

빅데이터 분석을 통한 프라이버시 인식에 관한 연구

이송이¹, 김성원², 이환수^{3*}

¹단국대학교 IT법학협동과정 석사과정, ²단국대학교 IT법학협동과정 박사과정,
³단국대학교 융합보안학과 교수

A Study on the Privacy Awareness through Bigdata Analysis

Song-Yi Lee¹, Sung-Won Kim², Hwan-Soo Lee^{3*}

¹Student, Department of Information Technology and Law, Dankook University

²Ph.D, Student, Department of Information Technology and Law, Dankook University

³Professor, Department of Convergence Security, Dankook University

요 약 4차 산업혁명시대에 접어들면서 정보기술의 발전으로 인해 다양한 편익과 함께 프라이버시 이슈에 대한 사회적 관심 또한 증가해 왔다. 빅데이터를 통한 개인의 프라이버시 침해 위협 가능성이 높아지게 됨에 따라, 프라이버시에 대한 관리 및 보호에 대한 학술적 논의 또한 활발해지기 시작하였다. 전통적인 관점의 프라이버시는 기본적 인권으로 다양한 차원에서 정의된 반면, 최근의 연구 동향에 따르면 대부분 온라인상에서의 개인정보보호라는 정보 차원의 프라이버시에 대해서만 주로 논의하고 있다. 이러한 제한적 논의는 이론적 개념과 실제 인식 간의 왜곡을 초래할 뿐만 아니라 프라이버시 개념의 학술적 정의 및 사회적 합의를 더욱 어렵게 만드는 요인이 될 수 있다. 이에 본 연구에서는 최근 1년 동안 포털 사이트 12,000건의 뉴스데이터를 바탕으로 온라인상에서 노출되는 프라이버시 개념을 분석하여 이론적 개념과 사회적으로 통용되는 개념의 차이를 비교 분석한다. 이러한 실증적 접근은 변화하고 있는 프라이버시 개념에 대한 이해와 국내 상황에 맞는 프라이버시 개념화를 위한 연구 방향을 제시할 수 있을 것으로 기대한다.

주제어 : 프라이버시, 인식, 빅데이터, 텍스트마이닝, 뉴스

Abstract In the era of the 4th industrial revolution, the development of information technology brought various benefits, but it also increased social interest in privacy issues. As the possibility of personal privacy violation by big data increases, academic discussion about privacy management has begun to be active. While the traditional view of privacy has been defined at various levels as the basic human rights, of the recent research trends are mainly concerned only with the information privacy of online privacy protection. This limited discussion can distort the theoretical concept and the actual perception, making the academic and social consensus of the concept of privacy more difficult. In this study, we analyze the privacy concept that is exposed on the internet based on 12,000 news data of the portal site for the past one year and compare the difference between the theoretical concept and the socially accepted concept. This empirical approach is expected to provide an understanding of the changing concept of privacy and a research direction for the conceptualization of privacy for current situations.

Key Words : Privacy, Awareness, Big data, Text-mining, News

*This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2018S1A5A8027174)

*Corresponding Author : Hwan-Soo Lee(hanslee992@gmail.com)

Received July 18, 2019

Revised August 20, 2019

Accepted October 20, 2019

Published October 28, 2019

1. 서론

인공지능, 3D 프린팅, 빅데이터 기술 등 4차산업혁명에 따른 기술발전으로 고도화되고 개인화된 서비스로 삶의 질은 과거에 비해 크게 향상되었다. 그러나 이러한 변화는 인간의 일자리 축소, 정보의 빈부격차, 보안 문제 등 다양한 사회적 이슈를 야기하기도 하였다. 이러한 사회적 이슈 중 기술발전에 따른 대표적 부작용으로 거론되는 것이 바로 개인의 프라이버시 침해에 관한 것이다[1]. 대표적인 예로 빅데이터 기술을 활용 시, 방대한 양의 정보를 필연적으로 처리하게 되어 프라이버시 침해 가능성이 높아진다[2]. 디지털화된 개인정보가 인터넷을 통하여 쉽게 확산될 수 있게 되면서, 프라이버시 침해 위협은 급격히 증가하게 되었고, 프라이버시 관리 및 보호에 대한 관심이 동시에 증가하면서 이에 대한 학술적 논의가 활발해지기 시작하였다[3].

국가적 차원에서 프라이버시의 보호를 위한 많은 논의가 이루어지고 있다. 미국은 2012년 인터넷 이용자의 프라이버시 보호를 위하여 새로운 프라이버시 프레임워크를 발표하였고, EU는 2016년 5월 일반개인정보보호규정(GDPR)을 제정한 바 있다[4]. 그 외 일본에서는 최근 개인정보보호법 일부를 개정하면서, 개인정보 개념을 명확히 하여 세밀한 개인정보보호를 위한 방안을 마련 중에 있다. 또한, 개인의 프라이버시 침해에 대한 불안감이 증폭됨에 따라 기업에서도 프라이버시 보호를 위하여 다양한 노력을 추진하고 있다. 구글에서는 차세대 안드로이드 운영체제상에서 몇 차례의 정보 유출로부터 흔들렸던 서비스 보안에 대한 신뢰성을 회복하고자 노력하였다[5]. 마이크로소프트사는 GDPR의 기준을 충족시키기 위해 도구, 시스템 및 프로세스를 재설계하는데 막대한 투자를 진행하는 등 다양한 노력을 계속하고 있다[6].

그러나 프라이버시에 대한 관심이 세계적으로 증가함에도 불구하고 대부분의 논의는 기술과 정보의 관점에서만 논의하는 경향이 있다. 이러한 접근은 프라이버시에 대하여 제한적인 논의 및 법률적 기본 개념의 가치를 왜곡하는 결과로 이어질 수 있다. 원래 프라이버시는 기본적인 인권으로 정의되는 개념으로 인간의 존엄과 가치, 표현의 자유 및 결사의 자유 그리고 통신의 자유 등 다수의 인권과 관련된 광범위한 개념이다. 또한 프라이버시는 환경에 따라서 다양한 의미를 갖는 매우 복잡한 개념이다[7]. 각국마다 사회적 배경과 정치 및 경제력의 차이에 따라 프라이버시의 개념 및 보호 범위에 대한 해석을 다르게 하고 있다. 그러나 최근의 프라이버시에 대한 연구들은 이러한 광범위한 특징과 복잡성을 정보라는 일원화된 차원에서 주로 논의하고 있다.

