

뉴스 데이터 토픽 모델링을 활용한 COVID-19 대유행 전후의 클라우드 보안 동향 파악*

이 선 우*, 이 재 우**

요 약

COVID-19 대유행으로 인해 많은 기업에서 재택근무를 도입했다. 하지만 재택근무 도입으로 기업의 민감한 정보에 접근하려는 공격 시도가 증가했고, 보안위협에 대응하기 위해 많은 기업에서 클라우드 서비스를 이용하기 시작했다. 본 연구는 COVID-19 대유행 전후의 국내 클라우드 보안 동향의 변화를 분석하기 위해 ‘클라우드 보안’ 키워드로 뉴스 데이터를 수집하여 LDA 토픽 모델링 기법을 사용했다. COVID-19 대유행 전에는 국내 클라우드 보안에 대한 관심이 낮아 추출한 토픽에서 대표성이나 연관성을 찾을 수 없었다. 다만, 현재 많은 연구가 이뤄지는 IT기술인 AI, IoT, 블록체인을 위해서는 높은 컴퓨팅 성능을 위해 클라우드의 도입이 필요하다는 것을 분석할 수 있었다. 반면, COVID-19 대유행 이후 추출된 토픽을 보면 국내에서 클라우드에 대한 관심이 증가했고, 이에 따라 클라우드 보안에 대한 관심이 향상된 것을 확인했다. 따라서 앞으로 계속 증가할 클라우드 서비스 사용량에 대비한 보안 대책을 수립해야 할 것이다.

Topic Modeling to Identify Cloud Security Trends using news Data Before and After the COVID-19 Pandemic

Soun U Lee*, Jaewoo Lee**

ABSTRACT

Due to the COVID-19 pandemic, many companies have introduced remote work. However, the introduction of remote work has increased attacks on companies to access sensitive information, and many companies have begun to use cloud services to respond to security threats. This study used LDA topic modeling techniques by collecting news data with the keyword 'cloud security' to analyze changes in domestic cloud security trends before and after the COVID-19 pandemic. Before the COVID-19 pandemic, interest in domestic cloud security was low, so representation or association could not be found in the extracted topics. However, it was analyzed that the introduction of cloud is necessary for high computing performance for AI, IoT, and blockchain, which are IT technologies that are currently being studied. On the other hand, looking at topics extracted after the COVID-19 pandemic, it was confirmed that interest in the cloud increased in Korea, and accordingly, interest in cloud security improved. Therefore, security measures should be established to prepare for the ever-increasing usage of cloud services.

Key words : Cloud security, LDA Topic Modeling, Big data, Trend Analysis

접수일(2022년 05월 31일), 수정일(2022년 06월 17일),
게재확정일(2022년 06월 20일)

* 중앙대학교/융합보안학과(주저자)

** 중앙대학교/산업보안학과(교신저자)

★ 본 연구는 과학기술정보통신부 및 정보통신기획 평가원 대학(CT)연구센터지원사업의 연구결과로 수행되었음.

(IITP-2022-2020-0-016555)

1. 서 론

2020년 2월 국내 COVID-19 확진자가 증가하면서 대규모 실내 집합 등을 금지하는 사회적 거리두기 정책이 시행되었고, 이에 따라 많은 기업에서 재택근무를 도입했다. 그러나 재택근무로 인해 몇 가지 문제가 발생했다. 기업은 민감한 사내 정보를 보호하기 위해 외부 인터넷의 접근을 차단하는 별도의 내부 네트워크를 운영한다. 하지만 재택근무 시 직원들이 개인 PC에서 사내 정보에 접근할 수 있어야 하므로 방화벽을 열어줘야 하지만, 이를 통해 공격자는 개인 PC를 통해 사내 네트워크로 침입하여 사내 정보에 접근하게 되면 복사와 사진 촬영 등을 통해 사내 정보를 외부로 유출하기 때문에 기업은 유출 여부를 확인할 수 없다. 실제로 SERVEONE 조사에 따르면 국내 바이오 기업들에 대한 공격 시도는 (그림 1)과 같이 2019년 12월 9건, 2020년 1월 16건, 2월 23건, 3월 53건, 4월 352건, 5월 401건으로 COVID-19 대유행이 시작된 이후 공격 시도가 계속 증가하였다 [1].



