Summer 9-11-2018

A Bibliometric Analysis of "Health Cloud" Scientific's Productions

Mahboobe Shokoohian
Department of Knowledge and Information Science, University of Isfahan, mahboobeshokohian@gmail.com

Asefeh Asemi
Department of Knowledge and Information Science, University of Isfahan, asemi@edu.ui.ac.ir

Ahmad Shabania
Department of Knowledge and Information Science, University of Isfahan, shabania@edu.ui.ac.ir

Mozafar Cheshme Sohrabi
Department of Knowledge and Information Science, University of Isfahan, afasemi@gmail.com

A Bibliometric Analysis of "Health Cloud" Scientific's Productions

Mahboobeh Shokoohian, PhD Std. mahboobehshokohian@gmail.com
Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran
Asefeh Asemi (Corresponding Author*), Email:* *asemi@edu.ui.ac.ir*
Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran
Doctoral School of Business Informatics, Corvinus University of Budapest, Hungary
Ahmad Shabani
Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran
Mozafar Cheshmeh Sohrabi
Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran

Abstract

Introduction: Cloud computing is an innovative paradigm meeting the user's demand for accessing a shared source comprising adjustable computational sources, such as servers and applied programs. An increase in the costs of information technology, emerging problems with updating software and hardware, and expanded storage volume, make it possible to utilize cloud-based health information cases. Organizations have focused on cloud platform-based services as a new opportunity to develop the software industry for healthcare. The aim of the research is to conduct a bibliometric study of the scientific productions on "health cloud".
Methodology: The present study, applied in nature, was conducted using a bibliometric and scientometric method. It was conducted in 2019 using PubMed and key portmanteaus over the period 2009-2019. Subjected to the application of input and output standards, 491 research papers were selected for analysis.
Findings: The findings revealed that the production of health cloud-focused papers over a decade, excluding those in 2017, had an upward trend. The USA, India, and China were the most productive in this respect. Having presented 5 papers on cloud computing, Costa, Lee, Malamateniou, Stoicu-Tivadar, Vassilacopoulos, writers, were most productive. The greatest co-occurrence was that of the words Internet, electronic health records, computer security, information storage and retrieval, algorithms, confidentiality, female, male, delivery of health care, computer communication networks, medical informatics, mobile applications, data mining, and health information exchange.
Conclusion: The results of the present study indicate the leading status of the USA in health cloud publications. In view of the recognition received for using cloud computing, the trend of the papers in the base was upward in nature. On analysis of the co-occurrence of words, the largest cluster was that of cloud computing with 6 items focused on: The Internet of Things (IoT), Electronic health record, healthcare, and e-health in one cluster, indicating the continuity of the issues.

Keywords: PubMed, Scientific Production, Cloud Computing, Health, Bibliometrics, Co-occurrence of Words.

Introduction

In view of the influence and developing new technologies, shifting of information systems towards digitizing their information, preparing the mechanisms required, making them web-based, using cloud computing, and the need for a new paradigm in the information environment, models need to be presented based on which powerful infrastructures can be developed.

Of the definitions for cloud computing, NIST has offered a comprehensive one: “cloud computing is a model for providing easy access based on the user's demand by network for a set of variable and configurable computational resources with minimal need for resource management or the need for direct intervention of service providers for the speed created and used“(Mell & Grance, 2011).

Of the cloud-based Deployment Models, reference can be made to Private Cloud developed by an organization for its private use, Hybrid Cloud developed by organizations with identical needs and sharing the same infrastructure to enjoy the benefits of cloud computing, Public Cloud for a multitude of clients using a shared infrastructure, and Hybrid Cloud for placing the infrastructure of an organization's private cloud in that of a public cloud (Schweitzer, 2011).

