

# 텍스트마이닝을 활용한 “잊힐 권리”의 토픽 분석\*

## Topic Analysis of the “Right to be Forgotten” Using Text Mining

이소현 (So-Hyun Lee)\*\*

구본진 (Bon-Jin Koo)\*\*\*

### 초 록

본 연구는 잊힐 권리와 관련한 뉴스 기사와 학술지 게재 논문을 대상으로 텍스트마이닝 분석을 활용하여 각 문서 내에 나타난 논점과 특성을 살펴보았다. 분석을 위해 ‘잊힐 권리’와 ‘잊혀질 권리’ 키워드를 검색어로 하여 2010년부터 2020년까지의 데이터를 수집하였다. 수집된 데이터를 대상으로 키워드 분석과 토픽모델링 분석을 수행한 결과, 지난 10년간 뉴스 기사와 학술지 논문에서 다루어진 쟁점은 크게 다르지 않으며, 접근 방법 또한 유사한 것으로 나타났다. 다만 뉴스 기사와 학술지 논문 간 비교를 통해 이들 간 공통적으로 나타나는 쟁점과 부분적인 쟁점의 차이가 있음을 확인하였다. 따라서 본 연구에서 도출된 쟁점을 중심으로 기록관리학 분야에서도 적극적인 논의가 이루어져야 할 필요가 있으며, 공통적인 쟁점들을 우선적으로 고려하되, 쟁점 상 이견이 존재하는 경우, 이를 다각적으로 논의하는 것이 필요하다고 볼 수 있다. 본 연구는 국내 기록관리학계에서 잊힐 권리와 관련된 논의가 이루어지고 있지 않은 현재의 상황에서 기록관리학 분야에서 잊힐 권리의 의미와 향후 발생할 수 있는 이슈를 도출해볼 수 있었다는데 의의가 있으며, 본 연구의 결과를 중심으로 기록관리학 분야에서 잊힐 권리에 대한 다양한 논의가 이루어지기를 기대한다.

### ABSTRACT

This study examined the issues and characteristics that appeared in news and journal articles related to the ‘right to be forgotten’ using text mining analysis. Data for analysis were collected from 2010 to 2020 with the keyword ‘right to be forgotten’. Keyword analysis and topic modeling analysis were performed on the collected data. As a result, in the last 10 years the issues about ‘right to be forgotten’ are not much different in news and journal articles and the approaches also are similar. However, it confirmed common issues and the partial difference between news and journal articles through comparison. Therefore in Archives and Records Management Studies, it is necessary to discuss derived in this study. In particular common issues are considered first but if there are differences in issues, it is needed to discuss them in various ways. This study is meaningful to understand the meaning and to draw issues that may arise in the future of the ‘right to be forgotten’. The results of this study will contribute to be variously discussed on the ‘right to be forgotten’ in Archives and Records Management Studies.

키워드: 잊힐 권리, 기록정보서비스, 텍스트마이닝, 토픽모델링, LDA

right to be forgotten, archival information services, text mining, topic modeling, LDA

\* 이 논문은 석사학위논문 수정·요약본임.

\*\* 부산대학교 도서관 사서(shlee8028@pusan.ac.kr) (제1저자)

\*\*\* 부산대학교 문헌정보학과 강사(bjkoo@pusan.ac.kr) (교신저자)

■ 논문접수일자: 2022년 5월 15일 ■ 최초심사일자: 2022년 5월 26일 ■ 게재확정일자: 2022년 6월 14일

■ 정보관리학회지, 39(2), 275-298, 2022. <http://dx.doi.org/10.3743/KOSIM.2022.39.2.275>

※ Copyright © 2022 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

## 1. 서론

### 1.1 연구의 배경과 목적

정보 기술의 급속한 발달로 인해 현대 사회에서 우리는 원하는 정보를 쉽고, 빠르고, 다양하게 접할 수 있는 기회를 가지게 되었다. 특히, 인터넷이 보편화되면서 현재의 일상적 행위 대부분이 온라인 상에서 이루어지고 있으며, 소셜네트워크서비스(SNS), 모바일 메신저 어플리케이션, 각종 커뮤니티 등 다양한 매체를 통해 문자, 음성, 영상 등 다양한 방식으로 자신을 표출하고 타인과 이를 공유할 수 있게 되었다. 이에 따라, 인터넷상에는 인터넷의 발달 이전에 축적된 지식과 정보보다 더 많은 양의 정보가 짧은 시간 동안 쌓이고 있을 뿐 아니라 자신의 생활의 모든 것이 디지털로 기록됨으로써 정보 자체의 중요성과는 무관한 다양한 범주의 정보가 서버에 축적되고 있다. 이처럼 인터넷에 올라간 개인 정보는 인터넷 검색엔진을 통해 자동으로 수집되고 색인되며, 얼마든지 검색을 통해 해당 정보를 획득할 수 있을 뿐만 아니라 빅데이터, 인공지능 등 고도로 발달된 기술로 인해 개인의 식별 가능성이 더욱 확대되고 있다.

인터넷 발달 초기에는 '알 권리'가 인간의 기본권으로 취급되고, 중요한 개념으로 인식되었으나, 한번 인터넷에 개방된 정보는 쉽게 삭제되지 않고, 심지어는 정보 주체의 사후에도 남아있게 되면서 개인의 프라이버시와 인격권 보호를

위해 '잊힐 권리'를 보장해야 한다는 주장이 제기되었다. 특히, 2010년에 스페인 변호사 마리오 코스테하 곤잘레스(Mario Costeja Gonzalez)가 1998년에 '라 방가르디아(La Vanguardia)'에 게재된 기사와 구글 검색 링크가 자신의 프라이버시 권리를 침해했다며 소송을 제기하였고, 스페인 법원이 이에 대한 법률적인 판단을 유럽사법재판소에 의뢰하면서 잊힐 권리가 글로벌 이슈로 급부상하게 되었다.

우리나라의 경우, 국민의 알권리를 보장하고 국정의 투명성을 확보하기 위해 정보공개제도를 마련하고 있으며, 국가기록원 기록정보센터의 열람서비스의 경우, '정보공개법령'에 의거하여 처리되고 있는 만큼 기록정보서비스도 포괄적인 의미의 정보공개라고 이해할 수 있다(남윤아, 임진희, 2016). 정보공개제도의 도입이 알권리의 구현을 목적으로 하고 있는 만큼 기록정보서비스 역시 국민의 알권리를 보장하기 위한 하나의 방편이라고 해석할 수 있는 것이다.

그러나 2010년 곤잘레스 사건 이후 알권리와 대척점에서 있다고 볼 수 있는 잊힐 권리가 부각되고, 나아가 정보화 사회의 발달로 개인 정보의 무방비 노출과 이로 인해 상당히 많은 피해들이 발생하면서 잊힐 권리의 성격이 개인 정보 자기 결정권, 즉, 자기 정보에 대한 관리 및 통제에 대한 권리로 확장되었고, 잊힐권리는 알권리보다 더 보호받아야 할 권리로 인식되고 있다. 그리고 이러한 인식의 변화는 알권리를 전제로 하고 있는 기록정보서비스에도 영

1) '잊힐 권리'는 '잊혀질 권리'로 표기되는 경우도 있는데, '잊히다'는 피동사이기 때문에 또다시 피동의 뜻을 나타내는 '-어지-'를 붙이는 것은 어법에 맞지 않다. 이에 따라, 본 연구에서는 연구를 진행하기 위해 참고한 문헌에서 '잊혀질 권리'로 표기된 것 외에는 모두 '잊힐 권리'로 통일하여 사용하였다.

향을 미칠 수 밖에 없으며, 따라서 기록관리 측면에서도 잊힐권리에 대한 연구와 논의가 반드시 필요하다.

잊힐 권리가 글로벌 이슈로 부상한 지 10여 년이 흘렀지만 국내에서는 법학, 언론학, 정보학 등 일부 학문 분야에서 논의가 이루어졌을 뿐 아직 기록관리학 분야에서는 잊힐 권리에 대한 논의가 이루어지지 않고 있다. 그러나 잊힐 권리를 ‘인터넷 상에서 자신과 관련된 정보의 삭제를 요구할 수 있는 권리’라는 개념을 포함하는 것을 고려한다면 결국 잊힐 권리는 기록의 유지와 삭제에 관한 것이기 때문에 기록관리의 측면에서도 잊힐 권리에 대한 논의가 필요하다고 볼 수 있다. 따라서 본 연구에서는 기록관리학 분야에서 잊힐 권리에 대한 논의가 시작될 경우, 다루어야 할 필요가 있는 논점들에 대해 파악하고자 하였으며, 나아가 기록정보서비스에 있어서 잊힐 권리가 가지는 의미를 고찰해보고자 하였다. 이를 위해 본 연구는 뉴스 기사와 학술지 논문을 대상으로 토픽모델링 분석을 통해 사회적 및 학문적으로 잊힐 권리와 관련하여 다루어졌던 주요 주제를 체계적이고 객관적으로 도출하고, 향후 기록관리학 관점에서 논의가 필요한 주요 주제들은 무엇인지 파악하고자 하였다.

## 1.2 선행연구

현재까지 기록관리학 분야에서 진행된 잊힐 권리에 대한 연구는 없으며, 법학, 언론학, 정보보안 분야 등에서 주로 연구가 이루어졌다.

먼저 최중선(2016)은 인터넷 자기게시물 접근배제요청권 가이드라인에 대한 분석을 통해

잊힐 권리가 인정되면 발생할 수 있는 법률적 논점으로 불명확한 도입 근거, 권리 객체의 협소성, 디지털 정보의 상속 여부, 잊힐 권리 상대방의 문제, 손해배상책임의 문제, 정보통신서비스 제공자에 의한 사적 검열 및 표현의 자유 침해, 공인의 정의 및 범위 등의 모호성, 공익과 상당한 관련성이 있는 경우 판단 기준, 개인정보 보호법과의 충돌 등 가이드라인의 법적 문제점을 제시하고, 잊힐 권리와 관련된 입법에 있어 고려해야 할 점을 제시하고 있다. 또한, 박용숙(2016)은 잊힐 권리와 관련하여 우리나라 역시 글로벌 수준에 적합한 법제의 정비가 필요하지만 충분한 연구와 논의를 거치지 못하고 있다는 점을 문제로 지적하며, 잊힐 권리가 표현의 자유, 알 권리 등의 기본권과의 조화를 이룰 수 있도록 심도있는 검토와 연구의 필요성을 주장하였다.

이영진(2013) 역시 잊힐 권리의 입법화 필요성에는 공감하지만 충분한 사회적 합의 없는 도입으로 인해 발생할 수 있는 알 권리 및 정보접근권과의 충돌, 공익적 기록의 삭제 가능성 등의 문제를 제시함으로써, 잊힐 권리의 법제화와 관련하여 해결해야 할 쟁점들을 제시하였고, 고은별, 최마리아, 최광희(2012)는 정보주체가 온라인 상에 존재하는 자신의 정보를 삭제 및 확산 방지할 수 있는 권리의 필요성이 대두되고 있지만 현행 국내법은 SNS를 통해 축적된 여러 행태 정보에 대한 보호가 이루어지고 있지 않아 법제화 시 다양한 충돌이 예상되므로 이를 예측하고, 이용자, 사용자 및 국가 간의 균등한 조화를 이룰 방법을 고려해보아야 한다는 점을 주장하였다.

