Citation Report
List of articles citing

Automatic Text Categorization in Terms of Genre and Author

DOI: 10.1162/089120100750105920
Computational Linguistics, 2000, 26, 471-495.

Source: https://exaly.com/paper-pdf/31935747/citation-report.pdf
Version: 2024-04-10

This report has been generated based on the citations recorded by exaly.com for the above article. For the latest version of this publication list, visit the link given above.
The third column is the impact factor (IF) of the journal, and the fourth column is the number of citations of the article.

#	Paper	IF	Citations
219	Combining Naive Bayes and n-Gram Language Models for Text Classification. *Lecture Notes in Computer Science*, **2003**, 335-350	0.9	54
218	Advances in Knowledge Discovery and Data Mining. *Lecture Notes in Computer Science*, **2003**,	0.9	44
217	The WATCHER Project: Building an Agent for Automatic Extraction of Language Resources from the Internet. **2003**, 18, 449-464		
216	Style mining of electronic messages for multiple authorship discrimination. **2003**,		47
215	Artificial Intelligence and Machine Translation in the XXIst Century*. **2003**, 10, 215-238		1
214	Language and task independent text categorization with simple language models. **2003**,		25
213	Music artist style identification by semi-supervised learning from both lyrics and content. **2004**,		12
212	On combining multiple clusterings. **2004**,		12
211	. **2004**,		20
210	Methods and Applications of Artificial Intelligence. *Lecture Notes in Computer Science*, **2004**,	0.9	0
209	Linguistic correlates of style. **2004**,		67
208	Co-trained support vector machines for large scale unstructured document classification using unlabeled data and syntactic information. **2004**, 40, 421-439		23
207	Augmenting Naive Bayes Classifiers with Statistical Language Models. **2004**, 7, 317-345		122
206	A quantitative analysis of lexical differences between genders in telephone conversations. **2005**,		13
205	Multiple sets of features for automatic genre classification of web documents. **2005**, 41, 1263-1276		64
204	Segmenting documents by stylistic character. **2005**, 11, 397-415		56
203	A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text. **2005**, 21, 259-274		43

202	AN APPLICATION OF NEURAL NETWORKS TO SEQUENCE ANALYSIS AND GENRE IDENTIFICATION. **2005**, 19, 199-215		
201	Automatic Genre Detection of Web Documents. *Lecture Notes in Computer Science*, **2005**, 310-319	0.9	5
200	Natural Language Processing IJCNLP 2004. *Lecture Notes in Computer Science*, **2005**,	0.9	1
199	Automatic source attribution of text: a neural networks approach.		
198	N-Gram Feature Selection for Authorship Identification. *Lecture Notes in Computer Science*, **2006**, 77-86	0.9	62
197	Integrating Features from Different Sources for Music Information Retrieval. **2006**,		5
196	Flexible length phrases in document classification. **2006**,		
195	. **2006**, 8, 564-574		86
194	Classification of Text, Automatic. **2006**, 457-462		20
193	Bibliography. 335-388		
192	Effective identification of source code authors using byte-level information. **2006**,		47
191	Towards practical genre classification of web documents. **2006**,		3
190	From fingerprint to writeprint. **2006**, 49, 76-82		88
189	Extracting key-substring-group features for text classification. **2006**,		20
188	Toward Unification of Source Attribution Processes and Techniques. **2006**,		1
187	AUTHORSHIP ATTRIBUTION BASED ON FEATURE SET SUBSPACING ENSEMBLES. **2006**, 15, 823-838		37
186	Automatic Turkish Text Categorization in Terms of Author, Genre and Gender. *Lecture Notes in Computer Science*, **2006**, 221-226	0.9	26
185	. **2007**,		1

