Antoine Doucet

List of Publications by Year
in descending order

84	992	777949 13	843174 20
papers	citations	h-index	g-index

90	90	90	564
all docs	docs citations	times ranked	citing authors

#	Article	IF	Citations
1	In-depth analysis of the impact of OCR errors on named entity recognition and linking. Natural Language Engineering, 2023, 29, 425-448.	2.1	5
2	Survey of Post-OCR Processing Approaches. ACM Computing Surveys, 2022, 54, 1-37.	16.1	80
3	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.	1.5	14
4	MELHISSA: a multilingual entity linking architecture for historical press articles. International Journal on Digital Libraries, 2022, 23, 133-160.	1.1	3
5	Introducing the HIPE 2022 Shared Task: Named Entity Recognition and Linking in Multilingual Historical Documents. Lecture Notes in Computer Science, 2022, , 347-354.	1.0	4
6	Assessing the impact of OCR noise on multilingual event detection over digitised documents. International Journal on Digital Libraries, 2022, 23, 241-266.	1.1	5
7	A Comprehensive Extraction of Relevant Real-World-Event Qualifiers for Semantic Search Engines. Lecture Notes in Computer Science, 2021, , 153-164.	1.0	2
8	Token-Level Multilingual Epidemic Dataset for Event Extraction. Lecture Notes in Computer Science, 2021, , 55-59.	1.0	2
9	Information Extraction from Invoices. Lecture Notes in Computer Science, 2021, , 699-714.	1.0	7
10	Determining image age with rank-consistent ordinal classification and object-centered ensemble. , 2021, , .		1
11	Deep multimodal learning for cross-modal retrieval: One model for all tasks. Pattern Recognition Letters, 2021, 146, 38-45.	2.6	11
12	A Multilingual Dataset for Named Entity Recognition, Entity Linking and Stance Detection in Historical Newspapers. , 2021, , .		14
13	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.	1.6	36
14	Event Detection with Entity Markers. Lecture Notes in Computer Science, 2021, , 233-240.	1.0	5
15	Evaluating the Robustness of Embedding-Based Topic Models to OCR Noise. Lecture Notes in Computer Science, 2021, , 392-400.	1.0	4
16	Named Entity Recognition Architecture Combining Contextual and Global Features. Lecture Notes in Computer Science, 2021, , 264-276.	1.0	5
17	Multilingual Epidemic Event Extraction. Lecture Notes in Computer Science, 2021, , 139-156.	1.0	0
18	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.	2.6	41

#	Article	IF	Citations
19	When to Use OCR Post-correction for Named Entity Recognition?. Lecture Notes in Computer Science, 2020, , 33-42.	1.0	9
20	Neural Machine Translation with BERT for Post-OCR Error Detection and Correction. , 2020, , .		20
21	Accessing and Investigating Large Collections of Historical Newspapers with the NewsEye Platform. , 2020, , .		6
22	Alleviating Digitization Errors in Named Entity Recognition for Historical Documents. , 2020, , .		17
23	Linking Named Entities across Languages using Multilingual Word Embeddings. , 2020, , .		4
24	Entity Linking for Historical Documents: Challenges and Solutions. Lecture Notes in Computer Science, 2020, , 215-231.	1.0	4
25	Multilingual Epidemiological Text Classification: A Comparative Study. , 2020, , .		10
26	An Extended Evaluation of the Impact of Different Modules in ST-VQA Systems. Lecture Notes in Computer Science, 2020, , 562-574.	1.0	0
27	Deep Statistical Analysis of OCR Errors for Effective Post-OCR Processing. , 2019, , .		33
28	Large Scale Analysis of Semantic and Temporal Aspects in Cultural Heritage Collection's Search. , 2019, , .		3
29	A lightweight and multilingual framework for crisis information extraction from Twitter data. Social Network Analysis and Mining, 2019, 9, 1.	1.9	15
30	An Analysis of the Performance of Named Entity Recognition over OCRed Documents. , 2019, , .		20
31	A Comparative Study for Classification of Skin Cancer. , 2019, , .		28
32	ICDAR 2019 Competition on Post-OCR Text Correction. , 2019, , .		29
33	Post-OCR Error Detection by Generating Plausible Candidates. , 2019, , .		6
34	Semantic Text Recognition via Visual Question Answering. , 2019, , .		2
35	Impact of OCR Quality on Named Entity Linking. Lecture Notes in Computer Science, 2019, , 102-115.	1.0	13
36	Document in Context of its Time (DICT). , 2019, , .		3