한편 잊힐 권리의 입법화보다는 현행법을 활

용하지는 의견도 있는데, 황창근(2016)은 잊힐 권리가 인터넷의 발달과 그로 인한 개인의 불편함에서 기인한 것으로 잊힐 권리에 대한 도입 필요성은 인정하지만, 추가적 입법의 필요성이 인정되지는 않는다고 주장하며, 잊힐 권리에 대한 논의가 시작된 EU보다 개인의 정보 관련 권리를 보호하는 제도가 발달해 있어 오히려 기본 법제와의 사이에서 혼선이 발생하지 않도록 신중한 접근이 필요하다고 지적하였다. 홍숙영(2014) 또한 잊힐 권리 보장을 위해서는 새롭게 입법을 하는 것보다는 현행법인 언론중재법을 수정하여 활용하는 것이 적절하다고 주장하며, 『개인정보보호법』 제36조 ‘정정·삭제요구권’을 비롯하여, 정보제공 동의 기간 등에 대해 제한을 두는 등의 규정이 필요함을 언급하고 있다.

최경진(2012)은 개인정보보호적 관점에서 보았을 때, 잊힐 권리를 보호해야하는 것은 분명하지만, 경제적 가치를 무시할 수 없기 때문에 현행 개인정보보호법에서 나타나는 잊힐 권리와 관련된 규정을 검토하여 개인정보의 보호와 안전한 활용의 조화라는 측면을 개선할 필요가 있다고 주장하였다. 또한 정재호(2014)는 잊힐 권리가 헌법상 인격권 등에 근거하여 새롭게 형성되는 권리로 언론의 자유를 위축시키면서까지 우월하게 인정될 권리는 아니나, 이후 성문화되어 언론의 취재 및 보도 활동에 위축 효과를 초래할 수 있으므로 언론계에서는 이에 대비하여 가이드라인을 제정할 필요가 있음을 주장하였다.

이상의 선행연구들을 종합해 보았을 때, 법학과 언론학을 중심으로 잊힐 권리가 논의되고 있으나 대부분의 연구가 잊힐 권리 입법화에

대한 찬반 논쟁과 입법 시 발생할 수 있는 문제점들에 대해 다루고 있으며, 특히, 모두 잊힐 권리에 대한 인정과 도입, 실현을 위해서 해결해야 할 쟁점들이 많이 있으며, 이들 쟁점에 대한 충분한 논의와 사회적 합의가 필요하다는 점을 공통적으로 주장하고 있다.

이처럼 우리나라의 경우, 일부 분야에서 잊힐 권리에 대한 논의가 이루어지고 있을 뿐 기록관리학 분야에서의 연구는 찾아볼 수 없는 실정이다. 그러나 선행연구들에서 거론된 잊힐 권리의 도입 시 발생할 수 있는 주요 쟁점들은 공공 기록에 대한 접근 제한, 알 권리와 의 형평성, 사전 검열 가능성 등의 문제로 인하여 기록관리 영역에서도 반드시 논의될 필요가 있는 주제들이다. 따라서 현재까지 각 분야에서의 잊힐 권리 관련 연구들에서는 어떤 쟁점들이 논의되었는지, 혹은 이외에 또 다른 쟁점들이 존재하는지 등을 확인한다면 향후 기록관리 영역에서 다양한 쟁점들을 종합적으로 고려하여 더 나은 논의를 위한 초석이 될 수 있을 것이다.

## 2. 이론적 배경

### 2.1 유럽에서의 잊힐 권리의 전개

잊힐 권리가 세계적인 이슈로 떠오른 것은 지난 2010년에 스페인 변호사인 마리오 코스테하 곤잘레스(Mario Costeja Gonzalez)가 구글을 상대로 소송을 제기하면서이다. 곤잘레스는 1998년 스페인의 ‘라 방가르디아(La Vanguardia)’ 신문에 자신의 사회보장 분담금 부채 내역과

재산 강제 매각 공고 기사가 실린 이후 2009년 까지 구글 검색엔진에 자신의 이름을 입력하면 해당 기사가 링크되어 있는 것을 발견하였다. 이에 곤잘레스는 해당 문제가 이미 해결되었음에도 불구하고 자신의 이름을 검색할 때 해당 기사가 검색되는 것에 대해, 스페인의 개인정보보호원(Agencia Española de Protección de Datos, AEPD)에 ‘라 방가르디아’ 신문의 해당 원문 삭제와 구글에서 자신과 관련 기사의 신문사 링크 삭제를 요청했다. AEPD는 ‘라 방가르디아’ 신문사에 대한 구제 신청은 기각하였으나, 구글에 대해서는 명예훼손의 가능성이 있다는 이유로 구글 검색 결과에 해당 기사의 링크를 삭제할 것을 명령하였다. 하지만 구글은 AEPD의 결정에 불복, AEPD와 곤잘레스를 스페인 고등법원에 제소하였고, 스페인 고등법원은 해당 결정의 법적 근거가 된 「1995년 유럽연합 개인정보보호지침」의 규정에 대한 판단을 유럽사법재판소에 의뢰하면서 잊힐 권리가 세계적인 이슈로 부상하게 되었다(김갑석, 2017; 박용숙, 2016; 박진우, 2014; 백수원, 2015).

유럽사법재판소는 “구글 검색 결과에 링크된 해당 웹사이트의 정보가 합법적인 경우에도 링크를 삭제할 의무가 있다.”며 곤잘레스의 손을 들어줬다. 이른바 잊힐 권리를 인정한 것이다. 유럽사법재판소의 판결 이후 구글에는 검색 결과의 삭제 요청이 이어졌으며, 독일에서는 흉악범죄자에게도 잊힐 권리를 보장해야한다는 판결을 내리기도 했다(장재은, 2019). 이후 2016년 5월 유럽집행위원회(European Commission,

EC)는 기존 EU 개인정보보호 지침을 대체하는 유럽연합 개인정보보호 일반규정(General Data Protection Regulation, GDPR)을 제정하였고, GDPR의 제 17조에 삭제권(잊힐 권리)을 성문화(成文化)하였다.

## 2.2 국내에서의 잊힐 권리의 논의

우리나라에서 잊힐 권리에 대한 논의가 본격적으로 이루어진 시점 역시 곤잘레스 판결이 국내에 알려진 이후인 2010년 즈음이다. 당시 전 세계적으로 잊힐 권리에 대한 관심이 급증하고 국내에서도 이에 대한 관심이 높아지자, 2011년에 방송통신위원회는 ‘국가 정보화전략위원회’에서 소셜네트워크서비스(SNS) 이용자가 본인의 게시물이나 콘텐츠에 대해 원할 경우 파기 또는 삭제할 수 있는 잊힐 권리의 도입을 검토한다는 내용을 포함한 「소셜플랫폼 기반의 소통·창의·신뢰 네트워크 사회 구현 전략」을 발표하였으나 이 발표에 부응하는 후속조치는 이루어지지 않았다.

그러나 이후 인터넷의 발달 및 스마트폰의 보편화와 함께 인터넷상에 저장, 유통되는 의뢰인의 개인정보를 수집하여 영구적인 파기를 대리해주는 신종 직업인 디지털 장의사가 등장하는 등 잊힐 권리에 대한 요구가 높아지기 시작했고, 2016년 법률 제17358호 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 제44조의2(정보의 삭제요청 등) 제1항<sup>2)</sup>이 추가되며 한국판 잊힐 권리 시행에 박차를 가하게 된다.

2) 정보통신망을 통하여 일반에게 공개를 목적으로 제공된 정보로 사생활 침해나 명예훼손 등 타인의 권리가 침해된 경우 그 침해를 받은 자는 해당 정보를 처리한 정보통신서비스 제공자에게 침해사실을 소명하여 그 정보의 삭제 또는 반박내용의 게재(이하 “삭제등”이라 한다)를 요청할 수 있다.

뒤이어 2016년 6월에는 방송통신위원회의 「인터넷 자기게시물접근배제요청권 가이드라인」이 시행됨에 따라 정보통신 서비스 이용자가 본인이 작성한 게시물에 대해 삭제를 요청할 수 있게 되었다. 해당 가이드라인은 헌법상의 개인정보 자기결정권, 행복추구권 및 사생활의 비밀과 자유, 그리고 정보통신망법상의 정보통신서비스제공자의 책무 등에 근거하여 이용자 본인이 인터넷상 게시한 게시물에 대하여 타인의 접근 배제를 요청할 수 있는 권리를 보장하는 것이 목적이다. 또한 해당 가이드라인에 따르면 법인을 제외한 자연인 누구든지 정보통신망을 이용한 게시판을 직접 관리 및 운영하는 정보통신서비스 제공자 또는 정보통신망을 이용한 검색 서비스를 제공하는 사업자를 대상으로 권리를 행사할 수 있다. 그러나 이 방송통신위원회의 2016년 가이드라인 이상의 잊힐 권리에 대한 법·제도적 뒷받침은 우리나라에선 현재까지는 없는 실정이다.

그럼에도 불구하고 잊힐 권리의 실현과 잊힐 권리를 보장하기 위한 움직임은 지금도 계속되고 있다. 2020년 1월, 'n번방 사건'으로 불리는 단체 채팅방을 통해 아동·청소년 성착취물 및 불법 음란물을 생성, 이를 거래하고 유포한 디지털 성범죄 사건이 세상에 알려졌다. 해당 사건의 피해자가 수만 명에 이를 것으로 추정되면서 법무부에서는 피해자들을 위한 구제책으로 검찰의 'AI(인공지능) 기반 불법촬영물 유포 탐지 및 피해자 지원 시스템'을 이용해 'n번방'에서 다른 사이트로 유출된 불법 영상물을 탐색·삭제하기로 했다(윤지원, 2020). 또한, 주민등록번호, 주소, 이름 등 민감한 개인정보가 유출된 피해자들의 경우, 이들의 성명과 주

민등록번호 변경에 필요한 법률적 조력도 실시할 계획이 있음을 밝혔다(김진주, 2020).

또한 최근 들어 이러한 범죄의 피해자뿐만 아니라 일상생활에서도 잊힐 권리가 실현되는 추세이다. 2020년 초부터 발생한 코로나19의 경우, 발생 초기에는 코로나19 확산 방지와 시민의 알 권리를 충족시키기 위해 확진자의 이동 경로를 각 지자체별로 공개해왔다. 그러나 이동 경로를 공개함으로써 확진자가 방문한 식당 및 카페 등에서 경제적 피해를 호소하고, 확진자의 동선 등 개인정보 유출로 인한 피해가 발생하면서 코로나19 확진자의 이동 경로를 비공개로 전환(서승진, 2020)하거나, 삭제(한지숙, 2020)하는 지자체가 늘어났으며, 한국인터넷진흥원(KISA)에서 동선 정보 탐지 지침과 작성 양식을 만들어 지자체에 배포하고 인터넷 사업자와 간담회를 통해 동선 정보의 자율 삭제를 요청(최태범, 2020)하는 등 정부 차원에서의 움직임도 있었다.

