

Donald Metzler

Email: metzler@google.com

Scholar: <https://scholar.google.com/citations?user=bmXpOd8AAAAJ&hl=en>

LinkedIn: <https://www.linkedin.com/in/donmetzler>

Last updated: 1 July 2025

Employment

- **Google, Mountain View, CA**

Senior Staff Research Scientist (March 2022 – Present)

Senior Staff Software Engineer (December 2016 – February 2022)

Staff Software Engineer (June 2014 - November 2016)

Senior Software Engineer (May 2012 - May 2014)

- Google DeepMind (2024 - Present)
Google Research (2019 - 2024) - Tech Lead and Manager of a group of ~40 Software Engineers and Research Scientists working on challenging problems at the intersection of machine learning, natural language processing, and information retrieval. Research focus areas include large language modeling foundations, generative information retrieval, safety/information quality, document understanding, and efficiency. Contributed to major launches across Gemini, Bard, Search, Cloud, and Chrome.
- G Suite (2016 - 2019)
Google Apps (2014 - 2016) - Tech Lead and Manager of Apps Search Quality and Intelligence team. Led a team of ~40 Software Engineers to improve the search and finding experience across the G Suite ecosystems (Gmail, Drive, Calendar, Groups, Sites, etc.). Led over 100 quality and feature launches, improving Gmail search quality by >100% and Drive search quality by >50%.
- Display Ads (2012 - 2014) - Tech Lead of the Rephil team. Rephil is a large-scale topic modeling system that is widely used across Google. Developed methods to incrementally update the model every few months instead of every few years, unlocking significant model quality and revenue gains.

- **Chomp, San Francisco, CA**

Technology Advisor (July 2011 – March 2012)

- Advised Cathy Edwards (co-founder and CTO) on state-of-the-art search practices.
- Other advisors included Ashton Kutcher and David Blei.
- Chomp was acquired by Apple in 2012.

- **University of Southern California, Los Angeles, CA**

Research Assistant Professor, Department of Computer Science (July 2011 – March 2012)

Computer Scientist, Information Sciences Institute (April 2010 – March 2012)

- Machine Reading -- Developed Mavuno, a scalable Hadoop-based "distributional similarity engine" that was used for a variety of machine reading tasks, including large-scale paraphrase acquisition and information extraction.
- Social media -- Investigated various natural language processing (noisy text normalization), data mining (automatically mining lexical variants), and information retrieval (automatic timeline construction) aspects of social media with a specific emphasis on crisis-related events.
- Learning to efficiently rank -- In a joint project with Jimmy Lin and Lidan Wang from the University of Maryland, developed a robust family of machine-learned ranking functions capable of trading off between effectiveness (result quality) and efficiency (speed).

- **Yahoo! Research, Santa Clara, CA**

Senior Research Scientist (May 2012 - May 2012)

Research Scientist (September 2007 – April 2010)

- As a member of the Search and Computational Advertising group at Yahoo! Research, worked closely with various product teams on a wide variety of real-world problems involving large-scale data sets.
- Designed and implemented highly effective text matching features for Web search. The features were deployed in various markets worldwide and yielded significant relevance gains, especially for long queries.
- Other notable projects included temporal query analysis, predicting when not to show ads, sponsored search ad matching for tail queries, enriched Web page representations, and ranking in match-making (e.g., dating) systems.

- **Microsoft, Redmond, WA**

Research Intern (June 2006 - August 2006)

- Worked with Susan Dumais and Chris Meek to develop novel similarity measures for short segments of text (e.g., query/query or query/ad similarity).

- **University of Massachusetts, Amherst, MA**

Research Assistant (September 2002 - August 2007)

- Markov Random Field Model - Developed a highly effective, formally motivated information retrieval model based on Markov Random Fields. The model can model dependencies between terms and incorporate arbitrary features in a unified way.
- Indri - Indri is an effective, efficient, flexible open source information retrieval toolkit. Led the design of the Indri query language and underlying retrieval model.
- TREC - Took part in the 2004, 2005, and 2006 Terabyte Tracks and the 2005 Robust Track. Systems developed were consistently among the top performing across a large international field of participants.

Education

- Ph.D., Computer Science, September 2007, University of Massachusetts, Amherst, MA
- M.S., Computer Science, May 2005, University of Massachusetts, Amherst, MA
- B.S., Computer Science and Mathematics, Summa Cum Laude, May 2002, Rose-Hulman Institute of Technology, Terre Haute, IN

Teaching Experience

- Instructor, CSCI 599: Applications of Natural Language Processing – Information Retrieval, University of Southern California, Spring 2012.
- Teaching Assistant, CMPSCI 646: Information Retrieval, University of Massachusetts, Amherst, Fall 2006.

Awards and Honors

- Member, ACM SIGIR Academy, 2025.
- Honorable Mention for Test of Time Award, SIGIR 2023.
 - Wang, L., Lin, J., and Metzler, D. "A Cascade Ranking Model for Efficient Ranked Retrieval," in the *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 105-114, 2011.
- Distinguished Member, Association for Computing Machinery (ACM), 2021.
- Test of Time Award, CIKM 2021.

