

발 간 등 록 번 호

11-1371029-000269-01



대한민국 웹자원 아카이빙(OASIS) 개선 및 활용성 강화 방안 연구

2024. 12.



문화체육관광부
국립중앙도서관

제 출 문

국립중앙도서관장 귀하

본 보고서를

「대한민국 웹자원 아카이빙(OASIS) 개선 및 활용성 강화 방안 연구」 과제의

최종보고서로 제출합니다.

2024. 12.

-
- | | |
|-------------|------------------------|
| ■ 위탁연구기관명 : | 한남대학교 산학협력단 |
| ■ 연구책임자 : | 박성희 (한남대학교 문헌정보학과 부교수) |
| ■ 공동연구원 : | 이만희 (한남대학교 컴퓨터공학과 교수) |
| | 김진섭 (아르고넷(주)) |
| | 박지우 (아르고넷(주)) |
-

이 연구는 2024년도 국립중앙도서관의 연구개발비로 수행되었으며, 본 연구에서 제시된 정책적 대안이나 의견 등은 국립중앙도서관의 공식 의견이 아니라 본 연구진의 개인 견해를 밝힙니다.

목 차

I. 연구의 개요

1.1 연구 배경 및 목적	3
1.2 연구 내용 및 방법	3
1.2.1 연구 내용	3
1.2.2 연구 방법	4
1.3 연구 과정: 연구 수행 절차	4

II. 웹 아카이빙 국제 동향 조사

2.1 국제 웹 아카이브 법제도 동향	9
2.1.1 IIPC 주요 회원국 웹 아카이브 제도 운영 사례	9
2.1.2 요약 및 시사점	32
2.2 국제 웹 아카이빙 기술 동향	32
2.2.1 수집(Capture)	33
2.2.2 재생과 색인(playback+indexing)	38
2.2.3 분석(Analysis)	42
2.2.4 웹 아카이빙 서비스(Service)	43
2.3 웹 아카이빙 우수 개선 사례	49
2.3.1 룩셈부르크 국립도서관 개선 사례	49
2.3.2 미국 의회도서관 개선 사례	51
2.3.3 우수 개선 사례 시사점	54
2.4 웹 아카이브 데이터 서비스 우수 개발 사례	54
2.4.1 Internet Archive	54
2.4.2 데이터 서비스 우수 개발 사례 시사점	58

III. 웹자원 아카이빙(OASIS) 시스템 데이터 현황 진단과 분석

3.1 오아시스 실무자 의견 조사	63
3.1.1 면담 조사 설계	63
3.1.2 면담 조사 분석 결과	64
3.1.3 요약 및 시사점	69

3.2 웹 아카이빙 전문가 의견 조사	70
3.2.1 면담 조사 설계	70
3.2.2 면담 조사 분석 결과	71
3.2.3 요약 및 시사점	80
3.3 오아시스 누리집 현황 진단	83
3.3.1 사용성 평가 설계	83
3.3.2 사용성 평가 결과	83
3.3.3 전반적 사용 소감 요약 및 시사점	94
3.4 우수 운영 사례 분석	96
3.4.1 미국 의회도서관(LC)	96
3.4.2 우수 운영 기관(미국 의회도서관) 사례 분석의 시사점	102
3.5 오아시스 누리집 성능 측정	102
3.5.1 성능 측정 도구	102
3.5.2 GTmetrix를 이용한 오아시스 누리집 성능 분석	104
3.5.3 오아시스 누리집 GTmetrix 성능 지표 요약 및 비교	119

IV. 시스템 개선 방안 및 프로토타입

4.1 오아시스 시스템 시사점 유형화 및 개선 방안	123
4.1.1 오아시스 누리집 시사점 유형화	123
4.2 오아시스 개선 프로토타입 시범적용	131
4.2.1 오아시스 누리집 프로토타입 시범 적용 논의 결과	131

V. 오아시스 개선 및 활용성 강화 추진 전략

5.1 오아시스 단기 기술 고도화 전략	143
5.1.1 수집 기술 고도화 방안	143
5.1.2 색인 및 검색 기술 고도화 방안	151
5.1.3 재생 기술 고도화	156
5.1.4 데이터 품질 보증(QA: Quality Assurance) 기술	158
5.1.5 데이터 서비스 고도화	160
5.2 시스템 증장기 개선(migration) 전략	163
5.2.1 기존 시스템 개선 사례	163
5.2.2 개선 방안(1안): 웹 데이터 수집·보존 개선을 위한 단기 기술 고도화	166
5.2.3 개선 방안(2안): 새로운 경량 구조(마이크로서비스) 도입	167

5.2.4 개선 방안(3안): 새로운 웹 아카이브 시스템 구조로의 전면적인 마이그레이션(안)	168
5.3 협력 및 홍보 방안	169
5.3.1 우리나라 관련 웹자원의 수집·보존 및 공유 강화를 위한 국제협력 방안	169
5.3.2 대국민 이용 활성화를 위한 오아시스 홍보 방안 제안	170

Ⅴ. 오아시스 비전 및 목표 수립

6.1 오아시스 비전 및 목표 수립	173
6.2 비전 수립 근거	173
6.2.1 우수 운영 기관의 비전과 전략 분석	173
6.2.2 국내외 웹 아카이빙 서비스 비전의 시사점	176
6.3 오아시스 비전 및 목표, 추진 전략	176

Ⅵ. 과제별 세부 추진 방안

7.1 정책 과제 및 실행 과제별 세부 추진 방안	181
7.2 정책 과제 및 실행 과제의 연차별 예산 소요 내역	196
7.2.1 추진 전략별 소요 예산	196

Ⅷ. 결론

8.1 제언	201
참고문헌	204
참고 웹사이트	204
부록A 오아시스 누리집 사용성 평가 결과 화면 모음	207
부록B GTmetrix 성능 지표 해설	233

표 목차

<표 1-1> 연구 수행 절차	5
<표 2-1> 호주 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거	11
<표 2-2> 캐나다 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거	17
<표 2-3> 프랑스 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거	23
<표 2-4> 독일 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거	25
<표 2-5> 일본 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거	27
<표 2-6> 뉴질랜드 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거	28
<표 2-7> 영국 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거	30
<표 2-8> 네덜란드 웹 아카이브 관련 납본제도의 특징 요약 및 참고자료	31
<표 2-9> 미국 웹 아카이브 관련 보관제도의 특징 요약 및 법적 근거	32
<표 2-10> 하드웨어 비교 - 이전 전 및 이전 후	50
<표 3-1> 실무자 심층 면담 질문 및 내용	64
<표 3-2> 오아시스 실무자 그룹: 심층 면담 질문 및 내용	64
<표 3-3> 실무자 심층 면담 주요 조사 결과	64
<표 3-4> 전문가 심층 면담 대상자	71
<표 3-5> 웹 아카이빙 학계 연구계 전문가 그룹: 면담 질문 및 내용	71
<표 3-6> 전문가 심층 면담 주요 조사 결과	72
<표 3-7> GTmetrix를 이용한 오아시스 누리집 성능 지표 비교 평가	119
<표 5-1> 하드웨어 비교 - 이전 전 및 이전 후	165
<표 5-2> 단일 아키텍처와 마이크로 서비스 아키텍처 비교(장단점)	168
<표 6-1> 오아시스 중장기 계획 추진 전략	177
<표 6-2> 오아시스 중장기 계획 정책 과제·실행 과제	177
<표 7-1> 오아시스 개선 및 활용성 강화 추진 전략별 소요 예산	196
<표 7-2> 오아시스 개선 및 활용성 강화 추진 전략별 정책 과제·실행 과제의 소요 예산	197

그림 목차

<그림 2-1> 호주 국립도서관 납본제도 소개 웹페이지	9
<그림 2-2> 연방 입법 등록부 웹페이지 (저작권법, 1968)	11
<그림 2-3> 호주 국립도서관에 관한 납본제도를 규정한 3부-호주 국립도서관(195CA - 195CJ)	12
<그림 2-4> 국제무역기구 TRIPS 협정 온라인 가이드	13
<그림 2-5> 호주 국가 edeposit 웹페이지	14
<그림 2-6> 캐나다 정의법(캐나다 연방법령 및 규칙) 웹사이트에 있는 캐나다도서관·기록관법 (S.C. 2004. c.11)	15
<그림 2-7> 프랑스 국립도서관 웹사이트에 안내되어 있는 웹 법적 납본 안내	20
<그림 2-8> Peter Stirling, et.al. 프랑스 디지털 납본제도에 대한 소개 및 안내자료, IFLA 2011, June 1, 2011	23
<그림 2-9> 독일 국립도서관법	24
<그림 2-10> 저작권의 예외 형태로 허용되는 사용을 규정한 온라인 법률 조회사이트	24
<그림 2-11> 독일 납본제도 법률	25
<그림 2-12> 법적 납본에 대해 설명하는 뉴질랜드 국립도서관 웹 아카이브 웹사이트	28
<그림 2-13> 디지털 법적 납본 10주년 기념 블로그	30
<그림 2-14> 오픈 소스 Heritrix3 Github 화면	33
<그림 2-15> 오픈 소스 Browsertrix Crawler 메인 화면	34
<그림 2-16> 오픈 소스 Pywb 메인 화면	35
<그림 2-17> brozzler 메인 화면	36
<그림 2-18> HTTrack 프로젝트 메인 화면	37
<그림 2-19> 오픈소스 OpenWayback 메인 화면	38
<그림 2-20> 오픈소스 SolrWayback 메인 화면	39
<그림 2-21> 오픈소스 OutbackCDX 메인 화면	40
<그림 2-22> 오픈소스 Web Archive Discovery의 메인 화면	41