이처럼 우리나라도 꾸준히 잊힐 권리에 대한 움직임이 이어져 오고 있을 뿐 아니라 한국정보화진흥원(2015)의 「2014년 정보문화 실태조사」 보고서에 따르면 설문조사 대상자의 64.8%가 표현의 자유나 알 권리보다는 개인의 프라이버시권이나 개인정보 자기결정권이 더 중요하다고 응답할 정도로 잊힐 권리의 보장에 대한 요구는 일찍이 존재했다. 그러나 현재까지 잊힐 권리와 관련하여 가이드라인 수준 이상의 제도적 뒷받침은 없으며, 특히 국내 기록관리학 분야에서는 관련 논의가 전혀 이루어지지 않고 있다.

기록정보서비스는 기록관리의 궁극적인 목적이자 결과이다(Puhg, 1992). 국가기록원 역

시 기록의 궁극적인 목적은 ‘활용’이라 밝히며, 기록물을 국민들이 보다 쉽고 편리하게 활용할 수 있도록 기록물 열람을 비롯해 콘텐츠 개발, 전시회 등 다양한 기록정보서비스를 제공하고 있다.<sup>3)</sup> 특히 국민의 알권리를 보장하고 국정 투명성을 확보하기 위해 제정된 정보공개제도를 기록정보서비스의 한 측면으로 본다면, 공공기록에 대한 잊힐 권리의 인정과 도입이 정보 접근을 제한하고, 알권리와 의 형평성을 훼손하는 등 기록정보서비스의 가치를 떨어트릴 수 있다. 따라서 기록정보서비스의 중요성이 어느 때보다도 더 강조되고 있는 현재의 실정에서 국내 기록관리 실정에 적합한 잊힐 권리에 대한 논의가 이루어져야 할 필요가 있다.

특히, 잊힐 권리와 자기정보 삭제 요구권을 강제할 수 있는 법안이라고 평가받고 있는 EU의 GDPR 역시 모든 개인정보에 대해 삭제권을 주장할 수 있는 것은 아니며, 공공이익 보장을 위한 기록 보관 목적, 연구를 위한 목적 등 공적인 이익을 위한 경우에는 예외적으로 적용되지 않는다는 점을 고려할 때(김나루, 2018), 국내에서도 잊힐 권리의 적용 범위 등에 대한 논의가 반드시 필요하다.

### 3. 연구 방법

#### 3.1 연구 자료

본 연구는 2010년 1월 1일부터 2020년 12월

31일 사이에 공개 및 게재된 뉴스 기사와 학술지 논문을 분석 자료로 사용하였다. 자료수집의 시점을 2010년 이후로 설정한 이유는 잊힐 권리가 본격적으로 언급되기 시작한 시점이 스페인 변호사 마리오 코스테하 곤잘레스(Mario Costeja Gonzalez)가 소송을 제기한 2010년 이후였기 때문에 이를 기준으로 자료 수집의 시작 시점을 정하였다. 뉴스 기사와 학술지 논문 모두 “잊힐 권리”와 “잊혀질 권리”라는 키워드를 검색어로 사용하였으며, 뉴스 기사는 국내 검색엔진인 네이버 뉴스(<https://news.naver.com>)를 통해, 학술지 논문은 RISS(<https://www.riss.kr>)를 통해 수집하였다.

뉴스 기사의 경우, 파이썬(python)의 웹 스크래핑 라이브러리인 Selenium과 BeautifulSoup4 패키지를 이용하는 웹 크롤링 기법을 적용하여 제목, 작성일, 언론사, 본문 내용을 수집하였으며, 수집된 뉴스 기사는 총 3,888건이었다. 이 중 기사 본문에서 잊힐 권리의 비중이 현저하게 가십성 연예 기사와 양정철 전 보좌관 관련 기사, 중복기사를 제외한 총 1,388건의 뉴스 기사를 분석 대상 자료로 선정하였다.

학술지 논문 검색 결과는 총 270건이었으며, 이 중 발표용 ppt 자료 등 학술지 논문의 형식을 갖추지 않은 자료를 제외한 246건의 학술지 논문을 분석 대상으로 하였다. 학술지 논문의 경우, 초록과 결론만을 분석에 활용하였는데, 이는 텍스트 분석의 특성상 개별 자료의 분량과 문헌 내에 등장하는 단어의 빈도가 분석 결과에 영향을 미치는 만큼 각 학술지의 분량에 따라 연구 결과가 편향될 수 있다고 판단하였

3) 국가기록원 홈페이지. <https://www.archives.go.kr/next/organ/greeting.do>

〈표 1〉 상위 10개 언론사 및 학술지 현황

순	뉴스 기사		학술지 논문	
	언론사명	기사 수	학술지명	논문 수
1	ZDNet Korea	79	정보법학	9
2	보안뉴스	74	憲法學研究	6
3	경향신문	49	홍익법학	6
4	뉴스1	44	언론과 법	5
5	동아일보	41	江原法學	5
6	법률신문	31	고려법학	4
7	매일경제	30	디지털융복합연구	4
8	노컷뉴스	22	법학논총(전남대학교)	4
9	조선비즈(Chosun Biz)	15	법학논총(단국대학교)	4
10	전자신문	13	외법논집	4

기 때문이다.

### 3.2 데이터 전처리

텍스트 분석에 있어 단어에 대한 정제 작업은 반드시 필요하다. 이에 따라 본 연구에서는 자료 정제를 위해 'R' 프로그램을 활용하여 형태소 분석(morphological analysis)과 데이터 전처리(preprocessing)를 실시하였다.

먼저 수집된 모든 자료를 textclean 패키지를 이용하여 한글과 띄어쓰기만을 남기고 모든 불필요한 데이터를 제거한 후, tidytext 패키지와 KoNLP 패키지를 이용하여 각 문장에서 명사만 추출하였다. tidytext 패키지는 텍스트 데이터를 토큰화하여 정돈된(tidy) 데이터로 변환시키는 데 사용하며, KoNLP 패키지는 한글 자연어 분석 패키지로 텍스트 데이터를 의미 단위(형태소 조합)로 정제하는 데 사용한다. 비정형 데이터를 분석하기 위해서는 컴퓨터가 분석할 수 있도록 데이터를 정리하기 위해 자연어처리(NLP) 패키지를 이용하는데, 한국어의 경우

체언과 용언의 변화가 다양하여 영어, 일어 등에서 이용하는 텍스트전처리 패키지를 동일하게 이용할 수 없어 한국어에 특화된 KoNLP 패키지를 이용한다(이종화, 이현규, 2015). 정돈된 데이터는 열은 변수를, 행은 관측치를 형성하며, 이를 통해 다양한 분석을 할 수 있다(Wickham, 2014). 이에 따라 tidytext 패키지를 이용하여 텍스트 데이터를 단어 단위로 끊어내고, 다음으로 KoNLP 패키지의 extractNoun 함수를 이용하여 형태소 분석을 실시, 본문에서 명사 부분만 추출하였다.

다음으로 전처리 작업을 통해 일반명사나 고유명사가 아닌데도 오분류되어 나타난 단어와 단위, 서수 등을 불용어 처리하고, 일부 혼용되는 유의어나 띄어쓰기 등이 맞지 않아 다른 단어로 인식하는 경우를 동일어로 일치시켰다.

### 3.3 분석 방법

이상의 자료를 대상으로 잇힐 권리의 동향 및 관련 논점을 파악하기 위해 키워드 분석과 LDA



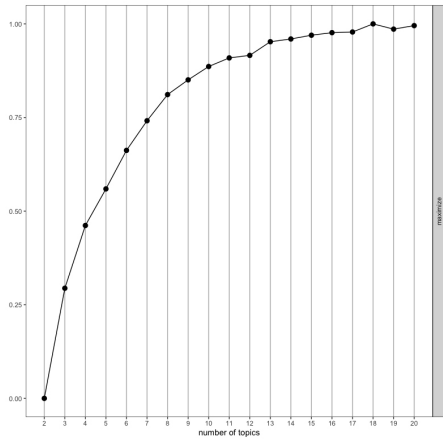
(Latent Dirichlet Allocation) 알고리즘에 기반한 토픽모델링 분석을 수행하였다.

키워드 분석은 TF-IDF(Term Frequency-Inverse Document Frequency)를 이용하였다. TF-IDF는 여러 문서로 이루어진 문서들이 있을 때 특정 단어가 문서 내에서 얼마나 중요한 것인지 나타내는 통계적 수치를 의미한다. TF-IDF를 통해 문서에서 흔하지 않으면서 비중있게 사용된 단어를 추출할 수 있으며, TF-IDF 값이 큰 단어는 해당 단어가 속한 문서의 주제 또는 의미를 결정지을 가능성이 크기 때문에 TF-IDF 값을 주요 키워드를 추출할 수 있는 척도로 활용할 수 있다.

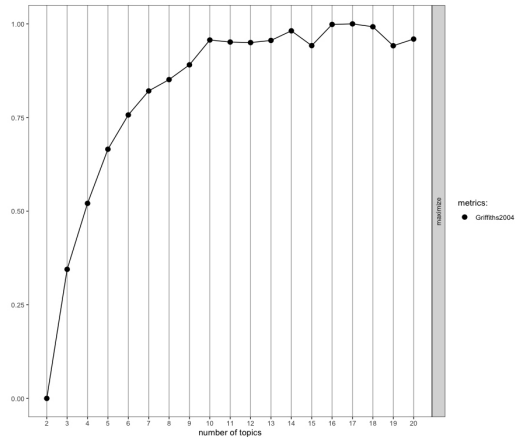
토픽모델링은 문서 집합 내에서 주제(topic)를 도출해줄 뿐만 아니라 개별 주제들이 어떻게 연결되어 있으며, 이들 주제가 시간에 따라 어떻게 변화하는지를 분석할 수 있다(Blei, 2012). 본 연구에서는 토픽모델링 중 LDA 방법을 이용하여 뉴스기사와 학술지 논문을 분석하였다. LDA는 토픽의 확률이 디리클레 분포(Dirichlet distribution) 형태의 사전 분포를 따른다는 것을 가정하는데, 이는 문서들은 토픽들의 혼합으로 구성되어 있고, 토픽들은 확률 분포에 기반하여 단어들을 생성한다고 가정하는 것을 의미한다. 따라서 LDA는 개별 문서는 토픽의 확률적 분포로 표현되고 각 토픽은 단어의 확률적 분포로 표현되는 구조를 이용해 문서 내의 단어들이 어떤 특정 토픽에 포함될 확률을 계산하는 생성확률 모델(generative probabilistic model)이다(Steyvers & Griffiths, 2007). 특히, LDA는 여러 개의 토픽이 내재되어 있는 많은 문서들을 분석에 사용할 수 있고, 같은 의미를 지닌 서로 다른 단어나 문맥에 따라 다른 의

미를 가지는 단어들을 분리하거나 통합하여 다룰 수 있다는 특징이 있다(Born et al., 2014). 즉, 텍스트 데이터 내에 단어들이 서로 연관성이 있다는 것을 가정하고 단어가 나타나는 환경에 따라 주제를 도출하기 때문에 이러한 LDA의 특징을 활용하면 SNS, 뉴스, 논문, 특허 등 비정형 텍스트 데이터의 잠재 주제를 파악하는데 유용하다.