The principal mission of the health system is to promote the health level and respond to the needs of the public and society. To this end, the health technology needs updating for equal distribution of health information among the public. The Healthcare Information System is the best cloud computing functional area for health. It can have a cloud-based design to overcome such problems as lack of interaction and sharing of information, in order to integrate the health information sources (Aijing & Jin, 2015). Transferring medical information to cloud can not only improve the individual health management, scalability, performance, and flexibility but also cut down on the costs. Therefore, cloud computing will create new treatment opportunities for patients (Mayo et al, 2017)

Publication of research papers on cloud computing over the last decade has seen considerable growth owing to its characteristics for enjoying health services in the e-world. In this context, the best tool for a systematic analysis of research areas, the release output of any topic, writer, country, etc. is a bibliometric analysis. Researchers, executives and decision-making institutions can garner good support for their decisions.

Nowadays, making use of topographic maps as a powerful tool to study the structure and dynamics of disciplines has assumed considerable importance. The maps reveal the structure of discipline by showing the interrelations of important areas within the discipline, and how the areas grow over time. Researchers can use bibliometric maps to achieve an understanding of the area in which they are working. Some maps, for instance, indicate the interrelations of writers and journals, and some other indicate the interrelations of keywords (Van Eck et al, 2010).

In the analysis of co-occurrence of words, the juxtaposition of words or concepts in a document indicates the contiguity of their content, where, by measuring the degree of co-occurrence, the conceptual structure of a particular area of research can be tracked. The conceptual network can be tracked by counting the occurrences of each thematic idiom juxtaposed with other idioms. According to the analysis used for the co-occurrence of words, if two idioms co-occur more frequently in a document, it is an indication of their conceptual interrelation. Moreover, the relationship between two areas of science can be extracted to trace the development of both (Callon et al, 1983).

A plethora of studies conducted on scientometrics in addition to a variety of research activities on cloud computing in recent years will be reviewed. In their paper "A Scientometric Analysis of Cloud Computing Literature", Voβ & Heilig (2014) did a study on 32620 unique keywords in Scopus over the period 2008-213. The writer of "Buyya R." from Australia and those from China and the USA had the largest number of papers on cloud computing. The keywords from virtualization, security, computer systems, grid computing, scalability virtual machines, distributed computer systems, Hadoop, data mining, the Internet, cloud services, and cloud computing had the highest frequency. In their paper, Jan, Wani & Hafiz (2015) conducted a scientometric analysis of cloud computing in the Web of Science over the period 2009-2013. They studied the growth pattern of cloud computing literature in 1879 publications of ten leading countries and those of India. The USA and China had the largest and India the smallest number of publications. The publication ns in this area had an upward trend over the period. Sivakumaren, Swaminathan & Karthikeyan (215) conducted a study titled "A Scientometric Study of the Growth and Development of Cloud Computing" having extracted the data from the Web of Science over the years 2001-2010. Analysis of 510 cloud-related cases indicated that the Buyya.R, a writer, and the US produced the largest number of papers in an upward order over time. Siddagangaiah (2017) did a bibliometric study of cloud computing productions in the Web of Science over a period of 21 years from1996 through 2017 to analyze the results, using such software as Histcite and VosViewer. The findings revealed that the writer, Buyya R. from Australia was the most productive in the area and on an upward trend over time. Hanumanth (2017) did a scientometric study of cloud computing in IEEE over the period 2013-2016, working on 101 research papers, with the findings indicating that the papers published over the years had an upward trend and that Buyya R. had the largest number of pulications.

A review of the backgrounds indicated that the research addressing and analyzing the studies conducted on cloud computing have analyzed papers indexed over different periods of time in the Web of Science, Scopus, and IEEE, and that heretofore no study has been conducted on the use of co-occurrence of words in tracing the structure of cloud computing in the PubMed. The major issue of the present study arises when, with respect to the significance of the issue, researchers need sufficient information on the scientific production in the area to develop their research. Furthermore, in view of the significance of PubMed on health, the principal aim of this study is to conduct a bibliometric study of "cloud health" in the PubMed over a period of 10 years from 2009 through 2019 and on the proportions of response to special aims. Establishing the growth trend of the scientific papers on cloud health, the writers and countries with the largest number of papers published in the area, the co-occurrence of words in the papers in area, constitute the special aims of the study.

Method

This study is applied in nature and based on a bibliometric method and analysis of co-occurrence of words. The research population for all the papers on cloud health over the past decade was in PubMed. The research was conducted in the following three steps.

