Coronavirus infection study: bibliometric analysis of publications on COVID-19 using PubMed and Dimensions databases

A.Kh. Khakimova[1,A], O.V. Zolotarev[2,B], M.A. Berberova[3,A,B,C]

A ANO «Scientific and Research Center for Information in Physics and Technique»
B ANO HE «Russian New University»
C ANO International Nuclear Safety Center

[1] ORCID: 0000-0001-9355-9249, aida_khatif@mail.ru
[2] ORCID: 0000-0001-6917-9668, ol-zolot@yandex.ru
[3] ORCID: 0000-0002-6357-7929, maria.berberova@gmail.com

Abstract

Information about the disease COVID-19 today dominates over any other and is accompanied by many publications. Bibliometrics is a modern tool for the analysis of publication activity. This article describes the results of bibliometric analysis and visualization of an array of publications [1] related to the new type and previous types of coronavirus. Some patterns have been identified. Analysis of publications in the PubMed database made it possible to isolate and cluster terms related to research on coronavirus infections. The most common keywords in PubMed are "coronavirus" and "COVID-19". Analysis of citation of publication activity during 2019-2020 on the topic of "coronavirus" based on the Dimensions database showed that although the authors of the most cited publications are Chinese doctors and researchers, these articles are published in journals in the UK and USA. A temporal analysis of citation showed that prior to the pandemic, journals published in the Netherlands and the United Kingdom were widely cited, while journals published in the United States were most actively cited during the COVID-19 pandemic. The most cited authors include specialists in the field of virology, immunology, medicine, working in Europe, Hong Kong, USA.

Keywords: COVID-19, bibliometric analysis, PubMed, Dimensions, SARS-CoV, MERS-CoV, SARS-CoV-2, publication activity.

1. Introduction

Bibliometrics has been defined as the use of statistical methods to analyze the mass of literature to identify the historical development of a field of research [2-4], as well as qualitative and quantitative research of publications. Bibliometric studies were carried out in many areas of medicine, such as ophthalmology [5], rheumatology [6], otolaryngology [7], nephrology [8], geriatrics [9], 4P-medicine [10], etc.

In foreign scientific literature, PubMed is often used for bibliometric analysis of medical publications. At the same time, bibliometric analysis considers the total (absolute) number of publications, while calculating the relative indicators, their dynamics for different periods of time, the amount of research funding. One of the areas of bibliometric analysis is content analysis, which includes identifying research trends [11-13].

In recent years, many bibliometric studies have been published in highly rated medical journals [14-18].

Chinese scientists analyzed publications about COVID-19 obtained from a PubMed search using the keyword "COVID-19". The authors reviewed 183 publications published from January 14 to February 29, 2020. The following parameters were analyzed: citizenship of researchers, place of work (hospitals, universities, research institutions), journals, types of research [19].

We conducted a bibliometric analysis of publications in the field of coronavirus research to identify the main areas of research, collaboration of scientists and organizations, and international research interaction.

2. Materials and methods

As in the previous work [18], the Dimensions [21] and PubMed databases were used to search for documents.

On the Dimensions platform, documents were searched using the keywords "coronavirus" in titles and annotations for the period of 2019-2020. We retrieved data on 2500 publications that were most relevant to the query.

A search query on the NCBI (National Center for Biotechnology Information) PubMed database included the term "coronavirus" in a search box using MeSH (Medical Subject Headings). Information was retrieved on 17,140 works available in PubMed (request date 10.04).

To visualize bibliometric networks, the VOSviewer 1.6.13 software [22] was used, which supports overlay and density visualizations [23].

In this work, we used the methods of regression analysis of data [24] to analyze the dependences of citation on various factors [25,26]. The models were built using the OLS (Ordinary Least Squares) method to estimate the possible parameters of a linear regression model, and a statistical prediction method was also used.

3. Results and discussion

3.1. Analysis of the dynamics of publication activity for a sample from the PubMed database

Analysis of the dynamics of publications from the PubMed database (Fig. 1) showed bursts of research activity timed to coincide with outbreaks of coronavirus infections. The first case of SARS-CoV virus disease was registered in 2002; an outbreak of MERS-CoV Middle East respiratory syndrome in 2015; the SARS-CoV-2 virus, which is responsible for the new type of pneumonia pandemic COVID-19, was identified in 2020 [27].

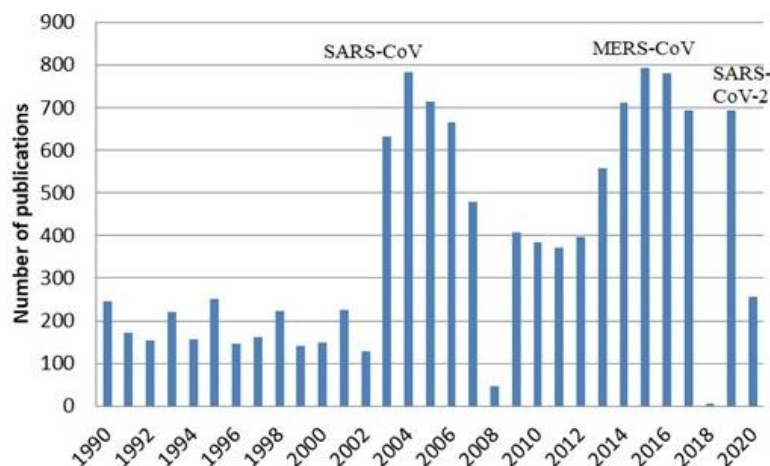Figure: 1. Dynamics of publications on the topic of "coronavirus" in the PubMed database for the period of 1990-2020

3.2. Analysis of the subject of publication activity based on a sample from the PubMed database

The study of the thematic orientation of publications from the PubMed database was carried out using the VOSviewer program to build a terminological map based on the joint occur-

rence of terms in the titles and annotations of articles. A limitation was set: the term must occur at least 15 times. There were selected 131 terms out of 2499. The terminological map was built using the author's keywords included 6 clusters, uniting 116 key concepts by thematic proximity (Fig. 2).

Figure: 2. The most frequent author's keywords (terms) in the publications related
to research in the field of "coronavirus" in the PubMed database, grouped into clusters

This map (Fig. 2) visualizes the frequency of use of author's terms and various variants of term combinations both within clusters and between clusters.

The size of the term circle reflects the number of publications in which the term was found, and the distance between the two terms gives an approximate indication of the relationship between the terms. The greater the number of publications in which both terms were found, the stronger the relationship between the terms. Clusters represent groups of terms that are relatively strongly related to each other [28].

