

사법분야 인공지능 발전을 위한 판결문 데이터 개선방안

- ‘판결서 인터넷열람 서비스’를 중심으로 -

박성미* · 이유나** · 최아리** · 안정민***

〈 목 차 〉

- I. 서론
- II. 전자정보로서의 판결문의 중요성
- III. 국내외 판결문 데이터 수집 및 공개 현황
- IV. ‘판결서 인터넷열람 서비스’ 판결문 생성과정 분석
- V. 인공지능 발전을 위한 ‘판결서 인터넷열람 서비스’ 개선방안
- VI. 결론

I. 서 론

공공기관들의 문서 작성이 전자화되고 국민의 공공데이터 이용권에 관한 관심이 높아지면서 법원에서 생성하는 판결문데이터 역시 공개대상으로 조명받고 있다. 민·형사 판결문¹⁾은 사회적 분쟁에 관한 판단을 내림으로써 사회적 갈등을 해결하고 공동체로서 국가가 지향하는 미래의 지표 역할을 수행하며, 법의 분석에 대해 학계에 중요한 화두를 던지기도 한다.²⁾ 이처럼 판결문은 재판에서 핵심적인 역할을 수행하는 문서로

* 고려대학교 정보보호대학원 박사과정, 주저자

** 한림대학교 정보법학과 석사과정

*** 한림대학교 정보법학과 교수, 교신저자

**** 이 논문은 2021년도 정부(경찰청)의 재원으로 지원받아 수행된 연구결과임 [내역사업명: AI 기반 범죄수사 지원 / 연구개발과제번호: PR10-02-000-21]

1) 판결문이란 확정된 판결의 내용을 서면으로 작성한 문서를 뜻하며 판결문(예: 대법원 행정예규 전자우편 등을 통한 판결문 제공에 대한 예규) 또는 판결서 (예: 판결서 인터넷열람 제도) 두 가지 표현을 혼용사용이 가능하다.

국민의 법적 이해와 권리 보장에 큰 영향을 끼치는 만큼 판결문에 대한 접근성을 보장하는 것은 투명한 법률 문화 정착의 시금석이라고 볼 수 있다. 그뿐만 아니라 최근 해외에서 급진적인 성장을 이루고 있는 리걸테크(Legal Tech) 시장과 같이 법률 업무 자동화 및 지원 산업분야에서 판결문 데이터 활용은 최대의 관심사이자 해결되어야 할 과제이다.³⁾

현재 우리나라에서는 국민의 알권리 충족, 재판 공개의 원칙을 이유로 확정된 사건의 판결서 등(일부 제한)을 ‘누구든지’ 열람 또는 복사하도록 하고 있으며⁴⁾, 전자우편, 법원도서관 특별열람실 방문 또는 확정 판결서 인터넷열람 등 다양한 방식으로 판결문을 공개하고 있다. 이러한 원칙과 노력에도 불구하고 2019년 1월부터 지금까지 시행되고 있는 인터넷열람제는 열람 1건당 1,000원의 수수료를 부과하면서도 제한적인 임의어 검색과 텍스트 검색이 불가능한 이미지 형태의 PDF(Image-only PDF)파일로 제공되어 학술적 연구를 위해서만 아니라 일반 국민에 대한 편리성에 있어서도 큰 걸림돌로 작용하고 있다.

2021년 7월 5일 「전자우편 등을 통한 판결문 제공에 관한 예규」⁵⁾의 개정을 통해 기계 판독이 가능한 판결서가 제공되기 시작하였으나 이전의 판결문에 대해서는 소급해서 적용되지 않아 현재 90% 이상의 공개된 판결문 데이터는 여전히 이미지 PDF의 형태로 제공된다. 민사사건의 경우 2023년부터 미확정 판결문도 공개하는 등 범위의 확대에 대한 긍정적인 논의가 있지만 형사사건의 공개범위에 대해서는 여전히 논란이 있다.⁶⁾ 판결문 공개에 관한 논의는 꾸준히 지속되어 왔고, 공개 시스템 성능 개선에 대해서는 이탄희 의원(의안번호: 2103299), 박주민 의원(의안번호: 2107718)의 형사

2) 양시훈, “판결에 대한 국민의 이해를 높이고 이를 알릴 수 있는 방안에 관한 연구”, 사법정책연구원, 2014, 28면.
3) 심재훈 변호사 “판결문 공개해 리걸테크 발전 기반 삼아야”, iT조선, 2020.06.17. 기사 (http://it.chosun.com/site/data/html_dir/2020/06/16/2020061603849.html, 2021.08.29 최종방문)
4) 형사소송법 제59조의 3 (형사판결서), 민사소송법 제163조의 2 (민사판결서)
5) 전자우편 등을 통한 판결문 제공에 관한 예규 제6조 (판결문의 제공방법) ① 법원행정처 등은 제공 대상 판결문 내용 중 이름·주민등록번호·주소 등 소송관계인의 개인정보에 관한 사항을 판결서 등의 열람 및 복사를 위한 비실명 처리 기준(재일 2014-2)에 따라 비실명 처리한 후 당해 사본을 전자우편·우편·모사전송 또는 직접제공의 방법 중 신청인이 원하는 제공방법에 따라 제공한다. 다만 전자우편의 방법에 의하여 판결문을 제공할 경우에는 판결문에 기재된 문자열 또는 숫자열이 검색어로 가능할 수 있는 파일로 변환하여 제공하여야 한다. [행정예규 제 1276호, 시행 2021.7.5.]
6) “미확정 판결서 공개…어디까지 왔나”, 법률신문, 2020.12.03. 기사(<https://www.lawtimes.co.kr/Legal-News/Legal-News-View?serial=166111>, 2021.9.10. 최종방문)

소송법 일부개정법률안에도 언급된 바 있다. 그러나 정작 국내에서 공개된 판결문의 건수에 대한 통계나 공공기관의 판결문 공개 시스템 구조 분석, 데이터 수집의 관점에서 구체적으로 어떤 기술적 한계가 있는지에 대한 연구는 많이 부족한 상황이다.

사법 데이터로서 가치 있는 판결문 제공이란 단순한 공개 이상의 것을 의미한다. 인공지능이나 기계학습의 사법분야로의 확대는 활용가능한 판결문 데이터를 필수요소로 한다. 이러한 관점에서 본 연구는 국내외 리걸테크 시장의 발전과 사례를 통하여 어떻게 판결문이 사법분야 인공지능 개발에 기여하는지 확인하고, 데이터로서 하급심(1·2심) 판결의 중요성과 공개 필요성을 살펴보았다. 이를 위해 현재 우리나라에 공개되어 있는 판결문 건수, 범위 및 형태 등을 미국·독일의 공공·민간 판결문 공개 사이트와 - 그 중에서도 특히 그 데이터의 인공지능 활용의 면에 중점을 두고 - 비교·분석하였다. 마지막으로 이러한 비교분석의 결과를 근거로 대법원에서 현재 운영되고 있는 '판결서 인터넷열람 서비스' 구조와 기능을 분석하고, 향후 인공지능 연구 활성화를 위한 개선사항을 결론으로 제시하고자 한다.

II. 전자정보로서의 판결문의 중요성

1. 판결문 기반 인공지능 연구 동향

리걸테크 또는 법정보기술 산업은 1873년 미국 법률 출판사의 직원이었던 Frank Shepard가 판례의 태도를 확인할 수 있는 인용 색인(Shepard's Citation)을 제작해 판매했던 것을 시작으로 한다.⁷⁾ 1970년대 초반 미국의 Lexis Nexis가 컴퓨터 보조를 통한 법률 연구 서비스⁸⁾를 선보이면서 본격적인 시장이 형성되었다. 2010년에는 미국 스탠포드 로스쿨과 컴퓨터공학과가 협력하여 법률 사무를 위한 기술 개발·연구 및 인재 양성을 위해 '코드엑스(CodeX) 프로젝트를 시작하였다.⁹⁾ 리걸테크 산업은 머신러

7) Esther Salmerón-Manzano, "Legaltech and Lawtech: Global Perspectives, Challenges, and Opportunities", Laws, 제 10권 제2호, 2021, 1면.

8) 법률 회사와 미국 도서관의 법률 데이터베이스를 연결하여 Ohio와 New York 주의 판례검색 서비스를 제공하였으며 1970년대 후반에는 전화접속(telephone dialer)을 통해 데이터베이스 접근이 가능한 서비스를 구축하였다.

9) 코드엑스 스탠포드 법정보학 연구소(CodeX Stanford Center for Legal Informatics)는 법률 분석 자동화

닝·딥러닝·자연어처리 기술 등의 발전에 따라 단순한 법령 검색기능을 넘어 법률 분석·재판 결과 예측·법률문서 작성 등의 분야까지 확장되게 된다.¹⁰⁾ 2019년을 기준으로 리걸테크 스타트업 투자금액은 영국 1억 1500만 달러(약 1300억 원), 미국은 19억6000만 달러(약 2조 2171억 원)에 달하였고,¹¹⁾ 법률서비스 소프트웨어 시장의 투자금액은 2015년 기준 약 38억 2,800만 달러(약 4조 4942억 원)에서 2019 기준 57억 6,300만 달러(약 6조 8158억 원)로 예상되어 높은 성장세를 보이고 있다.¹²⁾

국내 리걸테크 산업 또한 해외의 리걸테크 시장 확장에 힘입어 누적 투자 규모가 1200만 달러(약 132억 원)에 달하는 등¹³⁾ 인공지능 기술을 법률 데이터와 결합하는 서비스를 제공하고 있다. 국내외 법률 분야 인공지능 및 빅데이터 연구 중 대표적인 유형을 분석한 <표1>에서 국내 리걸테크 기업 역시 유사 판례 검색, 법적 검토 자동화, 문서 작성 자동화 등 해외 기업과 유사한 유형의 서비스를 제공하고 있음을 볼 수 있다. 특히 로톡(Law Talk)의 형량예측기는 1심 형사판결문 40만 건을 기반으로 인공지능 모델을 학습하였으며¹⁴⁾ 인텔리콘社의 유렉스(U-Lex) 인공지능 법령 검토 시스템은 300만 건의 판결문 및 법령 데이터를 딥러닝으로 학습하였다고 한다.¹⁵⁾ 스타트업인 엘박스(구 리걸텍)社의 엘박스(L-box)는 판결문 학습을 통하여 검색 정확도 및 효율성을 높이고 2021년 하반기까지는 100만 건의 판결문 데이터를 수집할 계획이라고 밝히고 있어 법률 분야에의 인공지능 기술 접목이 구현되고 있음을 할 수 있다.¹⁶⁾

(Computational Law), 블록체인 기반 법률 시스템 등 다양한 프로젝트를 진행하고 있으며 CaseText, Ravel 등 기존의 리걸테크 기업뿐만 아니라 해외의 연구기관과 협업을 통해 법률 분야와 기술을 융합하는 연구 분야를 이끌어가고 있다. (참조: <https://law.stanford.edu/codex-the-stanford-center-for-legal-informatics/codex-projects/>)

- 10) 이성진·이연주, “4차 산업혁명시대 법정기술의 현황과 발전방안”, 정보화정책, 제28권 제1호, 한국지능정보사회진흥원, 2021, 4면.
- 11) “시로 개인정보 거른 판결문 공개 필요...‘리걸테크’ 유니콘 탄생 첫걸음”, 뉴스웍스, 2020. 12. 8. 기사 (<https://www.newsworks.co.kr/news/articleView.html?idxno=506027>, 2021. 9. 10. 최종방문).
- 12) 현대경제연구원. “리걸테크(Legaltech) 산업 현황과 시사점”, 현대경제연구원 VIP Report 16-31(통권 669호), 2016, 4면.
- 13) “IT만남 법률...변호사 선임 무턱 낮추고 AI가 형량 예측”, 세계일보, 2021. 5. 2. 기사(<https://m.segye.com/view/20210502508132>, 2021. 9. 10. 최종방문).
- 14) 홈페이지 주소 (<https://www.lawtalk.co.kr/sentence-estimation>, 2021. 9. 10. 최종방문).
- 15) 홈페이지 주소 (<https://www.ulex.co.kr/>, 2021. 9. 10. 최종방문).
- 16) “서초동에 부는 AI 바람...리걸테크가 온다”, 중앙일보, 2020. 5. 30. 기사(<https://www.joongang.co.kr/>

