

국가기록포털 검색서비스 품질 점검을 위한 평가셋 구축 및 현황 분석

Test Set Construction for Quality Evaluation of NAK Portal's Search Service and the Status Analysis

나정호(Jeong Ho Na)¹, 소현기(Hyeon-Gi So)², 엄경록(Gyung Rok Yeom)³,
이정옥(Jung-Ok Lee)⁴, 오효정(Hyo-Jung Oh)⁵

E-mail: jhna2012@naver.com, 95dizyso@gmail.com, yumgr@iwaz.co.kr, jungok17@korea.kr,
ohj@jbnu.ac.kr



- 1 제1저자 전북대학교 일반대학원 기록관리학과 석사과정
- 2 전북대학교 일반대학원 기록관리학과 석사과정
- 3 ㈜ 아이와즈, 솔루션개발팀장
- 4 국가기록원 서비스정책과
- 5 교신저자 전북대학교 문헌정보학과 교수, 문화융복합아카이빙연구소 공동연구원

논문접수 2022.10.17
최초심사 2022.10.20
게재확정 2022.11.08

ORCID

Jeong Ho Na
https://orcid.org/0000-0001-5118-3239
Hyeon Gi So
https://orcid.org/0000-0001-9450-8806
Yeom Gyung Rok
https://orcid.org/0000-0001-5319-613X
Jung-Ok Lee
https://orcid.org/0000-0002-7894-8761
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.

초 록

기록을 관리하는 궁극적인 목적은 안전한 보존과 효율적 활용이다. 그러나 공공기록물 관리의 수장인 국가기록원의 검색서비스인 국가기록포털은 검색 시스템이 도입된 지 10년이 넘어 노후화되었고, 검색도구도 이원화되어 있다. 또한, 국가기록물을 열람하고자 하는 이용자의 검색 만족도를 충족시키지 못해 개선이 필요하다. 본 연구는 국가기록물 검색 고도화를 위한 사전 연구로, 현재의 국가기록원 검색시스템의 검색 품질 현황을 파악하는 것을 목적으로 한다. 이를 위해 국가기록원에서 사용 중인 검색도구인 CAMS와 국가기록포털의 현황을 분석하고 검색서비스 품질을 점검하기 위한 평가셋을 구축, 이용자 관점에서 국가기록포털의 검색 품질 평가를 실시하였다. 국가기록포털의 성능평가를 위한 척도로는 정확률과 재현율, F-score, MRR 등을 활용, 다양한 각도에서 분석하였다. 검색 품질 평가결과 전반적으로 검색 성능이 낮은 것으로 나타났으며, 특히 '주제별상세검색'의 경우 정확률, 재현율, MRR이 모두 낮게 나와 개선이 시급할 것으로 보인다. 본 연구를 통해 구축된 평가셋이 향후 국가기록원 검색시스템의 고도화 이후 시스템의 성능이 얼마나 개선되었는지를 객관적으로 측정할 수 있는 기틀로 활용되기를 기대한다.

ABSTRACT

The ultimate record management's purpose is preservation and utilization. However, the National Archives of Korea (NAK)s Portal has problems such as search system aging and search tools dualization. As a result, the users' search satisfaction is not satisfied, and the improvement demand increases. This study aimed to evaluate the NAK's search quality as a preliminary study for NAK search system advancement. To this end, we analyzed the current status of CAMS and NAK's Portal. Then, we established the test sets and evaluated the NAK's Portal quality from the user's point of view. Evaluation results were analyzed using Precision, Recall, F-score, and MRR. The analysis results showed that the overall search performance was low, particularly in the "advanced subject search," which showed low performance in Precision, Recall, and MRR. Thus, improvement is urgently needed. The test sets established for this study are expected to be used as a basis for objectively measuring the improvement of the search performance after the NAK search system advancement.

Keywords: 국가기록포털, 정보검색, 검색도구, 검색평가, 평가셋

National Archives of Korea (NAK) Portal, Information Retrieval, Finding Aids, Retrieval Evaluation, Test Set

•본 논문은 2022년 국가기록관리 활용기술 연구개발(R&D) 사업의 연구비를 지원받아 수행되었음.

1. 서론

1.1 연구목적

기록을 생산하고 수집하고 관리하는 궁극적인 목적은 기록의 안전한 보존과 효율적 활용을 위한 것으로, 이는 「공공기록물 관리에 관한 법률」 제1조에 명시되어 있다. 또한, 동법 제38조의2(영구기록물관리기관 보존기록물의 활용)에서 “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 정하고 있다. 「공공기관의 정보공개에 관한 법률」 제6조(공공기관의 의무)에서도 “공공기관은 정보의 적절한 보존 및 신속한 검색과 국민에게 유용한 정보의 분석 및 공개 등이 이루어지도록 정보관리체계를 정비하고, 정보공개 업무를 주관하는 부서 및 담당하는 인력을 적정하게 두어야 하며, 정보통신망을 활용한 정보공개시스템 등을 구축하도록 노력하여야 한다”고 정하고 있다. 이처럼 공공기관은 기록물을 체계적으로 관리할 뿐만 아니라 기록의 활용과 기록정보서비스 제공을 위해 힘써야 함을 확인할 수 있다.

국가기록원은 공공기록물의 효율적인 수집·보존·평가 및 활용에 관한 업무를 총괄하고 관장하는 우리나라 대표 영구기록물관리기관이다. 현재 국가기록원은 성남 나라기록관을 포함한 서울, 대전, 부산, 광주 5개소에 각각 기록정보센터를 운영하고 있으며 이 기록정보센터들을 활용해 기록의 열람과 제공을 추진하고 있다. 이 중 나라기록관의 국가기록정보센터가 기록정보센터들의 기록열람서비스를 총괄하고 있어 일반 이용자들은 중앙기록물관리시스템(Central Archives Management System, 이하 CAMS)과 연계된 국가기록포털(<https://www.archives.go.kr/>)을 통해 통합검색서비스를 이용할 수 있다(설문원, 2022). 그러나 2013년 국가기록포털 검색시스템이 통합 개편된 이후(국가기록원, 2014) 개선이 되지 않아 노후화되면서 여러 가지 문제점이 발생하고 있다. 이에 더해 국가기록포털 이용자들의 검색 만족도 역시 충족시키지 못하고 있다는 의견이 수렴되고 있다(장희정, 2012; 이효진, 김지현 2021).

특히 4차 산업혁명 시대를 맞아 원천 데이터의 가치와 중요성이 오르고 있어 원천 데이터로 사용할 수 있는 전자기록물에 대한 사회적 요구와 중요성이 증가하고 있다(노명환, 2020). 또한, 경제적인 측면에서도 국가기록원이 보유한 전자기록물의 생산 및 이관, 관리 비용 절감을 위해 효율적인 색인 설계 및 자료 검색이 필요하다. 이러한 사회적 요구에 대응하여 신속하고 정확한 기록정보서비스를 제공하기 위해서는 국가기록원의 검색시스템에 대한 개선이 무엇보다도 시급하며, 이를 위해서는 현재의 기록검색 서비스 품질을 점검하여 현황을 파악해야 한다.

본 논문은 상기한 요구, 즉 ‘국가기록물 검색 고도화’라는 궁극의 목표를 달성하기 위한 사전 연구로, 현재 국가기록원에서 일반 이용자들을 대상으로 제공하고 있는 국가기록포털의 검색 품질을 측정하고자 한다. 이를 위해 먼저 국가기록물 검색시스템의 현황을 분석하고 각 도구별 검색 결과를 평가하고자 한다. 특히, 검색 품질은 실제로 기록물 검색시스템을 이용하는 이용자 관점에 초점을 맞춰 이용자가 원하는 기록을 얼마나 잘 찾았는지를 나타내는 정확성과 효율성을 측정하는 평가셋에 기반해 연구를 진행하였다. 연구방법은 먼저 2장에서 국가기록원이 운영하고 있는 검색시스템인 CAMS와 국가기록포털의 현황 분석을 통해 각 시스템별 특징 및 제공하는 검색도구를 비교하여 차이점을 규명하고 문제점을 파악하였다. 3장에서는 시스템 이용자 관점에서 시스템을 평가하기 위한 평가셋의 구성요소와 실제 구축 과정을 기술하고, 4장에서 다양한 검색 품질 평가척도에 의한 측정 결과를 제시한다. 마지막으로 평가셋 구축 과정 및 품질 평가 과정을 통해 도출된 이슈와 개선방안을 논의하는 것으로 결론을 맺는다.

