

# 2024程序设计实践A - 期末大作业要求

作业提交截止时间 Jul 31, 2024

目标：通过该综合项目，全面掌握Scrapy数据爬取、数据处理与可视化、Web开发等技术，并具备整合多项技术解决实际问题的能力。

## 项目描述

设计并实现一个综合项目，要求使用Scrapy爬取数据，对数据进行处理与可视化，并通过一个Web应用展示采集到的信息和分析结果。

## 项目要求

1. 数据采集	<ul style="list-style-type: none"><li>使用Scrapy爬取指定网站的数据。</li><li>需要爬取的信息根据项目具体要求而定，如：标题、日期、内容、评分等。</li><li>数据存储到数据库（如SQLite或MongoDB）中。</li></ul>
2. 数据处理与分析	<ul style="list-style-type: none"><li>使用Pandas或其他数据处理工具对采集到的数据进行清洗和整理。</li><li>对数据进行统计分析或情感分析等，生成相应的分析结果。</li><li>使用Matplotlib或其他可视化工具对分析结果进行可视化展示。</li></ul>
3. Web应用开发	<ul style="list-style-type: none"><li>使用Flask或Django等Web框架开发一个Web应用。</li><li>实现一个Web界面，展示数据和分析结果。</li><li>提供搜索和过滤功能，例如根据特定字段进行搜索和筛选。</li><li>将数据可视化图表嵌入到Web页面中，提供分析结果的展示。</li></ul>
4. 文档与演示	<ul style="list-style-type: none"><li>撰写项目文档，详细描述项目的设计、实现过程和使用方法。</li><li>录制项目演示视频，讲解项目的功能和特点，并制作演示PPT。</li><li>完整材料打包（文件命名格式：学号_姓名_项目名称）上传至Int.xmu.edu.cn的期末大作业提交处</li></ul>
评估标准	<ul style="list-style-type: none"><li>数据采集的完整性和准确性。</li><li>数据处理与分析的合理性和准确性。</li><li>数据可视化的合理性和美观性。</li><li>Web应用的功能性、用户体验和界面设计。</li><li>项目文档的完整性和规范性。</li><li>项目演示视频的清晰度和表现力，包括PPT内容的详细讲解和实际项目功能的展示。</li></ul>

## 推荐的可选项目（也可自己拟定）

### 1. 城市天气数据采集与分析展示系统

- 数据采集：使用Scrapy爬取多个城市的天气数据。
- 数据处理与可视化：清洗整理数据，并生成温度变化图、湿度变化图等。
- Web应用开发：使用Flask或Django开发Web应用，展示天气信息和数据分析结果。

### 2. 全球新闻数据采集与分析展示系统

- 数据采集：使用Scrapy爬取多个新闻网站的新闻数据。
- 数据处理与可视化：清洗整理数据，并生成新闻发布频率图、新闻类别分布图等。
- Web应用开发：使用Flask或Django开发Web应用，展示新闻信息和数据分析结果。

### 3. 在线商品评论数据采集与情感分析系统

- 数据采集：使用Scrapy爬取某电子商务网站的商品评论数据。
- 数据处理与情感分析：清洗整理数据，进行情感分析，并生成情感分布图、评分与情感关系图等。
- Web应用开发：使用Flask或Django开发Web应用，展示商品评论信息和情感分析结果。

### 4. 房地产市场数据采集与分析展示系统

- 数据采集：使用Scrapy爬取某房地产网站的房源数据。
- 数据处理与分析：清洗整理数据，进行房价分布、面积分布等统计分析，并生成相关图表。
- Web应用开发：使用Flask或Django开发Web应用，展示房源信息和市场分析结果。

可以选择以上题目之一进行实践，也可以自行拟定项目，并确保项目包含数据采集、数据处理与可视化、Web应用开发三个部分。项目完成后，需撰写详细的项目文档，录制项目演示视频，讲解项目的功能和特点，并制作演示PPT。

下面是可选项目开发建议：

## 1. 城市天气数据采集与分析展示系统

项目描述：

设计并实现一个综合项目，要求学生使用Scrapy爬取多个城市的天气数据，对数据进行处理与可视化，并通过一个Web应用展示采集到的天气信息。

项目要求：

### 1. 数据采集：

- 使用Scrapy爬取某天气网站（如Weather.com、AccuWeather等）的城市天气数据。
- 需要爬取的信息包括：城市名称、日期、温度、湿度、天气状况等。
- 数据存储到数据库（如SQLite或MongoDB）中。