184	Bigrams of Syntactic Labels for Authorship Discrimination of Short Texts. **2007**, 22, 405-417		82
183	E-business and Telecommunication Networks. **2007**,		
182	Stylistic text classification using functional lexical features. **2007**, 58, 802-822		77
181	Language use reflects scientific methodology: A corpus-based study of peer-reviewed journal articles. **2008**, 75, 203-238		17
180	Examining the significance of high-level programming features in source code author classification. **2008**, 81, 447-460		28
179	Towards High-Quality Next-Generation Text-to-Speech Synthesis: A Multidomain Approach by Automatic Domain Classification. **2008**, 16, 1340-1354		16
178	Author identification: Using text sampling to handle the class imbalance problem. **2008**, 44, 790-799		76
177	Chat mining: Predicting user and message attributes in computer-mediated communication. **2008**, 44, 1448-1466		48
176	Web page genre classification. **2008**,		16
175	Artificial Intelligence: Theories, Models and Applications. *Lecture Notes in Computer Science*, **2008**,	0.9	1
174	An Examination of Genre Attributes for Web Page Classification. **2008**,		13
173	Genre identification of Chinese finance text using machine learning method. **2008**,		
172	Writeprints. **2008**, 26, 1-29		210
171	Genre and author detection in Turkish texts using artificial immune recognition systems. **2008**,		3
170	A Platform Framework for Cross-Lingual Text Relatedness Evaluation and Plagiarism Detection. **2008**,		3
169	MEL-IRIS: An Online Tool for Audio Analysis and Music Indexing. **2009**, 2009, 1-15		1
168	Study on feature selection in finance text categorization. **2009**,		1
167	Automatic Genre Classification by Using Co-training. **2009**,		

166	Computational methods in authorship attribution. **2009**, 60, 9-26	297
165	A survey of modern authorship attribution methods. **2009**, 60, 538-556	669
164	A Genre-Aware Approach to Focused Crawling. **2009**, 12, 285-319	9
163	Cognitive archaeology: Uses, methods, and results. **2009**, 22, 250-265	16
162	On combining multiple clusterings: an overview and a new perspective. **2010**, 33, 207-219	21
161	Mining writeprints from anonymous e-mails for forensic investigation. **2010**, 7, 56-64	78
160	Inferring gender of movie reviewers. **2010**,	57
159	Change of Word Characteristics in 20th-Century Turkish Literature: A Statistical Analysis. **2010**, 17, 167-190	6
158	Application of Cascading Rough Set-Based Classifiers on Authorship Attribution. **2010**,	2
157	NATURAL SPEAKING AND HOW TO ASSESS IT. **2010**, 14, 120	
156	Variable Length Character N-Gram Approach for Online Writeprint Identification. **2010**,	8
155	Demographic information classification exploiting spoken language. **2011**,	
154	Multilayered feedforward neural networks as a tool for distinction of the authors of texts. **2011**,	1
153	Detecting contrast patterns in newspaper articles by combining discourse analysis and text mining. **2011**, 21, 647-683	14
152	Machine Learning for Author Affiliation within Web Forums -- Using Statistical Techniques on NLP Features for Online Group Identification. **2011**,	3
151	Authorship Attribution Based on Specific Vocabulary. **2012**, 30, 1-30	42
150	Distinguishing venues by writing styles. **2012**,	4
149	Venue Recommendation: Submitting Your Paper with Style. **2012**,	24

