

# 2022 年度“楚怡杯”湖南省职业院校技能竞赛 赛项规程

## 一、赛项名称

1. 赛项名称：Python 程序开发
2. 赛项组别：高职高专组
3. 赛项归属：电子信息大类

## 二、竞赛内容

Python 程序开发赛项以企业真实项目为基础，采用市场主流软件开发架构和实际操作形式进行现场编程设计。竞赛采用“程序开发”、“爬虫应用”、“数据清洗”、“数据分析”、“数据可视化”等 5 种题型。主要涉及的知识和技能如下：

### 1. 程序开发

- (1) Python 编程基础；
- (2) HTML5、CSS3、JavaScript、ES6 基本语法；
- (3) Django 框架设计，包括服务发布、路由映射、模板编程、数据处理等；
- (4) 理解需求，规划业务流程，并编码实现功能；
- (5) 基于 RESTful 规范设计 API 接口；
- (6) MySQL 数据库存储。

### 2. 爬虫应用

- (1) 查看网页源码，分析网页结构，明确数据采集对象；
- (2) 爬虫知识体系与相关工具；
- (3) 规则文件数据和关系型数据库数据抓取以及数据同步。

### 3. 数据清洗

- (1) 利用 Numpy、Pandas 进行数据取值和选择；
- (2) 操作字符串进行数据去重、填充操作；
- (3) 对数据进行计算。

### 4. 数据分析

- (1) 编写 Python 程序，去除数据中所有的标点符号；
- (2) 数据结果存储等文件及文件夹操作。

### 5. 数据可视化

- (1) Matplotlib、pyecharts 绘图；
- (2) 基本图形的绘制；
- (3) 高级图形的绘制。

## 三、竞赛方式

个人赛。

## 四、竞赛时量

总时量：240 分钟。

## 五、名次确定办法

以竞赛总成绩从高到低排序确定名次，不设并列名次。

成绩相同的，按照竞赛项目题型的顺序对选手得分的高低进行排序，如果第一个题型能得到排序则按此确定名次，否则对后续题型得分依次进行排序，只要能区分名次就终止后续题型的判断。如果以上都相同，按照代码行数从少到多排序（以行数少为优，空行不计入行数）。

## 六、评分标准与评分细则

### 1. 评分标准

竞赛满分为 100 分。竞赛项目内容如下：

**程序开发（30%）：**程序开发模块重点考核参赛选手针对实际问题的程序编程、REST API 设计及数据库使用的能力。

**爬虫应用（20%）：**爬虫应用重点考核参赛选手对开源海量信息获取能力、爬虫相关知识体系、网页结构分析及数据存储的熟练程度与编程能力。

**数据清洗（15%）：**数据清洗模块重点考核参赛选手处理数据缺失、数据去重、数据转换及数据集成的能力。

**数据分析（20%）：**数据分析模块重点考核参赛选手对应用 Python 高级特性和数据分析软件包进行简单与复杂科学计算机数据分析的能力，以及考察选手基于实际应用选择分析的能力。

**数据可视化（15%）：**数据可视化模块重点考核参赛选手能够选择合适的图形表示数据分析结果的能力，包括基本图形和高级图形。

### 2. 评分细则

考核环节	考查点	分值	评分细则
程序开发 (30 分)	Python 编程基础，Python 基本语法、面向对象等知识	5	结果评分 (1) 运行结果完全达标(接口函数编写，Web 页面，业务逻辑，Django 后端，数据存储)：100% (2) 运行结果部分达标：按实现结果在总要求的百分比给分。 (3) 未实现：0%
	Web 技术开发，HTML5、CSS3、JavaScript 开发框架，ES6 基本语法、npm 模块管理	5	
	业务逻辑，对业务需求及功能的理解，规划业务流程，并编码实现功能	5	
	Django 框架，Django 服务发布、路由映射、模板编程、数据处理等功能设计；模型，视图，模板，认证等使用	5	
	RESTful 接口设计，基于 RESTful 规范设计 API 接口	5	
	数据存储，MySQL 数据库增删改查	5	
爬虫应用 (requests、urllib, 20 分)	使用开发者工具查看网页源码，分析网页结构，明确数据采集对象	5	结果评分 (1) 爬虫数据抓取按实现部分百分比给分，完全达标：100% (2) 未实现：0%
	构建数据采集请求，抓取网络数据	5	
	利用网络爬虫相关组件实现网络数据爬取	5	
	规则文件数据和关系型数据库数据抓取以及数据同步	5	
数据清洗 (15 分)	利用 Numpy、Pandas 进行数据取值和选择	5	结果评分 (1) 取值、去重、填充，完全达标：100% (2) 部分达标：按实现结果在总
	操作字符串进行数据去重、填充操作	5	

