本次我们选取的文本是红楼梦，以章节为颗粒度。

大作业的目标是建立一个小型的搜索引擎。首先对红楼梦进行分词，主要思想是根据tf-idf模型，向量空间中的相似度计算出和询问相关度最高的若干个章节。在搜索返回结果页面，尝试匹配出章节中与询问关键词相关的句子，并以高亮表示。123123