**2023年河北省职业院校技能大赛**

**“Python编程”赛项规程**

**一、赛项名称**

赛项编号：

赛项名称：2023年河北省职业院校学生技能大赛

“Python编程”赛项

赛项组别：高职组

竞赛形式：团体赛

赛项专业大类：电子信息大类

**二、竞赛目的**

（一）以大赛检验教育教学成果

本赛项以Python语言为考察语言，从语言基础与算法设计、网页与文本分析和数据处理与分析三方面出发设置考察任务，并以参赛选手任务的完成度、完成任务的质量以及参赛选手自身的职业素养作为竞赛指标，综合考察参赛选手的Python编程语言程序设计能力与运用Python编程语言解决实际问题的能力。

参加竞赛的每位选手需要具备相关的从业知识和技能、相应的程序设计能力和基本的计算机程序设计员职业素养，可以根据程序设计需求完成相对应的程序设计任务。能够熟练使用Requests、BeautifulSoup、Pandas、Numpy、Jieba、Pyecharts、Matplotlib等常见的Python包（库）来解决学习工作中的碰到的实际问题。

（二）以大赛促进教育教学改革

一方面，通过赛项可以考察选手的Python语言程序编写与设计能力，从而提高在校学生技能学习的积极性，激发在岗教职工崇尚技能的热情；另一方面，通过赛项可以促进产教融合、校企合作与产业发展，从而推动提升电子信息类专业人才整体水平。

（三）以大赛营造崇尚技能氛围

通过本赛项，在信息产业大力弘扬工匠精神，引导全社会尊重、重视、关心技能人才的培养和成长，宣传技能人才的重要贡献和重大作用，在全社会倡导“崇实尚业”之风，营造尊敬技能人才的社会氛围，让尊重劳动、尊重技术、尊重创造成为社会共识，激励广大青年走技能成才、技能报国之路。

**三、竞赛内容**

竞赛包括知识理论和上机实操两部分，均分为Python语言基础与算法设计、Python网页与文本分析、Python数据处理与分析三项竞赛内容。

具体如下：

|  |  |  |
| --- | --- | --- |
| 竞赛内容 | 相关知识 | 技能要求 |
| Python语言基础与算法设计 | Python基础语法和常见的算法与数据结构等 | 掌握Python基础语法和常见数据结构；  了解常见的计算机算法；  代码编写规范。 |
| Python网页与文本分析 | HTTP协议、浏览器原理、网页基础、爬虫基本原理、文件切分清洗与分类等 | 熟悉Python网页分析常用的Requests、bs4等包（库）；  掌握基于网页规则的信息抽取；  掌握Jieba、SnowNLP等文本分析库；  了解基本的自然语言语言解析规则与相关应用场景精通正则表达式。 |
| Python数据处理与分析 | 数据获取与清洗、数据分析、基础统计分析等 | 熟悉Numpy、Pandas、Pyecharts、Matplotlib等数据处理与分析相关的包（库）；  掌握基础的数据处理与分析方法；  了解数据分析处理实际常用应用算法； |

（1）Python语言基础与算法设计

本部分重点考察参赛选手对于Python基础语法和常见计算机算法的掌握情况与实际使用能力。赛项要求参赛选手能够根据任务要求，输出正确的计算结果。

任务考察点分为两方面，一方面考察参赛选手对于Python基础语法中的掌握，例如逻辑判断、循环、遍历、字符串、列表、字典、元组、集合、函数、类等相关知识点；另一方面考察参赛选手的算法设计能力，例如常见的初级排序算法、查找算法等相关知识点。

（2）Python网页与文本分析

本部分考察点分为两方面，一方面需要选手了解常用网页结构，熟练使用Python相关工具获取内置专题页面的指定数据；另一方面需要选手了解Python常用文本处理与分析的模块（库），能够根据任务要求完成指定操作并返回任务中所需要的数据。

（3）Python数据处理与分析

参赛选手可根据自己的专业方向与兴趣特长，合理选择考试任务方向合理分配考试时间，尽可能多的完成相关任务，得到任务期望输出的指定结果，提交准确高效的赛题解答代码。

本部分需要使用的Python相关数据处理与分析的第三方包（库），如Numpy和Pandas等，其均已内置于竞赛系统，无需参赛选手手动安装，参赛选手仅需要比赛时引入相关包（库）。

竞赛总时间为：300 分钟。

每支参赛队 2 名参赛选手，按照团队总成绩进行最终成绩排名。

**四、竞赛方式**

1.本赛项为团体赛，参赛组别为高职组。

2.竞赛由各校自行组队参赛。每所学校不超过 2 支参赛队，每支参赛队参赛选手 2 名，每支参赛队指导教师不超过 2 名。

**五、竞赛流程**

竞赛日程安排：

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | | **事项** | |
| 6.4 | 20:00前 | 裁判、仲裁、监督报到 | 住宿酒店 |
| 6.5 | 09:00-14:00 | 参赛队报到，安排住宿，领取资料 | 住宿酒店 |
| 09:00-12:00 | 裁判培训会议 | 会议室 |
| 14:30-15:00 | 开幕式 | 会议室 |
| 15:00-15:30 | 领队会 | 会议室 |
| 15:30-16:00 | 熟悉赛场 | 竞赛场地 |
| 16:30 | 检查封闭赛场 | 竞赛场地 |
| 17:00 | 参赛领队返回酒店 | 竞赛场地 |
| 6.6 | 7:30 | 参赛队前往学院 | 住宿酒店 |
| 08:00 | 参赛队到达竞赛场地前集合 | 竞赛场地前 |
| 08:10-08:30 | 检录 | 竞赛场地前 |
| 08:30-08:50 | 第一次抽签加密 | 一次抽签区域 |
| 08:50-09:10 | 第二次抽签加密 | 二次抽签区域 |
| 09:10-09:30 | 设备工具检查确认、题目发放 | 竞赛场地 |
| 09:30-12:00 | 竞赛 | 竞赛场地 |
| 12:00-13:00 | 午餐 | 竞赛场地 |
| 13:00-15:30 | 竞赛 | 竞赛场地 |
| 15:30-16:30 | 核分、复验、申诉 | 竞赛场地 |
| 17:00 | 闭幕式 | 会议室 |