	对数据进行计算	5	要求的百分比给分。 (3) 未实现: 0%
数据分析 (20 分)	编写 Python 程序, 去除数据中所有的标点符号	10	结果评分 (1) 运行结果完全达标(去标点, 代码编写正确): 100%, 否则按实现百分比给分 (2) 未实现: 0%。
	编写一个 Python 程序, 用户输入开始搜索的路径, 查找该路径下(包含子文件夹内)所有的指定格式的文件, 并创建一个文件用来存放所有找到的文件的路径	10	
数据可视化 (matplotlib、 Pyecharts, 15 分)	使用图形化工具绘制基本图形, 如柱状图、饼图、折线图、散点图等;	8	结果评分 (1) 运行结果完全达标: 100% (2) 坐标或数据显示错位给 30% (3) 未实现: 0%
	能使用图形工具绘制高级图形, 如地图、热力图、雷达图、词云图等;	7	

## 七、赛项相关设施设备技术参数

### 1. 竞赛设备

选手以个人为单位参赛, 每人需要 1 台 PC 机(用于运行服务端与客户端开发)。

裁判区域: 供裁判休息及工作场地。配电脑, A4 激光打印机 1 台, 桌椅, 饮水机, 纸杯, 文具用品等。

### 2. 硬件环境及配置

序号	类别	参数
1	服务器	2U 机架式(50 人/台) 处理器: 2 颗 Intel Xeon 银牌 4214R 内存: ECC DDR4 32GB*8 硬盘: 1TB SSD * 2(组建 RAID 1) RAID: PERC H730P 2G 缓存阵列卡 网卡: 4*1GbE 机架安装导轨 750W 热插拔(1+1)冗余电源
2	台式机	参赛选手用机, Intel 酷睿 i5 或以上级别处理器, 8GB 或以上内存, 256G 或以上固态硬盘, 千兆网卡。
3	网络	千兆网络环境

### 3. 软件环境

竞赛涉及的所有软件均为正版或开源软件, 如下表所示。

序号	软件类型	软件名称
1	PC 操作系统	Windows10
2	IDE 开发和调试工具	Visual Studio Code 2021 PyCharm Community Edition 2021 及以上
3	浏览器	火狐浏览器 谷歌浏览器
4	数据库环境	Navicat MySQL 5.7
5	Python3.6 库	json random urllib.request BS4 time

		csv tqdm string pandas matplotlib.pyplot IPython.display re os django3.1.7
6	其他	Microsoft Visio 2013 及以上 Microsoft Office 2010 及以上 福昕 PDF 阅读器

备注：具体设备由赛点提供。

## 八、选手须知

### 1. 选手自带工（量）具及材料清单

无需选手自带工具。

### 2. 主要技术规范及要求

该赛项主要涉及以下国家标准，参赛选手在实施竞赛项目中要求遵循如下规范：

序号	标准号	中文标准名称
1	DB21/T 2347.3-2014	信息技术行业职业技能 第3部分：软件开发
2	GB/T 32423-2015	系统与软件工程 验证与确认
3	GB/T 32424-2015	系统与软件工程 用户文档的设计者和开发者要求
4	GB 8566-1988	计算机软件开发规范
5	SJ/T 10367-1993	计算机过程控制软件开发规程
6	GB/T 36475-2018	软件产品分类
7	GB/T 36964-2018	软件工程 软件开发成本度量规范
8	GB/T 37691-2019	可编程逻辑器件软件安全性设计指南
9	GB/T 25000.2-2018	系统与软件工程 系统与软件质量要求和评价（SQuaRE）第2部分：计划与管理
10	GB/T 28174.1-2011	统一建模语言（UML）第1部分：基础结构
11	GB/T 11457-1995	软件工程术语
12	GB/T 16260.1-2006	软件工程 产品质量 第1部分：质量模型
13	GB/T 32421-2015	软件工程 软件评审与审核
14	GB/T 32423-2015	系统与软件工程 验证与确认
15	GB/T 30264.2-2013	软件工程 自动化测试能力 第2部分：从业人员能力等级模型
16	GB/T 32904-2016	软件质量量化评价规范
17	GB/T 30998-2014	信息技术 软件安全保障规范

