**《Python程序设计》**

课程设计报告

题目： 在线课程评价与推荐系统

**专 业： 人工智能 班 级： 2301**

**学 号： 2302190128**

**姓 名： 丁凯峰**

**小组成员： 朱正勇、马景星**

**指导教师： 王恒**

**成 绩 ：**

**数学与计算机学院**

**2024 年 6 月**

目录

[一、课程设计目的 1](#_Toc170684012)

[二、课程设计要求 1](#_Toc170684013)

[三、项目设计内容 2](#_Toc170684014)

[四、系统实现 4](#_Toc170684015)

[五、总结 9](#_Toc170684016)

[项目优点: 9](#_Toc170684017)

[项目缺点: 9](#_Toc170684018)

## 一、课程设计目的

掌握所学Python语言程序设计的方法，熟悉所学语言的开发环境及调试过程，熟悉所学语言中的数据类型， 程序控制、数据可视化，巩固和加深对理论课中知识的理解，提高学生对所学知识的综合运用能力。 通过综合设计要求达到下列基本技能：

1.培养查阅参考资料、手册的自学能力，通过视频学习课堂上没有涉及的知识， 通过独立思考深入钻研问题，学会自己分析、解决问题。

2.通过对所选题目方案分析比较，确立方案，编制与调试程序，初步掌握程序设计的方法，能熟练调试程序。

3.编程简练，可用，功能全面，并有一定的容错能力。用户界面良好，有较好的输出功能。 在完成课题基本要求后， 具有创新型设计，具有一定的实用价值。

4.根据个人的设计调试过程，撰写设计报告，参与项目答辩。

## 二、课程设计要求

1.查阅相关资料或实际调研，对课题进行需求分析、系统设计(例如：前台界面设计、中间功能模块设计，可视化展示设计)；建议使用Python3.7以上版本，Pycharm集成开发环境完成课题，做到程序运行实现所给功能、界面友好、代码可维护性好。

2. 课程设计结束时，提交完成的所有源程序、相关文件和可执行文件，认真撰写并完成《课程设计报告》，参加现场答辩，填写答辩表。

3.源程序及课程设计报告册不允许完全照抄网上内容，必须结合课程设计指导书中的要求对你查阅的资料消化、吸收、改进与创新。

4. 如果不在机房集中进行课程设计时，班长、学习委员、寝室长严格督促各位同学在图书馆完成资料的查阅及课程设计报告的撰写。

5. 不符合要求的程序、设计报告、抄袭的设计报告或源程序代码、在设计中完全未参与的将作不及格处理。

6.学生原则上以每组3人为单位，能力比较强的同学可以一人为单位，根据设计内容和要求、独立完成一个完整的可独立运行的软件系统。。

## 三、项目设计内容

1、课题名称

《在线课程评价与推荐系统》

2、课题概况（分析一下课题的背景。相关技术等）

随着网络技术的迅速发展，网上的学习资源愈发丰富，在提供优质学习资源

同时，也为人们快速寻找需要的资源制造了困难。幸运的是，编程语言与网络技术同步发展，网络爬虫技术以然高度成熟。通过网络爬虫，人们可以轻松获取到想要的信息，并通过合适的编程语言，使用合适的算法归纳整理这些信息，使其变得有序、便于查找。

