**在下划线上填写答案。**

# 根据学习大纲补充学习目标

**引言**

学习目标：了解网页爬虫的重要性和应用场景。

内容：

介绍人工智能在金融数据采集中的应用，网页爬虫在获取上市公司财务报告中的作用。

**网页爬虫基础理论**

学习目标：

内容：

什么是网页爬虫？

网页爬虫的工作流程和基本结构。

机器人协议（robots.txt）的作用和遵守方法。

**常用网页爬虫工具简介**

学习目标：

内容：

BeautifulSoup：用于解析HTML和XML文档。

Scrapy：一个快速、高效的网页爬虫框架。

Selenium：用于模拟浏览器操作，处理动态网页。

**环境搭建与工具安装**

学习目标：掌握网页爬虫工具的安装和环境配置。

内容：

安装Python和pip包管理工具。

安装并配置BeautifulSoup、Scrapy、Selenium。

**网页数据解析**

学习目标：

内容：

使用BeautifulSoup进行HTML解析。

使用XPath和CSS选择器提取数据。

实践操作：解析一个简单的网页数据。

**动态网页处理**

学习目标：掌握处理动态网页的方法。

内容：

使用Selenium模拟浏览器操作。

处理动态加载的数据，如Ajax请求。

实践操作：爬取一个动态加载的网页数据。

**Scrapy框架实战**

学习目标：

内容：

Scrapy项目的创建和基本配置。

编写爬虫脚本，设置抓取规则。

数据存储与导出：将数据保存为CSV、JSON等格式。

实践操作：构建一个Scrapy爬虫项目，获取上市公司财务报告。

**数据清洗与处理**

学习目标：

内容：

数据去重、缺失值处理。

数据格式转换与规范化。

实践操作：对获取的财务报告数据进行清洗和处理。

**综合项目实践**

学习目标：综合运用所学知识进行网页爬虫项目。

内容：

项目介绍：构建一个完整的财务报告爬虫项目。

数据采集：获取多个网站的财务报告数据。

数据处理与分析：清洗并分析获取的数据。

项目评审：展示爬虫成果，讲解实现思路和遇到的问题。

**总结与答疑**

学习目标：巩固所学内容，解决学员疑问。

内容：培训内容总结，关键点回顾，学员提问与解答。

**评估测试**

学习目标：评估学员掌握的网页爬虫技能。

内容：

理论测试：关于网页爬虫基本概念和工具使用的笔试。

实操测试：要求学员完成指定的网页爬虫任务。