**Spark课程设计**

**基于《权力的游戏》的spark数据分析**

**需求分析说明书**

1. 背景与目标

背景：近年来，数据分析技术的发展使得对大规模数据的分析变得更加高效和便捷。电视节目《权力的游戏》因其复杂的剧情和众多的角色受到了广泛关注，拥有大量可供分析的数据。通过对这些数据的深入分析，可以揭示出观众的喜好、剧情的发展趋势以及角色的受欢迎程度等有趣的信息。

目标：本项目旨在利用Spark技术，对《权力的游戏》的剧集数据进行全面的分析，并通过一个Web应用平台展示分析结果。具体目标包括:

(1).数据的收集、清洗与存储。

(2).数据分析与挖掘、提供有价值的信息

(3).数据的可视化展示，提供用户友好的交互界面，直观展示数据

2.功能需求：

2.1数据收集与处理：

（1）.**数据来源**：

《权力的游戏》全剧集剧本台词、各剧集信息,包括作者、编剧、收看人数、平均评分等。

（2）.**数据处理**：处理缺失值、重复值，确保数据质量

2.2数据展示与可视化：

(1).**静态展示**：词云图、柱状图、折现图、扇形图、玫瑰图等。

(2).**动态展示**：支持动态交互效果，如3D柱状图，时间线组合图。

(3).**多维度分析**：支持同一因素多维度分析，如观看人数、参评人数、平均评分与剧集间，导演、作者与剧集间，权游情节、人物台词的词云等。

2.3用户交互：

（1）**筛选功能**：用户可以根据不同分析要求进行数据查询，得到直观信息。

（2）**交互式图表**：图表支持鼠标悬停查看详细信息，提供丰富的交互体验与信息反馈。

3.系统架构：

（1）**前端：**使用LayuiMini框架实现响应式布局和丰富的UI组件，提升用户体验。通过Ajax与后端进行数据交互。

（2）**后端：**基于Flask框架，使用蓝图模式分离视图，Jinja模板引擎进行页面渲染。使用Flask-SQLAlchemy进行数据库操作。

（3）**数据处理：**使用Spark进行大规模数据处理，确保数据处理的高效性和准确性。

（4）**可视化：**使用Pyecharts进行数据的可视化展示，提供丰富的图表类型和交互效果。

4.详细设计：

4.1前端设计：

（1）页面布局：主页面包含导航栏、数据展示区和操作区，用户可以方便地进行数据查询和分析。

（2）组件设计：使用LayuiMini的预构建组件，如表格、表单、图表等，快速构建页面功能。

4.2后端设计：

（1）路由设计：使用Flask蓝图分离不同模块的路由，使代码结构清晰。

（2）数据接口：提供RESTful API接口，支持前端的数据请求。

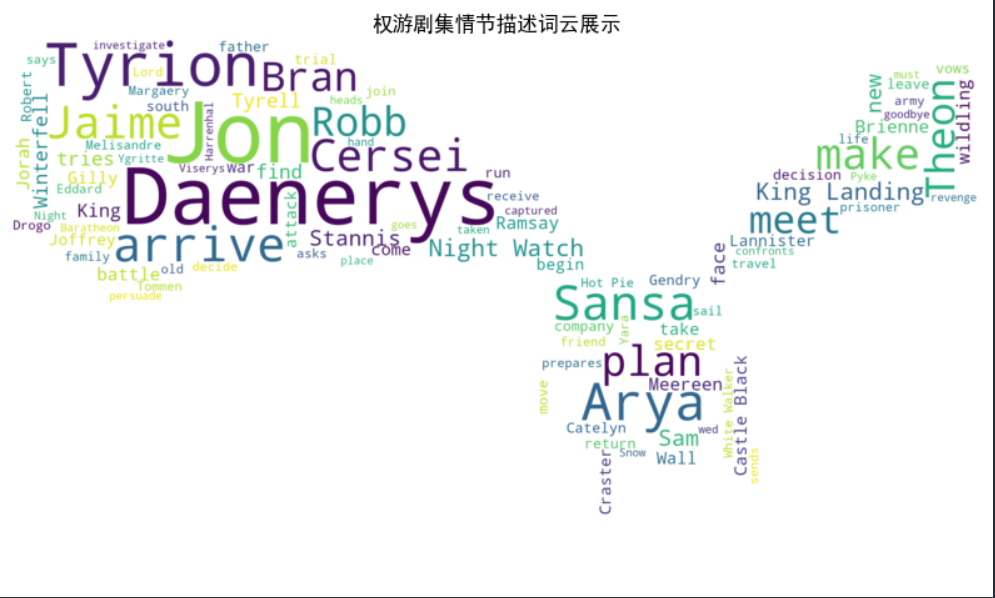
4.3数据处理设计：

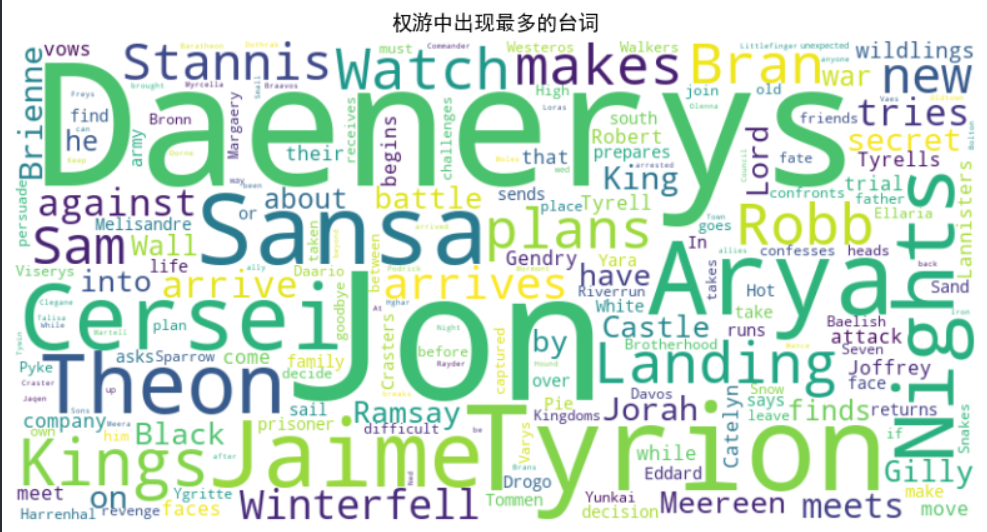
（1）数据清洗：使用Spark对原始数据进行清洗，包括缺失值处理、数据转换，按某属性进行分组等。

