《Python 网络爬虫程序技术》课程标准

课程类别:专业课
适用对象:计算机软件技术专业三年制学生
授课单位：软件学院
学分：3　　学 时：64
编写执笔人：　黄锐军　编写日期：2024-20
审定负责人：　薛国伟　审定日期：2024-26

一、课程定位和课程设计

（一）课程性质

本课程注重对学生职业能力和创新精神、实践能力的培养。本课程旨在对学生的程序设计思想和技能进行，培养学生利用主流 scrapy 框架进行爬虫项目的设计与开发。《Python 网页爬虫程序技术》课程是软件技术专业 Python 方向的专业课程，是融理论与实践一体化，教、学、做一体化的专业课程，是基于设计的工作过程系统化学习领域课程，是工学结合课程。

（二）人才培养中的作用

该课程是软件技术专业课程，目的是综合前面两年学习的知识技能，综合开发一个完整的软件项目。项目严格按照软件开发的流程进行，项目由学生自由组队完成，流程包括项目调研、编写项目需求、编写项目总体设计、编写代码、项目测试、项目部署与应用、编写项目使用说明书等环节。培养学生软件知识应用能力、编码能力、团队精神、综合开发能力。

（三）课程设计思路

课程开发遵循的基于工作过程导向的现代职业教育指导思想，课程的目标是网页爬虫程序开发职业能力培养。

课程教学内容的取舍和内容排序遵循以工作需求为目标原则，务求反映当前网页爬虫开发的主流技术和主流开发工具，同时重视软件工程的标准规范，重视业内工作过程中的即成约定，努力使学生的学习内容与目标工作岗位能力要求无缝对接。

课程设计的总体原则是“基于课程对应的就业岗位工作过程”，基于此原则课程组进行了充分的调研，过程如下：

第一步：岗位需求调研；

第二步：确定课程对应的岗位能力需求；

第三步：根据岗位能力需求确定教学目标；

第四步：明确课程的学习任务；

第五步：并根据任务设计学习活动

二、课程目标

（一）素质目标（含课程思政、职业素养、工匠精神、双创精神等）

课程教学中积极融入课程思政的元素，首先从项目的选题开始就注重项目的积极性与思想性，在项目开发过程中充分融入社会主义核心价值观的教育，注重教书育人，注重职业素养、工匠精神的培养。最后一个项目是学生自主选题的项目，充分调动学生积极性与创造性，培养学生双创精神。

（二）能力目标（含知识、技能目标、岗位核心能力等）

知识与能力目标：

本课程内容涵盖了对学生在“基本理论”、“基本技能”和“职业素质”三个层次的培养。以网页爬虫开发岗位必备的开发技能为重点并具备相应的理论基础的同时，注重综合职业素质的养成，课程采用启发诱导式教学，鼓励学生“勤于思考，勤于动手”。

1)基本理论要求：

- 掌握爬虫程序设计理念；
- 掌握数据提取与存储思想

2)基本技能要求：

- 熟练掌握 ullib 网页下载方法；
- 熟练掌握正则表达式选取数据的规则；
- 熟练掌握 BeautifulSoup 工具选择数据的方法；
- 熟练掌握 xpath 选择数据的方法；
- 熟练掌握 css 选择数据的方法；
- 熟练掌握递归爬取多个页面数据的方法；
- 熟练掌握深度优先爬取多个页面数据的方法；
- 熟练掌握广度优先爬爬取多个页面数据的方法；
- 熟练掌握多线程下载图像的方法；
- 熟练掌握数据存储方法

- 熟练掌握 scrapy 框架的应用
- 熟练掌握 scrapy 中 item、pipline 数据存储方法
- 熟练掌握 scrapy 中使用 xpath 查找数据的方法
- 熟练掌握 selenium 爬取动态数据的应用场景
- 熟练掌握 selenium 中查找数据的方法
- 熟练掌握 selenium 中等待元素的机制
- 熟练掌握 selenium 执行 JavaScript 程序的方法

3)职业素质要求：

- 能够完成真实业务逻辑向代码的转化；
- 能够独立分析解决技术问题；
- 自学能力强，能够快速准确地查找参考资料；
- 能够按照规范编写技术文档；
- 沟通能力强，能够与小组其他成员通力合作。

三、课程内容与要求

项目 1 爬取外汇网站数据

1.1 外汇网站项目任务
1.2 搭建爬虫程序开发环境
1.3 Flask 创建 Web 网站
1.4 Flask 网页模板
1.5 GET 与 POST 方法访问网页
1.6 正则表达式匹配数据
1.7 综合项目 爬取模拟外汇网站数据
1.8 实战项目 爬取实际外汇网站数据

