JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 23, No. 7, pp. 1297-1305, Jul. 2022

Check for updates

미국 특허와 기술체제분석을 이용한 데이터 익명화 기술의 동향과 전망

홍 은 아[1] · 김 준 엽[2] · 이 종 호[3*]
[1]숙명여자대학교 기후환경에너지학과 박사수료
[2]서울대학교 경제학부 박사수료
[3*]인하대학교 법학연구소 AI·데이터법센터 책임연구원

Trend and Prospect of Data Anonymization Technology Using US Patent and Technological Regime Analysis

Eun-Ah Hong[1] · Joon-Yub Kim[2] · Jong-Ho Lee[3*]

[1]Ph.D Candidate, Department of Climate, Environment and Energy Studies, Sookmyung Women's University, Seoul 104310, Korea
[2]Ph.D Candidate, Department of Economics, Seoul National University, Seoul 08826, Korea
[3*]Senior Researcher, AI·Data Law Research Center, Law Institute, Inha University, Incheon 22212, Korea

[요 약]

스마트폰, SNS 및 제4차 산업혁명의 등장은 데이터양을 급격히 증가시키고 있다. 본 연구는 데이터 활용의 열쇠가 되는 익명화 기술에 주목한다. 분석 자료는 미국 특허(1976-2021년)이고 텍스트 마이닝과 기술체제 분석을 이용한다. 익명화 특허는 2010년 이후 빠르게 증가하고 있으며, 대다수의 특허가 G06F(디지털 데이터 처리)에 속한다. 특허권자 다수는 미국, 독일과 일본에 속해있고 IBM이 가장 많은 특허를 갖고 있다. 통신, 카드, 은행, 플랫폼 기업들도 특허권자로 등장한다. 데이터 익명화는 과학기술논문에 덜 의존적이며, 특허권자의 분산도는 높고, 전유성은 낮다. 상대적으로 경쟁력 있는 선발자가 적어서 기회의 창이 존재하는 것으로 나타났다. 클라우드 분석에 따르면 "method", "data", "anonym*" 이 특허 제목에 많이 등장하며 "소프트웨어", "보호", "식별" 또는 "암호화", "의료"와 "서비스" 등이 빈번한 주제로 나타난다.

[Abstract]

The advent of smartphones, SNS, and the 4th industrial revolution is rapidly increasing the amount of data. This study focuses on anonymization technology. The analysis data is a US patent (1976-2021) and we use text mining and technological regime analysis. Anonymization has been increasing rapidly since 2010, and the majority of patents belong to G06F (digital data processing). The majority of assignees belong to the U.S., Germany, and Japan, with IBM holding the largest number of patents. Telecom, card, banking, and platform companies also appear as assignees. Data anonymization is less dependent on S&T papers, the degree of dispersion of assignees is high and appropriability is low. It was found that there is a window of opportunity because there are relatively few competitive incumbents. Cloud analysis shows that "method", "data", and "anonym*" appear frequently in patent titles, with frequent topics such as "software", "protection", "identification" or "encryption", "medical" and "services" appears as

색인어 : 디지털, 콘텐츠, 데이터 익명화, 기술체제, 미국특허
Keyword : Digital, Contents, Data Anonymization, Technological Regime, US Patent

http://dx.doi.org/10.9728/dcs.2022.23.7.1297

Received 18 June 2022; **Revised** 20 July 2022
Accepted 22 July 2022

***Corresponding Author; Jong-Ho Lee**

Tel: +82-32-860-8975
E-mail: jongholee@inha.ac.kr

Ⅰ. 서 론

클라우스 슈밥은 제4차 산업혁명이 지속성장을 위한 전 세계 미래의 새로운 물결이 될 것이라고 주장하였다[1]. 여기서 말하는 제4차 산업혁명은 사람과 사물이 유기적으로 결합하는 초연결(Hyper Connectivity)라고 정의된다. 이러한 초연결은 어떤 요소의 사용과 이동이 필수적인데 바로 그것은 데이터이다. 그러나 데이터가 원활하게 이동하고 공유되기 위해서는 안전함이라는 것이 전제되어야 한다. 여기에서 말하는 안전함이란 데이터가 개인정보를 포함하고 있기 때문에 개인이 식별될 수 없게 하거나 익명화하는 것이라고 정의할 수 있다.

데이터 익명화에 특히 민감한 분야는 보건 또는 금융 데이터와 같이 민감한 디지털 콘텐츠에서 가져온 데이터를 이용하는 경우이다. 이러한 유형의 데이터는 데이터 활용의 전제조건인 일부 데이터 익명화 기술 및 방법을 사용하여 익명화하는 동시에 데이터 프라이버시를 유지해야만 한다. 데이터 익명화에는 알고리즘 및 물리적 장비들뿐만 아니라 다양한 수단이 사용될 수 있다. [2]는 "데이터 익명화 기법은 원천 데이터에서 정보를 제거, 일반화, 작은 변화(노이지 추가) 또는 치환(연관성 바꾸기)"라고 하였다. 우리가 데이터 익명화에 관심을 갖기 시작한 것은 그리 오래되지 않았으며 최근 관련 연구가 증가하고 있다.

세계지적재산기구(World International Property Organization: WIPO)에 따르면, 특허는 어떤 일을 하는 새로운 방법 또는 문제에 대한 새로운 기술 솔루션을 제공하는 제품 또는 프로세스 등의 발명에 대해 부여된 배타적 권리이다[3]. 즉, 특허권은 소유자의 허가 없이 특허 발명을 상업적으로 제작, 사용, 배포, 수입 또는 판매하는 것을 중지하거나 방지할 수 있는 독점적인 권리를 제공한다. 물론, 이 권리는 특허가 출원되거나 부여된 국가 또는 지역에서만 유효하게 된다[4]. 그리고 일반적으로 출원일로부터 20년이라는 제한된 기간 동안 보호되고 있다[3]. 대다수의 국가들은 특허 협력 조약[3]에 따라 만들어진 "세계 특허 출원(PCT)"을 기반으로 하는 국가 특허 시스템을 사용하고 있다. WIPO는 1971년 스트라스부르 협정에 의해 설립된 국제특허분류시스템(IPC)을 사용하여 공개된 국제특허출원 데이터베이스를 유지관리하고 있으며 현재 전 세계 100개국 이상에서 사용되고 있다[3][5]. 특허 문서는 특허 제목, 설명, 단순 가족 ID, 발행/발행 연도, 출원/출원 연도, 우선 국가 코드, 양수인 국가, 양수인 원본/발명자, IPC 코드, CPC 코드 등과 같은 필드를 포함한다. 특허 정보를 이용한 연구는 혁신에 관심을 두는 슘페테리안 학파의 연구자들에게서 특히 많다. 특히, [6]-[9]의 최근 연구에서는 미국 특허 데이터를 이용하여 국가, 산업, 기업의 혁신시스템 수준을 측정하는 방법으로 특허 분석의 유용성을 확인시켜 준 바 있다. 더 나아가 [9]는 3차 산업혁명 기술과 4차 산업혁명 기술의 혁신수준을 비교하여 AI가 다른 기술들과 구별되는 혁신성이 높은 기술이라는 것을 보여주었다. 이렇게 기술동향 분석에는 특허가 매우 유용하게 이용되고 있다. 따라서 우리는 기존 연구자들이 제시한 연구방법론을 기반으로 익명화 기술의 기술적 특성과 기술적 관계에 대해 분석해 보고자 한다.

