국립중앙도서관

ISSUE PAPER

이슈페이퍼

국립중앙도서관 고신문의 온라인서비스 발전방안

Vol. 17
2023. 10.

문화체육관광부
국립중앙도서관

OPEN 출처표시 상업용금지 변경금지
공공누리 공공저작물 자유이용허락
'공공누리' 출처표시-상업적 이용금지-변경금지 조건에 따라 이용할 수 있습니다.
단, 표지, 일러스트레이션 및 사진은 제외됩니다.

국립중앙도서관 고신문의 온라인서비스 발전방안

권정임 사무관
(국립중앙도서관 온라인자료과)

Ⅰ. 들어가며 02

Ⅱ. 국립중앙도서관 고신문 디지털화 사업 02
1. 디지털화 사업의 추진 근거 02
2. 디지털화 사업의 필요성과 목적 04
3. 신문의 디지털화 현황 06

Ⅲ. 고신문 온라인서비스 플랫폼, 대한민국 신문 아카이브 08
1. 서비스 내용 08
2. 국내외 서비스 사례 10
3. 서비스 발전 방안 14

참고문헌 15

| 주요 키워드 |
대한민국 신문 아카이브, 고신문, 디지털컬렉션, 고신문 컬렉션, 인공지능

Vol. 17
2023. 10.

국립중앙도서관
ISSUE PAPER
이슈페이퍼

국립중앙도서관 **이슈페이퍼 제17호**

발행일 2023년 10월 31일
발행처 국립중앙도서관
발행인 국립중앙도서관장
주소 서울특별시 서초구 반포대로 201
전화 02-590-0799
팩스 02-590-0546
누리집 https://nl.go.kr
ISSN 2765-3005

• 본지에 실린 글의 내용은 집필자의 개인적인 견해이며, 국립중앙도서관의 공식적인 의견과 다를 수 있습니다.
• 본지의 저작권은 국립중앙도서관에 있으며, 사전 허락 없이 무단으로 복제·변경·배포할 수 없습니다.

국립중앙도서관 고신문의 온라인서비스 발전방안

권정임 사무관
(국립중앙도서관 온라인자료과)

요약

국립중앙도서관은 2013년부터 115억 7천만 원의 예산을 투입하여 한국 근현대 신문(1883~1960년 발행)에 대해 99종 381,010면의 디지털 원문을 구축했고, 신문기사 7,678,320건(2022. 12. 31. 기준)에 대한 DB 구축 및 신문별·일자별·기사별로 1,509만 건의 색인 DB를 구축하였고, 이를 누리집 '대한민국 신문 아카이브'를 통해 서비스하고 있다. 동 누리집은 디지털 원문과 일부 텍스트를 제공하고 있으며, 연구자·기자 등이 많이 찾는 국립중앙도서관의 여러 누리집 중에서 지명도가 꽤 높다. 디지털 원문 99종의 신문에 대한 간략 소개(해제)와 함께 기사에 대해 신문명과 발행일, 키워드와 문장으로 검색서비스를 제공하며 일부 신문에 대해서는 원문 텍스트를 제공하고 있어 복사 및 다운로드를 받아 활용할 수 있고, 기사별 LOD를 발행하고 있다.

고신문은 가로쓰기 및 세로쓰기로 되어 있고, 국한문 및 일어가 혼용되어 있어 이용에 불편함이 많으나, 대한민국 근현대사의 경제·사회·문화 전반을 알 수 있는 사료로서 가치가 높다. 따라서, 1960년 이전의 국내외 민간 신문사 및 교육·문화예술기관과의 협력을 통해 국립중앙도서관에 없는 고신문이나 결호를 발굴하고, 디지털화 지원을 통해 디지털 원문을 확보, 대한민국 신문 아카이브를 통해 서비스할 필요가 있다. 또한 인공지능 기술 등을 활용하여 고신문에 대한 이미지뿐만 아니라, 텍스트 DB, 번역 DB, 이미지 DB 등을 구축하고 검색할 수 있는 환경을 제공하고, 이용자가 시스템에 접속하여 텍스트 및 내용을 수정할 수 있는 시스템을 마련할 필요가 있다.

