

C Lee Giles

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/3320536/publications.pdf>

Version: 2024-02-01

316
papers

18,576
citations

47006

47
h-index

27406

106
g-index

323
all docs

323
docs citations

323
times ranked

11512
citing authors

#	ARTICLE	IF	CITATIONS
1	Face recognition: a convolutional neural-network approach. IEEE Transactions on Neural Networks, 1997, 8, 98-113.	4.2	2,444
2	Accessibility of information on the web. Nature, 1999, 400, 107-107.	27.8	1,118
3	Self-organization and identification of Web communities. Computer, 2002, 35, 66-70.	1.1	762
4	Searching the World Wide Web. Science, 1998, 280, 98-100.	12.6	736
5	Learning, invariance, and generalization in high-order neural networks. Applied Optics, 1987, 26, 4972.	2.1	587
6	Learning long-term dependencies in NARX recurrent neural networks. IEEE Transactions on Neural Networks, 1996, 7, 1329-1338.	4.2	551
7	Efficient identification of Web communities. , 2000, , .		539
8	CiteSeer. , 1998, , .		528
9	Digital libraries and autonomous citation indexing. Computer, 1999, 32, 67-71.	1.1	389
10	Learning and Extracting Finite State Automata with Second-Order Recurrent Neural Networks. Neural Computation, 1992, 4, 393-405.	2.2	353
11	Winners don't take all: Characterizing the competition for links on the web. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99, 5207-5211.	7.1	340
12	Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference. Machine Learning, 2001, 44, 161-183.	5.4	258
13	The Number of Scholarly Documents on the Public Web. PLoS ONE, 2014, 9, e93949.	2.5	218
14	CiteSeer. , 1998, , .		186
15	Name disambiguation in author citations using a K-way spectral clustering method. , 2005, , .		172
16	Co-ranking Authors and Documents in a Heterogeneous Network. , 2007, , .		172
17	Real-time automatic tag recommendation. , 2008, , .		171
18	The CHEMDNER corpus of chemicals and drugs and its annotation principles. Journal of Cheminformatics, 2015, 7, S2.	6.1	166

#	ARTICLE	IF	CITATIONS
19	Probabilistic models for discovering e-communities. , 2006, , .		161
20	The Real Power of Artificial Markets. Science, 2001, 291, 987-988.	12.6	160
21	Scholarly publishing in the Internet age: a citation analysis of computer science literature. Information Processing and Management, 2001, 37, 661-675.	8.6	157
22	Goos-HÄnchen shifts from absorbing media. Physical Review A, 1982, 25, 2099-2101.	2.5	148
23	Learning to Extract Semantic Structure from Documents Using Multimodal Fully Convolutional Neural Networks. , 2017, , .		145
24	Dynamic recurrent neural networks: Theory and applications. IEEE Transactions on Neural Networks, 1994, 5, 153-156.	4.2	143
25	Context and page analysis for improved Web search. IEEE Internet Computing, 1998, 2, 38-46.	3.3	140
26	Constructing deterministic finite-state automata in recurrent neural networks. Journal of the ACM, 1996, 43, 937-972.	2.2	128
27	Machine Learning Using a Higher Order Correlation Network. Physica D: Nonlinear Phenomena, 1986, 22, 276-306.	2.8	126
28	Exploring social annotations for information retrieval. , 2008, , .		124
29	Hot streaks in artistic, cultural, and scientific careers. Nature, 2018, 559, 396-399.	27.8	123
30	Automatic tag recommendation algorithms for social recommender systems. ACM Transactions on the Web, 2011, 5, 1-31.	2.5	117
31	Nonconvex Online Support Vector Machines. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33, 368-381.	13.9	117
32	Learning to Read Irregular Text with Attention Mechanisms. , 2017, , .		116
33	Active learning for class imbalance problem. , 2007, , .		112
34	Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101, 17599-17604.	7.1	109
35	A delay damage model selection algorithm for NARX neural networks. IEEE Transactions on Signal Processing, 1997, 45, 2719-2730.	5.3	107
36	TableSeer. , 2007, , .		105

