

Bibliographic Computer Science Indexing Review with Disease Covid 19

¹Andrianingsih Andrianingsih, ²Tri Wahyu Widyaningsih, ³Meta Amalya Dewi

¹Department of Information System, Universitas Nasional, Indonesia

²Department of Informatics Engineering, Tanri Abeng University, Indonesia

³Information Systems, BINUS Online Learning, Bina Nusantara University, Indonesia

¹andrianingsih@civitas.unas.ac.id,

²tri.widyaningsih@tau.ac.id,

³meta.dewi@binus.edu

Accepted 12 March 2022

Approved 08 June 2022

Abstract - Researchers in conducting their research use the search using the homepage of the publication, according to expertise, collaboration in research, and research interests. And at this time the Covid 19 pandemic, became a trending topic for researchers, in various scientific fields. This study classifies based on publications located on the homepage source namely Scopus and Google Scholar, by analyzing the following topics, namely Natural Language Processing, Text Mining, Remote Sensing, and Sentiment Analysis using Name Entity Recognition to detect and classify named entities in text and using occurrence and link strength methods. The results showed science index literature about diseases Covid 19, obtained that Scopus has the most equitable percentage, has a good occurrence and link strength among the five scientific fields, namely Natural Language Processing 23.81%.33%, Text Mining 19.05%, Remote Sensing 0 %, Sentiment Analysis 57.14 % then Google Scholar Natural Language Processing 51.35%, Text Mining 0 %, Remote Sensing 48.65 %, Sentiment Analysis 0 %

Index Terms: Information Extraction; Bibliographic indexing; Disease Covid 19

I. I. INTRODUCTION

Coronavirus disease (COVID 19) was first discovered in Wuhan China at the end of 2019 [1]. This type of virus is highly contagious and spread rapidly in various parts of China but also in Japan, Thailand, and South Korea in less than 1 month through respiratory droplets and close contact [2]. This is a significant threat to the global health of millions of lives worldwide [3] causing acute respiratory system disorders so it was officially

declared by the World Health Organization (WHO) as a global pandemic on March 11, 2020 [4]. Of course, this pandemic has caught the world's attention because of its uncontrolled spread causing a spike in cases to increase [5]. It was reported that as of April 30, 2020, there were 3.2 million confirmed cases with a total of 227,847 deaths in 185 countries [6]. Because the spread of this virus is very high and uncontrolled, a large number of studies have been carried out and have been published [7] for free to speed up research and assist the government in responding to the crisis [8]. Therefore, it is important to evaluate the literature with quantitative and qualitative values to obtain literature patterns. and identify gaps and use the results.

Unstructured data in the form of entities, relations, objects, events, and many other types are a process extracted from Information Extraction, to improve the data analysis in the form of entities, objects, relations, and the perspective of streaming data is very different from static data. Static data does not have any connection between the dynamic time of initial processing and subsequent processing. Publication from researcher's academic contains rich information, which enables many applications such as academic search bibliographic and citation analysis.

This research was conducted with bibliometric analysis of scientific publications that become useful tools to know the process of generation and development of knowledge, as well as to evaluate the quality of the field of science and the impact it brings in the academic area [9]. Besides, bibliometric analysis can be used to find out the mapping of research from the research that is being done, already done as well as future opportunities [10].

The purpose of this research is to find out the mapping of research into several scopes of technology by discussing several parameters of topics concerning Natural Language Processing, Text Mining, Remote Sensing, and Sentiment Analysis published during the pandemic. The research mapping process is carried out by the stages of the object selection process, calculating the objects interacting, and the normalization process, creating maps and displaying maps, and evaluating the map [11]. Vosviewer is used to display bibliometric map visualizations downloaded from the page: www.vosviewer.com. Bibliometric map views are visualized with Vosviewer based on author or journal name with co-citation data, or based on keywords with co-occurrence data with label map display, sketch and density, and clusters [12]. Clusters in maps from Vosviewer are presented with color differences. Each parameter is operated by a clustering algorithm that can be changed so that more or fewer clusters are generated [13].

II. RESEARCH METHOD

A. Data Sources and Methods

Bibliometrics is the utilization of factual strategies to dissect books, articles, and other distributions. Bibliometric strategies are habitually utilized within the field of library and data science. The sub-field of bibliometrics which concerns itself with the investigation of logical distributions is called scientometrics. Scientometrics may be a sub-field of informetric. Major investigation issues incorporate the estimation of the effect of investigation papers and scholar diaries, the understanding of logical citations, and the utilization of such suggestions for something idea in approach and administration settings.