<그림 2-23> 오픈소스 Jupyter Notebooks 메인 화면	43
<그림 2-24> Internet Archive의 Archive-It 메인 화면	44
<그림 2-25> 웹 아카이빙 서비스 MirrorWeb의 메인 화면	45
<그림 2-26> Webrecorder의 Conifer 메인 화면	46
<그림 2-27> Archiveweb.page 메인 화면	47
<그림 2-28> Browsertrix 메인 화면	48
<그림 2-29> IIPC 블로그에 게시된 룩셈부르크 국립도서관 웹 아카이브 개선 사례	49
<그림 2-30> BnL의 웹 아카이빙 시스템 마이그레이션(Pywb + OutbackCDX 워크플로)	51
<그림 2-31> LC블로그에 게시된 LC 웹 아카이브 개선 사례	52
<그림 2-32> 미국 LC의 새로운 웹 아카이브 액세스 시스템의 구조(재생SW(Pywb), 인덱스 서버(OutbackCDX) 및 WARC 저장소의 처리 흐름도)	53
<그림 2-33> 2023년 IIPC 총회의 온라인 세션에서 저자가 TrendMachine에 대해 발표하는 장면	55
<그림 2-34> 웹페이지 회복력에 대한 수학적 모델(시그모이드 함수)	56
<그림 2-35> 웹페이지 회복력 가시화 예시 화면	56
<그림 2-36> GitHub에 오픈 소스로 공개된 Internet Archive의 NewSum	57
<그림 2-37> TV 뉴스 요약 서비스 newsum docker image 실행 화면	58
<그림 3-1> 미국 의회도서관 웹 아카이브가 제공하는 보기 메뉴(List, Gallery, Grid, Slideshow)	96
<그림 3-2> 미국 의회도서관 웹 아카이브 슬라이드쇼 보기 화면	97
<그림 3-3> 미국 의회도서관 웹 아카이브 자동 음성 생성 서비스	97
<그림 3-4> 미국 의회도서관 웹 아카이브 공유 기능	98
<그림 3-5> 미국 의회도서관 웹 아카이브 저작권 및 접근 권한 도움말 제공	99
<그림 3-6> 미국 의회도서관 웹 아카이브 인용 기능	100
<그림 3-7> 미국 의회도서관 웹 아카이브 추천 기능	100
<그림 3-8> 미국 의회도서관 웹 아카이브 메타데이터 보기 기능	101
<그림 3-9> 웹 브라우저 개발자 도구(F12) > 성능 평가 도구(Lighthouse)를 통해 오아시스 메인	

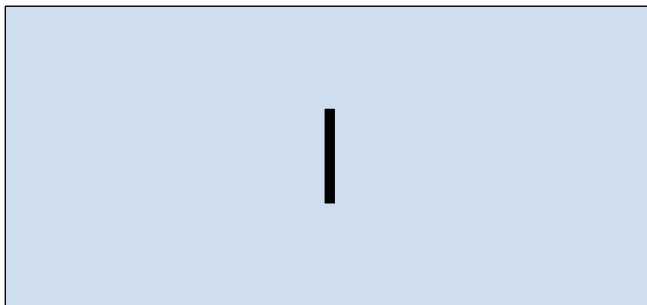
웹페이지(https://www.nl.go.kr/oasis/) 분석	103
<그림 3-10> PageSpeed Insight를 통해 오아시스 메인 웹페이지(https://www.nl.go.kr/oasis/) 분석	104
<그림 3-11> 오아시스 메인페이지 성능 분석 화면	105
<그림 3-12> 오아시스 검색메뉴페이지 성능 분석 화면	106
<그림 3-13> 오아시스 검색 메뉴 중 KDC 주제별 웹페이지 성능 분석 화면	107
<그림 3-14> 오아시스 검색 > 초성 웹페이지 성능 분석 화면	108
<그림 3-15> 오아시스 검색 > ‘월드컵’ 검색 성능 분석 화면	109
<그림 3-16> 오아시스 컬렉션 메뉴 웹페이지 성능 분석 화면	110
<그림 3-17> 오아시스 재난 아카이브 메뉴 웹페이지 성능 분석 화면	111
<그림 3-18> 소셜미디어 메뉴 웹페이지 성능 분석 화면	112
<그림 3-19> 웹트렌드 메뉴 웹페이지 성능 분석 화면	113
<그림 3-20> 컬렉션-웨이백 웹페이지 성능 분석 화면	114
<그림 3-21> 웨이백 재생-성공-금빛 웹사이트 성능 분석 화면	115
<그림 3-22> 웨이백 재생-월드컵파크 웹사이트 성능 분석 화면	116
<그림 3-23> 웨이백 재생_월드컵파크_20160629_서브 메뉴_하늘공원이용 웹페이지 성능 분석 화면	117
<그림 3-24> 연합뉴스(yna) 웨이백 재생 캘린더 화면 성능 분석 화면	118
<그림 5-1> 오픈 소스 Heritrix3 Github 화면	144
<그림 5-2> 오픈 소스 Browsertrix Crawler 메인 화면	145
<그림 5-3> Browsertrix-crawler docker image가 docker desktop에서 사용되고 있는 설치 화면	146
<그림 5-4> CLI 창에서 한남대학교 홈페이지를 수집하도록 명령 입력	146
<그림 5-5> Docker desktop속에 volume으로 crawls 폴더 밑에 수집된 웹페이지	147
<그림 5-6> WARC 파일뷰어인 Replay Webpage 실행 화면 (WARC 파일 로드 전)	148
<그림 5-7> WARC 파일뷰어인 Replay Webpage 실행 화면 (WARC 파일(test.wacz) 로드 후)	148

<그림 5-8> TEST.WACZ 파일 상세정보 제시 화면 (WACZ파일명, 수집일시, 웹페이지 제목 및 URL)	149
<그림 5-9> 수집된 WACZ 파일 재생 화면 (한남대학교 홈페이지: https://www.hannam.ac.kr) 화면	149
<그림 5-10> 오픈 소스 Pywb 메인 화면	150
<그림 5-11> Selenium 프로젝트 공식 홈페이지 화면	151
<그림 5-12> 오픈소스 SolrWayback 메인 화면	153
<그림 5-13> Elastic Search 프로젝트 공식 홈페이지 화면	154
<그림 5-14> 검색 기능에 Elastic Search를 사용하고 있는 인덱스 서버(The National Archive)	155
<그림 5-15> DB-Engine에서 제공하는 검색엔진 대중성 랭킹에서 1위인 Elastic Search ...	155
<그림 5-16> 오픈 소스 OpenWayback 메인 화면	157
<그림 5-17> 오픈 소스 Pywb 메인 화면	158
<그림 5-18> Internet Archive의 ChatGPT를 활용한 뉴스 요약 서비스 화면	162
<그림 5-19> Internet Archive의 ChatGPT를 활용한 뉴스 요약 서비스 화면(ChatGPT OpenAPI 허용 이용량 초과 에러 화면)	162
<그림 5-20> 웹페이지 회복력 가시화 예시 화면	163
<그림 5-21> 미국 LC의 새로운 웹 아카이브 액세스 시스템의 구조(재생SW(Pywb), 인덱스 서버(OutbackCDX) 및 WARC 저장소의 처리 흐름도)	164
<그림 5-22> BnL의 웹 아카이빙 시스템 마이그레이션(Pywb + OutbackCDX 워크플로)	166
<그림 5-23> 현행 단일구조 오아시스 개선 운영 개념도(방안1)	167
<그림 5-24> 단일 아키텍처 대 마이크로서비스 아키텍처 비교	167
<그림 5-25> 현행 단일구조(OASIS)와 경량 마이크로서비스 구조병행 운영을 통한 점진적 마이그레이션(방안2)	168
<그림 5-26> 새로운 웹 아카이브 시스템 구조로의 전면적인 마이그레이션(방안3)	169
<그림 부록A-1> 웹사이트 검색 화면	207
<그림 부록A-2> 웨이백 재생 무한 로딩	207
<그림 부록A-3> 썸네일 영상이 없을 때 표시되는 디폴트 이미지(OASIS로고)	208