#	ARTICLE	IF	CITATIONS
37	Identifying, Indexing, and Ranking Chemical Formulae and Chemical Names in Digital Documents. ACM Transactions on Information Systems, 2011, 29, 1-38.	4.9	102
38	Disambiguating authors in academic publications using random forests. , 2009, , .		101
39	An analysis of noise in recurrent neural networks: convergence and generalization. IEEE Transactions on Neural Networks, 1996, 7, 1424-1438.	4.2	100
40	Natural language grammatical inference with recurrent neural networks. IEEE Transactions on Knowledge and Data Engineering, 2000, 12, 126-140.	5.7	98
41	Recommending citations. , 2012, , .		98
42	Efficient topic-based unsupervised name disambiguation. , 2007, , .		96
43	Learning multiple graphs for document recommendations. , 2008, , .		95
44	Bi-LSTM-CRF Sequence Labeling for Keyphrase Extraction from Scholarly Documents. , 2019, , .		95
45	<i>Screenomics</i>: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them. Human-Computer Interaction, 2021, 36, 150-201.	4.4	93
46	Accessibility of information on the Web. Intelligence, 2000, 11, 32-39.	0.2	91
47	Adversary Resistant Deep Neural Networks with an Application to Malware Detection. , 2017, , .		90
48	Pruning recurrent neural networks for improved generalization performance. IEEE Transactions on Neural Networks, 1994, 5, 848-851.	4.2	88
49	Efficient Name Disambiguation for Large-Scale Databases. Lecture Notes in Computer Science, 2006, , 536-544.	1.3	84
50	Learning Chaotic Attractors by Neural Networks. Neural Computation, 2000, 12, 2355-2383.	2.2	81
51	Web Search--Your Way. Communications of the ACM, 2001, 44, 97-102.	4.5	80
52	How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies. Neural Networks, 1998, 11, 861-868.	5.9	78
53	Collaboration over time. , 2008, , .		77
54	Citation recommendation without author supervision. , 2011, , .		77

#	ARTICLE	IF	CITATIONS
55	A system for automatic personalized tracking of scientific literature on the Web. , 1999, , .		76
56	Fuzzy finite-state automata can be deterministically encoded into recurrent neural networks. IEEE Transactions on Fuzzy Systems, 1998, 6, 76-89.	9.8	74
57	An LDA-based Community Structure Discovery Approach for Large-Scale Social Networks. , 2007, , .		73
58	Extraction, Insertion and Refinement of Symbolic Rules in Dynamically Driven Recurrent Neural Networks. Connection Science, 1993, 5, 307-337.	3.0	72
59	A Strategic Perspective on Search Engines: Thought Candies for Practitioners and Researchers. Journal of Interactive Marketing, 2009, 23, 49-60.	6.2	71
60	Measuring Prerequisite Relations Among Concepts. , 2015, , .		71
61	AlgorithmSeer: A System for Extracting and Searching for Algorithms in Scholarly Big Data. IEEE Transactions on Big Data, 2016, 2, 3-17.	6.1	69
62	Inquirus, the NECI meta search engine. Computer Networks, 1998, 30, 95-105.	1.0	68
63	What's the code?. , 2002, , .		67
64	Indexing and retrieval of scientific literature. , 1999, , .		66
65	First-order versus second-order single-layer recurrent neural networks. IEEE Transactions on Neural Networks, 1994, 5, 511-513.	4.2	65
66	Autonomous citation matching. , 1999, , .		64
67	Automatic identification of informative sections of Web pages. IEEE Transactions on Knowledge and Data Engineering, 2005, 17, 1233-1246.	5.7	64
68	Multi-Scale Multi-Task FCN for Semantic Page Segmentation and Table Detection. , 2017, , .		64
69	Topic evolution and social interactions. , 2006, , .		63
70	A hierarchical naive Bayes mixture model for name disambiguation in author citations. , 2005, , .		62
71	Constructive learning of recurrent neural networks: limitations of recurrent cascade correlation and a simple solution. IEEE Transactions on Neural Networks, 1995, 6, 829-836.	4.2	59
72	Automatic extraction of informative blocks from webpages. , 2005, , .		59