〈표 1〉 국내외 판결문 분석 기반 리걸테크 유형 및 현황

분야	기능	상세설명 및 예시 (프로그램 또는 서비스명)
의사결정지원	법률 문서 검토	시에 판결문 내용을 학습시켜 자체적으로 법률문서를 검토 (주요 키워드 추출 및 위험요소 분석) 예시) KIRA 시스템(미국), eBrevia(미국), 알파로(한국)
	형량 예측	1심 형사 판결문 기반 형량 예측 지원 예시) 로톡 AI(한국)
	스마트 검색	사건사실, 쟁점(legal issue) 등 다양한 사용자 자료 기반 검색을 진행하여 관련 판례의 범위를 축소·기각 표시·관련 주요 판결 확인 등 사용자의 목적에 맞는 검색 가능 예시) FastCase(미국), CaseText(미국), 유렉스(한국), 엘박스(한국)
법적분석지원	법률 문서 작성	법률 전문가의 지식 기반으로 구성된 쟁점 별 법률문서 작성 지원 및 관련 판례 추천 예시) Compose(미국), 로폼(한국)
	사건 승률 분석	승소 확률, 적정 합의금, 재판 담당 판사의 과거 판결 분석 예시) Lex Machina(미국), Gavelytics(미국)
	검색 시각화	관련 판례 맵 생성 및 주요 판결 강조 예시) Ravel Law(미국), 엘박스(한국)

국내 민간분야의 리걸테크 산업은 외국에 비해 시장규모가 작고 법적 규제나 법조인들의 반발¹⁷⁾로 그 성장이 제한적으로도 보이나, 정부는 리걸테크 시장 활성화 및 법조영역으로의 인공지능 도입에 적극적인 태도를 보이고 있다. 특히 대법원에서 2018년에 발표한 ‘2024년 스마트법원 4.0’은 차세대 전자소송 시스템 구축사업으로 빅데이터 기반의 인공지능시스템이 소송기록을 분석하여 자동으로 쟁점 문장을 추출해주는 ‘지능형 쟁점추출’, 법원·사건명과 같은 메타데이터뿐만 아니라 주문과 판단 이유 등 실질적 내용을 담고 있는 판결문 초고를 제공하는 ‘지능형 판결문 작성’ 지원을 계획하고 있다.¹⁸⁾ 판결문, 소장, 답변서 등 기존 소송자료를 학습데이터로 사용하여 유사한 판례

article/23789440#home)(2021. 9. 10. 최종방문).

17) “로톡’사태는 시작일뿐... 다른 법률 서비스 스타트업도 고사위기”, 이데일리, 2021. 9. 8. 기사(<https://www.edaily.co.kr/news/read?newsId=01108646629178152&mediaCodeNo=257&OutLnkChk=Y>, 2021. 09. 10. 최종방문).

18) 단, 판결문 주문 작성에는 기술적인 어려움이 클 것이라고 보고 있는 추세다. “AI가 판결문까지 작성?...법원은 ‘신중 모드’”, 문화일보, 2019. 4. 9. 기사(<http://m.munhwa.com/mnews/view.html?no=2019040901032003009001>, 2021. 09. 10. 최종방문).

를 법관에게 제공하고 전문가가 분석한 판결의 쟁점을 지도학습을 통해 자동 추출, 중요내용을 자동으로 요약하여 판결문 작성에도 기여하겠다는 의도로 해석된다.

대법원은 2024년 서비스 개시를 목표로 이 시스템 구축에 약 2500억 원을 투자할 예정이며 이를 통해 사법 정보공개체계의 투명성과 효율적인 재판 환경을 보장하여 국민의 신뢰 및 접근성을 높여 줄 것으로 기대하고 있다. 법률 분야의 인공지능은 단순한 검색 시스템이 아니라 판결 및 소송기록을 분석하고 의사결정과정을 지원함으로써 법치주의의 근간인 예측가능성을 제공하고 법적 안정성 제고에 기여한다. 법의 해석은 결국 다양하고 많은 사례에 적용되고 축척되는 과정을 통해 확립된다. 때문에 판결문에 정립된 다양한 사실관계와 판단은 무엇보다 활용도가 높은 의미있는 데이터이며 법조 분야의 인공지능 도입을 위해서 필수불가결한 데이터인 것이다.

2. 하급심 판결문 데이터 확보의 중요성

법원의 판결은 사회적 갈등이 발생할 때 그 갈등을 실정법에 따라 해석하고 사회적 합의를 이끌어내는 작용을 한다. 우리나라는 원심과 항소심에서 사실관계를 판단하고 인정되는 사실관계에 의해 법리판단을, 최종심에서는 법적 판단을 그 대상으로 한다. 인공지능의 도입이 전사회적으로 확산되는 시점에서 사법 분야 인공지능 개발의 관점에서 데이터분석은 사실관계에 대한 분쟁으로부터 시작하므로 하급심의 판결은 데이터 축척과 분석에 있어서 최종심보다 더 중요하고 더 큰 가치를 가진다. 실제로 전문가들은 사례를 기반으로 유사 사건을 다루거나 분석하는 인공지능은 직접적 사실관계를 다루는 하급심이 필요하고¹⁹⁾, 쟁점 하나당 하급심 판결문 1,000건 가량이 필요하다²⁰⁾고 보는 등 하급심 확보 및 공개의 중요성을 강조하고 있다.

리걸테크 시장 관련자들은 국내 리걸테크 산업의 느린 성장요인으로도 ‘판결문 데이터’ 확보의 어려움을 꼽았으며, 상급심 위주의 제한된 판례 공개의 활용성은 낮다고 보고 있다.²¹⁾ 리걸테크의 선도국으로 불리는 미국은 연방법원 판결문을 즉시 공개하고, 주법원(State court) 판결문 역시 대부분 비식별화 없이 공개하고 있는 것에 비하여

19) “주목받는 리걸테크 산업”, 법률신문뉴스, 2021. 4. 26. 기사(<https://www.lawtimes.co.kr/Legal-News/Legal-News-View?serial=169586>, 2021. 8. 27. 최종방문).

20) 전정현·김병필, “인공지능과 법률 서비스 현황과 과제”, 저스티스, 제170권 제1호, 2019, 218-258면.

21) 앞의 기사(주5) 참조.

국내 판결문 데이터는 비교적 쉽게 접근할 수 있는 대법원 판결문조차 전체 대법원 판결의 약 3%가량만 공개되고 있다. 더욱이 하급심의 경우 약 0.003%의 수준만 공개되고 있어²²⁾ 법조계의 사건검토와 법리연구에 지장을 주고, 판결문 입수에 있어 인맥이 동원되어야 하는 등의 전관예우가 성행하여 지속적으로 문제가 되고 있는 현황이다.²³⁾

많은 데이터학습을 요구하는 사법 분야 인공지능 활성화와 재판의 투명성을 위해 판결문의 공개범위는 확대될 필요가 있다. 사회적으로 사법부의 역할이 증대되고 판례의 중요성이 높아짐에 따라 대법원뿐만 아니라 하급심의 판결문도 공개함으로써 데이터상의 불균형도 해소되어야 하는 문제다. 일반인의 경우 인터넷 등을 통해 하급심 판결결과를 찾아보고 이에 따라 소송 여부를 결정하는 경우가 많고,²⁴⁾ 판결문 검색 사이트를 활용하는 사람의 대부분은 추상적인 법리적 검토보다는 본인 사건과 유사한 범죄사실을 기준 삼아 유사사건을 찾거나 관련된 형량이나 판단 기준의 검색을 통해서 재판에서 충분한 방어권을 행사하고자 하는 사례가 증가하고 있다는 점을 고려해보면 하급심 판결의 공개필요성은 더욱 절실하다.

특히 인공지능 학습 데이터로 사용될 하급심은 1심과 2심의 관계성을 포함하여 수집되어야 활용도를 가진다. <표 2>에서 볼 수 있듯이 1심과 2심은 핵심적으로 다루어지는 내용의 차이도 존재하고, 1심에서 인정되었던 사실관계가 2심에서는 인정되지 않거나, 2심에서는 다뤄지지 않은 쟁점이 1심에는 피고인의 핵심주장인 경우도 있어 실제로 하나의 형사사건을 분석하기 위해서도 최종 판결뿐만 아니라 관련 판례들의 핵심 쟁점을 비교·분석하게 된다. 하급심 데이터는 이처럼 그 내용 자체뿐만 아니라 따름 판례 등 관련 판례와의 비교·분석의 연결고리이기 때문에 인공지능 적용을 위한 귀중한 자료이다.

22) 이탄희 의원 대표 발의 형사소송법 일부개정법률안(2020. 08. 26. 의안 제2103299호).

23) “판결문 공개 과감히 확대하라”, 법률신문, 2019. 10. 28. 기사(<https://m.lawtimes.co.kr/Content/Article?serial=156740>, 2021. 8. 27. 최종방문); “전국 법관 대표 ‘하급심 판결문 공개 확대해야’”, 이투데이, 2021. 1. 18.. 기사(<https://www.etoday.co.kr/news/view/1985806>, 2021. 8. 27. 최종방문)

24) 앞의 기사(주12) 참조.

〈표 2〉 국내 판결문 목차·내용 비교분석

항목	상세설명	1심 판결문 목차 및 내용	2심 판결문 목차 및 내용
메타 데이터	선고일, 사건번호 등 정보	○	○
이유	실제 판결내용으로 범죄유형, 범죄사실에 따라 약간의 차이가 있으나 대체로 유사한 순서로 이어짐	<ul style="list-style-type: none"> • 범죄사실 - 주요 범죄사실을 시간순으로 작성 	<ul style="list-style-type: none"> • 항소이유 요지 - 사실오인, 양형부당, 법리오해 등으로 분류되어 해당 항소이유와 관련된 사실관계 및 법리적 검토를 다룸
		<ul style="list-style-type: none"> • 증거의 요지 - 판결의 증거 목록 요약본 	<ul style="list-style-type: none"> • 판단 - 원심을 재판단하기 위해 쟁점 중심 사실관계 또는 법리적 해석 확인
		<ul style="list-style-type: none"> • 법령의 적용 - 판결에 적용된 법령 	<ul style="list-style-type: none"> • 결론 - 이유가 있다고 볼 때 원심파기, 이유가 없다고 볼 때 항소 기각
		<ul style="list-style-type: none"> • 피고인 주장에 대한 판단 - 주로 피고인의 주장 요지 중심으로 피고인의 주장과 사실관계를 판단 	<ul style="list-style-type: none"> • 다시 쓰는 판결 - 원심의 판결과 같은 부분을 그대로 인용, (원심파기 등의 이유로) 다시 써야 하는 부분만 수정
		<ul style="list-style-type: none"> • 양형이유 - 형량에 영향을 끼치는 유리한 또는 불리한 정황 	

Ⅲ. 국내외 판결문 데이터 수집 및 공개 현황

국내에서 인터넷을 통해 판결문을 수집하는 방법은 (1) 법원에서 운영하는 공식적인 열람서비스 사용과 (2) 민간분야에서 운영하는 사이트를 통해 제공하는 판결문을 검색 및 다운로드하는 방법이 있다. 이하에서는 판결문 열람이 가능한 국내외 사이트의 검색기능 및 특징 등을 분석하고 종합적으로 국내 법원·민간분야의 판결문 검색 사이트와 미국·독일 법원 사이트의 공개 데이터양과 제공되는 파일 형태 등을 비교하여 국내 법원에서 운영하고 있는 판결서 인터넷열람 서비스의 데이터 제공방식과 그 효용성을 알아보고자 한다.