1.2 선행 연구

검색 품질 측정과 관련된 연구는 검색 속도나 정확률, 재현율 등 시스템의 검색 성능을 평가셋에 기반하여 비교·분석하거나, 이용자 집단에게 검색 테스트를 수행시키고 설문조사나 인터뷰 등을 통해 사용성과 만족도를 평가하고

있었다. 검색시스템의 정확도와 만족도를 평가한 연구를 살펴보면 다음과 같다. 김현희(2007)는 FRBR 모형 기반 서지검색시스템의 검색 효율성을 평가하기 위해서 두 개의 실험 시스템을 구축하였고, 이 두 시스템을 재현율, 정확률 및 검색 시간을 측정 기준으로 하여 비교하였다. 박소연, 이준호(2008)는 국내 주요 검색 포털들의 통합검색 시스템을 중심으로 이용자들의 입력한 질의어와 실제 조회한 문서를 기록한 클릭 로그를 참고하여 평가셋을 구축, 검색 결과의 적합도 측면에서 분석하였다. 박민수와 현미환(2011)은 국가과학기술정보시스템을 대상으로 이용자 사용성 평가를 통해 시스템 개선사항을 도출하였다. 사용성 평가는 16명의 이용자를 대상으로 태스크 수행과 사후 인터뷰를 통해 진행하였다. 장인호(2011)는 온톨로지 기반 법률 검색시스템을 구축하고 기존의 키워드 기반 법률 검색시스템과 비교하여 검색 소요 시간과 검색 적합성을 중심으로 검색 성능을 평가하였다.

사용성 관점에서 기록관 시스템의 만족도에 초점을 맞춘 연구는 다음과 같다. 박민수, 서은경(2012)은 표준 기록관리시스템(Records Management system, 이하 RMS)에 대한 전반적인 진단이 필요하다고 보고 기록관리시스템을 사용하는 기록관 전문요원을 대상으로 기능성과 인터페이스에 대한 만족도를 평가하였다. 이경남(2013)은 공공기관에서 RMS를 사용하고 있는 기록관 전문요원을 대상으로 설문조사를 실시해 검색 기능 활용 현황 분석 및 기능 평가를 수행하였다. 이윤령과 이해영(2014)은 국가기록원을 중심으로 호주, 영국, 미국의 국립기록관에서 제공하는 온라인 검색도구 현황을 분석하고, 실제 이용자들을 선정하여 포커스 그룹 인터뷰를 통한 사용성 평가를 수행해 개선점을 도출하였다. 이효진과 김지현(2021)은 전문가 평가와 다양한 그룹의 이용자를 대상으로 사용성 평가를 수행해 ISO 9241-11:2018에서 제시하는 효과성, 효율성, 만족도를 중심으로 국가기록포털 검색시스템을 분석하고 개선방안을 제시하였다.

본 연구는 국가기록포털의 검색 서비스 품질을 측정하기 위해 규격화된 평가셋을 구축하고, 다양한 검색 평가지표에 의한 평가결과를 수치화한 최초의 연구이다. 특히, 일반 이용자의 특성을 반영한 질의셋을 구성하고, 이용자 관점에서 검색 결과로 실제 제시되는 포털 페이지를 기반으로 검색 성능을 측정하였다는 점에 연구 의의를 둔다. 본 연구를 통해 구축된 평가셋을 활용해 향후 다양한 방법론이 적용된 검색시스템의 고도화 이후 얼마나 개선되었는지를 객관적으로 측정할 수 있는 기틀로 활용되길 고대한다.

2. 국가기록물 검색도구 현황

국가기록물 검색시스템의 검색 품질 평가를 위해서는 먼저 현재 국가기록원 검색체계가 어떠한지 진단할 필요가 있다. 본 장에서는 국가기록원에서 운영되고 있는 검색시스템인 CAMS와 국가기록포털을 중심으로 연구진이 직접 기록원에 방문, 실무진과 함께 시스템별 특징 및 검색도구 현황을 분석하였다.

2.1 국가기록물 유형

현재 국가기록포털에서는 기록물 유형을 <표 1>과 같이 기록물의 형태와 내부 기준에 의한 구분으로 구별하고 있다. 기록물 형태는 말 그대로 기록물의 형태에 따라 구별되며, 다음과 같이 1) 일반문서, 2) 사진·필름류, 3) 녹음·동영상류, 4) 도면, 5) 카드류, 6) 국무회의록, 7) 정부간행물, 8) 일반도서, 9) 총독부간행물 나뉜다. 기록물 구분은 기록물의 출처나 내용에 따라 의미 있는 단위로 구별해 놓은 것으로, 1) 일반기록물, 2) 역사기록, 3) 총독부기록, 4) 해외기록, 5) 민간기록, 6) 시청각기록, 7) 간행물, 8) 일반도서, 9) 행정박물로 총 9가지로 분류되어 있다. 국가기록원에서 제공하고 있는 검색서비스 역시 이러한 구분에 의한 유형별 검색도구를 ‘상세검색’ 기능을 통해 제공하고 있으나, 일반 이용자가 이를 활용하기에는 너무 복잡하고 세부 구분 기준에 대한 설명이 부족해 잘 활용되고 있지 않다. 기록물 유형별 검색도구의 품질을 평가하기 위해 평가셋 구축 시 <표 1>의 기준에 의한 기록물 유형별 질의어

를 후보로 선정하였다.

<표 1> 국가기록물 유형 정의 기준

구분	기록물유형	구분	기록물유형
기록물형태	일반문서	기록물구분	일반기록물
	사진·필름류		역사기록
	녹음·동영상류		총독부기록
	도면		해외기록
	카드류		민간기록
	국무회의록		시청각기록
	정부간행물		간행물
	일반도서		일반도서
총독부간행물	행정박물		

2.2 국가기록물 내부 검색도구: CAMS

CAMS는 중앙기록물관리기관인 국가기록원에서 영구기록물 관리를 전자적으로 수행하는 시스템으로 기록물 유형과 관계없이 전자·비전자 기록물에 대한 통합적인 업무수행과 안전한 보존·관리, 기록정보의 대국민 서비스를 지원하기 위한 국가기록물 관리시스템이다(국가기록원, 2015). CAMS는 기술과 기술분류, 시소러스, 생산기관 변경정보 등 보존기록의 활용을 위한 영구기록물의 기능을 반영하고 있어 CAMS를 이용하는 업무 담당자와 CAMS와 연계된 국가기록포털을 이용하는 이용자들은 여러 가지 방법으로 국가기록원에서 제공하는 기록물을 검색할 수 있다(김장환, 이은별, 2015). CAMS에서 제공되는 검색도구는 <그림 1>과 같이 ‘기록물철상세검색’, ‘기록물건상세검색’, ‘신통합검색’으로 구분되며, 이 중 ‘기록물철/건상세검색’은 CAMS의 기록물 철/건 목록 DB를 탐색하는 방식으로 검색엔진을 사용하지 않는다. 반면 ‘신통합검색’은 A사의 검색엔진을 도입한 검색 기능으로, 세부적으로 ‘철/건통합검색’, ‘철검색’, ‘건검색’으로 나뉜다.



<그림 1> CAMS 검색도구

2.3 국가기록물 외부 검색도구: 국가기록포털



<그림 2> 국가기록포털 검색도구

국가기록포털의 검색도구는 <그림 2>와 같이 ‘통합검색’, ‘상세검색’, ‘원문검색’, ‘기술계층별검색’, ‘국정분야주제별검색’이 있으며 B사의 검색 엔진을 사용 중인 것으로 파악되었다. 이중 키워드 즉 질의어 기반 검색을 제공하는 ‘통합검색’, ‘상세검색’, ‘원문검색’, ‘기술계층별검색’과 ‘국정분야주제별검색’의 세부항목에 있는 ‘주제별상세검색’을 대상으로 현황 분석을 진행하였다.

‘통합검색’, ‘상세검색’, ‘원문검색’은 국가기록원이 보유하고 있는 공공기관이 생산한 30년 이상의 보존 기간을 가지는 중요기록물과 해외에서 수집한 한국 관련 기록물, 민간 소장자에게서 수집하거나 기증받은 기록물, 정부간행물 등 공개 기록물을 대상으로 한 검색서비스로 국가기록원 DB의 기록물에 관해서 검색할 수 있다(박상현, 김희섭, 2014). ‘기술계층별검색’은 기록물의 생산맥락에 근거하여 기록물의 집합을 계층으로 나누어 디렉토리 형태로 제공함으로써 전체적으로 기록물의 구조와 내용을 조망해 볼 수 있는 검색이다. 마지막으로 ‘주제별상세검색’은 주요 정책·제도, 국책사업 등 국정분야별로 주제를 선정·구조화하고, 주제별 주요 내용과 관련 기록물 및 목록을 연계하여 제공하고 있다.

