

# Marc Najork

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/7862157/publications.pdf>

Version: 2024-02-01

70  
papers

3,409  
citations

1478280

6  
h-index

1281743

11  
g-index

73  
all docs

73  
docs citations

73  
times ranked

1364  
citing authors

| #  | ARTICLE  | IF | CITATIONS |
|----|--|----|-----------|
| 1  | Search and Discovery in Personal Email Collections. , 2022, , .  |    | 1         |
| 2  | On Optimizing Top-K Metrics for Neural Ranking Models. , 2022, , .   |    | 4         |
| 3  | Revisiting Two-tower Models for Unbiased Learning to Rank. , 2022, , .   |    | 11        |
| 4  | Improving Cloud Storage Search with User Activity. , 2021, , .   |    | 2         |
| 5  | Ensemble Distillation for BERT-Based Ranking Models. , 2021, , .   |    | 1         |
| 6  | Bootstrapping Recommendations at Chrome Web Store. , 2021, , .   |    | 3         |
| 7  | Natural Language Understanding with Privacy-Preserving BERT. , 2021, , .   |    | 4         |
| 8  | A Stochastic Treatment of Learning to Rank Scoring Functions. , 2020, , .  |    | 33        |
| 9  | Beyond 512 Tokens: Siamese Multi-depth Transformer-based Hierarchical Encoder for Long-Form Document Matching. , 2020, , . |    | 44        |
| 10 | Adversarial Bandits Policy for Crawling Commercial Web Content. , 2020, , .  |    | 2         |
| 11 | Learning to Cluster Documents into Workspaces Using Large Scale Activity Logs. , 2020, , .                                 |    | 3         |
| 12 | Feature Transformation for Neural Ranking Models. , 2020, , .  |    | 13        |
| 13 | Permutation Equivariant Document Interaction Network for Neural Learning to Rank. , 2020, , .                              |    | 11        |
| 14 | Representation Learning for Information Extraction from Form-like Documents. , 2020, , .                                   |    | 58        |
| 15 | Semantic Text Matching for Long-Form Documents. , 2019, , .  |    | 49        |
| 16 | Revisiting Approximate Metric Optimization in the Age of Deep Neural Networks. , 2019, , .                                 |    | 37        |
| 17 | TF-Ranking. , 2019, , .  |    | 70        |
| 18 | Multi-view Embedding-based Synonyms for Email Search. , 2019, , .  |    | 12        |

| #  | ARTICLE  | IF  | CITATIONS |
|----|--|-----|-----------|
| 19 | Addressing Trust Bias for Unbiased Learning-to-Rank. , 2019, , .   |     | 33        |
| 20 | RiSER: Learning Better Representations for Richly Structured Emails. , 2019, , .                                       |     | 7         |
| 21 | Predictive Crawling for Commercial Web Content. , 2019, , .  |     | 3         |
| 22 | Uncovering Hidden Structure in Sequence Data via Threading Recurrent Models. , 2019, , .                               |     | 1         |
| 23 | Estimating Position Bias without Intrusive Interventions. , 2019, , .  |     | 91        |
| 24 | Learning Groupwise Multivariate Scoring Functions Using Deep Neural Networks. , 2019, , .                              |     | 39        |
| 25 | An Analysis of the Softmax Cross Entropy Loss for Learning-to-Rank with Binary Relevance. , 2019, , .                  |     | 40        |
| 26 | Online template induction for machine-generated emails. Proceedings of the VLDB Endowment, 2019, 12, 1235-1248.        | 2.1 | 1         |
| 27 | Position Bias Estimation for Unbiased Learning to Rank in Personal Search. , 2018, , .                                 |     | 156       |
| 28 | Hidden in Plain Sight. , 2018, , .   |     | 3         |
| 29 | Learning Effective Embeddings for Machine Generated Emails with Applications to Email Category Prediction. , 2018, , . |     | 2         |
| 30 | Anatomy of a Privacy-Safe Large-Scale Information Extraction System Over Email. , 2018, , .                            |     | 10        |
| 31 | The LambdaLoss Framework for Ranking Metric Optimization. , 2018, , .  |     | 65        |
| 32 | Semantic Location in Email Query Suggestion. , 2018, , .   |     | 0         |
| 33 | Learning with Sparse and Biased Feedback for Personal Search. , 2018, , .  |     | 4         |
| 34 | Web Crawler Architecture. , 2018, , 4608-4611.   |     | 2         |
| 35 | Web Spam Detection. , 2018, , 4677-4681.   |     | 0         |
| 36 | Learning from User Interactions in Personal Search via Attribute Parameterization. , 2017, , .                         |     | 33        |

