

方楠

基本信息 Basic information

籍贯：黑龙江哈尔滨

年龄：29 岁 (1995 年)

邮箱：16607557430@163.com

联系方式：16607557430

户口：深圳市南山区

毕业学校：南华大学核物理系

期望地区：广州/深圳

目前状况：离职

期望职业：爬虫工程师、后端工程师、全栈工程师

期望薪资：面议



受限于篇幅，个人简历网站有更详细描述和支撑材料 <https://xiwen.cpolar.cn/resume> (强烈推荐)

只接受 python 岗位，外包不考虑，谢谢

自我介绍 Self-evaluation

具有 8 年爬虫/后端/全栈工程师的工作经历。具有资深技术研发、商务洽谈、领导团队的能力，领导过大型团队且有成功的创业经历。**但扪心自问，还是最喜欢写代码的快感！**

技术专长

深度使用 Python 进行爬虫、后端开发

爬虫方向

- 主持 JS 逆向培训课程，了解 JS 逆向攻坚检测如 debugger/原型链/浏览器指纹/node 检测/AST 混淆 (平坦流大数组大对象) 反解/jsvmp，熟练使用浏览器 proxy 进行扣代码/补环境，有自己的一套常用的 Nodejs 补环境套餐，熟练使用油猴插件
- 深度理解 Scrapy 框架及原理，包括其 Middleware、Scheduler、Downloader。对 scrapy 源码有非常深入的了解，例如使用 asyncio 重构 Twisted 从而使其有更好的请求性能
- 熟悉 Selenium 的三大组件 webdriver/IDE/Grid，深入阅读过 webdriver-python 底层源码，熟练大规模使用 WebDriver-Remote 进行代码、下载器分离的云部署和本地部署
- 熟悉 Appium 框架，除了本地化机房搭建，还进行进行过猫池，云手机采集部署
- 认识众多数据供应商和黑科技，有稳定可靠的 IP 代理、虚拟号手机注册、邮箱账号供应商
- 自己的公司有自建数据中心(详见网页版简历支撑材料)，电商/店铺/客群/新闻数据可以直接有偿接入

后端方向

- 在小步公司，基于 Django 从 0 开发了完整的核心产品 Xworld，一人独立完成 70% 的接口开发，支持千级并发抢票活动、越南盾结算、智能合约。熟练使用 swagger 管理接口，熟练使用 django-admin 并进行二次开发
- 根据业务需求灵活使用设计模式进行框架开发。如工厂模式、装饰器模式、代理模式、观察者模式。设计的架构有前瞻性，能支撑产品经理频繁的横纵向拓展需求，不必大改且保持稳定
- 善于进行 Redis、MongoDB、MySQL 数据库瓶颈分析、资源估算和数据库调优，基于数仓思想的数据库设计与调优
- 数据可视化、监控和部署方案搭建，如 Prometheus、Grafana。熟练使用 docker 部署。

能力专长

- 能够以项目经理或技术总监的身份搭建和管理技术团队，作为引领者能够筛选团队人员，把握项目方向、进行项目控制、对项目结果负责。创业经历详见项目经验，**营收 1 亿元，纯利润百万级**。在探迹公司，实施**日入亿级**的采集系统，管理采集源 **5606** 个(详见简历网站支撑材料)，几乎涵盖了所有互联网上的公开数据。以 PM 身份管理技术开发团队 14 人(4 爬虫+1 前端+2 后端+1 大数据+1 算法+2 实习生+2 标注师)
- 技术栈相对自己来说广而不精，具有爬虫、后端、NLP 算法、前端、测试、产品多种技术栈，工作中往往担任橡皮泥岗，根据公司现有人员的情况查漏补缺，不局限于固定的岗位，对全新的技术充满了学习的热情
- 解决问题的过程中没有牛角尖精神，非常期待且擅长以整合、沟通、取巧方法快速绕过问题完成内外部需求。相信公司中的顽固大【锅】一定有巧妙的解决方法，并有多次以低成本解决公司顽固的产品、技术问题的经历