(그림 1) 국내 바이오 기업 공격 시도

기존에 동향 파악을 위해 흔히 사용한 방법은 델파이 기법이다. 델파이 기법은 전문가들의 합의에 의해 이뤄진다 [2][3][4]. 하지만 절차가 복잡하고 시간 및 비용이 많이 발생하며, 소수 전문가의 의견이 주된 영향을 발휘할 수 있다 [5].

이에 본 연구에서는 전문가들의 합의로 동향 파악하는 것이 아닌 국내 시장 동향을 파악할 수 있는 뉴스 데이터를 수집해 토픽 모델링을 했다. 토픽 모델링 [6]이란 비구조화된 문서에서 추상적인 토픽을 찾기 위한 통계적 모델이다. 사람이 문서 작성 시 토픽을 선정하고, 그에 관련된 단어들을 바탕으로 작성하게 되는데 이와 같은 과정을 역추적하여 문서의 토픽을 추출한다 [7].

본 연구에서 COVID-19 대유행 전후의 클라우드 보안 동향 변화를 분석한 결과 COVID-19 대유행 이전은 국내 클라우드 시장이 활성화되지 않았고, COVID-19 대유행 이후 국내 클라우드 시장이 커진 것을 확인할 수 있었다. 다만, 클라우드 서비스의 사용량이 증가할수록 보안에 대한 위협도 증가하였으나 이에 대한 보안 대책은 미비했다. 따라서 보안 대책을 수립하고 사용자에게 클라우드 보안의 중요성을 인식시켜야 한다.

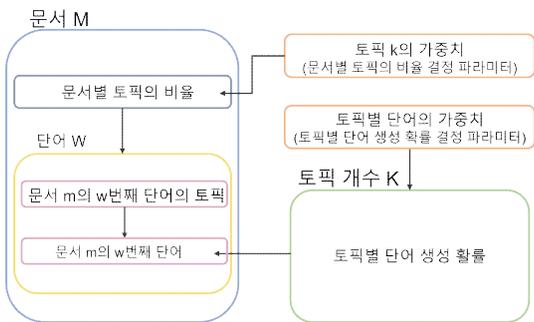
2. 관련 연구

2.1 토픽 모델링

토픽 모델링을 수행하는 알고리즘은 LSA (Latent Semantic Analysis) [8]와 LDA (Latent Dirichlet Allocation) [9] 기법이 존재한다. 문장 내 단어의 등장 횟수를 표현하여 Bow(Bag-of-words) 벡터로 행렬을 표현한 DTM(Document term matrix)이나, 의미 없는 단어들의 가중치를 낮게 계산하여 행렬에서 제거한 TF-IDF(Term Frequency-Inverse Document Frequency)의 경우 문서에서 개별 단어의 등장 횟수에 따라 중요도가 달라져 단어의 잠재적인 의미에 대한 고려는 간과되었다. 단어 내에 잠재된 의미를 도출하는 알고리즘이 LSA 모델로 각 단어가 갖는 의미가 있다고 가정한다. 토픽별 단어 개수에 따른 분포를 바탕으로 분석해 해당 문서가 어떤 토픽을 포함하는지 예측이 가능하다. 하지만 단어에 잠재된 의미를 파악하기 위해 가중치를 부여해야 하고, SVD(Singular Value Decom

position) 계산이 필요하다는 단점이 존재한다. LSA 모델의 단점을 보완한 (그림 2)의 LDA 모델은 여러 토픽들이 혼합되어 문서가 구성되어 있고, 각 토픽에 대한 확률 분포에 기반하여 단어를 생성한다고 가정한다.