1. 미국

미국의 경우 원칙적으로 연방법원 판결문은 선고 후 24시간 이내 공개되며, 뉴욕주나 캘리포니아주와 같은 일부 주법원에서도 판결문을 제공한다.²⁵⁾ 무엇보다 미국 법원은 국민의 알 권리와 언론의 자유, 법원의 투명성을 보장하기 위해 대부분 비실명화 작업을 거치지 않고 당사자의 이름 등 정보가 담긴 판결문을 공개할 뿐만 (개인 신상 정보는 일부 비식별화) 아니라 모든 법정기록을 통상적으로 전부 공개한다는 특징이 있다.

그 외에도 정부에서 운영하는 유료 법률 시스템인 PACER(Public Access to Court Electronic Records)를 통해 사건검색이 가능하다. PACER는 200개 이상의 연방법원의 사건 정보를 온라인으로 검색할 수 있는 연방 사법부의 시스템이다. 이 시스템은 사건 정보(사건번호, 사건당사자 등), 소송 절차 및 제출된 문서를 포함하는 기록 문서, 판단, 사건 관련 정보 검색, 사건 현황에 대한 정보, 법원 청문회 음성 파일 등을 제공하는 메타데이터 관련 데이터베이스를 유지하고 관리한다. PACER에서 요구하는 가격은 페이지 당 0.10달러 (약 117원)²⁶⁾부터 시작하여 서류(docket, motion, order, judgment, brief in a case 등)의 경우는 3달러(약 3,500원)까지 상한을 두고 있다. 서류를 제외한 PACER 검색결과, 사건보고서(report), 법정 소송 기록(transcripts of court proceedings)은 가격의 상한을 두지 않는다. MP3 형태로 제공되는 법원 기록 음성 파일은 파일 당 2.40달러(약 2,800원)이다.

검색 후 다운로드한 문서는 기본적으로 기계 판독 가능한 PDF²⁷⁾ 또는 HTML²⁸⁾ 형태로 제공된다. 또한, PACER는 검색 자동화하는 프로그램을 만드는 사용자를 위해 별도로 인증 API (Application Programming Interface, 애플리케이션 프로그래밍 인터페이스)도 제공하고 있다. 사용자는 PACER 사건 검색(PACER Case Locator, PCL) API를 통해 연방판결 또는 관련 당사자에 대한 색인을 검색할 수 있으며, PACER 내의 사건 정보검색, 고급검색에 대한 결과를 JSON²⁹⁾, XML³⁰⁾로 받을 수 있

25) 양시훈, 위의 연구보고서(주3), 85-87면.

26) HTML 파일일 경우 4,320 bytes가 1 페이지로 산정된다.

27) Portable Document Format, Adobe에서 만든 문서 파일로 문서의 텍스트를 선택, 복사 등이 가능하며 문서 보안성이 높다.

28) Hyper Text Markup Language, 웹문서를 작성할 때 사용하는 인터넷 문서파일이다.

다. 이런 API는 간단한 쿼리문을 통해 구조화(페이지 정보, 사건번호, 피고인 명 등 중요한 메타데이터가 추출하기 쉽게 표현)된 데이터 대량수집을 가능하게 할 뿐만 아니라 일반 텍스트(Plain text)보다 활용이 쉬워, 개발과 데이터 분석의 문턱을 낮추는 역할을 한다. 또한, PACER는 학술 연구자 및 비영리 단체, 취약계층, 무료(국선) 변호사 등에 해당하는 개인 또는 집단일 경우, 법원에 요청하여 위의 사용료를 내지 않는 면제 제도를 운영하고 있다. 면제 대상에 공공기관이나 언론사 등은 포함되지 않고, 개인 연구자 역시 연구과제가 오로지 학술적인 목적임을 명확히 해야 한다는 제한이 있으나, 면제 요청 서류 작성이 매우 간편하다.³¹⁾ 이와 같이 미국은 정부 차원에서 법률 데이터의 연구적 사용을 적극적으로 권장하고 데이터 제공 환경 또한 그런 정책에 맞춰 발전하고 있는 것으로 평가된다.

2. 독일

판례가 법원(法源)이 되는 영미법계 국가보다 대륙법계 국가는 대체로 중요하다고 판단한 판결만 공개하는 성향이 있다.³²⁾ 독일은 현재까지 판결문 공개는 각 법원의 재량에 맡기는 상황이다. 1992년 특허법원의 판결³³⁾ 이후로, 2017년 연방대법원 민사 제4부는 판결에서 ‘판결의 내용은 - 대부분 절차와 마찬가지로 - 공개된 것으로, ... 재판의 당사자는 본인에 관한 결정이 공개되지 않을 것이라고 근본적으로 배제할 수 없다.’³⁴⁾ 라고 다시 한번 판결문 공개에 비교적 적극적인 모습을 보였다. 한편, 연방대법

29) JavaScript Object Notation, 문자 기반의 데이터 포맷으로 최소한의 정보만을 가져 웹에서 데이터를 전송할 때 자주 사용된다.

30) Extensible Markup Language, 확장성 마크업 언어로 문서를 구조적으로 저장하여 문서를 재가공할 때 용이하다.

31) PACER 홈페이지의 연구자 사용료 면제 요청(<https://pacer.uscourts.gov/my-account-billing/billing/fee-exemption-request-researchers>) 페이지에 링크되어있는 PDF 문서에서 (1) 데이터를 사용할 법원, (2) 소속기관, (3) PACER 데이터 사용처 요약, (4) 사용 기간 등을 작성한 후 전송을 클릭하면 요청이 완료된다.

32) 백상준, “판결서 인터넷열람 제도의 개선현황과 향후과제”, 국회입법조사처, 2019

33) ‘법원은 적당한 방법(angemessener Weise)으로 판결문을 공개할 의무’를 지니며, 해당 판례에서 쟁점이 되었던 등록 상표 관련 판결은 일반 대중에게도 정보 가치가 있으므로, 공개되지 않았으나 비식별화처리된 판결을 요청할 경우 근본적으로 접근 가능해야 한다고 보았다. BPatG, 27 ZA (pat) 19/90, Beschluss vom 23. April 1991 (연방특허법원 제27부, 1991년 4월 23일 판결).

34) BGH, IV AR(VZ) 2/16, Beschluss vom 5. April 2017 (연방대법원 민사제4부, 2017년 4월 5일 판결);

원의 형사 제5부는 2018년 본인의 홈페이지 및 언론에 공개할 목적으로 이해관계가 없는 사인인 제3자가 형사판결문 제공을 요구하자 그 요청을 거부하여 판결문의 공개가 무조건적 권리가 아님을 강조하였다.³⁵⁾

연방대법원은 2000년 1월 1일부터 민·형사부 판결문을 대법원 홈페이지에서 글자 검색이 되는 PDF 형태로 제공하고 있으며, 베를린(Berlin) 주의 경우, 2021년 1월 1일을 기점으로 무료로 베를린 법원 판결문을 글자 검색 가능 PDF로 제공하고 있다. 바덴뷔르템베르크(Baden-Württemberg) 주는 2006년 1월 1일 기점으로 (비영리 목적일 경우) 법원 판결문을 HTML 형태로 제공하고 있고, 바이에른(Bayern) 주의 경우 주요 판결문을 무료로 공개하고 있으며 PDF, RTF, XML 형식을 지원한다. 헤센(Hessen)주에서는 각 법원에서 별도로 운영 중이던 판례 데이터베이스를 주(州)법원 판례 데이터은행(Landes rechtsprechungsdatenbank, LaReDa)³⁶⁾으로 통합시켜 운영 중이다. 헤센 주의 데이터은행은 판결문을 기계 판독이 가능한 PDF 또는 XML 형식(등록된 사용자만 해당)으로 무료로 제공하고 있으나, 상업적인 목적일 경우 건당 2.50 유로(약 3,450원)를 부과하고 있다.

위의 사례를 보아 독일은 미국처럼 공개를 기본원칙으로 하기보다는 해당 판결이 가지고 있는 정보의 가치, 보호해야 할 법익의 균형 등을 고려하여 선별적 공개를 계속할 것으로 보인다. 연방대법원과 각 주의 법원들이 판결문을 이미 상당수 공개 및 기계 판독 가능한 파일 형식으로 제공하고 있거나 효율적인 판결문 관리 및 공개를 위해 새로운 통합시스템을 구현하고 있다는 점을 고려할 때, 독일 역시 판결문 데이터 제공 및 활용에 긍정적인 방향으로 나아가고 있다고 평가할 수 있겠다.

Jessika Kallenbach, "Kein Anspruch auf Herausgabe eines Strafurteils an private Dritte [형사판결을 개인인 제3자에게 공개해야 할 권리는 없다]", *Anwaltsblatt*, 2018.09.11., (<https://anwaltsblatt.anwaltverein.de/de/anwaeltinnen-anwaelte/berufsrecht/bgh-anonymisierte-urteilsabschriften>) (2021.09.09. 최종접속)

35) BGH, V AR(Vs) 112/17, Beschluss vom 20. Juni 2018 (연방대법원 형사제5부, 2018년 6월 20일 판결); Jessika Kallenbach, 위의 기사(주34)에서 재검색.

36) "Bürgerservice Hessenrecht - Landesrechtsprechungsdatenbank(LaReDa) [헤센법 민원서비스 - 주(Land) 법원 판례 데이터 은행]", (<https://www.lareda.hessenrecht.hessen.de/bshe/search>, 2021.9.4. 최종방문).

3. 국내

가. 법원

국내에서 판결문을 열람할 수 있는 법원 사이트로는 ‘종합법률정보시스템’과 ‘판결서 인터넷열람 서비스’가 있다. 종합법률정보시스템은 법원도서관에서 운영하며 판결문을 무료 열람할 수 있으나, 판례집이나 공보를 통해 공개된 판례와 미간행 판례 중 주요 판결로 선정된 일부의 판결문을 공개하고 있는 시스템이다. ‘판결서 인터넷열람 서비스’는 법원도서관 특별열람실을 방문해 판결문을 검색하고 열람할 수 있었던 제도를 온라인상으로도 가능하게 한 서비스로, 종합법률정보시스템에 비해 더 많은 판례를 보유하나 1건당 수수료 1,000원을 부과하고 있는 유료서비스 방식으로 운영된다.

종합법률정보시스템은 판례, 법령, 문헌, 규칙·예규·선례 등 자료를 제공하고 있다.³⁷⁾ 대법원 판결문의 경우 대법원판례집(1953년~2009년), 법원공보(1973년~1995년), 판례공보(1996~현재) 등에서 공개된 판결문을 제공하며 하급심 판결문은 고등법원판례집(1961년~1983년), 하급심판결집(1984년~2003년), 각급(제1,2심) 판결공보(2003~현재) 등에 실린 판결문 외 미간행 판결문 중 법원도서관 측에서 주요 판결로 선정한 판결문을 공개하고 있다. 본문을 저장할 시 필요와 목적에 따라 PDF와 HWP 파일을 선택할 수 있으며 저장 범위 또한 본문 또는 일부만 저장할 수 있다. 또한, 국가법령정보센터에서 제공하는 오픈 API를 통해 판례목록, 본문 검색에 관한 결과를 XML로 받을 수 있다.

종합법률정보시스템은 이와 같이 무료로 파일을 저장할 수 있고 API를 통해 수집하는 것도 용이하나, 간행물에 실린 판결문 중심으로 주요 판결로 선정된 판결문만 공개하기에 공개된 판결문의 수가 극히 적다는 단점이 있다. 종합법률정보 사이트에서 검색할 수 있는 대법원 판결은 전체의 3.2%에 불과하고³⁸⁾, 온라인으로 공개되지 않은 1·2심 판결문이나 나머지 96.8%의 대법원 판결문을 보려면 법원도서관 컴퓨터를 사전에 예약하여 열람하여야 하는데 이 경우 출력·복사·촬영이 금지되고 있다.

