

Svitlana Vakulenko

Machine Learning Scientist
Amazon Alexa AI - Web Information
Barcelona, Spain
svvakul@amazon.com

RESEARCH INTERESTS

Conversational Search, Information Retrieval, Question Answering, Knowledge Representation

EDUCATION

Vienna University of Technology PhD in Computer Science with Honours Thesis: “Knowledge-based Conversational Search” (Supervisors: Prof. Axel Polleres, Prof. Maarten de Rijke)	October 2015 – November 2019
University of Tartu and Tallinn University of Technology MSc in Software Engineering Thesis: “Extraction of Process Models from Business Process Descriptions” (Supervisor: Prof. Marlon Dumas)	September 2009 – June 2011
Kyiv National Economic University BSc in Business Informatics with Honours	September 2005 – June 2009

RESEARCH POSITIONS

Amazon Alexa Machine Learning Scientist	December 2021 – present
University of Amsterdam Postdoctoral Researcher	October 2019 – October 2021
Apple Inc. Research Intern	June 2019 – September 2019
Vienna University of Economics and Business Project Staff	January 2017 – November 2019
webLyzard technology Applied Scientist	September 2014 – December 2016
University of Liechtenstein Predoctoral Researcher	September 2013 – August 2014
Karlsruhe Institute of Technology Project Staff	November 2011 – March 2012

RESEARCH PROJECTS

NWO Art DATIS artdatis.nl Information retrieval from historical archives	October 2019 – September 2020
AT FFG CommuniData communidata.at Open data for local communities	January 2017 – April 2019
EU Horizon 2020 InVID invid-project.eu Social media video verification	May 2016 – December 2016
EU FP7 PHEME pHEME.eu Rumour verification on social media	August 2014 – December 2016
EU FP7 LiquidPublication liquidpub.org Predicting conference ranking from bibliometric indicators	September 2010 – May 2011

RESEARCH VISITS

Stanford University Mark Musen's Lab (Prof. Axel Polleres)	June 2018
Federal University of the State of Rio de Janeiro Department of Applied Informatics (Prof. Kate Revoredo)	April 2018 – May 2018
University of Amsterdam ILPS (Prof. Evangelos Kanoulas)	July 2017 – September 2017

REFEREED PUBLICATIONS

I published 38 peer-reviewed articles, with 814 citations, and an h-index of 16 and an i10-index of 25.

1. **Svitlana Vakulenko**, Bill Byrne, Adria de Gispert. Uniform training and marginal decoding for multi-reference question-answer generation. Proceedings of the 26th European Conference on Artificial Intelligence (ECAI). 2023.
2. Xiaoyu Shen, **Svitlana Vakulenko**, Marco Del Tredici, Gianni Barlacchi, Bill Byrne, Adria de Gispert. Neural Ranking with Weak Supervision for Open-Domain Question Answering: A Survey. Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL). 2023.
3. Shaojie Jiang, **Svitlana Vakulenko**, Maarten De Rijke. Weakly Supervised Turn-level Engagingness Evaluator for Dialogues. Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR). 2023.
4. Mariya Hendriksen, **Svitlana Vakulenko**, Ernst Kuiper, Maarten de Rijke. Scene-Centric vs. Object-Centric Image-Text Cross-Modal Retrieval: A Reproducibility Study. Proceedings of the European Conference on Information Retrieval (ECIR). 2023.
5. Weijia Zhang, **Svitlana Vakulenko**, Thilina Rajapakse, Yumo Xu, Evangelos Kanoulas. Tackling Query-Focused Summarization as A Knowledge-Intensive Task: A Pilot Study. Proceedings of the Workshop on Generative Information Retrieval (Gen-IR@SIGIR). 2023.
6. Paul Owoicho, Jeffrey Dalton, Mohammad Aliannejadi, Leif Azzopardi, Johanne R Trippas, **Svitlana Vakulenko**. TREC CAsT 2022: Going Beyond User Ask and System Retrieve with Initiative and Response Generation. Proceedings of the NIST Text Retrieval Conference (TREC). 2022.

