

Andrew Yates — Curriculum Vitae

Senior Researcher
Max Planck Institute for Informatics
Campus E1.4, 66123 Saarbrücken
Germany

January 7, 2021
ayates@mpi-inf.mpg.de
<https://andrewyates.net>

Research Interests

Information Retrieval with particular emphasis on neural networks and the medical domain

Education

Georgetown University
Ph.D., Computer Science — co-advised by Dr. Ophir Frieder & Dr. Nazli Goharian
Dissertation: Identifying Real World Concepts in Social Media
Washington, DC, USA
2011 – May 2016

Illinois Institute of Technology
B.S., Computer Science
Chicago, IL, USA
2007 – 2011

Experience

Max Planck Institute for Informatics
Senior Researcher — with student supervision rights at Saarland University
Postdoctoral Researcher — supervised by Prof. Dr. Gerhard Weikum
Saarbrücken, Germany
Fall 2018 – present
Fall 2016 – Fall 2018

Twitter
Research Intern — supervised by Dr. Alek Kolcz
San Francisco, CA, USA
Fall 2015

Microsoft Research
Research Intern — supervised by Dr. Eric Horvitz & Dr. Ryen White
Redmond, WA, USA
Summer 2014 & 2015

Georgetown University
Research Assistant
Teaching Assistant — Introduction to Databases
Teaching Assistant — Text Mining & Analysis
Teaching Assistant — Information Retrieval
Teaching Assistant — Seminar in Search Systems
Teaching Assistant — Introduction to Data Mining
Teaching Assistant — Introduction to Information Systems
Washington, DC, USA
Fall 2011 – Spring 2016
Spring 2015
Fall 2013
Fall 2013
Spring 2013
Spring 2013 & 2015
Spring & Fall 2012

Orbitz Worldwide
Intern — Machine Learning Team
Chicago, IL, USA
Summer 2010, 2011, & 2012

Illinois Institute of Technology
Research Assistant — Information Retrieval Lab
Chicago, IL, USA
2009–2011

Awards

- Top Reviewer Award from *Information Processing & Management (IPM)* 2019
- Best Reviewer Award from *Conference on Empirical Methods for Natural Language Processing (EMNLP)* 2018
- Area Chair Favorite Paper Award from *International Conference on Computational Linguistics (COLING)* 2018
for *A Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions*
- Best Long Paper Award from *Conference on Empirical Methods for Natural Language Processing (EMNLP)* 2017
for *Depression and Self-Harm Risk Assessment in Online Forums*
- Best Poster Awards from *ICBI Biomedical Informatics Symposium at Georgetown University*
Second Place Award for *On Clinical Decision Support* 2014
First Place Award for *Mining Social Media for Healthcare* 2012
- Camras Scholarship from *Illinois Institute of Technology* for full tuition 2007–2011

Grants

- Google Cloud. COVID-19 research credits — \$3,000 USD value 2020
- Google Cloud. Research credits — \$10,000 USD value 2019–2020
- Google TensorFlow Research Cloud. Ten dedicated TPUs — \$660,000 USD value 2019–2020
- Max Planck Society. Co-applicant on internal GPU cluster grant for institute — €1,015,000 EUR 2019
- NVIDIA. Titan Xp GPU grant — \$1,199 USD value 2017
- SIGIR. Student travel grant — \$1,049 USD 2012

Ph.D. Student Supervision

- Anna Tiginova. Topic: information extraction for populating personal knowledge bases. 2018–present
- Ghazaleh Torbati. Topic: personalizing search & recommendations. Co-supervised with Gerhard Weikum. 2018–present

Ph.D. Committee Membership

- Committee member for Sean MacAvaney, Georgetown University, 2021 (expected).
- Examiner for Kashyap Popat, Saarland University, 2019.

