

# 국가기록포털 검색 품질 개선을 위한 파일럿 시스템 구축 및 실효성 검증\*

## Construction of Pilot System to Improve Search Quality in National Archives of Korea Portal and Effects Validation

소현기(Hyeon-Gi So)<sup>1</sup>, 염경록(Gyung Rok Yeom)<sup>2</sup>, 오효정(Hyo-Jung Oh)<sup>3</sup>

E-mail: 95dizyso@gmail.com, yumgr@iwaz.co.kr, ohj@jbnu.ac.kr



<sup>1</sup> 제1저자 전북대학교 일반대학원 기록관리학과 석사과정  
<sup>2</sup> (주) 아이와즈, 솔루션개발팀장  
<sup>3</sup> 교신저자 전북대학교 문헌정보학과 교수, 문화융복합아카이빙연구소 공동연구원

논문접수 2023.04.25  
최초심사 2023.05.07  
게재확정 2023.05.19

### ORCID

Hyeon-Gi So  
https://orcid.org/0000-0001-9450-8806  
Yeom Gyung Rok  
https://orcid.org/0000-0001-5319-613X  
Hyo-Jung Oh  
https://orcid.org/0000-0001-8067-2832

### © 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.

\* 본 논문은 2023년도 한국연구재단 연구비 지원에 의한 결과의 일부임 (NRF-2021R111A3047435).  
본 논문은 '2022년 국가기록관리 활용 기술 연구개발(R&D) 사업' 결과의 일부를 확장, 보완한 것임

### 초 록

국가기록원에서는 국민에게 소장하고 있는 공공기록물에 대한 접근점을 제공하기 위해 국가기록포털이라는 대국민 검색서비스를 운영 중이다. 그러나 지속적으로 검색 결과에 대한 이용자 만족도가 낮다는 의견들이 수렴되고 있으며, 그와 더불어 포털 이용률도 감소하고 있는 추세이다. 본 연구는 이러한 상황을 극복하기 위해 수행한 국가기록포털 검색서비스 품질 점검 연구의 후속 연구로, 국가기록포털의 문제점을 규명하고 그에 따른 개선안을 제안하고, 나아가 그 실효성을 검증하는 것을 목표로 한다. 선행된 품질 평가를 통해 국가기록포털의 주요 문제점으로 검색 도구의 저조한 성능, 검색 결과의 일관성 결여, 기본검색 기능 부재를 비롯한 6가지 문제점을 도출하였으며, 이에 대한 개선방안을 규명했다. 제안한 방안의 실효를 검증하기 위해 이 중에서 현실적으로 당장 도입 가능한 방안을 적용한 파일럿 시스템을 구축, 국가기록포털과 검색 성능 비교를 수행하였다. 평가 결과, 파일럿 시스템 검색 도구의 정확률, 재현율, MRR 모든 측면에서 유의미한 상승을 확인하였으며 그 효과를 입증하였다.

### ABSTRACT

The National Archives of Korea (NAK) operates the NAK Portal as a record search system. However, user search satisfaction is too low, and the number of visitors to the portal is gradually decreasing. This study identifies the portal's issues, proposes feasible improvements, and constructs a pilot system to validate the solutions. The preliminary assessment revealed six major issues, such as poor search tool performance and the lack of consistency in search results. After clarifying the improvement measures, a pilot system was established and compared with the National Records Portal. The evaluation showed significant performance improvements in the pilot system, such as Precision, Recall, and Mean Reciprocal Rank (MRR).

Keywords: 국가기록포털, 기록정보서비스, 정보검색, 검색평가, 파일럿 시스템  
National Archives of Korea (NAK) Portal, Record information services, Information retrieval, Retrieval evaluation, Pilot system

## 1. 서론

### 1.1 연구배경

국가기록원은 법률에 의거, 이용자에게 기록정보서비스 차원에서 적절한 검색 도구를 만들어 기록물을 제공할 의무가 있다. 먼저, 『공공기록물법』 제4조 2(공무원의 등의 의무)에서 “공공기관 및 기록물관리기관의 장은 기록물이 국민에게 공개되어 활용될 수 있도록 적극적으로 노력하여야 한다.”고 규정하고 있다. 동법의 제38조 2(영구기록물관리기관 보존기록물의 활용) 역시 “영구기록물관리기관의 장은 그 기관이 보존하고 있는 기록물의 공개 및 열람 편의를 제공하기 위하여 기록물을 정리(整理)·기술(記述)·편찬하고, 콘텐츠를 구축하는 등의 사업을 추진하여야 한다.”고 규정하고 있다. 또한 『공공기록물법 시행령』 제74조의 2(보존기록물의 온라인 검색·열람)를 통해 “영구기록물관리기관의 장은 해당 기관에서 보존하고 있는 기록물을 공공기관이 직접 검색·열람할 수 있도록 온라인서비스를 제공할 수 있다.”라고 명시하고 있다. 이에 따라 영구기록물 관리기관인 국가기록원은 기록물 검색 시스템 국가기록포털을 운영 중이다.

국가기록원은 “대국민 기록정보제공을 통해 지식정보사회의 선도를 목표 및 임무로 삼고 있다. 특히 정보화시대에 발맞추어 국가기록원은 공개 가능한 기록물들의 원본을 원문서비스를 통해 국가기록포털에서 제공 중이다. 실제로 이용자들은 주로 토지, 사건, 도면, 정보, 특정 인물에 대한 기록물들에 관심을 갖고 국가기록포털에 방문한다(진주영, 이해영, 2018). 그러나 다수의 연구를 통해 포털 이용자들의 검색 만족도가 그리 높은 편은 아님을 공통적으로 확인할 수 있다. 이용자들은 생소한 기록관련 용어와 계층구조, 가독성이 낮고 불친절한 인터페이스 등의 문제로 검색에 어려움을 겪고 있다(이윤령, 이해영, 2014). 또한 기록정보콘텐츠는 포털사이트 내의 기록물과 제대로 연계되지 않으며, 이용자가 자주 검색하는 기록물을 잘 찾아주지 못한다(백지연, 오효정, 2019). 이러한 이용자 불만은 서비스 이용률에 그대로 투영되어 국가기록포털 이용률은 점점 줄어들고 있는 상황이다. 일례로 2021년의 경우, 사회적 거리두기로 인해 사람들의 인터넷 사용량이 대폭 증가했음에도 불구하고(과학기술정보통신부, 2022), 국가기록포털의 2019년부터 2021년에 이르는 페이지뷰 수는 오히려 감소하는 추세를 보인다. 상기한 낮은 검색 만족도, 기록물연계 및 검색 성능 미흡, 이용자 수 감소 등의 문제와 더불어 국가기록포털 개선에 대한 요구는 지속적으로 제기되어 왔다(강례림, 2020; 김지현, 2012).

이에 국가기록원은 “국민의 알권리 보장 및 기록정보 서비스 강화”를 주요 과제로 설정하고 “주요기록의 적극적 공개 및 국민 친화형 서비스 확대”를 추진하기로 결정했다(국가기록원, 2022). 상기 목표를 달성하기 위해 규격화된 평가셋을 구축하고 수치화된 평가지표에 기반해 국가기록포털 검색서비스 품질 점검을 수행했다(나정호 외, 2022). 품질 점검과정에서 여러 문제점이 발견되었지만, 그중 제일 시급히 해결해야 할 문제로 국가기록포털에서 제공 중인 검색 도구들의 저조한 성능에 대한 개선이라는 결론을 얻었다. 본 연구는 2022년 수행된 국가기록포털 검색서비스 품질 점검 연구(나정호 외, 2022)의 후속연구로, 국가기록포털의 검색 품질 향상을 위한 다양한 방안을 제안하고, 개선방안이 적용된 파일럿 시스템을 구축, 검색 성능을 비교함으로써 그 실효성을 검증하는 것을 목표로 한다.

연구방법은 다음과 같다. 먼저, 2장에서 선행된 국가기록포털 품질 점검 결과로 도출된 문제점과 그에 대한 원인을 파악한다. 3장에서는 도출된 국가기록포털 문제점에 대한 개선방안을 제안하고 이를 검증하기 위한 파일럿 시스템 구축 과정에 대해 기술한다. 4장에서는 다양한 검색 품질 평가지표로 파일럿 시스템과 기존 국가기록포털의 성능 비교를 수행하여 개선방안의 실효성을 검증한다. 마지막으로, 본 연구의 의의 및 한계를 논의하고, 향후 국가기록포털 ‘검색서비스 고도화’에 도움이 될 만한 기술을 제시하는 것으로 결론을 맺는다.

## 1.2 선행연구

포털 검색 시스템 개선에 관한 연구들은 여러 관점에서 분석이 이루어졌다. 먼저 사용성 관점의 연구들은 이용자 집단으로부터 설문, 인터뷰 혹은 검색 로그를 수집해 분석하는 경우가 다수로, 박민수와 현미환(2011)은 과학기술정보시스템 웹 사용성 평가를 위해 사이트 이용 태스크 및 태스크 시나리오를 설정하였다. 이용자들이 태스크를 수행하는 동안의 검색 로그를 수집하여 효과성, 만족도, 오류 등의 항목으로 사용성에 대한 정량적 분석을 통해 개선 전후 사용성 평가를 측정했다. 씽크 어라운드와 사후 인터뷰를 통해 구두 데이터를 수집하고 정성적 분석을 수행하여 태스크별 개선점을 도출했다. 국가기록포털 관련 연구로는 진주영과 이해영(2018)이 국가기록원 웹사이트 이용자 질의 로그 분석을 통해 유입경로와 주로 사용하는 검색어를 규명하였다. 이를 기반으로 국가기록원 웹사이트 연계성 및 접근성 강화를 위해 포털사이트 적극적 노출, 모바일 서비스 활성화를 제안했으며 검색서비스 고도화 방안으로 검색어 추천서비스, 검색 가이드 제공을 제안했다.

국가기록포털 검색 도구 및 구조에 대한 개선을 논하는 연구들도 있었다. 이윤령과 이해영(2014)은 국가기록원과 해외기록관의 검색 도구별 이용자 그룹을 설정하여 비교를 수행했다. 이용자 중심의 개선방안으로 직관적인 계층정보 제공, 다양한 검색 옵션 제공, 메타데이터 정보 결과 노출, 기록정보콘텐츠 보강 및 개선 등을 제안하였다. 강운아 외(2022)는 해외 영구기록물 관리 기록관에서 제공하고 있는 온라인 포털의 연계구조를 분석하고 차용점을 도출함으로써 국가기록포털의 소장기록 및 내부 서비스, 그리고 콘텐츠의 접근성 및 활용성을 향상시키고, 검색 결과의 유기성을 부여하는 개선방안을 제안했다.