2.4 문제점

국가기록원에서 운영하는 두 검색시스템의 문제점은 다음과 같다. 첫째, CAMS와 국가기록포털에서 사용하는 검색엔진 및 서비스 방식이 다르다. CAMS는 A사의 검색엔진을 사용하고 있으며, 국가기록포털은 B사의 검색엔진을 사용하고 있어 지원되는 검색 기능과 활용하는 서비스 방식에서 차이를 가진다. 그러나 사용하는 검색엔진이 다르더라도 서비스 방식을 통일시키는 것으로 상이한 검색 결과가 나타나는 문제를 최소화할 수 있을 것으로 보인다.

둘째, 국가기록포털의 검색서비스는 2013년에 개편된 후 거의 10년이 되었으나 유지보수만 계속되고 있어(국가기록원, 2014), 질의 오류에 대한 적극적인 대처가 부족하며 지원되는 검색 편의 기능이 적어 검색 결과에 대한 이용자 만족도가 낮은 상황이다. 특히 10년이란 기간 내 지속된 유지보수 상황이 그때그때 돌출된 문제 해결을 위한 임시방편적인 대응만 있었을 뿐, 그에 대한 총괄적인 수정 내역이나 과정에 대한 이해가 매우 낮은 상황이다.

셋째, 검색 대상이 되는 CAMS 내 원천 데이터의 품질이 낮다. ‘2021년 국가기록원 주요통계연보’에 따르면 2021년 12월 31일 기준 국가기록원의 CAMS는 114,777,646건의 자료를 소장하고 있다. 그러나 중요 필드가 비어있거나 오류 값이 포함되는 등, 데이터의 품질 미비로 인해 국가기록원 검색시스템으로 기록물을 검색할 때 일부 기록물만 검색되거나 검색이 아예 되지 않을 뿐 아니라, 검색 결과 역시 오류를 내포하고 있는 경우가 다수

있다.

마지막으로 검색 결과에 대한 정확도 및 재현율 등 검색서비스 자체에 대한 품질이 낮다. 국가기록원 검색시스템은 복잡한 검색도구로 인해 검색 속도저하 및 서버 부하 문제가 발생하며, 검색도구 간의 연계성이 떨어져 원하는 정보를 찾기 위해 많은 노력과 시간이 소모되고, 소장기록물에 대한 접근성 및 활용성이 떨어진다. 이처럼 국가기록원의 검색시스템은 이용자가 원하는 기록을 찾기 어려운 구조로 되어있어 이용자 서비스에 불편을 초래하며(김지현, 2012), 기록관리 업무에도 많은 비용을 발생시키고 있다. 본 연구는 이러한 문제점을 내포하고 있는 국가기록포털의 현 상태의 품질을 객관적으로 평가하고 수치화하는데 주목적이 있다.

3. 국가기록포털 품질 점검을 위한 평가셋 구축

3.1 평가셋 구축 설계

검색서비스의 품질, 특히 검색 성능을 평가하기 위해서는 이용자의 요구가 반영된 평가셋 구축이 선행되어야 한다. 본 장에서는 일반 이용자 관점에서 현행 국가기록물 검색도구의 검색 성능에 대한 정량적·정성적 평가를 할 수 있는 평가셋 구축 과정에 대해 기술하고자 한다. 검색 품질 점검을 위한 평가셋은 문서집합, 질의셋, 정답셋으로 구성된다. 이중 문서집합은 2021년 12월 기준 국가기록원에서 관리하고 있는 기록물 114,777,646건(국가기록원, 2021)을 대상으로 삼는다.

질의셋은 이용자들의 정보요구를 표현한 것이다. 박미성(2009)은 질의셋을 구축할 때의 고려사항으로 첫째, 질의어는 반드시 문서집합의 주제를 포함해야 한다. 둘째, 질의어는 실제 정보요구를 정확하게 반영해야 한다. 마지막으로 질의어가 다양한 주제를 다루되 너무 광범위하거나 너무 좁은 개념이어서는 안 된다고 제시하고 있다. 이를 고려하여 본 연구에서는 최근 3년(2019년~2021년)간 국가기록포털에서 자주 입력된 상위 300개의 질의어를 대상으로, 고빈도 질의, 기록물 유형에 따른 질의, 생산기관 특화 질의 등, 다양한 관점에서 검색 성능평가를 위한 질의셋 후보를 선정하였다.

정답셋은 선정한 질의셋을 통해 국가기록물 검색도구를 대상으로 수작업 검색을 수행한 후 검색목적, 즉 해당 질의의 정보요구에 부합하는 기록물인지를 확인해 적합 문서를 정답문서로 수집한다. 그러나 현재 국가기록물을 검색하기 위한 도구는 2장에서 언급한 대로 CAMS와 국가기록포털로 나뉘어 있으며 서로 다른 검색엔진을 사용하고 있어, 그 결과 또한 상이하다. 이에 포괄적인 정답셋 구축을 위하여 먼저 CAMS 검색 결과로부터 적합도 평가를 통해 정답셋을 구축한 뒤, 국가기록포털에서 동일한 방식으로 적합도 평가를 수행해 CAMS에서 구축한 정답셋을 기반으로 평가셋을 확장하였다.

평가셋 구축에는 총 5명의 작업자가 참여하였다. 1명은 기록관리 박사과정으로 전문가 관점으로 참여, 2명은 기록관리에 대한 배경지식이 있는 이용자 관점으로, 나머지 2명은 기록물에 관한 지식이 없는 일반 이용자 관점을 대변하도록 구성하였다. 특히 질의셋 구축 및 정답 평가 시 질의어를 통해 찾고자하는 검색 목적을 구체화 하는 작업을 수행하였다. 질의별 검색목적 선정은 5명의 작업자가 모두 검색목적 선정 후 교차검증을 하였고, 적합도 평가는 전문가 및 일반인 그룹으로 나누어 두 그룹의 작업자가 평가 후 검증하는 방식으로 수행하였다.

3.2 질의어 선정 결과

질의셋 선정에 앞서 최근 3년간 국가기록포털의 이용자 질의로그를 분석한 결과, 이용자들이 실제로 포털 검색창에 입력한 질의어 대부분이 1~2단어로 구성되어 있어 이용자들의 정보요구를 명확히 나타내지 못하고 있었다. 또한,

동영상이나 지도 등 특정 유형의 기록물을 원하는 질의임에도 이에 대한 검색도구의 활용에는 별 차이가 없음을 파악하였다. 또한, 없었거나 새로 신설된 정부기관에 대한 질의도 상당 부분을 차지하고 있었다. 이 같은 특성을 반영하고 질의어 후보에 대한 검색목적의 구체화하기 위해 해당 질의어를 대상으로 이용자가 찾고자 하는 문서가 무엇인지, 어떤 기록이 검색 요구를 만족시킬 수 있는 정답일지 등을 다수의 작업자에게 임의로 설정하도록 하고, 이를 교차검증하여 일반인의 관점에서 대표적인 검색목적 상황을 선정하였다. 또한, 기록물 유형과 생산기관에 따른 정보요구를 특화하여 문서집합의 특성에 따른 질의어를 포함하였다. 기록물 유형 선정 기준은 <표 1>에 의거, 국가 기록포털의 '상세검색'에서 표시되는 기록물형태 항목 9종과 기록물구분 항목 9종을 참조하여 정의하였다. 생산기관 관련 질의는 생산기관변천정보와 기술계열서를 참조하여 선정하였다.

<표 2> 평가셋용 질의어 선정 결과

질의어	개수	질의어 특성	특성
QL-001~021	21	국가기록포털 고빈도 질의	짧은 질의어(1~2단어)로 구성되어 실제 검색목적이 구체화되지 않음
QL-022~049	28	기록물 유형별 질의	이용자가 찾고자 하는 기록물 유형을 구체화하여 선정
QL-050	1	생산기관 관련 질의	생산기관 변천 및 기술계열서 특화 검색어

<표 2>는 앞서 설계한 방법을 통해 선정된 질의어 선정 결과이다. 이용자가 실제로 자주 사용하는 질의어와 그 질의어를 통해 찾고자 하는 문서가 무엇인지 검색목적 설정, 최종 결과는 국가기록원과 협의하여 선정하였다. QL-001부터 QL-021은 고빈도 질의어로 이용자들이 국가기록포털에서 실제로 검색한 질의어로 2019년에서 2021년까지 검색 빈도가 높은 상위 300개 단어 중에서 선정하였다. QL-022부터 QL-049까지는 기록물 유형에 따른 질의어로 해당 기록물 유형을 포함하는 기록물을 검색할 수 있는 질의어 중에서 후보를 선정하였다. QL-050은 생산기관 검색을 위한 질의어로 생산기관 변천 및 기술계열서에 특화된 검색어이다. 최종적으로 선정된 평가셋용 질의어는 고빈도 질의 21건, 기록물 유형별 질의 28건, 생산기관 관련 질의 1건으로 총 50건으로, 구체적인 질의어 내용은 <표 3>에 예시하였다.

