

ESG 주요 연구 동향 분석: SCOPUS DB를 중심으로

Analysis of the ESG Research Trend : Focusing on SCOPUS DB

노규성(Kyoo-Sung Noh)*

국문초록 : 본 연구는 ESG(환경, 사회, 지배구조)에 대한 연구 동향을 분석함으로써 그 트렌드를 분석하고 기업과 투자자들이 ESG 정보를 활용하는 방안을 모색하는 방향을 제시하는 것을 연구 목적으로 한다. 이를 위해 비정형 데이터마이닝 기법 중 하나인 텍스트마이닝을 활용하여 분석하였다. SCOPUS 데이터베이스에서 2014년 1월부터 2023년 2월까지의 논문 초록을 수집하였으며, Economics, Econometrics and Finance 분야가 가장 많았다. 미국과 중국이 가장 많은 ESG 논문을 게재하였으며, 한국은 세계에서 6번째로 많은 논문을 게재하였다. 본 연구에서는 ESG의 주요 연구 동향을 텍스트마이닝 기법인 LDA, 토픽모델링 등의 방법을 사용하여 분석을 수행하였다는 점에서 의의가 있다. ESG가 특정한 분야가 아닌 다양한 분야에서 이루어지고 있다는 사실을 확인할 수 있었으며, ESG가 미치는 다양한 영향 요인과 파급효과에 대한 분석을 수행하였다는 점에서 기존 연구들과 차별성이 있다.

주제어 : ESG, 비정형 데이터마이닝, SCOPUS, 연구 동향, 정보

Abstract The purpose of this study is to analyze research trends on ESG (Environmental, Social, and Governance), and to present a direction for companies and investors to use ESG information. To this end, text mining, one of the atypical data mining techniques, was used for analysis. Thesis abstracts from January 2014 to February 2023 were collected from the SCOPUS database, and Economics, Econometrics and Finance were the most common. The United States and China published the most ESG papers, and Korea published the 6th most papers in the world. This study is meaningful in that it analyzed the main research trends of ESG using text mining techniques such as LDA and topic modeling. It was confirmed that ESG is being conducted in various fields, not in a specific field, and it is differentiated from previous studies in that it analyzed various influencing factors and ripple effects of ESG.

Key Words : ESG, unstructured data mining, SCOPUS, research trends, information

I. 서론

전 세계적으로 기후 변화에 대응을 하기 위해 파리기후 협약에서 250여개국은 지구의 평균적인 기온을 산업화와 대비하여 1.5도 이하로 억제하기로 약속을 하였다. EU는 2019년 12월 2050년 탄소중립을 실현하기 위해 30년간 온실가스 감축 목표를 상향 조정(40%~55%) 하였으며, 탄소국경세 법안의 초안을 발표하였다. 미국은 2020년 7월에 2050년에 Net Zero 달성 목표를 실현하기 위해 우선 전력 부문에서 2035년에 탄소배출 제로를 달성하고, 친환경차 산업 육성 및 기후변화 국제 공조를 주도하고 있다[1].

현재 한국의 주요 수출품목인 철강, 반도체, 석유 화학 등은 탄소국경세 이슈를 피할 수 없는 상황이다. 2019년 기준으로 주요국의 석탄 발전 비중을 살펴보면, 한국은 41.5%, 일본 30.9%, 독일 28.2%, 미국 23.6%, EU 17.5%로 한국이 석탄발전 비중이 가장 높은 것으로 나타났다. 이에 탄소국경세가 도입된다면 제품의 제조 원가가 낮더라도 생산과정에서 탄소를 많이 배출할 경우 결국에는 글로벌 가격 경쟁력에서 뒤처지게 될 것으로 보인다.

또한 현재 기업의 ESG 공시는 의무가 아니지만, 글로벌 추세로 확산되고 있다. EU에서는 종업원수가 250명 이상이고 자산이 2,000만 유로 이상, 매출이 4,000만 유로 이상 중에서 2가지 이상을 충족하는 기업은 상장 여부와 관련이 없이 2024년부터 공시기준에 따라 ESG를 보고해야 한다[2].

한국에서도 2025년부터 2조원 이상의 자산을 보유한 기업은 의미적으로 공시를 해야 하며, 2030년부터는 코스피 상장사가 의무적으로 공시를 할 예정이다.