주요 키워드 대한민국 신문 아카이브, 고신문, 디지털컬렉션, 고신문 컬렉션, 인공지능

Ⅰ. 들어가며

우리나라 최초의 근대신문은 1883년 10월 31일 대한제국에서 창간한 『한성순보』이다. 당시 개화파들은 신문을 통해 국민에게 외국의 사정을 널리 알려 개화사상을 고취시키려고 하였다. 최초의 민간신문은 1894년 4월 7일 서재필이 창간한 『독립신문』으로, 한글전용과 띄어쓰기가 되어 있는 점이 특징이다. 이 신문은 이후 민간신문 제작에 큰 영향을 미치며 민중계몽과 자주독립사상을 확립하는 데 기여하였다[1]. 이와 같이 신문은 전통적으로 경제, 정치, 문화 등 사회의 주요 이슈와 매일의 사건사고 등 사회의 변화를 가장 먼저 전달하고, 날씨, 스포츠, 연예 등 독자에게 일상의 읽을거리를 제공하는 주요 매체 중 하나다. 특히, 과거의 신문 기사는 한국 근현대사의 정치·경제·사회·문화의 변화와 발전을 살펴볼 수 있는 중요한 자료다. 지금 우리는 신문을 모바일·온라인으로 어디에서나 손쉽게 읽는 것을 당연하게 생각하고 있으나, 1990년대 이전까지는 가정 또는 직장에서 읽거나 지하철, 버스 등에서 대강 훑어보곤 했었다. 뉴스 콘텐츠 소비방식이 이렇게 빠른 변화를 거치게 된 것은 온라인이 보편화된 세계적 흐름과 맥을 같이하고 있다.

이와 같이 사회적인 변화에 따라 국립중앙도서관은 1945년 개관 때부터 납본, 구입 등을 통해 수집해 온 실물자료의 효율적 관리 및 보존, 온라인서비스를 하기 위해 1995년부터 실물자료를 디지털로 변환하는 「국가문헌의 보존 및 지식정보자원 공유기반 구축 사업(이하 디지털화 사업)」을 시작하였고, 고신문[2]은 2013년부터 진행해 오고 있다.

본 고에서는 국립중앙도서관의 디지털화 사업 전반에 대해 짚어보는 한편 고신문 디지털화 사업 현황 및 서비스에 대해 살펴보고자 한다. 이와 함께 국내외 다른 기관의 신문 디지털화 현황 및 서비스를 살펴보고 향후 발전방안에 대해 논의해 보고자 한다.

Ⅱ. 국립중앙도서관 고신문 디지털화 사업

1. 디지털화 사업의 추진 근거

국립중앙도서관은 국가대표도서관으로서 국내에서 발행된 자료를 후대에 이용할 수 있도록 과학적

1 네이버 두산백과 <한국의 신문 역사>. 출처: https://terms.naver.com/entry.naver?docId=1186930&cid=40942&categoryId=31755

2 국립중앙도서관에서의 고신문은 1887년~1960년 발행한 신문을 범위로 하고 있다.

보존과 체계적 관리를 담당하고 있다. 이를 위해 「도서관법」, 문화체육관광부의 「문화비전2023」 등과 같은 제도적 기반을 근거로 하여 디지털화 사업을 추진하고 있다(<표 1> 참조). 「도서관법」 제20조 제1항 제2호에서는 국립중앙도서관이 국내외 도서관자료를 수집·제공·보존하도록 규정하고, 문화체육관광부 「사람이 있는 문화: 문화비전 2030」 의제7 중 대표과제4에서는 문화기관 등의 다양한 문화콘텐츠를 아카이빙하고 플랫폼을 통해 대국민 서비스를 제공하도록 하고 있다. 이를 근거로 국립중앙도서관에서는 디지털화 사업과 관련 플랫폼[3]을 구축 및 서비스하고 있다.

<표 1> 국립중앙도서관 디지털화 사업 추진 근거

- 「도서관법」 제19조, 제20조
  제19조(국립중앙도서관의 설치 등) ① 국가를 대표하는 도서관으로 문화체육관광부장관 소속으로 국립중앙도서관을 둔다.
  제20조(업무) ① 국립중앙도서관은 다음 각 호의 업무를 수행한다.
  2. 국내외 도서관자료의 수집·제공·보존관리

- 「문화비전 2030(2018. 5. 16.)」 의제7의 대표과제4
  의제7(문화자원의 융합역량 강화) 중 문화자원의 기록·보전·활용 체계 개선(대표과제4(문화자원의 기록·보전·활용 체계 개선)
  - 국·공립 및 민간 문화기관 등의 아카이빙 플랫폼 지원 확대
  - 다양한 문화콘텐츠의 디지털 아카이브 구축 추진

또한 「국립중앙도서관 소장자료 디지털화 및 지식공유기반 구축에 관한 규정」('21. 1. 5. 제정, '23. 3. 24. 개정)에 따라 디지털화 대상자료, 디지털화 형식, 디지털 장서 등록 및 보존 등 디지털화 사업 전반에 걸쳐 효율적인 추진을 도모하고 있다(<표 2> 참조).