**六、竞赛试题**

竞赛样卷随竞赛规程同时公布，见附件：2023 年河北省职业院校技能大赛“Python编程”赛项样题。

**七、竞赛规则**

（一）竞赛报名

1. 各高职院校按照大赛组委会规定的报名要求，通过“报名系统”报名参赛。

2. 高职学生组参赛对象为全日制普通高等职业院校在校生和五年制高职四至五年级在校生。

3.参赛选手和指导教师报名，获得确认后不得随意更换。比赛前参赛选手和指导教师因故无法参赛，须由学校在赛项开赛前 10 个工作日出具书面说明，并按参赛选手资格补充人员并接受审核，经省大赛组委会办公室同意后予以更换。

（二）熟悉场地规则

1.为各参赛选手提供统一有序的熟悉场地时间安排，熟悉场地时限定在指定区域与路线，不允许进入比赛区。

2.熟悉场地时严禁与现场工作人员进行交流，不发表没有根据以及有损大赛整体形象的言论。

3.熟悉场地时严格遵守大赛各种制度，严禁拥挤，喧哗，以免发生意外事故。

（三）入场规则

1.参赛选手按赛项规定的时间准时到达赛场检录区集合。

2.裁判将对各参赛选手的身份进行核对。参赛选手须提供参赛证、身份证、经学校注册的学生证，证件上的姓名、年龄、相貌特征等应与参赛信息一致。

3.裁判检验参赛选手携带的随身物品，不允许携带任何具有通讯及存储功能的设备、纸质材料等物品，检查合格后方可进入赛场抽签区。

4.第一次加密时，参赛选手按已获得的抽签顺序号依次抽取参赛编号；第二次加密时，参赛选手凭已获得的参赛编号抽取比赛工位号；完成两次加密后，参赛选手在现场裁判的指挥下有序进入赛场，按已获得的比赛工位号就位。

（四）正式比赛规则

1.选手进入赛场后，必须听从现场裁判的统一安排和指挥。

2.参赛选手进入比赛工位号后，可查看比赛环境，如有问题可立即向现场裁判报告，但不可进行比赛任务的操作。

3.在现场裁判宣布比赛开始后，参赛选手才能开始动手完成比赛任务的操作。

4.比赛过程中，参赛选手必须严格遵守安全操作规程，确保人身和设备安全，并接受现场裁判和技术人员的监督和警示。

5.若认为比赛设备有问题需修复或更换时，应在赛场记录表的相应栏目填写修复问题及原因、更换设备原因、更换时间等并签比赛工位号确认后，由现场裁判和技术人员予以修复或更换。修复或更换工作完成后，现场裁判和技术人员须检验并将结果记录在赛场记录表的相应栏目中并由参赛选手签工位号确认。

6.在需要通电检查或调试设备时，应先报告现场裁判或技术人员，通电前的安全检测合格，获允许并派人监护后，才能通电检查或调试。

7.经现场裁判和技术人员检验，比赛设备确因设备、元器件故障或损坏而进行修复或更换时，从参赛选手报告现场裁判到完成修复或更换之间的用时，为比赛补时时间。

8.在比赛过程中，参赛选手不得随意离开比赛工位，不得与其他参赛选手和人员交流，不得无故提前结束比赛。参赛选手因故需提前终止比赛任务，应报告现场裁判，离场前在赛场记录表的相应栏目填写离场时间、离场原因并由现场裁判签名和选手签工位号确认。

9.在比赛过程中，严重违反赛场纪律影响他人比赛者，违反操作规程不听劝告者，越界影响他人者，有意损坏赛场设备或设施者，经现场裁判报告裁判长，经大赛组委会办公室同意后，由裁判长宣布取消其比赛资格。

（五）离场规则

1.比赛结束前15分钟，裁判长提示一次比赛剩余时间。

2.比赛结束信号给出，由裁判长宣布终止比赛。

3.裁判长宣布终止比赛后，参赛选手应停止全部竞赛任务的操作。竞赛过程中的所有资料和比赛用设备工具等均需整齐摆放在工作台上，不得带出赛场。

4.裁判长宣布终止比赛后，现场裁判组织、监督选手有序退出工位，等待评定比赛成绩。裁判长宣布离场时，现场裁判指挥选手统一离开赛场。

5.需要补时的参赛选手继续在自己的比赛工位上，在现场裁判宣布补时操作开始后，需要补时的参赛选手开始继续操作。当现场裁判宣布补时时间结束时，选手应停止操作，离开赛场。

