

문화유산 기록관리 분야 AI기술 적용 사례 -‘문화유산 찾아-ZOOM’을 중심으로-

Examples of AI Technology Applications in the Field of Cultural Heritage Record Management -Focusing on "Finding Cultural Heritage - ZOOM"-

백주현 (Ju hyun Baek)

E-mail: whiteju920@korea.kr

국립문화유산연구원 기록연구소



논문접수 2024.7.18
최초심사 2024.7.22
게재확정 2024.8.22

ORCID

Ju Hyun Baek
https://orcid.org/0009-0004-9582-4260

초 록

본고는 국립문화유산연구원 기록관에서 진행해온 ‘AI기반의 문화유산 연구기록물 학습데이터 및 검색시스템 구축’ 사례를 통해 최신 첨단기술과 기록관리 분야의 접목이 업무뿐 아니라 기록정보서비스에 새로운 가능성을 창출할 수 있을 것인지에 대한 적용방안 및 추진과정을 소개하고 있다. ‘문화유산 찾아-ZOOM’은 1973년부터 현재까지 문화유산 분야에서 발간한 간행물에 수록된 이미지를 학습데이터로 구축하여, 유사 이미지를 동시에 제시함으로써 연구 자료에 대한 사전 수요 예측이 가능하도록 선제적으로 제공하고 있는 시스템이다. 4차 산업혁명으로 인한 첨단기술과 기록관리 분야에 새로운 변화와 발전을 도모하고자 시도한 사례로, 기록관리, 문화유산 분야 연구자들 뿐만 아니라, 실무자와 일반대중에게도 유용한 정보로 활용되기를 바란다.

ABSTRACT

This study explores the integration of cutting-edge technology with records management, aiming to create new value not only in work processes but also in record information services. The research focuses on the case of constructing an “AI-based cultural heritage research record learning data and search system,” carried out by the National Research Institute of Cultural Heritage (NRICH) Archives, and analyzes user satisfaction results. “Discovering Cultural Heritage with ZOOM” is a system designed to proactively predict research data demand by constructing big data (learning data) from images (675,338 items) contained in 1,421 volumes of publications in the cultural heritage field, spanning from 1973 to the present, and simultaneously presenting 50 similar images. This initiative aims to foster change and development in the field of records management and cultural heritage in response to the Fourth Industrial Revolution’s advanced technologies. It is expected to provide valuable information to researchers, practitioners, and the general public alike.

Keywords: 국립문화유산연구원 기록관, 문화유산 찾아-ZOOM, 인공지능, 유사 이미지 검색
National Research Institute of Cultural Heritage, Archive, AI, image search

© 한국기록관리학회

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.

<https://jksarm.koar.kr>

1. 시작하는 말

1.1 배경 및 목적

과거에도 그래왔지만 기록관의 역할과 기능은 광범위해지고 있고, 사회적 변화와 요구는 끊임없이 증가하고 있다. 특히 4차 산업혁명으로 인한 각종 최신 기술의 등장은 사회 전반에 엄청난 발전을 이루었고, 국가 정부정책 및 혁신 종합계획을 보면 ‘디지털로 일하는 정부’의 역할로 인공지능, 빅데이터 등 디지털 기술발전에 따른 행정 업무 효율성 제고를 점점 더 강조하고 있다. 또한 각종 최신 기술을 활용한 분야별 업무서비스의 편의성·접근성 향상이 중점과제로 추진되고 있는 만큼 기록관리 업무 현장에서의 첨단기술의 적용과 중요성은 더욱 커질 것으로 보인다.

실제, 국립문화유산연구원(National Research Institute of Cultural Heritage, 이하 NRICH)¹⁾ 홈페이지(‘국가유산 지식e음’)에서 제공되는 연구기록물을 포함한 정보 및 콘텐츠의 다운로드 건수는 매년 증가하여 현재는 약 30만건을 넘는다.²⁾ 이와 같은 연구정보 데이터양의 폭발적인 증가와 제공되는 자료 유형의 다양화는 변화하는 정보의 홍수 속에서 이용자에게 신속하고 정확한 정보를 제공하고, 다양한 요구를 해결하기 위한 많은 관심과 노력을 기울이게 하는 결과를 가져왔다.

하지만, 이와 같은 일반국민의 높은 기대와 첨단기술이 보편화된 사회적 변화와는 달리, NRICH 기록관에서 보유하고 있는 약 100만 건에 대한 연구 자료를 찾는 데 소요되는 시간은 수주에서 수개월에 걸릴 정도로 예측이 불가하며 정확도도 낮았다. DB구축 전에는 기록연구사가 직접 수작업으로 보존상태가 열악한 앨범 및 인화사진, 슬라이드 필름을 일일이 확인하여 해당 연구기록물을 대조해가며 기록물을 제공하였다면, DB구축 이후에는 비교적 수월했지만 여전히 해당 기록물의 메타데이터 부족, 오류, 정확성 낮음으로 자료를 찾아 제공하는 데 한계가 있었다.

예를 들어, 문화유산 보존처리를 위해 특정 사찰의 ‘대웅전’ 과거 이미지를 찾고자 할 때, 해당 이미지의 메타데이터 부재와 과거 잘못된 편철 방법으로 인한 정보의 오류는 업무의 비효율성으로 이어지고, 대내외 서비스 제공에 걸림돌이 되었다. 이를 위해 대량의 기록물을 빠르게 처리하여 업무 효율과 정확도를 향상시키고, 검색되는 대상의 내용을 분석하여 관련된 맥락 정보들을 연결해 주는 서비스가 반드시 필요했다. 특히, NRICH 기록관에는 문화유산 현장사진(정기조사, 발굴조사 등), 실측 및 수리도면, 보존처리 기록카드, 모사도, 탁본 등 50여 년 동안 축적해온 연구기록물이 산재되어 있는 만큼 이를 효과적으로 관리하고, 검색하고, 공공에 서비스할 수 있는 첨단기술 적용이 시급했는데, 이는 대량의 기록물 중 특히, 시청각(이미지) 자료를 보유하고 있는 기관의 기록물관리 담당자라면 누구나 쉽게 공감할 것으로 생각한다.

