

일제강점기 일본어신문 『조선시보』의 메타데이터 요소 개발과 적용 연구*

Metadata Elements Design and Application for Japanese Newspaper ‘Chosunsibo’ Issued in Colonial Korea

노 지 현(Jee-Hyun Rho)**

<목 차>

I. 서론	1. 메타데이터 요소 도출과 요소 값 정의
II. 메타데이터 요소 개발을 위한 사전 검토	2. 검증 및 crosswalk 작성
1. 일제강점기 신문DB 구축 현황	3. 메타데이터 입력기의 개발
2. 『조선시보』의 형태적·내용적 특징	IV. 신문기사 메타데이터 개발을 위한 제안
III. 『조선시보』 메타데이터 요소의 개발	V. 결론

초 록

본 연구의 목적은 일제강점기 일본어 신문 중 자료의 방대한 양이나 사료로서의 가치에도 불구하고 아직까지 신문기사 검색이 불가능한 『조선시보』의 메타데이터 구축 방안을 마련하는데 있다. 이를 위해, 본 연구에서는 (1) 일제강점기 신문DB구축 현황과 메타데이터 요소를 분석하고, (2) 『조선시보』의 형태적·내용적 특성을 파악한 다음, (3) 메타데이터 설계를 위한 기본원칙을 정한 후 『조선시보』에 적용할 메타데이터 요소 및 인코딩 스킴, 통제어휘를 도출하였다. 이러한 메타데이터 설계(안)은 (1) 데이터 입력 테스트를 통한 검토 및 수정, (2) DB 전문기관의 자문, (3) 관련 표준 및 관련 사례와의 crosswalk 작성 등의 과정을 거쳐 확정되었다. 본 연구의 범위는 『조선시보』에 적합한 메타데이터 요소의 개발 및 이를 토대로 데이터의 입력을 위한 입력기 개발까지로 설정하였다.

키워드: 메타데이터, 메타데이터 요소 설계, 신문기사, 신문DB, DB구축, 조선시보, 일제강점기 신문

ABSTRACT

This study aims to develop metadata elements and vocabularies of ‘Chosunsibo’ issued in Korea during the Japanese colonial era. Despite the vast amount and historical values, newspaper articles in Chosunsibo are still unsearchable and difficult to access. To the end, this study includes as follows: (1) analyzed the DB construction and metadata elements of newspapers issued in Korea during the Japanese colonial, (2) investigated the morphological characteristics and content of ‘Chosunsibo’, and (3) derived metadata elements, encoding schemes and controlled vocabularies for ‘Chosunsibo’. The final draft was verified in three methods: (1) review and revise through data input tests, (2) consulting by a DB expert agency, and (3) drawing up of crosswalks with metadata standards and related cases. The scope of this study is to design the metadata schema of the historical newspaper ‘Chosunsibo’ and develop a data input system based on that schema.

Keywords: Metadata, Metadata element, Metadata element design, Digital newspaper, Chosunsibo, Chosunjiho

* 이 논문은 2017년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A5B4055721)

** 부산대학교 문헌정보학과 교수(jhrho@pusan.ac.kr)

•논문접수: 2019년 11월 18일 •최초심사: 2019년 11월 27일 •게재확정: 2019년 12월 11일

•한국도서관·정보학회지 50(4), 137-158, 2019. [http://dx.doi.org/10.16981/kliiss.50.201912.137]

I. 서론

신문은 시대의 역사를 고스란히 담고 있는 기록이다. 신문은 정치, 경제, 사회, 문화 등 각 분야의 실정은 물론이고 당대의 주요 사건사고와 오락·광고, 그리고 사회 각계각층의 다양한 여론을 담고 있어 역사적인 사실을 검증하거나 시대사 연구에 있어 중요한 자료로 활용되고 있다. 이러한 의미에서 볼 때 일제강점기에 발간된 신문은 일본의 조선 지배정책이나 생활상, 당대의 담론 등을 파악하기 위한 식민시기 연구에 있어 필수불가결한 자료라 할 수 있다. 이에 따라 국립중앙도서관을 비롯한 공공기관과 민간기관에서는 일제강점기에 발간된 신문에 대한 DB 구축사업을 지속적으로 추진하고 있다. 그러나 일제강점기 신문에 대한 연구 성과나 DB 구축이 주로 ‘국한문’ 신문이나 ‘민족지’적 성격의 신문에 집중되어 있어, 당시 조선에 거주하던 일본인이 발행한 일본어 신문에 대한 사회적 관심은 상대적으로 저조한 편이라 할 수 있다.

일제강점기 조선에서 일본인들의 언론 활동은 생각보다 활발하였고, 그것이 식민지 정치·경제·사회·문화에 미친 영향이 적지 않았다(전성현 등 2019, 370). 1876년 개항 이후 조선에 진출한 일본인들은 개항장을 중심으로 조선의 각 도시에 진출하여 일찍부터 현지 신문을 발행하였다. 1910년 이전에 조선 각 지방에서 발행된 일본인 신문은 무려 40여 종에 달하였으며, 1910년 강제병합 이후에 조선총독부의 기관지 역할을 했던 『매일신보』를 제외하고 대부분의 국한문 신문이 강제 폐간되면서 일본인이 발행한 일본어 신문은 오히려 전성기를 맞이하였다(홍순권 2012, 18). 이러한 일본인의 언론 활동은 1941년 전시동원체제를 강화하기 위해 실시한 ‘一道一紙制’에 따라 지역신문이 강제 통폐합되기 전까지 지속되었다. 상황이 이러한데도 일제강점기 일본어 신문에 대한 학계의 관심이 저조한 이유는 무엇보다도 자료의 ‘결핍’과 밀접한 관련이 있다. 현재까지 남아있는 일본어 신문은 경성에서 발행된 『경성일보』(1906~1945)와 『조선신문』(1908~1942), 그리고 부산에서 발행된 『부산일보』(1907~1945)와 『조선시보』(1894~1941) 등 단 4종에 불과하기 때문이다.

현존하는 4종의 일본어 신문에 대한 원문 구축은 이미 완료되어 웹을 통해 공개되고 있다.¹⁾ 그러나 신문기사에 대한 메타데이터는 『조선신문』과 『부산일보』만 구축되어 있어 실질적으로 신문기사 검색은 이 2종의 신문에 대해서만 가능한 상태이다. 그러다보니 일제강점기 일본인의 언론 활동과 시대적 상황에 총체적으로 접근하는데 한계가 있고, 특히 신문매체의 특성상 동일한 사건을 대하는 시각이나 사실정보의 전달이 신문을 발행한 주체에 따라 다르거나 특정 목적을 위해 의도적으로 왜곡하고 조작하는 경우도 있어 당시 발행되었던 다른 매체와의 교차

1) 4종의 일본어 신문에 대한 원문DB의 구축 주체가 상이한데, 『경성일보』는 한국학술정보(http://www.kstudy.com/kyungsung_new/)에서, 『조선신문』은 국립중앙도서관의 대한민국 신문 아카이브(<http://nl.go.kr/news/paper/sub0101.do>)에서, 그리고 『부산일보』와 『조선시보』는 국사편찬위원회 한국사데이터베이스(<http://db.history.go.kr/>)에서 각각 원문을 제공하고 있다.

검증이 필수적으로 이루어져야 하지만 이 또한 어려운 상황이라 할 수 있다. 앞으로 일제강점기에 일본인에 의해 발행된 일본어 신문에 대한 추가적인 발굴은 물론이고 이에 대한 DB 구축 사업을 보다 폭넓게 그리고 시급히 추진해 나가야 하는 이유가 바로 여기에 있다.

본 연구의 궁극적인 목적은 일제강점기 일본어 신문 중 자료의 방대한 양이나 사료로서의 가치에도 불구하고 아직까지 신문기사 검색이 불가능한 『조선시보』를 대상으로 메타데이터를 구축함으로써 개별 신문기사에 대한 효율적인 접근을 지원하는데 있다. 본 연구의 대상인 『조선시보』는 강제병합 직전 일본인의 인구 밀도가 가장 높았던 부산에서 일본인이 발행한 일본어 신문이다. 현존하는 다른 일본어 신문인 『경성일보』, 『조선신문』, 『부산일보』에 비해 발행 시기가 10여년 앞서며, 상업 일간지이지만 정치, 경제, 해운, 기술, 통신, 문예, 인사, 교육, 생활 등 전 주제 영역의 기사를 수록하고 있어 식민시기 조선에 거주한 일본인의 정치경제적 활동을 해명하는데 중요한 자료라 할 수 있다. 뿐만 아니라 부산권역을 중심으로 하는 일본인의 지역사회 동향을 반영하고 있어 지역사 연구는 물론이고 기존의 경성 중심의 근대사 연구를 전국으로 확장시키는데 근거 자료로도 활용할 수 있다. 그러나 『조선시보』는 근대 초기의 일본어로 표기되어 있어 자료 해석과 판독이 어렵다는 이유로 2009년 원문 구축이 완료된 이후에도 제대로 활용되지 못하고 있다. 이에 본 연구에서는 일제강점기 일본어 신문의 특성을 잘 반영하면서도 정확하고 상세한 메타데이터 구축 방안을 마련하고자 한다.