The "red" cluster is the largest. It contains 47 terms. This cluster has received the provisional name - "Genesis and Detection of Previous Types of Coronavirus", as it combines concepts related to:

– previously known varieties of coronavirus and animal carriers (Porcine Epidemic Diarrhea Virus), PEDV, bovine coronavirus, chicken, feline coronavirus, feline infectious peritonitis, pigs, porcine delta coronavirus, porcine gastrointestinal coronavirus, porcine transmissive virus (TGEV), swine flu);

– types of analysis (immunohistochemistry, PCR, phylogenetic analysis, RT-PCR, serology, ELISA (enzyme-linked immunosorbent assay));

– vaccines and immunity (antibody, coinfection, immune response, immunity, vaccine).

This (the first) cluster also includes general medical terms (apoptosis, diagnosis, diarrhea, isolation, mortality, pathogenesis, etc.).

The second cluster ("green") unites 18 terms and can be called "MERS-CoV", since the main terms refer to this coronavirus (camel, coronaviruses, dromedary camels, Hajj, MERS, MERS-CoV, Middle East, Middle East respiratory syndrome, Saudi Arabia, viruses, zoonoses).

The third cluster ("blue") includes 16 terms and can be designated as "epidemiology and virology" (chloroquine, coronavirus infections, diseases, epidemiology, infection, infection control, infectious diseases, infectious diseases, inflammation, novel coronavirus pneumonia, outbreak, pandemic , public health, quarantine, virology).

The fourth cluster ("yellow") includes 15 terms and can be called "COVID-19" (2019 novel coronavirus, 2019-NCoV, clinical characteristics, coronavirus disease, COVID-19, novel coronavirus, pneumonia, SARS-CoV-2, etc. .).

The fifth cluster ("purple") includes 14 terms and can be called "SARS" (coronavirus, influenza, influenza virus, interferon, respiratory infection, respiratory syncytial virus, respiratory virus, SARS, SARS coronavirus, SARS-CoV, severe acute respiratory syndrome (Severe Acute Respiratory Syndrome).

The sixth cluster ("pink") "Wuhan" includes 6 terms and refers to the outbreak of coronavirus in Wuhan (2019 novel coronavirus disease, China, zoonoses, Wuhan, severe acute respiratory syndrome coronavirus 2, respiratory infections).

VOSviewer also can display the time of occurrence of the most frequently used terms in research. The closer to blue, the older the research, the closer to yellow, the more modern the research. The results of the visual display of the temporal updating of the terms reflects the sequence of the development of epidemiological situations in connection with the varieties of coronavirus (Fig. 3). The cluster under the code name "Genesis and Detection of Previous Types of Coronavirus" was formed by 2016. Then were formed the clusters: "MERS-CoV", "SARS", "Epidemiology and Virology". The most recent emerging clusters were COVID-19 and Wuhan.

Figure: 3. The most common terms in publications related to research in the field of "coronavirus" on the PubMed database, time of appearance

VOSviewer can visualize keyword density (Fig. 4). The color of each node in the keyword density renderer depends on the density of the elements in that node. In other words, the color of a node depends on the number of elements in the vicinity of the node. Keywords in the yellow area appear more frequently; instead, keywords in the green area appear less frequently. It can be seen that the most frequently used keywords are "coronavirus" and "covid-19".

Figure: 4. The most common terms used in the field of "coronavirus" on the PubMed base.
(VOSviewer density visualization)

3.3. Visualization of citation networks and characteristics of the most cited articles based on a sample from the Dimensions database

To analyze citation and co-authorship networks, we used a selection of scientific papers from the Dimensions database.

To identify the most cited publications, a restriction was set: the publication must be cited at least 5 times. Out of 2500 articles, 493 were selected. As a result, 28 clusters were identified (Fig. 5), uniting 420 publications (the rest of the publications was not included in the clusters).

Figure: 5. The most cited publications on the topic of "coronavirus"
for 2019-2020 on the Dimensions database

The large number of clusters shows a variety of research areas. Large nodes represent influential publications. With the Fig. 5, the most cited articles are much superior to others in importance; for the subsequent analysis, we selected publications with more than 100 citations (there were 25 of them). We presented these publications in the form of density visualization (Fig. 6).
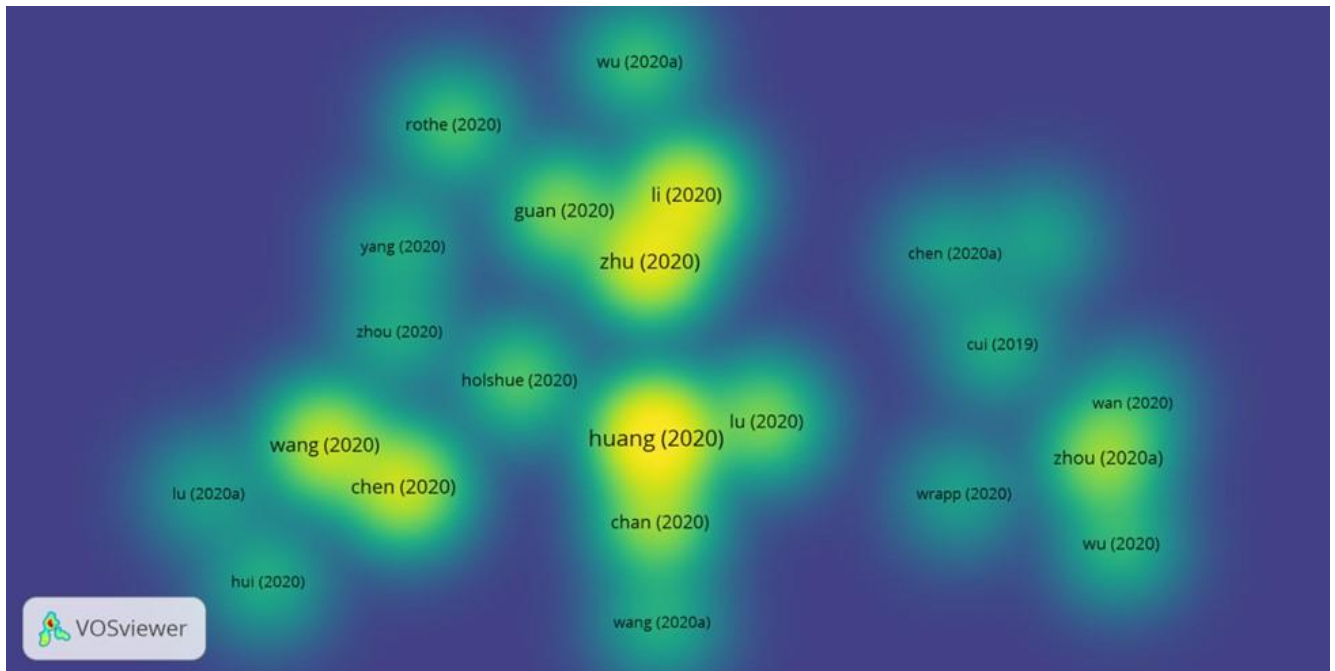
Figure: 6. The most frequently cited publications in the field of "coronavirus" for 2019-2020 (Dimensions Database Density Visualization)

Figure 6 demonstrates that researchers more often refer to the following works: "Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China" [29]; "A Novel Coronavirus from Patients with Pneumonia in China, 2019" [30]; "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus – Infected Pneumonia" [31]; Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: a Descriptive Study [32]; Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus – Infected Pneumonia in Wuhan, China [33].