이에, 본고에서는 NRICH 기록관에서 진행해 온 ‘AI기반의 문화유산 연구기록물 학습데이터 및 검색시스템 구축’ 사례를 통해 최신 첨단기술과 기록관리 분야의 접목이 실제 기록관 업무뿐 아니라 기록정보서비스에 새로운 가능성을 확인하는 사례로 소개하고자 한다. 이와 같은 시도는 변화하는 환경과 이용자 요구에 대한 적극적인 대응이기 때문이다. 본 사례를 통해 대부분의 ‘1인 기록관 체제’에 업무 효율을 높이고, 기록정보서비스의 만족도 역시 높아지기를 기대한다. 아울러 기록관리, 문화유산 분야 연구자들뿐만 아니라, 일반대중에게도 유용한 정보로 활용되기를 바란다. 다만, 소개하고자 하는 사례인 ‘문화유산 찾아-ZOOM’은 현재 고도화 및 후속 연구가 꾸준히 이루어지고 있는 만큼 각 분야에서 특성에 맞게 활용·확장하여, 발전시킬 필요가 있겠다.

1) 국립문화유산연구원은 ‘국립국립문화재연구소(1995년~2022년)’에서 ‘국립문화재연구원(2022년2월22일)’으로 변경되고, 2024년 5월 17일 문화재청이 ‘국가유산청’으로 변경됨에 따라 명칭이 변경되었다.

2) 최근 4년간 다운로드 건수: 2020년(211,967건), 2021년(241,651건), 2022년(227,869건), 2023년(357,433건)

1.2 선행연구

4차 산업혁명으로 첨단기술의 분야별 영향력이 확장됨에 따라 기록관리 분야에도 최신 기술을 적용하기 위한 다양한 방안들이 연구되었다. 안대진 외(2017)은 신기술을 활용하여 기록관리를 더 잘 할 수 있는 방안으로 빅데이터, 인공지능, 사물인터넷, 클라우드 컴퓨팅, 네 가지 기술들의 개념과 특징, 적용분야를 정리하고 대표적 기술들을 제시하여 관련 연구를 구체화하는 기초가 되었다. 특히 신기술을 이해하고 새로운 기록관리의 틀을 설계하는데 많은 도움이 되었다. 김인택 외(2017)은 4차 산업혁명의 핵심기술인 인공지능을 중심으로 기록관리에 적용할 만한 분야를 텍스트, 영상인식, 음성인식으로 도출하여 활용하기 위한 준비 내용과 해결 과제들을 제시하였다. 김태영 외(2018)은 지능형 기록정보서비스를 위한 선진 기술 현황 및 국내외 기록관, 도서관, 박물관 서비스 현황을 비교·분석하였다. 조민지(2020)는 4차 산업혁명 데이터 기반 사회에서 기록관리 기관의 정체성, 기록관리 직의 역할 등에 대한 전망과 방향을 강조하였다. 강윤아 외(2023)은 전자기록관리 업무의 효율화 및 기록정보서비스의 이용자 만족도를 위하여 기록관리 현업에서 생성형 AI를 활용할 수 있는 실무 방안을 업무 영역별로 제시하였다. 이들 연구는 첨단기술의 장점과 단점을 분석하여 기록관리 업무에 적용하려는 시도와 실현 가능한 방안을 제시하였다는 점에서 시사점을 주고 있다.

첨단기술의 기록관리 분야 실제 적용을 제시한 연구도 다수 선행되었는데, 이창희 외(2018)은 기록정보서비스를 제공할 수 있는 메신저 기반의 챗봇(chatbot) 프로토타입을 실제 개발하여 이론적 설계부터 개발까지의 절차를 제시하여 기록관과 이용자의 상호작용이 효과적임을 강조하였다. 김해찬술 외(2018)은 서울시의 결재문서를 ETRI의 엑소브레인을 통해 자동분류 테스트를 시도하였다. 그 결과 기록물 분류 업무에서의 인공지능 기술 적용 가능성을 확인하고, 향후 기록관리 기관에 적용하기 위한 고려사항은 제안하였다. 방재현(2018)은 기계학습 알고리즘 중에서 심층신경망을 이용해 기록 텍스트와 기록 이미지를 다루는 업무영역에 의사결정 지원가능성을 실무적으로 검토하였는데, 특히 기록 이미지 기술(Descriptin)을 위한 학습데이터를 국내 문화유산으로 선정하여 전이 학습(Transfer Learning) 방식의 실험을 진행한 결과물은 본 연구에 많은 도움이 되었다. 이은정(2019)은 동영상 기록관리시스템 중에서 구술기록을 관리·서비스 하는 기관의 시스템을 선택분석하여 문제점을 도출하였으며, 지능형 동영상 기록시스템 개발을 위한 다양한 방안을 제시하였다. 강윤아 외(2021)은 영구기록물관리기관에서 수행하고 있는 현행의 직제분석 업무 프로세스를 분석하고, 기계학습 기법을 비롯한 다양한 IT기술을 접목하여 자동화가 가능한 영역을 도출하였다. 백영미 외(2022)는 실무적 관점에서 기능분류체계를 점증적으로 개선한 사례를 통해 신기술활용을 위한 사전준비와 기대효과를 소개하였다. 임진술 외(2023)는 기관의 직제가 바뀌거나 업무가 신설될 때 해당 변화를 즉각적으로 기록물 분류체계에 반영할 수 있는 자동화 및 지능화 방안을 실제 적용하고 활용할 수 있는 방안을 제안하였다. 한편, 국가기록원에서는 2023년부터 “국가기록관리·활용기술 연구 개발 5개년 계획(2023-2027)”을 토대로 기록관리 통합플랫폼 개념 모델 설계 연구를 통해 디지털전환 대응 지능형 전자기록관리 기반 구축을 시작한 상태이다. 이후 기록물 빅데이터 활용을 위한 기반기술의 확립과 기록 모니터링, 보존기간 관리, 문서분류, 기록물 언어모델 등 기록관리에 필요한 인공지능 기술을 적용할 계획이다(김용일, 2023).

