

过程系统工程实验室

学术分享会之

——导师不会教你的小事

百度质量流量控制部

暑期实习分享

季一丁

2015/11/19

OUTLINE

- 项目组
 - 工作内容
 - 人员组成
- 实习项目内容
 - 项目介绍
 - 项目开发过程
- 同事和业余生活
 - 公司活动
 - 自己的生活
- 自己的一些经历和感受
 - 求职
 - 科研
 - 工作

项目组

- 百度盈利模式
 - 广告：广告主购买广告（展现率/点击率/关键词）
 - 流量：流量联盟导流，赚取流量分成/打折流量
- 流量作弊
 - 点击：竞争对手恶意攻击，赚取不当分成（与钱相关）
 - 检索：提高关键词排名，刷百度热搜，刷搜索框suggestion
- 流量质量控制部(Anti)
 - 前商务搜索部(Ecom)流量反作弊团队
 - 通过检索、点击日志进行作弊流量过滤
 - 保护广告主利益，为下游提供高质量流量

项目组

- 人员构成
 - 策略：针对线上作弊情况，研发流量反作弊算法
 - 工程：开发策略的支撑平台，内部工具，开发对外流量产品
- 工程师等级



异常refer离线挖掘

季一丁|Anti-Intern

CONTENT

- 背景分析
- 数据调研
- 简单策略
 - 代码内关键字匹配
- 模型策略
 - 提取特征训练模型
- 网页抓取策略
 - 内部平台支持
- 总结

互联网同事和生活

- 工作
 - 无固定上下班时间（10:00~21:00）
 - 每周部门例会、策略方向站会
 - 随时处理客户投诉，系统报警
- 80后/90后同事
 - 工作
 - 业余生活
- 生活
 - 周末不加班
 - 需要学习的东西还是很多
 - 北漂的压力



求职经历

- 研一准备简历，学习语言
 - 简历：不断完善
 - 考试和科研打好基础
- 研二看数据结构和机器学习相关资料，关注求职动态
- 研二下学期投暑期实习简历
 - 4~5月份，投了10余家公司，拿到百度和中行数据中心offer
 - 面试主要问实验室项目经历，简单的算法题目，语言基础，智力题，顺便考察非技术方面的综合能力
- 暑期实习
 - 学习业务流程
 - 学习linux操作
 - 学习shell和python编程

求职相关资料

- 语言类
 - Java编程思想
 - Shell编程指南
 - Python学习手册
- 算法和数据结构类
 - 数据结构与算法
 - 普林斯顿大学“算法”公开课 by Robert Sedgewick
 - 程序员面试金典 `cracking the code interview`
 - 刷题网站`leetcode`等
- 机器学习类
 - Andrew NG机器学习公开课（两个版本）
 - 统计学习方法
 - 机器学习实战
 - Hadoop实战
- 以及各种`csdn`博客，知乎问题，等等

一些感受

- 对工作，对毕业，对未来都应该早作打算
- 科研和工作并不冲突
- 工作以后依然是一个成长的阶段，保持学习的状态

一些建议

- 信心很重要
 - “只要你想去这个岗位，现在开始准备，你一定可以的。”
- 对待困难要有积极的态度
 - “问题解决不了是好事，代表你又要学到很多新的知识。”
- 核心竞争力是学习能力，最好可以用事实向招聘你的人证明
 - “我说我有学习能力，实习项目装在手机里一展示，面试官就明白了。”

祝各位科研顺利~

谢谢

季一丁

2015/11/19