그 외에도 효율적인 기록 탐색을 위한 검색 시스템 구축에 관한 연구들도 다수 진행되었다. 박상현과 김희섭(2014)은 국내 영구기록물 관리기관의 검색 시스템의 사용성을 향상시키기 위해 통합검색과 상호데이터 교환 기능을 갖춘 새로운 검색 시스템을 설계 및 구현하였다. 국내 환경에 적합한 EAD 요소를 선정하고, 이를 기반으로 한국 영구기록물 통합검색 시스템(KAIRS)을 설계한 후, 기존 영구기록물 관리기관의 시스템을 에뮬레이션하여 구현하였다. 이유빈과 이해영(2017)은 RIC(Record In Context) 개념모델과 온톨로지를 기반으로 시멘틱 검색 시스템의 인터페이스를 제안하고, 이용자들의 인터뷰를 통한 사용 경험 평가를 진행하였다. 최근 기록관리 분야에서 관리 대상으로 강조하고 있는 행정정보 데이터세트에 대한 검색 연구도 진행되었는데, 박순호와 김영길(2019)은 의료 분야 데이터 관리를 위해 공공데이터 기반의 질병 검색 시스템을 오픈소스 ElasticsearchTM를 활용하여 개발하였다.

최근에는 인공지능 기술과 검색 시스템의 결합이 이루어지고 있다. 애플 시리, 아마존 알렉사, 네이버 클로바와 같은 음성 AI가 거대 플랫폼 환경에서 검색보조 역할을 하고 있다(김철진, 2018; 송지성, 정다희, 2021). 이종원, 조우승, 김태현(2022)은 NTIS 이용자를 대상으로 KoBERT 모델 기반의 인공지능 챗봇을 도입하여 R&D 검색에 도움이 되는 자연스러운 대화형 검색 시스템을 제안하였다. 인공지능 기술 중 기록물 원문 처리에 필수적인 OCR(Optical Character Recognition)을 활용해 기존의 규격화된 색인에 의존한 검색과 이미지 열람에만 그쳤던 비전자기록의 데이터화도 가속화되고 있다(안세진, 황현호, 임진희, 2022).

이렇듯 이용자가 원하는 정보를 정확하고 편리하게 제공하기 위한 검색서비스 개선에 방점을 둔 다양한 연구가 진행되고 있는 상황과는 달리, 국가기록포털은 2013년 개편(국가기록원, 2015) 이후로 10년간 괄목할만한 개선사항이 없었다. 또한 검색대상 기록물의 원천 데이터가 부실하여 재정비가 시급한 상황이다(나정호 외, 2022). 본 연구는 검색대상 문서의 효과적인 탐색을 향상시키는 방향으로 개선방안을 제시하고, 그 효과를 실제 파일럿 시스템을 통해 구축하는 과정을 포함한다. 이를 통해 제안한 개선안의 실효성을 실증적인 방법으로 검증했다는 점에서 차별화를 둔다.

## 2. 국가기록포털 현황분석

### 2.1 국가기록포털

국가기록원에서는 국가기록포털을 통해 일반 국민들이 국가가 보존, 관리 중인 공공기록물을 쉽게 검색하고 편리하게 열람할 수 있도록 하는 온라인 통합검색서비스를 제공 중이다. 국가기록포털은 중앙기록물 관리시스템(Central Archives Management System, 이하 CAMS)과 연계되어 기록물을 제공하고 있다(설문원, 2022). CAMS는 『공공기록물법 시행령』 제2조에 의거, 국가기록원이 중앙기록물 관리기관으로서 영구기록물을 전자적으로 관리하는 시스템으로, 기록원 내부 검색 도구를 포함하고 있다. 이 시스템은 기록물 유형에 상관없이 전자 및 비전자기록물의 통합 처리와 안전한 보존 및 관리를 지원하며, 국민에게 기록정보서비스를 제공하기 위한 원시 DB를 관리한다(국가기록원, 2015).

2023년 1월 기준, 국가기록포털에서 제공하고 있는 검색 도구로 ‘통합검색’을 비롯해 ‘상세검색’, ‘원문검색’, ‘기술계층별검색’, ‘국정분야주제별검색’ 등이 있다<sup>1)</sup>. 이 중 ‘통합검색’, ‘상세검색’, ‘원문검색’은 국가기록원 CAMS에서 관리 중인 공공기관에서 생성된 30년 이상 보존 기간의 중요한 기록물, 해외 수집 한국 관련 기록물, 민간 소장자로부터 수집 또는 기증받은 기록물, 정부간행물 등의 기록물 중 공개 기록물을 대상으로 하는 검색 도구다(박상현, 김희섭, 2014). 특히 ‘원문검색’은 원문서비스를 제공하는 기록물 검색에 특화된 검색 도구이다. ‘기술계층별검색’은 기록물의 생산맥락에 따라 계층적으로 분류하여 디렉토리 형태로 제공하며, 기록물의 전반적인 구조와 내용을 살펴볼 수 있는 검색 도구이다. ‘주제별상세검색’은 국정 분야별 주요 정책, 제도, 국책 사업 등을 선정 및 구조화하고, 해당 주제와 관련된 기록물 및 목록과 연결해주는 검색 도구이다.

이렇듯 국가기록포털은 일반 이용자가 국가가 소장하고 있는 주요 기록물에 접근할 수 있는 가장 대표적인 도구임에도 불구하고, 이용자들의 검색 만족도를 충족시키지 못하고 있다는 의견이 수렴되고 있다(이효진, 김지현 2021; 장희정, 2012). 특히 2013년 시스템 개편 이후 이렇다 할 개선이 이루어지지 않고 있는 실정이다.

### 2.2 국가기록포털 검색 품질 점검

국가기록포털에 대한 검색 만족도를 측정한 연구가 다수 진행되었으며(이효진, 김지현 2021; 장희정, 2012; 최선희, 2008), 이들 연구 대부분은 일반 이용자들을 대상으로 직접 설문을 수행하거나 관찰 조사 방법을 통해 조사된 연구들이다. 이와는 달리 나정호 외(2022)에서는 국가기록포털의 검색 품질을 수치화된 지표를 통해 평가했다. 이를 위해 이용자 요구가 반영된 평가셋에 기반해 검색 품질을 측정하기 위한 다양한 평가지표를 적용, 객관적인 품질 점검에 주안점을 두었다.

선행연구에서 진행한 국가기록포털 검색 품질 점검은 키워드 검색이 가능했던 ‘통합검색’, ‘상세검색’, ‘원문검색’, ‘주제별상세검색’ 4가지 검색 도구를 대상으로 진행되었다. 품질 점검을 위해 먼저 대상 문헌, 질의셋, 정답셋으로 구성된 평가셋을 구축하고, 이를 기반으로 국가기록포털 품질 평가를 직접 수행하고, 검색 성능을 평가지표로 산출했다. 일반적으로 검색 시스템 성능 평가에 많이 사용되는 정확률(Precision)과 재현율(Recall), F-score, MRR(Mean Reciprocal Rank)을 사용하였다(이병길, 김희섭, 2013; Voorhees, 1999). <표 1>은 검색 도구별 품질 점검 결과로, 전반적으로 저조한 결과를 보였음을 알 수 있다. 본 연구에서는 선행연구를 통해 측정된 검색 품질을 개선하는데 주안점을 두고, 국가기록포털 품질 점검과정에서 도출된 문제점과 원인을 규명하고자 한다.

1) 2023년 4월 현재(투고시점)에는 국가기록포털 개편으로 ‘통합검색’, ‘상세검색’, ‘분야·유형별검색’, ‘기술계층별 검색’ 도구가 제공되고 있다.

〈표 1〉 검색 도구별 성능 평가 결과

검색 도구	결과 제시 질의 수	검색 제시 수 평균	찾은 정답 수 평균	정확률(Precision)		재현율(Recall)		Macro F-score
				Macro-Avg.	Micro-Avg.	Macro-Avg.	Micro-Avg.	
통합검색	47	21.32	6.84	38.50%	32.08%	57.73%	51.27%	42.83%
상세검색	44	18.44	5.3	35.12%	28.74%	44.13%	39.73%	34.59%
원문검색	24	8.44	2.18	21.02%	25.83%	16.89%	16.34%	15.19%
주제별 상세검색	22	6.96	2.48	23.00%	35.63%	21.68%	18.59%	15.69%

\* 볼드체: 가장 높은 수치, 이탤릭체: 가장 낮은 수치, 나정호 외(2022)

### 2.3 국가기록포털 문제점

선행된 검색 품질 평가에서 나타난 국가기록포털 문제점을 세부적으로 분석하면 다음과 같다. 첫째, 검색 도구의 성능이 전반적으로 저조하다. 특히, 검색 정확률이 낮게 평가되었는데, 검색 도구 중 평균적으로 성능이 가장 좋았던 ‘통합검색’의 Macro 평균 정확률이 38.50%였으며, ‘원문검색’의 경우 21.02%였다. 이렇게 낮은 정확도는 국가기록포털에서 제공하는 검색 결과에 오류가 다수 포함되어 있어, 이용자가 기대한 정보요구에 맞는 문서를 찾기 매우 어렵다는 의미이다. 또한 재현율의 측면에서도 ‘통합검색’ 성능만 50%를 넘겨 상대적으로 양호한 편일 뿐, ‘주제별 상세검색’이나 ‘원문검색’의 경우 20% 이하의 성능을 보여주고 있다. 이는 기록원이 이미 보유하고 있는 기록물임에도 이용자가 열람할 수 있는 접근점을 제시하지 못하고 있다는 방증이다.

둘째, 검색 결과의 일관성이 결여되어 있다. 국가기록포털의 검색 도구는 이용자 편의를 위해 각각의 목적에 따라 다른 결과를 제시해주도록 설계된다. 따라서 검색 도구 간 성능의 차이가 있을 순 있으나, 문제는 그 간극이 너무 심하다는 점이다. ‘통합검색’과 ‘원문검색’의 Macro 평균 정확률은 거의 2배 차이가 나며, Macro 평균 재현율은 거의 3배 차이가 난다. 또한 일반적인 검색 포털에서는 상세검색을 활용하여 검색 범위를 명확하게 설정함으로써 통합검색보다 검색 정확도가 높게 나타나는 경향이 있는데, 국가기록포털의 ‘상세검색’은 오히려 반대되는 결과를 보여준다.