### 3. 选手注意事项

(1) 参赛选手在比赛前应认真阅读赛项规程，严格按照赛项规程参加比赛，避免不必要失误。

(2) 各参赛选手应在竞赛开始前一天按照规定的时间段进入赛场熟悉环境。

(3) 各参赛选手不得统一着装，并不得穿有身份标识的服装。

(4) 参赛选手应按照规定时间抵达赛场，凭身份证、学生证，以及统一发放的参赛证，完成入场检录、抽签确定竞赛工位号，不得迟到早退。并按工位号入座，检查比赛所需竞赛设备齐全后选手签字方可开始参赛。选手在比赛中应注意随时存盘。竞赛期间

不准出场，竞赛结束后方开离场。

(5) 参赛选手不得私自携带任何竞赛软硬件工具（各种便携式电脑、各种移动存储设备等）、设计资源、通信工具进入考场。

(6) 参赛选手要严格遵守竞赛现场规则，如发现有冒名顶替等舞弊行为者，均取消竞赛资格。如遇到电脑或其他比赛用设备故障，可向裁判提出，获得及时解决。

(7) 竞赛过程中，各参赛选手不得与其他人员讨论问题，也不得向裁判、巡视员和其他必须进入考场的工作人员询问竞赛项目的操作流程和操作方法，如有竞赛题目文字不清、软硬件环境故障问题时，可向裁判员询问。

(8) 竞赛过程中除裁判和其他必须进入考场的工作人员外，任何其他非竞赛选手不得进入竞赛场地。

(9) 竞赛结束后，参赛选手要确认成功提交竞赛要求的文件，裁判员与参赛选手一起签字确认，参赛选手在确认后不得再进行任何操作。

(10) 其它未尽事宜，将在赛前向各领队做详细说明。

#### **4. 竞赛直播**

1. 赛点提供全程无盲点录像。
2. 可在赛点指定区域通过网络监控观摩比赛。

## **九、样题（竞赛任务书）**

2022 年度“楚怡杯”湖南省职业院校技能竞赛  
高职高专组电子信息类 Python 程序开发赛项

[时量：240 分钟，试卷号： ]

(样卷)

---

# 竞 赛 任 务 书

场次号：\_\_\_\_\_ 机位号（工位号、顺序号）：\_\_\_\_\_。

2021 年 12 月 日

## 一、 注意事项

1. 请根据大赛所提供的竞赛环境，检查所列的硬件设备、软件清单、材料 清单是否齐全，计算机设备是否能正常使用。

2. 竞赛结束前，将各试题代码整合，部署并运行到竞赛所提供的计算机设 备上，并提交一个完整的项目工程代码到大赛组委会提供的U 盘中，目录格式如 图1 所示。



3. 竞赛结束时，请将U 盘统一提交。

## 二、 竞赛环境

1. PC 机: 系统已安装 Python 相关环境、MySQL 数据库，用户名密码分别为: root/123456。

2. 从U 盘 “/01 程序开发/” 路径下 “01 工程代码” 文件夹中获取程序开发项 目工程代码，包括前端和后端两部分工程。U 盘中的工程代码可以直接本地导入、 编译运行和发布。

3. 根据 API 接口规范，利用 Python 语言编写 API 接口程序，编写完成后，运行前 端工程，检验程序编写是否正确。

4. Nginx 服务

服务器地址: `http://{本机 IP 或 127.0.0.1}:8088/Nginx` 位于服务器:

D:\\Nginx。

5. Tomcat 服务

服务器地址: `http://{本机 IP 或 127.0.0.1}:8081/Tomcat` 位于服务器:

D:\\Tomcat。

## 赛题说明

本套赛题包括五个部分: 第一部分为程序开发，第二部分数据获取题，第三部分为数据清洗题， 第四部分为数据处理题，第五部分为数据可视化题。请考生根据题目中所描述的需求，并参阅 U 盘中的辅助文档，自行设计，编码实现。

## 三、 软件组件

Python 编程语言及相关开发环境 (Python、Pycharm、anaconda)

Web 框架 (Django)

Python 数据分析组件 (numpy、Pandas)

Python 可视化组件 (matplotlib)

中文分词Python 库 ( jieba )

