# Justin Zobel

## List of Publications by Year
## in descending order

| 63 papers | 5,990 citations | 331259<br>21 h-index | 214527<br>47 g-index |
|---|---|---|---|
| 70 all docs | 70 docs citations | 70 times ranked | 6549 citing authors |

| # | Article | IF | Citations |
|---|---------|----|-----------|
| 1 | Benchmarks for measurement of duplicate detection methods in nucleotide databases. Database: the Journal of Biological Databases and Curation, 2023, 2023, . | 1.4 | 10 |
| 2 | Measurement of clustering effectiveness for document collections. Information Retrieval, 2022, 25, 239-268. | 1.6 | 5 |
| 3 | Immediate Text Search on Streams Using Apoptosic Indexes. Lecture Notes in Computer Science, 2022, , 157-169. | 1.0 | 2 |
| 4 | Exploring automatic inconsistency detection for literature-based gene ontology annotation. Bioinformatics, 2022, 38, i273-i281. | 1.8 | 0 |
| 5 | Automatic consistency assurance for literature-based gene ontology annotation. BMC Bioinformatics, 2021, 22, 565. | 1.2 | 2 |
| 6 | Document Clustering vs Topic Models: A Case Study. , 2021, , . |  | 1 |
| 7 | Quality Matters: Biocuration Experts on the Impact of Duplication and Other Data Quality Issues in Biological Databases. Genomics, Proteomics and Bioinformatics, 2020, 18, 91-103. | 3.0 | 14 |
| 8 | GeneMates: an R package for detecting horizontal gene co-transfer between bacteria using gene-gene associations controlled for population structure. BMC Genomics, 2020, 21, 658. | 1.2 | 9 |
| 9 | Generation of Synthetic Query Auto Completion Logs. Lecture Notes in Computer Science, 2020, , 621-635. | 1.0 | 6 |
| 10 | Exploring effective approaches for haplotype block phasing. BMC Bioinformatics, 2019, 20, 540. | 1.2 | 24 |
| 11 | Abstraction of query autoÂcompletion logs for anonymity-preserving analysis. Information Retrieval, 2019, 22, 499-524. | 1.6 | 4 |
| 12 | Automated assessment of biological database assertions using the scientific literature. BMC Bioinformatics, 2019, 20, 216. | 1.2 | 3 |
| 13 | Modeling User Actions in Job Search. Lecture Notes in Computer Science, 2019, , 652-664. | 1.0 | 3 |
| 14 | Search Effectiveness in Nonredundant Sequence Databases: Assessments and Solutions. Journal of Computational Biology, 2019, 26, 605-617. | 0.8 | 1 |
| 15 | Can machine translation systems be evaluated by the crowd alone. Natural Language Engineering, 2017, 23, 3-30. | 2.1 | 70 |
| 16 | Duplicates, redundancies and inconsistencies in the primary nucleotide databases: a descriptive study. Database: the Journal of Biological Databases and Curation, 2017, 2017, baw163. | 1.4 | 36 |
| 17 | Literature consistency of bioinformatics sequence databases is effective for assessing record quality. Database: the Journal of Biological Databases and Curation, 2017, 2017, . | 1.4 | 8 |
| 18 | Automated detection of records in biological sequence databases that are inconsistent with the literature. Journal of Biomedical Informatics, 2017, 71, 229-240. | 2.5 | 12 |

| # | Article | IF | Citations |
|---|---------|-----|-----------|
| 19 | A Taxonomy of Query Auto Completion Modes. , 2017, , . | | 14 |
| 20 | Learning Biological Sequence Types Using the Literature. , 2017, , . | | 2 |
| 21 | Evaluation of CD-HIT for constructing non-redundant databases. , 2016, , . | | 10 |
| 22 | A categorical analysis of coreference resolution errors in biomedical texts. Journal of Biomedical Informatics, 2016, 60, 309-318. | 2.5 | 7 |
| 23 | Coreference resolution improves extraction of Biological Expression Language statements from texts. Database: the Journal of Biological Databases and Curation, 2016, 2016, baw076. | 1.4 | 9 |
| 24 | Medical information retrieval: introduction to the special issue. Information Retrieval, 2016, 19, 1-5. | 1.6 | 26 |
| 25 | Supervised Learning for Detection of Duplicates in Genomic Sequence Databases. PLoS ONE, 2016, 11, e0159644. | 1.1 | 10 |
| 26 | High performance computing enabling exhaustive analysis of higher order single nucleotide polymorphism interaction in Genome Wide Association Studies. Health Information Science and Systems, 2015, 3, S3. | 3.4 | 24 |
| 27 | Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics, 2015, 31, 3350-3352. | 1.8 | 1,671 |
| 28 | Evaluation of a Machine Learning Duplicate Detection Method for Bioinformatics Databases. , 2015, , . | | 17 |
| 29 | SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Medicine, 2014, 6, 90. | 3.6 | 953 |
| 30 | Accurate and Robust Genomic Prediction of Celiac Disease Using Statistical Learning. PLoS Genetics, 2014, 10, e1004137. | 1.5 | 95 |
| 31 | Using Inter-file Similarity to Improve Intra-file Compression. , 2014, , . | | 3 |
| 32 | Storing a Collection of Differentially Compressed Files Recursively. , 2014, , . | | 1 |
| 33 | Supercomputing enabling exhaustive statistical analysis of genome wide association study data: Preliminary results. , 2012, 2012, 1258-61. | | 5 |
| 34 | Redesigning the string hash table, burst trie, and BST to exploit cache. Journal of Experimental Algorithmics, 2010, 15, . | 0.7 | 8 |
| 35 | Against recall. ACM SIGIR Forum, 2009, 43, 3-8. | 0.4 | 16 |
| 36 | Rank-biased precision for measurement of retrieval effectiveness. ACM Transactions on Information Systems, 2008, 27, 1-27. | 3.8 | 405 |

