

Mapping of Cross-Lingual Emotional Topic Model Research Indexed in Scopus databases from 2000-2020

Ibrahim Hussein Musa^{1,2*}, Ibrahim Zamit³

¹Department of Software Engineering, School of Computer Science and Engineering, Southeast University, Nanjing, CHINA.

²Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing, CHINA.

³Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, CHINA.

ABSTRACT

The scientific research on Cross-Lingual Emotional Topic models (CLETM) saw rising interest over the years. This study aims to identify the publication trends and growth potential of CLETM studies which will offer a better understanding and potential future research directions using bibliometric tools. All published articles related to 'Cross Lingual' or 'Emotional topic model' from Scopus were identified and analyzed using Bibliometrix R-package and VOSviewer software. A total of 1,188 publications were identified from 2000 to 2020 published in 429 journals contributed by 2529 authors with a 2.22 collaboration Index and 2.13 authors per document. Lecture Notes in Computer Science are most sources of published papers with 120 articles, *h*-index 12. The most active country was China with (TNP=145) documents. National Natural Science Foundation of China was the leading organization engaged in CLETM research funding. Li H who is affiliated with the department of Electrical and Computer Engineering, National University of Singapore was the most active author with 17 articles and *h*-index 9. Tsinghua University is the top author's affiliation with 30 articles. The findings of this study provide landmarks, baseline information on vital research productivity, and insights into the historical progressions of CLETM research.

Keywords: Cross lingual Emotional Topic models, Bibliometric analysis, Scopus, VOSviewer.

Correspondence

Ibrahim Hussein Musa

¹Department of Software Engineering, School of Computer Science and Engineering, Southeast University, Nanjing, CHINA.

²Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing 211189, CHINA.

Email id: ibrahim_hm80@yahoo.com

ORCID ID: 0000-0001-8060-6341

Received: 17-06-2022

Revised: 05-09-2022

Accepted: 27-09-2022

DOI: 10.5530/jscires.11.3.46

INTRODUCTION

Cross Lingual Emotional Topic Models (CLETM) is a machine learning technique used for discovering semantic topics from a document collection which provides a convenient way to analyze large amounts of unclassified text.^[1] It is considered as a useful application tool in recent years that has been employed in a variety of applications, such as information retrieval from digital libraries, data visualization, multilingual modeling, statistical inference, and linguistic understanding.^[2-4] Therefore, it is seen that topic models are a mathematical framework that assisted many users in the field of computer science to better understand a large number of document collections: not just to find individual documents but to further understand the general themes present in the collection, as well as new tools to explore and browse extensive collections of scholarly literature.^[2,5] Due to the absence of bibliometric research on-topic models and sentiment analysis, and given the importance posed on the

assessment of scientific production,^[6-8] it became an essential branch of informatics with high usage in various scientific fields.^[9,10] We sought out to make use of available bibliometric tools to analyze both the qualitative and quantitative attributes of published documents. The potential of digital libraries is not only in making documents more accessible but also in providing automated tools that can analyze the literature and help the readers better realize the term scientific contributions.^[11] Bibliometrics is a field of study that attempts to utilize bibliographic data of publications and their citation relations to evaluate and reveal the structure of previous research and disciplines^[12] such as recent published article in field of medicine in the progress in COVID-19,^[13] effects of COVID-19 Pandemic on Mental Health^[14] using of Artificial Intelligence and Machine Learning in Oncology,^[15] infectious disease,^[16,17] Ebola,^[18] and childhood obesity^[19,20] in order to understand the rapid growth of scientific research which is a difficult task and offers the road map of future research direction and challenges towards fill the future research gaps. Therefore, bibliometric studies stand out as a useful research technique to evaluate the continuous and rapidly evolving literature concerning CLETM and possibly identify future research directions.

Copyright

© The Author(s). 2022 This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In this study, we aimed to provide a comprehensive analysis of the research outputs in order to better assess the scientific research productivity regarding CLETM, and to characterize the high-impact articles, annual growth patterns of published documents, authorships, and authors scientific collaboration between researchers in the field of computer science and mathematics over the past 20 years. The main contributions of this study are given as follows. Bridge and provide further and broader understanding of the latest trends in CLETM global publication indexed in Scopus database from 2000–2020.

DATA AND METHODS

Selecting Study design

This study uses bibliometric analysis as a crucial scientific research approach adopted by many scientific scholars to monitor the research performance and scientific progress and also support appropriate policy actions for researchers or governments. The basic bibliometric variables, which presenting annual trends in a number of publications and citation times, number of authors, institutions, countries, journals, collaboration corresponding author analysis, and research hotspots, were assessed to provide researchers with a greater understanding of the documents published in the field of CLETM regarding past, current, and future directions of scientific research progress using bibliometric tools.

Data Sources

The data collected for this study was based on the retrieval of documents indexed in Scopus database (<http://www.scopus.com/>). Scopus is a unique database that can be used for bibliometric analysis.

