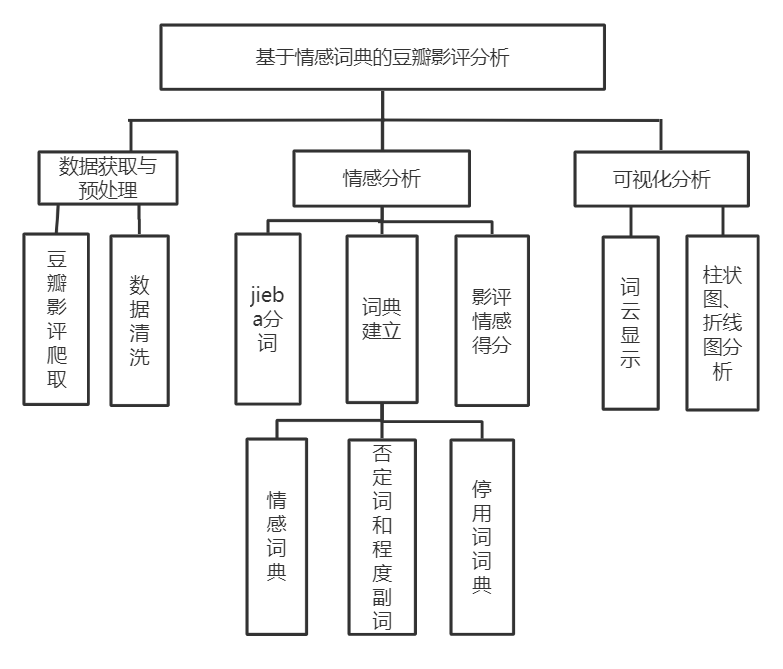
概要设计



1. **数据获取与预处理**

豆瓣影评爬取：选取爱情、动作、历史、喜剧、悬疑、战争共6类电影，每类选择10部电影，并对每部电影爬取2000条评论，存储至csv文件中。

数据清洗：对爬取到的电影短评数据作预处理，删除重复评论，删除空字段和特殊符号评论内容。

1. **情感分析**

jiaba分词：使用jieba进行中文分词，对比原评论，未去停用词的jieba分词评论，去停用词后的jieba分词评论。

词典建立：建立正面情感词典、负面情感词典、否定词词典、程度副词词典、停用词词典，并进行词典扩充处理，包括：设置程度副词权重，根据电影名、演员名等信息进行停用词扩充。

影评情感得分：编写python程序，使用情感词典对预处理后的电影评论进行情感分析。

1. **可视化方法**

柱状图：我们将每部电影的评分进行可视化，以影评情感得分为x轴，影评数量为y轴。生成相应的柱状图。

词云图：对每部电影根据分词和去停用词之后生成相应的词云图。