**信息检索 第1次上次实验**

1.实验名称：使用Lucene根据关键字快速搜索

2.实验日期：2019年9月3号

3.实验介绍：对中文维基百科的每个词条数据建立索引（一个txt文件对应一个ID号），在安装好Lucene的基础上，能够快速根据搜索的关键字找到对应词条数据的ID号

4.实验环境：JDK，lucene-6.6.3，Python3.0以上/Java，python包PyLucene

5.实验目的：（1）能够针对特定数据生成索引文件；

（2）能理解Lucene中的Document-Field结构的数据建模过程

6.实验步骤：（1）Java环境变量配置

（2）Lucene环境配置

（3）根据提供的维基百科中文语料分成一个txt一个词条数据

（4）用多种分词工具（thulac、jieba、LTP）对语料进行分词处理

（5）对分好的词条数据建立倒排索引

（6）安装PyLucene包

（7）使用PyLucene创建上面提供好的倒排索引

（8）使用PyLucene进行多种查询

7.实验效果：（1）命令行输入java -version后传Java版本信息，说明安装成功

（2）进行一个词的查询，显示多种分词下的实验结果

（3）进行多个词的并集查询，显示多种分词下的实验结果

（4）进行多个词的交集查询，显示多种分词下的实验结果

8.实验感想：做完本次实验的感想