This bibliometric data collection is done from Scopus and Google Scholar, and total data analyzed as many as a total of 2991 papers indexed with the keyword "Coronavirus with ...[topic]" or "Covid 19 with...[topic]". Given the similarity of virus types before, restrictions are made on data retrieval while pandemic with the search categories used is topics, titles, and abstracts. Analysis of research trends using Vosviewer software with weighting method used is occurrence to see a lot of research on the topic and link strength to show the connectedness between research topics. Both methods analyze data based on abstraction and author.

The use of named entity recognition in the publisher homepage has problems and complexities that are generally the same as those in English, especially if using a machine learning approach. Fundamental differences exist when rule-based methods are used for completion or using hybrid model approaches between

rule-based and machine learning. This approach will use unsupervised learning so that it does not require labeled data for the learning process.

The stages of the process of the proposed method are as follows:

Data preparation for sequential pattern mining: In this step are prepared sentences that have named entities in it to be able to be degenerate paternal at each appearance of the entity. To avoid the amount of pattern produced, the pattern extraction process is limited to 5 words before and after the appearance of the entity.

Sequential Pattern Mining: In this step, an algorithm will be applied to the existing learning data to produce the desired pattern.

Pattern Marching and Candidate Extract: Datasets for testing are prepared for custom testing with the resulting pattern. The results will be sorted according to the level of confidence and support.

Candidate Pruning: This process is carried out to improve the accuracy of the named entity produced.

B. Citation Mapping Result

Mapping is a process that allows one to identify knowledge elements and their configurations, dynamics, interdependencies, and interactions. Knowledge mapping is used for technology management purposes, which include the definition of research programs, decisions regarding technology activities, the design of knowledge base structures, and the creation of education and training programs. A Quotation Outline such as the citation mapping could be a graphical representation that appears the quotation connections (cited references and citing articles) between a paper and other papers utilizing different visualization apparatuses and procedures. The citation mapping instrument from Web of Information tracks an article's cited and cited by references through two eras. So citation mapping is a graphical representation that shows the citation relationships (cited references and citing articles) between a paper and other papers using various visualization tools and techniques. The citation mapping tool from Web of Knowledge tracks an article's cited and cited by references through two generations.

1. Search Web of Science Core Collection and select a record
2. Click on the link **View Citation Map** link
3. Choose Forward, Backward, or Forward and Backward, plus 1 or 2 generations
4. Click **Create Map** button.

In a paper conveyed earlier this year, Malcolm Tight examines the theoretical considerations around commonalities inside the approaches of communities of sharpening and Becher’s insightful tribes and districts. He conducts a co-citation examination of Higher Instruction ask approximately journals; centering on maker characters and ranges, subjects, theories and examinations, methodologies and procedures, appearing as a basic diagrammatical representation of his descriptive demonstrating. Comparable thoughts of ‘citation mapping’ have been investigated someplace else, particularly inside the typical sciences, and a shape has as of late been displayed in the citation and journal database ISI Web of Science. And Instinctively Originator W. Bradford Paley’s visualization of 800,000 coherent papers livelihoods maker citations to explore the intercontinental between science perfect models.

Related to bibliometrics, science mapping is a method of visualizing a field of science. This visualization is done by creating a landscape map that can display topics from science [14]. Information visualization is a vital portion of information science, and it is utilized in two fundamental parts of the information science cycle: at the starting of the introduction of information investigation and within the conclusion amid the result introduction. Indeed, even though the visualization procedures are the same, these two stages have diverse objectives. Information investigation begins from numbness and tries to get the information, to find covered-up realities, designs, or exceptions. Result introduction begins with information and tries to communicate the message in the clearest and most viable way conceivable. Hence, indeed although they share the same procedures, the objective and the beginning point are diverse.

In the downloaded text records from Scopus and Google Scholar, we performed the metadata analysis for data extraction. This included extraction of title, author, year, and computer science of topic.

