W John Wilbur

List of Publications by Year
in descending order

70 papers	2,106 citations	304368 22 h-index	243296 44 g-index
70 all docs	70 docs citations	70 times ranked	2093 citing authors

#	Article	IF	Citations
1	Tagging gene and protein names in biomedical text. Bioinformatics, 2002, 18, 1124-1132.	1.8	296
2	The automatic identification of stop words. Journal of Information Science, 1992, 18, 45-55.	2.0	218
3	PubMed related articles: a probabilistic topic-based model for content similarity. BMC Bioinformatics, 2007, 8, 423.	1.2	154
4	Evaluation of query expansion using MeSH in PubMed. Information Retrieval, 2009, 12, 69-80.	1.6	139
5	The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text. BMC Bioinformatics, 2011, 12, S3.	1.2	121
6	A theoretical basis for large coefficient of variation and bimodality in neuronal interspike interval distributions. Journal of Theoretical Biology, 1983, 105, 345-368.	0.8	116
7	New directions in biomedical text annotation: definitions, guidelines and corpus construction. BMC Bioinformatics, 2006, 7, 356.	1.2	112
8	Multi-dimensional classification of biomedical text: Toward automated, practical provision of high-utility text to diverse users. Bioinformatics, 2008, 24, 2086-2093.	1.8	87
9	Author name disambiguation for <scp>P</scp>ub<scp>M</scp>ed. Journal of the Association for Information Science and Technology, 2014, 65, 765-781.	1.5	68
10	An analysis of Stein's model for stochastic neuronal excitation. Biological Cybernetics, 1982, 45, 107-114.	0.6	52
11	Evaluating Relevance Ranking Strategies for MEDLINE Retrieval. Journal of the American Medical Informatics Association: JAMIA, 2009, 16, 32-36.	2.2	45
12	Optimal Training Sets for Bayesian Prediction of MeSH(R) Assignment. Journal of the American Medical Informatics Association: JAMIA, 2008, 15, 546-553.	2.2	41
13	LitSense: making sense of biomedical literature at sentence level. Nucleic Acids Research, 2019, 47, W594-W599.	6.5	37
14	Bridging the gap: Incorporating a semantic similarity measure for effectively mapping PubMed queries to documents. Journal of Biomedical Informatics, 2017, 75, 122-127.	2.5	33
15	Meshable: searching PubMed abstracts by utilizing MeSH and MeSH-derived topical terms. Bioinformatics, 2016, 32, 3044-3046.	1.8	32
16	Assisting manual literature curation for protein-protein interactions using BioQRator. Database: the Journal of Biological Databases and Curation, 2014, 2014, bau067-bau067.	1.4	28
17	BioCreative V BioC track overview: collaborative biocurator assistant task for BioGRID. Database: the Journal of Biological Databases and Curation, 2016, 2016, baw121.	1.4	28
18	Evolving use of ancestry, ethnicity, and race in genetics researchâ€"A survey spanning seven decades. American Journal of Human Genetics, 2021, 108, 2215-2223.	2.6	27

#	Article	IF	Citations
19	Non-parametric significance tests of retrieval performance comparisons. Journal of Information Science, 1994, 20, 270-284.	2.0	26
20	How to Get the Most out of Your Curation Effort. PLoS Computational Biology, 2009, 5, e1000391.	1.5	26
21	Spelling correction in the PubMed search engine. Information Retrieval, 2006, 9, 543-564.	1.6	25
22	The ineffectiveness of within-document term frequency in text classification. Information Retrieval, 2009, 12, 509-525.	1.6	25
23	The BioC-BioGRID corpus: full text articles annotated for curation of proteinâ€“protein and genetic interactions. Database: the Journal of Biological Databases and Curation, 2017, 2017, baw147.	1.4	24
24	Modeling actions of PubMed users with n-gram language models. Information Retrieval, 2009, 12, 487-503.	1.6	23
25	GENERATION OF A LARGE GENE/PROTEIN LEXICON BY MORPHOLOGICAL PATTERN ANALYSIS. Journal of Bioinformatics and Computational Biology, 2004, 01, 611-626.	0.3	19
26	Navigating information spaces: A case study of related article search in PubMed. Information Processing and Management, 2008, 44, 1771-1783.	5.4	17
27	The Synergy Between PAV and AdaBoost. Machine Learning, 2005, 61, 71-103.	3.4	15
28	BioC interoperability track overview. Database: the Journal of Biological Databases and Curation, 2014, 2014, bau053-bau053.	1.4	15
29	Deep learning with sentence embeddings pre-trained on biomedical corpora improves the performance of finding similar sentences in electronic medical records. BMC Medical Informatics and Decision Making, 2020, 20, 73.	1.5	15
30	Improving accuracy for identifying related PubMed queries by an integrated approach. Journal of Biomedical Informatics, 2009, 42, 831-838.	2.5	14
31	PubMed Phrases, an open set of coherent phrases for searching biomedical literature. Scientific Data, 2018, 5, 180104.	2.4	13
32	A THEMATIC ANALYSIS OF THE AIDS LITERATURE. , 2001, , .		13
33	Corpus-based statistical screening for content-bearing terms. Journal of the Association for Information Science and Technology, 2001, 52, 247-259.	2.6	12
34	Retro: concept-based clustering of biomedical topical sets. Bioinformatics, 2014, 30, 3240-3248.	1.8	12
35	Finding abbreviations in biomedical literature: three BioC-compatible modules and four BioC-formatted corpora. Database: the Journal of Biological Databases and Curation, 2014, 2014, bau044-bau044.	1.4	12
36	The importance of the lexicon in tagging biological text. Natural Language Engineering, 2006, 12, 335-351.	2.1	10