EU, 미국, 일본, 한국 등의 국가적인 차원에서 이루어지는 연구 및 보고서, 기업이나 컨설팅사 차원에서 이루어지고 있는 보고서, 학술적 차원의 ESG 연구 등 현재 ESG 관련 연구는 다양하게 이루어지고 있다.

이에 본 연구에서는 Elsevier사에서 제공하는 SCOPUS 학술 데이터베이스에서 ESG와 관련된 논문중에서 초록이 제공되고 있는 논문에 한정하여 데이터를 수집하고 이를 텍스트마이닝 기법을 통해 분석하여 ESG 연구 동향을 파악하고자 한다.

II. 이론적 배경

1. ESG

르네페세가 1979년 출간한 Economic Systems and Living Systems에서 강조한 내용이 ESG 개념의 이론적인 토대가 되었다[3]. ESG라는 용어는 UNGC/International

Finance Corporation(2004)의 “Who Cares Wins” Report에서 처음으로 공식 등장하였다[4].

ESG는 환경(E, Environmental), 사회(S, Social), 지배구조(G, Governance)의 머리글자를 딴 용어로 기업의 비재무적인 경영 요소를 의미한다. ESG는 사회적이고 환경적인 요인에 대해 조직의 노력을 측정이 가능한 지표로 정량화하는 과정이다[5]. ESG에 대해 살펴보면, ‘E’ 요인은 환경적 요인 영역으로 기후 변화와 관한 탄소배출 감소, 친환경 건축물, 친환경 설비, 생물의 다양성, 재생에너지, 에너지 효율성, 오염/폐기물 처리, 수질 및 토양 오염, 환경문제의 잠재적인 문제를 처리하는 방법 등의 내용이 주를 이루고 있다.

‘S’ 요인은 사회적 요인 영역으로 사람과 사회에 관한 광범위한 이슈를 다루고 있다. 기회 평등, 연합 자유, 결사 자유, 건강 및 안전, 지역 사회문제, 인권/노동문제 등의 내용이 주를 이루고 있다.

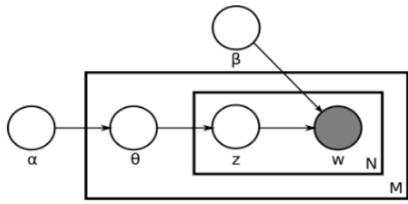
‘G’ 요인은 지배구조 요인 영역으로 경영에 관한 전반적인 사항에 대해 다루고 있다. 경영진, 경영 윤리, 공정 및 투명한 이사회, 이사회 독립성, 임원 관리, 뇌물/패, 정보공개 등의 내용이 주를 이루고 있다[6].

2. 비정형 데이터 마이닝

일반적으로 정형화(structured)되지 않는 데이터를 비정형(unstructured) 데이터라고 부른다. 이러한 비정형 데이터는 숨겨져 있는 패턴이나 경향을 발견하는데, 텍스트 마이닝과 같은 분석 기법을 통해 이루어진다[7]. 비정형데이터 마이닝 기법은 다양하지만, 그 중에서 가장 대표적으로 사용되는 기법이 바로 텍스트 마이닝 기법이다[8,9]. 텍스트 마이닝은 텍스트를 정형화하는 것이 반드시 필요한데, 정형화 하는 방법은 메타 데이터로 직접 태그를 달거나 텍스트 속의 단어의 한 부분에 대응하도록 한다[7].

3. LDA 토픽모델링

토픽 모델링은 주어진 문서내에 숨겨져 있는 주제가 무엇 인지를 분석하는 텍스트마이닝 기법 중의 하나이다. 이러한 토픽 모델링(Topic Modelling) 알고리즘 중에서 단순한 기법인 잠재 디리클레 할당인 Latent Dirichlet Allocation(LDA)는 여러 개의 토픽으로 혼합되어 있고, 구성되어져 표현이 되고, 각 토픽의 단어 분포로 구성된다는 개념에 기초하고 있는 기법이다[10,11]. 이러한 LDA 모델은 [그림 1]과 같다.