- Chapelle, O., Metzler, D., Zhang, Y., and Grinspan, P. “Expected Reciprocal Rank for Graded Relevance,” in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 2009.
- Best Paper Award Runner-Up, WSDM 2021
 - Bahri, D., Tay, Y., Zheng, C., Brunk, C., Metzler, D., and Tomkins, A. “Generative Models are Unsupervised Predictors of Page Quality: A Colossal-Scale Study,” in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 301-309, 2021.
- Test of Time Award, ECIR 2017.
 - Metzler, D., Dumais, S., and Meek, C. “Similarity Measures for Short Segments of Text,” in the *Proceedings of the 29th European Conference on Information Retrieval*, 16-27, 2007.
- Honorable Mention for Test of Time Award, SIGIR 2017.
 - Metzler, D. and Croft, W.B., “A Markov Random Field Model for Term Dependencies,” in the *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 472-479, 2005.
- Best Paper Honorable Mention, WSDM 2017.
 - Bendersky, M., Wang, X., Metzler, D., and Najork, M. “Learning from User Interactions in Personal Search via Attribute Parameterization,” in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 791-799, 2017.
- Senior Member, Association for Computing Machinery (ACM), 2012
- Runner-Up for Best Search Paper, WSDM 2012.
 - Bendersky, M., Metzler, D., and Croft, W.B. “Effective Query Formulation with Multiple Information Sources,” in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 443-452, 2012.
- Honorable Mention for Best Paper, SIGIR 2011.
 - Bendersky, M., Metzler, D., and Croft, W.B. “Parameterized Concept Weighting in Verbose Queries,” in the *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 605-614, 2011.
- Short-Listed for Best Interdisciplinary Paper Award, CIKM 2010.
 - Wang, L., Metzler, D., and Lin, J. “Ranking Under Temporal Constraints,” in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 79-88, 2010.
- Best Student Paper, SIGIR 2005.
 - Metzler, D. and Croft, W.B., “A Markov Random Field Model for Term Dependencies,” in the *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 472-479, 2005.
- Microsoft Live Labs Graduate Fellowship, 2006-2007

Patents

- US Patent 11,989,220. System for determining and optimizing for relevance in match-making systems. Fernando Diaz, Donald Metzler, Sihem Amer-Yahia. Assignee: Match Group LLC. Filed August 12, 2019, issued May 21, 2024.
- US Patent 11,886,976. Efficient decoding of output sequences using adaptive early exiting. Tal Schuster, Adam Joshua Fisch, Jai Prakash Gupta, Mostafa Dehghani, Dara Bahri, Vinh Quoc Tran, Yi Tay, Donald Metzler. Assignee: Google LLC. Filed July 14, 2023, issued January 30, 2024.

- US Patent 10,970,293. Ranking search result documents. Mike Bendersky, Marc Alexander Najork, Donald Metzler, Xuanhui Wang. Assignee: Google LLC. Filed August 26, 2019, issued April 6, 2021.
- US Patent 10,394,832. Ranking search results documents. Mike Bendersky, Marc Alexander Najork, Donald Metzler, Xuanhui Wang. Assignee: Google LLC. Filed October 24, 2016, issued August 27, 2019.
- US Patent 10,380,158. System for determining and optimizing for relevance in match-making systems. Fernando Diaz, Donald Metzler, Sihem Amer-Yahia. Assignee: Match Group LLC. Filed August 8, 2016, issued August 13, 2019.
- US Patent 9,953,185. Identifying query patterns and associated aggregate statistics among search queries. Mike Bendersky, Donald Metzler, Marc Alexander Najork, Dor Naveh, Vlad Panait, Xuanhui Wang. Assignee: Google LLC. Filed November 24, 2015, issued April 24, 2018.
- US Patent 9,449,282. System for determining and optimizing for relevance in match-making systems. Fernando Diaz, Donald Metzler, Sihem Amer-Yahia. Assignee: Match.com LLC. Filed July 1, 2010, issued September 20, 2016.
- US Patent 8,676,827. Rare query expansion by web feature matching. Donald Metzler, Lance Riedel, Evgeniy Gabrilovich, Vanja Josifovski. Assignee: Yahoo! Inc. Filed February 4, 2009, issued March 18, 2014.
- US Patent 8,521,731. Systems and methods for query expansion in sponsored search. Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Lance Riedel, Peter Ciccolo. Assignee: Yahoo! Inc. Filed July 9, 2008, issued August 27, 2013.
- US Patent 8,156,111. Identifying and expanding implicitly temporally qualified queries. Rosie Jones, Donald Metzler, Fuchun Peng. Assignee: Yahoo! Inc. Filed November 24, 2008, issued April 10, 2012.
- US Patent 8,005,845. System and method for automatically ranking lines of text. Tapas Kanungo and Donald Metzler. Assignee: Yahoo! Inc. Filed May 20, 2008, issued August 23, 2011.

Theses

- Metzler, D. *"Beyond Bags of Words: Effectively Modeling Dependence and Features in Information Retrieval,"* Ph.D. Dissertation, University of Massachusetts, Amherst. Amherst, MA, 2007.

Books

- Metzler, D. *"A Feature-Centric View of Information Retrieval,"* Information Retrieval Series, Springer, 2011.
- Croft, W.B., Metzler, D., and Strohman, T. *"Search Engines: Information Retrieval in Practice,"* Addison Wesley, 2009.

Book Chapters

- Bendersky, M, Metzler, D., Najork, M., and Wang, X. "Searching Personal Collections". *Information Retrieval: Advanced Topics and Techniques (1st ed.)*. Association for Computing Machinery, New York, NY, USA, 515–538, 2024.

Journal Articles

- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E., Hashimoto, T., Vinyals, O., Lian, P., Dean, J., Fedus, W. "Emergent

Abilities of Large Language Models,” in the *Transactions of Machine Learning Research (TMLR)*, 2022.

- Tay, Y., Dehghani, M., Bahri, D., and Metzler, D. “Efficient Transformers: A Survey,” in *ACM Computing Surveys*, 55(6), 2022.
- Metzler, D., Tay, Y., Bahri, D., Najork, M. “Rethinking Search: Making Domain Experts out of Dilettantes,” in *SIGIR Forum*, 55(1), 1-27, 2021.
- Bendersky, M., Wang, X., Najork M., and Metzler, D. “Search and Discovery in Personal Email Collections”, in *Foundations and Trends in Information Retrieval*, 15(1), 1–133, 2021.
- Metzler, D. and Croft, W.B. “Linear Feature-Based Models for Information Retrieval,” in *Information Retrieval*, 10(3), 257-274, 2007.
- Metzler, D. and Croft, W.B. “Analysis of Statistical Question Classification for Fact-based Questions,” *Information Retrieval*, 8(3), 481-504, 2005.
- Metzler, D. and Croft, W.B. “Combining the Language Model and Inference Network Approaches to Retrieval,” *Information Processing and Management Special Issue on Bayesian Networks and Information Retrieval*, 40(5), 735-750, 2004.