위의 선행연구들은 기록관리 분야에 최신 기술을 어떻게 접목할 수 있을 것인지에 대한 시사점과 기록관리 업무환경에 대한 적용 가능성을 전망하였다. 또한 이를 통해 예상되는 문제점을 해결하고, 한계로 지적되어 온 사항들을 개선하여 실제 업무에 적용할 수 있어야 함을 제안하고 있다. 이러한 선행연구와 연구동향을 바탕으로 본 사례보고에서는 최신 기술 접목을 통한 기록관리 업무 현장 편의성 제고 및 대내외 기록정보서비스 체계를 구현한 NRICH 기록관의 실제 사례를 소개하고자 한다.

2. AI기반의 문화유산 연구기록물 검색시스템 구축 현황

2.1 ‘문화유산 찾아-ZOOM 소개’

‘문화유산 찾아-ZOOM’은 NRICH 기록관에서 서비스하고 있는 ‘AI기반의 문화유산 연구기록물 학습데이터 및 검색시스템’으로 1973년부터 현재까지 문화유산 분야에서 발간한 간행물에 수록된 이미지를 학습데이터로 구축하여, 유사 이미지를 동시에 제시함으로써 연구 자료에 대한 사전 수요 예측이 가능하도록 선제적으로 제공하고 있는 시스템(ai.nrich.go.kr)이다. 로그인 등 별도의 조치 없이 일반국민 누구나 이용 가능하며, 대국민 공개(2023.12.1.) 이후 적극적인 관심과 확산-공유로 2024년 8월 현재까지 시스템 이용 누적횟수는 약 9,896회를 넘는다. ‘문화유산 찾아-ZOOM’은 당초 내부에서 ‘유사 이미지 검색시스템’으로 명명되다가 사용자가 부르기 편하고, 기억하기 쉽도록 별도의 명칭을 붙여 사용하게 되었다. ‘ZOOM’이 가지는 사전적 의미도 있지만, ‘ZOOM’의 한글 발음 표기인 ‘줌’을 통해 문화유산을 찾아준다는 의미도 포함하고 있다. ‘문화유산 찾아-ZOOM’의 검색 단계별 프로세스는 <그림 1>과 같다.



<그림 1> ‘문화유산 찾아-ZOOM’ 검색 단계별 프로세스

학습데이터의 결과는 유사한 이미지 검색에 대한 신뢰도를 담보할 수 있는 데이터(기록물)로 제공되어야 하므로 검색 대상이 되는 데이터를 다르게 하여 외부용과 내부용 두 가지 버전으로 구축하였다.³⁾ 첫째, 외부용은 현재 서비스 하고 있는 ‘문화유산 찾아-ZOOM’으로 찾고자 하는 문화유산 이미지와 유사한 연구기록물(사진, 도면 등)을 연구원에서 발간한 연구보고서 속에서 검색할 수 있으며, 특히 연구보고서 속 해당 이미지의 수록면(페이지) 확인이 가능하도록 하였다. 이를 위해 NRICH에서 발간하여 ‘문화유산 지식e음’에서 서비스 중인 간행물 총 1,421권을 선정하여, 보고서 내 이미지 약 675,338건을 추출하고 정제-분류-이미지 벡터화 등의 단계를 거쳐 학습된 데이터를 연관 사이트로 링크하여 정확성을 입증할 수 있도록 하였다. 검색 결과 이미지에 대한 상세정보(보고서 원문 등)를 검색할 수 있을 뿐만 아니라 이미지를 이용하여 재검색 하거나, 다른 이미지를 지정하여 검색할 수 있다. 또한 컴퓨터(PC), 모바일, 태블릿 등 어느 환경에서나 검색 결과를 확인 할 수 있도록 범용성을 확대하여 활용도를 높이고, 이용자의 편의성도 고려하였다.

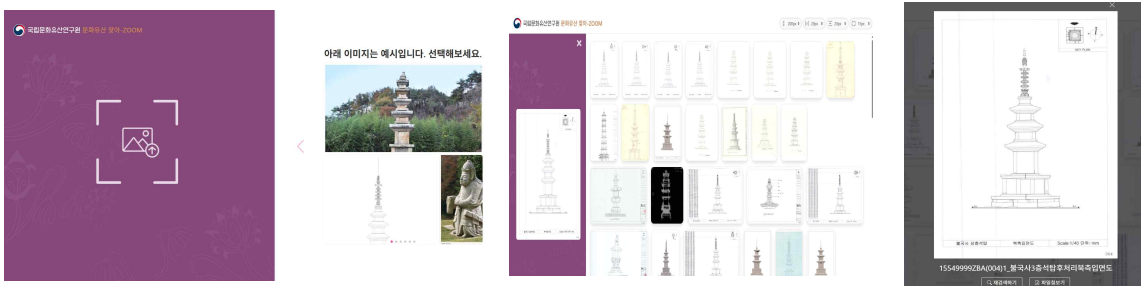
3) 내부용은 현재 업무상 필요한 자료를 검색하는 용도로 향후 데이터 정비 및 충분한 정제-검증을 통해 점차 공개해 나갈 계획이다.



