



数据科学基础

大作业说明

赵源

南京大学智能软件工程实验室

www.iselab.cn

2020年4月14日



数据说明

- user_id: 用户唯一标识ID
- final_score: 该题最终得分
- case_id: 题目ID
- case_type: 题目类型
- case_zip: 题目包
- upload_id: 提交记录ID
- upload_time: 上传时间
- code_url: 对应upload_id所提交的代码
- score: 对应upload_id提交得分



数据示例

```
[
  {
    "user_id": 49405,
    "cases": [
      {
        "case_id": 2081,
        "case_type": "字符串",
        "case_zip": "http://moocetest-site.oss-cn-shanghai.aliyuncs.com/target/%e5%ad%90%e4%b8%b2%e6%9f%a5%e6%89%be_1577868767370.zip",
        "final_score": 100,
        "upload_records": [
          {
            "upload_id": 255171,
            "upload_time": 1582681866412,
            "code_url": "http://moocetest-dev.oss-cn-shanghai.aliyuncs.com/data/answers/4238/49405/%E5%AD%90%E4%B8%B2%E6%9F%A5%E6%89%BE_1582681865671.zip",
            "score": 0
          },
          {
            "upload_id": 255173,
            "upload_time": 1582681873689,
            "code_url": "http://moocetest-dev.oss-cn-shanghai.aliyuncs.com/data/answers/4238/49405/%E5%AD%90%E4%B8%B2%E6%9F%A5%E6%89%BE_1582681872840.zip",
            "score": 100
          }
        ]
      }
    ]
  }
]
```



下载脚本示例

```
import json
import urllib.request,urllib.parse
import os

f =open('data.json',encoding='utf-8') #打开'data.json'的json文件
res=f.read()
data = json.loads(res) # 加载json数据
print(data)
# 取出json中第一个学生的cases数据
cases = data[0]['cases']
print(cases)
# 遍历做题信息
for case in cases:
    print(case["case_id"],case["case_type"]);
    filename = urllib.parse.unquote(os.path.basename(case["case_zip"])) # 获取文件名, url里对中文会urlencode, 解个码
    print(filename)
    urllib.request.urlretrieve(case["case_zip"],filename) # 下载题目包到本地
```

研究报告格式要求

小组信息：人数 + 学号_姓名_邮箱_Python练习完成题目数量 + 组员分工职责

1. 研究问题：你研究的问题，包括但不限于如何选择学生的编程能力评价指标；如何在编程过程中推荐学习代码；如何自动生成编程学习路径等等，请简述问题的应用场景。
2. 代码开源地址：给出分析代码的地址，并且解释代码与研究问题的对应关系，代码的实现逻辑。
3. 研究方法（重点）：包括你用到的数据分析方法、所使用的数据集，及其他你所用到的一切方法。这部分需要你的详细说明，要做到逻辑清晰且易理解。
4. 案例分析：针对研究问题的分析，格式不作限制。
5. 对这门课的意见和想对老师说的话，也欢迎感兴趣的同学们申请加入实验室的研究。
6. 附录：你认为需要补充到研究报告中帮助读者理解的数据、图表等。

注：以上为研究报告的标准格式，如果你有特别的思路，也可以适当调整报告的结构。

提交PDF格式，宋体小四，1.5倍行距。



研究问题示例

研究问题：自动推荐代码片段

代码开源地址：XXXXXX

研究方法：分析编程代码和对应测试用例的通过关系，找到对应的编程代码片段。

在进行该题的编写时，如果用户一直未通过该测试用例，则将该编程代码片段推荐给用户……

案例分析：XX排序

陈老师，助教太菜了，明年我来当助教吧

附录：XXXXXX



研究问题示例

研究问题：题型自动分类

代码开源地址：XXXXXX

研究方法：使用相似度计算方法来比较题目的实现代码，对于实现高度相似的题目，我们将其归为一类……

案例分析：XX排序和XX排序

陈老师，我想学python找工作……

附录：XXXXXX

