

Dr. Jelena Mitrović

**Nachwuchsforscherinnengruppe CAROLL
Lehrstuhl für Data Science**



[01IS20049] Abschlussbericht Forschungsprojekt CAROLL | Teil I. Kurzbericht

Vorhabenbezeichnung: **CAROLL: Computergestützte Rhetorik in Sozialen Medien und Recht**

Laufzeit des Vorhabens: 01.10.2020 – 30.09.2024

Berichtszeitraum: 01.10.2020 bis 30.09.2024

Autorin: Jelena Mitrović

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS20049 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin.

1. Aufgabenstellung

Künstliche Intelligenz (KI) und die Verarbeitung natürlicher Sprache (NLP) haben im letzten Jahrzehnt rasante Fortschritte gemacht. Word2Vec [1], GloVe [2], Sprachmodelle aus der BERT-Familie [3] haben sich als sehr effektiv erwiesen. Allerdings sind Sprachmodelle allein nicht in der Lage, den gesunden Menschenverstand und die Komplexität einiger NLP-Aufgaben zu bewältigen. Der Weg in die Zukunft besteht darin, auf hybride Ansätze zurückzugreifen, die sich auf automatische Schlussfolgerungen und ontologisches Engineering stützen, was zu einem wirklich tieferen Verständnis von Sprache führen kann. Dies war unser Ziel. Mit der BMBF Bekanntmachung „Förderung von KI-Nachwuchswissenschaftlerinnen“ vom 19. Juni 2019 wird das Ziel verfolgt, den Anteil qualifizierter Frauen in Führungspositionen der deutschen KI-Forschung zu steigern und den Einfluss von Wissenschaftlerinnen auf das Themengebiet nachhaltig zu stärken. Die CAROLL-Nachwuchsgruppe wurde gefördert durch das BMBF-Programm für Frauen in der KI - KI03 KI-Nachwuchswissenschaftlerinnen.

2. Ablauf des Vorhabens

Wir haben unsere Arbeit auf die Prinzipien der Computational Rhetoric gestützt, um zu untersuchen, wie rhetorische Figuren gesellschaftliche Probleme wie Hassrede, Propaganda, Fehlinformation und Medienbias beeinflussen. Wir haben computergestützte rhetorische Methoden im Kontext von sozialen Medien und Recht eingesetzt. Die Forschungsziele von CAROLL basierten auf den folgenden Forschungspfeilern, die in Abschnitt 3 näher erläutert werden:

1. Ontologische Modellierung von rhetorischen Figuren und Umsetzung von Lösungen komplexer gesellschaftlicher Phänomene
2. Legal Tech und Legal Natural Language Processing
3. Verbesserung der großen Sprachmodelle und Open Web Search

3. Wesentliche Ergebnisse

1. Das Gebiet der computergestützten Rhetorik steht im Mittelpunkt der CAROLL-Forschung. Ein wichtiger Teil davon ist die Modellierung von Ontologien rhetorischer Figuren in OWL und deren Verwendung zur Verbesserung lexikalischer Ressourcen und Werkzeuge für die Erkennung komplexer rhetorischer Figuren [4,5]. Diese Forschungslinie wurde im CAROLL vertieft, insbesondere durch die Entwicklung eines hochmodernen BERT-Sprachmodells, HateBERT [6], und anderer, ähnlicher Systeme zur Erkennung von Hassrede, beleidigender und anstößiger Sprache [7,8,9,10].

2. Unsere Arbeit im Bereich der Rechtstechnologien in CAROLL begann mit der ontologischen Modellierung von Rechtsbegriffen in SUMO und setzte sich fort mit der Verwendung verschiedener Systeme für maschinelles Lernen und tiefe neuronale Netze für den Entwurf und die Implementierung deutscher juristischer Entscheidungskorpora [11] und die Klassifizierung von Teilen des deutschen juristischen Schreibstils in deutschen Rechtsurteilen [12]. Wir haben das deutsche BERT-Modell für Named Entity Recognition (NER) [13], einen Datensatz mit deutschen juristischen Referenzannotationen [14] und einen Datensatz für GDPR-konforme NER für Datenschutzrichtlinien entwickelt. Wir untersuchten die semantische Ähnlichkeit zwischen deutschen Rechtstexten und verweisenden Gesetzen [15] und stellten die Herausforderungen und Überlegungen bei der Annotation von Rechtsdaten vor [16].

3. Die Forschung im Rahmen des Open Web Search [17], konzentrierte sich auf verschiedene Aspekte der Verbesserung von Web-Crawling- und Indexierungsansätzen sowie auf den Aufbau hochwertiger Datensätze zur Erleichterung neuer Paradigmen des maschinellen Lernens und der Large Language Models in Kombination mit Information Retrieval, z. B. Retrieval Augmented Generation, wie in einer unserer Veröffentlichungen zu sehen ist [18].