M.S. Thesis Supervision

- Thong Nguyen. Topic: explainable depression detection via clinical questionnaires. 2020
- Ahmed Sohail Anwari. Topic: keyword and category queries for searching social media. 2020
- Azin Zahrayi. Topic: known-item search with poor memory. 2020
- Ahmad Ahmadian Moghadam. Thesis: *Axiomatic Augmentation for Neural Information Retrieval*. 2019
- Maggie Moheb Mounir Sourial. Thesis: *Automatic Neural Network Optimization*. 2019
- Michael Hedderich. Thesis: *Semantic Embeddings for Paraphrase Detection and Reverse Dictionaries*. 2017

Research Assistants & Interns

- Ph.D. students: Sean MacAvaney (2017)
- M.S. students: Youmna Ismaeil (2020–present), Canjia Li (2020), Kevin M. Jose (2019–present), Siddhant Arora (2019), Thong Nguyen (2019), Maggie M. M. Sourial (2018)
- B.S. students: Xinyu Zhang (2020), Skanda Vaidyanath (2019–2020), Aditya Chetan (2019), Samarth Mehrotra (2019)

Teaching

- Summer 2020, *Machine Learning for Harvesting Health Knowledge*. Seminar. 9 students. Co-taught with Dr. Erisa Terolli and Dr. Patrick Ernst.
- Winter 2019/2020, *Information Retrieval and Data Mining*. Core lecture, 9 ECTS credits. 105 students. Co-taught (IR half) with Dr. Rishiraj Saha Roy (DM half).
- Summer 2019, *Topics in Neural Information Retrieval*. Advanced lecture, 6 ECTS credits. 24 students.

Invited Talks

- Goldman Sachs R&D, 2020. *Transformers for Text Ranking*.
- Blekinge Institute of Technology, Karlskrona, Sweden, 2018. *Soft Matching Terms and Phrases with a Neural IR Model*.
- Saarland Informatics Campus, Saarbrücken, Germany, 2018. *Incorporating Positional and Contextual Information into a Neural IR Model*.
- University of Glasgow, United Kingdom, 2017. *Incorporating Positional Information and Other Domain Knowledge into a Neural IR Model*.

Professional Service

- Editorial Board — Information Processing & Management (2019–present)
- Program Committee (Area Chair) — EMNLP 2019, ACL 2020, EACL 2021
- Program Committee (Reviewer) — AAAI (2018, 2019), ACL (2018, 2019), CIKM (2020), ECIR (2019, 2020, 2021), EMNLP (2018, 2020), LREC (2016, 2018, 2020), NAACL (2019), SIGIR (2021), WWW (2018, 2021)
- Workshop Organizer — Seventh Workshop on Computational Linguistics and Clinical Psychology, NAACL, 2021

Journal Publications

- J5. S. MacAvaney, [A. Yates](#), A. Cohan, L. Soldaini, K. Hui, N. Goharian, and O. Frieder. 2019. Overcoming low-utility facets for complex answer retrieval. *Information Retrieval Journal*, 395-418.
- J4. A. Cohan, S. Young, [A. Yates](#), and N. Goharian. 2017. Triaging Content Severity in Online Mental Health Forums. *Journal of the Association for Information Science and Technology*.
- J3. L. Soldaini, [A. Yates](#), and N. Goharian. 2017. Learning to Reformulate Long Queries for Clinical Decision Support. *Journal of the Association for Information Science and Technology*.
- J2. J. Parker, [A. Yates](#), N. Goharian, and O. Frieder. 2015. Health-related hypothesis generation using social media data. *Social Network Analysis and Mining*, 1-15.
- J1. L. Soldaini, [A. Yates](#), E. Yom-Tov, O. Frieder, and N. Goharian. 2015. Enhancing web search in the medical domain via query clarification. *Information Retrieval Journal*, 1-25.