<그림 2> ‘문화유산 찾아-ZOOM’ 사용자 메인화면 및 검색화면(외부용)

둘째, 내부용은 기록관담당자 및 내부 직원들만 접속할 수 있으며, 실제 업무에 활용할 수 있도록 NRICH 기록관에 보유하고 있는 이미지(사진, 도면 등 약 1,045,597건)를 대상으로 내부 서고관리시스템으로 연결되어 검색될 수 있도록 하였다. 외부용과 달리 찾고자 하는 이미지는 기록물이 편철되어 있는 과일철 정보로 연결되며, 이와 연관된 정보인 생산부서, 생산연도, 보존기간, 기록관 내 실물의 위치 정보를 확인할 수 있고, 다운로드 기능을 통해 관련 정보를 언제나 내려받아 활용할 수 있도록 하고 있다. 단, 활용범위가 업무용이나 단순 참고용이 아닌 외부 제공이나 상업적 이용 등의 경우에는 저작권 여부 등을 확인할 수 있도록 별도 지침을 안내하고 있다. 이를 통해 실제로 기록관 내 연구기록물을 찾아 제공하는데 소요되는 시간이 대폭 단축되었고, 내부 직원들이 기록연구사를 거치지 않고, 직접 이미지 검색을 통해 기록관에 있는 연구기록물을 찾아 활용할 수 있어 업무 효율성을 향상시켰다.

이밖에 내부용의 경우, 기록물 정수점검에도 유용하게 활용되고 있다. 검색 대상이 기록물 관리정보와 연계되어 있어 철 단위의 편철이 잘못된 경우를 바로잡을 수 있고, 메타정보 및 부가정보에 대한 오류를 수정할 수 있다. 특히, 과거의 문화유산 연구기록물은 대개 친철하지 못한 경우가 많다. 50년간 쌓아온 자료이기 때문에 누가 보더라도 매우 소중한 자료임에는 틀림없지만, 결과물로만 존재할 뿐이고, 그 생산 과정이 대부분 생략되어 맥락을 유추하기 어려운 기록물들이 상당수이다. 예를 들면, 오래된 흑백사진 속 발굴현장이 몇 년도에 어느 부서가, 왜, 그 지역의 유적을 발굴한 것인지에 대한 기록과 정보가 없어 일반 대중에 서비스를 하고 싶어도 한계가 있을 수밖에 없다. 기록물에 대한 메타정보의 부재는 활용의 효율성과 신뢰성에 큰 영향을 미친다. 즉, 메타데이터가 미비하거나 잘못된 경우 데이터를 올바르게 해석하여 활용하기가 어렵다. 따라서 향후 내부용 시스템은 자원과 지식의 한계가 보완될 수 있도록 내부 직원(연구자들)들의 참여를 높여 기록정보를 편집할 수 있고, 관련된 유용한 정보를 저장하여 활용할 수 있도록 확대해 나갈 계획이다.



<그림 3> ‘문화유산 찾아-ZOOM’ 사용자 메인화면 및 검색화면(내부용)

외부용과 내부용 모두 각각 대상이 되는 데이터를 학습시켜 유사도 순위가 높은 이미지가 검색되는 기술적 방법은 동일하나, 외부용은 일반국민을 대상으로 NRICH 홈페이지에서 텍스트 기반 검색이 아닌 이미지 검색을 통해 연구 정보를 얻을 수 있다면, 내부용은 업무 참고용으로 찾고자 하는 이미지의 기록관 보유 유무 확인 및 해당 기록물의 철·건, 위치 정보 등을 확인할 수 있어, 신속한 기록정보서비스 응대가 가능하다. 또한 기록의 생산자이자 이용자인 내부 직원의 참여를 촉진시켜 기록물의 신뢰성을 확보해 나갈 수 있다. 이렇듯 ‘문화유산 찾아-ZOOM’은 외부용과 내부용이 각각의 목적에 맞게 활용이 가능하며, 연구범위를 확장시키는데 도움을 줄 수 있다. 앞으로 지속적인 관심과 노력을 기울여 잘 관리된 양질의 기록이 모든 일반국민에게 공정하게 제공될 수 있도록 노력해 나갈 계획이다.

2.2 주요 추진경과

인공지능 기술들은 알고리즘 개발과 하드웨어 구축, 그리고 인공지능 학습에 상당한 자금과 시간이 소요된다. 직접 인공지능 인프라를 구축할 경우에는 해당 기관의 업무 효율성이나 서비스 수준을 향상시키는 데 꼭 필요한 인공지능 서비스를 기획하고 적합한 플랫폼을 신중히 선별해야 한다. 정확도 향상을 위한 학습계획을 세우고 샘플 기록물 또한 선정해야 한다(안대진 외, 2017). 이를 위해 NRICH 기록관에서는 연구기록물의 체계화 및 서비스 활용을 위한 전략계획(1~3단계)을 수립하여 예산확보 및 실행을 추진해왔다. 1단계(2021년)는 연구기록물의 빅데이터 구축을 위한 데이터 수집 및 전처리 표준안을 마련하고, 2단계(2022년)는 AI기반의 문화유산 데이터 분석 및 연구 기반정보 구축을 위한 예산 확보 및 대상 분석-선정-도입을 본격적으로 추진하였다. 아울러 기초 데이터 정비 및 서비스 제공 방법에 대한 계획을 수립하였다. 마지막 3단계(2023년)는 실제 시스템을 구축하고, 대국민 공개 이전에 실제 사용해 본 이용자를 중심으로 만족도 조사를 통해 시스템에 완성도를 높일 수 있었다.

물론, 업무를 추진하는데 있어 담당자의 의지와 의욕만으로는 어려운 점이 많다. 이 경우에는 인공지능이나 빅데이터 등 최신 기술과 전문분야 지식을 숙지해야 하는 부담감도 있었지만, 타 기관의 운영 사례나 시행착오에 대한 직접적인 사례가 없었기 때문에 실현시키기 까지 많은 노력을 기울일 수밖에 없었다. 또 대규모 예산이 투입되어야 하는 만큼 내부적으로도 AI기반의 연구기록물 활용에 대한 필요성과 타당성, 적정성 등이 충분히 검토되어야 하고, 예산 및 기타 추진 결과에 대한 자료들이 충분히 제시되어야 한다. 이를 위해 기록관에서는 관련기관 자문회의, 단계별 추진계획과 예상되는 문제점을 지속적으로 점검하고, 사전 설문조사를 통해 이용자 의견을 반영하고자 하였다. 또한 기록관에서 보유하고 있는 기록물에 대한 대내·외 요청건수(최근 3년간 2만여 건)가 매년 증가하는 점에 주목하여 ‘국정 비전 정책 실현’과 ‘인공지능 기술을 융합한 서비스 패러다임’의 전환 시도를 강조하며, 내부역량을 강화하였다.

