

Daniil Sorokin

UKP Lab, Hochschulstraße 10
64289 Darmstadt, Germany
✉ sorokin@ukp.informatik.tu-darmstadt.de
📄 daniilsorokin.github.io

Keywords: natural language processing, machine learning, semantic analysis, knowledge bases

Work Experience

- 04.2015 – now **Researcher**, *UKP Lab, TU Darmstadt*, Germany.
- Neural network approaches for open-domain question answering
 - Teaching “Foundations of Natural Language Processing” for 3 semesters
 - Supervised 7 bachelor and master theses
 - Organizing a regular research seminar and guest talks for 3.5 years
 - Candidate evaluation for PhD- and PostDoc-level positions for 2.5 years
- 11.2018 – 01.2019 **Research Intern**, *Factmata Ltd.*, London, UK.
- Automatic evidence-based fact checking and claim validation
- 01.2014 – 03.2015 **Researcher**, *Research center SFB 833, University of Tübingen*, Germany.
- Automatic semantic processing of German and English noun compounds
- 05.2012 – 12.2013 **Research Assistant**, *Computational Linguistics department, University of Tübingen*, Germany.
- Teaching assistant for the course “Programming with Java”
 - Development of RESTful web services for linguistic resources
- 10.2010 – 08.2011 **Software Engineer**, *PROMT Ltd.*, Saint Petersburg, Russia.
- Construction of multi-lingual parallel corpora from web data
 - Stress- and usability-tests of machine translation tools

Education

- 04.2015 – now **PhD in Computer Science**, *TU Darmstadt*, Germany.
(expected 08.2019) Supervisor: Prof. Dr. Iryna Gurevych, UKP Lab
Topic: Lexical-Semantic and World Knowledge for Semantic Parsing
- 10.2011 – 01.2014 **M.A. in Computational Linguistics**, *University of Tübingen*, Germany,
GPA: 4.0 (German 1.1 / 1.0).
Thesis: A Machine Learning Approach to the Semantic Classification of German Noun Compounds
- 09.2007 – 07.2011 **B.A. in Computational Linguistics**, *Saint Petersburg State University*, Russia, GPA: 4.0 (Russian 4.9 / 5.0).
Thesis: Automatic Identification of Morphological Types of Geographical Names and Their Representation in a Geo-information System
- Summer schools Machine Learning Summer School (2016, Max Plank Institute), The European Summer School in Logic, Language and Information (ESSLLI 2014)

Skills

- | | | | |
|-----------------------|--|------------|--|
| Programming languages | Main: Python, Java | Experience | Javascript, XML, CSS, \LaTeX , with: SPARQL, SQL, R, Haskell, Scala |
| Main frameworks | PyTorch, NumPy, scikit-learn, TensorFlow, Keras, Flask, Weka, CoreNLP | | |
| Languages | English, German, Russian | Misc. | Climbing, cycling, skiing |
| Awards | Outstanding Reviewer (ACL, 2018), Teaching Award in Computer Science (TU Darmstadt, 2017), Best Challenge Paper (ESWC, 2017), Fully-funded Master’s scholarship (DAAD, 2011) | | |
| Service | Reviewer: ACL’18–19; RepL4NLP’19; CONLL’18; EMNLP’18; GSCL’17 | | |

Selected Invited Talks

- 2019 **Encoding Knowledge with Graph Neural Networks**, *NLP Seminar of the Department of Computing*, Imperial College London.
Graph Neural Networks for Knowledge Base Question Answering, *South England NLP Meetup of the UCL Machine Reading group*, Alan Turing Institute, London.
- 2018 **Graph Neural Networks for Knowledge Base Question Answering**, *NLIP Seminar Series*, Computer Laboratory, University of Cambridge.
- 2017 **Knowledge Base Question Answering**, *Natural Language Processing and the Web Seminar*, TU Darmstadt.
Full list at <https://daniilsorokin.github.io/>

Main Publications

Daniil Sorokin and Iryna Gurevych. “Interactive Instance-based Evaluation of Knowledge Base Question Answering”. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018): System Demonstrations*. DEMO URL: <http://semanticparsing.ukp.informatik.tu-darmstadt.de:5000/question-answering/>. Brussels, Belgium, Nov. 2018, pp. 114–119. URL: <http://aclweb.org/anthology/D18-2020>.

Daniil Sorokin and Iryna Gurevych. “Mixing Context Granularities for Improved Entity Linking on Question Answering Data across Entity Categories”. In: *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics (*SEM 2018 @ NAACL 2018)*. New Orleans, LA, U.S., June 2018, pp. 65–75. URL: <http://aclweb.org/anthology/S18-2007>.

Daniil Sorokin and Iryna Gurevych. “Modeling Semantics with Gated Graph Neural Networks for Knowledge Base Question Answering”. In: *Proceedings of COLING 2018, the 27th International Conference on Computational Linguistics*. Santa Fe, NM, U.S., Aug. 2018, pp. 3306–3317. URL: <http://aclweb.org/anthology/C18-1280>.

Daniil Sorokin and Iryna Gurevych. “Context-Aware Representations for Knowledge Base Relation Extraction”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*. Copenhagen, Denmark, Sept. 2017, pp. 1785–1790. URL: <http://aclweb.org/anthology/D17-1188>.

Daniil Sorokin and Iryna Gurevych. “End-to-end Representation Learning for Question Answering with Weak Supervision”. In: *ESWC 2017 Semantic Web Challenges*. Portoroz, Slovenia, 2017.

Google Scholar profile (complete publications list): <https://scholar.google.de/citations?user=PRX-B-oAAAAJ>

References are available upon request.