近似抽取

2011011258　　计 13　谭志鹏

一、　实验目的

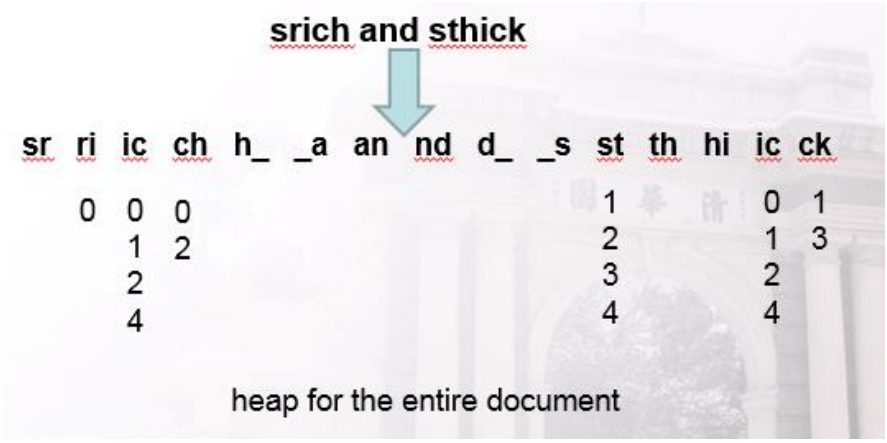
实现数据库近似抽取算法

二、　实验算法

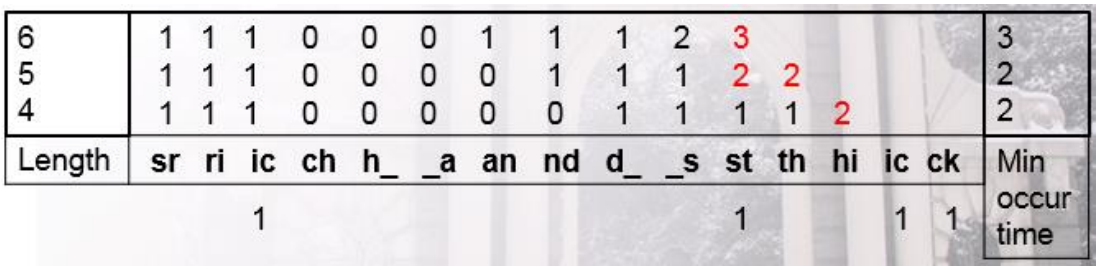
使用课件上介绍的基于单堆的抽取算法，具体步骤如下：

1、 为字典建立倒排列表

2、 对于需要进行抽取的文件，依据 q-gram 建立一个向量，并且对于其中在倒排表中存在的 gram 下面挂上相应的序列。



3、 每次找出其中最小的序列，然后使用一组向量来计算相同 gram 数，如果大于阈值则接受为 candidate，进行最终的验证。具体的可以只用一个向量，每次从下到上迭代计算，可以节省存储空间。



三、　实验总结

有了前两次的小作业，这次的实验实现起来的难度不大。不过实验中一开始对于约定的输入接口产生误解，使得一开始一直无法通过评测。不过在助教的热情帮助下，及时的找到了问题，顺利的通过了 OJ 评测。在这里要在此感谢助教了。