Search strategy

A search strategy was developed, and comprehensive literature on CLETM literature was performed in Scopus on the 3rd of September 2021. The study used the keywords: “cross-lingual” or “Emotional topic model”, to retrieve CLETM documents from the Scopus database published within the time span of 2000–2020. The keywords were searched in the article titles as to maximize the accuracy of the retrieved inquiry output. Regarding manuscript types, only English written documents, including research articles, conference papers, and review papers, were considered for analysis in this study (Figure 1). Two reviewers (IHM and IZ) independently screened the title to complete a list of the top 10 documents on the CLETM. Finally, all documents were downloaded in text file format (bib.txt, bib. ris, and CSV data format), and as a result, 1,188 publications related to CLETM were the subject of further analysis using the aforementioned bibliometric analysis techniques.

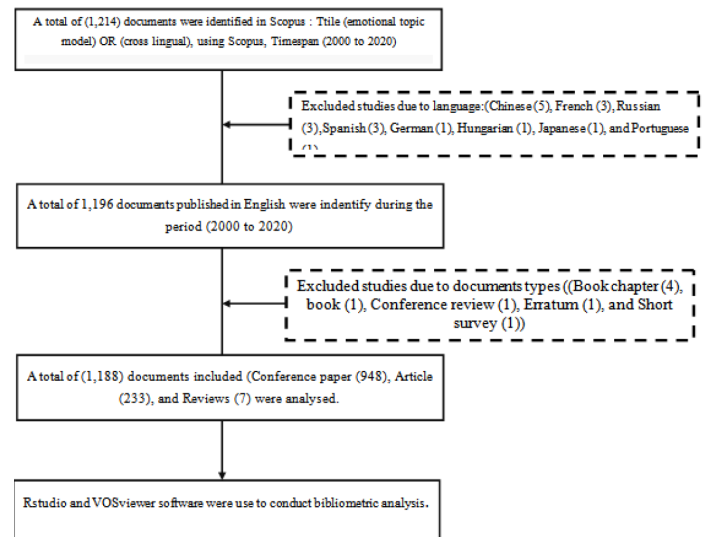


Figure 1: Data extraction process.

Data, analysis and visualization

The data was analysed via visualization software tools^[21] such as the “Biblioshiny app” (using R-studio cloud)^[22] and VOSviewer (version 1.6.6) package program (Leiden University, Leiden, The Netherlands) was used for mapping analysis and it also facilitates the visualization of dynamics and structure of information for the analysed documents.^[23]

RESULTS

Characteristics of the meta-data

The retrieved documents were published in 2000–2020, in 429 journals with 2,529 authors contributed, and Collaboration Index of 2.22 per document. According to the analysis, all documents received a total citation score of 11,562. The majority of the documents were conference papers with 948 (79.80%), followed by articles 233 (19.61%), and review papers 7 (0.59%) as presented in (Table 1). The Annual trends of publication and citation times during the study period was represented in Figure 2.

Most cited documents

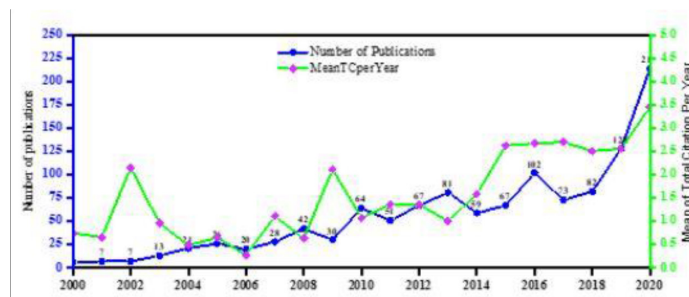
The top 10 cited documents are presented in Table 2. The top most cited paper for CLETM research is titled ‘Co-training for cross-lingual sentiment classification’ which received over 324 citations. This paper proposes to use the co-training approach to address the problem of cross-lingual sentiment classification,^[24] followed by ‘Learning a multilingual subjective language via cross-lingual projections’ with 232 citations.^[25] The later study proposed a new unified framework for monolingual and cross-lingual information retrieval.

Our findings show that LI H from the Department of Electrical and Computer Engineering, National University of Singapore is the most productive author with 17 published

Table 1: Descriptive characteristics of study on CLETM (2000-2020).

Description	Counts/ indices	Description	Counts/ indices
Timespan	2000:2020	Low income	1 (1.52)
Sources (Journals, Books, etc)	429	Lower-Middle income	9 (13.64)
Documents	1,188	High income countries	36 (54.55)
Funding agencies	159	Document Contents	
Authorship Affiliation	160	Keywords Plus (ID) ^a	4210
Corresponding Author's Country	50	Author's Keywords (DE) ^b	1693
Total citation score	11,562	Authors	
Average years from publication	6.57	Authors	2529
Average citations per documents	10	Author Appearances (AA) ^c	4164
Average citations per year per doc	1.531	Authors of single-authored documents	65
References	29527	Authors of multi-authored documents	2464
Document types, n (%)		Authors Collaboration	
Article	233 (19.61)	Single-authored documents	79
Conference paper	948 (79.80)	Documents per Author	0.47
Review	7 (0.59)	Authors per Document	2.13
Geographical contribution of research (n=66)		Co-Authors per Documents	3.51
Upper-Middle income	20 (30.3)	Collaboration Index (CI) ^d	2.22

^a: Frequency distribution of keywords associated with the document by Scopus; ^b: Frequency distribution of the authors' keywords; ^c: Number of author appearances; ^d: The scientific collaboration on the social process by which two or more researchers are work together sharing their intellectual and material resources to produce new scientific knowledge.