3 코리안메모리 서비스(https://nl.go.kr/memory), 대한민국 신문 아카이브(https://nl.go.kr/newspaper)

<표 2> 국립중앙도서관 소장자료 디지털화 및 지식공유기반 구축에 관한 규정 제3조, 제6조

• 「국립중앙도서관 소장자료 디지털화 및 지식공유기반 구축에 관한 규정」
제3조(대상자료) ① 디지털화 대상자료
1. 귀중자료
2. 산화 또는 열화로 영구보존이 필요한 자료
3. 「저작권법」 제31조제3항에 따른 발행일로부터 5년이 경과한 자료
4. 기증자료 유치, 국가 및 공공기관 발간 정책희귀자료, 인문예술자료 등 도서관 미소장 자료의 수집과 공동활용을 위한 정책적 결정에 따른 자료
제6조(디지털화 형식) 디지털로 변환하는 자료의 파일형식

구분			파일형식	해상도(1-2)/ 비트심도(3)/코덱(4)	비 고
1. 인쇄자료	텍스트	보존용	칼라, TIFF/24bit Color 칼라, JPEG/24bit Color	400-600ppi 300ppi	인쇄 자료(단행, 연속, 신문 등) Born-Digital의 경우 원 파일 형식 유지
		제공용	칼라, JPEG 칼라, 텍스트/목차-PDF	300ppi	아크로뱃 8.x 이상
		썸네일	칼라, JPEG/24bit Color	96ppi	
	지도	보존용	칼라, TIFF/24bit Color 칼라, JPEG/24bit Color	400-600ppi	Born-Digital의 경우 원 파일 형식 유지
		제공용	칼라, JPEG 칼라, 텍스트/목차-PDF	300ppi	아크로뱃 8.x 이상
		썸네일	칼라, JPEG/24bit Color	96ppi	
2. 이미지자료	사진	보존용	칼라, TIFF/24bit Color 칼라, JPEG/24bit Color	800-1000ppi	Born-Digital의 경우 원 파일 형식 유지
		제공용	칼라, JPEG 칼라, 텍스트/목차-PDF	400ppi	아크로뱃 8.x 이상
		썸네일	칼라, JPEG/24bit Color	96ppi	
3. 녹음자료	오디오	보존용	WAV(무압축)	24bit	48kHz
		제공용	MP3(손실압축)	16bit	44.1kHz
			AAC(손실압축)	16bit	48kHz
4. 영상자료	비디오	보존용	MXF(무압축)	MPEG4(영상) / PCM(음성)	• SD: 4:3 • HD: 16:9
		제공용	MPEG-4(손실압축)	H.264(MPEG-4/AVC) (영상) / AAC(음성)	• SD: 4:3 • HD: 16:9

※ 원자료의 형태, 열화 등 훼손정도, 파일 용량, 뷰어프로그램 성능 등을 고려하여 파일형식 및 해상도를 조정할 수 있음

2. 디지털화 사업의 필요성과 목적

1980년대까지 발행된 산성지로 된 실물자료들은 취약한 보존환경으로 인해 열화 및 산화가 진행되면서 종이의 부스러짐이 심했고, 이에 더해 낮은 수준의 인쇄기술로 인해 글자 판독이 곤란하다는 단점을 가지고

있었다. 또한 마이크로필름으로 변환한 자료들 역시 과거 기술의 한계로 인해 현재는 판독이 어려운 경우가 많았다. 이러한 문제들로 인해 자료를 디지털화하여 보존함으로써, 자료의 원본성과 신뢰성을 확보해야 한다는 필요성이 제기되었다. 1990년대 이전에 발행된 신문은 pH 4.0 이하의 산성지로, 이런 종류의 종이는 발행한 지 30년이 지나면 열화 및 훼손 등으로 원본을 온전히 보존하기 어렵기 때문에 디지털화가 시급하다. [그림 1]은 고신문의 실물 상태를 보여주고 있다.

고신문 마이크로필름에 곰팡이가 낀 상태

[그림 1] 고신문의 실물 상태

또한 최근 민간기관과 공공기관에서는 저작권이 해결된 디지털화 자료, 특히 이미지가 아닌 정제된 텍스트(text) 자료를 초거대 AI, 챗GPT 학습용 자료로써 요구하고 있다. 왜냐하면 이미지 자료는 초거대 AI, 챗GPT가 학습할 수 있는 텍스트 형태가 아니기 때문이다. 이러한 신기술들은 모바일·온라인으로 이용자와 소통함으로써 문제를 해결하는 생활의 편리를 제공한다고 볼 수 있는데, 도서관도 예외는 아니다. 코로나 이후 이용자들은 도서관에 오지 않거나 오더라도 대면하지 않고 온라인·비대면으로 자료를 이용하고자 하는 수요가 증가하였다. 국립중앙도서관의 경우, 2021년 온라인 자료의 이용률은 2020년에 비해 238%, 우편복사는 108%가 증가하였다. 이와 함께 디지털인문학 분야의 연구자들도 디지털 원문자료의 내용을 분석·연구하기 위해서 데이터 분석기법과 시각화 기법을 활용하는 등 연구를 위해서 저작권이 해결된 디지털화 자료의 이미지와 텍스트 데이터에 대한 요구가 증가하고 있다. 그리고 도서관 자료의 이용에 있어서도 Text PDF로 변환한 자료는 본문 내용 검색을 활용할 수 없기 때문에 이용자는 이미지를 1장씩 넘기며 확인해야 하는 불편이 존재하지만, 텍스트 데이터는 디지털 원문의 내용 분석에 활용될 수 있기 때문에 본문에 나타난 주제, 사건, 시간, 장소 등 풍부한 메타데이터를 작성하고, 이를 검색 요소로 제공함으로써 이용자는 더욱 많은 자료에 편리하게 접근할 수 있다는 장점이 있다.

