Sicheng Wang

One Miramar Street, La Jolla, CA 92092

(858)319-5237 ● ✉ siw029@ucsd.edu ● in sicheng-wang-

Education

University of California San Diego   **San Diego, CA**
- Master of Science in Computer Science, GPA: 3.82/4.00   *Sept. 2021 — Mar. 2023*

Tsinghua University   **Beijing, China**
- Bachelor of Engineering in Automation, GPA: 3.83/4.00   *Aug. 2017 — Jul. 2021*
- Awards: Philobiblon Scholarship (**Top 5%**), 2019 | Guangyao Scholarship (**Top 10%**), 2018

Experience

Stackline   **Seattle, WA**
Software Development Intern   *Jun. 2022 — Sept. 2022*
- Developed an Amazon Deals crawling pipeline with **Node.js** and used **GitLab CI/CD** to deploy it on **Amazon Web Services** to collect promotional information of **19** countries, crawling over **10 millions** deals per week.
- Triggered the pipeline execution automatically by using **Airflow DAG** to send messages to **Amazon SQS**.
- Reduced crawling overheads by **20%** through deducing deal product variation status and removing redundant crawls.
- Utilized **Amazon Kinesis Data Firehose** to stream the results parsed by **Cheerio** to **Amazon S3**.
- Discriminated the use of **AJAX** API call and deal ID URL request, **2**× the efficiency for over **90%** of deal ID crawls.
- Optimized the automatic captcha puzzle solving by using image sharpening(Laplacian Filter) to get clear boundaries.

University of California San Diego   **San Diego, CA**
Research Assistant   *Feb. 2020 — Jul. 2020*
CERT: Contrastive Self-supervised Learning for Language Understanding
- Proposed a novel language data augmentation technique with back translation to boost the contrastive training.
- Increased the accuracy by **1.1 %** on SST-2 task.
- Utilized **Docker** images and **Kubernetes** to conduct a comprehensive set of CERT experiments on SQuAD v1.1, SST-2 and MRPC of the General Language Understanding Evaluation (GLUE) benchmark.
- Published the paper and obtained **127 citations**: https://arxiv.org/pdf/2005.12766.pdf

Projects

Distributed Cloud Storage System(Go)   **Jan. 2022 — Mar. 2022**
- Developed a Dropbox-like cloud file system with Go, which allows clients to sync to and from cloud via **gRPC**.
- Constructed the **distributed** server part, with each server node contains MetaStore and BlockStore interfaces.
- Utilized the **RAFT** consensus algorithm to divide server nodes into leader and followers, intelligently electing the leader for managing the replicated log and ensuring data availability and consistency.

Sentiment Analysis & Filter System for Danmu Video Comments   **Dec. 2018 — Mar. 2019**
- Constructed a system that can analysis the emotion of Danmu(Bullet Screen) comments and filter the undesirable ones according to video viewers' emotional preference.
- Used **Python** web-crawling tool **Scrapy** to collect **40,000** video comments and segmented them into Chinese words.
- Applied the **LSTM** model and implemented **Pseudo-Labeling** learning algorithm as a semi-supervised approach to predict the 7 dimensional emotion of the extracted keywords.
- Utilized Bilibili's **XML** Danmu API to implement the real time video comments filtering function.
- Led a **3-member** team, won the second prize and Best Rookie Award (**Top 16/784**) in 37th THU Challenge Cup.

Travel Story Sharing Website(RESTful API, MongoDB, Express)   **Nov. 2021 — Dec. 2021**
P2P Network Chat System Based on Central Server(Python)   **Nov. 2019 — Dec. 2019**
Parallel computing: General Matrix Multiply(C++, SIMD, AVX2)   **Mar. 2022 — May. 2022**
Hotel Reservation and Management System(C++, Qt, SQLite)   **Jul. 2018 — Aug. 2018**
Nachos: Operating System Simulator(Java)   **Oct. 2021 — Dec. 2021**

Skills

- Programming Languages: Python, C/C++, C#, MATLAB, Javascript, Java, Haskell, Shell Script
- Framework and Tools: AWS, Kubernetes, Docker, SQL, Git, Gitlab CI/CD, MongoDB, Node.js, HTML, CSS, RESTful API, Pytorch, Spark, Postman