

个人简历

基本信息

姓名：韩侃侃

性别：男

手机：18713593730

邮箱：827435393@qq.com

英语水平：CET-4

就读学校：燕山大学

专 业：计算机科学与技术专业



教育背景

2014.9 -至今 燕山大学 本科

所获奖励

2014-2017 获得奖学金多次

2016 年 8 月 大学生创新训练计划项目优秀结题

专业技能

- 1) 熟练掌握 C 语言、C++以及 Python 三门编程语言，
- 2) 熟练掌握数据结构和算法；
- 3) 熟练使用 django 框架和 scrapy 爬虫；
- 4) 掌握分布式架构 Hadoop，多线程并行数据抓取；
- 5) 灵活运用 Redis，MongoDB 及 MySQL 进行数据混合存储策略存取；
- 6) 熟练使用 Linux，并在 deepin 环境下基于 nginx 搭建在线个人网站（www.ysuhankan.top 域名已备案）。
- 7) 擅长互联网数据的采集、集成及分析处理工作；

社会实践

2015 年 9 月至 2016 年 6 月 担任燕山大学话剧团副团长；

2015 年 6 月至 8 月 在石家庄苏宁国美门店当门徒；

2016 年 11 月至 2017 年 1 月 采集某二手网站百万数据进行整理和分析，找出与计算机专业相关商品，运用所学算法进行销量排名与统计，之后搭建个人微信公众号，QQ 专属群，利用信息不对称性，挣得人名币 **5 万**。

项目 1：2015/7-2016/8：智行题库教育系统

项目背景：智行题库系统旨在为老师找题、出题、组卷，学生找题、做题、模拟练习提供便捷实用、高效精准的教育互联网开放平台。

开发部署环境：阿里云服务器，Windows 7.0，Ubuntu 12.04，Python3.5，MongoDB，MySQL，PyCharm。

数据成果：1、部署上线智行在线题库系统；

2、采用多线程和分布式的方法抓取试题量 **10 万**。

任务描述：

- 1) 抓取某度文库，设计并实现“反爬虫策略”应对及破解机制，建立代理连接池，分析网站防爬策略，通过配置数据头 header，伪装浏览器，切换代理 ip，更变抓取间隔时间，以及失败尝试次数来应对“反爬虫策略”，并部署在多台服务器上，采用分布式方法进一步提升效率；
- 2) 基于 **Bootstrap** 搭建友好前端界面，基于 **Django** 实现学生注册，登录等功能，配合使用轻量级搜索引擎 **Elasticsearch** 实现查题找题功能；
- 3) 在 **Ubuntu 12.04** 环境下基于 **nginx+uwsgi** 部署上线题库系统；
- 4) 作为数据抓取负责人，组织成员对不同网站进行调研，并多次对组员进行有关数据源解析，数据存储，分布式爬虫和网络反爬虫机制的培训，带领组员抓取上百万数据。

项目 2：2016/11-2017/1：互联网大数据采集及分析

项目背景：以所学专业为基础，探索致富之路。

开发部署环境：Windows 10，Python3.5，MongoDB，MySQL，PyCharm。

数据成果：采用多线程的方法抓取二手网站信息超 **100 万**，通过整理分析有用信息最终获利 **人民币 5 万**。

任务描述：

- 1) 设计并实现“反爬虫策略”应对及破解机制，建立代理连接池，分析网站防爬策略，通过配置每个数据源的不同代理切换，更变抓取间隔时间及失败尝试次数来应对“反爬虫策略”，并采用多进程多线程方法进一步提升效率；
- 2) 实现增量式爬虫，检索条件设置成时间，并抽取出最新日期商品，通过 Web 工程的定时器调度，定时执行日更程序，实现自动日增增量采集；
- 3) 实现商品数据的去重工作，基于 Bloom Filter 方法高效处理数据，选取非关系型数据库 **Mogondb** 存储海量数据，基于 Python 库函数 **highcharts** 实现数据图形化展示。

- 4) 作为项目唯一负责人，独自一人完成此同一类型网站的调研与分析工作，在项目实施的路上遇到的问题逐一攻克。

项目 3：2017/7-2017/8：海同公司 HTML5 项目开发

项目背景：以企业为目标用户，将数据资源和互联网资源集成到一个信息管理平台之上，开发门户网站。

任务描述：

- 1) 作为项目组的组长，对项目进行整体规划和部署，并带领整个团队共同开发网站，主要开发前台和交互部分；
- 2) 综合利用 HTML5, CSS3, JS 实现网页布局美化，实现登陆，动画特效、轮播图、放大镜、商品评分、购物车等功能。

项目成果：实习期间，带领团队实现门户网站，电商网站和音乐播放器的开发。

项目 4：2017/9：学校学生信息的安全性研究

项目背景：学校关于学生信息处理不妥当，学生本身对自身信息不重视。

项目成果：1、获取燕山大学在校学生 **上万名** 学生隐私信息。

2、获得校方积极反馈，就此问题与校方一起协商解决。

任务描述：

- 1) 发现学校就学生账号信息存在严重信息泄露问题；
- 2) 为引起校方重视，利用此 BUG，写程序收集相关证据；
- 3) 待所有信息收集完毕，以发邮件方式取得校方高度重视。

自我评价

- 1) 极其爱好编程，具有扎实的互联网分布式信息采集分技术基础；
- 2) 逻辑思维能力强，曾获国家级数学奥赛三等奖；
- 3) 团队合作能力强，实习和课设过程中多次担任项目组组长；
- 4) 责任心强，按时保质保量完成任务；
- 5) 热爱生活，乐观积极。