Lam Thanh Do

linkedin.com/in/thanh-lam-do/
lamdo@illinois.edu
https://louisdo.github.io/
217-418-1659

EDUCATION

University of Illinois at Urbana-Champaign (UIUC) — Illinois, USA
Ph.D. Student (Expected date of graduation: 5/2029) — *8/2024-Present*

- **Research interest:** information retrieval, learned sparse retrieval, retrieval-augmented generation, cross-modal information retrieval, generative model, vision language model
- **GPA:** 3.88/4.0
- **Academic advisor:** Prof. Kevin Chang.

Hanoi University of Science and Technology — Hanoi, Vietnam
BSc. in Computer Science — *8/2018-5/2023*

PUBLICATIONS

List of publications on Google Scholar.

Preprint

1. **Lam Thanh Do**, Linh Van Nguyen, David Fu, Kevin Chen-Chuan Chang. CASPER: Concept-integrated Sparse Representation for Scientific Retrieval. *Under submission.* 2025

Conference Publications

1. **Lam Thanh Do**, Aaditya Bodke, Pritom Saha Akash, Kevin Chen-Chuan Chang. ERU-KG: Efficient Reference-aligned Unsupervised Keyphrase Generation. *The 63rd Annual Meeting of the Association for Computational Linguistics (ACL).* 2025.
2. **Lam Thanh Do**, Pritom Saha Akash, Kevin Chen-Chuan Chang. Unsupervised Open-domain Keyphrase Generation. *The 61st Annual Meeting of the Association for Computational Linguistics (ACL).* 2023.

RESEARCH EXPERIENCE

Forward Data Lab, UIUC — Illinois, USA
Research Assistant — *8/2024-present*

- **Description:** Working on learned sparse retrieval (LSR) algorithms for scientific documents.
- **Advisor:** Prof. Kevin Chang

The Impact Research Group, UIUC — Illinois, USA
Research Assistant — *6/2025-present*

- **Description:** Working on a framework for automatic RAG benchmarks construction with quality guarantee.
- **Advisor:** Prof. Wei-mei Hwu

Forward Data Lab, UIUC — Hanoi, Vietnam
Undergraduate Research Assistant (Part-time, remote) — *4/2021-12/2022*

- **Description:** Developed unsupervised keyphrase generation algorithms that achieved performance nearly equal to a supervised model.
- **Advisor:** Prof. Kevin Chang

INDUSTRY EXPERIENCE

Cazoodle Inc. — Hanoi, Vietnam
Lead Software Engineer — *5/2023-8/2024*

- **Co-founded** the Academic Online project (***Explore***, ***Public Profile***) with my advisor. This project collects and combines scholarly data from various sources to help users understand the research landscape. **Led a team** of 9 members throughout its development.
- The project are currently being beta-tested by **seven US institutions**.
- **Advisor:** Prof. Kevin Chang

iCOMM Media & Tech, Jsc — Hanoi, Vietnam
Software Engineer — *12/2022-5/2023*

- **Topic:** News Reports Generator for the Vietnamese Language

Viettel Group — Hanoi, Vietnam
Software Engineer — *12/2021-10/2022*

- **Topic:** Recommendation System for ***TV360***

Pixta Inc. Hanoi, Vietnam
Software Engineer *5/2020-4/2021*
- Worked on a Learning-to-Rank system for *Pixtastock* alongside 2 other members.
- Achieved a **15% increase in click-through rate** and a **4% increase in conversion rate** in the first A/B test. The project was eventually deployed to production and is **currently in use**.

Pixta Inc. Hanoi, Vietnam
Software Engineer Intern *9/2019-5/2020*
- Designed an Automatic Image Tag Annotation system. The tags are used by the *Pixtastock* search engine. The project was deployed to production and is **currently in use**.

TEACHING EXPERIENCE
- *Lecturer*, VietAI *(10/2023 - Present)*

PROFESSIONAL SERVICES
1. Reviewer: ACL

SKILLS
1. **Programming languages:** Python, C/C++, SQL.
2. **Technologies:** Numpy, Scikit-learn, PyTorch, HuggingFace, Git, AWS, Docker, Linux, LaTeX.
3. **Languages:** English *(fluent)*, Vietnamese *(native)*.

RELEVANT COURSES
- **CS483 - Applied Parallel Programming** by Prof. Volodymyr Kindratenko *(UIUC - Spring 2026)*
- **ECE534 - Random Processes** by Prof. Olgica Milenkovic *(UIUC - Spring 2026)*
- **CS512 - Data Mining Principles** by Prof. Jiaxuan You *(UIUC - Fall 2025)*
- **CS546 - Advanced Topics in NLP** by Prof. Dilek Hakkani-Tur *(UIUC - Fall 2025)*
- **CS510 - Advanced Information Retrieval** by Prof. ChengXiang Zhai *(UIUC - Spring 2025)*
- **CS447 - Natural Language Processing** by Prof. Julia Hockenmaier *(UIUC - Spring 2025)*
- **CS412 - Machine Learning** by Prof. Arindam Banerjee *(UIUC - Fall 2024)*
- **CS410 - Text Information Systems** by Prof. Pablo D Robles Granda *(UIUC - Fall 2024)*

REFERENCE
1. **Prof. Kevin Chang**, Full Professor at the Siebel School of Computing and Data Science - University of Illinois Urbana-Champaign (UIUC), USA *kcchang@illinois.edu*
2. **Prof. Wen-mei Hwu**, Professor and Sanders-AMD Chair of Electrical and Computer Engineering - University of Illinois Urbana-Champaign (UIUC), USA *w-hwu@illinois.edu*