Viet D. Lai

vietl@uoregon.edu
+1-541-515-2996
https://laiviet.github.io/
Google Scholar

Education

University of Oregon 2018-Present
Eugene, Oregon, USA
Ph.D. in Computer Science
Advisor: Thien Huu Nguyen

Japan Advanced Institute of Science and Technology 2016-2018
Ishikawa, Japan
Master in Computer Science
Advisor: Minh Le Nguyen

Posts and Telecommunications Institute of Science and Technology 2011-2015
Hanoi, Vietnam
Bachelor in Computer Science

Work Experience

Research Intern 6/2022-9/2022
Adobe Research, San Jose, California, USA
Project: Reinforcement Learning with Generative Model for Punctuation Restoration

Research Intern 3/2022-6/2022
Educational Testing Service, Princeton, New Jersey, USA
Project: Automatic Scoring of Teacher in Online Learning Environment using Video Transcript

Research Intern 6/2021-12/2021
Adobe Research, San Jose, California, USA
Project: Video Transcript Understanding (punctuation restoration, chitchat detection, question answering)

Research Assistant 4/2017-3/2018
National Institute of Informatics, Tokyo, Japan
Project: Abstract Meaning Representation for Legal Document

Network Engineer Intern 6/2015-12/2015
Cisco System, Hanoi, Vietnam

Software

Trankit: A multilingual NLP Toolkit (`http://nlp.uoregon.edu/trankit`)
Trankit is a light-weight Transformer-based Toolkit for multilingual Natural Language Processing (NLP). It provides a trainable pipeline for fundamental NLP tasks over 100 languages, and 90 pre-trained pipelines for 56 languages.

FourIE: An event extraction system (`http://nlp.uoregon.edu/fourie`)
FourIE annotates text for entity mentions (names, pronouns, nominals), relations, event triggers and argument roles.

Segmentano: An annotation tool for text segmentation
(`http://nlp.uoregon.edu:8002/?project=pr`)
Segmentanno is a fast light-weight interactive general purpose annotation tool for text segmentation. It is designed to speed up the annotation for text segmentation by minimizing the mouse activities. It supports various level of text units (e.g. token, clause, sentence, and paragraph) and various tasks such as sequence labeling, text classification.

Patents

Viet Dac Lai and Franck Dernoncourt. Low resource event detection (pending), 2022

Selected Publications

- Viet Dac Lai, Amir Pouran Ben Veyseh, Minh Van Nguyen, Franck Dernoncourt, and Thien Huu Nguyen. Meci: A multilingual dataset for event causality identification. In *COLING*, 2022
- Viet Dac Lai, Amir Pouran Ben Veyseh, Franck Dernoncourt, and Thien Huu Nguyen. BehancePR: A punctuation restoration dataset for livestreaming video transcript. In *NAACL*, 2022
- Viet Dac Lai, Amir Pouran Ben Veyseh, Franck Dernoncourt, and Thien Huu Nguyen. BehanceCC: A chitchat detection dataset for livestreaming video transcripts. In *LREC*, 2022
- Viet Dac Lai, Amir Pouran Ben Veyseh, Franck Dernoncourt, and Thien Huu Nguyen. A new dataset for identifying question-answer pairs in video transcripts. In *LREC*, 2022
- Viet Dac Lai, Franck Dernoncourt, and Thien Huu Nguyen. Learning prototype representations across few-shot tasks for event detection. In *EMNLP*, 2021
- Viet Dac Lai, Minh Van Nguyen, Heidi Kaufman, and Thien Huu Nguyen. Event extraction from historical texts: A new dataset for black rebellions. In *ACL-IJCNLP 2021*, 2021
- Viet Dac Lai, Minh Van Nguyen, Thien Huu Nguyen, and Franck Dernoncourt. Graph learning regularization and transfer learning for few-shot event detection. In *SIGIR*, pages 2172–2176, 2021
- Minh Van Nguyen, Viet Dac Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen. Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. In *ACL (System Demo)*, pages 80–90, 2021
- Viet Dac Lai, Tuan Ngo Nguyen, and Thien Huu Nguyen. Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. In *EMNLP*, pages 5405–5411, 2020
- Viet Dac Lai, Franck Dernoncourt, and Thien Huu Nguyen. Exploiting the matching information in the support set for few shot event classification. *PAKDD*, 2020
- Viet Dac Lai, Tuan Ngo Nguyen, and Thien Huu Nguyen. Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. In *EMNLP*, pages 5405–5411, 2020
- Viet Dac Lai and Thien Nguyen. Extending event detection to new types with learning from keywords. In *WNUT Workshop (EMNLP)*, 2019

Achievements and Awards

Adobe Research Fellowship 2022
Adobe Inc.

Outstanding Demo Paper Award 2021
The 16th Conference of the European Chapter of the Association for Computational Linguistics

Best Graduate Teaching Assistant 2021
Department of Computer and Information Science, University of Oregon

Monbukagakusho Honors Scholarship 2018
Ministry of Education, Culture, Sports, Science and Technology, Japan

JAIST Scholarship 2017
Japan Advanced Institute of Science and Technology, Japan

Professional Academic Services

Reviewer
Reviewing for top-tier conferences, workshops, and journals

- ACL Rolling Review (2021-present, monthly)
- Neurocomputing Journal (2021)
- ACM Transactions on Information Systems (2021)
- The 28th International Conference on Computational Linguistics (COLING 2020)
- The AAAI-21 Workshop on Scientific Document Understanding (SDU@AAAI 2020, 2021)
- The 10th CCF International Conference on Natural Language Processing and Chinese Computing (NLPCC2021)

Workshop Organizer
Organizing workshops co-located with top-tier conferences on Natural Language Processing

- The COLING-2022 First Workshop on Transcript Understanding
- SemEval-2022 Shared Task 12 at NAACL-HLT
- The AAAI-2022 Workshop On Video Transcript Understanding