

## Project-3 CBIR

**提交方式：**邮件发送到 [20110240015@fudan.edu.cn](mailto:20110240015@fudan.edu.cn)

**邮件主题：**高级大数据解析 Project-3

**附件：**姓名+学号.pdf

报告中只需包含题号和相应的代码，以及运行结果

**Deadline: May-15 5:00pm**

### **Task :**

1. 自己实现 LSH 算法, 并实现三种不同的距离度量所对应的哈希函数。(euclidean distance, Hamming distance, Cosine distance, Jaccard distance )
2. 假设此时的数据是 numpy 随机生成的 1 万个维度为 48 的数据(random seed=4), 有一个数据 query (random seed=10), 利用上述的 LSH 算法找到最接近的 50 个向量。计算这些向量和 query 之间的距离。（距离按照上述实现的三种不同的距离度量方式）
3. 对 E2 求得得 50 个向量以及 query 利用合适的方式可视化结果