148	Detecting authorship deception: a supervised machine learning approach using author writeprints. **2012**, 27, 183-196		35
147	Representation models for text classification. **2012**,		12
146	Implicit Group Membership Detection in Online Text: Analysis and Applications. *Lecture Notes in Computer Science*, **2012**, 222-230	0.9	1
145	Cross-Genre Authorship Verification Using Unmasking. **2012**, 93, 340-356		26
144	Writing with Style: Venue Classification. **2012**,		2
143	Online writeprint identification via Multi-PRM. **2012**,		
142	Author recognition by Abstract Feature Extraction. **2012**,		0
141	Plag-Inn: Intrinsic Plagiarism Detection Using Grammar Trees. *Lecture Notes in Computer Science*, **2012**, 284-289	0.9	1
140	Content-based mobile spam classification using stylistically motivated features. **2012**, 33, 364-369		21
139	Vocabulary Richness Measure in Genres. **2013**, 20, 339-349		29
138	Gender, writing and ranking in review forums: a case study of the IMDb. **2013**, 35, 645-664		13
137	. **2013**,		1
136	Recognition of word collocation habits using frequency rank ratio and inter-term intimacy. **2013**, 40, 4301-4314		6
135	Countering Plagiarism by Exposing Irregularities in Authors' Grammar. **2013**,		
134	Closing the loop: Assisting archival appraisal and information retrieval in one sweep. **2013**, 50, 1-10		
133	Mathematical Modelling of the Pattern of Occurrence of Words in Different Corpora of the Hindi Language*. **2013**, 20, 1-12		6
132	Distant Listening to Gertrude Stein's 'Melanctha': Using Similarity Analysis in a Discovery Paradigm to Analyze Prosody and Author Influence. **2013**, 28, 582-602		2
131	A Study on Chinese Quantitative Stylistic Features and Relation Among Different Styles Based on Text Clustering* Address correspondence to: Hou Renkui, Laboratory of Computational Linguistics, School of Humanities, Tsinghua University, Beijing 100084, China. E-mail: hourk0917@163.comView all notes. **2014**, 21, 246-280		10

130	A theoretical framework to identify authentic online reviews. **2014**, 38, 634-649		24
129	Text style analysis using trace ratio criterion patch alignment embedding. **2014**, 136, 201-212		1
128	Naïve Bayes classifiers for authorship attribution of Arabic texts. **2014**, 26, 473-484		25
127	Mining language variation using word using and collocation characteristics. **2014**, 41, 7805-7819		3
126	Computational Constancy Measures of Texts—Yule's K and Rényi's Entropy. *Computational Linguistics*, **2015**, 41, 481-502	2.8	19
125	Review Authorship Attribution in a Similarity Space. **2015**, 30, 200-213		5
124	Learning to classify short text from scientific documents using topic models with various types of knowledge. **2015**, 42, 1684-1698		42
123	A Comparison of Text Categorization Methods. **2016**, 5, 31-44		1
122	Tekstgenres analyseren op lexicale complexiteit met T-Scan. **2016**, 38, 263-304		1
121	A Time Series Model of the Writing Process. *Lecture Notes in Computer Science*, **2016**, 128-142	0.9	1
120	. **2016**,		1
119	Stylometric model for detecting oath expressions: A case study for Quranic texts. **2016**, 31, 1-20		5
118	Universality of Stylistic Traits in Texts. **2016**, 143-155		1
117	An Application of a Profile-Based Method for Authorship Verification: Investigating the Authenticity of Pliny the Younger's Letter to Trajan Concerning the Christians. **2016**, fqw001		1
116	A feature selection method for author identification in interactive communications based on supervised learning and language typicality. **2016**, 56, 175-184		9
115	Music genre classification from Turkish lyrics. **2016**,		2
114	Overview of PAN'16. *Lecture Notes in Computer Science*, **2016**, 332-350	0.9	12
113	Classifying True and False Hebrew Stories Using Word N-Grams. **2016**, 47, 629-649		2