Conference Papers

- Balog, K., Metzler, D., and Qin, Z. “Rankers, Judges, and Assistants: Towards Understanding the Interplay of LLMs in Information Retrieval Evaluation,” in the *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025.
- Zhang, X., Lu, J., Tran, V., Schuster, T., Metzler, D., and Lin, J. “Tomato, Tomahto, Tomate: Measuring the Role of Shared Semantics among Subwords in Multilingual Language Models,” in the *Findings of the North American Chapter of the Association for Computational Linguistics (NAACL Findings)*, 2025.
- Gao, Y., Alon, D., and Metzler, D. “Impact of Preference Noise on the Alignment Performance of Generative Language Models,” in the *Proceedings of the Conference on Language Modeling (COLM)*, 2024.
- Baumgärtner, T., Gao, Y., Alon, D., and Metzler, D. “Best-of-Venom: Attacking RLHF by Injecting Poisoned Preference Data,” in the *Proceedings of the Conference on Language Modeling (COLM)*, 2024.
- Gao, Y., Ma, J., Korotkov, I., Hall, K., Alon, D., and Metzler, D. “OpenMSD: Towards Multilingual Scientific Documents Similarity Measurement,” in the *Proceedings of the The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC/COLING)*, 2024.
- Schuster, T., Lelkes, A., Sun, H., Gupta, J., Berant, J., Cohen, W., and Metzler, D. “SEMQA: Semi-Extractive Multi-Source Question Answering,” in the *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2024.
- Qin, Z., Jagerman, R., Hui, K., Zhuang, H., Wu, J., Shen, J., Liu, T., Liu, J., Metzler, D., Wang, X., and Bendersky, M. “Language Models are Effective Text Rankers with Pairwise Ranking Prompting,” in the *Findings of the North American Chapter of the Association for Computational Linguistics (NAACL Findings)*, 2024.
- Pradeep, R., Hui, K., Gupta, J., Lelkes, A., Zhuang, H., Lin, J., Metzler, D., and Tran, V. “How Does Generative Retrieval Scale to Millions of Passages?,” in the *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.
- Tay, Y., Wei, J., Chung, H.W., Tran, V., So, D., Shakeri, S., Garcia, X., Zheng, H., Rao, J., Chowdhery, A., Zhou, D., Metzler, D., Petrov, S., Houlsby, N., Le, Q., and Dehghani, M.

“Transcending Scaling Laws with 0.1% Extra Compute,” in the *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

- Mehta, S., Gupta, J., Tay, Y., Dehghani, M., Tran, V., Rao, J., Najork, M., Strubell, E., and Metzler, D. “DSI++: Updating Transformer Memory with New Documents,” in the *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.
- Tay, Y., Dehghani, M., Abnar, S., Chung, HW., Fedus, W., Rao, J., Narang, S., Tran, V., Yogatama, D., and Metzler, D. “Scaling Laws vs Model Architectures: How Does Inductive Bias Influence Scaling?,” in the *Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP Findings)*, 2023.
- Drozdov, A., Zhuang, H., Dai, Z., Qin, Z., Rahimi, R., Wang, X., Alon, D., Iyyer, M., McCallum, A., Metzler, D., and Hui, K. “PaRaDe: Passage Ranking using Demonstrations with Large Language Models,” in the *Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP Findings)*, 2023.
- Milbauer, J., Louis, A., Hosseini, M. J., Fabrikant, A., Metzler, D., and Schuster, T. “LAIT: Efficient Multi-Segment Encoding in Transformers with Layer-Adjustable Interaction,” in the *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.
- Bahri, D., Zheng, C., Tay, Y., Metzler, D., and Tomkins, A. “Surprise: Result List Truncation via Extreme Value Theory,” in the *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023.
- Tay, Y., Dehghani, M., Tran, V., Garcia, X., Wei, J., Wang, X., Chung, HW, Bahri, D., Schuster, T., Zheng, S., Zhou, D., Houlsby, N., Metzler, D. “UL2: Unifying Language Learning Paradigms,” to appear in the *Proceedings of the Eleventh International Conference on Learning Representations (ICLR)*, 2023.
- Schuster, T., Chen, S., Buthpitiya, S., Fabrikant, A., and Metzler, D. “Stretching Sentence-pair NLI Models to Reason over Long Documents and Clusters,” in the *Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP Findings)*, 2022.
- Gupta, J., Tay, Y., Kamath, C., Tran, V., Metzler, D., Bavadekar, S., and Gabrilovich, E. “Dense Feature Memory Augmented Transformers for COVID-19 Vaccination Search Classification,” in the *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2022.
- Schuster, T., Fisch, A., Gupta, J., Dehghani, M., Bahri, D., Tran, V., Tay, Y., and Metzler, D. “Confident Adaptive Language Modeling,” in the *Proceedings of the 36th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.
- Tay, Y., Tran, V., Dehghani, M., Ni, J., Bahri, D., Mehta, H., Qin, Z., Hui, K., Zhao, Z., Gupta, J., Schuster, T., Cohen, W., and Metzler, D. “Transformer Memory as a Differentiable Search Index,” in the *Proceedings of the 36th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.
- Zamani, H., Bendersky, M., Metzler, D., Zhuang, H., and Wang, X. “Stochastic Retrieval-Conditioned Reranking,” in the *Proceedings of the 8th The ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR)*, 2022.
- Lees, A., Tran, V., Tay, Y., Sorensen, J., Gupta, J., Metzler, D., and Vasserman, L. “A New Generation of Perspective API: Efficient Multilingual Character-level Transformers,” in the *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2197–3207, 2022.