<표 3> 평가셋용 질의어 예시

구분	번호	질의어명	검색목적	검색목적구체화	비고
고빈도 질의	QL-001	토지조사부	개인의 재산, 경력 (조상땅찾기)	특정 지역(경기도 수원군 음 덕면 운석리)에 소유하고 있는 조상땅을 찾기 위해	2021년 1월~2월, 4월~6월, 10월 많이 찾은 검색어

	QL-021	공역	공역 지정	공역의 변화 과정을 조사하기 위해 공역 지정 내역을 조사	'향로공역'의 ' ' 때문에 검색이 잘되지 않을 수 있기에 질의어를 분리한 항목
기록물 유형별 질의	QL-022	박정희 암살	박정희 대통령 암살 사건	박정희 암살 사건에 대한 자료조사를 위한	기록물구분: 해외기록물 기록물형태: 일반문서

	QL-049	평창 동계 올림픽 기념 배지	평창 동계 올림픽 기념 배지 목록	평창 동계 올림픽 기념 배지엔 어떤 것들이 있었는지 조사하기 위한	기록물구분: 행정박물 기록물형태: 기념류
생산기관 관련 질의	QL-050	미래창조 과학부 연구	연구개발사업보고서	주요 연구개발사업 성과에 대한 자료조사	2013년 3월 23일 신설되었다가, 2017년 7월 26일 해산됨

3.3 정답셋 구축

3.3.1 정답셋 구축 과정

정답셋은 주어진 문서집합 내에서 질의의 검색목적에 만족시키는 모든 문서이다. 그러나 국가기록원에서 관리하고 있는 모든 문서를 대상으로 질의에 적합한 문서인지를 판별하기 위해서는 기초적인 검색이 수반된다. 국가기록포털에서 서비스하는 국가기록물은 원칙적으로 CAMS 시스템에서 관리하고 있는 기록물과 동일한 것으로, 문서집합 내에 산재해 있는 정답을 선정하기 위해서는 먼저 CAMS 검색이 선행되어야 한다고 판단하였으며, 최종적으로는 이용자들에게 실제 제공되고 있는 국가기록포털 결과를 반영하는 순으로 진행하였다.

정답셋을 위한 정보요구 적합도 평가는 검색된 문서 제목을 대상으로 진행하였다. 이용자는 일반적으로 검색시스템을 이용할 때 검색 질의에 적합한 문서인지 판단 시 제목을 우선하는 경향이 있다(박소연, 이준호, 전지운, 2006). 또한, 국가기록물 검색시스템에서 온라인으로 전문(full text) 열람을 지원하지 않는 문서는 별도의 서비스를 이용해야 한다는 문제가 있다. 따라서 현실적으로 모든 기록물의 원문을 확인하기 어려워 검색된 문서의 제목을 기준으로 해당 문서가 검색목적에 부합하는지에 대한 적합도 평가를 진행, 도구별로 5개의 적합문서를 수집하였다.

이러한 설계를 기반으로 선정한 총 50개의 질의어에 대해 각 시스템의 검색도구별로 원질의(original query) 그대로 검색을 수행하고 검색된 기록물을 '적합문서'와 '부적합문서'로 평가하였다. 1차 검색 결과가 검색목적에 만족하지 못하면 질의 목적에 맞게 질의어를 확장하거나 결과 내 재검색을 통해 2차, 3차 평가를 수행하였으며, 이때 포함한 추가 검색 질의어를 질의셋에 명기하였다. 이후, 각 시스템의 검색도구별 검색 결과 화면을 기반으로 상위 5개의 적합문서를 취합하여 정답셋을 구축하였다.

3.3.2 CAMS 적합성 평가 및 1차 정답셋 구축

기초 정답셋은 선정한 질의셋을 활용하여 CAMS 검색시스템을 대상으로 적합성 평가를 수행한 결과에 기반해 구축되었다. CAMS의 적합성 평가는 CAMS의 검색도구인 '기록물철상세검색', '기록물건상세검색', '신통합검색'('철/건통합검색', '철검색', '건검색') 총 3개의 검색도구에 대해 평가를 수행하였다.

<그림 3>은 CAMS 검색도구 적합성 평가를 통해 구축된 정답셋 중 일부를 예시한다. 정답셋은 <표 2>의 50개 질의어에 대해 각 검색도구별로 상위 5개의 정답문서를 추출하여 정답셋을 구축하였으며 기록물 유형, 정답문서 순위, 총 검색문서 개수, 관리번호, 철/건제목 등을 포함하였다. <표 4>는 CAMS를 통해 구축된 검색도구별 정답셋 개수를 정리한 표로, '기록물철상세검색' 115개, '기록물건상세검색' 154개, '신통합검색' 596개로 총 866개의 정답문서로 구성되었다.

질의어 번호	질의명	검색목적	검색도구	건				기록물 유형				
				정답 순위	검색 결과	관리번호	제목					
QL-001	토지 조사부	경기도 수원군 음덕면 온석리에 소유하고 있는 조상땅을 찾기 위해	기록물철 상세검색	-	-	-	-	일반 문서류				
			기록물건 상세검색	-	-	-	-					
			신통합검색	-	-	-	-					
QL-002	씨랜드	씨랜드 화재사고의 원인파 그에 따른 대책을 조사	기록물철 상세검색	-	-	-	-	일반 문서류				
			기록물건 상세검색	1	40	BA0817719	씨랜드 화재사고 관련 어린이시설에 대한 안전대책 수립·추진					
								
				11		DA0771766	씨랜드 화재사고관련 학생안전 지도대책					
			신통합검색	2	10	BA0817719	씨랜드 화재사고 관련 어린이시설에 대한 안전대책 수립·추진					
				3		DA0771766	씨랜드 화재사고 관련 학생안전 지도대책					
								
				8		DA0567385	화성 씨랜드 화재사고 관련 재난관리행정 개선대책보고					
			QL-050	미래 창조 과학부	연구개발사업 보고서	기록물철 상세검색	-		-	-	-	-
						기록물건 상세검색	-		-	-	-	
신통합검색	5	10				DM00068133	2016 미래창조과학부 주요 연구개발사업 성과분석보고서					
	7					FM00033299	2015 미래창조과학부 주요 연구개발사업 성과분석보고서					
	8					DM00054376	2013회계연도 성과보고서 (미래창조과학부)					
	9		DM00054433	2013회계연도 결산보고서 (미래창조과학부) 별책: 미래창조과학부 성과보고서								

<그림 3> CAMS 검색을 통해 구축된 정답셋 예시

<표 4> CAMS 검색도구별 정답셋 구축 개수

구분	CAMS 검색도구		
	기록물철상세검색	기록물건상세검색	신통합검색
정답문서	115	154	597
총합	866		

3.3.3 국가기록포털 적합성 평가 및 정답셋 확장

궁극적으로 이용자에게 제공되는 검색 결과는 국가기록포털 시스템 결과이다. 1차 구축된 정답셋을 검증 및 확장하기 위해 국가기록포털에서도 CAMS 검색시스템과 동일한 방식의 2차 적합성 평가를 수행, 기존의 정답셋을 확장하였다. 국가기록포털의 적합성 평가는 국가기록포털에서 제공하는 검색서비스인 ‘통합검색’, ‘상세검색’, ‘원문검색’, ‘기술계층별검색’, ‘주제별상세검색’ 5개의 검색도구를 대상으로 평가를 수행하였다. 검색도구별로 상위 5개의 문서를 선정하여 구축한 정답셋은 <그림 4>와 같으며, 국가기록포털을 통해 구축된 검색도구별 정답셋은 ‘통합검색’ 379개, ‘상세검색’ 327개, ‘원문검색’ 84개, ‘기술계층별검색’ 110개, ‘주제별상세검색’ 123개로 총 1,023개의 정답문서로 구성되었다(<표 5> 참조).