The morphological analysis allowed tagging of data potential use. Various issues with Covid 19 disease were encountered while tagging the data, which are described as follows:

Author field: Computer science part with Covid 19 topic names are usually made up of two or three parts. It is not always clear which indicator criteria for disease Covid 19. A definition was created in Covid 19 disease names, but the punctuation mark (,) was not given: review, role, outbreak, diagnosis, approach, detection, chest x-ray, and pneumonia

Types of documents in the database, we found the following types of documents Computer Science with Covid 19 disease:

- Scopus

- Google Scholar

For each type of document, we identified the mandatory fields, and other field values occurring in the database. The information extraction algorithm and the retrieval logic were based on these field values. A Sample of mandatory fields for each document is tabulated in Table 1.

Table 1. List of documents with total record paper

No.	Bibliographic Indexing	Quantity	Year	Title	Keyword	Abstract	Topic
1	Scopus	738	2020	✓	✓	✓	✓
2	Google Scholar	4000	2020	✓	✓	✓	✓

Based on the data table above obtained a total of 4738 articles, where Scopus obtained 738 (15.57%) and Google Scholar 4000 (84.42%) the three are accessed through API Publish or Perish with the resulting type in the form of conference and article as shown in the image below:

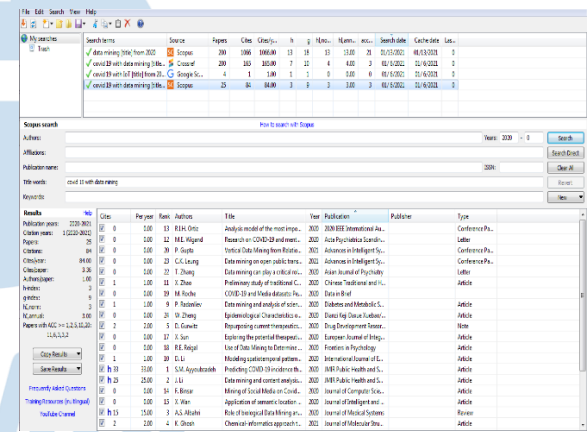


Fig. 1. Access Scopus and Google Scholar data through the Publish or Perish API

The results of the article data in table 1 above are then imported one by one into Vosviewer with txt format and compiled, inserted the title of the five fields of science filtered with topics related to Covid 19 and then carried out weighting using occurrence and total link strength which is then presented in the table below:

Table 2. Article 5 computer topic areas on google scholar

Keyword	Paper	Link	Total Link
NPL for Covid 19	1000	976	976
Text Mining for Covid 19	1000	432	4737
Remote Sensing	1000	2228	6180
Sentiment Analysis	1000	1284	4603

II. ARTICLE'S MAPPING

A. Google Scholar

1.1 NLP for Covid 19

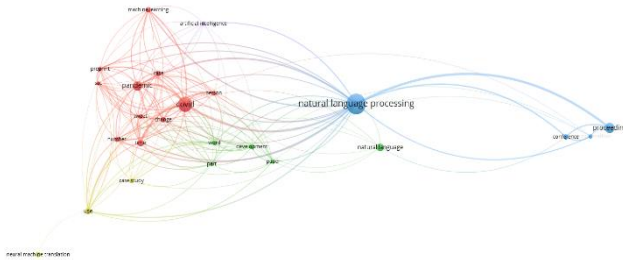


Fig. 2. Map visualization link topics NLP for covid 19

Vosviewer analysis showed the connectedness of the NLP field resulting in 4 clusters based on color differences in figure 2 that were related to natural language, artificial intelligence, and machine learning. In this mapping, it can be concluded that there is no research link between NLP with Text Mining, Remote Sensing, and Sentiment Analysis.

1.2 Text Mining for Covid 19

The field of Text Mining has different research links to the NLP field with 3 clusters in red, green, and blue that are closely related topics around text, system, and classification. So it can be concluded that there are still text mining research opportunities with drug, risk, and country, as well as with 3 other fields, which are visualized in figure 4 below:

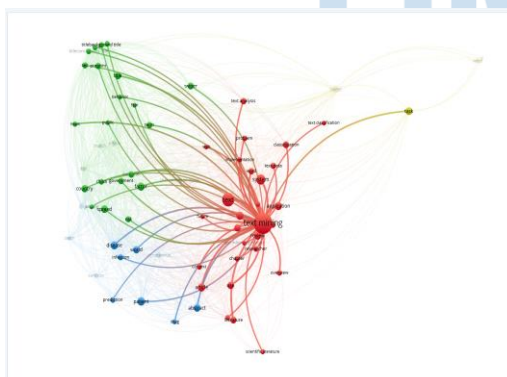


Fig. 3. Text mining for covid 19 field topic link visualization map

1.3 Remote Sensing for Covid 19



Fig. 4. Remote sensing for covid 19 field topic link visualization map

Vosviewer produces a research mapping analysis into 5 clusters with red, blue, green, and yellow and displays a strong correlation with information in purple, including topics of accuracy, change, factor, an erratum.