## 四、 赛题

## 第一部分：程序开发（30分）

### 第1题：实现员工登录功能（10分）

#### 【任务要求】

本任务需要实现员工登录功能，如果登录成功，进入首页，否则，提示员工登录失败。

#### 【任务描述】

1. 登录页面如图1-1所示，帐号信息（用户名：admin，密码：123456）。
2. 观察登录页面布局，填补页面代码和后端程序代码，实现当管理员填写完登录信息以后，进入登录后首页，如图1-2所示。

【工程代码】从U盘 “/01 程序开发/” 路径下 “01 工程代码” 文件夹中获取程序开发项目工程代码。

#### 【API接口】

接口名称	API
修改员工信息 API	http://127.0.0.1:8000/login

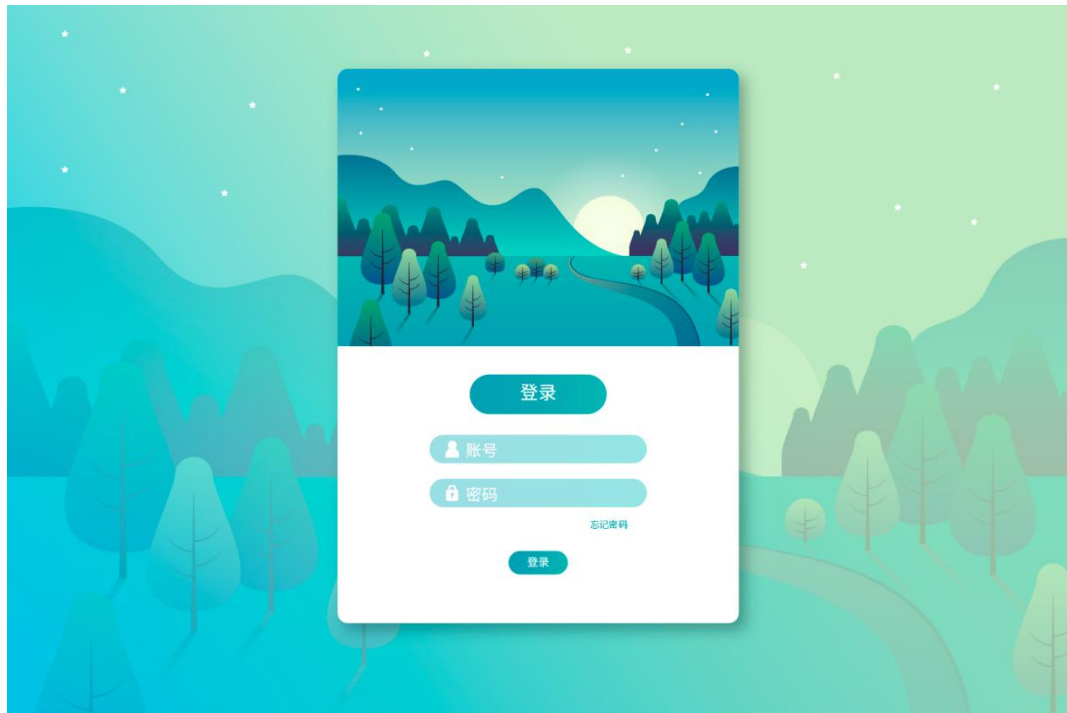


图1-1 首页





图1-2 管理员登录页面

## 第2题：实现增加员工信息功能（10分）

### 【任务要求】

本任务需要实现如下功能，管理员登录之后，在首页选择“增加员工信息”选项以后，进入增加员工信息页面，管理员在页面添加信息后，提交给后端程序进行保存。

### 【任务描述】

1. 员工信息管理网站（用户名：admin，密码：123456），登录页面如图1-1所示，进入首页。
2. 观察首页布局，如图1-2所示，填补页面代码，实现当管理员选择“增加员工信息”选项时，进入“增加员工信息页面”（add.html）。
3. 观察增加员工信息页面，如图1-3所示，填补add.html页面代码和后端程序代码，保证只有root用户（root用户的学号为1）才能进行增加员工，在增加员工信息以后，点击“保存”按钮以后，把员工信息保存到数据库。

## 请在下列信息栏中填入员工的基本信息

员工号码:	<input type="text"/>
员工姓名:	<input type="text"/>
员工电话:	<input type="text"/>
员工地址:	<input type="text"/>
员工院系:	<input type="text"/>
员工专业:	<input type="text"/>
<input type="button" value="取消"/> <input type="button" value="确定"/>	