LDA 모델의 과정은 먼저 M개의 문서에서 단어 W를 추출하고 각 문서 m에서 추출된 단어 w에 임의의 토픽 번호를 부여한다. 그 후 문서와 토픽 내 단어 분포를 계산하기 위해 문서 내 임의의 토픽 번호의 개수에 토픽 k의 가중치 값을 더한다. 여기서 토픽 k의 가중치 값은 0이 아닌 값을 지정해 주어야 한다. 계산 결과를 임의로 배정한 단어 w에 다시 할당해 준 후, 각각의 m 문서에 단어 w가 나타날 확률을 계산하는 과정을 반복하여 가장 높은 확률을 가진 토픽에 해당 단어와 문서를 분류하는 것으로 LDA 기법이 이루어진다.



(그림 2) LDA 모델

2.2 LDA 토픽 모델링을 활용한 동향 분석 연구

Yang 외 2인 [10]은 스마트 팩토리 연구의 발전과 도입을 위해 과거부터 현재까지의 연구 동향을 파악할 필요성을 느끼고, 국가적 사례를 바탕으로 LDA 토픽 모델링을 활용해 스마트 팩토리 연구 동향을 분석했다. 그 결과 해당 연구가 특정 국가에서만 이뤄지는지 국제적으로 이뤄지는지 확인할 수 있었다.

Park 외 2인 [11]은 과학기술의 동향을 예측하기 위해 LDA 토픽 모델링을 활용해 미국 특허

문서에서 AI 세부기술을 20가지 추출하여 Hot 기술과 Cold 기술로 분류했다. 그 결과 Hot 기술은 최근 이슈인 딥러닝과 관련된 토픽이 추출됐고, Cold 기술은 1990년대 후반부터 등장한 머신러닝에 필요한 기술과 연관성이 있었다.

Lee 외 4인 [12]은 DT(Digital transformation)의 시장이 빠르게 변화하고 성장하기 때문에 기술 연구 및 개발을 분석하기 위해 99개의 기사에 대해 토픽 모델링을 수행했다. 이를 통해 DT를 효과적으로 수행하기 위한 기본 지식을 추출했고, 첨단 제조 공업에서의 DT 연구 및 개발에 도움을 주었다.

Sun 외 1인 [13]은 교통 연구는 과학과 공학에서 중요하기 때문에 LDA 토픽 모델링 기법을 사용해 16년간 주요 교통 저널에 게재된 17,163건을 분석하여 50개의 핵심 주제를 추출했다. 이를 통해 지속 가능한 교통 및 무동력 이동수단에 대한 주제가 최근 점점 주목을 받고 있다는 것을 확인했다.

Choi 외 1인 [14]은 물류 분야의 기술 동향을 분석하기 위해 USPTO(United States Patent and Trademark Office)에 등록된 물류 관련 특허를 LDA 토픽 모델링 기법을 사용해 분석했다. 전통적인 물류 영역인 해운 서비스, 컨테이너, 제품정보시스템 등은 줄어들고, 데이터베이스와 센싱을 통한 기술이 주목을 받고 있다는 것을 확인했다.

Choi 외 2인 [15]은 개인 정보 프라이버시에 대한 연구 동향을 살펴보기 위해 Scopus DB에서 34년 동안 출판된 2356개의 문서를 LDA 토픽 모델링 기법을 사용해 분석했다. 그 결과 디지털 시대 개인 정보 프라이버시에 대한 우려가 증가했고, 알고리즘을 통한 개인 정보의 프라이버시를 보장하기 위한 연구가 많이 진행되고 있다는 것을 확인했다.

서예령 외 2인 [16]은 COVID-19 대유행 시기를 3개의 기간으로 분류하여 LDA 토픽 모델링 기법을 사용해 각 기간 별 마스크 정책에 대한 언론보도의 변화를 분석했다. 이를 통해 향후 COVID-19 대유행과 같은 위기 상황에서 거시적

인 담론 형성을 통해 사회 전체의 통합적인 대응에 언론이 기여할 수 있는 점을 제시했다.