6.所有参赛选手在完成比赛之后，均应该整理比赛工位及其周边的清洁，使之符合职业规范。

（六）成绩评定

竞赛系统根据设定的评分标准进行自动评分，确定参赛选手完成比赛成果的各项评分。

（七）成绩公布

本赛项评判、检测和成绩统计完成后，承办院校指派专人负责将比赛成绩及时录入网上竞赛管理系统，并公示。最终奖次以正式获奖文件为准。

（八）竞赛要求

1.竞赛过程中，不得使用任何搜索引擎查询，同时不得使用 QQ、微信等各类通讯 APP 用于比赛内容交流，一经发现取消比赛资格，成绩作废。

2.竞赛过程中，参赛选手不得携带任何笔纸等工具，（黑色水笔一支、空白A4 草稿纸一张均已在工位放置好），不得携带任何电子、纸质等资料，不按要求准备经组委会一次警告不予改正的，取消比赛资格，成绩作废。

**八、竞赛环境**

竞赛场地中的每个竞赛工位内设有操作平台。

每个工位内，配有单独带漏电保护空气开关的220伏交流电源，工位内的电缆线应符合安全要求。每个竞赛工位按照600W的用电负荷来规划准备。每个竞赛工位面积5～6㎡，工位之间由隔板隔开，以确保参赛队之间互不干扰。竞赛工位标明工位号，并配备竞赛平台和技术工作要求的软、硬件环境。环境标准要求保证赛场采光（大于 500lux）、照明和通风良好，为每支参赛队提供笔、纸张等工具，提供饮用水，提供一个垃圾箱。竞赛全天使用内网进行比赛，不连接公网环境。

**九、竞赛设备**

竞赛设备由主办方统一提供，具体见下表。

| 品名 | 规格要求说明 |
| --- | --- |
| 参赛选手计算机 | 配置要求：酷睿I3双核3.0以上CPU；4G以上内存；100G以上硬盘。预装Windows7及以上操作系统；预装谷歌浏览器；预装录屏软件；预装搜狗等输入法；预装Python3.9及以上编程环境 |
| 裁判等工作人员计算机 | 配置要求：酷睿I5双核3.0以上CPU；8G以上内存；100G以上硬盘；千兆网卡。  预装Windows7以上操作系统；预装谷歌、火狐浏览器；预装搜狗输入法等； |
| 网络连接设备 | 提供网络布线、千兆交换机、千兆分布式无线路由器 |
| 现场布置 | 工位隔断、环境布置、桌椅等 |
| 竞赛服务器 | 配置要求：英特尔至强E5系列八核以上CPU；16GB以上内存；500G以上硬盘；千兆网卡。  预装Windows Server 2012 R2操作系统 |

**十、成绩评定**

（一）评分标准

Python编程竞赛得分分为知识理论部分和上机实操部分，两者分数比例设置为2:8。

实操中每个竞赛试题任务的最终得分等于任务得分 \* 任务难度系数，难度系数分为容易、中等、困难三个等级。

1.程序设计

每道程序设计任务的测试用例，分为公开测试用例和隐藏测试用例两种，每个任务满分为100分，由任务基础分和任务完成分两部分构成。

点击“提交代码”按钮，系统显示提交结果、提交次数、执行

用时，并计算该任务的任务基础分和任务完成分。

2.提交结果

点击“提交代码”按钮，提交结果为“全部通过”“部分通过”“未通过”3种状态：

全部通过：测试用例全部通过，程序无报错，程序执行时间未超过规定上限。

部分通过：测试用例未全部通过，程序无报错，程序执行时间未超过规定上限。

未通过：程序运行异常或程序执行时间超过规定上限。

3.任务基础分

任务基础分总分为60分。

点击提交代码，提交结果显示“全部通过”时，任务基础分为60分，并计算任务完成分。

点击提交代码，提交结果显示“部分通过”时，根据测试用例通过数量计算相应任务基础分，而任务完成分为0分。

例如某参赛选手某个任务测试用例部分通过，提交次数为3，系统计算任务基础分=42分，则该任务最终得分=任务基础分42分+任务完成分0分=42分。

4.任务完成分

任务完成分最高为40分。

当提交结果显示“部分通过”和“未通过”时，任务完成分为0分。只有提交结果显示“全部通过”，系统才能计算任务完成分。

任务完成分= 40 \* 0.8 ^ (提交次数-1），任务完成分最低为0分。

5.竞赛总成绩

参赛选手竞赛总成绩= ∑各个任务最终得分\*难度系数，竞赛总成绩由完成的程序设计任务数量和任务难度决定，竞赛总成绩保留2位小数。

6.成绩评定

（1）成绩评定：评分方式为系统根据程序运行结果评分。

（2）成绩复核：为保障成绩评判的准确性，裁判组将对赛项总成绩排名前30%的所有参赛选手的成绩进行复核；对其余成绩进行抽检复核，抽检覆盖率不得低于15%。如发现成绩错误，请以书面方式及时告知裁判长，由裁判长更正成绩并签字确认。复核、抽检错误率超过5%的，裁判组将对所有成绩进行复核。

（3）赛项最终得分：按参赛选手最终完成的有效任务计分，最终成绩经复核无误，由裁判长签字确认后公布。

**十一、申诉与仲裁**

1.各参赛队对不符合赛项规程规定的设备、工具、材料、计算机软硬件、竞赛执裁、赛场管理及工作人员的不规范行为等，可向赛项仲裁组提出申诉。

2.申诉主体为参赛队领队。

3.申诉启动时，参赛队以该队领队签字同意的书面报告的形式递交赛项仲裁组。报告应对申诉事件的现象、发生时间、涉及人员、申诉依据等进行充分、实事求是的叙述。非书面申诉不予受理。

