

大数据实习项目

一、题目

网易云音乐网站数据分析

二、研究主要内容

研究内容：

- - 分析单曲排行榜
 - 上榜歌曲
 - 歌手名
 - 评论数
 - 歌曲情感分析
 - 评论情感分析
 - *歌曲风格分析
 - 单曲榜emo指数
 - 热门歌曲的地域分布

评论关键词

根据用户的地理位置信息，统计热门歌曲在不同地区的受欢迎程
- - 分析歌单排行榜
 - 上榜歌单
 - 标签分析
 - 播放量
 - 收藏数
 - 评论数
 - 歌单情感分析
 - 评论情感分析
 - 歌单榜emo指数

评论关键词
- - 分析专辑排行榜
 - 上榜专辑
 - 专辑价格
 - 购买数
 - 评论数
 - 专辑情感分析
 - 评论情感分析
 - 专辑榜emo指数

评论关键词
- 相似歌曲推荐
- 分地区活跃用户emo指数分析
- 分地区单曲榜单
- 分年龄活跃用户emo指数分析
- 分性别活跃用户emo指数分析

研究方法：

- 开发爬虫程序采集网易云音乐数据
- 使用redis构造分布式爬虫
- 使用hadoop存储网络数据
- 使用spark分析网络数据

研究目标：

通过对网易云音乐网站数据分析，可以洞察用户需求、优化用户体验、提升音乐推荐效果，同时也为音乐产业链的各个环节提供决策支持，推动音乐行业的发展。

三、主要技术指标

1. 系统需基于B/S结构
2. 采用传统的三层结构方式进行解耦开发
3. 严格按照软件工程的要求实施课题的调研，分析，设计，测试，部署和各阶段管理
4. 数据分析需采用hadoop或者spark