다음 Ⅱ장은 제4차 산업혁명 시대에서 익명화의 대상이 되는 데이터(빅데이터)가 어떤 특징을 갖고 있으며 왜 중요한지에 대하여 살펴보고, Ⅲ장은 특허 분석에 대한 방법론, 그리고 Ⅳ장은 분석결과를 논의하며, 마지막으로 Ⅴ장은 연구의 결과를 정리한다.

Ⅱ. 빅데이터, 특허분석방법론, 그리고 익명화연구

1970년대에 시작된 제3차 산업혁명은 정보통신기술을 기반으로 아날로그에서 현재의 디지털 기술까지 변화하는 디지털 혁명을 시작으로 데이터 혁명으로 이어지고 있다. 제3차 산업혁명의 발전에는 개인용 컴퓨터, 인터넷, 그리고 ICT기술이 포함된다. 그리고 새로이 등장한 제4차 산업혁명은 기술이 사회와 심지어 인간의 신체와 융합하는 새로운 방식으로 등장하고 있다. 제4차 산업혁명은 로봇 공학, 인공 지능, 나노기술, 양자 프로그래밍, 생명 공학, IoT, 3D프린팅 및 자율주행 차량 등을 비롯한 다양한 분야에서 새로운 방식으로 나타나고 있다. 슈밥은 4차 산업혁명이 기술 발전에 의해 특징지어졌던 이전 세대의 혁명과는 근본적으로 다른 점이 있다고 언급하였다[1]. 새로운 기술들은 전세계 사람들을 웹에서 연결하고 조직, 비즈니스 등을 효율적으로 향상시키는 커다란 잠재력을 갖고 있다. 이 과정에서 디지털 콘텐츠(즉, 데이터)가 무엇보다 중요한 요소로서 강조된다. 모든 플랫폼에는 데이터가 기반이기 때문이다.

특히 빅데이터는 기존 방식으로는 저장 및 관리, 분석이 거의 불가능할 정도로 규모가 크고, 변화의 속도가 빠르며, 다양한 형태를 갖는 데이터를 일컫는다. [10]은 빅데이터의 주된 특성으로 크기(volume), 다양성(variety), 속도(velocity) 등 3V를 제시하였다. 첫 번째 V는 크기인 Volume을 말한다. 기업 데이터, 감지기 데이터, SNS 데이터 등의 규모가 페타바이트(petabyte, 1 PB = 1024TB)나 엑사바이트(exabyte, 1 EB = 1024 PB) 이상으로 커지고 있다. 두 번째는 다양성인 Variety이다. 데이터는 관계형 데이터베이스 등과 같은 정형(structured) 데이터에서부터 HTML, XML, JSON 등과 같은 반정형(semi-structured) 데이터, 이미지, 비디오, SNS, 감지기 데이터 등과 같은 비정형(unstructured) 데이터까지 모든 형태를 가지기 때문에 다양하다. 마지막은 속도라 불리는 Velocity이다. 데이터 수집 및 가공, 분석 등 연속적인 과정을 실시간 또는 특정 시기에 처리할 수 있는 데이터 처리 능력을 말한다. 그리고 최근 빅데이터의 기존 특성인 3V에 진실성(veracity) 혹은 가치(value)를 추가하여 4V, 둘 다 추가하여 5V, 시각화(visualization)까지 추가하여 6V 등으로 확대하기도 하였다. 여기서 진실성인 Veracity는 의사 결정이나 기업 활동에 활용될 수 있도록 진실하고 정확해야 한다

는 것을 말한다. 그리고 가치인 Value는 비즈니스에 실현될 궁극적 가치에 중점을 둔다는 것을 의미한다. 마지막으로 시각화인 Visualization은 사용자 친화적인 시각적 기능을 통하는 방식을 말한다.

빅데이터가 의미를 갖기 위해서는 큰 규모의 데이터를 수집하는 것을 넘어 이러한 데이터를 통해 통찰력(insight)을 얻고 실제 기업 활동에 활용하여 가치를 창출할 수 있어야 한다. 웹사이트 검색 통계, SNS 데이터를 분석 등을 통해 시장 예측 및 상품 개발을 하고 소비자의 방문 및 구매 패턴을 분석하여 마케팅 전략을 수립하며, 제조 과정에서 발생하는 감지기 데이터를 이용하여 불필요한 작업 제거 및 개선으로 생산성 향상 등을 꾀해야 한다는 것이다. 이런 빅데이터를 이용하기 위한 기술은 수집·공유 단계, 저장·관리 단계, 처리 단계, 분석 단계, 지식 시각화 단계 등 각 처리 프로세스마다 다양한 기술이 존재한다. 그런데 이러한 복잡한 처리 과정이 필요한 데이터를 자유롭게 활용하기 위해서는 데이터의 원천소유자에게 허락을 받는 것이 중요하고 그 데이터를 익명화하여 원천이 어디인지 알 수 없게 해야 대중에게 개방이 가능하다.

