

布尔查询

硬件&软件

Mibook Air i5-6200U + python 3.7.1

实验内容

本次实验要求实现一个简单的布尔查询系统，数据来源为某年3w+条tweets，要求根据tweets内容建立postinglist并实现简单的布尔查询，两项的 AND 和 OR

实验步骤

- 实验数据有200多条重复的，先去重
- 对于每条tweet，使用正则表达式整理出其中的单词
- 根据整理的单词构建 termId -> docId
- 查询语义的实现

实验中遇到的问题

每次运行都要从文档中存取数据。每次启动都要几秒钟。

语义分析的时候少写了一个赋值，每个句子只返回第一个单词对应的tweets。。。

其他

感觉，可以使用redis?。。好像Map-Reduce也可以用redis做= =，不太了解，有机会看看