[그림 1] LDA Model

LDA가 기본적으로 가정하는 문서의 생성과정은 말뭉치의 전체 문서 개수, 전체 토크스, 관찰이 가능한 변수 등의 정보를 가지고 하이퍼 파라미터를 제외한 나머지 모든 잠재 변수를 추정 과정을 수행한다.

4. 선행연구

ESG에 대한 선행연구는 여러 방향에서 이루어지고 있는데, 여기에서는 연구 동향에 관한 선행연구로 국한하여 정리하고자 한다.

유현주(2023)은 글로벌 ESG 공시 표준화 대비, 공급망 ESG 리스크 대비, 비즈니스의 친환경 수준 식별, ESG 관련 분쟁 경계, ESG Washing 모니터링 강화 등의 ESG 10대 트렌드에 대해 전망을 하였다[12]. 이가은과 최영준(2023)은 2013년~2022년까지 Web of Science의 저널 중에서 346편을 대상으로 토크모델링과 네트워크 분석을 수행하였다. 분석결과, 연도별로 중요한 연구주제에 차이가 있다는 점을 찾아내었다[13].

양찬열(2022)은 2020년~2022년까지의 뉴스 데이터를 활용하여 ESG 경영에 대한 동향을 연구하였는데, ESG와 ESG 경영 키워드가 지속적으로 증가를 하고 있고 관련 연구 논문도 급증하고 있다고 주장하였다[14]. 변영조와 우승한(2022)은 국내외 ESG연구 동향을 파악하였는데, 전반적으로 관련 연구가 증가를 하고 있다는 사실을 확인하였다. 그러나 국내외와 국외의 경우는 특정한 데이터베이스 한 곳만을 대상으로 하였다는 한계점이 있다[5].

III. 연구방법 및 절차

1. 연구절차

본 연구는 SCOPUS 학술 데이터베이스에서 추출한 ESG 연구를 비정형데이터마이닝의 주요 기법 중 하나인 텍스트마이닝 기법을 활용하여 중요 토크를 도출하고자 한다 [15,16]. 본 연구의 연구 절차는 문제 정의 → 데이터 수집 → 데이터 전처리 → 데이터 분석 → 결과 해석(시각화)으로

진행한다[17,18].

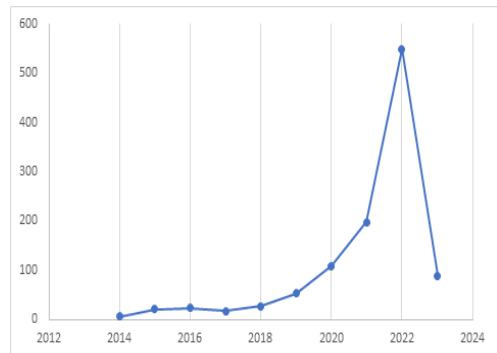


[그림 2] Research Process

먼저 문제 정의에서는 범위를 주제를 정하고, 관련 데이터에 대한 수집의 범위를 사전에 정하는 것으로 이를 통해 분석을 하고자 하는 데이터를 수집하게 된다. 수집한 데이터는 데이터 전처리(정제) 과정을 거쳐 분석하기에 적합하게 만든다[1,9]. 그리고 전처리된 데이터를 키워드 분석, Latent Dirichlet Allocation(LDA) 등의 분석과 시각화를 통해 수행하였다[20].

2. 데이터 수집과 전처리

분석 대상 데이터는 ESG와 관련된 SCOPUS 데이터베이스의 논문 초록이다. 분석 기간은 2014년 1월 ~ 2023년 2월까지이다. open access 논문은 1,127개가 검색이 되었고 2020년부터 논문수가 증가하기 시작하여 2022년에는 549건의 논문이 발행되었다. 그만큼 ESG에 대한 관심이 전 세계적으로 증가를 하고 있다는 것을 보여주고 있다.

