
빅데이터의 데이터마이닝과 저작권법상 일시적 복제*

이 지 호**

목 차

- I. 서언
- II. 빅데이터의 개념
 - 1. 빅데이터의 정의
 - 2. 빅데이터의 처리 · 분석 과정
- III. 빅데이터의 처리 · 분석과정에서의 저작권 문제
 - 1. 빅데이터 자원의 저작물성
 - 2. 빅데이터 자원 수집 구간에서의 데이터 전송
 - 3. 빅데이터 분산 처리 과정에서의 데이터 저장
- IV. 결론

* 이 논문에서 빅데이터 분석에 관한 기술적 이해는 상당 부분 한국저작권위원회 이진태 선임연구원의 선행연구 성과에 의존하고 있음을 밝힌다. 이진태, “빅데이터 활성화와 저작권문제: 하둡(Hadoop)을 중심으로”, 계간저작권, 제102호(2013), 136-173면 참조.

** 성균관대학교 법학전문대학원 학술연구분야 박사과정.

초록

IT기술의 발달과 스마트 기기의 보급으로 전 세계적으로 온라인상에서 수많은 정보가 생성되고 공유된다. 이렇게 발생하는 엄청난 자료들을 일컬어 ‘빅데이터’라고 하는데, 이 많은 양의 데이터가 트렌드 분석이나 미래예측 분석을 하는 정보 자원의 원석으로 활용되면서 작년에 이어 올해에도 IT분야만이 아니라 경제 산업의 핫 키워드로 부상하고 있다. 반면, 빅데이터 활용의 긍정적 파급효과 이면에서 발생할 수 있는 법률적 리스크를 우려하는 견해도 있다. 이에 따라 본 원고에서는 빅데이터 활용과 관련한 저작권 문제를 확인해 보았다. 우선, 빅데이터의 활용하기 위해서는 검색기술에 의해서 온라인상 수많은 데이터를 수집하고 정보를 추출해내는 데이터마이닝의 과정을 거치게 되는데, 이때 대량의 정보를 분산처리하는 시스템을 통한 데이터마이닝 과정에서 저작권으로 보호되는 데이터의 일시적 저장이 발생한다. 우리나라 저작권법은 최근 한미자유무역협정의 이행법안으로 개정된 저작권법에서 일시적 복제를 명문화하고 있다. 따라서 빅데이터 활용을 위한 데이터마이닝 과정에서 발생하는 일시적 저장이 우리 저작권법상 일시적 복제에 해당하는지, 해당한다면 합법적인 이용이 가능한지에 대하여 복제의 개념(법 제2조 제2호)과 일시적 복제에 대한 면책 규정(법 제35조의2)을 중심으로 검토하고, 빅데이터의 원활한 활용을 위한 해결 방안을 현행 저작권법의 체계 안에서 모색해보았다.

주제어

빅데이터, 데이터마이닝, 데이터베이스, 일시적 복제, 일시적 복제의 면책 요건, 빅데이터와 저작권

I. 서언

빅데이터는 작년에 이어 '13년에도 정보통신기술업계뿐만 아니라 경제산업 전반에 걸친 핫키워드로 급부상하였으며¹⁾ 앞으로도 그 열풍을 이어갈 것으로 보인다. 빅데이터란 인터넷 환경과 스마트 기기의 발달로 발생하는 대용량 디지털 데이터를 말한다.

EMC가 발표한 자료에 따르면 2011년에 생성된 디지털 정보량은 무려 1.8제타바이트²⁾로, 매 2년마다 2배씩 증가한다고 한다. 이는 대한민국 국민 모두가 17만 847년 동안 쉬지 않고 매 분마다 트위터에 3개의 글을 게시할 경우 생성되는 데이터량이다. 또 한 사람이 쉬지 않고 4,700만년 동안 시청할 수 있는 2천억개 이상의 고화질(HD) 영화(상영시간 2시간 기준)의 용량과도 같다. 이 같은 데이터가 매년 2배씩 증가해 2020년에는 관리해야 할 정보량이 50배 급증한다고 한다. 그야말로 '빅' 데이터의 시대인 것이다.

초기에는 이렇듯 기하급수적으로 누적되어 쌓여가는 거대한 데이터가 저장 매체를 낭비하는 골칫거리였지만, IT기술의 발달로 통신서비스가 공급자에서 소비자로의 단방향적서비스에서 공급자와 소비자가 정보를 주고 받는 양방향서비스로, 또 이러한 정보를 스스로 처리하고 분석하여 필요한 자료로 만들어주는 지능형(이른바 '스마트')서비스로 진화함에 따라 이러한 정보처리의 원석이 되는 데이터가 주목받기 시작했고, 보다 많은 표본은 보다 정확한 분석이나 예측을 가능하게 만들기 때문에 거대한 양의 빅데이터가 기업의 자산이자 차별화 전략으로 각광받게 된 것이다.

따라서 다양한 민간과 공공영역에서 앞다투어 빅데이터의 분석 기술과 활용 방안을 모색하고 있다.³⁾ 하지만 높은 관심에 비하여 아직까지는 그 효용성에 대

1) 조성훈, "2013년 IT메가트렌드? 빅데이터와 클라우드 주목", 머니투데이뉴스,

<<http://www.mt.co.kr/view/mtview.php?type=1&no=2012092510044087904&outlink=1>>, 검색일: 2013. 7. 30.

2) 데이터의 단위: 킬로바이트(KB) < 메가바이트(MB) < 기가바이트(GB) < 테라바이트(TB) < 페타바이트(PB) < 엑사바이트(EB) < 제타바이트(ZB) < 요타바이트(YB)

3) 황태호, "삼성·SK텔레콤·KT, 빅데이터 활용 프로젝트 가동", 전자신문,

<<http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=004&oid=030&aid=0002184515>>, 검색일: 2013. 7. 30.

한 의문도 존재하므로 빅데이터로부터의 가치 창출을 위하여 그 전략적인 가치를 인식한 다각적인 시각에서의 균형 잡힌 접근이 필요할 것이다. 근래 들어 우려의 목소리가 높아지는 개인정보보호에 관한 이슈⁴⁾를 보더라도 빅데이터의 효율적인 활용을 위하여 적절한 법규 환경의 조성이 시급한 과제이다. 이 중 저작권법과 관련하여 검토하여 볼 때, SNS를 통하여 이용자들이 생산한 수많은 정보들이 빅데이터의 일부를 구성하는데, 그 정보들은 일부 저작물을 포함하고 있다. 따라서 저작권으로 보호되는 데이터들을 활용한 빅데이터 처리에서 발생할 수 있는 저작권적 이슈사항에 대한 검토가 필요하다. 이에 본고에서는 빅데이터의 개념과 빅데이터의 활용을 위한 정보처리·분석과정에서의 저작권법상의 권리 침해 문제 발생 여부와 규제의 필요성 등에 대하여 알아보하고자 한다.

II. 빅데이터의 개념

1. 빅데이터의 정의

빅데이터에 대하여 다양한 분야에서 그 정의를 내리고 있는데, 시장조사 전문기관인 IDC는 “대규모의 다양한 데이터로부터 수집, 검색, 분석을 신속하게 처리하여 경제적인 가치발굴을 수행하도록 설계된 차세대 기술 및 아키텍처”⁵⁾라고 정의하고 있으며, 기존의 관리 방법이나 분석 체계로는 처리하기 어려운 막대한 양의 정형 또는 비정형 데이터 집합,⁶⁾ 보통 수십에서 수천 테라바이트 정도의 거대한 크기를 갖고 여러 가지 다양한 비정형 데이터를 포함하고 있으며 생성-유통-소비(이용)가 몇 초에서 몇 시간 단위로 일어나 기존의 방식으로는 관리와 분석이 매우 어려운 데이터의 집합⁷⁾ 등으로 정의하고 있다. 과거 빅데이

4) 김현주, “개인정보보호 탓에 걸린 빅데이터] ‘빅데이터와 프라이버시’ 두 마리 토끼 잡으려면 … 개인정보 개념부터 명확하게”, 매경이코노미,

〈<http://news.mk.co.kr/newsRead.php?year=2013&no=472942>〉, 검색일: 2013. 7. 30.

5) IDC, 빅데이터 분석: CIO를 위한 미래지형적 아키텍처, 기술 그리고 로드맵, IDC analyze the future, 2011, 1면.

6) 한국정보통신기술협회(TTA), 정보통신용어사전.