이 경우, 토픽의 개수를 연구자가 지정하여야 하는데 본 연구에서는 여러 모델의 성능 지표를 비교해 최적값을 찾는 하이퍼파라미터 튜닝을 통해 적당한 토픽의 수를 결정하였으며, 토픽의 개수를 2개에서 20개까지 비교하여 복잡도(perplexity) 값을 도출하였다. 복잡도 값은 특정 모델의 성능 지표로, 실제로 관측되는 값을 얼마나 잘 예측하는지를 평가할 때 사용된다(김갑철, 노현중, 2019). 복잡도 그래프는 토픽의 수가 증가하다가 등락을 반복하는 형태로 나타나며, 등락을 반복하기 시작하는 지점이 토픽의 수를 늘려도 성능이 크게 향상되지 않는 지점이라고 해석할 수 있다. 따라서 성능이 비슷하면 토픽의 수를 늘려도 의미가 없으므로 해당 지점에서 토픽의 수를 결정한다. 분석 결과, 토픽의 개수가 신문기사의 경우, 13개, 학술지 논문의 경우, 10개일 때까지 성능 지표가 점진적으로 증가하다가 그 이후부터 등락을 반복하는 것을 확인할 수 있으므로 최종적으로 각각 13개와 10개의 토픽을 선정하는 것이 가장 적절하다고 볼 수 있다(〈그림 1〉, 〈그림 2〉 참조).



〈그림 1〉 토픽 수 선정 결과(신문 기사)



〈그림 2〉 토픽 수 선정 결과(학술지 논문)

## 4. 연구 결과

### 4.1 키워드 분석 결과

분석에 사용된 키워드의 수는 뉴스 기사(1,388건) 545,669개, 학술지 논문(246건) 409,819개였으며, <표 2>는 이들 키워드를 TF-IDF 분석을 통해 사용 빈도 상위 30개의 키워드를 제시한 것이다. “권리”, “삭제”, “인터넷”, “검색”, “서비스” 등의 14개 키워드가 뉴스 기사와 학술지 논문에서 모두 빈번하게 사용된 것으로 나타났다.

뉴스 기사의 경우, 가장 많이 언급된 키워드는 “권리”로 총 7,234회 등장하였다. 이어서 “삭제”(5,775회), “인터넷”(3,964회), “개인정보”(3,891회), “검색”(3,428회), “데이터”(3,058회) 순으로 나타났다. 이를 통해 뉴스 기사에서는 인터넷상에서 개인정보와 관련된 기록을 삭제할 수 있어야 하며, 이를 알 권리와 동등하게 누릴 수 있는 권리로 인정받아야 함에 대한 내용을 주요 논점으로 다루고 있음을 알 수 있다.

특히 학술지 논문에서는 거의 언급되지 않았던 “기업”, “요청”, “개인정보”, “구글” 등의 키워드가 뉴스 기사에서는 빈번하게 사용된 것으로 나타났는데, 이는 현재 잇힐 권리로 인해 많은 국가의 정부에서 기업의 디지털 서비스에 대한 규제가 이루어지고 있는 만큼 이러한 단어가 많이 사용된 것으로 유추해볼 수 있다.

반면 학술지 논문의 경우, “개인”이 가장 많이 언급(1,620회) 되었으며, 이어서 “권리”(1,293회), “삭제”(518회), “자유”(420회), “주체”(360회), “인터넷”(355회) 순으로 나타났다. 학술지 논문의 경우, “주체”, “인격”, “기본권”, “프라이버시”, “법제”, “규정” 등의 키워드가 빈번하게 사용된 것으로 나타나 잇힐 권리와 관련된 법·제도 마련을 위한 연구가 주로 진행되고 있음을 알 수 있다.

### 4.2 토픽모델링 결과

뉴스 기사와 학술지 논문을 대상으로 토픽모

〈표 2〉 TF-IDF 상위 빈도 키워드

뉴스 기사				학술지 논문			
순위	단어	빈도	TF-IDF	순위	단어	빈도	TF-IDF
1	권리	7,234	44,879,411	1	개인	1,620	17,484,057
2	삭제	5,775	35,821,673	2	권리	1,293	13,945,250
3	인터넷	3,964	24,588,371	3	삭제	518	5,587,477
4	<b>개인정보</b>	3,891	24,134,860	4	자유	420	4,531,599
5	검색	3,428	21,264,854	5	<b>주체</b>	360	3,881,617
6	<b>데이터</b>	3,058	18,971,496	6	인터넷	355	3,831,331
7	<b>기업</b>	2,934	18,200,473	7	결과	302	3,258,560
8	디지털	2,848	17,666,347	8	침해	290	3,119,513
9	<b>요청</b>	2,713	16,827,309	9	기술	281	3,025,903
10	서비스	2,516	15,603,944	10	<b>인격</b>	274	2,951,818
11	자신	2,450	15,195,043	11	공개	271	2,917,983
12	<b>게시</b>	2,330	14,450,320	12	검색	254	2,740,374
13	<b>구글</b>	1,856	11,508,829	13	<b>기본권</b>	253	2,726,591
14	<b>사진</b>	1,854	11,501,617	14	<b>이용</b>	239	2,573,798
15	기사	1,786	11,074,212	15	서비스	227	2,447,088
16	자유	1,771	10,986,804	16	요구	227	2,447,088
17	기술	1,742	10,804,172	17	디지털	227	2,440,002
18	<b>온라인</b>	1,707	10,584,047	18	<b>가치</b>	225	2,418,713
19	침해	1,676	10,392,400	19	기사	220	2,368,859
20	판결	1,671	10,362,842	20	자신	207	2,224,985
21	<b>기록</b>	1,651	10,237,004	21	<b>프라이버시</b>	203	2,181,390
22	결과	1,618	10,036,398	22	<b>내용</b>	197	2,122,941
23	요구	1,600	9,924,550	23	<b>법제</b>	192	2,071,485
24	<b>국내</b>	1,554	9,639,887	24	<b>적용</b>	192	2,064,110
25	<b>이용자</b>	1,543	9,572,186	25	판결	188	2,027,141
26	<b>개인정보보호</b>	1,479	9,171,248	26	<b>해결</b>	188	2,027,141
27	<b>관리</b>	1,377	8,538,248	27	<b>규정</b>	188	2,019,728
28	<b>생각</b>	1,359	8,426,943	28	<b>도입</b>	184	1,982,569
29	<b>사건</b>	1,331	8,256,125	29	<b>고려</b>	183	1,967,661
30	공개	1,328	8,232,772	30	<b>검토</b>	178	1,915,274

\* 굵은 글씨는 상위 빈도 키워드 중 각 자료에서만 나타난 키워드임

텔링 분석을 수행한 결과는 〈표 3〉, 〈표 4〉와 같다. 토픽별 주요 키워드 10개와 이를 토대로 선정한 토픽명을 제시하였으며, 이와 함께 해당 토픽이 대표적으로 나타난 문서의 수를 제시하였다. 토픽 모델링 결과는 단순히 토픽이 등장하는 뉴스 기사나 학술지 논문의 양만으로

는 도출된 주제의 중요성 여부를 판단할 수 없다는 판단 하에 도출된 토픽 순으로 제시하였으나 뉴스 기사나 학술지 논문의 양이 완전히 의미가 없다고 할 수 없으므로 빈도순 결과를 함께 제시하였으며, 각 토픽의 명칭은 단어들이 가지는 연계성을 고려하여 연구자들이 부여

한 후, 각각의 토픽과 가장 관련이 큰 뉴스 기사 및 학술지 논문을 직접 확인, 실제 문맥을 고려하여 확정하였다.

4.2.1 뉴스 기사 분석

뉴스 기사에 대한 토픽모델링 분석 결과는 <표 3>과 같다.

토픽 1은 “의견”, “도입”, “한국”, “개인정보 보호”, “전문가”, “법제화”, “방안” 등의 키워드를 포함하고 있는 것으로 나타났다. 토픽 1과 관련된 기사의 내용을 살펴보면, 잊혀질 권리를 국내에 적용하는 문제에 대하여 관련 법제의 도입까지 고려해야 할 사항이 많다는 내용,

법제화 고려 시 표현의 자유, 알권리 등을 종합적으로 판단해야 한다는 내용의 기사들이 모여 있어, 주로 잊힐 권리와 관련된 법안 도입 방안에 대한 내용을 나타내고 있으므로 ‘잊힐 권리 관련 법제화 방안’으로 명명하였다.

토픽 2는 “피해자”, “사건”, “범죄”, “피해”, “처벌”, “유포”와 같은 키워드가 포함되어 있으며, 토픽 2와 관련도가 높은 기사들의 내용은 주로 디지털 성범죄, 성 착취 사건이었던 일명 ‘n번방 사건’ 등 성범죄와 아동학대 등 범죄 사건의 피해자와 피해자 가족들이 입고 있는 2차 피해에 대한 기사들이었다. 따라서 해당 뉴스 기사의 의미와 맥락을 고려하여, ‘범죄 피해자’

<표 3> 뉴스 기사 토픽모델링 결과

토픽	토픽명	핵심 키워드	기사수(%)
1	잊힐 권리 관련 법제화 방안	의견, 도입, 한국, 개인정보보호, 고려, 주체, 전문가, 방송, 법제화, 방안	137(9.87)
2	범죄 피해자	피해자, 사건, 범죄, 피해, 처벌, 영상, 여성, 경찰, 불법, 유포	83(5.98)
3	디지털 장의사	업체, 동영상, 흔적, 사이트, 친구, 대표, 사이버, 계정, 페이스북, 확산	177(12.75)
4	GDPR(General Data Protection Regulation)	제공, 규정, 유럽연합, 데이터, 정책, 시행, 관리, 위반, 개인정보보호, 이용	82(5.90)
5	‘새 정부 개인정보정책 방향과 정보인권을 위한 정책토론회’	국민, 회의, 발언, 대통령, 정부, 지적, 국회, 참석, 인사	47(3.39)
6	DAS(Digital Aging System)과 강원도조례	사업, 지원, 계획, 관계자, 미래, 최초, 소멸, 관리, 시스템, 홈페이지	82(5.90)
7	블록체인	사용, 기능, 사용자, 이용, 공유, 저장, 방식, 제공, 대표, 서버	87(6.27)
8	잊힐 권리 인정 사례	기사, 사법재판소, 보도, 언론, 소송, 링크, 판단, 이름, 법원, 스페인	178(12.82)
9	EU의 실리콘밸리 기업 규제에 대한 미국의 반발	미국, 정부, 국가, 페이스북, 프랑스, 영국, 시작, 규제, 광고, 이용자	101(7.28)
10	잊힐 권리 법제화 논란	의미, 존재, 우리나라, 현실, 생각, 과정, 상황, 판단, 핵심, 목적	87(6.27)
11	한국정보화진흥원(NIA) ‘2015년 개인정보보호 10대 트렌드 전망’ 보고서	활용, 변화, 환경, 경제, 발전, 이슈, 안전, 분석, 보안, 빅데이터	65(4.69)
12	인터넷 자기계시물 접근배제 요청권	게시물, 포털, 방송, 자기, 본인, 통신위원회, 이용자, 공개, 게시, 사업자	152(10.96)
13	(분석 제외)	생각, 기억, 세상, 시대, 이야기, 인간, 오늘, 마음, 중요, 일상, 사람들	110(7.93)

로 토픽명을 정하였다.