2.3 사용자 의견수렴

NRICH 기록관에서는 ‘문화유산 찾아-ZOOM’의 대국민 공개를 앞두고, 의견수렴을 위해 2023년 9월 14일부터 2023년 9월 16일까지(3일간) 경주화백컨벤션센터에서 열린 ‘2023 세계국가유산업전⁴⁾’에 참가하였다. 대형 멀티비전(터치형)을 활용해 부스에 방문한 관람객이 직접 시스템을 체험해 볼 수 있도록 하였으며, 서비스를 체험한 방문객을 대상으로 이용 만족도와 향후 이용가능성, 개선사항 등을 알아보았다. QR코드 접속 후 2개 항목 8개

4) 국가유산분야 전문전시회로 2017년부터 추진되었다. 2023년에는 ‘우리 유산의 새로운 시작, 모두가 누리는 미래가치’라는 주제로 96개 업체가 참가하였고, 331개 전시부스가 운영되었다. 전체 관람객은 5,007명으로 집계되었다.

문항에 대한 설문조사를 실시하였는데, 응답자는 총100명(남 30명, 여 70명)으로 응답자 연령은 10대 4명, 20대 39명, 30대 28명, 40대 16명, 50대 이상 13명으로 골고루 분포되어 있었다. 특기할만한 부분은 참여자 중 약42%는 평소 NRICH 기록관(기록물)에 관심이 많아서 방문하였다고 응답했고, 사전 보도자료 등을 통해 '문화유산 찾아-ZOOM'에 관심있는 언론사와 유관기관에서도 직접 방문하여 시연에 참여하였다.

시스템을 직접 체험해 본 이용자의 93%는 정보 검색의 정확도에 대하여는 만족한다는 응답을 하였고, 정보 검색의 충실성에 대하여는 91%가 만족한다는 응답을 하였다. 제공되는 화면 구성의 편의성에 대한 만족도에 대하여는 90%가 만족한다고 응답하였다. 향후 홈페이지에 서비스할 경우 이용 계획이 있는지에 대한 질문에는 71%가 많이 이용할 것 같다는 응답을 하였고, 24%는 필요할 때만 이용할 것 같다는 응답을 하였다. 마지막으로 연구자료 활용에 있어 최신 기술 적용에 대하여는 편리하고, 효과적이고 반드시 필요하다는 의견이 94%로 매우 높음을 알 수 있었다. 주요 설문 내용에 대한 결과는 다음과 같다.

<표 1> '문화유산 찾아-ZOOM' 이용자 주요 설문결과

[문항 5] '유사 이미지 검색시스템' 정보 검색의 만족도: 만족(93%)					
	매우만족	만족	보통이다	불만족	매우불만족
응답자 수	53	40	6	0	1
[문항 6] '유사 이미지 검색시스템' 정보 검색의 충실성: 만족(91%)					
	매우만족	만족	보통이다	불만족	매우불만족
응답자 수	53	40	6	0	1
[문항 7] '유사 이미지 검색시스템' 화면 구성 등 편의성: 만족(90%)					
	매우만족	만족	보통이다	불만족	매우불만족
응답자 수	59	31	9	1	0
[문항 8] 향후 홈페이지에서 서비스할 경우 이용 계획: 이용(95%)					
	많이 이용	필요시 이용	보통이다	이용하지 않음	전혀 이용안함
응답자수	71	24	4	1	0
[문항 9] 최신 기술 적용에 대한 개인적인 의견: 효과적(94%)					
	반드시 필요	편리하고 효과적	잘모르겠다	영향이 없다	전혀 영향이 없다
응답자수	71	23	5	0	1

이밖에 출처가 명확한 연구 자료를 찾을 수 있고, 찾고자 하는 정보를 정확하고 편리하게 얻을 수 있어 유용하며, 찾고자 하는 자료에서 확장된 검색을 통해 다른 연구 자료를 알게 되는 프로세스가 매우 유용하다는 긍정적인 의견이 많았다. 시스템의 개선사항 및 기타의견으로는 영문 버전 지원 필요, 검색 이미지와 관련된 해시태그 제공 등에 대한 제안이 있었다. 더불어, 유료서비스 혹은 회원가입이 필요한 시스템인지에 대한 질의도 꽤 있었는데, 최신 기술을 도입한 유용한 검색시스템을 국가기관에서 선제적으로 선보이는 것에 대한 낮췄과 복잡한 회원가입 절차 등을 염려한 반가운 표현으로 이해할 수 있었다.



<그림 4> 2023 국제산업전 NRICH 기록관 부스 참가 및 ‘문화유산 찾아-ZOOM’ 공개 시연

세계국가유산산업전 참가 및 설문조사 결과를 통해 ‘문화유산 찾아-ZOOM’에 대한 만족도와 새로운 첨단기술과 기록관리 분야에 대한 관심도를 파악할 수 있었다. 또한 시스템에 대한 잠재적인 이용자를 발굴하는 데 의미가 있었다. 이렇게 첨단기술을 활용하여 필요한 정보를 쉽게 검색하고, 활용할 수 있게 한 사례를 경험한 일반 이용자의 요구와 기대는 앞으로 계속 높아질 수밖에 없을 것이다. 예전처럼 수작업을 통해 이용자에게 제공하는 데 걸리는 시간과 노력을 생각한다면, 최신 기술은 단 몇 초 만에 이를 해결해 줄 수 있으니 말이다. 이런 시대적 흐름과 기술의 발전에 따라 기록관의 역할과 기록정보서비스 방향성을 고민해 볼 필요가 있다.