（2）数据存储：清洗后的数据按特性有些存储在非关系型数据库中，有些进行导出按CSV文件格式存储，方便后续查询和分析。

4.4可视化设计：

**(1)词云图**：通过绘制词云图寻找权游中的核心人物，进行人物剧情的分析，主要展示的数据是所有剧集的剧本情节描述和人物台词词频统计。





**(2)面积图**：绘制面积图展示权游各季的观看人次的变化，通过看面积图的变化趋势反应权游的受欢迎程度。

**(3)折现图**：绘制折线图展示权游各季参与评价人数的变化，参评人数也是反应剧集爆火程度的重要标准，通过观察折现图折线的变化趋势反应权游的受欢迎程度。



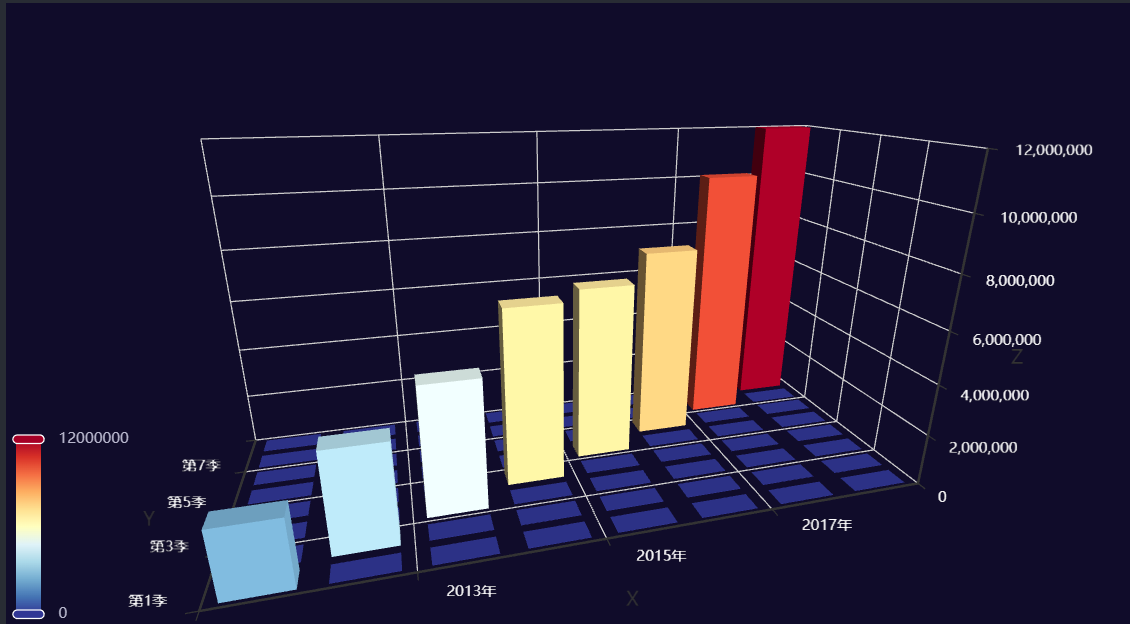
**(4)时间线仪表盘图**：通过利用时间线组件将多个仪表盘图整合在一起，每个仪表盘图展现了权游当季剧集的平均IMDB评分，以10分为满分，0分为最低分进行评价，展现当季的好评率。