셋째, ‘기본검색’ 기능이 제공되지 않는다. 일반적인 포털서비스에서의 ‘통합’검색은 질의어 분석을 통해 가장 유사한 문서를 찾아주는 기본검색 기능을 하고, 이용자가 그 결과를 한정시킬 수 있도록 확장도구를 지원한다. 그러나 국가기록포털의 ‘통합검색’은 오직 기록물 철과 건의 ‘제목’에 대한 검색만 이루어지고 기록물의 내용이나 생산기관에 대한 검색은 이루어지지 않는다. 또한 복합 질의어에 대한 토큰화가 제대로 이루어지지 않아 질의어가 조금이라도 검색대상 문서의 키워드와 일치하지 않으면 적절한 검색 결과가 나타나지 않는다.

넷째, 검색어 추천 기능이 부족하다. 국가기록포털을 방문하는 일반 이용자들은 학술 연구자들과 달리 “일반적으로 자신이 찾고자 하는 정보에 접근하는 방식에 대한 이해도가 낮고, 실제로 찾고자 하는 정보가 무엇인지 정확하게 인지하지 못하는 경우가 많다(설문원, 2022).” 따라서 국가기록포털은 일반 이용자의 정보 검색을 도와주기 위해 다수의 이용자가 찾은 정보에 대한 검색어 랭킹 정보나, 이용자 질의분석을 통한 추천 검색어와 연관 검색어를 제공해야 한다. 그러나 현재의 국가기록포털에서는 이러한 기능이 제공되지 않고 있으며, 검색 도구 중 ‘원문검색’에서만 일부 ‘자주 찾는 검색어’를 제공하고는 있으나 이용자들의 요구를 시의 적절하게 반영하기 위한 지속적인 업데이트는 이루어지지 않고 있다.

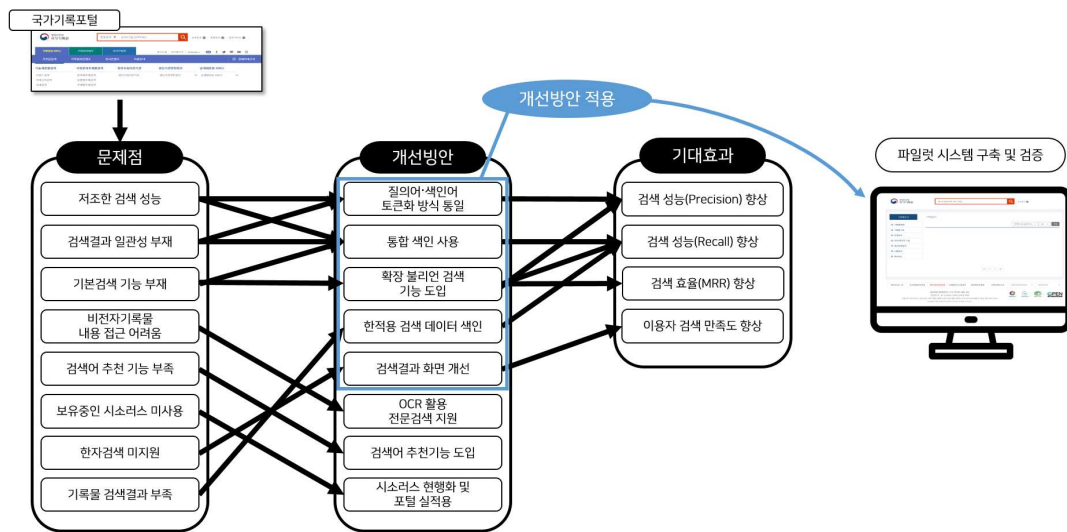
다섯째, 질의어 확장 기능이 부재하다. 일반 국민들의 경우, 행정 기관 체계나 공공업무 기능에 대한 이해가 낮은 편이다. 또한 공공기록물에 사용되는 많은 행정용어에 대해 낯설어하고 실제 질의로써의 활용도도 매우 낮다. 따라서 이용자가 입력한 질의어에 포함된 일반 어휘를 기록 관련 용어로 확장 해석해야 하며, 이러한 기능을 구현하기 위해서는 유의어 사전이나 시소러스와 같은 별도의 지식자원 활용이 수반되

어야 한다. 한편 국가기록원에서는 이미 이 같은 지식자원 중 하나인 기록시소러스를 구축, 2017년 기준 일반시소러스 461,584개, 기능시소러스 48,086개의 용어를 확보(국가기록원, 2018)했음에도 불구하고 현재 포털에는 적용되지 않고 있다.

무엇보다도 가장 근본적인 문제로는 비전자기록물에 대한 접근이 어렵다는 점이다. 현재 국가기록원에서 공개대상인 디지털화된 비전자기록물들은 CAMS와 연계하여 국가기록포털에서 '원문제공서비스'를 통해 원문 열람이 가능하다. 그러나 문제는 원문제공서비스에서 제공하는 문서는 이미지 또는 PDF 형식의 스캔본으로, 기계가 읽을 수 있도록 텍스트로 전사화된 문서가 아니다. 따라서 기록물 제목을 통해서만 문서를 찾을 수밖에 없으며, 문서 제목 안에 기록물과 관련된 키워드가 나타나지 않는 경우, 이를 찾기는 거의 불가능하다. 이러한 문제를 해결하기 위해서는 기록물 원문에 대해 기계가독형(machine-readable) 전처리 과정이 필수 수반되어야 한다. 그 밖에도 다수의 기록물 제목에 한자가 사용되고 있음에도, 한자 검색을 지원하지 않아 탐색이 어렵다는 점, 기록물 검색 결과가 담고 있는 정보가 부족하다는 점, 별도의 검색 결과 정렬기능이 없다는 점 등이 있다.

### 3. 파일럿 시스템 구축

#### 3.1 검색 품질 개선방안 도출



〈그림 1〉 국가기록포털 문제점 및 개선방안

본 장에서는 2장에서 도출된 국가기록포털 검색 시스템의 문제점에 대한 개선방안을 제시하고, 그 실효성을 검증하기 위한 파일럿 시스템 구축 과정에 관해 기술한다. <그림 1>은 도출된 문제점에 대한 개선방안과 그에 따른 기대효과를 종합하여 도식화한 그림으로, 자세한 설명은 다음과 같다. 국가기록포털 검색 시스템에서 해결해야 할 가장 시급한 문제는 낮은 성능이다. 특히 정확률이 매우 낮은 편으로, 이는 사용자가 입력한 검색어를 처리하는 방식과 검색대상 기록물로부터 색인어를 추출 방식의 불일치로 인함이다. 또한 도구별 검색 방식 차이로 인해 검색 결과 역시 서로 불일치한 정보를 제시하고 있다. 이를 해결하기 위해서는 질의어와 색인어 추출 방식을 통일하고 일관성 있는 정책을 적용하는 것이 가장 급선무이다. 또한 기록물의 특성에 따른 검색 후보 필드를 선별하고, 실제 색인 대상 데이터 유형에 따라 동일한 토큰화를 수행해야 한다.

두 번째 개선방안은 통합색인 방식 적용이다. 검색 도구별로 실제 검색대상 필드를 세부 분석한 결과, 상기 지적한 질의어/색인어 처리 불일치 외에도 CAMS에서 관리 중인 기록물 DB 필드별로 서로 다른 색인어 추출 방식을 적용하거나 여러 색인 필드에 산재, 분절되어 저장되고 있었다. 또한 유사한 기능을 수행하는 도구들임에도 불구하고 서로 다른 필드를 대상으로 검색하고 있음이 파악되었다. 본 연구에서는 이러한 문제를 해결하기 위해서 ‘통합색인’ 방식을 제안한다. 여기서 ‘통합’이란 CAMS에서 관리 중인 검색대상 기록물 건과 철 제목으로부터 일관된 색인어 추출 규칙을 적용하여 색인어를 추출하고 이를 통합된 색인 필드에 저장함으로써, 키워드가 분산되어 있어도 검색이 가능하게 하는 방식이다. 이러한 통일된 색인 방식을 통해 모든 검색 메뉴에서 일관성 있는 결과를 보여주며, 검색 결과 필터링 기능 활용 시 상위 계층 문서 정보를 활용해 하위 계층 문서 정보 이용을 가능하게 한다.

세 번째 개선방안은 ‘확장 불리언’ 검색 기능 도입이다. 다른 여타의 포털시스템과 달리 국가기록포털의 ‘통합검색’은 이용자 질의어를 모두 ‘AND’처리 하고 있다. 따라서 질의 키워드 중 하나만 부재하더라도 검색 결과가 도출되지 않는다. 특히 모든 키워드가 매칭된 문서임에도 상위에 올라오지 않는 문제가 도출되었다. 이에 대한 대안으로 확장 불리언 검색 기능을 도입, 특정 검색 질의어가 존재하지 않더라도 가중치 알고리즘 랭킹을 통해 가장 유사한 문서를 검색해주는 기능을 통합검색의 기본검색으로 적용하는 방안을 제안한다. 또한 모든 키워드가 매칭된 문서가 상위에 올라오도록 검색질의어(query)를 구성함으로써 재현율과 정확률을 개선하고 더불어 검색 효율 향상을 꾀한다.

네 번째 개선방안은 검색어 추천기능 도입이다. 국가기록포털에서 보유 중인 이용자 질의 로그를 활용해 이용자들이 자주 검색하는 인기 검색어를 분기별로 최신화한다면, 자주 찾는 기록물에 대한 검색 가이드 및 추천검색어 컬렉션의 접근점으로 인기 추천어를 제공하여 이용자 검색 편의성을 증대시킬 수 있다. 또한 이용자가 2차 검색으로 사용한 검색어를 통해 연관 검색어를 제공함으로써 정보 탐색을 위해 검색어 선택에 고민하는 시간을 줄이고, 각 정보에 적절한 검색어로 변경하여 이용자의 검색 만족도를 증대시킬 수 있다.