工作与项目经历 Work experience

2023.5—2024.10

广州探迹科技有限公司

资深 python 爬虫工程师

公司主要产品是卖拓客软件，入职时任平台架构资深爬虫工程师，后升职作为行业版的技术负责人 PM

工作职责

- **任职平台架构时**，对公司各项目组提供技术支持，负责**招投标数据、政策数据、联系方式数据采集攻坚**。创新性的研发了一套**智能采集配置器 AISpider**，能够对于长尾源（就是源数量很多但是单个采集难度较小的数据），能够**自动识别召回新源（无中生有）**，能够使用标题识别算法+列表翻页识别算法+文本密度正文识别器，**机器自动反写出配置模板需要的 Xpath 路径**，能够自动对采集结果进行**质控，得到其优先级，保证其准确性**。文档链接 <https://xiwen.cpolar.cn/resume/file/>

AISpider爬虫模板生成.pdf	29-Nov-2024 02:46	4M
AISpider盘源工具.pdf	29-Nov-2024 02:46	4M
AISpider质控平台.pdf	29-Nov-2024 02:47	3M
Thumbs.db	29-Nov-2024 02:24	12K
支撑材料-5606采集渠道截图.png	13-Nov-2024 03:57	121K
支撑材料-AISpider工作流程图.png	20-Nov-2024 06:24	109K
支撑材料-AISpider盘采集源277621.png	20-Nov-2024 06:45	68K
支撑材料-探迹服务器管理数.jpg	19-Sep-2024 07:18	16K
支撑材料-日入亿级采集系统截图.png	13-Nov-2024 03:55	105K
支撑材料-简洁周报.xlsx	20-Nov-2024 04:21	891K
支撑材料-采集系统侧架构图.png	23-Aug-2024 16:42	381K
万楠的行业版PM工作指北.pdf	20-Nov-2024 04:26	786K

访问链接后的截图

- [平台架构]招投标识别召回 **18098** 个源网站，采集或配置数据源网站 **10669** 个源网站，对标千里马反向对比指标达到 95%
- [平台架构]政府源识别召回 **113024** 个源网站，采集或配置数据源网站 **1996** 个源网站，政策数据完整度超越市面上所有竞品
- [平台架构]联系方式源采集 **15085** 个源网站(门户或黄页 **4529**/采购网站 **4750**/群 **2987**/招聘 **2819**)，联系方式数据完整度合规超越市面所有竞品
- [平台架构]管理标注师、实习生 **22** 人，从中简拔人才升职初级爬虫工程师 **2** 人
- **任职行业版 PM 时**，直接带领团队 14 人，背一亿年度业绩目标，解决产品版本迭代、框架优化、招聘人员管理、客诉维稳等工作
- [行业版 PM]建立数据导向的质控体系，量化管理商标、专利、建筑四库、化妆品、安许证、税务、待续办证书、百度地图、大众点评、品牌、拟在建 **11** 大主要行业版模块
- [行业版 PM]优化爬虫框架，落实责任到线+责任到人，在满足行业版采集量日均 1.2 亿(提升 55%+)的产品需求，将云服务器从 80+降低到 77 个实例
- [行业版 PM]实践重点源采集攻坚，百度地图(完整度 100%，脏数据比例低于 5%)、大众点评(单账号日采集量 1000+，包括 APP/微信小程序/支付宝小程序/H5 四源，维度包括新店标签和营业执照)

工作业绩

1. **公司超千人，研发总人数超过 200 人，资深/专家工程师 title 比例不超过 10%**
2. **入职三个月即提前转正，年终绩效评优 A(+2 月年终奖)且加薪 8%(有银行流水及公司证明，详见网页版简历)**
3. **行业版业绩目标一个亿，截至 7 月 30 日(7/12)完成 45.4%**

2021.12—2023.5

深圳小步网络科技有限公司

高级 python 工程师

公司主要产品是一款 Web3.0 资讯网赚产品《Xworld》

工作职责

- 负责 1200+渠道的资讯、短视频、新闻爬虫、google-play 游戏爬虫的研发工作

- 负责 Xworld 产品的后端工作，账号注册与登陆、越南盾法币充值、中心化钱包系统、去中心化钱包系统(部分)、游戏服务、租号系统、短信、push 推送、拉新裂变、反馈模块的研发工作
- 负责了 Appium 安卓自动化操作的研发工作