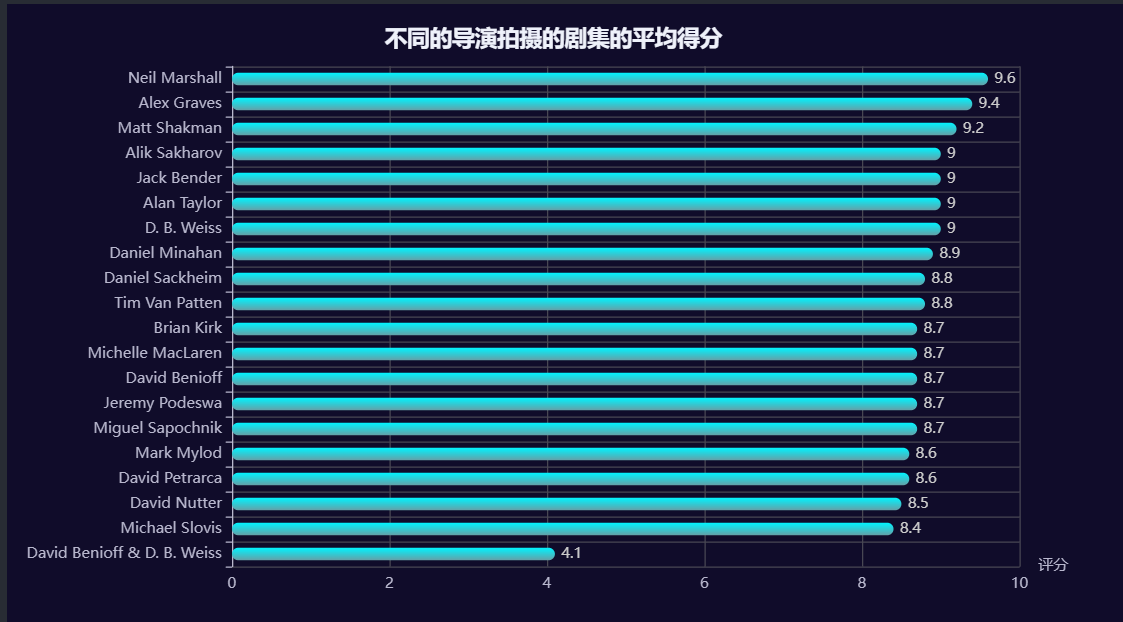


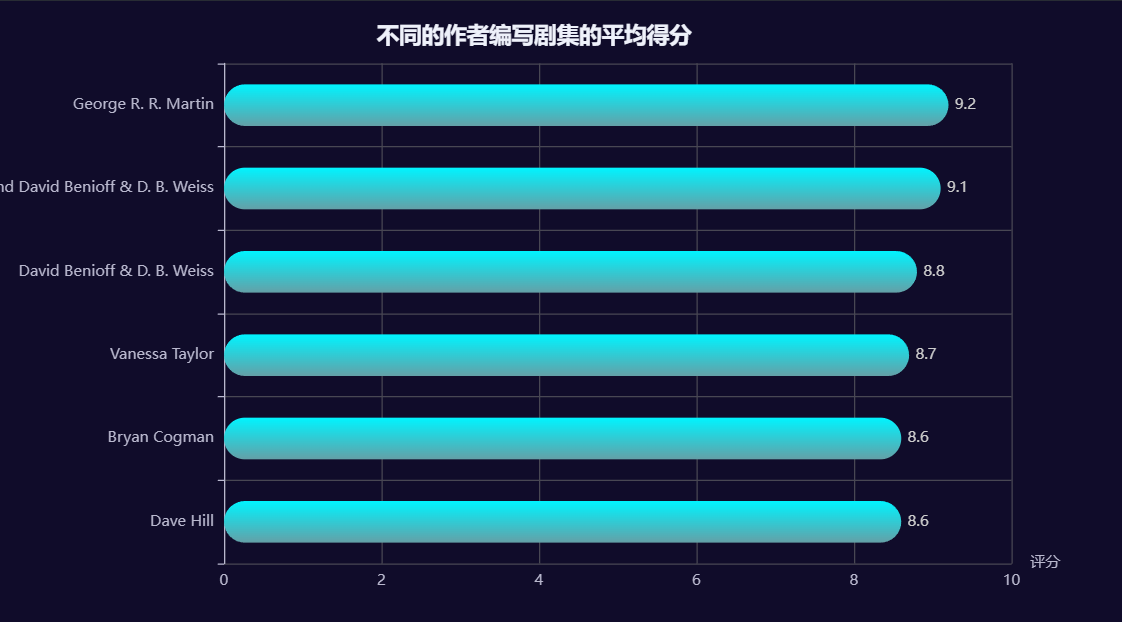
**(5)3D柱状图**：绘制3D柱状图展现权游按年份发布的剧集的的受欢迎程度，提供了一个立体的视角，帮助用户从年份和季数两个维度分析和理解《权力的游戏》各个季数的观看人数表现，更加立体的让用户感受。