3、开发环境（根据自己情况修改）：

（1）操作系统：window10及以上版本

（2）Python版本：Python3.12.4

（3）开发工具： PyCharm

（4）Python模块：urllib, pandas, matplotlibbs4, openpyxl, tkinter, re,

matplotlib

4、功能要求（根据自己情况修改）

（1）爬取在线课程平台的课程信息和学生评价

（2）分析课程受欢迎程度

（3）基于用户评价推荐相关课程

（4）提供课程搜索和筛选功能

5、系统分析（各部分的需求分析）

（1）爬取在线课程平台的课程信息和学生评价

需要获取网站信息，使用合适的方法分析、检索需要的信息并将其存放至excel表格中

（2）分析课程受欢迎程度

读取（1）获取的数据，对其进行分析归纳，计算并对比不同类型的课程的占比

（3）基于用户评价推荐相关课程

利用用户报名数量及好评率等指标，推荐优质课程

（4）提供课程搜索和筛选功能

设计一个模块，为用户筛选出需要的课程

6、系统设计（各部分的设计思路）

（1）爬取在线课程平台的课程信息和学生评价

a.使用urllib的request模块爬取网页源码

b.使用bs4的Beautifulsoup模块分析网页源码

c.使用re模块创建正则匹配对象，在网页源码中查找想要的信息

d.使用openpyxl的Workbook模块将信息保存到excel表格中

（2）分析课程受欢迎程度

a.使用pandas库读取数据并将其转化为列表对象

b.使用列表推导式提取对应分类的数据

c.将不同类型的数据按照数量多少排序

d.使用matplotlib画出饼状图，直观表现出不同类型的课程的受欢迎程度

（3）基于用户评价推荐相关课程

获取课程的报名人数和报名人数，推荐优质课程

（4）提供课程搜索和筛选功能

获取用户提供的关键字，查找课程，将相应课程反馈给用户

1. 小组分工

小组成员：丁凯峰、朱正勇、马景星

具体分工：丁凯峰：网络爬虫及用户界面实现（爬取在线课程平台的课程信

息和学生评价、基于用户评价推荐相关课程、

提供课程搜索和筛选功能）

朱正勇：分析课程受欢迎程度（数据可视化）

马景星：ppt制作及答辩、辅助完成部分代码

## 四、系统实现

import re  
import tkinter as tk  
import urllib.error  
import urllib.request  
  
from bs4 import BeautifulSoup  
from openpyxl import Workbook  
  
import datavisualization # 将朱正勇的可视化代码修改为一个模块导入主程序，在主程序中输出可视化结果  
  
# 创建正则规则对象，用于匹配查找想要的数据  
findLink = re.compile(r'href="(.\*?)"') # 课程视频链接  
findImg = re.compile(r'src="(.\*?)"') # 课程封面图片链接  
findTitle = re.compile(r'title="(.\*?)"') # 课程名称  
findRate = re.compile(r'好评率(\d\*.?)') # 课程好评率  
findNum = re.compile(r'(\d\*.?)人报名') # 报名人数  
  
  
def askURL(url: str):  
 """@获取网页源码"""  
 header = {  
 "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/125.0.0.0 Safari/537.36"  
 } # 伪装自己为浏览器，绕开反爬虫程序  
 req = urllib.request.Request(url, headers=header)  
 html = ""

# 检查是否爬取失败， 并反馈失败原因  
 try:  
 response = urllib.request.urlopen(req)  
 html = response.read().decode('utf-8')  
 except urllib.error.URLError as e:  
 if hasattr(e, "code"):  
 print(e.code)  
 if hasattr(e, "reason"):  
 print(e.reason)  
 else:  
 print(html)  
  
  
def getDate(h: str) -> list:  
 """@获取课程数据"""  
 data = [] # 定义空列表用于存放数据  
 soup = BeautifulSoup(h, "html.parser")  
 for item in soup.find\_all("section", class\_="course-card-expo-wrapper"):  
 # print(item.prettify())  
 item = str(item)  
 link = "https://ke.qq.com/" + re.findall(findLink, item)[0]  
 img = re.findall(findImg, item)[0]  
 title = re.findall(findTitle, item)[0]  
 rate = re.findall(findRate, item)  
 if len(rate): # 部分课程未进行学员打分统计  
 rate = rate[0]  
 else:  
 rate = '0%'  
 num = re.findall(findNum, item)[0]  
 if num[-1] == '万':  
 num = int(num[:-1]) \* 10000  
 else:  
 num = int(num)  
 # print(num)  
 data.append([title, link, img, rate, num])  
 return data  
  
  
def savaData(data: list, savepath: str):  
 """@将数据保存到excel表格中"""  
 wb = Workbook() # 创建workbook工作对象  
 sheet = wb.active # 创建工作表  
 col = ('课程名称', '课程链接', '课程封面链接', '课程好评率', '课程报名人数')  
 for i in range(0, 5):  
 sheet.cell(1, i + 1, col[i])  
 for i in range(0, len(data)):  
 # print(f'第{i+1}条')  
 d = data[i]  
 for j in range(0, 5):  
 sheet.cell(i + 2, j + 1, d[j])  
  
 wb.save(savepath)  
  
  
def course\_recommend(data: list, root: tk.Tk):  
 """@ 课程推荐，在新窗口展示"""  
 data.sort(key=lambda x: x[-1], reverse=True)  
 d = data[:50]  
 d.sort(key=lambda x: int(x[-2][0:-1]), reverse=True)  
 d1 = d[:10]  
 # print(\*d1, sep='\n')  
 new\_face = tk.Toplevel(root) # 创建子窗口  
 new\_face.geometry("700x850")  
 new\_face.title("课程推荐")  
 txt = tk.Text(new\_face)  
 txt.place(rely=0, relheight=1.0, relwidth=1.0)  
 i = 1

# 在子窗口输出推荐的课程及其相关信息  
 for line in d1:  
 course\_name = line[0]  
 course\_link = line[1]  
 course\_num = line[-1]  
 s = ("第%d个课程\n课 程 名 称 :\t%s\n课 程 链 接 :\t%s\n课程报名人数:\t%d\n\n\n" %  
 (i, course\_name, course\_link, course\_num))  
 txt.insert(tk.END, s)  
 i += 1  
 txt.insert(tk.END, "基于报名人数和好评率综合推荐")  
 close\_but = tk.Button(new\_face, text="关闭", command=new\_face.destroy, height=2, width=12)  
 close\_but.place(relx=0.85, rely=0.01)  
  