질의어 번호	질의명	검색목적	검색도구	건				기록물 유형				
				정답 순위	검색 결과	관리번호	제목					
QL-001	토지 조사부	경기도 수원군 음덕면 온석리 에 소유하고 있는 조상땅을 찾기 위해	통합검색	-	-	-	-	일반 문서류				
			상세검색	-	-	-	-					
			원문검색	11	15	CJD0004409	토지조사부(경기도수 원군음덕면온석리)					
			기술계층별 검색	-	-	-	-					
			주제별상세 검색	-	-	-	-					
QL-002	씨랜드	씨랜드 화재사고의 원인과 그에 따른 대책을 조사	통합검색	4	114	DA0567386	화성 씨랜드화재 관련 보고	일반 문서류				
				6		DA0567385	화성 씨랜드 화재사고 수습					
								
				14		DA0567383	화성 씨랜드 화재사고 상황보고					
			상세검색	3	94	DA0567386	씨랜드 청소년 수련읍집 현황					
				4		DA0567386	화성 씨랜드화재 관련 보고					
								
				9		DA0567383	화성 씨랜드 화재사고 상황보고					
			원문검색	-	-	-	-					
			기술계층별 검색	-	-	-	-					
			주제별상세 검색	-	-	-	-					
			QL-050	미래 창조 과학부		통합검색	1		9	DM00054433	2013회계연도 결산보고서 (미래창조과학부) 별책: 미래창조과학부 성과보고서	정부 간행물
							2			DM00054376	2013회계연도 성과보고서 (미래창조과학부)	
...										
8	FM00033299	2015 미래창조과학부 주요 연구개발사업 성과분석보고서										
상세검색	5	695				DM00065419	[2015년 과학기술종합 조정지원사업] 과학기 술기본계획 연도별 추진실적 점검 및 시행계획 수립					
	7					DM00065385	[생명연구지원확보관 리및활용사업 최종보 고서] 유전자변형마우 스(GEM) 기반구축 최종보고서					
					
	12					DM00065394	[줄기세포연구사업 최 종보고서] 줄기세포의 고민감도 표지자 개발 및 난치성 신경계/심 장질환에서 생체 내 자 기공명영상 추적 기반 기술 연구					
원문검색	-	-				-	-					
기술계층별 검색	-	-				-	-					
주제별상세 검색	-	-				-	-					

〈그림 4〉 국가기록포털 검색을 통해 구축된 정답셋 예시

<표 5> 국가기록포털 검색도구별 정답셋 구축 개수

구분	국가기록포털 검색도구				
	통합검색	상세검색	원문검색	기술계층별검색	주제별상세검색
정답문서	379	327	84	110	123
총합	1,023				

질의어번호	질의명	검색도구	정답셋		검색도구별 개수	총 합계
			관리번호	제목		
QL-001	토지 조사부	CAMS	CJD0004409	토지조사부(경기도수 원군읍덕면온석리)	2	2
			CJD0004409	토지조사부(경기도수 원군읍덕면온석리)		
QL-002	씨랜드	포털	DA0567386	화성 씨랜드화재 관련 보고	15	33
			DA0567385	화성 씨랜드 화재사고 수습		
				
		CAMS	DA0036795	씨랜드 관련 문서사본	18	
			DA0771766	화성 씨랜드 사안		
			BA0817719	씨랜드화재사고관련 (2-2)		
DA0567388	씨랜드 상황관리 보고서 (2-1)					
QL-003	붕괴 사고	포털	C11M02772	삼풍백화점 붕괴사고 백서	19	26
QL-050	미래 창조 과학부	포털	DM00054433	2013회계연도 결산보고서 (미래창조과학부) 별책: 미래창조과학부 성과보고서	14	14
			DM00054376	2013회계연도 성과보고서 (미래창조과학부)		
				
			FM00020771	신혈관질환 분자시스템생물학연구 (최종보고서)		

<그림 5> 최종 취합된 정답셋 예시

<표 6> 최종 정답셋 구축 결과

구분	검색도구	정답문서	검색도구별 합계	중복개수	총계
CAMS	기록물철상세검색	115	866	894	995
	기록물건상세검색	154			
	신통합검색	597			
국가기록포털	통합검색	379	1,023		
	상세검색	327			
	원문검색	84			
	기술계층별검색	110			
	주제별상세검색	123			

<그림 5>는 최종 정답셋의 예시로, CAMS에서 구축한 정답셋(<그림 3>)을 기반으로 국가기록포털에서 선정한 정답문서(<그림 4>)를 병합한 형태이다. <표 6>은 최종 구축된 정답셋 결과로, CAMS로부터 수집된 866개 정답과 국가기록포털 1,023개의 정답 중 관리번호와 기록물 제목이 중복된 894개를 제외한 총 995개로 구성되었다.

4. 평가셋 기반 국가기록포털 품질 평가

4.1 평가방법 및 평가지표

국가기록포털의 검색 성능평가는 다음과 같은 방법으로 진행하였다. 먼저 구축된 평가셋을 기반으로 국가기록포털의 '건' 기록물 검색 결과의 품질을 점검하였다. 기록물 '건'을 대상으로 평가한 이유는 일반 이용자가 기록의 계층에 대한 이해도가 낮고 국가기록포털의 검색도구에 대한 적절한 설명이 미흡하기 때문으로(이효진, 김지현, 2021), 본 연구에서는 일반 이용자가 국가기록포털에서 기록물을 탐색할 때, 익숙하지 않은 개념인 주제나 군, 철보다 검색 시 가장 먼저 제시되는 화면인 건기록물 검색 결과에서 검색목적에 부합하는 기록물을 찾을 것이라 가정하였다. 평가를 진행한 검색도구는 건기록물 검색을 제공하지 않는 '기술계층별검색'을 제외하고 '통합검색', '상세검색', '원문검색', '주제별상세검색'을 대상으로 선정하였다.

평가는 평가자들이 평가셋에 포함된 질의어를 입력하여 검색을 실행하고, 검색 결과 화면에 출력된 상위 30개의 결과, 즉 세 번째 페이지까지를 대상으로 기록물관리번호와 기록물 제목을 확인한다. 이를 사전에 구축한 정답셋의 기록물관리번호와 제목과의 대조를 통해 기록물의 적합성 여부를 평가하였다. 이때 1차 검색 결과에서 적합문서가 검색되지 않거나 가장 처음 검색된 적합문서가 하위에 위치할 경우, 질의셋에 명기된 추가 질의어를 참조해 검색목적에 맞는 질의어 확장이나 결과 내 재검색을 통해 2차, 3차 평가를 실시하였다. 평가 대상 결과를 상위 30개로 한정하는 이유는 일반적으로 이용자가 필요한 정보를 찾을 때, 검색 결과에서 상위 세 번째 페이지까지를 확인하는 경향이 있기 때문이다(이상진, 정재승, 2011).

이후, 검색 품질 테스트 결과를 바탕으로 다양한 평가지표를 활용하여 현행 국가기록포털의 검색 성능을 분석하였다. 본 연구는 개별질의에 대해서는 일반적으로 검색시스템 성능 평가척도로 가장 많이 사용되는 정확률(Precision)과 재현율(Recall), MRR(Mean Reciprocal Rank)을 사용하였다(이병길, 김희섭, 2013; Voorhees, 1999). 전반적인 검색 품질에 대한 종합적인 평가지표로는 개별질의에 대한 최종 정확률과 재현율은 각각 Macro 평균과 Micro 평균을 사용하여 50개의 질의어에 대한 검색도구별 정확률, 재현율의 평균을 산출하고 이 둘의 통합지표인 F-score를 측정, 그 결과를 분석하였다(Sebastiani, 2002).

4.2 성능평가 결과

4.2.1 검색 품질 평가

<표 7>은 평가 대상인 4개의 검색도구별 검색 성공 여부를 정리한 것으로, 수치상으로는 50개의 질의에 대한 정답을 도구별로 채점했으므로 200개(50개 질의 x 4개 도구)의 검색 결과를 평가한 결과이다. 가장 대표적인 검색도구인 '통합검색'을 평가한 결과, 원 질의어 자체를 그대로 입력한 1차 검색을 통해 성공한 질의는 37개, 재검색 과정을 통해 성공한 질의는 9개로, 전체 94% (47개)의 질의에 대해 상위 30개 결과에서 검색목적에 만족시켰음을 알 수 있다. '상세검색' 도구를 통해 1차 검색으로 검색목적에 만족한 경우는 32개, 재검색을 포함하면 전체 44개 (88%)의 질의에 대해 검색목적에 부합하는 결과를 제시하였다.