판결서 인터넷열람 서비스³⁹⁾는 2013년 1월 1일부터 확정된 형사사건 판결서와

37) 홈페이지 주소(<https://glaw.scourt.go.kr/>, 2021. 9. 10. 최종방문).

38) 대법원 판결의 3.2%, 각급 법원 판결의 0.003%가 종합법률정보 시스템에 공개되어있다고 볼 수 있다. 이탄희 의원 발의 민사소송법 일부개정법률안.(2020. 08. 26. 의안 제2103298호).

2015년 1월 1일부터 확정된 민사사건 판결서를 전자적으로 제공하는 서비스다. 해당 법원 홈페이지에만 공개되었던 이전 제도와 다르게 한 홈페이지에서 모든 법원의 판결서를 검색할 수 있게 되었다는 점에서 진일보한 법률 서비스라고 할 수 있다. 특히 열람이 제한되어 있지 않은 확정 판결서를 공개하고 있어 종합법률정보시스템보다 더 많은 데이터양을 보유하고 있다는 특징이 있다.

나. 민간분야

국내 최대 법률 포털 사이트라고 할 수 있는 로앤비(LawnB)⁴⁰는 법무법인 태평양의 자회사에서 시작하여 2012년 톰슨 로이터에 인수된 온라인 법률 데이터베이스 사이트다.⁴¹ 로앤비 공식 사이트에 의하면 로앤비는 법령정보, 행정자료, 주석서 등 법률 문헌, 법조인 정보, 생활법률 등을 제공하고 있으며, 하급심 판례를 비롯하여 총 23만여 건의 판례를 제공한다. 판례 검색 기능은 검색어와 제외어를 입력하여 기본검색이 가능하고 사건번호, 판시사항, 재판요지, 전문, 사건명, 참조조문, 검색범위(공간 여부, 사건종류, 법원등급, 선고일자)를 상세 설정하여 검색할 수 있다. 로앤비는 재판 경과를 연결하여 참조판례, 따름판례 등을 확인할 수 있는 장점이 있을 뿐만 아니라 파일 저장 시 PDF, HWP, MS Word, TXT 형식을 모두 지원하고 있다. 다만 민간에서 운영하는 사이트인 만큼 회원등급별 제공범위의 차이가 있는데 무료회원의 경우에는 판례 본문 중 일부만 이용 가능하며 개인회원인 경우 ID 1개당 1개월에 99,000원 회원비를 부담해야 한다.

또 다른 민간분야의 판례검색 사이트인 케이스노트(CaseNote)⁴²는 판례검색 및 열람서비스를 제공한다. 판례검색 기능은 판례번호 또는 키워드를 입력하여 기본검색이 가능하고 제외어, 법원등급, 사건종류, 선고일자를 상세 설정하여 검색할 수 있다. 검색결과로 제공하는 판결문은 TXT, PDF, 이미지 PDF 형태로 받을 수 있다. 회원등급

39) 홈페이지 주소(https://www.scourt.go.kr/portal/information/finalruling/peruse/peruse_status.jsp, 2021. 9. 10. 최종방문).

40) 홈페이지 주소(<https://www.lawnb.com/>, 2021. 9. 10. 최종방문).

41) "법률AI에 정보 입력하자... '당신 형량은 2년'", 매일경제, 2020. 9. 6. 기사(<https://www.mk.co.kr/news/business/view/2020/09/919950/>, 2021.9.1. 최종방문).

42) 홈페이지 주소(<https://casenote.kr/>, 2021. 9. 10. 최종방문).

별 제공범위의 차이가 있는데, 프로(PRO) 회원일 경우 총 26만 건 판례검색 및 열람을 할 수 있으며, 판결문 구입이 필요할 경우 법원에 사본제공신청을 하여 월 최대 30건의 판결문을 신청할 수 있다. 프로 회원 가입비는 ID 1개당 1개월에 42,900원이다.⁴³⁾

아래 <표 3>은 로앤비와 케이스노트 외 국내 판결문 검색 사이트와 판결문을 기반으로 분석 서비스를 제공하는 국내 기업들의 데이터 보유량과 유형을 분석한 표이다.

<표 3> 국내 민간분야 법률 데이터 수집량 비교표

서비스명(기업명)	데이터 수집량	데이터 유형	공개여부
로앤비(통스로이더)	23만 건	법원 판례	(판결문 일부) 무료 공개 (판결문 전체) 회원제 공개
케이스노트 (케이스노트)	37만 건	법원 판례(26만), 학술논문(11만)	(판결문 일부) 무료 공개 (판결문 전체) 회원제 공개
엘박스(엘박스)	25만 건	법원 판례	(판결문) 공개
유렉스(인텔리콘)	300만 건 (판례 외 문서 포함)	법원 판례, 법령조례 등	(판결문) 공개
리걸엔진(까리용)	350만 건 (판례 외 문서 포함)	법원 판례, 법령해석 자료 등	(판결문) 공개
로톡 형량예측서비스 (로앤컴퍼니)	40만 건	형사 1심 판결문 (2012년~2020년)	(서비스 [형량예측]) 공개 (판결문) 비공개
인텔리켓(베링랩)	4,800만 건 (판례 외 문서 포함)	법률, 특허 문서	(서비스 [법률문서번역]) 회원제 공개 (판결문) 비공개

대다수 검색사이트에서 보유하고 있는 판결문 건수는 심급의 구분 없이 23~26만 건 정도라고 볼 수 있으며 데이터 활용 및 분석 중심 서비스를 제공하는 업체의 경우 서비스의 목적에 맞는 특정 문서를 중심으로 대량수집하고 있다는 것을 알 수 있다.

43) 매월 최대 30개의 판결문 구입 신청(수수료 30,000원)이 가능하다는 점과 실제로 판결서 인터넷열람 사이트에서 다운로드받은 PDF파일도 함께 제공한다는 점 등을 고려하면 현재 케이스노트는 회원가입비 중 약 70% 정도를 데이터 확보에 투자하고 있다고 볼 수 있다.

4. 국내외 판결문 사이트 하급심 판결문 공개 및 결과 비교

본 연구에서는 위에서 언급한 사이트들이 보유하고 있는 하급심의 양과 실제로 제공되는 데이터의 형식을 비교함으로써 국내 공공기관에서 운영하는 판결문 열람 제도의 실효성을 평가하고자 미국과 독일의 판결문 공개 결과를 비교해 보았다. 국내는 종합법률정보, 판결서 인터넷열람 서비스, 로앤비, 케이스노트 PRO에서 검색하였고 독일의 경우 통일된 판결문 열람사이트가 없는 관계로 독일연방 법무부의 전자 법률 정보시스템 구축사업에서 파생된 법률 포털 Juris 기반 검색을 하였다. 미국 PACER의 경우 검색에도 요금을 부과하여 로앤비와 동일회사에서 운영하는 Westlaw 사이트를 사용하였다.⁴⁴⁾

검색어로는 국내 사이트에서는 ‘형법 제250조’를 사용하였으며 독일 사이트인 Juris에서는 유사법령인 독일형법(StGB) 제211조(Mord 모살)와 제212조(Totschlag 고살)를, 미국 사이트인 WestLaw에서는 특정 법 조항 대신 각각 Murder (모살), Manslaughter(고살)와 ‘conviction(유죄선고)’의 키워드 조합을 사용하였다.⁴⁵⁾

〈표 4〉 국내외 판결문 사이트 '살인' 검색 결과 비교표 (최종 검색일자: 2021.09.04.)

사이트 명	종합 법률정보	판결문 인터넷열람	로앤비	케이스노트 PRO	Juris (독일)	Westlaw (미국)
운영 기관	법원	법원	투스 로이터	케이스노트	juris GmbH	투스 로이터
검색어	형법 제250조	형법 제250조	형법 제250조	형법 제250조	StGB §§ 211, 212	murder & conviction, manslaughter & conviction

44) 미국의 대표적인 법률 리서치 엔진은 Westlaw, LexisNexis, Bloomberg Law 등이 있으나 이 중에서 Westlaw가 미국내 시장점유율이 가장 높고, 국내 법원도서관과 판례 데이터교류협정을 체결하여 검색기능 평가의 대상으로 선정하였다.

“법원도서관-웨스트로 판례데이터 교류협정 체결”, 법률신문, 2011. 4. 11. 기사

(<https://m.lawtimes.co.kr/Content/Article?serial=56925>, 2021. 9. 10. 최종방문).

45) 미국의 보통법은 통상적으로 모살죄를 1급 모살죄(first-degree murder)와 2급 모살죄(second-degree murder)와 고살죄를 고의고살죄(voluntary manslaughter)와 비고의고살죄(involuntary manslaughter)로 구분하나, 다른 검색결과와 유사한 조건을 맞추기 위해 모범형법전(Model Penal Code) 기준인 모살(murder)와 고살(manslaughter)을 검색키워드로 사용하였다. 이경재, “미국형법상 살인죄의 유형과 그 내용”, 홍익법학, 18권 1호, 2017, 38면 참조.

사이트 명	종합 법률정보	판결문 인터넷열람	로앤비	케이스노트 PRO	Juris (독일)	Westlaw (미국)
사건 종류	형사	형사	형사	형사	Strafrecht	Crimes and Criminal Procedure
법원 등급	고등법원 지방법원	고등법원, 지방법원 (지원포함)	고등법원, 지방법원	고등법원, 지방법원	고등지방법원 (OLG), 지방법원(LG)	State, Federal (Supreme court 제외)
검색 결과	123건 (98건 고등, 25건 지방)	6,054건 (1,356건 고등, 4,698건 지방)	1,220건 (389건 고등, 831건 지방)	632건 (172건 고등, 460건 지방)	441건 (149 OLG, 292 LG)	19,613건 (2,094 연방, 17,519 주) *모살, 고살 각각 9,800건 가량
지원 형식	HWP, PDF	이미지 (Image-only) PDF*	PDF, HWP, TXT, MS Word	PDF, TXT	PDF, RTF	MS Word, RTF, PDF, WordPerfect
비용	무료	1건당 1,000원	회비 월 99,000원	회비 월 42,900원	회비 월 75 유로(약 104,000원)부터	회비 월 252 달러(약 295,000원)부터

*2021년 상반기 기준

위 결과를 봤을 때 우리나라의 경우 판결문과 관련 문서를 전면공개(실명포함)하는 미국과 비교해 볼 때 공개건수가 최소 3배 이상 차이가 난다는 걸 볼 수 있다. 이는 판례법과 관습법을 근간으로 하는 영미법계인 점과 사건 발생 건수의 차이라고도 해석할 수 있으나, 리걸테크 시장 규모나 기술의 발전 등을 고려할 때 아쉬운 부분이다. 판결서 인터넷열람 서비스는 공개하는 하급심 건수로만 보면 약 6,504건으로 우리나라와 여건이 비슷한 독일의 경우보다 더 많고, 상업형 국내 법률검색사이트와 비교하더라도 (형사 하급심 기준) 가장 많은 데이터양을 보유하고 제공하고 있는 것으로 보인다. 그러나 판결서 인터넷열람 서비스는 기계 판독이 불가능한 이미지 PDF 파일 형태로 제공하고 있어 인공지능 학습용으로는 가치가 전혀 없다고 할 수 있다. 이런 점에서 위 결과만 비교할 때, 독일에 비하여 공개건수가 비교적 많게 보이나, 오히려 전자정보의 관점에서는 종합법률정보가 제공하는 123건에 불과하여 독일의 441건에 비하여 현저히 적다. 또한, 다른 서비스에서는 적어도 2종류 이상의 파일 형식을 제공하는 데 비해 판결서 인터넷열람 서비스는 PDF뷰어를 통해 자료를 보여주고 있어 전자

도서관과 유사하게 오로지 '보기'에만 치중이 되어 있다고 할 수 있다. 통합적인 데이터 관리 시스템을 구축한 선구적인 모습과 달리 데이터 제공과 활용에 소극적인 태도를 보인다고 할 수 있다.

