课程项目第一阶段

Java08组：许芳芳 郭转转 李欣

1. 项目题目：

基于爬虫技术的社交网络平台的好友推荐系统

1. 项目简介：

利用爬虫技术给豆瓣社交网络的用户推荐好友。根据豆瓣中每个用户喜欢的书，电影，音乐，关注的小站，常去的小组，豆列，关注的人，和被关注的人这些作为推荐的依据，向用户推荐他们有可能及其感兴趣的人或者认识的人。其中依据元素的具体提取信息对推荐系统的作用：

1. 喜欢的书，电影，音乐，豆列：这些我们需要爬取的是其标签和其数量。例如音乐的标签：古典，流行，欧美，日韩等。爬取每个标签的数量的作用是为了在后面对数据处理进行推荐时作为权重来设置推荐的优先推荐等。
2. 关注的小站，常去的小组：这个是爬取小站和小组的id号。若两个用户关注的小站，和小组是一样的，进行设置权值，推荐好友。
3. 关注的人，以及被关注的人：爬取每个用户的关注的人的id和被关注的人的id。例如有ABCD四个用户，A关注了B，A关注了C，D关注了A，则我们可以将D推荐给B或者C。
4. 项目流程



1. 技术调研
2. 对于现有的爬虫技术的调研：可以查找相关文献以及博客进行爬虫技术的了解以及下一个阶段的具体实施和操作。
3. 对于一些推荐算法的调研：由于我们要做好友推荐，所以需要掌握相关的算法调研，这也是整个项目的核心，根据所获得的数据以及数据的作用来进行相关算法的查找，以便后序工作顺利进行。