

Antoine Doucet

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/555380/publications.pdf>

Version: 2024-02-01

84
papers

992
citations

687363

13
h-index

752698

20
g-index

90
all docs

90
docs citations

90
times ranked

513
citing authors

#	ARTICLE	IF	CITATIONS
1	Building engagement for MOOC students. , 2014, , .		86
2	Survey of Post-OCR Processing Approaches. ACM Computing Surveys, 2022, 54, 1-37.	23.0	80
3	Improving Skin-Disease Classification Based on Customized Loss Function Combined With Balanced Mini-Batch Logic and Real-Time Image Augmentation. IEEE Access, 2020, 8, 150725-150737.	4.2	41
4	AI outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-CNN architecture with custom mini-batch logic and loss function. Scientific Reports, 2021, 11, 17485.	3.3	36
5	Deep Statistical Analysis of OCR Errors for Effective Post-OCR Processing. , 2019, , .		33
6	ICDAR 2019 Competition on Post-OCR Text Correction. , 2019, , .		29
7	A Comparative Study for Classification of Skin Cancer. , 2019, , .		28
8	Multilingual event extraction for epidemic detection. Artificial Intelligence in Medicine, 2015, 65, 131-143.	6.5	27
9	Impact of OCR Errors on the Use of Digital Libraries: Towards a Better Access to Information. , 2017, , .		27
10	ICDAR2017 Competition on Post-OCR Text Correction. , 2017, , .		22
11	Overview of the INEX 2007 Book Search track. ACM SIGIR Forum, 2008, 42, 2-15.	0.5	22
12	Setting up a competition framework for the evaluation of structure extraction from OCR-ed books. International Journal on Document Analysis and Recognition, 2011, 14, 45-52.	3.4	20
13	An Analysis of the Performance of Named Entity Recognition over OCRed Documents. , 2019, , .		20
14	Neural Machine Translation with BERT for Post-OCR Error Detection and Correction. , 2020, , .		20
15	Document summarization based on word associations. , 2014, , .		19
16	Evaluating the Impact of OCR Errors on Topic Modeling. Lecture Notes in Computer Science, 2018, , 3-14.	1.3	17
17	Find it! Fraud Detection Contest Report. , 2018, , .		17
18	Adaptive Edit-Distance and Regression Approach for Post-OCR Text Correction. Lecture Notes in Computer Science, 2018, , 278-289.	1.3	17

#	ARTICLE	IF	CITATIONS
19	Alleviating Digitization Errors in Named Entity Recognition for Historical Documents. , 2020, , .		17
20	Non-contiguous word sequences for information retrieval. , 2004, , .		16
21	Dating Color Images with Ordinal Classification. , 2014, , .		15
22	A lightweight and multilingual framework for crisis information extraction from Twitter data. Social Network Analysis and Mining, 2019, 9, 1.	2.8	15
23	A Multilingual Dataset for Named Entity Recognition, Entity Linking and Stance Detection in Historical Newspapers. , 2021, , .		14
24	Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians. Journal of the Association for Information Science and Technology, 2022, 73, 225-239.	2.9	14
25	Overview of INEX 2013. Lecture Notes in Computer Science, 2013, , 269-281.	1.3	14
26	ICDAR 2009 Book Structure Extraction Competition. , 2009, , .		13
27	Impact of OCR Quality on Named Entity Linking. Lecture Notes in Computer Science, 2019, , 102-115.	1.3	13
28	Overview of the INEX 2008 Book Track. Lecture Notes in Computer Science, 2009, , 106-123.	1.3	13
29	Overview of the INEX 2011 Books and Social Search Track. Lecture Notes in Computer Science, 2012, , 1-29.	1.3	13
30	ICDAR 2011 Book Structure Extraction Competition. , 2011, , .		12
31	ICDAR 2013 Competition on Book Structure Extraction. , 2013, , .		11
32	Computational generation and dissection of lexical replacement humor. Natural Language Engineering, 2016, 22, 727-749.	2.5	11
33	Deep multimodal learning for cross-modal retrieval: One model for all tasks. Pattern Recognition Letters, 2021, 146, 38-45.	4.2	11
34	Overview of the INEX 2010 Book Track: Scaling Up the Evaluation Using Crowdsourcing. Lecture Notes in Computer Science, 2011, , 98-117.	1.3	11
35	Automatic discovery of word semantic relations using paraphrase alignment and distributional lexical semantics analysis. Natural Language Engineering, 2010, 16, 439-467.	2.5	10
36	Detecting prominent microblog users over crisis events phases. Information Systems, 2018, 78, 173-188.	3.6	10