<그림 부록A-4> 썸네일 영상이 없을 때 표시되는 디폴트 이미지(문서이미지)	208
<그림 부록A-5> 미리보기 영상이 흰 바탕의 영상과 오아시스 영상이 통일되지 않음	209
<그림 부록A-6> 미리보기가 제시되지 않은 이미지 결과 제시 화면	209
<그림 부록A-7> 전용 뷰어 프로그램 설치 안내 화면	210
<그림 부록A-8> 동영상에 대해 '웹사이트 보기' 버튼	210
<그림 부록A-9> 국립중앙도서관 내 자료 이용 가능 안내 메시지	211
<그림 부록A-10> 보존된 웹페이지의 이미지 로딩을 실패한 웨이백 재생 화면	211
<그림 부록A-11> 컬렉션 > 2022 카타르월드컵	211
<그림 부록A-12> 수집 건수 (확대영상) 자료유형별 아이콘이 너무 작아 구분이 잘 안 됨	212
<그림 부록A-13> 보기 형식 중 12개 자료보기 화면	212
<그림 부록A-14> 링크 표시가 안 되어있는 웹사이트 URL	212
<그림 부록A-15> 검색 > 초성보기 > 'ㄱ'으로 브라우징한 결과 화면	213
<그림 부록A-16> 검색 > KDC주제별보기> 철학에서 잡지, 사전은 아카이브 자료가 소수(0건)임	213
<그림 부록A-17> 아카이빙된 정보테크 웹사이트 정보	214
<그림 부록A-18> URL(정보테크.한국) 링크가 깨어진 웹사이트	214
<그림 부록A-19> 웹사이트가 Error로 나오는 needles.co.kr 웹사이트	215
<그림 부록A-20> '월드컵' 검색결과 자료유형으로 웹사이트, 누리집, 웹페이지, 뉴스, 웹사이트로 구분되어 있음	215
<그림 부록A-21> '월드컵' 검색결과 자료유형은 웹사이트, 누리집, 웹페이지, 뉴스, 웹사이트로 구분되지 않고 full-text로만 제시되고 있음	216
<그림 부록A-22> 'SNDSOFT' 검색결과 자료유형은 웹사이트(2)로 검색된 결과	216
<그림 부록A-23> 컬렉션에 대해 'SNDSOFT' 검색실행, 전체 0건 검색된 결과	217
<그림 부록A-24> 재난아카이브에 대해 'SNDSOFT' 검색 실행, 전체 0건 검색된 결과	217
<그림 부록A-25> 웹트렌드 분석에 대해 'SNDSOFT' 검색 실행, 전체 1건 검색된 결과	218
<그림 부록A-26> '검색'메뉴에서 작은 돋보기 아이콘과 비교되게 넓은 면적의 상세검색 버튼	218

<그림 부록A-27> 야후 검색 창외 경우 돋보기는 검색을 의미하는 아이콘이고 버튼 없이 엔터로 검색 실행	218
<그림 부록A-28> 야후와 유사하게 돋보기 아이콘은 앞쪽에 검색레이블 대신 사용되었고, 검색실행은 엔터(enter)나 아래쪽 'Google 검색' 버튼을 눌러 실행	219
<그림 부록A-29> 상세검색에서 발행년도 날짜 선택 컨트롤	219
<그림 부록A-30> Windows11 날짜 선택 컨트롤(월 선택)	219
<그림 부록A-31> Windows11 날짜 선택 컨트롤(연도 선택)	219
<그림 부록A-32> 12개씩 보기 화면	220
<그림 부록A-33> 24개씩 보기 화면	220
<그림 부록A-34> 컬렉션 > 3.1운동 및 임시정부 100주년 기념 전체보기 화면 및 출처 (https://www.together100.go.kr)	221
<그림 부록A-35> https://www.together100.go.kr 접속 오류 화면	221
<그림 부록A-36> 주제별 컬렉션 선택 시 원본 이미지 로딩시간 같은 미리보기 이미지 로딩시간	222
<그림 부록A-37> 한국영화 <출처> 한국영화 100년 기념사업 홈페이지 정보	222
<그림 부록A-38> 한국영화 100년 기념사업 홈페이지 링크 접속 결과(503 Service Temporarily Unavailable)	223
<그림 부록A-39> 컬렉션 메뉴에서 검색 버튼(돋보기 모양)이 동작하지 않음	223
<그림 부록A-40> 재난아카이브 > 콘텐츠목록 > 이미지 메뉴 선택 시 동영상 자료 결과제시	223
<그림 부록A-41> 재난아카이브 > 콘텐츠목록 > 동영상 메뉴 선택 시 이미지 자료 결과제시	224
<그림 부록A-42> 재난 분류 중 직관적으로 이해가 어려운 해외재난 분류 체계(해외재난>해외)	224
<그림 부록A-43> 2022년이 최신이며 엠바고에 대한 부가 설명이 없는 화면	224
<그림 부록A-44> 접근 금지된 질병관리본부 URL	225
<그림 부록A-45> 재난아카이브 자료 선택 시 화면이 두 번 그려지면서 깜빡이는 현상	225
<그림 부록A-46> 재난아카이브 분류(재난유형별) 해외 3건	226

<그림 부록A-47> 재난아카이브 분류(지역별) 해외 90건	226
<그림 부록A-48> 소셜미디어 정부기관 설명에서 URL이 링크화가 되어 있지 않음	226
<그림 부록A-49> 가입일보다 최초 트윗일이 빠른 사례 1(합천군)	226
<그림 부록A-50> 가입일보다 최초 트윗일이 빠른 사례 2(노사발전재단)	226
<그림 부록A-51> 고용노동부의 공식 트위터를 소개하는 기관 문양의 사진은 선명한 반면, 수집된 트윗에서의 고용노동부의 문양 사진은 선명하지 않고 이미지가 깨져있음	227
<그림 부록A-52> 소셜미디어 > 환경부 > 자세히 보기 URL (링크화되어 있지 않음)	227
<그림 부록A-53> 앞 그림의 환경부 자세히 보기 URL(https://bit.ly/2GY9iue) 복사 웹브라우저 주소창에 붙여넣기	228
<그림 부록A-54> 소셜미디어 > 환경부 트윗 중 하나(1/817 페이지 중 제일 마지막 트윗)로 자세히 보기 링크가 링크화가 안 되어 있음	228
<그림 부록A-55> 앞 그림 URL(http://bitly.kr/P3kIV_pic.twitter.com/IXjlliFAJV)을 복사 붙여넣기 한 결과	229
<그림 부록A-56> '키워드 비중 및 '수집년도 구분' 화면에 가려진 안내 메시지('2016년부터 ...')	229
<그림 부록A-57> 중복된 라디오 버튼	230
<그림 부록A-58> 웹트렌드 도움말 설명	230
<그림 부록A-59> 단편적인 정보를 포함한 키워드 비중 도움말	230
<그림 부록A-60> 웹트렌드 키워드 검색 결과 화면	231
<그림 부록A-61> URL링크 접속 실패 화면	231
<그림 부록A-63> 웹트렌드 > 웹트렌드 활용예제 메뉴 선택 시 화면(왼쪽 사이드 메뉴 제공에 주목)	232
<그림 부록A-62> 웹트렌드 > 웹트렌드 분석 메뉴 선택 시 화면(왼쪽 사이드 메뉴 미제공에 주목)	231
<그림 부록A-64> 웹트렌드 > 태그 클라우드 메뉴 선택 시 화면(왼쪽 사이드 메뉴 미제공에 주목)	232



연구의 개요

1. 연구의 개요

1.1 연구 배경 및 목적

- 본 연구의 목적은 2004년부터 수집 서비스를 시작한 대한민국 웹 아카이빙 서비스 오아시스(OASIS: Online Archiving & Searching Internet Sources)가 올해로 20주년을 맞이하면서, 현재 오아시스 시스템의 전반적인 상황을 진단하고 다음 20년을 준비하기 위한 단기, 중장기 계획을 마련할 필요가 있음
- 지난 20년 동안 국립중앙도서관은 오아시스를 통해 국가 지식 디지털 자원으로서의 웹자원을 후대에 계승 보존하려는 목표 하에 웹자원의 수집 보존, 서비스를 진행하여 왔으며, 오아시스 시스템의 웹자원의 양적 성장은 눈에 띄게 증가하였음. 이에 다음 20년은 서비스의 품질을 개선하여 활용도를 높이는 데 좀 더 초점 맞추어 진행할 시기임
- 따라서 본 연구는 웹 아카이빙의 국내외 법적, 제도적, 기술적 동향을 조사하여 국내외 환경 변화를 분석하고, 오아시스 시스템 현황 진단을 통해 개선점을 도출하여 새로운 서비스를 위한 프로토타입 시범 적용, 오아시스 시스템 단, 중장기 개선 방안 도출하고 웹 아카이빙 국제적 협력 체계 구축, 홍보활동 강화 및 활용성 강화 방안을 수립하기 위함

