本周主要任务：

熟悉构建搜索引擎工具Lemur

了解检索模型

师兄建议使用Lucene工具，Lucene的文档资料丰富，更新较快，易于二次开发

已完成：

1、Lemur环境搭建 转Lucene

2、Lucene环境搭建

Lucene 虽然有python版本但版本较老，没能装成功，文档资料较少。师兄建议用java版本的，版本新且文档资料丰富，本身也是java开发的。

3、现已熟悉Lucene的demo实现机制

4、Lucene的检索模型 是布尔模型和向量空间模型结合。了解它的评分算法，可以进行修改自定义评分机制（利于在检索的结果中找到更合适的结果）

5、对之前豆瓣，猫眼，微博的数据进行处理，形成文档目录

6、编写对数据文档目录建立索引的代码，正在编写检索的代码

后面的问题：

1、检索的结果是词或句子所在的文档，通过Lucene可以得到在文档中的位置和出现的频次，但没有返回相关的句子。这个需要进一步了解之后进行处理。

2、根据数据建立的文档，未放弃主要数据，文档包含电影信息及电影评论信息。考虑到问题的不同，对应的文档不同，是否是在一些特定文档上加权值（已了解如何加权，有利于结果排序）。