本论文的主要研究内容如下：

了解基于Spark的TapTap游戏数据分析系统的基本架构，掌握系统的开发方法，包括系统开发基本流程、开发环境的搭建、测试与运行等。系统功能图如下图1所示。

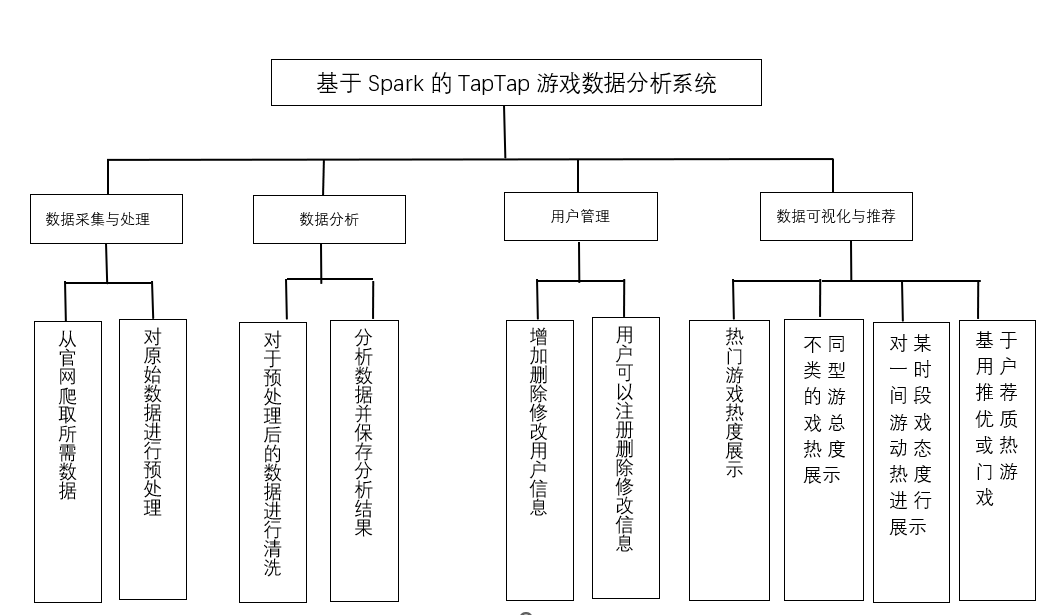


图1 系统功能图

主要功能如下：

（1）用户管理模块：用户能够注册、登录及修改个人信息，查看热门游戏及攻略信息。

（2）数据采集与处理模块：主要通过Python编程，爬取Tap Tap社区中游戏热门榜、热玩榜以及游戏的标签、评分等数据，同时删除冗余和无用信息，以用于大数据分析。

（3）数据分析模块：

①类型分析：对爬取的数据进行梳理并分析不同标签游戏的数据榜单，例如策略、单机、休闲、卡牌等不同版块。获取游戏中下载数、关注数、评价数等信息内容并分析。

②动态分析：分析最新动态内有关游戏的图文、视频和帖子，将各种动态的发帖时间、游戏出处、讨论数，点赞数量和游戏动态数量进行数据分析，分析某一时间段游戏动态热度并进行排名。

③游戏推荐：根据数据挖掘得到的信息，对所有游戏信息、游戏动态进行热度总结，按照不同权重和热度递增的方式筛选出不同游戏类型排名前十的游戏，点击进入不同的类型，系统会以最新和最热的方式进行游戏推荐。

（4）数据可视化模块：主要利用Echarts插件，对类型分析、动态分析、游戏推荐三个模块中数据分析的内容进行可视化展示。