전통적인 프라이버시 연구들은 포괄적인 개념을 구체화하거나 체계화하는 노력을 해 왔다. L.Brandeis[8]는 프라이버시를 ‘혼자 있을 권리(right to be left alone)’로 정의하였으며, Westin[9]은 프라이버시란, 어떠한 환경에서든지 자신의 신체, 태도와 행위를 타인에게 얼마만큼 노출시킬 것인가를 자유롭게 선택할 수 있는 자유라고 정의하였다[9]. 또한, Edward Bloustone[10]은 프라이버시를 인격권의 법익으로써 인격의 침해, 개인의 자주성, 존엄과 완전성을 보호하는 것이라고 정의하였다. 그러나 최근의 프라이버시 연구들은 정보 프라이버시에 집중되어 있는 것을 확인할 수 있다. 예를 들어 손영화[11]는 빅데이터 시대의 프라이버시 침해 예방 방안 설명에 있어서 개인정보에 한하여 개인정보보호를 위한 해결방안을 제시하였으며, 오태원과 유지연[3]은 정보사회에서의 프라이버시란 디지털화된 개인정보가 인터넷상에서 유통될 때, 개인정보에 대하여 개인이 주체적 권리자로서 행동할 수 있도록 그 권리의 범위가 확장되어 보다 적극적인 권리로서 변화되었다고 설명한다. 최초의 프라이버시가 불법적인 침해로부터 사생활을 보호하는 것을 의미했던 것에 반하여, 최근의 프라이버시 개념은 대부분 온라인 상에서의 개인정보보호라는 기술이나 정보적 관점의 프라이버시에만 집중하고 있다[12].

우리나라는 미국이나 유럽 등에 비해 프라이버시에 대한 논의가 상대적으로 부족한 상황이다. 수 천년 동안 농경을 바탕으로 공동체 생활을 해 온 역사적인 배경으로 인하여 프라이버시의 개념이 발달함에 어려움이 있었던 것이 한 이유이다[13]. 이러한 이유로 우리나라에서는 프라이버시의 개념을 정의하는 것이 사실상 어렵다. 우리 「헌법」 제17조는 “모든 국민은 사생활의 비밀과 자유를 침해받지 아니한다.”고 규정하여, 개인의 사생활의 비밀과 자유를 보장하고 있다. 이는 타인으로부터 사생활이 침해받지 않도록 하거나 자신의 사생활이 공개되지 않도록 할 수 있는 소극적 권리뿐만 아니라, 오늘날 정보화 사회 속에 포함된 자신의 정보에 대하여 자율적으로 통제 가능한 적극적 권리까지도 포함하는 것이다[14]. 이 법률 조항이 프라이버시 개념을 일부 포함하고는 있으나 국내법상에서는 명시적으로 정의하거나 관련된 원칙이나 권리가 무엇인지 구체적으로 제시하고 있지 않다.

프라이버시 개념과 관련하여 학술적으로나 실무적으로 명확히 정의하기 힘든 상황에서 실제 사람들이 인식하고 있는 프라이버시에 대한 연구는 거의 이루어지지 못하였다. 이론적 개념과 실제 인식의 간극에 대한 실증적 접근은 변화하고 있는 프라이버시 개념에 대한 이해와 국내 상황에 맞는 프라이버시 개념화를 위한 학술적 방향을 제시할 수

있을 것으로 예상된다. 이에 본 연구에서는 온라인상에서 노출되는 프라이버시의 개념을 분석하여 이론적 개념과 실제 통용되는 개념의 차이를 살펴보고 시사점을 도출하고자 한다. 이를 위해 기존 연구를 바탕으로 정의된 프라이버시의 개념을 분석 및 정리하고, 빅데이터 분석기법을 활용하여 오늘날 우리나라에서 프라이버시의 개념에 관한 인식을 고찰한다. 또한, 도출된 결과를 통해 현재 정의된 프라이버시 개념의 한계점을 제시하고 현시점에 적합한 프라이버시의 개념 정의의 필요성에 대하여 논의한다.

2. 관련 연구

2.1 프라이버시 시각의 다양성

프라이버시는 학문 분야, 국가, 문화, 개인 및 상황에 따라 다르게 정의되고 인식될 수 있는 복잡한 개념이다. Cooley[15]판사는 1888년에 처음으로 프라이버시의 개념에 대하여 인격의 불가침성(personal immunity)으로서 자신의 신체적인 안전을 위한 ‘홀로 있을 권리(Right to be let alone)’로 정의하였다. 이후 1890년에 Warren과 Brandeis[8]는 ‘The right to privacy’라는 논문에서 프라이버시의 개념을 이전과는 다른 새로운 권리로써 헌법을 통하여 보호되어야 한다고 주장하였다. 이러한 주장을 바탕으로 1903년에 뉴욕 주에서는 프라이버시보호법이 최초로 제정되었고, 1965년에 비로써 미국 연방대법원을 통하여 프라이버시가 헌법적인 권리로 인정되었다[16]. 초기의 프라이버시는 자신의 사생활에 대하여 함부로 공개 당하거나 침해당하지 아니할 소극적 권리로 정의되어왔다.

그러나 인터넷의 발달에 따라 개인의 의도와는 다르게 사생활이 공개되고 개인에 대한 정보가 유포되는 심각한 문제가 발생하기 시작하였다. 이러한 사회변화는 프라이버시 개념의 변화에도 영향을 주었다. Westin[17]은 “개인, 그룹 또는 조직이 자신의 정보를 언제, 어떻게, 어느 정도를 타인에게 공개할 것인가에 대하여 스스로 결정할 수 있는 권리”도 프라이버시의 범위에 포함 시켜야 한다고 주장하였다[17]. 프라이버시의 개념이 ‘혼자 있을 권리’로서의 소극적 개념에서 점차 자기에 관한 정보의 유통에 대하여 통제하는 권리인 ‘자기 정보의 통제권’으로서의 적극적 개념으로 확장되기 시작한 것이다[16].