터는 규모와 기술 측면에서 발전하였으나, 가치와 활용효과 측면으로 의미가 확대되는 추세이며 규모(Volume), 다양성(Variety), 복잡성(Complexity), 속도(Velocity)의 공통된 특징을 가지고 있으며 4개 요소가 충족될수록 빅데이터에 적합하다.⁸⁾

〈표 1〉 빅데이터의 4가지 구성요소⁹⁾

구분	주요 내용
규모(Volume)	· 기술적인 발전과 IT의 일상화가 진행되면서 해마다 디지털 정보량이 기하급수적으로 폭증 ⇒ 제타바이트(ZB) 시대로 진입
다양성(Variety)	· 로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터 종류의 증가 · 텍스트 이외의 멀티미디어 등 비정형화된 데이터 유형의 다양화
복잡성(Complexity)	· 구조화되지 않은 데이터, 데이터 저장방식의 차이, 중복성 문제 등 · 데이터 종류의 확대, 외부 데이터 활용으로 관리대상 증가 · 데이터 관리 및 처리의 복잡성이 심화되고 새로운 기법 요구
속도(Velocity)	· 사물정보(센서, 모니터링), 스트리밍 정보 등 실시간성 정보 증가 · 실시간성으로 인한 데이터 생성, 이동(유통) 속도의 증가 · 대규모 데이터 처리 및 가치 있는 현재정보(실시간) 활용을 위해 데이터 처리 및 분석 속도가 중요

또한 이하의 글에서는 ‘빅데이터’에 대한 개념은 넓은 의미인 ‘기존의 관리 방법이나 해석 체계로는 처리하기 어려운 막대한 양의 정형 또는 비정형 데이터 집합’을 말하는 것으로 하고, ‘데이터’란 하기 ‘빅데이터 자원’의 개별 요소들을 의미하는 것으로 정의하고 논의를 이어가고자 한다.

7) 함유근·채승병, 빅데이터, 경영을 바꾸다, 삼성경제연구소, 2012, 36면.

8) 정지선, “신가치창출 엔진, 빅데이터의 새로운 가능성과 대응 전략”, IT & Future Strategy, 제18호(2011), 8면.

9) 정지선, 위의 글(주 8).

〈표 2〉 빅데이터 자원의 분류¹⁰⁾

생성 주체	컴퓨터 생산 데이터	사람 생산 데이터	관계 데이터
	· 어플리케이션 서버 로그 (웹사이트, 게임 등) · 센서 데이터(날씨, 물, 스마트그리드 등) · 이미지, 비디오(트래픽, 보안 카메라 등)	트위터, 블로그, 이메일, 사진, 게시판 글 등	페이스북, 링크드인 등
유형	정형	반정형	비정형
	DB에 저장된 구조적 데이터	웹문서, 메타 데이터, 센서 데이터, 공정 컨트롤 데이터, 콜 상세 데이터 등	소셜 데이터, 문서, 오디오, 비디오, 동영상, 이미지 등

〈표 3〉 빅데이터의 활용 사례

분야	활용사례
의료	진료기록을 활용한 질병 예방 및 응급사항 대비
마케팅	고객의 니즈 및 트렌드 분석
재난	지진, 홍수 등 재해 정보 분석으로 사회 안정망 구축
정치	유권자층의 정치 유관 검색어를 통한 동향 분석 및 각 정당 지지율 예측

2. 빅데이터의 처리 · 분석 과정

1) 빅데이터의 분산 처리

빅데이터를 바라보는 시각은 크게 두 부류로 나뉜다. 하나는 데이터 처리의 방법을 다루는 기술이고, 다른 하나는 가치를 창출하기 위한 분석이다. 인터넷의 대중화는 수많은 웹문서를 생성해냈고, 정보의 홍수 속에서 가장 빠르고 정

10) 정지선, “빅데이터 시대의 데이터 자원 확보와 품질 관리 방안”, IT & Future Strategy, 제3호(2012), 한국정보화진흥원, 4면.

확한 결과를 얻기 위해 검색엔진이 등장하였다. 특히 구글은 전 세계의 모든 웹 문서를 수집하고 페이지랭크(page rank)라고 하는 가중치 부여 알고리즘을 통해 최고의 검색엔진으로 자리 잡았다. 2000년대 초반 구글의 고민은 무한히 생성되는 웹문서를 기존의 시스템과 달리 비교적 저렴하게 수집 및 저장하고 페이지랭크의 계산 기간을 최단으로 단축하는 것이었다. 이를 해결하기 위해 나온 것이 바로 분산 저장 방식을 채택한 “구글파일시스템(GFS)”과 대규모 병렬처리 프레임워크인 “맵리듀스(mapreduce)”이다.¹¹⁾ 분산 시스템상에서 대용량 데이터 처리 분석을 지원하는 오픈소스 소프트웨어 프레임워크로 구현한 것이 “하둡(Hadoop)”이며, 현재 빅데이터 처리기술의 산업상 표준으로 자리잡아가고 있다.

〈그림 1〉 빅데이터 관련 기술의 구조¹²⁾



11) 이현중, “빅데이터 하둡 플랫폼의 활용”, 한국통신학회논문지, 제29권 제1호(2012), 43면.

12) 이지훈 · 김도균, “빅데이터 환경을 위한 하둡”, 마이크로소프트웨어,

〈http://www.imaso.co.kr/?doc=bbs/gnuboard.php&bo_table=article&wr_id=40766〉, 검색일: 2012. 7. 31.

빅데이터의 분석 기술은 통계학, 전산학, 기계학습 데이터마이닝 분야에서 사용되던 기법의 알고리즘을 대규모 데이터 처리에 맞도록 수정·보완하여 빅데이터 처리에 적용시키고 있으며, SNS에서 발생하는 반정형·비정형 데이터의 증가로 인하여 분석기법들 중에서 텍스트마이닝, 오피니언마이닝, 소셜네트워크 분석, 군집 분석 등이 주목받고 있다.¹⁴⁾

(1) 텍스트마이닝(text mining)

텍스트마이닝 기술체계는 자연어 처리, 정보추출, 시각화, 데이터베이스 그리고 기계학습의 분야를 포함하고 있다. 텍스트마이닝에서 가장 일반적으로 사용하는 기법은 특성벡터(feature vector)를 이용하는 것으로, 이 방법은 특성추출(feature extraction) 과정을 통하여 텍스트에 대한 특성벡터를 생성하게 된다. 따라서 텍스트 분석의 기반이 되는 것이 바로 특성추출에 의한 특성벡터이며 이의 통계수치는 각 분석기법들의 근거가 되는 것이다. 특성추출은 텍스트에서 중요한 용어(term)를 인식하여 추출해내는 것으로, 추출된 용어들은 일반적으로 단어의 원형(word)으로 변형되어 특성벡터를 구성하게 된다. 이러한 특성벡터는 문서를 분류하거나 요약하는 데 기초 정보로 사용되며 특성의 중요성을 나타내는 가중치 함수와 지지도 함수의 계산은 단어가 발생한 위치와 발생한 횟수에 기반을 둔다.¹⁵⁾

(2) 데이터마이닝(data mining)

데이터마이닝은 데이터베이스의 데이터처럼 정형화된 데이터를 대상으로 처리하기 때문에 텍스트마이닝에 비하여 특성 간의 연관성 파악이나 규칙 생성 등 매우 다양하고 강력한 알고리즘들이 많이 개발되고 있다. 많은 데이터 가운데 숨겨져 있는 유용한 상관관계를 발견하여 미래에 실행 가능한 정보를 추출해 내고 의사 결정에 이용할 수 있으므로 데이터베이스 마케팅의 핵심기술이라고 할

14) 이진형, “데이터 빅뱅, 빅데이터의 동향”, 방송통신전파채널, 통권 제47호(2012), 45면.

15) 최윤정·박승수, “웹 콘텐츠의 분류를 위한 텍스트마이닝과 데이터마이닝의 통합 방법연구”, 인지과학, 제13권 제3호(2002), 33-46면 재구성.

수 있다. 예를 들어, 상품판매 실적에 있어 판매시간, 연령층, 품목 등 함께 수집할 수 있는 데이터들과의 상관관계를 반영하여 마케팅에 반영하는 것이다.

Ⅲ. 빅데이터의 처리 · 분석 과정에서의 저작권 문제

1. 빅데이터 자원의 저작물성

위에서 살펴본 빅데이터의 기술적 처리 과정은 데이터가 분산파일시스템(HDFS)의 서버(네임노드)에서 분산 저장되고, 텍스트마이닝과 데이터마이닝 등의 분석과정을 거쳐 이용자에게 맞춤형 정보를 추출하여 전달하는 일련의 과정이라고 할 수 있다. 즉, 데이터에 대한 접근 및 수집 · 저장과 요약의 과정에서 실제 물리적인 데이터의 복제가 수반되므로, 이때 처리되는 데이터가 저작권법상 보호받는 저작물이라면 네임노드에의 복제가 발생하고, 각 데이터노드 간의 전송이 발생할 수 있는 것이다.

우리 저작권법은 “저작물은 인간의 사상 또는 감정을 표현한 창작물”이라고 규정하고 있다(법 제2조 제1호). 따라서 빅데이터를 이루고 있는 개별 데이터들은 그 데이터의 특성에 따라 저작물의 해당 여부를 검토할 수 있을 것인데, <표 2>에서 생산주체가 컴퓨터인 데이터는 ‘인간’의 사상이나 감정을 표현한 것이 아니므로 저작물로 보호받지 못한다. 또, 사람에 의하여 수집 · 생성 되었다고 하더라도 현실 세계로부터 관찰이나 측정에 의해서 얻은 가공되지 않은 수치, 문자, 사실 등을 의미하는 정형 데이터는 ‘창작적인 표현’이 아니기 때문에 저작물로 보호 받지 못한다. 날씨정보, 교통정보, 각종 통계, 기업의 재무기록 등이 이에 해당될 것이다.

그러나 생산주체가 사람인 데이터 중 비정형 · 반정형 데이터는 그 창작성의 여부에 따라 저작물로 보호 가능하다. 트위터, 페이스북 등의 SNS서비스 활성화로 개인이 자신의 감정을 표현하여 생성하는 글, 그림, 사진, 음악, 동영상 등의 데이터가 이에 해당될 것이다.