#	ARTICLE	IF	CITATIONS
73	Time-delay neural networks: representation and induction of finite-state machines. IEEE Transactions on Neural Networks, 1997, 8, 1065-1070.	4.2	56
74	Talking Helps: Evolving Communicating Agents for the Predator-Prey Pursuit Problem. Artificial Life, 2000, 6, 237-254.	1.3	56
75	Using Prerequisites to Extract Concept Maps from Textbooks. , 2016, , .		55
76	Are your citations clean?. Communications of the ACM, 2007, 50, 33-38.	4.5	54
77	CiteSeerX: AI in a Digital Library Search Engine. AI Magazine, 2015, 36, 35-48.	1.6	54
78	INFORMATION SCIENCE: Enhanced: Cybertools and Archaeology. Science, 2006, 311, 958-959.	12.6	53
79	Automatic tag recommendation for metadata annotation using probabilistic topic modeling. , 2013, , .		52
80	Rule revision with recurrent neural networks. IEEE Transactions on Knowledge and Data Engineering, 1996, 8, 183-188.	5.7	51
81	Discovering Temporal Communities from Social Network Documents. , 2007, , .		51
82	Multi-scale FCN with Cascaded Instance Aware Segmentation for Arbitrary Oriented Word Spotting in the Wild. , 2017, , .		51
83	Architecture of a metasearch engine that supports user information needs. , 1999, , .		50
84	Automated analysis of images in documents for intelligent document search. International Journal on Document Analysis and Recognition, 2009, 12, 65-81.	3.4	50
85	Learning a class of large finite state machines with a recurrent neural network. Neural Networks, 1995, 8, 1359-1365.	5.9	48
86	Detecting research topics via the correlation between graphs and texts. , 2007, , .		46
87	Can't see the forest for the trees?. , 2013, , .		44
88	CiteSeerx. , 2006, , .		43
89	Discovering missing links in networks using vertex similarity measures. , 2012, , .		42
90	Tangent-CFT. , 2019, , .		42

#	ARTICLE	IF	CITATIONS
91	Stable Encoding of Large Finite-State Automata in Recurrent Neural Networks with Sigmoid Discriminants. <i>Neural Computation</i> , 1996, 8, 675-696.	2.2	41
92	Exploring web scale language models for search query processing. , 2010, , .		39
93	EXPERIMENTAL COMPARISON OF THE EFFECT OF ORDER IN RECURRENT NEURAL NETWORKS. <i>International Journal of Pattern Recognition and Artificial Intelligence</i> , 1993, 07, 849-872.	1.2	38
94	Efficient record-level wrapper induction. , 2009, , .		37
95	CiteSense. , 2008, , .		36
96	Automatic Extraction of Data from Bar Charts. , 2015, , .		36
97	Distractor Generation for Multiple Choice Questions Using Learning to Rank. , 2018, , .		36
98	Automatic Detection of Pseudocodes in Scholarly Documents Using Machine Learning. , 2013, , .		35
99	Fresnel reflection and transmission at a planar boundary from media of equal refractive indices. <i>Applied Physics Letters</i> , 1982, 40, 210-212.	3.3	33
100	On the distribution of performance from multiple neural-network trials. <i>IEEE Transactions on Neural Networks</i> , 1997, 8, 1507-1517.	4.2	33
101	How are you feeling?: A personalized methodology for predicting mental states from temporally observable physical and behavioral information. <i>Journal of Biomedical Informatics</i> , 2017, 68, 1-19.	4.3	33
102	RefSeer: A citation recommendation system. , 2014, , .		32
103	Topic segmentation with shared topic detection and alignment of multiple documents. , 2007, , .		31
104	A sparse gaussian processes classification framework for fast tag suggestions. , 2008, , .		31
105	Identifying table boundaries in digital documents via sparse line detection. , 2008, , .		31
106	Measuring conference quality by mining program committee characteristics. , 2007, , .		30
107	AckSeer. , 2012, , .		30
108	Towards building a scholarly big data platform: Challenges, lessons and opportunities. , 2014, , .		30