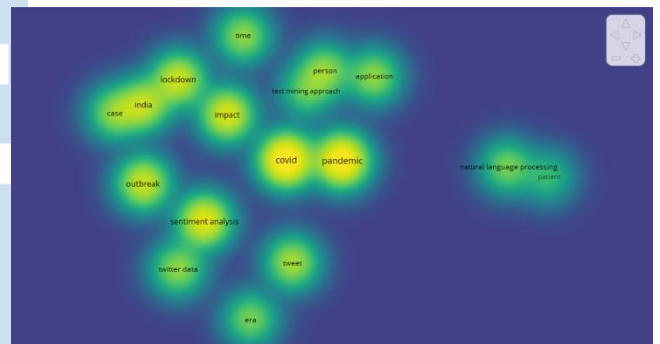


Fig. 5. Sentiment Analysis for Covid 19 field visualization map

In the field of sentiment analysis, the map displays very strong relationships including topics of neural networks, sentiment classification, models, papers, text, and algorithms.

B. Scopus

The second analysis was carried out on article data on Scopus with the following data:

Table 3. List of Articles on Scopus

No.	Field of Computer Science	Number of Papers
1	Natural Language Processing	138
2	Text Mining	200
3	Remote Sensing	200
4	Sentiment Analysis	200

The visualization in the table above can be seen in Figure 6 which is a network visualization and Figure 7 shows the density visualization.

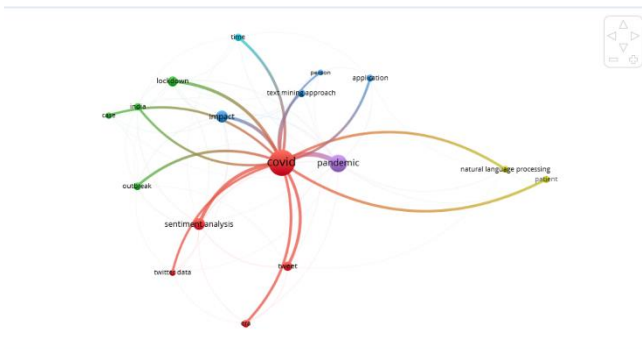


Fig. 6. Network Visualization on Scopus Articles

In the picture above the topic of covid has a strong network with pandemics and there is a network with Twitter data, sentiment analysis, outbreak, and impact, and there is a correlation with the text mining approach and with natural language processing.

Results	Help	Links	Per year	Rank	Authors	Info	Year	Publication	Publisher	Type
Publication year: 2020-2021										
Citation count: 1 (2020-2021)			105	16,000	1 A. Abd-Alhameed	Top outcomes of passenger during t...	2020	Journal of Medical Internet...		Article
Pages: 268			59	60,000	2 B.K. Tan	Studies of novel coronavirus disease...	2020	International Journal of Edu...		Article
Citations: 769			63	60,000	3 Y. Thaw	Chinese public's attitude to the C...	2020	Journal of Medical Internet...		Article
Outcomes: 76797			17	11,000	4 X. Han	Using social media to mine and an...	2020	International Journal of Inf...		Article
Outcomes: 215			53	53,000	5 M.S. Ghadi	Impacts of COVID-19 on global tra...	2020	Systems Management Res...		Article
Outcomes: 647			21	21,000	6 A. Alsoyouni	A bibliometric analysis of covid-19...	2020	Sakhalinskaya (Sakhalinsk)		Article
Outcomes: 22			73	70,000	7 L. Maeker	Why do we think together: how covid...	2020	European Journal of Inf...		Article
Outcomes: 22			59	16,000	8 A. Alshabab	Using physicians' opinions on so...	2020	Journal of Medical Health and Sa...		Article
Outcomes: 1440			16	6,000	9 Q. Chen	LiCovid: An open database of CO...	2021	Medical Health Research		Article
Outcomes: 22			19	11,000	10 J. Hatan	Key factors influencing covid-19 i...	2020	Sustainability (Switzerland)		Article
Outcomes with Alt: 1,123,118,376			17	12,000	11 J. Hatan	Systems pharmacological study fo...	2021	Pharmacoinformatics		Article
Alt: 10,743,847			12	2,000	12 A. Sarkar	Self-reported COVID-19 symptoms...	2020	Journal of the American M...		Article
			13	12,000	13 J. Tomarova	Coronavirus and people with int...	2020	Journal of Intellectual Disa...		Article
			11	11,000	14 Y. Wang	Investigating risk and crisis commu...	2021	Complexity in Human Behav...		Article
			10	10,000	15 M. Koh	How loneliness is related about in...	2020	Journal of Psychiatric Rese...		Article
			9	16,000	16 D. Pan	The influence of COVID-19 on exp...	2020	Risk-Reduction		Article
			8	6,000	17 M.E. Shuhada	How digital marketing performance...	2020	IFFA Advances		Article