Step I: The results were confined to using advanced search, the search formula selected (Table 1), the time frame of May 30th '09-May 27th '19, and the English language.

Table 1. Adavanced search formula on cloud health in PubMed over 2009-2019

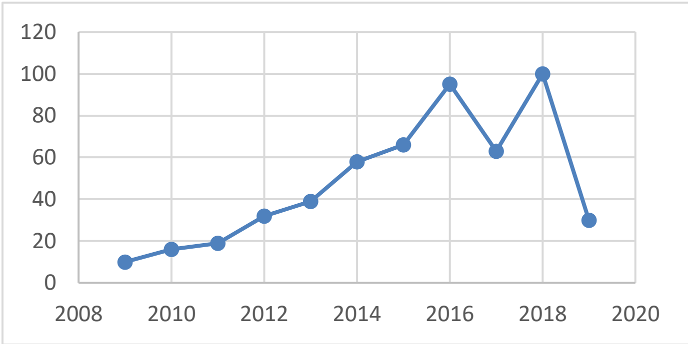
فرمول جستجو
("electronic health record"[All Fields] OR "Health Care Information System"[All Fields] OR "Personal Health Record"[All Fields]) AND ("cloud computing"[MeSH Terms] OR ("cloud"[All Fields] AND "computing"[All Fields]) OR "cloud computing"[All Fields]) AND ("2009/05/30"[PDat] : "2019/05/27"[PDat] AND English[lang])

Step II: The data obtained from the PubMed based on the research questions, i.e. establishing the production growth trend, the most productive writers, the most productive countries, and the most frequent words were prepared for analysis. First, all the records extracted from PubMed were inputted in a specific format to VOSviewer software. Then, the table for co-occurrence of words and their maps were created. The excel software was used to trace the tables and graphs.

Step III: At this point, the maps and data were interpreted. VOSviewer is some free software used to develop and observe bibliometric maps capable of showing the maps of writers or of publications based on equally cited data or maps of keywords based on the data for the co-occurrence (Van Eck & Waltman, 2010).

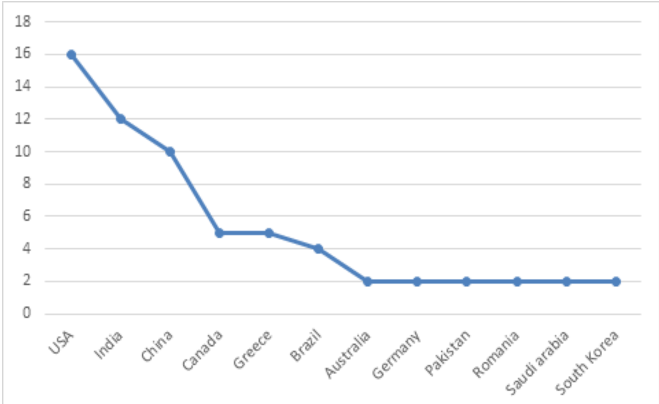
Findings

The findings (Graph 1) indicated that the growth of papers from 2009 through 2016 had an upward trend but a downward trend in 2017, and then took on an upward trend.



Graph 1. Production trend for papers on cloud in PubMed over 2009-2019

As for the leading countries in cloud health, the results indicated that the US, India, and China were most productive and Iran zero productive, respectively (Graph 2).