IV. '판결서 인터넷열람 서비스' 판결문 생성과정 분석

현재 법원에서 제공하는 '판결서 인터넷열람 서비스'는 형사판결문 하급심 원문을 공개하고 있으나, 앞에서 설명한 바와 같이 유료로 공개하면서도 대부분 기계 판독이 불가능한 이미지 PDF 파일 형태로만 제공하고 있다. 최근 민사소송법 개정⁴⁶⁾으로 인해 2023년 1월부터 민사판결문이 기계 판독 가능한 PDF파일로 제공하게 되면서 「전자우편 등을 통한 판결문 제공에 관한 예규」도 개정되어⁴⁷⁾ 2021년 7월 5일을 기준으로 게시⁴⁸⁾되는 민·형사 판결문은 텍스트 검색이 가능한 파일로 제공되게 되었다. 그러나 본 예규는 소급적용이 되지 않아 약 200만 건⁴⁹⁾의 형사판결문이 여전히 기계 판독이 불가능한 이미지 PDF로 남겨 되었다. 판결서 인터넷열람 서비스에서 제공되는 자료가 판결서뿐만 아니라 증거목록과 기록목록도 포함된다는 점을 고려한다면 더 많은 양의 데이터가 마땅한 활용방안 없이 사장되는 상황이다.

본 연구는 '판결서 인터넷열람 서비스'가 국내 사법 분야 인공지능 기술 성장에 기여하기 위해서 판결문 공개비율 상향과 함께 예규가 적용되지 않는 판결문 데이터까지 쉽게 호환될 수 있는 시스템으로 개선될 필요가 있다고 보았다. 이하에서는 원문 판결

46) 시행령정 민사소송법 제162조의2(판결서의 열람·복사) ② 제1항에 따라 열람 및 복사의 대상이 되는 판결서는 대법원 규칙으로 정하는 바에 따라 판결서에 기재된 문자열 또는 숫자열이 검색어로 기능할 수 있도록 제공되어야 한다. <신설 2020. 12. 8.> [시행일 : 2023. 1. 1.]

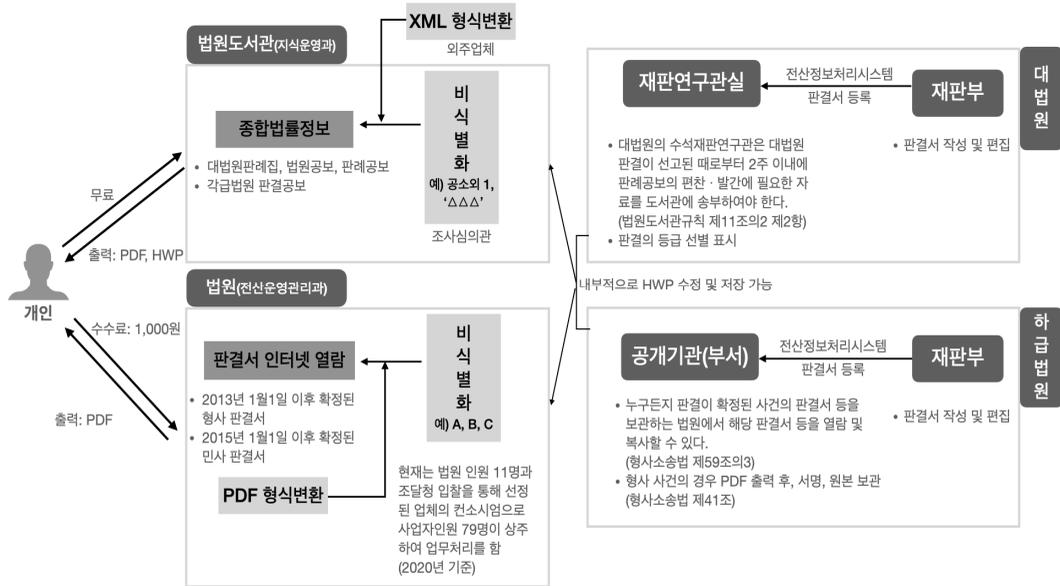
47) 위 예규에서의 판결문 제공방식은 전자우편뿐만 아니라 판결서 인터넷열람 서비스도 포함하여 2021년 7월 5일 이후로 제공되는 판결문은 문자열 검색이 가능하다.

48) 선고일이 아닌 판결문을 PDF로 변환 시 자동으로 삽입되는 문구의 게시일자 기준인 것으로 보인다. 게시일자는 사용자가 열람사이트에서 해당 PDF를 구매하거나 열람한 날짜와 관계없는 (대부분 판결 선고일과 1년 이내) 날짜로 PDF 상단에서 확인할 수 있다.

49) 사법정책연구원의 연구보고서(이민형, "형사재판절차에서의 개인정보보호에 관한 연구" 335면의 [표 14] '형사 판결서 공개 열람 현황')에 따르면 2019년까지 가장 많은 공개건수는 263,648건(2017년)이었고 가장 적은 공개건수는 209,142건(2013년)으로, 매년 평균 공개건수를 약 233,817건으로 가정할 때 2021년 7월까지 약 2,000,000건의 판결문이 텍스트 추출이 불가능한 이미지 PDF로 제공되었다고 본다.

문이 일반인에게 인터넷으로 공개되기까지의 과정에 대해 구체적으로 살펴봄으로써 현재와 같이 판결문이 이미지 PDF 파일로 계속 제공되어야 할 정책적 이유가 있는지, 그리고 텍스트 데이터로 활용(변환)하는 것이 기술적으로 가능한지의 여부를 분석해 보았다.

먼저 판결문 데이터를 인공지능에 활용할 수 있는 방법을 모색하기 위해 우리나라 판결서 인터넷열람 서비스에서 판결문이 이미지 PDF 파일로 제작되어 제공되는 과정을 <그림 1>과 같이 설명해보았다. 판결문 데이터가 제공되는 과정 중 비식별화 과정 등 일부 단계를 설명한 자료는 있었으나 법원에서 개인까지 판결문이 생성되고 변환되는 전반적인 흐름에 관한 참고문헌은 극히 드문 관계로 이를 파악하기 위해 검증된 전문가와 인터뷰를 통해 확인하여 심층 분석하였다.⁵⁰⁾



판결문이 열람사이트를 통해 사용자에게 제공되기까지 법원의 심급과 중요도에 따라 여러 기관을 거치게 된다. 먼저 재판부는 한컴社의 한글 프로그램을 사용하여 판결문을 텍스트 데이터로 작성, HWP 파일 형태로 저장한다⁵¹⁾. 작성된 판결문은 전산 정보

50) 이 흐름도는 법원실무자 등과 인터뷰 및 국회토론회 자료 등을 통해 판결문 작성과정과 공개현황을 직접 파악하여 작성한 것이다.

처리시스템을 통해 등록되며, 재판서의 원본이 되는 문서는 형사소송법 제41조(재판서의 서명)에 의거하여 이 등록된 판결문을 PDF로 변환한 후 전자 서명(형사판결문의 경우 직접 서명)한 전자문서다.

이후 대법원의 경우 판례공보 편찬을 위해 재판관연구실에서 대법원 판결이 선고된 후 2주 이내에 판례공보의 편찬·발간에 필요한 자료를 법원도서관 조사심의관실로 송부한다.⁵²⁾ 이때 재판관연구실에서 각 판결문의 중요도에 따라 크게 A, C, D, X급으로 분류하여 보내고,⁵³⁾ 이 중요도에 따라 판례공보에 판례를 실을지에 대한 여부를 조사심의관실에서 등급을 확정지어 결정한다. 조사심의관실에서 등급이 분류된 판결은 법원도서관 내규인, 「판례 편찬물 등의 비실명 처리 지침」에 따라 조사심의관들에 의해 ‘소외인’, ‘△△△’ 등으로 표시하는 방식으로 비실명화 처리가 된다. 이후 종합법률정보에 등록하기 위해 비실명화 처리된 판결문을 외주업체를 통해 XML 형태로 변환작업⁵⁴⁾을 한다. 이렇게 처리된 판례공보는 종합법률정보에서 사용자가 무료로 웹사이트에서 텍스트, PDF, HWP의 형태로 판결문에 접근할 수 있다.

대법원 판결문의 처리 과정과 달리 하급법원의 판결문은 「판결서 등의 열람 및 복사를 위한 비실명 처리 기준(재일2014-2)」에 따라 심급별, 주소, 차량번호, 금융정보, 전화번호, 별지 등을 알파벳 또는 특정 부호 등으로 대체하는 작업을 거치게 된다. 우선적으로 인공지능 기반 비식별 시스템을 통해 자동 비식별화를 진행하고, 민사 법원 행정처 재판사무국의 ‘비실명화 및 전자소송 사건 스캔 업무 지원사업소’에서 검수 인력이 대량의 정보를 유형 구분 없이 A, B, C 등 알파벳 대문자로 수동으로 대체하여 2차 비식별화작업을 수행하고 있다.⁵⁵⁾ 이렇게 비식별화 과정을 거친 판결문의 선고일, 법원명, 사건번호, 사건명, 주문 그리고 이유 일부는 임의어 검색 대상이 되며, 사용자가 수수료 1,000원을 지불하면 이미지 PDF의 형태로 받게 된다.

51) 이용재, “현행 판결문 공개제도 검토”, 판결문 공개 확대를 위한 국회토론회(2019.10.25.), 7면.

52) 이기리, “판결문의 비실명처리”, 법원사람들, 2017년 12월호, 법원행정처 공보관실, 30면.

53) A급은 전원합의체 및 대법원판례집에 실을 만한 판례이고, C급은 선례적 가치가 있어 실을 만한 판결, D급은 중요한 법리를 포함하고 있으나, 기존 판시가 있었거나 이미 수록된 적 있었던 판결, X급은 참조할 만한 내용이 없는 판결로 분류한다.

54) 여기서 XML의 형태 변환작업은 판례공보가 종합법률정보에 게시되기 위해 데이터의 구조와 의미를 정의하는 것이다. 앞의 글 (주50) 참조.

55) 이민형, “형사재판절차에서의 개인정보보호에 관한 연구”, 사법정책연구원, 2021, 15면 참조.

이상과 같이 판결서 인터넷열람 서비스의 비식별화 과정이나 검색과정을 보면 법원 내에서는 판결문이 텍스트 검색이 가능한 데이터로 저장되고 공유된다는 것을 알 수 있다. 기술적으로도 이 데이터의 출력형태가 반드시 이미지 PDF이어야 하지 않음도 확인할 수 있다. 아래는 법원에서 사용 중인 PDF 변환 도구와 판결문 PDF를 분석한 표로 실제로 기계 판독이 가능한 PDF 변환의 가능성을 실험해본 결과이다.

〈표 5〉 판결문 PDF 파일 및 변환 도구 분석표

사건명	게시일	선고일	PDF 버전	PDF 변환 도구 버전	기계 판독 가능 여부
인양지원 2021고합23	2021-06-18	2021-04-30	1.7	v2.1.45.0 (e435b1f,dirty)	× (게시일자 문구도 드래그 불가)
서울서부지법 2020고단3268	2021-06-25	2021-05-12	1.7	v2.1.47.0 (e154fb2)	×
서울북부지법 2020고합495	2021-07-06	2021-05-04	1.4	v2.1.47.0 (e154fb2)	○
서울중앙지법 2017노4066	2021-08-18	2021-07-01	1.4	v2.1.47.0 (e154fb2)	○
전주지법 2021노270	2021-09-01	2021-07-01	1.4	v2.1.48.0 (b8363f1)	○
광주지법 2021고단1729	2021-09-14	2021-08-12	1.4	v2.1.48.0 (b8363f1)	○

PDF변환에서 사용되는 도구는 ePapyrus社의 pdfcmd.exe로 열람서비스에서 다운로드받은 PDF 파일의 PDF변환소프트웨어(Producer) 정보를 통해 확인할 수 있다. pdfcmd.exe는 여러 이미지 파일(jpg, png, bmp 등)이나 doc, docx, hwp파일 등을 pdf로 변환해주는 도구로 게시일자 기준 2021년 9월 이후 법원에서 사용하고 있는 버전은 v2.1.48.0이다. 기계 판독이 되는 PDF 변환을 위해 사용된 것은 v2.1.47.0 이상의 버전이며 PDF 1.7버전으로 변환되는 이미지 PDF와 다르게 하위 버전인 PDF 1.4버전으로 출력되는 것을 알 수 있다.⁵⁶⁾

56) 이는 pdfcmd.exe의 특징으로 jpg 파일 등 이미지를 PDF로 저장할 때는 자동으로 PDF 1.7버전으로 설정되었으나 한글 에디터의 가상 프린터를 사용하여 생성한 (텍스트 검색이 가능한) PDF일 경우 PDF 1.4버전으로 저장되었다.