4.提出申诉应在赛项比赛结束后一小时内提出。超过一小时不予受理。

5.赛项仲裁组在接到申诉报告后的一小时内组织复议，并及时将复议结果以书面形式告知申诉方。申诉方对复议结果仍有异议，可由领队向大赛仲裁工作组提出申诉。大赛仲裁工作组的仲裁结果为最终结果。

6.申诉方不得以任何理由拒绝接收仲裁结果；不得以任何理由采取过激行为扰乱赛场秩序。仲裁结果由申诉人签收，不能代收；如在约定时间和地点申诉人离开，视为自行放弃申诉。

7.申诉方可随时提出放弃申诉。

**十二、竞赛视频**

1．赛场内部署录像设备，能实时录制并保存赛场情况。

**十三、竞赛须知**

（一）指导老师须知

1.各指导老师要发扬良好道德风尚，听从指挥，服从裁判，不弄虚作假。指导老师经报名、审核后确定，一经确定不得更换。

2.对申诉的仲裁结果，指导老师应带头服从和执行，还应说服选手服从和执行。

3.指导老师应认真研究和掌握本赛项比赛的技术规则和赛场要求，指导选手做好赛前的一切准备工作。

4.指导老师应在赛后做好技术总结和工作总结。

（二）参赛选手须知

1.参赛选手在报名获得审核确认后，原则上不再更换，如在筹备过程中，选手因故不能参赛，所在学校需出具书面说明并按相关规定补充人员并接受审核；竞赛开始后，不得更换参赛选手。

2.各参赛选手要注意饮食卫生，防止食物中毒。

3.各参赛选手要发扬良好道德风尚，听从指挥，服从裁判，不弄虚作假。

4.各参赛选手在比赛期间，应保证自身的安全，防止交通事故和其它意外事故的发生。

5.参赛选手应遵守比赛规则，尊重裁判和赛场工作人员，自觉遵守赛场秩序，服从裁判的管理。

6.参赛选手应佩戴参赛证，带齐身份证、已注册学生证。在赛场的表现，应体现自己良好的职业习惯和职业素养。

7.进入赛场前须将手机、手表、耳机等一切与比赛无关物品交赛场相关人员保管，不能带入赛场。未经检验的工具、电子通讯与储存器件和其他不允许带入赛场物品，一律不能进入赛场。

8.比赛过程中不准互相交谈，不得大声喧哗；不得有影响其他选手比赛的行为，不准有任何作弊行为。

9.参赛选手在比赛的过程中，应遵守安全操作规程，文明操作。通电调试设备时，应经现场裁判许可，在技术人员监护下进行。

10.需要更换设备时，应向现场裁判报告，并在赛场记录表上填写更换设备原因，核实从报告到更换完成的时间并签工位号确认，以便后续补时之用。

11.比赛过程中需要去洗手间，应报告现场裁判，由裁判或赛场工作人员陪同离开赛场。

12.未完成比赛任务，因病或其他原因需要终止比赛离开赛场，需经裁判长同意，在赛场记录表的相应栏目填写离场原因、离场时间并签工位号确认后，方可离开；离开后，不能再次进入赛场。

13.裁判长发出停止比赛的指令，选手（包括需要补时的选手）应立即停止操作进入通道，在现场裁判的指挥下离开赛场到达指定的区域等候评分。需要补时的选手在现场裁判的安排下进行补时，待补时结束后应立即停止操作。

14.如对裁判员的执裁有异议，可在比赛结束后1小时内向赛项仲裁组以书面形式提出申诉。

15.遇突发事件，立即报告裁判和赛场工作人员，按赛场裁判和工作人员的指令行动。

（三）工作人员须知

1.工作人员必须服从赛项组委会统一指挥，佩戴工作人员标识，认真履行职责，做好服务赛场、服务选手的工作。

2.工作人员按照分工准时上岗，不得擅自离岗，应认真履行各自的工作职责，保证竞赛工作的顺利进行。

3.工作人员应在规定的区域内工作，未经许可，不得擅自进入竞赛场地。如需进场，需经过裁判长同意，核准证件，有裁判跟随入场。

4.如遇突发事件，须及时向裁判长报告，同时做好疏导工作，避免重大事故发生，确保竞赛圆满成功。

5.竞赛期间，工作人员不得干涉及个人工作职责之外的事宜，不得利用工作之便，弄虚作假、徇私舞弊。如有上述现象或因工作不负责任的情况，造成竞赛程序无法继续进行，由赛项组委会视情节轻重，给予通报批评或停止工作，并通知其所在单位做出相应处理。

（四）裁判员须知

1.裁判员执裁前应参加培训，了解比赛任务及其要求、考核的知识与技能，认真学习评分标准，理解评分表各评价内容和标准。不参加培训的裁判员，取消执裁资格。

2.裁判员执裁期间，统一佩戴裁判员标识，举止文明礼貌，接受参赛人员的监督。

3.遵守执裁纪律，履行裁判职责，执行竞赛规则，信守裁判承诺书的各项承诺。服从赛项专家组和裁判长的领导。按照分工开展工作，始终坚守工作岗位，不得擅自离岗。

4.裁判员有维护赛场秩序、执行赛场纪律的责任，也有保证参赛选手安全的责任。时刻注意参赛选手操作安全的问题，制止违反安全操作的行为，防止安全事故的出现。

5.裁判员不得有任何影响参赛选手比赛的行为，不得向参赛选手暗示或解答与竞赛有关的问题，不得指导、帮助选手完成比赛任务。

6.公平公正的对待每一位参赛选手，不能有亲近与疏远、热情与冷淡差别。

7.选手有检查设备、更换器件或补充耗材的要求时应予以满足。检查设备或更换设备由赛场技术人员进行，更换后设备的应检测无误后继续投入比赛。赛场技术人员须在赛场记录表上记录需要更换的原因、要求更换到更换完毕的用时、并要求参赛选手签工位号确认。

