1. 爬虫爬取

学校网站地址为

http://www.sdcet.cn/xfw/channels/ch00387/，利用python爬虫技术，将图中标红的标题抓取下来，并写入到title.txt文件中。

将上述标题中的正文提取出来，保存到title\_text.txt文件中

将上边提取出来的数据利用python清洗出以“条例”结尾的标题【就抓取第1页】

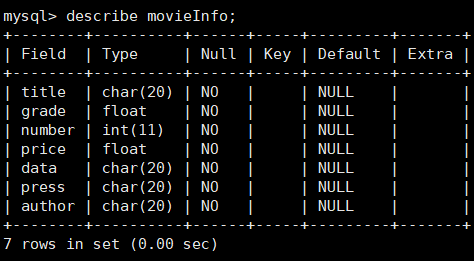


（1）爬取<https://book.douban.com/top250?icn=index-book250-all>网站的图书信息，主要包括图书名（只需要爬取中文名称即可）、作者、出版社、出版日期、价格、评分及评价人数，将所爬取的输入保存为文本文件。

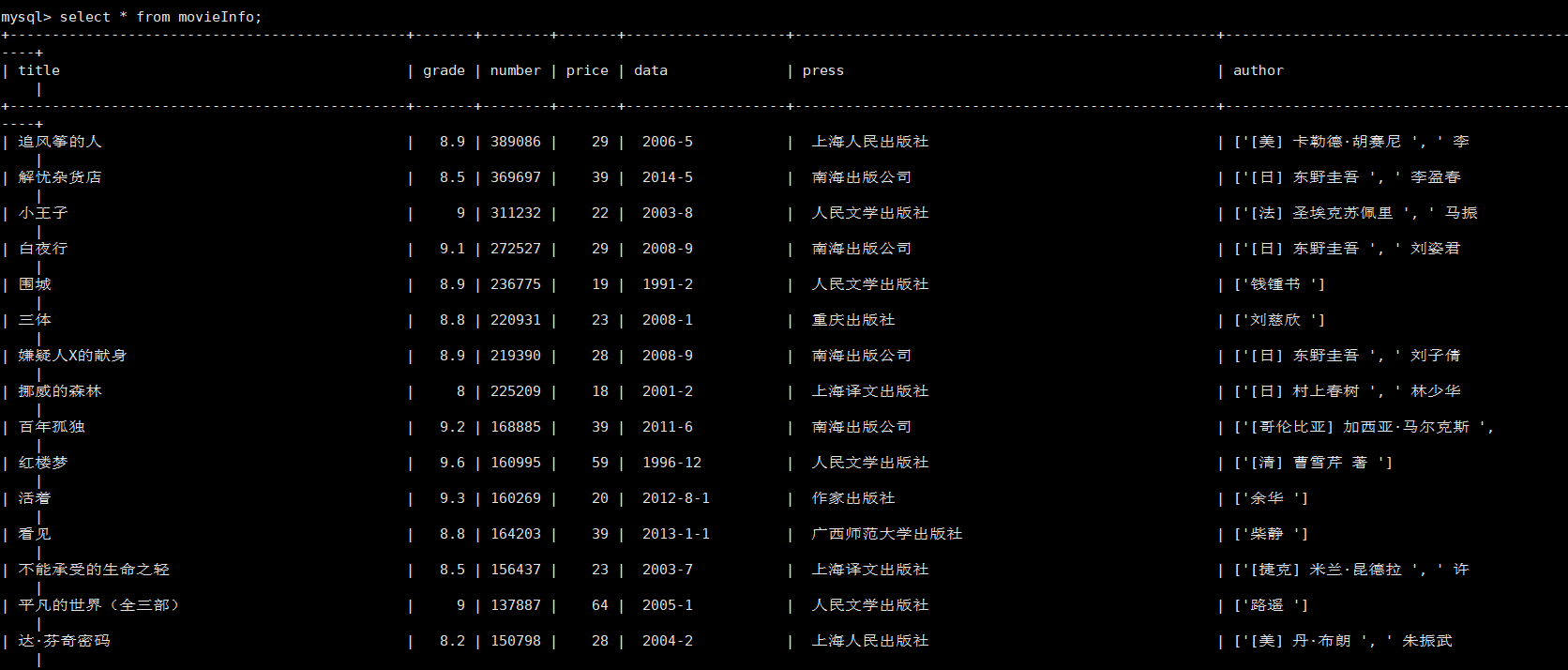


（2）将（1）中爬取结果保存到mysql数据库中。

数据库格式设置：



保存结果：



（3）对（1）中的数据进行处理，求出样本网页的图书平均分，结果保留两位小数。

1. 网址[http://theater.xtime.com/ Beijing/](http://theater.xtime.com/China_Beijing/) 有北京市场电影放映信息，网页样本保存于www1目录中，以获取市场正在放映的电影名称、电影类型、电影评分、上映场次并写入到ans0101.csv文件中，不要索引。代码文件名为：ans0101.py。
2. 网址<http://bor8848.com/> 中有电影市场票房信息，网页样本保存于www2目录中，分析样本文件，综合利用正则表达式和字符串处理算法，截取票房概况表格信息，写入到ans0102.csv文件中，不要索引。并求出实时票房总额，写入到ans0102.txt文件中。