3. ‘문화유산 찾아-ZOOM’의 특징

3.1 기능적 특징

‘문화유산 찾아-ZOOM’은 홈페이지에서 서비스 중인 발간 간행물(1,421권)을 선정하여, 보고서 내 이미지 추출 및 정제 작업을 실시하였다. 여기서 추출된 이미지는 약 675,338건으로 학습 모델을 생성하는데 적합하지 않은 이미지(예를 들면, 아이콘, 중복사진 등)를 제외한 정제된 데이터에 대하여 27종으로 나누어 분류하였다. 이미지 분류에는 다양한 기준이 있겠지만, 본 사례의 경우에는 문화유산의 특성을 고려해 형태별로 분류하고, 그 안에서 또다시 유형별, 크기별로 구분하여 기준을 만들었다. 이렇게 정제된 이미지는 유사 이미지 검색을 위해 벡터 값으로 변환하여 DB에 저장하고, 27종으로 분류된 84,595건은 학습데이터 활용 이미지로 이미지 검색(외부용)에 활용되었다. 학습데이터 이미지 분류 현황은 다음과 같다.

<표 2> 학습데이터 이미지 분류 현황(27종)

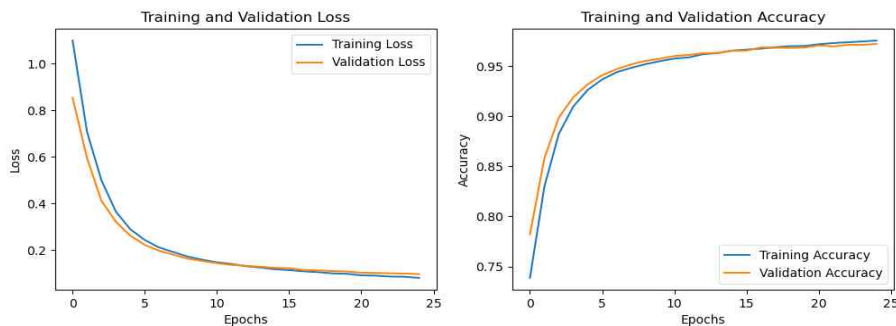
구분	분류	구분	분류	구분	분류	구분	분류
1	건축_근대(부분)	8	그림	15	불상	22	유물_중형(목재)
2	건축_근대(전경)	9	도면_건축물	16	식물사진	23	유물_중형(석재)
3	건축_석조(부분)	10	도면_그림	17	옷	24	지형지도
4	건축_석조(전경)	11	도면_사물	18	유물_대형(금속)	25	책_서류
5	건축_한옥(부분)	12	동물사진	19	유물_소형(금속)	26	표
6	건축_한옥(전경)	13	목판_석판	20	유물_소형(목재)	27	풍경
7	그래프	14	무형_행사	21	유물_소형(석재)		

‘문화유산 찾아-ZOOM’의 이용 방법은 찾고자 하는 이미지를 끌어 놓기(drag and drop) 혹은 브라우징, 예제

이미지를 선택하는 방식으로 검색할 수 있으며, 다양한 형식의 이미지 파일(jpg, png, tiff 등)에 대하여 검색이 가능하도록 제공하고 있다. 검색 결과 목록은 검색 시 사용한 이미지와 유사한 이미지를 유사도 순으로 화면에 표시한다. 이후, 검색된 이미지가 수록된 간행물(원문PDF) 및 문화유산 지식e음의 원문정보 페이지의 열람이 가능하고, 관련 사이트로 연계하여 확장된 결과물을 얻을 수 있도록 하였다. 또한, 검색되는 화면을 이용자 편의에 맞게 구성할 수 있는데 화면의 우측 상단에 있는 조정 기능을 통하여 검색된 이미지의 크기와 이미지 간의 간격(좌우, 위아래)을 조정하여 배열할 수 있고, 검색된 이미지의 테두리를 각지게 혹은 둥글게 조절하여 원하는 자료를 효율적으로 탐색할 수 있도록 하였다.

3.2 기술적 특징

본 사례에서는 유사 이미지 분류를 위해 CNN(Convolutional Neural Network) 기반의 EfficientNetB4모델⁵⁾을 사용하였다. CNN은 이미지나 영상, 음성을 분석하기 위해 패턴을 찾는데 유용한 알고리즘으로 데이터에서 이미지를 직접 학습하고, 패턴을 사용하여 이미지를 분류하는 신경망의 한 종류이다. 일반적인 DNN은 기본적으로 1차원의 데이터를 사용해서 학습해야 한다. 이렇게 1차원으로 바꾸면 데이터 손실이 일어나는데, 이를 해결한 게 CNN이다(이동준 외, 2022). 이와 함께 EfficientNetB4 모델을 활용하여 최적화된 유사 이미지가 검색될 수 있도록 하였다. EfficientNetB4 모델을 기반으로 구축된 84,595건의 학습데이터를 활용한 생성 이미지 분류 모델의 정확도는 약97.2339%이다. 이 결과는 다음 <그림 5>와 같이 훈련 셋(Train set)과 검증 셋(Validation set)의 손실 값(Loss)과 정확도(Accuracy)의 추이에서 살펴볼 수 있다. Training Accuracy와 Valiation Accuracy가 일치하며, 일정 수준부터 손실 값과 정확도에 변화가 없음을 알 수 있다. 이는 27종에 따라 정의된 학습데이터와 검증의 결과가 일치한 결과값으로 적합하게 구축된 것으로 이해할 수 있다.



<그림 5> 이미지 분류 모델 정확도 결과

4. 맺음말

본 사례는 NRICH 기록관에서 추진한 '문화유산 찾아-ZOOM'을 중심으로 문화유산 연구기록물 분야와 AI기반의 첨단기술 실제 적용이 실제 업무와 기록정보서비스에 새로운 가치를 창출할 수 있을 것인지에 대한 가능성과

5) EfficientNet은 Compound Scaling을 통해 기존보다 훨씬 적은 파라미터수 가지고 정확성과 효율성을 향상시킨 Image Classification 모델(SOTA, State-Of-The-Art)이다.