기존 연구들을 통해 LDA 토픽 모델링이 동향 파악에 사용되었고, 최근 동향 파악을 통해 앞으로 다가올 변화에 유연하게 대처할 수 있다. 따라서 LDA 토픽 모델링 기법을 사용해 본 연구 주제인 클라우드 보안에 대한 동향을 파악하고자 한다.

3. 연구방법

3.1 연구 프레임워크

본 연구에서 실험은 Google에서 제공하는 Colaboratory(Colab) 환경에서 수행했고, 데이터 분석을 위해 Intel(R) Xeon(R) CPU @ 2.20GHz와 13GB의 메모리를 사용했다. 본 연구의 연구 프레임워크는 (그림 3)과 같다. 한국언론진흥재단에서 운영하는 뉴스빅데이터 분석 서비스인 BIGKinds★를 통해 HTML과 XML 문서에서 원하는 데이터를 추출할 수 있는 라이브러리인 BeautifulSoup을 사용하여 2016년부터 2022년 5월 4일까지의 ‘클라우드 보안’ 키워드의 뉴스 데이터 24,381건을 수집했다. 연도별 클라우드 보안 동향의 변화를 확인하기 위해 <표 1>과 같이 2016년 766건, 2017년 1368건, 2018년 4095건, 2019년 8215건, 2020년 4502건, 2021년 4189건, 2022년 1246건으로 분류했다.



(그림 3) 연구 프레임워크

<표 1> 연도별 분석 데이터

수집	연도별	데이터 개수
BIG Kinds	2016년	766 건
	2017년	1,368 건
	2018년	4,095 건
	2019년	8,215 건
	2020년	4,502 건
	2021년	4,189 건
	2022.01.01~2022.05.04	1,246 건
	총합	24,381 건

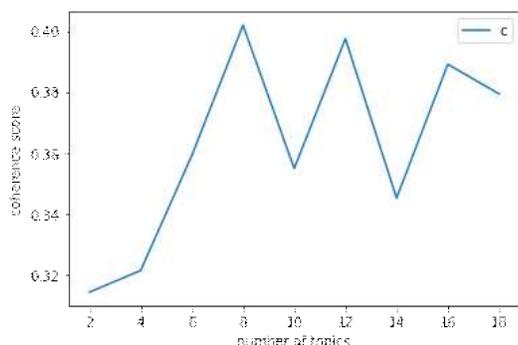
3.2 데이터 전처리

본 연구에 사용한 국내 뉴스 데이터에서 동향 분석을 방해하는 숫자, URL, 특수문자, 전치사 및 대명사를 제거해 하나의 단어에 대한 어간 추출 작업을 해야 한다 [17]. 한국어 자연어 처리를 위해 파이썬에서 제공하는 KoNLpy Library를 사용하여 연도별 분류한 데이터로부터 0에서 9까지 숫자, 대소문자를 모두 포함한 영단어, '@', '~', '%'와 같은 기호 및 특수문자를 제거했다. 그리고 KoNLpy Library의 okt.nouns()를 사용하여 한글에서 형태소를 분석해 명사 단어를 추출하는 단계를 거쳐 ‘은, 는, 이, 가, 을, 를, 와, 과’와 같은 불필요한 조사와 ‘이것, 저것’과 같은 불필요한 대명사를 제거했다.

3.3 LDA 토픽 모델링 설계

본 연구에서 LDA 토픽 모델링은 파이썬 gensim 패키지로 수행했다. Newman, David, et al.[18]에서 모델이 실제 의미 있는 결과를 내는지 확인하기 위해 사용하는 gensim 패키지의 CoherenceModel 연산을 통해 토픽 내에서 단어의 의미가 일치하는지 수치로 계산하는 일관성 점수를 산출하여 토픽 개수를 정했다. 가장 높은 일관성 점수를 가진 토픽 수 값이 해당 문서의 적절한 토픽 개수이다.