Authors of the article [29] represent a large number of specialized medical institutions in China, including hospitals in Wuhan (The Central Hospital of Wuhan, Zhongnan Hospital of Wuhan University, Yin-tan Hospital), medical universities and colleges in China (Capital Medical University, Peking University Joint Center for Life Sciences, Peking Union Medical College Tsinghua University School of Medicine (Beijing); Huazhong University of Science and Technology (Wuhan); hospitals (China-Japan Friendship Hospital, Beijing Ditan Hospital, Peking University First Hospital, Peking University People's Hospital (Beijing)); research centers (Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Clinical and Research Center of Infectious Diseases etc.).

The article reports on the epidemiological, clinical, laboratory and radiological characteristics, treatment, and clinical outcomes of 41 patients with laboratory-confirmed 2019-nCoV infection hospitalized in Wuhan by January 2, 2020. Researchers communicated directly with patients or their families to clarify epidemiological data and data about the symptoms [29].

The authors of the article "A Novel Coronavirus from Patients with Pneumonia in China, 2019" [30] present: Beijing (NHC Key Laboratory of Biosafety, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, and the Department of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University); Wuhan (Wuhan Jinyintan Hospital, the Division for Viral Disease Detection, Hubei Provincial Center for Disease Control and Prevention, the Center for Biosafety Mega-Science, Chinese Academy of Sciences); Jinan (the Shandong First Medical University, Shandong Academy of Medical Sciences).

The article reports on a novel coronavirus that was detected in hospitalized patients in Wuhan, China in December 2019 and January 2020.The beta coronavirus was detected by unbiased sequencing in samples from patients with pneumonia. Epithelial cells of the human respiratory tract were used to isolate a new coronavirus named 2019-nCoV [30].

The authors of the article "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus – Infected Pneumonia" [31] represent different provinces of China: Hubei (Hubei Provincial Center for Disease Control and Prevention, Jingzhou Center for Disease Control and Prevention); Sichuan; Hunan; Henan; Inner Mongolia; Liaoning; Guizhou; Jiangxi. The authors also represent Beijing, Shanghai, Hong Kong.

The article analyzed data on the first 425 confirmed cases of new coronavirus infection in Wuhan as of January 22, 2020, to determine the epidemiological characteristics of the disease. The authors described the characteristics and estimated the epidemic doubling time and baseline reproductive number. The authors have shown that human-to-human transmission has occurred among close contacts since mid-December 2019 [31].

The authors of the article [32] mainly represent Wuhan institutions (different departments of Wuhan Jinyintan Hospital, State Key Laboratory of Virology, State Key Laboratory of Virology, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences); as well as Shanghai (Ruijin Hospital, etc.).

The authors sought to clarify the epidemiological and clinical characteristics of 2019-nCoV pneumonia. In this retrospective study, they included all confirmed cases of 2019-nCoV at Wuhan Jinyintan Hospital from January 1 to January 20, 2020. Epidemiological, demographic, clinical and radiological features and laboratory data were analyzed. Of the 99 patients with 2019-nCoV pneumonia, 49% had a history of eating seafood from the Huanan market. The authors suggested that 2019-nCoV infection is more likely to affect older men with underlying medical conditions and can lead to severe and even fatal respiratory diseases such as acute respiratory distress syndrome [32].

 All authors of the article "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China" [28] represent Zhongnan Hospital of Wuhan University, Wuhan, Hubei.

The authors describe the epidemiological and clinical characteristics of Novel Coronavirus-Infected Pneumonia (NCIP) for a series of 138 consecutive hospitalized patients with confirmed NCIP at Zhongnan Hospital of Wuhan University, from January 1 to 28, 2020. Hospital transmission has been suspected as a putative transmission mechanism for those infected medical workers and hospitalized patients. Coronavirus transmission, presumably related to hospitalization, was suspected in 41% of patients [33].

3.4. The most influential (cited) journals - identification based on a sample from the Dimensions database

To build a citation network by sources, we applied a restriction so that the source had at least 7 publications. 74 out of 743 sources satisfied the condition. 70 sources were identified in 3 clusters (Fig. 7).
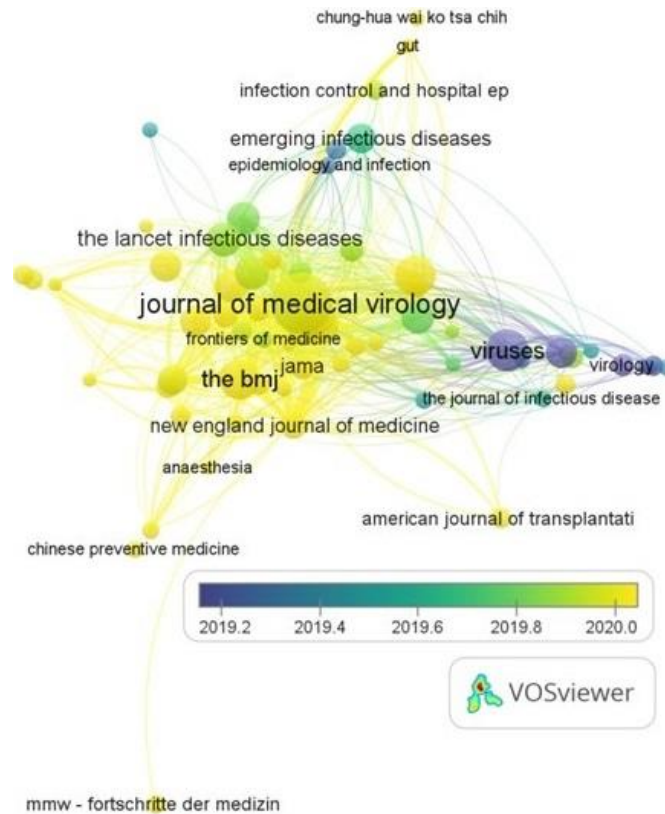
Figure: 7. The most frequently cited sources in the field of "coronavirus"
for 2019-2020 on the Dimensions Database

The clustering results are shown in Table 1.

