

텍스트마이닝 통한 데이터 3법 개정을 기준으로 개인정보에 대한 인식 변화 연구

이연수, 이예슬, 이옥한*
숙명여자대학교, *숙명여자대학교

{Pluto2002p99, lystanz}@sookmyung.ac.kr, *uhlee@sookmyung.ac.kr

A Study on Changes in perception of Personal Information through Text Mining

Lee Yeon Su, Lee Ye Seul, Lee Uk Han*
Sookmyung Women's Univ., *Sookmyung Women's Univ.

요 약

본 논문은 개인정보에 대한 인식의 변화를 연구하기 위해 뉴스 빅데이터를 시기별로 구분하여 넷마이너(NetMiner)로 분석하였다. 뉴스에 나타난 개인정보 관련 자료를 시기별로 구분하여 키워드의 변화를 워드 클라우드(Wordcloud)를 통해 제시함으로써 향후 개인정보의 의의와 방향성을 제시하고자 한다.

I. 서 론

현대사회는 초연결사회로 일컬어지는 4차 산업혁명의 시대이다. 4차 산업혁명 시대에 들어서 새로운 산업과 서비스를 창출하기 위한 주요 자원으로서 데이터의 시장가치가 높아지고 있다.[1] 문제는 '21세기 원유'라고 지칭되는 이들 빅데이터가 개개인에 관한 정보의 집합물이기에 빅데이터의 활용 가능성을 높일수록 개인정보의 침해 위험성도 그만큼 높아진다는 것이다.[2] 빅데이터 산업의 발전 측면에서 보자면, 개인정보를 가급적 두텁게 보호하고자 했던 그간의 우리나라 개인정보 보호법제가 산업성장의 가장 큰 장애물이 되어 버렸고, 이에 수년간 개인정보의 손쉬운 활용을 위해 관련 법령을 개정해달라는 유관 기관과 학계, 기업의 요청이 끊임이 없었다.[2]

이와 같은 시대적 요구에 따라 우리나라에서 2018년 11월 일명 데이터 3법(개인정보보호법, 신용정보법, 정보통신망법)이 발의되었다. 이후 2020년 1월 9일 국회 본회의를 통과하였고, 2월 4일 개정되어 8월부터 정식 시행되었다. 데이터 산업의 활성화를 위한 발판이 되고자 했던 데이터 3법은 가명정보 도입을 통한 데이터 이용 활성화, 개인정보 보호체계 일원화, 마이데이터 등 금융분야 데이터 신산업 도입, 전문기관을 통한 데이터 결합 지원 등이 주요 내용이다.[3] 이외에도 데이터 산업의 미래적 중요성을 인지하여 소위 '데이터 뉴딜 정책'과 데이터 기반 산업 및 경제 육성을 차세대 국정사업으로 지정하여 국가적 역량을 집중하고 있다.[4]

이처럼 데이터 3법의 개정에 따라 개인정보의 사회적 인식이 변화했을 것임을 예측해볼 수 있다. 따라서 본 연구에서는 개인정보의 의의가 데이터 3법의 개정을 기준으로 사회적으로 어떠한 인식의 변화가 이루어졌는지에 대해 텍스트 마이닝(Text Mining)을 통해 살펴보고자 한다.

II. 연구방법

본 연구의 분석대상은 뉴스 기사로 선정하며, 전국일간지, 경제일간지, 지역일간지, 방송사, 전문지 등을 포함하였다. 뉴스 기사는 한국언론진흥재단에서 구축한 빅카인즈(BIG KINDS)에서 '개인정보'를 키워드로 추출하였다. 2018년 2월 4일부터 2022년 2월 4일까지 데이터 3법의 개정 시점을 기준으로 두 시기로 구분하였으며, 총 30,889건의 뉴스 데이터를 수집하였다. 분석 대상에 대한 시기별 뉴스 데이터 정보는 아래(표 1)와 같다.

표 1. 분석 대상 뉴스 건수(건)

시기	건수
1기 데이터 3법 개정 전 (2018.02.04-2020.02.03)	13,246
2기 데이터 3법 개정 후 (2020.02.04-2022.02.04)	17,643

본 연구를 위해 빅데이터 분석 프로그램 넷마이너(NetMiner)를 활용하였다. 데이터 정제는 사용자 지정 사전을 통해 유의어, 지정어, 제외어 설정을 통해 이루어졌다. 먼저, 유의어의 경우 '배달의 민족'과 '배달의민족'이 있는 경우 띄어쓰기로 인해 '배달'과 '민족' 혹은 '배달의민족' 세 가지 경우로 나누어 분석될 수 있어 관련 용어를 묶어서 유의어로 지정하였다. 지정어의 경우 '신용정보', '암호화폐' 등과 같이 단어와 단어가 합쳐져 하나의 고유명사로 사용하는 단어들을 지정어로 설정하였다. 마지막으로 텍스트 내에서 상투적으로 등장하는 '최근', '보도', '소속' 등의 단어들을 제외어로 설정하였다.

데이터 정제 후 넷마이너를 통해 단어 빈도분석을 실시하여 빈출 상위 500개 단어를 워드 클라우드로 시각화 하였으며, 상위 키워드 20개를 추출하여 의미를 분석하였다.

III. 연구결과

시기별 워드 클라우드 결과는 다음(그림 1, 2)과 같다. 빈출 상위 500 개의 단어에 대해 워드 클라우드를 시각화 하였으며, 시기별로 상위 키워드 20 개에 대한 의미를 분석하였다.