★ <https://www.bigkinds.or.kr/>



(그림 4) 토픽 개수에 따른 일관성 점수

LDA 토픽 모델링 수행 시 필요한 샘플링 반복 횟수는 [16]에서 적용한 500회로 설정했고, CoherenceModel 연산을 수행하여 일관성 점수가 가장 높은 토픽 개수를 선정했다. (그림 4)의 연산결과 COVID-19 대유행 전인 “2016년부터 2019년” 뉴스 데이터 문서는 토픽 개수를 10개로 선정했고, COVID-19 대유행 후인 “2020년부터 2022년 5월 4일” 뉴스 데이터 문서 토픽 개수를 8개로 선정했다.

파이썬에서는 pyLDAvis Library를 통해 LDA 토픽 모델링의 학습된 결과를 시각화할 수 있도록 지원하고 있다. (그림 5)와 (그림 6)에서 원의 개수가 일관성 점수로 선정된 문서 내 토픽의 개수이다. 원의 크기가 클수록 해당 토픽이 문서 내 차지하는 비중이 크다는 것을 의미하며, 원들 사이의 위치가 멀수록 토픽을 이루고 있는 단어들이 서로 연관성이 낮아 대표적인 특징을 표현하는 토픽을 추출할 수 없고, 원들 사이의 위치가 가까울수록 토픽을 이루고 있는 단어들이 비슷하여 대표적인 특징을 표현하는 토픽을 찾을 수 있다.

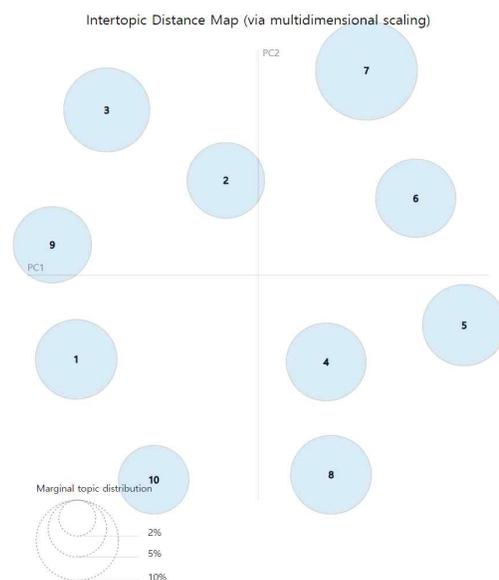
4. 연구 결과

4.1 COVID-19 대유행 이전 분석 결과

<표 2>와 <표 3>의 ‘토픽 번호’는 일관성 점수로 선정한 토픽 개수를 토대로 각 토픽에 임

의로 부여한 번호이고, ‘문서 내 비율’은 해당 토픽 번호가 문서 내 차지하는 비율이며, 각 토픽별 상위 10개 단어를 작성했다.

<표 2>에서 데이터 문서 내 차지하는 토픽들의 비율이 큰 차이를 보이지 않고, (그림 5)에서 10개의 토픽을 표현하는 각 원과의 거리가 멀어 대표적인 특징으로 표현되는 토픽을 찾을 수 없었다. <표 2>의 다른 토픽에서도 ‘보안’ 단어가 추출됐지만, 추출된 다른 단어를 보았을 때 해당 토픽이 클라우드 보안에 대한 토픽이라고 분석할 수 없었다. <표 2>의 토픽 2번에서 ‘보안’, ‘솔루션’, ‘해킹’, ‘유출’ 단어가 추출된 것을 확인했고, 토픽 2번을 클라우드 보안의 토픽으로 볼 수 있으나 데이터 문서 내 차지하는 비율이 8.9%로 10개의 토픽 중 9위로 낮은 순위이다. 따라서 COVID-19 대유행 전 국내 클라우드 보안에 대한 관심이 낮은 것으로 볼 수 있다. 하지만 토픽 3번에서 ‘AI’, ‘IoT’, ‘차세대’, ‘블록체인’, ‘인공지능’, ‘개발’ 단어가 추출되어 높은 컴퓨팅 성능을 필요로 하는 IT 기술을 위해서는 클라우드의 도입이 필요하다는 것을 알 수 있다.