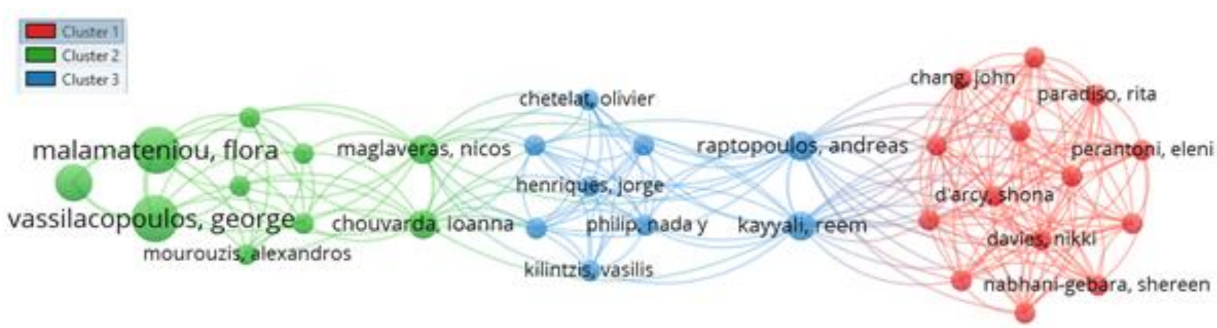


Graph 2. Frequency distribution of most productive countries in the world on cloud health in PubMed over 2009-2019

To develop a Coauthorship network in VosViewer, the minimum number of papers published by writers, one writer was taken into consideration and 2176 writers were analyzed. A cluster analysis of the network of fellow researchers indicated that the network comprised three clusters in red, blue, and green. Carlos Costa, Sungyoung Lee, Flora Malamateniou, Lacramioara Stoicu-Tivadar, George Vassilacopoulos, writers with5 papers, had the largest number of papers (Table 2). In Graph 2, the network of cooperation of the most productive writers on cloud health was traced using VosViewer in three clusters: green, red, and blue.

Table 2. Frequency of the most productive writers on cloud health in PubMed over 2009-2019

total link strength	documents	author	
26	5	costa, carlos	1
15	5	lee, sungyoung	2
15	5	malamateniou, flora	3
10	5	stoicu-tivadar, lacramioara	4
10	5	vassilacopoulos, george	5
19	4	jiang, xiaoqian	6
9	4	lopez-coronado, miguel	7
8	4	lupse, oana sorina	8
5	4	mcgregor, carolyn	9
3	4	oliveira, jose luis	10
22	3	koufi, vassiliki	11
18	3	lai, feipei	12
12	3	ohno-machado, lucila	13
7	3	ribeiro, luis s	14
6	3	serban, alexandru	15
6	3	viana-ferreira, carlos	16



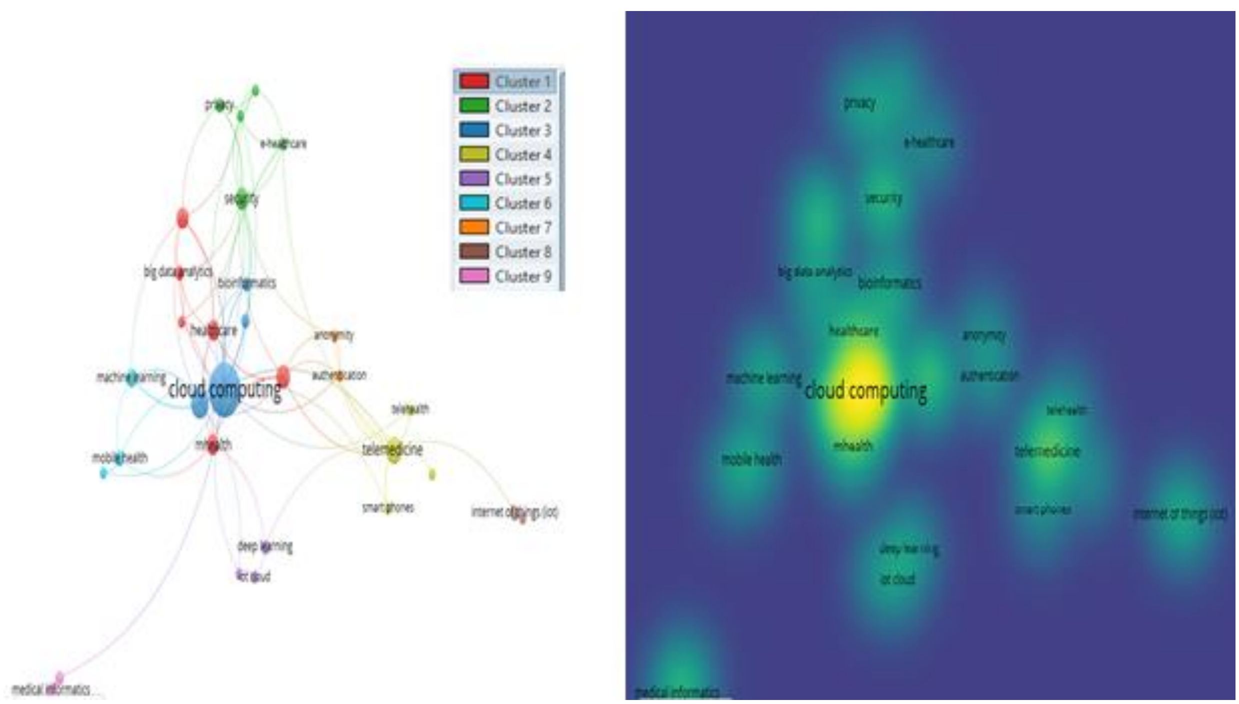
Graph 2. Coauthorship network on cloud health in PubMed over 2009-2019