'원문검색'과 '주제별상세검색' 도구를 통해 검색에 성공한 질의는 각각 24개(48%)와 22개(44%)로, 두 도구 모두 절반 넘는 질의에 대해 정답을 제시하지 못한 것으로 평가되었다. 이를 200개의 건기록물 검색 결과셋 전체로 환산하여 살펴보면, 검색에 성공한 질의와 검색에 실패한 질의는 각각 137개(68%)와 63개(32%)로 평가되어 전체 검색도구가 절반이 넘는 질의에 대해 정답을 제시한 것으로 나타났다.

<표 7> 검색도구별 검색에 성공한 질의어 개수

구분		도구별 적합문서가 검색된 질의어 개수(건기록물)				전체	종합
		통합검색	상세검색	원문검색	주제별상세검색		
검색 성공	1차 검색	38	32	21	21	137	49
	재검색	9	12	3	1		
검색실패		3	6	26	28	63	1
총합		50				200	50

<표 8> 검색도구별 질의어에 따른 성능 평가 결과

검색도구	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

4개의 도구 중 어느 하나라도 정답을 찾은 경우를 종합해보면, 전체 50개 질의 중 49개의 질의가 최소 하나 이상의 검색도구에서 정답을 찾은 것으로 나타났다. 검색에 실패한 나머지 1개의 질의는 기록물구분이 ‘행정박물’로 건기록물이 없는 질의였기 때문에 사실상 모든 질의어가 4개의 검색도구를 활용하여 최소 한번은 검색에 성공하였음을 의미한다.

<표 8>은 건기록물에 대한 각 검색도구별 정확률과 재현율을 비교한 것으로, 각 질의별 정확률 혹은 재현율을 더해 전체 질의 수 (50개)로 나누어 Macro 평균과 전체 검색 결과(최대 상위 30개) 혹은 전체 정답문서 중 검색된 정답 수를 평균 낸 Micro 평균으로 나누어 제시했다. 또한 Macro 평균 정확률과 재현율을 기준으로 통합지표인 F-score를 측정하였다.

평가결과, ‘통합검색’의 Macro 평균 정확률은 38.50%로 가장 높은 수치를 보였고 Micro 평균 정확률은 32.08%로 평가되었다. ‘상세검색’은 Macro 평균 정확률 35.12%, Micro 평균 정확률은 28.74%로 ‘통합검색’에 비해 다소 낮게 나타났다. 특히 ‘원문검색’의 경우 Macro 평균 정확률과 Micro 평균 정확률이 각각 21.02%와 25.83%로, 매우 낮은 성능을 보였다. ‘주제별상세검색’은 Macro 평균 정확률은 23.00%로 낮은 반면 Micro 평균 정확률은 35.63%로 가장 높게 측정되었다.

건기록물에 대한 검색도구별 평균 재현율을 살펴보면 다음과 같다. ‘통합검색’의 Macro 평균 재현율은 57.73%, Micro 평균 재현율은 51.27%로, 각 평균 정확률보다 모두 높게 측정되었다. ‘상세검색’의 Macro, Micro 평균 재현율은 각각 44.13%와 39.73%로, ‘통합검색’과의 평균 정확률 편차보다 더 크게 하향 평가되었다. ‘원문검색’은 Macro 평균 재현율이 16.89%, Micro 평균 재현율이 16.34%로, 네 도구 중 가장 낮은 재현율을 보였게 측정되었으며, ‘주제별상세검색’의 Macro, Micro 평균 재현율은 각각 21.68%, 18.59%로, ‘원문검색’과 유사하다.

정확률과 재현율을 종합적으로 반영한 F-score로 살펴보면, ‘통합검색’과 ‘상세검색’은 각각 42.83%, 34.59%로 측정되었다. 이들 도구는 모두 정확률에 비해 재현율이 높은 양상을 보였다. 반면 ‘원문검색’과 ‘주제별상세검색’은 각각 15.19%와 15.69%로 ‘통합검색’과 ‘상세검색’에 절반에도 못 미치는 성능을 보였으며 이 두 도구 모두 재현율이 현저히 낮은, 즉 전체 정답 중 찾은 정답의 수가 매우 저조한 것으로 평가되었다.

도구별 평가 결과를 보다 세부적으로 살펴보면, ‘통합검색’과 ‘상세검색’에서는 Macro 평균 정확률이 Micro

평균 정확률보다 높게 측정되었는데, 이는 특정 질의에서 정답을 많이 찾았거나 못 찾는 등의 쓸림 현상이 크지 않고 전체적으로 정답을 찾은 양상이 유사함을 함의한다. 반면, '원문검색'과 '상세검색'은 Micro 평균 정확률이 더 높게 나와 특정 질의에 국한하여 정답을 많이 찾았음을 의미한다. 특히 '주제별상세검색'의 경우 Micro 평균 정확률과 Macro 평균 정확률의 차이가 가장 크게 나타나 그 편차가 심한 것으로 평가되었는데, 검색 결과 제시 질의 수 자체가 전체 50개 중 22개에 그친다는 점이 주요한 원인이다. 이는 '주제별상세검색'이 국정분야별로 주제를 선정하여 구조화하고 그 주제와 관련된 기록물을 수작업으로 연계해주는 방식으로 구성되어 특정 주제와 연계된 기록물의 수가 미흡해 기록물이 정확하게 검색이 되지 않아 발생하는 문제로 추정된다.

특이한 점은 '상세검색'의 평균 정확률과 평균 재현율이 '통합검색' 결과보다 낮게 나타난 점으로, 일반적인 포털이 '상세검색'을 통해 검색범위를 더 명확히 함으로써 검색 정확도가 높게 나타나는 양상과는 반대이다. 이는 국가기록포털에서의 '상세검색' 방식이 이례적임을 방증한다. 또한 '원문검색'의 재현율이 현저히 낮은 점도 눈에 띄는데, 이 역시 포털 이용자가 일반적으로 '원문검색'하면 떠올리는 전문(Full text) 검색이 아닌, 원문이미지가 제공되는 기록물 내에서 메타데이터 검색을 통해 결과를 제공하는 검색 도구이기 때문으로, 애초 국가기록원에서 보존하고 있는 전체 기록물 수에 비해 원문이미지가 제공되는 기록이 적고 메타데이터 자체가 부실한 경우가 많음에서 기인한 것으로 파악된다.

〈표 8〉을 본 평가가 국가기록포털에서 제시한 검색 결과 전체를 대상으로 평가한 것이 아닌, 검색 결과 중 최대 세 번째 페이지까지의 결과, 즉 상위 30개의 결과를 평가한 것을 감안해 재해석해보면, 이용자가 국가기록포털의 '통합검색'을 통해 어떤 질의를 검색하는 경우, 상위 30등 내에 검색 목적에 부합한 결과를 제시할 확률은 38.5%이고 그 결과가 전체 정답의 57.73%에 해당, 검색 만족도는 42.83%로 절반에 못 미치는 수준이다. '주제별상세검색'의 경우, 전체 50개 중 22개의 질의에 대해서만 정답을 제시해 재현율이 21.68%로 매우 낮은 뿐만 아니라 정확률 역시 23%로 저조하다. 이는 이용자들이 일반적인 포털서비스에 기대하는 성과와는 편차가 다소 큰 수치로, 국가기록포털의 검색 성능에 대한 개선이 시급함을 다시금 확인시키는 결과이다.

4.2.2 검색 효율 평가

정확률이나 재현율이 검색 성능의 품질을 정보요구를 얼마나 충족시켰는지의 효과(effectiveness)를 측정하는 지표라면, MRR은 정보요구를 얼마나 쉽게 충족시킬 수 있는지 즉 효율(efficiency)을 측정하는 지표이다. MRR은 검색 결과 중 정답이 최초로 나타난 순위의 역수를 평균을 낸 것으로, 상위에 정답을 제시하면 높은 점수, 즉 숫자가 높을수록 좋은 성능임을 의미한다.

〈표 9〉는 건기록물에 대한 검색도구별 MRR 측정 결과이다. 먼저 검색도구별로 정답문서가 제시된 질의 수를 기준으로 살펴보면 '통합검색'은 전체 50개의 질의 중 검색 결과를 제시한 질의는 47개로, 그 중 정답문서가 1등으로 제시된 수는 33개이다. '통합검색'에서 최초로 정답문서가 제시된 순위는 평균 2.48위였으며 MRR은 0.77이다. 이는 평균 3위 이내 정답이 하나 이상 있을 확률이 매우 높음을 의미한다.

