

# Mike Zhang

[jjzha.github.io](https://github.com/jjzha) | [github.com/jjzha](https://github.com/jjzha) | [linkedin.com/in/mikejj-zhang](https://www.linkedin.com/in/mikejj-zhang)  
mikejj.zhang(at)gmail(dot)com | +31 6 52895785 | Groningen, The Netherlands

## EDUCATION

### UNIVERSITY OF GRONINGEN | MA INFORMATION SCIENCE

Sep. 2018 - Jan. 2020 | Groningen, NL

Cum. GPA: 8.4 / 10.0 | Cum Laude

**Master Thesis:** *Translationese in the WMT: The Effect of Translationese in Machine Translation Test Sets*

- Researched the effect of translated text in MT, and its influence on the evaluation of MT systems.
- Looked into different grammatical structures in terms of perplexity in language models. Part-of-speech tags obtained with UDPipe and language models created with KenLM.
- Built upon prior research and presented results at WMT19 (Conference on Machine Translation).

### UNIVERSITY OF GRONINGEN | BSC INFORMATION SCIENCE

Sep. 2015 - Jun. 2018 | Groningen, NL

Cum. GPA: 7.6 / 10.0

**IASP Exchange Program:** *The Chinese University of Hong Kong, GPA: 8.0/10.0*

**Bachelor Thesis:** *From Relations to Falsehood: Labeled Bilexical Dependencies for Deception Detection.*

- Constructed an ML model and researched best approaches to address deception in Dutch product reviews.
- Implemented SVMs with part-of-speech- and dependency tags as features, obtained the syntactic features via the Alpino parser, reached up to 76.5% accuracy.

## EXPERIENCE

### DATAPROVIDER.COM | DATA ENGINEER

Jan. 2020 - Present | Groningen, NL

- Structuring unstructured company data with Python and Elastic.
- Building classifiers to predict 70M+ companies' NAICS taxonomy codes.
- Part of the Data Science team, researching applied methods for projects from stakeholders.

### DATAPROVIDER.COM | DATA SCIENTIST INTERN

Sep. 2019 - Dec. 2019 | Groningen, NL

- Developed a Machine Learning model for the generation of existing domain names with Python.
- Implemented Natural Language Processing algorithms (language modelling, word embeddings, character-based recurrent neural networks).
- Obtained 100,000+ existing domain names for scraping.

### UNIVERSITY OF GRONINGEN | TEACHING ASSISTANT

Apr. 2019 - Jan. 2020 | Groningen, NL

- *SOMINDW07, Fundamentals of Machine Learning*; shown theoretical and practical basis for ML experiments to Bachelor students.
- *LIX016M05, Learning from Data*; taught Postgraduates to master key concepts of ML and academically report their findings.
- *LIX017B05, Social Media*; taught a class of 85 Bachelor students how to do quantitative analysis on Twitter data.

## EXTRACURRICULAR

### ACL 2019 | SECOND REVIEWER

Apr. 2019 | Florence, IT

Co-reviewed an Association for ACL 2019 paper with Dr. A. Toral Ruiz.

### ASCI | TREASURER, SECRETARY

Apr. 2019 | Groningen, NL

- Part of the Career- and Travel committee of the Information Science Student Association.
- Organized ten career workshops and lectures by successfully contacting and inviting representatives from companies, through verbal and written communication.
- Organized a professional trip to Barcelona, Spain with 32 students. Visited companies such as Softonic and Ubisoft. Responsible for the travel budget of 10,000+ EUR.

## COURSEWORK

### GRADUATE (SELECTED)

Computational Semantics  
Semantic Web Technology  
Learning from Data  
Communication Technology

### UNDERGRADUATE (SELECTED)

Information Retrieval  
Statistics  
Machine Learning  
Computational Grammar  
Search Engines  
Programming in Python

## SKILLS

### PROGRAMMING

Python • R • Java • SQL

### TOOLS

Git • Elastic • Agile •  $\text{\LaTeX}$

### LANGUAGES

Dutch (Native) • English (Fluent) • Spanish (B1) • Mandarin (HSK2)

## CERTIFICATES

**Machine Learning** by Stanford University on Coursera (LNDVQ8TKJZTC)