

Klim Zaporojets

Machine learning researcher — natural language processing, information extraction and knowledge graphs.

Open to research and applied-ML roles in industry and academia.

klimzaporojets@gmail.com · +45 22 32 84 63 · klimzaporojets.github.io · github.com/klmzaporojets · Google Scholar · ORCID 0000-0003-4988-978X · Aarhus, Denmark

RESEARCH AND PROFESSIONAL POSITIONS

Marie Skłodowska-Curie Postdoctoral Fellow

04/2024 – 04/2026

Department of Computer Science, Aarhus University, Denmark — Algorithms, Data and Artificial Intelligence group

- Independently funded research programme (€230,000) on updating knowledge graphs from evolving textual sources such as news and Wikipedia.
- Concurrent Guest Researcher, INDE Lab, University of Amsterdam.

Postdoctoral Researcher

01/2023 – 12/2023

Department of Computer Science, Aarhus University, Denmark

- Hospital@Night (H@N), an interdisciplinary project with health-IT specialists at Systematic and clinicians at Aalborg University Hospital.
- Entity summarization in knowledge graphs and integration of structured knowledge with unstructured clinical text; contributed to a clinical decision-support system matching patient information to procedural guidelines.

Doctoral Researcher

02/2018 – 12/2022

IDLab and Text-to-Knowledge (T2K) group, Ghent University, Belgium

- EU Horizon 2020 CPN project: information extraction and entity linking from news text, integrating textual (Wikipedia) and structured (Wikidata) sources.
- Led development of the DWIE and TempEL benchmark datasets.

Computational Linguist

06/2015 – 02/2018

Actonomy, Belgium

- Semantic matching between CVs and job vacancies for a commercial HR product.
- Pointwise mutual information and neural models for HR knowledge-graph completion; Tableau and Matplotlib dashboards surfacing labour-market skill-demand trends.

Research Assistant

05/2014 – 01/2015

Information Extraction and Synthesis Laboratory, University of Massachusetts Amherst, USA

- With Prof. Andrew McCallum. Conditional random fields for automatic identification of sections and references in scientific articles.
- Knowledge extraction from materials-science literature, in collaboration with AI2, the Materials Project, and Prof. Elsa Olivetti's group at MIT.

Software Engineer

01/2009 – 06/2013

MercadoLibre, Inc., Argentina

- Scrum team on checkout and fraud prevention at Latin America's largest e-commerce platform (NASDAQ: MELI).
- Training, testing and deployment of highly scalable machine learning models — decision trees, logistic regression and neural networks — with feature selection and model pruning, scalable SQL/PL-SQL, and load balancing and caching via Nginx.

RESEARCH VISITS

Visiting Research Fellow

03/2022 – 05/2022

University of Copenhagen, Denmark — host: Prof. Isabelle Augenstein

- Effect of temporal knowledge evolution on entity linking. Output: TempEL, NeurIPS 2022 Datasets and Benchmarks Track.

EDUCATION

PhD in Computer Science

02/2018 – 12/2022

Ghent University, Belgium

- Thesis: *Neural Approaches to Entity-Centric Information Extraction*.
- Supervisors: Prof. Chris Develder, Prof. Thomas Demeester.

MS in Computer Science

09/2013 – 05/2015

University of Massachusetts Amherst, USA

- GPA 3.9/4.0. Supervisor: Prof. Andrew McCallum.
- Funded by the Argentine Presidential Fellowship (Fulbright).

Systems Engineer

05/2002 – 12/2008

Instituto Universitario Aeronáutico, Córdoba, Argentina

- GPA 9.26/10 (top 5%). Thesis: *Virtualization of Distributed Systems with Service-Oriented Architecture*.

GRANTS, AWARDS AND HONOURS

Marie Skłodowska-Curie Actions Postdoctoral Fellowship – €230,000

04/2024 – 04/2026

European Commission, Horizon Europe. Grant No. 101146515.

Senior Area Chair Highlights Award, ACL 2025

2025

For Influences on LLM Calibration; selected for oral presentation.

Circle U. European University Alliance – RAI4KG – €10,000

12/2024 – 12/2025

Responsible AI for Knowledge Graphs, with University of Vienna and King's College London. Funded the pre-ACL 2025 Vienna workshop.

Research Foundation Flanders (FWO), long stay abroad – €6,000

03/2022 – 05/2022

Grant No. V412922N.

Argentine Presidential Fellowship (Fulbright) – €50,000

09/2013 – 05/2015

Graduate studies in the United States.

Best Grade Point Average Award

2006

Instituto Universitario Aeronáutico, Argentina.

BIBLIOMETRICS

24 publications listed below. **401 citations**, **h-index 11**, **i10-index 11** (Google Scholar). 397 of those citations are from 2021 onwards. Most cited: *DWIE* (118), *Injecting Knowledge Base Information* (50), *RISec* (36).

PUBLICATIONS

Reverse-chronological. Name in bold; * denotes equal contribution.

