Jonas Rieger

rieger@statistik.tu-dortmund.de
https://jonasrieger.github.io/

EDUCATION

Doctoral degree *[Doktor der Naturwissenschaft]***, Dr. rer. nat.** | *Statistics* — Sept. 2022
TU Dortmund University — Dortmund, Germany
- Reliability evaluation and an update algorithm for the latent Dirichlet allocation

Master of Science, M. Sc. | *Statistics* — Nov. 2018
TU Dortmund University — Dortmund, Germany

Bachelor of Science, B. Sc. | *Statistics* — Oct. 2016
TU Dortmund University — Dortmund, Germany

PREVIOUS AND CURRENT POSITIONS

NLP Scientist | *[50% position]* — since Oct. 2022
Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI) — Hamburg, Germany
- Media Research Methods Lab (MRML)

Postdoc | *[50% position]* — since Oct. 2022
TU Dortmund University — Dortmund, Germany
- Department of Statistics: Chair of Business and Social Statistics
- Teaching in the amount of 2 hours per week

Doctoral student — Dec. 2018 – Sept. 2022
TU Dortmund University — Dortmund, Germany
- Department of Statistics: Chair of Business and Social Statistics
- Teaching in the amount of 4 hours per week

AFFILIATIONS

Member of the TU Dortmund Young Academy — since June 2023
Member of DoCMA (Dortmund Center for Data-based Media Analysis) — since Dec. 2018

FUNDING

TU Dortmund Young Academy — June 2023
The Era of ChatGPT: Evaluation and Regulation of Large Language Models — 5000€ for a GPU Workstation

PROJECTS

NEAR — since Oct. 2023
Narrative Economics Alliance Ruhr — MERCUR

FLACA — since Oct. 2022
Few-Shot Learning for Automated Content Analysis in Communication Science — BMBF

GADMO — since Oct. 2022
German-Austrian Digital Media Observatory — EU

Research Interests

Methodological research

- Evaluation of topic models (i.a., quality and reliability) and other NLP systems
- Parameter efficiency (e.g., adapters) for large language models in few-shot scenarios
- Model selection and parameter tuning of topic models
- Update algorithms and monitoring settings for topic models
- Detection of structural breaks, events and narratives in text corpora

Software engineering in R

- Author and maintainer of rollinglda and ldaPrototype, co-author and contributor of tosca and spINAR

Applied research

- Text corpus-based indicators
- Content analysis of texts and tweets of political parties and parliamentarians
- Argument mining in news and social media debates
- Characteristics of disinformation and fact-checks

Service to the Research Community

Reviewing

- Advances in Statistical Analysis
- Communication Methods and Measures
- Computational Intelligence
- Educational and Psychological Measurement
- Statistical Papers: DOI: 10.1007/s00362-019-01126-7

Invited Talks

ZPID Lecture Series　　　　　　　　　　　　　　　　　　　　　　　Dec. 2022
Keep rollin'! The abilities for monitoring growing corpora using RollingLDA　　　Trier, Germany

Contributions to Conferences and Workshops

DiTox'23 Workshop @LDK 2023　　　　　　　　　　　　　　　　　Sep. 2023
Debunking Disinformation with GADMO: A Topic Modeling Analysis　　　　Vienna, Austria
of a Comprehensive Corpus of German-language Fact-Checks

Statistische Woche 2023　　　　　　　　　　　　　　　　　　　　Sep. 2023
Bekämpfung von Desinformation durch GADMO: Analyse eines umfassenden　Dortmund, Germany
deutschsprachigen Faktencheck-Korpus mithilfe von Topic Modellen

Statistische Woche 2023　　　　　　　　　　　　　　　　　　　　Sep. 2023
Scrutinizing ChatGPT against Few-Shot Learning with Adapter Extensions　Dortmund, Germany
and XLM-RoBERTa: A Case Study on Identifying Claims, Arguments and their Stance
in the German News Media Debate on Arms Deliveries to Ukraine

ECREA PolComm 2023　　　　　　　　　　　　　　　　　　　　Aug. 2023
Beyond "Master Frames": A Semi-automated Approach to　　　　　　　　Berlin, Germany
Studying Viewpoint Diversity of the Media Discourse

DGPuK 2023　　　　　　　　　　　　　　　　　　　　　　　　　May 2023
Few-shot learning for automated content analysis:　　　　　　　　　　Bremen, Germany
Efficient coding of arguments and claims in the debate on arms deliveries to Ukraine

