Derek Dong
Nanjing University of Science and Technology
Changping district Beijing city
Email: derekdongsir@163.com
Phone: 18205089562

EDUCATION

Bachelor of Intelligence Science and Technology   **September 2012 – June 2016**
Nanjing University of Science and Technology

WORK EXPERIENCE

PYTHON CRAWLER ENGINEER   **August 2017 – June 2019**
Nanjing City
- Use the python crawler technology to crawl the data of the target website according to the collection requirements book.
- Participate in the design of crawler frameworks such as asynchronous crawlers and distributed crawlers, and formulate anti-reverse crawling means for the anti-climbing encountered.
- Clean the collected data through big data technology like Hadoop, Spark

PYTHON DEVELOPER ENGINEER   **June 2016– August 2017**
Nanjing City
- Participate in the needs analysis and logical process design of the company's Web project part of the business;
- Participate in project deployment, big data environment construction, big data cleaning

PROJECT EXPERIENCE

PUBLIC INFORMATION WEBSITE CRAWLING   **June 2018– June 2019**
- Project Description: The company's main business, crawling public information on government and institutional websites
- Main work: crack all kinds of difficult anti-climbing, such as slider verification code, js encryption, app collection, etc., design and write aiohttp asynchronous crawler framework, deploy distributed crawler through Redis-scrapy, collect data volume to reach hundreds of millions of scale per month, Clean data with big data technology.

SOCIAL NETWORKING SITES CRAWLING   **November 2017– June 2018**

- Project Description: Crawling user information on social networking sites such as Renren.com, Weibo, Century Jiayuan, Lily.com
- Main work: use scrapy for large-scale data collection, solve the image verification code, dynamic cookies, custom fonts and other anti-climbing means encountered during the collection process. The data collection amount reaches 60 million per month

CV DATA CRAWLING   **August 2017– November 2017**

- Project Description: Climb the company's talent resume data according to the company's collection requirements table

- Main work: use requests to crawl some websites with low difficulty in crawling, solve the simple anti-climbing mechanism by writing IP proxy pool, loading random headers, etc., and the information collection amount reaches 20 million per month;

BOOKWORM RECOMMENDATION WEBSITE   **October 2016 – August 2017**

- Project Description: Organize and classify online book resources in the big data environment, and accurately push the books and information of interest to users through statistical analysis of user access information.

- Main work: Participate in project deployment, use nginx, uwsgi to do dynamic separation and load balancing, use Redis for caching;
Use Hadoop to set up HDFS and HBase clusters to collect user access information. Use hive, sqoop technology to analyze the user access log, and dump the processed result to the MySQL database.

EMPLOYEE MANAGEMENT SYSTEM   **July 2016 – October 2016**

- Project Description: The management system provides administrator login and employee login interface. The administrator can add, delete, and change employee information, employee accounts, and department information. Employees can log in to query personal account information.

- Main work: Participated in the needs analysis of the project;
Participate in the preparation of the login registration module code, including picture verification code, mobile phone number login, email activation asynchronous request to determine whether the account has been registered;
Participate in the design and preparation of employee information classification display, batch deletion, account information modification, etc., using Ajax to dynamically update page information.

TECHNICAL SKILLS

- Proficiency in basic Python grammar, OOP programming thought, multi-process, multi-threading, coroutine
- Proficiency in HTML,CSS,js,jquery,Ajax,Bootstrap, skilled in using ,jquery for dom operations;
- Proficiency in the use of relational database MySQL, deep understanding of SQL optimization, transaction control;
- Proficiency in the use of Django and MTV design patterns, enabling rapid development of web backend programs;
- Proficiency in the use of Linux, skilled in using uwsgi, nginx for load balancing and dynamic separation
- Familiar with the use of Redis and its persistence mechanism, can quickly build a distributed Redis cluster
- Familiar with git tools for version control and team collaboration
- Proficient in hdfs for distributed storage, hive for data cleaning, sqoop with MySQL for data interaction.
- Familiar with Scrapy crawler framework, and distributed spider with scrapy-redis
- Familiar with the use of the aiohttp asynchronous request framework and wrap it
- Participated in the design and writing of distributed crawlers
- proficiency with common anti-crawl mechanisms and cracking methods, such as slider verification code, js encryption, custom fonts, etc.
- Skilled in the cleaning of offline data such as word, excel, familiar with modules such as python-docx pywin32 pandas numpy
- Familiar with spark to do data deduplication and merging
- Familiar with pyecharts for data visualization and data reporting
- Understand machine learning, have a certain understanding of linear regression algorithms, logistic regression algorithms, clustering algorithms, etc.
- Familiar with common data structures such as stacks, queues, trees, etc., as well as common sorting, lookup algorithms

Acknowledgment

Thank you for taking the time to read my resume and look forward to having the opportunity to work with you.