『조선시보』 메타데이터 요소 개발을 위한 절차는 다음과 같이 설정하였다: (1) 일제강점기 신문DB 구축 현황과 메타데이터 요소를 분석하고, (2) 일본어 신문인 『조선시보』의 형태적·내용적 특성을 파악한 다음, (3) 『조선시보』에 적용할 메타데이터 요소 및 인코딩 스킴(encoding scheme)을 개발하고, (4) 관련 표준 및 관련 사례와의 비교 및 상호운용성(interoperability)의 확보를 위해 crosswalk를 작성하였다. 메타데이터 설계 과정에서 특히 중점을 두고자 한 것은 향후 신문기사 검색 및 이용을 고려하여 (1) 기사제목이나 발행일자 뿐만 아니라 관련 인물, 주제, 시대, 사건, 지리/장소 등 다양한 접근점을 제공하기 위한 방안과 (2) 근대 초기 일본어의 해독이 어려운 이용자들을 위해 일본어 원어와 한글 표기를 병기하는 방안을 마련하는 것이었다. 연구의 범위는 『조선시보』에 적합한 메타데이터 요소의 개발 및 이를 토대로 데이터의 입력을 위한 입력기 개발까지로 설정하였다. 연구진행 과정에서 언어적·내용적 전문성을 확보하기 위해 본 연구는 일본어 전문가 및 근현대사 연구자들과의 협업으로 진행하였다.

II. 메타데이터 요소 개발을 위한 사전 검토

1. 일제강점기 신문DB 구축 현황

4 한국도서관·정보학회지(제50권 제4호)

신문은 부피가 크고 지질이 열악해 오래 보관하기가 곤란하며 다른 매체에 비해 기대수명이 높지 않은 편이다. 또한 종이신문은 이용이 불편하고 수장 공간이나 보존 환경의 문제를 야기하기 때문에 일반적으로 마이크로매체나 PDF, TIFF 파일과 같은 전자 형태로 보존하면서 신문기사 검색 및 원문서비스를 제공하는 방식을 취하고 있다. 일제강점기에 발행된 근대 신문은 현재 국립중앙도서관의 「대한민국 신문 아카이브」, 국사편찬위원회의 「한국사데이터베이스」, 독립기념관의 「한국독립운동정보시스템」 등 세 곳에서 원문 DB와 신문기사에 대한 메타데이터의 구축과 통합을 주도하고 있다. 그 외에 도서관과 신문사, 민간 DB업체에서 원문만 혹은 원문과 메타데이터를 자체적으로 구축하여 제공하는 경우도 있다. 이렇게 구축된 DB는 한국언론재단 「빅카인즈(Big Kinds)」의 ‘고신문 검색’과 「한국역사정보통합시스템」을 통해서도 서비스되고 있다.

1910년 강제병합부터 1945년 광복에 이르기까지 일제강점기에 발행된 근대 신문은 아직까지 어떤 신문이 발행되었는지에 대한 기초적인 데이터조차 정리되지 않아(미즈노 나오키 2007, 254) 발행 현황을 전체적으로 조망하기가 어렵다. 다만 추이를 살펴보면, 강제병합 직후에 민족지 성격의 신문들이 모두 폐간 조치되고, 『매일신보』(국문판), 『경성일보』(일문판), 『The Seoul Press』(영문판)와 같은 총독부 기관지만 남거나 『신한국보』(하와이), 『독립신문』(상하이), 『대한독립신문』(만주) 등 외국에서 교포신문이 주로 발행되었던 것으로 나타난다. 이와 대조적으로, 식민지 조선에서는 『조선신문』, 『경남일보』, 『경성신보』, 『부산일보』, 『조선상공신문』 등 일본인이 발행한 신문이 전성기를 누렸다. 1919년 3·1 운동을 기점으로 『조선독립신문』과 같은 지하신문에서부터 『조선일보』, 『동아일보』, 『시대일보』, 『중외일보』, 『민중신문』 등 총독의 시정을 비판하거나 민족자주독립사상을 고취하는 민간지가 발행되었다. 이외에 『기독교신보』, 『천주교회보』, 『기독교신문』과 같은 종교 신문도 발행되었다.

이들 일제강점기 근대 신문은 2019년 10월 현재 국립중앙도서관의 「대한민국 신문 아카이브」에서 14종, 국사편찬위원회의 「한국사데이터베이스」에서 8종, 독립기념관의 「한국독립운동정보시스템」에서 14종, 그리고 한국학술정보(KISS)에서 1종 등 총 37종을 DB로 구축하여 제공하고 있다.²⁾ 이 가운데, 『조선시보』를 비롯한 8종의 신문에 대해서는 원문 이미지만 제공되고 있을 뿐 메타데이터 구축이 아직 진행되지 않아 신문기사 검색이 불가능한 상황이다.

신문기사에 적용된 메타데이터는 구축 주체에 따른 차이를 보이고 있다. 동일한 주체에 의해 DB구축이 추진된 경우에도 대상 신문에 따라 상세성이나 기술 범위에 차이가 있는데, 대

2) 「시대일보」 → 「중외일보」 → 「중앙일보」 → 「조선중앙일보」와 같이 신문의 체호가 변경되어 간행된 신문은 이를 별개의 신문으로 간주하여 산출한 결과이다. 또한, 원본의 소장처와 원문 이미지 구축 기관이 상이한 경우도 있다. 예를 들어, 「중외일보」의 경우 원본은 종로도서관에 소장되어 있지만, 한국언론진흥재단에서 이를 원문 이미지 파일로 변환하였고, 국립중앙도서관에서 메타데이터를 구축한 후 「대한민국 신문 아카이브」를 통해 신문기사 검색과 원문을 제공하고 있다. 이 경우, 해당 신문의 DB 구축기관은 최종적인 서비스 기관으로 간주하였다. 한편, 현재 제공되는 각 신문의 DB 구축 범위는 자료의 소실이나 훼손으로 인해 실제 신문의 발행기간과 상이한 경우가 많다.

체로 세 기관에서 적용하고 있는 일제강점기 신문의 메타데이터 요소는 다음 <표 1>과 같다. 세 기관에서 공통적으로 적용하는 요소는 <기사제목>, <신문명>, <발행일> 등 3개 요소에 불과하며, 나머지는 구축 주체에 따라 차이가 있다. 동일한 성질의 데이터에 대한 적용 방식에서도 차이가 드러나는데, 가령 국립중앙도서관에서는 <기사제목>과 <부제목> 요소를 구분하고 있지만 국사편찬위원회와 독립기념관에서는 <부제목>을 따로 두지 않고 <기사제목>에 통합하여 적용하고 있다. 신문기사의 위치 정보도, 국립중앙도서관에서는 <기사위치>에 해당 지면과 단의 위치를 함께 수록하고 있지만, 국사편찬위원회에서는 <면수>와 <단수>로 분리하여 적용하고 있음을 알 수 있다. 세 기관 중에서 개인의 이름, 단체명, 지명, 회사조합 등에 대한 색인어를 구축하는 사례는 국사편찬위원회의 「한국사데이터베이스」가 유일하다.

<표 1> 일제강점기 신문기사에 대한 구축 주체별 메타데이터 적용 요소

국립중앙도서관 「대한민국 신문 아카이브」	국사편찬위원회 「한국사데이터베이스」	독립기념관 「한국독립운동정보시스템」
기사제목	기사제목	기사제목
부제목		
신문명	신문명	신문명
발행처		
발행일	발행년월일	발행일
기사위치	면수 단수	
기사유형	기사형식*	
기사주제		
주기사항 ³⁾		
관련 주제어		
키워드		
	주제분류*	
	색인어(이름, 단체, 지명 등)	
	발행구분	
		생산정보(신문명+제○호)
		요약정보
	발신지(장소)	

* 표시는 일제강점기 신문 전체가 아니라 일부 신문에만 적용되는 요소임

데이터의 표현방식에서도 세 기관 간의 차이는 뚜렷하다. 신문기사의 핵심 주제어에 대해, 국립중앙도서관에서는 「국립중앙도서관 주제명표목표」를 적용하여 ‘주제어휘’(예: ‘농업[農業]’, ‘정치(권력)[政治]’, ‘의술[醫術]’)로 표현하고 있지만, 국사편찬위원회에서는 일제강점

3) 주기사항에는 일본연호로 표기된 발행일자와 서력의 한자표기, 신문의 창간일, 원문의 소장처나 원문 이미지 구축 기관 등에 대한 정보 등이 포함되어 있다.

기 신문기사 분류를 위해 자체 개발한 「주제분류표」⁴⁾를 적용하여 ‘디렉토리’ 형태(예: ‘사회 > 여성 > 여성단체 > 기타여성단체’)로 주제를 나타내고 있기 때문이다. 독립기념관에서는 이러한 주제어 데이터 없이 <요약정보>를 제공하고 있다. <발행일> 또한 서력으로 환산하여 표기하거나 서력과 일본연호를 함께 표기하는 등의 차이도 있다. 이외에도, 같은 일본어 신문이라 하더라도 일본어로 쓰여진 기사제목을 일본어로만 표기하거나(예: 『조선신문』), 일본어 대신 한자의 한글음이나 한글로 번역하여 표기한 사례도 있다(예: 『부산일보』). 국한문 신문도 한자 이체자를 정자나 한글로 변환하여 적용하는 등, 일제강점기에 발행된 근대 신문에 대해 적용되는 메타데이터는 적용 요소는 물론이고 데이터의 표현방식에서도 적지 않은 차이를 보이고 있다.

