1. 数据的获取问题

接受

解决方案：通过爬虫爬取一定数量的非结构化数据（文本、图像或音频）构造测试数据集供测试使用，初定数据规模在1W左右

1. 上层程序的测试问题，

接受

解决方案：

衡量标准：检索时间和检索准确性（准确率、召回率）

具体实施：（1）部署原始lucene架构，从测试集中选取测试数据，在原始架构上使用测试数据作为query，将检索结果以及运行时间作为groundtruth。（2）通过运行改进后的程序，使用相同的测试数据作为query，将检索结果以及运行时间与groundtruth进行对比从而获得改进效果。