1.2 연구 내용 및 방법

1.2.1 연구 내용

- 대한민국 웹 아카이빙(오아시스) 개선 및 활용성 강화 방안 연구 세부 내용은 다음과 같음.
 - 웹 아카이브 구축 관련 국제 동향 및 국내외 운영 사례 조사
 - 국제인터넷보존컨소시엄(IIPC) 등 웹 아카이브 구축과 관련된 기술적·법적 국제 동향 조사
 - 미국, 영국, 프랑스 등 국내외 주요 웹 아카이브 운영 사례 조사
 - 웹자원 아카이빙(OASIS) 시스템·데이터 현황 진단과 분석
 - 웹사이트 수집(포괄적 수집) 현황 조사 및 분석
 - 오아시스 누리집 서비스 현황 조사 및 분석
 - 웹자원 수집·보존 기술 및 시스템 개선 방안 제안
 - 웹 데이터 수집·보존 개선을 위한 기술 고도화
 - 시스템 중장기 개선 방안 제안
 - 웹 데이터 서비스 개선을 위한 프로토타입 개발 및 시범 적용
 - 신규 데이터 서비스 개발을 위한 프로토타입 개발 및 시범 적용
 - 웹 아카이빙 글로벌 협업 및 홍보 방안 제안 등

- 우리나라 관련 웹자원의 수집·보존 및 공유 강화를 위한 국제 협력 방안 제시
- 대국민 이용 활성화를 위한 오아시스 홍보 방안 제안
- 오아시스 비전 수립, 추진 전략, 세부 추진 과제 도출 및 세부 추진 방안 제시

1.2.2 연구 방법

○ 문헌 조사 연구

- IIPC 회원국 디지털 납본제도 및 저작권법 동향 조사
- 국내외 웹 아카이빙 우수 운영 사례 자료
- 웹 아카이빙 관련 기술 도입 사례

○ 면담 조사 (심층 인터뷰 및 포커스 그룹 인터뷰)

- 조사 대상: 웹 아카이브 이용자 47명(유효 표본 47건), 오아시스 실무자 6명
- 조사 방법: 온라인 심층 인터뷰, 포커스 그룹 인터뷰 기법
- 조사 내용: 오아시스 위상, 오아시스 수집, 데이터, 색인, 검색 및 서비스 요구사항 및 문제점, 개선 방안

○ 전문가 자문

- 자문 대상: 웹 아카이빙 분야 전문가 3명(포괄적 수집 전문가, 오아시스 시스템 전문가, 재난아카이브 전문가 등 문헌정보학 교수들로 구성)
- 자문 내용: 오아시스 현황 진단과 개선 및 활용성 방안의 방향, 비전, 추진 전략, 세부 추진 과제 적절성 검토

1.3 연구 과정: 연구 수행 절차

○ 본 연구는 다음과 같은 5단계에 따라 과업을 수행하도록 함

- 1단계(국내외 동향 분석) : 문헌 연구와 현황 자료 조사를 통해 국제 웹 아카이빙 현황 및 연구에 필요한 법, 제도, 운영 사례, 신기술 개발 및 도입 사례 등을 조사
- 2단계(이용자 및 실무자, 전문가 의견 수렴) : 면담 조사 및 자문회의를 통한 실무자 및 전문가 의견 반영
- 3단계(개선 방안 및 활용성 강화 방안 제시) : 오아시스 개선 방안, 오아시스 비전과 전략 및 중장기 개선 계획 수립을 통한 결과 도출
- 4단계(홍보 방안 및 협력 방안 제시) : 활용성 강화를 위한 대국민 홍보 방안 및 국제적 협력 체계 제시
- 5단계(보고) : 보고회(3회)

<표 1-1> 연구 수행 절차

단계	연구 내용	세부 연구	추진방법
준비 단계	연구 진행 협의	<ul style="list-style-type: none"> 과업 내용의 검토 및 추진 내용 파악 연구 방향 설정 및 연구 진행 방법 협의 	과업 지시서 검토 발주처 의견 청취
1단계 국제 동향 파악	이론적 연구	<ul style="list-style-type: none"> IIPC 회원국 납본법 및 저작권법 조사 주요 IIPC 회원국 웹 아카이빙 정책 분석 ✓ 미국, 영국, 호주, 프랑스 등 웹 아카이빙 정책 등 	문헌 자료 조사
	사례 조사	<ul style="list-style-type: none"> 국내외 주요 웹 아카이빙 시스템 및 운영 현황 사례 분석 ✓ 우수 웹 아카이빙 시스템 분석 ✓ 우수 웹 아카이빙 개선 사례 분석 ✓ 웹 아카이빙 기술 현황 조사 및 분석 	현황 자료 조사
	이용자 면담 조사	<ul style="list-style-type: none"> 오아시스 이용자 대상으로 온라인 면담 조사 실시 (47명) 	이용자 요구 조사 및 오아시스 운영자 면담 전문가 자문
2단계 이용자 및 실무자 의견 수렴	오아시스 수집 및 관리 운영자 인터뷰, 전문가 의견 조사	<ul style="list-style-type: none"> 웹자원 수집 및 오아시스 운영 관리자를 대상 온라인 서면 인터뷰 실시 웹 아카이빙 분야 전문가 자문 	
	오아시스 개선점 도출 및 프로토타입 시범 적용	<ul style="list-style-type: none"> 오아시스 개선 방안 및 시스템 개선 로드맵 제시 신규 서비스 개발을 위한 프로토타입 개발 및 시범 적용 ✓ 오아시스 누리집 추가 서비스 제안 ✓ 오아시스 누리집 디자인 개선 제안 	개선점에 따른 개선 방향 및 실현을 위한 전략 과제 제시
3단계 개선점 도출	오아시스 개선점 도출 및 프로토타입 시범 적용		
4단계 활용성 강화 방안 도출	홍보 방안 및 국제 협력 체계 수립	<ul style="list-style-type: none"> 우리나라 관련 웹자원의 수집·보존 및 공유 강화를 위한 국제 협력 방안 대국민 이용 활성화를 위한 오아시스 홍보 방안 제안 	홍보 방안 제시 및 국제 협력 체계 제안
5단계 보고	연구 보고	착수 보고회, 중간 보고회, 완료 보고회	보고회

II

웹 아카이빙 국제 동향 조사

2. 웹 아카이빙 국제 동향 조사

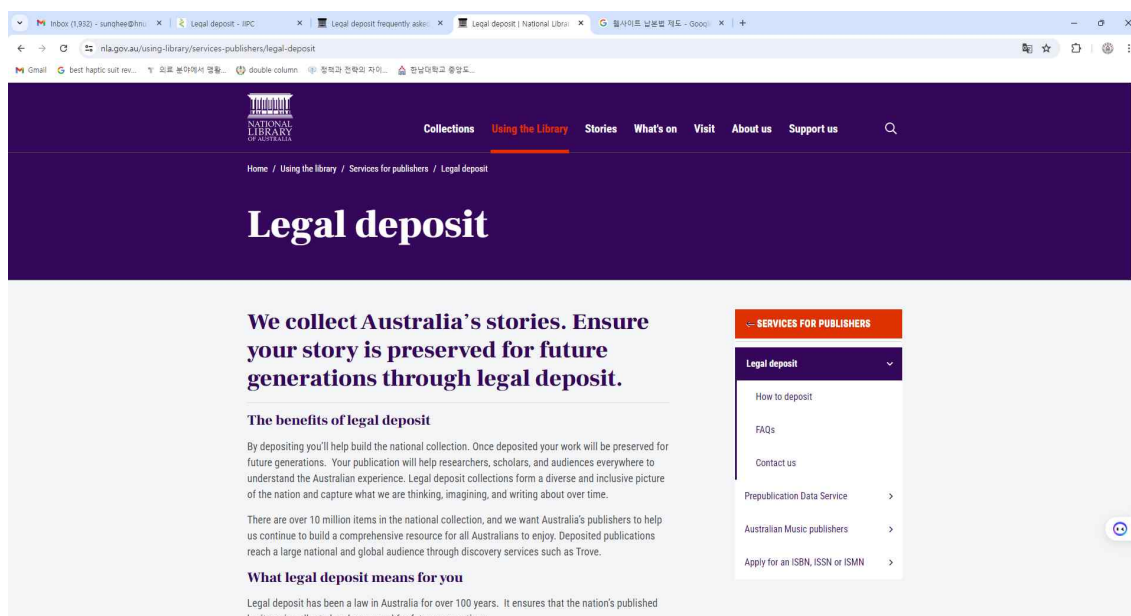
2.1 국제 웹 아카이브 법제도 동향

2.1.1 IIPC 주요 회원국의 웹 아카이브 제도 운영 사례

- 본 절에서는 주요 IIPC 회원국들의 웹 아카이브 납본법 제도의 현황을 살펴보고, 우리나라 웹 아카이브 오아시스에 적용할 수 있을 시사점을 도출함