### 2. 数据处理与可视化：

- 使用Pandas或其他数据处理工具对采集到的数据进行清洗和整理。
- 使用Matplotlib或其他可视化工具对各城市的天气数据进行可视化分析。
- 生成相关图表，如温度变化图、湿度变化图等。

### 3. Web应用开发：

- 使用Flask或Django等Web框架开发一个Web应用。
- 实现一个Web界面，展示城市天气信息。
- 提供搜索和过滤功能，例如根据城市名称、日期等进行搜索和筛选。
- 将数据可视化图表嵌入到Web页面中，提供天气数据分析结果的展示。

#### 4. 文档与演示：

- 撰写项目文档，详细描述项目的设计、实现过程和使用方法。
- 进行项目演示，展示项目的功能和特点。

#### 项目步骤建议：

##### 1. 数据采集：

- 选择目标网站，分析网页结构，确定要爬取的信息。
- 使用Scrapy编写爬虫，爬取目标网站的城市天气数据。
- 将爬取到的数据存储到数据库中。

##### 2. 数据处理与可视化：

- 使用Pandas加载并处理数据，进行必要的清洗和整理。
- 使用Matplotlib生成数据可视化图表，如温度变化图、湿度变化图等。

##### 3. Web应用开发：

- 使用Flask或Django创建Web应用，设计前端页面布局。
- 编写视图函数，处理城市天气信息的展示和搜索功能。
- 将数据可视化图表嵌入到Web页面中。

##### 4. 文档与演示：

- 撰写详细的项目文档，包括项目介绍、技术方案、实现步骤、使用说明等。
- 进行项目演示，展示项目的主要功能和特点。

## 2. 全球新闻数据采集与分析展示系统

#### 项目描述：

设计并实现一个综合项目，要求学生使用Scrapy爬取全球新闻网站的新闻数据，对数据进行处理与可视化，并通过一个Web应用展示采集到的新闻信息。

项目要求：

#### 1. 数据采集：

- 使用Scrapy爬取多个新闻网站（如CNN、BBC、Reuters等）的新闻数据。
- 需要爬取的信息包括：新闻标题、发布时间、作者、新闻内容、新闻类别等。
- 数据存储到数据库（如SQLite或MongoDB）中。

#### 2. 数据处理与可视化：

- 使用Pandas或其他数据处理工具对采集到的数据进行清洗和整理。
- 使用Matplotlib或其他可视化工具对新闻数据进行分析 and 可视化。
- 生成相关图表，如新闻发布频率图、新闻类别分布图等。

#### 3. Web应用开发：

- 使用Flask或Django等Web框架开发一个Web应用。
- 实现一个Web界面，展示新闻信息。
- 提供搜索和过滤功能，例如根据新闻标题、发布时间、新闻类别等进行搜索和筛选。
- 将数据可视化图表嵌入到Web页面中，提供新闻数据分析结果的展示。

#### 4. 文档与演示：

- 撰写项目文档，详细描述项目的设计、实现过程和使用方法。
- 录制项目演示视频，讲解项目的功能和特点，并制作演示PPT。

项目步骤建议：

#### 1. 数据采集：

- 选择目标新闻网站，分析网页结构，确定要爬取的信息。
- 使用Scrapy编写爬虫，爬取目标新闻网站的数据。
- 将爬取到的数据存储到数据库中。

#### 2. 数据处理与可视化：

- 使用Pandas加载并处理数据，进行必要的清洗和整理。
- 使用Matplotlib生成数据可视化图表，如新闻发布频率图、新闻类别分布图等。

#### 3. Web应用开发：

- 使用Flask或Django创建Web应用，设计前端页面布局。

- 编写视图函数，处理新闻信息的展示和搜索功能。

- 将数据可视化图表嵌入到Web页面中。

#### 4. 文档与演示：

- 撰写详细的项目文档，包括项目介绍、技术方案、实现步骤、使用说明等。

- 制作演示PPT，包含项目的主要内容和展示。

- 录制项目演示视频，讲解PPT中的内容，展示项目的功能和特点。

### 3. 在线商品评论数据采集与情感分析系统

#### 项目描述：

设计并实现一个综合项目，要求学生使用Scrapy爬取某电子商务网站的商品评论数据，对数据进行情感分析与可视化，并通过一个Web应用展示采集到的评论信息和情感分析结果。