| #  | ARTICLE   | IF  | CITATIONS |
|----|---|-----|-----------|
| 37 | Quick Access. , 2017, , .   |     | 17        |
| 38 | Email Category Prediction. , 2017, , .  |     | 8         |
| 39 | Web Crawler Architecture. , 2017, , 1-4.  |     | 1         |
| 40 | Web Spam Detection. , 2017, , 1-5.  |     | 0         |
| 41 | Learning to Rank with Selection Bias in Personal Search. , 2016, , .  |     | 176       |
| 42 | Using Machine Learning to Improve the Email Experience. , 2016, , .   |     | 3         |
| 43 | Debugging a Crowdsourced Task with Low Inter-Rater Agreement. , 2015, , .   |     | 15        |
| 44 | Are Some Tweets More Interesting Than Others? #HardQuestion. , 2013, , .  |     | 21        |
| 45 | The Power of Peers. Lecture Notes in Computer Science, 2011, , 497-502.   | 1.0 | 1         |
| 46 | Web Crawling. Foundations and Trends in Information Retrieval, 2010, 4, 175-246.                                    | 5.8 | 261       |
| 47 | Querying the Web Graph. Lecture Notes in Computer Science, 2010, , 1-12.  | 1.0 | 1         |
| 48 | The scalable hyperlink store. , 2009, , .   |     | 11        |
| 49 | Web Crawler Architecture. , 2009, , 3462-3465.  |     | 23        |
| 50 | Web Spam Detection. , 2009, , 3520-3523.  |     | 8         |
| 51 | Efficient and effective link analysis with precomputed salsa maps. , 2008, , .                                      |     | 7         |
| 52 | Computing Information Retrieval Performance Measures Efficiently in the Presence of Tied Scores. , 2008, , 414-421. |     | 33        |
| 53 | Using Bloom Filters to Speed Up HITS-Like Ranking Algorithms. , 2007, , 195-201.                                    |     | 3         |
| 54 | Detecting spam web pages through content analysis. , 2006, , .  |     | 413       |

| #  | ARTICLE  | IF  | CITATIONS |
|----|--|-----|-----------|
| 55 | Adversarial information retrieval on the web (AIRWeb 2006). ACM SIGIR Forum, 2006, 40, 27-30.                                | 0.4 | 4         |
| 56 | Detecting phrase-level duplication on the world wide web. , 2005, , .  |     | 83        |
| 57 | A large-scale study of the evolution of Web pages. Software - Practice and Experience, 2004, 34, 213-237.                    | 2.5 | 117       |
| 58 | Spam, damn spam, and statistics. , 2004, , .   |     | 198       |
| 59 | A large-scale study of the evolution of web pages. , 2003, , .   |     | 210       |
| 60 | Efficient URL caching for world wide web crawling. , 2003, , .   |     | 51        |
| 61 | High-Performance Web Crawling. Massive Computing, 2002, , 25-45.   | 0.4 | 48        |
| 62 | Breadth-first crawling yields high-quality pages. , 2001, , .  |     | 200       |
| 63 | Web-based algorithm animation. , 2001, , .   |     | 6         |
| 64 | Performance limitations of the Java core libraries. Concurrency and Computation: Practice and Experience, 2000, 12, 363-373. | 0.6 | 4         |
| 65 | On near-uniform URL sampling. Computer Networks, 2000, 33, 295-308.  | 3.2 | 145       |
| 66 | Performance limitations of the Java core libraries. , 1999, , .  |     | 18        |
| 67 | Measuring index quality using random walks on the Web. Computer Networks, 1999, 31, 1291-1303.                               | 3.2 | 78        |
| 68 | Mercator: A scalable, extensible Web crawler. World Wide Web, 1999, 2, 219-229.  | 2.7 | 365       |
| 69 | Cube: Eine dreidimensionale visuelle Programmiersprache. Informatik Aktuell, 1993, , 340-345.                                | 0.4 | 0         |
| 70 | Roles and their role in posing recursive queries. Information Systems, 1990, 15, 173-186.                                    | 2.4 | 0         |