Conference Publications

- C37. S. Naseri, J. Dalton, [A. Yates](#), and J. Allan. 2021. CEQE: Contextualized Embeddings for Query Expansion. In *European Conference on Information Retrieval (ECIR '21)*.
- C36. G. Torbati, [A. Yates](#), and G. Weikum. 2021. You Get What You Chat: Using Conversations to Personalize Search-based Recommendations. In *European Conference on Information Retrieval (ECIR '21)*.
- C35. X. Zhang, [A. Yates](#), and J. Lin. 2021. Comparing Score Aggregation Approaches for Pretrained Neural Language Models. In *European Conference on Information Retrieval (ECIR '21)*.
- C34. A. Tiginova, P. Mirza, [A. Yates](#), and G. Weikum. 2020. RedDust: a Large Reusable Dataset of Reddit User Traits. In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC '20)*.

- C33. A. Tiginova, [A. Yates](#), P. Mirza, and G. Weikum. 2020. CHARM: Inferring Personal Attributes from Conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP '20)*.
- C32. G. H. Torbati, [A. Yates](#), and G. Weikum. 2020. Personalized Entity Search by Sparse and Scrutable User Profiles. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*.
- C31. [A. Yates](#), K. M. Jose, X. Zhang, and J. Lin. 2020. Flexible IR pipelines with Capreolus. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20)*.
- C30. Z. Zheng, K. Hui, B. He, X. Han, L. Sun, and [A. Yates](#). 2020. BERT-QE: Contextualized Query Expansion for Document Re-ranking. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings (Findings of EMNLP '20)*.
- C29. M. A. Hedderich, [A. Yates](#), D. Klakow, and G. Melo. 2019. Using Multi-Sense Vector Embeddings for Reverse Dictionaries. In *Proceedings of the 13th International Conference on Computational Semantics-Long Papers (IWCS '19)*.
- C28. S. MacAvaney, [A. Yates](#), A. Cohan, and N. Goharian. 2019. CEDR: Contextualized embeddings for document ranking. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*.
- C27. S. MacAvaney, [A. Yates](#), K. Hui, and O. Frieder. 2019. Content-Based Weak Supervision for Ad-Hoc Re-Ranking. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*.
- C26. K. Popat, S. Mukherjee, [A. Yates](#), and G. Weikum. 2019. STANCY: Stance Classification Based on Consistency Cues. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP '19)*.
- C25. A. Tiginova, [A. Yates](#), P. Mirza, and G. Weikum. 2019. Listening between the lines: Learning personal attributes from conversations. In *The World Wide Web Conference (WWW '19)*.
- C24. [A. Yates](#) and M. Unterkalmsteiner. 2019. Replicating Relevance-Ranked Synonym Discovery in a New Language and Domain. In *European Conference on Information Retrieval (ECIR '19)*.
- C23. A. Cohan, B. Desmet, [A. Yates](#), L. Soldaini, S. MacAvaney, and N. Goharian. 2018. SMHD: a Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING '18)*. **Area Chair favorite paper.**
- C22. K. Hui, [A. Yates](#), K. Berberich, and G. Melo. 2018. Co-PACRR: Deep Retrieval with Contextualized Local Matching. In *Conference on Web Search and Data Mining (WSDM '18)*.
- C21. C. Li, Y. Sun, B. He, L. Wang, K. Hui, [A. Yates](#), L. Sun, and J. Xu. 2018. NPRF: A Neural Pseudo Relevance Feedback Framework for Ad-hoc Information Retrieval. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP '18)*.
- C20. S. MacAvaney, [A. Yates](#), A. Cohan, L. Soldaini, K. Hui, N. Goharian, and O. Frieder. 2018. Characterizing question facets for complex answer retrieval. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*.
- C19. K. Popat, S. Mukherjee, [A. Yates](#), and G. Weikum. 2018. DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP '18)*.
- C18. K. Hui, [A. Yates](#), K. Berberich, and G. Melo. 2017. A Position-Aware Deep Model for Relevance Matching in Information Retrieval. In *Empirical Methods for Natural Language Processing (EMNLP '17)*.
- C17. K. Hui, [A. Yates](#), K. Berberich, and G. Melo. 2017. Position-Aware Representations for Relevance Matching in Neural Information Retrieval. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*.
- C16. L. Soldaini, N. Goharian, and [A. Yates](#). 2017. Denoising Clinical Notes for Medical Literature Retrieval with Convolutional Neural Model. In *Conference on Information and Knowledge Management (CIKM '17)*.
- C15. [A. Yates](#), A. Cohan, and N. Goharian. 2017. Depression and Self-Harm Risk Assessment in Online Forums. In *Empirical Methods for Natural Language Processing (EMNLP '17)*. **Best paper award.**
- C14. [A. Yates](#), N. Goharian, and O. Frieder. 2016. Learning the Relationships between Drug, Symptom, and Medical Condition Mentions in Social Media. In *International AAAI Conference on Web and Social Media (ICWSM '16)*.
- C13. [A. Yates](#), J. Joselow, and N. Goharian. 2016. The News Cycle's Influence on Social Media Activity. In *International AAAI Conference on Web and Social Media (ICWSM '16)*.
- C12. [A. Yates](#), A. Kolcz, N. Goharian, and O. Frieder. 2016. Effects of Sampling on Twitter Trend Detection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC '16)*.
- C11. L. Soldaini, A. Cohan, [A. Yates](#), N. Goharian, and O. Frieder. 2015. Retrieving medical literature for clinical decision support. In *European Conference on Information Retrieval (ECIR '15)*.
- C10. [A. Yates](#), N. Goharian, and O. Frieder. 2015. Extracting Adverse Drug Reactions from Social Media. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '15)*.
- C9. A. Cohan, L. Soldaini, [A. Yates](#), N. Goharian, and O. Frieder. 2014. On Clinical Decision Support. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB '14)*.