프라이버시에 대한 논의는 인문학, 사회과학, 공학 등 다양한 학문 분야에서 진행되어왔다. 법학 분야에서는 1890년 Warren과 Brandeis[8]가 발표한 “The Right to

Privacy”를 통하여 프라이버시의 개념을 ‘홀로 남겨져야 할 권리’로써 처음으로 독립적 권리로 정의하였다. 하나의 인 권으로 프라이버시가 보호받기 시작하면서 오늘날까지도 헌법학에서는 프라이버시의 정의 및 보호 범위에 대하여 구체적으로 확정하기 위한 논쟁을 진행하고 있다[16]. 반면에 경영학 또는 경제학에서는 프라이버시를 하나의 상품으로써 다루고 있다. 정보의 주체자인 개인은 자신의 정보를 제공함에 따라 받게 되는 기대이익(금전적 대가처럼)과 교환 가능한 수단으로서 프라이버시를 사고팔 수 있는 거래 가능한 도구로 바라본다[18]. 그 밖에 심리학에서는 프라이버시를 주로 개인의 느낌이나 감정에 관한 것으로 정의하고 있으며, 철학에서는 통제의 관점에서 제한적인 접근 또는 격리로 프라이버시의 개념을 정의하고 있다[19]. 이렇듯 프라이버시를 바라보는 시각과 접근의 차이로 인해 합의된 정의가 존재하지 않고 연구 분야나 목적에 맞추어 그 개념이 차별적으로 정의되고 있는 실정이다.

2.2 프라이버시 개념 변화의 역동성

프라이버시의 개념은 국가별 시대 흐름에 따라 변화되어왔다. 미국의 경우, 1890년 Warren과 Brandeis[8]가 프라이버시를 본격적으로 논의하기 시작하였다. 이 시기에는 개인의 생각 및 감정을 어느 정도까지 전달할 것인지 선택할 권리를 주로 강조하였다. 이후 1960년대 들어서면서 프라이버시와 관련된 많은 논문들이 발표되었고, 법원은 프라이버시의 범위에 대하여 광범위하게 수용하기 시작하였다[16]. 연방대법원의 Whalen vs. Roe 사건에서 프라이버시의 결정과 공개를 분리하여 바라보기 시작했고, 판결에서는 프라이버시의 보호 범위를 ‘중요한 결정을 독자적으로 하는 것에 대한 이익’과 ‘개인적인 사안(personal matters)을 공개하지 않는 것에 대한 이익’의 두 가지 이익과 관련된다고 판단하였다. 다시 말해서 ‘사생활 결정에 관한 프라이버시’와 ‘사생활 공개에 관한 프라이버시’로 구별하여 바라보기 시작한 것이다[16]. 이처럼 미국은 프라이버시의 개념에 공간적, 사생활 결정 및 사생활 공개에 관한 내용을 포함하였고, 그 맥락에 따라서 확장하고 분리하는 등 적극적으로 개념을 재정의해왔다.

유럽의 경우, 프라이버시의 개념을 주로 개인의 데이터, 즉 개인정보 중심으로 논의해왔다. 최근에는 모든 유럽 시민들의 개인정보보호를 위하여 ‘일반개인정보보호규정’(GDPR)을 제정하여 권한을 부여받은 해당 지역의 조직만이 데이터에 접근할 수 있도록 규정하도록 강화하고 있다[20]. GDPR의 입법 목적은 “자연인의 기본적 권리와 자유

로써 프라이버시” 보호를 강조해 왔으나, 2016년에 와서는 ‘개인정보보호’가 강조되는 형태로 변화되기 시작하였다[20]. 그 영향으로 브라질에서는 일반데이터보호법(General Data Protection Law, GDPR), 중국에서는 핵심 인프라에서부터 데이터 거버넌스에 이르기까지 모든 요소를 포괄한 개인정보의 보호를 위한 규칙으로 데이터 기밀 보호법 등이 발표되기도 하였다[21].

우리나라의 경우, 앞서 언급한 바와 같이 「헌법」 제17조의 규정을 프라이버시의 보호에 관한 규정으로 파악하고 있으며, IT 환경에 발맞춘 프라이버시의 보호를 위한 자기정보통제의 관점에서 「개인정보 보호법」이나 「정보통신망법」 등 법률에서 법제화가 시도되었다[22]. 미국은 헌법으로는 프라이버시를 보장하지 않지만, 판례를 통하여 프라이버시를 하나의 기본권으로 인정하고 주 정부 등에서 「소비자 프라이버시법」(Consumer Privacy Act) 등 이에 상응하는 법을 제정하고 있다[23, 24]는 점과는 차이가 있다. 이는 프라이버시의 보호 범위 및 그 정도에 대하여 광범위하고, 구체적인 미국과는 달리 우리나라는 프라이버시의 보호에 관한 보호가 정보 분야에서의 개인정보자기결정권과 관련한 것으로 한정되는 등 정도가 부족함을 알 수 있다. 특히 우리나라는 급속한 기술발전에도 불구하고 프라이버시 보호를 위한 법률제정 속도는 매우 더딘 상황이다. 불과 몇 년까지만 하더라도 헌법재판소는 「헌법」 제17조 사생활의 비밀과 자유 위반을 이유로 위헌 또는 인용 결정을 내린 사례가 많지 않았다[25]. 대규모의 정보 유출 사고 등 이에 관련하여 발생할 수 있는 프라이버시 유출의 위험성이 사회문제로 대두되면서 오늘날 우리나라는 비로소 프라이버시의 본질을 파악하기 위한 연구가 시작되었다[25]. 미국이나 유럽 등과 문화적 환경이 다른 우리나라에서는 프라이버시라는 개념이 사실상 익숙치않고 이를 학술적으로 정의하는 것 또한 더 어려운 일이다. 실제 프라이버시라는 외래어를 한글로 표현하기 어려운 것도 같은 맥락에서 이해할 수 있다. 따라서 실제 우리나라의 헌법에서 보호하고자 하는 프라이버시의 본질과 정의가 무엇인지 파악할 필요가 있다. 또한, 기존의 프라이버시를 바라보는 방법론적 관점의 문제점을 도출하고, 현실점에 적합한 프라이버시란 무엇인지 논의할 필요가 있다.

2.3 프라이버시 이론의 복잡성

인터넷이 등장하면서 개인의 정보와 관련한 프라이버시 유출에 따른 민감성이 확대[26]됨에 따라 프라이버시에 대한 태도와 행동에 대하여 새로운 관점으로 접근하는 이론들이 연구되고 있다. 프라이버시 계산이론은 소비자 개인의

정보제공과 관련한 의사결정을 이해하기 위하여 발전된 이론이다[27]. Smith, et al.[28]에 따르면 프라이버시 계산이론은 프라이버시를 하나의 상품으로 바라보는 이론이다. 예를 들면, 소비자가 기업으로부터 자신의 정보를 제공할 것을 요구받았을 경우, 소비자는 개인정보를 제공하면서 얻을 수 있는 이익과 위험에 대하여 평가하게 된다. 이러한 프라이버시 계산 과정을 거친 후, 소비자가 생각하는 프라이버시를 제공하여 얻게 되는 이익이 프라이버시를 제공함으로써 받을 수 있는 위험보다 크거나 최소한의 균형을 이루게 된다면 소비자는 개인정보를 제공하자는 의사결정을 하게 된다[27]. 이후로도 프라이버시 계산이론을 바탕으로 소비자가 개인의 정보를 제공하게 되는 행동과정을 설명하기 위하여 다양한 연구가 진행되고 있다[27].