#	ARTICLE	IF	CITATIONS
109	An Empirical Evaluation of Rule Extraction from Recurrent Neural Networks. <i>Neural Computation</i> , 2018, 30, 2568-2591.	2.2	30
110	Attractive Periodic Sets in Discrete-Time Recurrent Networks (with Emphasis on Fixed-Point Stability) <i>Tj ETQq0 0 0 ggBT /Overlock 10 Tf</i>	2.2	29
111	Designing for e-science: Requirements gathering for collaboration in CiteSeer. <i>International Journal of Human Computer Studies</i> , 2009, 67, 297-312.	5.6	29
112	Similar researcher search in academic environments. , 2012, , .		29
113	A generalized topic modeling approach for automatic document annotation. <i>International Journal on Digital Libraries</i> , 2015, 16, 111-128.	1.5	29
114	Concept Hierarchy Extraction from Textbooks. , 2015, , .		29
115	What's there and what's not?. , 2005, , .		28
116	Automatic extraction of table metadata from digital documents. , 2006, , .		28
117	Ranking authors in digital libraries. , 2011, , .		28
118	Figure Metadata Extraction from Digital Documents. , 2013, , .		28
119	Analysis of lexical signatures for improving information persistence on the World Wide Web. <i>ACM Transactions on Information Systems</i> , 2004, 22, 540-572.	4.9	26
120	Identifying Content Blocks from Web Documents. <i>Lecture Notes in Computer Science</i> , 2005, , 285-293.	1.3	25
121	A large-scale study of robots.txt. , 2007, , .		25
122	Cloud Computing: A Digital Libraries Perspective. , 2010, , .		25
123	Online Person Name Disambiguation with Constraints. , 2015, , .		25
124	Social Bookmarking for Scholarly Digital Libraries. <i>IEEE Internet Computing</i> , 2007, 11, 29-35.	3.3	24
125	Better Naive Bayes classification for high-precision spam detection. <i>Software - Practice and Experience</i> , 2009, 39, 1003-1024.	3.6	23
126	A hybrid approach to discover semantic hierarchical sections in scholarly documents. , 2015, , .		23

#	ARTICLE	IF	CITATIONS
127	Training Second-Order Recurrent Neural Networks using Hints. , 1992, , 361-366.		23
128	Extracting query modifications from nonlinear SVMs. , 2002, , .		22
129	Rule-based word clustering for document metadata extraction. , 2005, , .		22
130	Improving the Table Boundary Detection in PDFs by Fixing the Sequence Error of the Sparse Lines. , 2009, , .		22
131	Measuring the web crawler ethics. , 2010, , .		22
132	An algorithm search engine for software developers. , 2011, , .		22
133	Ranking experts using author-document-topic graphs. , 2013, , .		22
134	PDFMEF. , 2015, , .		22
135	Automatic categorization of figures in scientific documents. , 2006, , .		21
136	The Ethicality of Web Crawlers. , 2010, , .		21
137	Information incorporation in online in-Game sports betting markets. , 2003, , .		20
138	Collaborative Filtering with Maximum Entropy. IEEE Intelligent Systems, 2004, 19, 40-48.	4.0	20
139	Extraction and search of chemical formulae in text documents on the web. , 2007, , .		20
140	The evolution of a crawling strategy for an academic document search engine. , 2012, , .		20
141	CSSeer. , 2013, , .		20
142	Near duplicate detection in an academic digital library. , 2013, , .		20
143	Continual Learning of Recurrent Neural Networks by Locally Aligning Distributed Representations. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31, 4267-4278.	11.3	20
144	Finding topic trends in digital libraries. , 2009, , .		19

#	ARTICLE	IF	CITATIONS
145	Social network document ranking. , 2010, , .		19
146	Brewster angles for magnetic media. Journal of Infrared, Millimeter and Terahertz Waves, 1985, 6, 187-197.	0.6	18
147	Mining, indexing, and searching for textual chemical molecule information on the web. , 2008, , .		18
148	ASCOS++. ACM Transactions on Knowledge Discovery From Data, 2015, 10, 1-26.	3.5	18
149	Hybrid Deep Pairwise Classification for Author Name Disambiguation. , 2019, , .		18
150	SNDocRank. , 2010, , .		17
151	Can back-of-the-book indexes be automatically created?. , 2013, , .		17
152	ASCOS. , 2013, , .		17
153	Neural Networks And Hybrid Intelligent Models: Foundations, Theory, And Applications. IEEE Transactions on Neural Networks, 1998, 9, 721-723.	4.2	16
154	ChemXSeer. , 2007, , .		16
155	Determining Bias to Search Engines from Robots.txt. , 2007, , .		16
156	Improving algorithm search using the algorithm co-citation network. , 2012, , .		16
157	Table of contents recognition and extraction for heterogeneous book documents. , 2013, , .		16
158	Scholarly big data information extraction and integration in the CiteSeer<sup>χ</sup> digital library. , 2014, , .		16
159	Deep Learning Can Predict Laboratory Quakes From Active Source Seismic Data. Geophysical Research Letters, 2021, 48, e2021GL093187.	4.0	16
160	A service-oriented architecture for digital libraries. , 2004, , .		15
161	Efficient Multiclass Boosting Classification with Active Learning. , 2007, , .		15
162	Design and evaluation of awareness mechanisms in CiteSeer. Information Processing and Management, 2008, 44, 596-612.	8.6	15