메타데이터를 구현하는 스키마로 국립중앙도서관에서는 MODS를 사용하고 있다. 국사편찬위원회와 독립기념관은 텍스트 형태의 데이터만 제공하고 있어 표준 스키마의 적용 여부에 대한 확인은 불가능하다. MODS를 사용하는 국립중앙도서관의 일제강점기 신문기사에 대한 메타데이터는 <표 1>에 제시된 요소 외에, 자료유형(<genre>), 발행지(<place>), 발행간기(<frequency>), 언어(<language>), 파일 수와 크기(<physicalDescription> <extent>), 이용대상자(<targetAudience>), 소장기관(<physicalLocation>), 메타데이터 생성일자(<recordCreationDate>), 메타데이터 수정일자(<recordChangeDate>) 등도 포함하고 있다. 또한 장소, 언어, 날짜 등에는 국제적으로 널리 사용되는 표준 인코딩 스키마를 적용하고 있음을 확인할 수 있다.

2. 『조선시보』의 형태적·내용적 특징

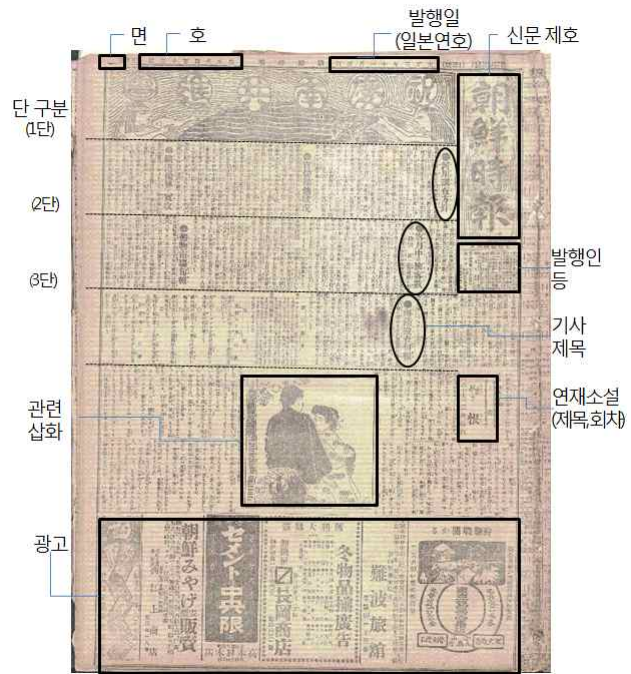
『조선시보』 신문기사의 메타데이터 구축을 위해서는 『조선시보』의 형태적·내용적 특징에 대한 분석도 선행될 필요가 있다. 『조선시보』는 부산에 본사를 둔 ‘조선시보사’에서 창간한 일본어 상업 일간지이다. 1892년 12월 5일에 창간된 『부산상황』이 『동아무역신문』으로 개제되고, 이를 명성황후 시해 사건의 실행 멤버였던 아다치 겐조(安達謙藏)가 인수한 후 제호를 『조선시보』로 바꾸어 1894년 11월 21일부터 1941년 5월까지 발행하였다. 발행 초기인 1910년경의 기록에 따르면, 부산지역에 거주한 일본인 가구의 1/5 정도가 구독할 만큼 발행 부수가 많았으며, 조선 이외에 일본이나 청 등에게까지도 보급되었다(채백 2012, 96-97; 212). 조선 내에서도 전국 주요 도시에 지사를 설치하여 전국적인 배포를 시도하였다. 그러나 같은 지역에서 발행된 『부산일보』의 시세 확장과 1941년에 실시된 총독부의 언론 통폐합 방침에 따라 『부산일보』로 통합되면서 폐간되었다. 『조선시보』의 원본은 현재 부산광역시립

4) 국사편찬위원회의 한국사데이터베이스에서는 일제강점기 자료의 분류를 위해 자체적으로 <주제분류표>를 개발하여 사용하고 있다. 근현대신문에 적용되는 <주제분류표>는 9개 대주제, 82개 중주제, 413개 소주제, 1,858개 최하위주제로 구성되어 있다.

시민도서관에 소장되어 있으며, 이에 대한 원문 이미지는 국사편찬위원회에서 구축하여 「한국사데이터베이스」를 통해 제공되고 있다. 현재까지 남아 있는 『조선시보』는 1914년 11월 2일부터 1940년 8월 30일까지의 발간분이며, 지면으로 약 17,287면, 기사 건수로 약 345,740건의 분량에 이른다.⁵⁾

『조선시보』는 당시의 다른 신문에 비해 상업기사, 생활정보기사 및 광고의 비중이 특히 높은 편이다. 이는 『조선시보』가 일본 농민의 부산 이주를 장려하고 부산을 일본에 크게 알리는 언론의 역할을 담당하였기 때문으로 추측된다. 전체적인 내용은 사실을 전달하는 보도기사와 논설·사설, 연재기사, 기고문, 행사나 인사 등을 알리는 단신, 연재소설 등으로 구성되어 있다. 표기 문자는 근대 초기의 일본어로 기록되어 있어 일본어 전문가가 아니면 해석이 어려운 경우도 있다. 일본어 신문이므로 한국어는 일본문자로 표기되어 있으나, 광고에 일본어와 한국어가 병기된 사례도 있다. 발행 면수는 대체로 4면 또는 6면이지만, 최소 1면에서 최대 18면으로 발행되기도 하였다. 또한 부산지역을 근거로 한 지역신문이지만 부산 이외의 지역에 위치한 지국에서 기사를 작성하여 신문의 한 면을 <경북판>, <양산판>, <경성판> 등의 제호로 발행하였다는 점이 특이점이다.

신문의 형태적 특징을 좀 더 자세히 살펴보면, 1면에는 신문 제호와 발행인·인쇄인 등의 정보가 기록되어 있고, 모든 면의 상단에 발행일(일본연호 사용), 호수 및 면수, 신문발행 인가사항 등이, 그리고 하단에 광고가 배치되어 있다. 단 구분은 별도로 표시되어 있지 않으나 구분 선이 있어 상단부터 1단, 2단, 3단 등으로 구분이 가능하다. 또한 신문기사의 제목은 ● 표시 다음에 나타나기 때문에 식별이 용이한 편이다. 신문기사의 제목은 한 줄 또는 두 줄로 표기되어 있으며, 제목보다 크기가 작은 활자나 ▲ 표시 다음에 부제가 나오는 경우도 있다. 그밖에 신문기사나 연재소설 등과 관련된 삽화나 사진이 존재하거나, 삽화나 사진에 제목과 작성



<그림 1> 『조선시보』의 지면 구성과 내용

5) 『조선시보』의 발행 분량은 부산시립시민도서관에 보관된 원본 및 국사편찬위원회 「한국사데이터베이스」에서 제공하는 원문 이미지를 대상으로 연구팀에서 추정된 추정치이다. 요코하마 신문 박물관 등 일본 현지에도 『조선시보』 원본 일부가 남아 있는 것으로 파악되고 있으나, 누락본의 소장여부는 아직 확인되지 않고 있다.

자의 이름이 부기된 경우도 있다. 조선시보의 지면 구성과 기사 내용은 <그림 1> 및 <그림 2>와 같다.



<그림 2> 「조선시보-경북판」(왼쪽) 및 기사제목의 표기 사례(오른쪽)

Ⅲ. 『조선시보』 메타데이터 요소의 개발과 적용

1. 메타데이터 요소 도출과 요소 값 정의

일제강점기에 발간된 다수의 신문에 적용된 메타데이터와 『조선시보』의 형태적·내용적 특징에 대한 분석을 토대로 하여, 『조선시보』 신문기사에 대한 메타데이터 요소를 개발하였다. 요소 개발에 앞서, 연구팀은 연석회의를 통해 다음과 같은 기본 원칙을 설정하였다.

- (1) 데이터의 성격을 고려하여 신문기사와 광고 메타데이터는 분리하여 설계한다.
- (2) 기사제목이나 발행일자 뿐만 아니라 기사유형, 주제, 시대, 관련 인물, 사건, 지리/장소 등 다양한 접근점을 제공할 수 있도록 데이터 항목을 세분화하여 설계한다.
- (3) 언어적으로 난해한 근대 초기의 일본어 신문이므로 일본어 해독이 어려운 이용자들을 위해 주요 요소(제목, 관련 인물 등)에 대해서는 일본어 원어와 한글 표기(일본어의 한국어 음, 일본한자에 대한 한글음)를 병기한다.
- (4) 데이터의 일관성을 유지할 필요가 있는 요소(기사유형, 주제, 시대, 언어, 지역 등)에 대해서는 통제어휘를 적용한다.