그림 1. 1 기 워드 클라우드 결과



1 기 상위 키워드 20 개는 다음과 같다. 정보(5,991), 보호(2,832), 데이터(2,796), 페이스북(2,774), 경제(2,569), 규제(2,563), 금융(2,479), 기술(2,429), 정부(2,366), 산업(1,999), 국회(1,916), 미국(1,757), 활용(1,751), 인공지능(1,714), 보안(1,650), 관리(1,444), 한국(1,432), 공개(1,357), 혁신(1,291), 정책(1,257) 이다.

1 기 상위 키워드에 대한 분석 결과는 해당 시기 미국의 경우 주식 시장에서 구글, 페이스북 등 기업의 데이터 독주에 따른 우려의 목소리가 등장하였다. 이에 따른 디지털 산업에서의 감독 및 규제와 기업의 자립과 데이터 활용의 갈등이 대두되었다.

한편 우리나라의 경우 인공지능 신산업 및 정보기술의 발달에 따라 데이터 활용 시대에 대한 활발한 논의가 이루어졌다. 그 결과 2020 년 1 월 데이터 3 법이 국회 본회의를 통과하였으며, 정부 주도의 공공 서비스 정책, 과학적 의사결정 체계 구축, 데이터 산업을 핵심 정책 과제로 선정하였다. 특히 데이터 3 법 중 신용정보법에서의 마이데이터 도입에 따라 개인정보가 데이터로서 금융권에서의 활용 가능성을 높이게 되었다.

그림 2. 2 기 워드 클라우드 결과



2 기 상위 키워드 20 개는 다음과 같다. 정보(6,056), 보호(3,662), 정부(3,021), 경제(2,992), 데이터(2,678), 기술(2,503), 금융(2,420), 인공지능(2,405), 중국(2,395), 미국(2,006), 보안(2,004), 개인정보보호위원회(1,979), 한국(1,899), 디지털(1,714), 활용(1,679), 코로나바이러스(1,642), 공개(1,641), 인증(1,506), 사업(1,485), 정책(1,392) 이다.

2 기 상위 키워드에 대한 분석 결과는 우선 개인정보 유출 및 보호에 관한 문제가 활발히 논의됨을 파악할 수

있다. 미국의 경우 구글이 이용자들을 속이고 지속적으로 이용자의 위치 추적을 한 혐의로 소송을 당하였다. 중국의 경우 중국의 글로벌 숏폼(short form) 모바일 비디오 플랫폼인 틱톡(TikTok)에서 개인정보 유출에 따른 개인정보 침해 문제가 제기되어 국제적으로 논란이 발생했다. 또한 신종 코로나 사태라는 국가적 재난 상황이 발생함에 따라 개인정보의 공적 차원에서 이용이 불가피하게 발생하였다. 특히 개인의 이동 동선이 유출되어 개인이 특정됨에 따라 사생활 침해가 심각한 문제로 논의되었다.

한편 데이터 3 법의 시행에 따라 정보주체의 주권 강화와 데이터 경제 활성화 정책이 활발히 시행되었다. 또한 개인정보 유출 등에 대한 감독기구가 개인정보보호위원회로 일원화됨에 따라 중앙행정기관으로서 역할이 상당해졌음을 확인할 수 있다. 공공기관 개인정보 관리 수준 진단 평가 실시, 페이스북 운영사 메타를 상대로 개인정보 제 3 자 무단 제공에 대한 법정 분쟁을 진행했다. 특히 메타와의 소송은 개인정보보호위원회가 기업을 상대로 한 첫 법정 분쟁으로서 그 관심이 컸던 만큼 키워드 분석이 의미 있게 진행되었음을 알 수 있다.

IV. 결 론

본 연구는 뉴스 기사에서 개인정보라는 키워드를 중심으로 관련 빈출 키워드를 시기별로 추출함으로써 개인정보의 인식 변화에 대해 분석하였다. 그 결과, 초연결사회에서의 개인정보 의미가 데이터 3 법의 개정을 기준으로 변화했음을 확인할 수 있었다.

4 차 산업혁명 시대에서 개인정보는 인격권으로서 보호 대상에 해당하는 헌법적 가치와 데이터로서 활용의 범위에 해당하는 물적 가치를 동시에 지니고 있다. 특히 급격한 사회 변화에 따라 개인정보의 데이터로서 가치는 핵심적 역할을 수행하게 되었음을 연구를 통해 알 수 있다. 따라서 개인정보의 보호와 활용 간의 균형적 규제 체제를 위한 노력이 지속되어야 할 것이다.

참 고 문 헌

- [1] Jeong-Soo, K. "A Study on the Relationship between Legislation and Protection of Fundamental Rights in Intelligence Information Society-focus on the Protection of Personal Information in Data 3 Act-", pp.339-367, Nov. 2020.
- [2] Seung Hee, J. "Main Changes and Criminal Issues of the Revisions of Three Data Privacy Laws in 2020", Soongsil Law Review, pp.267-296, Jan. 2021.
- [3] Taewon, L., Sunghoon, C. "A Study on Social issues Related to Personal Information Protection Act based on Text Analysis", pp.27-48, Aug. 2020.
- [4] Eun-Chan, K., Eun-Young, K., Hyo-Chan, L., Byung-Joon, Y. "The Details and Outlook of Three Acts Amendment in South Korea: With a Focus on the Changes of Domestic Financial and Data Industry", pp.49-72, Sep. 2021.