〈표 9〉 검색도구별 질의어에 따른 MRR

검색도구	결과 제시 질의 수	1등 정답 질의 수	최초 정답문서 순위 평균	MRR	
				결과 제시 질의	전체 질의
통합검색	47	33	2.48	0.77	0.72
상세검색	44	30	2.18	0.77	0.68
원문검색	24	14	3.7	0.69	0.33
주제별상세검색	22	16	1.8	0.82	0.36

‘상세검색’은 50개 중 44개에 대해 결과를 제시하였고, 그 중 1등으로 정답 제시에 성공한 질의가 30개로 파악되었다. 최초 정답문서 순위는 통합검색 2.48 보다 높은 2.18로 나타났으나 MRR은 통합검색과 동일한 0.77로 측정되었다. ‘원문검색’은 전체 50개 질의 중 24개만 결과를 제시하여 응답률이 낮을 뿐 아니라, 1등 질의와 최초 정답문서 순위, MRR 수치 모두 각각 14와 3.7, 0.69로 가장 낮은 성능을 보였다.

‘주제별상세검색’은 22개 질의에 대해 결과를 제시하여 응답율은 가장 낮았으나, 제시된 결과 중 1등에 정답을 제시한 질의 수가 16개, 최초 정답문서 순위와 MRR이 각각 1.8과 0.82로 나타나. 검색 효율 지표로는 가장 높게 평가되었다. 그러나 정답문서를 제시하지 못한 질의를 포함하여 전체 50개 질의를 대상으로 MRR을 산출하면 ‘통합검색’이 0.72로 가장 높게 평가되었으며, ‘상세검색’은 0.68로 나타나 통합검색보다 낮은 점수를 보였다. ‘주제별상세검색’과 ‘원문검색’은 각각 0.33과 0.36으로 매우 낮게 측정되었다.

<표 8>과 <표 9>를 종합적으로 살펴보면, 먼저 정확도 측면에서 ‘통합검색’ 도구의 경우, 평균 정확률과 재현율 역시 각각 30%와 50%대로. 일반 포털서비스에 비하면 그리 높지 않은 수치를 보였다. 검색 효율 측면에서는 평가 대상 50개 질의 중 47개의 질의에 대해 검색 결과를 제시하였고, 그 중에서 33개의 질의에 대해서는 1등에 정답을 제시하는데 성공. MRR을 0.7 대를 보이는 것으로 평가되었다. 특히 최초정답 문서 평균 순위가 2.48로, 이러한 결과는 평가셋으로 구성된 다양한 질의에 대해 첫 번째 화면에서 1개 이상의 정답을 찾을 확률, 특히 1등으로 찾을 확률이 66%(33/50)임을 알 수 있다.

‘상세검색’의 경우, ‘통합검색’에 비해 정확률 향상을 기대하였으나 그렇지 못했고, 검색 범위를 좁혀 검색하는 특성으로 인해 결과 제시 질의 수가 44개로 줄은 반면 최초 정답 순위에서 약간의 (2.48 → 2.18) 성능 향상을 보였으나 MRR 척도에서는 같은 수치를 보여 유의미한 차이는 아닌 것으로 파악되었다. 반면 ‘원문검색’과 ‘주제별상세검색’의 경우, 전체 50개 질의 중 검색 결과를 제시한 경우가 각각 24개와 22개로 절반에도 못 미치는 응답 결과를 보였다. 그러나 ‘주제별상세검색’의 경우, 제시된 검색 결과 중 72%(16/22)가 1등에 정답을 제시하는데 성공, 0.82의 높은 MRR을 보였으며, 이는 특화된 질의에 대해서는 검색 만족도가 높을 수 있음을 보여준다.

분석 결과를 총평해보면, 국가기록포털의 전반적인 검색 정확도는 낮은 편이다. 특히 정확률 측면에서 모든 검색 도구가 40%를 밑도는 성능을 보여 이에 대한 개선이 시급하다. 특히 검색 만족도를 나타내는 F-score의 최고치가 43%를 밑돌며, ‘원문검색’과 ‘주제별검색’의 경우 15% 선에 그치는 수치는 일반 이용자의 기대에 부응하지 못하는 결과이다. 그에 비해 재현율은 ‘통합검색’이 50%는 상회하는 성능을 보이고 검색 효율 측면에서도 상위 등수에서 정답을 최소 1개 이상 찾을 확률이 다소 높은 것으로 파악되었다. 그러나 이 역시 정답셋 수집 시 각 도구별 상위 5등까지의 정답을 모수로 선택했기 때문에 발생한 왜곡으로 실제 체감 성능은 이보다 낮을 것으로 예상된다.

또한, 각 검색도구 도구별 기대치와 실제 성능 간의 간극이 크고, 질의별 특성에 따라 도구별 대응 결과가 달라 이에 대한 개선이 필요함을 알 수 있었다. 그러나 이는 상위 30등까지의 검색 결과만을 평가한 점, 애초 정답셋을 구축할 때 활용한 기초검색 결과에 정답이 포함되어 있지 않은 경우를 고려하지 못한 점 등의 한계가 있다. 다음 절에서는 본 평가 과정을 통해 도출된 문제점을 기술하였다.

4.3 국가기록포털 검색 품질 평가 한계점

본 연구는 국가기록포털에서 제공하고 있는 검색도구의 품질을 측정하는 연구로, 규격화된 평가셋에 기반해 다양한 평가척도를 통해 정량적인 수치로 검색 성능을 평가하였다. 그러나 그 과정에서 다음과 같은 한계점이 도출되었다. 첫째, 질의어 선정 과정에서 이용자의 의견이 완전히 반영되었다고 말할 수는 없다. 비록 국가기록포털을 통해 수집된 이용자 질의로그에 기반해 고빈도 질의어를 선별하였으나, 해당 질의어는 1~2단어로 구성되어 있어 검색목표가 불분명해 검색 품질 평가를 위해서는 구체적인 정보요구 설정이 필요하였다. 이를 위해 이용자 집단을 선정하여 설문조사를 하는 등의 직접적인 의견 청취 과정과 공통의 정보요구에 대한 수렴이 요구되었으나 현실적으로 수행하

기 어려워 다수의 작업자가 교차검증을 통해 일반적이고 보편적인 수준의 시나리오를 수집하였다.

정답셋 역시, 국가기록원이 보존하고 있는 기록물 전체를 대상으로 정답을 수집하진 못했다. 1억 건이 넘는 모든 국가기록물 목록을 검토하고 원문을 열람해서 후보 정답을 추출하는 것은 현실적으로 불가능하기 때문에 국가기록원 내부 직원들이 활용하고 있는 CAMS 검색도구를 활용해 기초 정답 후보를 선별하였으며, 도구별로 상위 5개까지 수집하였다. 따라서 5위 이후에 나타난 정답이거나 애초에 CAMS 검색도구의 오류로 인해 찾아지지 않은 숨어있는 기록의 경우에는 정답셋에 반영되지 못했다.

셋째, 품질 점검이 국가기록포털의 검색 결과에 국한되어 진행되었다. 검색 품질은 검색시스템의 색인 정책과 검색 결과를 제시하는 순위화, 즉 랭킹 정책에 따라 좌우된다. 그러나 국가기록물 검색시스템 내부의 정확한 랭킹 알고리즘을 알 수가 없어 검색 결과로 제시된 화면만으로 평가를 진행하였다. 또한, 검색 성능 채점의 경우에도 검색 결과 전체가 아닌 이용자의 일반적인 검색 행위 패턴을 고려해 상위 30위까지로 한정하여 평가한 점이 한계로 여겨진다.

5. 결론

본 연구는 노후화된 국가기록포털 검색시스템의 개선을 위해 국가기록원에서 운영하는 검색시스템의 현황을 파악하고, 국가기록포털에서 제시하는 검색 결과의 품질을 평가하는 것을 목적으로 하였다. 이를 위해 먼저 CAMS와 국가기록포털에서 제공하는 기록물 유형과 시스템별 검색도구의 현황을 분석하였고, 검색 결과의 품질을 점검하기 위한 평가셋을 구축하였다. 문서집합은 국가기록원에서 관리하는 기록물 114,777,646건을 대상으로 하였으며, 이용자 관점에서의 품질 평가를 위해 실제 이용자들이 국가기록포털에서 검색한 질의어를 대상으로 질의셋을 선정, 2차에 걸친 정답셋 수집 과정을 거쳐 최종 평가셋을 구축하였다. 국가기록포털의 검색 품질 평가는 전통적인 검색 평가척도인 정확률과 재현율을 대상으로 Macro/Micro 평균치와 통합지표인 F-score를 제시하였고 검색 효율을 평가하기 위해 MRR을 측정하였다.