Table 1 - The most frequently cited sources in the field of "coronavirus" for 2019-2020 on the Dimensions Database

№	Source	Number of publications	Number of citations	Norm. number of citations	Country	Impact Factor
			1 кластер			
1	American Journal of Roentgenology	13	42	5.7751	USA	3.1610
2	American Journal of Transplantation	13	11	1.5125	USA	6.4930
3	British Journal of Anaesthesia	7	2	0.275	United Kingdom	6.1990
4	BioScience Trends	9	151	20.7629	Japan	1.6860
5	Chinese Journal of Contemporary Pediatrics	12	10	1.375	China	0.1900
6	Chinese Journal of Epidemiology	7	118	16.2253	China	0.9048
7	Chinese Journal of Pediatrics	13	22	3.0251	China	0.2100
8	Chinese Journal of Preventive Medicine	9	4	0.55	China	0.7737
9	European Journal of Nuclear Medicine and Molecular Imaging	7	26	3.5751	Germany	7.2770
10	Frontiers of Medicine	9	26	4.3277	China	1.8470
11	International Journal of Antimicrobial Agents	15	122	16.7753	Netherlands	1.5310
12	International Journal of Biological Sciences	11	9	1.2375	Canada	4.0670
13	International Journal of Environmental Research and Public Health	15	15	2.4388	Switzerland	2.4680
14	Journal of the American Medical Association (JAMA)	20	1073	147.5402	USA	51.2730
15	JMIR Public Health and Surveillance	12	0	0	Canada	2.3900

№	Source	Number of publications	Number of citations	Norm. number of citations	Country	Impact Factor
16	Canadian Journal of Anesthesia / Journal canadien d'anesthésie	10	51	7.0126	Canada	2.1270
17	Journal of Clinical Medicine	19	178	24.4754	Switzerland	5.6880
18	Journal of Korean Medical Science	17	75	10.3127	South Korea	1.7160
19	Journal of Medical Virology	108	731	101.5178	USA	2.0490
20	Korean Journal of Radiology	21	29	3.9876	South Korea	3.7300
21	MMW Fortschritte der Medizin	3	0	0	Germany	0.0300
22	Morbidity and Mortality Weekly Report (MMWR)	8	70	9.6252	USA	14.8740
23	The New England Journal of Medicine	21	2386	328.08100	USA	70.6700
24	PLoS ONE	10	10	2.2531	USA	2.7760
25	Radiology	26	593	81.539	USA	7.6080
26	British Medical Journal (BMJ)	50	139	19.1128	United Kingdom	27.6040
27	The Lancet	51	3275	450.3207	United Kingdom	59.1020
28	The Lancet Respiratory Medicine	12	179	24.6129	United Kingdom	22.9920
29	World Journal of Pediatrics	7	63	8.6627	Germany	1.1690
30	中华结核和呼吸杂志	33	40	5.5001	China	2.0360
2 кластер						
31	Antiviral Research	8	44	7.5553	Netherlands	4.1300
32	Archives of Virology	8	4	1.0518	Germany	2.1340
33	Communicable diseases intelligence (Comm Dis Intell)	8	7	0.9625	Australia	1.0000
34	Emerging Microbes & Infections	26	71	12.5223	United Kingdom	6.2120
35	Eurosurveillance	35	436	60.3274	France	7.4210
36	Frontiers in Microbiology	8	43	11.3064	Switzerland	4.2590
37	Infection, Genetics and Evolution	8	57	9.8446	Netherlands	2.6110
38	Journal of Microbiology and Biotechnology	8	8	2.3664	Netherlands	1.9750
39	Journal of Microbiology, Immunology and Infection	13	42	8.2838	Taiwan	2.4550
40	Journal of Virology	30	201	38.551	USA	4.3240
41	Methods in molecular biology	12	3	0.4125	USA	10.7100
42	Nature	48	606	83.3265	United Kingdom	43.0700
43	Pathogens	8	11	1.638	Switzerland	3.4050
44	Poultry Science	8	5	1.3147	USA	2.0270
45	Proceedings of the National Academy of Sciences of the United States of America	7	64	11.6852	USA	9.5800
46	Science	11	210	28.8755	USA	16.8100
47	Journal of Infectious Diseases	10	44	8.4334	Netherlands	5.0450
48	Transboundary and Emerging Diseases (Transbound Emerg Dis)	13	44	11.5693	Germany	1.9500
49	Veterinary Microbiology	32	96	18.8449	Netherlands	2.7910
50	Veterinary Record	8	12	3.0298	United Kingdom	2.0500
51	Virology	15	3	0.7888	Germany	2.6570
52	Virus Research	15	21	5.3963	Netherlands	2.7360
53	Viruses	9	10	2.0022	Switzerland	3.8110
3 кластер						
54	Chinese Medical Journal	29	74	10.1752	China	1.0530
55	Chung-Hua Wai Ko Tsa Chin (Chinese Journal of Surgery)	8	2	0.275	China	
56	Clinical Infectious Diseases	37	193	28.1687	United Kingdom	9.1170
57	Emerging Infectious Diseases	32	279	44.3841	USA	7.4220
58	Epidemiology and Infection	10	33	7.4226	United Kingdom	2.0750
59	Gut	7	7	0.9625	United Kingdom	17.9430

№	Source	Number of publications	Number of citations	Norm. number of citations	Country	Impact Factor
60	Infection Control & Hospital Epidemiology	12	19	3.8669	United Kingdom	3.0840
61	International Journal of Infectious Diseases	33	329	45.3638	Netherlands	3.5380
62	Journal of Clinical Microbiology	6	13	1.7875	USA	4.9590
63	Journal of Hospital Infection	16	99	16.4978	United Kingdom	3.7040
64	Journal of Infection	30	117	16.0878	Italy	5.0990
65	Journal of Infection and Public Health	12	14	3.1794	Netherlands	2.4870
66	Journal of Travel Medicine (J Travel Med)	13	171	23.5129	United Kingdom	4.1550
67	Microbes and Infection	12	3	0.4125	Netherlands	2.6690
68	The Lancet Infectious Diseases	7	53	7.2876	United Kingdom	27.5160
69	Travel Medicine and Infectious Disease	3	8	0.6634	Netherlands	4.8680

The parameters given in the table indicate the number of cited publications of the source, the number of citations of the publications of the source (normal and normalized), the country of origin of the source, and the impact factor of the journal. The normalized document citations are calculated as the ratio of the number of citations per document to the average number of citations for all documents published in the same year and included in the data provided by VOSviewer. The normalization is intended to correct the fact that older publications took longer to receive citations than more recent publications.

As seen from Fig. 7, the first cluster includes the most recent cited sources (late 2019 - early 2020), the third - those cited in the mid-2019, and the second - those cited at the beginning of 2019.