To locate the important research issues, distribute the writers' keywords, and trace the maps for the co-occurrence of the words in this area, first, all the records extracted from PubMed were inputted to VosViewer. The software is capable of tracing the CoWord map based on the topic keywords, abstract, and MeSH. In the CoWord analysis, the co-occurrence threshold for keywords is determined. In this study, the minimum co-occurrence for each word in MeSH has been considered 10 times and in the topic 3 times. Obviously, small areas where the words have not achieved the co-occurrence threshold, are deleted from the results of the analysis, although the difficulty is diminished in the CoWord analysis by lowering the co-occurrence threshold and more keywords are recovered. However, the bulk of words of low value significantly increases in the map, making it difficult to understand the map. Considering the above co-occurrence threshold for the target words, the software identified 58 words in the text and 37 words in the headings of the documents under investigation, and then the CoWord map of this number of words was traced using the above software. It should be noted that the lines between the concepts on the map indicate the interrelations between the concepts, and that the keywords with more interrelations are closer to each other on the map. Moreover, the size of the circles indicates the magnitude of the existing knowledge about each concept. The colors of the circles indicate the clusters where the keywords are located. As seen in Graph 3, the greatest focus of the target texts is on the topics of cloud computing, Electronic health record, health information transfer, human, and gender with larger circles than those for other concepts.

In analyzing such maps, the location of each concept on the map will yield no specific information; rather, it is the relative distance of each concept from other concepts that specifies the relations between concepts. The distances of the concepts on the maps indicate to what extent the existing texts have expressed the relations between the two concepts and have measured their impacts on each other. If the concept of cloud computing is located close to the concept of cloud health, it means that in the existing texts, the focus of attention has been on their impacts on each other. However, when the distances of cloud computing concepts from the Internet of Things is seen as great, it means that in the existing literature, there is less attention paid to the impacts of the concepts on one another.

The most frequently used words in cloud health based on the co-occurrence of author keywords were: topics and parenthetical occurrences, cloud processing (82) Big data(21), telemedicine(19), security(9), electronic health records(8), healthcare(8), privacy(7), machine learning(6), medical informatics(5), mobile health(5) (Table 3).

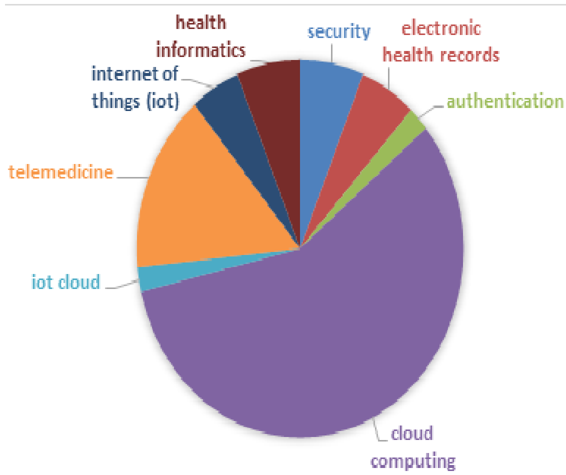
Overall, 31 items in 9 clusters specify the map for the co-occurence of the author keywords on cloud health (Graph 3). To thematically classify each cluster, in each cluster, the most frequent keyword was selected as the cluster name. The proportion that each cluster has in forming the CoWord network is shown in Graph 4. The largest cluster of cloud computing with 6 items, which, along with the Internet of Things, Electronic health record, healthcare, and e-health, forming another cluster, are suggestive of the contiguity of issues.