또, 저작권법은 “소재를 체계적으로 배열 또는 구성한 편집물로서 개별적으

로 그 소재에 접근하거나 그 소재를 검색할 수 있도록 한 것”을 ‘데이터베이스’로 규정하고 저작권법의 체계 안에서 보호하고 있다(법 제2조 제19호). 따라서 현실 세계로부터 관찰이나 측정에 의해서 얻은 가공되지 않은 수치, 문자, 사실 등을 의미하는 정형 데이터가 그 개별 데이터로서는 창작성이 결여되어 저작권의 보호대상이 아닐지라도 수집 및 처리의 과정에서 인적·물적인 투자를 하여 체계적으로 배열 또는 구성되었고, 이용자들이 개별적으로 그 소재에 접근하거나 소재를 검색할 수 있다면 이는 저작권법상 데이터베이스로 보호될 수 있다.

그러므로 빅데이터 처리·분석 과정에서 자원이 되는 데이터 중 저작물로 보호 가능한 비정형 데이터나, 누군가 인적·물적 자본을 투자하여 구축해놓은 저작권법상 보호되는 데이터베이스라면 빅데이터 처리를 위한 분산처리 시스템에서의 이용과정에서 복제 및 전송의 문제가 발생할 수 있다.

2. 빅데이터 자원 수집 구간에서의 데이터 전송

인터넷 및 SNS서비스를 통해 실시간으로 발생하는 엄청난 데이터의 양으로 인해, SNS서비스 기업들이 개별적으로 구축하고 운영하던 컴퓨팅 자원과 서비스를 통해 모든 데이터를 처리하는 데 무리가 있는 경우, 기업들은 클라우드 컴퓨팅을 통해 빅데이터를 처리하여 기업의 IT운영비용을 절감하고 협업 환경을 개선한다.¹⁶⁾ 이때, 이용자들이 저작물을 생성하여 저장하는 서버와 분산처리와 분석(데이터마이닝)이 이루어지는 서버가 달라질 수 있으며, 빅데이터 분석 전문기관으로 이관될 수도 있다.¹⁷⁾ 이때, 저작물이 포함된 데이터가 수집되는 기관으로부터 분석처리 전문기관의 서버에 전송되어 복제될 것인데, 이러한 이용으로 복제권 및 전송권¹⁸⁾ 침해에 대한 문제가 제기 될 수 있다.

16) 정교일 외 4인, “빅데이터와 정보보안”, 한국정보기술학회지, 제10권 제3호(2011), 19면.

17) 국내통신업체들을 중심으로 이미 SKTelecom은 인터넷, 뉴스, 블로그, 카페, 눈(페이스북, 트위터, 미투데이) 등을 분석해서 제공하는 SMAT Insight 서비스를 제공하고 있으며, LG유플러스는 LG의 소셜서비스인 와글의 정보를 이용해서 빅데이터 분석을 해주는 서비스를 올해 말 제공할 예정이다. 이진태, “빅데이터의 저작권문제”, 한국저작권법학회 2012년 추계학술대회, 2012, 43면.

18) 저작권법은 “공중송신 중 공중의 구성원이 개별적으로 선택한 시간과 장소에서 접근할 수 있도록 저작물 등을 이용에 제공하는 것”을 ‘전송’으로 규정하고 있으며(법 제2조 제10호), 이때 ‘공중’이란 특정 다

한편, 텍스트마이닝 과정에서 논문이나 소설 등의 원본 내용을 요약하여 독자가 원본을 접하지 않더라도 해당 어문저작물의 내용을 모두 파악할 수 있도록 제공하는 것은 시장적인 수요를 대체하여 원본의 경제적인 가치를 현저히 떨어뜨릴 수 있다. 또한 그 요약 형태에 따라 저작권법상 2차적 저작물로 성립하여 2차적 저작물작성권 침해문제가 발생할 수도 있다. 다만, 이와 같은 전송 및 복제, 2차적 저작물작성권에 대한 침해는 특별히 빅데이터의 분석·처리 과정에서 발생하는 특수한 저작권적 문제라고 분류할 수는 없고 온라인상의 저작물 이용과정에서 일반적으로 발생하는 문제이며, 이에 대한 연구 및 분쟁 사례가 다수 존재하므로 본고에서는 빅데이터 활용에 있어 발생할 수 있는 저작권적 문제 중 데이터 분산·처리 과정에서의 일시적 저장을 중심으로 논하고자 한다.

3. 빅데이터 분산 처리 과정에서의 데이터 저장

일반 기업들도 빅데이터 분석을 위하여 오픈소스 기술(RSS, Open API)을 통하여 온라인상 공개된 데이터에 직접 접근하여 이용하는 경우도 존재하는데,¹⁹⁾ 이때 일반적으로 이용자의 동의가 필요하지는 않다. 현행 저작권법은 접근권을 규정하고 있지는 아니하기 때문에 공개된 자료에 접근하거나 검색하는 것에는 문제가 없기 때문이다. 다만, 데이터를 분석하기 위해서는 우선 조직 내부에 분산된 정형데이터와 외부에 분산된 비정형데이터를 로그 수집기, 크롤링, 센싱 등의 방법으로 수집하고, 데이터 검색을 통해 수집을 마쳤다면 ‘저장’ 단계에 들어선다.²⁰⁾ 그렇다면, 개인 이용자가 해당 기업의 서버에 저장을 하였고, 검색에 공개를 동의하였다하더라도 단순히 지인들과 일상을 나누기 위한 목적으로 생

수인을 포함한 불특정 다수인을 말하므로(법 제2조 제32호), 일반 대중에게 공개되는 것이 아니라 계약 당사자인 기업 사이에 이루어지는 전송이라 할지라도 침해를 구성할 수 있을 것이다.

19) 독일 식칼제조·판매 업체인 행켈은 칼의 판매량이 계속 떨어지자 트위터의 수백만 건의 글을 분석하여 칼에서 나는 냄새가 판매량 하락의 원인임 파악하여 제품의 향을 변경한 후 판매량을 회복하였다. 황시영, “SNS 읽어라, 소비자 마음 보인다”, 매경이코노미,

〈<http://news.mk.co.kr/newsRead.php?year=2011&no=708899>〉, 검색일: 2013. 7. 31.

20) 빅데이터 저장은 대량의 데이터를 파일 형태로 저장할 수 있는 기술과 비정형 데이터를 정형화된 형태로 저장하는 기술이 중요하다. 김지선, “빅데이터 확보 첫 단추는 ‘검색과 수집’”, 디지털 타임즈, 〈http://www.dt.co.kr/contents.html?article_no=2012051602011060746002〉, 검색일: 2013. 7. 31.

성한 저작물이 대기업의 체계적인 시스템에 의하여 다시 수집되어 저장되고 해당 기업의 경제적 가치 창출을 위한 정보 분석에 쓰이는 것은 아무런 문제가 없는 것일까? 이러한 사실을 인지한 후 과연 모든 이용자들이 그 사실을 알기 전과 다름없이 SNS서비스에 저작물의 이용에 동의를 할 것인가?

특히, 기존의 데이터 처리·분석 기술은 조건에 따라 구조화된 정형 데이터만을 추출하였다면, 빅데이터 처리 기술로 각광받는 “하둡”의 분산처리 기술은 큰 용량의 데이터 자체를 64MB의 크기로 쪼개어 저장한 후 직접 해당 데이터에서 원하는 정보를 추출하기 때문에 데이터의 직접적인 저장이 발생한다. 예를 들어, 기존의 데이터처리 방식은 영화의 정보를 검색하고자 할 경우에 개봉 시기, 투자비용, 감독, 배우 등 세부 조건을 기술한 스키마를 만들고, 구조화하여 데이터베이스가 만들어지고 검색시 결과를 제공하는 형식이라면, 하둡은 비정형 데이터인 영화 파일 자체를 64MB의 크기로 나누어 데이터노드에 저장하여 읽어 들여 영화 파일 내의 정보를 직접 검색하여 결과를 제공하는 것이다. 물론 결과의 제공에 있어 저작물의 보호되는 표현(영상저작물) 그 자체를 이용하는 것이 아니라 그 안에 내재된 사실 정보 및 아이디어를 추출해 내는 것이지만,²¹⁾ 데이터 처리 과정에서의 영화 파일의 ‘저장’이라는 과정이 수반되기 때문에 이러한 ‘저장’이 저작권법상 복제에 해당하는지, 해당한다면 저작권 침해에 이르는지 검토가 필요하다.

1) 저작권법상 복제의 개념

저작권법은 저작권자에게 자신의 저작물을 복제하거나 타인이 복제할 수 있도록 허락하는 독점 배타적인 권리를 부여하고 있는데, 이때 ‘복제’는 “인쇄·사진촬영·복사·녹음·녹화 그 밖의 방법으로 일시적 또는 영구적으로 유형물에 고정하거나 다시 제작하는 것을 말하며, 건축물의 경우에는 그 건축을 위한 모형 또는 설계도서에서 따라 이를 시공하는 것을 포함한다.”고 규정하고 있다(법제2조 제22호). 즉, 복제의 방법이나 수단에는 제한이 없으므로 컴퓨터의 하드

21) 저작권법상 사실정보나 아이디어는 저작권의 보호범위에 있지 아니하므로, 이를 이용하는 것은 저작권 침해를 구성하지 아니한다. 대법원 1993. 6. 8. 선고, 93다3073 판결.

