

# COVID-19 주요 연구 동향 분석: SCOPUS DB를 중심으로

YI ZHAO<sup>1</sup>, 손진현<sup>2\*</sup>

<sup>1</sup>충북대학교 경영정보학과 박사, <sup>2</sup>선문대학교 경영학과 교수

## Analysis of the COVID-19 Research Trend : Focusing on SCOPUS DB

YI ZHAO<sup>1</sup>, Jinhyeon Sohn<sup>2\*</sup>

<sup>1</sup>Ph.D., Department of MIS, Chungbuk National University

<sup>2</sup>Professor, Department of Business Administration, Sunmoon University

**요약** 본 논문의 목적은 COVID-19의 최근 주요 연구 동향을 파악하는데 있다. 이에 Elsevier에서 제공하는 해외 학술 DB인 SCOPUS를 활용하여 최근 3년간(2020~2022) COVID-19의 연구 동향을 파악하고자 한다. 빈도 분석 결과, covid가 7,248건, pandemic이 4,974건, study가 3,313건, research가 2,137건, crisis가 1,777건 순으로 그 중요도가 나타났다. 트렌드 분석결과, covid와 pandemic은 꾸준히 연구가 진행되고 있으나, study, research, crisis는 최근 들어 연구가 다소 감소하고 있는 것으로 나타났다. LDA 토픽 모델링 분석 결과, 중요한 토픽은 'covid19, pandemic' 인 것으로 나타났다. 이는 COVID-19에 대한 연구가 일상생활뿐만이 아니라 기업이나 조직에서도 필요하기 때문에 의학 이외의 다른 학문 분야에서 중요하다는 것을 보여준다. 특히 COVID-19가 그 어느 때보다도 중요해지는 상황에서 COVID-19의 연구가 미치는 영향 및 파급효과에 관한 관심이 지속적으로 이루어질 것으로 보인다.

**키워드** : COVID-19, 팬데믹, 트렌드, SCOPUS, LDA

**Abstract** The purpose of this study is to identify the major research trends of COVID-19 in recent times. In addition, we would like to use SCOPUS, an overseas academic database provided by Elsevier, to understand the research trends of COVID-19 in the last three years (2020–2022). As a result of frequency analysis, covid 7,248 cases, pandemic 4,974 cases, study 3,313 cases, research 2,137 cases, crisis 1,777 cases appeared in order of importance. As a result of the trend analysis, we found that studies on covid and pandemic are progressing steadily, but those on study, research, and crisis have decreased somewhat recently. As a result of LDA topic modeling analysis, the important topics were found to be 'covid19, pandemic'. This shows that research on COVID-19 is important not only in everyday life, but also in companies and organizations, and therefore in other academic fields besides medicine. When (the study of)COVID-19 becomes more important than ever, there seems to be an ongoing interest in the impact and ramifications of COVID-19 research.

**Key Words** : COVID-19, Pandemic, Trend, SCOPUS, LDA

\*Corresponding Author : Jinhyeon Sohn(jhsohn@sunmoon.ac.kr)

Received December 30, 2022

Accepted February 20, 2023

Revised January 9, 2023

Published February 28, 2023

## 1. 서론

2019년 말경에 발생한 원인 미상의 폐렴이 WHO(세계보건기구)에 보고되면서 COVID-19가 전세계적으로 유행을 하기 시작하였다[1]. 2023년 1월 7일을 기준으로 전세계 확진자수는 664,887,301명, 사망자는 6,709,637명, 완치자(격리해제)는 601,420,028명으로 나타나고 있다. 2021년 10월부터 전 세계적으로 백신 접종자 수 증가, 사회적 거리두기, 국가간의 이동 제한, 여행 제한 등을 통해 확진자 수가 점차 감소 추세에 있었다. 또한 오미크론이 대세종이 되기 시작하면서 기존의 델타 변이에 비해 관련 증상이 비교적 가볍게 나타나면서 전 세계적으로 코로나가 종식이 될 수 있다는 기대감을 주었다.