토픽 3은 “업체”, “동영상”, “흔적”, “사이트” 등이 핵심 키워드로 도출되었다. 이는 일명 ‘DAS(Digital Aging System)’라고 불리는 디지털 소멸 시스템과 관련된 단어들이면서 디지털 장의사를 설명할 수 있는 키워드이기도 하였다. 따라서 정확한 내용의 파악을 위해 주요 뉴스 기사를 확인한 결과 역시 디지털 장의사, 디지털 세탁업 등에 대한 내용과 웹 상의 정보를 삭제하기 위해 정보 주체 사망 시 사망자가 웹 상에 남긴 다양한 흔적의 상속 여부 문제 등을 다루고 있어 ‘디지털 장의사’를 토픽명으로 정하였다.

토픽 4는 “규정”, “유럽연합”, “데이터”, “정책”, “시행”, “개인정보보호” 등의 단어가 포함된 것으로 나타났으며, 토픽 4에 해당하는 기사들의 주요 내용은 EU의 ‘GDPR(General Data Protection Regulation)’ 법안 통과 및 시행에 관한 기사들로, 이에 따라 토픽 4의 이름을 ‘GDPR(General Data Protection Regulation)’이라 명명하였다.

토픽 5에 포함되어 있는 주요 키워드는 “국민”, “회의”, “발언”, “대통령”, “정부”, “지적”, “국회” 등이었는데, 이들 키워드는 2013년 2월에 박근혜 정부가 새로 들어서면서 사회문제가 되고 있는 개인정보 침해·남용에 따른 문제 해결 방안을 논의하는 토론회와 관련한 뉴스 기사에서 자주 등장하였으므로, 해당 토픽을 ‘새 정부 개인정보정책 방향과 정보인권을 위한 정책토론회’로 명명하였다.

토픽 6은 “사업”, “지원”, “미래”, “최초”, “소멸”, “관리”, “시스템” 등의 단어들로 구성되어 있으며, 여기에서는 디지털 소멸 시스템 DAS(Digital Aging System)의 상용화 추진과 지

자체 최초로 잊힐 권리를 인정한 바 있는 강원도에서 디지털 소멸 시스템인 DAS을 도입한 내용을 포함하고 있어, ‘DAS와 강원도 조례’로 토픽명을 정하였다.

토픽 7의 토픽명은 ‘블록체인’으로, 주로 등장한 단어는 “사용”, “기능”, “사용자”, “이용”, “공유”, “저장”, “방식” 등이었다. 관련성이 높은 기사들의 주요 내용을 확인한 결과, 정보를 분산하여 저장하는 블록체인 기술은 데이터 삭제가 불가능한 것이 특징이기 때문에 잊힐 권리를 침해할 수 있다는 지적과 함께 등장한 단어들이었다.

토픽 8에는 “기사”, “사법재판소”, “보도”, “언론”, “소송”, “판단”, “법원”, “스페인” 등의 단어가 등장했다. 포함되는 주요 기사들의 내용은 마리오 코스테하 곤잘레스와 구글 스페인의 소송 건에 관해 유럽사법재판소가 잊힐 권리를 인정한 것을 비롯하여 여러 국가에서 잊힐 권리에 대한 판결이 이어진 것에 대한 내용들로, 따라서 해당 토픽명을 ‘잊힐 권리 인정 사례’로 명명하였다.

토픽 9에 등장하는 주요 키워드는 “미국”, “정부”, “국가”, “페이스북”, “프랑스”, “영국”, “시작”, “규제”, “광고”, “이용자”였다. 이들 단어가 포함된 기사들은 유럽연합이 ‘GDPR’을 통해 잊힐 권리를 인정하면서 이로 인해 구글을 비롯한 여러 실리콘밸리 기업이 유럽연합 시장에서 각종 규제를 받는 데에 대한 미국 정부의 항의에 대한 내용을 나타내고 있어, 토픽명은 ‘EU의 실리콘밸리 기업 규제에 대한 미국의 반발’로 정하였다.

토픽 10은 ‘잊힐 권리 법제화 논란’으로 정하였는데, “의미”, “존재”, “우리나라”, “현실” 등

의 단어가 포함되어 있으며, 주요 기사의 내용은 한국형 잊혀질 권리에 대한 제안, 우리나라의 실정에 맞는 ‘정보 갱신권’의 도입, 잊힐 권리의 법제화를 위한 사회적 의견 수렴의 필요성 등이었다.

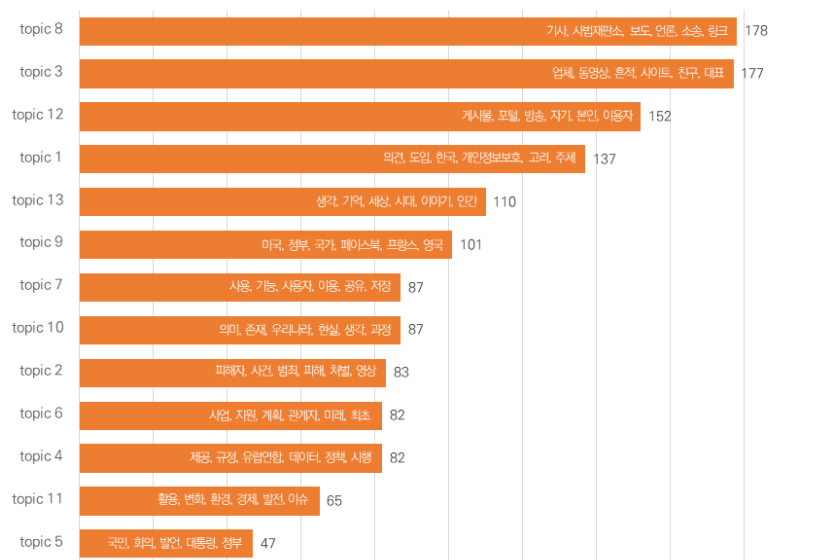
토픽 11에서는 “활용”, “변화”, “환경”, “경제”, “발전”, “이슈”, “안전”, “분석”, “보안”, “빅데이터”의 키워드가 포함되어 있었는데, 이 단어들은 한국정보화진흥원(NIA) ‘2015년 개인정보보호 10대 트렌드 전망’ 보고서에 관련된 내용이 포함된 기사에 나타나는 단어들로 한국정보화진흥원(NIA) ‘2015년 개인정보보호 10대 트렌드 전망 보고서’를 토픽명으로 정하였다.

토픽 12는 “게시물”, “포털”, “방송”, “자기”, “본인”, “통신위원회”, “이용자”, “공개” 등이 주요 키워드였으며, 기사들의 주요 내용은 방송통신위원회의 ‘잊혀질 권리 가이드라인’ 논

의, ‘인터넷 자기게시물 접근배제요청권 가이드라인’의 초안 공개 및 시행과 관련된 내용이 주를 이루고 있어, 해당 토픽명을 ‘인터넷 자기게시물 접근배제 요청권’으로 명명하였다.

마지막으로 토픽 13의 경우 “생각”, “기억”, “세상”, “시대”, “이야기”, “인간”, “오늘”, “마음”, “중요”, “일상”, “사람”들과 같은 단어가 등장하였으나, 이들 단어들을 포함하는 기사들의 경우, 그 맥락상 연결되는 하나의 주제가 존재하지 않아 해당 토픽은 분석에서 제외하였다.

가장 많은 뉴스 기사에서 등장한 토픽은 토픽 8 ‘잊힐 권리 인정 사례’로 178개 기사에서 등장하였으며, 전체의 12.82%를 차지하였다. 다음으로 많이 나타난 토픽은 토픽 3인 ‘디지털 장의사’(177개, 12.75%)였으며, 이어서 토픽 12 ‘인터넷 자기게시물 접근배제 요청권’이 152개(10.95%) 기사에서 등장하였다(〈그림 3〉 참조).



〈그림 3〉 뉴스 기사 빈도순 토픽과 주요 단어

4.2.2 학술지 논문 분석

다음으로 학술지 논문에 대한 토픽모델링 분석 결과는 <표 4>와 같다.

토픽 1에 포함된 단어들은 “법률”, “방안”, “개인정보”, “제도”, “조화”, “주체”, “프라이버시” 등이었다. 토픽 1에 포함된 주요 문헌들은 잊혀질 권리(잊힐 권리) 및 개인정보보호에 대한 필요성과 국내외적 이슈 분석을 통해 국내에서의 잊힐 권리 법제화의 의미와 법제화를 위한 시사점을 제시하고 있으므로 ‘잊힐 권리 관련 법제화 방안’으로 토픽명을 정하였다.

토픽 2는 “삭제”, “판단”, “사실”, “내용”, “판결” 등이 주요 키워드였으며, 이들 키워드를 포함하는 문헌들의 경우, 유럽연합 내 잊힐 권리에 관한 판결과 시사점을 포함하여, 잊힐 권리 인정의 타당성을 유럽연합의 판례를 중심으로 분석하는 내용을 포함하고 있어, 해당 토픽명을 ‘유럽연합 내 판결’로 명명하였다.

토픽 3은 “영향”, “분석”, “결과”, “가능성”,

“인식”, “중심”, “데이터”, “수집”, “반영”, “역할”이 핵심 키워드로 등장하였는데, 각 키워드만 봤을 때 이들의 의미가 다소 모호한 측면이 있다. 그러나 토픽 3을 구성하는 주요 문헌들의 내용을 확인한 결과, 잊혀질 권리에 대한 언론 보도 프레임 분석, SNS 환경에서의 잊혀질 권리 구현 방안 등 빅데이터 분석을 수행했거나 빅데이터 환경에서 잊힐 권리의 구현에 관련된 연구들이 핵심 문헌으로 등장하고 있어, 해당 내용의 전체적인 맥락을 고려하여, ‘빅데이터 분석’으로 토픽명을 정하였다.