#	ARTICLE	IF	CITATIONS
163	Large scale author name disambiguation in digital libraries. , 2014, , .		15
164	Chemical entity extraction using CRF and an ensemble of extractors. Journal of Cheminformatics, 2015, 7, S12.	6.1	15
165	Scalable algorithms for scholarly figure mining and semantics. , 2016, , .		15
166	A Web Service for Author Name Disambiguation in Scholarly Databases. , 2018, , .		15
167	Towards Better Understanding of Academic Search. , 2016, , .		15
168	How communication can improve the performance of multi-agent systems. , 2001, , .		14
169	Analysis of lexical signatures for finding lost or related documents. , 2002, , .		14
170	Boosting the Feature Space: Text Classification for Unstructured Data on the Web. IEEE International Conference on Data Mining, 2006, , .	0.0	14
171	Segregating and extracting overlapping data points in two-dimensional plots. , 2008, , .		14
172	A classification scheme for algorithm citation function in scholarly works. , 2013, , .		14
173	A Web Service for Scholarly Big Data Information Extraction. , 2014, , .		14
174	Distractor Generation with Generative Adversarial Nets for Automatically Creating Fill-in-the-blank Questions. , 2017, , .		14
175	A Comparison of On-Line Computer Science Citation Databases. Lecture Notes in Computer Science, 2005, , 438-449.	1.3	14
176	Automatic acknowledgement indexing. , 2005, , .		13
177	Supporting distributed scientific collaboration: Implications for designing the CiteSeer collaboratory. , 2007, , .		13
178	Personalized ranking for digital libraries based on log analysis. , 2008, , .		13
179	Modeling and visualizing geo-sensitive queries based on user clicks. , 2008, , .		13
180	Graph-based seed selection for web-scale crawlers. , 2009, , .		13

#	ARTICLE	IF	CITATIONS
181	MobiSNA. , 2009, , .		13
182	Locality and attachedness-based temporal social network growth dynamics analysis: A case study of evolving nanotechnology scientific collaboration networks. Journal of the Association for Information Science and Technology, 2010, 61, 964-977.	2.6	13
183	Finding algorithms in scientific articles. , 2010, , .		13
184	Researcher homepage classification using unlabeled data. , 2013, , .		13
185	Unifying Adversarial Training Algorithms with Data Gradient Regularization. Neural Computation, 2017, 29, 867-887.	2.2	13
186	Extracting Semantic Relations for Scholarly Knowledge Base Construction. , 2018, , .		13
187	Large Scale Subject Category Classification of Scholarly Papers With Deep Attentive Neural Networks. Frontiers in Research Metrics and Analytics, 2020, 5, 600382.	1.9	13
188	CiteSeerĭ. , 2006, , .		12
189	A metadata generation system for scanned scientific volumes. , 2008, , .		12
190	Smart Library: Identifying Books on Library Shelves Using Supervised Deep Learning for Scene Text Reading. , 2017, , .		12
191	Text Extraction from Smartphone Screenshots to Archive in situ Media Behavior. , 2017, , .		12
192	Clustering Scientific Literature Using Sparse Citation Graph Analysis. Lecture Notes in Computer Science, 2006, , 30-41.	1.3	12
193	Graph based crawler seed selection. , 2009, , .		12
194	Distributed error correction. , 1999, , .		11
195	oreChem ChemXSeer. , 2010, , .		11
196	Mining and indexing graphs for supergraph search. Proceedings of the VLDB Endowment, 2013, 6, 829-840.	3.8	11
197	Learned Neural Iterative Decoding for Lossy Image Compression Systems. , 2019, , .		11
198	Automatic Generation of Headlines for Online Math Questions. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34, 9490-9497.	4.9	11