따라서 국가문헌에 대한 디지털화 사업이 조속히 이루어져야 하는 이유는 다음과 같이 정리할 수 있다. 첫째, 실물자료의 안정적 보존과 글로벌 지식정보 생태계 속에서 지속가능한 활용을 위해서이다. 둘째, 최근 소개된 초거대 AI, 챗GPT의 등장으로 인한 본문 내용분석, 텍스트 데이터 가공 및 분석 등 수요자 중심의 서비스 제공과 국가 데이터 경제 및 디지털 융합학문 연구와 데이터 수요에 부응하는 서비스를 지원하기 위해서이다. 셋째, 모바일·온라인이 서비스의 중심이 된 지금의 이용자 수요에 빠르게 대응하기 위해서이다.

3. 신문의 디지털화 현황

국립중앙도서관에는 납본·기증 등의 경로를 통해 매년 50만여 책(점)의 다양한 실물(도서·비도서) 자료들이 신규로 입수되고 있다. 매해 자료의 중복과 디지털화 대상 제외자료[4](사전·참고서·수험서, 전자자료로 발행된 자료 등)로 인한 편차가 발생하기는 하나, 이 중 매년 20만여 책이 신규로 디지털화를 해야 할 대상 물량에 해당된다.

한편 국립중앙도서관은 1995년부터 2023년까지 1,400억 원의 예산을 투입하여 디지털화 사업을 추진하고 있다. 2023년 131억 4천만 원의 예산에 따른 디지털화 대상 물량이 9만 1천 책인데, 해마다 물가와 인건비의 상승 및 2022년 이후 예산도 조금씩 감소되는 추세임을 감안할 때, 향후 디지털화 물량이 지속적으로 감소될 것으로 예상된다. 따라서 한정된 예산과 저작권 이슈 등을 고려하여, 국가적으로 보존 및 이용 가치가 높은 귀중서, 희귀자료, 오래된 자료들을 중심으로 우선순위를 정하여 추진하고 있다. 신문도 동일한 원칙으로 디지털화 하고 있는데, 1차적으로는 1960년까지의 신문 중 다른 기관에서 디지털화를 하지 않았거나 유료·무료 관계없이 전자신문으로 제공 중인 것은 제외하고 있다[5].

국립중앙도서관은 신문에 있어서는 우선적으로 고신문을 이미지로 디지털화하고 서비스하는 사업에 중점을 두고 있다. 2013년에 미래창조과학부(현 과학기술정보통신부)와 한국정보화진흥원(현 한국지능정보사회진흥원)의 국가 DB 사업에 "귀중본 고신문 DB구축 사업"이 선정되어, 1939년부터 1949년까지 발행된 국민신보, 독립신문, 대한독립신문(민보), 독립신보, 민중일보, 조선중앙일보(서울석간), 중앙신문 등 7종, 14만 6,736건의 기사를 DB로 구축하기 시작하였다. 이렇게 디지털로 구축된 신문은 처음에는 국립중앙도서관 디브러리[6](https://www.dibrary.net)를 통해 서비스하였는데, 이때부터 신문의

4 국립중앙도서관은 「소장자료 디지털화 지침(1999)」에서 디지털화 제외 대상자료를 정하고 있다(p. 41).

5 디지털화 사업 예산 추이: '20~'21(연간 150억 원) → '22~'23(연간 137억 원) → '24(127.8억 원 예상)

6 디브러리는 2015년 4월에 서비스를 중단하였다.