그러나 새로운 변이들이 발견되기 시작하면서 2022년 7월부터 전 세계적으로 확진자 수가 급증하는 추세이다. 오미크론 하위 변이 BA.2.75 하위변종 BA.4와 BA.5 등이 등장하게 되면서 새로운 방역체계를 위한 각국의 노력이 이루어지고 있다.

COVID-19로 인해 전 세계 국가들은 비대면이 일상화 되고, 경제 침체, 소비자 물가 지수 상승 등으로 인한 피해가 발생하고 있으며[2], 이러한 피해를 최소화하고 팬데믹 위기를 극복하기 위해서 각국 정부에서는 효과를 극대화할 수 있는 다양한 대응책을 마련하고 있다.

2022년 말 기준으로 전 세계에서 특정 국가를 제외하고서는 워드 코로나 정책을 시행하면서, 많은 변화가 일어나고 있다. 실외에서 마스크를 해제하고, COVID-19 진단검사도 간소화되는 추세이다. 이에 따라 국내 및 해외의 여행도 자유로워지고 있다.

코로나가 일상생활 뿐만 아니라 연구자들에게도 많은 영향을 미치고 있다. 비대면이 일상화되면서, 원격으로 회의, 교육 등이 이루어지고 있는 상황에서 과거와는 다른 다양한 협력을 통해 연구를 수행하고 있다.

이에 본 연구에서는 SCOPUS DB에서 COVID-19와 관련된 연구 논문을 추출하고, 초록이 있는 논문 중에서 Business, Management and Accounting 논문에 한정하여 연구 동향을 파악하고자 한다.

## 2. 선행연구

### 2.1 COVID-19

세계보건기구(WHO)는 COVID-19에 대해 전염병 최고단계인 '팬데믹(Pandemic)'을 선포했다[1]. 국내에서

는 2020년 2월 23일 COVID-19와 관련하여 위기 경보 단계를 최고단계인 심각으로 격상하고 대응체계를 대폭 강화했다. 심각 단계에서는 정부에서 보다 적극적인 COVID-19 확산 방지를 위해 해외 유입 차단 및 사회적 격리와 같은 강력한 대응조치를 추진했다[3].

COVID-19에 걸리게 되면, 대부분은 경미한 증상을 보이지만, 경우에 따라서는 심하게 아픈 경우도 있다. 그러나 고령자와 특정한 기저질환이 있는 경우에는 COVID-19의 중증질환 위험이 더욱 크며, 전 세계적으로 많은 사람이 사망하였다. COVID-19 백신이 개발되어 시판되고 있는데, 화이자 모더나 등의 백신은 안전하고 효과적이며, 최근에는 변이 바이러스를 잡을 수 있는 백신을 개발하여 보급하고 있다.

### 2.2 비정형 데이터 마이닝

데이터는 정형 데이터와 비정형 데이터로 분류를 할 수 있다. 일반적으로 수치로 된 데이터의 경우가 정형 데이터이며, 정형화되지 않는 데이터를 비정형 데이터라고 한다. 비정형 데이터는 데이터 속에 숨겨져 있는 패턴을 발견하거나 텍스트 마이닝과 같은 분석 방법을 통해서 이루어진다[4].

비정형데이터 마이닝은 텍스트마이닝, 오피니언 마이닝, 웹 마이닝, 소셜네트워크 분석 등 다양하다. 비정형 데이터 마이닝 중에서 가장 대표적인 방법이 바로 텍스트 마이닝이다[5,6].

텍스트 마이닝 분석을 수행하기 위해서는 먼저 텍스트 정형화가 필요하며, 일반적으로 메타 데이터로 직접 태그를 달거나 텍스트 속의 단어나 스피치의 한 부분에 대응하도록 하는 정형화 방법을 택한다[4].