1) 호주

- 2016년 2월부터 온라인 콘텐츠를 포함한 디지털 자원까지 납본 적용하며, 웹사이트를 포함함
- 호주의 납본제도에 대하여 호주 국립도서관의 누리집에 소개되어 있음
- 출처: <https://www.nla.gov.au/using-library/services-publishers/legal-deposit>



<그림 2-1> 호주 국립도서관 납본제도 소개 웹페이지

- 웹페이지는 다음과 같은 내용들을 포함함
 - 납본제도의 이점(The benefits of legal deposit)
 - 납본제도의 의미(What legal deposit means for you)
 - 납본 방법

- 호주 전역 납본
- 납본 관련 안내자료

○ 납본제도(Legal Deposit) 자주 묻는 질문(FAQ)

- 다음 링크에서 납본제도에 대한 “자주 묻는 질문(FAQ)”을 확인해 볼 수 있음
- 출처 :

<https://www.nla.gov.au/using-library/services-publishers/legal-deposit/legal-deposit-frequently-asked-questions#websites>

○ 다음과 같은 웹사이트 납본 및 보존에 대한 질문이 있음

납본을 통해 어떻게 웹사이트를 수집하는지?

공개 이용 가능한 웹사이트가 납본 수집 대상입니다. 국가 장서에 포함하기로 결정한 발행된 웹사이트를 자동으로 보존합니다. 암호로 보호되거나 유료 기반의 웹 자료는 수집하지 않습니다.

귀하의 공개된 웹사이트를 수집하는 데 어려움이 있다면, 직접적으로 귀하에게 연락할 수도 있습니다. 온라인 자원의 게시자에게 제공하는 공지에서 자세한 정보를 찾을 수 있습니다.

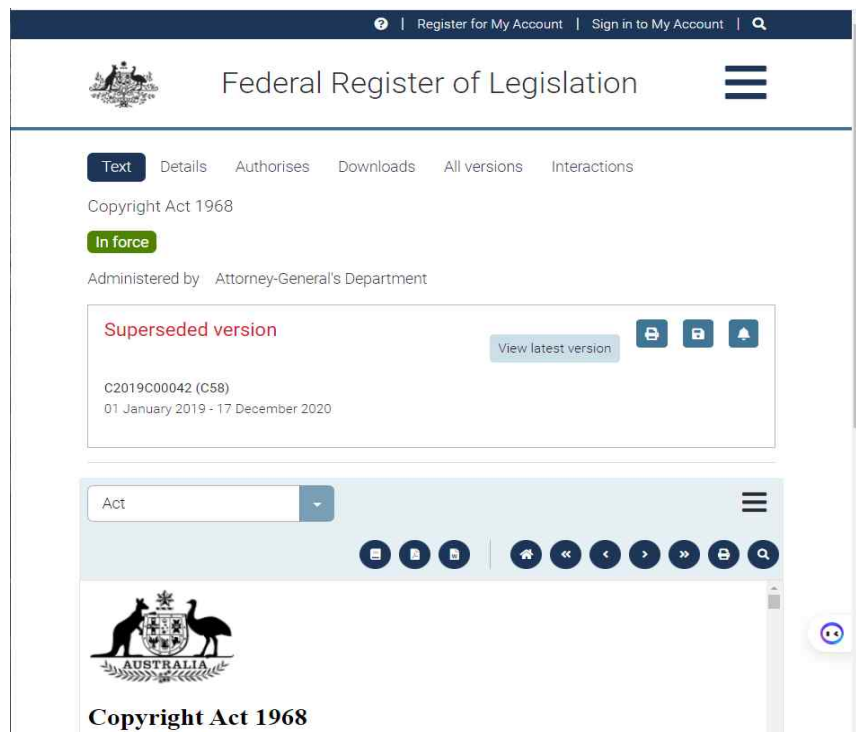
보존된 웹사이트는 호주 웹 아카이브를 통해 대중에게 공개되며, Trove에 보존되면 시간이 지나 변경하거나 사라질 수 있는 콘텐츠를 연구자들이 볼 수 있도록 허용합니다.

웹 보존에 대해 더 자세히 알아보시다.

○ 호주 웹 아카이브 관련 납본제도의 특징 및 법적 근거를 요약하면 다음 <표 2-1>과 같음

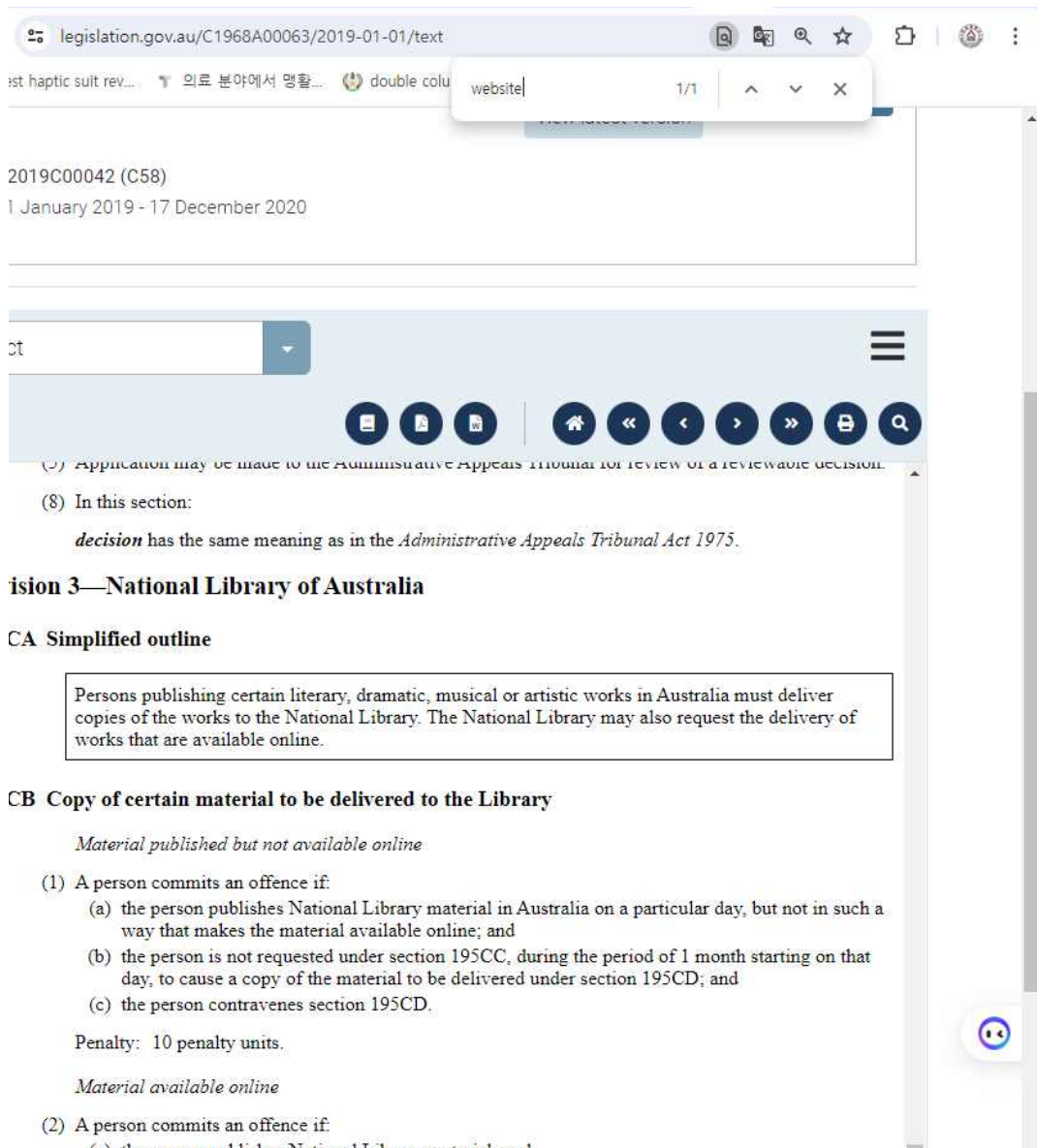
<표 2-1> 호주 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거