원문을 집에서 읽을 수 있게 되었다. 2016년 9월에는 한국언론진흥재단과 업무협약을 체결하고 매일신문, 황성신문, 조선중앙일보 등 한국언론진흥재단이 구축한 고신문 15종, 82,055면, 1,473,481건의 신문기사 DB를 넘겨받았다. 이후 2016년 대한민국 신문 아카이브(https://nl.go.kr/newspaper)를 개발하고 2017년부터 본격적으로 동 플랫폼에서 서비스함으로써 기자·연구자·창작자 등이 매일 평균 367회 접속하는 대표적인 디지털 기반 신문 서비스로 자리잡고 있다. 2018년부터는 국립중앙도서관이 직접 사업으로 예산을 편성하고 디지털화 사업을 추진 중에 있다. 2013년부터 2022년까지 115억 7천만 원의 예산을 투입하여 99종 381,010면을 디지털화했고, 신문이 주로 날짜별, 기사별로 소비되는 특성을 반영하여 DB로 구축된 신문기사 7,678,320건('22. 12. 31. 기준)에 대해 신문별·일자별·기사별로 1,509만 건의 색인 DB를 구축하여 서비스의 편리함을 더하고 있다. 다만, 조선일보, 동아일보, 경향신문, (구)서울신문, 한국일보 등은 해당 신문사에서 DB를 구축하여 유료·무료 서비스를 하고 있어 디지털화 사업에서 제외하고 있다. <표 3>, <표 4>는 국립중앙도서관의 소장자료와 고신문의 디지털화 현황을 보여주고 있다.

<표 3> 국립중앙도서관 소장자료 디지털화 DB 구축 현황: 2022. 12.

구분	고신문	고문헌 (고서·고지도)	일반서	잡지·학술지	비도서	관보	합계
콘텐츠 (건)	99종 7,678,320건	97,130건	1,490,321건	189,987건	250,001건	336,239건	10,041,998건
	76.5%	0.9%	14.9%	1.9%	2.5%	3.3%	100%

<표 4> 국립중앙도서관 고신문 디지털화 구축 현황: 2022. 12.

① 구축 대상	② 구축(22년까지)	③ 민간 구축*	④ 구축 잔여물량 (①-(②+③))
135종	99종	5종	31종

* 조선일보, 동아일보, 경향신문, (구)서울신문, 한국일보

그리고 한국언론진흥재단에서 텍스트 DB로 구축한 한성순보 등 13종, 28,680면, 364,655건의 신문기사에 대해서도 대한민국 신문 아카이브에서 함께 서비스하고 있다. 다만, 이들 텍스트는 AI OCR 기반으로 구축한 것이 아니고 한국언론진흥재단에서 일일이 사람의 손으로 입력하여 구축한 것이다.[7]

7 텍스트 DB 구축(13종): 남선신문, 남조선민보, 대한매일신보, 대한매일신보(국한문), 대한매일신보(영문), 독립신문(서재필), 독립신문(영문), 매일신문, 한성순보, 한성주보, 협성회회보, 황성신문(이중 대한매일신보, 마산일보 2종은 텍스트 DB 구축이 완료되지 않고 잔여물량이 있음)

III. 고신문 온라인서비스 플랫폼, 대한민국 신문 아카이브[8]

1. 서비스 내용

국립중앙도서관은 대한민국 신문 아카이브(https://nl.go.kr/newspaper/)를 통해 디지털화 사업으로 구축한 신문에 대해 이용자가 편리하게 찾아 활용할 수 있도록 서비스하고 있다. [그림 2]는 대한민국 신문 아카이브 서비스의 메인화면이다.

[그림 2] 대한민국 신문 아카이브 서비스

8 https://nl.go.kr/newspaper/. 검색일: 2023. 9. 30.

대한민국 신문 아카이브는 구축한 신문 99종에 대한 간략 소개(해제)와 함께 신문의 이미지와 기사에 대해 신문명과 발행일, 키워드와 문장으로 검색할 수 있게 하는 서비스이다. 신문의 텍스트가 있는 경우에는 텍스트를 제공하고 있어 복사 및 다운받아 활용할 수 있으며, 기사별 LOD(Linked Open Data)를 발행하고 있다([그림 3], [그림 4] 참조).

신문기사 이미지 신문기사 텍스트

[그림 3] 신문기사의 이미지 및 텍스트 서비스

三國 아파트 倉庫 火災 at lod.nl.go.kr
http://lod.nl.go.kr/resource/CNTS-00094996967

dcterms:accessRights	nlk:license0
nlon:antecedentSource	reformatted digital (xsd:string)
nlon:audienceNote	일반이용자 (xsd:string)
nlon:datePublished	2023-09-07T01:06:41 (xsd:dateTime)
dcterms:description	발행일 : 19391216 (xsd:string) 원소장처 : 한국연구원 (xsd:string) 창간일 : 19641427, 一千九百五年八月十一日 (xsd:string) 한국언론진흥재단에서 원문 파일을 기증받아 이미지 파일로 변환하였음 (xsd:string)
bibframe:extent	이미지파일 10개 (5M) (xsd:string)
dc:format	image/jpeg (xsd:string)
nlon:genre	신문 (xsd:string)
dcterms:hasFormat	nlk:KSE000002737
nlon:holdingInstitution	국립중앙도서관 (xsd:string)
nlon:keyword	기사 (xsd:string) 내자정삼국아파트 (xsd:string) 사회 (xsd:string)
rdfs:label	三國 아파트 倉庫 火災 (xsd:string)
nlon:medium	전자자료(Image) (xsd:string)
nlon:newsPosition	2면 10단 (xsd:string)
nlon:otherNumber	084 (xsd:string)
nlon:publicationPlace	한국 (xsd:string)
nlon:reproductionNote	서울: 국립중앙도서관, 2017 (xsd:string)
dcterms:title	三國 아파트 倉庫 火災 (xsd:string)
nlon:titleOfHostItem	每日新報 (xsd:string)
rdf:type	nlon:ElectronicJournal nlon:OnlineMaterial
nlon:typeOfResource	nlk:text
nlon:uci	G701:B-00094996967 (xsd:string)

