

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	Detecting spam web pages through content analysis. , 2006, , .		413
2	Mercator: A scalable, extensible Web crawler. World Wide Web, 1999, 2, 219-229.	2.7	365
3	Web Crawling. Foundations and Trends in Information Retrieval, 2010, 4, 175-246.	5.8	261
4	A large-scale study of the evolution of web pages. , 2003, , .		210
5	Breadth-first crawling yields high-quality pages. , 2001, , .		200
6	Spam, damn spam, and statistics. , 2004, , .		198
7	Learning to Rank with Selection Bias in Personal Search. , 2016, , .		176
8	Position Bias Estimation for Unbiased Learning to Rank in Personal Search. , 2018, , .		156
9	On near-uniform URL sampling. Computer Networks, 2000, 33, 295-308.	3.2	145
10	A large-scale study of the evolution of Web pages. Software - Practice and Experience, 2004, 34, 213-237.	2.5	117
11	Estimating Position Bias without Intrusive Interventions. , 2019, , .		91
12	Detecting phrase-level duplication on the world wide web. , 2005, , .		83
13	Measuring index quality using random walks on the Web. Computer Networks, 1999, 31, 1291-1303.	3.2	78
14	TF-Ranking. , 2019, , .		70
15	The LambdaLoss Framework for Ranking Metric Optimization. , 2018, , .		65
16	Representation Learning for Information Extraction from Form-like Documents. , 2020, , .		58
17	Efficient URL caching for world wide web crawling. , 2003, , .		51
18	Semantic Text Matching for Long-Form Documents. , 2019, , .		49

#	ARTICLE	IF	CITATIONS
19	High-Performance Web Crawling. Massive Computing, 2002, , 25-45.	0.4	48
20	Beyond 512 Tokens: Siamese Multi-depth Transformer-based Hierarchical Encoder for Long-Form Document Matching. , 2020, , .		44
21	An Analysis of the Softmax Cross Entropy Loss for Learning-to-Rank with Binary Relevance. , 2019, , .		40
22	Learning Groupwise Multivariate Scoring Functions Using Deep Neural Networks. , 2019, , .		39
23	Revisiting Approximate Metric Optimization in the Age of Deep Neural Networks. , 2019, , .		37
24	Learning from User Interactions in Personal Search via Attribute Parameterization. , 2017, , .		33
25	Addressing Trust Bias for Unbiased Learning-to-Rank. , 2019, , .		33
26	Computing Information Retrieval Performance Measures Efficiently in the Presence of Tied Scores. , 2008, , 414-421.		33
27	A Stochastic Treatment of Learning to Rank Scoring Functions. , 2020, , .		33
28	Web Crawler Architecture. , 2009, , 3462-3465.		23
29	Are Some Tweets More Interesting Than Others? #HardQuestion. , 2013, , .		21
30	Performance limitations of the Java core libraries. , 1999, , .		18
31	Quick Access. , 2017, , .		17
32	Debugging a Crowdsourced Task with Low Inter-Rater Agreement. , 2015, , .		15
33	Feature Transformation for Neural Ranking Models. , 2020, , .		13
34	Multi-view Embedding-based Synonyms for Email Search. , 2019, , .		12
35	The scalable hyperlink store. , 2009, , .		11
36	Permutation Equivariant Document Interaction Network for Neural Learning to Rank. , 2020, , .		11

#	ARTICLE	IF	CITATIONS
37	Revisiting Two-tower Models for Unbiased Learning to Rank. , 2022, , .		11
38	Anatomy of a Privacy-Safe Large-Scale Information Extraction System Over Email. , 2018, , .		10
39	Email Category Prediction. , 2017, , .		8
40	Web Spam Detection. , 2009, , 3520-3523.		8
41	Efficient and effective link analysis with precomputed salsa maps. , 2008, , .		7
42	RiSER: Learning Better Representations for Richly Structured Emails. , 2019, , .		7
43	Web-based algorithm animation. , 2001, , .		6
44	Performance limitations of the Java core libraries. Concurrency and Computation: Practice and Experience, 2000, 12, 363-373.	0.6	4
45	Learning with Sparse and Biased Feedback for Personal Search. , 2018, , .		4
46	Adversarial information retrieval on the web (AIRWeb 2006). ACM SIGIR Forum, 2006, 40, 27-30.	0.4	4
47	Natural Language Understanding with Privacy-Preserving BERT. , 2021, , .		4
48	On Optimizing Top-K Metrics for Neural Ranking Models. , 2022, , .		4
49	Using Bloom Filters to Speed Up HITS-Like Ranking Algorithms. , 2007, , 195-201.		3
50	Using Machine Learning to Improve the Email Experience. , 2016, , .		3
51	Hidden in Plain Sight. , 2018, , .		3
52	Predictive Crawling for Commercial Web Content. , 2019, , .		3
53	Bootstrapping Recommendations at Chrome Web Store. , 2021, , .		3
54	Learning to Cluster Documents into Workspaces Using Large Scale Activity Logs. , 2020, , .		3

#	ARTICLE	IF	CITATIONS
55	Learning Effective Embeddings for Machine Generated Emails with Applications to Email Category Prediction. , 2018, , .		2
56	Improving Cloud Storage Search with User Activity. , 2021, , .		2
57	Adversarial Bandits Policy for Crawling Commercial Web Content. , 2020, , .		2
58	Web Crawler Architecture. , 2018, , 4608-4611.		2
59	Uncovering Hidden Structure in Sequence Data via Threading Recurrent Models. , 2019, , .		1
60	Ensemble Distillation for BERT-Based Ranking Models. , 2021, , .		1
61	Querying the Web Graph. Lecture Notes in Computer Science, 2010, , 1-12.	1.0	1
62	The Power of Peers. Lecture Notes in Computer Science, 2011, , 497-502.	1.0	1
63	Web Crawler Architecture. , 2017, , 1-4.		1
64	Online template induction for machine-generated emails. Proceedings of the VLDB Endowment, 2019, 12, 1235-1248.	2.1	1
65	Search and Discovery in Personal Email Collections. , 2022, , .		1
66	Roles and their role in posing recursive queries. Information Systems, 1990, 15, 173-186.	2.4	0
67	Semantic Location in Email Query Suggestion. , 2018, , .		0
68	Cube: Eine dreidimensionale visuelle Programmiersprache. Informatik Aktuell, 1993, , 340-345.	0.4	0
69	Web Spam Detection. , 2017, , 1-5.		0
70	Web Spam Detection. , 2018, , 4677-4681.		0