일반적으로 데이터를 처리하는 기술은 특허로서 나타나게 된다. 특허는 코카콜라의 조리법처럼 공개될 때 복제가 수월한 아이디어를 제외하면 대부분의 기술을 보호하는 역할을 한다. 즉, 기업, 산업, 국가의 경쟁력을 결정하는 미래 사회의 핵심 요소라고 할 수 있다. 이러한 특허는 Competitive Technical Intelligence Report, White Space Analysis 또는 Technical Gap Analysis 같은 다양한 방식으로 분석되고 있다. 특허 분석은 특정 분야를 이해하는 데 필요한 유용한 정보를 추출하기 위해 방대한 특허 데이터 세트를 사용하는 연구이다[11]. 특허 분석을 통해 특허의 공개 추세 또는 특허의 출원 추세에 대한 통찰력을 얻을 수 있으며, 특허 양수인 또는 특허를 출원한 회사 및 그 수 그리고 특허가 국가에 걸쳐 확산되는 방식 등에 연구가 이루어지고 다른 접근 방식도 자주 사용된다[12]. 예를 들어, [13]은 특허 분석을 위한 키워드 전략에 초점을 맞추고 특허 분석을 위한 키워드 선택 및 처리에 대한 지침을 제공했으며, [14]는 특허 문서의 분석 및 요약을 위한 방법론을 제시하였다. 텍스트 마이닝 및 시각화 기반 접근 방식은 방대한 문헌에서 특허 내용을 분석하는 데에도 사용되었다[15]. [16]은 정보통신기술과 관련된 표준에서 기술적 융합을 확인하였다. 그들은 소셜 네트워크와 연관 규칙 분석을 적용하였다. [17]은 트렌드 분석과 네트워크 기반 연구와 키워드 기반 연구를 결합한 방법을 사용하여 발광 다이오드 및 무선 광대역 분야 관련 특허를 분석하였다. 그리고 미국의 가상화 기술 개발을 탐색하고 기술 수명 주기, 양수인 조직 및 국가, 특허 분류 및 특허 인용을 분석하는 데도 사용된 바 있다[18]. [19]는 네트워크 분석을 사용하여 Green ICT의 기술 보급과 혁신주체의 다양성을 조사하였다. 그리고 [6]은 미국특허 데이터를 이용하여 5개 지수인 기술의 독창성, 기술의 현지화, 기술의 다각화, 기술수명주기, 기술의 집중화를 측정하여 혁신체제의 중요성을 강조한 바 있다. [6]의 연구를 발전시킨 바 있는 [7]은 5개 지표가 [20]에서 제시한 Economic complexity Index와 비교할 때 상대적으로 외부 환경에 독립적이라는 것을 확인하였다. [8]은 국가별 혁신체제의 형태가 다르다는 것을 보이며 유형별 성장에 차이가 있음을 보였다. 그리고 [9]는 ICT기술과 비교시 AI기술은 혁신성이 크고 기존 기술의 발전형태보다는 새로운 형태의 기술에 가깝다는 결론을 제시하였다. 이러한 다양한 특허분석 연구에도 불구하고 우리가 아는 한, 데이터 익명화와 관련된 특허에 대한 분석은 [21]을 제외하면 거의 없다. 물론 [2]에서 익명화에 대한 지침 등을 규정하였으나 관련 기술에 대한 분석을 수행하지는 않았다. 따라서 본 연구에서 수행되는 익명화 관련 특허 기술의 분석은 해당 분야의 개념을 이해하는데 도움이 될 것이고 개인 정보 보호 수단에 대해 더욱 많은 정보를 제공할 것이다.

Ⅲ. 연구방법론

본 연구에서 특허분석을 위한 단계는 다음 4단계로 구성된다. (i) 특허 선택, (ii) 기술 분야에 대한 분석, (iii) 특허권자인 국가와 기업에 대한 분석, 그리고 (iv) 기술체제 측정 및 텍스트 마이닝 분석이다. 특허 검색 및 선택을 위한 자료는 미국특허청의 등록특허 데이터를 이용한다. 미국 특허청의 등록 특허 데이터베이스는 미국에서 등록된 특허에 대한 원 특허문서를 텍스트 파일형태로 제공하고 있다. 데이터 익명화와 관련된 특허를 분석하기 위해 abstract에서 "data"를 포함하는 특허 중에서 “anonymizing”, “anonymization”, “anonymized”, “anonymizy”, “anonymize”와 영국식 철자인 “anonymisation” 같은 단어 중 하나가 포함된 특허를 추출하였다[21]. 그리고 데이터 익명화와 비교를 하기 위한 대조군으로서 데이터 보호에 관한 특허를 추가로 분석한다. 데이터 익명화와 마찬가지로 abstract에서 "data"를 포함하는 특허 중에서 protect를 포함하는 특허를 대상으로 한다. 분석 기간은 1976년 1월 1일 이후 2021년 12월 31일까지 USPTO에서 등록된 모든 특허를 대상으로 한다. 약 8백만건의 특허와 1억여건의 인용정보가 포함된다. 특허는 상태에 따라 활성, 비활성-거부, 거부, 정지 또는 비활성-철회/양도 등으로 구분할 수 있는데, 우리의 분석에서는 활성화 여부와는 상관없이 원 특허문서에 존재하는 모든 특허를 대상으로 하였다.

데이터 익명화에 대한 특허의 기술적 내용을 판별하기 위해서 우리는 기존 연구인 [21]이 국제특허분류(International patent classification: IPC)를 사용한 것과 달리 CPC(Cooperative patent system)를 사용하기로 한다. CPC는 IPC와 섹션 측면에서는 차이가 없고 더 세부적으로 구분된 것이 특징이다. 물론 큰 분류에서는 사실상 차이가 없다. CPC는 Section, Class, Subclass, Group, Subgroup로 세분화되는데 특허지표 생성의 기술적 한계가 존재하기 때문에 우리는 Subclass 단위에서 특허를 추출하고 분류하여 분석한다.

우리는 익명화 관련 특허의 기술체제를 측정하고 기존 연구들과 비교를 위하여 [9]에서 사용된 지표들을 도입한다. Technological cycle time(기술수명주기: TCT), Localization(기술의 현지화), Originality(기술적 창의성), Scifi(과학기술논문 인용도), Generality(기술의 범용성), Appropriability(전유성) 등을 추정한다[9]. 각 지표들은 다음과 같이 측정될 수 있다. 먼저 TCT는 원 특허가 인용한 특허들이 얼마나 최근 특허인지 그 기간의 차이를 측정하는 것이다. 최근 특허를 인용할수록 기존 특허를 검토하는 시간이 줄어들어 후발자가 추격하는데 우월하다고 볼 수 있다[6].

$$TCT_i = (\text{원특허의 등록년도} - \text{인용된 특허의 등록년도})/N \quad (1)$$

여기서, TCT_i는 i특허의 등록연도에서 i특허가 인용한 특허의 등록연도를 차감한 값을 의미한다. 그리고 인용된 특허의 수로 나누어 평균 기술수명주기를 측정한다. 다음 Localization(현지화)은 인용한 특허에서 자국 특허의 비율을 측정하는 것이다. 자국 특허 인용 비율이 높을수록 기술의 외주화를 지양하고 기술적 독립성이 높다는 것을 의미한다.