Graph 3. A. Co-occurrence map- B. Overall compressed map of topics based on author keywords

Table 3. The occurrence of most frequent author keywords on cloud health in PubMed over 2009-2019

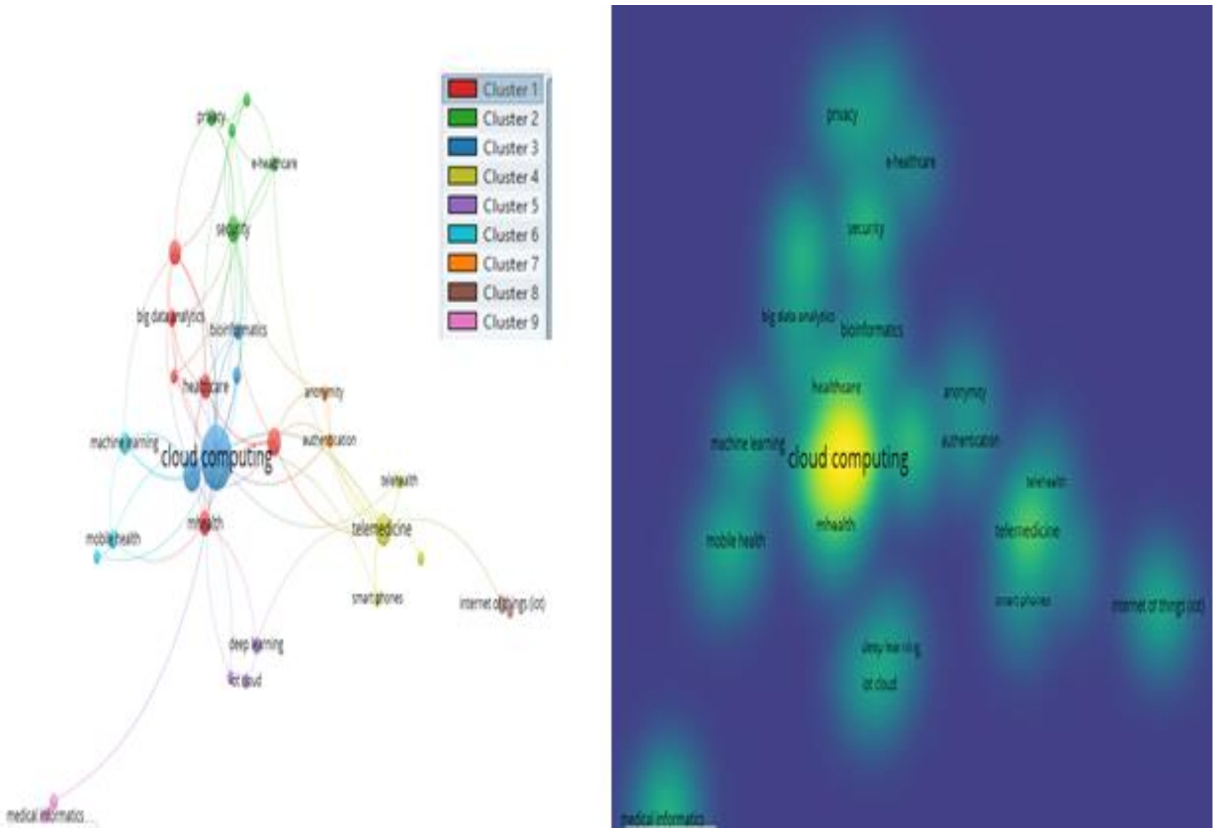
total link strength	occurrences	keyword	rank
57	82	cloud computing	1
20	21	big data	2
19	19	telemedicine	3
14	9	security	4
9	8	electronic health records	5
6	8	healthcare	6
7	8	M health	7
8	7	privacy	8
7	6	machine learning	9
1	5	medical informatics	10
5	5	mobile health	11
6	4	big data analytics	12
6	4	bioinformatics	13
3	4	deep learning	14
3	4	distributed computing	15
3	4	health informatics	16
2	4	internet of things	17
6	3	access control	18
5	3	authentication	19