전통적인 프라이버시와 관련 연구는 태도와 행동의 일관성이라는 가정을 바탕으로 진행되어왔다. 반면에 프라이버시 역설이론은 이러한 가정을 부정한 이론으로 제시되었다[27]. 프라이버시 역설이론이란 사용자가 프라이버시에 대한 염려가 높음에도 불구하고 프라이버시 보호를 위한 행동을 그만큼 하지 않거나 작은 이익을 위해 개인의 정보를 제공한다는 것이다[27]. 높은 프라이버시 염려에도 불구하고 사용자가 개인의 정보를 제공하는 원인은 사용자가 특정 서비스 이용을 위하여 또는 서비스 제공자로부터 더 많은 이익을 얻기 위하여 사용자의 정보를 제공해야 하기 때문이다. 자신의 개인정보를 제공함에 따라 발생할 수 있는 프라이버시 침해나 공개한 개인정보의 2차적 사용 등과 같은 위험은 감수해야 한다[27].

프라이버시를 상대적 개념으로 정의하려는 노력 또한 활발히 진행되어 왔다. Altman[29]은 프라이버시에 대하여 ‘자신에 대한 접근권의 선택적 통제(selective control of access to the self)’로 정의하였다. 이전의 프라이버시를 바라보는 수동적 관점에서 벗어나 프라이버시를 바라보는 관점에 있어서 변증법적이며 역동적인 속성을 강조하였다[29]. 이러한 관점에 의하면 사람들은 프라이버시를 자신의 상황에 맞추어 개방과 폐쇄라는 스펙트럼을 따라 타인의 접근 가능성에 대하여 최적화할 수 있는 경계조절 과정(boundary regulation process)에 따라 프라이버시 개념을 정의할 수 있게 된다. 유사한 관점에서 Petronio[30]은 ‘커뮤니케이션 프라이버시 관리(CPM: Communication Privacy Management)’ 이론을 주장 하였다. 이에 따르면 프라이버시 경계(privacy boundaries)는 유동적으로 개인적인 것이 될 수도 있으며, 때로는 집단적인 것이 될 수도 있다[29]. 이러한 프라이버시의 경계선을 설정하는 기준은 문화, 성별, 동기, 맥락 및 모험-이익 비율 등 여러 가지 요

인에 영향을 받을 수 있다[30].

최근에 Zhou and Piramuthau[31]는 다차원적인 프라이버시 개념과 이러한 프라이버시의 개념적 역동성을 설명할 수 있는 맥락적 프라이버시 이론을 제시하였다. 맥락적 프라이버시 이론에 따르면, IoT 환경에서 개인화된 프라이버시를 어떻게 보호할 수 있는지에 대하여 초점을 맞춰 상황, 장소, 시간의 3가지 축에 따라서 프라이버시의 침해 범위에 해당하는지 판단할 수 있다[31]. 상황 축은 주어진 활동에 대한 상황을 의미한다. 예를 들어 의사에게 진료를 받는 상황 등이 이에 해당한다. 같은 상황에서도 장소나 시간에 따라서 프라이버시 경계는 달라질 수 있다. 장소는 물리적·가상적 공간을 동시에 포함하는 개념을 의미하며 시간은 시간의 흐름 또는 변화의 순간에 따라 다양한 의미로 영향을 줄 수 있다. 맥락적 프라이버시 이론은 프라이버시에 민감한 사람과 덜 민감한 사람을 구분할 수 있는데 이러한 이론을 바탕으로 개인의 정보제공에 대한 요청 범위를 설정할 수 있다.

앞서 살펴본 것처럼 프라이버시라는 개념은 학문 분야, 국가, 연구자 별로 다양한 관점에서 접근해 왔고 그 복잡성과 역동성에 대해 많은 논의가 있었다는 것을 알 수 있다. 그러나 활발한 학술적 논의에 반해 사람들이 인식하는 프라이버시는 어떻게 변해왔는지에 대해 실증적으로 살펴본 연구는 드물다. 대부분의 연구들이 전통적으로 정의되어온 개념을 바탕으로 연구가설에 대한 검증 차원에서 논의가 있었지만 사회에서 통용되는 개념에 대한 분석이나 논의는 매우 제한적이다.

3. 연구방법

3.1 연구문제

본 연구는 프라이버시 개념에 있어서 제한적으로 접근하고 있는 기존 연구들의 한계점을 지적하고, 향후 프라이버시 관련 연구 방향 제시를 위한 탐색적 연구이다. 기존 연구들을 통해 논의된 전통적인 프라이버시의 개념과 현 시대의 프라이버시 개념의 차이를 살펴봄으로써 프라이버시 개념의 학술적 재 정의와 사회적 합의의 필요성을 확인하는 것이 본 연구의 핵심 주제이다. 온라인 상의 뉴스 빅데이터를 바탕으로 프라이버시 연관 단어를 사회연결망 분석 기법으로 분석하여 현 사회가 주목하고 있는 프라이버시 이슈를 도출한다. 분석 결과를 통해 이론적 개념과 실제 인식 간의 차이를 확인하고 향후 프라이버시 관련 연구의 방향을 제시하는데 본 연구의 의의가 있다.

3.2 측정 도구

본 연구는 온라인 상에서의 활용되고 있는 프라이버시 관련 단어 분석을 통해 프라이버시에 대한 사회적 인식 현황을 파악하여 현시대에 맞는 프라이버시의 개념 제정의 필요성을 강조하는데 그 목적을 두고 있다. 이를 위해 대표적인 포털사이트인 네이버(Naver), 다음(Daum), 구글(Google)에서 프라이버시와 함께 등장하는 단어들을 수집하여 프라이버시라는 용어가 어떻게 활용되고 있는지를 살펴보았다. 데이터 수집 도구로써 텍스트(Textom) 서비스를 이용하였고, 분석 채널은 뉴스로 한정하였다. 자료검색을 위한 키워드는 ‘프라이버시’를 사용하였다. 자료의 분석 기간은 1년으로 2018년 1월 1일부터 2018년 12월 31일까지 연구대상 기간으로 한정하였다. 수집된 데이터는 월별 1000건 총 14,628건의 문서 데이터가 수집되었다. 분석데이터 정보는 Table 1과 같다. 네트워크 구조 분석 시 노드 수가 많은 경우 시각화 표현에 어려움이 있어, 상위 100개의 키워드를 중심으로 살펴보았다.