图1-3增加员工信息页面

【工程代码】从U盘 “/01 程序开发/” 路径下 “01 工程代码” 文件夹中获取程序开发项目工程代码。

【API接口】

接口名称	API
添加员工 API	http://127.0.0.1:8000/add

### 第3题：实现修改员工信息功能（10分）

【任务要求】

本任务需要实现如下功能，管理员登录之后，在首页选择“修改员工信息”选项以后，进入修改员工信息页面，展示员工已有信息，管理员在页面修改信息后，提交给后端程序进行保存。

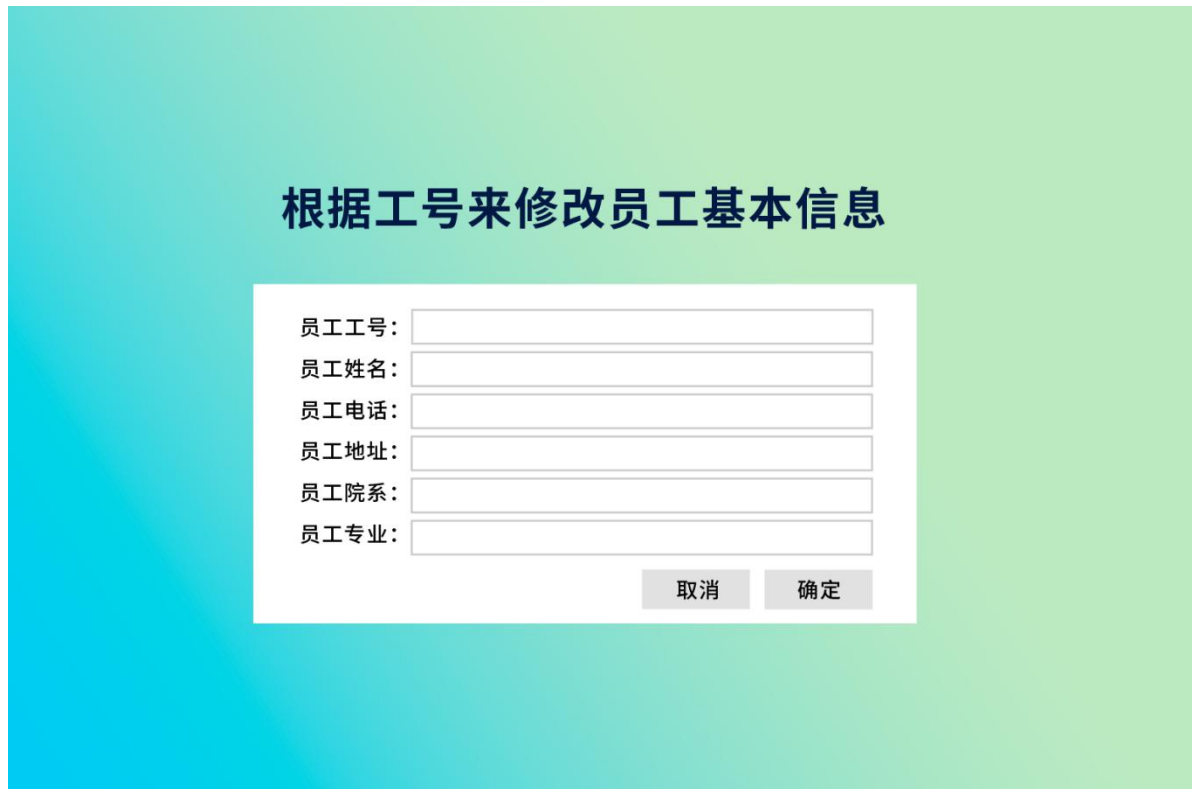
【任务描述】

3. 登录员工信息管理网站（用户名：admin，密码：123456），登录页面如图1-1所示，进入首页。
4. 观察首页布局，如图1-2所示，填补页面代码，实现当管理员选择“修改员工信息”选项时，进入“修改员工信息页面”（update.html）。
5. 观察修改员工信息页面，如图1-4所示，填补update.html页面代码和后端程序代码，在管理员修改完信息以后，点击“保存”按钮以后，把员工信息保存到数据库。

【工程代码】从U盘 “/01 程序开发/” 路径下 “01 工程代码” 文件夹中获取程序开发项目工程代码。

【API接口】

接口名称	API
------	-----



**根据工号来修改员工基本信息**

员工工号:

员工姓名:

员工电话:

员工地址:

员工院系:

员工专业:

图1-4 修改员工信息页面

## 第二部分：爬虫应用（20分）

### 第4题：利用爬虫抓取热门书籍数据（10分）

#### 【任务要求】

基于Python爬虫代码从指定读书网站抓取热门书籍数据，爬取内容包括：书籍名称，书籍评分，书籍图片。

#### 【任务描述】

6. 进入书籍网站，首页如图2-1所示，观察网页布局，使用浏览器阅读网页源码，分析所要爬取数据项的页面位置及关系，为编写数据爬取代码做准备。

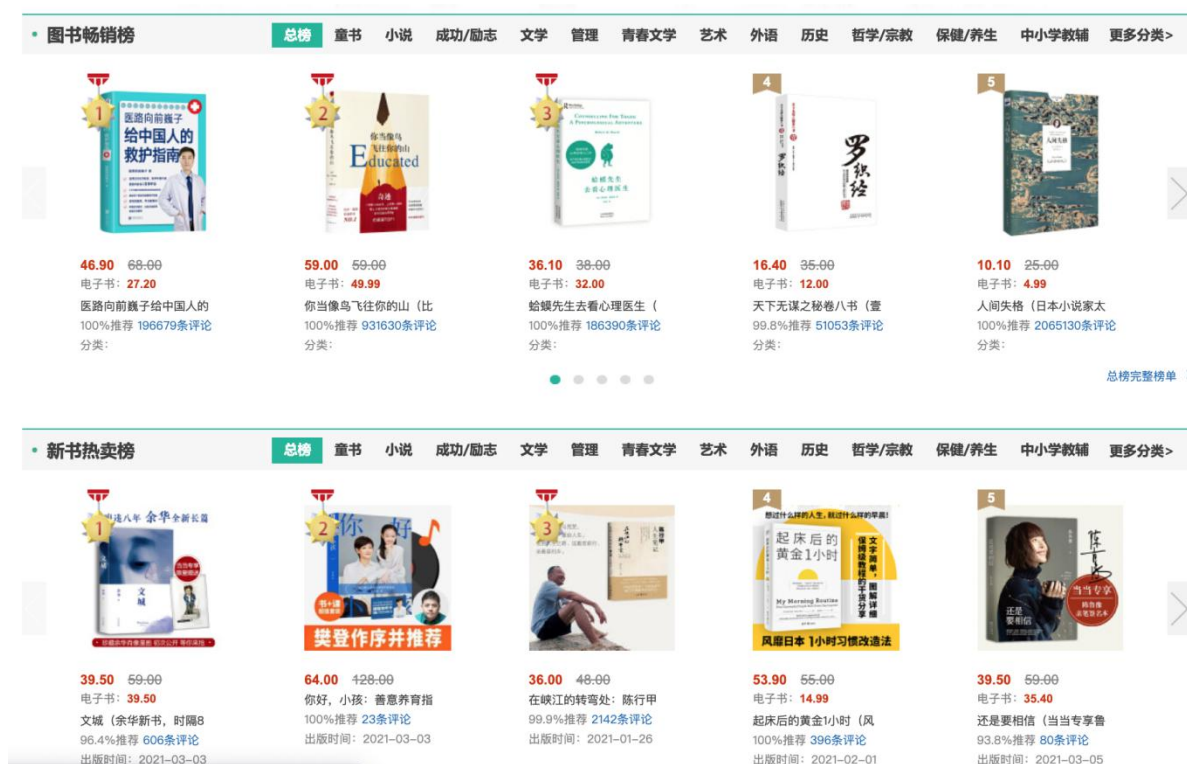


图2-1 热门书籍页面

- 从U盘 “/01 数据获取/” 路径下 “01 工程代码” 文件夹中获取程序开发项目工程代码。
- 按照代码中的任务点要求, 补充Python爬虫代码, 实现爬取首页“书籍名称”、“书籍图片”、“书籍评分”, 并分别保存至本地磁盘 “D:\\home\\书籍名称\\书籍mingci.txt”、“D:\\home\\书籍图片\\书籍图片.txt”、“D:\\home\\书籍评分\\书籍评分.txt” 文件中。

