Mike Zhang

mikz@itu.dk • jjzha.github.io • github.com/jjzha

Research Interests

Focus: Working on automated high-quality Information Extraction from unstructured text with real-life use cases that have societal impact.

Areas: Natural Language Processing (Information Extraction, Active Learning, Distant Supervision, Transfer Learning).

Education

IT University of Copenhagen – Copenhagen, Denmark   Aug. 2023 (Expected)
Ph.D. Computer Science
- · Advised by Prof. Barbara Plank.

University of Groningen – Groningen, the Netherlands   Jan. 2020
*M.A. Information Science, **cum laude***
- · Advised by Dr. Antonio Toral Ruiz.
- · **Thesis:** The Effect of Translationese in Machine Translation Test Sets.

University of Groningen – Groningen, the Netherlands   Jun. 2018
B.Sc. Information Science
- · Advised by Prof. Barbara Plank.
- · **Thesis**: From Relations to Falsehood: Labeled Bilexical Dependencies for Deception Detection.
- · **Exchange Semester** at The Chinese University of Hong Kong (博文約禮), Hong Kong SAR

Professional Experience

Dataprovider.com B.V. – Groningen, the Netherlands   Jan. 2020 – Aug. 2020
Data Engineer
- · Built a classifier predicting SIC codes for company names, using Python and scikit-learn;
- · Improved the preprocessing steps for structuring 72M company addresses;
- · Maintained code quality with PEP8 standards, basic unit- and integration tests, Git CI/CD, and code reviews.

Dataprovider.com B.V. – Groningen, the Netherlands   Aug. 2019 – Dec. 2019
Research Engineer Intern
- · Developed multiple algorithms in Python for the generation of domain names;
- · Obtained 1M+ existing and undiscovered domain names for web extraction and indexing;
- · Integrated models to production using Elasticsearch and NGINX.

Peer-Reviewed Publications, *equal contribution

4. **M. Zhang** and B. Plank. *Cartography Active Learning.* To appear in Findings of the Association for Computational Linguistics: EMNLP, 2021. [[pdf](#)]

3. K. N. Jensen,* **M. Zhang**,* and B. Plank. *De-identification of privacy-related entities in job postings.* In Proceedings of the Nordic Conference on Computational Linguistics (NoDaLiDa), 2021. [[pdf](#)]

2. **M. Zhang** and A. Toral. *The Effect of Translationese in Machine Translation Test Sets.* In Proceedings of the Fourth Conference on Machine Translation (WMT), 2019. [[pdf](#)]

1. **M. Zhang**, R. David, L. Graumans, and G. Timmerman. *Grunn2019 at SemEval-2019 Task 5: Shared Task on Multilingual Detection of Hate.* In Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval), 2019. [[pdf](#)]

Core Competencies

Programming: Python (PyTorch, scikit-learn, Pandas, NumPy), R, Java, HTML
Tools: Git, Elasticsearch, Agile, Scrum, LaTeX
Languages: Dutch (Native), English (Fluent), Spanish (Basic), Mandarin (Basic), Danish (Basic)

Services

Reviewer

Association for Computational Linguistics Rolling Review (ACL Rolling Review)	2021
The 7th Workshop on Noisy User-generated Text (W-NUT)	2021
The SIGNLL Conference on Computational Natural Language Learning (CoNLL)	2021
The 57th Annual Meeting of the Association for Computational Linguistics (ACL)	2019

Volunteer

The Conference on Empirical Methods in Natural Language Processing (EMNLP)	2021
Computational Linguistics in the Netherlands (CLIN)	2019

Talks

Talk at the NLP & CL group of the University of Groningen (GroNLP)	2021

Teaching

IT University of Copenhagen

Research Project, M.Sc. in Computer Science (**Supervision**)	Fall 2021
Communicating State-of-the-art NLP Research to a Broader Audience (**PhD Course**)	Fall 2021
Introduction to Natural Language Processing and Deep Learning (**Senior Teaching Assistant**)	Spring 2021

University of Groningen

Learning from Data (**Head Teaching Assistant**)	Fall 2019
Machine Learning (**Head Teaching Assistant**)	Fall 2019, 2020
Social Media (**Teaching Assistant**)	Spring 2019