디스크나 서버 등의 전자적 기록매체에 저장하는, 이른바 디지털 복제도 저작권 법상 복제에 포함된다. 따라서 빅데이터를 수집하여 분산 처리시 발생하는 저장도 서버라는 물리적인 공간에 저장되는 것이라면 유형물로의 고정인 복제에 해당한다고 보아야 할 것이다. 하지만, 빅데이터 분석 과정시 발생하는 분산 처리는 데이터가 포함한 정보를 추출하는 목적으로 정보처리 속도향상을 위하여 작은 크기의 디지털 파일로 나누어 일시적으로 저장되어 처리되는 것이며, 저작물에 포함된 정보만을 추출하는 것이지 실질적인 저작물 이용이 이루어지지 않는다는 점에서 일반적으로 저작권자의 복제권이 미치는 범위인지 의문이 생길 수 있다.

(1) 일시적 복제

일시적 복제란 디지털 저작물을 이용하기 위해 컴퓨터상에서 작동하게 되면 자동적으로 주기억장치인 램(RAM)에 일시적으로 데이터가 저장되는 것을 말한다. 이것은 컴퓨터의 특성상 컴퓨터에서 현재 사용 중인 모든 프로그램이나 데이터는 램(RAM)을 거치게 되어 있기 때문에 발생하는 결과로 볼 수 있다. 인터넷 검색이나 홈페이지, 블로그 등에서 사진, 음악, 동영상 등을 접속하면 이용자들이 일부러 저장하지 않아도 PC에 저절로 일시 보관되며, 어떠한 자료가 인터넷상에서 전송되는 경우 OSP나 네트워크 운영자의 소프트웨어는 자동적으로 그 자료의 복제물을 램(RAM)에서 만든다.²²⁾

이렇듯 디지털 저작물이 컴퓨터상 일시적 저장과 관련하여 저작권법상 복제에 해당하는 것인가 불 것인가 하는 문제는 저작권 분야에서 오랜 시간 논의되어 왔다. 왜냐하면 그것이 저작물의 정상적인 이용과정에서 단지 부수적(incidental)으로 발생하고 순간적(transient)인 기간 동안 존속하며 전원이 공급되면 중단되어 사라져버리는 등 기술적인 측면에서 기존의 저작권법 체계상의 복제와는 차이가 있고, 이를 보호할 경우, 자칫 저작권자에게 기존의 저작권 보호체계에서는 인정되지 않는 사용권 내지 접근권을 부여하는 것과 같은 효과가 유발될 것이라는 우려가 제기되었기 때문이다.²³⁾ 이와 관련하여 우리나라는 저

22) 최상필, “일시적 저장과 공정이용”, 인터넷법률, 통권 제43호(2008), 법무부, 24면.

저작권법상 복제 규정상 일시적 복제나 영구적 복제에 대한 구별은 없었으나, 램(RAM)에의 일시적 저장 등은 전원이 차단됨과 동시에 사라져 버리는 것으로 ‘유희물예의 고정’으로 보기 어려워 복제로 인정하기 어렵다는 주장이 지배적이었다.²⁴⁾ 그러나, 2012년 한·미 자유무역협정에 따른 저작권법의 개정으로 명문의 규정으로 일시적 복제를 인정하게 되었다.²⁵⁾

해외 주요국의 일시적 복제에 대한 입법례를 살펴보면, EU는 우리나라와 같이 명문의 규정으로 일시적 복제에 대한 정의 규정을 두고 있고²⁶⁾ 미국과 일본은 복제에 대한 정의는 규정하지 않았지만 예외 규정을 통하여 일시적 복제에 대한 보호를 하고 있다.²⁷⁾ 이에 따라 미국과 일본의 경우에는 ‘일시적 저장’을

23) 임원선, “일시적 복제의 보호입법에 관한 연구”, 계간저작권, 통권 제93호(2011), 5면.

24) 이해완, 저작권법, 2007, 박영사, 207면.

25) 한·미 FTA협정문 제18.4조 제1항. ‘일시적 저장’의 취급에 대하여 “각 당사국은, 저작자·실연자 및 음반제작자가 어떠한 방식이나 형태로, 영구적 또는 일시적으로(전자적 형태의 일시적 저장을 포함한다), 그의 저작물·실연 및 음반의 모든 복제를 허락하거나 금지할 권리를 가진다.”

26) Directive 2001/29/EC of European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society, Article 2. Reproduction right: Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part:

27) The Digital Millennium Copyright Act, Article. 117. Limitations on exclusive rights: Computer programs⁵⁴

(a) Making of Additional Copy or Adaptation by Owner of Copy. – Notwithstanding the provisions of section 106, it is not an infringement for the owner of a copy of a computer program to make or authorize the making of another copy or adaptation of that computer program provided:

(1) that such a new copy or adaptation is created as an essential step in the utilization of the computer program in conjunction with a machine and that it is used in no other manner, or

(2) that such new copy or adaptation is for archival purposes only and that all archival copies are destroyed in the event that continued possession of the computer program should cease to be rightful.

(c) Machine Maintenance or Repair. – Notwithstanding the provisions of section 106, it is not an infringement for the owner or lessee of a machine to make or authorize the making of a copy of a computer program if such copy is made solely by virtue of the activation of a machine that lawfully contains an authorized copy of the computer program, for purposes only of maintenance or repair of that machine, if –

(1) such new copy is used in no other manner and is destroyed immediately after the

모두 일시적 복제로 인정하지는 않고 ‘순간적·과도적’인 것으로 볼 수 있는 일정한 유형의 경우에는 여전히 복제의 개념에 포함되지 않는 것으로 논의하고 있다.²⁸⁾²⁹⁾

우리 저작권법도 복제를 규정함에 있어 여전히 ‘유형물への 고정’을 요건으로 하고 있으므로 ① 램(RAM)에서의 일시적 저장, ② 캐싱 등에 의한 하드디스크에의 일시적 저장이 모두 포함되는 것으로 봄이 입법 경위, 목적 등에 비추어 타당하다. 다만, 개정법의 일시적 복제에 개념에 해당하기 위해서는 비록 일시적이라 하더라도 사회통념상 유형물への ‘고정’에 해당하는 것으로 볼 수 있는 최소한의 저장 시간을 요한다고 할 수 있을 것이다.³⁰⁾

위와 같은 논의에 따라 일시적 복제의 특징 세 가지를 정리하면 다음과 같다. 첫째, 휘발성. 전원 공급이 중단되면 기억장치에 저장되었던 정보가 사라진다. 둘째, 일시존속성. 복제물이 일시적·과도적으로만 존속하는 성질을 말한다. 셋째, 부수성 또는 불가피성. 이는 컴퓨터 등에서 저작물을 사용하는 과정에서 기술적으로 부수적으로 그리고 불가피하게 이루어지는 성질을 말한다. 컴퓨터 등에서 저작물을 이용하거나 네트워크를 통하여 저작물의 송신 또는 수신하는 과정에서 이루어지는 버퍼링이나 캐싱의 경우처럼 저작물의 이용이나 송신과정에서 효율성이나 안정성을 높이기 위한 목적으로 이루어지는 것이 있다.³¹⁾

(2) 빅데이터 분산·처리 과정의 일시적 복제의 발생 여부

maintenance or repair is completed; and

(2) with respect to any computer program or part thereof that is not necessary for that machine to be activated, such program or part thereof is not accessed or used other than to make such new copy by virtue of the activation of the machine.

28) 미국의 연방 항소법원은 케이블TV회사가 콘텐츠를 제공하면서 자신의 회사 서버에 0.1초 정도의 데이터를 저장하여 일시적 복제가 문제가 된 사건을 통해 저작물이 버퍼에 ‘일시적으로 짧은 시간 동안(transitory period)’에 존재하는 것이 ‘고정’되었다고 할 수 없으므로 복제물을 만들지 않았다고 판시하였다. *CartoonNetworkLP v. CSC Holdings, Inc.* 536 F.3d 121 (2d Cir. 2008)

29) 일본은 2006년 1월의 문화심의회 저작권분과회보고서에서 일시적 저장을 ‘순간적·과도적인 저장’과 ‘그 이외의 것’으로 나눈 다음, 전자의 ‘순간적·과도적’ 저장은 복제가 아니라고 하고, 후자는 일시적 고정으로 복제라고 정리하였다. 이해완, *저작권법*, 제2판, 2012, 박영사, 334면.

30) 이해완, 위의 책(주 29), 338면.

31) 임원선, 앞의 논문(주 23), 7-8면.

빅데이터 분산·처리과정에서 발생하는 할 수 있는 복제는 일반적으로 저작물을 하드디스크에 저장하는 복제는 아니다. 맵리듀스로 데이터를 처리하기 위해서는 64MB의 블록의 메모리에 읽어 들어서 처리하게 되고, 맵함수에 의해 처리된 64MB의 블록 데이터의 중간 결과값은 리듀스 함수로 처리되기 전에 로컬 디스크에 임시로 저장된다. 또한, 검색결과 추출 등의 정보처리를 위한 저장으로 그 이용이 부수적이라 할 수 있을 것이다.