**Figure 2:** Annual trends of publication and citation times.

articles, followed by Vuli I from 'Language Technology Lab, University of Cambridge' with 17 published articles. The third most prolific author was Zhang Y with 15 published articles. Furthermore, the top 10 authors whom have published more than 10 papers on Cross-Lingual Emotional Topic Model are presented in Table 3.

Output analysis of top 10 countries wise publication

A total of 66 countries contributed in 1,188 papers. China tops the list with around 145 publications followed by The United States of America (USA) with 58 papers, and Germany with 46 papers. These three countries are found to be the most productive countries for research in the field of CLETM. Additionally, our results showed that China is ranked first in terms of the number of documents and number of citations (Table 4).

Output analysis of top 10 most cited Journal source

The 1,188 documents on CLETM were published in 429 Journals, the top 10 most published journals are listed in Table 5. Conference proceeding papers which published as Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) was the greatest number of publications 120 followed by Proceedings of the Annual Conference of the International Speech Communication Association, Interspaced with 38 papers, and Ceur Workshop Proceedings with 35 papers.

The top 10 Affiliation and funding agencies

The study also exhibits the top 10 affiliations and funding agencies. Tsinghua University is the top author affiliation institution with 30 authors (18.75%), followed by The University of Edinburgh 27 (16.88%). China's National Natural Science Foundation is the top funding agency for CLETM research with 97 articles (61.01%), followed by European Commission with 55 articles (34.59%) as presented in (Table 6).

Keyword Occurrence Analysis

Figure 3 depicts the result of the keywords plus analysis which unveils the most used keywords in CLETM literature which can enable the identification of research themes and topics that have been heavily studied by researchers and documented in Scopus database during the past 20 years. The top 10 frequent

Table 2: Top 10 most cited documents on CLETM.

Authors	Title	Year	TC
Wan X	Co-training for cross-lingual sentiment classification	2009	324
Mihalcea R <i>et al.</i>	Learning multilingual subjective language via cross-lingual projections	2007	232
Vulić I Moens M.-F.	Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings	2015	171
Agirre E., <i>et al.</i>	SemEval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation	2016	154
Lavrenko V <i>et al.</i>	Cross-lingual relevance models	2002	152
Artetxe M <i>et al.</i>	A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings	2018	145
Zhou X <i>et al.</i>	Attention-based LSTM network for cross-lingual sentiment classification	2016	135
Conneau A, Lample G.	Cross-lingual language model pretraining	2019	118
Täckström O <i>et al.</i>	Cross-lingual word clusters for direct transfer of linguistic structure	2012	112
Duong L., <i>et al.</i>	Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser	2015	110

TC: Total citations score

Table 3: Top 10 most prolific authors on CLETM.

Authors	Affiliations	<i>h</i> _index	TC	TNP	%
LI H	Department of Electrical and Computer Engineering, National University of Singapore,	9	168	17	1.43
Vulić I	Language Technology Lab, University of Cambridge	9	392	17	1.43
Zhang Y	Institute of Advanced Technology, Westlake Institute for Advanced Study Information	6	180	15	1.26
Ji H	Rensselaer Polytechnic Institute, NY, USA	6	112	12	1.01
Li J	Department of CST, Tsinghua University, Beijing, China	5	152	12	1.01
Wang W	School of computer Science and Engineering Northerneasten University, Shenyang, China	7	99	12	1.01
Wang Z	Amazon AWS	7	314	12	1.01
Zhang L	Institute AIFB, Karlsruhe Institute of Technology, Germany	5	75	12	1.01
Xu J	Sohu, Inc., Beijing, China	7	215	11	0.93
Glava G	Data and Web Science Group, University of Mannheim, Germany	6	99	10	0.84

TC: Total Citations; TNP: Total Number of Publications

Table 4: Top 10 countries wise publications on CLETM.

Country	Pubs (%)	Freq (%)	SCP	MCP	TC
China	145 (12.21)	25.39	113	32	1378
USA	58 (4.88)	10.16	49	9	1122
Germany	46 (3.87)	8.06	28	18	560
Japan	33 (2.78)	5.78	28	5	209
Spain	33 (2.78)	5.78	26	7	224
UK	27 (2.27)	4.73	20	7	327
India	25 (2.10)	4.38	22	3	115
Italy	15 (1.26)	2.63	14	1	187
Hong Kong	14 (1.18)	2.45	8	6	137
Czech Republic	13 (1.09)	2.28	11	2	40

TC: Total Citations; MCP: Multiple Country Publication; SCP: Single country publications; Pubs: Publications

keywords were as follows: “cross-lingual” (589) times, “computational linguistics” (324) times, “natural language processing systems” (319) times, “semantics” (258) times, “information retrieval” (177) times, “translation (languages)” (163) times, “linguistics” (150) times, “speech recognition” (118) times, target language” (115) times, and “machine translations” (114) times among others.