7. **Svitlana Vakulenko**, Johannes Kiesel, Maik Fröbe. SCAI-QReCC Shared Task on Conversational Question Answering. Proceedings of the Language Resources and Evaluation Conference (LREC). 2022.
8. Georgios Sidiropoulos, **Svitlana Vakulenko**, Evangelos Kanoulas. On the Impact of Speech Recognition Errors in Passage Retrieval for Spoken Question Answering. Proceedings of the Conference on Information and Knowledge Management (CIKM). 2022.
9. Mariya Hendriksen, Maurits J. R. Bleeker, **Svitlana Vakulenko**, Nanne van Noord, Ernst Kuiper, Maarten de Rijke. Extending CLIP for Category-to-Image Retrieval in E-Commerce. Proceedings of the European Conference on Information Retrieval (ECIR). 2022.
10. **Svitlana Vakulenko**, Evangelos Kanoulas, Maarten de Rijke. A Large-Scale Analysis of Mixed Initiative in Information-Seeking Dialogues for Conversational Search. ACM Transactions on Information Systems. Special Issue on Conversational Search and Recommendation. 2021.
11. **Svitlana Vakulenko**, Nikos Voskarides, Zhucheng Tu, Shayne Longpre. A Comparison of Question Rewriting Methods for Conversational Passage Retrieval. Proceedings of the European Conference on Information Retrieval (ECIR). 2021.
12. **Svitlana Vakulenko**, Shayne Longpre, Zhucheng Tu, Raviteja Anantha. Question Rewriting for Conversational Question Answering. Proceedings of the International ACM Conference on Web Search and Data Mining (WSDM). 2021.
13. Niels Rouws, **Svitlana Vakulenko**, Sophia Katrenko. Dutch SQuAD and Ensemble Learning for Question Answering from Labour Agreements. Proceedings of the Joint International Scientific Conferences on AI BNAIC/BENELEARN. 2021.
14. Georgios Sidiropoulos, Nikos Voskarides, **Svitlana Vakulenko**, Evangelos Kanoulas. Combining Lexical and Dense Retrieval for Computationally Efficient Multi-hop Question Answering. Proceedings of the Workshop on Simple and Efficient Natural Language Processing (SustainLP@EMNLP). 2021.
15. Julien Rossi, **Svitlana Vakulenko**, Evangelos Kanoulas. VerbCL: A Dataset of Verbatim Quotes for Highlight Extraction in Case Law. Proceedings of the Conference on Information and Knowledge Management (CIKM). 2021.
16. Vera Provatorova, Samarth Bhargav, **Svitlana Vakulenko**, Evangelos Kanoulas. Robustness Evaluation of Entity Disambiguation Using Prior Probes: the Case of Entity Overshadowing. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2021.
17. Raviteja Anantha, **Svitlana Vakulenko**, Zhucheng Tu, Shayne Longpre, Stephen Pulman, Srinivas Chappidi. Open-Domain Question Answering Goes Conversational via Question Rewriting. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT). 2021.
18. **Svitlana Vakulenko**, Evangelos Kanoulas and Maarten de Rijke. An Analysis of Mixed Initiative and Collaboration in Information-Seeking Dialogues. Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). 2020.
19. **Svitlana Vakulenko**, Nikos Voskarides, Zhucheng Tu, Shayne Longpre. Leveraging Query Resolution and Reading Comprehension for Conversational Passage Retrieval. TREC. 2020.
20. **Svitlana Vakulenko**, Shayne Longpre, Zhucheng Tu, Raviteja Anantha. A Wrong Answer or a Wrong Question? An Intricate Relationship between Question Reformulation and Answer Selection in Conversational Question Answering. Proceedings of the Workshop on Search-Oriented Conversational AI (SCAI @ EMNLP). 2020. *Best paper award*.
21. Vera Provatorova, **Svitlana Vakulenko**, Evangelos Kanoulas, Koen Dercksen, Johannes M. van Hulst. Named Entity Recognition and Linking on Historical Newspapers: UvA.ILPS & REL at CLEF HIPE 2020. CLEF (Working Notes). 2020.
22. **Svitlana Vakulenko**, Javier D. Fernandez, Axel Polleres, Maarten de Rijke and Michael Cochez. Message Passing for Complex Question Answering over Knowledge Graphs. Proceedings of the International Conference on Information and Knowledge Management (CIKM). 2019.