	특징	법적 근거
호주	<ul style="list-style-type: none"> 호주의 납본 조항은 1968년 호주 연방 저작권법(ss195CA-195CJ)의 일부를 구성 2016년 2월부터 납본 조항이 온라인 콘텐츠를 포함한 디지털 자료에 적용 온라인 자료가 납본 대상이 되려면 저작권이 해당 법에 따라 존속해야 하며(따라서 .au 웹 도메인으로 제한되지 않음) 1960년 국립도서관법에 명시된 도서관의 기능에 따라 국립도서관 관장은 해당 자료가 국립도서관 장서에 포함되어야 한다고 간주 출판사는 국립도서관 관장이 요청할 때 온라인 자료를 보관해야 함 해당 법에서는 납본 요청이 전자 통신 및 사용자 에이전트(로봇) 요청을 통해 이루어질 수 있다고 명시하고 있음 납본 조항 자체는 접근권을 언급하지 않음. 그러나 1968년 저작권법 제200AB항은 일반적으로 도서관에서 제공하는 종류의 서비스를 포함하여 도서관을 유지 또는 운영할 목적으로 도서관에 대한 저작권 면제를 제공함. 이 면제는 TRIPS¹⁾ 협정 제13조에 따라 이루어져야 함 	<ul style="list-style-type: none"> 납본 FAQ <ul style="list-style-type: none"> 1968년 저작권법(ss.195CA-195CJ)의 납본 조항(<그림 2-2>, <그림 2-3> 참조) 1968년 저작권법(s.200AB)에 따른 도서관 면제 세계무역기구(WTO) TRIPS 협정 제13조(<그림 2-4> 참조)



<그림 2-2> 연방 입법 등록부 웹페이지 (저작권법, 1968)

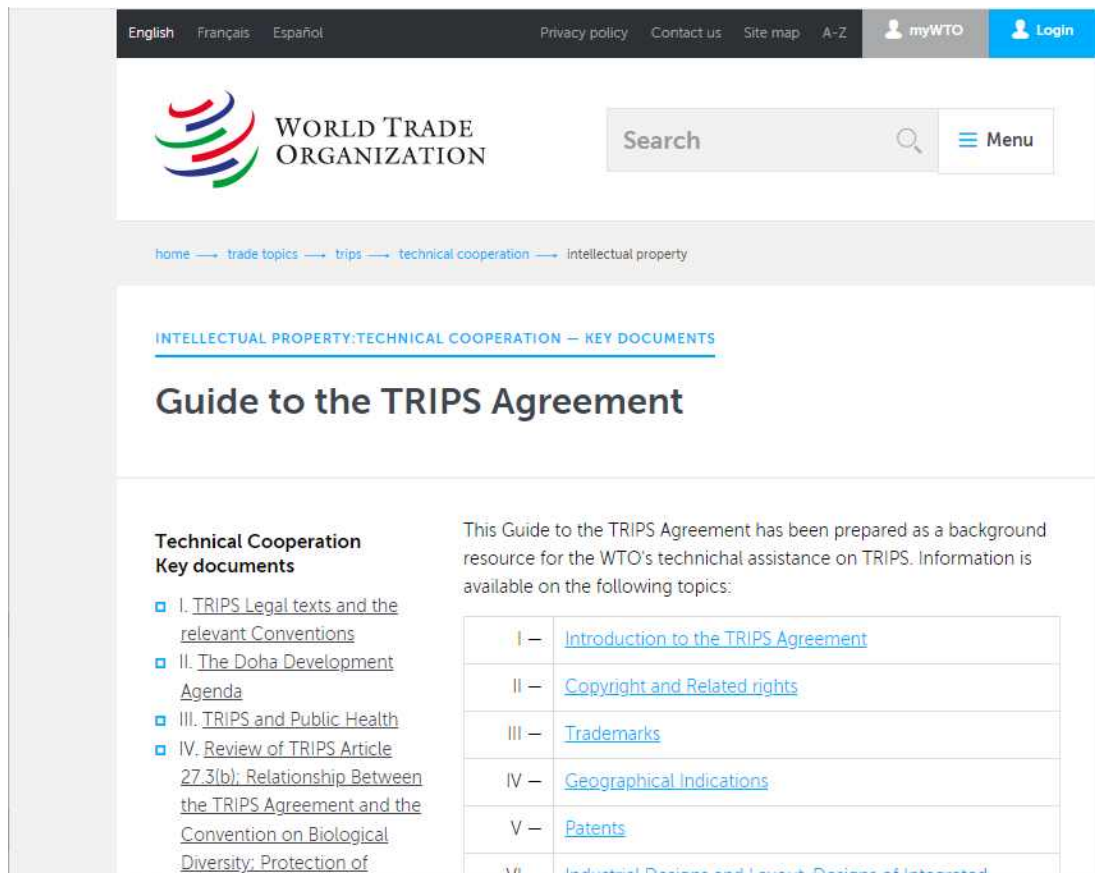
1) WORLD TRADE ORGANIZATION. TRIPS(Trade-Related Intellectual Property Rights): 무역관련 지적재산권. https://www.wto.org/english/tratop_e/trips_e/trips_e.htm



<그림 2-3> 호주 국립도서관에 관한 납본제도를 규정한 3부-호주 국립도서관(195CA - 195CJ)

○ TRIPS 협정 제13조

- 무역관련 지적재산권 협정(TRIPS: Trade-Related Aspects of Intellectual Property Rights)으로 전문과 총 7부 및 73개의 조문으로 구성되어 있으며, 지적재산권의 거의 모든 분야를 망라함. 제1부에서 총론적 규정으로 내국민대우, 최혜국 대우 등의 일반원칙을 다루고, 전통적 의미에서의 지적재산권 자체의 보호에 관한 규정은 제2부에서 다루고 있음
- 출처: WORLD TRADE ORGANIZATION. TRIPS: Trade-Related Aspects of Intellectual Property Rights. https://www.wto.org/english/tratop_e/trips_e/trips_e.htm