그러나 기계 판독이 가능한 PDF로 변환하는 기능은 최신 버전이 아니더라도 존재했던 것으로 보인다. 위의 표에서 볼 수 있듯이 v2.1.47.0로 변환된 서울서부지법 2020 고단3268 판결문은 이미지 PDF이나 같은 버전의 도구를 사용한 서울북부지법 2020 고합495 판결문은 텍스트 추출이 가능한 PDF로 제공된다. 일반적인 상용도구의 경우 기능을 추가할 때 버전을 업데이트하여 관리한다는 점을 고려할 때 기계 판독이 가능한 PDF 생성만을 위한 대대적인 시스템 변화가 있었다고 볼 수 없다. 따라서 지금까지 기계판독이 가능한 PDF를 생성하지 않은 것은 기술적인 어려움이 아닌 단순히 정책적인 문제였던 것으로 판단된다. 실제로 ePapyrus社에서 공개하는 Papyrus 5 (PDF 뷰어)를 사용해보았을 때 동일한 PDF 변환 도구를 이용하여 HWP파일을 기계 판독 가능한 PDF로 변환할 수 있었다.⁵⁷⁾ 이와 같은 사정을 고려하면 데이터베이스에 저장된 200만 건의 형사판결문도 기술적으로 별다른 어려움 없이 이미지 PDF가 아니라 문자 내용복사가 가능한 PDF 문서로 변환하고 그 과정을 충분히 자동화할 수 있을 것으로 생각된다.

법원은 공개되는 판결문이 이미지 PDF 형태로 제공되는 이유로 편집·수정으로 인한 악용 가능성을 언급하고 있으나⁵⁸⁾, 이는 해시값(Hash value) 사용⁵⁹⁾ 등 무결성 검사나 문서 제한을 통해 일차적으로 방지 가능할 뿐만 아니라 예규가 개정되어 판결문이 기계 판독 가능한 PDF로 제공되게 된 이상 더는 유효한 정책적 이유로 볼 수 없다. 변환되는 기점인 2021년 7월 이후 공개건수가 다시 200만 건만큼 축적되기까지 대략 8~9년이 걸린다는 점을 고려할 때 이미지 PDF로 보관된 판결문의 변환은 반드시 선행되어야 하는 작업이다.

만약 2021년 7월 이전의 판결문이 지금과 같은 이미지 형태의 PDF로만 제공될 수밖에 없다면 인공지능의 활용을 위해서는 (1) 수동으로 이미지 PDF 판결문 전문을 타

57) 2010 한컴오피스 한글 버전 8.0.0.466으로 생성한 HWP파일과 pdftocmd v2.1.1.2032을 사용하여 진행하였다. Papyrus 5에 포함된 pdftocmd 도구가 법원에 제공되는 것의 구버전임을 고려할 때 이미 기계 판독이 가능한 PDF 변환 도구임에 불구하고 수정·편집 방식을 위해 HWP 파일 또는 텍스트 파일인 판결문을 이미지 (JPG, PNG 등)로 변환, 변환된 이미지를 pdftocmd를 사용해 다시 PDF로 생성하고 있었던 것으로 보인다.

58) 송오섭, “판결서 열람·검색제도 개선을 위한 몇 가지 고려사항”, 판결문 공개 확대를 위한 국회토론회 (2019.10.25), 32면.

59) 임의의 길이를 가진 데이터를 특정 알고리즘을 사용해 고정된 길이의 데이터로 매핑하여 도출하는 값으로 다 운받은 파일에 실제 제공되는 파일과 같은지 확인할 수 있는 문자열이다. 문자 하나만 추가되어도 문자열이 달라져 무결성 검사에 사용된다.

이핑하거나, (2) 판결문 구입 신청 서비스를 제공하는 민간 판례검색서비스를 이용하는 방법 또는 (3) OCR(Optical Character Recognition, 광학 문자 인식)을 사용하여 기술적으로 우회하는 방법을 생각해 볼 수 있다. 그러나 (1)과 같이 수동으로 타이핑을 하는 것은 과도한 인력과 시간이 낭비되며, (2)와 같은 구입은 국가가 제공하는 판결문을 유료 민간서비스로 구독해야 하며 불필요한 많은 비용을 발생시킨다는 문제점이 있다.

(3)과 같은 기술적 우회방법을 사용할 경우 이미지 파일인 판결문을 대량으로 텍스트 데이터로 자동변환하는 것은 가능하다. OCR은 이미지(사진) 속 문자의 위치를 찾는 기술과(Text Detection, 문자 검출) 어떤 문자인지 자동으로 알아내는 기술(Text Recognition, 문자 인식)로 구성되고 일반적으로 문자 검출에는 인공지능 모델인 CNN(Convolutional Neural Network)계열의 모델을 그리고 문자 인식에는 RNN(Recurrent Neural Network) 계열의 모델을 사용한다. 이렇듯 OCR에 사용되는 핵심 기술은 큰 차이가 없으나 모델의 성능은 학습에 사용되는 데이터의 특징에 맞춰지기 때문에 추출하려는 대상을 잘 인식해내는 모델을 찾아 사용해야 한다. 한글 인식의 경우 무료로 공개된 OCR 모델을 사용할 수도 있으나 한영혼용의 문제가 있어⁶⁰⁾ 한글에 최적화된 기업형 OCR을 사용해야 판결문 데이터처리에 용이하다. 따라서 OCR을 사용하여 데이터 변환을 진행할 때는 인식 오류의 문제가 있을 뿐만 아니라 문법오류 수정과 같은 부차적인 과정을 거쳐야 하고, 불가피하게 상업용 OCR 모델을 사용할 경우 추가비용이 발생할 수밖에 없다. 이런 사정을 모두 고려한다면 판결서 인터넷열람 서비스의 이미지 PDF 형태 제공방식을 유지한다는 것은 기술적으로 쉽게 해결할 수 있음에 불구하고 합리적인 이유가 없는 채 판결문의 전자정보로서의 가치를 떨어뜨리고 있는 것으로 보인다.

V. 인공지능 발전을 위한 ‘판결서 인터넷열람 서비스’ 개선방안

1. 데이터 제공 방식 변경

앞에서 언급한 바와 같이 현재 보관된 파일은 90% 이상이 기계 판독이 불가능한

60) 한글과 영문자 알파벳이 혼용되어 한 단어 또는 문장에 사용될 경우 인식오류가 발생하는 경우를 말한다.

이미지 PDF이다. 따라서 인공지능이나 기계학습을 위해 데이터를 활용하기 위해서는 OCR 등 추가적인 기술을 사용하여 데이터를 추출해야 하는 실용적 어려움이 있다. 그러나 앞서 IV장에서 확인한 바와 같이 법원 전산시스템에 등록된 판결문은 시스템적으로 충분히 기계 판독이 가능한 형태로 변환할 수 있으며 예규 개정 전 이미지 형태의 판결문을 제공한 것은 오로지 정책적인 판단에 의한 것으로 보인다.

판결문 데이터 제공방식에 관한 문제를 정책적인 관점에서 보기 위해 먼저 판결문이 공공데이터의제공및이용활성화에관한법률(이하‘공공데이터법’이라고 한다)제2조 제2호에서 정의하고 있는‘공공데이터’에 해당하는지를 살펴보았다. 공공데이터법 제2조 제2호에서‘공공데이터’란 공공기관이 법령 등에서 정하는 목적을 위해 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보를 의미하는데, 공개된 판결문은 이미 법원 전산정보처리시스템에 전자적 방식으로 보관하고 있다는 점을 고려하면 공공데이터의 전형적인 유형이라고 할 수 있다. 그렇다면 공공데이터법 제2조 제4호⁶¹⁾에 따라 이용자는 판결문 정보를 기계 판독이 가능한 형태로 접근할 수 있거나 전달받아야 할 것이다.

기계 판독이 가능한 판결문 제공과 관련해서는 2020년 8월 26일자에 이탄희 의원 등 22인이 발의한 형사소송법 일부개정법률안(의안번호: 2103299) 등 여러 차례 언급된 바가 있다.⁶²⁾ 발의안에서는 판결문에 대한 정보 접근성을 높이는 것을 사법비용을 절감하고 사법절차의 공정성과 투명성을 확보하는 조치로 보고 있으며 이러한 근거는 행정예규의 개정이유에서도 확인된다.⁶³⁾ 그런데도 판결서 인터넷열람 서비스는 개인정보보호·내용 수정 및 편집 방지라는 이유로 이미지 PDF로만 제공했고, 텍스트 검색이 가능한 PDF로 공개가 결정 된 이후에도 지난 10여 년간 이미지화되어 축적된 판결문 데이터에 관해서는 아무런 조치가 취해지고 있지 않는 상황이다.

61) 공공데이터의 제공 및 이용 활성화에 관한 법률 제2조(정의) 4. “제공”이란 공공기관이 이용자로 하여금 기계 판독이 가능한 형태의 공공데이터에 접근할 수 있게 하거나 이를 다양한 방식으로 전달하는 것을 말한다.

62) 개정형사소송법안 제59조의3(확정 판결서등의 열람·복사) ② 제1항에 따라 열람 및 복사가 제한되지 아니한 판결서(일부만 제한된 경우도 포함한다)는 대법원 규칙에서 정하는 바에 따라 판결서에 기재된 문자열 또는 숫자열이 검색어로 기능할 수 있도록 「공공데이터의 제공 및 이용 활성화에 관한 법률」 제2조에 따른 기계 판독이 가능한 형태로 제공되어야 한다. (의안번호:2103299)

63) 전자우편 등을 통한 판결문 제공에 관한 예규 1. 개정이유
○ 전자우편으로 제공하는 판결문의 접근성을 보다 높이기 위해 기계 판독이 가능한 형태로 판결문을 제공하도록 개정함 [행정예규 제1276호, 시행 2021. 7. 5.]

이미지 PDF로 변환시켜 놓은 기존의 데이터를 다시 기계 판독이 가능한 형태로 제공하기 위해 추가적인 자동화 시스템을 도입하는 등 예산상 혹은 행정적인 부담도 고려요소가 될 수 있다. 그러나 판결서 인터넷열람 서비스의 수수료 납부액은 2019년에는 9,000만원, 2020년에는 13,000만원으로 수납금액이 큰 폭으로 증가하고 있으며⁶⁴⁾, 판결서 인터넷열람 서비스가 보유하고 있는 판결문 데이터양이 국내에서 가장 크기 때문에 사용자는 지속적으로 유입될 것이 예상된다. 이는 판결서 인터넷열람 서비스가 미국의 PACER처럼 방대한 법률정보 플랫폼으로 성장하여 지능형 정보시스템 관련 산업의 발전에 큰 기여할 가능성이 있음과 비용의 상쇄가능성을 시사하는 것이다. 이러한 성장 가능성을 고려해서라도 양질의 데이터 수집과 활용을 위한 노력이 필요하며, 제공되는 데이터를 인공지능 학습데이터 구축에 적합한 형식으로 변경하는 조치도 필요하다고 할 것이다.