국가기록포털에서 제공하는 검색도구의 검색 성능 평가 결과는 다음과 같다. 먼저, 검색성공여부에서는 4개의 검색도구를 통해 50개의 질의어가 최소 한번은 정답문서를 찾은 것으로 나타났다. 그러나 세부적으로 보면, ‘통합검색’과 ‘상세검색’의 평균 정확률의 경우 30%대, ‘원문검색’과 ‘주제별상세검색’은 20% 초반대로, 전체적으로 저조하다. 통합지표인 F-score로는 ‘통합검색’ 도구에서 최고치인 43%, ‘원문검색’에서 최저치인 15%를 보였다. 특히, ‘원문검색’과 ‘주제별상세검색’은 특정 질의어에서만 정답문서가 제시되는 경향이 심한 것으로 평가되었다. 검색 효율 지표인 MRR 상에서도 전체 50개 질의를 기준으로 ‘통합검색’, ‘상세검색’, ‘원문검색’, ‘주제별상세검색’ 순으로 나타나 ‘원문검색’과 ‘주제별상세검색’의 검색 성능과 추천 성능이 모두 낮게 평가되어 개선이 시급할 것으로 보인다.

한편, 본 연구에서는 국가기록포털을 통해 수집된 질의를 활용하였지만, 질의가 너무 짧아 검색목적이 불분명해 작업자의 교차검증을 통해 일반적인 검색 시나리오를 설정하였다. 실제 이용자집단을 선정하여 설문조사나 인터뷰 등을 통해 질의어를 선정한다면 평가셋 구축의 타당성을 높일 수 있을 것으로 예상된다. 이에 더해 국가기록원이 보존하고 있는 국가기록물 전체를 대상으로 정답셋을 수집하지는 못했으며, 일반 이용자의 검색 행위 패턴을 중심으로 두어 전체 검색 결과가 아닌 상위 30개의 문서만 평가하였다는 한계가 있다.

본 연구는 규격화된 평가셋에 기반해 국가기록원의 검색 서비스 품질을 측정한 첫 사례로, 질의 유형에 따른 검색도구별 특성을 파악하거나 성능 하락의 원인 규명을 통한 검색 서비스 고도화 이후의 성능 향상 효과를 정량적으로 수치화하는데 활용할 수 있다. 특히 연구 결과물의 공개를 통해 향후 국가기록물 검색서비스 개선을 위한 많은 연구에서 활용되길 기대한다.

참고문헌

- 국가기록원 (2014). 2013년 국가기록 백서. 대전: 국가기록원.
- 국가기록원 (2015). 2014년 국가기록 백서. 대전: 국가기록원.
- 국가기록원 (2021). 2021년 국가기록원 주요통계연보. 대전: 국가기록원.
- 김장환, 이은별 (2015). 국회기록정보 통합관리시스템 개발 방향에 관한 연구. 한국기록관리학회지, 15(2), 103-136.
<https://doi.org/10.14404/JKSARM.2015.15.2.103>
- 김지현 (2012). 국가기록원 기록정보서비스에 대한 이용자 인식에 관한 연구. 한국기록관리학회, 12(1), 167-187.
<https://doi.org/10.14404/JKSARM.2012.12.1.167>
- 김현희 (2007). FRBR 모형 기반 서지검색시스템의 검색 효율성 평가 연구. 한국도서관·정보학회지, 38(3), 223-246.
<https://doi.org/10.16981/kliss.38.3.200709.223>
- 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, 7-44.
<https://doi.org/10.23035/kaics.2020.1.11.007>
- 박미성 (2009). INEX Book Search 트랙의 실험 고찰. 한국도서관·정보학회지, 40(4), 199-225.
<https://doi.org/10.16981/kliss.40.4.200912.199>
- 박민수, 서은경 (2012). 표준 기록관리시스템에 대한 사용자 만족도 분석. 한국기록관리학회지, 12(1), 189-207.
<https://doi.org/10.14404/JKSARM.2012.12.1.189>
- 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 통합검색시스템 설계 및 구현. 한국기록관리학회지, 14(1), 101-124.
<https://doi.org/10.14404/JKSARM.2014.14.1.101>
- 박소연, 이준호 (2008). 주요 검색 포털들의 통합 검색 서비스 비교 평가. 한국도서관·정보학회지, 39(1), 265-278.
<https://doi.org/10.16981/kliss.39.1.200803.265>
- 박소연, 이준호, 전지운 (2006). 지식 검색 서비스 개선을 위한 문서의 적합도 및 신뢰도 분석. 한국문헌정보학회지, 40(2), 299-314.
- 설문원 (2022). 국가기록원의 기록제공서비스에 관한 탐색적 연구. 한국기록관리학회지, 22(3), 103-124.
<http://dx.doi.org/10.14404/JKSARM.2022.22.3.103>
- 이경남 (2013). 표준 기록관리시스템 검색 기능 평가. 기록학연구, 37, 273-305.
<https://doi.org/10.20923/kjas.2013.37.273>
- 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>
- 이상진, 정재승 (2011). 검색엔진 최적화(SEO) 기법을 활용한 전자무역 마케팅 성과에 관한 실증연구. 통상정보연구, 13(1), 3-28.
<https://doi.org/10.15798/kaici.13.1.201103.3>
- 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>
- 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>
- 장인호 (2011). 온톨로지 기반 법률 검색시스템의 구축 및 평가에 관한 연구. 한국문헌정보학회지, 45(2), 345-366.
<https://doi.org/10.4275/KSLIS.2011.45.2.345>
- 장희정 (2012). 국가기록관 웹사이트의 평가에 관한 연구. 한국기록관리학회지, 12(2), 51-70.
<https://doi.org/10.14404/JKSARM.2012.12.2.051>
- Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys, 34(1),

1-47. <https://doi.org/10.1145/505282.505283>

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)

Chang, InHo (2011). Developing and evaluating an ontology-based legal retrieval system. Journal of the Korean Society for Library and Information Science, 45(2), 345-366.

<https://doi.org/10.4275/KSLIS.2011.45.2.345>

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>

Kim, Jang-hwan & Lee, Eun Byol (2015). A Study on the Development of the National Assembly Archives and Records Integrated Management System. Journal of Korean Society of Archives and Records Management, 15(2), 103-136. <https://doi.org/10.14404/JKSARM.2015.15.2.103>

Kim, Jihyun (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>

Kim, Hyun Hee (2007). An Experimental Study on the Retrieval Efficiency of the FRBR Based Bibliographic Retrieval System. Journal of Korean Library and Information Science Society, 38(3), 223-246.

<https://doi.org/10.16981/kliss.38.3.200709.223>

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, Hyojin & Kim, Jihyun (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, Kyung Nam (2013). Evaluation of search functions of the standard records management systems. The Korean Journal of Archival Studies, 37, 273-305. <https://doi.org/10.20923/kjas.2013.37.273>

Lee, Sang Gin & Chung, Jason (2011). An Empirical Study on the Marketing Performance of e-Tradeusing Search Engine Optimization. International Commerce and Information Review, 13(1), 3-28.

<https://doi.org/10.15798/kaici.13.1.201103.3>

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>

National Archives of Korea (2014). National Archives of Korea White Paper 2013. Daejeon: National Archives of Korea.

National Archives of Korea (2015). National Archives of Korea White Paper 2014. Daejeon: National Archives of Korea.

National Archives of Korea (2021). National Archives of Korea Statistical Yearbook 2021. Daejeon: National Archives of Korea.

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, 7-44.
<https://doi.org/10.23035/kaics.2020.1.11.007>
- Park, Mi-Sung (2009). Task Review of INEX Book Search Track. *Journal of Korean Library and Information Science Society*, 40(4), 199-225. <https://doi.org/10.16981/kliss.40.4.200912.199>
- Park, Minsoo & 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, Min Soo & Seo, Eun-Gyoung (2012). Analyzing User Satisfaction of the Standard Records Management System. *Journal of Korean Society of Archives and Records Management*, 12(1), 189-207.
<https://doi.org/10.14404/JKSARM.2012.12.1.189>
- 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, Soyeon, Lee, Joon-Ho & Jeon, Jiwoon (2006). Evaluation of the documents from the web-based question and answer service. *Journal of the Korean society for library and information science*, 40(2), 299-314.
- Park, Soyeon & Lee, Joon-Ho (2008). Comparative evaluation of the unified search services provided by major Korean search portals. *Journal of Korean library and information science society*, 39(1), 265-278. <https://doi.org/10.16981/kliss.39.1.200803.265>
- Seol, Moon-Won (2022). An Exploratory Investigation of Archival Reference Services in the National Archives of Korea. *Korean Society of Archives and Records Management*, 22(3), 103-124.
<http://dx.doi.org/10.14404/JKSARM.2022.22.3.103>