<그림 2-4> 국제무역기구 TRIPS 협정 온라인 가이드

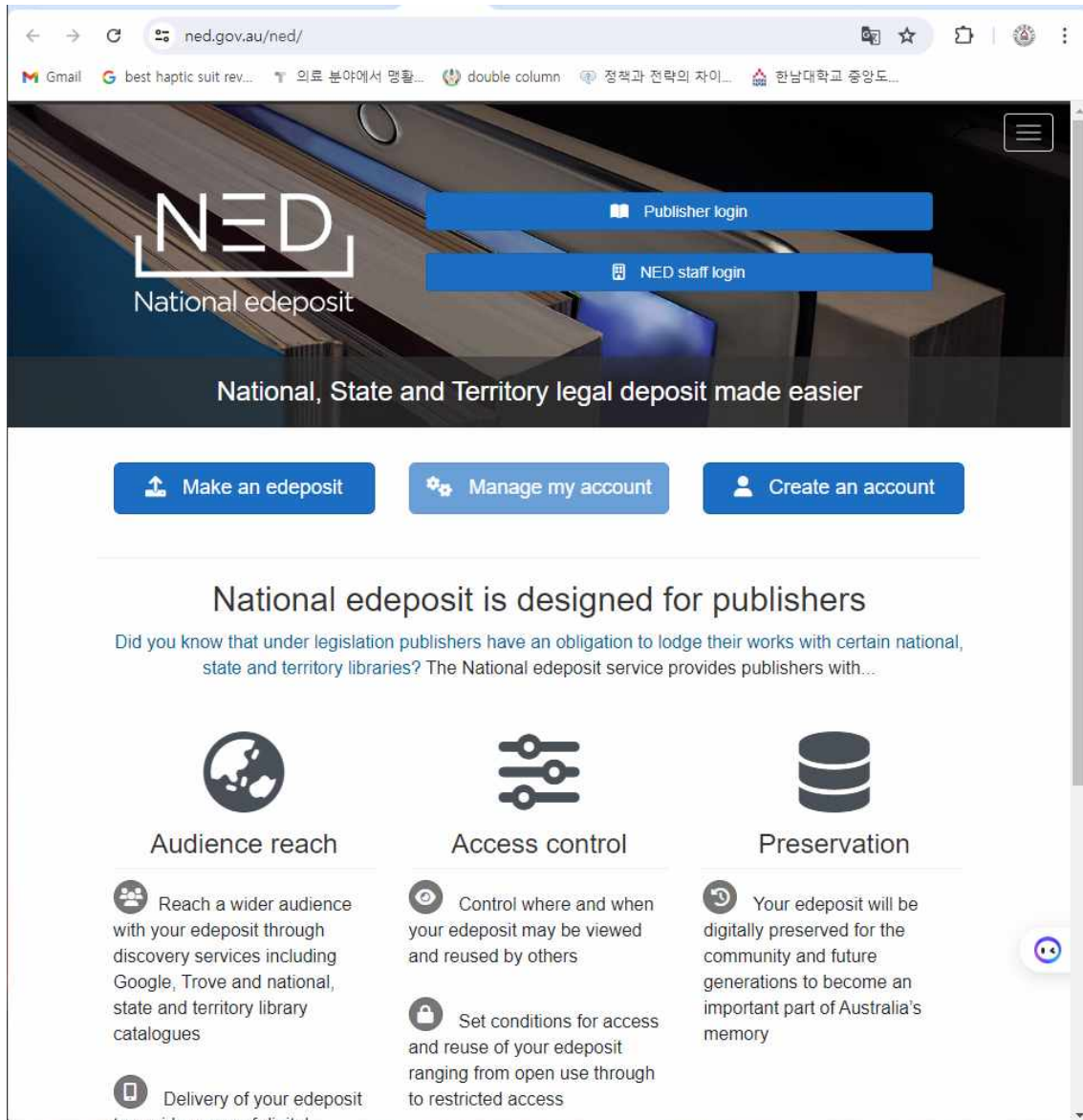
○ edeposit 프로젝트

- 전자납본 웹사이트를 통한 납본서비스 제공
- 출처: <https://ned.gov.au/ned/>
- 호주 출판사 및 개인이 회원가입 후 로그인을 통해 납본 가능
- NED(National E-Deposit)는 다음과 같은 기능 제공
 - ▷ **이용자 접근(Audience Reach)**
 - ▶ 구글, Trove, 국립, 주정부, 지역 도서관 목록 검색 서비스를 통해 더 많은 사람들에게 전달할 수 있음
 - ▶ 더 확대된 디지털 장치에 전자 납본 자료를 전달함
 - ▷ **접근 통제(Access control)**
 - ▶ 전자 납본 자료가 다른 사람에 의해 언제, 어디서 보여지고 사용될지 통제할 수 있음
 - ▶ 공개 이용에서부터 이용 제한까지 전자 납본 자료의 재사용과 이용에 대한 조건을 설정할 수 있음
 - ▶ → (시사점) 이용 및 재사용에 대한 다양하고 세밀한 관리가 가능함

- ▶ 지적재산권은 국립 전자 납본 서비스에 참여할 때 보호됨
- ▶ → 납본자의 지적재산권을 보장

▷ 보존(Preservation)

- ▶ 전자 납본 자료는 중요한 호주의 기억 일부로 보존되어 후대에 전승될 것임



<그림 2-5> 호주 국가 edeposit 웹페이지

2) 캐나다

○ 법적 근거

- 인터넷을 통한 선별 수집(Sampling from Internet)

- ▶ 캐나다도서관·기록관법 8조 2항은 다음과 같이 규정하고 있음

(2) 전(1)항 (a)에서 언급된 권한을 실행하고 보존목적을 위해, 사서 및 기록관리관은 그들이 적합하다고 생각하는 시기와 방식으로 인터넷 혹은 유사매체를 통하여 제한 없이 공중이 접근할 수 있도록 캐나다에 유익한 기록 자료의 대표적인 견본을 수집할 수 있다.²⁾

- ▶ 법률 원문은 다음 웹사이트에서 확인할 수 있음



<그림 2-6> 캐나다 법률정보(캐나다 연방법령 및 규칙) 웹사이트에 있는
캐나다도서관·기록관법 (S.C. 2004. c.11)

출처: <https://laws-lois.justice.gc.ca/eng/acts/L-7.7/index.html>

2) 출처: 국립중앙도서관 도서관연구소, “주요국 도서관법령집(14. 캐나다도서관·기록관법), 도서관연구소 자료집 4, pp. 187-204
<https://oak.go.kr/nl-ir/bitstream/2020.oak/394/1/%EC%A3%BC%EC%9A%94%EA%B5%AD%EB%8F%84%EC%84%9C%EA%B4%80%EB%B2%95%EB%A0%B9%EC%A7%91-1.pdf>

- **납본제**

- ▶ 캐나다도서관·기록관법 10조(1)~(3)항은 다음과 같이 규정하고 있음

법정 납본(Legal Deposit)

출판물의 납본(Deposit of publications)

제10조 (1) 규정에 의하여 캐나다에서 이용되는 출판물을 제작하는 출판업자는 자비로 출판물의 복제본 2부를 사서 및 기록관리관에게 제출해야 한다. - 사서 및 기록관리관은 다음과 같이 그 수령을 확인한다.

- (a) (b)에서 언급되는 것과 다른 경우에는 이용이 가능한 날로부터 7일 이내
- (b) (2)항 (d)에서 규정된 부류에 속한 출판물의 경우, 사서 및 기록관리관으로부터 서면요구를 받고 나서 7일 이내, 또는 그 요청에서 언급한 연장기일 이내

규정(Regulations)

- (2) 장관은 아래와 같은 조항들을 포함하여 이 장의 목적을 위하여 규정을 제정할 수 있다.
 - (a) “출판업자”를 정의하는 것
 - (b) 종이와 다른 매체를 사용하는 출판업자들이 그들의 출판물을 사서 및 기록관리관이 이용할 수 있도록 하기 위해 취해야 하는 모든 조치에 관한 규정
 - (c) 오직 복제본 한 부만이 요구되는 출판물의 부류를 규정하는 것. 그리고
 - (d) (1)항의 의무조항이 오직 사서 및 기록관리관으로부터의 서면요청에 의해서만 적용받는 출판물의 부류를 규정하는 것

소유권(Property)

- (3) 이 장에 의거하여 사서 및 기록관리관에게 제공되는 출판물은 여왕 폐하에게 귀속하며 캐나다도서관·기록관의 소장 장서를 형성한다.
- (4) 이 장 목적을 위하여, 출판물의 모든 재판, 개정판 또는 변형은 별개의 출판물로 간주되어야 한다.

- **지적재산권법**

- ▶ 지적재산권법(R.S.C., 1985, c. C-42)의 예외 조항으로 30조, 5항에 허용규칙을 규정하고 있음
- ▶ 캐나다도서관·기록관법의 제8조(2)항, 제10조(1)항, 제11조(2), 방송법 제2조(1)에 정의된 내용에 대해 예외를 규정함
- ▶ 아래에 제시된 법령의 국문 번역은 본 연구 수행을 목적으로 연구자에 의해 번역되었으므로 법적 효력이 없음
- ▶ 출처: <https://laws-lois.justice.gc.ca/eng/acts/C-42/FullText.html>

30.5 다음은 캐나다도서관·기록관법에 따른 캐나다 사서 및 기록관리관에 대하여는 저작권 침해가 아니다.
(a) 해당 법 제8조(2)항에 따라 보존을 목적으로 대표적인 견본을 수집하여 저작물이나 기타 대상의 복제본을 작성하는 행위
(b) 해당 법 제10조(1)항에 따라 통신을 통해 제공되는 해당 법 제2조에 정의된 출판물 복제본을 보존하는 행위
(c) 해당 법 제11조의 목적을 위해 해당 법 제11조(2)항에 정의된 녹음물의 복제본을 작성하는 행위 또는
(d) 방송법 제2조(1)항에 정의된 방송사업자가 전기통신을 통해 저작물 또는 기타 대상을 공중에게 전달할 때, 다음과 같은 저작물 또는 기타 대상의 복제본을 제작하는 행위

- ▶ 현재 웹 아카이빙에 대한 법적 제도는 모든 형태의 출판물에 대해 납본제도를 2004년부터 시행하고 있으며, 온라인 자료에 대해서는 사서나 기록관리관에게 “선별 수집(8조(2)항)”이라는 개념으로 지적재산권에 대한 예외를 적용하고 있음
- ▶ 10조(2)(b)에 사서나 기록관리관이 간행물에 접근하기 위해 취하는 모든 조치를 존중한다고 하여 패스워드나 기술적 정보 요청도 포함함
- ▶ 하지만, 이용자들의 보존된 자료에 대한 접근에 대해서는 명시하고 있지 않음

<표 2-2> 캐나다 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거

	특징	법적 근거
캐나다	<ul style="list-style-type: none"> ♦ 2004년부터 웹 보존을 허용하고 온라인 간행물에 대한 납본을 확대함 ♦ 비밀번호/기술 정보를 요구할 수 있음 ♦ 법률은 공공 접근 조항을 다루지 않음 	<ul style="list-style-type: none"> ▪ 캐나다 도서관 및 기록보존소법(8(2)항) ▪ 섹션 10 및 규정