$$Localization_i = \frac{n_{ii}}{n_i} \quad (2)$$

여기서, n_i는 i특허가 인용한 특허 총 특허수이고, n_{ii} 는 i특허가 인용한 특허 중에서 국적이 같은 특허의 수이다. 다음 Originality는 기술적 창의성을 의미하며 얼마나 다양한 분야의 특허를 인용했는지를 측정한다. 이 지표 값이 클수록 다양한 분야의 특허를 인용하여 기술적 창의성이 크다고 볼 수 있다.

$$Originality_i = 1 - \sum_{k=1}^{N_i} \left(\frac{NCited_{ik}}{NCited_i} \right)^2 \quad (3)$$

여기서, k는 Subclass 단위의 CPC, $NCited_{ik}$는 k는 Subclass k에 속해 있는 i특허가 인용한 특허수, $NCited_i$는 i특허가 인용한 특허수이다. 그리고 HHI(Herfindal - Hershman Index: 허핀달-허쉬만 지수)는 수식 (3)과 동일한 수식을 사용하여 특허분류를 대체하여 특허권자의 이름으로 특허권자의 분산도를 측정한다. 다음 Scifi는 특허 출원시 특허를 제외한 비특허 논문을 이용한 횟수를 측정한 것이다. 기술이 고도화될수록 과학기술논문인용도가 증가한다는 기존 연구의 증거가 있다[9].

$$Scifi_{it} = N_{it} \quad (4)$$

여기서 N_{it}는 인용된 과학기술논문의 개수를 의미한다. 다음 Generality(범용성)는 현 특허가 후행 특허의 어떤 분야에서 인용되고 있는 지를 측정하는 지표이다. 즉, 범용성이 높다는 것은 다양한 분야에서 인용되고 있는 범용성이 높은 특허라고 생각할 수 있다.

$$Generality_i = 1 - \sum_{k=1}^{N_i} \left(\frac{NCiting_{ik}}{NCiting_i} \right)^2 \quad (5)$$

여기서, k는 Subclass 단위의 CPC, $NCiting_{ik}$는 k가 Subclass k에 속해 있는 i특허가 인용한 특허수, $NCiting_i$는 i특허가 인용한 특허수이다. 그리고 Appropriability는 전유성이라고 정의되며, 이익 창출이 가능한 지를 살펴보는 지표로서 개발되었다. 여기서는 전체 인용한 특허에서 자기 특허의 비율을 측정하여 self-citation을 측정하기 위함이다.

$$Appropriability_i = \sum_{t=t_1}^{t_n} SC_{it} / \sum_{t=t_1}^{t_n} TC_{it} \quad (6)$$

여기서 SC_{it}는 t년도의 i특허에 의한 자기인용 특허수, TC_{it}는 t년도의 i특허에 의한 총인용 특허수를 말하는데 특허권자의 이름을 기준으로 하며, n은 분석기간이다. 우리는 위 지표들을 이용하여 데이터 익명화와 관련된 특허들이 어떤 기술체제 특성을 갖고 있는지 그리고 다른 기술들과 어떤 차별성이 있는지를 분석한다. 그리고 각 지표 중 TCT, Originality, Localization, Generality, 과학기술논문 인용도는 해당 연도의 모든 특허의 평균값으로 해당 값을 나누어 추세에 따른 변화를 최소화시켰다.

그리고 텍스트 마이닝 접근 방식은 데이터 익명화와 관련된 특허에서 등장하는 주요 주제를 감지하는 데 활용되고 있다. 특허 제목에 대한 텍스트 마이닝은 데이터 익명화와 관련된 가장 자주 등장한 주제를 결정하는 데 사용된다. 이 방식은 보통 단어의 가변성의 크기를 줄이기 위해서 필터링, 표제어 추출 또는 형태소 분석과 같은 다양한 접근 방식을 사용한다.

Ⅳ. 분석결과

4-1 익명화 관련 특허권의 소유와 기술 분류 동향

분석에 이용된 USPTO의 특허데이터는 1976년 1월 1일에서 2021년 12월 31일 간을 대상이다. 분석의 편의를 위하여 특허권자의 이름과 국적은 첫 번째 특허권자를 기준으로 한다[6]. 다음 그림은 1976년에서 2021년 사이의 데이터 보호 및 익명화와 관련된 특허수 추세를 나타내고 있다(그림 1). 데이터 보호와 관련된 특허수가 급격히 증가하는 모습을 보이는 가운데 익명화와 관련된 특허수는 2010년대 이후 서서히 증가하는 것으로 나타났다. 1976년~2021년 사이 데이터 보호 관련 특허는 13,431건이고 익명화 관련 특허는 581건이다. 추출된 특허들의 특허번호를 비교해 보면, 익명화 기

술 581건 중 데이터 보호 특허와 겹치는 특허는 50건에 불과하였다. 익명화 기술이 데이터 보호보다는 개인정보 보호에 더 가까운 측면이 있기 때문에 두 범주에 차이가 나타나는 것으로 볼 수 있다.

다음 그림은 검색된 특허가 속해있는 주요 CPC 특허분류를 나타낸 것이다(그림 2). 익명화 관련 특허는 대체로 3가지 CPC 섹션에 집중되어 있음이 밝혀졌는데, A섹션(인간에게 필요한 기술), G섹션(물리학) 및 H섹션(전기)이다. 이 중 G06F로 분류된 특허가 264건, H04L이 107건, G06Q가 58건, G16H가 45건, A61B가 14건 등의 순이다. 여기서 G06F는 Electric digital data processing(전기 디지털 데이터 처리), H04L은 Transmission of digital information(디지털 정보 전송), G06Q는 Data processing systems or methods(데이터 처리 시스템 또는 방법), G16H는 Healthcare informatics(보건 정보과학), 그리고 A61B는 Diagnosis;Surgery;Identification(진단; 수술; 식별)로 구성된다. 특허의 기술 분류를 보다 직관적으로 살펴보기 위해서 클라우드 분석을 수행하였다[22].