#	Article	IF	Citations
37	Syntactic sentence compression in the biomedical domain: facilitating access to related articles. Information Retrieval, 2007, 10, 393-414.	1.6	10
38	How to interpret PubMed queries and why it matters. Journal of the Association for Information Science and Technology, 2009, 60, 264-274.	2.6	10
39	Identifying named entities from PubMedÂ® for enriching semantic categories. BMC Bioinformatics, 2015, 16, 57.	1.2	10
40	PubMed Text Similarity Model and its application to curation efforts in the Conserved Domain Database. Database: the Journal of Biological Databases and Curation, 2019, 2019, .	1.4	10
41	Natural language processing pipelines to annotate BioC collections with an application to the NCBI disease corpus. Database: the Journal of Biological Databases and Curation, 2014, 2014, bau056-bau056.	1.4	8
42	BioC implementations in Go, Perl, Python and Ruby. Database: the Journal of Biological Databases and Curation, 2014, 2014, bau059-bau059.	1.4	8
43	A thematic analysis of the AIDS literature. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 2002, , 386-97.	0.7	8
44	A comparison of group and individual performance among subject experts and untrained workers at the document retrieval task. Journal of the Association for Information Science and Technology, 1998, 49, 517-529.	1.2	7
45	Non-word identification or spell checking without a dictionary. Journal of the Association for Information Science and Technology, 2004, 55, 169-177.	2.6	7
46	Finding related sentence pairs in MEDLINE. Information Retrieval, 2010, 13, 601-617.	1.6	7
47	Global term weights for document retrieval learned from TREC data. Journal of Information Science, 2001, 27, 303-310.	2.0	6
48	Better synonyms for enriching biomedical search. Journal of the American Medical Informatics Association: JAMIA, 2020, 27, 1894-1902.	2.2	6
49	Retrieval testing by the comparison of statistically independent retrieval methods. Journal of the Association for Information Science and Technology, 1992, 43, 358-370.	1.2	5
50	Improving a gold standard: treating human relevance judgments of MEDLINE document pairs. BMC Bioinformatics, 2011, 12, S5.	1.2	5
51	BioC viewer: a web-based tool for displaying and merging annotations in BioC. Database: the Journal of Biological Databases and Curation, 2016, 2016, baw106.	1.4	5
52	MeSH-based dataset for measuring the relevance of text retrieval. , 2018, , .		5
53	A strategy for assigning new concepts in the MEDLINE database. AMIA ... Annual Symposium proceedings, 2005, , 395-9.	0.2	5
54	Modeling Text Retrieval in Biomedicine. , 2005, , 277-297.		4

#	Article	IF	Citations
55	Identifying Abbreviation Definitions Machine Learning with Naturally Labeled Data. , 2010, , .		4
56	A Field Sensor: computing the composition and intent of PubMed queries. Database: the Journal of Biological Databases and Curation, 2018, 2018, .	1.4	4
57	Stochastic Gradient Descent and the Prediction of MeSH for PubMed Records. AMIA ... Annual Symposium proceedings, 2014, 2014, 1198-207.	0.2	4
58	Improving a Gold Standard: Treating Human Relevance Judgments of MEDLINE Document Pairs. , 2010, 2010, 491-498.		3
59	Identifying well-formed biomedical phrases in MEDLINEÂ® text. Journal of Biomedical Informatics, 2012, 45, 1035-1041.	2.5	3
60	Retrieval testing with hypergeometric document models. Journal of the Association for Information Science and Technology, 1993, 44, 340-351.	1.2	2
61	The statistics of unique native states for random peptides. Biopolymers, 1998, 38, 447-459.	1.2	2
62	A Study of the Morpho-Semantic Relationship in Medline. The Open Information Systems Journal, 2013, 6, 1-12.	0.1	2
63	Using MEDLINE as a knowledge source for disambiguating abbreviations in full-text biomedical journal articles. , 0, , .		1
64	Characterizing RNA Secondary-Structure Features and Their Effects on Splice-Site Prediction. , 2007, , .		1
65	Comparison of Two Methods for Finding Biomedical Categories in Medline. , 2011, , .		1
66	Discovering themes in biomedical literature using a projection-based algorithm. BMC Bioinformatics, 2018, 19, 269.	1.2	1
67	The statistics of unique native states for random peptides. , 1996, 38, 447.		1
68	The dimensions of indexing. AMIA ... Annual Symposium proceedings, 2003, , 714-0.	0.2	1
69	An EM Clustering Algorithm which Produces a Dual Representation. , 2011, , .		0
70	PDC - a probabilistic distributional clustering algorithm: a case study on suicide articles in PubMed. AMIA Summits on Translational Science Proceedings, 2020, 2020, 259-268.	0.4	0