

杨洁坤

数据挖掘, PYTHON, C/C++

Download PDF

jiekunyang@gmail.com

151-9189-5258

个人信息

男, 1988年3月, 山东烟台人

技术博客

<http://www.MissPy.com>

主要技能

数据挖掘

熟悉爬虫的设计与实现, 了解自然语言分析、组群识别、数据可视化、图论等算法。拥有相关实际项目经验, 可以承担复杂分析任务。

科学计算

熟悉数学建模和主流算法, 熟悉 Matlab, Maple 等计算软件。能够利用计算机快速解决科学研究相关领域的问题, 尤其是信号处理、数据可视化等。

团队合作

拥有小型开发团队、学生社团管理经验。熟悉学习型组织、矩阵管理等管理理论。能够通过寻找双赢/共赢的方式建立高效、友好的合作关系, 提升效率。

技术技能

1. 编程语言。熟练掌握 python。能够使用 c/c++, javascript, shell, C# 完成日常工作。了解 Lisp。
2. 开发环境。以 linux 为主, 控 git / vim, 了解 gcc, make 等。能够使用 vs2010 / eclipse 完成日常工作。
3. 敏捷开发。熟练掌握 OOP, 熟悉设计模式、单元测试。熟悉 Python 代码的重构与性能优化。
4. 其他。熟悉 HTTP 协议, 了解 TCP。熟悉 MySQL, 了解 MongoDB。熟悉正则表达式。

工作经历

软通动力

2013.9-今

华为 OSS 存量

团队主要负责根据服务器端记录的数据, 评估、预测网络状况。针对已经存在或即将出现的问题, 给出解决方案和预期结果。预期结果主要包括预期收益和量化后的用户体验提升。

个人工作产出:

1. 代码性能优化。借鉴函数式编程的思想, 增加一个 Cache 层缓存计算过程。结合 Python 修饰器的语法, 可以实现原有代码的最小改动 (一般只要 1 行)。优化后, 运行时间降低 70%。
2. 优化项目已有 Python 代码。重构部分质量风险较高的模块。代码行数降至 30%, 平均圈复杂度由 20+ 降至 4。
3. 通信基站群覆盖范围可视化展示。根据一组散点的经纬度与覆盖半径, 计算最小覆盖边界并输出至 Mif / KML 文件, 可分别导入 MapInfo / Google map 查看最小覆盖区域。负责算法+编码实现。

中国(南京)未来网络

2013.4-2013.7

SDN 网络控制台-开发

1. 网络拓扑结构的可视化展示。
2. 基于 libvirt (虚拟化工具的函数库) 的二次开发。

易思博

2011.11-2012.9

华为 SDP (业务部署与运营平台)-交付组

1. 设计并实现 Linux 性能监控与可视化显示脚本。针对性能测试的关注点, 探索出新的滤波算法。通过该工具成功定位出长期遗留的隐性 bug, 有效提升了产品的稳定性。
2. 基于有限状态机和宽搜算法, 建立用例设计模型, 可以用更少的用例覆盖更全面的场景, 并且, 用例的规模基本接近。
3. 设计并实现轻量级的半自动化 SOAP 接口测试用具, 有效提升接口测试的效率和准确度。

教育经历

西北大学

本科, 2007.8-2011.7

1. 美国大学生数学建模竞赛 Honorable Mention (top 42.73%).
2. 《科研知识管理与协作平台》获西北大学"挑战杯"创业计划竞赛二等奖。
3. 陕西省大学生骨干培养学校, 第四期学员(陕西各高校主要学生干部, 约150人)。期间, 在陕西省宝鸡市凤翔县彪角镇政府挂职锻炼, 镇长助理。
4. 西北大学大学生科技协会主席, 主要负责组织各项课外科技学术活动。期间组建网络部门, 建立宣传网站, 并发起定期举办的小型学术沙龙活动。

个人项目

人人网信息抓取与数据挖掘

2013

Python, MySQL

编写爬虫自动登录人人网并抓取好友列表、状态、个人资料等信息。绘制好友关系拓扑图, 并进行网络结构和自然语言分析。

代码 github 地址: <https://github.com/JackonYang/renren>

Star 19

Fork 17

电子书管理软件

2014

Python

对本地电子书进行扫描, 建立基于 MD5 码的个人图书信息库。可以分享、合并他人的图书信息库。当信息库足够大时, 我们不再需要手动整理本地电子书。
记录个人阅读期间的操作, 可以形成个人阅读 DNA 图谱。

代码 github 地址: <https://github.com/JackonYang/bookhub>

linux性能监控与图像显示

2012

python, shell

根据日常性能监控需求, 对 sar 命令进行封装和二次开发, 方便数据收集。
优化滤波算法, 根据性能测试的常见问题对数据进行滤波并做图显示, 方便发现bug。

代码 github 地址: <https://github.com/JackonYang/performance-monitoring>