### 第5题: 利用爬虫抓取书籍评论数据 (10分)

#### 【任务要求】

基于Python爬虫代码从指定书籍网站抓取书籍详情数据, 爬取内容包括: 评论作者、评论内容、评论评分、评论标题、评论日期。

#### 【任务描述】

- 进入书籍网站首页, 点击任一书籍, 进入书籍评论页面, 如图2-2所示, 观察网页布局, 使用浏览器阅读网页源码, 分析所要爬取数据项的页面位置及关系, 为编写数据爬取代码做准备。

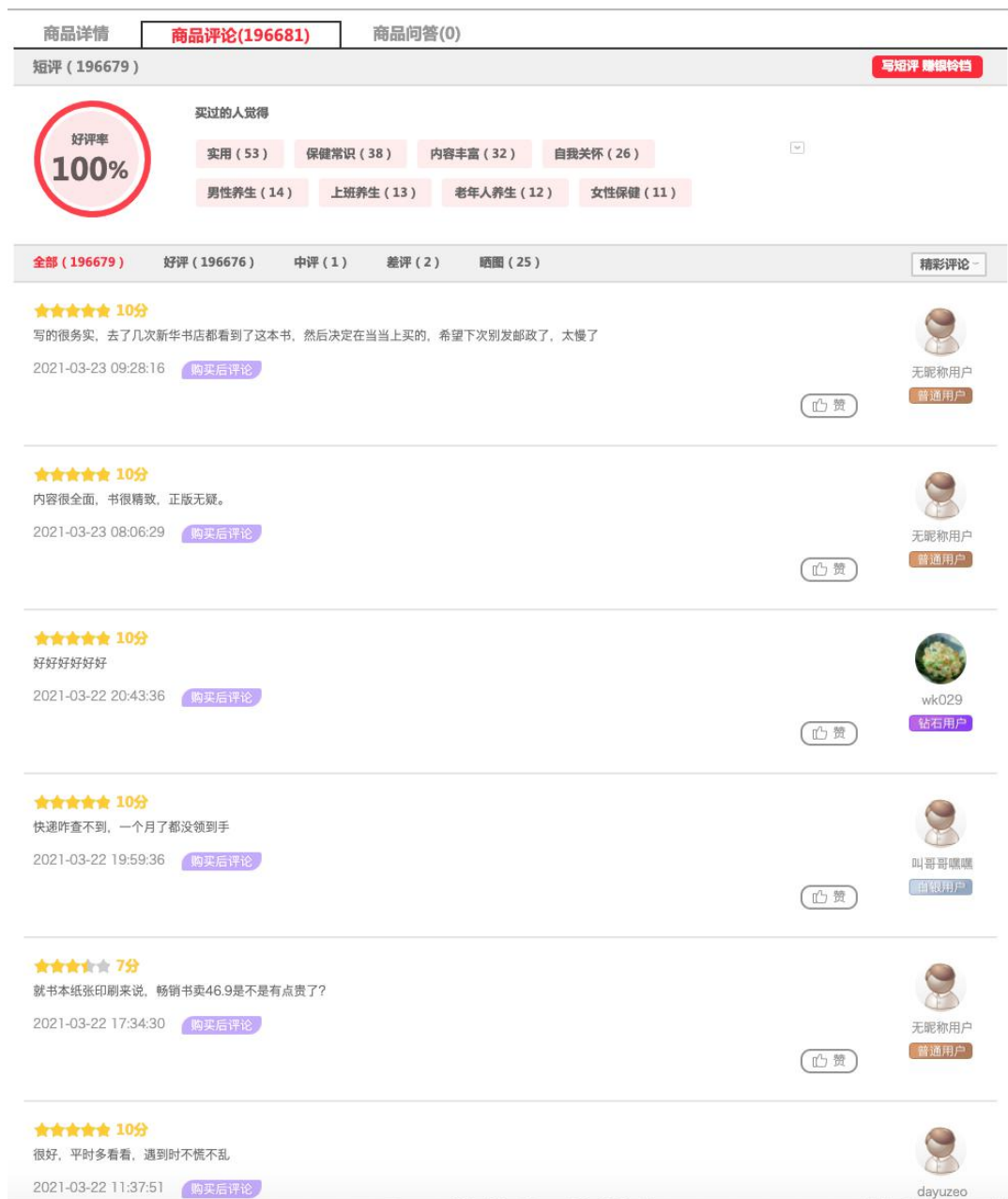


图2-2 电影评论页面

2. 从U盘 “/01 数据获取/” 路径下 “02 工程代码” 文件夹中获取程序开发项目工程代码。
3. 按照代码中的任务点要求, 补充Python爬虫代码, 实现爬取每个书籍的“评论作者”、“评论内容”、“评论评分”、“评论标题”、“评论日期”, 并保存至本地磁盘 “D:\\home\\书籍评论\\书籍评论.csv” 文件中。