- He, Y., Zheng, H., Tay, Y., Gupta, J., Du, Y., Aribandi, V., Zhao, Z., Li, Y., Chen, Z., Metzler, D., Cheng, H., Chi, E. "HyperPrompt: Prompt-based Task-Conditioning of Transformers," in the *Proceedings of the International Conference on Machine Learning (ICML)*, 2022.
- Zamani, H., Diaz, F., Dehghani, M., Metzler, D., and Bendersky, M. "Retrieval-Enhanced Machine Learning," in the *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022.
- Tay, Y., Tran, V., Ruder, S., Gupta, J., Chung, H., Bahri, D., Qin, Z., Baumgartner, S., and Metzler, D. "Charformer: Fast Character Transformers via Gradient-based Subword Tokenization," in the *Proceedings of the Tenth International Conference on Learning Representations (ICLR)*, 2022.
- Aribandi, V., Tay, Y., Schuster, T., Rao, J., Zhen, H., Mehta, S., Zhuang, H., Tran, V., Bahri, D., Ni, J., Gupta, J., Hui, K., Ruder, S., and Metzler, D. "ExT5: Towards Extreme Task Scaling for Transfer Learning," in the *Proceedings of the Tenth International Conference on Learning Representations (ICLR)*, 2022.
- Tay, Y., Dehghani, M., Rao, J., Fedus, W., Abnar, S., Chung, H., Narang, S., Yogatama, D., and Metzler, D. "Scale Efficiently: Insights from Pretraining and Finetuning Transformers," in the *Proceedings of the Tenth International Conference on Learning Representations (ICLR)*, 2022.
- Bahri, D., Jiang, H., Tay, Y., and Metzler, D. "SCARF: Self-Supervised Contrastive Learning using Random Feature Corruption," in the *Proceedings of the Tenth International Conference on Learning Representations (ICLR)*, 2022.
- Shen, Y., Tay, Y., Zheng, C., Bahri, D., Metzler, D., and Courville, A. "StructFormer: Joint Unsupervised Induction of Dependency and Constituency Structure from Masked Language Modeling," in the *Proceedings of the The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, 2021.
- Tay, Y., Dehghani, M., Gupta, J., Aribandi, V., Bahri, D., Qin, Z., and Metzler, D. "Are Pretrained Convolutions Better than Pretrained Transformers?," in the *Proceedings of the The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, 2021.
- Aribandi, V., Tay, Y., and Metzler, D. "Are Model Diagnostics Reliable?," in the *Proceedings of the The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, 2021.
- Tay, Y., Bahri, D., Metzler, D., Juan, D., Zhao, Z., and Zheng, C. "Synthesizer: Rethinking Self-Attention for Transformer Models," in the *Proceedings of the International Conference on Machine Learning (ICML)*, 2021.
- Tay, Y., Dehghani, M., Aribandi, V., Gupta, J., Pham, P., Qin, Z., Bahri, D., Juan, D., and Metzler, D. "OmniNet: Omnidirectional Representations from Transformers," in the *Proceedings of the International Conference on Machine Learning (ICML)*, 2021.
- Tay, Y., Dehghani, M., Abnar, S., Shen, Y., Bahri, D., Pham, P., Rao, J., Yang, L., Ruder, S., and Metzler, D. "Long Range Arena : A Benchmark for Efficient Transformers," in the *Proceedings of the Ninth International Conference on Learning Representations (ICLR)*, 2021.
- Tay, Y., Zhao, Z., Bahri, D., Metzler, D., and Juan, D. "HyperGrid Transformers: Towards A Single Model for Multiple Tasks," in the *Proceedings of the Ninth International Conference on Learning Representations (ICLR)*, 2021.
- Bahri, D., Tay, Y., Zheng, C., Brunk, C., Metzler, D., and Tomkins, A. "Generative Models are Unsupervised Predictors of Page Quality: A Colossal-Scale Study," in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 301-309, 2021.

- Tay, Y., Bahri, D., Yang, L., Metzler, D., and Juan, D. "Sparse Sinkhorn Attention," in the *Proceedings of the International Conference on Machine Learning (ICML)*, 9438-9447, 2020.
- Qin, Z., Chen, S., Metzler, D., Noh, Y., Qin, J., and Wang, X. "Attribute-based Propensity for Unbiased Learning in Recommender Systems: Algorithm and Case Studies," in the *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2359-2367, 2020.
- Chen, S., Qin, Z., Wilson, Z., Calaci, B., Rose, M., Evans, R., Abraham, S., Metzler, D., Tata, S., and Colagrosso, M. "Improving Recommendation Quality in Google Drive," in the *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2900-2908, 2020.
- Qin, Z., Cheng, Y., Zhao, Z., Chen, Z., Metzler, D., and Qin, J. "Multitask Mixture of Sequential Experts for User Activity Streams," in the *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3083-3091, 2020.
- Bahri, D., Tay, Y., Zheng, C., Metzler, D., and Tomkins, A. "Choppy: Cut Transformer for Ranked List Truncation," in the *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1513-1516, 2020.
- Qin, Z., Li, Z., Bendersky, M., and Metzler, D. "Matching Cross Network for Learning to Rank in Personal Search," in the *Proceedings of The Web Conference*, 2835-2841, 2020.
- Li, R., Qin, Z., Wang, X., Chen, S., and Metzler, D. "Stabilizing Neural Search Ranking Models," in the *Proceedings of The Web Conference*, 2725-2732, 2020.
- Tay, Y., Bahri, D., Zheng, C., Brunk, C., Metzler, D., and Tomkins, A. "Reverse Engineering Configurations of Neural Text Generation Models," in the *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 275-279, 2020.
- Chen, S., Wang, X., Qin, Z., and Metzler, D. "Parameter Tuning in Personal Search Systems," in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 97-105, 2020.
- Meng, Y., Karimzadehgan, M., Zhuang, H., and Metzler, D. "Separate and Attend in Personal Email Search," in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 429-437, 2020.
- Li, P., Qin, Z., Wang, X., and Metzler, D. "Combining Decision Trees and Neural Networks for Learning-to-rank in Personal Search," in the *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2032-2040, 2019.
- Li, C., Zhang, M., Bendersky, M., Deng, H., Metzler, D., and Najork, M. "Multi-view Embedding-based Synonyms for Email Search," in the *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 575-584, 2019.
- Ashkan, A. and Metzler, D. "Revisiting Online Personal Search Metrics with the User in Mind," in the *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 625-634, 2019.
- Tran, B., Karimzadehgan, M., Pasumarthi, R., Bendersky, M., Metzler, D. "Domain Adaptation for Enterprise Email Search," in the *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 25-34, 2019.
- Gupta, J. Qin, Z., Bendersky, M., and Metzler, D. "Personalized Online Spell Correction for Personal Search," in the *Proceedings of the the World Wide Web Conference (WWW)*, 2785-2791, 2019.
- Shen, J., Karimzadehgan, M., Bendersky, M., Qin, Z., and Metzler, D. "Multi-Task Learning for Email Search Ranking with Auxiliary Query Clustering," in the *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM)*, 2127-2135, 2018.