This page shows information obtained from the SPARQL endpoint at http://lod.nl.go.kr/sparql.
As Turtle | As RDF/XML | As JSON | As N3 | As nTriple

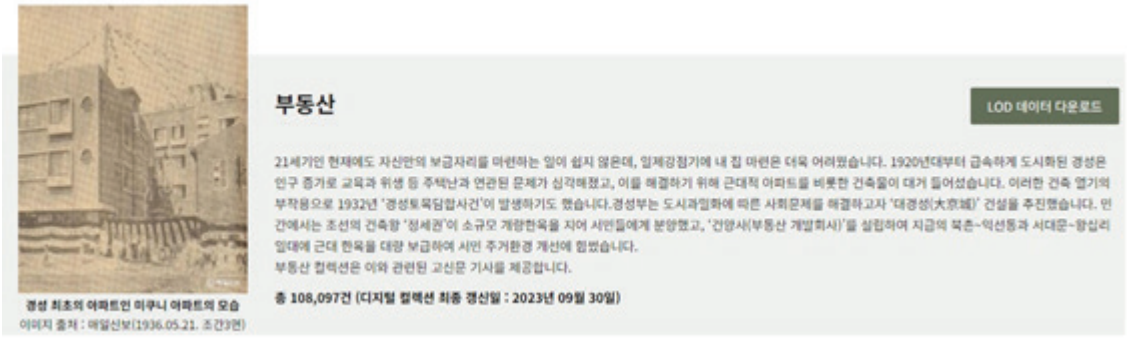
- Powered by National Library of Korea Linked Open Data using OntoBase (Triple Store)

[그림 4] 대한민국 신문 아카이브 LOD 발행

더불어 신문 브라우징을 통해 좀 더 다양한 기사에 편리하게 접근할 수 있도록 하고 있다. 즉, 주제 6개(사회·문화·정치 등), 유형 12개(기사·광고·소설 등), 소재 20,896개(의약품·책·화장품 등), 인물

5,083명(이승만·이순신·이광수 등), 단체 1,992개(국제연합·조선은행 등), 장소 5,667건(조선·부산·미국·서울역 등)에 대해 AND·OR·NOT 조건으로 신문 및 기간을 선택하여 검색할 수 있다.

한편, 시대별 4건(대한제국, 일제강점기, 군정기, 대한민국)과 부동산·물가·전쟁 등 13개의 주제에 대해 디지털 컬렉션을 구축하여 서비스하고 있는데, 신문을 통해 그 시대의 사회상을 엿볼 수 있어 흥미롭다([그림 5] 참조).



대한민국 최초 아파트 '경성(지금의 서울) 미쿠니 아파트의 모습'(매일신보, '36. 5. 21. 조간3면)

[그림 5] 디지털 컬렉션 서비스

2. 국내외 서비스 사례

가. 네이버 뉴스 라이브러리[9]

네이버 뉴스 라이브러리(https://newslibrary.naver.com/)는 2011년 4월 7일부터 경향신문(1946. 10. 6.~1999. 12. 31.), 동아일보(1920. 4. 1.~1999. 12. 31.), 매일경제(1966. 3. 24.~1999. 12. 31.), 조선일보(1920. 3. 9.~1999. 12. 31.), 한겨레(1988. 5. 15.~1999. 12. 31.) 등 주요 일간지 5종의 이미지와 텍스트 DB를 구축하여 서비스하고 있다. 날짜·키워드·상세(신문사·지면·섹션 등) 검색과 검색 결과에 따른 원문보기 및 한글변환 기능, 텍스트 보기 및 마이스크랩 기능을 제공하고 있다([그림 6] 참조).

9 https://newslibrary.naver.com/. 검색일: 2023. 10. 9.