### 第三部分：数据清洗（15分）

#### 第6题：利用Pandas对书籍数据进行清洗

##### 【任务要求】

利用Pandas函数完成对书籍数据进行清洗, 包括数据缺失值处理、重复数据处理、异常值处理、数据类型转换。

##### 【任务描述】

1. 从U盘 “/02 数据清洗/” 路径下 “03 工程代码” 文件夹中获取程序开发项目工程代码。
2. 按照代码中的任务点要求，补充Python数据清洗代码，实现如下数据清洗任务：
  - i. 检查缺失数据项
  - ii. 对缺失 “评论评分” 数据项进行均值插补
  - iii. 对缺失 “评论作者” 数据项的那条记录进行删除
  - iv. 对冗余数据记录进行删除
  - v. 把清洗后的数据保存至本地磁盘 “D:\\home\\清洗数据结果\\清洗数据结果.txt” 文件中。

#### 第四部分：数据分析（20分）

##### 第7题：编写Python程序对书籍评论数据进行处理。（10分）

###### 【任务要求】

编写Python程序，去除书籍评论中所有的标点符号。

###### 【任务描述】

1. 从U盘 “/03 数据处理/” 路径下 “04 工程代码” 文件夹中获取程序开发项目工程代码。
2. 按照代码中的任务点要求，实现下列任务：
  - i. 补充Python数据清洗代码，去除给定数据记录中电影评论数据项中的标点符号 “，” 和 “、”。
  - ii. 把处理后的数据保存至本地磁盘 “D:\\home\\数据处理结果” 文件夹中。

##### 第8题：编写Python程序，查找目录下的指定文件格式。（10分）

###### 【任务要求】

编写一个Python程序，用户输入开始搜索的路径，查找该路径下（包含子文件夹内）所有的指定格式的文件，并创建一个文件用来存放所有找到的文件的路径。

###### 【任务描述】

1. 从U盘 “/03 数据处理/” 路径下 “05 工程代码” 文件夹中获取程序开发项目工程代码。
2. 按照代码中的任务点要求，实现下列任务：
  - i. 补充Python代码，用户输入开始搜索的路径，查找该路径下（包含子文件夹内）所有的word、pdf、txt格式文件，并创建一个文件（fileList.txt）存放所有找到的文件的路径。
  - ii. 运行代码，输入查找目录 “/03 数据处理/”
  - iii. 把运行结果文件保存至本地磁盘 “D:\\home\\数据处理结果\\fileList.txt” 文件中。

#### 第五部分：数据可视化（15分）

**第9题：编写Python程序对书籍评论数据进行统计分析并进行可视化展示。（10分）**

**【任务要求】**

编写Python程序，使用numpy和Pandas对电影短评数据进行数据统计并使用matplotlib进行可视化展示，然后创建一个文件用来存放统计和可视化结果。

**【任务描述】**

1. 从U盘 “/04 数据分析与可视化/” 路径下 “06 工程代码” 文件夹中获取程序开发项目工程代码。
2. 按照代码中的任务点要求，实现下列任务：
  - i. 补充Python代码，完成对评分进行频次统计并绘制折线图。
  - ii. 补充Python代码，对评论内容生成词云图。
  - iii. 运行代码，把统计分析的结果保存至本地磁盘 “D:\\home\\数据分析与可视化结果” 文件夹中。

**第10题：编写Python程序，对书籍评论数据进行数据分析并进行可视化展示。（5分）**

**【任务要求】**

编写Python程序，使用numpy和Pandas进行电影短评数据进行分析并进行可视化展示，并创建一个文件用来存放数据分析和可视化结果。

**【任务描述】**

1. 从U盘 “/04 数据分析与可视化/” 路径下 “07 工程代码” 文件夹中获取程序开发项目工程代码。
2. 按照代码中的任务点要求，实现下列任务：
  - i. 补充Python代码，分析书籍上架后每天的评论数据走势，按天统计评论的个数，并且按照天进行排序，然后绘制折线图。
  - ii. 补充Python代码，分析每天评分的比例，并绘制柱状图。
  - iii. 运行代码，并把运行结果文件保存至本地磁盘 “D:\\home\\数据分析与可视化结果” 文件夹中。