2. 검색기능의 개선

이상과 같은 기계 판독이 가능한 판결문 서비스 제공과 더불어 판결문 데이터를 이용한 인공지능 활성화를 위해서는 열람서비스의 기능을 개선해야 할 사항들이 몇 가지 있다. 우선 검색기능의 선고일자(확정일자) 기간 범위 설정 개선이 필요하다. 현재 판결서 인터넷열람 서비스는 최대 2년의 기간을 검색범위로 제한하고 있다. 사용자가 원하는 검색기간 범위가 2년 이상이 될 경우 2년 단위에 맞춰 재검색을 진행해야 하는 번거로움이 있다. 예를 들어 2016년에서 2019년까지의 결과를 얻기 위해서는 먼저 2016.1.1.부터 2017.12.31.사이의 범위를 검색하고 2018.1.1. 사이의 2019.12.31. 사이의 결과를 총 두 번 검색해야 한다는 것이다. 만일 찾고자 하는 판결문의 정확한 연도를 모를 경우라면 여러 차례의 검색이 반복되어야 한다.

간단한 키워드 검색 외 필터링 기능이 없는 것도 이용의 불편을 초래한다. 판결서 인터넷열람 서비스에서 제공하는 검색결과는 선고일자(확정일자), 법원명(사건번호), 사건명, 주문, 이유 등으로 구분되어 사용자에게 제공된다. 이는 데이터베이스 상 판결문 데이터가 스키마⁶⁵⁾로 분류되어 한 사건이 한 개의 레코드(행)로 저장되어 있다고 볼

64) 국회예산정책처, 형사소송법 일부개정법률안 [비용추계서 미첨부 사유서] (의안번호: 2107718, 의뢰인:박주민 의원), 3면.

65) 데이터베이스 상의 스키마는 테이블의 구조 특성을 뜻한다.

수 있다. 일반적으로 스키마의 필드(열)는 스키마값으로 필터링 할 수 있다. 예를 들어 동일한 키워드 ‘살인’을 포함하고 있는 판결문들을 사건명(살인등[형사], 강도살인[형사] 등)에 따라 필터링할 수 있는 것이다. 또한 주문, 이유 부분이 구분되는 것은 판결문의 내용도 항목 범위로 나눠졌음을 의미한다. 앞서 봤듯이 연도 필터 기능은 2년 단위로 사용할 수밖에 없고, 사건번호 검색은 정확한 번호를 알지 못하면 사용할 수 없다. 법령 검색 또한, ‘형법 제250조’를 검색했음에도 불구하고, ‘공직선거법 제250조 제2항, 각 형법 제307조 [...]’가 검색되는 등 정확한 검색이 불가능하고, 사건명은 필터 기능이 없어 불필요하게 많은 결과로 이어진다.

마지막으로 임의어(키워드) 검색이 단순히 문자열 검색에 불과해 발생하는 검색결과 부정확성의 문제가 있다. 실제로 ‘살인’이라는 임의어를 검색하면 강제추행 사건임에 불구하고 판결 전문에 ‘[...] 몇 살인 것 같냐고 물어보고 저에게 몇 살이냐고 [...]’⁶⁶⁾에 ‘살인’이 언급되어 검색결과에 포함되는 경우가 발생하는 것이다. 또한, 데이터베이스 상 주문, 이유가 분리되어 있음에도 불구하고 키워드 검색 기능은 전체를 검색 대상으로 하여 주문 부분에서 유무죄의 여부를 확인하기 위해 검색해도, 이유 부분에 포함된 ‘무죄’가 검색되는 등 어느 위치에서 검색된 결과인지 특정하기 어렵다.

이와 같이 판결서 인터넷열람 서비스는 제한적인 기간 범위 설정, 사건명의 검색과 필터링 기능의 부족, 단순 텍스트 매칭의 키워드 검색 기능의 문제점들이 있다. 동일하게 법원에서 운영하는 종합법률정보 서비스를 비교해 본다면 판결서 인터넷 열람서비스보다 한층 고도화된 검색기능을 제공하고 있다. 종합법률정보의 검색엔진은 단순 검색과 상세검색 뿐만 아니라 검색어 자동완성, 연관검색어, 최근검색이력, 키워드맵 기능을 제공하고, 상세검색은 위에서 지적했던 여러 문제들을 해결해내고 있다.

종합법률정보에서는 제한 없는 선고일자 범위설정, 사건명을 가진 판결만을 검색, 판결문의 내용의 항목범위(판시사항, 판결요지, 판결전문) 설정도 가능하다. 관련어 찾기로 질의어(검색어)를 입력하여 검색하면 법원도서관에서 구축한 법률분야 관련어집의 데이터를 통해 관련어질의어 및 관계, 관련어 항목 목록이 나와 쉽게 검색하도록 도와준다. 이러한 기능을 가능케 하는 검색엔진은 데이터베이스에서 사용자가 지정한 키워드 또는 문자에 해당하는 항목을 검색하고 식별하는 프로그램이다.⁶⁷⁾ 더 향상된 양질

66) 서울서부지방법원, 2021.04.12.선고, 2020노1613

67) "search engine, n.". OED Online. Oxford University Press. <https://www.oed.com/view/Entry/83275825?redirectedFrom=search+engine>, 2021.09.10. 최종방문)

의 검색 결과를 도출하기 위해 검색엔진에 자연어처리 기술인 형태소 분석, 말뭉치 데이터, 개체명 인식, 정보추출 등을 접목하면 사용자 의도나 상황에 맞는 정보를 일상 용어 중심으로 또는 검색어를 유의어나 동의어 형태로 변환하여 검색할 수 있다. 기존에 인터넷열람 서비스는 위의 '살인' 검색예시처럼 문장이 아닌 단어로만 이루어진 조합으로 검색되었지만, 자연어처리를 적용하면 '소주를 마시고 운전을 하였다.'라는 문장에서 '소주'를 주류의 객체로, '운전'을 행위의 객체로 인식하고, 문장에서의 단어를 의미분석하여 '음주운전', '알코올', '혈중알코올농도', '음주', '주류' 등으로 변환되어 검색된 더 많은 결과값을 도출한다. 만일 위의 문장에서 '운전'이라는 단어에 가중치를 둔다면 도로교통법과 관련된 판례가 나오게 하는 것이다.

종합법률정보의 키워드맵은 입력한 검색어에 대한 카테고리별 연관 검색어를 텍스트 형태의 목록식으로 보여주는데, 더 나아가 시각화된 검색결과는 법령과 판례, 판례 등의 연결성을 이해하는 데 효과적이다. 도식화된 검색결과로 연결되는 법령 또는 판례를 따라 더 상세한 검토와 추론을 하는데 도움이 된다. 따라서 판결서 인터넷 열람서비스의 검색엔진에도 종합법률정보처럼 다양한 기능을 가진 검색엔진을 추가하고 텍스트 데이터를 자연어처리, 시각화, 인공지능의 발전된 정보기술을 반영하여 법률 분야의 고유한 검색엔진을 구축해야 할 필요가 있다. 향후 이러한 기술로 법률분야에 특화된 데이터셋이 구축된다면 인공지능 학습용 데이터로 검색서비스에 활용될 수 있을뿐 아니라 사용자의 질문을 해석하고 그에 적절한 법률 내용을 자동으로 응답하는 챗봇 서비스 기능 향상에까지 도움을 줄 수 있을 것이다.

3. 사용자 편의성 개선

사용자의 편의성을 위해 다음과 같은 사항이 개선되어야 한다. 먼저 디지털 약자에 대한 배려가 필요하다. 대법원은 노약자나 장애인 등 문자로만 이루어진 데이터에 쉽게 접근하기 어려운 이들을 위해 2019년 종합법률정보시스템에 음성변환(Text-to-Speech, TTS) 서비스를 구축하였다.⁶⁸⁾ 그러나 TTS란 기본적으로 텍스트 분석 후 그에 맞는 언어적 해석을 통하여 음성으로 변환하는 것이기 때문에 별도의 절차 없이 텍스트를 추출할 수 없는 이미지 PDF와 같은 문서에는 적용할 수가 없다. 현

68) "이피피루스, 대법원에 시각장애인용 뷰어 서비스 구축...음성 변환지원 적용", 전자신문, 2019.08.27. 기사 (<https://m.etnews.com/20190827000143>, 2021.9.10. 최종방문)

재로서 디지털 약자들이 판결서 인터넷열람 서비스를 통해 공개된 약 200만 건의 판결문에 접근할 방법은 매우 제한적인 셈이다. 이를 해결하기 위해서라도 기계판독이 가능한 데이터 제공은 필수적이라 할 것이다.

수수료 과금체계도 해결되어야 할 문제점이다. 현재는 1건 당 1,000원의 수수료가 부과됨에 불구하고 결제하기 전에는 전체보기가 불가하여 실제로 필요한 판결문인지 평가하기가 어렵다. [더보기]를 클릭해도 검색 키워드 주변의 문단 일부⁶⁹⁾만 보일 뿐, 실제로 유사사건의 기준이 될 수 있는 범죄사실 등을 보기 위해서는 여러 키워드를 시도하여 관련 내용을 추가로 검색하는 방법 또는 1,000원의 수수료를 결제 후 확인하는 방법에 없어 확인을 위해 매번 구매해야하는 큰 제한이 있다. 그뿐만 아니라, 장바구니에 담을 수 있는 판결문은 100건임에도 불구하고, 한 번에 결제할 수 있는 판결문은 최대 20건으로 제한되어 관련성 있는 데이터 확보가 더욱 어렵다.

열람 허용 기간도 빅데이터 환경에 맞게 조정될 필요가 있다. 현재는 판결문 최초 열람 후 24시간 이내에만 열람할 수 있으며, 그 시간 내 다운로드를 받지 않았다면 열람이 더 불가능한 문제가 발생한다. 이는 열람하고자 하는 판결문 대상이 몇 건이 되지 않는다면 큰 문제로 이어지지 않을 수 있지만, 실제로 판결문을 분석하고자 하는 목적이거나 데이터 학습을 위하여 수집하려는 경우라면 열람 후 저장이 되지 않는 등 오류가 발생하는 경우도 생길 수 있다. 이러한 열람 제한 또한 사용자 편의성과 데이터 확보 효율성의 저해요인으로 작용하고 있어 이에 대한 개선이 필요하다.

VI. 결 론

판결문 공개범위 확대, 기계 판독 가능한 데이터 제공 및 하급심 판례 공개는 리걸테크 시장과 학계는 물론 국회에서도 여러 차례 논의되어 온 사안이다. 다행스럽게도 이와 같은 노력의 결과로 최근 민사소송법·전자우편 등을 통한 판결문 제공에 관한 예규를 개정해 지금부터의 판결문에 대해서는 기계학습이 가능해졌지만 여전히 사법데이터로서 큰 가치를 가지는 형사판결문은 특히 공개 진행이 미진한 상태이다.

사법부 내외에서 기계 판독에 대한 논의는 활발하게 진행되고 있는 반면 아직까지

69) 약 800자에서 1000자(공백포함) 사이로 제공된다.

실제 사용자의 관점에서 편의성 개선에 반드시 다뤄져야 할 제공 판결문의 형태, 검색 기능 고도화, 시각장애인 등 디지털 약자에 대한 배려, 데이터와 무관하게 일괄 부과 되는 수수료, 다양한 데이터 확보를 위한 대량 다운로드 가능성 등 인공지능 기반을 마련하기 위해 실무에서 필요한 구체적인 논의는 전혀 이루어지지 않고 있다.