8.赛场中选手出现的所有问题，如：违反赛场纪律、违反安全操作规程、提前离开赛场等，都应在赛场记录表上记录，并要求选手签工位号确认。

9.严格执行竞赛项目评分标准，做到公平、公正、真实、准确，杜绝随意打分；对评分表的理解和宽严尺度把握有分歧时，请示裁判长解决。严禁利用工作之便，弄虚作假、徇私舞弊。

10.竞赛期间，因裁判人员工作不负责任，造成竞赛程序无法继续进行或评判结果不真实的情况，由赛项组委会视情节轻重，给予通报批评或停止裁判资格，并通知其所在单位做出相应处理。

**附件-样卷**

**2023 年河北省职业院校技能大赛**

**“Python编程”赛项样题**

**模块一 语法基础与算法分析**

语法基础与算法分析任务包含任务详情、代码编辑区和控制台。每个算法设计任务都具有详细的任务说明、提交要求和测试用例。通过限定学生的输入和输出，对学生的输入和输出条件加以限定，学生能够更加准确高效地完成算法程序设计任务。

学生完成代码编写后，可以点击运行程序查看程序运行结果。学生也可以通过程序运行结果优化代码。

点击提交代码，可以讲程序运行结果上传，系统自动记录代码提交前的调试次数、程序执行时间等。

**样题1**：

定一个整数数组 nums 和一个目标值 target，请你在该数组中找出“和”为目标值的两个整数，并返回他们的数组下标。你可以假设每种输入只会对应一个答案，数组中元素不能重复输入。请根据上面的要求，编写相应的算法，将右边的函数 twoSum() 补充完整。