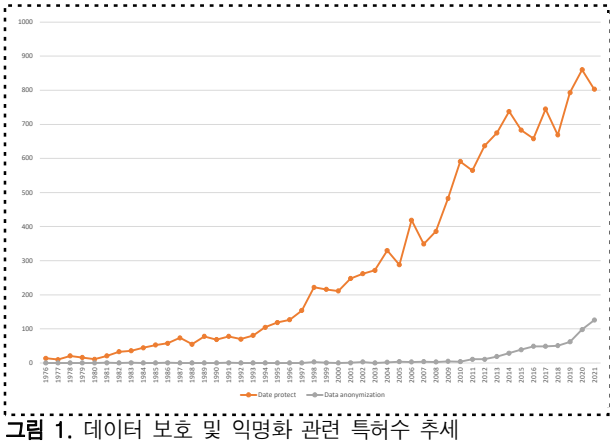

그림 1. 데이터 보호 및 익명화 관련 특허수 추세
Fig. 1. Trends in the number of patents related to data protection and anonymization

그림 2. 익명화 관련 특허의 특허분류
Fig. 2. Patent classification of anonymization-related patents

클라우드 분석은 텍스트에서 가장 흔한 단어를 단어의 크기와 상대적 빈도와 관련하여 시각화하기 때문에 가장 많이 발생하는 주제를 시각화하는 일반적인 방법이다. 따라서 더 자주 발생하는 단어가 더 크게 나타난다. 즉, 데이터 익명화 분야의 기술이 특정 분야인 G06F에 집중되어 있다는 것을 의미한다. 단순히 빈도수로 보면 G06F에 속한 기술이 다른 분야에 속한 기술들 모두를 합친 것과 거의 유사한 수치를 나타내고 있다.

다음 표는 첫 번째 특허권자를 기준으로 특허권자의 국적을 나타낸 것이다(표 1). 미국 특허를 대상으로 분석했기 때문이기도 하지만 기술적으로 대규모 IT기업이 존재하는 미국의 특허가 409건으로 전체의 70%를 차지하고 있으며, 독일이 34건으로 2위, 일본 24건, 캐나다 20건, 아일랜드 11건, 중국 7건, 프랑스 7건 등이었다. 한국은 2건에 그치고 있다. 분석에 따르면 특허 활동이 여러 국가에 걸쳐 있는 것으로 나타났지만 미국, 일본 및 독일이 80%이상을 점유하고 있었다. 1976년에서 2021년 사이의 기간 동안 국가별 특허 소유 현황을 간략하게 보여줍니다. 분석한 특허데이터가 미국 특허를 기반으로 했기 때문에 미국 특허가 가장 많은 것은 당연하지만, 유독 그 차이가 컸다. 주어진 기간에 5개 이상의 특허를 할당한 유럽 국가는 독일(34개), 프랑스(7개), 스위스(6개) 등 이다. 발명인은 존재하지만 특허권자는 명시하지 않은 특허도 23개나 되며 기타 유럽국가도 21, 그리고 기타 국가들은 19개로 나타난다.

표 1. 국가별 익명화 관련 특허수
Table 1. Number of patents related to anonymization by country

Country	Number of Patents	Ratio (%)
USA	409	70.4
Germany	34	5.9
Japan	24	4.1
Canada	20	3.4
Ireland	11	1.9
China	7	1.2
France	7	1.2
Swiss	6	1.0
Non-country	23	4.0
Europe countries	21	3.6
Other countries	19	3.3
Total	581	100.0

다음 표는 익명화 관련 특허의 특허권자별 보유현황을 나타낸다(표 2). 최소 특허 8개 이상을 가진 특허권자를 대상으로 정리하였다.

표 2. 특허권자별 익명화 관련 특허수
Table 2. Number of patents related to anonymization by patent holder

Holder	Nations	Number of Patents	Ratio (%)
IBM	USA	68	11.7
CipherCloud	USA	18	3.1
MS	USA	18	3.1
SAP	Germany	16	2.8
AT&T	USA	12	2.1
ADP	Ireland	10	1.7
Privacy Analytics	Canada	10	1.7
AMAZON	USA	8	1.4
Other	-	421	72.5
Total	-	581	100.0

특허권자는 회사, 학술 기관 및 개인 등을 모두 포괄하는데, 관측 기간 중 데이터 익명화와 관련된 단순 특허수가 가장 많은 기관은 미국의 IBM으로 68개이며, CipherCloud, MS, SAP, AT&T가 뒤를 잇고 있다. 상위 21개 특허권자 중에서 독일, 아일랜드, 캐나다, 일본을 제외하면 다른 국가가 없는 것도 특징이다. 다만 기타 특허권자의 비율이 72.5%에 달하며 소수의 특허를 보유한 특허권자가 많다는 것이 확인된다. 표 1과 표 2를 통해 알 수 있는 것은 익명화 기술 특허가 국가별로는 편중되어 있는 것으로 나타나지만 내부적으로 보면 다수의 기업들이 특허를 보유하고 있어 상대적으로 기술적 분산도가 높은 것으로 볼 수 있다. 즉, 뚜렷한 선발자가 존재하지 않는다는 것이다. 이 결과는 기술체제 분석을 통해 확인할 수 있다.

4-2 익명화 관련 특허의 기술체제 분석

다음 표는 익명화 관련 기술 특허의 기술체제를 측정하고 데이터 보호 관련 기술과의 차이를 나타낸 것이다(표 3). 데이터를 보호하기 위해서는 익명화, 암호화, 네트워크를 보호하는 보안 등 다양한 방법이 동원된다. 데이터 보호는 디지털 콘텐츠 데이터가 범람하는 21세기에 중요한 이슈이다. 특히 그 중에서 개인정보 보호를 위한 익명화 기술은 데이터를 원활히 사용하기 위한 필수적 조건으로 주목받고 있다. 따라서 기존 선발자들이 존재하는 데이터 보호 기술 분야와 비교하여 익명화 기술 분야에 후발자를 위한 기회의 창이 존재하는지를 살펴보는 것은 기업의 입장에서 중요하다. 기회의 창이란 기술적 패러다임의 전환이 있을 시 기존 기술의 선발자와 새로운 기술에 진입하려는 후발자가 모두 동일한 출발선 상에 서있다는 것을 말한다[6]. 따라서 두 기술 집단 간의 비교는 익명화 기술이 새로운 기술인지 여부와 후발자에게 경쟁 기회가 있는지 여부를 확인하는데 중요한 자료가 된다.