추진과정을 소개하고 있다. 사실 딥러닝을 이용한 이미지 분석을 통한 유사 이미지 검색 기술은 생소한 기술은 아니다. 이미 많은 분야에서 광범위하게 활용되고 있기도 하다. NRICH 기록관은 이러한 최신 기술을 도입하여 보유하고 있는 연구기록물 중 시청각 이미지 자료를 바탕으로 기록관 업무에 필요한 부분을 충족시키면서 새로운 기록정보서비스를 구현하였다. 이는 기존 연구기록물의 오프라인 제공 및 활용에 따른 공간적 제약을 해소하고, 검색 시간과 정확도를 단축하여 열람 및 기록정보서비스 기능을 강화한 사례라고 볼 수 있다. 뿐만 아니라 이미지에 잠재되어 있는 다양한 메타 정보(형태, 모양, 색상, 무늬 등)를 효과적으로 활용해 기존 키워드처럼 검색 자료로 활용하였다는 점에서 의미가 있다.

이를 통해 기록관 업무 담당자는 시간의 효율성과 데이터의 정확성에서 도움을 받을 수 있었고, 이용자들은 각자의 취향에 맞는 정보를 취사선택할 수 있다. 그러나 향후 추진해야 할 과제도 많이 있다. 첫째, 학습데이터의 외연 확대이다. NRICH에서 발간한 간행물 이외의 타 기관 간행물, 논문, 문화유산 자료 등으로 확대하고 관련 기관들 간의 협력을 통해 DB를 확충확장할 필요가 있다. 당장 국가유산청에서 서비스 중인 간행물 5,830권이 대상이 될 수 있다. 둘째, 이미지의 Resize, Rotate, Flip 등의 데이터 전처리를 통한 학습데이터 확충으로 검색 성능을 고도화하는 것이다. 모든 데이터 분석 과정이 그러하듯 이미지 분석에서도 전처리가 매우 중요한데, 이 단계에서는 원본 이미지에 대한 불필요한 요소를 제거하고 분석에 적합한 형태로 변환하여야 한다. 이러한 전처리 과정의 정밀화는 학습을 좀 더 효과적으로 만들어준다. 셋째, '문화유산 찾아-ZOOM'은 이용자가 직접 유적, 유물 등을 촬영한 사진을 검색하였을 때도 유사한 이미지를 찾아낼 수 있는데 현장감 있는 검색 대상을 감안하여 정확도를 제고해야 한다. 넷째, 현재 27종의 분류체계에서의 정확도, 정합성 등을 고려하여 세분화할 필요가 있다. 이미지가 분류체계에서 벗어날 경우 잘못된 정보를 제공할 수 있기 때문이다. 이러한 기술적인 부분과 시스템을 안정화하기 위해서는 지속적인 지원과 연구가 필수적이다.

'문화유산 찾아-ZOOM'은 최신 기술을 기록관리 분야에 접목시켜 서비스하고 있는 사례로 시스템 운영은 아직 1년이 채 되지 않는다. 많은 기대가 모이기도 하지만 과제도 산적해 있다. 앞으로 서비스를 발전시키기 위해서는 단순 검색 성능뿐 아니라 서비스 측면에서 좀 더 편리한 기능을 만들어 내고 어떻게 하면 사용자 요구와 양방향의 기록정보서비스를 구현해 낼 수 있을지 끊임없이 고민하고 발전시켜야 할 것이다. 이러한 변화를 수용하고 도전하는 것은 우리 모두에게 중요한 과제가 될 것이다. 이 사례가 참고가 되어 향후 각 분야에서 심도 있는 후속 연구가 이어지기를 기대한다.

참고문헌

- 강운아, 박태연, 김현진, 오효정 (2021). 생산기관 직제분석 자동화 및 공통 활용 방안. 한국기록관리학회지, 21(4), 81-99. <https://doi.org/10.14404/JKSARM.2021.21.4.081>
- 강운아, 오효정 (2023). 전자기록관리 업무 및 기록정보서비스에서의 생성형 AI 기술 활용. 한국기록관리학회지, 23(4), 179-200. <https://doi.org/10.14404/JKSARM.2023.23.4.179>
- 김용일 (2023). 지속가능한 AI기술개발 및 기록관리분야 적용을 위한 검토 (11-1741050-000032-14). 국가기록원.
- 김인택, 안대진, 이해영 (2017). 인공지능을 활용한 지능형 기록관리 방안. 한국기록관리학회지, 17(4), 225-250. <https://doi.org/10.14404/JKSARM.2017.17.4.225>
- 김태영, 강주연, 김진, 오효정 (2018). 지능형 기록정보서비스를 위한 선진 기술 현황 분석 및 적용 방안. 한국기록관리학회지, 18(4), 149-182. <https://doi.org/10.14404/JKSARM.2018.18.4.149>