举例

给定 nums = [2, 7, 11, 15]，target = 9；

nums[0] + nums[1] = 2 + 7 = 9，返回[0, 1]。

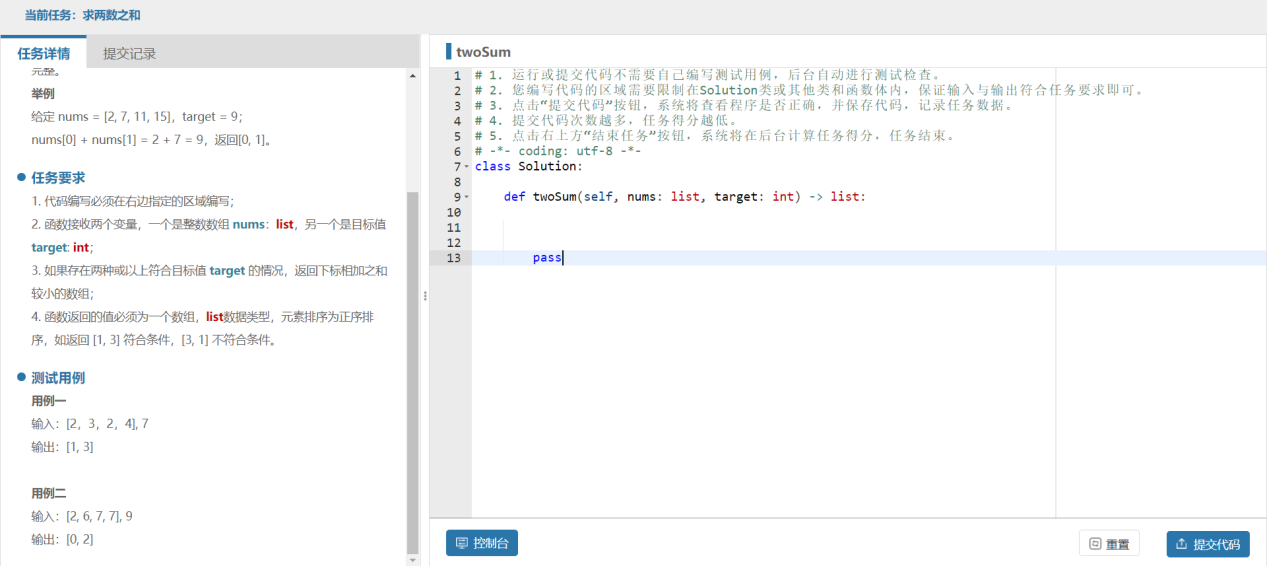
任务要求

1. 代码编写必须在右边指定的区域编写；

2. 函数接收两个变量，一个是整数数组 nums：list，另一个是目标值 target: int；

3. 如果存在两种或以上符合目标值 target 的情况，返回下标相加之和较小的数组；

4. 函数返回的值必须为一个数组，list数据类型，元素排序为正序排序，如返回 [1, 3] 符合条件，[3, 1] 不符合条件。



**样题2：**

**任务详情**

给定一段英文句子，判断句中单词拼写是否满足以下规则。

除以下特殊情况外，句子中第一个单词首字母必须大写，其它所有单词小写：

1. 如果句中的某个单词或短语，字母全部为大写，则该单词或短语拼写正确。比如“USA”、“UK”、“JUST DO IT”等；

2. “Python”、“Java”、“MachineLearning”、“DataMining”四个单词必须为双引号中给出的形式，否则拼写不正确；

3. 如果句中单词为“数字+字母”的混合形式，比如“5G”，该单词所有字母全部大写。

**任务要求**

1. 输入英文句子为 str 类型，输出为 bool 类型；

2. 如果句子满足规则要求，程序返回结果为 True；否则返回 False。

**测试用例**

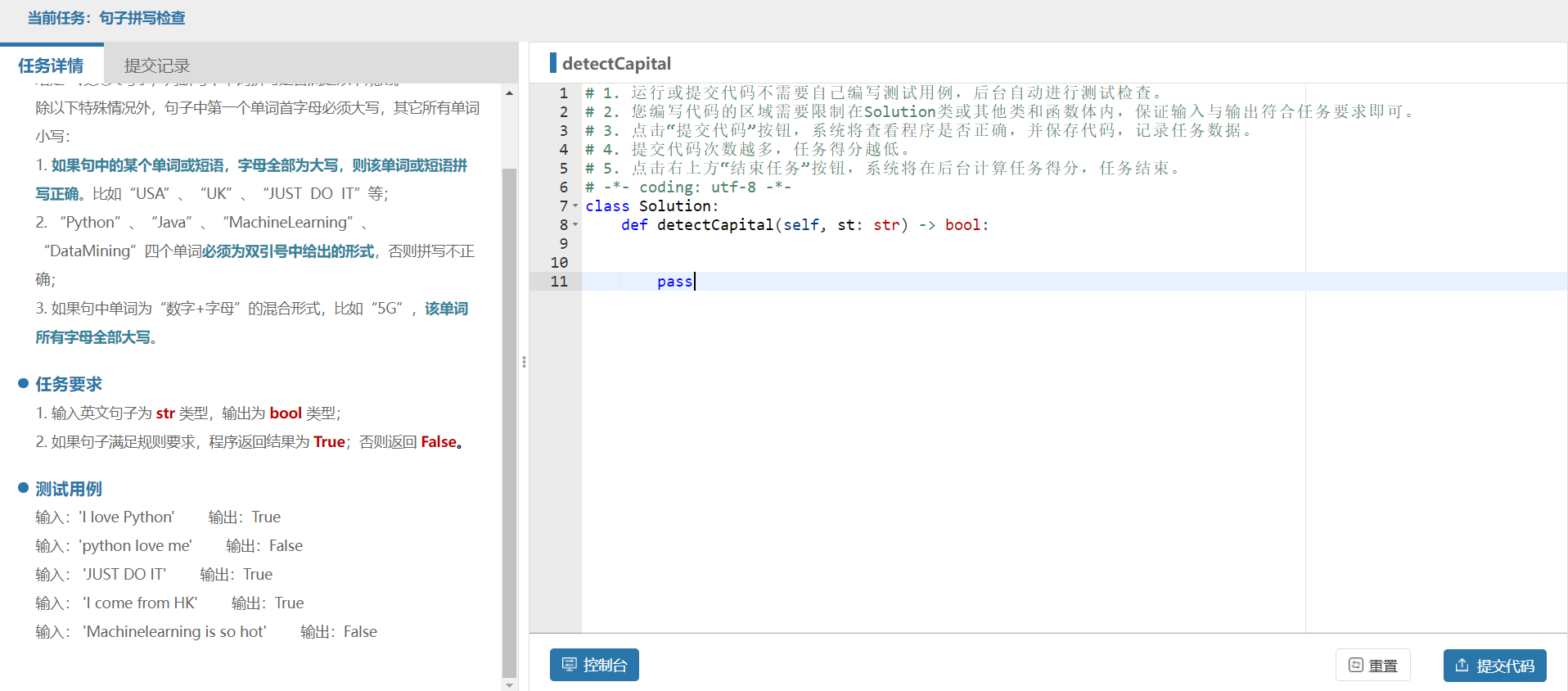
输入：'I love Python' 输出：True

输入：'python love me' 输出：False

输入： 'JUST DO IT' 输出：True

输入： 'I come from HK' 输出：True

输入： 'Machinelearning is so hot' 输出：False



**模块二 网页处理与文本分析**

网页处理与文本分析设计任务，主要锻炼学生的对HTTP基础、网页基础、正则表达式、爬虫的基本原理的掌握，能够熟练使用Request、Beautiful Soup等爬虫库。