- Bendersky, M., Wang, X., Najork, M., and Metzler D. "Learning with Sparse and Biased Feedback for Personal Search," in the *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 5219-5223, 2018.
- Bendersky, M., Wang, X., Metzler, D., and Najork, M. "Learning from User Interactions in Personal Search via Attribute Parameterization," in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 791-799, 2017.
- Wang, X., Bendersky, M., Metzler, D., and Najork, M. "Learning to Rank with Selection Bias in Personal Search," in the *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 115-124, 2016.
- Balasubramanian, A., Balasubramanian, N., Huston, S., Metzler, D., and Wetherall, D. "FindAll: A Local Search Engine for Mobile Phones," in the *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, 277-288, 2012.
- Wang, H., Cai, C., Philpot, A., Latonero, M., Hovy, E., and Metzler, D. "Data Integration from Open Internet Sources to Combat Sex Trafficking of Minors," in the *Proceedings of the 13th Annual International Conference on Digital Government Research (dg.o)*, 246-252, 2012.
- Metzler, D., Cai, C., and Hovy, E. "Structured Event Retrieval over Microblog Archives," in the *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT)*, 646-655, 2012.
- Bendersky, M., Metzler, D., and Croft, W.B. "Effective Query Formulation with Multiple Information Sources," in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 443-452, 2012.
- Metzler, D. and Cai, C. "USC/ISI at TREC 2011: Microblog Track," in the *Proceedings of the 2011 Text REtrieval Conference (TREC)*, 2011.
- Metzler, D., Hovy, E., and Zheng, C. "An Empirical Evaluation of Data-Driven Paraphrase Generation Techniques," in the *Proceedings of the The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*, 546-551, 2011.
- Bendersky, M., Metzler, D., and Croft, W.B. "Parameterized Concept Weighting in Verbose Queries," in the *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 605-614, 2011.
- Asadi, N., Metzler, D., Elsayed, T., and Lin, J. "Pseudo Test Collections for Learning Web Search Ranking Functions," in the *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1073-1082, 2011.
- Asadi, N., Lin, J., and Metzler, D. "Cross-Corpus Relevance Projection," in the *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 1163-1164, 2011.
- Wang, L., Lin, J., and Metzler, D. "A Cascade Ranking Model for Efficient Ranked Retrieval," in the *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 105-114, 2011.
- Elsayed, T., Asadi, N., Metzler, D., Wang, L., and Lin, J. "UMD and USC/ISI: TREC 2010 Web Track Experiments with Ivory," in the *Proceedings of the 2010 Text REtrieval Conference (TREC)*, 2010.
- Lang, H., Metzler, D., Wang, B., and Li, J. "Improved Latent Concept Expansion Using Hierarchical Markov Random Fields," in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 249-258, 2010.
- Wang, L., Metzler, D., and Lin, J. "Ranking Under Temporal Constraints," in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 79-88, 2010.