项目 2 爬取名言网站数据

2.1 名言网站项目任务
2.2 BeautifulSoup 装载 HTML 文档
2.3 BeautifulSoup 查找 HTML 元素
2.4 BeautifulSoup 遍历文档元素
2.5 BeautifulSoup 使用 CSS 语法查找
2.6 综合项目 爬取模拟名言网站数据
2.7 实战项目 爬取实际名言网站数据

项目 3 爬取天气网站数据

3.1 天气网站项目任务
3.2 lxml 装载 HTML 文档
3.3 xpath 查找 HTML 元素
3.4 综合项目 爬取模拟天气网站数据

3.5 实战项目 爬取实际天气网站数据

项目 4 爬取旅游网站数据

4.1 旅游网站项目任务
4.2 网站树的爬取路径
4.3 爬取多页面网站数据
4.4 Python 实现多线程
4.5 爬取旅游网站图像
4.6 综合项目 爬取模拟旅游网站数据
4.7 实战项目 爬取实际旅游网站数据

项目 5 爬取图书网站数据

5.1 图书网站项目任务
5.2 scrapy 创建爬虫程序
5.3 scrapy 通过 BeautifulSoup 爬取数据
5.4 scrapy 通过 xpath 查找元素
5.5 scrapy 爬取关联网页数据
5.6 scrapy 通过 xpath 爬取数据
5.7 scrapy 通过管道存储数据
5.8 综合项目 爬取模拟图书网站数据
5.9 实战项目 爬取实际图书网站数据

项目 6 爬取商城网站数据

6.1 商品网站项目任务
6.2 selenium 编写爬虫程序
6.3 selenium 查找 HTML 元素
6.4 selenium 实现用户登录
6.5 selenium 爬取 Ajax 网页数据
6.6 selenium 等待 HTML 元素
6.7 综合项目 爬取模拟商城网站数据
6.8 实战项目 爬取实际商城网站数据

四、学时分配表（以课题或知识单元编排）

序号	模 块 名 称	学 时	其中	
			讲授	实践
1	项目 1 爬取外汇网站数据	8	4	4
2	项目 2 爬取名言网站数据	8	4	4
3	项目 3 爬取天气网站数据	8	4	4
4	项目 4 爬取旅游网站数据	12	6	6
	项目 5 爬取图书网站数据	12	6	6

	项目 6 爬取商城网站数据	16	8	8
	总计	64	32	32

注：12 学时作为考核、系统测试、验收、发布、讲评等

五、实践教学

序号	模 块 名 称	学 时
1	项目 1 爬取外汇网站数据	4
2	项目 2 爬取名言网站数据	4
3	项目 3 爬取天气网站数据	4
4	项目 4 爬取旅游网站数据	6
5	项目 5 爬取图书网站数据	6
6	项目 6 爬取商城网站数据	8
	总计	32

六、考核方式及成绩评定

1．基本思路

该课程是面向工作过程的工学结合特点的课程，因此重视学生的知识与能力的培养与训练，采用完成程序结果与过程考核相结合的方法。

2．课程设计的评分标准（共 100 分）

考核内容	考核细节名称	考核权重	权重
学习情境考核	项目 1 Python 爬虫基础	10%	60%
	项目 2 爬取名人名言数据	10%	
	项目 3 爬取天气预报数据	20%	
	项目 4 多线程爬取网站图像	20%	
	项目 5 爬取图书网站数据	20%	
	项目 6 爬取商城网站数据	20%	
学习过程考核	上课出勤	25%	40%

	学习态度	25%	
	合作精神	25%	
	组织协调	25%	
合计			100%

（二）独立实践项目

本课程无独立实践项目。

七、必要说明

（一）课程开设的基本条件

本课程需要 Python3.6 + PyCharm IDE。

（二）建议使用的教材及教学参考书

[教材]

《Python 爬虫项目教程(微课版》第 2 版，黄锐军编写，人民邮电出版社，2025

[参考书]

《Python 程序设计》赵家刚主编，人民邮电出版社，2016 年版

大纲编订部门：软件学院　　执笔人：　黄锐军

大纲审订部门：软件学院　　审订部门负责人：蔡铁

编订日期：　2024-08-26　　修订日期：　2024-08-26

深圳信息职业技术大学