XIAOYI DUAN

⋄ 408-470-0153 ⋄ xiaoyid@andrew.cmu.edu ⋄ Mountain View, CA
⋄ LinkedIn: xiaoyid ⋄ Website: http://xiaoyidolly.github.io/

EDUCATION

Master in Software Engineering (2-year PhD student) Sep. 2016 - May 2019
Research Area: Service Computing ECE Department, Carnegie Mellon University
Bachelor, Master in Software Engineering Aug. 2009 - June 2016
Research Area: Data Management SE Institute, East China Normal University, China

SKILLS

Web Service & Application	Python/Flask, Java/Play, JavaScript/Node.js, jQuery, Ajax
Data Processing	Hadoop, MongoDB, Neo4j, MySQL, PostgreSQL
Natural Language Processing	wordNet, NLTK, PyTorch
Cloud Computing & Tools	AWS, Azure, Jupyter, Docker, Git, Jira

RESEARCH PROJECTS

Knowledge Graph Construction for Earth Science June 2017 - Current
***NASA project**, building a knowledge graph for earth science publications* Student Leader

· Analyzed unstructured text data for **NLP** problems, i.e., semantic entity identification (**CRF**, **word2vec**) (17% improvement over heuristic methods), topic modeling (**LLDA**) and word disambiguation (**ontology**)
· Designed a **Flask** pipeline for crawling html, cleaning and processing unstructured text, and extracting features
· Collaborated with linguists to extend **Stanford parser** to build a lexical parser for earth science domain and visualize extracted relations with **D3.js**
  – *(Best Student Paper Award)* **Xiaoyi Duan**, Jia Zhang, Rahul Ramachandran, et at. "A Neural Network-Powered Cognitive Method of Identifying Semantic Entities in Earth Science Papers", in Proceedings of IEEE International Conference on Cognitive Computing (ICCC), 2018, San Francisco, CA, USA, pp. 9-16.

Evaluation and Knowledge Network Query System June 2017 - Dec. 2017
***NSF/NASA project**, an evaluation system to collect user evaluations* Student Leader

· Developed **RESTful APIs** in **Flask** for data analysis tasks and **Play** for the user-oriented system, with data interaction among **MySQL**, **MongoDB**, and **Neo4j**
· Designed data synchronization mechanism to speedup data loading and ensure data consistency
· Built **dockers** for development usage and deployment usage with **Jenkins**

Collaborative Scientific Workflow Aug. 2016 - May 2017
***NSF/NASA project**, a real-time system to collaborate on the scientific workflow* Student Leader

· Extended PROV **data model** to store collaborative design history of **scientific workflow**
· Developed RESTful APIs in **Node.js** to query the **provenance** in neo4j
· Lead the team to build a real-time collaboration system with management of workflow history and user group
  – **Xiaoyi Duan**, Jia Zhang, Qihao Bao, et at. "Linking Design-Time and Run-Time: A Graph-Based Uniform Workflow Provenance Model", in Proceedings of IEEE International Conference on Web Services (ICWS), 2017, pp. 97-105.

OTHER SELECTED PUBLICATIONS

Qihao Bao, Jia Zhang, **Xiaoyi Duan**, et al. "A Fine-Grained API Link Prediction Approach Supporting Mashup Recommendation", in Proceedings of IEEE International Conference on Web Services (ICWS), 2018, pp. 220-228.

Xiaoyi Duan, Cheqing Jin, Xiaoling Wang, et al. "Real-time Personalized Taxi-Sharing", in Proceedings of International Conference on Database Systems for Advanced Applications (DASFAA), 2016, pp. 451-465.

ACTIVITIES AND AWARDS

Student Volunteer, 2018/2017 IEEE Services Congress	July 2018 and June 2017
Excellent Award, Microsoft Imagine Cup Contest	Jan. 2014
Second Place Award (2/660), China Big Data Techniques Innovation and Entrepreneurship Contest	Dec. 2013