23. **Svitlana Vakulenko**, Kate Revoreda, Claudio Di Ciccio and Maarten de Rijke. QRFA: A Data-Driven Model of Information Seeking Dialogues. Proceedings of the European Conference on Information Retrieval (ECIR). 2019. *Best paper award (User track)*.
24. Sophia Keyner, Vadim Savenkov and **Svitlana Vakulenko**. Open Data Chatbot. Proceedings of the Extended Semantic Web Conference (ESWC). 2019.
25. Kathrin Figl, Samuel Kießling, Christiane Rank, **Svitlana Vakulenko**. The Effect of Fake News Flags on the Believability of Social Media Posts. Proceedings of the International Conference on Information Systems (ICIS). 2019.
26. **Svitlana Vakulenko**, Maarten de Rijke, Michael Cochez, Vadim Savenkov and Axel Polleres. Measuring Semantic Coherence of a Conversation. Proceedings of the International Semantic Web Conference (ISWC). 2018. *Spotlight paper*.
27. **Svitlana Vakulenko**, Ilya Markov and Maarten de Rijke. Conversational Exploratory Search via Interactive Storytelling. Proceedings of the Workshop on Search-Oriented Conversational AI (SCAI @ ICTIR). 2017.
28. **Svitlana Vakulenko** and Vadim Savenkov. TableQA: Question Answering on Tabular Data. Proceedings of SEMANTiCS Conference. 2017. *Best paper nominee (Poster and Demo track)*.
29. **Svitlana Vakulenko**, Lyndon Nixon, and Mihai Lupu. Character-based Neural Embeddings for Tweet Clustering. Proceedings of the Workshop on Natural Language Processing for Social Media (SocialNLP@EACL). 2017.
30. Albert Weichselbraun, Stefan Gindl, Fabian Fischer, **Svitlana Vakulenko**, and Arno Scharl. Aspect-based extraction and analysis of affective knowledge from social media streams. In IEEE Intelligent Systems. 2017.
31. Sebastian Neumaier, Vadim Savenkov and **Svitlana Vakulenko**. Talking Open Data. Proceedings of the Extended Semantic Web Conference (ESWC). 2017.
32. **Svitlana Vakulenko**, Albert Weichselbraun, and Arno Scharl. Detection of valid sentiment-target pairs in online product reviews and news media articles. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI). 2016.
33. **Svitlana Vakulenko**, Max Göbel, Arno Scharl and Lyndon Nixon. Know which Way the Wind Blows: Visualising the Propagation of News on the Web. Proceedings of the Workshop on Recent Trends in News Information Retrieval (NewsIR@ECIR). 2016.
34. Gregory Katsios, **Svitlana Vakulenko**, Anastasia Krithara and Georgios Paliouras. Towards Open Domain Event Extraction from Twitter: REVEALing Entity Relations. Proceedings of the Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DERIVE@ESWC). 2015.
35. **Svitlana Vakulenko**, Oliver Müller, Jan vom Brocke. Enriching iTunes App Store Categories via Topic Modeling. Proceedings of the 35th International Conference on Information Systems (ICIS). 2014.
36. Nadine Reuter, **Svitlana Vakulenko**, Jan vom Brocke, Stefan Debortoli, Oliver Müller. Identifying the Role of Information Systems in Achieving Energy-Related Environmental Sustainability using Text Mining. Proceedings of the European Conference on Information Systems (ECIS). 2014.
37. Andrea Herbst, Alexander Simons, Jan vom Brocke, Oliver Müller, Stefan Debortoli, **Svitlana Vakulenko**. Identifying and characterizing topics in enterprise content management: a latent semantic analysis of vendor case studies. Proceedings of the European Conference on Information Systems (ECIS). 2014.
38. Peep Küngas, Siim Karus, **Svitlana Vakulenko**, Marlon Dumas, Cristhian Parra, Fabio Casati. Reverse-engineering conference rankings: what does it take to make a reputable conference? Scientometrics. 2013.

TECHNICAL REPORTS

1. Gustavo Penha, **Svitlana Vakulenko**, Ondrej Dusek, Leigh Clark, Vaishali Pal, Vaibhav Adlakha. The Seventh Workshop on Search-Oriented Conversational Artificial Intelligence (SCAI'22). SIGIR.

- 2022.
2. **Svitlana Vakulenko**, Ondrej Dusek, Leigh Clark. Report on the 6th Workshop on Search-Oriented Conversational AI (SCAI 2021). SIGIR Forum. 2021.
 3. Krisztian Balog, Lucie Flekova, Matthias Hagen, Rosie Jones, Martin Potthast, Filip Radlinski, Mark Sanderson, **Svitlana Vakulenko**, Hamed Zamani. Common Conversational Community Prototype: Scholarly Conversational Assistant. Dagstuhl Seminar 19461 on Conversational Search. 2020.
 4. **Svitlana Vakulenko**, Vadim Savenkov and Maarten de Rijke. Conversational Browsing. Vienna University of Economics and Business. 2020.
 5. **Svitlana Vakulenko**, Sebastian Neumaer, Vadim Savenkov. AT FFG CommuniData Project Deliverable 3.1: Dataset Discovery Plugin. Vienna University of Economics and Business. 2017.
 6. Lyndon Nixon, **Svitlana Vakulenko**. EU Horizon 2020 InVID Project Deliverable 2.1: Social Media Filtering and Extraction. MODUL Technology. 2016.
 7. **Svitlana Vakulenko**, Arno Scharl. EU FP7 PHEME Project Deliverable 3.2: Algorithms for Implicit Information Diffusion Networks Across Media. MODUL University Vienna. 2016.
 8. **Svitlana Vakulenko**, Sudhir Agarwal. CoAggregator: User-Driven Approach to Web Aggregation. Karlsruhe Institute of Technology. 2012.

