

作业 2

- 1、 题目：检索及评估系统
- 2、 目的：检索及评估系统是信息检索中不可缺少的组成部分，检索的质量直接影响到系统的性能。主要包括评价文档和用户查询相关性的匹配策略、查询结果的排序方法、用户相关度反馈和系统评估方法。这个作业的目的是使大家对各种检索匹配算法、系统评估方法有更加充分的理解，并发现它们中所蕴含的重点、难点问题及当前的解决方法。同时是对大家所完成的作业 1 的的质量的检验。如果索引质量差，那么检索结果不会令人满意的。
- 3、 资源
文档集，问题集，答案集
- 4、 任务：
 - 1、 query 处理（与文档处理采用相同方法：切词，选择关键词等）。
 - 2、 用 SVM 方法，计算一个查询与一个文档集合中各文档的相似度，并排序输出结果（对问题 10 和 20 输出详细结果，参见 Result（Vector），其他问题只记录文档排序即可）。
 - 3、 应用查询扩展方法对 query 处理后，对同一问题再次进行检索，并排序输出结果（对问题 10 和 20 输出详细结果，参见 Result（Vector），其他问题只记录文档排序即可）。
 - 4、 应用 MRR (Mean Reciprocal Rank, 参见第一章)方法，计算你所建立的系统的 MRR 值。
- 5、 要求：
程序：结构清楚，代码优化，注释清晰
报告：见项目报告格式及要求
编程环境：任选
- 6、 提交内容及方式：提交源程序、检索结果(query10 和 query20) 及报告。时间是 2015 年 1 月 5 号。
- 7、 参考资料：
英文文档的检索结果 [Example-DocIndex\(tf\)](#) , [Example-DocIndex\(tfidf\)](#)
[Example-DocInvert\(tf\)](#), [Example-DocInvert\(tf*idf\)](#), [Result\(Vector\)](#)