비정형데이터 마이닝의 과정은 탐색과정, 이해과정, 분석과정의 3단계를 거치고 있다. 탐색에서는 질의, 집합 연산, 재귀, 팽창 등의 작업을 수행하고, 이해에서는 분배, 군집화, 분류 등의 과정을 거치게 되며, 분석과정에서는 패턴, 상관관계 등의 과정을 수행하게 된다[7].

### 2.3 LDA 토픽모델링

LDA(Latent Dirichlet Allocation)는 토픽모델링 알고리즘 중에서 가장 단순한 기법이며, 문서가 여러 개 토픽으로 혼합 및 구성되어 표현이 되고, 각각의 토픽이 단어 분포로 구성이 된다는 개념에 기초를 하고 있다.

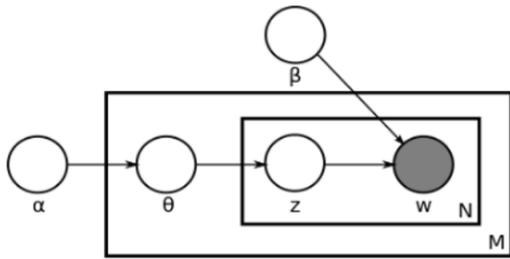


Fig. 1. LDA model

LDA는 문서를 표현하고 있는 잠재적 토픽과 단어가 디히클레 분포에 기반하여 토픽에 포함될 가능성을 추정하고, 문서 내에 잠재된 토픽 단어들을 집합으로 추출하는 알고리즘이다[8,9]. LDA 모델은 Fig. 1에서 보는 바와 같다.

### 3. 연구방법 및 절차

#### 3.1 연구절차

본 연구는 SCOPUS DB에서 추출한 COVID-19에 대한 연구를 텍스트마이닝 기법을 활용하여 주요 토픽을 파악하고자 한다[10,11]. 본 연구의 연구 절차는 Fig. 2에서 보는 바와 같으며, 문제 정의 → 데이터 수집 → 데이터 전처리(정제) → 데이터 분석 → 결과 해석으로 구분된다 [12,13].



Fig. 2. Research process

문제 정의에서는 본 연구의 주제에 대한 정의로서 주제, 데이터 수집 범위를 정하는 것이며, 이를 통해 데이터 수집을 하게 된다. 데이터를 수집한 이후 분석하기 전에 데이터 정제 및 전처리 과정을 수행하게 된다[14].

정제된 데이터는 빈도 및 키워드 추출, LDA(Latent Dirichlet Allocation) 토픽모델링, 텍스트마이닝 등의 분석과 시각화 단계로 진행하였다[15].

#### 3.2 데이터수집과 전처리

본 연구를 위한 분석 대상 데이터는 COVID-19 관련 SCOPUS DB의 논문 초록 데이터이다.

분석 기간을 2020년부터 선정하게 된 이유는 2020년 1월부터 전 세계적으로 코로나가 확산하기 시작하면서, 많은 변화가 일어났기 때문이다. 또한 SCOPUS DB에서 논문 검색을 수행한 결과 2019년까지는 COVID-19에 대한 연구가 거의 이루어지지 않았기 때문에 COVID-19에 대한 연구 동향을 파악하기 위해서 2020년부터 데이터를 수집하게 되었다.

분석 기간은 2020년 1월 ~ 2022년 12월까지이며, 302,130개의 open access 논문이 검색되었고 2020년을 기점으로 논문 수가 증가하고 있는 것으로 나타났다.

논문 유형별(Document type)로 분류하면, article은 273,509건, review는 38,824건, letter는 31,251건, conference는 24,021건 순으로 나타났다.