TEACHING

- Main lecturer of the “Language, Speech and Dialogue Processing” BSc course, University of Amsterdam, SS 2021.
- Co-lecturer of the “Intelligent Interactive Systems” MSc course, University of Amsterdam, WS 2020.
- Co-lecturer of the “Information Retrieval 2” MSc course, University of Amsterdam. 2020.
- Main lecturer of the “Principles of Software Programming” BSc course, WU Wien, WS 2017.
- Co-organizer and lecturer of the Django Girls Vienna workshop for female and non-binary. April 2017.
- Main lecturer of the “Introduction to Programming” MSc course, University of Liechtenstein, SS 2014.

INVITED TALKS

- | | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| Conversational Search: Interactive Information Retrieval through a Conversational User Interface
ILLC Computational Linguistic Seminar, Netherlands | 23 February 2021 |
| Question Rewriting for Conversational Question Answering
Montreal NLP Consortium Meetup, Canada | 25 November 2020 |
| Message Passing for Complex Question Answering over Knowledge Graphs
Search Engines Amsterdam Meetup, Netherlands | 31 January 2020 |
| Message Passing for Complex Question Answering over Knowledge Graphs
Vrije Universiteit Amsterdam, Netherlands | 9 December 2019 |
| Message Passing for Complex Question Answering over Knowledge Graphs
Apple Inc., USA | 12 September 2019 |
| Measuring Semantic Coherence of a Conversation
Stanford University, USA | 6 June 2018 |

Conversation Mining: Process Mining for Conversational Data Federal University of the State of Rio de Janeiro, Brazil	24 April 2018
Semantic Coherence for Conversational Browsing of a Knowledge Graph University of Bonn, Germany	31 January 2018
CommuniData: Project Report GovCamp Vienna, Austria	1 December 2017
Conversational Exploratory Search via Interactive Storytelling ILPS Seminar, Netherlands	5 September 2017
Memory Networks for Question Answering on Tabular Data Grammarly Meetup Kyiv, Ukraine	7 June 2017

STUDENT SUPERVISION

2024 (planned)

- Shaojie Jiang. PhD thesis: Engagement and Text Quality in Conversational Systems. Promotor: Maarten de Rijke. Co-promotores: Svitlana Vakulenko, Ruqing Zhang.

2023

- Philipp Christmann. PhD internship at Amazon Barcelona: Rethinking External Knowledge Access in Retrieval-Augmented Generation for Long-form Question Answering. Mentor: Svitlana Vakulenko. Manager: Adria de Gispert.

2021

- Maria Diea. MSc thesis: Disentangled Heterogeneous Graph Learning for Link Prediction. *Amsterdam Data Science Thesis Awards Honourable Mention.*
- Leila Talha. MSc thesis: Dialogue Act Modelling as Neural Machine Translation for Joint Dialogue Act Segmentation and Prediction on Automated Transcripts.
- Niels Rouws. BSc thesis: Extractive Question Answering for Dutch Labour Agreements. *Published in BNAIC 2021.*
- Marvin Ong. BSc thesis: A Comparison of Different Named Entity Recognition Architectures Trained on Non-standard Entities.
- Juno Prent. BSc thesis: A Systematic Comparison of Pre-Processing Methods for Bias Mitigation.
- Florian van der Steen. BSc thesis: A Systematic Comparison of In- and Post-Processing Methods for Bias Mitigation.

2020

- Arvid Lindström. MSc thesis: Dialogue Act Classification using Inductive Graph Learning.
- Alex Ketelaars. BSc thesis: Language Style Matching in Dialogue Generation Models.
- Bart van Vulpen. BSc thesis: Integrating a Language Style Matching Objective into Deep Neural Networks for Dialogue Response Generation.

2019

- Yoan Bachev. MSc thesis: Social Media News Chatbot.
- Matthew Minson. MSc thesis: Deep Learning Models for Question Answering.
- Sophia Keyner. BSc thesis: Open Data Chatbot. *Published in ESWC 2019.*

2018

- Eric Bruch. MSc thesis: Open Data Gamification: Quizz Generator.
- Paul Rudolf Zahrl. BSc thesis: Real-time Job Recommendations from Twitter.
- Christiane Rank. BSc thesis: Rumor Detection Systems and Trustworthiness of Social Media Posts.