**样题1：**

下方的网页为微信公共号数据分析展示页，网页数量为1。

请根据网页给出的信息，完成以下任务，将右边的函数 weixinData() 补充完整。

给定原创排行榜里任一微信公众号的名称，请判断该公众号是否满足以下条件：

1.“头条平均阅读”数 > 90000

2.“原创平均阅读”数 > 80000

3.“预估活跃粉丝”数 < 300万

如果满足，函数 weixinData() 返回大写英文单词“YES”，否则返回“NO”。

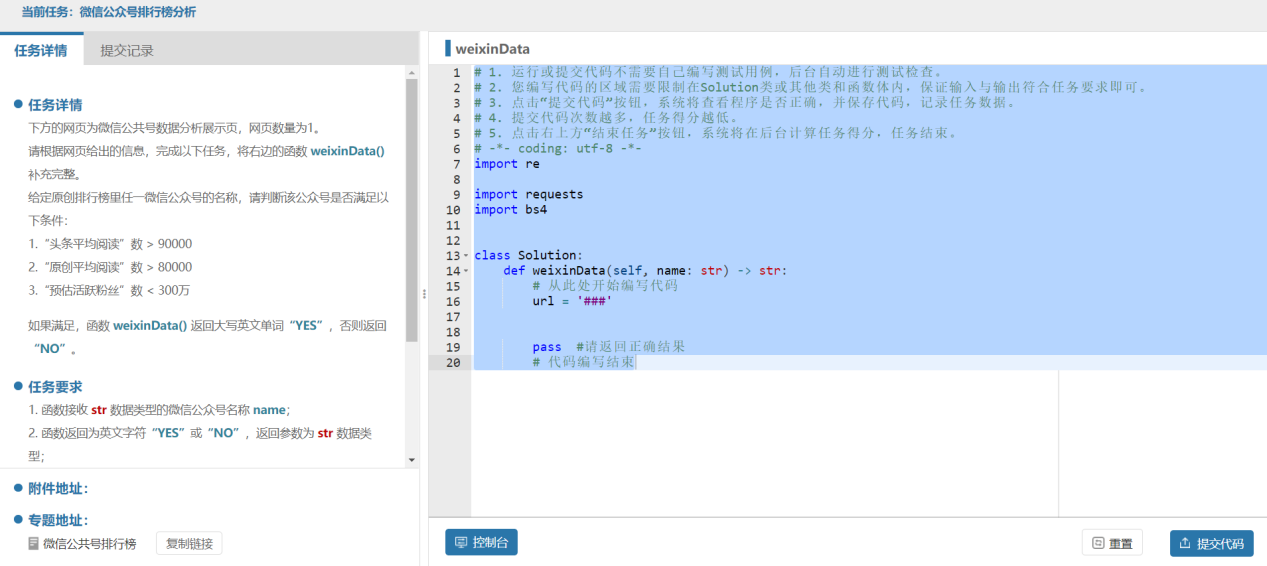
**任务要求**

1. 函数接收 str 数据类型的微信公众号名称 name；

2. 函数返回为英文字符“YES”或“NO”，返回参数为 str 数据类型；

3. 不得修改函数 weixinData() 的名称；

4. 题目所需网站链接已经在下方给出，编码方式为UTF-8。





**样题2：**

**任务详情**

根据提供的商品文本信息，对商品的标题内容、属性内容和描述内容进行分词（不包含标题和商品的属性名），完成文本分析工作。

请根据以上要求，将以下所需的函数补充完整。

本任务提供了jieba 中文分词库和 requests 库。

**任务要求**

1. 构建函数 wordTfidf()，对商品中除停用词外的关键词，计算其TF-IDF值；

2. 返回文本中 TF-IDF 最大的前5个关键词，返回结果为 list 数据类型；

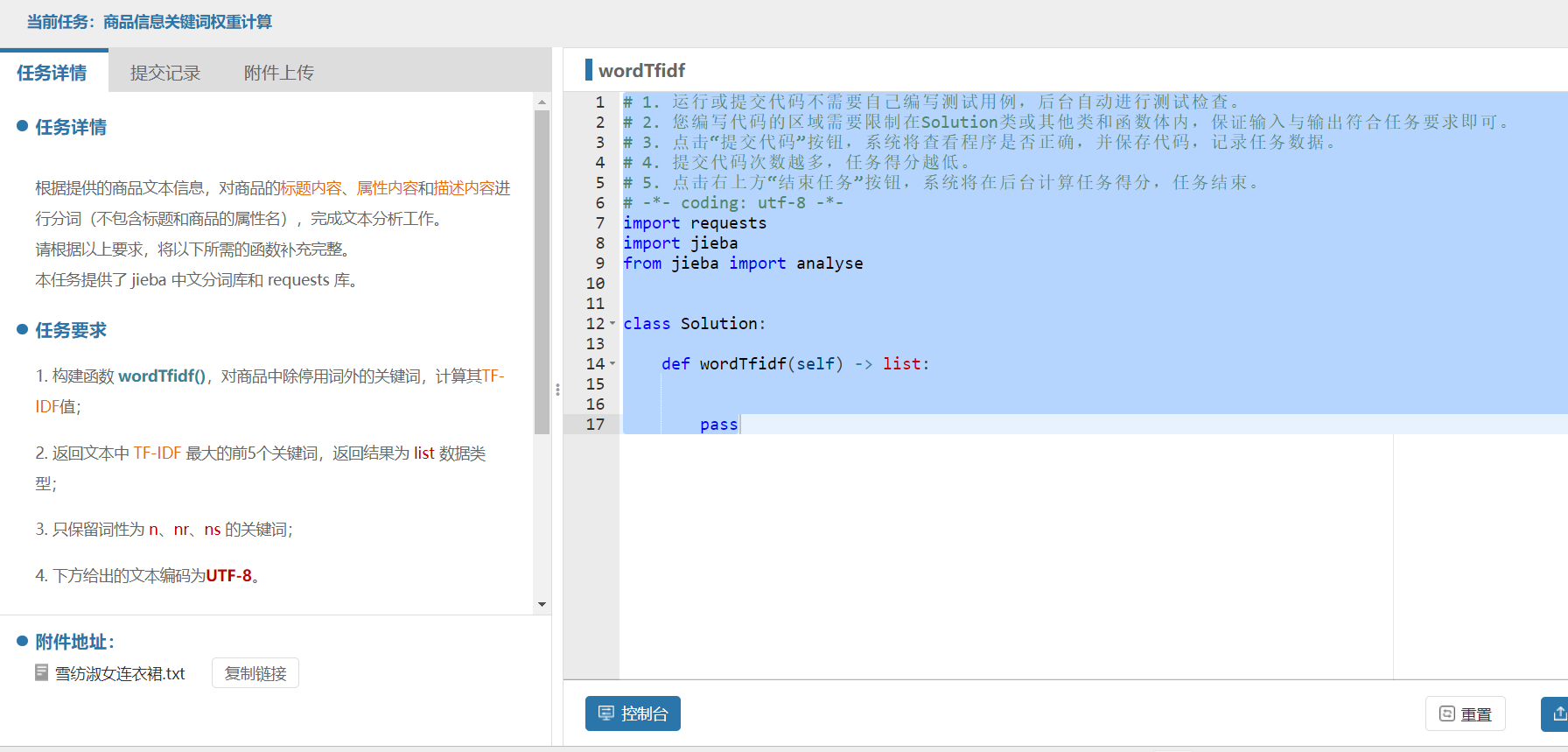
3. 只保留词性为 n、nr、ns 的关键词；

4. 下方给出的文本编码为UTF-8。

**测试用例**

输入：任务中商品信息

输出：['裙子', '\*\*\*', '\*\*\*', ...]



**模块三 数据处理与分析**

数据处理任务主要培养学生数据获取与清洗、商务数据分析、基础统计分析的能力，能够熟练掌握对Numpy、Pandas等数据分析库的应用。实训系统内置了多个大数据文件，可以通过程序直接读取并进行处理。

**样题1：**

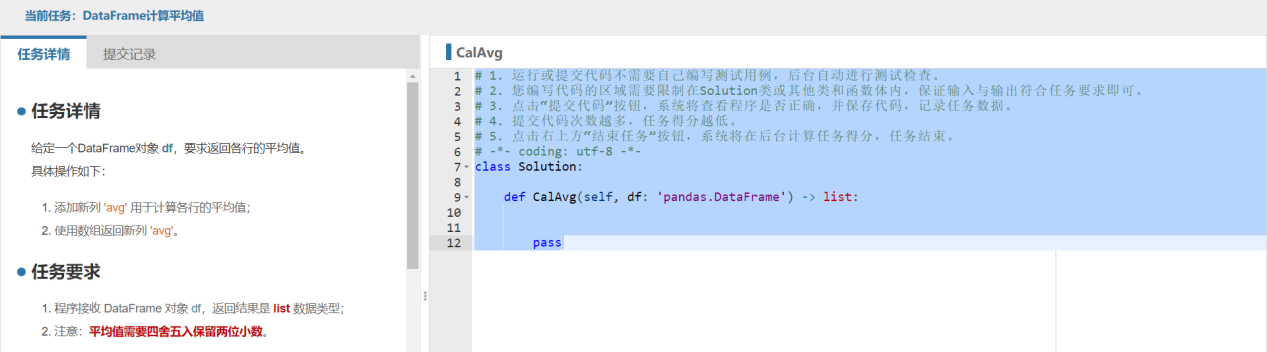
**任务详情**

给定一个DataFrame对象 df，要求返回各行的平均值。  