Conceptual structure map analysis

Figure 4, shows the Conceptual structure of keyword was use to represent the conceptual structure of the current literature on CLETM for capturing an article content with greater depth of understand the scientific concepts in CLETM research over the past 20 years. to the analysis of CLETM mapping

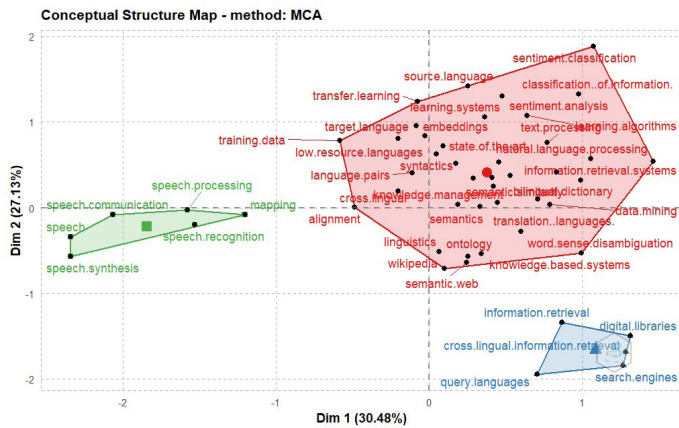


Figure 4A

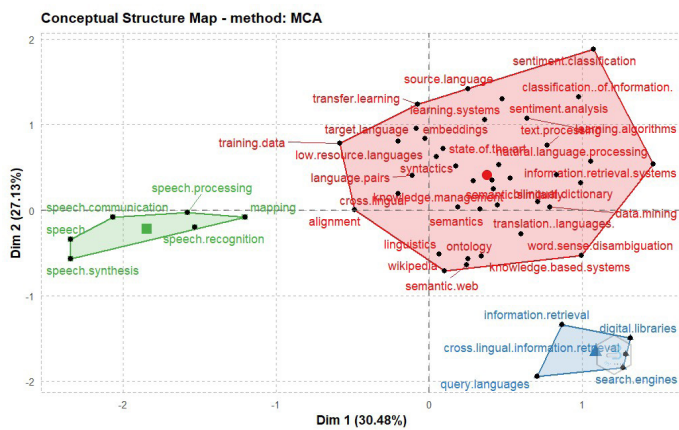


Figure 4B

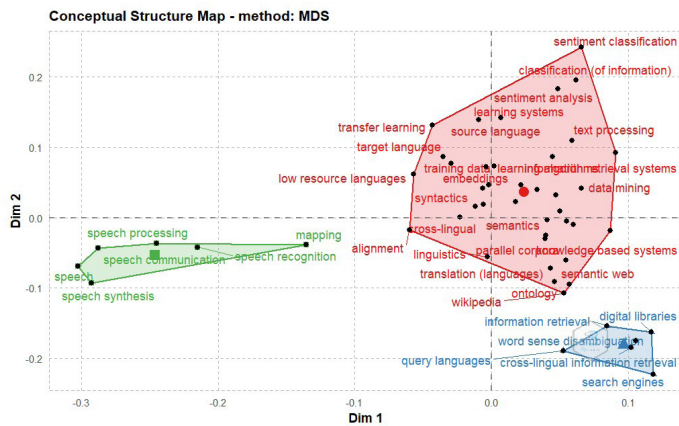


Figure 4C

Figure 4: Conceptual structure map of CLETM themes using techniques of Correspondence Analysis (A), multiple Correspondence Analysis (B), and Multidimensional Scaling (C) method, for ($n=50$ Keyword).

based on the article title by using thematic evolution of topics reported that caught the reader's attention and identified the extent to which topics are related to each other (Figure 6B).

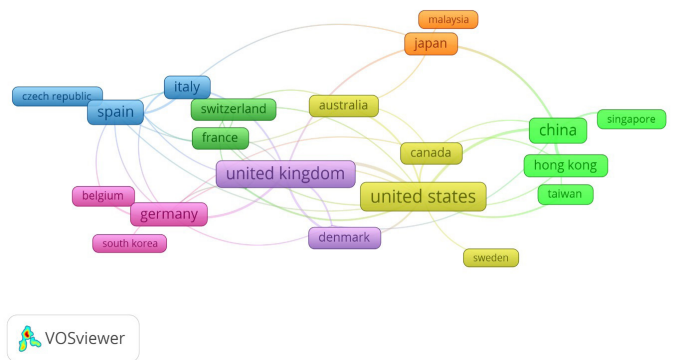


Figure 5: Co-authorship analysis between countries based on the Links (L) and Total Link Strength (TLS) between two countries.