법원이 선진적으로 기획한 스마트법원 4.0이나 종합법률 API 개발, 각 법원에서 있던 판결문 데이터를 통합하여 제공하는 판결서 인터넷열람 서비스 등을 보아도 우리나라가 정책적으로도 법률 데이터 활용에 대해 선진적인 태도로 접근하고 있다고 할 수 있다. 그러나 실제로 보유하고 제공할 수 있는 데이터양과 달리 일반 국민에게 공개되는 정보량, 관련 정보 검색을 위한 편의성, 데이터 확보 과정 등에 대한 지원은 조금 더 관심을 가지고 깊이 있게 논의될 필요가 있다.

최근 국내의 리걸테크 시장과 기존 변호인들과의 갈등이 최고조로 오르면서 리걸테크 기업들이 존폐의 기로에 선 가운데 법조계에서는 판결문 데이터베이스화에 대한 부정적 의견들도 나오고 있다. 인공지능을 통한 의사결정, 소송결과 예측, 문서 자동 작성 등을 전문가가 아닌 수치로 된 데이터에만 의존할 시 법률 사무 수행에 큰 지장이 간다는 입장이다. 그렇지만 기술한 바와 같이 현재의 약 200만 건의 판결문을 이미지 PDF 형태로 제공한다고 하여 기술적으로 이를 우회하는 방법이 없는 것도 아닌 반면 판결문 데이터 활용에는 큰 제한을 가져온다. 특히 하급심 판결의 제공은 인공지능 개발 활성화뿐만 아니라 기초 사법연구에도 매우 중요한 문제인 만큼 본 연구를 통해 제기된 문제들이 조속히 개선되어 사법분야에도 인공지능 기술이 활성화되기를 기대한다.

주제어 : 판결서 인터넷열람 서비스, 리걸테크, 인공지능, 판결문, 텍스트마이닝

*논문접수:2021.09.29. *심사개시:2021.10.10.

*논문수정:2021.10.22. *게재확정:2021.10.23.

參攷文獻

I. 국내 문헌

- 김승래, AI시대 리걸테크의 발전과 미래 법률시장의 변화 모색, 법이론실무연구 제8권 제3호, 한국법이론실무학회, 2020.
- 양시훈, 판결에 대한 국민의 이해를 높이고 이를 알릴 수 있는 방안에 관한 연구, 사법정책연구원, 2015.
- 이경재, 미국형법상 살인죄의 유형과 그 내용, 홍익법학 제 18권 제 1호, 2017.
- 이민형, 형사재판절차에서의 개인정보보호에 관한 연구, 사법정책연구원, 2021.
- 이성진·이연주, 4차 산업혁명시대 법정기술의 현황과 발전방안, 정보화정책 제28권 제1호, 한국지능정보사회진흥원, 2021.
- 전정현·김병필, 인공지능과 법률 서비스 현황과 과제, 저스티스 제170권 제1호, 2019.

II. 국외 문헌

- Esther Salmerón-Manzano, Legaltech and Lawtech: Global Perspectives, Challenges, and Opportunities, Laws 제 10권 제2호, 2021.

III. 국내 판례

- 서울서부지방법원, 2021.04.12. 선고, 2020노1613

IV. 국외 판례

- BPatG, 27 ZA (pat) 19/90, Beschluss vom 23. April 1991 (독일 연방특허법원 제27부, 1991년 4월 23일 판결).
- BGH, IV AR(VZ) 2/16, Beschluss vom 5. April 2017 (독일 연방대법원 민사제4부, 2017년 4월 5일 판결).

BGH, V AR(Vs) 112/17, Beschluss vom 20. Juni 2018 (연방대법원 형사제5부, 2018년 6월 20일 판결)

V. 기타

송오섭, “판결서 열람·검색제도 개선을 위한 몇 가지 고려사항”, 판결문 공개 확대를 위한 국회토론회(2019.10.25).

이용재, “현행 판결문 공개제도 검토”, 판결문 공개 확대를 위한 국회토론회(2019.10.25.).

이기리, “판결문의 비실명처리”, 법원사람들, 2017년 12월호, 법원행정처 공보관실.

이기리, “판례공보와 종합법률정보”, 법원사람들, 2017년 8월호, 법원행정처 공보관실.

현대경제연구원, “리걸테크(Legaltech) 산업 현황과 시사점”, 현대경제연구원 VIP Report 16-31 (통권 669호), 2016.

백상준, “판결서 인터넷열람 제도의 개선현황과 향후과제”, 국회입법조사처, 2019

이탄희 의원 대표 발의 민사소송법 일부개정법률안.(2020. 08. 26. 의안 제 2103298호).

이탄희 의원 대표 발의 형사소송법 일부개정법률안 (2020. 08. 26. 의안 제 2103299호).

국회예산정책처, 형사소송법 일부개정법률안 [비용추계서 미첨부 사유서] (의안번호: 2107718, 의뢰인:박주민 의원).

AI가 판결문까지 작성?...법원은 '신중 모드', 문화일보, 2019. 4. 9 기사 (<http://m.munhwa.com/mnews/view.html?no=2019040901032003009001>, 2021. 09. 10. 최종방문).

AI로 개인정보 거른 판결문 공개 필요...‘리걸테크’ 유니콘 탄생 첫걸음, 뉴스웍스, 2020. 12. 8 기사 (<https://www.newsworks.co.kr/news/articleView.html?idxno=506027>, 2021. 9. 10. 최종방문).

IT만남 법률...변호사 선임 무턱 낮추고 AI가 형량 예측, 세계일보, 2021. 5. 2 기사 (<https://m.segye.com/view/20210502508132>, 2021. 9. 10. 최종방문).

‘로톡’사태는 시작일뿐...다른 법률 서비스 스타트업도 고사위기, 이데일리, 2021. 9. 8 기사 (<https://www.edaily.co.kr/news/read?newsId=01108646629178152&mediaCodeNo=257&OutLnkChk=Y>, 2021. 09. 10. 최종방문).

미확정 판결서 공개...어디까지 왔나, 법률신문, 2020.12.03. 기사(<https://www.lawtimes>.

- co.kr/ Legal-News/Legal-News-View?serial=166111, 2021. 9. 10. 최종방문).
- 법률AI에 정보 입력하자...“당신 형량은 2년”, 매일경제, 2020. 9. 6 기사 (<https://www.mk.co.kr/news/business/view/2020/09/919950/>, 2021. 9. 1. 최종방문).
- 서초동에 부는 AI 바람...리걸테크가 온다, 중앙일보, 2020. 5. 30 기사(<https://www.joongang.co.kr/article/23789440#home>, 2021. 9. 10. 최종방문)
- 심재훈 변호사 "판결문 공개해 리걸테크 발전 기반 삼아야", iT조선, 2020.06.17. 기사 (http://it.chosun.com/site/data/html_dir/2020/06/16/2020061603849.html, 2021.08.29. 최종방문).
- 이파피루스, 대법원에 시각장애인용 뷰어 서비스 구축...음성 변환지원 적용, 전자신문, 2019.08.27. 기사(<https://m.etnews.com/20190827000143>, 2021.9.10. 최종방문).
- 전국법관대표 “하급심 판결문 공개 확대해야”, 이투데이, 2021. 1. 18 기사 (<https://www.eteday.co.kr/news/view/1985806>, 2021. 8. 27. 최종방문).
- 주목받는 리걸테크 산업, 법률신문뉴스, 2021. 4. 26 기사 (<https://www.lawtimes.co.kr/Legal-News/Legal-News-View?serial=169586>, 2021. 8. 27. 최종방문).
- 판결문 공개 과감히 확대하라, 법률신문, 2019. 10. 28 기사 (<https://m.lawtimes.co.kr/Content/Article?serial=156740>, 2021. 8. 27. 최종방문).
- Jessika Kallenbach, “Kein Anspruch auf Herausgabe eines Strafurteils an private Dritte [형사판결을 개인인 제3자에게 공개해야 할 권리는 없다]”, Anwaltsblatt, 2018.09.11 기사 (<https://anwaltsblatt.anwaltverein.de/de/anwaeltinnen-anwaelte/berufsrecht/bghanonymisierte-urteilsabschriften>, 2021.09.09. 최종방문)

[국문 초록]

사법분야 인공지능 발전을 위한 판결문 데이터 개선방안 - ‘판결서 인터넷열람 서비스’를 중심으로 -

박성미 · 이유나 · 최아리 · 안정민

빅데이터·인공지능 분야에 정부의 집중적 투자와 공공기관의 전면 디지털화가 추진되면서 법원과 같은 공공기관에서 생산하는 데이터 확보에 관한 관심도 높아지고 있다. 판결문의 경우, 민사사건은 2023년 1월 1일을 기점으로 미확정 판결문과 기계 판독 가능한 파일을 제공하는 등 그 제약이 줄어들고 있지만 형사판결문은 여전히 확보가 어려운 상황이다. 사법 데이터로서 가치 있는 판결문 제공이란 단순한 공개 이상의 것을 의미한다. 인공지능이나 기계학습의 사법분야로의 확대는 활용가능한 판결문 데이터를 필수요소로 한다. 특히 사법 분야 인공지능 개발을 위해서는 사실관계의 쟁점을 다루는 하급심 판결문 데이터 확보가 중요한데, 국내외 공공·민간 판결문 제공 사이트의 성능과 데이터 보유량을 비교해 보았을 때 현재 대한민국 법원에서 운영하는 ‘판결서 인터넷열람 서비스’에서 제공하는 데이터양은 상대적으로 많음에도 불구하고 기술적 제한으로 데이터의 활용가능성의 측면은 부진한 상태다.

이러한 관점에서 본 연구는 국내외 리걸테크 시장의 발전과 사례를 통하여 어떻게 판결문이 사법분야 인공지능 개발에 기여하는지 확인하고, 데이터로서 하급심(1·2심) 판결의 중요성과 공개 필요성을 살펴보았다. 이를 위해 현재 우리나라에 공개되어있는 판결문 건수, 범위 및 형태 등을 미국·독일의 공공·민간 판결문 공개 사이트와 - 그 중에서도 특히 그 데이터의 인공지능 활용의 면에 중점을 두고 - 비교·분석하였다. 마지막으로 이러한 비교분석의 결과를 대법원에서 현재 운영되고 있는 ‘판결서 인터넷 열람 서비스’구조와 기능을 분석하고, 향후 인공지능 연구 활성화를 위한 개선사항을 결론으로 제시하고자 한다.

주제어 : 판결서 인터넷열람 서비스, 리걸테크, 인공지능, 판결문, 텍스트마이닝

[Abstract]

The 'Online Access to Judgment' Service in Korea: A Study on Improving Judgment Data for the Development of Legal AI(Artificial Intelligence)

Park, Sung-Mi* · Lee, Yu-Na** · Choi, A-Ri** · Ahn, Jungmihn J.***

Over the past few years, the government has increased investment in artificial intelligence and big data by promoting the complete digitization of public agencies. This digitization along with digital transformation has brought more attention to accessing court documents particularly the judgments and opinions. It has been announced that as of 2023, all civil cases will be available to the general public in machine-readable format. However, there is no special plan announced for the criminal judgments and lower courts' decisions which had always been hard to obtain.

Acquiring lower courts decisions is crucial to building AI-powered criminal law systems, as those decisions focus on disputed facts. Compared to other legal search systems in the private sector and other countries, Korean 'Online Access to Judgment' holds a relatively extensive database. However, the usage has been very restricted due to technical limitations and not providing machine-readable file has made it difficult to use this court judgment data to train artificial intelligence.

This study analyzes the systematic structure of the online access to judgment service to see if it is feasible to produce data in machine-readable format. And the paper compares legal database of the United States and Germany to show what aspects of Korean system needs to be

* Phd Candidate, School of Cybersecurity, Korea University, 1st Author

** Master's Candidate, Legal Informatics and Forensic Science, Hallym University

*** Professor, Legal Informatics and Forensic Science, Hallym University, Corresponding author

improved. By suggesting how the current technical limitations can be fixed to accommodate a better user experience and to increase user convenience, we hope to open up opportunities for legal AI service in Korea.

Key words: Online Access to Written Judgment, Legal Tech, Artificial Intelligence, Court records, Text mining