Jun Hu

hu-jun.com | jun.hu@columbia.edu | (646) 271-1419

EDUCATION

COLUMBIA UNIVERSITY
M.S. in Computer Science
Jan 2017 - May 2018
New York, NY

TSINGHUA UNIVERSITY
Ph.D. in Biology
Sep 2006 - Jan 2012
Beijing, China

WUHAN UNIVERSITY
B.S. in Biology
Sep 2002 - Jul 2006
Wuhan, China

COURSEWORK

Databases
Advanced Database Systems
Big Data Analytics
Advanced Big Data Analytics
Machine Learning
Artificial Intelligence
Analysis Of Algorithms
Computational Graphics
Microservice/Cloud Apps
Natural Language Processing

PUBLICATIONS

Research Papers: 49
Citations: > 700 times
Full list: bit.ly/2xVZYic

SKILLS

Java • Python
Linux • Docker
Graph DB • Elasticsearch
HTML/Sass • JavaScript
TensorFlow • Scikit-learn
Git • LaTeX

LINKS

Portfolio: hu-jun.com
Linkedin: /hu-jun
Github: /vibrioh

STATUS

Permanent Resident via EB-1A
(Alien of Extraordinary Ability)

WORK EXPERIENCE

GRAPHEN, INC — New York, NY
Principal Research Scientist, Big Data — Sep 2018 — Current
Research Scientist — Jan 2018 – Sep 2018
Contributed to the IBM-ICBC FinTech project ($10 M).
Participated in the BOC Cybersecurity project ($0.8 M).
Collaborated with Harvard on a cancer classification project (non-profit).
Developed Client module & maintained REST layer of our graph system.

ICAHN SCHOOL OF MEDICINE AT MOUNT SINAI — New York, NY
Postdoctoral Fellow — Apr 2012 - Dec 2016
Oversaw research progress of the laboratory graduate students.
lead 8 projects and published 19 research articles.
Established a pipeline of large scale sequencing analysis in the laboratory.

PROJECTS

ICBC NON-PERFORMING LOANS PREDICTION
IBM GBS provided an AI solution to ICBC ($10 M).
Graphen provided graph analytics solution to IBM ($2 M).
Learn More: http://hu-jun.com/prj/icbc.html
Dockerized graph database system and applications into containers.
Integrated Kafka/Elasticsearch/Asynchronous/REST/GraphDB/HAProxy.
Deployed the distributed graph system on production environment.
Designed data schema and formulated loading procedure.
Transformed business requirements into analytical scripts.
Conceptualized the dataset to a fake data generator for off-site.
Automated daily ingestion and computation workflow in Shell/Cron.
Consolidated high availability by tuning HAproxy/NGINX/uWSGI.
Strengthened the robustness by implement of main/backup auto switch.
Customized RESTful API iteratively for specific business logic.
Co-authored dynamic graph UI backend under IBM Spring Framework.
Implemented .doc/.xls report generating module with FreeMarker.
Facilitated user experience by implemented the Client module in Python.
Trained users in ICBC for graph database knowledge and operations.

BOC NY ADAPTIVE CYBERSECURITY MONITORING
Graphen provided a Cybersecurity AI solution to Bank of China New York branch ($0.8 M).
Learn More: http://hu-jun.com/prj/boc.html
Determined features selection for the advanced persistent threat.
Deployed standalone graph database system.
Created data schema, loading and graph feature scripts for ML models.
Co-authored display UI backend in Python, with Flask, Elasticsearch, Redis.
Drafted SIT/UAT test scripts and documentation.

B-CELL RECEPTOR BASED CANCER CLASSIFICATION
A non-profit research project with Harvard X. Shirley Liu Lab.
Totally 4241 patient samples were investigated.
Learn More: http://hu-jun.com/prj/cancer.html
Performed NLP n-gram method in amino acid sequences.
Explored Neural networks model with GPU acceleration.