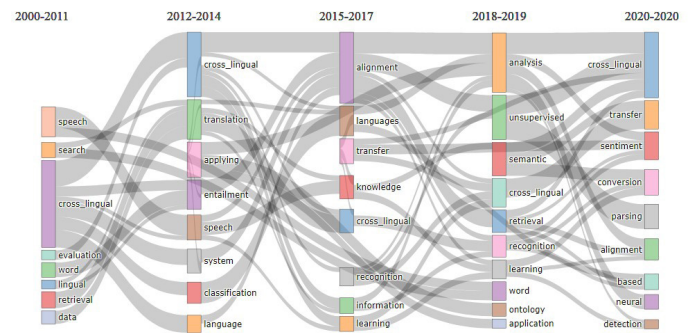


Figure 6A

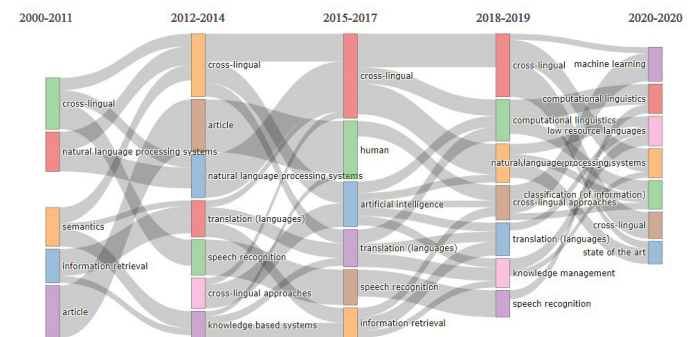


Figure 6B

Figure 6: Representative thematic evolution Sy diagram based on keyword Plus (A) and Article title (B).

Collaboration analysis between countries and institution on CLETM

To uncover new knowledge and determine the collaboration between researchers within the top 50 institutions and countries, R software was used. The analysis showed that

Tsinghua University, Nanyang Technological University, National University of Singapore, Soochow University, Microsoft Research Institution were located in cluster one with closeness (0.005, 0.004, 0.004, 0.004, and 0.0004). The analysis of collaborations between countries also shows that (China, the USA, Hong Kong) are located in cluster one with closeness between countries as (0.012, 0.014, and 0.011) as can be seen in supplementary Table S1.

DISCUSSION

The effect that Bibliometrics has had in the past years is significant weather that was in governing, policymaking, or trying to better understand certain scientific fields.^[26] The data for this study was retrieved from Scopus, this database provides different *h*-index ratings for authors which are needed to track citations and determine the impact of publications.^[27] A total of 1188 documents have been selected from the Scopus database. The retrieved documents were published during 2000–2020. The study focused on the comprehensive analysis

of the global publication output and outlined possible future directions to the researchers just venturing into the field of CLETM by providing sufficient information on the growth and development of the literature, information on active authors, journals, countries, institutions, funding agencies, as well as complete keyword analysis for terms most frequently used in CLETM research.

The findings showed that research surrounding CLETM was steadily increasing and reached its highest peak in 2010 and 2020. The retrieved publications on CLETM received a high number of citations with an average of 10 citations per document, which is indicative of a large number of readers and scholars. The most-reported published documents in CLETM literature were published as conference papers 79.80%, followed by full research articles 19.61% and review papers 0.59%. These findings shows that researchers in the field of CLETM prefer to publish their work as conference papers which they believe can gain more attention from the community rather than another type of documents.

Table S1: Collaboration analysis of Institutions and countries.

Collaboration analysis of top 25 countries					Collaboration analysis of top 25 Institutions				
Institutions	Cluster	Betweenness	Closeness	PageRank	Country	Cluster	Betweenness	Closeness	PageRank
Tsinghua University	1	162.434	0.005	0.053	China	1	72.239	0.012	0.092
Nanyang Technological University	1	51.359	0.004	0.024	USA	1	194.882	0.014	0.105
National University of Singapore	1	32.309	0.004	0.032	Hong Kong	1	4.194	0.011	0.032
Institute for Info COMM Research	1	29.083	0.004	0.023	Germany	2	193.452	0.014	0.095
Soochow University	1	0.000	0.004	0.008	UK	2	443.756	0.015	0.119
Microsoft Research	1	53.004	0.004	0.021	Japan	3	58.986	0.012	0.030
University of Edinburgh	2	86.490	0.004	0.048	Spain	4	109.498	0.013	0.046
IDIAP Research Institute	2	39.614	0.004	0.019	India	5	7.896	0.011	0.017
Saarland University	2	33.723	0.004	0.019	Italy	6	58.777	0.013	0.036
Nagoya Institute of Technology	2	36.000	0.004	0.029	France	7	23.285	0.012	0.026
Northwest Normal University	2	0.000	0.003	0.008	Singapore	8	1.389	0.011	0.026
Carnegie Mellon University	3	17.680	0.004	0.027	Belgium	9	0.000	0.010	0.009
University of California	3	41.120	0.004	0.030	Netherlands	10	26.258	0.011	0.033
University of Southern California	3	64.708	0.004	0.038	Switzerland	11	12.019	0.012	0.027
Columbia University	3	33.837	0.004	0.028	Malaysia	12	0.373	0.008	0.007
University of Michigan	3	172.144	0.005	0.038	Australia	13	18.159	0.011	0.023
University of Illinois at Urbana-Champaign	3	6.588	0.004	0.012	Canada	14	8.493	0.011	0.025
City University of New York	3	20.130	0.004	0.022	Czech Republic	15	0.000	0.009	0.007
University of Illinois	3	35.608	0.004	0.028	Ireland	16	10.372	0.011	0.017
University of Cambridge	4	126.636	0.005	0.048	Iran	17	3.067	0.009	0.008
University of Mannheim	4	0.000	0.004	0.024	Finland	18	0.467	0.010	0.011
University of Copenhagen	4	89.936	0.004	0.028	Slovenia	19	1.117	0.008	0.008
Department of Computer Science	4	0.000	0.004	0.015	Denmark	20	23.118	0.011	0.025
University of Amsterdam	4	0.000	0.004	0.008	Egypt	21	0.000	0.009	0.010
Johns Hopkins University	5	0.000	0.003	0.013	Israel	22	0.000	0.010	0.008