따라서 상기 일시적 복제의 세 가지 특성에 비추어 빅데이터 분산·처리 과정에서의 데이터의 일시적 저장은 저작권법상 일시적 복제에 해당한다고 볼 수 있다. 다만, 빅데이터 처리 기술의 핵심은 그 처리 속도에 있는 것으로, 데이터를 64MB로 쪼개어 동시에 병렬 처리하는 과정에서 발생하는 일시적인 저장이라든가 유형물체의 ‘고정’이라고 볼 만한 저장 시간으로 인정될 것인지에 대한 의문은 남는다.³²⁾ 일반적으로 영화 한편의 디지털 파일의 크기가 700MB라고 가정할 때, 해당 영화 파일을 64MB로 쪼개어 저장한다면 약 10개 정도의 파일로 나뉘고, 이 10개의 파일을 읽어 들여 정보를 추출하는 동안 데이터가 컴퓨터나 서버에 일시 저장되는 시간은 일반적으로 ‘영화’라는 저작물을 실질적으로 이용(재생이나 복제 등)하기 위하여 일시적으로 저장되는 시간보다 현저히 짧은 시간이 될 것이기 때문이다.

따라서 만일 빅데이터 활용을 위한 데이터 분석·처리 과정에서 발생하는 분산처리 시스템상의 일시적 저장이 ‘고정’에의 요건을 충족하지 못한다면 저작권법상의 일시적 복제에 해당하지 않을 수도 있다. 하지만, 현재 빅데이터 활용을 위한 데이터마이닝 과정에 대하여 유형물체의 고정 여부에 대한 명확한 분석이나 법원의 판단 등이 존재하지 않기 때문에, 일괄적으로 데이터마이닝 과정에서의 일시적 저장이 저작권법상 일시적 복제에 해당하지 않는다고 단언하기는 어려울 것으로 생각된다. 이하의 글에서는 데이터마이닝 과정이 일시적 복제에 해당하는 상황을 전제하여 일시적 복제에 대한 면책 규정 및 공정이용 규정의

32) 미국의 연방 항소법원은 케이블TV회사가 콘텐츠를 제공하면서 자신의 회사 서버에 0.1초 정도의 데이터를 저장하여 일시적 복제가 문제가 된 사건을 통해 저작물이 버퍼에 ‘일시적으로 짧은 시간 동안(transitory period)’에 존재하는 것이 ‘고정’ 되었다고 할 수 없으므로 복제물을 만들지 않았다고 판시하였다. *CartoonNetworkLP v. CSC Holdings, Inc.*, 536 F.3d 121 (2d Cir. 2008).

적용 여부를 검토해 보고자 한다.

(3) 일시적 복제에 대한 면책규정

일시적 복제를 일반적 복제와 동일한 수준에서 저작권자의 복제권의 범위 안에 두게 되면 그 특성상 이용자들의 정당한 이용이 제한받을 소지가 많다. 예컨대 인터넷 이용시 웹브라우저를 통해 저작물의 열람 및 검색과정에서 자동적으로 발생하는 경우와 통신망을 통한 시스템 캐싱(system caching)과 같이 순간적이고 우발적으로 발생하는 경우 등에 대해서도 복제권 침해 분쟁이 발생할 수 있기 때문이다. 따라서 한미자유무역협정에서 우리 정부는 일시적 복제에 대하여 비교적 포괄적인 예외를 설정할 수 있도록 합의하였고, 이에 따라 저작물 이용과정에서의 일시적 복제가 가능하도록 “컴퓨터에서 저작물을 이용하는 경우에는 원활하고 효율적인 정보처리를 위하여 필요하다고 인정되는 범위 안에서 그 저작물을 그 컴퓨터에 일시적으로 복제할 수 있다. 다만, 그 저작물의 이용이 저작권을 침해하는 경우에는 그러하지 아니하다.”라는 예외규정을 신설하였다(저작권법 제35조의2).

① 컴퓨터에서 저작물을 이용하는 경우

컴퓨터 환경에서의 저작물 이용(주된 이용)을 위해 부수적으로 일시적 복제가 이루어지는 경우일 것을 요하는 취지로, 이때 저작물의 ‘이용’은 일시적 복제를 제외한 다른 형태의 이용을 말하며, 반드시 저작권재산권의 구체적 지분권에 해당하는 이용만 포함되는 것이 아니라 DVD시청, 정보의 검색·열람 등과 같이 지분권에 포함되지 않는 일반적인 의미의 이용도 포함되는 것으로 보아야 할 것이다.³³⁾ 즉, 컴퓨터에서의 DVD, CD의 재생은 주된 이용이며, 이때 램에의 일시적 저장이나 컴퓨터에서 인터넷 검색시 일부 내용이 램이나 캐시 파일로 하드디스크 임시 저장 폴더에 복제되는 것 등이 부수적 이용(일시적 복제)이라고 할 수 있을 것이다.

33) 池村總, 著作権法コンメンタル, 別冊—平成21年 改正解説, 2010, 勁草書房, 126면; 이해완, 앞의 책(주 29), 513면에서 재인용.

② 원활하고 효율적인 정보처리를 위하여 필요하다고 인정되는 범위

컴퓨터 등에서 저작물을 송신받아 이용하거나 또는 컴퓨터 내의 저장매체나 그밖의 저장매체에 저장된 저작물을 이용하는 경우에, 버퍼링(buffering)이나 캐싱(caching) 등을 포함하여 이를 원활하고 효율적으로 처리하는 데 필요한 범위 내의 일시적 복제가 모두 해당되며, 이에 해당하기 위하여 그러한 일시적 복제가 반드시 합법적 이용에 부수적이어야 한다거나 불가피한 것일 필요도 없다.³⁴⁾

③ 그 저작물의 이용이 저작권을 침해하지 아니하는 경우

저작물의 이용이 저작권을 침해하지 아니하는 경우는 저작권자의 허락을 받았거나, 저작권의 제한사유에 해당하여야 한다. 그런데, 이 단서에서 말하는 “그 저작물의 이용”은 본문 규정의 ‘주된 이용’을 말하는 것이고, ‘부수적 이용(일시적 복제)’을 말하는 것은 아니라고 해석된다. 왜냐하면 만약 이 단서 규정의 “그 저작물의 이용”이 본문의 일시적 복제행위를 의미하는 것이라고 해석한다면, 본문에서 그러한 일시적 복제행위를 허용한다고 규정하고 있는 것은 그 복제행위가 원칙적으로는 저작재산권자의 복제권을 침해하는 행위가 될 것이지만 본 조 본문 규정에 의하여 특별히 그 복제권을 제한한 취지라고 할 것인데, 이를 단서에서 다시 저작권의 침해라고 규정하는 것이 되어 본문과 단서가 논리적으로 모순되기 때문이다.³⁵⁾ 따라서, 단서 규정은 저작물의 ‘주된 이용’ 저작권법상의 저작재산권을 침해하는 경우에는 그에 따라 이루어지는 일시적 복제를 굳이 예외로 하여야 할 이유가 없으므로 면책하지 않겠다는 취지이다. 그렇다면 어떠한 경우에 일시적 복제가 ‘부수적 이용’에 해당하지 아니할까? 현재의 디지털 기술이나 저작물의 이용방법에 있어 실례를 찾기는 힘들 것이다. 따라서 해당 요건이 불필요하다는 의견도 존재한다.³⁶⁾

이와 관련한 해외 주요국의 일시적 저장에 의한 복제 예외 규정 입법례를 참고하여 볼 때, EU의 경우 저작권 지침(Copyright Directive) 제2조에서 일시적

34) 한국저작권위원회, 한·미 FTA 이행을 위한 개정 저작권법 설명자료, 2011, 6면.

35) 이해완, 앞의 책(주 29), 514면.

36) 민경재, “한·미 FTA 이행을 위한 개정 저작권법 고찰”, 법학논총, 제32권 제3호(2012), 93면.

저장이 저작권자의 복제권의 범위에 포함됨을 명시한 후, 제5조 제1항에서 일시적 저장에 의한 복제가 허용되는 경우와 관련하여 “잠정적 또는 부수적으로 일어나고 기술적 과정의 필수적이고 본질적인 것으로서 그 유일한 목적이 저작물 또는 그 밖의 대상물에 대하여 ① 네트워크상의 매개자에 의한 제3자들 사이의 송신 또는 ② 합법적인 이용을 하게 하는데 있는 일시적 복제 행위는 복제권으로부터 제외된다.”라고 규정하고 있다.³⁷⁾ 독일 및 영국의 저작권법 역시 공통적으로 “순간적이거나 부대적인(부수적인) 일시적 복제행위와 기술적인 절차의 필수적이며 본질적인 부분을 나타내는 일시적 복제행위가 허용되는 것은, 그 유일한 목적이 저작물 혹은 여타 보호대상물을 ① 중개자를 통한 제3자들 사이의 네트워크에서 전송하거나, ② 적법한 이용인 경우이면서 독자적으로 경제적인 의미를 갖지 않는 경우를 들고 있다.”³⁸⁾ 우리나라의 저작권법은 유일목적이나 독자적으로 경제적인 의미에 대한 명시적인 언급은 하고 있지 않지만, 일시적 복제가 ‘주된 이용’으로 성립되는 실례가 존재하지 않는 현재는 ‘부수적 이용’이 아닌 경우를 판단함에 있어, 이러한 해외의 입법례를 참고할 수 있을 것이라 생각한다.

