Lab 1: Topic-wise Web crawler: 话题

- 互联网垂直话题爬虫
  - 话题：打折/优惠/羊毛、女神男神、留学信息、英文写作技巧、壁纸图片视频音乐、创意、头条热点新闻、笑话趣闻逗13 、八卦、流行美妆服饰、生活窍门
  - 全互联网爬虫
    - 源网页格式较多
    - 需要注册、认证；反爬虫
    - 爬虫能爬到的帖子有限
    - 难度较大
  - BBS、贴吧、天涯、微博、人人格式转换工具
    - 将看到的好的帖子链接、标签、内容保存
    - 利用javascript等技术
    - 类似于人人改造器
  - 微信公众号热门文章抓取工具
    - 热门指标？点赞数 or 阅读量
    - 利用javascript等技术
    - 输出格式的展示、整合

Lab 1: Topic-wise Web crawler：展示

- 自建服务器、展示网页、推荐排序
  - 推荐展示
    - 优先级排序
  - 随机展示
    - 随机，但是刷新之后不能重复

Project requirement

- 组队：每队不超过**4**人
  - 每个具体topic，先选先得
  - 之后由教师分配
- **2015年3月25日0:01AM提交**（邮件发送给我）
- 当日下午，给助教展示
- 提交完整可以运行的代码、文档、**demo**网页
  - HTML注意和手机端的适配
- 评分标准: **internal review + public review**
  - 基本分：浏览器适配，手机适配，助教评分
  - 扩展分：推荐排序是否有道理、影响力（PV、转发数）