

音乐推荐下载器

一、考核摘要

1. 任务概要
2. 语言
3. 工具推荐

二、项目需求

1. 基础功能
2. 拓展功能（可选）
3. 分包思路建议
4. 项目要求
5. 项目提交

三、知识点说明

1. 知识清单
2. 学习建议

四、考核时间

1. 具体期限
2. 建议时间分配

五、评估

六、写在最后

音乐推荐下载器

一、考核摘要

1. 任务概要

- 利用原生爬虫，模拟QQ音乐搜索、下载功能，并将下载的数据同步到mysql
- 利用提供的数据集，根据推荐算法，给用户推荐其可能喜欢的歌手

2. 语言

python3.X

3. 工具推荐

- 推荐python环境管理：Anaconda（若官网下载过慢，可到清华镜像站下载：<https://mirrors.tuna.tsinghua.edu.cn/anaconda/archive/?C=M&O=D>）
- 推荐python编译器：PyCharm（社区版够用）
- 推荐MySQL管理工具：Navicat for MySQL（安装可参考<http://www.downcc.com/soft/322714.html>）

二、项目需求

1. 基础功能

1. 模拟QQ音乐网页中的搜索和下载功能：

- 输入歌手或歌名搜索出相关的歌曲列表，然后可选下载。
- 可以将歌曲下载到本地。下载到本地的命名方式为：歌名-歌手-id，下载信息同步到MySQL数据库，用于检测是否重复下载。

2. 模拟QQ音乐的推荐系统：

- 根据用户的历史记录，给用户推荐其可能喜欢的歌手，并可以选择下载他们的歌曲。
- 推荐算法自选，至少实现两种，包括但不限于：无脑随意推荐、根据热度推荐、基于用户的协同过滤算法、基于物品的协同过滤算法。
- 根据需要划分数据集，并从准确率、召回率、覆盖率三方面比较不同推荐算法的性能。

2. 拓展功能（可选）

- 设计一个GUI交互界面，可用tkinter、PyQt等库
- 爬取歌曲相关评论，并进行文本数据分析，如提取高频词绘制词云等
- 专辑歌单或mv搜索、下载
- 多线程异步下载
- 爬取网易云音乐（包含了加密算法，较难于QQ音乐）
- 自行发挥

3. 分包思路建议

- MusicSpiderByName
 - utils（工具文件夹，用来放各种自己写的函数和类）
 - db_tools（数据库相关文件夹）
 - src（存放爬虫的文件夹）
 - recs（存放推荐算法的文件夹）
 - ...

4. 项目要求

- 在readme文件（txt\Markdown）写清项目思路、说明外加库
- 注意面向对象的编程
- 注意代码命名规则
 - 类名 首字母大写：Music
 - 属性名 单个单词小写：music
 - 属性名 多个单词除第一个，每个单词字母大写：myMusic
 - 数据库命名 全部小写，下划线隔开：my_music
 - 所有命名英文！！：music √ yinyue ×

5. 项目提交

- 笔记提交：每人注册一个博客账号（若已有可继续使用），用于记录所学知识点，每隔一定时间（自己把握，最好每隔两三天）总结一次学习的内容、问题或者进度。（学习笔记也会作为评价标

准的一部分)

- 文件提交：将项目打包成压缩文件在4月19号前提交（源代码、readme文件）

三、知识点说明

1. 知识清单

- python3入门基础
- 爬虫原理及相关知识
- 数据库SQL基础语法
- 推荐系统算法原理

2. 学习建议

- 推荐菜鸟教程、廖雪峰网站等python入门基础学习
- 爬虫建议实战、在案例中学习。没有爬虫基础的同学建议先尝试一些简单的爬虫。可以参考崔庆才的《Python 3网络爬虫开发实战》
- 推荐系统可以学习项亮的《推荐系统实践》第一二章
- 各类视频资源可以自行查看网课：网易云、慕课、腾讯课堂等
- 各类博客资源科自行查看，如：CSDN、博客园等

四、考核时间

1. 具体期限

- 2020年3月26日~2020年4月19日

2. 建议时间分配

- python基础：4天左右
- 推荐系统相关知识：4天左右
- 数据库知识：3天左右
- 爬虫知识：4天左右
- 基本实现代码：1周左右
- 完善优化：3天左右

五、评估

评估内容	比例
基本功能实现	50%
程序健壮性	10%
答辩	15%
拓展功能	15%
学习笔记	10%

六、写在最后

- 坚持就是胜利!!!
- 考核结果不会以最终功能实现为唯一标准，会考虑师弟师妹们的基础、学习是否扎实、学习速度、积极性、性格等等~~~