

Curriculum Vitae

Judith van Stegeren

November 10, 2021

1 Personal facts

Full name Judith Emma van Stegeren
Home Lili Bleekerstraat 3
6533 HH Nijmegen
The Netherlands
+31 6 45 83 67 16
Born September 29, 1990, Utrecht, The Netherlands
Nationality Dutch
Email judith@jd7h.com

2 Experience

- January 2018 – present. (Defense planned on 25th of March 2022)
PhD Student, Human Media Interaction at University of Twente.
Research topic: Natural language generation for role-playing video games.
Supervisors: Dr. Mariët Theune, Prof. Dr. Dirk Heylen
Research project: DATA2GAME, collaboration between University of Twente, Thales, Tilburg University and Veiligheidsregio Twente funded by NWO. Published 8 peer-reviewed papers about natural language generation, natural language processing, sentiment analysis, data engineering, open data, text in video games. Developed various natural language processing prototypes. Gave talks at international conferences. Developed part of a course on Twitter bots. Developed research software and datasets for various colleagues as side-projects.
- January 2016 – December 2017.
Security Specialist, National Cyber Security Centre, Ministry of Security and Justice, NL.
Member of the Computer Emergency Response Team for the Dutch government. Daily activities: incident response, assessing the impact of security vulnerabilities, monitoring developments in the security field. Other responsibilities: monitoring research and innovation in operational security technology, sharing knowledge by giving technical talks, writing a newsletter about developments in the field, inviting speakers and organizing training sessions.
- August 2013 – February 2014.
Teaching assistant for course "Logic for AI students", Radboud University Nijmegen.
First-year course on propositional and predicate logic, languages and machines, discrete mathematics and modal logic. Activities: assisting the students with their homework during exercise classes, marking and correcting homework and exams.

3 Education

- MSc. in Computer Science, *Bene Meritum*, Radboud University Nijmegen, January 2016. Specialization in Mathematical Foundations of Computer Science, minor in Digital Security. Thesis topic: automata learning, model-checking, correctness; supervisor: Prof. Dr. F.W. Vaandrager.
- BSc. in Computer Science, *Bene Meritum*, Radboud University Nijmegen, August 2013.

4 Publications (selection)

These publications give a good overview of the topics I researched during my PhD. For a full list of my peer-reviewed publications, see <https://judithvanstegeren.com/publications/>. All publications are freely available online via open access.

1. Judith van Stegeren and Jakub Myśliwiec. “Fine-tuning GPT-2 on annotated RPG quests for NPC dialogue generation”. In: *Proceedings of the 16th International Conference on the Foundations of Digital Games (FDG) 2021*. FDG '21. Online: ACM, Aug. 2021
Topic: natural language generation, deep learning, Transformers, large language models, GPT-2, machine learning, video games
2. Lorenzo Gatti and Judith van Stegeren. “Improving Dutch sentiment analysis in Pattern”. In: *Computational Linguistics in the Netherlands Journal* 10 (Dec. 2020)
Topic: research software, sentiment analysis, Pattern, natural language processing, Dutch
3. Judith van Stegeren and Mariët Theune. “Fantastic Strings and Where to Find Them: The Quest for High-Quality Video Game Text Corpora”. In: *Proceedings of the 2020 Workshop on Intelligent Narrative Technologies*. Oct. 2020
Topic: data engineering, data extraction, parsing, open data, corpora, video games
4. Thérèse Bergsma, Judith van Stegeren, and Mariët Theune. “Creating a Sentiment Lexicon with Game-Specific Words for Analyzing NPC Dialogue in The Elder Scrolls V: Skyrim”. English. In: *Workshop on Games and Natural Language Processing*. Marseille, France: European Language Resources Association, May 2020, pp. 1–9. ISBN: 979-10-95546-40-5. URL: <https://www.aclweb.org/anthology/2020.gamnlp-1.1>
Topic: sentiment analysis, machine learning, natural language processing, domain specific language, underresourced languages, video games

5 Extracurricular activities (selection)

1. Owner, Datakami research & development. 2020–present.
2. Chair of investing club Rucessie. 2020–present.
3. Invited speaker, Radboud University. 2016–present.
Activities: giving talks about a variety of topics¹ for bachelor students, master students and scientific staff members of the Science Faculty.
4. Volunteer, Stichting #RU. 2010–present.
Activities: organizing events, organizing training sessions, coaching, giving talks.
5. Member of the Women in CyberSecurity Capture The Flag competition team. 2016–2017.
Activities: organising training sessions, participating in Capture the Flag tournaments, notably the CTF at Hack in The Box 2016.
6. Owner, Meteoriet Design. Graphic design. 2009–present.