표 3. 기술체제 분석 (익명화 vs. 데이터 보호)
Table 3. The Analysis of Technological regime (anonymization vs. data protection)

	Anony (A)	Protect (B)	(H0:A-B=0) t -value
TCT	0.81 (0.06)	0.79 (0.01)	0.44
Originality	0.98 (0.09)	1.02 (0.02)	-0.43
Localization	1.38 (0.07)	1.15 (0.02)	3.03***
Generality	1.28 (0.86)	1.13 (0.02)	1.69
Scifi. Citation	0.66 (0.08)	0.92 (0.05)	-2.87**
HHI	0.32 (0.06)	0.07 (0.02)	4.15**
Appropriability	0.41 (0.01)	0.80 (0.003)	-4.21**

Note: 1) Anony is an abbreviation for the data anonymization.
2) Protect is an abbreviation for the data protection.
3) Standard errors in parentheses; * p≦0.10, ** p≦0.05, *** p≦0.01

두 기술 집단의 비교에는 이분산을 가정한 두 표본 t-테스트를 적용하였다. 익명화 기술과 데이터 보호 기술은 기술수명주기, 기술적 창의성, 기술의 현지성과 기술의 범용성에서 큰 차이가 없는 것으로 나타났다. 두 기술 모두 디지털 자료에 대한 보호 또는 변환을 목적으로 하기 때문에 같은 세대의 기술 범주에 속한다고 생각해 보면 이해에 큰 무리가 없다. 다만, 과학기술논문인용도와 특허권자 집중도, 전유성 측면에서는 차이가 컸다. 익명화 기술 관련 특허가 상대적으로 비특허 자료를 적게 이용하고 있는데 익명화 이슈가 등장한 시기가 오래되지 않았기 때문에 관련 연구의 수가 상대적으로 부족하기 때문이다. 그리고 특허권자 분산도는 익명화 기술 분야에서 더 높은데 이는 아직 익명화 기술 분야는 새로이 떠오르는 기술로서 상대적으로 집중하고 있는 기업이 적다는 것을 의미한다. 그리고 전유성은 특허권자의 자기인용 정도를 측정하는 것으로서 데이터 보호에 비하면 상대적으로 낮았다. 익명화 기술의 특허권자의 분산도는 상대적으로 높지만 전유성은 낮기 때문에 다수의 특허권자가 외부의 기술을 이용하여 기술을 개발하고 있다는 것으로 해석할 수 있다. 반면, 데이터 보호 관련 특허는 분산도는 낮은 대신 전유성이 높아서 다양한 특허권자들이 자신만의 기술로서 경쟁하는 독점적 경쟁시장에 가깝다고 추정할 수 있다.

그리고 본 연구의 분석 결과는 [9]와 연계된 데이터 셋을 사용하였기 때문에 기존 연구와 직접적으로 비교가 가능하다. [9]에서 제시한 4차 산업혁명이나 3차 산업혁명 기술들과 비교하면, AI관련 기술과 유사한 수준의 기술수명주기 값을 나타내고 있어 다른 4차 산업혁명 기술 대비 길게 나타났다. 그리고 창의성은 1.38로 다른 기술대비 높았고 일반성은 큰 차이가 없었다. 과학기술논문 인용도를 보면 본 연구에서는 0.66으로 기존 연구에서 주요 4차 산업혁명 기술의 평균값 0.85에 비해 상대적으로 낮았다. 익명화 기술과 데이터 보호 관련 기술은 기술수명주기 측면에서 보면 4차 산업혁명 기술에 가깝지만 다른 기술체제 값에서는 특정 분야에 해당하는 기술이라고 확정할 수 있는 근거는 없었다.

후발자에게 중요한 기회의 창 존재 여부는 기술의 창의성, 기술수명주기, 그리고 기술의 현지화 등으로 평가해 볼 수 있다. 우선 기술수명주기가 짧으면 기술의 변화가 빠르기 때문에 후발자에게 유리하다는 것은 기존 연구에서 증명된 바 있다 [6]. 본 연구에서는 익명화 기술의 기술수명주기가 조금 길었다 하더라도 통계적으로 유의성이 떨어져 기술수명주기로 판단은 어려웠다. 다만, 데이터 보호 기술이나 익명화 기술 모두 데이터 기반 산업에서 핵심을 담당하는 기술로서 신기술에 속하는 분야이기 때문에 기술수명주기가 짧은 것은 두 분야 모두에서 기회의 창이 열려있다는 것을 의미한다. 다만, 익명화 기술 분야에서 기술의 현지화 수준이 높고 자기인용 비율이 상대적으로 낮기 때문에 뚜렷한 기술적 우위를 가진 기업이 존재하지 않는다는 것을 의미하며 데이터 보호 기술보다는 기회의 창이 더 넓다고 볼 수 있다. 더불어 두 기술의 기술체제 특성이 차이가 존재하기 때문에 기술의 동질성 여부를 분석해 보면 581개 익명화 기술 중 데이터 보호와 중복되는 기술은 50개 정도로 10%가 되지 않는 것으로 나타났기 때무에 두 분야의 선발자가 중복되지 않을 가능성도 높다는 것을 의미한다.

4-3 특허 주제 식별을 위한 텍스트 마이닝

우리는 형태소 접근 방식을 사용하여 특허 제목에서 나타나는 대표적인 단어들을 추출하기 위하여 텍스트 마이닝 분석을 사용한다. 익명화 기술이 어떤 키워드로서 등장하고 어떤 분야에서 가장 많이 활용되고 있는 지를 확인해 볼 수 있다. 즉, 다음 그림은 데이터 익명화와 관련된 특허의 제목에서 가장 많이 사용되는 어간 또는 구문을 보여준다(그림 3). 분석된 결과를 보면, 익명화 기술이라는 용어를 제외하고 computing, medical 등이 눈에 띈다. 즉, 익명화 기술이 의료분야에서 많이 적용되는 컴퓨터 기술이라고 간단히 설명될 수 있다.

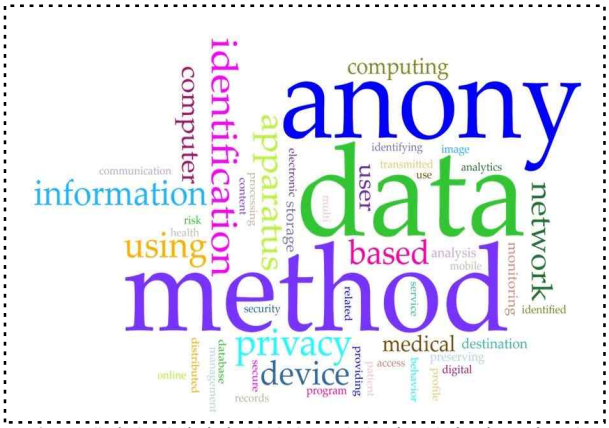

그림 3. 데이터 익명화와 관련된 특허의 제목에서 자주 사용되는 단어의 클라우드

Fig. 3. Cloud of frequently used words in the titles of patents related to data anonymization

그리고 특허의 제목에서 가장 많이 등장한 단어는 "data"이다. 또한 “method”, “anony*”, “identification” 순으로 가장 자주 사용되는 문구이다. 또한 “data”가 291건, “method”가 241건, “anony*”기 231건으로 세 단어가 대다수를 차지하고 있다. 다음 주제 그룹도 확인되었는데, (i) "정보" 또는 "장치"와 같은 물리적 장비와 관련된 주제; (ii) "프로그램", "프로세스" 및 "분석" 또는 "관리"와 같은 소프트웨어와 관련된 주제; (iii) "프라이버시", "정보" 와 같은 특허의 목표와 관련된 주제 (iv) "컴퓨터", "의학" 또는 "네트워크"와 같은 구현 영역과 관련된 일부 특정 주제가 도출되었다.