#	ARTICLE	IF	CITATIONS
199	CiteSeerX. , 2019, , .		11
200	K-SVMMeans: A Hybrid Clustering Algorithm for Multi-Type Interrelated Datasets. , 2007, , .		10
201	On identifying academic homepages for digital libraries. , 2011, , .		10
202	Iterative Graph Feature Mining for Graph Indexing. , 2012, , .		10
203	Searching online book documents and analyzing book citations. , 2013, , .		10
204	Detecting Arbitrary Oriented Text in the Wild with a Visual Attention Model. , 2016, , .		10
205	Improving Offline Handwritten Chinese Character Recognition by Iterative Refinement. , 2017, , .		10
206	Adversarial Training for Community Question Answer Selection Based on Multi-Scale Matching. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33, 395-402.	4.9	10
207	BBookX. , 2015, , .		10
208	Rule-based word clustering for text classification. , 2003, , .		9
209	Learning metadata from the evidence in an on-line citation matching scheme. , 2006, , .		9
210	Generative models for name disambiguation. , 2007, , .		9
211	Extracting Author Meta-Data from Web Using Visual Features. , 2007, , .		9
212	Determining the sexual identities of prehistoric cave artists using digitized handprints. , 2010, , .		9
213	Updating Graph Indices with a One-Pass Algorithm. , 2015, , .		9
214	Inventor Name Disambiguation for a Patent Database Using a Random Forest and DBSCAN. , 2016, , .		9
215	Improving Researcher Homepage Classification with Unlabeled Data. ACM Transactions on the Web, 2015, 9, 1-32.	2.5	9
216	Farm-to-table: A Situation Awareness Model for Food Safety Assurance for Porous Borders. Comprehensive Reviews in Food Science and Food Safety, 2005, 4, 31-33.	11.7	8

#	ARTICLE	IF	CITATIONS
217	Deriving knowledge from figures for digital libraries. , 2007, , .		8
218	BotSeer: An Automated Information System for Analyzing Web Robots. , 2008, , .		8
219	SimSeerX. , 2014, , .		8
220	Migrating a Digital Library to a Private Cloud. , 2014, , .		8
221	Text extraction and retrieval from smartphone screenshots. , 2018, , .		8
222	Online Learning of Deep Hybrid Architectures for Semi-supervised Categorization. Lecture Notes in Computer Science, 2015, , 516-532.	1.3	8
223	Curve Separation for Line Graphs in Scholarly Documents. , 2016, , .		8
224	ROUTING IN RANDOM MULTISTAGE INTERCONNECTIONS NETWORKS: COMPARING EXHAUSTIVE SEARCH, GREEDY AND NEURAL NETWORK APPROACHES. International Journal of Neural Systems, 1992, 03, 125-142.	5.2	7
225	eBizSearch. , 2003, , .		7
226	An architecture for creating collaborative semantically capable scientific data sharing infrastructures. , 2006, , .		7
227	Automatic searching of tables in digital libraries. , 2007, , .		7
228	HSN-PAM: Finding Hierarchical Probabilistic Groups from Large-Scale Networks. , 2007, , .		7
229	Independent informative subgraph mining for graph information retrieval. , 2009, , .		7
230	Personalized Feed Recommendation Service for Social Networks. , 2010, , .		7
231	Scholarly big data. , 2013, , .		7
232	Automatic Knowledge Base Construction from Scholarly Documents. , 2017, , .		7
233	A Supervised Learning Approach To Entity Matching Between Scholarly Big Datasets. , 2017, , .		7
234	Follow The Curve: Arbitrarily Oriented Scene Text Detection Using Key Points Spotting And Curve Prediction. , 2020, , .		7