- (5) 데이터 입력과정에서의 오류를 방지하고 입력의 편의를 위해 각 요소에 대한 정의, 어휘, 적용방식, 사례 등을 포함하는 ‘입력 지침’을 작성하여 활용한다.
- (6) 향후 유사 신문기사 DB(한국 고신문 및 일제 강점기 신문 등)와의 상호운용성 확보를 위해 가급적 메타데이터 표준과 인코딩 스킴을 준용한다.

이러한 원칙을 토대로 하여 신문기사의 제 요소를 발행 정보, 내용 정보, 관리 정보로 구분한 후 메타데이터 요소(안)을 일차적으로 도출하였다. 이어, 일정 기간 동안에 실제 적용해보는 테스트 과정을 거쳐 요소의 추가 및 보완 등 수정작업이 이루어졌다.⁶⁾ 최종적으로 도출한 『조선시보』 신문기사에 대한 메타데이터 요소는 다음 <표 2>와 같다. 각 요소에 대해 필수와 반복 등의 적용사항을 설정하였고, 인코딩 스킴을 적용하는 요소에 대해서는 해당 사항을 표시하였다.

<표 2> 『조선시보』 신문기사의 메타데이터 요소(최종)

구분	속성(한글)	설명	필수*	반복**	인코딩
발행 정보	발행일자_원문	신문에 기재된 일자 (예: 昭和10-03-03)	M	NR	
	발행일자_양력	신문에 기재된 일자의 서력 일자 (예: 1935-03-03)	M	NR	W3C DTF
	호(號)	신문이 발행된 호수 (예: 9685)	M	NR	
	면(面)	해당 기사가 수록된 면수 (예: 1)	M	NR	
	단(段)	해당 기사가 수록된 단수 (예: 3)	O	NR	
	판	해당 기사가 수록된 판의 이름 (예: 경북판)	O	NR	
내용 정보	제목_원문	해당 기사의 제목	M	NR	
	제목_한글번역	한글로 번역한 제목	M	NR	
	부제목_원문	해당 기사의 부제목	O	R	
	부제목_한글번역	한글로 번역한 부제목	O	R	
	기사유형	기사의 유형 구분	M	NR	자체개발
	삽화	해당 기사와 관련된 삽화의 존재 유무	O	NR	
	주제	기사의 내용과 관련 있는 주제	O	R	자체개발
	시대	기사의 내용과 관련된 시대 구분	O	R	자체개발
	지역	기사의 내용과 관련된 국가 구분	O	R	ISO3166
	지역_키워드	기사의 내용과 관련된 국가 내 세부 지역명	O	R	
	언어	기사를 작성한 언어	O	R	RFC3066
	사건	기사의 배경이 된 주요 사건이나 행사 등	O	NR	
	키워드	기타 키워드(자유 기술)	O	NR	
	관련 인물_원어	기사의 내용과 관련된 인물의 이름(원어) (예: 寺内正毅)	O	R	
관련 인물_일본음	기사의 내용과 관련된 인물의 이름(일본음을 한글로 표기) (예: 데라우치 마사타케)	O	R		

6) 요소 추가 및 보완의 필요성은 데이터 입력 테스트를 통해 제기되었는데, 이 과정에서 기존에 정의되지 않았던 ‘판’, ‘삽화’ 등의 요소가 추가되고, ‘시대’, ‘지역’, ‘기사유형’, ‘집필자’ 등에 적용하는 데이터 값에 대한 보완이 있었다. 자세한 내용은 “2. 검증 및 crosswalk 작성”에 서술하였다. <표 2>는 이러한 검증 결과를 반영한 최종 설계안에 해당한다.

10 한국도서관·정보학회지(제50권 제4호)

	관련 인물_한글음	기사의 내용과 관련된 인물의 이름(한자에 대한 한글음으로 표기) (예: 사내정의)	O	R	
	요약	기사 내용을 간추린 요약문	O	NR	
	집필자 구분	기사를 작성한 집필자의 신분(기자, 독자, 작가)	O	NR	
관리 정보	집필자명	기사를 작성한 집필자의 이름			
	소장처	신문 원본을 소장하는 기관	M	R	
	입력주기	판독 불가능, 표기오류 등 설명 어구	O	NR	
	원문 링크	해당 기사가 수록된 지면에 대한 원문 링크 정보	M	NR	
	ID	레코드 ID	M	NR	
	작성자	데이터 입력자의 이름	M	NR	
	입력일자	메타데이터를 최초로 입력한 일자	M	NR	W3C DTF
수정일자	메타데이터를 최종 수정한 일자	O	NR	W3C DTF	

* 필수항목의 M(Mandatory)은 필수 요소, O(Optional)는 선택 요소를 의미함

** 반복항목의 R(Repeatable)은 반복 가능, NR(Non-Repeatable)은 반복 불가를 의미함

<표 2>에서 보듯이, 기사유형, 주제, 시대 등 3개의 요소에 대해서는 요소에 적용할 통제어휘를 연구팀에서 자체 개발하였다. 자체 개발을 결정한 주된 이유는 기존 사례가 일체강점기 신문기사에 적합하지 않기 때문이거나 관련 사례가 없어 통제어휘를 직접 개발하여 적용하는 것이 데이터의 일관성 유지를 위해 보다 효율적일 것으로 판단되었기 때문이다. 가령, 한국언론진흥재단 「빅카인즈」의 고신문 검색에서 ‘신문기사의 유형’ 항목을 사용하고 있지만⁷⁾ 이를 본 연구의 대상인 일체강점기 신문에 적용하기에는 지나치게 상세하거나 유형 구분이 균등하지 않아 『조선시보』의 기사 내용에 맞도록 수정 전개를 필요가 있었다. 뿐만 아니라 국립중앙도서관의 「대한민국 신문 아카이브」에서도 ‘기사유형’ 요소를 적용하고 있지만 실제로 존재하는 데이터 값이 ‘기사’ 이외에는 없었으며, 국사편찬위원회 「한국사데이터베이스」의 경우 ‘사설 > 정치’와 같이 주제분류에 기사유형을 혼재하여 적용하거나 일부 신문에 한해서만 기사유형 요소를 적용하고 있었다. 이러한 이유로, 언론학 분야의 자료와 『조선시보』 신문기사에 대한 내용분석 결과를 참고하여 ‘기사유형’ 요소에 대한 통제어휘는 연구팀에서 재정의하여 사용하였다.

이에 더해, 신문기사의 ‘주제’에 해당하는 통제어휘의 개발도 필요하다고 판단되었다. 앞서 언급하였듯이, 국립중앙도서관의 「대한민국 신문 아카이브」에서는 주제명표목표를 이용한 ‘주제어’를, 그리고 국사편찬위원회의 「한국사데이터베이스」에서는 자체 개발한 주제분류표의 카테고리를 적용하고 있다. 그러나 국립중앙도서관의 주제명표목표는 주지하다시피 신문 기사에 적용하는데 한계가 있고, 국사편찬위원회에서 근현대신문에 적용하는 주제분류는 지

7) <빅카인즈>의 고신문 검색에서 제공하는 기사유형 카테고리는 다음과 같이 설정되어 있다: ① (연재)소설, ② 강좌, ③ 광고, ④ 기고, ⑤ 기사, ⑥ 날씨, ⑦ 독자란, ⑧ 만평, ⑨ 사설, ⑩ 사진, ⑪ 사진그림, ⑫ 소설, ⑬ 시, ⑭ 알림, ⑮ 잡보, ⑯ 편성표, ⑰ 해외단신, ⑱ 기타 (한국언론진흥재단, “빅카인즈 고신문 검색”, <https://www.bigkinds.or.kr/v2/news/oldNews.do> [인용 2019. 10. 20]).

나치게 상세하고 복잡하여 검색에 전혀 활용되지 못하고 있는 상황이었다. 이에 연구팀에서는 「한국사데이터베이스」에 적용된 4계층의 ‘주제분류’ 카테고리를 참고하되 데이터 검색과 식별을 위해 2계층 수준으로 간략화하고, 이를 주제어 어휘의 형태로 변환하여 적용하였다. 이밖에, 신문기사와 직접적으로 관련된 ‘시대’ 데이터에 대해서도 한국과 일본에 공통으로 적용할 수 있는 통제어휘를 개발하여 적용하였다. 『조선시보』 신문기사에 적용할 기사유형, 주제, 시대, 지역의 통제어휘 결과는 다음 <표 3>~<표 5>와 같다.