  
def course\_seek(root: tk.Toplevel, inp: tk.Entry):  
 """@ 课程查找"""  
 key\_word = inp.get() or "未输入"  
 txt = tk.Text(root)  
 txt.place(relx=0.1, rely=0.2)  
 if key\_word == "未输入":  
 txt.insert(tk.END, "请输入关键字以查询课程")  
 else:  
 flag = 1  
 for d in da:  
 if key\_word in d[0]:  
 course\_name = d[0]  
 course\_link = d[1]  
 course\_num = d[-1]  
 s = ("课 程 名 称 :\t%s\n课 程 链 接 :\t%s\n课程报名人数:\t%d\n\n\n" %  
 (course\_name, course\_link, course\_num))  
 txt.insert(tk.END, s)  
 flag = 0  
 if flag:  
 txt.insert(tk.END, "抱歉! 未查到您要找的课程, 请重新输入关键字查询!")  
  
  
def seek\_face(root: tk.Tk):  
 """@ 课程搜索界面"""  
 new\_face = tk.Toplevel(root)  
 new\_face.geometry("700x800")  
 new\_face.title("课程搜索")  
 label = tk.Label(new\_face, text="课程搜索", font=("华文新魏", 30))  
 label.place(relx=0.35, rely=0.01)  
 label1 = tk.Label(new\_face, text="请输入关键词搜索课程", font=("SimHei", 12))  
 label1.place(relx=0.25, rely=0.12)

# 在子窗口创建搜索框  
 input1 = tk.Entry(new\_face, textvariable=tk.StringVar())  
 input1.place(relx=0.25, rely=0.15)  
 seek\_but = tk.Button(new\_face, text="搜索课程", height=1, width=10, command=lambda: course\_seek(new\_face, input1))  
 seek\_but.place(relx=0.5, rely=0.14)  
 close\_but = tk.Button(new\_face, text="关闭", command=new\_face.destroy, height=2, width=12)  
 close\_but.place(relx=0.85, rely=0.9)  
  
  
def interface(text: str):  
 """@用户主界面"""  
 root = tk.Tk() # 创建界面对象  
 root.geometry("700x400") # 设置界面大小  
 label = tk.Label(root, text=text, height=1, width=30, font=("华文新魏", 30))  
 label.place(relx=0.2, height=80, width=430)  
 txt = tk.Text(root)  
 txt.place(rely=0.95, relx=0.25, relheight=0.1, relwidth=0.5)  
 txt.insert(tk.END, "本程序获取课程来自腾讯课堂(https://ke.qq.com/)")  
 bt1 = tk.Button(root, text="推荐课程", height=2, width=15, command=lambda: course\_recommend(da, root))  
 bt1.place(relx=0.33, rely=0.2)

# 创建功能按键  
 visual\_but = tk.Button(root, text="数据饼图", height=2, width=15, command=datavisualization.show)  
 visual\_but.place(relx=0.53, rely=0.2)  
 seek\_but = tk.Button(root, text="搜索课程", height=2, width=15, command=lambda: seek\_face(root))  
 seek\_but.place(relx=0.33, rely=0.35)  
 quit\_but = tk.Button(root, text="退出程序", height=2, width=15, command=root.destroy)  
 quit\_but.place(relx=0.53, rely=0.35)

# 将页面保持住  
 root.mainloop()  
  
  
if \_\_name\_\_ == '\_\_main\_\_':

""" 因为js异步渲染的关系, 无法爬取到完整源码, 网上的解决方案都不成功  
 所以没有使用askURL函数来获取， 只是直接复制网页源码存放到txkt.html文件

中"""  
 # Url = "https://ke.qq.com/"  
 # html = askURL(Url)  
 # print(html)  
 with open('txkt.html', 'r', encoding='utf-8') as f:  
 html = f.read()  
 da = getDate(html)  
 # print(\*da, sep='\n')  
 savaData(da, "腾讯课堂课程数据.xlsx")  
 da.sort(key=lambda x: x[-1], reverse=True)  
 print('爬取成功!\n启动用户界面', end='')  
 interface(text="在线课程评价与推荐系统")

## 五、总结

在项目中，结合在网上刷题的经验，我为函数形参添加了类型标注，降低代码灵活性，提高代码的稳定性和可读性。在实现项目的过程中，我逐步巩固所学知识，并为实现新功能学习新的知识，如爬虫的实现、thkinter库绘制GUI以实现用户界面；建立起初步的项目开发思维。同时，我还认识到了当前所学知识以的不足，不能更完美高效的解决在开发中出现的问题，为此花费了大量时间。纸上得来终觉浅，觉知此事要躬行。我想，这就是本次项目给我带来的最大的收获，课本上的知识仅仅是冰山一角，Python这门编程语言还有很多地方等着我们去发掘、去探索、取学习、去运用。

### 项目优点:

项目分离文件少，整体性高

### 项目缺点:

项目文件少，过于集中，牵一发而动全身，不方便修改调试；

爬取网页部分未使用动态爬取，而是复制网页源码存放到文件中，不利于复用，并导致数据不能随网页更新更新