

Mike Zhang

mikz@itu.dk · +31 6 52895785 (personal), +45 7218 5204 (work) · Copenhagen, Denmark

 [jjzha.github.io](https://github.com/jjzha) ·  github.com/jjzha ·  linkedin.com/in/jjzha

Education

IT University of Copenhagen

Aug. 2023 (Expected)

PhD Computer Science | Supervisor: Prof. Dr. Barbara Plank

Copenhagen, Denmark

- Extracting relevant competencies from multilingual job vacancy data to create better job matching;
- Core research fields include Information Extraction, Active Learning, Distant Supervision.

University of Groningen

Jan. 2020

MA Information Science, cum laude | Supervisor: Dr. Antonio Toral Ruiz

Groningen, the Netherlands

- **Thesis:** *Translationese in the WMT: The Effect of Translationese in Machine Translation Test Sets*. Grade: 9.0/10.0.

University of Groningen

Jun. 2018

BSc Information Science | Supervisor: Prof. Dr. Barbara Plank

Groningen, the Netherlands

- **Thesis:** *From Relations to Falsehood: Labeled Bilexical Dependencies for Deception Detection*. Grade: 8.0/10.0;
- **Exchange Semester** at The Chinese University of Hong Kong (博文約禮), Hong Kong SAR (Spring 2018).

Professional Experience

Dataprovider.com B.V.

Jan. 2020 – Aug. 2020

Data Engineer

Groningen, the Netherlands

- Built a classifier predicting SIC codes for company names, using Python and scikit-learn;
- Improved the preprocessing steps for structuring company addresses;
- Maintained code quality with PEP8 standards, basic unit- and integration tests, Git CI/CD, and code reviews.

Dataprovider.com B.V.

Aug. 2019 – Dec. 2019

Data Scientist Intern

Groningen, the Netherlands

- Developed 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.

Publications, *equal contribution

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

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

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	Spring 2021

University of Groningen

Learning from Data	Fall 2019
Machine Learning	Fall 2019, 2020
Social Media	Spring 2019