다섯 번째 개선방안으로는 시소러스 현행화 및 포털 검색 실적용을 제안한다. 시소러스는 기록물 분야에 특화된 지식베이스 구축에 초석이 되는 중요한 용어사전으로, 향후 인공지능 기술 적용 시 검색 품질 향상에 도움을 줄 수 있다. 이를 위해 기구축한 국가기록 시소러스를 현행화하고 관리도구를 개발해 지속적인 유지보수를 수행해야 한다. 현행화 및 관리도구 개발 완료 후에는 우선적으로 질의어 확장에 도움을 주는 유의어, 관련어 사전을 구축하고 서비스에 반영한다. 더 나아가 이용자가 사전을 통해 직접 질의어를 선택해 확장하는 것이 아닌 시스템에 의한 자동 확장 검색 기능이 도입된다면 이용자 편의성과 검색 재현율이 증대될 것으로 기대된다.

여섯 번째 개선방안은 OCR 기술을 활용한 전문(Full-Text)검색 지원이다. 국가기록원은 이미 2019년부터 OCR 기술의 기록관리 업무에 적용 가능성에 대한 검토를 연구했고, 2020년에는 연구개발 사업을 통해 데이터셋 구축 및 OCR 모델 개발을 완료했다(국가기록원, 2021a). 2022년 12월 31일 기준, 디지털화된 비전자기록물 수(건)는 총 99,952,533건 중 37,685,200건으로, 이 중 문서 형식의 기록물에 대한 디지털화는 20,007,038건이며 지속적으로 디지털화 작업이 진행되고 있다(국가기록원, 2023). 이렇듯 국가기록원에서 확보 중인 OCR 모델을 통해 디지털화된 스캔 문서들로부터 색인 대상 텍스트를 추출하여 기계가독형 문서로 변환시킨 후, 전문 내용이나 내용 요약에 대한 색인을 생성하여 통합색인으로 관리한다면 검색 재현율 향상에 효과적이다. 다만, 국가기록포털은 CAMS에서 공개하는 정보만을 제한적으로 서비스해야 한다는 점을 고려해, 사전에 『정보공개법』 제9조(비공개 대상 정보)에서 규정한 국가안보, 통일, 외교관계, 개인 정보 등의 비공개 정보들은 제외하고 색인해야한다.

그 밖에도, 한적(漢籍)문서를 찾기 위해 한적용 검색 데이터 구축 및 색인을 통해 검색 재현율을 향상시키거나 이용자 검색만족도 향상을 위한 검색 결과 화면 개선을 제안한다. 현재 국가기록포털의 검색 결과목록은 ‘기록물 제목’, ‘생산기관’, ‘생산연도’, ‘관리번호’만 제공되며 이용자 입장에서 궁금해할 만한 정보가

부족하다. 따라서 주제, 내용 요약과 같은 기록물의 일부 정보를 제공할 필요가 있으며, 기록물 원문의 경우 ‘기록물 뷰어’를 통해서만 보는 것이 아닌 커서를 올렸을 때 간략하게 확인 가능한 미리보기 기능이 필요하다.

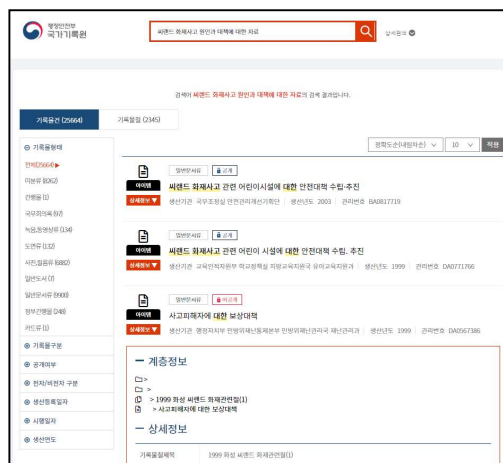
### 3.2 파일럿 시스템 구축

이번 절에서는 3.1절에서 도출한 개선방안을 실제 검색 시스템에 적용하고, 그 실효성을 검증하기 위한 파일럿 시스템 구축 과정에 대해 기술한다. 다만 도출된 개선방안 중 OCR 적용을 통한 전문 입수 방안은 기록원 내부 업무 프로세스에 대한 대대적인 조정이 선행되어야 한다는 제약이 있어 제외하였다. 더불어 시소러스 활용 및 검색어 추천 등의 방안 역시 시스템 적용을 위해서는 사전에 관련 지식자원을 구축하는 비용과 시간이 소요되며, OCR 적용과 마찬가지로 전체 기록원 내부 정책 수립이 선행되어야 하는 한계가 있어 제외하였다. 결과적으로 제도적, 현실적 판단에 따라 검색엔진 및 질의 인터페이스 개선 등 당장 적용 가능한 개선방안을 중심으로 파일럿 시스템을 구축하였다(<그림 1>의 파란 박스 참조).

먼저 검색 시스템 구축을 위해서는 검색대상 문서 집합 설정이 필요하다. 2021년 12월 기준, 국가기록원에서 관리 중인 기록물은 총 114,777,646건(국가기록원, 2021b)이다. 객관적인 평가를 통한 국가기록포털과의 비교를 위해서는 전체 데이터를 확보하여 동일한 환경을 구축하여야 하나 기록원 정책상 보안의 문제로 데이터 반출이 불가하였으며, 이 정도로 방대한 데이터를 처리하기 위해서는 시간과 비용적인 어려움이 있었다. 따라서 파일럿 시스템 구축용 데이터는 선행연구(나정호 외, 2022)에서 수행된 검색 평가 결과와의 비교를 우선 목표로 삼고, 선행연구에서 선정한 50개의 질의어에 대한 130,713건의 데이터를 색인 범위로 선정하였으며 전처리와 변환 과정을 거쳐 색인 데이터로 구축하였다. 데이터 변환 과정에서 기록물 제목과 기록물 정보에 대한 통합색인 필드를 생성하고, 데이터를 통합하였다.

파일럿 시스템에서 사용한 검색기는 ‘Lucene’기반 오픈소스 검색엔진 ‘ElasticsearchTM’를 차용하였다. 특히 질의어가 모두 일치하는 문서가 높은 순위에 노출되고, 입력 질의어의 순서에 따라 추가적인 가중치를 부여하여 이용자에게 관련성 높은 결과를 제시하도록 순위화하는 모듈을 구현하였다. 또한 ElasticsearchTM에서 제공하는 질의어 인터페이스 확장 기능 활용하여 제안한 확장 불리언 검색 기능을 구현하였다.

한편, 기존 국가기록포털은 ‘기본검색’의 역할을 하는 도구가 없으며 그 역할을 해야 하는 ‘통합검색’이 기록물에 대한 종합적인 검색이 아니라 단순히 제목을 검색하는 기능만을 해, 일반 이용자가 질의어 키워드 선정 및 원하는 질의 결과를 얻기 어려웠다. 본 파일럿 시스템에서는 일반 이용자가 일반 단순 키워드 나열



<그림 2> 파일럿 시스템 검색 결과 화면



〈그림 3〉 '상세검색' 화면 예시

뿐만 아니라 질의 내 조사, 띄어쓰기가 혼합된 자연어 형식을 포함하는 다양한 형태의 질의에 대응할 수 있도록 질의어 처리 모듈을 고도화하였으며, <그림 2>는 자연어 입력 결과에 대한 예시이다.

<그림 3>은 개선된 '상세검색' 화면 예시로, '상세'라는 이름에 걸맞게 검색의 정확도를 올리는 기능을 수행하도록 설계하였다. 검색영역을 '전체', '기록물 제목', '기록물 정보'로 나누었는데, 특히 '기록물 정보'는 기록물 제목 외에 내용 요약, 명칭, 지명 등을 통해 검색되도록 설정하였다. 또한 국가기록포털의 문제점 중 '상세검색'을 기본 설정만으로 수행하였을 때 '통합검색'에 비해 검색 '정확률'이 오히려 낮아지는 문제를 해결하고자 하였다. 이를 위해, '상세검색'의 기본검색 기능이 작동하는 검색영역 전체에 대해 '통합검색'과 동일한 색인 필드를 통해 검색되도록 설계하였다. 그 밖에 검색 외의 한자검색 기능을 원활하게 작동시키기 위해 동형이음어 데이터를 생성하여 색인했다. 또한 검색 결과 화면에 대한 개선으로 검색 결과를 가중치, 서순, 생산일에 따라 정렬하는 기능을 추가하고, 기록물의 검색 가시성을 높이기 위해 매칭된 질의 키워드를 강조하는 하이라이팅(highlighting) 기능을 도입하였다. 하이라이팅은 제목뿐만 아니라 기록물 정보에도 적용되도록 하였다.

1차 시스템 구축이 완료된 후, 본격적인 검색 성능 평가에 앞서 시스템이 원활하게 작동하는지 확인하기 위해 사전 조사를 수행하였다. 그 결과, 연구진이 설계한 기능들은 의도대로 원활하게 작동되었으나, 질의와 전혀 연관성이 없는 기록물이 검색 결과 상위에 등장하는 문제가 발생하였다. 원인은 조사(助詞)에 부여된 가중치가 높아 중요 키워드보다 우선해서 매칭되어 발생한 것으로, 질의어 토큰화가 과하게 적용되어 주요 키워드까지 분절하는 오류가 있었기 때문이었다. 이를 보완해 2차 구축 시에 조사에 대한 가중치를 낮추고 키워드에 대한 가중치를 높이는 방식으로 오류를 해결하였다.

## 4. 개선방안 검증을 위한 성능 평가 및 비교

### 4.1. 검색 성능 평가방법

#### 4.1.1 평가방법

본 절에서는 제안된 개선방안의 실효성 검증을 위해 기존 국가기록포털과 파일럿 시스템의 성능을 비교하고자 한다. 평가 질의는 앞서 선행된 국가기록포털 검색 품질 평가(나정호 외, 2022)에서 선정한 동일한 질의어(50개)를 대상으로 진행하였다. 평가자는 선행된 검색 품질 평가와 마찬가지로 참여한 5인의 기록물

에 대한 기초적인 배경지식을 가진 인원들로 구성했다. 다만, 평가 대상 기록은 ‘건’을 기준으로 한정하였는데, 그 이유는 기록물 ‘철’ 검색의 경우, 일반 이용자가 기록물 계층구조에 대한 이해도가 낮으며(이효진, 김지현, 2021), 국가기록포털 검색 결과 화면에서 가장 먼저 제시되는 내용 역시 건 기록물 검색 결과이기 때문이다. 또한 색인 범위 한정으로 인해 성능 평가대상 검색 도구는 파일럿 시스템에 구현한 ‘통합검색’과 ‘상세검색’으로 한정하여 진행하였다. 특히 ‘상세검색’에 대한 평가는 국가기록포털 정확률의 간극에 대한 개선의 정도를 파악하기 위해 기본 설정으로 검색을 수행하였다.