Fig. 8. Access NLP on Scopus through the Publish or Perish API

Fig. 9. Research visualization map on Scopus

On the topic of the NLP field, it has several links to 5 other studies, including covid, tweets, and the NLP application itself. It can be concluded that there are still many research opportunities in the NLP field, especially related to the COVID 19 pandemic.

2.2 Text Mining

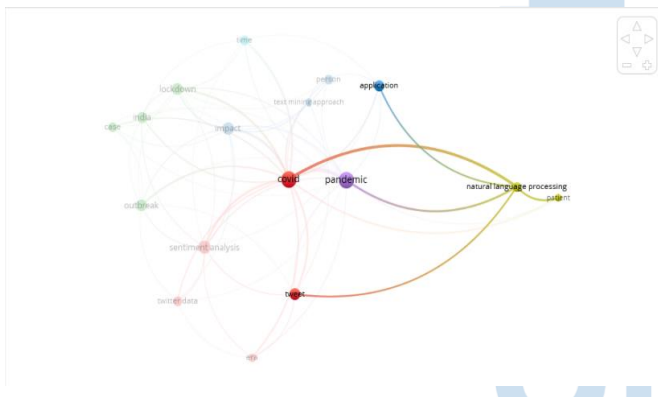


Fig. 7. Density Visualization of Scopus Articles

Figure 7 above shows a density map which is the result of an analysis using all articles on the development of covid 19 research for the fields of Natural Language Processing, Text Mining, Remote Sensing, and Sentiment Analysis on Scopus, both related and unrelated. And it turns out that the visualization results show that there are still few with the information that the red is the closer it is, and the green is the less often it is done.

2.1 NLP search for covid 19 on Scopus is accessed via API Publish or Perish can be seen in the image below

Results	Help	Links	Per year	Rank	Authors	Info	Year	Publication	Publisher	Type
Publication year: 2019-2021										
Citation count: 1 (2019-2021)			30	30,000	1 N. Zheng	Predicting COVID-19 in China Cities...	2020	IEEE Transactions on Cyber...		Article
Pages: 238			30	60,000	2 M.A.A. Ibrahim	Natural language processing on personal...	2020	Journal of Reproductive He...		Article
Citations: 274			23	23,000	3 M. Sulekar	Deep Sentiment Classification and...	2020	IEEE Journal of Biomedical...		Article
Outcomes: 47410			22	22,000	4 Y. Zheng	The terminology counts as a potential...	2020	Clinical Chemistry and Labo...		Article
Citations: 1,588			17	17,000	5 J.L. Rodriguez	Clinical Characteristics and Progn...	2020	Journal of Medical Internat...		Article
Outcomes: 678			17	17,000	6 M.K. Syed	An artificial intelligence and NLP b...	2020	International Journal of Ac...		Article
Outcomes: 9			11	11,000	7 J. Liu	Natural language processing metho...	2020	Journal of Medical Internat...		Article
Outcomes: 9			10	10,000	8 S. Shari	Medical, Commercial and Progn...	2020	Proceedings of the 18th In...		Conference Pap...
Outcomes: 9			9	9,000	9 T.K. Mackay	Dig data, natural language process...	2020	JAMA Public Health and Sur...		Article
Outcomes: 9			9	9,000	10 S. Liu	An Artificial System for Detectin...	2020	IEEE Access		Article
Outcomes: 7			7	7,000	11 A. Naser	Natural language processing for so...	2020	Journal of Medical Internat...		Article
Outcomes: 6			6	6,000	12 Y. Gao	An Emotion Case Model using Med...	2021	Chaos, Solitons and Fractals		Article
Outcomes: 6			6	6,000	13 M. Oshun	Application of topic modeling to L...	2020	Studies in Health Technol...		Conference Pap...
Outcomes: 5			5	5,000	14 L. Liu	COVID-19: Associated with lang...	2021	Current Eye Research		Article
Outcomes: 4			4	4,000	15 S. Alshali	COVID-19 Public Sentiment Insig...	2021	Complexity, Materials and...		Article
Outcomes: 4			4	4,000	16 G. Roy	Analyzing one-day user trends dur...	2021	Russian Reception Researc...		Article
Outcomes: 4			4	4,000	17 C.A.H. Buhari	Global Abatement of COVID-19 T...	2021	Journal of Healthcare Infor...		Article