Table 1. Research Data

Category	Content
Target	Naver(news) 7,152 counts, Daum(news) 7,266 counts, Google(news) 210 counts
Period	January 1, 2018 ~ December 31, 2018
Tool	Textom, Ucinet 6.645, NetDraw
Key word	Privacy

3.3 측정 방법

텍스트마이닝이란, 텍스트를 바탕으로 데이터베이스에 나타난 언어를 바탕으로 기존에 알려지지 않은 패턴과 다양한 지식 및 새로운 정보를 얻을 수 있는 분석 기술이다[32]. 본 연구에서는 텍스트를 활용하여 텍스트의 핵심단어들을 추출한 후에 메트릭스를 구성하였다. 수집된 데이터에서 프라이버시와 관련 없는 단어인 주목, 장점, 일부 등을 제외하였으며, 블록체, 블로체인인, 모바 등 불완전한 단어를 제거하였다. 또한 있, 위하, 대하, 민, 후, 공 등 조사나 문장부호 등을 제거하고 정제하여 상위 100개 단어를 선정하였다. 소셜네트워크 상에 내재되어 있는 프라이버시에 대한 관계를 시각화하여 프라이버시에 대한 인식을 파악하기 위해 Ucinet6.645를 통해 단어들 간에 연결 구조 파악 및 연결 중심성을 분석하였다. 연결 중심성 분석을 통하여 특정한 주제에 대해 구조적으로 파악할 수 있다. Netdraw를 사용하여 프라이버시와 관련된 단어들의 네트워크를 시각

화하고 유사점을 가진 단어들의 군집을 나타내기 위해 CONCOR분석을 실시하였다. 이에 본 연구에서는 네이버(Naver), 다음(Daum), 구글(Google)에서 ‘프라이버시’ 단어를 포함한 뉴스의 자료를 수집 및 데이터 정제 과정을 통해 도출해 낸 단어를 바탕으로 이들 간의 관계성을 살펴보았다.

4. 분석결과

4.1 키워드 분석결과

2018년 1월 1일부터 2018년 12월 31일(총 1년) 동안 프라이버시 키워드를 중심으로 네이버(Naver), 다음(Daum), 구글(Google)의 뉴스에서 추출된 단어 수는 총 30,208개였다. 2018년 1월 1일부터 2018년 12월 31일(총 1년)까지 프라이버시에 대한 키워드 분석을 실시한 결과 총 100개의 단어로 다음 Table 2와 같이 정리하였다.

Table 2. Keyword Frequency

Word	Frequency	Word	Frequency
Protection	2184	Regulation	444
Data	1711	Apple	442
Security	1693	Life	438
Technology	1475	Reinforcement	438
Personal Data	1421	Settlement	438
Facebook	1339	Structure	414
Privacy	1253	Release	409
Information	1209	Officetels	403
Guarantee	1185	Platform	393
Space	1096	Naver	384
Generation	1036	Cloud	384
Problem	1033	Application	379
Offer	1015	Progress	379
Plan	1013	Affair	377
Personal	965	World	375
Block Chain	956	Artificial Intelligence	371
Application	947	Solution	361
Infringement	902	Participant	359
Company	895	Environment	358
Service	876	Global	356
Open	873	Organization	353
Utilization	865	Big Data	351
Installation	723	Occurrence	350
User	723	Explanation	349
Function	721	Emphasis	342
Necessity	712	Cryptocurrency	339

Use	703	Announcement	333
Google	702	House	332
Personal Privacy	692	View	328
Streets	692	Premium	326
Safety	689	Victim	323
Human	644	Oneself	322
America	618	Plan	321
Concern	613	Share	321
Arrangement	597	Internet	318
Utilize	573	Private life	314
Consideration	563	Independence	309
Operate	554	China	308
Apartment	545	User	305
Customer	528	Convenience	304
Introduction	519	Smart Phone	303
Think	513	Zuckerberg	303
Development	511	Korea	303
Basement	507	Patient	297
Family	487	Representative	295
Digital	481	Domestic	294
privacy Protection	465	Support	293
Leakage	455	Content	286
Management	454	Build	285
Government	445	Residents	283

프라이버시에 대한 연결 중심성 분석을 실시한 결과 Table 3과 같이 정리하였다.

Table 3. Keyword Centrality

단어	중심성	단어	중심성
Protection	0.03014	Family	0.009373
Data	0.029014	Representative)	0.009208
Security	0.027193	Artificial Intelligence	0.009075
Facebook	0.02557	App	0.008943
Technology	0.024576	China	0.008943
Personal Data	0.023781	Consideration	0.008843
Problem	0.022523	Settlement	0.008777
Information	0.021297	Reinforcement	0.008512
Open	0.021198	Share	0.008446
Offer	0.020966	Privacy Protection	0.00838
Company	0.020204	Picture	0.008347
Personal	0.019707	World	0.008347
Human	0.019508	Korea	0.008313
Block Chain	0.018283	Cloud	0.008148
Service	0.018051	Opinion	0.007982
Space	0.018018	Occurrence	0.007982

Use	0.017852	Patient	0.007949
Application	0.017289	Plan	0.007916
America	0.016726	Life	0.007817
Guarantee	0.016627	Internet	0.007784
Necessity	0.016461	Community	0.007784
Utilization	0.015898	Street	0.00775
User	0.015435	Solution	0.00775
Infringement	0.015302	Platform	0.007684
Google	0.014739	Affair	0.007552
Function	0.014739	Law	0.007519
Concern	0.01401	Support	0.007485
Customer	0.013282	Demand	0.007254
Plan	0.01295	Leakage	0.007254
Development	0.012586	Company	0.00722
Installation	0.012255	Privacy	0.007187
Utilize	0.012156	Products	0.007187
Safety	0.012122	Domestic	0.007121
Generation	0.012089	Female	0.007055
Explanation	0.011891	Crypto currency	0.007055
Apartment	0.011824	Policy	0.006989
Introduction	0.011493	Officetels	0.006856
Announce	0.011195	House	0.006856
Apple	0.010963	Build	0.006823
Regulation	0.010864	User	0.00679
Operate	0.010798	Big Data	0.006657
Government	0.010499	Investigation	0.006657
House/Home	0.0104	Global	0.006624
Management	0.010268	Trade	0.006558
Release	0.010234	Code	0.006392
Progress	0.010201	Advertisement	0.006359
a Participant	0.009771	Expansion	0.006326
Oneself	0.009572	Respect	0.00626
Emphasis	0.009473	Business	0.006227
Digital	0.009406	Naver	0.006161