토픽 4의 핵심 단어들은 “개인”, “자유”, “표현”, “논의”, “언론”, “사회” 등이었다. 이들 키워드가 등장한 문헌들의 내용은 인터넷과 언론 기사의 데이터베이스 축적이 보도 기사의 유효기간을 무한대로 만들었을 뿐 아니라 언제든지 개인의 과거를 현재화할 수 있는 도구가 되었음을 지적하며, 언론 기록의 삭제 권리 부여와 관련된 동향 분석 및 언론 보도 내용의 제한을

<표 4> 학술지 논문 토픽모델링 결과

토픽	토픽명	핵심 키워드	논문수(%)
1	잊힐 권리 관련 법제화 방안	법률, 방안, 개인정보, 제도, 조화, 주체, 프라이버시, 모색, 다양, 충돌, 법제화, 문제	26(10.57)
2	유럽연합 내 판결	삭제, 판단, 사실, 내용, 판결, 우리나라, 인정, 기본권, 형량, 주장	48(19.51)
3	빅데이터 분석	영향, 분석, 결과, 가능성, 인식, 중심, 데이터, 수집, 반영, 역할	26(10.57)
4	언론 보도의 잊힐 권리	개인, 자유, 표현, 논의, 언론, 마련, 사회, 권의, 역사, 국민	29(11.79)
5	디지털 기술과 프라이버시 보호	범위, 해결, 상황, 문제, 비교, 유통, 실현, 디지털, 방법, 논의	23(9.35)
6	법외에 관한 잊힐 권리	시행, 지위, 단계, 수단, 보존, 미국, 등장, 공개, 적절, 여기	14(5.69)
7	블록체인	기술, 인터넷, 요구, 자신, 사회, 이용, 기록, 이용자, 서비스, 위험	25(10.16)
8	개인정보보호	보호, 개인정보, 공개, 적용, 행위, 처리, 목적, 문제, 개인정보보호, 요건	17(6.91)
9	GDPR(General Data Protection Regulation)	인격, 제정, 입법, 정보보호, 규정, 기업, 유럽연합, 확보, 행사, 우선	26(10.57)
10	(분석 제외)	연구, 다음, 생각, 부분, 기준, 결과, 대표, 제시, 강조	21(8.54)

위한 제도 도입을 위한 논의가 대다수이다. 따라서 이는 ‘언론 보도의 잊힐 권리’로 토픽명을 정할 수 있다.

토픽 5에서는 “범위”, “해결”, “상황”, “실현”, “디지털”, “방법” 등의 단어가 포함되어 있고, 주요 문헌들의 내용이 온라인 공간에서의 개인정보 유통과 이로 인한 명예훼손 문제를 해결하기 위한 연구들이 주를 이루고 있어, ‘디지털 기술과 프라이버시 보호’로 토픽명을 정하였다.

토픽 6은 “시행”, “지위”, “단계”, “수단”, “보존”, “미국”, “등장”, “공개” 등의 단어가 포함되어 있었는데, 관련성이 높은 문헌들의 내용을 확인한 결과, 성범죄자의 신상공개제도와 그 문제점, 그리고 범죄자의 잊혀질 권리의 인정에 대한 부분을 주제로 다루고 있었다. 따라서 해당 토픽명은 ‘범죄에 관한 잊힐 권리’로 명명하였다.

토픽 7에 포함되어 있는 단어들은 “기술”, “인터넷”, “요구”, “서비스” 등으로 잊힐 권리와 관련된 핵심 기술인 블록체인과 관련 있는 단어들로 구성되어 있으며, 문헌의 내용 역시 블록체인 기술에서 개인정보보호 관련 논점과 블록체인을 통해 구현한 시스템에서 개인정보를 보호하기 위한 방안과 관련된 연구들이 주를 이루어, ‘블록체인’으로 토픽명을 정하였다.

토픽 8은 ‘개인정보보호’로 토픽명을 정하였다. 토픽 8에서는 개인정보의 보호와 관련된 단어들이 “보호”, “개인정보”, “공개”, “적용”, “행위” 등의 단어가 핵심 키워드로 등장하였으며, 주로 개인정보보호, 잊혀질 권리 등에 대한 법·제도적 고찰 관련 문헌이 주를 이루고 있다.

토픽 9에서는 “인격”, “제정”, “입법”, “정보

보호”, “규정”, “기업”, “유럽연합” 등의 단어가 포함되어 있었으며, 관련성이 높은 문헌들로는 잊힐 권리에 대한 유럽의 논의를 검토하고 이를 국내에 적용할 때 참고할 수 있도록 유럽의 개인 정보 일반 규정의 내용을 다루고 있는 연구가 대부분이다. 따라서 토픽명은 ‘GDPR’로 하였다.

마지막으로 토픽 10에서 나타난 주요 단어들은 “연구”, “다음”, “생각”, “부분”, “기존”, “결과” 등으로, 이러한 단어들은 학술지 논문에 일반적으로 사용되는 단어들이며 잊힐 권리와 관련하여 직접적인 연관이 없다고 생각하여 이를 제외하였다.

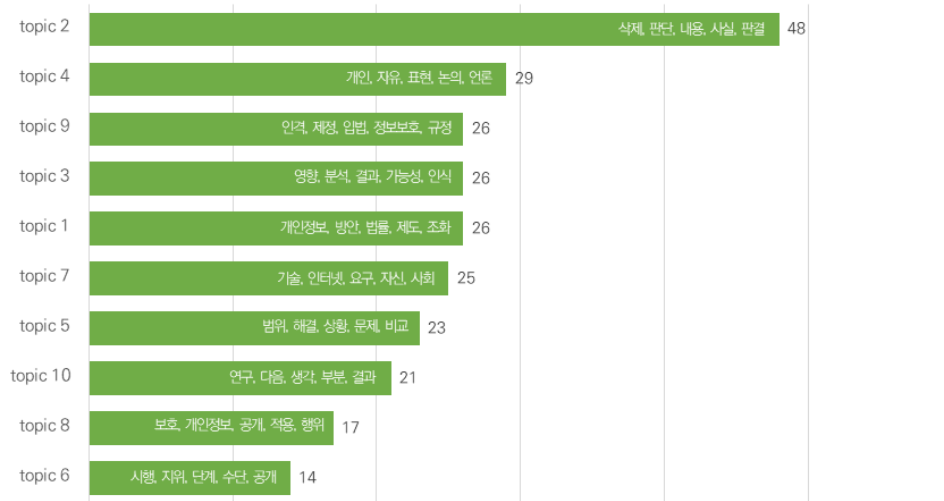
학술지 논문에서 가장 많이 등장한 토픽은 토픽 2 ‘유럽연합 내 판결’이었으며, 다음으로 토픽 4 ‘언론에 나타난 잊힐 권리’, 토픽 9 ‘GDPR’, 토픽 3 ‘빅데이터’, 토픽 1 ‘법제화’ 순으로 나타났다(〈그림 4〉 참조).

#### 4.2.3 뉴스 기사와 학술지 논문 비교

잊힐 권리에 대해 뉴스 기사와 학술지 논문에 나타난 논점에 어떠한 차이가 있는지 살펴보기 위해, 뉴스 기사와 학술지 논문에 나타난 토픽을 단어 구성과 대표 기사 및 논문의 내용을 고려하여 비교하였다.

비교 결과, 뉴스 기사와 학술지 논문에서 등장하는 토픽들을 범주화할 수 있었는데, 크게 법안 및 규정, 잊힐 권리의 실현을 위한 기술 구현 및 적용, 인격권 및 개인정보자기결정권, 사례 및 해석, 기타로 나누어 볼 수 있으며, 등장하는 토픽들이 해당 범주들에서 크게 벗어나지 않아 현재까지 언론과 학계 모두 비슷한 수준의 논의가 이루어져 오고 있음을 확인할 수 있었다.





〈그림 4〉 학술지 논문 빈도순 토픽과 주요 단어

〈표 5〉 뉴스 기사와 학술지 논문 간 토픽 비교

항목	뉴스 기사		학술지 논문	
	토픽명	항목당 기사 수 (%)	토픽명	항목당 논문 수 (%)
법안 및 규정	1. 잊힐 권리 관련 법제화 방안	306 (22.05)	1. 잊힐 권리 관련 법제화 방안	52 (21.14)
	4. GDPR		9. GDPR	
	10. 잊힐 권리 법제화 논란			
잊힐 권리의 실현을 위한 기술 구현 및 적용	3. 디지털 장의사	346 (24.93)		48 (19.51)
	6. DAS		5. 디지털 기술과 프라이버시 보호	
	7. 블록체인		7. 블록체인	
인격권 및 개인정보 자기결정권	2. 범죄 피해자의 잊힐 권리	235 (16.93)		60 (24.39)
	12. 인터넷 자기게시물 접근배제 요청권		4. 언론 보도의 잊힐 권리	
			6. 범죄자의 잊힐 권리	
			8. 개인정보보호	
사례 및 해석	8. 잊힐 권리 인정 사례	279 (20.10)		48 (19.51)
	9. EU 기업 규제에 대한 미국의 반발		2. 유럽연합 내 판결	
기타	5. 정책토론회	112 (8.07)		26 (10.57)
	11. 한국정보화진흥원 보고서			
			3. 빅데이터 분석	

\* 초록색 음영은 뉴스 기사와 학술지 논문에 공통적으로 나타난 토픽을 표시한 것임

하지만 각 항목에 포함된 토픽들이 가장 많이 나타나는 문서의 수를 기준으로 살펴보면 언론과 학계에서 주로 관심을 두는 항목에는 차이가 있는 것으로 나타났다.

먼저 뉴스 기사의 경우, '잊힐 권리의 실현을 위한 기술 구현 및 적용'에 대한 내용을 가장 많이 다루고 있었는데, 이는 디지털 시대를 맞아 온라인과 오프라인의 경계가 모호해지면서 점점 더 표현의 자유보다 잊힐 권리에 대해 많은 사람들이 주목하고 있다는 것을 알 수 있다. 뿐만 아니라 잊힐 권리 관련 기술의 도입과 개발 전망에 대한 문제가 지속적으로 제기되고 있으며, 잊힐 권리의 도입을 일종의 산업으로 인식하는 시각에 주목하고 있음을 확인할 수 있다.

반면 학술지 논문의 경우, '인격권 및 개인정보자기결정권'에 대한 내용을 가장 많이 다루고 있었다. 구체적으로 잊힐 권리의 법제화 등 제도적 측면에서 정보주체의 대상이나 삭제할 수 있는 정보 등에 대한 논란을 다루고 있으며, 이는 잊힐 권리에 대한 연구가 예상되는 다양한 한계와 문제점을 제기하고 대응방안을 모색하고자 하는 연구가 대부분이라고 볼 수 있다. 그러나 그 내용을 들여다보면 구체적인 해결방안의 제시보다는 해결을 위한 쟁점이나 검토 사항을 제시한 연구가 대부분으로 아직까지는 심도 있는 논의가 이루어지지 않았다고 판단할 수 있다.