The ‘Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)’ are very much active in publishing documents on CLETM over the past years. The published documents had a relatively high reported scientific contribution and ranking having around 100 documents receiving 827 citations score. Thus, based on the analysis and evidence reported, it seems that researchers are particularly interested in publishing their research as lecture notes in these domains and subsequently refer to these notes for reference in other publications.

The most active author in this field is LI H from the Department of Electrical and Computer Engineering, National University of Singapore. Furthermore, the collaboration analysis published in the field indicates that (Tsinghua University, Nanyang Technological University, National University of Singapore, Institute for Information Research, Soochow University, and Microsoft Research) were close in collaboration over the past period. The study ranks China as the pioneering country in CLETM research. As a consequence, the ongoing funding and support from the National Natural Science Foundation of China led to Chinese authors having the maximum number of publications in this domain.

The keyword occurrence analysis mainly focused on ‘cross-lingual’, ‘computational linguistics’, ‘natural language processing systems’, ‘semantics’, ‘information retrieval’, ‘translation (languages)’, ‘linguistics’, ‘speech recognition’, ‘target language’, and “machine translations” among others. These top 10 keywords can be used to identify future research hotspots for CLETM. The and analyze the conceptual areas using three model shows the distribution of the topic and means the attention of the researchers to the CLETM subject theme of the study.

In addition, cross-lingual, Speech recognition research has received more attention during the 2000–2020 time slice in relation to the thematic evolution analysis. Co-training for cross-lingual sentiment classification article published by Wan, 2009,^[24] and Learning a multilingual subjective language via cross-lingual projections published by Mihalcea *et al.* 2007 attracted the interest of most scientists, and they had the highest citations.^[28] These articles originally introduced CLETM to the international scientific community.

The analysis of the top cited articles recognized the article published by Wan *et al.*, 2009 as the most cited with more than 324 times.^[24] While other 10 most cited articles offer new idea in CLETM research.^[25,28–35] Thereafter, research in the cross-lingual Emotional topic model was increased after the year 2010 and Year 2020. Moreover, highly cited articles are very different from ‘ordinary’ cited articles^[36] Since citation is used as a key indicator of research quality, highly cited publications

are positively correlated with the *h*-index of the author, institution, and country.^[37] Regarding top funding agencies, most of the research concerning CLETM was funded by the National Natural Science Foundation of China. Therefore, the findings further highlights that there is need to enhance research in the area of CLETM, and increase collaborations among different authors for future research.

Previous studies argued that analysis of citation for an author is a good factor to analyze the impact and usability of research done by that researcher.^[38–41] Significant correlations were noted between the number of citations and the years since publication, Number of countries, Number of authors, and Authors *h*-index. In our analysis, we noticed that based on the number of articles and citation score, authors Vulic I, Zhang Y, and Li H from developed countries dominated the list. This could imply the usability and relevance of CLETM research in developed countries’ communities. Based on the total link strength (TLS) the United States, United Kingdom, Germany, and China showed high occurrence between them in CLETM research. Despite its many advantages, our study has some limitations, which can be considered in the scope of future research in the CLETM study. First, we only used one database Scopus to obtain the released publications. Therefore, other published documents not indexed in Scopus are not included in the analysis. Furthermore, we only included articles, conferences, and review papers in the analysis. In future bibliometric analysis, researchers may consider using more diverse data sources. Future bibliometric studies might consider using other databases such as Web of Science, Google Scholar to provide a more comprehensive overview of research productivity in the field of CLETM.