Table 1. Subject area

Area	Frequency	Percent
Medicine	234,921	46.5%
Social Sciences	63,909	12.7%
Biochemistry, Genetics and Molecular Biology	36,750	7.3%
Computer Science	36,421	7.2%
Engineering	28,432	5.6%
Immunology and Microbiology	27,188	5.4%
Environmental Science	23,634	4.7%
Nursing	18,102	3.6%
Business, Management and Accounting	17,687	3.5%
Psychology	17,661	3.5%

논문 주제 분야는 Medicine 관련 주제의 논문이 전체 논문 중에서 차지하는 비중이 가장 높게 나타났다. 그다음으로는 Social Sciences, Biochemistry, Genetics and Molecular Biology, Computer Science 순으로 그 중요도가 나타났다(Table 1).

논문 발행 국가로는 United States(미국)가 97,406건, United Kingdom(영국)이 39,414건, China(중국)이 37,090건, India(인도)가 30,320건, Italy(이탈리아)가 28,983건 순으로 나타났다.

소속기관(affiliation)으로는 Harvard Medical School이 가장 많은 것으로 나타났으며, University of Toronto, Inserm, University of Oxford, University College London 순으로 나타났다(Table 2).

본 논문에서는 subject area를 Business, Management and Accounting으로 한정하고 Document type을 article로 한정했다. 선택된 6,784개의 논문 중 키워드를





pandemic이 중요한 주제로 도출되었다.

본 논문에서는 Business, Management and Accounting으로 한정하고 연구동향을 분석하였다. 그 결과 covid, pandemic에 대한 연구가 주를 이루고 있음을 알 수 있었으며, crisis와 outbreak을 중심으로 다양한 주제에 대한 연구가 지속적으로 이루어지고 있음을 알 수 있었다. COVID-19의 변이가 지속적으로 발생하고 있는 가운데, 향후 연구에서는 위드 코로나와 포스트 코로나가 비즈니스와 경영 분야에 미치는 영향에 대한 다양한 관점의 연구가 지속적으로 이루어질 것으로 보인다.

Table 3. LDA topic

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
pandemic	pandemic	covid19	covid19	covid19
during	data	study	pandemic	pandemic
work	during	crisis	data	research
analysis	study	pandemic	crisis	study
study	tourism	during	impact	health
impact	social	impact	results	during
different	research	research	social	model
findings	results	results	health	social
new	factors	online	resilience	effects
tourism	analysis	public	study	findings
management	management	business	supply	public
research	these	number	model	using
world	using	role	using	global
positive	risk	approach	during	future
health	global	financial	research	media
countries	authors	can	countries	article
development	covid19	findings	different	work
industry	strategies	economic	economic	effect
market	lockdown	time	practices	tourism

## 5. 결론

본 연구는 COVID-19의 최근 주요 연구 동향을 살펴 보고자 SCOPUS DB에서 논문의 제목을 검색하여 비정형 데이터 마이닝 분석 기법을 활용하여 분석을 수행하였다.

먼저 빈도 분석결과, covid가 7,248건으로 가장 높은 빈도로 나타났으며, pandemic이 4,974건, study가 3,313건, research가 2,137건 crisis가 1,777건, data가 1,711건 순으로 중요한 것으로 나타났다.

트렌드 분석 결과, covid와 pandemic은 꾸준히 연구가 진행되고 있으나, study, research, crisis는 최근 들어 감소하고 있는 것으로 나타났다. LDA 토픽 모델링 결과, covid19와 pandemic이 중요한 것으로 나타났다.

본 연구에서는 COVID\_19의 최근 주요 연구 동향을

빅데이터 방법을 통해 분석을 수행하였다는 점에서 시사점이 있다. 또한, COVID-19의 연구가 의학 분야뿐만이 아니라 Business, Management and Accounting 분야, social science 분야 등 다양한 분야에서도 COVID-19가 미치는 다양한 영향과 파급효과 분석을 수행하고 있다는 점에서 기존 연구와 차별성을 갖는다고 하겠다.