[그림 6] 네이버 뉴스 라이브러리

나. 호주 국립도서관 누리집, TROVE의 신문 서비스[10]

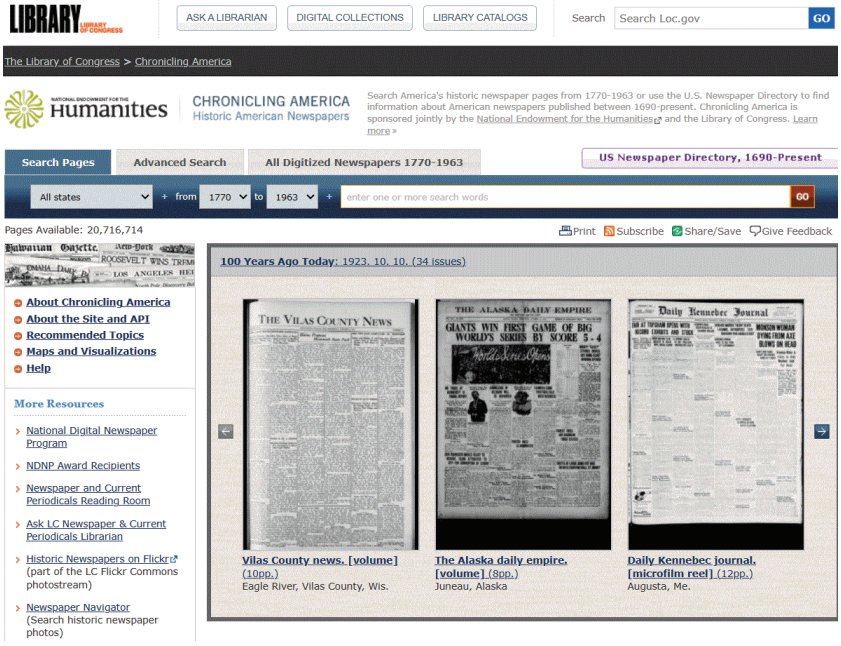
호주 국립도서관은 Albur and District Historical Society, NSW(뉴사우스웨일스) 지역문화기금 등 다양한 기금과 실물자료를 소장하고 있는 기관(예를 들어, NSW 주립도서관 등)으로부터 신문 마이크로필름 또는 신문 실물을 제공받아 디지털화를 하고 누리집 TROVE(https://www.trove.nla.gov.au)를 통해 무료로 읽을 수 있도록 서비스하고 있다. 현재 저작권이 자유로운 1803년에서 1954년까지 발행된 1,000개 이상의 신문에 대한 수백만 페이지의 디지털화된 이미지와 텍스트를 제공하고 있다. 또한 누구라도 접근하여 뉴스 기사의 텍스트를 수정할 수 있도록 '텍스트 수정 가이드'를 누리집으로 제공하고 있다([그림 7] 참조).

[그림 7] 호주 TROVE 신문 서비스

10 https://www.trove.nla.gov.au. 검색일: 2023. 10. 9.

다. 미국의회도서관의 Chronicling America[11]

Chronicling America의 NDNP(National Digital Newspaper Program, 전국 디지털 신문 프로그램)는 1770년에서 1963년까지 미국에서 발행된 역사적인 신문의 디지털화 자료에 대한 접근을 제공하는 웹사이트이다. 미국의회도서관은 2003년부터 NEH(National Endowment for the Humanities)와 파트너십을 체결하고, 미국의 모든 주와 지역의 도서관(공공/대학), 기록관, 역사학회 등 50여 개 처에 대하여 디지털화한 신문의 서지정보를 제공하고, 선택적 보존을 위한 기술지원을 통해 영구적으로 디지털화한 신문에 접근이 가능하게 하기 위한 목적으로 NDNP 프로젝트를 시작하였다. 기계학습 기술을 활용하여 1900년부터 1963년까지 156만 개의 신문 속 그림, 사진, 일러스트, 만화 등을 추출하여 신문 내비게이터(Newspaper Navigator)에서 검색(장소·기간·키워드)할 수 있도록 제공하고 있다([그림 8], [그림 9], [그림 10] 참조).

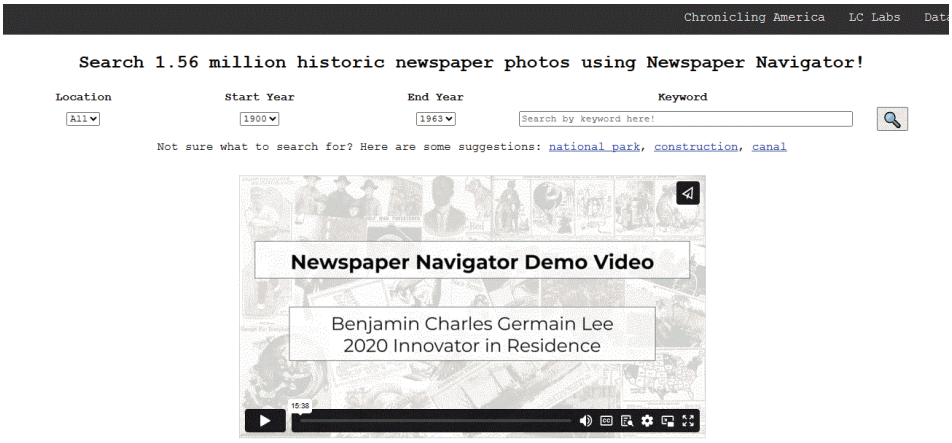


[그림 8] 미국의회도서관의 신문 서비스

11 https://chroniclingamercialoc.gov/about. 검색일: 2023. 10. 10.