Fig. 8 shows that the first cluster is dominated by journals published in the USA, in the second cluster - in the USA and the Netherlands, in the third cluster - in the UK and in the Netherlands.
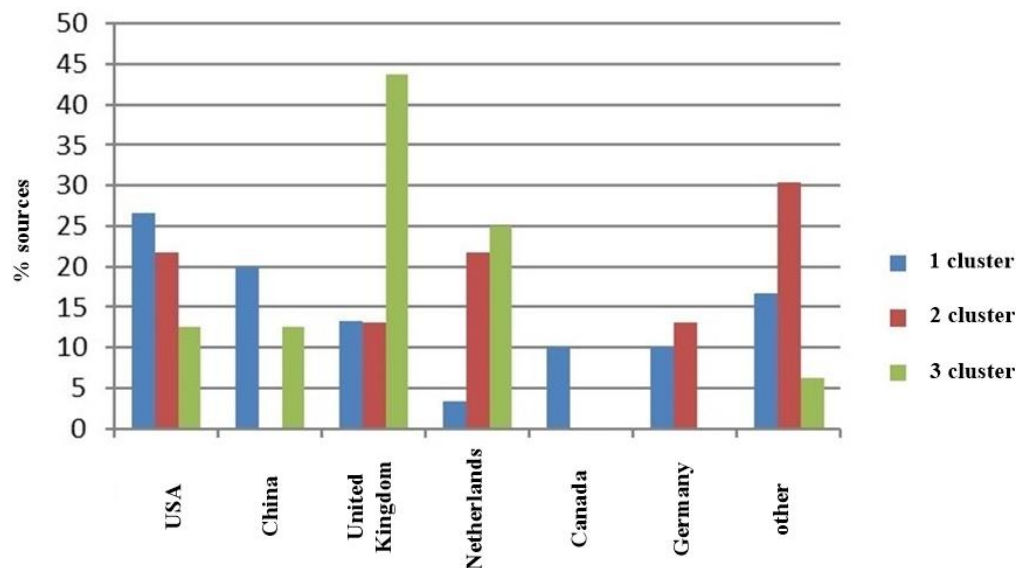
Figure: 8. Number of sources by cluster (in%), considering the country publishing the journal

The average impact factor of journals in the 1st cluster is 10.48, the second - 6.22, and the third - 6.65.

We excluded sources with a normalized number of citations less than 10 from further consideration.

The absolute leader in citation is The Lancet (450,3207). In second place is The New England Journal of Medicine (328,081). In third place is the Journal of the American Medical Association (JAMA) (147.5402). In fourth place is the Journal of Medical Virology (101.5178). The rest of the sources can be divided into groups (sources are listed in descending order):

– from 100 to 50 (Nature, Radiology, Eurosurveillance);
– from 50 to 30 (International Journal of Infectious Diseases, Emerging Infectious Diseases, Journal of Virology);
– from 30 to 20 (Science, Clinical Infectious Diseases, The Lancet Respiratory Medicine, Journal of Clinical Medicine, Journal of Travel Medicine, BioScience Trends);
– from 20 to 10 (British Medical Journal, Veterinary Microbiology, Chinese Journal of Epidemiology, International Journal of Antimicrobial Agents, Journal of Hospital Infection, Journal of Infection, Emerging Microbes & Infections, Proceedings of the National Academy of Sciences of the United States of America, Transboundary and Emerging Diseases, Frontiers in Microbiology, Journal of Korean Medical Science, Chinese Medical Journal).

There are 8 of the 28 highly cited journals published in the USA, 8 - in the UK, 3 - in the Netherlands, 1 - in France, 2 - in Switzerland, 1 - in Japan, 2 - in China, 1 - in Italy, 1 - in Germany, 1 - in South Korea.

3.5. The most influential (cited) authors - identification based on a sample from the Dimensions database

To find the most cited authors, we applied the search term: the author must have at least 10 publications. There were 16 such authors out of 11035. These results were presented in the form of density visualization (Fig. 9). Authors in yellow areas are cited more often; on the contrary, authors in green areas are cited less frequently.
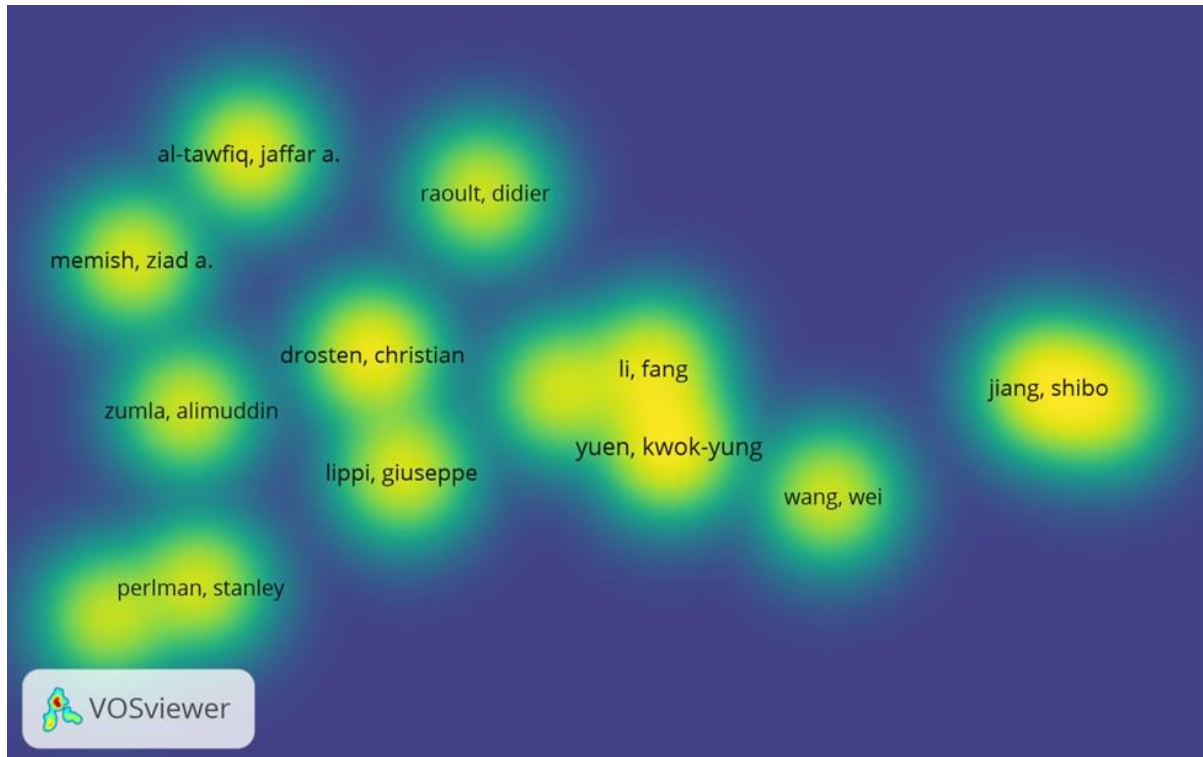
Figure: 9. The most frequently cited authors in the field of "coronavirus" for 2019-2020 on the Dimensions Database (Density visualization)