#### 项目要求：

##### 1. 数据采集：

- 使用Scrapy爬取某电子商务网站（如亚马逊、京东、淘宝等）的商品评论数据。

- 需要爬取的信息包括：商品名称、评论者、评论内容、评分、评论日期等。

- 数据存储到数据库（如SQLite或MongoDB）中。

##### 2. 数据处理与情感分析：

- 使用Pandas或其他数据处理工具对采集到的数据进行清洗和整理。

- 使用情感分析工具包（如TextBlob、VADER等）对评论数据进行情感分析，判断评论的正面、负面或中性情感。

- 使用Matplotlib或其他可视化工具对情感分析结果进行可视化展示，如情感分布图、评分与情感关系图等。

##### 3. Web应用开发：

- 使用Flask或Django等Web框架开发一个Web应用。

- 实现一个Web界面，展示商品评论信息和情感分析结果。

- 提供搜索和过滤功能，例如根据商品名称、评分、评论日期等进行搜索和筛选。

- 将数据可视化图表嵌入到Web页面中，提供情感分析结果的展示。

##### 4. 文档与演示：

- 撰写项目文档，详细描述项目的设计、实现过程和使用方法。

- 录制项目演示视频，讲解项目的功能和特点，并制作演示PPT。

项目步骤建议：

#### 1. 数据采集：

- 选择目标电子商务网站，分析网页结构，确定要爬取的信息。
- 使用Scrapy编写爬虫，爬取目标网站的商品评论数据。
- 将爬取到的数据存储到数据库中。

#### 2. 数据处理与情感分析：

- 使用Pandas加载并处理数据，进行必要的数据清洗和整理。
- 使用TextBlob或VADER进行情感分析，添加情感标签（正面、负面、中性）。
- 使用Matplotlib生成数据可视化图表，如情感分布图、评分与情感关系图等。

#### 3. Web应用开发：

- 使用Flask或Django创建Web应用，设计前端页面布局。
- 编写视图函数，处理商品评论信息和情感分析结果的展示和搜索功能。
- 将数据可视化图表嵌入到Web页面中。

#### 4. 文档与演示：

- 撰写详细的项目文档，包括项目介绍、技术方案、实现步骤、使用说明等。
- 制作演示PPT，包含项目的主要内容和展示。
- 录制项目演示视频，讲解PPT中的内容，展示项目的功能和特点。

## 4. 房地产市场数据采集与分析展示系统

项目描述：

设计并实现一个综合项目，要求学生使用Scrapy爬取房地产网站的房源信息，对数据进行分析与可视化，并通过一个Web应用展示采集到的房源信息和市场分析结果。

项目要求：

#### 1. 数据采集：

- 使用Scrapy爬取某房地产网站（如链家、贝壳、Zillow等）的房源数据。
- 需要爬取的信息包括：房源名称、地址、价格、面积、房型、发布日期等。

- 数据存储到数据库（如SQLite或MongoDB）中。

## 2. 数据处理与分析：

- 使用Pandas或其他数据处理工具对采集到的数据进行清洗和整理。
- 对房源数据进行统计分析，如房价分布、面积分布、各区域房源数量等。
- 使用Matplotlib或其他可视化工具生成相关图表，如房价分布图、面积分布图、各区域房源数量图等。

## 3. Web应用开发：

- 使用Flask或Django等Web框架开发一个Web应用。
- 实现一个Web界面，展示房源信息和市场分析结果。
- 提供搜索和过滤功能，例如根据房源名称、地址、价格、面积等进行搜索和筛选。
- 将数据可视化图表嵌入到Web页面中，提供市场分析结果的展示。

## 4. 文档与演示：

- 撰写项目文档，详细描述项目的设计、实现过程和使用方法。
- 录制项目演示视频，讲解项目的功能和特点，并制作演示PPT。

## 项目步骤建议：

### 1. 数据采集：

- 选择目标房地产网站，分析网页结构，确定要爬取的信息。
- 使用Scrapy编写爬虫，爬取目标网站的房源数据。
- 将爬取到的数据存储到数据库中。

### 2. 数据处理与分析：

- 使用Pandas加载并处理数据，进行必要的清洗和整理。
- 对房源数据进行统计分析，如房价分布、面积分布、各区域房源数量等。
- 使用Matplotlib生成数据可视化图表，如房价分布图、面积分布图、各区域房源数量图等。

### 3. Web应用开发：

- 使用Flask或Django创建Web应用，设计前端页面布局。
- 编写视图函数，处理房源信息和市场分析结果的展示和搜索功能。
- 将数据可视化图表嵌入到Web页面中。

### 4. 文档与演示：

- 撰写详细的项目文档，包括项目介绍、技术方案、实现步骤、使用说明等。
- 制作演示PPT，包含项目的主要内容和展示。
- 录制项目演示视频，讲解PPT中的内容，展示项目的功能和特点。