112	Authenticating the writings of Julius Caesar. **2016**, 63, 86-96	46
111	Literary writing style recognition via a minimal spanning tree-based approach. **2016**, 61, 145-153	8
110	A Unified Approach to Authorship Attribution and Verification. **2016**, 70, 232-242	3
109	Analysis on Chinese quantitative stylistic features based on text mining. **2016**, 31, 357-367	3
108	Graph vs. bag representation models for the topic classification of web documents. **2016**, 19, 887-920	10
107	Multilingual author profiling on Facebook. **2017**, 53, 886-904	28
106	Information Retrieval and Text Mining Technologies for Chemistry. **2017**, 117, 7673-7761	124
105	Syntactic methods for topic-independent authorship attribution. **2017**, 23, 789-806	1
104	Language for Specific Purposes and Corpus-based Pedagogy. **2017**, 248-264	7
103	Effects of various preprocessing techniques to Turkish text categorization using n-gram features. **2017**,	5
102	Genre Identification Based on SFL Principles: The Representation of Text Types and Genres in English Language Teaching Material. **2017**, 1, 373-392	3
101	Sociolinguistic Features for Author Gender Identification: From Qualitative Evidence to Quantitative Analysis. **2017**, 24, 65-84	5
100	Authorship Attribution for Social Media Forensics. **2017**, 12, 5-33	92
99	Document embedding approach for efficient authorship attribution. **2017**,	1
98	A stylometric analysis on Bengali literature for authorship attribution. **2017**,	8
97	Moving Away from the Implicit Teaching of Genres in the L2 Classroom. **2018**, 2, 351-373	2
96	On De-anonymization of Single Tweet Messages. **2018**,	1
95	Do all roads lead to Rome?: Modeling register variation with factor analysis and discriminant analysis. **2018**, 14, 233-273	51

94	An ensemble scheme based on language function analysis and feature engineering for text genre classification. **2018**, 44, 28-47	105
93	A Topic Drift Model for authorship attribution. **2018**, 273, 133-140	4
92	On the role of syntactic dependencies and discourse relations for author and gender identification. **2018**, 105, 87-95	7
91	Masking topic-related information to enhance authorship attribution. **2018**, 69, 461-473	12
90	Automatic Gender and Identity Recognition in Annotated Multimodal Face-to-face Conversations. **2018**,	1
89	Word2vec and Clustering based Twitter Sentiment Analysis. **2018**,	2
88	Text-based Authorship Identification - A survey. **2018**,	3
87	An Effective and Scalable Framework for Authorship Attribution Query Processing. **2018**, 6, 50030-50048	8
86	A Computational Approach for Authorship Attribution on Multiple Languages. **2018**,	1
85	From Genre Classification to Aspect Extraction: New Annotation Schemas for Book Reviews. **2018**,	2
84	Feature Selection in Time Aware Authorship Attribution. **2018**,	
83	Stylometric analysis of classical Arabic texts for genre detection. **2018**, 36, 842-855	4
82	Chat biometrics. **2018**, 7, 454-466	1
81	Automatic Selection of Parallel Data for Machine Translation. **2018**, 146-156	1
80	Multilingual SMS-based author profiling: Data and methods. **2018**, 24, 695-724	7
79	Intrinsic Author Verification Using Topic Modeling. **2018**,	10
78	A multi-language writer identification method based on image mining and genetic algorithm techniques. **2019**, 23, 7655-7669	1
77	Author identification with feature transformation method. **2019**,	

76	The Algorithm of Document Classification of Research and Education Institution Using Machine Learning Methods. **2019**,	0
75	Assessing the Applicability of Authorship Verification Methods. **2019**,	5
74	Deep learning model for unstructured knowledge classification using structural features. **2019**, 1	1
73	Example of Application of n-grams: Authorship Attribution Using Syllables. **2019**, 27-39	1
72	Text-Based Detection of the Risk of Depression. **2019**, 10, 513	14
71	Improving author verification based on topic modeling. **2019**, 70, 1074-1088	16
70	Syntactic n-grams in Computational Linguistics. **2019**,	14
69	One-Class Classification with Deep Autoencoder Neural Networks for Author Verification in Internet Relay Chat. **2019**,	3
68	Selecting the Most Probable Author of Asocial Posting in Online Media. **2019**,	1
67	Robust stylometric analysis and author attribution based on tones and rimes. **2020**, 26, 49-71	9
66	Improved algorithms for extrinsic author verification. **2020**, 62, 1903-1921	6
65	Bucketed common vector scaling for authorship attribution in heterogeneous web collections: A scaling approach for authorship attribution. **2020**, 46, 683-695	1
64	Yapay Zeka Algoritmalarindan Word2vec ile M□eri Bildirimlerinin Kategorilendirilmesi : Categorizing Customer Notifications with an Artificial Intelligence Method Word2vec. **2020**,	
63	Classification of regional and genre varieties of Chinese: A correspondence analysis approach based on comparable balanced corpora. **2020**, 26, 613-640	3
62	The Discriminativeness of Internal Syntactic Representations in Automatic Genre Classification. **2021**, 28, 138-171	1
61	Machine Learning and Feature Selection for Authorship Attribution: The Case of Mill, Taylor Mill and Taylor, in the Nineteenth Century. **2021**, 9, 7143-7151	2
60	An Extensive Text Mining Study for the Turkish Language. **2021**, 272-306	
59	Tags, Borders, and Catalogs. **2021**, 5, 1-29	3