- C8. [A. Yates](#), N. Goharian, and O. Frieder. 2014. Relevance-Ranked Domain-Specific Synonym Discovery. In *European Conference on Information Retrieval (ECIR '14)*.
- C7. [A. Yates](#), J. Parker, N. Goharian, and O. Frieder. 2014. A Framework for Public Health Surveillance. In *Ninth International Conference on Language Resources and Evaluation (LREC '14)*.
- C6. E. W. Burger, H. Federoff, O. Frieder, N. Goharian, and [A. Yates](#). 2013. Social media communications networks and pharmacovigilance: SequelAE-2.0. In *Proceedings of the IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom '13)*.
- C5. J. Parker, Y. Wei, [A. Yates](#), N. Goharian, and O. Frieder. 2013. A framework for detecting public health trends with Twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM '13)*.
- C4. J. Wilberding, [A. Yates](#), M. Sherr, and W. Zhou. 2013. Validating Web Content with Sensor. In *Annual Computer Security Applications Conference (ACSAC '13)*.
- C3. [A. Yates](#) and N. Goharian. 2013. ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. In *European Conference on Information Retrieval (ECIR '13)*.
- C2. [A. Yates](#), N. Goharian, and O. Frieder. 2013. Graded relevance ranking for synonym discovery. In *Proceedings of the 22nd international conference on World Wide Web companion (WWW '13 Companion)*.
- C1. J. Parker, [A. Yates](#), N. Goharian, and W. G. Yee. 2012. Efficient estimation of aspect weights. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '12)*.