6 Technical proficiencies (short version)

- Operating systems: Windows, Linux (Arch Linux, Debian/Ubuntu, Kali), iOS, Android
- Languages: Python, Java, C/C++, Haskell, Clean, Processing, JavaScript and R. Bash, Powershell. HTML, CSS, LaTeX, Markdown, SQL, and XML.
- Data gathering: urllib, beautifulsoup, selenium, scrapy, various APIs
- Statistical analysis: pandas, numpy, scipy, R, jupyter
- Machine learning: scikit-learn, pytorch

¹For an overview of my talks, see: <https://judithvanstegeren.com/publications/#talks>

- Natural language processing: NLTK, Pattern, SpaCy, Transformers
- Visualization: matplotlib, seaborn, R, GraphViz, DOT, TikZ, Prometheus, Grafana, Processing
- Web applications: HTML, CSS, javascript, Django, Flask, Hugo, Jekyll, sqlite, MySQL
- Cloud: Debian, SLURM, Google Cloud Platform.

7 Technical proficiencies (long version)

I'm familiar with Windows, multiple flavors of Linux, iOS and Android. I use Linux for development and other technical tasks. I run Arch Linux on my personal laptops and manage multiple VPSes (virtual private servers) that typically run Debian. In the past I've also used Ubuntu, Linux Mint and Kali Linux for projects in the information security domain. I heavily use bash (Linux commandline) and related tools such as grep, awk, sed, ssh, scp, etc. On Windows, I will automate repetitive tasks with Powershell. I use git for collaborations and version-control.

I am comfortable with the following programming languages: Python, Java, C/C++, Haskell, Clean, Processing, JavaScript and R. I also know my share of mark-up and query languages such as HTML, CSS, LaTeX, Markdown, SQL, and XML. Which programming language I use depends on the goals of the project. Since my current projects involve data gathering, data processing and analysis, I tend to use Python.

Most of my projects start with data gathering, so I am proficient in writing web scrapers and extracting data using APIs. For web scraper development I have used Python's urllib with BeautifulSoup for parsing, scrapy, and selenium for web scrapers that have to function in a realistic browser environment. I have worked with the APIs of Spotify, Twitter, Facebook, Last.fm, Instagram, Amazon.com, Bol.com, GoodReads, and openlibrary. Additionally I have developed web scrapers to extract information from websites such as ah.nl (grocery stores), funda.nl (real estate market), CBS.nl (Statistics Netherlands), KVK (Chamber of Commerce), Yahoo Finance, Bloomberg and EuroNext.

For statistical analysis, I use Python's numpy, scipy and pandas, or the language R. I prefer to do data exploration in Jupyter notebooks. For machine learning, I use scikit-learn. I am currently studying PyTorch to increase my implementation skills for deep learning. For fast, basic, rule-based natural language processing I use NLTK and Pattern, and for more complex problems I use SpaCy and Huggingface's transformers library. I choose a machine learning approach based on multiple parameters of the project: where will the model be deployed, how large is the relevant data, does the machine learning model have to be fully transparent, will the model function fully autonomously or under supervision, etc.

For data visualization, I use matplotlib and seaborn in Python, or R. In the past I have used GraphViz, the DOT language, and TikZ for complex graph making. I have also created various dashboards by combining my web scrapers with Prometheus and Grafana. The dashboards are updated in real-time, and have monitoring, critical failure detection and email alerting functionality.

I have experience with both frontend and backend development of web applications. Apart from basic web development in HTML, CSS, and Javascript, I have used static site generators Jekyll (written in Ruby) and Hugo (written in Go). I am currently developing an open source book management web application in Django (Python). I have created APIs and small web applications with flask (Python). My web applications use sqlite and MySQL/MariaDB databases.

In terms of remote computing, distributed computing and cloud computing: I have extensive experience managing Debian VPSes. Among others, I run my own webserver (nginx) and git-server (gitolite). Most of my research code was deployed on the High Performance Computing cluster at the University of Twente (SLURM with Ubuntu controller nodes). I also have some experience with Google Cloud Platform, and I'm currently investigating deploying machine learning models on Azure and Amazon Web Services.