The most cited authors are Christian Drosten (German virologist specializing in coronaviruses; head of the Berlin Charité Institute of Virology), Kwok-Yung Yuen (Hong Kong microbiologist, physician and surgeon), Alimuddin Zumla (British Zambian professor of infectious diseases and international health at the Medical University of London University College), Stanley Perlman (Professor of Microbiology and Immunology, Professor of Pediatrics, University of Iowa), Lanying Du (Head of Viral Immunology Laboratory, Lindsley F. Kimball Research Institute), Shibo Jiang (Shanghai School of Medicine, Fudan University in Shanghai, China), Fang Li (Associate Professor, Department of Veterinary and Biomedical Sciences, University of Minnesota), Ziad A. Memish (Undersecretary of Public Health in Saudi Arabia), Jaffar A. Al-Tawfiq (Adjunct Professor of Medicine, Indiana University School of Medicine).

3.6. Multivariate citation analysis based on Dimensions

For the analysis, the following parameters were identified (Table 1): *PubNum* - the number of publications, *CiteNum* - the number of citations, *Country* - the country of publication, *ImpactFactor* - the impact factor of the publication. All of the three analyzed clusters were combined. As a result of the study, 3 models were built, and a forecast of publication activity was made, depending on many factors.

A model 1 was built. OLS method was used for the analysis. The analysis was carried out by country, considering countries, the number of citations, and the impact factors of journals. Dependent variable is *PubNum*, regressors: *CiteNum*, *Country*, *ImpactFactor*. The following regression equation was built:

$$PubNum = 18.3 + 0.02 * CiteNum - 0.47 * Country - 0.35 * ImpactFactor \qquad (1)$$

The most significant variable in this equation is the *CiteNum* variable. It is significant even at the 1% significance level. The highest P-value (significance level, extreme value of the statistic) is obtained for the *Country* variable. As a result, it is difficult to distinguish how significant indicators are on the number of publications by country.

Then a model 2 was built. In this model, the number of citations, *CiteNum*, was chosen as the dependent variable; the analysis was carried out taking into account countries, impact factors of journals, and the number of publications. *Country* is significant at 10% significance level, *ImpactFact* and *PubNum* are significant at 1% significance level. As a result, the following regression equation was obtained:

$$CiteNum = -234 + 12.52 * Country + 28.94 * ImpactFact + 7.67 * PubNum \qquad (2)$$

It can be concluded that the number of citations depends on the country of publication, the impact factor of the journal and the number of publications. At the same time, it is the impact factor of the journal that influences the number of citations to a greater extent, to a lesser extent - the country of publication, and even to a lesser extent - the number of publications in the journals of this country.

And then a model 3 was built. By removing the *Country* regressor (as insignificant), the following regression equation was obtained for the model:

$$CiteNum = -170 + 28 * ImpactFact + 7.56 * PubNum \qquad (3)$$

The country of publication matters, but not as significant as the impact factor of the journal. Based on the results of the model of the dependence of the number of citations on the selected regressors, a forecast for the dependent variable *CiteNum* was built.
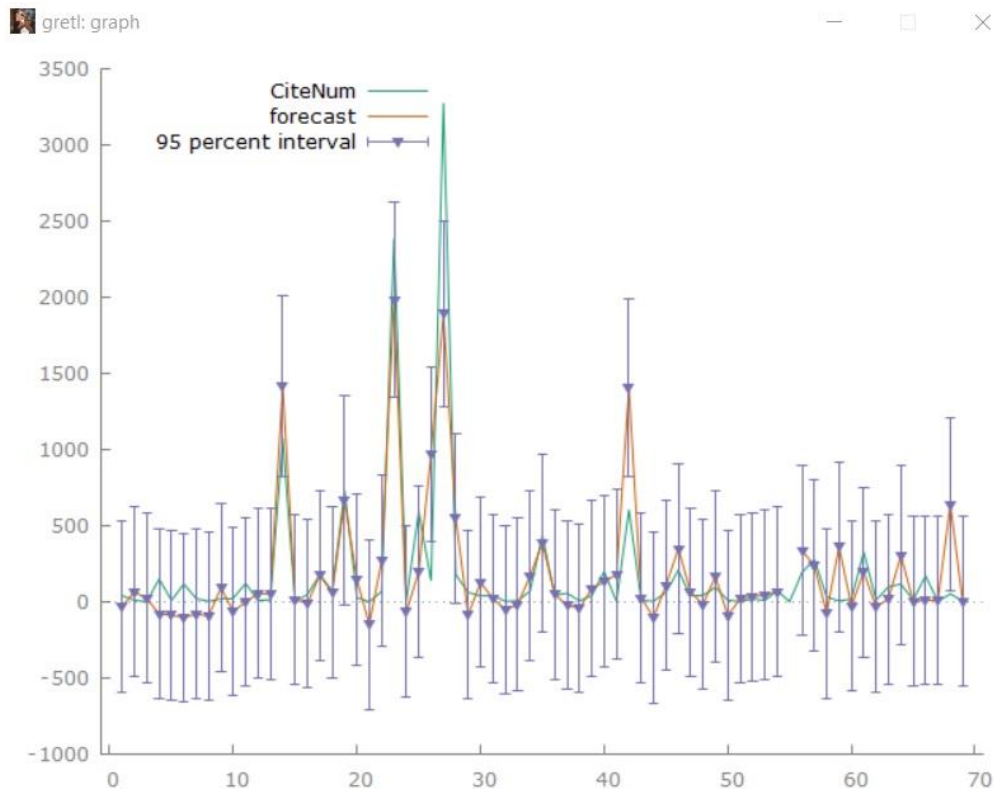
Figure: 10. Citation forecast depending on the impact factor of the publication

The figure shows the actual values of the number of citations and the citation forecast. The figure shows that, depending on several factors, the predicted values are somewhat smoothed. Most of the publications is with an impact factor in the range from 15 to 30. The result is not very much dependent on the country of publication and the number of publications.

The values of the predictive function are combined on a single graph with the most significant actual parameters, which made it possible to determine the most preferable values of the "impact factor" parameter to maximize the number of citations.

4. Conclusion

This study performed bibliometric analysis and visualization of scientific publications related to the novel and previous coronavirus types. The results obtained can be summarized as follows.

Obvious bursts of publication activity in the biomedical literature are associated with epidemics of the previously known coronaviruses SARS-CoV, MERS-CoV and the new type of SARS-CoV-2.

