下载后端网页 ●使用 beautifulsoup 提取网页的数据 ●存储提取的数据	工具选择数据，掌握 find_all 等常用方法	
6	使 用 xpath css 选择数据	培养学生的团队精神和服务意识 培养学生的自主学习能力	●搭建前端开发环境 ●搭建后端静态网页 ●利用 urllib 下载后端网页 ●使用 xpath 、css 提取网页的数据 ●存储提取的数据	使用 xpath 、css 选择复杂的数据	6
7	使 用 scrapy 编写网页爬虫程序	培养学生的团队精神和服务意识 培养学生的自主学习能力	●搭建 scrapy 开发环境 ●搭建 web 后台网页 ●使用 scrapy 爬取网页文件 ●使用 xpath 、css 获取特征数据	能够使用 scrapy 网页爬取的工作流程爬取单个网页的某几个特征数据	6
8	使 用 item 、pipeline 实现数据序列化与存储	培养学生的团队精神和服务意识 培养学生的自主学习能力	●搭建 scrapy 开发环境 ●搭建 web 后台网页 ●使用 scrapy 爬取网页文件 ●使用 item 、pipeline 提取与存储数据	能够使用 scrapy、pipeline 进行数据提取与数据存储	6
9	使 用 scrapy 实现网页递归爬取	培养学生的团队精神和服务意识 培养学生的自主学习能力	●搭建 scrapy 开发环境 ●搭建 web 后台众多关联网页 ●使用 scrapy 爬取多层嵌套与关联的网页文件	能够使用 scrapy 中 spider 的网页递归爬取循环，实现数据的提取与存储	4

			●使用 item 、pipeline 提取与存储数据		
复习、考试					4
合计学时					48

4 教学实施建议

4.1 教学组织建议

① 教学方法多样化，教学内容真实化。建议教师在家偶尔过程中通过案例激发学生思考,基于真实的第三方数据和抓取的外部数据来布置任务,驱动教学，从而提高学生的学习积极性，提高学生实操能力。

② 教学手段现代化。利用多媒体、网络平台、信息系统、视频录像等现代化手段，强化实际操作技能的训练，提高课堂教学效率。

③ 教学组织团队化。运营管理实践工作都是以团队形式完成的，教学过程中同样采取分组方式来组织实践教学。每项实践活动都是一个完整的工作过程，因此都可以成立类似于企业的一个工作小组；每个小组由 5~6 人组成，小组工作要按企业化运作，实行组长负责制；并在班级或年级内开展小组竞赛，培养学生的团队协作能力和职业意识，提高学生的管理能力。

4.2 教学评价建议

4.2.1 课程内容评价要点

序号	单元（模块）	考核标准	权重比例%
1	课堂学习	包括出勤、课堂表现及课堂积极回答问题等	15
2	课后作业	是否按时、按质、按量完成教师布置的课后练习	15
3	课堂实训	能否实操出课堂练习	20
4	期末考试	由教师评定的笔试成绩	50

4.2.2 课程评价方法和内容

评价类型	评价方法	评价内容

职业素养（10%）	过程性评价（10%）	到课考勤，学习及工作态度、安全意识、质量观念、合作精神、敬业精神等纳入职业素养考核，在具体考核指标中体现。
理论知识（50%）	过程性评价（20%）	主要是课堂提问、平时作业、单元测验、期中测验等。
	终结性评价（30%）	主要是期末考试，评价综合专业理论知识掌握和运用能力，由计算机随机命题或人工命题组成标准试卷，尽量与国家临床医学检验技士职称资格考试接轨。
职业技能（40%）	过程性评价（20%）	实训报告、实际操作过程评价。
	终结性评价（20%）	建议考核核心技能项目 参照技能考核标准与要求，编制核心技能项目的评分标准，评分标准应涵盖操作规范性、结果准确性、人文关怀、沟通交流、操作安全等。

4.3 教材选用

1、教材选用建议

为了让学生掌握职业岗位工作所需的技术知识，顺利实施职业技能训练，授课承担部门应选用近几年出版的全国优秀的高职规划教材，并且采用项目驱动式的编写思路为宜。

2、教材编写建议

为了使教材适合高职教育以及现代技术发展快、创新多的特点，突出强调理论教学与实践操作紧密结合的一体化教学模式，自编教材应以“项目导向，任务驱动”为主线。

4.4 课程主讲教师和教学团队要求说明

本课程要求任课教师首先牢固树立中国特色社会主义理想信念，践行社会主义核心价值观，自觉增强立德树人、教书育人的荣誉感和责任感，学为人师，行为世范。最好由具有双师型素质的高学历的教程承担。要求教师具有扎实的专业知识和丰富的相关行业实际工作经验，具有一定职业教学教学能力，能够开展课

程教学改革和科学研究。

4.5 课程思政要求

全面推进课程思政建设，发挥好专业课程的育人作用。专业课程教学过程以专业知识和技能为载体，加强思想政治教育，充分发挥课堂主渠道功能，努力发掘课程中立德树人的要素，与思想政治理论课同向同行，形成协同效应。本专业课程思政具体要求如下。

1、课程教学与爱国主义教育相结合

通过选择优秀典型的行业企业案例、视频题材等重要思政教育内容，激发爱国热情，培养家国情怀。在专业教师引导之下，通过我国 IT 行业和大数据技术应用发展成就和实力的展示，开展爱国主义教育、中国梦教育，增强学生的国家认同感与民族自豪感。

2、课程教学与团队合作精神相结合

专业核心课程实训教学过程中，以实训任务为载体，以工作小组为单元，引导学生将企业本职工作经历融入学习过程，调动学习积极性，重点强调项目成员团队合作的原动力和凝聚力，树立了正确的集体观，培养团队合作精神。

3、课程教学与职业素养培养相结合

通过实践教学环节和企业经历，结合企业生产实际和行业人才素养需求，引入企业对优秀员工必备素质和基本规范的要求,引导学生自觉实践相关行业的职业精神和职业规范，增强职业责任感，培养学生良好的职业品德、职业纪律及职业责任心，教育学生爱岗敬业、讲究诚信、精益求精，在潜移默化中提高了学生未来岗位的适应能力。

4、课程教学与高职学生学情相结合

高职院校学生普遍基础薄弱、学习主动性不强，在这样的学情下，课程教学中教师应实时自我反思和自我总结，不断完善教学手段，增强学生的学习兴趣，提升学生的信心，提高学生的专业能力。

5、课程教学与实际项目案例相结合

教学中，引入实际企业或公司案例，通过理论课程教学结合实际项目案例的教学模式，引导学生提升自我意识、养成良好的职业精神和职业规范，在实际项目案例中不断总结自己、提升自我、提升团队作战意识和团队协作能力。

4.6 课程教学环境和条件要求

主要能够满足正常的课程教学、实习实训所需的专业教室、实训室。

1、专业教室基本条件

配备交互智能教育平板、黑（白板）、多媒体计算机、投影设备、音响设备，互联网接入或WIFI环境，并具有网络安全防护措施。安装应急照明装置并保持良好状态，符合紧急疏散要求、标志明显、保持逃生通道畅通无阻。

序号	教学场地	设施配置	功能
1	投影室	投影仪、相关软件等	公共课程教学
2	多媒体机房	电脑、投影仪、相关软件等	专业课理实一体化教学

4.7 教学资源的开发与利用

1、常规教学文件

常规教学文件应包括：授课计划、教案、讲稿、教学课件等资料。

2、教学资源

应建立适合教师教学的《教学案例库》和适合学生自主学习的《导学手册》和《习题集》。

4.8 其它