具体操作如下：添加新列 'avg' 用于计算各行的平均值；使用数组返回新列 'avg'。

**任务要求**

程序接收 DataFrame 对象 df，返回结果是 list 数据类型；

注意：平均值需要四舍五入保留两位小数。



**样题2：**

**任务详情**

下方是某店铺在 2 月份的订单数据，共 28010 行、7 列。 在该数据中，每一行代表一笔订单，每一笔订单可以 看做是一位顾客（买家），价格相同的产品可视为同一产品；

请编写程序，完成以下任务：

程序传入参数 area（收货地址），要求计算该地区订单的’订单的付款转化率’、’买家全额支付的转化率’、’买家实际支付总金额’、’客单价’、’销量（订单量）最多的产品的价格’、’买家实际支付总金额在所有地区实际支付总金额中的占比’，并将结果以列表形式返回。

**任务要求**

程序接收 str 类型的 area，返回的结果为 list 数据类型；

返回的结果中，列表中各个数据的数据类型为 str，其中转化率数据及占比数据需以百分比的形式输出，并且需要保留小数点后2位，如’13.14%’，其余结果不需要以百分比形式输出，但需要保留小数点后2位；

如果某地区所有买家的买家实际支付金额为 0，即该地区没有付费行为。则该地区的’客单价’、’销量（订单量）最多的产品的价格’需返回字符串’本地区无销量’，其他结果则正常计算；

如果该产品（订单）被全额退款，那么该产品不计入销量。

参考公式：

付款转化率 = 付款订单数 / 总订单数；

买家全额支付的转化率 = 买家全额支付的订单数 / 总订单数；

客单价 = 实际支付金额 / 支付买家数；

**测试用例**

输入：area = ‘河南省’

输出：[‘81.99%’, ‘61.18%’, ‘56354.03’, ‘92.84’, ‘37.00’, ‘2.96%’]

解释：河南省’订单的付款转化率’为’81.99%’，’买家全额支付的转化率’为’61.18%’，’买家实际支付总金额’为’56354.03’，’客单价’为’92.84’，’销量（订单量）最多的产品的价格’为’37.00’，’买家实际支付总金额在所有地区实际支付总金额中的占比’为’2.96%’

输入：area = ‘江苏省’

输出：[‘86.78%’, ‘66.75%’, ‘159359.18’, ‘109.22’, ‘37.00’, ‘8.38%’]