토픽의 경우, 뉴스 기사와 학술지 논문에 공통적으로 나타난 토픽은 '개인정보보호 관련 법제화 방안', 'GDPR', '블록체인'으로, 이들 토픽에 대해서는 언론과 학계 모두 유사한 관심을 가지고 있는 것으로 판단할 수 있다.

특히, 'GDPR' 토픽과 '블록체인' 토픽의 경우, 잊힐 권리에 대해 뉴스 기사와 학술지 논문 모두 유사한 논점을 가지고 있음을 알 수 있다. 'GDPR'의 경우, EU의 새로운 개인정보보호법인 GDPR의 통과 및 시행, 국내외의 적용에 대한 내용을 주로 다루고 있으며, '블록체인' 토픽에서도 블록체인 기술이나 블록체인 기술을 적용한 시스템에서 발생할 수 있는 개인정보 침해와 관련된 쟁점을 다루고 있다.

하지만 '잊힐 권리 관련 법제화 방안' 토픽의 경우, 해당 토픽을 구성하는 단어에는 다소 차이가 있는데, '법제화 방안'과 관련하여 뉴스 기사의 경우, "의견", "도입", "한국", "고려" 등의 단어들 포함되어 국내에 잊힐 권리의 적용을 위한 법안 도입에 있어 고려해야 할 사항들에 대해 주목하고 있는 반면, 학술지 논문의 경우, "법률", "방안", "조화", "모색", "충돌" 등의 단어들 포함되어 잊힐 권리의 필요성에 대한 연구 및 잊힐 권리의 법리적 해석과 문제 분석 등에 대한 연구가 주로 수행되어 왔음을 알 수 있다.

한가지 주목할 부분은 인격권 및 개인정보자기결정권 항목에서는 그 대상을 달리 바라보고 있다는 점이다. 뉴스 기사의 경우, 각종 범죄사건의 피해 당사자 및 그 가족들의 잊힐 권리에 대한 내용을 주로 언급하고 있다. 그러나 학술지 논문의 경우, 범죄자의 잊힐 권리와 언론 보도 기록의 삭제에 대한 내용을 연구 주제로 다루고 있는 경우가 많았다. 이는 특정 개인의 사생활과 인격권 침해 방지를 위하여 잊힐 권리를 인정하는 경우, 그 대상이 되는 정보주체가 비록 범죄를 저지른 또는 저질렀던 사람이라 할지라도 예외 없이 잊힐 권리가 인정되어야

하며, 이를 위해서는 관련 법제의 개선 및 한계 설정 등이 필요하다는 것이다. 또한 범죄자의 잊힐 권리를 인정하는 경우, 그 대상 기록은 범죄 관련 정보 일체를 포함하게 된다. 즉, 언론기관의 범죄 사실 보도 및 온라인 상에 게시된 게시물 등에 범죄 관련 사실이 포함되어 있는 경우를 모두 고려하여야 하므로 언론 보도의 잊힐 권리 역시 이에 초점을 맞춘 연구가 많은 것으로 보인다. 아직 기록관리 분야에서 해당 내용과 관련된 논의가 이루어진 적은 없으나 범죄 관련 기록의 삭제는 국민의 알권리 및 표현의 자유와 충돌할 수 밖에 없는 문제이자, 기록의 유지와 삭제에 관한 것이라는 점에서 기록관리 측면에서도 시사하는 바가 크다.

## 5. 결 론

본 연구는 잊힐 권리와 관련한 뉴스 기사와 학술지 게재 논문을 대상으로 텍스트마이닝 분석을 활용해 각 문서 내에 나타난 논점과 특성을 살펴보았다. 구체적으로 키워드 분석을 통해 각 문헌에서 사용된 주요 단어들을 파악함으로써 주로 다루고 있는 논점을 확인하였으며, 다음으로 토픽모델링 분석을 활용해 문서 내에 내재된 토픽을 추출하고 잊힐 권리와 관련된 관심과 주요 주제를 살펴보았다.

먼저 키워드 분석의 경우, “권리”, “삭제”, “인터넷”, “검색”, “서비스” 등의 14개 단어가 뉴스 기사와 학술지 논문에서 모두 빈번하게 사용된 것으로 나타났다. 이들 단어가 포함된 문헌의 주요 내용은 인터넷 상에서 개인정보와 관련된 기록을 삭제할 수 있어야 하며, 이를 알

권리와 동등하게 누릴 수 있는 권리로 인정해야 한다는 것으로, 뉴스 기사와 학술지 논문에서 모두 다루고 있는 관점으로 볼 수 있다.

그러나 해당 자료에서만 등장하는 단어도 상당수 존재하여 뉴스 기사와 학술지 논문 각각에서만 다루고 있는 내용도 파악할 수 있었는데, 먼저 뉴스 기사의 경우, 학술지 논문에서는 거의 언급되지 않았던 “기업”, “요청”, “개인정보”, “구글” 등의 키워드가 빈번하게 사용되었다. 이는 현재 잊힐 권리로 인해 많은 국가의 정부에서 기업의 디지털 서비스에 대한 규제가 이루어지고 있는 만큼 이와 관련된 기사가 많이 보도되고 있는 것으로 판단된다. 반면 학술지 논문의 경우, “주체”, “인격”, “기본권”, “프라이버시”, “법제”, “규정” 등의 단어가 빈번하게 사용된 것으로 나타나 잊힐 권리와 관련된 법·제도 마련을 위한 연구가 주로 진행되고 있음을 알 수 있다.

다음으로 토픽모델링 분석 결과, 뉴스 기사에서는 총 12개의 분석 대상 토픽이 추출되었고, ‘잊힐 권리 인정 사례’, ‘디지털 장의사’, ‘인터넷 자기게시물 접근배제 요청권’ 순으로 많은 기사 수를 포함하고 있었다. 학술지 논문의 경우에는 총 9개의 분석 대상 토픽이 추출되었으며, ‘유럽연합 내 판결’, ‘언론에 나타난 잊힐 권리’ 토픽을 포함한 학술지 논문이 상대적으로 더 많이 나타났다.

뉴스 기사와 학술지 논문에서 나타난 토픽 간 유사점과 차이점을 살펴본 결과는 다음과 같다. 먼저 각 자료에서 등장하는 토픽들을 범주화 한 결과, 법안 및 규정, 잊힐 권리의 실현을 위한 기술 구현 및 적용, 인격권 및 개인정보 자기결정권, 사례 및 해석의 범주들에서 크게

벗어나지 않아 현재까지 언론과 학계 모두 비슷한 수준의 논의가 이루어져 오고 있음을 확인할 수 있었다.

하지만 해당 범주 내에서도 언론과 학계에서 주로 관심을 갖는 토픽에는 차이가 있었는데, 뉴스 기사의 경우, '잊힐 권리의 실현을 위한 기술 구현 및 적용'에 대한 내용을 가장 많이 다루고 있어, 점점 더 많은 사람들이 잊힐 권리에 대해 주목하고 있다는 것을 알 수 있을 뿐만 아니라 잊힐 권리의 도입을 일종의 산업으로 인식하는 시각에 주목하고 있음을 확인할 수 있다. 반면 학술지 논문은 잊힐 권리의 법제화 등 제도적 측면에서 정보주체의 대상이나 삭제할 수 있는 정보 등에 대한 논란 등 '인격권 및 개인정보자기결정권'에 대한 내용을 가장 많이 다루고 있었다. 이는 잊힐 권리에 대한 연구가 예상되는 다양한 한계와 문제점을 제기하고 대응방안을 모색하기 위한 것으로 보인다.

토픽의 경우, '개인정보보호 관련 법제화 방안', 'GDPR', '블록체인' 토픽들은 뉴스 기사와 학술지 논문에 공통적으로 나타난 토픽으로, 언론과 학계 모두 유사한 관심을 가지고 있는 것으로 판단할 수 있다. 다만 다른 두 토픽과 달리 '개인정보보호 관련 법제화 방안' 토픽은 해당 토픽을 구성하는 키워드들이 다소 상이하서, 같은 토픽이라도 언론과 학계에서 주목하는 쟁점에 차이가 있음을 알 수 있다.

또한 뉴스 기사와 학술지 논문에서 뚜렷한 차이를 보이는 토픽도 발견할 수 있었다. 잊힐 권리의 대상이라고 볼 수 있는 정보주체에 대하여 뉴스 기사에서는 각종 범죄사건의 피해 당사자 및 그 가족들의 잊힐 권리에 대한 내용을 주로 언급하고 있으나 학술지 논문의 경우,

범죄자의 잊힐 권리와 언론 보도 기록의 삭제에 대한 내용을 연구 주제로 다루고 있는 경우가 많았다.

본 연구의 결과가 시사하는 바는 다음과 같다.

첫째, 본 연구에서 도출된 이슈들을 중심으로 하여 기록관리 분야에서도 적극적인 논의가 필요하다. 잊힐 권리에 대한 뉴스 기사와 학술지 논문은 큰 범주 내에서는 비슷한 내용의 논의가 이루어지고 있는 것으로 나타났으며, 이는 기록관리 분야에서도 동일하게 논의될 수 있는 주제들이다. 따라서 해당 주제들에 관심을 유지하고 지속적인 논의가 이루어져야 할 필요가 있다. 특히 뉴스 기사를 통해 잊힐 권리에 대한 대중들의 공감 정도와 인식을 엿볼 수 있다고 볼 때 큰 범주 내에서 유사한 내용의 논의가 이루어진다는 것은 사회적 이슈들이 학계에서 논의되는 이슈로 확장되고 있다고 해석할 수 있으므로 잊힐 권리에 대한 주제가 현실적 문제 인식에서 시작되어야 한다는 점을 생각해 볼 수 있다.

둘째, 뉴스 기사와 학술지 논문 간 공통적으로 나타나는 주제들을 우선적으로 고려할 필요가 있다. 잊힐 권리와 관련하여 중요하고 필요한 주제라면 우선적으로 논의되어야 할 필요가 있는데 뉴스 기사와 학술지 논문에서 모두 나타나는 주제는 사회적 및 학문적으로 모두 관심을 받고 있다고 볼 수 있다. 따라서 아직까지 기록관리학 분야에서 잊힐 권리와 관련된 논의가 이루어지지 않은 만큼 해당 주제를 우선적 논의 대상으로 고려하여, 이와 관련된 연구 과제를 진행함으로써 기록관리학 분야에서 잊힐 권리에 대한 함의를 제공하기 위한 기반을 마련할 수 있을 것이다.

셋째, 쟁점상 이견이 존재하는 경우, 이를 다각적으로 논의하는 것이 필요해 보인다. 분석 결과, 뉴스 기사와 학술지 논문 간 유사한 토픽이라도 그 내용과 관점에 차이가 있는 경우가 있는 것으로 나타났다. 특히 법·정책적 차원에서 잊힐 권리의 대상 및 적용 범위에 대해 서로 강조하고 있는 부분이 상이하다는 것은 주목할 만하다. 따라서 이견이 존재하는 쟁점의 경우 다각적으로 살펴볼 필요가 있다.