(4) 공정이용에 의한 면책 규정

한미자유무역협정에 따른 저작권법 개정을 통하여 저작물 이용에 대한 포괄적인 면책 이 새롭게 도입되어 “저작물의 통상적인 이용 방법과 충돌하지 않고 저작자의 정당한 이익을 부당하게 해치지 않는 경우에는 보도·비평·교육·연구 등을 위하여 저작물을 이용할 수 있다.”고 규정하고 있는데(저작권법 제35조

37) Copyright Directive, Article 5 Exceptions and limitations

1. Temporary acts of reproduction referred to in Article 2, which are transient or incidental [and] an integral and essential part of a technological process and whose sole purpose is to enable:

(a) a transmission in a network between third parties by an intermediary, or

(b) a lawful use of a work or other subject-matter to be made, and which have no independent economic significance, shall be exempted from the reproduction right provided for in Article 2.

38) 최상필, 앞의 논문(주 22), 14면.

의 3), 일시적 복제도 본 조항에 따라 공정한 이용이 가능하기 때문에 상기 일시적 복제에 대한 면책요건이 결여되는 경우에는 본 조항에의 적용가능성을 다시 살펴 볼 수 있다. 저작물의 이용행위가 공정이용에 해당하는가 여부는 동조 제2항의 각 호의 사항들을 고려하여 판단한다.

① 영리성 또는 비영리성 등 이용의 목적 및 성격

공정이용을 판단하기 위한 첫 번째 요건을 판단하기 위한 구체적인 요소로는 i) 상업적 목적인지의 여부, ii) 교육적 목적인지의 여부, iii) 변형적 이용인지의 여부, iv) 사회적 이익이 있는지의 여부 등이 고려된다. 이때, 상업적 목적 자체가 공정이용 법리를 부정하는 결정적인 요인은 아니라고 볼 수 있으며,³⁹⁾ 새로운 저작물이 단순히 원저작물을 대체하는 것인지, 아니면 목적 또는 상이한 성격을 추구하는지 ‘변형적(transformative)’ 이용인지의 여부도 중요하게 판단된다.

② 저작물의 종류 및 용도

공정이용을 판단하기 위한 두 번째 요건은 저작물의 성격이다. 저작물의 성격을 판단하는 이유 또한 공정이용의 근본이념에서 찾아볼 수 있는데, 모든 유형의 저작물 중에는 순수창작물과 달리 공익을 위하여 접근성이 더 많이 인정되어야 하는 저작물이 존재한다. 이용된 저작물이 사실적이라기보다 창작적일수록 저작권으로 보호받을 가능성이 높다. 예를 들어 정부의 법안 설명 자료를 이용하는 경우와 개별 창작자의 뮤직비디오를 이용하는 경우를 생각해 볼 수 있는데, 이런 경우라면 정부의 법안 설명 자료를 이용하는 경우가 공정이용으로 판단될 가능성이 높을 것이다.⁴⁰⁾

③ 이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성

39) Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 584(1994).

40) 이규호 · 서재권, 공정이용 판단기준 도출을 위한 사례연구, 한국저작권위원회, 2009, 33면.

공정이용을 판단하기 위한 세 번째 요건은 이용되는 저작물의 양이다. 타인의 저작물을 많이 복제하면 할수록 저작권 침해에 가깝게 판단된다. 이는 단순히 공정이용을 판단하기 위한 하나의 요건으로만 작용하는 것이 아니라, 네 번째 요건인 저작물의 시장이나 가치에 미치는 영향과 결부되어 작용한다. 즉, 이용된 저작물의 양이 증가하는 만큼, 새롭게 만들어진 저작물이 원저작물을 대체할 가능성이 높아지고, 원저작물의 시장에서의 경쟁력은 저하될 수밖에 없다.⁴¹⁾

④ 저작물의 이용이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향

공정이용을 판단하기 위한 네 번째 요건은 저작물의 시장가치에 대한 대체수요가 가능한지의 여부로, 원저작물을 이용하여 창작된 새로운 저작물의 등장으로 경쟁관계가 발생하여 원저작물의 가치가 하락한다면 공정이용에 해당하지 않는다고 해석될 수 있다.

이러한 공정한 이용의 요건에 대입해 볼 때, 빅데이터 처리시 저작권으로 보호되는 데이터를 ① 상업적인 목적으로 이용한다 하더라도, ② 저작물을 그대로 이용하는 것이 아니라 정보만을 추출하거나 변형 후 이용되며, ③ 저작물의 상당한 양이 복제 된다 하더라도 그 저작물의 본래 목적의 이용(영화 관람이나, 음악 감상 등)이 수반되는 것은 아니므로, ④ 원본의 원저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치를 대체할 만한 이용이라고 판단하기는 어렵다. 따라서 상업적인 목적에 있어서는 개별 사안에 따라 판단하여야 하겠지만 일반적으로 상업적 목적을 가지지 않는 공공부문에의 활용을 위한 빅데이터 자원 데이터마이닝 과정 등에 발생하는 저작물의 일시적 복제는 공정한 이용으로 해석될 수 있을 것이라 생각한다. 또한 실질적으로 공정이용에 관한 명확하고 구체적인 개념 정립이 없는 국내의 법적 현실을 고려하고, 공정이용의 도입에도 불구하고 일시적 복제에 대한 면책 규정이 별도로 존재한다는 것 자체가 공정한 이용에의 법리에 의한 포괄 적용보다 개별 규정이 효과적일 것이라는 검토에 따른 입법의 결과로 이해되기에 일시적 복제에 관하여는 제35조의2에 대한 접근

41) 이규호 · 서재권, 앞의 책(주 40), 33면.

이 좀 더 합리적일 것으로 보인다.

(5) 소결

일시적 복제의 면책요건에 따라 빅데이터 처리·분석 과정을 대입해 볼 때, ① 컴퓨터 환경에서의 정보의 추출을 위하여 데이터의 검색·열람 등의 주된 이용시 부수적으로 발생하고, ② 많은 데이터의 처리를 위한 분산 저장이므로 원활하고 효율적인 정보처리를 위하여 필요하다고 인정되는 범위이며, ③ 데이터를 수집하여 분산처리하고 정보를 읽어 들여 결과값으로 표현하는 일련의 과정에서 저작권으로 보호되지 아니하는 정보만을 추출하는 것이므로, 그 자체로 실질적으로 저작물의 침해행위(주된 이용)가 있다고 보기도 어렵다. 따라서 일반적인 웹브라우징의 과정과 유사하며, 데이터의 분산처리로 저장되는 저작물의 일시적인 복제 행위(부수적 이용)로 볼 수 있으므로, 현행 우리나라 저작권법상의 일시적 복제에 대한 면책 규정 적용에 큰 무리가 없을 것으로 생각된다.

그렇지만 일반적인 이용자들이 웹브라우징을 통해 정보를 추출하는 속도와는 비교할 수 없는 결과물이 산출되며 이것을 수익으로 연결하기 위하여 대량의 데이터를 처리하기 위한 전문 시스템을 갖추어 놓고 상업적으로 정보를 추출해 내는 데이터마이닝 과정이 일반 이용자들의 웹브라우징과 동일한 ‘부수적 이용’에 그치는 것으로 보아 일시적 복제에 대한 면책규정 적용이 가능한 것인가에 대하여 석연찮은 부분이 있다.

특히, 빅데이터가 출현하기 이전 정보 검색의 효율화를 위하여 활용되어 왔던 데이터베이스의 경우 인적·물적 자원을 투자하여 저작권법으로 보호를 받는 대상물인데, 이러한 데이터베이스의 전부 또는 상당 부분을 단시간에 읽어 들여 정보를 제공하는 서비스를 산업화함으로써 기존 데이터베이스 제작자들이 자신들의 데이터베이스 검색 제공을 통해 올리던 광고 수익 등에 감소가 발생한다면 이것이 ‘주된 이용이 저작권 침해를 구성하는 행위’가 아니라고 볼 수만은 없지 않을까?