<표 3> ‘기사유형’ 구분을 위한 통제어휘

구분	통제어휘	설명
기사 유형	사실기사	5W1H에 따라 발생한 사건이나 인지된 현상을 사실적으로 전달하는 기사
	해설기사	발생한 사건이나 인지된 현상에 대한 정보 전달이 아니라 설명이나 분석, 배경이 되는 정보, 예상되는 영향 등을 전할 목적으로 작성된 기사
	기획기사	특정 사건이나 이슈에 대한 심층보도 형태의 기사 일반적으로 기획/시리즈/연재/특집/집중점검 등으로 구분되어 있는 경우가 많음
	인터뷰기사	특정인을 인터뷰한 내용을 정리하여 작성한 기사
	사설	<사설>이라는 표제 하에 작성된 기사
	칼럼	내부 필진 또는 외부 필진의 글이나 기고문
	독자투고	독자가 작성한 글을 채택하여 수록한 기사
	문예창작*	독자나 기고자가 창작한 시, 수필, 소설, 희곡 등과 같은 문학작품
	기타	위의 기사에 해당되지 않는 기사

* ‘문예창작’은 신문기사 유형에 해당하지 않으나, 1900년 전후 일본 신문에서 ‘신문 소설’이 융성했던 시대적 배경을 고려하여 연구팀에서 추가한 것임. 당시 일본 신문들은 고정 독자의 확보를 위해 독자의 흥미를 끌 수 있는 신문 소설을 연재했는데, 『조선시보』에서도 이러한 문학류가 중요한 비중을 차지하고 있음.

<표 4> ‘주제’ 구분을 위한 통제어휘

구분	통제어휘	
	1계층	2계층
주제 구분	경제	경제일반, 경제단체, 농업, 임축수산업, 광공업, 토목건설업, 서비스업, 소비, 유통/운송, 화폐/금융, 무역(수출입), 상업/성인/시장
	교육	교육일반, 교육과정, 교육단체, 교육행정, 학교, 학교경영/교원/학생
	문화	문화일반, 문학, 언어, 철학/사상, 종교, 예술, 스포츠, 출판, 언론/방송, 민속/풍속, 문화재, 여가/오락
	사회	사회일반, 가족/가정생활, 노동, 범죄/비리, 보건/의료, 분규/분쟁, 의식주, 인구, 사회산업/사회복지/생활개선, 여성, 재해/사건/사고, 집단/계층
	산업	산업일반, 공공시설, 사회산업시설, 전기/가스/수도, 교통/통신
	전쟁	정세, 교전, 군사동원, 무기
	정치	정치일반, 정치단체, 국방/군사정책, 경찰, 법, 행정, 선거, 외교/국제관계
	기타	기타일반, 날씨, 인사/인물, 지리

<표 5> '시대' 구분을 위한 통제어휘

구분	통제어휘	<한국(조선) 기준>	<일본 기준>
시대	고대	고조선, 신라, 고구려, 백제, 가야	야마토(大和)시대, 아스카(飛鳥)시대, 나라(奈良)시대, 헤이안(平安)시대
	중세	고려	가마쿠라(鎌倉)시대, 무로마치(室町)시대, 남북조(南北朝)시대, 전국(戰國)시대, 아즈치 모모야마(安土桃山)시대
	근대	조선 ~ 해방 전(1945)	에도(江戸)시대, 메이지(明治)시대, 다이쇼(大正)시대, 쇼와(昭和)시대

2. 검증 및 crosswalk 작성

본 연구에서 도출한 메타데이터 요소 및 요소 값에 대한 타당성 검증은 다음 세 가지 방법을 통해 이루어졌다: (1) 실제 데이터의 적용을 통한 검증, (2) DB 전문기관의 자문(컨설팅), (3) 일제강점기 신문기사 검색서비스를 제공하는 기존 시스템 및 메타데이터 표준과의 crosswalk 작성.

먼저, 『조선시보』 신문 기사를 수집하여 검토하면서 앞서 도출한 데이터 요소를 직접 적용해보는 과정을 거쳤다. 근대 일본어로 작성된 신문기사의 정확한 판독을 위해 이 과정은 일본어 해독 능력과 일본문화 및 역사적 소양을 갖춘 전공자(대학원생)에 의해 진행되었다. 검토에 앞서, 본 연구에서 도출한 데이터 항목과 적용방법을 비롯하여 전체적인 메타데이터 구축 과정과 향후 활용방안에 대한 사전 교육을 2차에 걸쳐 실시하였다. 이후 신문기사 건별로 나누어 세밀하게 살펴 보면서 입력될 데이터 요소와 요소 값에 대한 검토는 물론이고, 일본어 번역 및 표기 문제 등에 대해서도 종합적으로 논의하였다. 그 결과, 일부 요소의 추가나 요소 값의 보완 등 미세한 수정이 필요한 것으로 나타났지만,⁸⁾ 메타데이터의 설계 자체는 대체로 적절하다고 평가되었다. 다만, 표기의 일관성이나 일본어 번역 오류 등 실제 데이터를 구축하는 과정에서 다양한 문제가 발생할 수 있을 것으로 예측되어 향후 철저한 품질 관리가 무엇보다 중요한 것으로 나타났다.

이어 수정(안)을 마련한 다음, 유사 분야의 DB 관리와 웹 서비스 구축 경험이 있는 전문가로부터 자문을 받았다. 자문은 메타데이터 속성 정의서 및 통제어휘에 대한 코드 테이블, 속성과 코드의 관계도 등 DB 구조 문서를 활용하여 진행하였다. 결과적으로, 본 연구의 메타데이터 설계안은 『조선시보』 신문기사 DB로서 뿐만 아니라 이전에 구축된 신문DB와의 상호운용성을 고려하여도 타당한 것으로 결론 내려졌다. 그러나 데이터 품질관리를 위해 검수 단

8) 검증 과정에서 『조선시보』에서 빈번히 나타나는 '지방관'과 '삽화'에 관한 정보를 기술할 별도의 요소와 1900년 전후 일본사회에서 열광적 인기를 얻었던 '신문 소설'을 기사유형에 추가할 필요성이 제기되었다. 이에 더해, 데이터 관리를 위해 메타데이터 생성일자과 수정일자의 추가와 입력 작업의 효율을 위해 '시대'와 '지역'에 적용할 요소 값의 간소화 방안 등이 논의되었다.

계별 검수자명이나 검수일자 등을 보완할 것을 제안 받았다.⁹⁾

마지막으로, 본 연구에서 제안한 메타데이터 항목을 현재 일제강점기 신문기사에 대한 검색서비스를 제공하고 있는 국립중앙도서관의 「대한민국 신문 아카이브」와 국사편찬위원회의 「한국사데이터베이스」, 독립기념관의 「한국독립운동정보시스템」 메타데이터와의 crosswalk를 작성하였다. 여기에, 향후 확장성을 고려하여 메타데이터 표준 중 전 세계적으로 가장 널리 사용되고 있는 Dublin Core 메타데이터 스키마를 추가하여 비교하였다. crosswalk는 다양한 메타데이터 요소를 의미적으로 상호 연결시킴으로써 상이한 표준들 간의 공통점과 차이점을 효과적으로 드러낼 뿐만 아니라 메타데이터 스키마들 간의 상호운용성을 확보하기 위한 목적에서 사용하는 대표적인 접근방법이라 할 수 있다(남태우·이승민 2018, 289). 이에 본 연구에서 설계한 메타데이터 요소 및 구조에 대한 타당성을 확보하기 위한 방법으로서는 물론이고, 향후 통합검색시스템의 구축이나 서로 다른 메타데이터 스키마의 데이터를 통합하고자 할 때 의미적 변환을 위한 기초자료로 활용할 수 있도록 crosswalk를 작성하였다.¹⁰⁾ 그 결과는 다음 <표 6>과 같다.

<표 6> 유사 시스템 및 Dublin Core 메타데이터와의 crosswalk 결과

본 연구 『조선시보』	대한민국 신문 아카이브	한국사 데이터베이스	한국독립운동 정보시스템	Dublin Core
발행일자_원문	(주기사항)			
발행일자_양력	발행일	발행년월일	발행일	date_issued
호(號)	(기사위치)		생산정보	
면(面)	기사위치	면수		source
단(段)		단수		source
관				
제목_원문	기사제목	기사제목	기사제목	title
제목_한글번역				
부제목_원문	부제목			title_alternative
부제목_한글번역				
기사유형	기사유형	기사형식		type
삼화				
주제	기사주제	주제분류		subject
시대				coverage_temporal
지역		발신지		coverage_spatial
지역_키워드				coverage_spatial

9) 이 제안에 대해서는 연구팀 내부 회의를 통해 잠정적으로 반영하지 않는 것으로 결정하였다. 그 이유는 데이터의 실제 구축과정과 깊은 관련이 있는데, 현재 『조선시보』 신문기사 메타데이터 구축에서는 검수자가 육안으로 검수하여 수정사항을 입력자에게 전달하고 입력자가 재확인하여 최종 반영하는 절차를 밟고 있기 때문이다.

