Yifu Wang

PHONE: (626) 226 6282 EMAIL: yifuw@andrew.cmu.edu
GITHUB: github.com/wangyifu LINKEDIN: www.linkedin.com/in/yifuwang

EDUCATION

M.S. Computer Science, Carnegie Mellon University, School of Computer Science **May / 16**

Courses: Advanced Data Structure and Algorithms, Machine Learning, Cloud Computing, Big Data System in Practice, Search Engine

Teaching: Data Structures for Application Programmers, JAVA for Application Programmers

B.S. Software Engineering, Dalian University of Technology **June / 14**

Courses: Operating System, Compiler Techniques, Database Systems, Computer Network, Computer Architecture, Data Structure and Algorithms

WORK EXPERIENCE

Software Engineer Intern at *Quixey, Inc.* May / 15 - Aug / 15

- Implemented Chinese query tokenizer based on **Early parser**.
- Using **chart parse** to store intermediate result.
- Applied **dynamic programming** in candidate path elimination.
- **40% faster** and **1.6% gain of DCG score** compared with previous tokenizer.
- Integrated into Quixey app search and auto suggestion in China market.

SDET Intern at *VMware, Inc.* Aug / 13 - Feb / 14

- Implemented **image comparison plugin** for automation testing framework using **C#**.
- Developed several web applications based on **LAMP(Linux, Apache, MySQL, PHP)** stack.
- Visualize product development progress, employee KPI, and bug report.
- Generate report in pdf and excel format for team meeting.

PROJECTS

DOTA2 COUNTER-PICK:
- Hero recommendation engine for Dota 2 game. Written in python.
- Generated feature matrix from 30000 matches using **urllib, numpy, pandas**.
- Applied machine learning algorithms including **logistic regression**, **K-nearest neighbors**, **SVM with RBF kernel** using **sci-kit** package.
- Achieved **70% accuracy** in predicting match outcomes.

BIO-ASK:
- Question answering system in medical field based on **UIMA architecture**.
- Implemented document retrieval algorithm based on **Stanford NLP** and **Lingpipe**.
- Developed result evaluation system using **BM25** and **GMAP**.

TARTAN ENGINE:
- Simple search engine based on **lucene**.
- Supports both **unstructured** and **structured** query.
- Implemented operators: **AND, OR, NEAR, WINDOWS, SYN**.
- Implemented several retrieval models: **BM25, Indri, Ranked Boolean**.

DRAW SOMETHING:
- Online graphic multiplayer game runs on **Linux**.
- Developed **graphic** part and **network communication** part using **C++** and **Qt**.
- Used **bezier curve** to represent and serialize canvas.
- Implemented **multi-thread TCP server** using Qt to support concurrent connections.

SKILLS

Languages: JAVA, Python, C++, shell script(bash, awk), Javascript, PHP, Scala, R, Matlab

Other Technologies: Web Development, MySQL, AWS, Hadoop, Elastic Mapreduce, ElasticSearch, Mongo, Mahout, Lucene, Git