#	ARTICLE	IF	CITATIONS
235	Specialized Research Datasets in the CiteSeerX Digital Library. D-Lib Magazine, 2012, 18, .	0.5	7
236	ChartReader: Automatic Parsing of Bar-Plots. , 2021, , .		7
237	Using recurrent neural networks to learn the structure of interconnection networks. Neural Networks, 1995, 8, 793-804.	5.9	6
238	CiteSeer-API. , 2004, , .		6
239	Workload analysis for scientific literature digital libraries. International Journal on Digital Libraries, 2008, 9, 139-149.	1.5	6
240	Real-time data pre-processing technique for efficient feature extraction in large scale datasets. , 2008, , .		6
241	Financial Entity Record Linkage with Random Forests. , 2016, , .		6
242	HESDK: A Hybrid Approach to Extracting Scientific Domain Knowledge Entities. , 2017, , .		6
243	Cleaning Noisy and Heterogeneous Metadata for Record Linking across Scholarly Big Datasets. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33, 9601-9606.	4.9	6
244	The complexity of language recognition by neural networks. Neurocomputing, 1997, 15, 327-345.	5.9	5
245	Designing efficient sampling techniques to detect webpage updates. , 2007, , .		5
246	Web crawler middleware for search engine digital libraries. , 2012, , .		5
247	CiteSeerX-2018: A Cleansed Multidisciplinary Scholarly Big Dataset. , 2018, , .		5
248	Storybase: Towards Building a Knowledge Base for News Events. , 2015, , .		5
249	Designing clustering-based web crawling policies for search engine crawlers. , 2007, , .		4
250	A Fast Preprocessing Method for Table Boundary Detection: Narrowing Down the Sparse Lines Using Solely Coordinate Information. , 2008, , .		4
251	Measuring user preference changes in digital libraries. , 2008, , .		4
252	SNDocRank. , 2010, , .		4

#	ARTICLE	IF	CITATIONS
253	The predictive value of young and old links in a social network. , 2013, , .		4
254	Towards the Discovery of Diseases Related by Genes Using Vertex Similarity Measures. , 2013, , .		4
255	Extracting Researcher Metadata with Labeled Features. , 2014, , .		4
256	Classifying and ranking search engine results as potential sources of plagiarism. , 2014, , .		4
257	CiteSeerX data. , 2016, , .		4
258	Information Extraction for Scholarly Digital Libraries. , 2016, , .		4
259	On the Use of Similarity Search to Detect Fake Scientific Papers. Lecture Notes in Computer Science, 2015, , 332-338.	1.3	4
260	Shapley Homology: Topological Analysis of Sample Influence for Neural Networks. Neural Computation, 2020, 32, 1355-1378.	2.2	4
261	Keyphrase Extraction in Scholarly Digital Library Search Engines. Lecture Notes in Computer Science, 2020, , 179-196.	1.3	4
262	Name-Ethnicity Classification and Ethnicity-Sensitive Name Matching. Proceedings of the AAAI Conference on Artificial Intelligence, 2012, 26, 1141-1147.	4.9	4
263	Classification of source code archives. , 2003, , .		3
264	Modeling the author bias between two on-line computer science citation databases. , 2005, , .		3
265	Image annotation by hierarchical mapping of features. , 2007, , .		3
266	WebKDD/SNAKDD 2007. SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, 2007, 9, 87-92.	4.0	3
267	Entity resolution using search engine results. , 2012, , .		3
268	Scaling SeerSuite in the Cloud. , 2013, , .		3
269	Defending Against Adversarial Samples Without Security through Obscurity. , 2018, , .		3
270	TextContourNet: A Flexible and Effective Framework for Improving Scene Text Detection Architecture With a Multi-Task Cascade. , 2019, , .		3