4.2 네트워크 시각화 분석결과

네트워크 내의 연결 관계 및 패턴을 시각적으로 도출하고 군집화하기 위하여 CONCOR분석을 실시하였다. 상위 주요 단어빈도를 이용해 CONCOR분석을 실시한 결과는 Fig. 1과 같이 나타난다. 프라이버시 관련 키워드는 ① 인터넷 발달과 관련된 용어(스마트폰/ 네트워크/ 영상/ 사진/ 클라우드 등), ② 정책 및 규제에 관한 용어(정부/ 규제/ 유럽/ 중국/ 영국/ 준수 등), ③ 프라이버시 관련 사업을 시작한 기업명칭(네이버/ 삼성전자/ 페이스북/ 서비스/ 구글 등), ④ 4차산업 관련 용어(블록체인/ 빅데이터/ 플랫폼/ 암호화폐 등), ⑤ 주거 관련 용어(집/ 아파트/ 조망/ 독립/ 단독주택/ 회사/ 오피스텔/ 호텔 등), ⑥ 개인의 사생활과 연

관된 용어(좌석/ 개별/ 존중 등), 그리고 기타 ⑦ 개인정보 피해 사례에 관한 용어(환자/ 의원/ 감독/ 사회 등), ⑧ 2차 피해(2차/ 피해자/ 가족/ 여성 등) 총 8개의 그룹이 형성되었다.

5. 논의 및 결론

본 연구는 빅데이터 분석을 활용하여 프라이버시의 인식에 대하여 알아보고, 이를 바탕으로 현 시점에 적절한 프라이버시의 개념 정의 방향을 살펴보고자 하였다. 분석결과에 따르면 ‘데이터’, ‘보호’, ‘규제’, ‘기술’, ‘개인정보’ 등의 단어들 이 상대적으로 많이 추출된 것을 확인할 수 있었다. 즉, 4차산업혁명 시대로 접어들며 기술발전에 따라 개인정보의 가치가 재평가되면서, 그 활용과 보호의 필요성이 많이 논의되고 있음을 확인할 수 있는 결과이다[33]. 프라이버시 관련 키워드에 대한 CONCOR분석을 실시한 결과, 프라이버시는 인터넷·스마트기기 등 기술 환경의 변화와 함께 개인의 사생활을 주제로 소통할 수 있는 영상·사진 등의 키워드와 군집화를 이루고 있다. 이는 기존의 소통 방식이 아날로그 중심에서 디지털 중심의 환경으로 변화하고 있음을 의미하며, 스마트기기 및 소셜미디어의 급성장과 영향력의 확산을 나타내면서[34], 개인 등이 가지고 있는 프라이버시에 관한 인식의 확장을 보여주고 있다.

이와 함께 우리 정부를 비롯한 세계 각국의 다양한 ‘규제·정책’에 관한 키워드가 군집화를 보이며, 이는 최근 유럽인의 개인정보 통제권 확대와 기업의 책임성 강화를 목적으로 시행된 일반개인정보보호법(GDPR)과 이에 영향을 받은 미국 캘리포니아주의 소비자 프라이버시법(Consumer Privacy Act), 브라질의 일반데이터보호법(General Data Protection Law, GDPL), 그리고 중국의 데이터 기밀보호법 등의 발효에 관한 기업과 개인의 관심을 보여준다[35]. 또한 프라이버시와 관련한 각종 기업의 기술개발과 사업모델 등을 축으로 한 군집이 나타나며, ‘인공지능’, ‘클라우드’, ‘디지털’, ‘플랫폼’, ‘개발’ 등의 단어들 이 추출된 것으로 보아, 최근 개인정보(프라이버시) 및 빅데이터를 활용한 사업 모델의 증가[36, 37]에 영향을 받았음을 나타내는 결과로 판단할 수 있다. 이는 정부의 정책 및 규제와 산업계의 프라이버시 관련 동향 등에 관련된 것이며, 이와 대칭을 이루는 군집으로 주거·생활환경 및 개인 프라이버시, 사생활 침해 등에 관한 각각의 군집이 나타났다. 먼저 ‘입주민’, ‘생활’, ‘아파트’, ‘오피스텔’, ‘설계’, ‘거리’ 등의 단어 군집은 ICT 기술과 주거환경의 융합에 따른 스마트홈 등 주거환경에 관한 개인의 관심과 사생활 침해의 방지나 방해배제 등 삶의 질