본 연구의 결과 및 시사점을 토대로 볼 때 기록관리 분야에서도 기록정보서비스에 있어 잊힐 권리를 적용하는 대상이나 수준에 대한 논의가 필요하며, 기록의 삭제나 기록 검색에서의 배제 등과 같은 잊힐 권리의 방법에 대한 차원의 논의도 이루어져야 할 필요가 있다. 나아가 본 연구의 결과로 도출된 주요 주제들은 잊힐 권리와 관련된 연구 주제를 도출하는 데 참고가 될 수 있으며, 지속적인 고민과 논의를 통해 올바른 연구 방향을 수립하는 데 도움을 줄 수 있을 것이다.

본 연구는 텍스트 마이닝 분석을 활용하여 뉴스 기사와 학술지 논문을 분석함으로써 잊힐 권리에 대한 주요 주제를 분석하였다. 특히 토

픽모델링 분석을 통해 잊힐 권리에 대한 주요 관심과 쟁점이 무엇인지 구체적으로 파악할 수 있었으며, 뉴스 기사와 학술지 논문 간 비교를 통해 이들 간 유사점과 차이점을 파악하고, 부분적으로 이들 간에 차이가 있음을 확인하였다. 또한 이 연구는 잊힐 권리에 대한 논의가 전무한 상태인 기록관리학 분야에서 잊힐 권리의 의미와 향후 발생할 수 있는 논점을 가늠해볼 수 있다는데 그 의의가 있다.

다만 학술지 논문의 수가 상대적으로 부족하여 뉴스 기사와 학술지 논문 간 토픽의 특징이나 차이를 세밀하게 비교, 분석하기에는 한계가 있었다. 또한 후속연구로 시계열 분석을 통해 뉴스기사와 학술지 논문에 나타나는 주제가 언급된 시기와 비율을 분석함으로써 시기별, 문서별 토픽 변화 추이를 좀 더 세밀하게 파악할 필요가 있다. 그러나 잊힐 권리를 보장하기 위한 국가적 차원의 움직임이나 사회적·학문적 논의가 이루어지고 있는 현실에서 기록관리학 분야에서도 더이상 이와 관련한 논의를 배제할 수 없으며, 본 연구에서 도출된 내용들을 중심으로 기록관리학 분야에서의 잊힐 권리에 대한 다양한 논의가 이루어지기를 기대한다.

## 참 고 문 헌

- 고은별, 최마리아, 최광희 (2012). 개인정보보호 측면 중심의 잊혀질 권리에 관한 분석. 한국경영정보학회 학술대회논문집, 460-464.
- 김갑석 (2017). 유럽에서의 잊힐 권리의 전개와 한국에서의 법제 제도화의 방안. 유럽헌법연구, 25, 53-78.
- 김갑철, 노현중 (2019). 토픽모델링을 활용한 국내 지역 지리교육 연구 동향 분석. 사회과교육, 58(4),

- 49-67. <https://doi.org/10.37561/sse.2019.12.58.4.49>
- 김나루 (2018). 빅데이터 시대의 잊힐 권리에 관한 최근의 입법 동향과 법적 과제. *비교법연구*, 18(1), 7-43.
- 김진주 (2020.4.2). 검찰 '박사방' 피해자 개명·주민번호 변경 지원. *한국일보*,  
출처: <https://www.hankookilbo.com/News/Read/202004021498060568>
- 남윤아, 임진희 (2016). 국가기록원 기록정보센터의 열람서비스 제공 절차 및 특징에 관한 질적 연구. *기록학연구*, 50, 177-229. <https://doi.org/10.20923/kjas.2016.50.177>
- 박용숙 (2016). 한국에서의 잊혀질 권리에 관한 소고. *江原法學*, 49, 49-101.  
<https://doi.org/10.18215/kwlr.2016.49..49>
- 박진우 (2014). 이른바 '잊힐 권리'에 관한 헌법적 조명. *세계헌법연구*, 20(2), 155-183.
- 백수원 (2015). EU사법재판소의 '잊혀질 권리' 관련 판결에 따른 과제: 유럽과 미국의 논의를 겸하여. *미국헌법연구*, 26(2), 125-161.
- 서승진 (2020.4.1). 속초시 '코로나 확진자 다녀간 곳' 주홍글씨 지운다. *국민일보*,  
출처: <http://news.kmib.co.kr/article/view.asp?arcid=0014430585&code=61121111&cp=nv>
- 윤지원 (2020.3.24). 법무부 "N번방 가해자 끝까지 추적...범죄단체 조직죄도 가능." *경향신문*,  
출처: <https://www.khan.co.kr/national/court-law/article/202003241628001>
- 이영진 (2013). "잊혀질 권리" vs "알 권리". *言論仲裁*, 127, 44-49.
- 이중화, 이현규 (2015). Data Dictionary 기반의 R Programming을 통한 비정형 Text Mining Algorithm 연구. *한국산업정보학회논문지*, 20(2), 113-124. <https://doi.org/10.9723/jksis.2015.20.2.113>
- 장재은 (2019.11.28). 독일 최고법원 "살인자에게도 '잊힐 권리' 있다" 판결. *연합뉴스*,  
출처: <https://www.yna.co.kr/view/AKR20191128115900009>
- 정재호 (2014). 언론 현장에서 바라본 잊혀질 권리. *言論仲裁*, 132, 22-27.
- 최경진 (2012). 잊혀질 권리-개인정보 관점에서. *정보법학*, 16(2), 97-120.
- 최중선 (2016). 잊힐 권리의 법적논점에 관한 연구: 인터넷 자기게시물 접근배제요청권 가이드라인을 중심으로. *미디어와 인격권*, 2(1), 63-113.
- 최태범 (2020.6.10). 코로나 '잊혀질 권리'...14일 지난 동선정보 지운다. *머니투데이*,  
출처: <https://news.mt.co.kr/mtview.php?no=2020061011362539954>
- 한국정보화진흥원 (2015). 2014년 정보문화실태조사.
- 한지숙 (2020.6.11). 확진자 잊혀질 권리 찾아 준 '송파구 인터넷방역단', 전국으로 확산. *헤럴드경제*,  
출처: <http://news.heraldcorp.com/view.php?ud=20200611000155>
- 홍숙영 (2014). 기사삭제 요구권과 잊혀질 권리-기사삭제의 인정기준에 관한 논의. *디지털융복합연구*, 12(12), 13-22. <https://doi.org/10.14400/JDC.2014.12.12.13>
- 황창근 (2016). 잊혀질 권리의 국내 적용과 법제화의 한계. *홍익법학*, 17(1), 299-327.

- <https://doi.org/10.16960/jhr.17.1.201602.299>
- Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84  
<https://doi.org/10.1145/2133806.2133826>
- Born, J., Scheihing, E., Guerra, J., & Cárcamo, L. (2014). Analysing Microblogs of Middle and High School Students Participating in Kelluwen. *European Conference on Technology Enhanced Learning 2014*. [https://doi.org/10.1007/978-3-319-11200-8\\_2](https://doi.org/10.1007/978-3-319-11200-8_2)
- Puhg, M. J. (1992). *Providing Reference Services for Archives & Manuscripts*. 설문원 역(2004). 기록정보서비스, 서울: 진리탐구.
- Steyvers, M. & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch eds, *Handbook of latent semantic analysis: A road to meaning*. New Jersey: Lawrence Erlbaum Associates, 427-448.
- Wickham, H. (2014). “Tidy Data.” *Journal of Statistical Software*, 59(1), 1-23.  
<https://doi.org/10.18637/jss.v059.i10>

• 국문 참고문헌에 대한 영문 표기  
(English translation of references written in Korean)

- Baek, Soo-Won (2015). Review on the CJEU Judgment related to ‘Right to be Forgotten’: With the Discussions of Europe and U.S.. *Study on The American Constitution*, 26(2), 125-161.
- Choi, Jong-Sun (2016). A study on the legal issues of the right to be forgotten: a focus on the exclusion request access rights of internet self-post guidelines. *Journal of Media and Defamation Law*, 2(1), 63-113.
- Choi, Kyoung-Jin (2012). Right to be forgotten-a personal information perspective. *Journal of Korea Information Law*, 16(2), 97-120.
- Chung, Chae-Ho (2014). The right to be forgotten from the press. *Press Arbitration*, 132, 22-27.
- Gim, Gab-Seok (2017). The development of forgotten rights in Europe and legislation of legal system in Korea. *European Constitution*, 25, 53-78.
- Hong, Sook-Yeong (2014). Right to delete and right to be forgotten -discuss on the condition of the right to delete. *Journal of Digital Convergence*, 12(12), 13-22.  
<https://doi.org/10.14400/JDC.2014.12.12.13>
- Hwang, Chang-Geun (2016). The Limitation of domestic application and legislation of the right to be forgotten. *Journal of Hongik Law Review*, 17(1), 299-327.  
<https://doi.org/10.16960/jhr.17.1.201602.299>

- Kim, Gap-Cheol & Noh, Hyun-Jong (2019). Research trends of regional geography education using topic modeling. *Social Studies Education*, 58(4), 49-68.  
<https://doi.org/10.37561/sse.2019.12.58.4.49>
- Kim, Na-Roo (2018). The latest legislation trends and legal issues on the right to be forgotten in the era of big data. *The Journal of Comparative Law*, 18(1), 7-43.
- Ko, Eun-Byul, Choi, Maria, & Choi, Kwang-Hee (2012). An analysis on the forgotten rights of personal information protection. *The Korea Society of Management Information Systems International Conference*, 460-464.
- Lee, Jong-Hwa & Lee, Hyun-Kkyu (2015). A study on unstructured text mining algorithm through R programming based on data dictionary. *Journal of the Korea Society Industrial Information System*, 20(2), 113-124. <https://doi.org/10.9723/jksis.2015.20.2.113>
- Lee, Young-Jin (2013). "Right to be Forgotten" vs "Right to Know". *Press Arbitration*, 127, 44-49.
- Nam, Yeun-Ah & Yim, Jin-Hee (2016). A qualitative study on the procedures and characteristics of reading service provision at center of archival information of national archives of Korea. *The Korean Journal of Archival Studies*, 50, 177-229.  
<https://doi.org/10.20923/kjas.2016.50.177>
- National Information Society Agency (2015). An investigation on Information culture 2014.
- Park, Jin-Woo (2014). The constitutional review on the so-called right to be forgotten. *World Constitutional Law Review*, 20(2), 155-183.
- Park, Yong-Sook (2016). A study on the right to be forgotten in Korea with a focus on checking the current discussions and exploring future directions. *Kangwon Law Review*, 49, 49-101.  
<https://doi.org/10.18215/kwlr.2016.49..49>