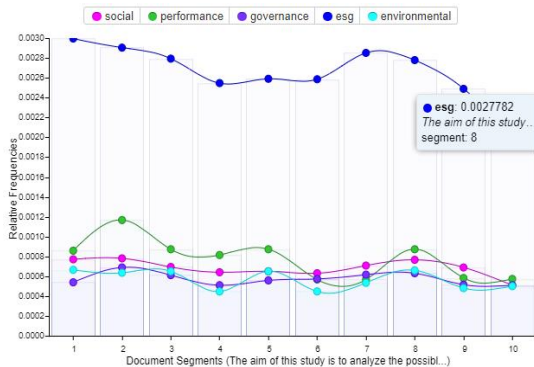


[그림 3] The number of papers

논문 유형별(Document type)로 분류하면, article은 965건, conference paper는 106건, book chapter는 53건, review는 33건 순으로 그 중요도가 나타났다.

2. 트렌드 분석

중요한 단어인 esg, social, performance, government, environmental의 트렌드 분석을 수행한 결과는 [그림 6]과 같다. 이를 통해 볼 수 있듯이, 전반적으로 모든 키워드들이 증감을 반복하고 있는 것으로 보인다.



[그림 6] Trend(voyant-tools)

3. LDA 토픽모델링 결과

토픽모델링 결과는 Table 3에서 보는 바와 같다. 1번 토픽의 중요 키워드로는 esg, social, environmental, study, firms 등이 있으며, 토픽의 주제는 ‘esg’와 ‘social’ 임을 유추할 수 있다.

2번 토픽의 중요 키워드로는 esg, performance, governance, social, study 등이 있으며, 토픽의 주제는 ‘esg’와 ‘performance’ 임을 유추할 수 있다.

3번 토픽의 중요 키워드로는 esg, corporate, social, financial, performance 등이 있으며, 토픽의 주제는 ‘esg’와 ‘corporate’ 임을 유추할 수 있다.

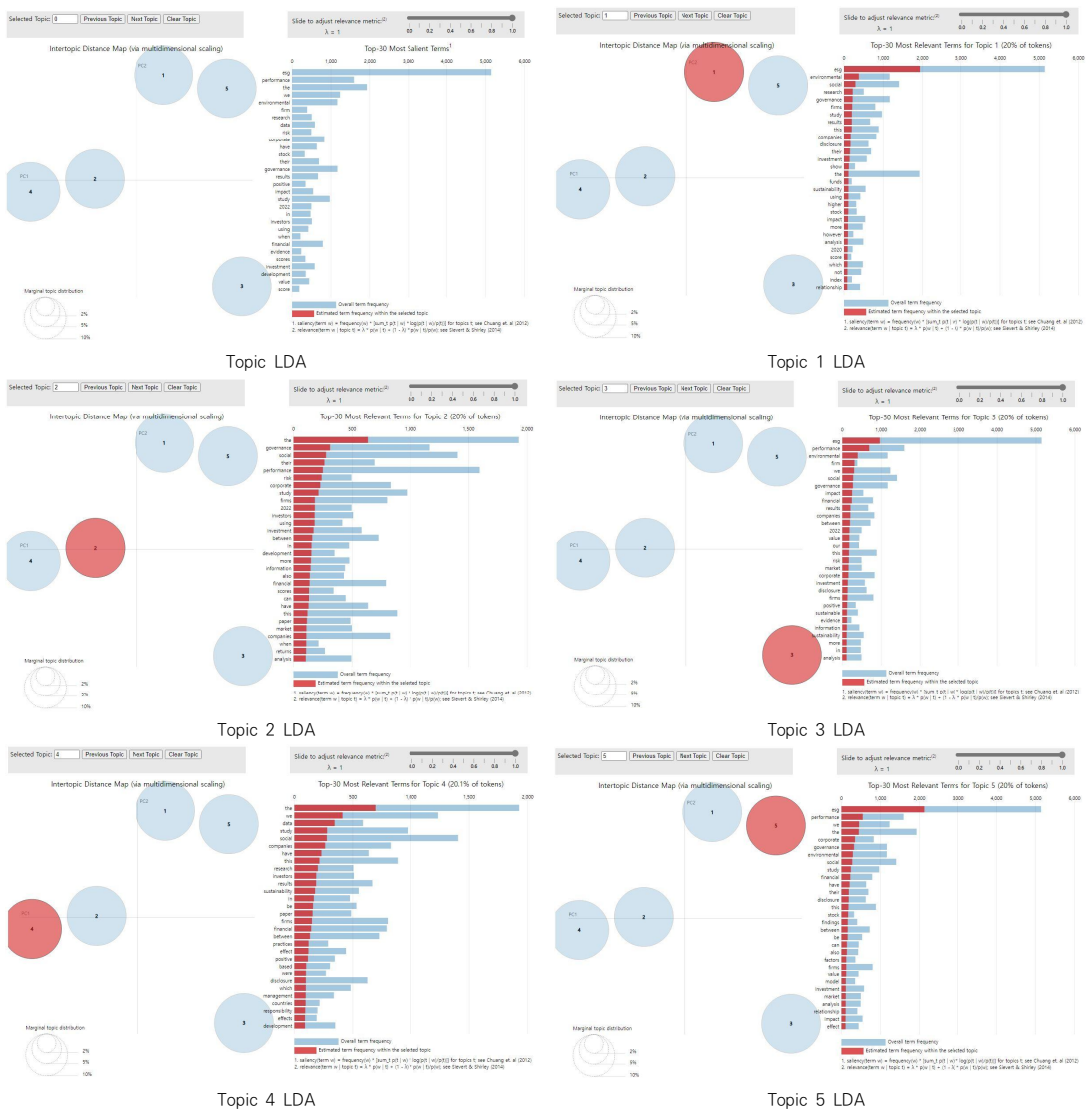
4번 토픽의 중요 키워드로는 esg, performance, environmental, governance, analysis 등이 있으며, 토픽의 주제는 ‘esg’와 ‘performance’ 임을 유추할 수 있다.

