**程序设计小组作业·项目方案书**

**项目题目：**豆瓣数据管理大师

**团队名称：**瑞士牛奶巧克力队

**队长：**董律成521120910137

**成员：**张艺蕾521120910128 刘浩521120910147

**项目简介：**

利用爬虫技术，从豆瓣网站挖掘数据，目前打算从短评（从评论判断用户的情感态度并预测电影评分、评论词频制作词云图等）、电影评分（电影评分与导演、演员等的相关性、分析随年份的电影畅销趋势等）和地理位置（观影类型的地域差异、某部电影不同地域的评分差异等）三个方式入手，分析豆瓣电影数据。整个Project分为三个模块，分别是数据爬取、数据分析和app用户GUI界面制作。

1. 数据爬取：requests, lxml(etree), beautiful soup, scrapy(可能)
2. 数据分析：matplotlib.pyplot, seaborn, numpy, pandas, sklearn, wordcloud, dlib
3. GUI界面：tkinter, pyinstaller