- Elsayed, T., Lin, J., and Metzler, D. "When Close Enough Is Good Enough: Approximate Positional Indexes for Efficient Ranked Retrieval," in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 1993-1996, 2011.
- Wang, L., Lin, J., and Metzler, D. "Learning to Efficiently Rank," in the *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 138-145, 2010.
- Diaz, F., Metzler, D., and Amer-Yahia, S. "Relevance and Ranking in Online Dating Systems," in the *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 66-73, 2010.
- Bendersky, M., Gabrilovich, E., Josifovski, V., and Metzler, D., "The Anatomy of an Ad: Structured Indexing and Retrieval for Sponsored Search," in the *Proceedings of the 19th International World Wide Web Conference (WWW)*, 101-110, 2010.
- Broder, A., Gabrilovich, E., Josifovski, V., Mavromatis, G., Metzler, D., Wang, J. "Exploiting Site-Level Information to Improve Web Search," in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 1393-1396, 2010.
- Bendersky, M., Metzler, D., and Croft, W.B. "Learning Concept Importance Using a Weighted Dependence Model," in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 31-40, 2010.
- Carterette, B., Gabrilovich, E., Josifovski, V., and Metzler, D. "Measuring the Reusability of Test Collections," in the *Proceedings of the ACM Conference on Web Search and Data Mining (WSDM)*, 231-240, 2010.
- Lin, J., Metzler, D., Elsayed T., and Wang, L., "Of Ivory and Smurfs: Loxodonta MapReduce Experiments for Web Search," in the *Proceedings of the 2009 Text REtrieval Conference (TREC)*, 2009.
- Chapelle, O., Metzler, D., Zhang, Y., and Grinspan, P. "Expected Reciprocal Rank for Graded Relevance," in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 2009.
- Metzler, D., Zaragoza, H. "Semi-Parametric and Non-Parametric Term Weighting for Information Retrieval," in the *Proceedings of the 2nd International Conference on the Theory of Information Retrieval (ICTIR)*, 42-53, 2009.
- Metzler, D., Novak, J., Cui, H., and Reddy, S. "Building Enriched Document Representations using Aggregated Anchor Text," in the *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 219-226, 2009.
- Metzler, D., Jones, R., Peng, F., and Zhang, R. "Improving Search Relevance for Implicitly Temporal Queries," in the *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 700-701, 2009.
- Broder, A. Ciccolo, P., Gabrilovich, E., Josifovski, V., Metzler, D., Riedel, L., and Yuan, J. "Online Expansion of Rare Queries for Sponsored Search," in the *Proceedings of the ACM Conference on the World Wide Web (WWW)*, 511-520, 2009.
- Zhang, R., Chang, Y., Zheng, Z., Metzler, D., and Nie, J. "Search Engine Adaptation by Feedback Control Adjustment for Time-sensitive Query," in the *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT)*, 165-168, 2009.
- Broder, A. Ciaramita, M., Fontoura, M., Gabrilovich, E., Josifovski, V., Metzler, D., Murdock, V., and Plachouras, V. "To Swing or not to Swing: Learning when (not) to Advertise," in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 1003-1012, 2008.

- Turtle, H. and Metzler, D., "CIIR Experiments for TREC Legal 2007," in the *Proceedings of the 2007 Text REtrieval Conference (TREC)*, 2007.
- Metzler, D. "Generalized Inverse Document Frequency," in the *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, 399-408, 2008.
- Metzler, D., Strohman, T., and Croft, W.B., "A Statistical View of Binned Retrieval Models," in the *Proceedings of the 30th European Conference on Information Retrieval (ECIR)*, 175-186, 2008.
- Metzler, D. "Automatic Feature Selection in the Markov Random Field Model for Information Retrieval," in the *Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM)*, 253-262, 2007.
- Metzler, D and Croft, W.B. "Latent Concept Expansion Using Markov Random Fields," in the *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 311-318, 2007.
- Metzler, D. "Using Gradient Descent to Optimize Language Modeling Smoothing Parameters" in the *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 687-688, 2007.
- Metzler, D., Dumais, S., and Meek, C. "Similarity Measures for Short Segments of Text," in the *Proceedings of the 29th European Conference on Information Retrieval (ECIR)*, 16-27, 2007.
- Diaz, F., and Metzler, D. "Pseudo-Aligned Multilingual Corpora," in the *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, 2727-2732, 2007.
- Metzler, D., Strohman, T., and Croft, W.B., "Indri at TREC 2006: Lessons Learned From Three Terabyte Tracks," in the *Proceedings of the 2006 Text REtrieval Conference (TREC)*, 2006.
- Metzler, D., "Estimation, Sensitivity, and Generalization in Parameterized Retrieval Models," in the *Proceedings of the 15th ACM Conference on Information and Knowledge Management (CIKM)*, 2006.
- Metzler, D., and Croft, W.B. "Beyond Bags of Words: Modeling Implicit User Preferences in Information Retrieval," in the *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 2006.
- Diaz, F. and Metzler, D. "Improving the Estimation of Relevance Models Using Large External Corpora," in the *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 154-161, 2006.
- Metzler, D., Strohman T., Zhou, Y., and Croft, W.B., "UMass Robust 2005: Using Mixtures of Relevance Models for Query Expansion," in the *Proceedings of the 2005 Text REtrieval Conference (TREC)*, 2005.
- Metzler, D., Strohman T., Zhou, Y., and Croft, W.B., "Indri at TREC 2005: Terabyte Track," in the *Proceedings of the 2005 Text REtrieval Conference (TREC)*, 2005. Strohman, T., Metzler, D., Turtle, H., and Croft, W.B., "Indri: A Language-Model Based Search Engine for Complex Queries," in the online *Proceedings of the International Conference on Intelligence Analysis*, 2005.
- Metzler, D., Bernstein, Y., Croft, W.B., Moffat, A., and Zobel, J. "Similarity Measures for Tracking Information Flow," in the *Proceedings of the 14th ACM Conference on Information and Knowledge Management (CIKM)*, 517-524, 2005.
- Metzler, D. and Croft, W.B., "A Markov Random Field Model for Term Dependencies," in the *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 472-479, 2005.
- Metzler, D., Bernstein, Y., Croft, W.B., Moffat, A., and Zobel, J., "The Recap System for Identifying Information Flow," *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 678-678, 2005.

- Metzler, D., Strohman T., Turtle H., and Croft, W.B., “Indri at TREC 2004: Terabyte Track,” in the *Proceedings of the 2004 Text REtrieval Conference (TREC)*, 2004.
- Metzler, D., Lavrenko, V., and Croft, W. B., “Formal Multiple-Bernoulli Models for Language Modeling,” *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 540-541, 2004.
- Metzler, D., and Manmatha, R., “An Inference Network Approach to Image Retrieval,” in the *Proceedings of the International Conference on Image and Video Retrieval (CIVR)*, 42-50, 2004