1. P. Zhang, Y. Jiang, **K. Zaporojets**, C. Cao, P. Groth, **TimeRoute: Time-Aware Modality Routing and Diffusion for Multi-Modal Recommendation** Submitted to the ACM International Conference on Web Search and Data Mining (WSDM 2027).
2. P. Zhang, **K. Zaporojets**, C. Cao, J.H. Huang, P. Groth, **Time Imprint: Learning Time-Aware Representations in Multi-Modal Knowledge Graphs** Submitted to the ACM International Conference on Web Search and Data Mining (WSDM 2027).
3. **K. Zaporojets**, D. Daza, E. Barba, I. Assent, R. Navigli, P. Groth, 2026. **EMERGE: A Benchmark for Updating Knowledge Graphs with Emerging Textual Knowledge** Submitted to the Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS 2026). [Code and dataset]
4. F. Polat, D. Daza, P. Zhang, **K. Zaporojets**, P. Groth, 2026. **Select, Don't Train: The Benefits of Modular Entity Disambiguation with LLM-Based Selection**. To appear in proceedings of the 25th International Semantic Web Conference (ISWC 2026).

5. A. Bainson, K. Mortensen, **K. Zaporojets**, D. Mottin, P. Karras, 2026. **Sankofa: Online Query-adaptive Dynamic Graph Summaries**. To appear in proceedings of the VLDB Volume 19 (VLDB 2026). [[Code and dataset](#)]
6. P. Zhang, **K. Zaporojets**, J. Liu, J. H. Huang, P. Groth, **Are a Thousand Words Better Than a Single Picture? Beyond Images - A Framework for Multi-Modal Knowledge Graph Dataset Enrichment**, in *Proceedings to 23rd European Semantic Web Conference (ESWC 2026)* (Dubrovnik, Croatia). [[Code and dataset](#)]
7. M. Li, P. Zhang, W. Xing, Y. Zheng, **K. Zaporojets**, J. Chen, R. Zhang, Y. Zhang, S. Gong, J. Hu, X. Ma, Z. Liu, P. Groth, M. Worring, **A survey of large language models for data challenges in graphs** *Expert Systems with Applications*, Vol. 298, 1 March 2026. [[Curated literature list](#)]
8. Y. Xia, PHL de Araujo, **K. Zaporojets**, B. Roth, **Influences on LLM Calibration: A Study of Response Agreement, Loss Functions, and Prompt Styles** in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics: ACL 2025 (Oral presentation and SAC Highlights Award)* (Vienna, Austria). [[Code](#)]
9. Y. Xia, **K. Zaporojets**, B. Roth, **Learn to pick the winner: Black-box ensembling for textual and visual question answering**, in *Proceedings of the 21st Conference on Natural Language Processing: KONVENS 2025 (Oral presentation)*, [[Dataset and code](#)]
10. P. Zhang, C. Cao, **K. Zaporojets**, P. Groth, **CYCLE: Cross-Year Contrastive Learning in Entity-Linking**, in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management: CIKM 2024* (Boise, United States). [[Code](#)]
11. V. Kougia, A. Sedova, A.J. Stephan, **K. Zaporojets**, B. Roth, **Analysing zero-shot temporal relation extraction on clinical notes using temporal consistency**, in *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing: BioNLP-WS 2024 co-located with EMNLP 2024* (Miami, Florida). [[Code](#)]
12. S. Javadi, A. Moradan, M. Sorkhpar, **K. Zaporojets**, D. Mottin, I. Assent, **Wiki Entity Summarization Benchmark** (2024 preprint), [[Dataset and code](#)]
13. S. Kiros Bitew, V. Schelstraete, **K. Zaporojets**, K. Van Nieuwenhove, R. Meganck, C. Develder, **Personality Style Recognition via Machine Learning: Identifying Anacritic and Introjective Personality Styles from Patients' Speech**, *Computational Linguistics in the Netherlands Journal*, Vol. 13, 01 March. 2024.
14. K. D'Oosterlinck, F. Remy, J. Deleu, T. Demeester, C. Develder, **K. Zaporojets**, A. Ghodsi, S. Ellershaw, J. Collins and C. Potts, **BioDEX: Large-Scale Biomedical Adverse Drug Event Extraction for Real-World Pharmacovigilance**, in *Findings of the 2023 Conference on Empirical Methods in Natural Language Processing: EMNLP 2023* (Singapore). [[BioDEX dataset and code](#)]
15. Y. Jiang, **K. Zaporojets**, J. Deleu, T. Demeester, C. Develder, **CookDial: a dataset for task-oriented dialogs grounded in procedural documents**, *Applied Intelligence*, 2023. [[CookDial dataset and code](#)]
16. **K. Zaporojets**, **Neural Approaches to Entity-Centric Information Extraction**, PhD thesis, Faculty of Engineering and Architecture, Ghent University, Dec. 2022. ISBN 978-94-6355-663-7.
17. **K. Zaporojets**, L.A. Kaffee, J. Deleu, T. Demeester, C. Develder, I. Augenstein, **TempEL: Linking Dynamically Evolving and Newly Emerging Entities**, in *Proceedings of the NeurIPS Datasets and Benchmarks Track 2022* [[TempEL dataset and code](#)]
18. **K. Zaporojets**, J. Deleu, T. Demeester and C. Develder, **Towards Consistent Document-Level Entity Linking: Joint Models for Entity Linking and Coreference Resolution**, in *Proceedings of the ACL 2022 (oral presentation)* (Dublin, Ireland) [[poster](#) | [code and AIDA+ dataset](#) | [video](#)]
19. S. Verlinden*, **K. Zaporojets***, J. Deleu, T. Demeester and C. Develder, **Injecting knowledge base information into end-to-end joint entity and relation extraction and coreference resolution**, in *Findings of the ACL: ACL-IJCNLP 2021* (Online Event) 1-6 Aug. 2021. (* equal contribution) [[poster](#) | [code](#) | [video](#)]
20. **K. Zaporojets**, J. Deleu, C. Develder and T. Demeester, **DWIE: an entity-centric dataset for multi-task document-level information extraction**, *Information Processing and Management*, Vol. 158, No. 4, Jul. 2021. [[DWIE dataset and code](#)]