이러한 시사점에도 불구하고, 본 연구에서는 특정한 학술 DB만을 활용하여 분석을 수행하고, 다양한 분석 방법을 고려하지 못했다는 한계점을 지닌다. 이에 향후 연구에서는 SCOPUS DB외에도 SPRINGER, Science Direct 등 다양한 학술 제공 사이트의 DB도 활용하고, 분석 방법론과 비교 분석 등의 연구가 필요할 것으로 보인다.

## REFERENCES

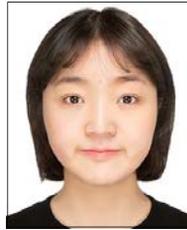
- [1] Kim, J., Shin, D., & Kim, H. (2021). Analysis of Major COVID-19 Issues Using Unstructured Big Data. *Knowledge Management Research*, 22(2), 145-165.
- [2] Suh, K. D. (2021). A Study on deduction of important factors for new infectious diseases through big data analysis. *Journal of Industrial Convergence*, 19(3), 35-40.
- [3] Central Disaster Management. Headquarters. (2020). Coronavirus Infectious Disease-19. Central Disaster Management. Headquarters regular briefing
- [4] KASOM(2021). Introduction to Big Data. kwangmoonkag.
- [5] Dumais, S. (1998). Using SVMs for text categorization. *IEEE Intelligent Systems*, 13(4), 21-23.
- [6] Wang, L. L., & Lo, K. (2021). Text mining approaches for dealing with the rapidly expanding literature on COVID-19. *Briefings in Bioinformatics*, 22(2), 781-799.
- [7] SDPM(2020). NCS management big data analysis. Wowpass.
- [8] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- [9] Griffiths, T. L., & Steyvers, M. (2004). Finding

scientific topics. Proceedings of the National academy of Sciences, 101(suppl\_1), 5228-5235.

- [10] Park, E. M., & Seo, J. H. (2020). Analysis of Research Trends in Technology Innovation: Focus on SCOPUS DB. Journal of Convergence for Information Technology, 10(8), 120-126.
- [11] Kim, I. G., Park, S. T., Park, S. S., Kim, M. S., Park, J. C., & Jiang, J. (2021). Analysis on Research Trends in Sport Facilities: Focusing on SCOPUS DB. Journal of Industrial Convergence, 19(6), 11-19.
- [12] Park, S. T., & Liu, C. (2020). A study on topic models using LDA and Word2Vec in travel route recommendation: focus on convergence travel and tours reviews. Personal and Ubiquitous Computing, 1-17.
- [13] Li, G., Dai, J. S., Park, E. M., & Park, S. T. (2017). A study on the service and trend of Fintech security based on text-mining: Focused on the data of Korean online news. Journal of Computer Virology and Hacking Techniques, 13(4), 249-255.
- [14] Ko, M. H., Park, E. M., & Park, S. T. (2018). Mining the open science: Themes and trends 10 years of open science. Journal of Advanced Research in Dynamical and Control Systems, 10(1), 263-270.
- [15] Park, S. T., Lee, S. W., & Kang, T. G. (2018). A study on the trend of cloud service and security through text mining technique. International Journal of Engineering & Technology, 7(2.33), 127-132.

YI ZHAO

[정회원]



- 2012년 8월 : 충북대학교 경영정보학과(학사)
- 2014년 8월 : 충북대학교 경영정보학과(석사)
- 2022년 8월 : 충북대학교 경영정보학과(박사)

- 관심분야 : 위기정보 및 뉴스, 소비자 행동, 클라우드 서비스, MIS
- E-Mail : zhaoyi207@gmail.com

손 진 현(Jinhyeon Sohn)

[정회원]



- 1986년 2월: 서울대학교 수학과(이학사)
- 1991년 2월: KAIST 산업공학과(공학석사)
- 1997년 2월: KAIST 산업공학과(공학박사)

- 1997년 3월~현재 : 선문대학교 경영학과 교수
- 관심분야 : 수송망설계, Network 이론, 빅데이터 분석
- E-Mail : jhsohn@sunmoon.ac.kr