Graph 4. The proportions of the clusters forming the co-occurrence network of author keywords

To determine the co-occurrence of MeSH words, the greatest concentration of the target texts on the topics with parenthesized occurrence were human (283) and cloud computing (196)

Internet(114), electronic health records(96), computer security(88), information storage and retrieval(82), algorithms(60), confidentiality(51), female(42), male(42), delivery of health care(39), computer communication networks(34), medical informatics(28), mobile applications(26), data mining(21), health information exchange(20) (Table 4). The concepts on the map enjoy a relatively good distribution, indicating that researchers in cloud health have addressed a variety of issues. The distance between the concepts indicates to what extent the interrelations of the concepts and their mutual effects in the texts have been explored. For instance, if the concepts: cloud computing, data mining, medical information, and software are placed close to one another, it means that their effects on one another have been the focus of attention in the texts. Overall, the co-occurrence map of the MeSH words in cloud health indicates 58 words in 5 clusters. The largest cluster includes 20 words and the largest keyword is "human".

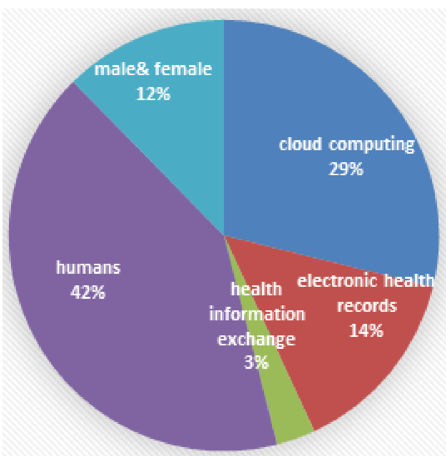


Graph 5. A. Co-occurrence map - B. Overall compressed map based on MeSH words

The graph of the co-occurrence map for the MeSH words in cloud health (Graph 5) shows 9 clusters. In this map, each of the colors used represents one cluster. To thematically classify each cluster, in each cluster, the most frequent word was selected as the cluster name. The proportion that each cluster has in forming the CoWord network is shown in Graph 6.

Table 4. The occurrence of the most frequent MeSH words on cloud health in PubMed over 2009-2019

total link strength	occurrences	keyword	rank
1290	283	humans	1
901	196	cloud computing	2
556	114	internet	3
506	97	electronic health records	4
472	88	computer security	5
388	82	information storage and retrieval	6
333	60	algorithms	7
292	51	confidentiality	8
284	42	female	9
279	42	male	10
171	39	delivery of health care	11
170	34	Computer communication etworks	12
144	28	medical informatics	13
193	27	middle aged	14
125	26	mobile applications	15
170	24	adult	16
100	21	data mining	17
100	21	databases, factual	18
116	20	health information exchange	19



Graph 6. The proportions of the clusters forming the co-occurrence network of MeSH words

Conclusion

The emergence of the concept "cloud computing" is traced back to the 1960s when John McCarthy said in a speech: "Computations may some day become a general means of organizing." (De Bruin & Floridi, 2017). The trend of the research results indicated that the growth trend of the papers in this area was upward, and then downward only in 2017. This is consistent with the studies by Sivakumaren, Swaminathan, and Karthikeyan (2015), Siddagangaiah (2017), and Hanumanth (2017) who achieved an upward trend in their results on cloud computing publications.

The United States, India, and China had the largest number of publications respectively: Sivakumaren, Swaminathan, and Karathikyan (2015) referred in their results to USA, and Voß &

Heilig (2014) referred to USA and China, and Jan, Wani, and Hafiz (2015) referred to USA as the most productive countries on cloud computing, which is consistent with the results of this study.