평가는 선행연구와 마찬가지로 총 3차례에 걸쳐 진행되었으며, 채점은 평가자들이 질의셋을 참고하여 질의어를 입력하고, 검색 결과로 제시된 상위 30개의 결과를 대상으로 하였다. <표 2>는 본 연구에서 보강한 파일럿 시스템 평가의 질의셋 예시이다. 기존에 선행된 국가기록포털 평가 시 1차 검색은 ‘질의어명’으로 단순 키워드 검색을 수행하였다. 2차, 3차 검색은 1차 검색에서 도출된 결과의 양상에 따라 두 가지 방법으로 이루어졌는데, 먼저 검색된 기록물의 양이 많아 상위 30개 결과 내에서 정답 문서를 찾을 수 없는 경우에는 범위를 좁혀서 기록물을 찾기 위해 ‘결과 내 재검색’ 기능을 활용하여 검색을 수행하였다. 만약 검색된 기록물이 적거나 없는 경우에는 <표 2>에서와 같이 기대정보가 반영될 수 있도록 질의어를 수정하는 방식으로 진행하였다.

<표 2> 파일럿 시스템 평가 질의셋 예시

구분	번호	질의어	1차 검색 질의 (질의어명)	2차 검색 질의(검색목적 구체화)	3차 검색 질의(기대한 정보)
고빈도 질의	QL-001	토지조사부	토지조사부	특정 지역(경기도 수원군 음덕면 온석리)에 소유하고 있는 조상땅을 찾기 위해	경기도 수원군 음덕면 온석리 일대의 토지조사부
	...	...	...	...	...
	QL-021	공역	공역	공역의 변화 과정을 조사하기 위해 공역 지정 내역을 조사	시간 흐름에 따른 공역 지정 정보
기록물 유형별 질의	QL-022	박정희 암살	박정희 암살	박정희 암살 사건에 대한 자료조사를 위함	박정희 암살과 관련한 자료
	...	...	...	...	...
	QL-049	평창 동계 올림픽 기념 배지	평창 동계 올림픽 기념 배지	평창 동계 올림픽 기념 배지에 어떤 것들이 있었는지 조사하기 위함	평창 동계 올림픽 기념 배지 목록
생산기 관 관련 질의	QL-050	미래창조 과학부 연구	미래창조 과학부 연구	주요 연구개발사업 성과에 대한 자료조사	주요 연구개발사업 성과에 대한 자료 조사

\* 나정호 외(2022) 재구성

개선된 파일럿 시스템 평가의 경우, 1차 검색은 기존평가와 동일한 방식으로 원질의 그대로 질의어명 검색하였다. 그러나 2차, 3차 검색의 경우, 구축된 파일럿 시스템에서는 조사나 띄어쓰기가 섞인 자연어 형식의 검색이 가능하도록 개선되었기 때문에 문장 형식의 구체적인 질의를 작성을 허용하였다<sup>2)</sup>. 2차 검색은 검색목적의 구체화하여 문장 형식의 질의로, 3차 검색은 기록물을 통해 얻고자 하는 정보를 대상으로 ‘기록물 정보’에 색인된 ‘내용 요약’을 통해 정보를 찾도록 질의를 변경하여 검색을 수행하였다.

검색 결과에 대한 판정은 1차, 2차, 3차 검색 중 어느 하나에서라도 정답문서를 얻었을 경우 검색성공, 얻지 못하였을 경우 검색실패로 분류했다. 파일럿 시스템 평가 과정의 구체적인 예시를 들면, ‘QL-021’의 ‘공역’의 경우, 1차 검색은 ‘질의어명’인 ‘공역’으로 검색했다. 그러나 1차 검색에서 정답문서를 얻지 못함에 따라 2차 검색으로 ‘검색목적 구체화’에서 정리된 ‘공역의 변화과정을 조사하기 위해 공역지정 내역을 조사’

2) 작성된 자연어 문장 형식의 질의어를 기존 국가기록포털에 입력한 경우, 선행평가 결과보다 못한 성능을 보였음

로 검색을 수행했다. 2차 검색에서도 정답문서를 얻지 못해, 3차 검색으로 ‘시간 흐름에 따른 공역 지정 정보’로 검색을 수행한 결과, 정답문서를 얻었으며, 최종 ‘검색성공’으로 분류되었다. 50개의 질의에 대한 검색 성능 평가가 완료된 후, 평가 결과를 바탕으로 다양한 검색 성능 평가지표를 통해 각 시스템의 성능을 수치화하여 검색 성능 비교를 진행했다.

4.1.2 정답셋 확장

파일럿 시스템 검색 성능 평가에 앞서 선행 국가기록포털 품질 점검에서 사용된 정답셋의 적합성을 점검했다. 기존의 정답셋은 CAMS와 국가기록포털의 검색 도구를 통해 도출된 정답문서들로 구축되었다. 그러나, 파일럿 시스템에서는 도출된 개선안이 적용됨에 따라 검색 결과로 제시되는 기록물의 양이 크게 증가하고, ‘기록물 정보’에 대한 검색필드가 추가되면서 기존 국가기록포털에서 검색되지 않던 정답문서들이 추가로 확인되었다. 따라서 기존의 정답셋에 파일럿 시스템에서 새롭게 등장한 정답문서들을 추가하여 정답셋을 보강하였으며, <그림 4>는 확장된 정답셋 예시이다.

질의어번호	질의명	검색 시스템	정답셋		검색도구별 개수	총 합계
			관리번호	제목		
QL-001	토지 조사부	CAMS	CJD0004409	토지조사부(경기도수원군읍면온석리)	2	4
			CJD0004409	토지조사부(경기도수원군읍면온석리)		
		파일럿 시스템	CJD0004409	경기도수원군읍면온석리199-417	2	
			CJD0004409	경기도수원군읍면온석리1-198		
QL-002	씨랜드	국가 기록 포털	DA0567386	화성 씨랜드화재 관련 보고	14	38
			DA0567385	화성 씨랜드 화재사고 수습		
			...	...		
			DA0036795	씨랜드 관련 문서사본		
		CAMS	DA0771766	화성 씨랜드 사안	18	
			BA0817719	씨랜드화재사고관련 (2-2)		
			...	...		
		파일럿 시스템	DA0567388	씨랜드 상황관리 보고서 (2-1)	6	
			DA1299957	씨랜드 청소년수련회 집 관련 예산절이정정		
			DA0771766	씨랜드 화재사고 관련		
QL-049	평창 동계 올림픽 폐기물 배지	CAMS	DH50008159	회기별 배지 : 골드색 상 엠블럼	11	22
		파일럿 시스템	DH50008163	2018평창동계올림픽대회 회기별 배지 : 수호암	1	
QL-050	미래 창조 과학부	국가 기록 포털	DM00054433	2013회계연도 결산보고서 (미래창조과학부) 별책: 미래창조과학부 성과보고서	9	11
			DM00054376	2013회계연도 성과보고서 (미래창조과학부)		
			...	...		
		파일럿 시스템	DM00068133	2016 미래창조과학부 주요 연구개발사업 성과분석보고서	2	
			DM00068133	2013회계연도 결산보고서 (미래창조과학부) 별책: 미래창조과학부 성과보고서		
			DM00068133	2013회계연도 결산보고서 (미래창조과학부) 별책: 미래창조과학부 성과보고서		

<그림 4> 확장된 정답셋 예시

<표 3>은 확장된 정답셋 통계로, 파일럿 시스템의 ‘질의어 검색’을 통해 398개, ‘검색목적 구체화’를 통해 385개 ‘기대한 정보 검색’을 통해 383개의 추가 정답문서를 얻었다. 총 정답셋은 1,445개로, 기존 CAMS에서 수집된 866개, 국가기록포털에서 수집된 1,023개, 파일럿 시스템에서 수집된 1,166개의 정답문서 중 관리번호와 기록물 제목이 중복되는 기록 1,608개를 제외하여, 결과적으로 기존 995개(나정호 외, 2022) 대비 450개의 정답문서가 추가되었다.

〈표 3〉 최종 정답셋 구축 결과

구분	검색 도구	정답문서	검색 도구별 합계	중복개수	총계
CAMS	기록물철상세검색	115	866	1,608	1,445 (기존 995, 나정호 외, 2022)
	기록물건상세검색	154			
	신통합검색	597			
국가기록포털	통합검색	379	1,023		
	상세검색	327			
	원문검색	84			
	기술계층별검색	110			
	주제별상세검색	123			
파일럿 시스템	질의어 검색	398	1,166		
	목적 검색	385			
	기대한 정보 검색	383			

#### 4.2 개선방안 실효성 검증을 위한 성능 비교

검색 성능에 대한 평가 결과 분석 역시 선행된 연구와 같은 방식으로 수행하였다. <표 4>는 기존 국가기록포털과 파일럿 시스템의 평가 결과를 비교한 것으로, 파일럿 시스템의 성능이 전반적으로 향상한 것을 볼 수 있다. 특히 ‘상세검색’과 ‘통합검색’ 사이의 정확률 간극이 해결되었는데, 이는 기존 포털이 전체 50개 질의 중 각각 45개, 37개의 질의에 대해서만 검색 결과를 제시한 반면, 파일럿 시스템은 49개의 질의가 정답문서를 제시한 것으로도 알 수 있다. 검색에 실패한 질의(1건)의 특성을 살펴본 결과, 이용자가 요구한 기록물의 유형 구분이 ‘행정박물’로 되어있어 애초 정답셋에 ‘건’ 단위의 기록물이 없는 질의로 파악되었다. 따라서 파일럿 시스템은 결과 제시가 가능한 모든 질의에 대해 검색 결과를 제시했음을 의미한다.

