

EDUCATION

IT University of Copenhagen

Ph.D. Natural Language Processing

Copenhagen, DK

2020–Current

- Supervisor: Prof. Barbara Plank.

University of Groningen

M.A. Information Science, *Cum Laude*

Groningen, NL

2018–2020

- Thesis: *Translationese in the WMT: The Effect of Translationese in Machine Translation Test Sets*.
Supervisor: Dr. Antonio Toral Ruiz.

University of Groningen

B.S. Information Science

Groningen, NL

2015–2018

- Thesis: *From Relations to Falsehood: Labeled Bilingual Dependencies for Deception Detection*.
Supervisor: Prof. Barbara Plank
- Exchange Semester at The Chinese University of Hong Kong, Hong Kong SAR during Spring 2018

EXPERIENCE

IT University of Copenhagen

PhD Fellow

Copenhagen, DK

Sep. 2020–Current

- Extracting information from multilingual job vacancy data with time-shift, as a result to create better job matching.
- *De-identification of Job Postings*: De-identification of personal data in Job Postings.
- Keywords: *active learning, distant supervision, cross-lingual learning, continual learning, domain shift, annotation, (E/I)SCO*

Dataprovider.com

Data Science Intern, Data Engineer

Groningen, NL

Aug. 2019–Aug. 2020

- *Companies*: Classification of companies according to the SIC taxonomy codes.
- *Domain Discovery*: Generation of undiscovered existing domain names, by exploiting domain names already found.
- Keywords: *extreme multiclass classification, distant supervision, SIC, NAICS, ensemble learning, web domain, deep learning*

PUBLICATIONS

- [1] K. N. Jensen, **M. Zhang**, and B. Plank, “De-identification of privacy-related entities in job postings”, *NoDaLiDa 2021*, p. 210, 2021.
- [2] **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*, 2019, pp. 391–395.
- [3] **M. Zhang** and A. Toral, “The effect of translationese in machine translation test sets”, in *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, 2019, pp. 73–81.

TEACHING

Teaching Assistant at IT University of Copenhagen

Spring 2021

Introduction to Natural Language Processing and Deep Learning (BSSEYEP1KU)

Teaching Assistant at University of Groningen

Spring 2019, Fall 2019, Fall 2020

Social Media (LIX017B05)

Learning from Data (LIX016M05)

Machine Learning (SOMINDW07)

SKILLS

Computer Skills

Python, R, HTML, Java

Libraries:

PyTorch, Pandas, Matplotlib, NumPy

Languages:

Dutch (native), English (near-native), Spanish (basic), Mandarin Chinese (basic), Danish (basic)

PROFESSIONAL SERVICE

Reviewer:

CoNLL 2021, WMT 2019