Clustering of author's keywords of publications in the PubMed database showed that by now a large cluster of terms has been formed, which can be conventionally called "Genesis and detection of previous types of coronavirus". The cluster includes specific concepts related to animal carriers of different types of coronavirus; terms denoting modern methods of analysis in the field of virology; terms related to vaccines, viruses, and immunity.

Clusters related to MERS-CoV and SARS have been formed. Currently, clusters are being formed, conventionally called by us "COVID-19" and "Wuhan". The first of them includes terms close to clinical manifestations. The second is more geographic. The most common keywords in PubMed are "coronavirus" and "covid-19".

We conducted a citation analysis based on publication activity during 2019-2020 on the subject of "coronavirus" based on the Dimensions database.

Of the most cited 25 articles, 8 were published in The Lancet (UK), 5 in the New England Journal Of Medicine (USA), and 2 in Nature (UK). One highly cited article was published by: JAMA (USA), International Journal of Infectious Diseases (Netherlands), Journal of Virology (USA), Nature Reviews Microbiology (UK), Eurosurveillance (France), Science (USA), The Lancet Respiratory Medicine UK) , Journal of Medical Virology (USA).

Consequently, 12 of the 25 most cited publications are published in journals published in the UK, 9 in the USA. Although, as further analysis showed, the authors of the most cited publications are Chinese doctors and researchers.

The authors of the most cited articles represent a large number of specialized medical institutions in China, including hospitals and research centers in Wuhan, medical universities and scientific organizations in various provinces of China, as well as centers in Beijing, Shanghai, and Hong Kong.

The articles report:
– on the epidemiological, clinical, laboratory and radiological characteristics and treatment and clinical results of 41 patients with laboratory-confirmed 2019-nCoV infection hospitalized in Wuhan by January 2, 2020 [29];
– on the isolation of a new coronavirus from the epithelial cells of the respiratory tract of hospitalized patients in Wuhan, China [30];
– on evidence of human-to-human transmission of the virus [31];
– on the clarification of the epidemiological and clinical characteristics and assessment of the predisposition of populations to pneumonia 2019-nCoV [32];
– on the epidemiological and clinical characteristics of the disease and the likelihood of nosocomial transmission [33-35].

The most cited journals on the subject of "coronavirus" are (listed in descending order of citation):
– The Lancet (which published 8 of the 25 most cited articles);
– The New England Journal of Medicine (which published 5 of the 25 most cited articles);
– Journal of the American Medical Association (JAMA) (which published 1 article out of 25 most cited).

The top 10 cited journals also include: Journal of Medical Virology; Nature; Radiology; Eurosurveillance; International Journal of Infectious Diseases; Emerging Infectious Diseases; Journal of Virology; Science; Clinical Infectious Diseases; The Lancet Respiratory Medicine.

A temporal analysis of citation showed that during the development of the coronavirus pandemic, journals published in the United States were most actively cited. In 2019, magazines published in the Netherlands and the United Kingdom were widely cited before the pandemic developed.

The most cited authors include specialists in the field of virology, immunology, medicine, working in Europe, Hong Kong, USA. So, the most cited author is a German virologist specializing in coronaviruses, Christian Drosten. The top 10 cited authors include only one researcher from China (Shibo Jiang, Shanghai Medical School, Fudan University in Shanghai). It is worth noting that the top 10 authors include Deputy Minister of Health of Saudi Arabia Ziad A. Memish.

In the future, we plan to develop a methodology for visualizing scientific data using the following methods and programs:
– the author's method for calculating the semantic similarity of documents,
– "t-SNE" method for calculating 3D coordinates of terms using a similarity matrix,
– automatic programs for detecting semantic similarities between terms and creating lines between spheres,

– WebVR methods for 3D visualization using calculated data.

We used a similar approach to visualize the three-dimensional cyberspace of scientific works [8].

Multivariate spatial analysis of publications was carried out. When constructing a forecast, a non-standard approach was used, which is based not on time series, as is usually the case in forecasting, but on the estimation of interval maximization of parameters. A forecast of the most probable maximum citation values was constructed depending on several factors, including the most significant one - the impact factor of the journal, as it turned out.

Acknowledgments

References

1. Michael Charnine, Konstantin Kuznetsov and Oleg Zolotarev. Multilingual Semantic Cyberspace of Scientific Papers Based on WebVR Technology. Proceedings of the International 2018 Conference on Cyberworlds. Singapore, 3-5 October.2018. P. 435-438.

2. Young H: Glossary of Library and Information Science. 1983, Chicago: American Library Association Google Scholar. Indian J Ophthalmol. 2015 Jan;63(1):54-8. DOI: 10.4103/0301-4738.151471.

3. Zolotarev O.V. Methods and domain modeling tools. Proceedings of the conference «The Civilization of Knowledge: Problems of Social Communications» - Moscow: RosNOU, 2012. p. 71-72.

4. Klimenko S., Khakimova A., Charnine M., Zolotarev O., Merkureva N. Semantic approach to visualization of research front of scientific papers using web-based 3D graphic. Proceedings of the 2018 International Conference Web 3D. The 23rd International Proceedings - Web3D 2018: 23rd International ACM Conference on 3D Web Technology 23, 3D Everywhere. 2018. C. 20.

5. Mansour AM, Mollayess GE, Habib R, Arabi A, Medawar WA. Bibliometric trends in ophthalmology 1997-2009. Semin Arthritis Rheum. 2017 Jun; 46(6): 828-833. DOI: 10.1016/j.semarthrit.2016.12.002.

6. Redondo M, Leon L, Povedano FJ, Abasolo L, Perez-Nieto MA, López-Muñoz F.A bibliometric study of the scientific publications on patient-reported outcomes in rheumatology. Clin Otolaryngol. 2017 Dec;42(6):1338-1342. DOI: 10.1111/coa.12910.

7. Saunders TFC, Rymer BC, McNamara KJ. A global bibliometric analysis of otolaryngology: Head and neck surgery literature. G Ital Nefrol. 2016 Nov-Dec;33(6). pii: gin/33.6.10.

8. Torrisi AM, Granata A. Bibliometric indicators of nephrology journals: strengths and weaknesses. [Article in Italian] Geriatr Gerontol Int. 2017 Feb;17(2):357-360. DOI: 10.1111/ggi.12880.

9. Ang HM, Kwan YH. Bibliometric analysis of journals in the field of geriatrics and gerontology. J Neuropsychiatry Clin Neurosci. 2015 Fall;27(4):354-61. DOI: 10.1176/appi.neuropsych.15010024.

10. Khakimova, A., Yang, X., Zolotarev, O., Berberova, M., Charnine, M. Tracking knowledge evolution based on the terminology dynamics in 4p-medicine. International Journal of Environmental Research and Public Health, 2020, 17(20), c. 1-19, 7444.