데이터베이스와 관련된 우리나라 법원의 판례를 보더라도 원칙적으로 전부 또는 상당부분을 복제하여야 침해가 성립되는 규정에도 불구하고 “데이터베이

스의 본질적인 부분을 특정한 목적으로 위하여 반복적으로 이용하여 데이터베이스의 경제적인 이익을 침해한 것으로 볼 수 있다.”는 판단을 내린 바 있다.⁴²⁾

해외의 경우에도 일본은 2008년 6월 정부 지적재산전략본부에서 결정된 “지식재산추진계획2008”에서 정보처리를 위한 기반적 기술이 되는 영상, 음성, 언어, 웹 분석기술 등의 연구개발에 관하여 그 연구개발의 과정에서 이루어지는 정보 이용에 대해서는 저작권법상 과제가 있다고 지적되자 정보처리 기술 기반이 되는 연구개발을 위한 저작물의 정보 분석이 가능하다는 저작권법 제47조 7의 규정⁴³⁾을 신설하면서 단서를 통하여 정보 분석을 행하는 자에게 제공하기 위하여 작성된 데이터베이스의 저작권에 대해서는 그 제한 사유가 아니라고 하여 데이터베이스 저작물 시장과 저촉하지 않는다는 조치를 취하고 있는데, 정보 분석분야의 연구개발은 저작물의 표현 그 자체를 이용하는 것이 아니라 그 정보나 아이디어를 추출하는 것에 지나지 않지만 ‘데이터베이스 저작물’ 그 자체를 제3자가 이용하는 경우에는 해당 데이터베이스 저작물의 권리자 이익을 해하는 우려가 있다는 점을 감안한 부분이라고 생각되기 때문이다.⁴⁴⁾

42) 서울고등법원 2010. 6. 9. 선고, 2009나96309 판결. 이 사건은 판매되는 물가정보지의 정보를 데이터 파일로 만든 다음 이를 연동할 수 있는 프로그램을 제작하여 유료 회원들에게 배포한 사안에 대하여 데이터베이스 저작권침해를 인정한 것으로, 법원은 “피고는 단순히 원고 물가정보지의 개별소재를 사용한 것에 불과하여 데이터베이스의 상당한 부분을 복제한 것이라고 할 수 없다고도 주장하나, 앞에서 본 바와 같이 피고가 원고의 홈페이지에서 검색 순위가 1위부터 100위 사이에 이르는 중요한 가격정보만을 추출해 피고 데이터 파일을 만든 이상, 이를 두고 단순히 원고 물가정보지의 개별 소재를 사용한 것에 불과하다고 할 수 없을 뿐만 아니라, 인정사실에 따르면, 원고의 수익은 물가정보지 판매수익과 물가정보지에 게재되는 광고수익으로 구성되어 있고, 기업이 공사내역서에 물가정보의 정확한 근거를 밝히기 위해서는 물가정보가 수록된 물가정보지의 발행 호수와 쪽수를 기입해야 하기 때문에 그 기입을 위해서는 해당 물가정보지를 구입할 수밖에 없는데, 이러한 점은 저작권법 제93조 제2항 단서에서 정한 바와 같이, 데이터베이스의 개별 소재 또는 그 상당한 부분에 이르지 못하는 부분의 복제 등이라 하더라도 반복적이거나 특정한 목적을 위하여 체계적으로 함으로써 당해 데이터베이스의 통상적인 이용과 충돌하거나 데이터베이스제작자의 이익을 부당하게 해치는 경우에 해당한다고 볼 수 있으므로 이러한 관점에서 보더라도 피고는 원고 물가정보지의 상당한 부분을 복제한 것으로 볼 수밖에 없다.”라고 판시하였다.

43) 일본 저작권법 제47조의 7(정보해석을 위한 복제 등) 저작물은, 전자계산기에 의한 정보해석(다수의 저작물 기타의 대량의 정보로부터, 당해 정보를 구성하는 언어, 음, 영상 기타의 요소와 관련된 정보를 추출, 비교, 분류 기타의 통계적인 해석을 행하는 것을 말한다. 이하 이 조에서 같다)을 하는 것을 목적으로 하는 경우에는, 필요하다고 인정되는 한도에서 기록매체에의 기록 또는 번안(이에 의하여 창작한 2차 저작물의 기록을 포함한다)을 할 수 있다. 다만 정보해석을 하는 자의 이용에 제공하기 위해 작성된 데이터베이스저작물에 대하여는 그러하지 아니하다.

44) 作花文雄, 詳解著作権法, ぎょうせい, 2010, 401면.

또한 벨기에 법원은 세계적인 웹 검색엔진 회사인 구글의 뉴스 검색 서비스에 대하여 뉴스 본문의 캐시메모리에의 저장이 복제권을 침해한다고 인정하는 것과는 별도로 인터넷에서 검색된 기사의 체계적 복제와 구글 캐시 저장은 언론사 데이터베이스의 상당 부분을 추출하는 것으로 볼 수 있으므로 캐시 기능 링크를 제공하는 것은 데이터베이스의 해당 부분 재이용이며, 데이터베이스권을 침해하였다고 판단하였다.⁴⁵⁾

한편, 영국에서도 출판협회가 연구목적이라도 데이터마이닝을 저작권 제한 사유로 규정한다면 경제적 손실을 일으킬 수 있다고 경고하고 있는데, 이에 대한 근거 중 하나는 현재 계약을 통해 이미 데이터마이닝을 하고 있는 경우를 들고 있다.⁴⁶⁾ 즉, 데이터마이닝을 위한 정보의 접근은 저작권법상의 권리는 아니더라도 이미 계약법상 상업적인 가치를 가지고 있다고 유추해볼 수 있는 것이다.

그러므로 빅데이터의 활용을 위한 정보의 접근 및 처리 방식에 있어서의 일시적 복제에 대한 면책규정의 적용여부에 대하여 데이터베이스의 경우에는 일반저작물과 구분하여야 할 실익에 대한 좀 더 면밀한 검토가 필요할 것으로 생각한다.

IV. 결론

세계 각국은 공공분야의 빅데이터 활용에 활발한 움직임을 보이고 있다. 미국의 빅데이터 추진전략(Big Data Initiative), 영국의 The Foresight Horizon Scanning Centre, EU의 FuturICT와 iKnow 프로젝트 등이 그것이다.⁴⁷⁾ 그러나 빅데이터 활용에 있어서 현재 저작권적 이슈가 부각된 경우는 미비하지만 개념적으로 사람이 아닌 검색기술에 의해서 수많은 데이터를 수집해야하는 빅데이

45) Google v. Copiepresse, No. 06/10.928/C(Tribunal de premiere instance de Bruxelles); 차준영·공병훈, "검색엔진의 캐싱(caching)과 뉴스 저작권에 관한 연구", 계간 저작권, 통권 제97호(2012), 211면에서 재인용.

46) 한국저작권위원회, Copyright Issue Report, 2011-13호, 2011, 2면.

47) 한국정보화진흥원 빅데이터 연구센터, "선진국의 데이터기반 국가미래전략 추진현황과 시사점", IT & Future Strategy, 제2호(2012), 4면.

터의 특성상 관리자가 공개(개방)는 하였으나 공유를 허락하지 않은 데이터들도 무차별적으로 수집하여 이용함으로써 저작권법적인 분쟁이 발생할 가능성을 염려하는 시선도 있다.⁴⁸⁾ 그렇지만 저작권법상의 법률적 리스크로 인하여 공공부문 등 산업상 빅데이터 활용이 저작권법으로 인하여 둔화된다면 이 역시 사회적으로 큰 손실이 아닐 수 없다. 수많은 데이터의 일부를 구성하고 있는 저작물로 인하여 빅데이터 활용으로 무한하게 흐를 수 있는 정보의 흐름이 잠식되어 버릴 수 있는 것이다.

이러한 관점에서 빅데이터의 정보 처리 분석 과정에 따른 저작권적인 문제 발생 여부를 검토한 결과 빅데이터 활용을 위한 데이터마이닝 과정에서 발생하는 일시적 복제 등의 저작권적 문제 발생가능성을 확인하였다. 하지만 현재는 일시적 복제의 개념이 일관되게 정립되지 않았고 관련 판례가 많이 축적되지 않았기에 선불리 침해를 판단하기 어려운 점이 있었다. 그러나 빠르게 변화하는 IT 기술 환경하에서 적용사례가 누적될 틈 없는 잦은 법규의 제·개정은 오히려 안정적인 산업발전을 위한 법규 환경 조성에 누가 될 수도 있을 것이라는 개인적인 의견과, 상기의 일시적 복제의 면책 규정 및 공정이용 적용 요건을 분석을 근거로 데이터마이닝 과정에서 발생하는 저작권 문제는 현행 저작권법적인 체계 안에서의 규율이 가능할 것이라는 결론에 이르렀다.

우선, 빅데이터의 데이터 마이닝 과정에서 저작물의 일시적 저장은 추후 일시적 복제에 대한 판례와 연구가 축적되어 개념이 정립되는 과정에서 일시적 복제로 성립되지 아니할 가능성도 부정할 수 없다. 또, 일시적 복제에 해당한다 하더라도 이는 저작물로 보호되는 데이터의 실질적인 이용이 아닌 정보처리에 주된 목적을 가진 기술구현에 수반되는 부수적인 복제일 뿐, 그 자체로 저작권자의 권리범위가 미쳐야 하는 독립적인 이용행위로 보기 어려워 현행 일시적 복제의 면책 규정에서의 적용이 가능하다. 물론 대용량 데이터의 처리를 위한 경제적인 투자와 산업적인 규모가 존재하고, 이익 창출이 가능하지만 이는 저작권법에서 저작재산권을 규정하여 목적하고 있는 일반적인 저작물의 이용으로 인한 경제적 이익으로 단정 짓기는 어렵기 때문이다. 그리고 추후 일시적 복제가 주된

48) 김종원, “빅 데이터와 저작권의 상생”, 디지털타임스,

(http://www.dt.co.kr/contents.html?article_no=2012092602010931699002) 검색일: 2012. 7. 30.

이용행위가 되는 서비스가 개발되거나 부수적인 이용자체가 ‘독자적이고 경제적인’ 이용행위로 인정되어 일시적 복제에 대한 면책규정의 적용이 어렵다고 가정하여도 복제권 침해 면책에 관한 일반규정인 포괄적 공정이용의 법리 안에 다시 한 번 함의될 수 있기 때문이다.