Workshop Papers

- Pradeep, R., Hui, K., Gupta, J., Lelkes, A., Zhuang, H., Lin, J., Metzler, D., and Tran, V. “Understanding Generative Retrieval at Scale,” in the *SIGIR Workshop on Generative Information Retrieval (Gen-IR)*, 2023.
- Metzler, D., and Hovy, E. “Mavuno: A Scalable and Effective Hadoop-Based Paraphrase Acquisition System,” in the *Proceedings of the KDD Workshop on Large-scale Data Mining: Theory and Applications (LDMTA)*, 2011.
- Gouws, S., Hovy, D., and Metzler, D. “Unsupervised Mining of Lexical Variants from Noisy Text,” in the *Proceedings of the EMNLP Workshop on Unsupervised Learning in NLP (UNSUP)*, 2011.
- Gouws, S., Metzler, D., Cai, C., and Hovy, E. “Contextual Bearing on Linguistic Variation in Social Media,” in the *ACL Workshop on Language in Social Media (LSM)*, 2011.
- Metzler, D. and Kanungo, T. “Machine Learned Sentence Selection Strategies for Query-Biased Summarization,” in the *Proceedings of the ACM SIGIR Learning to Rank for Information Retrieval Workshop*, 2008.

Technical Reports

- Gemini Team. “Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities,” Google DeepMind Technical Report, 2025.
- Cui, H., Reddy, S., and Metzler, D. “QueryText – Using Queries and Clicks to Improve Text Matching for Web Search,” Yahoo! Labs Technical Report, 2009.
- Metzler, D., Jones, R., Peng, F., and Zhang, R. “An Analysis of Implicitly Year Qualified Queries,” Yahoo! Labs Technical Report, 2009.
- Metzler, D., “Estimation, Sensitivity, and Generalization in Parameterized Retrieval Models (Extended Version),” CIIR Technical Report, 2006.
- Donald Metzler, “Direct Maximization of Rank-based Metrics,” CIIR Technical report.
- Strohman, T., Metzler, D., Turtle, H., and Croft, W.B., “Indri: A Language-Model Based Search Engine for Complex Queries (Extended Version)”. CIIR Technical Report.

Research Grants

- Key Personnel, DARPA Grant, “Situational Awareness for Social Media: Theories, Models and Algorithms,” University of Southern California, \$6,206,500, 2012-2015.
- PI, NSF Grant, “EAGER: Constructing, Indexing, and Searching Super-Enriched Document Representations in the Cloud,” University of Southern California, \$250,00, 2011-2013.
- PI, NSF Grant, “III: EAGER: Automatically Building Test Collections Using Implicit Relevance Signals from the Web,” \$150,000, University of Southern California, 2011-2012.
- PI, Amazon Web Services Research Grant, University of Southern, California, \$4,000, 2011.
- PI, Yahoo! Faculty Research Engagement Grant, “User Representations for Search, Advertising, and Content Applications”, \$7,000, University of Southern California, 2010-2011.

Open Source Software

- **Mavuno** (<https://github.com/metzlerd/mavuno>), Primary Developer. Mavuno is an open source, modular, scalable text mining toolkit built upon Hadoop. It supports basic natural language processing tasks (e.g., part of speech tagging, chunking, parsing, named entity recognition), is capable of large-scale distributional similarity computations (e.g., synonym, paraphrase, and lexical variant mining), and has information extraction capabilities (e.g., instance and semantic relation mining). It can easily be adapted to new input formats and text mining tasks.
- **Ivory** (<http://lntool.github.com/Ivory/>), Major Contributor. Ivory is a Hadoop-based retrieval toolkit that supports various retrieval models, including the Markov Random Field model for IR. The project is a joint effort with Jimmy Lin and his colleagues at the University of Maryland.
- **Indri** (<http://www.lemurproject.org/indri/>), Contributor. Indri is an efficient, scalable search engine with a robust query language. The project synthesizes and enhances the Lemur and Inquiry search tools.

Workshops Organized

- Benedict, G., Zhang, R., Metzler, D., Yates, A., and Jiang, Z. "Gen-IR @ SIGIR 2024: The Second Workshop on Generative Information Retrieval," full day workshop held in conjunction with the *47th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024.
- Benedict, G, Zhang, R., and Metzler, D. "The 1st Workshop on Generative Information Retrieval (Gen-IR)," full day workshop held in conjunction with the *46th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023.
- Diaz, F., Hovy, E., King, E., Li, J., Metzler, D., Moens, M., Tang, J., and Zhang, L. "The 3rd Workshop on Social Web Search and Mining (SWSM2011): Analysis of User Generated Content Under Crisis," full day workshop held in conjunction with the *34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2011.

Tutorials Organized

- Bendersky, M., Wang, X., Najork, M., and Metzler, D. "Search and Discovery in Personal Email Collections," half-day tutorial presented at the *15th ACM International Web Search and Data Mining Conference (WSDM)*, Phoenix, AZ, 2022.
- Wang, L., Lin, J., Metzler, D., and Han, J. "Learning to Efficiently Rank on Big Data," half-day tutorial presented at the *23rd International World Wide Web Conference (WWW)*, Seoul, Korea, 2014.
- Metzler, D. and Kurland, O. "Empirical Methods for Information Retrieval," half-day tutorial presented at the *35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Portland, OR, 2012.
- Metzler, D. and Lavrenko, V. "Probabilistic Models for Information Retrieval," full-day tutorial presented at the *32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Boston, MA, 2009.