[그림 9] 미국의회도서관의 NDNP

Newspaper Navigator 첫화면



검색결과

[그림 10] 미국의회도서관의 Newspaper Navigator

3. 서비스 발전 방안

고신문은 형태적으로 볼 때 세로쓰기와 가로쓰기로 되어 있고, 한글, 한글고어, 한문, 일본어가 혼용되어 있다. 따라서 기사 및 키워드를 추출하여 DB로 구축하는 작업은 한문과 근현대 일본어를 읽고 이해할 수 있는 능력이 필요한 높은 수준의 작업이다. 고신문은 대한민국 근현대를 다양한 관점에서 연구 및 교육하기 위한 대상자료로 활용되고 있고, 기사 작성 및 콘텐츠 제작의 소재 발굴 등에 필요한 귀중한 원천데이터이다. 또한 고신문에 실린 각종 그림 사진, 일러스트, 만화, 표, 광고 등을 통해서도 그 당시의 사건·사고뿐 아니라 관심사와 이슈 등을 엿볼 수 있다.

따라서 1960년 이전에 설립된 신문사 및 교육·문화기관과 협력하여 국립중앙도서관에 없는 고신문이나 결호를 발굴하고, 디지털화 지원을 통해 디지털 원문을 확보하여 대한민국 신문 아카이브에서 서비스할 필요가 있다. 그리고 인터넷 아카이브(Internet Archive, https://archive.org)와 같이 원문을 제공하고 있는 사이트를 통해서도 대한민국이 미소장하고 있는 고신문의 원문을 확보하여 서비스할 필요가 있다. 『선봉』[12]이나 『해조신문』[13]과 같이 1945년 해방 이전에 해외에서 발간된 고신문을 발굴하고 디지털원문을 확보하여 서비스하려는 노력 또한 필요하다. 그리고 인공지능 기술을 활용하여 고신문의 글자 자체에 대한 텍스트 DB와 내용에 대한 번역 DB를 각각 구축하는 한편, 신문에 있는 의미있는 이미지들을 DB로 구축하고 검색 및 내려받기할 수 있는 시스템을 갖춘다면 연구자, 기자, 교육자 및 창작자들에게 활용도가 높은 플랫폼이 될 것이다. 텍스트 DB, 내용에 대한 번역 DB 및 이미지 DB의 구축에는 많은 시간과 예산이 수반되기 마련이다. 누구라도 접속하여 뉴스기사의 텍스트, 번역한 내용을 수정하고 업로드할 수는 시스템과 함께 '텍스트·내용 수정 가이드'를 누리집으로 제공한다면 작업자들은 좀 더 쉽게 동 작업을 수행할 수 있을 것이다.

이와 같이 고신문 서비스 종수를 늘리고 결호를 메꾸는 노력과 함께 다양한 분야에서 활용될 수 있도록 첨단 디지털기술을 활용하여 고신문 전반에 관한 내용 이미지와 키워드 검색을 제공한다면 대한민국의 근현대를 조망하는 주요 정보원으로써 명실상부한 대한민국 신문 아카이브 포털로 자리매김하게 될 것이다.

12 『선봉』은 중국 연해주에서 1923년 3월 1일 창간, '三月一日'이라는 제호로 창간되었고, 4호부터 선봉으로 제호를 변경하여 발간되었다. 1938년부터는 카자흐스탄에서 '레닌기치'라는 제호로 발행되다가, 1990년 폐간되었다(이희용, 2017).

13 『해조신문』은 1908년 2월 26일 러시아 블라디보스토크에서 창간된 최조의 한글신문이다(이희용, 2017).

참고문헌

네이버 뉴스라이브러리. https://newslibrary.naver.com

네이버 두산백과 <한국의 신문 역사>. https://terms.naver.com

대한민국 신문 아카이브. https://nl.go.kr/newspaper/

미국의회도서관 신문 DB. https://chroniclingamercialoc.gov/about/

미국의회도서관 전국 디지털 신문 프로그램. https://www.loc.gov/ndnp

미국의회도서관 LABS. https://news-navigator.labs.loc.gov

이희용(2017. 4. 25.). “이희용의 글로벌 시대: 동포 언론인 단체 통합에 거는 기대”. 연합뉴스, 출처: https://yna.co.kr/view/AKR20170424146400371?nput=1195m

호주국립도서관 트로브. https://www.trove.nla.gov.au