〈표 4〉 질의어에 따른 검색 성능 평가 결과 개선방안 도입 전후 비교(건 기록물)

검색 시스템	검색 도구	결과 제시 질의 수	검색 제시 수 평균	찾은 정답 수 평균	정확률(Precision)		재현율(Recall)		Macro F-score
					Macro-Avg.	Micro-Avg.	Macro-Avg.	Micro-Avg.	
국가 기록 포털	통합 검색	45	23	7.34	37.54%	31.91%	45.17%	41.84%	41.00%
	상세 검색	37	18.52	5.4	33.42%	29.15%	40.02%	30.07%	36.42%
파일럿 시스템	통합 검색	49	25.22	8.08	37.83% (▲ 0.77 %)	32.03% (▲ 0.38 %)	52.68% (▲ 16.63 %)	46.97% (▲ 12.26 %)	44.03% (▲ 7.39 %)
	상세 검색	49	25.22	8.08	37.83% (▲ 13.20 %)	32.03% (▲ 9.88 %)	52.68% (▲ 31.63 %)	46.97% (▲ 56.20 %)	44.03% (▲ 20.90 %)

<표 4>를 보다 세부적으로 살펴보면, 모든 평가지표가 개선되긴 하였으나 정확률 측면에서는 다소 아쉬운 수준의 향상을 보여주었다. 기존 국가기록포털의 ‘통합검색’과 ‘상세검색’의 Macro 평균 정확률은 각각 37.54%, 33.42%이고, Micro 평균 정확률은 각각 31.91%, 29.15%로 평가되었다. 이에 비해 개선된 파일럿 시스템의 ‘통합검색’과 ‘상세검색’의 Macro 평균 정확률은 37.83%로, 기존 대비 각각 0.77%, 13.20%만큼 개선되었으며, Micro 평균정확률은 32.03%로, 역시 각각 0.38%, 9.88%로 개선되었다. 개선 폭을 살펴보면, ‘상세검색’의 정확률은 대략 10% 상당의 향상 폭을 보인 반면 그에 비해 ‘통합검색’의 상승효과는 매우 미미한 것으로 파악되었다.

이러한 원인은 파일럿 시스템의 검색 결과 문서 수가 국가기록포털보다 많기 때문이다. 정확률은 검색된

는 결과 문서 수가 클수록 줄어드는 경향이 있는데, 국가기록포털은 전체 50개 질의에 대해 결과를 제시한 질의 수가 통합검색은 45개, 상세검색은 37개의 결과를 제시했으며, 평균 문서 수는 각각 23개, 18.52개였다. 반면, 파일럿 시스템은 50개 질의 중 49개에 대해 결과를 제시하였으며, 평균 25.22개로 기존 포털보다 6.7개 더 많은 문서를 검색, 정답 후보로 제시하였다. 반면 파일럿 시스템의 상위 30개 결과 문서 중 찾은 정답 수 평균이 8.08개로 기존 포털 정답 수보다 많다. 이는 더 많은 검색 결과를 제시하면서도 동시에 이용자가 체감하는 검색 정확도는 유지되었음을 의미하며, 이 같은 결과는 높은 재현율 상승으로 이어졌다.

재현율 측면에서 파일럿 시스템은 정확률 대비 매우 유의미한 상승을 보여주었다. 국가기록포털의 ‘통합검색’과 ‘상세검색’의 Macro 평균 재현율은 각각 45.17%, 40.02%이고, Micro 평균 재현율은 각각 41.84%, 30.07%로 평가되었다. 이에 비해 파일럿 시스템의 Macro 평균 재현율은 52.68%로, ‘통합검색’ 대비 16.63%, ‘상세검색’ 대비 31.63% 대폭 개선되었다. 파일럿 시스템의 Micro 평균재현율 역시 46.97%로, 각각 12.26%와 56.20%의 큰 상승 폭을 보였다. 이러한 개선은 확장 불리언 검색 기능 도입으로 인해 기존 포털에서는 찾지 못했던 정답문서를 많이 획득하였고, 더불어 질의에 적합한 문헌을 상위로 제시하는 순위화 모델에서 기인한다. 또한 통합색인 방식의 적용으로 기록물 제목뿐만 아니라 다른 기록물 메타 정보 활용이 재현율 향상에 기여했다. 두 지표가 상승함에 따라 조화평균인 F-score 역시 향상되었다. 국가기록포털의 Macro F-score는 ‘통합검색’은 41.00%, ‘상세검색’은 36.42%로 평가된 것에 반해, 파일럿 시스템 검색 도구는 44.03%로, 각각 7.39%, 20.90%만큼 상승했다. 이렇듯 검색 품질과 관련된 모든 평가지표가 향상된 결과는 제시한 개선방안이 국가기록포털 검색 성능 개선에 실효가 있음을 방증한다.

<표 5> 질의어에 따른 MRR(건 기록물)

검색 시스템	검색 도구	결과 제시 질의 수	1등 정답 질의 수	최초정답문서 순위 평균	MRR	
					전체 질의	결과 제시 질의
국가 기록포털	통합검색	45	27	3.2	0.64	0.71
	상세검색	37	21	2.32	0.52	0.70
파일럿 시스템	통합검색	49	39	2.02 (▲ 1.18)	0.83 (▲ 28.13 %)	0.85 (▲ 19.72%)
	상세검색	49	39	2.02 (▲ 0.3)	0.83 (▲ 59.62 %)	0.85 (▲ 21.43 %)

<표 5>는 파일럿 시스템에 대한 검색효율성 지표인 MRR 측정 결과를 보여준다. 국가기록포털의 ‘통합검색’과 ‘상세검색’의 전체 50개 질의에 대한 MRR이 각각 0.64, 0.52로 집계되었다. 이 중 검색 결과를 제시한 질의만으로 평가 대상을 한정할 경우, 결과 제시 질의 수에 대한 MRR은 각각 0.71, 0.70로 측정되었다. 파일럿 시스템은 검색 효율성 측면에서도 큰 폭의 개선을 보였는데, 전체 질의에 대한 MRR이 0.83으로 ‘통합검색’ 대비 28.13%, ‘상세검색’ 대비 59.62%의 상승효과를 얻었다. 결과 제시 질의 수로 한정할 MRR은 0.85로 각각 19.72%, 21.43%만큼 상승했다. 두 경우 모두 유의미한 상승을 보였지만, 전체 질의에 대한 MRR의 상승 폭이 좀 더 높은 이유는 앞서 정확률 비교에서 설명한 바와 같이 국가기록포털이 정답문서를 제시한 질의 수가 파일럿 시스템보다 적었기 때문이다. 특히 정답문서가 1등으로 나온 질의 수와 최초 정답 문서 순위 평균 역시 기존 국가기록포털 대비 파일럿 시스템이 월등히 높아, 확장 불리언 검색 기능 도입으로 적합 문헌의 순위를 올린 것이 검색 효율성 측면에서도 유효함을 입증한다.

결론적으로, 색인어·질의어 토큰화 방식 통일, 통합색인 구축, 확장 불리언 검색 기능 도입 수준의 시스템 개선만으로도 검색 성능 면에서 의미 있는 상승효과를 보였다. 검색 성능 측면에서 정확률은 수치상으로는 큰 상승이 없었으나, 체감하는 정확도가 향상되었을 것으로 암시되며, 재현율은 최대 50% 이상의 증가율을 보였다. 검색 효율성 측면 역시 MRR의 향상으로 정답문서를 찾기 위해 뒷 페이지까지 확인해야 하는 상황이 줄어들 것으로 예측된다. 이렇듯 파일럿 시스템을 통해 검색 정확성 및 효율 측면 모두에서 유의미한 상승효과를 입증함으로써 제안된 개선방안의 실효성을 검증하였다.

### 4.3 연구 한계점 및 향후 연구 방향

한편, 파일럿 시스템에 다양한 개선방안을 적용했음에도 불구하고 최종 검색 성능은 여전히 60%를 밑도는 수준으로, 이는 다른 일반 포털 성능에 비해 낮은 편이다. 검색에 실패하거나 검색 정확도가 떨어진 질의 및 검색 결과를 세부적으로 분석한 결과, 다음과 같은 문제가 존재함을 파악하였다. 첫째, 숫자 및 특수문자나, 영문 질의에 대한 고려가 부족했다. 검색 성능 평가에서 '6.25 전쟁 사진'과 'WTO 분쟁'과 같은 질의어의 성능이 국가기록포털에 비해 파일럿 시스템에서 상대적으로 낮게 측정되었는데, 원인 분석 결과, 특수문자나 영문 대소문자에 대한 색인어 및 질의어 처리에 대한 기능이 구현되지 않았음에서 기인한 것으로 파악되었다.

둘째, 검색되는 문서의 양이 기존 포털대비 방대해졌다. 국가기록포털은 정답문서를 가지고 있음에도 동일되지 않은 토큰화 방식 및 복잡한 색인구조로 인해 검색되지 않는 문제점이 있었다. 이 문제를 해결하기 위해, 파일럿 시스템은 해당 질의와 관련된 기록물을 최대한 많이 찾는 방식으로 설계되었다. 이러한 상황에서 자연어 형식의 질의까지 수용하도록 확장됨에 따라 이번에는 역으로 핵심 키워드와 연관성이 낮은 기록물들도 함께 검색되는 문제가 제기되었다. 2차, 3차 검색 평가에서 평균적으로 검색된 기록물 수는 13,572건으로, 파일럿 시스템의 전체 문서 집합 중 약 10%에 해당한다. 따라서 검색 결과에 대한 적절한 컷오프(Cut-Off) 지점을 설정하여 검색 결과의 허수를 줄일 필요가 있다고 판단된다.

셋째, 일반 사용자 입장에서 편의성에 대한 평가가 이루어지지 않았다는 아쉬움이 있다. 키워드 하이라이팅, 정렬기능 개선, 기록물 상세정보 보강 등의 검색 결과 화면에 대한 일부 개선이 이루어졌지만, 본 연구가 검색 결과에 대한 성능 향상에 초점을 맞추어 진행되다 보니, 구현한 부가 기능이 실제 이용자에게 얼마만큼 편리성을 제공하였는지 측정하지 못했다. 향후, 기존 사용성 관점 평가에서 제안한 “이용자에게 익숙한 인터페이스와 검색기능의 제공, 도움말 지원 확대, 일관성 있는 웹 인터페이스 제공”(이효진, 김지현, 2021)과 같은 내용들을 보강 및 적용하여 이용자 측면에서 서비스적인 개선방안에 대해 검증할 필요가 있다.