CONCLUSIONS AND FUTURE RECOMMENDATIONS

The literature on CLETM had been continuously growing for the last ten years. We here analyzed the literature published from 2000 to 2020 and found it was produced by 2529 authors across 66 countries and published in 429 sources indexed in Scopus. The National Natural Science Foundation of China funded the greatest number of studies National Natural Science Foundation of China. The study provides indicators for uncovering vital research hot spots in the field of CLETM. In addition, the study delivers further information on country collaborations for future researchers based on the single country publication and multiple country publication. Following the geographical distribution of CLETM research, the analysis showed over 50 % of research on CLETM was produced by high-income countries. Several developing countries might be facing massive challenges in the field; therefore, further analysis on empirical research encompassing low-income countries, lower-middle income countries would

enrich the basis for proposing and implementing solid policy measures aimed at promoting CLETM in a wider range of contexts.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest.

FUNDING

Ibrahim Zamit acknowledge support from the ANSO Scholarship for Young Talen.

REFERENCES

- Alghamdi R, Alfalqi K. A Survey of Topic Modeling in Text Mining. *Int J Adv Comput Sci Appl*. 2015;6(1). doi: 10.14569/ijacsa.2015.060121
- Boyd-Graber J, Hu Y, Mimno D. Applications of topic models. *Found Trends Inf Retr*. 2017;11(2-3):143-296. doi: 10.1561/15000000030
- Li J, Chen X, Hovy E, Jurafsky D. Visualizing and understanding neural models in NLP. 2016 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol NAACL HLT 2016 - Proc Conf. 2016. doi: 10.18653/v1/n16-1082
- Wallach HM. Structured Topic Models for Language. Doctor. 2008.
- Boyd-Graber J, Yuening H, Mimmo D. Applications of Topic Models Jordan. *Found. Trends Human-Computer Interact*. 2015.
- Merediz-Solá I, Bariviera AF. A bibliometric analysis of bitcoin scientific production. *Res Int Bus Financ*. 2019;50:294-305. <https://doi.org/10.1016/j.ribaf.2019.06.008>
- Muhuri PK, Shukla AK, Janmaiya M, Basu A. Applied soft computing: A bibliometric analysis of the publications and citations during (2004–2016). *Appl Soft Comput J*. 2018;69:381-92. <https://doi.org/10.1016/j.asoc.2018.03.041>
- Borgman CL. Communication and Collaboration Scholarly Communication and Bibliometrics. *Annu Rev Inf Sci Technol*. 2002.
- Xu Z, Yu D, Wang X. A bibliometric overview of International Journal of Machine Learning and Cybernetics between 2010 and 2017. *Int J Mach Learn Cybern*. 2019;10(9):2375-87. <https://doi.org/10.1007/s13042-018-0875-9>
- Beattie PD, Bishop JM. Self-Localisation in the "Scenario" Autonomous Wheelchair. *J Intell Robot Syst Theory Appl*. 1998;22(3):255-67. <https://doi.org/10.1023/a:1008033229660>
- Mann GS, Mimno D, McCallum A. Bibliometric impact measures leveraging topic analysis. *Proc ACM/IEEE Jt Conf Digit Libr*. 2006;65-74. <https://doi.org/10.1145/1141753.1141765>
- Banshal SK, Uddin A, Singhal K, Singh VK. Computer science research in India: A scientometric study. 12th IEEE Int Conf Electron Energy, Environ Commun Comput Control (E3-C3), INDICON 2015. 2016. <https://doi.org/10.1109/INDICON.2015.7443320>
- Akintunde TY, Chen S, Musa TH, Amoo FO, Adediji A, Ibrahim E, Tassang AE, Musa IH, Musa HH. Tracking the progress in COVID-19 and vaccine safety research—a comprehensive bibliometric analysis of publications indexed in Scopus database. *Human Vaccines & Immunotherapeutics*. 2021;17(11):3887-97. DOI: 10.1080/21645515.2021.1969851
- Akintunde TY, Musa TH, Musa HH, Musa IH, Chen S, Ibrahim E, Tassang AE, Helmy MS. Bibliometric analysis of global scientific literature on effects of COVID-19 pandemic on mental health. *Asian Journal of Psychiatry*. 2021;63:102753.
- Musa IH, Musa TH, Zamit I, Okeke M. Artificial Intelligence and Machine Learning in Oncology: Historical Overview of Documents Indexed in the Web of Science Database. *Eurasian J Med Oncol* 2021;5:239-48.
- Musa TH, Akintunde TY, Musa IH, Mohammed LA, Tassang AE, Musa HH. Rift valley fever: Thematic analysis of documents indexed in the Web of Science Core Collection database. 2022. <https://doi.org/10.21037/aoi-21-9>
- Gatasi G, Musa TH, Odjidja EN. Bibliometric analysis of the top 100 cited articles on HIV/AIDS. *Ann Infect*. 2021;5:6.
- Kawuki J, Yu X, Musa TH. Bibliometric Analysis of Ebola Research Indexed in Web of Science and Scopus (2010-2020). *Biomed Res Int*. 2020. doi: 10.1155/2020/5476567