아울러 저작권자의 이용허락에 의하여 합법적인 이용이 가능한 저작권법 제46조 1항에 의하여도 이용자들의 생산한 데이터 이용에 대한 동의로도 합법적인 이용이 가능하다. 앞에 예를 들었던 SNS서비스 기업처럼 회원가입시 데이터의 이용에 대한 약관⁴⁹⁾을 고지하고 있기 때문에, 합법적인 동의 절차를 거쳐 이용자들이 생산한 데이터의 이용이나 이관이 가능할 수 있다.⁵⁰⁾ 또는 서비스 업체들의 노력으로 이용자들이 하여금 자신의 정보를 이용함을 동의하거나 동의하지 않도록(opt-in, opt-out) 선택할 수 있는 기술적인 환경설정도 가능할 것이다.

마지막으로, 저작권으로 보호되는 데이터베이스에 대한 접근에 대한 경제적 가치 및 그 이용에 관하여는 좀 더 검토가 필요할 것으로 제안하였지만, 이 역시 일시적 복제의 면책 요건과는 별개로 접근통제의 기술적 보호조치를 규정하고 있는 현행 저작권법하에서 데이터베이스제작에 대한 실질적인 보호방안이 존재하지 않는 것은 아니다.⁵¹⁾ 따라서, 좀 더 깊이 있는 검토는 필요하지만 데이터베이스에 대한 데이터마이닝을 제재할 제도적인 장치가 시급한 것은 아니라고 판

49) 물론 온라인상 수많은 저작물의 저작자의 이용동의를 일일이 받는 것은 쉬운일은 아니지만, 대표적인 SNS서비스인 트위터의 경우 약관을 통해(5. 귀하의 권리) 약관은 이용자들의 콘텐츠를 트위터사가 다른 회사, 단체, 개인들에게 제공할 수 있도록 정하고 있고, 페이스북은 이용자의 클릭 스트림, 교류활동, 사진이나 동영상의 시간, 날짜, 장소 등의 정보, 휴대기기 정보, GPS 등의 센서 데이터를 수집할 수 있음을 고지하고 있다. <<https://twitter.com/tos>> 검색일: 2013. 7. 30.

50) 다만, 우리나라 약관규제법은 제6조 제1항에서 “신의 성실에 원칙을 위반하여 공정성을 잃은 약관 조항은 무효이다.”라고 규정하고 있으며, 동조 제2항에서는 “고객에게 부당하게 불리한 조항, 고객이 계약의 거래형태 등 관련된 모든 사정에 비추어 예상하기 어려운 조항, 계약의 목적을 달성할 수 없을 정도로 계약에 따르는 본질적 권리를 제한하는 조항 등은 공정성을 잃은 것으로 추정된다.”고 규정하고 있으므로, 해당 약관의 공정성여부에 대하여는 별도의 검토가 필요할 것이다.

51) 저작자는 기술적 보호장치를 우회하도록 개발된 장치로부터 보호받을 자격이 있다. 합법적으로 접근할 수 있는 사람이라도 기술적 보호조치의 우회방지로 말미암아 데이터베이스의 배열이나 선택의 상당부분을 취하지 않으면서 콘텐츠를 취하지 못하도록 금할 수 있다. 데이터베이스에 도입된 기술적 보호수단을 우회하지 않으면 사실상 데이터베이스에 접속할 수 없기 때문이다. 구대환, “현행 저작권법과 유렵지침에 의한 데이터베이스 보호의 문제점”, 지식재산연구, 제1권 제1호(2006). 132면.

단된다. 다만, 현재 진행형인 IT기술에 발맞추어 저작권법상 적절한 규제가 필요한 상황을 배제할 수는 없으므로 현행 저작권법 체계하에서의 해결 방안을 모색하는 한편 새로운 법리의 필요성도 배제하지는 않는 지속적인 연구와 점검이 필요할 것이다.

참고문헌

〈국내 단행본〉

이규호 · 서재권, 공정이용 판단기준 도출을 위한 사례연구, 한국저작권위원회, 2009.

이해완, 저작권법, 박영사, 2007.

_____, 저작권법, 제2판, 박영사, 2012.

함유근 · 채승병, 빅데이터, 경영을 바꾼다, 삼성경제연구소, 2012.

한국저작권위원회, 한·미 FTA 이행을 위한 개정 저작권법 설명자료, 2011.

_____, Copyright Issue report, 2011-13호, 2011.

IDC, 빅데이터 분석: CIO를 위한 미래지형적 아키텍처, 기술 그리고 로드맵, IDC analyze the future, 2011.

〈해외 단행본〉

作花文雄, 詳解著作権法, 第4版, ぎょうせい, 2010.

〈국내 학술지〉

구대환, “현행 저작권법과 유럽지침에 의한 데이터베이스 보호의 문제점”, 지식재산연구, 제1권 제1호(2006).

이진태, “빅데이터 활성화와 저작권문제: 하둡(Hadoop)을 중심으로”, 계간저작권, 통권 제102권(2013).

이진형, “데이터 빅뱅, 빅데이터의 동향”, 방송통신전파채널, 통권 제47호(2012).

이현종, “빅데이터 하둡 플랫폼의 활용”, 한국통신학회논문지, 제29권 제1호(2012).

임원선, “일시적 복제의 보호입법에 관한 연구”, 계간저작권, 통권 제93호(2011).

정교일 외 4인, “빅데이터와 정보보안”, 한국정보기술학회지, 제10권 제3호(2012).

정지선, “빅데이터 시대의 데이터 자원 확보와 품질 관리 방안”, IT & Future Strategy, 제3호(2012), 한국정보화진흥원,

_____, “신가치창출 엔진, 빅데이터의 새로운 가능성과 대응 전략”, IT & Future Strategy, 제18호(2011), 한국정보화진흥원.

차준영 · 공병훈, “검색엔진의 캐싱(caching)과 뉴스 저작권에 관한 연구”, 계간 저작권, 통권 97권(2012).

- 최상필, “일시적 저장과 공정이용”, 인터넷법률, 통권 제43호(2008), 법무부.
- 최윤정·박승수, “웹 콘텐츠의 분류를 위한 텍스트마이닝과 데이터마이닝의 통합 방법 연구”, 인지과학, 제13권 제3호(2002).
- 한국정보화진흥원 빅데이터 연구센터, “선진국의 데이터기반 국가미래전략 추진현황과 시사점”, IT & Future Strategy, 제2호(2012).

〈인터넷 자료〉

- 김지선, “빅데이터 확보 첫 단추는 ‘검색과 수집’”, 디지털타임즈,
 〈http://www.dt.co.kr/contents.html?article_no=2012051602011060746002〉, 검색일: 2013. 7. 31.
- 김현주, “[개인정보보호 탓에 걸린 빅데이터] ‘빅데이터와 프라이버시’ 두 마리 토끼 잡으려면 ... 개인정보 개념부터 명확하게”, 매경이코노미,
 〈<http://news.mk.co.kr/newsRead.php?year=2013&no=472942>〉,
 검색일: 2013. 7. 30.
- 조성훈, “2013년 IT메가트랜드? 빅데이터와 클라우드 주목”, 머니투데이뉴스,
 〈<http://www.mt.co.kr/view/mtview.php?type=1&no=2012092510044087904&outlink=1>〉, 검색일: 2013. 7. 30.
- 이지훈·김도균, “빅데이터 환경을 위한 하둡”, 마이크로소프트웨어,
 〈http://www.imaso.co.kr/?doc=bbs/gnuboard.php&bo_table=article&wr_id=40766〉, 검색일: 2012. 7. 31.
- 황시영, “SNS 읽어라, 소비자 마음 보인다”, 매경이코노미,
 〈<http://news.mk.co.kr/newsRead.php?year=2011&no=708899>〉, 검색일: 2013. 7. 31..
- 황태호, “삼성·SK텔레콤·KT, 빅데이터 활용 프로젝트 가동”, 전자신문,
 〈<http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=004&oid=030&aid=0002184515>〉, 검색일: 2013. 7. 30.

〈국내 판례〉

- 대법원 1993. 6. 8. 선고, 93다3073 판결.
- 서울고등법원 2010. 6. 9. 선고, 2009나96309 판결.

〈해외 판례〉

Cartoon Network LP v. CSC Holdings, Inc. 536 F.3d 121 (2d Cir. 2008).

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 584(1994).

Big Data, Data Mining and Temporary Reproduction

Jl HO, LEE

Abstract

Spread of smart applications with IT development makes us create various digital contents and share them through internet. We may call these contents as data. Today, Big Data has become a hot keyword all over the world in that Big data may be defined as collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. Big Data is useful basic resources for analysing phenomenon to draw a certain output. However, There are some legal issues we need to discuss, especially in copyright sector. In a course of data mining through hadoop which is functioning distribute data processing, temporary reproduction is inevitably accompanied. And the issue is the subject of temporary reproduction here is works(including database) protected by copyright act. And the issue here is the subjects of temporary reproduction may include copyrighted works including database. This paper review the legal issues focusing on the temporary reproduction occurred in a process of datamining under Korea Copyright Act subparagraph 22 of Article 2 and paragraph 2 of Article 35. And it also examines other temporary reproduction issues related to Big Data such as fair use doctrine.

Keywords

Bigdata, Dataminig, Temporary Reproduction, Exceptions and limitations, Bigdata and Copyright