한편, <그림 1>의 설명에서 언급한 바와 같이, 본 연구에서는 기록원 내부의 정책 재정립이 필요하거나 원문 전체를 입수해야 하는 등의 매우 높은 비용과 시간을 수반하는 개선방안은 적용하지 못했다는 한계가 있었다. 향후 연구 방향은 다음과 같다. 먼저, OCR 기술을 활용한 개선방안으로 전문 내용을 통해 색인데이터 자동 구축을 하고자한다. OCR 기술을 통해 추출된 전문을 바탕으로 TextRank(Mihalcea & Tarau, 2004)와 같은 알고리즘을 활용하여 자동으로 기록물 요약 및 태그를 생성한다. 또한 기록물 제목과 내용의 괴리가 심할 경우 GPT(Generative Pre-trained Transformer)를 활용하여 대체 제목을 생성한다. 상기 작업을 통해 비전자기록물의 검색 재현율 향상을 도모하고자 한다. 현재, 국가기록원에서 개발한 타자체의 OCR모델의 탐지율이 98.9%, 인식률이 95.3%이므로 원활하게 전문 내용을 추출이 가능할 것으로 전망된다(국가기록원, 2021a).

시소러스를 활용한 질의어 자동 확장을 위해서는 국가기록원이 기보유 중인 시소러스 사전 현행화와 관리도구 구축이 선행되어야 한다. 갱신된 시소러스를 활용하여 이용자가 입력한 검색어와 연관된 동의어, 관련어 단어목록을 직접 선택하는 수동 질의 확장 기능을 구현하고, 더 나아가 자동질의 확장을 통해 등가 관계를 가진 단어가 포함된 문서까지 검색 범위를 확장하는 방식을 제안한다. 이를 통해 정답문서를 포함하는 검색 결과가 늘어날 것이며 궁극적으로 재현율 개선에 효과가 있을 것으로 기대한다.

## 5. 결론

최근 들어 국가기록포털에서 직접 검색한 후 기록 원문을 신청하는 온라인 열람이 늘어나고 있다(설문원, 2022). 또한 4차 산업혁명 시대를 맞아 원천 데이터의 가치와 중요성이 높아져 전자기록물에 대한 사회적

요구가 증가하고 있다(노명환, 2020). 국가기록포털은 이러한 요구를 충족시키는 가장 대표 검색 도구임에도 불구하고, 그 성능과 편리성이 일반 이용자의 기대치에는 미치지 못하는 실정이었다. 본 연구는 국가기록포털 검색 품질 개선을 궁극의 목적으로, 선행된 품질 평가 과정에서 도출한 국가기록포털의 문제점들을 규명했다. 첫째, 검색 도구의 성능이 전반적으로 저조하여, 검색 정확도와 재현율이 낮아 이용자가 원하는 문서를 찾기 어려웠다. 둘째 검색 결과의 일관성이 결여되어, 동일한 질의에 대한 검색 도구 간 성능의 차이가 컸다. 셋째, 기본검색 기능이 제공되지 않아, 기록물의 내용이나 생산기관에 대한 검색이 제한적이었다. 넷째, 검색어 추천 기능이 부족하여, 일반 이용자의 정보 검색에 대한 지원이 미흡했다. 다섯째, 질의어 확장 기능이 없어, 이용자가 입력한 일반 어휘를 기록 관련 용어로 확장 해석하지 못했다. 마지막으로, 근본적으로 비전자기록물에 대한 접근이 어려워 기록물 원문에 대한 기계가독형 전처리 과정이 필요한 상황이었다.

상기한 문제점을 해결을 위해 다음과 같은 개선방안들을 제안했다. 첫째, 질의어·색인어 추출 방식을 통일함으로써 검색 결과의 일관성 문제를 해결한다. 둘째, 통합색인 방식을 적용하여 검색대상 기록물 건과 철 제목으로부터 일관된 방식으로 색인어를 추출하고, 통합색인 필드에 저장하여 검색 결과의 일관성을 부여해준다. 셋째, 확장 불리언 검색 기능을 도입하고 가중치에 따른 랭킹 알고리즘을 통해 가장 유사한 문서를 검색 해주는 기능을 기본검색으로 적용한다. 넷째, 검색어 추천기능을 도입하여 이용자 질의 로그를 활용해 인기 검색어를 분기별로 최신화하고, 연관 검색어를 제공한다. 다섯째, 시소러스 현행화 및 포털 검색 실적용을 통해 구축한 유의어 사전을 활용하여 수동 질의 확장을 지원하고, 더 나아가 질의어 자동 확장 기능을 구현한다. 여섯째, 스캔 문서에서 OCR 기술을 활용하여 색인 대상 텍스트를 추출하고, 이를 기계가독형 문서로 변환하고, 전문 내용이나 내용 요약에 대한 색인데이터로 생성하여 전문검색 지원하는 방안을 제안했다. 다만 본 연구에서는 제안된 개선안의 실효성 검증을 위해 현실적, 제도적 제약이 있는 방안을 제외한 나머지를 적용하여 파일럿 시스템을 구축하였다.

파일럿 시스템 평가는 선행된 품질 평가와 같은 방식으로 진행되었고, 평가 결과는 역시 기존의 국가기록포털에서 측정된 정확률, 재현율, MRR 등을 평가지표로 산출하여 개선안 적용 전, 후의 성능을 비교하였다. 국가기록포털 대비 파일럿 시스템은 건 단위 검색 결과 제시가 가능한 모든 질의에 대해 정답문서를 제시하였으며, 최초 정답문서제시 순위 역시 상승했다. 평가지표의 경우, 파일럿 시스템은 정확률 측면에서는 미미한 향상을 보였으나, 재현율과 MRR에서 높은 상승 폭을 보였다. 이러한 평가 결과를 통해 본 연구에서 파일럿 시스템에 적용한 개선방안들이 품질개선에 유효함을 입증하였다.

한편, 본 원고 투고 시점인 2023년 4월 기준 국가기록포털에 디자인 변화와 전체적인 UI 개선이 있었다. 분석 결과, 파일럿 시스템에서 검증한 개선방안 중 하나인 확장 불리언 검색 기능이 ‘상세검색’에 적용된 것으로 확인되었으며, 한적검색 역시 원활하게 기능하는 것도 확인되었다. 이 밖에도 추가로 공개된 2023년 국가기록원 업무계획에 따르면, “통합색인 구축, 개인별 검색이력 관리, 인기검색어 추천” 등의 개선이 예정되어 있다(국가기록원, 2023). 향후 연구방향으로는 국가기록원의 이러한 노력에 부흥하는 차원에서 이번 연구를 통해 도출되었으나 아직 적용되지 못한 개선방안, 예컨대 OCR기술을 활용한 전문검색 지원이나 시소러스 현행화 등을 실제 포털에 적용하기 위한 실무 협의 및 후속 연구가 필요하다.

본 연구는 국가기록포털의 문제점을 구체적인 검색 성능 수치를 통해 규명하여 그에 따른 개선안을 도출하고, 실제 시스템을 구축함으로써 기존 시스템과 비교를 통해 그 효과를 검증했다는 점에서 의의가 있다. 이러한 일련의 과정 및 결과가 국가기록원과 같은 중앙정부 관리기관뿐 아니라 지방 영구기록물 관리기관이나 준현용기록을 관리하는 기록관 단의 검색 시스템 개선을 도모하는 연구의 밑거름이 되길 기대한다.

## 참고문헌

- 강례림 (2020). 국가기록원 기록정보서비스 개선방안에 관한 연구. 석사학위논문, 이화여자대학교 문헌정보학과.
- 강운아, 조영준, 김민정, 오효정 (2022). 국가기록물 접근성 및 활용성 향상을 위한 국가기록포털 연계 개선방안. 정보관리학회지, 39(4), 99-125. <https://doi.org/10.3743/KOSIM.2022.39.4.099>
- 과학기술정보통신부 (2022). 2021 인터넷이용실태조사.
- 국가기록원 (2015). 2014년 국가기록백서. 대전: 국가기록원.
- 국가기록원 (2018). 2018년 소장기록물 정리(整理)·기술(記述) 추진 계획. 대전: 국가기록원.
- 국가기록원 (2021a). 디지털화 기록의 문자인식(OCR) 기술 적용 및 활용방안. 대전: 국가기록원.
- 국가기록원 (2021b). 2021년 국가기록원 주요통계연보. 대전: 국가기록원.
- 국가기록원 (2022). 2022년 주요 업무 추진계획. 대전: 국가기록원.
- 국가기록원 (2023). 2023년 주요 업무 추진계획. 대전: 국가기록원.
- 국가기록원 [발행년불명]. 국가기록원 목표 및 업무. 출처: <https://www.archives.go.kr/next/neworgan/vision.do>
- 김지현 (2012). 국가기록원 기록정보서비스에 대한 이용자 인식에 관한 연구. 한국기록관리학회지, 12(1), 167-187. <https://doi.org/10.14404/JKSARM.2012.12.1.167>
- 김철진 (2018). 알렉사 기반의 음성서비스와 RESTfull API 연동을 위한 아키텍처 및 개발 프로세스. 한국지식정보기술학회 논문지, 13(3), 341-350. <https://doi.org/10.34163/jkits.2018.13.3.005>
- 나정호, 소현기, 영영록, 이정옥, 오효정 (2022). 국가기록포털 검색서비스 품질 점검을 위한 평가셋 구축 및 현황 분석. 한국기록관리학회지, 22(4), 25-43. <https://doi.org/10.14404/JKSARM.2022.22.4.025>
- 노명환 (2020). 4차 산업혁명 시대 데이터 아카이브와 기록관리의 길: 국가기록원의 행정정보 데이터세트 기록관리 실행방안에 대한 비판적 검토와 중·장기 차원의 제안. 기록과 정보·문화 연구, 11, 7-43. <https://doi.org/10.23035/kaics.2020.1.11.007>
- 박민수, 현미환 (2011). 웹 사용성 개선에 관한 연구 - 국가과학기술정보시스템을 중심으로 -. 한국비블리아학회지, 22(4), 5-19. <https://doi.org/10.14699/kbiblia.2011.22.4.005>
- 박상현, 김희섭 (2014). EAD 기반 영구기록물 통합검색 시스템 설계 및 구현. 한국기록관리학회지, 14(1), 101-124. <https://doi.org/10.14404/JKSARM.2014.14.1.101>
- 박순호, 김영길 (2019). 오픈 소스를 활용한 공공 데이터 기반의 질병 검색 시스템 구현. 한국정보통신학회논문지, 23(11), 1,337-1,342. <https://doi.org/10.6109/jkiice.2019.23.11.1337>
- 백지연, 오효정 (2019). 국가기록원 질의로그 빅데이터 기반 이용자 정보요구 유형 분석. 정보관리학회지, 36(4), 183-205. <https://doi.org/10.3743/KOSIM.2019.36.4.183>
- 설문원 (2022). 국가기록원의 기록제공서비스에 관한 탐색적 연구. 한국기록관리학회지, 22(3), 103-124. <https://doi.org/10.14404/JKSARM.2022.22.3.103>
- 송지성, 정다희 (2021). 인공지능 기반 네이버 앱 검색 서비스 사용성 연구. 한국디자인문화학회지, 27(1), 197-207. <https://doi.org/10.18208/ksdc.2020.27.1.197>
- 안세진, 황현호, 임진희 (2022). 종이기록 데이터화를 위한 AI-OCR 적용 사례연구. 정보관리학회지, 39(3), 165-193. <https://doi.org/10.3743/KOSIM.2022.39.3.165>
- 이병길, 김희섭 (2013). 새마을운동 기록물의 개체기반 온톨로지 검색시스템 설계 및 평가. 한국기록관리학회지, 13(3), 67-97. <https://doi.org/10.14404/JKSARM.2013.13.3.067>
- 이유빈, 이해영 (2017). 온톨로지 기반의 기록물 검색 시스템을 위한 인터페이스 제안. 한국기록관리학회지, 17(1), 217-244. <https://doi.org/10.14404/JKSARM.2017.17.1.217>
- 이윤령, 이해영 (2014). 온라인 검색도구의 개선 방안에 관한 연구 : 국립기록관 검색도구에 대한 평가를 중심으로. 한국기록관리학회지, 14(1), 75-100. <https://doi.org/10.14404/JKSARM.2014.14.1.075>
- 이종원, 조우승, 김태현 (2022). KoBERT 기반 대화형 검색 시스템 설계 및 구현. 한국지식정보기술학회 논문지, 17(5),



