

# 2022年夏\_信息检索大程要求

---

## 大程具体要求

1. 要求实现 **倒排索引**以及 **向量空间模型**。
2. 同时支持 **布尔查询**、**通配查询**、**拼写矫正**、**短语查询**、**同义词扩展**、**基于快速评分的Top K查询**、**索引压缩**、和**词典索引** 功能。(按照添加功能的强弱适当加分)
3. 分小组完成，每组最多4人，分组情况见课程网页。（如有疑问，请及时致信助教[11921034@zju.edu.cn](mailto:11921034@zju.edu.cn)）
4. 用C/C++/Java/Python编程，不要求图形化界面（控制台程序即可）。
5. 对于大程中要求的功能，不能直接调用开源代码中的API（如Lucene等）。
6. 大程使用的Reuters语料库为：[点击下载](#)
7. 组内成员得分相同，希望大家能均匀分工，合力完成。
8. 写一个文档说明，写明实现了哪些功能、程序用法以及大程分工情况。
9. 最后一节课将进行大程的课程展示，请小组成员做好相关的展示材料提前发给助教。
10. 具体的代码和文件请在课程结束后的一周内提交至课程网站。