58	Deep Feature Generation for Author Identification.		
57	Binary background model with geometric mean for author-independent authorship verification. 016555152110077		
56	Genre annotation for the Web.		1
55	Semantic Indexing of 19th-Century Greek Literature Using 21st-Century Linguistic Resources. **2021**, 13, 8878		
54	Literary Genre Recognition among Polish Blog Posts. **2021**, 192, 1071-1080		
53	Large Scale Unstructured Document Classification Using Unlabeled Data and Syntactic Information. *Lecture Notes in Computer Science*, **2003**, 88-99	0.9	2
52	A Profile-Based Method for Authorship Verification. *Lecture Notes in Computer Science*, **2014**, 313-326	0.9	20
51	Authorship Attribution with Very Few Labeled Data: A Co-training Approach. *Lecture Notes in Computer Science*, **2014**, 657-668	0.9	2
50	Feature Selection for Enhanced Author Identification of Turkish Text. **2016**, 371-379		3
49	Overview of the PAN/CLEF 2015 Evaluation Lab. *Lecture Notes in Computer Science*, **2015**, 518-538	0.9	23
48	Who Wrote the Web? Revisiting Influential Author Identification Research Applicable to Information Retrieval. *Lecture Notes in Computer Science*, **2016**, 393-407	0.9	17
47	Improving Cross-Topic Authorship Attribution: The Role of Pre-Processing. *Lecture Notes in Computer Science*, **2018**, 289-302	0.9	11
46	Supporting the Cybercrime Investigation Process: Effective Discrimination of Source Code Authors Based on Byte-Level Information. **2007**, 163-173		1
45	A Classifier System for Author Recognition Using Synonym-Based Features. **2007**, 839-849		15
44	Tensor Space Models for Authorship Identification. *Lecture Notes in Computer Science*, **2008**, 239-249	0.9	9
43	Using Grammar-Profiles to Intrinsically Expose Plagiarism in Text Documents. *Lecture Notes in Computer Science*, **2013**, 297-302	0.9	4
42	Classifying Papers from Different Computer Science Conferences. *Lecture Notes in Computer Science*, **2013**, 529-541	0.9	3
41	Empirical Evaluations Using Character and Word N-Grams on Authorship Attribution for Telugu Text. **2015**, 613-623		1