#	ARTICLE	IF	CITATIONS
37	Multilingual Epidemiological Text Classification: A Comparative Study. , 2020, , .		10
38	When to Use OCR Post-correction for Named Entity Recognition?. Lecture Notes in Computer Science, 2020, , 33-42.	1.3	9
39	Advanced document description, a sequential approach. ACM SIGIR Forum, 2006, 40, 71-72.	0.5	9
40	Overview of the INEX 2009 Book Track. Lecture Notes in Computer Science, 2010, , 145-159.	1.3	8
41	Any Language Early Detection of Epidemic Diseases from Web News Streams. , 2013, , .		7
42	Information Extraction from Invoices. Lecture Notes in Computer Science, 2021, , 699-714.	1.3	7
43	Applying Semantic Web Technologies for Improving the Visibility of Tourism Data. , 2015, , .		6
44	Post-OCR Error Detection by Generating Plausible Candidates. , 2019, , .		6
45	Accessing and Investigating Large Collections of Historical Newspapers with the NewsEye Platform. , 2020, , .		6
46	An efficient any language approach for the integration of phrases in document retrieval. Language Resources and Evaluation, 2010, 44, 159-180.	2.7	5
47	Language-independent multi-document text summarization with document-specific word associations. , 2016, , .		5
48	Feature Selection for Document Flow Segmentation. , 2018, , .		5
49	Event Detection with Entity Markers. Lecture Notes in Computer Science, 2021, , 233-240.	1.3	5
50	Added-Value of Automatic Multilingual Text Analysis for Epidemic Surveillance. Lecture Notes in Computer Science, 2013, , 284-294.	1.3	5
51	Named Entity Recognition Architecture Combining Contextual and Global Features. Lecture Notes in Computer Science, 2021, , 264-276.	1.3	5
52	In-depth analysis of the impact of OCR errors on named entity recognition and linking. Natural Language Engineering, 2023, 29, 425-448.	2.5	5
53	Assessing the impact of OCR noise on multilingual event detection over digitised documents. International Journal on Digital Libraries, 2022, 23, 241-266.	1.5	5
54	Every Word has its History. , 2018, , .		4

#	ARTICLE	IF	CITATIONS
55	DataTourism: Designing an Architecture to Process Tourism Data. , 2016, , 751-763.		4
56	Enhancing Keyword Search with a Keyphrase Index. Lecture Notes in Computer Science, 2009, , 65-70.	1.3	4
57	DAnIEL: Language Independent Character-Based News Surveillance. Lecture Notes in Computer Science, 2012, , 64-75.	1.3	4
58	Linking Named Entities across Languages using Multilingual Word Embeddings. , 2020, , .		4
59	Entity Linking for Historical Documents: Challenges and Solutions. Lecture Notes in Computer Science, 2020, , 215-231.	1.3	4
60	Evaluating the Robustness of Embedding-Based Topic Models to OCR Noise. Lecture Notes in Computer Science, 2021, , 392-400.	1.3	4
61	Introducing the HIPE 2022 Shared Task: Named Entity Recognition and Linking in Multilingual Historical Documents. Lecture Notes in Computer Science, 2022, , 347-354.	1.3	4
62	SentiML++: An Extension of the SentiML Sentiment Annotation Scheme. Lecture Notes in Computer Science, 2015, , 91-96.	1.3	3
63	Enhancing Table of Contents Extraction by System Aggregation. , 2017, , .		3
64	Large Scale Analysis of Semantic and Temporal Aspects in Cultural Heritage Collection's Search. , 2019, , .		3
65	Identification of Microblogs Prominent Users during Events by Learning Temporal Sequences of Features. , 2015, , .		3
66	Document in Context of its Time (DICT). , 2019, , .		3
67	MELHISSA: a multilingual entity linking architecture for historical press articles. International Journal on Digital Libraries, 2022, 23, 133-160.	1.5	3
68	Novel Query Suggestions. , 2014, , .		2
69	Unsupervised Crisis Information Extraction from Twitter Data. , 2018, , .		2
70	Semantic Text Recognition via Visual Question Answering. , 2019, , .		2
71	A Comprehensive Extraction of Relevant Real-World-Event Qualifiers for Semantic Search Engines. Lecture Notes in Computer Science, 2021, , 153-164.	1.3	2
72	Token-Level Multilingual Epidemic Dataset for Event Extraction. Lecture Notes in Computer Science, 2021, , 55-59.	1.3	2

#	ARTICLE	IF	CITATIONS
73	Overview of the INEX 2007 Book Search Track (BookSearchâ€™07). Lecture Notes in Computer Science, 2007, , 148-161.	1.3	2
74	XML-aided phrase indexing for hypertext documents. , 2008, , .		1
75	Temporal Reconciliation for Dating Photographs Using Entity Information. , 2015, , .		1
76	Exploiting Social Annotations to Generate Resource Descriptions in a Distributed Environment: Cooperative Multi-Agent Simulation on Query-Based Sampling. The Review of Socionetwork Strategies, 2017, 11, 83-93.	1.5	1
77	Determining image age with rank-consistent ordinal classification and object-centered ensemble. , 2021, , .		1
78	Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR'15). , 2015, , .		1
79	Report on the Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR) Tj ETQq1 1 0.784314 rgBT /Overbo 0.5 1		1
80	EXTIRP: Baseline Retrieval from Wikipedia. Lecture Notes in Computer Science, 2006, , 115-120.	1.3	1
81	A Proposal for a Multilingual Epidemic Surveillance System. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2010, , 343-348.	0.3	0
82	An Extended Evaluation of the Impact of Different Modules in ST-VQA Systems. Lecture Notes in Computer Science, 2020, , 562-574.	1.3	0
83	New Tasks on Collections of Digitized Books. Lecture Notes in Computer Science, 2008, , 410-412.	1.3	0
84	Multilingual Epidemic Event Extraction. Lecture Notes in Computer Science, 2021, , 139-156.	1.3	0