| # | ARTICLE | IF | CITATIONS |
|---|---------|-----|-----------|
| 37 | SICO: A System for Detection of Near-Duplicate Images During Search. , 2007, , . | | 8 |
| 38 | Does topic metadata help with Web search?. Journal of the Association for Information Science and Technology, 2007, 58, 613-628. | 2.6 | 12 |
| 39 | Efficient plagiarism detection for large code repositories. Software - Practice and Experience, 2007, 37, 151-175. | 2.5 | 64 |
| 40 | A pipelined architecture for distributed text query evaluation. Information Retrieval, 2007, 10, 205-231. | 1.6 | 88 |
| 41 | Compression techniques for fast external sorting. VLDB Journal, 2007, 16, 269-291. | 2.7 | 9 |
| 42 | Inverted files for text search engines. ACM Computing Surveys, 2006, 38, 6. | 16.1 | 797 |
| 43 | Searchable words on the Web. International Journal on Digital Libraries, 2005, 5, 99-105. | 1.1 | 33 |
| 44 | Fast on-line index construction by geometric partitioning. , 2005, , . | | 37 |
| 45 | Chinese OOV translation and post-translation query expansion in chinese--english cross-lingual information retrieval. ACM Transactions on Asian Language Information Processing, 2005, 4, 57-77. | 0.8 | 14 |
| 46 | An architecture for effective music information retrieval. Journal of the Association for Information Science and Technology, 2004, 55, 1053-1057. | 2.6 | 8 |
| 47 | Methods for identifying versioned and plagiarized documents. Journal of the Association for Information Science and Technology, 2003, 54, 203-215. | 2.6 | 220 |
| 48 | Efficient single-pass index construction for text databases. Journal of the Association for Information Science and Technology, 2003, 54, 713-729. | 2.6 | 73 |
| 49 | Burst tries. ACM Transactions on Information Systems, 2002, 20, 192-223. | 3.8 | 113 |
| 50 | Selfâ€adjusting trees in practice for large text collections. Software - Practice and Experience, 2001, 31, 925-939. | 2.5 | 29 |
| 51 | Compression techniques for Chinese text. Software - Practice and Experience, 1998, 28, 1299-1314. | 2.5 | 2 |
| 52 | Inverted files versus signature files for text indexing. ACM Transactions on Database Systems, 1998, 23, 453-490. | 1.5 | 249 |
| 53 | How reliable are the results of large-scale information retrieval experiments?. , 1998, , . | | 359 |
| 54 | Passage retrieval revisited. ACM SIGIR Forum, 1997, 31, 178-185. | 0.4 | 35 |

| # | Article | IF | Citations |
|---|---------|----|-----------| 
| 55 | A Formal Model for Representation and Querying of Structured Documents. Journal of Systems Integration, 1997, 7, 31-46. | 0.1 | 1 |
| 56 | Performance in Practice of String Hashing Functions. , 1997, , . |  | 40 |
| 57 | Filtered document retrieval with frequency-sorted indexes. Journal of the Association for Information Science and Technology, 1996, 47, 749-764. | 1.2 | 66 |
| 58 | Filtered document retrieval with frequency-sorted indexes. , 1996, 47, 749. |  | 4 |
| 59 | Filtered document retrieval with frequency-sorted indexes. , 1996, 47, 749. |  | 63 |
| 60 | Finding approximate matches in large lexicons. Software - Practice and Experience, 1995, 25, 331-345. | 2.5 | 94 |
| 61 | Adding compression to a full-text retrieval system. Software - Practice and Experience, 1995, 25, 891-903. | 2.5 | 72 |
| 62 | Rejoinder to letter from Stephen P. Harter. Journal of the Association for Information Science and Technology, 1993, 44, 213-213. | 1.2 | 0 |
| 63 | A model for word clustering. Journal of the Association for Information Science and Technology, 1992, 43, 616-627. | 1.2 | 8 |