21. **K. Zaporojets**, G. Bekoulis, J. Deleu, T. Demeester and C. Develder, **Solving Arithmetic Word Problems by Scoring Equations with Recursive Neural Networks**, *Expert Systems with Applications*, Vol. 174, 15 Jul. 2021. [code]
22. Y. Jiang, **K. Zaporojets**, J. Deleu, T. Demeester and C. Develder, **Recipe Instruction Semantics Corpus (RISeC): Resolving Semantic Structure and Zero Anaphora in Recipes** in Proc. 1st Conf. Asia-Pacific Chapter of the Assoc. Comput. Linguist. and 10th Int. Joint Conf. Natural Lang. Processing (AACL-IJCNLP 2020), Online, 4-7 Dec. 2020, pp. 821-826. [dataset]
23. S.K. Bitew, G. Bekoulis, J. Deleu, L. Sterckx, **K. Zaporojets**, T. Demeester and C. Develder, **Predicting suicide risk from online postings in Reddit: The UGent-IDLab submission to the CLPsych 2019 Shared Task A**, in Proc. 6th Ann. Workshop on Comput. Ling. Clin. Psychol. (CLPsych 2019) at NAACL-HLT 2019, Minneapolis, MN, USA, 6 Jun. 2019, pp. 158-161.
24. **K. Zaporojets**, L. Sterckx, J. Deleu, T. Demeester and C. Develder, **Predicting psychological health from childhood essays: The UGent-IDLab CLPsych 2018 shared task system**, in Proc. 5th Ann. Workshop on Comput. Ling. Clin. Psychol. (CLPsych 2018) at NAACL-HLT 2018, New Orleans, LA, USA, 5 Jun. 2018, pp. 119-125.

TEACHING

Teaching Assistant — Machine-Learning Based Natural Language Processing (Eo61340)

09/2020 – 12/2022

Ghent University, Belgium

- Content design, creation of assignments, and grading.

SUPERVISION AND MENTORING

Co-supervised **6 PhD students and 11 MSc theses across 6 institutions** (University of Amsterdam, University of Vienna, Aarhus University, Indiana State University, Politecnico di Torino, Ghent University). This supervision resulted in peer-reviewed publications at ACL (oral presentation, SAC Highlights Award 2025), CIKM, ESWC and *Expert Systems with Applications*.

PhD defence jury member, Ghent University, 02/2024 — Yiwei Jiang, *Towards grounded and robust task-oriented dialogue systems*.

ACADEMIC SERVICE — REVIEWING

Conferences: NeurIPS (2022–2024), ICML 2026 (**gold reviewer**), ICLR 2026, ACL (2024–2025), EMNLP (2023–2025), NAACL 2024, ECML-PKDD 2024, ECAI 2024, CIKM 2024.

Journals: IEEE Transactions on Neural Networks and Learning Systems, *Expert Systems with Applications*, IEEE Access, *Scientometrics*.

Additional reviewing during doctoral studies at Ghent University (2018–2022).

TECHNICAL SKILLS

Programming	Python (10+ yrs) · Java (7+ yrs, Sun Certified Programmer) · SQL (4+) · Scala (2+) · Groovy (2+) · JavaScript
Machine learning	PyTorch (6+ yrs) · Pandas (6+) · scikit-learn (5+) · NLTK (5+) · spaCy (4+) · Matplotlib (5+) · PyTorch Geometric · NetworkX · TensorFlow
Data and retrieval	Spark · Lucene · UIMA · Tableau · R · Matlab/Octave
Databases	Oracle (4+ yrs) · MySQL · Neo4j
Infrastructure	Linux (10+ yrs) · Git (10+) · Jenkins (5+) · Docker · Nginx · Flask · JUnit

LANGUAGES

Spanish (native) · Russian (native) · English (proficient) · Ukrainian (intermediate) · Dutch (intermediate) · Danish (beginner)