本周主要完成：

1、构建索引及检索代码的实现（Java）

2、Lucene评分算法：BM25算法（现为lucene默认）,布尔模型，向量空间模型（lucene以前用的）；语言模型有：狄利克雷语言模型和线性插值语言模型

3、数据库中数据转成能建索引的文档(lucene 检索的结果不能直接获得在文档中的句子，因此一条短文本一个文档，检索出文件路径读取对应的文本)

4、微信与Lucene检索进行连接（用Python3调用java的类和方法）

遇到的问题：

1、向量空间模型lucene源码中没有代码实现，只有类和方法的声明。（查看源码，可以通过自定义类来继承，也就是可以自己实现）

2、简单测试检索的性能，文档在14万左右，建立索引时间在1分钟（之前处理的数据是文档少，文档的内容多，建立索引时间在几秒的时间，但不能准确获得检索得到的句子）

3、通过简单的检索测试，BM25检索的结果较准确，排名的评分在数据上有明显的差别。

4、Python调用java，使用的工具Pjnius。尝试了Jpype,PJ4等。因为是通过c\c++转的，环境很不好弄，windows下总出问题。现只在自己电脑的ubantu上配置好了

下一步：

1、暂时以检索排第一的结果，作为微信机器人的回答结果；重复的问题，回答的答案是不变的，这个需要改变。

2、检索模型中，目前就BM25的检索结果最好，只看结果其他的与其相比就又明显的差距。向量空间模型尝试一下自己实现

3、lucene的模型分词是按一个字一个字进行分词的。尝试是否能自定义分词的方法（使用目前较好的分词库，jieba分词等）