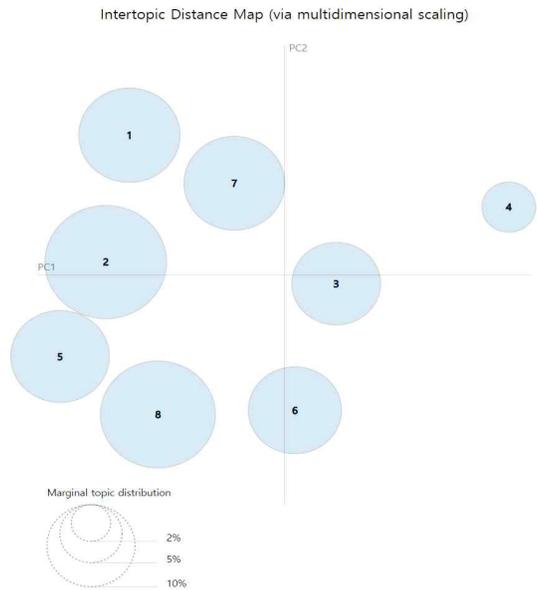
(그림 5) COVID-19 대유행 이전 시각화 결과

4.2 COVID-19 대유행 이후 분석 결과

(그림 6)을 보면 데이터 문서 내 대부분의 비율을 차지하는 토픽 1번, 2번, 5번, 7번, 8번 원의 위치가 서로 가까워 대표적인 특징으로 표현되는 토픽들을 찾을 수 있었다. 이를 <표 3>에서 확인하면 토픽 2번, 7번, 5번에서 ‘서비스’, ‘기업’, ‘운영’, ‘고객’, ‘업무’, ‘솔루션’, ‘사업’, ‘매출’, ‘성장’ 등의 단어를 통해 국내에서 클라우드를 COVID-19 대유행 이후 많이 도입한 것으로 분석할 수 있다. 추출된 단어들의 연관성을 분석해 클라우드 보안에 대한 토픽으로 분석한 토픽 6번과 3번의 경우 COVID-19 대유행 이전과 비교하여 데이터 문서 내 차지하는 비율이 8.9%에서 토픽 6번과 3번을 합친 21.3%로 증가했다. 각 토픽에 대한 단어도 ‘보안’, ‘정보’, ‘네트워크’, ‘보호’, ‘인증’, ‘공격’, ‘해킹’, ‘사이버’ 등 다양해진 것을 확인했고, 이는 클라우드 보안에 대한 관심이 증가하였다고 볼 수 있다. 또한 <표 3>에서 COVID-19 대유행 시 클라우드의 활용성에 대해 문서 내 비율은 3.7%로 8개의 토픽 중 8위로 가장 낮은 순위를 차지하지만, ‘의료’, ‘병원’, ‘헬스’, ‘환자’ 단어가 추출된 것을 보아 실제 COVID-19 대유행 이후 의료기관에서 클라우드에 대한 관심이 높아졌음을 알 수 있다.

2016년부터 2022년 5월 4일까지의 국내 클라우

드 보안 시장 동향을 분석한 결과 COVID-19 대유행 이전인 2016년부터 2019년까지 4년간은 국내에서 클라우드 보안에 대한 관심이 낮은 것을 확인할 수 있었다. 하지만 COVID-19 대유행 이후인 2020년부터는 국내 클라우드 도입이 활성화됨에 따라 클라우드 보안에 대한 토픽들도 다양하게 추출되었다.



(그림 6) COVID-19 대유행 이후 시각화 결과

<표 2> COVID-19 대유행 이전 LDA 토픽 모델링 결과

기간	토픽 번호	문서 내 비율 (%)	상위 10개 단어
2016 ~ 2019	7	15.3%	클라우드, 서비스, 보안, 시장, 금융, 인증, 솔루션, 공략, 공공, nhn
	3	10.9%	기술, AI, IoT, 시대, 차세대, 블록체인, 기업, 인공지능, 개발, CCTV
	5	10.2%	클라우드, 보안, 서비스, 도입, 하이브리드, 금융사, 전환, 전용, 금융, AWS
	1	9.8%	SD, 서비스, AI, 기업, SW, 베트남, 보안, 솔루션, 세계, 블록체인
	8	9.6%	클라우드, 기업, 로봇, 플랫폼, 더존비즈온, 글로벌, 국내, 데이터, 보안서비스, 코스콤
	6	9.5%	클라우드, SD, 시장, 역원, 분기, 파트너, 사업, 플랫폼, 영업익, 브로드밴드