#	ARTICLE	IF	CITATIONS
271	Effectively Searching Maps in Web Documents. Lecture Notes in Computer Science, 2009, , 162-176.	1.3	3
272	SNAKDD 2008 social network mining and analysis postworkshop report. SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, 2008, 10, 74-77.	4.0	3
273	COVIDSeer. , 2020, , .		3
274	Building an Accessible, Usable, Scalable, and Sustainable Service for Scholarly Big Data. , 2021, , .		3
275	What NARX networks can compute. Lecture Notes in Computer Science, 1995, , 95-102.	1.3	2
276	Alternative discrete-time operators: an algorithm for optimal selection of parameters. IEEE Transactions on Signal Processing, 1999, 47, 2612-2615.	5.3	2
277	A learning based model for headline extraction of news articles to find explanatory sentences for events. , 2005, , .		2
278	Estimating the web robot population. , 2010, , .		2
279	The feasibility of investing in manual correction of metadata for a large-scale digital library. , 2014, , .		2
280	"Building a search engine for algorithms" by Suppawong Tuarob, Prasenjit Mitra, and C. Lee Giles with Martin Vesely as coordinator. SIGWEB Newsletter: the Newsletter of ACM's Special Interest Group on Hypertext and Hypermedia, 2014, 2014, 1-9.	0.6	2
281	NMRexSeer: Metadata extraction and search for large scale Nuclear Magnetic Resonance (NMR) experimental data. , 2015, , .		2
282	Improving Similar Document Retrieval Using a Recursive Pseudo Relevance Feedback Strategy. , 2016, , .		2
283	A Neural State Pushdown Automata. IEEE Transactions on Artificial Intelligence, 2020, 1, 193-205.	4.7	2
284	Modeling Updates of Scholarly Webpages Using Archived Data. , 2020, , .		2
285	Using an education oriented digital library to organize and present classes in computing and information. SIGCSE Bulletin, 2003, , .	0.1	1
286	Next generation CiteSeer. , 2004, , .		1
287	Query Expansion Using Topic and Location. , 2007, , .		1
288	Formal definitions of web information search. Proceedings of the American Society for Information Science and Technology, 2007, 43, 1-1.	0.2	1

#	ARTICLE	IF	CITATIONS
289	A Non-parametric Approach to Pair-Wise Dynamic Topic Correlation Detection. , 2008, , .		1
290	Towards Click-Based Models of Geographic Interests in Web Search. , 2008, , .		1
291	Error-driven generalist+experts (edge). , 2008, , .		1
292	Learning to rank graphs for online similar graph search. , 2009, , .		1
293	Reports of the 2016 AAAI Workshop Program. AI Magazine, 2016, 37, 99-108.	1.6	1
294	BBookX. , 2016, , .		1
295	The Sibling Neural Estimator: Improving Iterative Image Decoding with Gradient Communication. , 2020, , .		1
296	Guest Editorial: Scholarly Big Data. IEEE Transactions on Emerging Topics in Computing, 2021, 9, 200-203.	4.6	1
297	Document Domain Randomization for Deep Learning Document Layout Extraction. Lecture Notes in Computer Science, 2021, , 497-513.	1.3	1
298	Routing in Random Multistage Interconnection Networks. , 1994, , 37-61.		1
299	Learning a Deep Hybrid Model for Semi-Supervised Text Classification. , 2015, , .		1
300	Searching the web (keynote address). , 1999, , .		0
301	Don't move. , 2003, , .		0
302	Learning term-relationships for ontology construction. , 2005, , .		0
303	A clustering method for web data with multi-type interrelated components. , 2007, , .		0
304	Intelligent Parsing of Scanned Volumes for Web Based Archives. , 2007, , .		0
305	Efficient user preference predictions using collaborative filtering. , 2008, , .		0
306	Metadata extraction and indexing for map search in web documents. , 2008, , .		0

#	ARTICLE	IF	CITATIONS
307	Finding a Haystack in Haystacks – Simultaneous Identification of Concepts in Large Bio-Medical Corpora. , 2008, , .		0
308	2013 international workshop on computational scientometrics. , 2013, , .		0
309	Crowd-sourcing Web knowledge for metadata extraction. , 2014, , .		0
310	Document Analysis and Retrieval Tasks in Scientific Digital Libraries. Communications in Computer and Information Science, 2015, , 3-20.	0.5	0
311	Automatically Generating a Concept Hierarchy with Graphs. , 2015, , .		0
312	Scholarly Big Data Knowledge and Semantics. , 2016, , .		0
313	Deep Learning, Grammar Transfer, and Transportation Theory. Lecture Notes in Computer Science, 2021, , 609-623.	1.3	0
314	An Entropy Metric for Regular Grammar Classification and Learning with Recurrent Neural Networks. Entropy, 2021, 23, 127.	2.2	0
315	Equivalence in Knowledge Representation. International Series on Computational Intelligence, 1999, , .	0.1	0
316	Large Scale Scene Text Verification with Guided Attention. Lecture Notes in Computer Science, 2019, , 260-275.	1.3	0