

# MINH PHAM

Los Angeles, California

☎ 385-238-5583 ✉ [minhpham1811@gmail.com](mailto:minhpham1811@gmail.com) Website: <https://minhptx.github.io/>

## RESEARCH INTERESTS

---

- Automatic Data Integration
- Table-text Understanding
- Deep Learning
- Knowledge Graph Construction

## EDUCATION

---

### University of Southern California

Sep 2015 – Now

*Ph.D. Candidate in Computer Science*

- Relevant Courses: Machine Learning, Building Knowledge Graphs, Advanced NLP, Representation Learning
- GPA: 3.88/4.0
- Advisors: Craig A. Knoblock and Muhao Chen

### Ho Chi Minh City University of Technology

Sep 2009 – Jan 2014

*Bachelor of Engineering in Computer Science*

- Thesis Topic: Wikipedia-based Entity Disambiguation
- GPA: 8.5/10
- Advisors: Tru Hoang Cao

## SELECTED PUBLICATIONS

---

**Pham, M.**, Knoblock, C., Chen, M., Vu, B. and Pujara, J., *SPADE: A Semi-supervised Probabilistic Approach for Detecting Errors in Tables*. In 30th International Joint Conference on Artificial Intelligence (IJCAI 2021).

**Pham, M.**, Knoblock, C. and Pujara, J., *Learning Data Transformation with Minimal User Effort*. In 2019 IEEE International Conference on Big Data (IEEE BigData 2019).

**Pham, M.**, Alse, S., Knoblock, C., and Szekely, P., *Semantic Labeling: A Domain-Independent Approach*. In 15th International Semantic Web Conference (ISWC 2016).

Gil, Y., Garijo, D., Khider, D., Knoblock, C.A., Ratnakar, V., Osorio, M., Vargas, H., **Pham, M.**, Pujara, J., Shbita, B. and Vu, B., *Artificial intelligence for modeling complex systems: taming the complexity of expert models to improve decision making*. In ACM Transactions on Interactive Intelligent Systems 2021

## RESEARCH EXPERIENCE

---

### Robust and Proactive Error Detection and Correction

Dec 2019 – Now

*Research Assistant, Center on Knowledge Graphs, USC/ISI*

- Develop an approach to verify information in tabular data with table-to-text controlled text generation and natural language inference
- Developed a semi-supervised error detection approach that leverages Probabilistic Soft Logic and deep neural networks to detect syntactic errors in tables (IJCAI 2021)

### MINT Project – <http://mint-project.info>

Sep 2017 – Dec 2019

*Research Assistant, Center on Knowledge Graphs, USC/ISI*

- Designed and developed a modular transformation pipeline to unify domains-specific data from different scientific formats including GPM, GLDAS or ISRIC
- Developed a novel unsupervised approach to transform string values between different formats without labeled examples (IEEE BigData 2019)

### PRINCESS Project – <https://cra.com/projects/princess>

Jan 2016 – Sep 2017

*Research Assistant, Center on Knowledge Graphs, USC/ISI*

- Developed a machine learning approach for domain-independent semantic labeling, which can achieve comparable results on different domains without fine-tuning (ISWC 2016)

### John von Neumann Institute, Vietnam National University

Sep 2014 – May 2015

*Research Assistant*

- Improved an existing learning-based entity linking system by candidate searching and rule-based coreference resolution.

## PROFESSIONAL EXPERIENCE

---

### **Nuance Communications Inc.**

**Jun 2018 - Aug 2018**

*Research Intern, Artificial Intelligence and Language Lab*

- Developed an unsupervised approach for an internal entity resolution problem using Probabilistic Soft Logic and automatic data profiling

### **DataFirst JSC Co.**

**Feb 2015 - Aug 2015**

*Research Programmer*

- Extracted real estates' information from millions of Vietnamese online listings with high accuracy for market analysis.

### **East Agile**

**Jun 2014 – Aug 2014**

*Software Engineer Intern*

- Developed and maintained an in-house video sharing platform using Ruby on Rails, CoffeeScript, HTML5 and CSS.

## TEACHING EXPERIENCE

---

### **University of Southern California**

**2016 - 2021**

*Teaching Assistant, DSCI 558: Building Knowledge Graphs*

*Los Angeles, California*

- Designed and evaluated course examinations, written assignments, and weekly quizzes
- Presented several sessions of lectures & research seminars to the class

## AWARDS AND SCHOLARSHIPS

---

### **Best Paper Award, ISI Graduate Student Symposium, University of Southern California**

**2019**

*Title: Learning Data Transformations with Minimal User Effort*

### **Vietnam Education Foundation (VEF) Fellowship for Ph.D. study in US**

**2015**

*\$54,000 for 35 selected Fellows in the whole country*

## TECHNICAL SKILLS

---

- **Machine Learning:** PyTorch, Tensorflow, Keras, Scikit-learn, Snorkel
- **Languages:** Python, Java, C++, SQL
- **Semantic Web:** RDF, Turtle, SPARQL
- **High Performance Computing:** Dask, Spark
- **Databases:** ElasticSearch, MongoDB

## LEADERSHIP / EXTRACURRICULAR

---

### **Vietnam Education Foundation Fellows and Scholars Association (VEFFA)**

**2016 - 2017**

*Board of Executives*

- Organized VEFFA annual conferences and events.
- Organized mock interviews for more than 60 Vietnam Education Foundation (VEF) Scholarship applicants.
- Led a group of 20 mentors to support VEF Scholarship applicants in preparing their applications.

### **PAKDD 2015 Conference & ACML 2014 Conference**

**2014 - 2015**

*Website Administrator and Volunteer*

- Designed and managed conference website.
- Monitored presentation sessions in the conference.