5번 토픽의 중요 키워드로는 esg, financial, study, value, environmental 등이 있으며, 토픽의 주제는 ‘esg’와 ‘financial’ 임을 유추할 수 있다.

한편, LDA 분석 결과를 시각화 것은 [그림 7]과 같다.

[표 3] LDA topic

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
esg	esg	esg	esg	esg
social	performance	corporate	performance	financial
environmental	governance	social	environmental	study
study	social	financial	governance	value
firms	study	performance	analysis	environmental
governance	data	positive	companies	companies
results	corporate	market	paper	social
disclosure	results	firms	impact	investment
companies	risk	investors	information	development
stock	paper	disclosure	sustainable	investors
investment	firms	risk	sustainability	performance
analysis	environmental	environmental	findings	sustainability
sustainability	companies	between	effect	results
between	investment	findings	relationship	between
significant	research	impact	using	reporting
based	scores	companies	practices	impact
relationship	impact	information	factors	relationship
sustainable	value	management	data	data
research	investors	study	research	market



[Fig 7] LDA Topic Modeling

4. 논의

단어빈도 분석 결과 esg, performance, social, governance, environmental, study, financial 순으로 그 중요도가 나타났으며, Link 분석 결과, esg가 가장 중심에 있으며, performance와 social을 중심으로 한 연구가 주를 이루고 있는 것으로 나타났다. 트렌드 분석결과 esg와 performance가 중요한 트렌드로 나타났다.토픽 모델링 결과, esg와 performance가 중요한 주제도 도출되었다.

이상의 분석 결과 esg, performance, social,

environmental에 대한 연구가 주를 이루고 있음을 알 수 있었으며, esg와 performance를 중심으로 다양한 주제에 대한 연구가 이루어지고 있음을 알 수 있다.

전세계적으로 탄소 중립에 대한 논의가 지속적으로 이루어지고 있고, ESG에 대한 다양한 변화의 흐름이 지속되고 있는 가운데, 향후 연구에서는 ESG가 산업 분야, 경제 분야, 기후, 국가 등의 다양한 관점에서의 연구가 지속적으로 이루어질 것으로 보인다.

V. 결론

본 연구는 ESG에 대한 주요한 연구 동향을 살펴보기 위해 논문 제공 사이트인 SCOPUS 데이터베이스에서 ESG라는 주제어를 입력하여 논문을 검색하고 초록이 제공되는 논문을 중심으로 텍스트마이닝 기법을 활용하여 분석을 수행하였다.

