

完成了词偏移倒排索引，单个词查询，多个词查询，任意数目词组查询，临近查询，以及上述查询的混合查询。

并且获得了每一个查询到的词在文件中的字偏移以便打印出查询结果附近的语句。我是在提供的资料基础上修改的。

主要遇到的问题：

- 1、对于词偏移倒排索引比较容易解决，在分词时加入一个新的变量wordIndex，每扫描到一个词完成时就加一并与词一起存入元组即可。
- 2、对于短语查询，我使用“-”来区分短语和非短语。对于短语的查询就像视频里讲的那样，将词偏移与查询语句里的偏移做对比。
- 3、临近查询，我发现python中的集合是一个非常好的工具，用集合把满足的范围表示出来，再用另一个词的词偏移求交集即可，这样比原来比大小的办法好许多。
- 4、这次主要遇到的麻烦之一是打印查询结果附近的语句，因为查询基于词偏移，而打印附近的语句则意味着需要获得字偏移。为了达到这个效果需要同时储存字偏移和词偏移，在完成词偏移的查询后再通过词偏移导出字偏移。这样做效率很低且占了不少储存空间，不过我还没有想到别的办法。如果只是需要查找查询在哪个文档里就不需要费劲了，字偏移的部分可以全部舍去。
- 5、遇到的第二个麻烦是混合查询，这里混合查询达到的效果是：用空格分割的各个查询结果相与。如果结果里只有查询在哪个文档里就比较好解决了，但是结果里还包含了每个查询在每个文档里的字偏移，同一个文档里还可能不止一个，试了不少办法，本来想在每个查询内解决这个问题，但是并不成功。最后通过把所有查询结果放在一起并用集合处理文档标识，再将不需要的文档pop掉。