- 김해찬술, 안대진, 임진희, 이해영 (2017). 기계 학습을 이용한 기록 텍스트 자동분류 사례 연구. *정보관리학회지*, 34(4), 321-344. <https://doi.org/10.3743/KOSIM.2017.34.4.321>
- 방재현 (2018). 지능형 아카이브 시스템을 위한 기계 학습 기술 적용 방안 연구: 심층신경망 적용을 중심으로. 박사학위논문, 한국의국어대학교.
- 백영미, 문원중, 이혁준, 오효정 (2022). 실무적 관점에서 본 점증적 기능분류체계 개선 사례연구. *디지털문화아카이브지*, 5(2), 31-45. <https://doi.org/10.23089/jdca.2022.5.2.002>
- 안대진, 임진희 (2017). 제4차 산업혁명 기술의 기록관리 적용 방안. *기록학연구*, (54), 211-248. <http://dx.doi.org/10.20923/kjas.2017.54.211>
- 이동준, 전승재, 이동휘 (2022). 이미지 분류를 위한 딥러닝 기반 CNN모델 전이 학습 비교 분석. *한국정보통신학회 2022년도 춘계종합학술대회 논문집*, 26(1), 370-373.
- 이은정 (2019). 지능형 동영상기록 시스템 개발에 관한 연구:한국학중앙연구원 현대한국구술자료관 시스템을 중심으로. 석사학위논문, 명지대학교.
- 이창희, 이해영, 김인택 (2018). 기록정보서비스를 위한 메신저 기반의 챗봇 프로토타입 개발 연구: 명지대학교 대학사료실을 중심으로. *정보관리학회지*, 35(3), 215-244. <https://doi.org/10.3743/KOSIM.2018.35.3.215>
- 임진솔, 한희정, 오효정 (2023). 공공기관 기록물 분류체계 재정비를 위한 지능화 방안: L 기관 사례를 중심으로. *정보관리학회지*, 40(2), 137-156. <https://doi.org/10.3743/KOSIM.2023.40.2.137>
- 조민지 (2020). 기록관리직 다시 생각하기 - 제4차 산업혁명 시대의 기록, 기록화 -. *기록과 정보·문화 연구*, (11), 77-113. <https://doi.org/10.23035/kaics.2020.1.11.077>
- 국립문화유산연구원 (2022). 국가유산 지식이음. 출처: <https://portal.nrich.go.kr>
- 국립문화유산연구원 (2023). 문화유산 찾아-ZOOM. 출처: ai.nrich.go.kr
- 세계국가유산산업전 (2024). 세계국가유산산업전. 출처: <https://heritage-korea.com>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- An, Dae-jin ; Yim, Jin-hee (2017). Application of 4th Industrial Revolution Technology to Records Management, (54), 211-248. <http://dx.doi.org/10.20923/kjas.2017.54.211>
- Baek, YoungMi, Moon, WonJong, Lee, HyukJun, & Oh, Hyo-Jung (2022). A Case study on Incremental Improvement of the functional Classification System from a Practical Perspective. *Journal of D-Culture Archives*, 5(2), 31-45. <https://doi.org/10.23089/jdca.2022.5.2.002>
- Bang, Jae Hyun (2018). A study on application of machine learning for intelligent archive system: Focusing on application of deep neural network. Doctoral dissertation, Graduate School of Hankuk University of Foreign Studies, Korea.
- Heritage Korea (2024). Heritage Korea. Available: <https://heritage-korea.com>
- Jo, Min-Ji (2020). Rethinking Archival Positions: Records and Archiving in the Era of the 4th Industrial Revolution. *The Korean Journal of Archival, Information and Cultural Studies*, (11), 77-113. <https://doi.org/10.23035/kaics.2020.1.11.077>
- Kang, Yoona & Oh, Hyo-Jung (2023). The Use of Generative AI Technologies in Electronic Records Management and Archival Information Service. *Journal of Korean Society of Archives and Records Management*, 23(4), 179-200. <https://doi.org/10.14404/JKSARM.2023.23.4.179>

- Kang, Yoona, Park, Tae-Yeon, Kim, Hyun-Jin, & Oh, Hyo-Jung (2021). Automation and Common Utilization Plans of Job and Organization Analysis of Producing Institutions. *Journal of Korean Society of Archives and Records Management*, 21(4), 81-99. <https://doi.org/10.14404/JKSARM.2021.21.4.081>
- Kim, Hae-Chan-Sol, An, Dae-Jin, Yim, Jin-Hee, & Rieh, Hae-Young (2017). A Study on Automatic Classification of Record Text Using Machine Learning. *Journal of the Korean Society for Information Management*, 34(4), 321-344. <https://doi.org/10.3743/KOSIM.2017.34.4.321>
- Kim, In-Taek, An, Dae-Jin, & Rieh, Hae-Young (2017). Intelligent Records and Archives Management That Applies Artificial Intelligence. *Journal of Korean Society of Archives and Records Management*, 17(4), 225-250. <https://doi.org/10.14404/JKSARM.2017.17.4.225>
- Kim, Tae-Young, Gang, Ju-Yeon, Kim, Geon, & Oh, Hyo-Jung (2018). A Study on the Current Status and Application Strategies for Intelligent Archival Information Services. *Journal of Korean Society of Archives and Records Management*, 18(4), 149-182. <https://doi.org/10.14404/JKSARM.2018.18.4.149>
- Kim, Yong Il (2023). Review for sustainable AI technology development and application of records management field (11-1741050-000032-14). National Archives of Korea.
- Lee, Chang-Hee, Rieh, Hae-Young, & Kim, In-Taek (2018). Development of Prototype Chatbot Based on Messenger App for Archival Reference Services: With Focus on the Archives & Records Center of Myongji University. *Journal of the Korean Society for Information Management*, 35(3), 215-244. <https://doi.org/10.3743/KOSIM.2018.35.3.215>
- Lee, Dong-jun, Jeon, Seung-Je, & Lee, DongHwi (2022). CNN model transition learning comparative analysis based on deep learning for image classification. *Proceedings of Korea Institute of Information and Communication Engineering 2022 at the Spring Conference*, 26(1), 370-373.
- Lee, Eun-jung (2019). A Study on the Development of Intelligent Video Recording System. Master's thesis, Myongji University, Korea.
- Lim, Jin-Sol, Han, Hui-Jeong, & Oh, Hyo-Jung (2023). An Intelligent Approach for Reorganization Record Classification Schemes in Public Institutions: Case Study on L Institution. *Journal of the Korean Society for Information Management*, 40(2), 137-156. <https://doi.org/10.3743/KOSIM.2023.40.2.137>
- National Research Institute of Cultural Heritage (2022). Knowledge of national heritage. Available: <https://portal.nrich.go.kr>
- National Research Institute of Cultural Heritage (2023). Finding Cultural Heritage - ZOOM. Available: ai.nrich.go.kr