Yang Bai

baiyang94@ufl.edu | 352 284 2346 | tonyby.github.io | Linkedin

EDUCATION

UNIVERSITY OF FLORIDA
Gainesville, FL, USA
2019 - 2023(Expected)
PhD. Computer Science
GPA: 3.96/4.0

2016-2018
MSc. Electrical and Computer Eng.
GPA: 3.5/4.0

SICHUAN UNIVERSITY
2012-2016 | Chengdu, China
BSc. Micro Electronics

SKILLS

LANGUAGES

Python :	*Expert*
Java :	*Expert*
SQL/SPARQL :	*Expert*
C/C++ :	*Intermediate*
JavaScript :	*Intermediate*
F# :	*Intermediate*

TOOLS

Machine Learning:	PyTorch
	TensorFlow
	Scikit-learn
	Hugging Face
Data processing:	Pandas
	Oracle DB
Others:	Docker
	Akka.NET
	Flask
	JUnit
	Googel Test

COURSEWORKS

Elements of Machine Intelligence
Deep Learning for Computer-Graphics
Applied Machine Learning
Trustworthy Machine Learning
Distributed Operating System
Programming Language Principles
Database Management System
Database System Implementation
Analysis of Algorithms
Advanced Data Structures
Computer Networks

EXPERIENCE

Data Science Research lab at University of Florida
| Research Assistant| 2019-present
Under the supervision of **Dr. Daisy Zhe Wang**, working on **machine learning** and **knowledge graph**-based data-science research. **Individual contributor and team lead** in the **DARPA** sponsored project "Active Interpretation of Disparate Alternatives(**AIDA**)", an alternative hypotheses search engine over event-enteric knowledge graphs. Our system achieves **top performance** at the **NIST TAC SM-KBP2020 evaluation**.

- Developed a query decomposition algorithm to break graph queries into atomic sections and a composition algorithm to combine the individual results.
- Developed a two-level graph searching algorithm to explore knowledge graphs at both mention-level and cluster-level improving the final F1 score by **25%**.
- Developed a graph clustering algorithm to differentiate alternative hypotheses by measuring both structural and semantic distance between candidates, which improves the original cluttering quality(v-measure) by **20%**.

COURSE PROJECTS

Realtime Face Mask Detection (Python, C++, PyTorch, CUDA)
- Implemented and evaluated various state-of-the-art object detection models including vanilla CNN, Faster R-CNN, and YOLOv3 state-of-the-art models
- Built a new face-mask identification dataset(5.5K images) by combining and unifying the formats of various open-source face-mask data, including synthetic and real-world data.

Database System Implementation (C++, POSIX Threads)
- Developed a database management system for Linux, using C++.
- Implemented parser functions to read the SQL queries using LEX&YACC.
- Implemented relational operations and a query optimizer using POSIX threads.

Compiler Implementation (Java, JUnit, ASM)
- Implemented a compiler for a custom programming language, using Java.
- The compiler implementation includes both the front end (Scanner, Parser, Semantic Analysis) and a back end (Target Code Generation).
- Used the ASM byte code framework to help with the target code generation.

PUBLICATIONS **Google Scholar**

More Than Reading Comprehension: A Survey on Datasets and Metrics of Textual Question Answering
Yang Bai,D.Wang **arXiv 2021**

GAIA at SM-KBP 2020 - A Dockerized Multi-media Multi-lingual Knowledge Extraction, Clustering, Temporal Tracking and Hypothesis Generation System
M.Li,...,**Yang Bai**,..., D.Wang **TAC 2020**