MUFin'23 Workshop @AAAI 2023　　　　　　　　　　　　　　　　Feb. 2023
Early Warning Systems? Building Time Consistent Perception Indicators　Washington, DC, USA
for Economic Uncertainty and Inflation Using Efficient Dynamic Modeling

SDP'22 Workshop @COLING 2022　　　　　　　　　　　　　　　　Oct. 2022
Finding scientific topics in contionuously growing text corpora　　Gyeongju, Republic of Korea

Statistische Woche 2022	Sep. 2022
Monitoring consistent topics in continuously growing scientific text corpora	Münster, Germany
Text2Story'22 Workshop @ECIR 2022	Apr. 2022
Dynamic change detection in topics based on rolling LDAs	Stavanger, Norway
DAGStat 2022	Mar. 2022
Improving the reliability of LDA results using LDAPrototype as selection criterion	Hamburg, Germany
EMNLP 2021	Nov. 2021
RollingLDA: An Update Algorithm of Latent Dirichlet Allocation to Construct Consistent Time Series from Textual Data	Punta Cana, Dominican Republic
EDML'20 Workshop @ECML PKDD 2020	Sep. 2020
Assessing the Uncertainty of the Text Generating Process using Topic Models	Online
NLDB 2020	June 2020
Improving Latent Dirichlet Allocation: On Reliability of the Novel Method LDAPrototype	Online
Statistische Woche 2019	Sep. 2019
Quantifizierung der Stabilität der Latent Dirichlet Allocation mithilfe von Clustering auf wiederholten Durchläufen	Trier, Germany
DGPuK 2019	May 2019
Softwaretools für die Kommunikationsforschung	Münster, Germany
DAGStat 2019	Mar. 2019
Measuring Stability of Replicated LDA Runs	Munich, Germany

PUBLICATIONS

Dissertation

- Rieger (2022). "Reliability evaluation and an update algorithm for the latent Dirichlet allocation". *TU Dortmund University*. DOI: 10.17877/DE290R-22949.

Peer-reviewed publications

- Rieger, Hornig, Flossdorf, Müller, Mündges, Jentsch, Elmer (2023). "Debunking Disinformation with GADMO: A Topic Modeling Analysis of a Comprehensive Corpus of German-language Fact-Checks". In: *Proceedings of the 1st Workshop on Disinformation and Toxic Content Analysis*. URL: https://github.com/GADMO-EU/DiTox2023/blob/master/paper.pdf.
- Rieger, Hornig, Schmidt, Müller (2023). "Early Warning Systems? Building Time Consistent Perception Indicators for Economic Uncertainty and Inflation Using Efficient Dynamic Modeling". In: *Proceedings of the 3rd Workshop on Modelling Uncertainty in the Financial World*. URL: https://github.com/JonasRieger/mufin23/blob/master/paper.pdf.
- Bittermann, Rieger (2022). "Finding scientific topics in continuously growing text corpora". In: *Proceedings of the 3rd Workshop on Scholarly Document Processing*, pp. 7–18. URL: https://aclanthology.org/2022.sdp-1.2.
- Lange, Rieger, Benner, Jentsch (2022). "Zeitenwenden: Detecting changes in the German political discourse". In: *Proceedings of the 2nd Workshop on Computational Linguistics for Political Text Analysis*. URL: https://old.gscl.org/en/arbeitskreise/cpss/cpss-2022/workshop-proceedings-2022.
- Rieger, Lange, Flossdorf, Jentsch (2022). "Dynamic change detection in topics based on rolling LDAs". In: *Proceedings of the Text2Story'22 Workshop*. CEUR-WS. URL: http://ceur-ws.org/Vol-3117/.
- Rieger, Jentsch, Rahnenführer (2021). "RollingLDA: An Update Algorithm of Latent Dirichlet Allocation to Construct Consistent Time Series from Textual Data". In: *Findings Proceedings of the 2021 EMNLP-Conference*. ACL, pp. 2337–2347. DOI: 10.18653/v1/2021.findings-emnlp.201.
- von Nordheim, Rieger, Kleinen-von Königslöw (2021). "From the Fringes to the Core - An Analysis of Right-Wing Populists' Linking Practices in Seven EU Parliaments and Switzerland". In: *Digital Journalism*, pp. 1–19. DOI: 10.1080/21670811.2021.1970602.
- von Nordheim, Koppers, Boczek, Rieger, Jentsch, Müller, Rahnenführer (2021). Die Entwicklung von Forschungssoftware als praktische Interdisziplinarität. In: *M&K Medien & Kommunikationswissenschaft* 69, pp. 80–96. DOI: 10.5771/1615-634X-2021-1-80.