Workshop Publications

- W7. X. Zhang, [A. Yates](#), and J. Lin. 2020. A Little Bit Is Worse Than None: Ranking with Limited Training Data. In *Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing (SustaiNLP '20)*.
- W6. S. Arora and [A. Yates](#). 2019. Investigating Retrieval Method Selection with Axiomatic Features. In *1st Interdisciplinary Workshop on Algorithm Selection and Meta-Learning in Information Retrieval (AMIR '19)*.
- W5. M. Unterkalmsteiner and [A. Yates](#). 2019. Expert-sourcing Domain-specific Knowledge: The Case of Synonym Validation. In *2nd Workshop on Natural Language Processing for Requirements Engineering and NLP Tool Showcase (NLP4RE '19)*.
- W4. S. MacAvaney, B. Desmet, A. Cohan, L. Soldaini, [A. Yates](#), A. Ziriky, and N. Goharian. 2018. RSDD-Time: Temporal Annotation of Self-Reported Mental Health Diagnoses. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic (CLPsych '18)*.
- W3. [A. Yates](#) and K. Hui. 2017. DE-PACRR: Exploring Layers Inside the PACRR Model. In *SIGIR 2017 Workshop on Neural Information Retrieval (Neu-IR '17)*.
- W2. [A. Yates](#), N. Goharian, and O. Frieder. 2013. Extracting adverse drug reactions from forum posts and linking them to drugs. In *Proceedings of the SIGIR Workshop on Health and Discovery (HSD '13)*.
- W1. W. G. Yee, [A. Yates](#), S. Liu, and O. Frieder. 2009. Are web user comments useful for search?. In *Proceedings of the 7th SIGIR Workshop on Large-Scale Distributed Systems for Information Retrieval (LSDS-IR '09)*.

Demonstration Publications

- D3. A. Tiginova, P. Mirza, [A. Yates](#), and G. Weikum. 2021. CHARMED: Exploring Personal Knowledge Extraction from Conversations. In *Conference on Web Search and Data Mining (WSDM '21)*.
- D2. [A. Yates](#), S. Arora, X. Zhang, W. Yang, K. M. Jose, and J. Lin. 2020. Capreolus: A Toolkit for End-to-End Neural Ad Hoc Retrieval. In *Proceedings of the 13th International Conference on Web Search and Data Mining (WSDM '20)*.
- D1. W. G. Yee, [A. Yates](#), O. Frieder, and A. Moehrl. 2010. eduKEN: a tool for fine-grained video comment collection and analysis. In *Proceedings of the 19th international conference on World wide web (WWW '10)*.

Shared Task Publications

- S7. C. Li and [A. Yates](#). 2020. MPIO at the NTCIR-15 WWW-3 Task. In *15th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-15)*.
- S6. S. Mehrotra and [A. Yates](#). 2019. MPIO at TREC CAst 2019: Incorporating Query Context into a BERT Re-ranker. In *Text REtrieval Conference (TREC '19)*.
- S5. [A. Yates](#). 2019. MPIO at the NTCIR-14 CENTRE task. In *14th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-14)*.
- S4. [A. Yates](#). 2019. MPIO at the NTCIR-14 WWW-2 Task. In *14th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-14)*.
- S3. S. MacAvaney, [A. Yates](#), and K. Hui. 2017. Contextualized PACRR for Complex Answer Retrieval. In *Text REtrieval Conference (TREC '17)*.
- S2. L. Soldaini, A. Cohan, [A. Yates](#), N. Goharian, and O. Frieder. 2014. Query Reformulation for Clinical Decision Support Search. In *Text REtrieval Conference (TREC '14)*.

S1. [A. Yates](#), D. DeBoer, H. Yang, N. Goharian, S. Kunath, and O. Frieder. 2012. (Not Too) Personalized Learning to Rank for Contextual Suggestion. In *Text REtrieval Conference (TREC '12)*.

Technical Reports

T1. C. Li, [A. Yates](#), S. MacAvaney, B. He, and Y. Sun. 2020. PARADE: Passage representation aggregation for document reranking. *arXiv preprint arXiv:2008.09093*.

Books

- J. Lin, R. Nogueira, and [A. Yates](#). Pretrained Transformers for Text Ranking: BERT and Beyond. Morgan & Claypool Publishers. (Publication expected in early 2021)

Tutorials

- [A. Yates](#), R. Nogueira, and J. Lin. Pretrained Transformers for Text Ranking: BERT and Beyond. 2021. Half-day tutorial at *Conference on Web Search and Data Mining (WSDM '21)*. (to appear)
- [A. Yates](#), R. Nogueira, and J. Lin. Pretrained Transformers for Text Ranking: BERT and Beyond. 2021. Half-day tutorial at *Conference of the North American Chapter of the ACL (NAACL '21)*. (to appear)