Die Sichtbarkeit unserer Forschung wurde deutlich erhöht, was sich in zahlreichen Artikeln in den Medien widerspiegelt, z.B. in *Forschung in Bayern*, *idw-Informationsdienst Wissenschaft*, *Universität Passau Digital Research Magazine* <https://www.digital.uni-passau.de/en/stories/2020/jelena-mitrovic/>. Der CAROLL-Workshop am 31.03.2023, der zeitgleich mit dem 5th Passau Data Science Summit (PasDaS) stattfand, war das Ergebnis dieser erhöhten Sichtbarkeit. Dr. Jelena Mitrović hielt zahlreiche Vorträge über die Forschung der CAROLL-Gruppe, z. B. auf dem *KI-Seminar des tschechischen Instituts für Informatik, Robotik und Kybernetik*; sie war eine eingeladene Rednerin auf der *8th International ScaDS.AI Summer School 2022*; zuletzt nahm Dr. Mitrović am *Dagstuhl-Seminar über Grand Challenges for Research on Privacy Documents* teil.

ANHANG 1

Bibliographie

- [1] T Mikolov, I Sutskever, K Chen, G S Corrado and J Dean (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, vol. 26, p. 3111–3119.
- [2] J Pennington, R Socher and C Manning (2014). Glove: Global vectors for word representation. In *EMNLP*.
- [3] J Devlin, C Ming-Wei, K Lee and K Toutanova (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*.
- [4] M Mladenović, J Mitrović, C Krstev, R Stanković, (2017). Using Lexical Resources for Irony and Sarcasm Classification. *The 8th Balkan Conference in Informatics, Skopje, Macedonia*.
- [5] J Mitrović, M Mladenović and C Krstev, (2016). A Language-independent Model for Adding New Semantic Relations to a WordNet. *Global WordNet Conference, Bucharest, Romania*.
- [6] T Caselli, V Basile, J Mitrović, M Granitzer, (2021). HateBERT: Retraining BERT for Abusive Language Detection in English. *WOAH Workshop @ ACL 2021*.
- [7] S Kannan and J Mitrović (2022). Hate speech and Offensive Content Detection in Hindi Language using C-BiGRU. *Forum for Information Retrieval Evaluation (Working Notes) (FIRE)*.
- [8] R Stanković, J Mitrović, D Jokić, C Krstev, (2020). Multi-word Expressions for Abusive Language Detection in Serbian. In *Proceedings of the 28th International Conference on Computational Linguistics COLING 2020*.
- [9] B Lewandowska-Tomaszczyk, S Žitnik, A Bączkowska, C Liebeskind, J Mitrović and G Valunaite Oleskeviciene, (2021). LOD-connected offensive language ontology and tag set enrichment. *SALLD @LDK 2021 – 3rd Conference on Language, Data and Knowledge*.
- [10] D Colla, T Caselli, V Basile, J Mitrović, M Granitzer, (2020). GruPaTo at SemEval-2020 Task 12: Retraining mBERT on Social Media and Fine-tuned Offensive Language Models. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval)*.
- [11] S Urchs, J Mitrović. M Granitzer, (2021). *Design and Implementation of German Legal Decision Corpora, ICAART, Vienna, Austria*.
- [12] S Urchs, J Mitrović, M Granitzer, (2020). Towards Classifying Parts of German Legal Writing Styles in German Legal Judgments. In *10th International Conference on Advanced Computer Information Technologies (ACIT)*.
- [13] H Darji, J Mitrović, M Granitzer, (2023). German BERT Model for Legal Named Entity Recognition. *International Conference on Agents and Artificial Intelligence ICAART*.
- [14] H Darji, J Mitrović, M Granitzer, (2023). A Dataset of German Legal Reference Annotations. *19th International Conference on Artificial Intelligence and Law - ICAIL 2023*.
- [15] H Darji, J Mitrović, M Granitzer, (2023). Exploring Semantic Similarity between German Legal Texts and Referred Laws. *Communications in Computer and Information Science Book Series, Springer*.
- [16] H Darji, J Mitrović, M Granitzer, (2023). Challenges and Considerations in Annotating Legal Data: A Comprehensive Overview. *JURIX 2023*.
- [17] M Granitzer, S Voigt, F, Noor Afshan, M Golasowski, C Guetl, D Hiemstra, J Martinovič, J Mitrović et. al, (2023). Impact and development of an Open Web Index for open web search. *Journal of the Association for Information Science and Technology, JASIST* <https://doi.org/10.1002/asi.24818>.
- [18] L Caspari, KG Dastidar, S Zerhoudi, J Mitrović, M Granitzer, (2024). Beyond Benchmarks: Evaluating Embedding Model Similarity for Retrieval Augmented Generation Systems. *ArXiv, abs/2407.08275*.