10) crosswalk 작성에서 Dublin Core 메타데이터 요소는 한정어(qualifier)의 적용까지 고려하였으며, Getty연구소에서 제공하는 “Metadata Standards Crosswalk”를 참조하여 작성하였다. Getty Research Institute. Introduction to Metadata: Crosswalk. https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html [cited 2019. 10. 24]

14 한국도서관·정보학회지(제50권 제4호)

언어	언어*			language
사건				subject
키워드				subject
관련 인물_원어		색인어_이름		subject
관련 인물_일본음				subject
관련 인물_한글음				subject
요약			요약정보	description_abstract
집필자 구분				
집필자 명				contributor
소장처	소장기관*			source
입력주기				
원문 링크				
레코드 ID				identifier
작성자				
입력일자	생성일자*			date_created
수정일자	수정일자*			date_modified
	신문명	신문명	신문명	
	발행처			publisher
	창간일			
	발행지*			
	발행간기*			
		발행구분		
	자료유형*			
	이용대상자*			audience
	파일수와 크기*			format_extent

*표시는 검색화면이 아닌 XML 구문을 통해 확인가능한 요소임

crosswalk를 작성한 결과, 공통 요소는 ‘발행일자’, ‘기사위치(면, 단)’, ‘기사제목’, ‘기사유형’, ‘주제’이며, 적어도 3개의 메타데이터 스키마에 적용된 요소는 ‘지역’, ‘언어’, ‘요약’, ‘소장처’, ‘입력일자’, ‘수정일자’ 등인 것으로 나타났다. 이러한 결과를 토대로 하여, 본 연구에서 설계한 일제강점기 일본어 신문인 『조선시보』의 메타데이터 스키마가 가지는 의미를 도출한 결과는 다음과 같다.

첫째, 『조선시보』 메타데이터는 유사 시스템에서 적용하고 있는 메타데이터 요소를 모두 수용하고 있다. 다만, 본 연구에서 설계한 메타데이터는 『조선시보』 신문기사에만 한정 적용하는 것이어서 신문 자체에 대한 데이터(예: 신문명, 신문의 창간일자, 발행처, 발행 간기(일간, 주간), 발행 구분(조간, 석간) 등)를 포함하고 있지 않다는 차이를 보인다. 그러나 본 연구에서 설계한 메타데이터에 『조선시보』 신문 자체에 대한 collection-level 메타데이터 스키마를 추가로 적용한다면, 이들 요소의 수용도 충분히 가능하다고 할 수 있다.¹¹⁾

11) 신문 자체에 관한 정보(종정보)는 신문기사 메타데이터에 반복적으로 기술하기보다는, 이처럼 collection-level 메타데이터를 생성한 후 연결하여 활용하는 방법이 데이터의 관리나 검색 측면에서 더욱 유용하다고 볼 수 있다. 이러한 collection-level 메타데이터에는 신문명, 창간일자, 발행처, 간기, 판본을 비롯하여 제호 변경에 관한 이력정보, 원문의 제공년도 등을 기술할 수 있다.

둘째, 보다 효율적인 접근을 지원하기 위해 『조선시보』 메타데이터는 다양한 요소(특히, 신문기사의 주제에 해당하는 요소를 ‘시대’, ‘지역’, ‘사건’, ‘관련 인물’, ‘키워드’ 등)를 세분 전개하고 있다. 이들 요소는 향후 검색서비스를 구현할 때 다각적인 접근점으로 활용될 수 있어 이용자의 입장에서 보다 정확하고 빠른 신문기사 검색이 가능할 것으로 기대된다.

셋째, 일제강점기 일본어 신문임을 고려하여 주요 요소(제목, 부제목, 관련 인물, 발행일자)에 대해서는 한국어와 일본어(원어)의 적용을 구분하고 있다. 이는 표기 방식이나 사용언어를 입력자가 임의로 적용하지 않도록 함으로써 데이터의 일관성과 정확성을 유지하는데 도움이 된다.

넷째, 『조선시보』 메타데이터에는 유사 시스템에 존재하지 않는 판, 삽화, 집필자(기사 작성자)와 같은 요소가 추가 적용되어 있다. 이는 『조선시보』 신문기사에 대한 실제적 검토를 토대로 한 것으로써 반복적으로 중요하게 다루어진 데이터를 요소 설계에 반영한 것이라 할 수 있다.

다섯째, 『조선시보』 메타데이터 설계에서 확장 전개한 요소(한국어와 일본어 표기 구분이나 기사위치를 ‘면’과 ‘단’으로 구분한 항목, 그리고 ‘주제’, ‘시대’, ‘지역’, ‘사건’ 등으로 세분한 주제적 요소 등)는 한정어(qualifier)의 사용이나 플랫폼(metadata registry)을 이용하여 충분히 통합 관리할 수도 있다. 따라서 다른 스키마로의 통합이나 재사용에서도 전혀 문제가 되지 않는다고 볼 수 있다.

3. 메타데이터 입력기의 개발

메타데이터 구축 대상인 『조선시보』는 약 48년 동안 발행된 일간지로서 신문기사의 분량만 무려 35만 건에 이를 것으로 추정된다. 따라서 excel 프로그램 등 일반적인 스프레드시트를 이용한 데이터의 입력과 관리는 불가능할 것으로 예측되어, 연구팀에서는 데이터 구축의 편의성과 효율성, 안정성을 고려하여 데이터 입력기를 자체 개발하여 사용하는 것으로 결정하였다. 이에 따라 MySQL과 PHP를 기반으로 하는 open CMS인 WordPress 4.9 버전을 사용하여 『조선시보』 메타데이터 입력기를 직접 개발하였다. 입력기의 개발에 앞서, 연구팀에서는 논의를 거쳐 입력기에 구현할 세부 기능을 먼저 도출하였다. 주요 내용은 다음 <표 7>과 같다.

<표 7> 입력기에 구현할 주요 기능

기능	설 명
권한 설정	입력자와 관리자의 권한을 구분하여 관리함. 입력자는 데이터의 신규 입력을, 관리자는 입력된 데이터에 대한 검수를 담당함. 데이터 입력 및 검색 등에서는 차이가 없으나 입력자는 자신이 입력한 데이터에 대한 수정만 가능하며, 관리자는 전체 데이터의 수정이 가능하도록 권한을 구분하여 설계함
코드 관리	앞서 개발한 통제어휘를 코드 값으로 설계함. 코드는 필요할 경우 변경이나 추가가 가능하도록 설계

16 한국도서관·정보학회지(제50권 제4호)

	하되, 변경된 코드는 기 적용된 데이터에 일괄 적용되도록 구현함. 코드를 적용하는 데이터 항목에는 설정된 코드 값을 제시함
필수 항목 설정	앞서 개발한 메타데이터 요소를 입력 항목으로 구현하되, 필수/선택 요소를 반영하여 필수 입력 항목을 지정함. 필수 항목 미입력 시 데이터의 저장은 불가능함.
검색 기능	입력된 데이터를 다양한 항목(발행일자, 입력일자, 입력자명, 기사제목, 키워드 등)으로 검색하거나 정렬하는 기능을 구현함
실시간 데이터 확인	입력된 데이터는 데이터의 일관성과 통일성을 위해 입력자들이 자유롭게 상호 참조하고 교차 검증할 수 있도록 실시간으로 제공함
알림 기능	입력과정에서 발생한 문제나 질문을 관리자에게 즉각 알리고, 관리자가 이를 수시로 확인하여 검토할 수 있도록 입력기 내에 쌍방향 커뮤니케이션 기능을 구현함
메타데이터와 원문 이미지의 연결	입력자나 관리자가 입력된 데이터를 수시로 확인하고 검수할 수 있도록 입력된 데이터와 함께 원문 이미지를 연계하여 제시함
통계 생성	입력자별, 발행일자별, 적용된 코드별로 해당되는 데이터의 수를 자동 산출함(예: 기사유형별 데이터 통계)

이러한 내용을 반영하여 개발한 입력기는 <그림 3>과 같다. 입력기의 화면구성은 입력 편의를 고려하여 기본정보, 인물정보, 원문이미지로 구분하였으며, 통제어휘를 적용하는 요소에 대해서는 해당 요소에 설정된 데이터 값을 입력화면에 직접 제시하였다. 또한 자체 개발한 통제어휘와 표준 인코딩 스키를 적용하는 주제, 기사유형, 시대, 지역, 언어 요소에 대해서는 ‘코드관리’ 기능을 통해 요소 값의 수정 및 추가가 가능하도록 구현하고(화면 왼편), 이를 토대로 실시간 자동 통계 생성이 가능하도록 구현하였다. 신문기사에 등장한 인물에 대해서는 ‘인물정보’ 항목에 원어, 일본음의 한글 표기, 일본한자의 한글음 표기로 구분하여 입력하도록 설계하였으며, 효율적인 검수를 위해 입력 화면에서 원문 이미지를 바로 확인할 수 있도록 구현하였다.

① 기본정보(일부)

② 인물정보

기사(News)에 등장한 인물 정보

	원문	일본음	한국음	
1	香推源太郎	가시이 겐타로	항추 원태랑	+ -
2	吉田秀治郎	요시다 히데지로	길전 수치랑	+ -

Add New

③ 원문 이미지



<그림 3> 『조선시보』 메타데이터 입력기

한편, 데이터를 생성하는 단계에서부터 데이터의 정확성과 일관성을 유지하기 위해서는 입력기의 개발과 더불어 각 요소별 상세한 입력 지침도 작성하여 활용할 필요가 있다. 자료 훼손으로 관독이 불가능하거나 오류를 내포하고 있는 경우, 문자 입력이 불가능한 경우 등에 대한 처리는 물론이고 근대 일본어로 표기된 인명이나 지명에 대한 일본어 및 한국어의 표기 방식, 한자 이체자의 표기방식 등을 사전에 정하여 일관성 있게 적용할 필요가 있기 때문이다. 이에 본 연구에서 설계한 모든 메타데이터 요소에 대해 정의와 적용지침, 사례, 그리고 통제어휘를 적용하는 경우 각 어휘에 대한 설명 등을 명시한 <『조선시보』 메타데이터 입력 지침>을 작성하였다. 이러한 입력지침은 데이터의 품질관리를 위한 장치로서는 물론이고, 향후 어플리케이션 프로파일(application profile)을 구축하는데 기초자료로도 활용될 수 있을 것이다.