V. 결 론

본 연구는 데이터 익명화와 관련하여 미국 특허청에서 제공하는 특허문서를 기반으로 분석을 수행한다. 분석기간은 1976년부터 2021년까지이며 데이터 익명화와 관련된 581개의 특허가 추출되었다. 또한 대조군으로서 익명화와 관련이 깊은 데이터 보호 기술을 추가하였는데 약 13,000여 건이었고 익명화 기술과는 50건 정도만 겹치게 나타났다. 실증 분석은 (i) 데이터 익명화 및 보호에 관한 특허 검색, (ii) 익명화 특허의 기술분류 분석, (iii) 익명화 특허의 특허권자와 특허권자가 속한 국가에 대한 분석, 그리고 (iv) 기술체제에 대한 분석과 주제 식별을 위한 텍스트 마이닝의 4단계로 수행되었다. 우리 분석은 디지털 콘텐츠의 활용과 개인정보 보호 사이에서 어떤 기술적 흐름이 있는지에 대한 의문에 답변을 제공한다. 또한 익명화 관련 기술의 특허 환경에 대한 상세한 정보를 제공한다.

첫 번째 연구 목표는 데이터 익명화 기술의 개발 현황, 즉 특허 동향을 파악하는 것이다. 특허수는 2011년 이후 서서히 증가하기 시작하여 2014년을 기점으로 급격히 증가하였다. 그리고 2019년에 급성장하였는데 이는 4차 산업혁명에 대한 관심이 대중으로 확대되면서 개인정보 보호에 대한 관심이 커졌기 때문이다. 데이터의 양이 적은 경우에는 수작업으로 익명화 작업이 가능하였으나 데이터가 빅데이터라는 새로운 분야로 발전하면서 기술의 중요성도 커졌기 때문이다. 두 번째 연구 목표는 익명화와 관련한 특허들이 보호되는 기술 분류에 대하여 조사하는 것이다. 즉, 유럽과 미국에서 사용되고 점점 확대되고 있는 CPC라고 불리는 특허분류 시스템에 기반을 두어 분석하는 것이다. 익명화와 관련된 특허는 대다수가 CPC 기준으로 G에 해당하는 물리학에서 추출되었다. 특히 G06F(전기 디지털 데이터 처리)라는 특허클래스에서 발견되었다. 특히 세부적으로 보면, 정보 검색 및 데이터베이스 구조에 맞는 디지털 컴퓨팅 또는 데이터 처리 장비 또는 방법 등에 사용되는 기술로서 정의된다. 즉 발명의 초점이 정보의 디지털화에 따른 데이터 프라이버시 보호가 주요 관심사였다는 것을 확인할 수 있었다. 세 번째 연구 목표는 익명화 혁신에 대해 특허를 보유한 국가와 조직을 구별해 내는 것이다. 분석에 따르면 익명화 기술은 다양한 국가에서 보유하고 있

지만 상당수가 미국, 독일, 일본 등 기술 선진국에서 나타나고 있는 것을 확인하였다. 분석기간 중 IBM이 가장 많은 특허를 갖고 있었으며, CipherCloud, MS, SAP, AMAZON, 카드회사, 은행에서 관련 기술들을 보유하고 있는 것으로 나타났다. 네 번째 연구 목표는 기술체제를 분석하는 것이다. 본 연구에서 추출한 익명화 기술과 데이터 보호 기술을 비교하고 기존 연구에서 수행된 기술과 어떤 차이가 있는지 살펴보았다. 익명화와 데이터 보호 기술은 겹치는 특허가 50건 정도 밖에 존재하지 않지만 기술수명주기, 창의성, 현지화, 범용성 등의 지표에서 큰 차이가 없었다. 다만 과학기술논문인용은 데이터 보호 측면에서 더 높았고 특허권자 분산도는 익명화 분야에서 더 높았다. 그리고 자기인용은 데이터 보호 분야에서 더 높았다. 익명화 기술의 출현 시점이 최근 10년 이내이고 관심도가 높아진 시점은 최근 5년 이내라고 본다면 다양한 기업들이 진입하였으나 아직 경쟁력 있는 선도 기업은 없다는 것을 의미한다. 더불어 [9]의 연구 결과와 비교하면, 익명화 기술과 데이터 보호 기술을 4차 산업혁명 기술이라고 볼 근거는 약하였다. 이 두 기술은 정보의 디지털화가 촉진되면서 그 중요성이 높아지고 있지만 데이터 보호 기술은 이미 70년대에도 존재하던 오래된 기술이었고 익명화 기술은 상대적으로 늦게 태동하였지만 AI처럼 다른 기술과 격차를 나타내는 기술이라고 볼 근거는 없었다. 그리고 익명화와 관련된 특허의 주제로 가장 많이 등장하는 이슈를 식별하였다. 가장 많이 사용된 단어는 “데이터”, “방법”, “익명” 이었다. 물리적 장비, 소프트웨어, 보호, 식별, 암호화 또는 개인 정보 보호, 커뮤니티, 의료 또는 서비스와 같은 특정 주제도 등장하였다.

본 연구는 다양한 측면에서 익명화 기술의 현황과 기술체제를 살펴보았지만, 또한 한계도 존재한다. 익명화 기술추출 과정에서 데이터와 익명화를 모두 포괄하는 특허만 검색하는 키워드 추출방식을 이용하였기 때문이다. 키워드 추출방식은 다양한 특허분석연구에서 사용되고 있는 것이 현실이지만 기술적 특성을 반영하지 못할 수 도 있다는 한계가 있다. 기술이라는 것이 어떤 한 분야에서만 이용되는 것도 아니고 특정 키워드가 없이도 제품이나 후행 기술에 사용될 수 있다는 가능성을 감안하지 못하기 때문이다. 이는 특허의 특성으로 인한 것으로서 익명화 기술에 대한 상세한 필수 기술정보를 분석하여 필요한 기술에 대한 특허만 추출하는 방법으로 보완이 필요하다.

감사의 글

이 논문은 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF- 2020S1A5C2A02093223)

참고문헌

[1] K. Schwab, *The fourth industrial revolution*, New York, NY: Crown Business, 2017.