- 1,081-1,088. <https://doi.org/10.34163/jkits.2022.17.5.028>
- 이효진, 김지현 (2021). 국가기록원 웹사이트 검색서비스의 사용성 평가를 통한 개선방안 연구. *한국기록관리학회지*, 21(3), 187-215. <https://doi.org/10.14404/JKSARM.2021.21.3.187>
- 장희정 (2012). 국가기록관 웹사이트의 평가에 관한 연구. *한국기록관리학회지*, 12(2), 51-70. <https://doi.org/10.14404/JKSARM.2012.12.2.051>
- 진주영, 이해영 (2018). 국가기록원 웹사이트 유입경로와 이용자 검색어 분석. *정보관리학회지*, 35(1), 183-203. <https://doi.org/10.3743/KOSIM.2018.35.1.183>
- 최선희 (2008). 국가기록포털 이용 분석을 통한 기록정보서비스 개선 방안 연구. 석사학위논문, 연세대학교 대학원 문헌정보학과.
- Mihalcea, R. & Tarau, P. (2004). TextRank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, 404-411.
- Voorhees, E. M. (1999). TREC-8 Question Answering Track Report. *Proceedings of the 8th Text Retrieval Conference*, 77-82.

• 국문 참고자료의 영어 표기

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

- Ahn, Se-Jin, Hwang, Hyeon-Ho & Lim, Jin-Hee (2022). A Case Study on the Application of AI-OCR for Data Transformation of Paper Records. *Journal of the Korean Society for Information Management*, 39(3), 165-193. <https://doi.org/10.3743/KOSIM.2022.39.3.165>
- Baek, Ji-Yeon & Oh, Hyo-Jung (2019). User Information Needs Analysis based on Query Log Big Data of the National Archives of Korea. *Journal of the Korean Society for Information Management*, 36(4), 183-205. <https://doi.org/10.3743/KOSIM.2019.36.4.183>
- Choi, Seon-Hee (2008). Improving the archival reference services through a use analysis of the national archives portal service. Master's thesis, Graduate School of Yonsei University, Library and Information Science.
- Jang, Hee-Jung (2012). A Study on Evaluation of National Archives Websites. *Journal of Korean Society of Archives and Records Management*, 12(2), 51-70. <https://doi.org/10.14404/JKSARM.2012.12.2.051>
- Jin, Joo-Young & Rieh, Hae-Young (2018). Analysis of Users' Inflow Route and Search Terms of the Korea National Archives' Web Site. *Journal of the Korean Society for Information Management*, 35(1), 183-203. <https://doi.org/10.3743/KOSIM.2018.35.1.183>
- Kang, Rye-Rim (2020) A Study on the Improvement of Archival Reference Service in the National Archives of Korea. Master's thesis, Graduate School of Ewha Womans University, Library and Information Science.
- Kang, Yoo-na, Jo, Young-jun, Kim, Min-Jung & Oh, Hyo-Jung (2022). Advancement Plans for Linkage of National Archives Portal Service to Improve Accessibility and Usability of National Records. *Journal of the Korean Society for Information Management*, 39(4), 99-125. <https://doi.org/10.3743/KOSIM.2022.39.4.099>
- Kim, Chul-Jin (2018). An Architecture and Development Process for Integration of Alexa-based Voice Service and RESTful API. *Journal of Knowledge Information Technology and Systems*, 13(3), 341-350. <https://doi.org/10.34163/jkits.2018.13.3.005>
- Kim, Ji-Hyun (2012). A Study on Users' Perception of Reference Services in National Archives of Korea. *Journal of Korean Society of Archives and Records Management*, 12(1), 167-187. <https://doi.org/10.14404/JKSARM.2012.12.1.167>

- Lee, Byung Gil & Kim, Heesop (2013). Design and Evaluation of an Individual Instance-based Ontology Retrieval System for Archival Records of the "Saemaul Movement". *Journal of Korean Society of Archives and Records Management*, 13(3), 67-97. <https://doi.org/10.14404/JKSARM.2013.13.3.067>
- Lee, Hyo-Jin & Kim, Ji-Hyeon (2021). A Study on Improvement Plans for National Archives of Korea Website's Search Service through Its Usability Evaluation. *Journal of Korean Society of Archives and Records Management*, 21(3), 187-215. <https://doi.org/10.14404/JKSARM.2021.21.3.187>
- Lee, Jong-Won, Jo, Woo-Seung & Kim, Tae-Hyun (2022). Design and Implementation of an Interactive Search System Based on KoBERT. *Journal of Knowledge Information Technology and Systems*, 17(5), 1,081-1,088. <https://doi.org/10.34163/jkits.2022.17.5.028>
- Lee, Yoon-Ryeong & Rieh, Hae-Young (2014). A Study on the Improvement Direction for Online Finding Aids: Based on the Assessment of National Archives. *Journal of Korean Society of Archives and Records Management*, 14(1), 75-100. <https://doi.org/10.14404/JKSARM.2014.14.1.075>
- Lee, Yu-Been & Rieh, Hae-Young (2017). A Suggestion of Interface for Ontology-Based Record Retrieval System. *Journal of Korean Society of Archives and Records Management*, 17(1), 217-244. <https://doi.org/10.14404/JKSARM.2017.17.1.217>
- Ministry of Science and ICT (2022). Survey on Internet use in 2021.
- Na, Jeong-Ho, So, Hyeon-Gi, Yeom, Gyung-Rok, Lee, Jung-Ok & Oh, Hyo-Jung (2022). Test Set Construction for Quality Evaluation of NAK Portal's Search Service and the Status Analysis. *Journal of Korean Society of Archives and Records Management*, 22(4), 25-43. <https://doi.org/10.14404/JKSARM.2022.22.4.025>
- Noh, Meung-Hoan (2020). The Way for Data Archive and Records/Archive Management in the 4th Industrial Revolution Era : Critical Reviews and Mid- and Long-term Proposals for the National Archives' Administrative Information Dataset Records/Archive Management Implementation Plan. *The Korean Journal of Archival, Information and Cultural Studies*, 11, 7-43. <https://doi.org/10.23035/kaics.2020.1.11.007>
- NAK (2015). White Paper on National Records, 2014. Daejeon: National Archives of Korea.
- NAK (2018). Plan to promote the organization of collection records in 2018. Daejeon: National Archives of Korea.
- NAK (2021a). Application and utilization of character recognition (OCR) technology of digitized records. Daejeon: National Archives of Korea.
- NAK (2021b). National Archives Major Statistical Yearbook 2021. Daejeon: National Archives of Korea.
- NAK (2022). Major business initiatives for 2022. Daejeon: National Archives of Korea.
- NAK (2023). Major business initiatives for 2023. Daejeon: National Archives of Korea.
- Park, Min-Su & Hyun, Mi-Hwan (2011). Usability of the National Science and Technology Information System. *Journal of the Korean BIBLIA Society for Library and Information Science*, 22(4), 5-19. <https://doi.org/10.14699/KBIBLIA.2011.22.4.005>
- Park, Sang-Hyun & Kim, Heesop (2014). Design and Implementation of an EAD-based Integrated Retrieval System for the Archives in Korea. *Journal of Korean Society of Archives and Records Management*, 14(1), 101-124. <https://doi.org/10.14404/JKSARM.2014.14.1.101>
- Park, Sun-Ho & Kim, Young-Kil (2019). Implementation of Disease Search System Based on Public Data using Open Source. *Journal of the Korea Institute of Information and Communication Engineering*, 23(11), 1,337-1,342. <https://doi.org/10.6109/jkiice.2019.23.11.1337>
- Seol, Moon-Won (2022). An Exploratory Investigation of Archival Reference Services in the National

Archives of Korea. Journal of Korean Society of Archives and Records Management, 22(3), 103-124.  
<https://doi.org/10.14404/JKSARM.2022.22.3.103>

Song, Ji-Sung & Jeong, Da-Hee (2021). A Study on the Usability of AI-based Naver App Search Service.  
Journal of the Korean Society of Design Culture, 27(1), 197-207.  
<https://doi.org/10.18208/ksdc.2020.27.1.197>