11. Huffman, M. D., Baldridge, A., Bloomfield, G. S., Colantonio, L. D., Prabhakaran, P., Ajay, V. S., . . . Prabhakaran, D. (2013). Global Cardiovascular Research Output, Citations, and Collaborations: A Time-Trend, Bibliometric Analysis (1999-2008). Plos One, 8(12), 7. DOI: 10.1371/journal.pone.0083440

12. Menendez-Manjon, A., Moldenhauer, K., Wagener, P., & Barcikowski, S. (2011). Nano-energy research trends: bibliometrical analysis of nanotechnology research in the energy sector. Journal of Nanoparticle Research, 13(9), 3911-3922. DOI: 10.1007/s11051-011-0344-9

13. Sooryamoorthy, R. (2010). Medical research in South Africa: a scientometric analysis of trends, patterns, produc-tivity and partnership. Scientometrics, 84(3), 863-885. DOI: 10.1007/s11192-010-0169-9

14. Aggarwal A., Lewison G., Idir S., Peters M., Aldige C., Boerckel W., Boyle P., Trimble E.L., Roe P., Sethi T+2 more. 2016. The state of lung cancer research: a global analysis. Journal of Thoracic Oncology 11:1040-1050

15. Almeida-Guerrero A., Olaya-Gomez J.C., Sanchez-Ramirez N., Murillo-Garcia D.R., Cardona-Ospina J.A., Lagos-Grisales G.J., Rodriguez-Morales A.J. 2018. Mitigation of the global impact of Lassa fever: have we investigated enough about this Arenavirus? - a bibliometric analysis of Lassa Fever research. Travel Medicine and Infectious Disease

16. Baek S., Yoon D.Y., Lim K.J., Cho Y.K., Seo Y.L., Yun E.J. 2018. The most downloaded and most cited articles in radiology journals: a comparative bibliometric analysis. European Radiology 28:1-7

17. Bruggmann D., Pulch K., Klingelhofer D., Pearce C.L., Groneberg D.A. 2017. Ovarian cancer: density equalizing mapping of the global research architecture. International Journal of Health Geographics 16 Article 3

18. Khan N.R., Saad H., Oravec C.S., Norrdahl S.P., Fraser B., Wallace D., Lillard J.C., Motiwala M., Nguyen V.N., Lee SL+10 more. 2018. An analysis of publication productivity during residency for 1506 neurosurgical residents and 117 residency departments in North America. Neurosurgery 83:217-227

19. J. Lou, S.-J. Tian, S.-M. Niu, X.-Q. Kang, H.-X. Lian, L.-X. Zhang, J.-J. Zhang. Coronavirus disease 2019: a bibliometric analysis and review. Eur Rev Med Pharmacol Sci. Year: 2020. Vol. 24 - N. 6. Pages: 3411-3421. DOI: 10.26355/eurrev_202003_20712.

20. A.Kh. Khakimova, O.V. Zolotarev, M.A. Berberova. Visualization of bibliometric networks of scientific publications on the study of the human factor in the operation of nuclear power plants based on the bibliographic database Dimensions. Scientific Visualization, 2020, volume 12, number 2, pages 127 – 138. DOI: 10.26583/sv.12.2.10, E-ISSN:2079-3537.

21. Hook DW, Porter SJ, Herzog C. Dimensions: building context for search and evaluation. Front Res Metr Anal. 2018 Aug 23;3:23. DOI: 10.3389/frma.2018.00023.

22. Van Eck NJ, Waltman L. Visualizing bibliometric networks. In: Ding Y, Rousseau R, Wolfram D, editors. Measuring scholarly impact: Methods and practice. Berlin: Springer; 2014.

23. Waltman L, Van Eck NJ, Noyons ECM. A unified approach to mapping and clustering of bibliometric networks. Journal of Informetrics. 2010;4(4):629–635. DOI: 10.1016/j.joi.2010.07.002.

24. J.Johnston, J.DiNardo, Econometrics Methods, 4th edition, McGraw-Hill, 1997.

25. Zolotarev, O., Solomentsev, Y., Khakimova, A., Charnine, M. Identification of semantic patterns in full-text documents using neural network methods. CEUR Workshop Proceedings, 2019, 2485, c. 276-279.

26. Irina V. Galina1, Michael M. Charnine, Nikolai V. Somin, Vladimir G. Nikolaev, Yulia I. Morozova, Oleg V. Zolotarev. Method for Generating Subject Area Associative Portraits: different Examples. Proceedings of the 2015 International Conference on Artificial Intelligence (ICAI 2015), vol.I, WORLDCOMP'15, July 27-30, 2015. Las Vegas Nevada, USA, v.I, pp.288-294.

27. https://en.wikipedia.org/wiki/Coronavirus

28. Van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics. 2017;111(2):1053–1070. DOI:10.1007/s11192-017-2300-7

29. Huang, Chaolin, Wang, Yeming, Li, Xingwang, Ren, Lili, Zhao, Jianping et al. (2020). Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. The Lancet, 395(10223), 497-506.

30. Zhu, Na; Zhang, Dingyu; Wang, Wenling; Li, Xingwang; Yang, Bo et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. (2020). New England Journal of Medicine, 382(8), 727-733.

31. Li, Qun; Guan, Xuhua; Wu, Peng; Wang, Xiaoye; Zhou, Lei et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia (2020). New England Journal of Medicine, 382(13), 1199-1207.

32. Chen, Nanshan; Zhou, Min; Dong, Xuan; Qu, Jieming; Gong, Fengyun et al. Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: a Descriptive Study. (2020). The Lancet, 395(10223), 507-513.

33. Wang, Dawei; Hu, Bo; Hu, Chang; Zhu, Fangfang; Liu, Xing et al. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China (2020). JAMA, 323(11), 1061-1069.

34. Khakimova, A., Yang, X., Zolotarev, O., Berberova, M., Charnine, M. Tracking knowledge evolution based on the terminology dynamics in 4p-medicine. International Journal of Environmental Research and Public Health, 2020, 17(20), c. 1-19, 7444.

35. O.V.Zolotarev, A.Kh.Khakimova, M.A. Berberova, V.P.Zolotareva, «Analysis of the dependence of the ruble exchange rate volatility on the oil market in a pandemic», CPT2020 Computing for Physics and Technology. The 8th International Conference on Computing for Physics and Technology (CPT2020). Conference Proceedings (2020), Nizhny Novgorod, Russia, May 11-15, 2020, pages 249-253. CEUR-WS.org/Vol-2763/CPT2020_paper_s3-7.pdf.
https://doi.org/10.30987/conferencearticle_5fce2772f12764.07821914.