Costa, Lee, Malamateniou, Stoicu-Tivadar, Vassilacopoulos, had 5 papers, the largest number of publications, on cloud computing. In the studies by Swaminathan and Karathikyan (2015), Siddagangaiah (2017), Hanumanth (2017), Voß & Heilig (2014), voted for Buyya.R as the most productive writer. The findings were suggestive of the absence of Iranian writers among the most productive writers.

The results of the study by Voß & Heilig (2014) indicated that keywords, virtualization, security, computer systems, grid computing, scalability virtual machines, distributed computer systems, Hadoop, data mining, the Internet, cloud services, and cloud computing had the highest frequency. In this study, the greatest co-occurrence was that of the words Internet, electronic health records, computer security, information storage and retrieval, algorithms, confidentiality, female, male, delivery of health care, computer communication networks, medical informatics, mobile applications, data mining, health information exchange. The results of the study indicated that the words "human", "cloud computing", "Electronic health record", "telemedicine", and "Internet" are not only keywords but also have a strong link to other keywords on the co-occurrence map of words.

Overall, each word on the co-occurrence map of words can represent a concept or research area in cloud computing. In certain cases, where words have close relations are placed on a single cluster. However, occasionally, like cloud health, the combination of the words themselves form a new area.

References:

Aijing, L., & Jin, Y. (2015). Design of the Hospital Integrated Information Management System Based on Cloud Platform. The West Indian medical journal, 64(5), 521–526. https://doi.org/10.7727/wimj.2016.057

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: an introduction to co-word analysis. Social Science Information, 22(2), 191–235. Available from: http://ijlis.org/img/2017_Vol_7_Issue_3/1-7.pdf

De Bruin, B., & Floridi, L. (2017). The Ethics of Cloud Computing. Science and engineering ethics. 23(1), 21–39. https://doi.org/10.1007/s11948-016-9759-0

Hanumanth, R. (2017). Scientometric analysis of "IEEE Transactions on Cloud Computing: 2013-2016. International Journal of Library and Information Studies. 7(3). Available from: http://ijlis.org/img/2017_Vol_7_Issue_3/267-273.pdf

Heilig, L. & Voß, S. (2014). A Scientometric Analysis of Cloud Computing Literature. IEEE Transactions on Cloud Computing, 2(3), 266-278. https://doi.org/10.1109/TCC.2014.2321168

Jan, R., Wasim, R., & Hafiz, O. (2015). Scientometric Analysis of Cloud Computing. Library Philosophy and Practice (e-journal). Available from: http://digitalcommons.unl.edu/libphilprac/1273

Mayo, R., Pearson, K., Avrin, D. & Leung, J. (2017). The Economic and Social Value of an Image Exchange Network: A Case for the Cloud. Journal of the American College of Radiology, 14(1), 130–134. https://doi.org/10.1016/j.jacr.2016.07.026

Mell, P. & Grance T. (2011). The NIST Definition of Cloud Computing. National Institute of Standards and Technology. Retrieved from: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Schweitzer, E. J. (2011). Reconciliation of the cloud computing model with US federal electronic health record regulations. Journal of the American Medical Informatics Association: JAMIA, 19(2), 161–165. https://doi.org/10.1136/amiajnl-2011-000162

Siddagangaiah, K.N. (2017). Global Research Productivity in Cloud Computing: A Bibliometric Study. International Journal of Library and InformationStudies, 7(3), 1-7. http://ijlis.org/img/2017_Vol_7_Issue_3/1-7.pdf

Sivakumaren, K.S., Swaminathan, S., & Karthikeyan, G. (2015). Growth and development of publications on cloud computing: A scientometric analysis. International Journal of Information Library and Society, 1 (1), 37-43. Available from: http://www.publishingindia.com/GetBrochure.aspx?query=UERGQnJvY2h1cmVzfC8xMjY3LnBkZnwvMTI2Ny5wZGY=

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Van Eck, N. J., Waltman, L., Noyons, E. C., & Buter, R. K. (2010). Automatic term identification for bibliometric mapping. Scientometrics, 82(3), 581–596. https://doi.org/10.1007/s11192-010-0173-0