IV. 신문기사 메타데이터 개발을 위한 제언

본 연구는 실제 데이터의 구축을 전제로 하여 『조선시보』 신문기사의 메타데이터를 개발하는데 목적을 두고 있다. 연구를 수행하기에 앞서, 유사 사례를 참조하기 위해 폭넓게 조사하였으나 적용된 메타데이터 요소의 확인이 불가능하거나 제한적으로만 접근이 가능할 뿐이었다. 관련 연구 또한 1886년 근대 신문의 창간부터 1960년까지 발간된 고신문 검색을 위한 ‘필수’ 메타데이터 요소를 제안한 연구(김희순 2014)와 인터넷 신문을 대상으로 기사 작성단계에 따른 메타데이터 구축방안을 제안한 연구(정성숙 2012) 정도에 불과하여 일제강점기 일본어 신문인 『조선시보』에 적용하는데 한계가 있었다. 뿐만 아니라 국외의 사례나 연구도 주로 국가도서관 등이 주도하여 방대한 양의 신문DB를 통합 구축하는 사례에 해당하는 것이어서 본 연구에서 참조하기에 범위가 너무 광범위하였다. 이에 연구팀은 이 연구의 결과를 향후 참조모델로 활용하고, 나아가 국내 신문 메타데이터의 표준화에 기여할 수 있도록 메타

데이터 개발의 전 과정을 가능한 자세하게 제시하고자 하였다. 지금부터는 『조선시보』 메타데이터를 개발하는 과정에서 연구팀이 얻은 경험을 토대로 하여 향후 신문기사 메타데이터의 개발과 활용 방안을 제안해 보고자 한다.

앞서 살펴본 것처럼 국내에서 발행된 일제강점기 신문DB는 국립중앙도서관과 국사편찬위원회, 독립기념관, 대학 연구소, 상용DB 업체, 언론사 등 다양한 주체들이 개별적으로 구축해 온 결과물이라 할 수 있다. 구축 주체가 이렇듯 분산되어 있고 이를 조정하거나 관리하는 책임 있는 주체가 부재하다보니 현재 구축된 신문 메타데이터는 요소와 어휘, 구조는 물론이고 내용적 품질 면에서 구축 주체별로 혹은 신문 종별로 적지 않은 차이를 보이고 있다.¹²⁾ 이와 같은 데이터의 편차 혹은 이질성은 향후 데이터를 활용하거나 통합시스템을 구축하는데 있어 커다란 장애가 될 수 있다. 기술적인 방법으로 이질성을 최소화하거나 상호운용성을 확보하기 위한 다양한 방안이 제안되고 있지만, 아직까지 포괄적이고 완전한 형태의 이상적인 방법이 개발되지는 않았기 때문이다. 이러한 이유로 연구팀에서는 신문기사 메타데이터의 효율적인 구축과 활용을 위해 다음 두 가지 방안을 제안하고자 한다.

첫 번째는 신문기사를 위한 ‘범용’ 메타데이터 표준을 개발하여 활용하는 방안이다. 이를 위해서는 우리보다 7~8년 늦게 신문DB 구축에 착수한 미국의 경우를 참조해 볼 수 있다. 미국의 경우 1690년 이후 미국에서 발행된 역사적 신문의 DB구축을 위해 국가디지털신문사업(National Digital Newspaper Program)을 추진하고 있다. 국가인문학기금(National Endowment for the Humanities)의 후원으로 미의회도서관에서 2005년부터 추진하고 있는 이 사업은 「Chronicling America」(<http://chroniclingamerica.loc.gov/>)를 통해 서비스되고 있다. 사업에 참가하여 신문DB 구축을 전담할 주체는 州별로 프로젝트 형식의 공모를 통해 선정하고 있는데, 선정된 기관(주로 주립도서관, 대학, 역사학회 등)에서는 미의회도서관에서 개발하여 제공하는 기술명세서(technical specification)를 기반으로 사업을 추진하고 있다. 이 기술명세서에는 신문DB에 적용될 포괄적 수준의 메타데이터 표준은 물론이고 신문의 디지털화 방식, OCR(Optical Character Recognition)을 통한 텍스트 추출 방법 등이 자세하게 설명되어 있다. 따라서 사업 참여 기관들은 효율적이고도 일관된 방식으로 신문DB를 구축할 수 있고, 미의회도서관에서는 각기 다른 주체에 의해 구축된 DB를 데이터의 손실 없이 손쉽게 통합하여 활용하고 있다(Library of Congress 2019).¹³⁾

12) 이미 구축 완료된 신문DB 중에는 데이터의 품질 면에서 심각한 문제를 가진 사례도 있다. 신문기사 메타데이터로 구축되는 데이터가 기사제목, 발행일, 신문명 외에 어떠한 데이터도 포함되어 있지 않은 사례가 있기 때문이다.

13) 미국 내에서 발행된 모든 신문을 대상으로 하는 DB구축 사업인 만큼, 메타데이터 표준은 (1) title information (신문종별), (2) container(마이크로필름 릴과 같은 종이신문에 대한 대체 매체), (3) issue(발행호수), (4) page (발행호수의 특정 면), (5) article(신문기사) 등 5개 개체로 구분되어 있으며, 각 개체별로 요소(속성)가 상세하게 마련되어 있다. 신문DB에 적용되는 메타데이터 요소 사전은 “The National Digital Newspaper Program: Technical Guidelines for Applicants” (Library of Congress 2019)을 참조하시오.

이러한 미국의 사례를 참조한다면, 국내에서도 만시지탄의 감이 있긴 하지만 지금이라도 신문기사에 적용할 범용 메타데이터 표준의 개발을 서두를 필요가 있다. 특히 일제강점기 신문에 대한 발굴과 이에 따른 DB구축이 앞으로도 지속될 것이며, 나아가 각기 다른 주체에 의해 구축된 다양한 신문DB에 대해 국가 차원에서 통합서비스 환경을 구축해 나가고 있다는 점을 고려한다면, 현재와 같이 상이한 형식으로 구축된 메타데이터의 통합이나 후속적인 질 관리에 에너지를 쏟기보다는 사전에 범용 메타데이터 표준을 개발하여 이를 적용하도록 유도하는 것이 훨씬 효과적인 방안이라 할 수 있다. 이러한 범용 메타데이터 표준은 현재 일제강점기 신문DB 구축을 주도하고 있는 세 기관(국립중앙도서관, 국사편찬위원회, 독립기념관)의 협의 하에 또는 국내에서 다양한 자료유형에 대한 메타데이터 표준을 선도하고 있는 국립중앙도서관이 주도하여 충분히 개발할 수 있을 것으로 보인다. 신문DB 구축을 위한 범용 메타데이터 표준이 개발되면, DB구축을 담당하는 기관들에서는 이를 준용하되 신문의 특수한 상황이나 구축 주체의 요구를 반영하여 확장하여 사용할 수 있을 것이다.¹⁴⁾

두 번째 방안은 분산 구축된 메타데이터의 이질성을 최소화하기 위한 자구책에 해당한다. top-down 방식으로 범용 메타데이터 표준을 개발하여 적용하는 것이 당장 어렵거나 불가능한 상황이라면, 향후 구축 예정인 신문DB는 그에 앞서 개발·적용된 유사 메타데이터 스키마를 폭넓게 참조하고 수용할 필요가 있다. 본 연구팀에서는 『조선시보』 메타데이터 요소와 통제어휘의 적용을 위해 이미 구축된 일제강점기 신문DB를 참조하였다. 이 과정에서 동일하거나 유사한 의미로 사용된 요소나 어휘는 가급적 『조선시보』 메타데이터에 활용하고자 하였으며, 『조선시보』 메타데이터에서 확장 또는 변용 적용한 요소에 대해서는 crosswalk를 작성하여 다른 시스템에 적용된 메타데이터와 어떠한 차이가 있는지를 명료하게 보여주하고자 하였다. 이는 다른 주체에 의해 구축된 메타데이터와의 의미적 연결을 통해 상호운용성을 확보하고, 향후 데이터의 통합 시 본 연구팀에서 개발한 메타데이터를 보다 안정적으로 재사용할 수 있도록 하기 위함이었다.