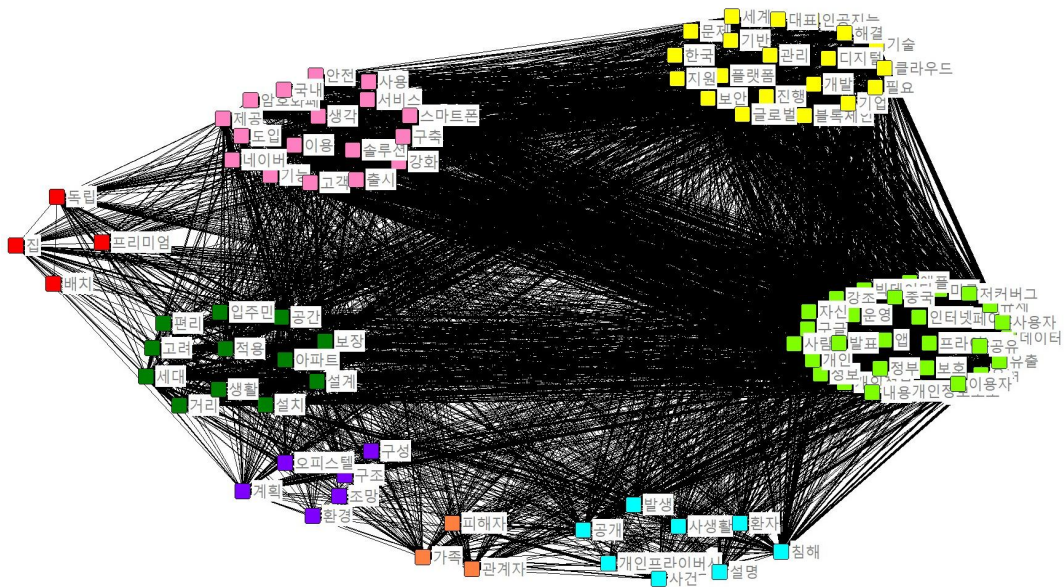


Fig. 1. Keyword Network Visualization

개선을 위한 개인의 열망을 보여주는 결과로 볼 수 있다. 그 밖에도 ‘개인 프라이버시’, ‘침해’, ‘발생’, ‘사건’, ‘공개’, ‘2차 피해’, ‘가족’, ‘여성’ 등이 각각 군집화를 이루고 있으며, 이는 최근 기업 등에 의한 개인 프라이버시의 무단도용이나 개인정보의 유출사건 등 프라이버시와 관련한 각종 사건에 대한 개인의 흥미가 증가함을 나타내고, 성별에 따라 프라이버시에 관한 염려 수준에 차이가 있음을 추정할 수 있다[38].

그러나 결과와 같이 정책 및 규제나 기업의 프라이버시 관련 동향 등의 군집과 주거·생활·환경·사생활 등의 군집이 대칭을 이루고 있는 것은 ‘프라이버시’에 관한 전통적인 개념과 변화 또는 강화의 기로에 있는 현시점의 새로운 ‘프라이버시’ 개념이 서로 적절하게 융합하지 못하였음을 보여주며, 최근의 논의가 기술 및 개인정보 관련 규제와 정책을 중심으로 편중되어 논의되고 있음을 극명히 드러내고 있다. 특히 개인의 사생활을 중심으로 연결되는 생활·주거 환경이나 개인정보의 침해 또는 유출 문제는 앞서 언급되는 군집의 화제성에 비하여 상대적으로 낮은 군집성을 보여준다.

본 연구 결과를 통하여 단어빈도 및 연결 중심성을 파악할 수 있었으며, CONCOR분석을 통하여 높은 순위의 단어들과 다양하게 형성된 단어별 그룹을 추출할 수 있었다. 연구를 통해 도출된 결과에 따르면 기술의 발달과 함께 ‘프라이버시’에 관한 개인·기업·정부의 인식이 충분히 성숙해가고 있으나, 그 성숙은 기술적 측면과 사이버상의 개인정보 통제에 관한 것으로 편중되어 나타나고 있으며, 특히 개인

과 기업 등 산업계가 개인정보에 관하여 가지고 있는 인식의 차이를 극명하게 보여준다. 그러므로 전통적 의미에서의 프라이버시에 대한 인식과 사회적·시대적·기술적인 의미에서의 프라이버시의 변화를 적절히 파악하여 현대에 적합한 프라이버시의 개념을 재정립한다면 프라이버시의 보호에 긍정적인 영향이 있을 것으로 판단된다[39].

본 연구 학술적 의의는 프라이버시에 대한 포괄적 문헌 연구와 빅데이터를 바탕으로 한 실증 분석을 통해 오늘날 프라이버시의 개념이 전통적인 프라이버시의 개념에 비해 제한적으로 논의되고 있다는 것을 확인하였다는데 있다. 또한 최근의 연구나 온라인 상의 답변들이 프라이버시 개념과 실제 사람들이 인식하고 있는 개념 사이에 차이가 있음을 밝힘으로써 변화하고 있는 프라이버시 개념에 대한 이해와 현 상황에 맞는 프라이버시의 법률적 개념정의의 필요성에 대하여 제시하였다는 점에서도 의의가 있다. 그러나 온라인 상의 빅데이터 분석을 통해 프라이버시에 대한 개념과 사회적 인식 현황을 살펴보았다는 점에서는 큰 의의가 있지만 1년 동안의 뉴스 데이터만으로 연구 결과를 도출하였다는 점에서 한계가 있다. 향후 연구에서 분석을 위한 키워드를 다양화하고 기간과 수집 범위를 확장한다면 보다 의미 있는 연구가 될 것으로 기대한다.