- Kawuki J, Ghimire U, Papabathini SS, Obore N, Musa TH. A bibliometric analysis of childhood obesity research from China indexed in Web of Science. *J Public Heal Emerg*. 2021. doi: 10.21037/jphe-20-95
- Musa TH, Akintunde TY, Gatasi G, Ghimire U, Kawuki J, Musa HH. A bibliometric analysis of the 100 top-cited articles on global malnutrition indexed in Web of Science. *J Public Heal Emergency* 2021;0-2
- Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools for conducting bibliometric analysis in science: An up-to-date review. *Prof la Inf*. 2020;29(1). <https://doi.org/10.3145/epi.2020.ene.03>
- Dervis H. Bibliometric analysis using bibliometrix an R package. *J Scientometr Res*. 2019;8(3):156-60. <https://doi.org/10.5530/JSCIRES.8.3.32>
- Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 2010;84(2):523-38. <https://doi.org/10.1007/s11192-009-0146-3>
- Wan X. Co-training for cross-lingual sentiment classification. *ACL-IJCNLP 2009 - Jt Conf 47th Annu Meet Assoc Comput Linguist 4th Int Jt Conf Nat Lang Process AFNLP, Proc Conf*. 2009;235-43. <https://doi.org/10.3115/1687878.1687913>
- Vulic I, Moens MF. Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. *SIGIR 2015 - Proc 38th Int ACM SIGIR Conf Res Dev Inf Retr*. 2015;363-372. <https://doi.org/10.1145/2766462.2767752>
- Boyce R, Rosch R, Finlayson A, Handuleh D, Walhad SA, Whitwell S, Leather A. Use of a bibliometric literature review to assess medical research capacity in post-conflict and developing countries: Somaliland 1991–2013. *Tropical Medicine & International Health*. 2015;20(11):1507-15. <https://doi.org/10.1111/tmi.12590>
- De Groote SL, Raszewski R. Coverage of Google Scholar, Scopus, and Web of Science: A case study of the h-index in nursing. *Nurs Outlook*. 2012;60(6):391-400. <https://doi.org/10.1016/j.outlook.2012.04.007>
- Mihalcea R, Banea C, Wiebe J. Learning multilingual subjective language via cross-lingual projections. *ACL 2007 - Proc. 45th Annu Meet Assoc Comput Linguist*. 2007;976-83.
- Ganguly D, Leveling J, Jones GJF. Cross-lingual topical relevance models. 24th Int. Conf Comput Linguist. - Proc. COLING 2012 Tech. Pap. 2012;927-42.
- Agirre E, Banea C, Cer D, Diab M, Gonzalez Agirre A, Mihalcea R, Rigau Claramunt G, Wiebe J. SemEval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation. *SemEval 2016 - 10th Int Work Semant Eval Proc*. 2016. <https://doi.org/10.18653/v1/s16-1081>
- Zhou X, Wan X, Xiao J. Attention-based LSTM network for cross-lingual sentiment classification. *EMNLP 2016 - Conf Empir Methods Nat Lang Process Proc*. 2016. <https://doi.org/10.18653/v1/d16-1024>
- Garneau N, Godbout M, Beauchemin D, Durand A, Lamontagne L. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings: Making the method robustly reproducible as well. *Lr. 2020 - 12th Int Conf Lang Resour Eval Conf Proc*. 2020
- Täckström O, McDonald R, Uszkoreit J. Cross-lingual word clusters for direct transfer of linguistic structure. *NAACL HLT 2012 - 2012 Conf North Am Chapter Assoc Comput. Linguist Hum Lang Technol Proc Conf*. 2012.
- Guo J, Che W, Yarowsky D, Wang H, Liu T. Cross-lingual dependency parsing based on distributed representations. *ACL-IJCNLP 2015 - 53rd Annu Meet Assoc Comput Linguist 7th Int Jt Conf Nat Lang Process Asian Fed Nat Lang Process Proc Conf*. 2015;1234-44. <https://doi.org/10.3115/v1/p15-1119>
- Ballesteros L, Croft B. Dictionary methods for cross-lingual information retrieval. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 1996. <https://doi.org/10.1007/bfb0034731>
- Aksnes DW. Characteristics of highly cited papers. *Res Eval*. 2003;12(3):159-70. <https://doi.org/10.3152/147154403781776645>
- Aksnes DW. Citation rates and perceptions of scientific contribution. *J Am Soc Inf Sci Technol*. 2006;57(2):169-85. <https://doi.org/10.1002/asi.20262>
- van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*. 2017;111(2):1053-70. <https://doi.org/10.1007/s11192-017-2300-7>
- Iqbal W, Qadir J, Tyson G, Mian AN, Hassan S ul, Crowcroft J. A bibliometric analysis of publications in computer networking research. *Scientometrics*. 2019;119(2):1121-55. <https://doi.org/10.1007/s11192-019-03086-z>
- Lopes RM, Faria DJG dos S de, Fidalgo-Neto AA, Mota FB. Facebook in educational research: a bibliometric analysis. *Scientometrics*. 2017;111(3):1591-621. doi: 10.1007/s11192-017-2294-1
- Gangarde R, Sharma A, Pawar A. Bibliometric survey of privacy of social media network data publishing *Libr Philos Pract*. 2019.