Guozhen She

GitHub: github.com/hazelnutsgz Email:15307130224@fudan.edu.cn LinkedIn:https://www.linkedin.com/in/hazelnutsgz/

EDUCATION

Fudan University, Shanghai, China 2015.9-
Bachelor of Computer Science (expected in 07.2020, **one year delay** because of a surgery)

GPA (overall): **3.55/4.0**; Ranking: **21/117**
Body:Computer Architecture(A)| Computer Network(A-)|Computer System(A)|Database Implementation(A-)|Operating System(B+)
Brain: Data Structure(B+)|Distributed System(A)|Linear Algebra(A)
Mouth: C Programming(A)|C++ Programming(A-)|Web Development(A)
Metaphysics: Neural Network and Deep Learning(A-)

SELECTED PROJECTS

Multiple Pattern Text Matching Tool (https://github.com/hazelnutsgz/NaiveACAutomation)
Implement the **Aho–Corasick** automation(Trie with failed pointers), facilitating the text matching in multiple pattern.
Interactive Visualization of Coauthor Affiliation(https://github.com/hazelnutsgz/NaiveScholarMap**)**
Construct the co-author affiliation graph based on yearly data crawling from google scholar, then build an interactive web service by D3.js library, for comparison and analysis of co-author affiliations in different years.
Monitoring Service of WeChat Group(https://github.com/DaShiLar/Naive-WeChat-Monitor)
Build a **monitoring** backend service for all WeChat groups of the user, which captures the real-time chatting information(video, text, voice), and stores them to database and filesystem in backend. Also a front-end web UI is provided to users for authorization and inspection. The system is implemented in a **multi-processes** architecture for **the isolation** of different users and utilization of multi-cores on backend server.

INDUSTRY EXPERIENCES

Microsoft Research Asia, System and Network Group | Research Intern Jan.2019-

- Real-time Bot Detection System for Azure Cloud Service
  - Implement a Golang web service to **build the preprocessing pipeline** of daily network log data(8,000,000) from Bing, which parse the raw log into heterogeneous structure hosted on a distributed file system.
  - Based on that, provide an analytic service(React, D3.js) for visualizing and understanding data. For conservation of memory footprint on VM, the data is **fetched on demand** and cached as the data structures in memory, guaranteeing the data at **hotspot** would stay longer in memory to accelerate the analytic efficiency.
  - Implement an algorithm to generate the behavior-based images for each request session, then develop a **CNN-based** model by PyTorch to detect bot behavior by classifying the images generated, which reaches **94.3%** accuracy on labeled Bing log data. Furthermore, build an interactive T-SNE & PCA service to visualize and validate the model.

Intel Asia-Pacific R&D, Open Source Technology Center | SDE Intern Aug.2018-Nov.2018

- **Contribute code** to StarlingX(**OpenStack** Foundation), assist in deploying the StarlingX on bare-metal devices.
- Build a **rule-based** command line tool which migrate code from python2 to python3.
- Assist the colleagues to setup the **compiling farm** based on K8S for building of StarlingX project.
- Implement a static graph-based algorithm for **package dependency analysis** in the project.

Wish | SDE Intern Jan.2018-Apr.2018

- Develop an info adjustment service, using Tornado framework at backend, and backbone.js at front end.
- Build an adaptive notification service for accounts out of credits. Based on that, designed an algorithm to **spot zombie users**.
- Build **a channel search service** equipped with multiple filters to assist users to select the proper channel on their own conditions.

ACADEMIC EXPERIENCES.

Fudan University, Advisor: Dr. **Yang Chen** | Research Assistant April.2017-

- **LinkedIn** website:
  - Build a **cookie-based** crawling system to scraping profiles retaining, which **imitate** normal user behavior for anti-crawling.
  - Implement automatic script to **expand the personal LinkedIn connection**s based on LinkedIn Recommendation.
  - Scrape the profiles of connections **concurrently** utilizing multiple mock accounts at the crawling system, then write an **error-tolerant** parsers to generate JSON-like profile for each profile URL.
- **Google Scholar** website
  - Crawl the Google Scholar profile, integrated with **anti-captcha** service. Use machine learning method to detect fake profile.
  - Crawl data from MS Academic, DBLP, Google Scholar, build the **heterogeneous graph** containing both authors and papers for given conferences in different year. Utilize **Graph Convolution Network(GCN)** to evaluate the conference.
- Qingyun Go
  - Build a **geo-based** social App. The communication between backend and frontend is hosted on HTTPS protocol based on RESTFUL API. The backend is FastCGI integrated with C++ code, while the frontend is a javascript runtime with asynchronous API. Then develop some streaming analytic tools to monitor the status of service and the behavior of users.

MISCELLANY

Interest: **Archeology** (on Computer Science), Soccer(DM), **Road Cycling**(ITT)
Programming Language: Python, Java, Golang, C, C++, JavaScript, MATLAB(wanderer), Rust(dabbler),
Framework & Library: Tornado(Python), D3.js(JS), Node.js(V8), System Call(C), Tensorflow, PyTorch