Fig. 10. Access Text mining on Scopus through the Publish or Perish API

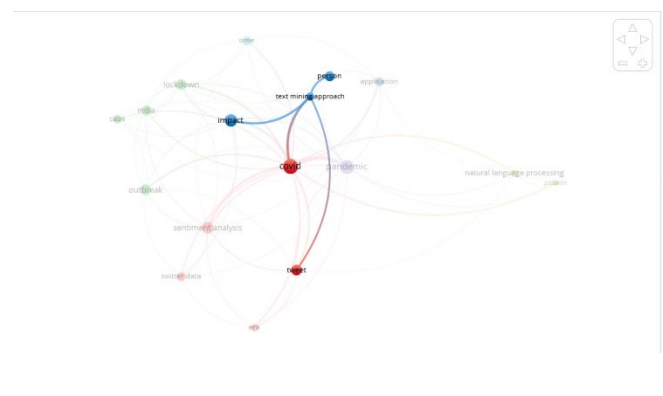


Fig. 11. Map visualization link topic text mining field on Scopus

Figure 9 above shows that Text Mining research has the strongest link with the topic of covid dan tweet. There are only 2 clusters, namely the red cluster with the topic of covid and tweets and the blue cluster with the topic of impact and person. This topic still has enormous research opportunities, especially related to covid 19.

2.3 Sentiment Analysis

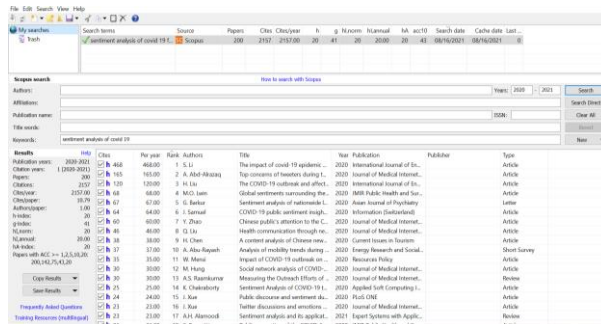


Fig. 12. Access sentiment analysis on Scopus through the Publish or Perish API

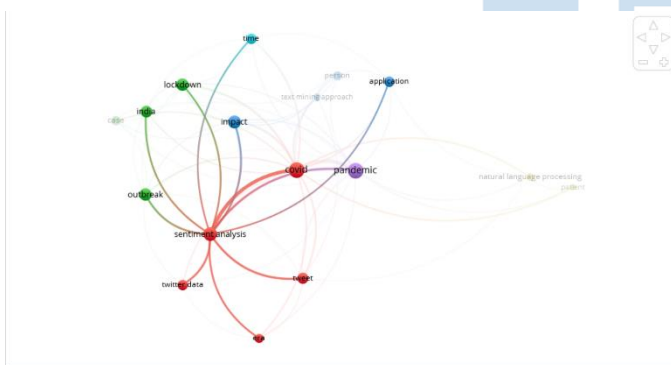


Fig. 13. Map visualization link topic sentiment analysis field on Scopus

This topic has links to other research, with the strongest link strength on Twitter data, tweets, covid, and its applications. There are 4 clusters, namely the red cluster with the topic of covid, tweet, Twitter data, and era, the green cluster with the topic of the outbreak, India, lockdown, the blue cluster with the topic of impact and application, and there is 1 pandemic topic in the purple cluster.