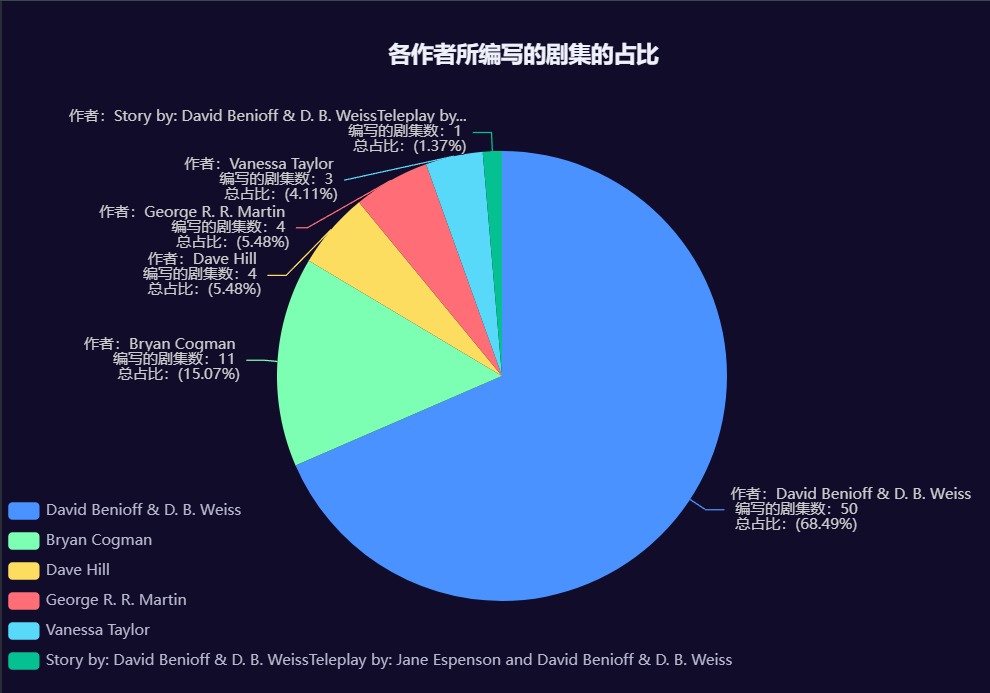
**(6)折线柱状组合图**：通过利用折线图和柱状图的组合，综合展示了观众在不同月份的观看行为、参与评价情况和剧集评分，提供了一个全面的视角来分析这些指标之间的关系。通过这种组合图表，可以观察到某个月的观看人数变化是否与评分变化相关联，或是参与评价人数的多少对评分有何影响，从而帮助理解观众行为和剧集质量变化趋势。