工作业绩

- 完成了整体产品从 0 到 1 的后端研发工作，占产品后端总接口数量的 70%。
- 产品研发成功后放量，支持了高峰期约 10w 日活的用户。期间解决了如支付系统、抢票系统等难题。在公司研发绩效考评中排名前 20%
- 工作第八个月时，公司通过了我幅度 30% 的加薪申请(有银行流水和加薪邮件证明，详见网页版简历)

2021.05—2021.11

(深圳分部)济南喜文信息科技有限公司

唯一法人

公司产品是一款面向就业培训重点人群的智能学习机

工作职责

- 核心研发团队 8 人，公司 高峰时总计 22 人，包括劳务派遣员工 50 人
- 国家会将五险一金中的失业险的结余金额，以职业技能培训补贴的形式层层下发，用以免费的职业技能培训。其中重点人群泛指偏远居住、待业农民、低学历人群等，重点人群的职业技能培训一直以来是行业痛点。传统解决方式是老师驱车至乡村，以“米油面”等物质诱惑让学生参加培训。实际操作中学生积极性不高、完课率很低。学校成本居高不下，政府加大预算投入效果也微乎其微，由于疫情，国家提倡线上+线下相结合的方式。但重点人群不会使用手机、学习积极性不高亟待解决。
- 我们结合 **android 开发、手机硬件开发、数据采集、自动化 RPA 操作技术**。使得学生开机即学习、在线考试，无需繁琐操作。完成学习后，手机会“解锁”成为一部手机，赠送给学生。大大提高了学生的报名积极性。
- 我们的“学习机模式”获得了教育局的资质认定和官方推广。使用此“智能学习机”产品在黑龙江省学校推广获得了成效，国家政策期合作多家学校**共计招生 11341 人，每位学生财政补贴 6240 元**(在疫情相对较弱后，补贴降低至 3120 元)。完课率达到 90.24%，我公司抽取财政补贴的 10%作为技术服务费，手机费用和教学费用由学校自行支付。
- 技术上实现了自动化 RPA 手机操作技术和手机锁技术，适用于市面上 300-1000 元 90%以上型号的手机。完善的管理后台，支持三千台手机心跳的同时并发处理。管理后台打通 PC/微信公众号，单老师可负责 200+学生的学习管理而游刃有余。数字化向上对接教育局、劳动部门，符合国家部分审核补贴款发放时尤为省工省力。实现了学生、学校、管理痛点的一箭三雕。

工作业绩

- 政府最终发放了培训补贴款，公司成功回款盈利超百万，疫情结束后项目结束

2020.08—2021.04

联易融数字科技集团有限公司

python 爬虫工程师

主要产品是服务于各家金融机构的供应链金融的 SaaS 管理软件

工作职责

- 负责现有 WebDriver 池的维护和升级。基于现有系统，借鉴 zookeeper 选举机制的思想，完成 WebDriver 池系统分发策略的优化。提高了分发效率和鲁棒性。此池服务器(虚拟机)320 台，带宽 1000M+
- 基于现有系统，增加调用优先级功能，从硬件隔离和调用隔离两个角度，使得资源分配更加合理。借鉴 nginx 的 ip-hash 的负载均衡思想，使得同一业务无需配置，总是调用业务环境适合的 driver 池和代理池，提高了采集业务的成功率
- 负责创新型项目《AIC 智能资讯采集》，用人认知“新闻网站”的思想办法(研发了搜索框识别算法、翻页识别算法、NLP 正文识别算法等进行协助)，**用 AI 代替爬虫研发代码。该项目只需填写资讯网站首页地址，不需要针对网站编写 css/xpath/re 就可以做到爬虫采集数据。**做到了 1 个模板、2 人维护、采集 1000 个网站的目标
- 《AIC 智能资讯采集》适配招投标信息和新闻结构信息，采集模板支持关键词搜索模板、深度优先模板和广度优先模板 3 个模板，准确率达到 90.4%，召回率 70%以上。上线截至离职采集去重新闻 3000w+(占同时期入库总量的 95%)，采集速度填满闲时带宽，丰富了公司舆情产品的语料水平