Table 2. Analysis of Scholar and Scopus

Computer Science Topic	Google Scholar		Scopus	
	Number of topic links	%	Number of topic links	%
Natural Language Processing	19	51,35%	5	23.81%
Text Mining	0	0%	4	19,05%
Remote Sensing	18	48,65%	0	0%
Sentiment Analysis	0	0%	12	57,14%
Total Link	37	100%	21	100%

IV. CONCLUSION

The analysis was carried out on Scopus and Google Scholar data sources for several scientific fields with the topics of Natural Language Processing, Text Mining, Remote Sensing, and Sentiment Analysis. The keywords used are about covid 19 which is related to the four scientific fields. By using Occurrence and Length Strength, it is concluded that at Google Scholar the most research related to Covid 19 is on the topic of Natural Language Processing and Remote Sensing, so there are still many research opportunities for the topic of Text Mining and Sentiment Analysis. Meanwhile, in Scopus, the most research related to Covid 19 is the topic of Sentiment Analysis, Natural Language Processing, and Text Mining, so there are still research opportunities on the topic of Remote Sensing.

REFERENCES

- [1] Wu T, Ge X, Yu G, Hu E. Open-source analytics tools for studying the COVID 19 coronavirus outbreak. medRxiv 2020:2020.02.25.20027433 [doi: 10.1101/2020.02.25.20027433]
- [2] Ren L-L, Wang Y-M, Wu Z-Q, Xiang Z-C, Guo L, Xu T, et al. Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. Chin Med J. 2020; published online February 11. DOI: 10.1097/CM9.0000000000000722.
- [3] Zhou Y, Chen L. Twenty-Year Span of Global Coronavirus Research Trends: A Bibliometric Analysis. Int J Environ Res Public Health. 2020;17(9):3082.
- [4] Hayat Ouassou, Loubna Kharchoufa, dkk The Pathogenesis of Coronavirus Disease 2019 (COVID 19): Evaluation and Prevention July 2020 Journal of Immunology Research Vol 2020 (New Insights into Immune-based Diagnosis,

- Therapy, and Prophylaxis for Infectious Diseases 2020):7
DOI: 10.1155/2020/1357983
- [5] Cucinotta, D.; Vanelli, M. WHO Declares COVID 19 a Pandemic. *Acta Biomed.* **2020**, *91*, 157–160. [CrossRef] [PubMed]
- [6] John Hopkins University and Medicine. Covid 19 Dashboard by the Center for Systems Science and Engineering at John Hopkins University. Available online: <https://coronavirus.jhu.edu/map.html> (accessed on 26 April 2020).
- [7] Bhattacharya S, Singh S. Visible Insights of the Invisible Pandemic: A Scientometric, Altmetric and Topic Trend Analysis. arXiv preprint arXiv:200410878. 2020.
- [8] Commission on a Global Health Risk Framework for the Future. *The Neglected Dimension of Global Security: A Framework to Counter Infectious Disease Crises*; National Academies Press: Washington, DC, USA, 2016.
- [9] Capobianco-uriarte M, Casado-belmonte P, Mar G, Ter E. 2019. A Bibliometric Analysis of International. DOI: 10.3390/su11071877
- [10] Indah Hartati, dkk, Analisa Bibliometrik Publikasi Ilmiah Bertema Biorefineri Biomassa Berlignoselulosa, Inovasi Teknik Kimia. Vol. 5, No.1, April 2020, Hal 42-49, ISSN 2527-614X, e-ISSN 2541-5891
- [11] N. J. V. Eck, Methodological Advances in Bibliometric Mapping of Science, Utrecht: Erasmus University Rotterdam, 2011.
- [12] Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538.<http://doi.org/10.1007/s11192-009-0146-3>
- [13] Leydesdorff, L., & Rafols, I. 2012. Interactive Overlays: A New Method for Generating Global thb Maps from Web-of-Science Data. *Journal of Informetrics*, *6*, 318–332
- [14] Royani, Y., Bachtar, M., Tambunan, K., Tupan, Sugiharto. 2013. Pemetaan Karya Tulis Ilmiah LPNK: Studi Kasus LIPI dan BPPT (2004-2008). *Journal Dokumentasi dan Informasi*, *34* (1) Juni
- [15] [15] N.J.V. Eck, Waltman Ludo. 2020. Vosviewer Manual version 1.6.16, Universiteit Leiden, CWTS Meaningful metrics,


 The logo for Universitas Muhammadiyah Negeri (UMN) is displayed in a light blue, stylized font. It consists of the letters 'U', 'M', and 'N' in a bold, sans-serif typeface. The 'U' and 'M' are connected at the top, and the 'N' is positioned to the right. The logo is centered horizontally and partially overlaps the bottom of the text area.