[2] G. Cormode and D. Srivastava, "Anonymized Data: Generation, models, usage," in *Proceedings of 2010 IEEE 26th International Conference on Data Engineering*, pp. 1211-1212, April 2010. https://doi.org/10.1109/ICDE.2010.5447721

[3] World Intellectual Property Organization(WIPO), *Guide to the IPC. WIPO*, 2015.

[4] Patent Lens. Patent Laws Around the World, 2016. http://www.bios.net/daisy/patentlens/ip/around-the-world.html

[5] J. Kim and S. Lee, "Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO," *Technological Forecasting and Social Change*, Vol. 92, pp.332-45, 2015. https://doi.org/10.1016/j.techfore.2015.01.009

[6] K. Lee, *Schumpeterian analysis of economic catch-up: Knowledge, path-creation, and the middle-income trap*, Cambridge, UK: Cambridge University Press, 2013.

[7] K. Lee and J. Lee, “National innovation systems, economic complexity, and economic growth: country panel analysis using the US patent data,” *Journal of Evolutionary Economics*. VoL.30, NO.4, pp.897-928, 2020. https://doi.org/10.1007/s00191-019-00612-3

[8] K. Lee, J. Lee, and J. Lee, "Variety of National Innovation Systems (Nis) and Alternative Pathways to Growth Beyond the Middle-Income Stage: Balanced, Imbalanced, Catching-up, and Trapped Nis." *World Development*, Vol.144, 2021. https://doi.org/10.1016/j.worlddev.2021.105472

[9] J. Lee, and K. Lee, "Is the Fourth Industrial Revolution a Continuation of the Third Industrial Revolution or Something New under the Sun? Analyzing Technological Regimes Using Us Patent Data," *Industrial & Corporate Change*, Vol. 30, No.1, pp.137-159, 2021. https://doi.org/10.1093/icc/dtaa059

[10] D. Laney, 3D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Note, 6, 2001.

[11] M. Sinha, and A. Pandurangi, *Guide to Practical Patent Searching and How to Use Patseer for Patent Search and Analysis*, Baner, India: Gridlogics, 2016.

[12] E. Grant, M. Van den Hof, and E. Gold. "Patent Landscape Analysis: A Methodology in Need of Harmonized Standards of Disclosure." *World Patent Information*, Vol.39, pp.3-10, 2014. https://doi.org/10.1016/j.wpi.2014.09.005

[13] H. Noh, Y. Jo, and S. Lee, "Keyword Selection and

Processing Strategy for Applying Text Mining to Patent Analysis." *Expert Systems with Applications*, Vol.42, No.9, pp.4348-60, 2015.
https://doi.org/10.1016/j.eswa.2015.01.050

[14] S. Brügmann, N. Bouayad-Agha, A. Burga, S. Carrascosa, A. Ciaramella, M. Ciaramella, J. Codina-Filba, et al. "Towards Content-Oriented Patent Document Processing: Intelligent Patent Analysis and Summarization." *World Patent Information*, Vol.40, pp.30-42, 2015.
https://doi.org/10.1016/j.wpi.2014.10.003

[15] A. Abbas, L. Zhang, and S. Khan. "A Literature Review on the State-of-the-Art in Patent Analysis." *World Patent Information*, Vol.37, pp.3-13, 2014.
https://doi.org/10.1016/j.wpi.2013.12.006

[16] Han, Eun Jin, and So Young Sohn. "Technological Convergence in Standards for Information and Communication Technologies." *Technological forecasting and social change*, Vol.106, pp.1-10, 2016.
https://doi.org/10.1016/j.techfore.2016.02.003

[17] J. Choi, and Y. Hwang. "Patent Keyword Network Analysis for Improving Technology Development Efficiency." *Technological Forecasting and Social Change*, Vol.83 ,pp.170-182, 2014.
https://doi.org/10.1016/j.techfore.2013.07.004

[18] S. Ju, M. Lai, and C. Fan, "Using Patent Analysis to Analyze the Technological Developments of Virtualization." *Procedia-Social and Behavioral Sciences*, Vol. 57, PP.146-154, 2012.
https://doi.org/10.1016/j.sbspro.2012.09.1168

[19] G. Cecere, N. Corrocher, C. Gossart, and M. Ozman. "Technological Pervasiveness and Variety of Innovators in Green Ict: A Patent-Based Analysis." *Research Policy*, Vol.43, No.10, pp.1827-1839, 2014.
https://doi.org/10.1016/j.respol.2014.06.004

[20] R. Hausmann, C. Hidalgo, S. Bustos, M. Coscia, and A. Simoes. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Cambridge, MA: Mit Press, 2014.

[21] M. Bach, J. Pivar, and K. Dumičić1, "Data anonymization patent landscape," *Croatian Operational Research Review*, Vol.8, No.1, pp.265–281, 2017.
https://doi.org/10.17535/crorr.2017.0017

[22] A. Spindler, S. Leone, M. Nebeling, M. Geel, and M. Norrie, Using synchronised tag clouds for browsing data collections. in *Proceedings of the 23rd international conference on Advanced information systems engineering*, pp.214-228, 2011.

홍은아(Eun-Ah Hong)

2004년 : 서울시립대학교 (경제학사)
2006년 : 서울시립대학교 대학원 (경제학석사)
2009년 : University of Nebraska 대학원 (이학석사)
2014년 : 서울대학교 환경대학원 (박사과정수료)

2012년～현 재: 한국환경산업기술원 환경피해예방실 책임연구원
2020년～현 재: 숙명여자대학교 기후환경에너지학과 박사과정
※관심분야 : 디지털플랫폼, 기후변화, 인증 등

김준엽(Joon-Yub Kim)

2007년 : 한국외국어대학교 (중국외교통상학사)
2011년 : 성균관대학교 대학원 (경제학석사)

2011년～2016년: ㈜에코시안 경제분석가
2017년～2017년: ㈜엔베스트 데이터분석가
2019년～2019년: 한국중소기업벤처연구원 연구원
2012년～현 재: 서울대학교 경제학부 박사과정
※관심분야 : 머신러닝, 딥러닝, AI, 특허

이종호(Jong-Ho Lee)

2004년 : 서울시립대학교 (경제학사)
2006년 : 서울시립대학교 대학원 (경제학석사)
2009년 : University of Nebraska 대학원 (이학석사)
2018년 : 서울대학교 대학원 (경제학박사 -기술경제학, 경제추격론)

2018년～2019년: 연세대학교 바른ICT연구소 박사후연구원
2019년～2020년: 서울대학교 대학혁신센터 선임연구원
2021년～현 재: 인하대학교 법학연구소 AI·데이터법센터 책임연구원
※관심분야 : AI, 데이터, 혁신, 특허, 디지털플랫폼 등