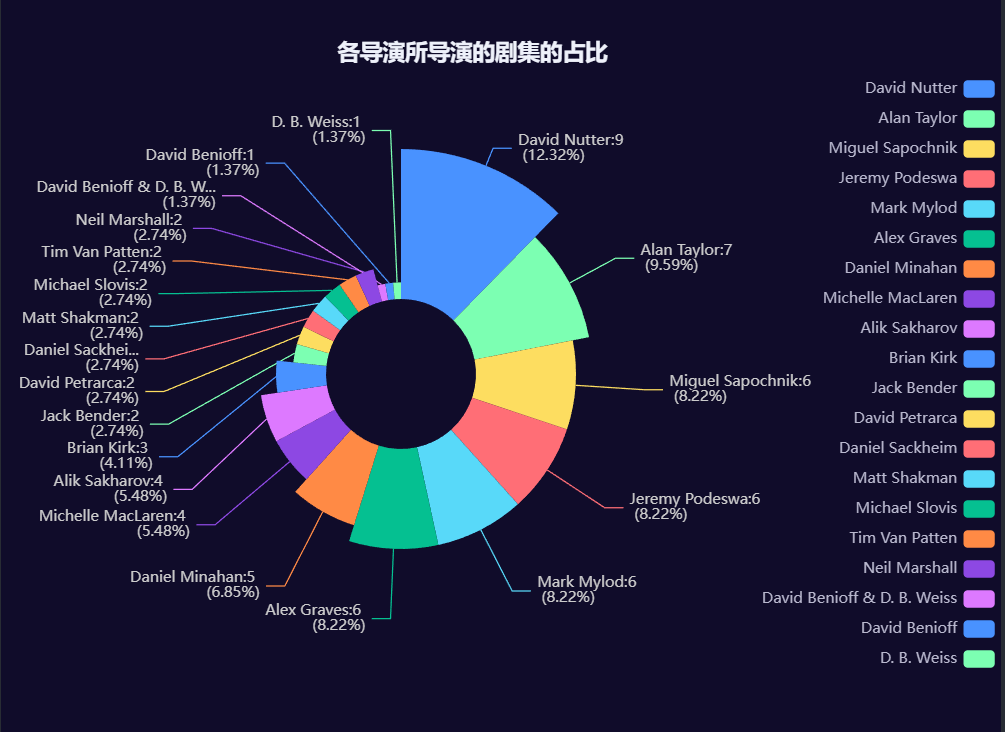


**(7)水平渐变柱状图**：绘制柱状图展现了权游不同作者编写剧集的平均IMDB评分，和不同导演拍摄剧集的平均IMDB评分，提供了一个直观的视角，帮助用户从不同导演、不同作者编写的剧集剧本的角度分析和理解《权力的游戏》剧集的评分差异。



**(8)饼图：**绘制饼图展现了《权力的游戏》中各作者编写剧集数量的占比，帮助用户从不同作者的角度分析和理解《权力的游戏》剧集的编写数量及其占比。



**(9)玫瑰饼图：**利用玫瑰饼图组件，该图表展现了《权力的游戏》中各导演所导演剧集数量的占比，帮助用户从不同导演的角度分析和理解《权力的游戏》剧集的导演数量及其占比。

5.总结

本文基于项目的系统设计说明书编写而成，包含了项目的背景、需求、系统架构和详细设计。开发本项目旨在通过对《权力的游戏》数据的分析和展示，提供一个功能强大且用户友好的数据分析系统。希望通过这个系统，用户能够深入了解这部剧集背后的数据故事。