Journal Editorial Boards

- ACM Transactions on Information Systems (TOIS), 2012-2014, 2024-present
- ACM Transactions on the Web (TWEB), 2021-present
- Information Retrieval Journal, 2012-2014

Conference Steering Committees

- International Conference on the Theory of Information Retrieval (ICTIR), 2014-2016

Conference Committee Chair

- Awards Chair, ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2025
- Program Chair, ACM Conference on Web Search and Data Mining (WSDM), 2014
- Program Chair, International Conference on the Theory of Information Retrieval, 2013
- Program Chair, Open Research Areas in Information Retrieval (OAIR/RIAO), 2013
- Best Paper Chair, International Conference on the Theory of Information Retrieval (ICTIR), 2011
- Publicity Chair, ACM Conference on Information and Knowledge Management (CIKM), 2011
- Poster Chair, ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2009

Conference Area Chair / Senior Program Committee Member

- ACM Conference on Information and Knowledge Management (CIKM), 2012, 2016-2018, 2022-2023
- ACM Conference on Web Search and Data Mining (WSDM), 2021, 2024
- ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2011-2012, 2017-2025
- ACM The Web Conference (TheWebConf), 2022
- Asian Information Retrieval Societies Conference (AIRS), 2013
- Conference on Artificial Intelligence (AAAI), 2012
- Conference on Empirical Methods for Natural Language Processing Conference (EMNLP), 2022
- Conference on Neural Information Processing Systems (NeurIPS), 2023-2025
- European Conference on Information Retrieval (ECIR), 2022-2023
- International Conference on Machine Learning (ICML), 2025
- International World Wide Web Conference (WWW), 2011

Conference Program Committee Member

- ACL Rolling Review (ARR), 2021
- ACM Conference on Information and Knowledge Management (CIKM), 2011-2012
- ACM Conference on Web Search and Data Mining (WSDM), 2008-2013, 2022
- ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2007-2011
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2010-2012
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2007, 2011, 2013
- Annual Meeting of the Association for Computational Linguistics (ACL), 2013, 2020-2021
- Asian Information Retrieval Societies Conference (AIRS), 2011
- Conference on Artificial Intelligence (AAAI), 2021
- Conference on Empirical Methods for Natural Language Processing Conference (EMNLP), 2008-2011, 2020-2021
- Conference on Neural Information Processing Systems (NeurIPS), 2021
- Design of Experimental Search & Information RETrieval Systems (DESIREs), 2020-2021, 2024
- European Association for Computational Linguistics Conference (EACL), 2008, 2012
- European Conference on Information Retrieval (ECIR), 2008, 2012

- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), 2020
- International AAAI Conference on Weblogs and Social Media (ICWSM), 2012
- International Conference on Computational Linguistics (COLING), 2010
- International Conference on Machine Learning (ICML), 2008, 2010, 2012
- International Conference on the Theory of Information Retrieval (ICTIR), 2011
- International Joint Conference on Natural Language Processing (IJCNLP), 2011
- International World Wide Web Conference (WWW), 2008, 2010-2014

Workshop Program Committees

- Analyzing Microtext (AAAI 2013 Spring Symposium)
- Application of the Web Workshop (WWW 2012)
- Beyond Binary Relevance Workshop (SIGIR 2008, 2009)
- Enriching Information Retrieval Workshop (SIGIR 2011)
- Information Retrieval for Advertising Workshop (SIGIR 2008, 2009)
- Learning to Rank Workshop for Information Retrieval (SIGIR 2007, 2008)
- Redundancy, Diversity, and Interdependent Document Relevance Workshop (SIGIR 2009)
- Workshop on Crowdsourcing for Search and Data Mining (WSDM 2011)
- Workshop on Crowdsourcing for Search Evaluation (SIGIR 2010-2011)
- Workshop on Enriching Information Retrieval (SIGIR 2011)
- Workshop on Feature Generation and Selection for Information Retrieval (SIGIR 2010)
- Workshop on Narrative Understanding, Storylines, and Events (ACL 2020)
- Workshop on Query Representation and Understanding (SIGIR 2011)
- Workshop on Real-Time Analysis and Mining of Social Streams (ICWSM 2012)
- Workshop on Semantic Matching in Information Retrieval (SIGIR 2014)
- Workshop on The People's Web meets NLP: Collaboratively Constructed Semantic Resources and their Applications to NLP (EACL 2012)
- Workshop on User Modeling for Web Applications (WSDM 2011)
- Workshop on Web Search Result Summarization and Presentation (WWW 2009)

Other Conference Committees

- Best Paper Awards Committee Member, European Conference on Information Retrieval (ECIR), 2012
- Demo PC Member, Human Language Technologies Conference (HLT), 2008
- Doctoral Consortium PC Member, ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2013
- Perspective Paper Track, ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2024
- Student Session PC Member, Annual Meeting of the Association for Computational Linguistics (ACL), 2011
- Tutorial PC Member
 - ACM Conference on Web Search and Data Mining (WSDM), 2012
 - ACM International Conference on Research and Development in Information Retrieval (SIGIR), 2010-2011, 2013
 - European Conference on Information Retrieval (ECIR), 2013
 - International Conference on the Theory of Information Retrieval (ICTIR), 2015
- Workshop PC Member

- ACM Conference on Research and Development in Information Retrieval (SIGIR), 2024
- ACM Conference on Web Search and Data Mining (WSDM), 2011
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2012
- European Conference on Information Retrieval (ECIR), 2013
- Human Language Technologies Conference (HLT), 2008
- International Conference on Learning Representations (ICLR), 2023

Journal Reviewer

- ACM Transactions on Intelligent Systems and Technology (TIST)
- ACM Transactions on Knowledge Discovery from Data (TKDD)
- ACM Transactions on Information Systems (TOIS)
- Communications of the ACM (CACM)
- Data Mining and Knowledge Discovery (DMKD)
- Electronics and Telecommunications Research Institute Journal (ETRI)
- Foundations and Trends in Information Retrieval (FnTIR)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- Information Processing and Management (IP&M)
- Information Processing Letters (IPL)
- Journal of Artificial Intelligence Research (JAIR)
- Journal of the American Society for Information Science and Technology (JASIST)
- Journal of Machine Learning Research (JMLR)
- Pattern Recognition Letters (PATREC)

Other Service

- European Research Council Reviewer (2023, 2024)
- National Science Foundation Panelist (2008, 2011)
- National Science Foundation Ad Hoc Reviewer (2010, 2012)