

Yitian Huang

<http://journal.ethanshub.com>

Cell: +1-(703) 395-3877 | Email: wfgydbu@163.com | WeChat: wfgydbu (2020) | [Github](#)

EDUCATION

The George Washington University

Washington DC

Master of Science in Computer Science, GPA: 3.97/4.0

May 2017

- Areas of Focus: Operating System, Cybersecurity and Database
- Relevant Coursework: Operating System, Database System II, Computer Security, Cryptography, Algorithms

Nanjing Tech University

Nanjing, China

Bachelor of Science in Computer Science

June 2012

- Graduate Thesis: Design and Implementation of the Clinical Expert System
- Relevant Coursework: Programming in C++, Software Engineering, Principles of Compiling

EXPERIENCE

Python Web Crawl Related Projects (all for research purpose):

- Distributedly and fully crawled site comic.kukudm.com:
 - (Plan A) Used *scrapy* and *scray_redis* framework. Queued URLs and stored crawled content in Redis. Max number of requests that were successfully processed reached to 270,000 per hour.
 - (Plan B) Used *pyspider* framework, launched 1 instance, 15 concurrent threads/instance, max number of requests that were successfully processed reached to 70,000 per hour.
- Crawled site weibo.com's Follower and Info pages:
 - Simulated weibo's login process to get session cookies.
 - Used logged user as a seed, recursively crawled each user's Follower page and Info page.
 - Used *requests* and *bs4* libs to retrieve and parser web pages. Used *BloomFilter* and *DB Index* to filter duplicated URLs. Used *sqlite3* for storage.
- Crawled users' order history at site taobao.com:
 - Simulated taobao's login process by using username/password/captcha combinations or scanning qrcode to get session cookies.
 - Crawled user's order history and stored in *sqlite3*, used *PrettyTable* for result display.
- Other crawler projects on [Github](#).

Python – ohHTMLToMarkdown Library

- A Python Library for converting a part or whole HTML page to Markdown format.
- Used *bs4* to parser HTML doc into blocks. Defined functions to parser each HTML tag: either directly converting it to Markdown format or continuing parser it recursively(DFS) or discarding it. Currently supported converting most HTML tags. Testing objects included Jianshu, Zhihu and my Blog, all tests passed.
- Source code on [Github](#), or searching *ohHTMLToMarkdown* in PyPI for installation.

Python – ohIPPool

- A dynamic proxy IP pool implemented by *Redis* and *Flask*.
- Highly custom *Configure* module to meet different deployment requirements. Module *Retriever* retrieves larger number of free proxy IPs and uses Redis ordered Set to manage them. Module *Validator* to guarantee IPs' quality in pool by scoring them. Module *Cleaner* cleans low-quality or disabled IPs routinely. Provided Web API to get one proxy IP by using a very simple GET request.
- Source code on [Github](#).

Python – A Content Display Site

- Used *Django* framework. Build a website for displaying crawled content coming from different Python crawlers. Currently used posts crawled from tieba.baidu.com as samples.
- Source code on [Github](#).

Course Project: TIME Action Tracking System**Individual Developer****The George Washington University, U.S***Summer 2016*

- Implemented a WordPress plugin that provides action tracking service for individuals including managing tasks and projects, generating statistic charts in various styles (Line, Pie, Column etc.), UI configurations and importing/exporting data. The website is accessible online, and codes are accessible on GitHub.
- Used CSS3 and HTML for UI design; PHP, JavaScript and interfaces provided by WordPress for backend codes and MySQL database for storage.

TICA Energy Data Management Ltd.**Software Engineer****Nanjing, China***July 2013 – April 2014*

- Participated in company's core product named Energy Saving Management System, which aims to achieve real time monitoring of regional energy consumption through various physical devices and generate energy report routinely as references for energy experts to develop energy-saving potentials.
- Used C and C++ language. Developed DBA (DataBase Agent) module for storing formatted data transmitted from DP into database and retrieving data from database for frontend team use; DP (Data Process) module for formatting data received from devices. Designed table structures for a MySQL Cluster.
- Worked as a technical assistant for colleagues, including deploying mail/file/printer/scanner/web/firewall servers, designing and configuring company's network structure and answering general technical questions.

China Soft International Ltd.**Test Engineer****Nanjing, China***March 2012 – May 2013*

- Participated in a joint project named Services Control Gateway (SCG), which is built as a third-party charge system for Telcos and provides additional services for subscribers.
- Participated in product testing, including building, reviewing and executing automated and manual test cases, by using self-developed test toolkits on Suse Linux.
- Provided online technical support mainly about product deployment for engineers at Kuwait station.

OBJECTIVE

I'm seeking for a Junior/Senior full-time Software Engineer position starting on August in Shanghai, China with emphasis on Web Crawl or Backend Development areas using Python or any other languages.

Software Skills:

- **Programming Languages:** Python 3, C/C++, HTML5, CSS, JavaScript, PHP, Shell
- **Framework:** Django, Flask, WordPress
- **Systems:** Windows, Linux, AWS/EC2
- **DB:** MySQL, MySQL Cluster, Oracle, SQLite, Redis

Language: Fluent in English and Mandarin