

# 物联网信息处理实验五

## 实验要求：

使用自己所擅长的程序语言编写布隆过滤算法(近似判断一个查询数据是否属于集合  $S$ )。

## 实验步骤：

- 1: 以 stream\_for\_bm.txt 文件的内容做为集合  $S$  的元素(元素为  $1-10^7$ ) (共  $10^4$  个元素)，构建一个比特数组将  $S$  的元素存储下来；
- 2: 读取 stream\_for\_query.txt 文件中的内容，对于每个数据判断其是否属于集合  $S$ ；
- 3: 精确算法，使用一个数组记录下  $S$  中的内容，然后对于每一个查询元素(stream\_for\_query.txt 中的内容)，判断其是否属于  $S$ ；
- 4: 设计并实现书上 107 页（第二版）的布隆过滤器，来近似查询每个元素是否属于  $S$ ；
- 5: 使用 10000 个比特，每个元素使用 1 个哈希函数，计算此时布隆过滤器的错误率（误是错误率）；
- 6: 使用 50000 个比特，每个元素使用 4 个哈希函数，计算此时布隆过滤器的错误率（误是错误率）；

## 实验考察要求：

- 1: 分析并讲解自己所编写程序；
- 2: 分析第 5 步与第 6 步，误差有区别的原因；
- 3: 不允许组队，如果在考察中，发现两人的程序完全相同，则两人本次的实验成绩为 0；