- Rieger, Jentsch, Rahnenführer (2020). "Assessing the Uncertainty of the Text Generating Process Using Topic Models". In: *Proceedings of the ECML PKDD 2020 Workshops*. Vol. 1323. CCIS. Springer, pp. 385–396. DOI: `10.1007/978-3-030-65965-3_26`.
- Rieger (2020). "ldaPrototype: A method in R to get a Prototype of multiple Latent Dirichlet Allocations". In: *Journal of Open Source Software* 5.51, p. 2181. DOI: `10.21105/joss.02181`.
- Rieger, Rahnenführer, Jentsch (2020). "Improving Latent Dirichlet Allocation: On Reliability of the Novel Method LDAPrototype". In: *Natural Language Processing and Information Systems*, NLDB 2020. Vol. 12089. LNCS. Springer, pp. 118–125. DOI: `10.1007/978-3-030-51310-8_11`.
- von Nordheim, Rieger (2020). "Im Zerrspiegel des Populismus - Eine computergestützte Analyse der Verlinkungspraxis von Bundestagsabgeordneten auf Twitter". In: *Publizistik* 65, pp. 403–424. DOI: `10.1007/s11616-020-00591-7`.

Selected non-peer reviewed publications (preprints, working papers and datasets)

- Rieger, Jentsch, Rahnenführer (2022). "LDAPrototype: A Model Selection Algorithm to Improve Reliability of Latent Dirichlet Allocation". Submitted to: *Knowledge and Information Systems*. Preprint available at Research Square. DOI: `10.21203/rs.3.rs-1486359/v1`.
- Lange, Rieger, Jentsch (2022). "Lex2Sent: A bagging approach to unsupervised sentiment analysis". Preprint available at arXiv. DOI: `10.48550/arXiv.2209.13023`.
- Shrub, Rieger, Müller, Jentsch (2022). "Text data rule - don't they? A study on the (additional) information of Handelsblatt data for nowcasting German GDP in comparison to established economic indicators". In: *Ruhr Economic Papers* #964. DOI: `10.4419/96973128`.
- Jentsch, Mammen, Müller, Rieger, Schötz (2021). "Text mining methods for measuring the coherence of party manifestos for the German federal elections from 1990 to 2021". In: *DoCMA Working Paper* #8. DOI: `10.17877/de290r-22363`.
- Rieger, von Nordheim (2021). "corona100d - German-language Twitter dataset of the first 100 days after Chancellor Merkel addressed the coronavirus outbreak on TV". In: *DoCMA Working Paper* #4. DOI: `10.17877/DE290R-21911`.

TEACHING EXPERIENCE

Introduction to Topic Modeling | *Seminar (English)*	SuSe 2023
University of Bremen	

Text as Data | *Lecture (English)*	WiSe 2022/23 – 2023/24
Data Mining Cup | *Seminar (English)*	SuSe 2019 – 2023
Einführung in LATEX | *Compact course (German)*	SuSe 2019 – 2022
Fallstudien I | *Seminar (German)*	WiSe 2021/22
Schätzen und Testen | *Organization (German)*	WiSe 2021/22
Nichtparametrische Verfahren | *Exercise (German)*	WiSe 2020/21
Text Data meets Econometrics | *Seminar (English)*	WiSe 2020/21
Entscheidungstheorie - Statistik VI | *Exercise (German)*	SuSe 2020
Wahrscheinlichkeitstheorie - Statistik V | *Exercise (German)*	WiSe 2019/20
Textdatenanalyse | *Seminar (German)*	SuSe 2019
TU Dortmund University	

SUPERVISED THESES

Master

- Comparison of Diachronic Embeddings with Pre-trained Model Embeddings for Historical Texts
  (Priyanka Madiraju, 2023)
- Text Data-based Nowcasting of German GDP Growth Using Newspaper Data
  (Yuliya Shrub, 2022)
- Resampling strategies for unsupervised sentiment analysis using lexicon-based text embedding methods
  (Kai-Robin Lange, 2021)

Bachelor

- Comparison of Active Learning techniques for the benefit of data set generation in the field of text mining
  (Jannik Bloß, 2023)