#	Article	IF	Citations
37	Detecting prominent microblog users over crisis events phases. Information Systems, 2018, 78, 173-188.	2.4	10
38	Every Word has its History. , 2018, , .		4
39	Evaluating the Impact of OCR Errors on Topic Modeling. Lecture Notes in Computer Science, 2018, , 3-14.	1.0	17
40	Unsupervised Crisis Information Extraction from Twitter Data. , 2018, , .		2
41	Find it! Fraud Detection Contest Report. , 2018, , .		17
42	Feature Selection for Document Flow Segmentation. , 2018, , .		5
43	Adaptive Edit-Distance and Regression Approach for Post-OCR Text Correction. Lecture Notes in Computer Science, 2018, , 278-289.	1.0	17
44	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.	1.0	1
45	Impact of OCR Errors on the Use of Digital Libraries: Towards a Better Access to Information. , 2017, , .		27
46	ICDAR2017 Competition on Post-OCR Text Correction. , 2017, , .		22
47	Enhancing Table of Contents Extraction by System Aggregation. , 2017, , .		3
48	Language-independent multi-document text summarization with document-specific word associations. , 2016, , .		5
49	Computational generation and dissection of lexical replacement humor. Natural Language Engineering, 2016, 22, 727-749.	2.1	11
50	DataTourism: Designing an Architecture to Process Tourism Data. , 2016, , 751-763.		4
51	Report on the Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR) Tj ETQq1 1 0.784314 rgBT /Over	0.4	1
52	Applying Semantic Web Technologies for Improving the Visibility of Tourism Data. , 2015, , .		6
53	Temporal Reconciliation for Dating Photographs Using Entity Information. , 2015, , .		1
54	Multilingual event extraction for epidemic detection. Artificial Intelligence in Medicine, 2015, 65, 131-143.	3.8	27

#	Article	IF	Citations
55	SentiML++: An Extension of the SentiML Sentiment Annotation Scheme. Lecture Notes in Computer Science, 2015, , 91-96.	1.0	3
56	Identification of Microblogs Prominent Users during Events by Learning Temporal Sequences of Features. , 2015, , .		3
57	Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR'15). , 2015, , .		1
58	Building engagement for MOOC students. , 2014, , .		86
59	Dating Color Images with Ordinal Classification. , 2014, , .		15
60	Novel Query Suggestions. , 2014, , .		2
61	Document summarization based on word associations. , 2014, , .		19
62	Any Language Early Detection of Epidemic Diseases from Web News Streams. , 2013, , .		7
63	ICDAR 2013 Competition on Book Structure Extraction. , 2013, , .		11
64	Added-Value of Automatic Multilingual Text Analysis for Epidemic Surveillance. Lecture Notes in Computer Science, 2013, , 284-294.	1.0	5
65	Overview of INEX 2013. Lecture Notes in Computer Science, 2013, , 269-281.	1.0	14
66	Overview of the INEX 2011 Books and Social Search Track. Lecture Notes in Computer Science, 2012, , 1-29.	1.0	13
67	DAnIEL: Language Independent Character-Based News Surveillance. Lecture Notes in Computer Science, 2012, , 64-75.	1.0	4
68	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.	2.7	20
69	ICDAR 2011 Book Structure Extraction Competition. , 2011, , .		12
70	Overview of the INEX 2010 Book Track: Scaling Up the Evaluation Using Crowdsourcing. Lecture Notes in Computer Science, 2011, , 98-117.	1.0	11
71	An efficient any language approach for the integration of phrases in document retrieval. Language Resources and Evaluation, 2010, 44, 159-180.	1.8	5
72	Automatic discovery of word semantic relations using paraphrase alignment and distributional lexical semantics analysis. Natural Language Engineering, 2010, 16, 439-467.	2.1	10

#	Article	IF	Citations
73	Overview of the INEX 2009 Book Track. Lecture Notes in Computer Science, 2010, , 145-159.	1.0	8
74	A Proposal for a Multilingual Epidemic Surveillance System. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2010, , 343-348.	0.2	0
75	ICDAR 2009 Book Structure Extraction Competition. , 2009, , .		13
76	Overview of the INEX 2008 Book Track. Lecture Notes in Computer Science, 2009, , 106-123.	1.0	13
77	Enhancing Keyword Search with a Keyphrase Index. Lecture Notes in Computer Science, 2009, , 65-70.	1.0	4
78	XML-aided phrase indexing for hypertext documents. , 2008, , .		1
79	Overview of the INEX 2007 Book Search track. ACM SIGIR Forum, 2008, 42, 2-15.	0.4	22
80	New Tasks on Collections of Digitized Books. Lecture Notes in Computer Science, 2008, , 410-412.	1.0	0
81	Overview of the INEX 2007 Book Search Track (BookSearchâ€™07). Lecture Notes in Computer Science, 2007, , 148-161.	1.0	2
82	Advanced document description, a sequential approach. ACM SIGIR Forum, 2006, 40, 71-72.	0.4	9
83	EXTIRP: Baseline Retrieval from Wikipedia. Lecture Notes in Computer Science, 2006, , 115-120.	1.0	1
84	Non-contiguous word sequences for information retrieval. , 2004, , .		16