그러나 현실적으로 각기 다른 신문DB에 적용된 메타데이터 요소의 의미나 구문에 접근하기란 쉽지 않았다. 메타데이터 요소가 실제 어떠한 의미로 사용되었으며, 인코딩 방식이나 필수/선택 등이 어떻게 적용되었는지에 대한 신뢰성 있는 정보를 제공해 주는 ‘어플리케이션 프로파일’의 확보가 불가능하였기 때문이다. 어플리케이션 프로파일은 각기 다른 스키마에 적용된 데이터 요소의 의미와 기능을 상세히 정의하고, 이들 요소를 특정 상황에 맞게 활용할 수 있도록 공개하는 역할을 한다(남태우·이승민 2018, 294). 따라서 유사한 정보자원에 대한 메타데이터 설계에 참고하고, 동시에 설계 단계에서부터 다른 DB와의 상호운용성을 제고할

14) 국립중앙도서관에서는 OAK(Open Access Korea)를 이미 이러한 방식으로 운용하고 있다. OAK에 참여하는 기관 리포지터리에서는 OAK에서 제공하는 메타데이터를 활용하되, 참여기관에서 원하는 메타데이터 요소를 자유롭게 확장하여 사용하도록 허용하고 있다. 뿐만 아니라 참여기관에서 데이터의 생성 및 검색에 활용할 수 있도록 오픈소스 프로그램을 무료로 제공하고 있는데, 참여기관에서는 이 프로그램을 활용하되 기관 특성에 맞게 커스터마이징하여 사용할 수 있다.

수 있는 기반을 마련하기 위해서는 메타데이터 요소, 구문, 구조를 포함하는 어플리케이션 프로파일을 구축하여 상호 활용하는 것이 반드시 필요하다고 할 수 있다. 이렇게 될 때 분산 구축된 신문DB의 이질성은 그나마 제거될 수 있을 것이며, 결과적으로 효율적인 통합시스템의 구축도 가능하게 될 것이기 때문이다.

V. 결론

본 연구는 일제강점기 일본어 신문 중 자료의 방대한 양이나 사료로서의 가치에도 불구하고 아직까지 신문기사에 대한 DB구축이 이루어지지 않은 『조선시보』를 대상으로 하여, 메타데이터 구축을 위한 기반을 마련하는데 목적을 두었다. 이를 위해 일제강점기에 발간된 다수의 신문에 적용된 메타데이터와 『조선시보』의 형태적·내용적 특징에 대한 분석을 토대로 하여 『조선시보』 신문기사에 적용할 메타데이터 요소를 개발하였다. 본 연구에서 개발한 『조선시보』 메타데이터는 이미 구축된 일제강점기 신문DB에 적용된 메타데이터 요소를 모두 수용하면서도 다양한 요소(특히, 신문기사의 내용에 해당하는 요소)를 세분 전개하고 있다는 특징을 가진다. 또한 일제강점기 일본어 신문임을 고려하여 주요 요소(제목, 부제목, 관련 인물, 발행일자)에 대해 한국어와 일본어(원어)의 병기 방안을 메타데이터 요소에 반영하고, 판, 삽화, 집필자(기사 작성자)와 같은 요소를 확장 적용하고 있다는 점에서 다른 신문DB와 차별성을 보이고 있다.

『조선시보』 메타데이터의 개발 과정에서 연구팀이 특히 중점을 둔 것은 향후 검색서비스를 구현할 때 다각적인 접근점을 마련함으로써 정확하고 빠른 신문기사 검색이 가능하도록 하는데 있었다. 이를 위해 기사제목과 발행일자 등과 같은 기초적인 정보 외에 기사유형, 주제, 시대, 관련 인물, 사건, 지리/장소, 기타 키워드 등으로 데이터 항목을 세분화하여 설계하였다. 더불어, 데이터의 표현방식에 있어 입력자들 간의 편차를 줄이면서 향후 검색효율을 제고하기 위해 통제어휘나 인코딩 스킴을 적절히 사용하여 데이터를 일관되게 표현하고자 하였다. 이러한 기본적인 설계를 바탕으로 하되 데이터의 실제 구축에서 무엇보다 중요한 것은 근대 초기의 난해한 일본어로 쓰여진 『조선시보』 신문기사의 내용을 정확하게 판독하여 적용하는 것임을 깨닫고, 연구팀에서는 일본어 해독능력과 일본문화 및 근현대사에 대한 전문성을 갖춘 전공자들이 DB구축에 참여하는 것으로 계획을 수립하였다.

현재 진행되고 있는 『조선시보』 신문기사에 대한 DB 구축이 완료되면, 본 연구팀에서는 다음과 같은 후속연구를 진행할 계획이다. (1) 한·일 양국어와 한자음을 이용하여 다양한 항목으로 검색할 수 있도록 검색시스템을 구현함으로써 일본어를 해독하지 못하는 연구자들도 『조선시보』에 손쉽게 접근할 수 있도록 한다. (2) 『조선시보』를 제외한 3종의 일본어 신문에 집중된 기존의 식민지·제국사 연구 성과를 확장하고 새로운 지식을 창출하는데 기여한다. 이

를 위해 『조선시보』의 시대적·내용적 특징을 규명하는 것은 물론이고, 『조선시보』가 가지는 학술적 가치를 관련 연구자들에게 적극적으로 알릴 것이다. (3) 『조선시보』를 통해 일제강점기의 정치, 경제, 문화의 유입과 수탈의 흐름, 모던도시 부산의 양상을 다각도로 파악하고, 특히 경성에 비해 충분히 연구되어 있지 않은 부산권역에 대한 지역연구를 심화하는데 기여한다. 본 연구의 결과가 일제강점기 일본어 신문인 『조선시보』에 대한 학계의 관심을 유도하고, 나아가 제대로 활용되지 못하고 있는 다양한 정보자원을 발굴하여 그에 대한 DB 구축사업을 폭넓게 진행하는데 기초자료로 활용되기를 기대한다.

참고문헌

- 국립중앙도서관. 대한민국 신문 아카이브. <<http://nl.go.kr/newspaper/sub0101.do>> [인용 2019. 9. 11].
- 국사편찬위원회. 한국사데이터베이스. <<http://db.history.go.kr/>> [인용 2019. 9. 11].
- 김희순. 2014. 『우리나라 고신문의 서비스 방안 연구』. 석사학위논문, 중앙대학교 교육대학원 문헌정보교육전공.
- 미즈노 나오키. 2007. 식민지기 조선의 일본어 신문. 『역사문제연구』, 18: 253-266.
- 송주형. 2016. 온라인 신문 아카이브 연구: 국내외 구축 사례를 중심으로. 『기록학연구』, 48: 93-139.
- 송지형. 2010. 1945년 이전 신문 자료의 온라인 활용 방법. 『도서관보』, 132: 45-62.
- 남태우, 이승민. 2018. 『정보자원의 기술과 메타데이터』. 개정판. 서울: 한국도서관협회.
- 전성현 외. 2019. 식민지 ‘지역언론’에서 ‘제국언론’으로 - 신문체제와 지면을 통해 본 『조선시보』의 특징과 의미. 『항도부산』, 37: 365-396.
- 정성숙. 2012. 『오마이뉴스 아카이빙 연구: 기록 메타데이터 재설계를 중심으로』. 석사학위논문, 한남대학교 대학원.
- 채백. 2012. 『부산언론사 연구』. 부산: 산지니.
- 한국언론재단. 빅카인즈 고신문 검색. <<https://www.bigkinds.or.kr/v2/news/oldNews.do>> [인용 2019. 10. 20].
- 한국학술정보. 경성일보. <http://www.kstudy.com/kyungsung_new/> [인용 2019. 9. 11].
- 홍순권. 2012. 日帝強占期 新聞史 研究의 現狀과 向後의 課題 - 식민지 일본인 경영의 신문 연구의 진척을 위한 제언. 『석당논총』, 52: 1-26.
- Getty Research Institute. Metadata Standards Crosswalk. <https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html> [cited 2019. 10. 24].

Library of Congress. “About Chronicling America”.

<<https://chroniclingamerica.loc.gov/about/>> [cited 2019. 10. 16].

Library of Congress. 2019. The National Digital Newspaper Program(NDNP) Technical Guidelines for Applicants: 2020–22 Awards. <https://www.loc.gov/ndnp/guidelines/NDNP_202022TechNotes.pdf> [cited 2019. 10. 16].

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Hong, Soon Kwon. 2012. “The Situation of the Historical Research on the Newspapers under the Rule of Japanese Imperialism and its Henceforth Tasks.” *Journal of Seokdang Academy*, 52: 1–26.
- Jeon, Sung Hyun et. al. 2019. “Characteristics and Meaning of 『Chosenjiho』 through the Newspaper System and Contents Under the Rule of Japanese Imperialism.” *Hangdo Busan*, 37: 365–396.
- Jeong, Seong–Suk. 2012. *A Study on Archiving of Oh My News : Focused on the Metadata Record Redesign*. M.A. Thesis, Hannam University.
- Kim, HeeSoon. 2014. *A Study on the Service Method of Korea Old newspapers*. M.A. Thesis, Chung–Ang University.
- Mizuno, Naoki. 2007. “The Japanese Newspaper in Colonial Korea.” *Critical Studies on Modern Korean History*, 18: 253–266.
- Nam, Taewoo and Seungmin Lee. 2018. *Resource Description and Metadata*. revised edition. Seoul : Korean Library Association.
- Song, Zoo Hyung. 2016. “A Study on the Online Newspaper Archive : Focusing on Domestic and International Case Studies.” *The Korean Journal of Archival Studies*, 48: 39–139.