工作业绩

- 该项目属于创新型项目，当时并无先例可循，个人负责完成了这个项目提议立项、计划、研发、验收全过程。立足公司现有资源，完成了当时属于行业领先的指标和人效**
- 绩效成绩为 A(按制度应该是前 30%，实际情况对员工保密)，年终奖系数为+3 个月工资(有银行流水及公司证明，详见网页版简历)**

2018.10—2020.08

深圳市比一比网络科技有限公司

技术负责人

该公司是 ToB 的定制化开发公司，产品主要由甲方确定

工作职责

- 担任旅游局技术负责人及核心研发，利用 scrapy 框架采集途家、猪八戒等 6 家民宿网站信息，后结合百度地图 API 开发相应前后端场景，最终负责项目提前完成，节约工期 25%
- 担任银达项目技术负责人及核心研发，设计项目表结构和核心消费者模型，对项目中的核心问题：
 - 无法采集《建设通》数据渠道接手开发破解，最终解决了数据采集问题，完成回款
- 担任银泰保理智能机器人技术负责人及核心研发，利用 Webdriver 技术完成自动化《中登网》登录、完成《发票验真平台》的 JS 破解、完成 PDF-OCR 技术及验证码 OCR 识别，完成手机短信自动收发从而完成中登网自动登录。最终结合客户场景接入对方已有系统，做到业务流程 0 打扰，高效研发节约工期 50%，客户认可度高，项目结束后，对方老板致电我方老板对项目进行赞扬并后续继续合作。
- 在公司开拓新业务线，发起并担任资讯业务线负责人，完成对深圳报业、南方报业、羊城晚报、珠海日报、佛山报业的技术支撑拜访工作。资讯流产品形态、技术后台等设计工作。包括基于 appium+miniproxy 的手机通知采集系统、基于 scrapy+探索词+基于 dom 树密度正文提取+bloomFilter 去重的全量爬虫系统。基于 scrapy-redis 下拉流数据爬虫系统。基于采集视觉中国

与图像 hash 比对的版权图片识别系统。并成交(包括标书撰写、投标和讲标)深圳报业项目，最终如期完成。

工作业绩

1. 在公司工作的两年期间，连续两次升职(初级开发工程师->高级开发工程师->技术负责人)
2. 公司在我任职的前半年通过了我幅度为 83%的加薪申请(有银行流水及公司证明，详见网页版简历)

2016.12—2017.11 北京搜狐新媒体信息技术有限公司 算法实习生，中级 JAVA 爬虫工程师
主要产品是搜狐新闻客户端

工作职责:

- 负责渠道 ROI 模型的设计与搭建。
- ROI 模型通过对新投放渠道前 3 天、前 7 天、前 30 天数据三个区间段的数据变化情况，分析出该投放渠道的投放潜力、得到定量分析、识别掺量假量。
- 使得运营对投放的管理达到量化级别，避免了人工决策带来的主观误差。
- 在后台自动完成监控，实时预警报警。
- 参与《你身边的头条》其中政务资讯抓取部分，使用 WebDriver 补充地域性政务新闻内容。

工作业绩:

1. 通过数学建模竞赛国家二等奖特招进入搜狐，成功实习转正

求职意向及资料要求 Job Intention

- **不接受外包**
- **如有公司官网的，请告知公司官网。或发送一些公司的产品资料。**我将会认真学习贵公司发展历史和产品情况，准备贵公司相关的经历和资料以供精准匹配面试。
- **不接受更换研发语言，只接受 python 技术栈。**个人可以熟练阅读、维护 java、go、vue 等语言，对接重构都可以接受。方向可以是爬虫、后端、大数据等不限。

教育背景 Education

2013.9-2017.6

南华大学

核物理

本科