40	Using Compression-Based Language Models for Text Categorization. **2003**, 141-165	43
39	Register in computational language research. **2019**, 1, 100-135	6
38	Source Code Authorship Analysis For Supporting the Cybercrime Investigation Process. **2010**, 470-495	6
37	Lexical Challenges in the Intersection of Applied Linguistics and ANLP. 50-72	7
36	SQSE: A Measure to Assess Sample Quality of Authorial Style as a Cognitive Biometric Trait. **2021**, 3, 583-596	
35	Music Performer Verification Based on Learning Ensembles. *Lecture Notes in Computer Science*, **2004**, 122-131	0.9
34	Automatic Classification of Web documents According to their Styles. **2004**, 11B, 555-562	
33	Training a Genre Classifier for Automatic Classification of Web Pages. **2007**, 15, 305	1
32	Authorship Attribution under the Rules of Evidence: Empirical Approaches in a Layperson's Legal System. **2009**, 15,	1
31	The Rest of the Story: Finding Meaning in Stylistic Variation. **2010**, 79-112	2
30	Korean Mobile Spam Filtering System Considering Characteristics of Text Messages. **2010**, 11, 2595-2602	2
29	Searching, Translating and Classifying Information in Cyberspace. **2011**, 62-75	
28	An Empirical Evaluation of SVM on Meta Features for Authorship Attribution of Online Texts. *Lecture Notes in Computer Science*, **2013**, 28-37	0.9
27	A Comparison of Keywords in the Dynamic Psychology of Jung, Swedenborg and Freud. **2015**, 3,	
26	Using Content-Based Features for Author Profiling of Vietnamese Forum Posts. **2016**, 287-296	1
25	Using Frequent Fixed or Variable-Length POS Ngrams or Skip-Grams for Blog Authorship Attribution. **2016**, 63-74	
24	Word Embedding Based on Low-Rank Doubly Stochastic Matrix Decomposition. *Lecture Notes in Computer Science*, **2018**, 90-100	0.9
23	A Comparative Survey of Authorship Attribution on Short Arabic Texts. *Lecture Notes in Computer Science*, **2018**, 479-489	0.9

22	Türkçe Kelime Yazarlarında Yapay Sinir Ağlarıyla Yazar ve Gazete Tahmin Etme. **2019**, 10, 45-56		0
21	TÜRKÇE METİNLERİN SINIFLANDIRMA BAŞARISINI ARTIRMAK İÇİN YENİ BİR YÖNTEM İNERSİ. **2019**, 24, 125-136		
20	Forensic Analysis Recognition. **2020**, 1-18		
19	Lost in Text: A Cross-Genre Analysis of Linguistic Phenomena within Text. *Ijcol*, **2020**, 6, 47-61	0.1	
18	Enriching feature engineering for short text samples by language time series analysis. *EPJ Data Science*, **2020**, 9,	3.4	1
17	Authorship Attribution by Functional Discriminant Analysis. *Lecture Notes in Computer Science*, **2020**, 438-449	0.9	
16	An Ensemble of Ensembles Approach to Author Attribution for Internet Relay Chat Forensics. *ACM Transactions on Management Information Systems*, **2020**, 11, 1-25	2	2
15	Authorship Analysis Approaches. *International Series on Computer Entertainment and Media Technology*, **2020**, 45-56	0.3	
14	Identification of Urdu Ghazal Poets using SVM. *Mehran University Research Journal of Engineering and Technology*, **2020**, 38, 935-944	0.6	2
13	Analysis of English text genre classification based on dependency types. *Current Issues in Linguistic Theory*, **2021**, 258-270	0.2	
12	Chapter´6. A register variation perspective on varieties of English. *Studies in Corpus Linguistics*, **2021**, 143-178	0.3	
11	Zombies, Conspiracies, and Esthetics in Crisis: A Computational Engagement with Genre Esthetics. **2021**, 231-259		
10	Role of Machine Learning in Authorship Attribution with Select Stylometric Features. *Lecture Notes in Networks and Systems*, **2022**, 920-932	0.5	
9	Estimating Intelligence Quotient Using Stylometry and Machine Learning Techniques: A Review. *Big Data Mining and Analytics*, **2022**, 5, 163-191	4.3	0
8	An Extensive Text Mining Study for the Turkish Language. **2022**, 690-724		
7	A Scientometric Study of the Stylometric Research Field. **2022**, 9, 60		1
6	[illegible] **2022**, 34, 3-20		0
5	Identifying Genre of a Book from Its Summary Using Machine Learning Approach. **2023**, 831-836		0

4	Predictive Analysis for Text Classification: Discrete Units in Company Registration Discourse. **2022**, 27, 229-252	0
3	A Survey on Authorship Analysis Tasks and Techniques. **2022**, 17, 153-167	0
2	Algorithm to Correct the Bigram Method to Identify an Author's Text. **2023**, 15, 245-254	0
1	A study on the construction of die-casting production prediction model by machine learning with Taguchi methods. 1-11	0