**《信息检索导论》大作业**

2017年秋季

**大作业** (1-7个选课学生组成1队)，完成以下任务(12月份末考查)：

新闻及评论搜索：定向采集不少于4个中文社会新闻网站或频道，实现这些网站新闻信息及评论信息的自动爬取、抽取、索引和检索。

1. 新闻网页数目不少于10万页。每个新闻网页及其评论能在1天内更新。
2. 支持关键词检索及通配符检索。检索时间平均不超过2秒。
3. 能按相关度、时间、热度(需要自己定义)等属性对检索结果进行排序。能对评论的贬褒进行分析。
4. 具备查询自动补齐、相关搜索推荐、snippet生成、结果预览(鼠标移到相关结果，能预览)等功能。
5. 每条检索结果下面可以对相似新闻进行查找。
6. 首页中列举当前最热的社会新闻。

**提交内容：**

1. 设计文档1份(内容包含并不仅仅限于设计方案、自我测试情况、创新点、经验总结等等)；
2. 汇报PPT 1份；
3. 程序源码；
4. 能够支持运行的部分语料(由于语料较大，暂不需要提交全部语料)。

**考查方法：**

现场或远程运行并测试系统，提问，打分。

**说明：**

不反对使用开源软件。但在设计文档中需要说明。

**评分标准：**

1. 功能部分满分60分，每缺少一项功能扣5分，最多扣60分。
2. 性能部分满分20分。
3. 材料部分满分10分。
4. 界面部分满分10分。
5. 创新性方面提供10分加分，总分加满100分为止。