2017

- Amar Taukeer. MSc thesis: Visualizing Geo-located Open Data.
- Alexander Meisl. BSc thesis: Open Data for Social Science: Red Cross Use Case.
- Donatella Novakovic. BSc thesis: Semantic Literature Review on Rumour Detection.

COMMUNITY SERVICE

2024

- Head of the advisory board for the 8th Search-Oriented Conversational Artificial Intelligence workshop (SCAI 2024) co-located with CHIIR 2024.

2023

- Invited panelist in the career panel of the Doctoral Consortium at the 26th European Conference on Artificial Intelligence (ECAI 2023)
- Program Committee member: 37th AAAI Conference on Artificial Intelligence (AAAI 2023), 32nd ACM International Conference on Information and Knowledge Management (CIKM 2023).
- Mentor at MigraCode Barcelona.

2022

- Main organiser of the 7th Search-Oriented Conversational Artificial Intelligence workshop (SCAI 2022) co-located with SIGIR 2022.
- Co-organiser of the Conversational Assistance track in TREC (TREC CAsT 2022).
- Co-organiser of the 2nd DialDoc Workshop.
- Program Committee member: Semantics and Knowledge track at The Web Conference 2022.

2021

- Main organiser of the 6th Search-Oriented Conversational Artificial Intelligence workshop (SCAI 2021).
- Program Committee member: Dialogue and Interactive systems track at the Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2021), Information Retrieval, Search and Question Answering Track at the conference of the European Chapter of the Association for Computational Linguistics (EACL 2021), Workshop Towards a new generation of Personalized Intelligent Conversational Agents (PICA 2021), held in conjunction with ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2021).
- Member of the SIGIR Artifact Evaluation Committee.
- Reviewer for the Conference on Computer-Supported Cooperative Work and Social Computing (CSCW), ACM Transactions on Information Systems (TOIS).

2020

- Program Committee member: Dialogue and Interactive Systems Track of the Annual Meeting of the Association for Computational Linguistics (ACL 2020), Dialog and Interactive Systems track of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2020), Dialogue - Interactive Systems Track of the International Conference on Computational Linguistics (COLING 2020), Workshop for Natural Language Processing Open Source Software (NLP-OSS@EMNLP 2020), In-Use Track of the International Semantic Web Conference (ISWC 2020).
- Selection committee member: Amsterdam Data Science Thesis Award, Student Grants for the Confer-

ence and Labs of the Evaluation Forum (CLEF 2020).

- Reviewer for Automated Knowledge Base Construction (AKBC 2020), ACM Transactions on Information Systems (TOIS).

2019

- Program Committee member: Resources and Evaluation Track of the Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2019), Reproducibility Track of the International Semantic Web Conference (ISWC 2019).
- Participant of the Dagstuhl Seminar 19461 on Conversational Search.
- Reviewer for Automated Knowledge Base Construction (AKBC 2019), International Conference on Information Systems (ICIS 2019), Transactions on Information Systems (TOIS), Knowledge and Information Systems (KAIS), et cetera.

HONORS AND AWARDS

- Diversity & Inclusion Award for attending the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020
- ACM Grant for attending the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). 2020
- SIGIR Student Travel Grant for attending the International Conference on Information and Knowledge Management (CIKM). 2019
- Student grant for attending the European Conference on Information Retrieval (ECIR). 2019
- Travel grant from the Austrian Research Association funding the research visit at Stanford. 2018
- Travel grant for attending the Workshop on Search-Oriented Conversational AI (SCAI). 2017
- Stipend for attending the IS Womens Network Workshop: “Advancing Women In IS”. 2014
- Scholarship for attending the Research-Centered Summer School at NCSR Demokritos. 2014
- Scholarship with a full funding of the preparatory phase at the Graduate School of Computer Science, Saarland University. 2012 - 2013
- Scholarship for attending Baltic Summer School at Kaunas University of Technology. 2010
- EU Eastern Partnership scholarship from the Open Estonia Foundation. 2010
- Scholarship from the Estonian Ministry of Foreign Affairs with a monthly allowance for the duration of my Master’s studies. 2009 - 2011

TECHNICAL SKILLS

Programming languages: Python, JavaScript, Java.

Machine learning: PyTorch, Tensorflow, Keras.

Databases: SQL, NoSQL, SPARQL.

Other: Git, LaTeX.

LANGUAGES

Ukrainian (Native), English (Fluent), Russian (Fluent), German (Advanced), Spanish (Basic).