



Education

- 2018–2023 **PhD studies**, *Humboldt-Universität in Berlin*, Berlin
Grade summa cum laude
- 2016–2018 **MSc Computer Science**, *Humboldt-Universität in Berlin*, Berlin
Grade 1.0
- 2013–2016 **BSc Computer Science**, *Freie Universität*, Berlin, Grade 1.4
- 2010–2013 **BA Philosophy/English studies**, *University of Bamberg*, Bamberg, Grade 1.3
- 2000–2009 **Abitur (equivalent to A level)**, *Ehrenbürg-Gymnasium*, Forchheim, Grade 1.8

Professional Experience

Research

- 11/2022– **Post-doctoral Researcher**, *Human-centric NLP & Machine Learning*, LMU Munich
Post-doctoral researcher in the Artificial Intelligence & Computational Linguistics lab. Developed machine learning and visualization techniques to semi-automatically detect annotation errors and ambiguous instances in gold standard datasets. Teaching and supervision in machine learning and NLP
- 11/2018–11/2022 **Doctoral Researcher**, *Text Mining for Pathway Curation*, Humboldt-Universität zu Berlin
PhD studies on developing machine-learning methods to extract structured information from biomedical text. Devised and implemented diverse machine-learning methods to process language, graphs and time series. Developed and supported user-friendly applications based on the research findings. Teaching and supervision in machine learning and NLP
- 07/2018–10/2018 **Researcher**, *Trend Detection in Biological and Biomedical Patent Documents*, Humboldt-Universität zu Berlin
Joint project with Bayer R&D Information. Developed methodology and evaluation framework for trend detection in biomedical patent documents based on large-scale representation learning of patent texts and outlier detection. Unpublished
- 11/2017–02/2018 **Student Research Assistant**, *Friedrich-Alexander-Universität*, Erlangen-Nürnberg
Deep Learning for biomedical time-series segmentation

- 04/2016– **Student Research Assistant**, *Humboldt Universität zu Berlin, Knowledge Management in Bioinformatics Group*, Berlin
10/2018 Machine Learning for Natural Language Processing and Genetics
- 02/2017 – **Research Intern**, *Max Dehlbrück Center, Computational Regulatory Genomics Group*, Berlin
05/2017 Deep Learning for Regulatory Genomics
- 02/2017 – **Contribution to Project**, *Free University*, Berlin
05/2017 Automated Higher Order Logic Theorem Proving in Philosophy
-

Open Source Contributions

- 11/2022– **Main Developer**, *Large-scale Information Extraction*, [PEDL+](#)
Main developer of PEDL+. Python library and command-line application for extracting protein-centric information from over 25 million scientific publications with machine learning
- 11/2021– **Open Source Research Contributor**, *Biomedical Language Modelling*, [BigBio](#)
Core contributor in the Huggingface BigScience biomedical working group. Part of a four-person team that designed, organized and carried out a hackathon that standardized over 125 biomedical NLP datasets with the help of roughly 50 participants. Led team of five for creating an massive multi-task learning model on the datasets.
- 04/2020– **Open Source Contributor**, *Natural Language Processing*, [flair](#)
current Co-lead of the HunFlair project that added biomedical Named Entity Recognition models to the flair NLP library. Designed and implemented data and model architecture.
-

Other

- 03/2015– **Student Employee**, *Carmeq GmbH*, Berlin
12/2015 Software Engineering in Java and Ruby
- 04/2014– **Student Employee**, *Freie Universität, Physics Department*, Berlin
03/2015 Software Engineering and System Administration

Awards & Competitions

- 2021 **First Place at Shared Task**, *DrugProt - BioCreative VII*, Shared task on extracting chemical-protein relations from the biomedical literature, First ranking team out of 30
- 2021 **Second Place at Shared Task**, *BioNLP - MEDIQA 2021*, Shared task on summarizing consumer health questions, Second ranking team out of 23
- 2019 **First Place at Shared Task**, *CLEF eHealth Multilingual Information Extraction*, Shared task on code assignment for German animal experiment summaries, First ranking team out of six
- 2018 **Award for Best Master's Thesis**, *Department of Computer Science, Humboldt-Universität zu Berlin*, NLPprolog – Reasoning with Weak Unification for NLP, One out of two awarded theses

Languages

- German native
English fluent

Computer Skills

Languages	Python, Java
Machine Learning Frameworks	PyTorch, huggingface transformers, scikit-learn, flair, Keras, Tensorflow 1.0
Other	GNU/Linux, Docker, Git, L ^A T _E X, IntelliJ IDEs, VIM, Microsoft Office

Publications (selected)

For a full list, see [Google Scholar](#)

- [1] R. Litschko, M. Müller-Eberstein, R. van der Goot, L. Weber, and B. Plank, “Establishing trustworthiness: Rethinking tasks and model evaluation,” in *EMNLP 2023 (accepted)*, 2023.
- [2] L. Weber, F. Barth, L. Lorenz, *et al.*, “Pedl+: Protein-centered relation extraction from pubmed at your fingertip,” *Bioinformatics (accepted)*, 2023.
- [3] L. Weber-Genzel, R. Litschko, E. Artemova, and B. Plank, “Donkii: Can annotation error detection methods find errors in instruction-tuning datasets?” *CoRR*, vol. abs/2309.01669, 2023. arXiv: [2309.01669](#).
- [4] L. Weber and B. Plank, “ActiveAED: A human in the loop improves annotation error detection,” in *Findings of the Association for Computational Linguistics: ACL 2023*, Jul. 2023.
- [5] J. A. Fries, L. Weber, N. Seelam, *et al.*, “BigBio: A Framework for Data-Centric Biomedical Natural Language Processing,” in *Thirty-Sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, New Orleans, Louisiana, USA, Oct. 2022.
- [6] L. Weber, J. Münchmeyer, S. Garda, and U. Leser, “Extend, don’t rebuild: Phrasing conditional graph modification as autoregressive sequence labelling,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021.
- [7] L. Weber, M. Sängler, J. Münchmeyer, M. Habibi, U. Leser, and A. Akbik, “HunFlair: an easy-to-use tool for state-of-the-art biomedical named entity recognition,” *Bioinformatics*, vol. 37, no. 17, Jan. 2021.
- [8] L. Weber, K. Thobe, O. A. Migueles Lozano, J. Wolf, and U. Leser, “PEDL: extracting protein–protein associations using deep language models and distant supervision,” *Bioinformatics*, vol. 36, no. Supplement_1, Jul. 2020.
- [9] L. Weber, J. Münchmeyer, T. Rocktäschel, M. Habibi, and U. Leser, “HUNER: improving biomedical NER with pretraining,” *Bioinformatics*, Jun. 2019, btz528.
- [10] L. Weber, P. Minervini, J. Münchmeyer, U. Leser, and T. Rocktäschel, “NLProlog: Reasoning with weak unification for question answering in natural language,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy: Association for Computational Linguistics, Jul. 2019.
- [11] M. Habibi, L. Weber, M. Neves, D. L. Wiegandt, and U. Leser, “Deep learning with word embeddings improves biomedical named entity recognition,” *Bioinformatics*, vol. 33, no. 14, Jul. 2017.