

豆瓣电影可视化数据格式

一、 电影

ID 电影名 评分 年份 类型 地区 简介

ID 是一个正整数，是电影的唯一标示

评分是一个正整数，0~10.0

年份是一个整数，如 2004，2012

类型为一个整数，这里大致可定义如下：

0	剧情	1	爱情
2	喜剧	3	犯罪
4	冒险	5	奇幻
6	动作	7	动画
8	悬疑	9	家庭
10	惊悚	11	科幻
12	战争	13	其他

地区可简单分为如下：

0	美国	1	英国
2	法国	3	日本
4	香港	5	德国
6	中国大陆	7	意大利
8	韩国	9	其他

电影名和简介都是一个字符串。

二、 用户

ID 用户名 看过 想看 影评

ID 是一个正整数，是用户的唯一标示

用户名是一个字符串

看过、想看和影评都是一个正整数，表示电影个数

三、 关系

ID1 ID2 评分 短评 影评

ID1 是用户 ID

ID2 是电影 ID

评分是用户对电影评分，0~5

短评是评分时给定的，影评是比较正式的影评

四、 数据使用

1. 根据用户喜爱对用户聚类。
2. 为用户推荐他喜欢的电影。