

# 侯捷

侯捷 | (+86) 131-7624-4325 | [hj.13.new@gmail.com](mailto:hj.13.new@gmail.com) | 哈尔滨工业大学 | 信息安全(本科) | 后台开发工程师 | 数据研发工程师

## 基本信息

- 侯捷 男 本科 信息安全
- 手机 : 13176244325
- Emial : [hj.13.new@gmail.com](mailto:hj.13.new@gmail.com) / [z.g.13@163.com](mailto:z.g.13@163.com)
- 技术博客 : [houjie13.com](http://houjie13.com)
- Github : [h-j-13](https://github.com/h-j-13)

## 实习经历

### 腾讯 TEG(技术工程事业群)

2018.6 — 2018.8  
信息安全部 用户策略组

前期基于faiss开发了亿级海量高纬向量相似检索模块，学习scala使用spark进行亿级用户关系链挖掘及社群划分  
后期参与部门实习生mini项目，担任PM及内容安全部分的开发工作；协调开发进度，设计实现了一整套保证项目内容安全的子系统

### 哈尔滨工业大学 网络与信息安全技术研究中心

2015.10 — 2018.6  
域名体系安全技术研究室 恶意域名组

作为研究室核心成员，主要负责域名安全相关工作，包括域名基础数据获取、恶意域名挖掘、关联分析、基础数据统计分析； 另外，负责研究室数据库的维护、设计与优化等工作；作为技术负责人带领团队进行项目开发，向研究室主任负责

## 教育经历

### 哈尔滨工业大学

2015.9 — 2019.6  
信息安全 (本科)

## 项目经历

### 方圆

2018.7 — 2018.8  
腾讯实习生mini项目

个人负责部分及特性：  
PM，系统架构设计，内容安全

- 基于Nginx负载均衡结合多Tornado实例的异步后端框架
- 基于DFA实现的高效文本过滤和多模型文本分类的文本审核模块
- 基于主动色情图片识别和多级高效相似图片检索实现的视频审核模块
- 基于多维数据的用户特征分析，结合决策树判定用户性质，动态生成打击策略

### 非法域名监测系统

2017.9 — 2018.4  
( 此项目为研究室承接国家某部某中心的重点研发计划)  
旨在对于非法域名进行监测、关联分析与数据挖掘

个人负责部分及特性：  
系统架构、数据库架构、域名WHOIS数据获取与处理、数据中间层与后台开发

- 基于水平、垂直拆分的MySQL数据库与同步ElasticSearch服务器作为存高性能储架构
- 基于tornado构建的HTTPS的REST api数据中间层结合基于token的无状态身份认证机制

- 海量数据的关联分析，整体情况的统计分析与整体趋势态势感知

系统目前收录及监控了约**4463万**条域名及关联数据,最终以**优异表现通过验收**.

## 非法域名挖掘与画像系统

2016.12 — 2017.07

第十届全国大学生信息安全竞赛 作品赛 作品

个人负责部分及特性：

PM，系统架构设计，数据获取与分析，后台与部分机器学习

基于域名与页面关联关系的非法域名分析与挖掘系统，旨在通过域名多元信息对域名进行精准画像，进而分析抽象获取非法域名相关特征，使用基于决策树的机器学习模型进行域名性质判定，通过多过程多方式进行恶意域名挖掘

- 采集与挖掘的域名相关信息包括：域名WHOIS/WHOWAS信息，IP、ICP、地理位置、CDN情况、页面解析情况、页面关联关系
- 通过WHOIS注册信息、站点间连接关系及关键词搜索进行数据反查与挖掘
- 通过基尼指数与决策树对域名进行非法可疑值量化与性质判别

总决赛 **全国三等奖**

## 域名WHOIS数据获取系统

2015.11 - 2018.3

旨在尽可能多的获取域名及域名WHOIS信息

个人负责部分及特性：

基础数据获取，WHOIS数据获取，数据库维护优化，网站后台

- 基于AMQP协议下的RabbitMQ框架通过Python实现了一个高性能、高效的分布式域名WHOIS数据获取引擎
- 通过底层SOCKS代理、核心数据多源轮询更新、基于RFC 3912 协议等方式深度优化了WHOIS获取过程，**核心代码完全原创**，不依赖任何第三方库及第三方网站数据
- 采用动态解析模板，可将形式各异的WHOIS数据解析出统一的特征项 并基于文本分类，编辑距离等算法对地理位置和电话、邮箱等相关字段进行了深度解析
- 采用水平、垂直拆分的MySQL数据库设计；多级缓存的设计，保证了数据的高可用性
- 额外的基于日志的检测系统可以获取系统详细情况、监测系统异常变动并即使发送邮件提醒、以便日常维护与开发

系统获取了超过**1.2亿的域名数据**（超过全球域名总量的1/3）与**1.5亿条域名WHOIS\WHOWAS数据**，覆盖超过1500个顶级域，1800个WHOIS服务器 数据支撑实验室其他相关项目.

## 专业技能

熟练掌握Python，有部分并发、分布式、数据库、网络编程、爬虫和机器学习相关开发经验

有丰富的Linux使用经验，熟练掌握git、vim、linux基本命令与配置经验，熟悉shell脚本

有tornado，flask(Python WEB框架)后台及RESTful风格API开发经验，了解简单的HTML与JS

熟悉常用的算法、数据结构，熟悉C++，有简单的qt及socket开发经验

本科期间维护与优化实验室核心数据库，对MySQL数据库略有理解与心得及较多的实践及优化经历 有简单的分布式MySQL搭建与维护 and ElasticSearch、Redis开发经历

## 自我描述

域名安全、后台开发、数据处理与分析领域经验比较丰富，对于新技术及开源社区兴趣浓厚，[github](#)和[个人博客](#)更新活跃，有过多次带领团队与主导项目开发的经历

自学能力较强，踏实肯干，抗压能力强，技术尚可，追求优雅的代码，有较好的数学、英语基础 希望在技术氛围浓厚的地方工作.

## 期望工作

后台开发工程师 / Python开发工程师 / 数据研发工程师