먼저 빈도 분석결과, esg가 5,475건으로 가장 높은 빈도로 나타났으며, performance가 1,608건, social이 1,420건, governance가 1,195건, environmental이 1,178건, study가 975건, financial이 931건 순으로 중요한 것으로 나타났다. 트렌드 분석 결과, 중요한 단어인 esg, social, performance, government, environmental의 트렌드 분석을 수행한 결과는 다음과 같다. 전반적으로 모든 키워드들이 증감을 반복하고 있는 것으로 나타났다. LDA 토픽 모델링 결과, esg와 performance이 중요한 것으로 나타났다.

본 연구에서는 ESG의 주요 연구 동향을 텍스트마이닝 기법인 LDA, 토픽모델링 등의 방법을 사용하여 분석을 수행하였다는 점에서 의의가 있다. ESG가 특정한 분야가 아닌 다양한 분야에서 이루어지고 있다는 사실을 확인할 수 있었으며, ESG가 미치는 다양한 영향 요인과 파급효과에 대한 분석을 수행하였다는 점에서 기존 연구들과 차별성이 있다.

다만, 학술 데이터를 제공하는 많은 데이터베이스 중에서 특정 데이터베이스만 대상으로 분석을 수행하고, 다양한 분석 기법을 활용하지 못했다는 한계점이 있다. 향후 연구에서는 SCOPUS 데이터베이스 뿐만 아니라 SPRINGER, Wiley Online, SAGE 등의 논문 제공 사이트와의 결합분석을 시도하고 나아가 다양한 분석 기법을 통한 비교 등의 연구가 필요할 것으로 사료된다.

REFERENCES

- [1] <https://esg.krx.co.kr/contents/01/01020200/ESG01020200.jsp>
- [2] https://www.pwc.com/kr/ko/publications/acc/samil-sgc-accg_vol18.pdf
- [3] P. René. (1979). L'économique et le vivant. Payot, «pbp.
- [4] <https://documents1.worldbank.org/curated/en/444801491483640669/pdf/113850-BRI-IFC-Breif-whocares-PUBLIC.pdf>
- [5] Y. Byun & S. H. Woo (2022). Research Trend on ESG Management of Corporation. *Clean Technology*, 28(2), 193-200. DOI : 10.7464/KSCT.2022.28.2.193
- [6] Y. J., Kwon (2023). *An Analysis on the Impact of ESG Value Formation Factors on Rent*. Kangwon National University.
- [7] KASOM (2021). *Introduction to Big Data*. kwangmoonkag.
- [8] S. Dumais (1998). Using SVMs for text categorization. *IEEE Intelligent Systems*, 13(4), 21-23.
- [9] L. L. Wang & K. Lo (2021). Text mining approaches for dealing with the rapidly expanding literature on COVID-19. *Briefings in Bioinformatics*, 22(2), 781-799.
- [10] E. M. Park & J. H. Seo (2020). Analysis of Research Trends in Technology Innovation : Focus on SCOPUS DB. *Journal of Convergence for Information Technology*, 10(8), 120-126.
- [11] Y. ZHAO & J. H., Sohn (2023). Analysis of the COVID-19 Research Trend : Focusing on SCOPUS DB. *Journal of Industrial Convergence*, 21(2), 17-23.
- [12] H. J. Yu (2023). 10 ESG trends forecast for 2023, POSRI Issue Report.
- [13] J. E. Li & Y. J. Choi (2023). Analysis of major research trends in ESG using topic modeling. *Korean Corporation Management Review*. 30(2), 133-166.
- [14] C. Y., Yang (2022). A study on ESG management trends through news big data analysis. *Sustainability Journal*, 1(1), 1-17.
- [15] E. M. Park & J. H. Seo (2020). Analysis of Research Trends in Technology Innovation: Focus on SCOPUS DB. *Journal of Convergence for Information Technology*, 10(8), 120-126.
- [16] I. G. Kim, S. T. Park, S. S. Park, M. S. Kim, J. C. Park & J. Jiang (2021). Analysis on Research Trends in Sport Facilities: Focusing on SCOPUS DB. *Journal of Industrial*

Convergence, 19(6), 11-19.

- [17] S. T., Park & C., Liu (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.
- [18] G. Li, J. S. Dai, E. M. Park & S. T. Park. (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.
- [19] M. H. Ko, E. M. Park & S. T. Park (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.
- [20] S. T. Park, S. W. Lee & T. G. Kang (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.