	4	9.4%	보안, 최대, 시대, 매출, 클라우드, 기업, ICT, 아마존, AWS, 서밋
	9	9.1%	빅데이터, 데이터센터, 삼성전자, 전자, 생태계, AI, 보안, 블록체인, 암호화폐, 플랫폼
	2	8.9%	디지털, 보안 , 전환, 혁신, 기업, 글로벌, 솔루션 , MOU, 해킹 , 유출
	10	7.3%	상위, 중국, 페이스북, 구글, 영업이익, IoT, 포럼, 상승률, 플랫폼, 에너지

<표 3> COVID-19 대유행 이후 LDA 토픽 모델링 결과

기간	토픽 번호	문서 내 비율 (%)	상위 10개 단어
2020 ~ 2022 /05 /04	2	19%	클라우드, 서비스 , 데이터, 기업 , 제공, 플랫폼, 운영 , 고객 , 관리, 보안
	8	17%	디지털, 산업, 기업, AI, 분야, 지원, 데이터, 사업, 정부, 교육
	1	13.2%	미국, 기업, 중국, 세계, 한국, 게임, 데이터, 시장, 글로벌, 구글
	7	13.1%	업무 , 솔루션 , 기업 , 서비스 , 제공, 환경, 관리, 지원, 클라우드, 보안
	5	12.6%	사업 , 서비스 , 매출 , 성장 , 기업 , 시장, 대표, 보안, 확대, 플랫폼
	6	11.2%	보안 , 솔루션 , 정보 , 네트워크 , 보호 , 서비스, 기업, 인증 , 제공, 관리
	3	10.1%	보안 , 인증 , 정보 , 공격 , 서비스, 해킹 , 클라우드, 사이버 , 이용, 금융
	4	3.7%	의료 , 반도체, 병원 , 종목, 장비, 헬스 , ETF, 클라우드, 시장, 환자

5. 결 론

본 연구는 2016년부터 2022년 5월 4일까지의 ‘클라우드 보안’ 키워드로 검색한 뉴스 데이터를 수집하여 COVID-19 대유행이 시작된 2020년을 기준으로 기간을 나누어 LDA 토픽 모델링을 수행했다. 분석 결과 COVID-19 대유행 이전인 2016년부터 2019년까지는 국내 클라우드 보안의 대표성을 나타낼 토픽을 찾을 수 없었고, 클라우드 보안에 대한 관심도 낮았다. 하지만 COVID-19 대유행이 시작된 이후인 2020년부터 기업에서 서비스 및 운영에 클라우드 서비스를 사용하면서 국내 클라우드 시장이 크게 성장한 것을 확인할 수 있었고, 클라우드 보안에 대한 관심이 증가했다.

COVID-19와 같은 전 세계의 전염병 대유행은 새로운 IT 기술의 급격한 확산을 초래한다. COVID-19의 경우 바이러스 확산을 통제하는 사회

두기 정책으로 인한 재택근무의 대비책으로 다양한 분야에 클라우드 서비스의 도입을 초래했다. 본 연구에서 알 수 있듯이 새로운 IT 기술의 도입 시 해당 기술의 보안 측면도 함께 관심도가 높아진다. 따라서 새로운 IT 기술을 기존 시스템에 도입할 시 보안 대책을 충분히 고려해야 할 것이다.

ACKNOWLEDGEMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2022-2020-0- 01655) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

참고문헌

- [1] SERVEONE, “재택근무 확대, 기업정보 해킹으로부터 보호하자”, 2020
- [2] Y. S. Choi and S. C. Baek, and H. I. Kwon, “Study on Ubiquitous City Revitalization Plan Via u-City Project”, *The Journal of Internet Electronic Commerce Research* 8.3: 183-209, 2008.
- [3] 최용석, 백승철, 권혁인, “텔파이기법을 이용한 U-city 사업의 핵심성공요인 도출”, *인터넷전자상거래연구*, 8.3: 183-209, 2008.
- [4] 김종석, 강진원, “텔파이기법을 이용한 제 4차 산업혁명의 주요 기술변화 분석”, *사회과학연구*, 57.1: 205-234, 2018.
- [5] J. S. Part and S. G. HONG and J. W. KIM, “A study on science technology trend and prediction using topic modeling”, *Journal of the Korea Industrial Information Systems Research*, 22.4: 19-28, 2017.
- [6] 남춘호, “일기자료 연구에서 토픽모델링 기법의 활용가능성 검토”, *비교문화연구*, 22, 2016.
- [7] D. M. Blei, “Probabilistic topic models”, *Communications of the ACM*, 55(4), 77-84, 2012.
- [8] S. Deerwester and S. T. Dumais and G. W. Furnas and T. K. Landauer and R. Harshman, “Indexing by latent semantic analysis”, *Journal of the American society for information science*, 41(6), 391-407, 1990.
- [9] D. M. Blei and A. Y. Ng and M. I. Jordan, “Latent dirichlet allocation”, *Journal of machine Learning research*, 3(Jan), 993-1022, 2003.
- [10] H. L. Yang and T. W. Chang and Y. Choi, “Exploring the research trend of smart factory with topic modeling”, *Sustainability*, 10.8: 2779, 2018.
- [11] J. S. Park and S. G. Hong and J. W. Kim, “A study on science technology trend and prediction using topic modeling”, *Journal of the Korea Industrial Information Systems Research*, 22.4: 19-28, 2017.
- [12] C. H. Lee and C. L. Liu and A. J. Trappey and J. P. Mo and K. C. Desouza, “Understanding digital transformation in advanced manufacturing and engineering: A bibliometric analysis, topic modeling and research trend discovery”, *Advanced Engineering Informatics* 50 : 101-128, 2021
- [13] L. Sun and Y. Yin, “Discovering themes and trends in transportation research using topic modeling”, *Transportation Research Part C: Emerging Technologies*, 77: 49-66, 2017.
- [14] D. Choi and B. Song, “Exploring technological trends in logistics: Topic modeling-based patent analysis”, *Sustainability*, 10.8: 2810, 2018
- [15] H. S. Choi and W. S. Lee and S. Y. Shin, “Analyzing research trends in personal information privacy using topic modeling”, *Computers & Security*, 67: 244-253 2017.
- [16] 서예령, 고금석, 이재우, “빅데이터 LDA 토픽 모델링을 활용한 국내 코로나 19 대유행 기간 마스크 관련 언론 보도 및 태도 변화 분석”, *한국정보통신학회논문지*, 25.5: 731-740, 2021.
- [17] 이상연, 이건명, “토픽 모델링을 이용한 덧

글 그래프 기반 소셜 마이닝 기법”, 한국지능시스템학회 논문지, 24.6: 640-645, 2014.

[18] D. Newman and J. H. Lau and K. Grieser and T. Baldwin, "Automatic evaluation of topic coherence.", Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics. 2010.

————— [저 자 소 개] —————



이 선 우 (Soun U Lee)
중앙대학교 융합보안학과 석사과정
관심분야 : 네트워크 보안, 인공지능, 블록체인, 사이버 물리 시스템
email : sounulee@gmail.com



이 재 우 (Jaewoo Lee)
2006년 서울대학교 컴퓨터공학부(학사)
2008년 서울대학교 컴퓨터공학부(석사)
2017년 University of Pennsylvania, Ph.D in Computer and Information Science
2018년~현재 중앙대학교 산업보안학과 조교수
관심분야 : 실시간 시스템, 사이버 물리 시스템 보안
email : jaewoolee@cau.ac.kr