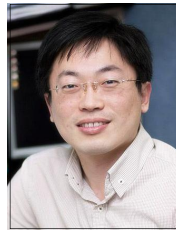
REFERENCES

- [1] K. Lee. (2009). Increased Internet privacy violation. *The Science Times*, <https://www.sciencetimes.co.kr>
- [2] G. B. Kwon, H. J. Lee & I. Kim. (2018). The Study on Legislation Process of GDPR(General Data Protection Regulation) and Subsequent Trend in EU System, *Studies on American Constitution*, 29(1), 1-38.
- [3] T. Oh & J. Yu. (2011). Change of privacy concept in social network service (SNS) environment, *KISDI*, 23(4), 26-42.
- [4] J. L. Yu. (2011). Personal Information Protection in Digital Era - Reviewing Personal information protection Act, *Journal of Digital Convergence*, 9(6), 81-90. UCI : G704-002010.2011.9.6.022
- [5] K. Lee. (2014). Privacy Protection in the Big Data Age, *Company Ethics Briefs*, 11, 1-26.
- [6] M. Park. (2018). GDPR, Microsoft's commitment to privacy and customer data management, *Microsoft*, <https://news.microsoft.com/ko-kr>
- [7] Y. Jung. (2001). Review of the Theory of Privacy in Modern Constitutional Law, *Journal of Cybercommunication Academic Society*, 7, 214-267.
- [8] S. D. Warren & L. D. Brandeis (1890). The right to privacy, *Harvard Law Review*, 4(5), 193-220.
- [9] A. F. Westin. (1967). *Privacy and Freedom*, Atheneum. New York: Ig Publishing.
- [10] E. J. Bloustein. (1964). Privacy as an Aspect of Human Dignity: An answer to Dean Prosser, *New York University Law Review*, 39, 962-1007.
- [11] Y. H. Son. (2014). The Protection of Personal Information in the Era of Big Data, *Business Law Review*, 28(3), 355-393. UCI : G704-001299.2014.28.3.003
- [12] G. Moon. (2019). Survive the privacy of the Internet, *Security News*, <https://www.boannews.com>
- [13] S. Sung. (2004). Cyber space and privacy, *KISDI*.
- [14] Supreme Court. (1998. 7.24.). Sentence, 96Na42789 Judgement.
- [15] T. M. Cooley. (1888). *A Treatise on the Law of Torts or the Wrongs Which Arise Independently of Contract*, Chicago : Callaghan and Company.
- [16] K. Cho. (2005). Statutory Development of Privacy in the United States, *Sungkyunkwan Law Review*, 17(3), 75-113. UCI : G704-001437.2005.17.3.013
- [17] A. Westin. (1970). *Privacy and Freedom*. (first edition 1967). Atheneum, New York.
- [18] J. Kim & S. Kim. (2015). A Study on Privacy Paradox from the Behavioral Economics Perspective, *Informatization Policy*, 22(3), 16-35. DOI : 10.22693/NIAIP.2015.22.3.016
- [19] J. B. Young. (1978). *Privacy*. Wiley, Chichester, UK, Cambridge University. DOI : <https://doi.org/10.1017/S0047279400009223>
- [20] I. S. Ham. (2016). The Enactment of 2016 General Data Protection Regulation(GDPR) in the EU and its Implications, *Chonnam Law Review*, 36(3), 411-453. UCI : G704-002100.2016.36.3.018
- [21] ITWorld. (2018). *Chinese data confidentiality protection law came into effect similar to China's GDPR in May*, <http://www.itworld.co.kr>
- [22] M. Cho & H. Lee. (2017). Improvement Issues of Personal Information Protection Laws through Meta-Analysis, *Journal of Digital Convergence*, 15(9), 1-14. DOI : <https://doi.org/10.14400/JDC.2017.15.9.1>
- [23] H. Jung. (2016). A Study on a Unique and Distinctive Culture(Law and Understanding) Regarding the Protections for Privacy in the U.S., *Ilkam Law Review*, 35, 277. DOI : 10.35148/ilsilr.2016.35.271
- [24] B. C. Kim. (2014). A study on Utilization of Big Data Based on the Personal Information Protection Act, *Journal of Digital Convergence*, 12(12), 87-92. DOI : 10.14400/JDC.2014.12.12.87
- [25] C. Chang. (2016). An Argument for Making Privacy a Working Constitutional Right - Referring to Daniel Solove's Pragmatic Theory -, *Journal of Media Law, Ethics and Policy Research*, 15(3), 1-30. UCI : G704-001997.2016.15.3.005
- [26] C. W. Park & J. W. Kim. (2015). An Empirical Research on Information Privacy and Trust Model in the Convergence Era, *Journal of Digital Convergence*, 13(4), 219-225. DOI : <http://dx.doi.org/10.14400/JDC.2015.13.4.219>
- [27] J. Kim & S. Kim. (2014). A Study on Privacy Paradox between Privacy Concern and Information Disclosure behavior: Focus on Privacy Calculus Theory, *Entrue Journal of Information Technology*, 13(3), 139-152. UCI : G704-001673.2014.13.3.007
- [28] H. J. Smith, T. Dinev & H. Xu. (2011). Information privacy research: an interdisciplinary review. *MIS Quarterly*, 35(4), 989-1016. DOI : 10.2307/41409970
- [29] I. Altman. (1975). *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*, Brooks/Cole Publishing Company, Monterey, California.
- [30] S. Petronio. (2007). Translational research endeavors and the practices of communication privacy management. *Journal of Applied Communication Research*, 35(3), 218-222. DOI : <https://doi.org/10.1080/00909880701422443>
- [31] W. Zhou & S. Piramuthu. (2015). Information Relevance Model of Customized Privacy for IoT. *Journal of Business Ethics*, 131(1), 19-30. DOI : 10.1007/s10551-014-2248-y
- [32] A. Hotho, A. Nürnbergger & G. PaaB. (2005). A brief survey of text mining. *In Ldv Forum*, 20(1), 19-62.

- [33] H. K. Kong & H. J. Jun & S. U. Yoon. (2018). A Study on the Privacy Policy of Behavioral Advertising, *Journal of the Korea Convergence Society*, 9(3), 231-240. <https://doi.org/10.15207/JKCS.2018.9.3.231>
- [34] Y. S. Jeong. (2014). Tracking Analysis of User Privacy Damage using Smartphone, *Journal of Convergence for Information Technology*, 4(4), 13-18. <https://doi.org/10.22156/CS4SMB.2014.4.4.013>
- [35] D. Lee. (2019). *One year of GDPR implementation, a different change than expected*, ITWorld, <http://www.itworld.co.kr>
- [36] N. H. Sung, (2010). *The Future of Digital Privacy*. LG Economy Research Institute, <http://www.lgeri.com/report/view.do?idx=16658>
- [37] K. A. Kim. (2018). A Rise on the Privacy Protection based on Secret Code : Focus on Homomorphic Encryption. Boannews, <https://www.boannews.com/media/view.asp?idx=74664>
- [38] H. S. Lee. (2016). Personal Characteristics and Information Privacy Concerns. *Asia-pacific Journal of Multimedia Service Convergent with Art, Humanities, and Sociology*, 6(9), 267-276. <http://dx.doi.org/10.14257/AJMAHS.2016.09.46>
- [39] J. W. Mun. (2013). *Protection of Privacy : From myth to norm*. A Policy Reports for the Improvement of Legal Protection System in Personal Information.

이 환 수(Hwan-Soo Lee)

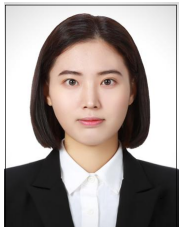
[장학원]



- 2014년 2월 : KAIST 기술경영과 (공학박사)
- 2017년 3월 ~ 현재 : 단국대학교 행정법무대학원 융합보안학과 조교수
- 관심분야 : ICT보안, 프라이버시, 빅데이터, 사용자행동
- E-Mail : hanslee992@gmail.com

이 송 이(Song-Yi Lee)

[학사학원]



- 2013년 2월 : 단국대학교 법학과(법학사)
- 2018년 9월 ~ 현재 : 단국대학교 IT법학협동과정 (석사과정)
- 관심분야 : 민법
- E-Mail : 94sylee@naver.com

김 성 원(Sung-Won Kim)

[장학원]



- 2015년 2월 : 단국대학교 법학과(법학사)
- 2017년 2월 : 단국대학교 일반대학원 BK21+ 지식재산·정보보호법 특화 전문인재양성사업단 IT법학협동과정(법학석사)
- 2015년 10월 ~ 2018년 5월 : 정보통신산업진흥원 선임연구원
- 관심분야 : 지식재산권법
- E-Mail : redmoon357@gmail.com