3) 프랑스

○ 납본제도

- 법적 근거

- ▶ 입법부 문화유산법 L131, L132, L133에서, 제1장: 납본의 목적과 범위(L131-1조 ~ L131-2조)와 제2장: 납본의 방식 및 조직. (L132-1조 ~ L132-7조), 제3장: 처벌 조항(L133-1조)을 규정하고 있음
- ▶ 행정부 시행령 R131-1부터 R133-1-1에서, 제I장: 납본의 목적과 범위 (R131-1조 - R131-7조)와 제2장: 납본 방식 및 조직(R132-1조- R132-45조), 제III장: 처벌 조항(R133-1 ~ R133-1-1)을 규정하고 있음

- ▶ 웹이나 인터넷이라는 매체에 대한 구체적이고 명시적인 언급은 없고 다음과 같은 조항들이 있음

○ 법률 제L131-2조 납본 방식

제L131-2조

- 인쇄형, 그래픽, 사진, 음반, 시청각, 멀티미디어 자료는 그들의 기술적 제작과정, 출판 전파방법을 불문하고 공중이 이용할 수 있게 된 때에 법정 납본의 대상이 된다. 다만 영상상영관에서 최초의 이용을 목적으로 하는 영상제작물은 영화법 제L211-1조에 따라 영상제작물(영화)을 상영할 수 있는 비자(승인)를 받은 이후 최초로 상영하는 때에 법적 기탁 의무가 적용된다.
- 소프트웨어 및 데이터베이스는 그것을 전달하는 매체의 물리적인 특성과 관계없이 배포를 통해 대중에게 공개되는 때에 법적 기탁의 의무가 적용된다.
- 또한 매체의 성격을 불문하고 일반대중에게 전자적으로 소통되는 기호, 신호, 문서, 이미지, 음향 또는 메시지도 의무 납본(법정 납본)의 대상이다.

○ 법률 제L132-1조 납본 방식

제L132-1조

법적 납본이란 자료를 납본기관에 제출하거나 제한된 수의 복제본을 우송료 없이 송부하는 것을 말한다.

국무회의 법령으로 다음의 사항을 정한다

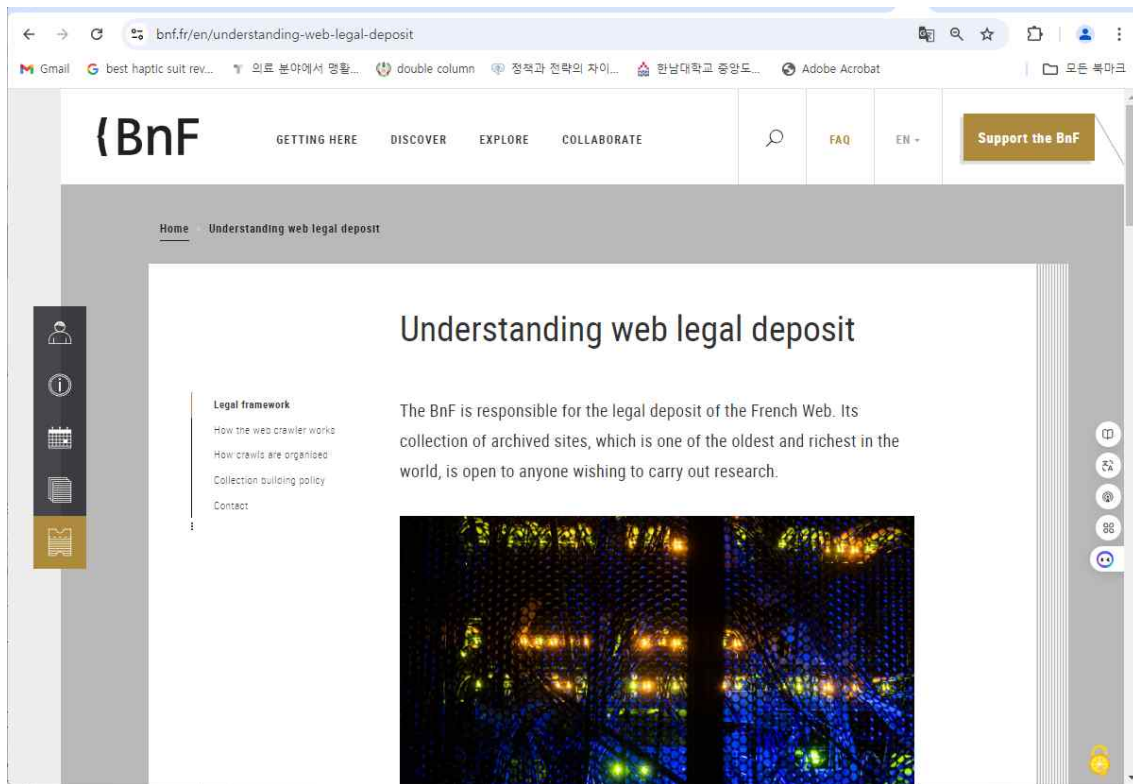
- a) 음성 라디오 또는 텔레비전 방송을 전제로 하는 프로그램의 방송 녹음을 포함하여 다른 수단에 의해 납본 의무가 충족될 수 있는 상황
- b) 제L.132-2조에 언급된 각 범주의 사람들에 대한 구체적인 적용 조건 및 이들 중 특정 사람들이 납본 의무로부터 면제될 수 있는 조건
- c) 수집 및 보존이 제L. 131-1조에 정의된 목적과 관련하여 충분히 부합하지 않는 것으로서, 수집 및 보존할 자료의 제출 의무에서 제외될 수 있는 예외 범위
- d) 모든 문서를 수집하고 보존할 필요 없이, 제L. 131-1조에 정의된 목적을 달성할 수 있는 경우, 제출할 문서를 선택하는 방식

○ 법률 제L132-2조: 납본 제출 의무자

제L131-2조에 언급된 제출 의무는 다음 사람에게 있다.

- a) 디지털 형식을 포함하여 인쇄, 그래픽 또는 사진 자료를 출판하거나 수입하는 자
- b) 상기 a에 언급된 자료를 인쇄하는 자
- c) 디지털 형식을 포함하여 소프트웨어 또는 데이터베이스를 출판, 생산 또는 수입하는 자
- d) 출판하는 자 또는 출판사가 없는 경우 디지털 형식을 포함하여 음반을 제작하거나 주문하는 자 및 수입하는 자
- e) 영화 자료를 제작하는 자, 수입된 영화 자료와 관련하여 그 자료를 저장하는 매체의 특성을 불문하고 이를 배포하는 자, 그리고 대중의 사적 사용을 목적으로 비디오 영상물을 출판하고 수입하는 회사로 구성된다. 디지털 형식을 포함한 영화 자료의 복제;
- f) 언론자유에 관한 법률(법률 제86-1067호, 1986년 9월 30일)의 의미 내에서 라디오 및 텔레비전을 서비스하는 자
- g) 위에 언급된 것 이외의 비디오 영상물과 디지털 형식을 포함하여 더 이상의 상업적 이용의 대상이 되지 않고 TV로 방영되는 비디오 영상물을 출판하는 자, 출판자가 없는 경우 제작 또는 주문하는 자, 수입하는 자
- h) 디지털 형식을 포함한 멀티미디어 자료를 출판하는 자 또는 출판자가 없는 경우 멀티미디어 자료를 생산하거나 수입하는 자
- i) 앞서 언급한 1986년 9월 30일 법률 제86-1067호 제2조의 두 번째 단락의 의미 내에서 전자 수단을 통해 대중에게 전달하기 위해 기호, 신호, 문서, 이미지, 음성, 메시지를 게시하거나 제작하는 자

외부에서 출판되거나 생산된 문서를 국내로 반입하는 사람은 이 조항의 의미 내에서 수입자로 간주한다.



<그림 2-7> 프랑스 국립도서관 웹사이트에 안내되어 있는 웹 법적 납본 안내
출처: <https://www.bnf.fr/en/understanding-web-legal-deposit>

○ 시행령 제R132-23조: 프랑스 국립도서관 납본 대상:

- 본 시행령 조항은 제3편 법적 납본(R131-1부터 R133-1-1까지)의 제2장 납본 방식 및 조직(R132-1조부터 R132-45조까지) 제4항의 전자적 수단을 통한 공공 통신 서비스의 법적 납본(R132-23조부터 R132-23-2조까지)에서 디지털 자원, 특히, 웹자원의 법적 납본에 대하여 규정하고 있음

- 출처:

https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006074236/LEGISCTA000024240034/#LEGISCTA000024240034

제R132-23조

2011년 12월 19일 제정 법령 n°2011-1904 - 예술. 8

프랑스 국립도서관 납본 대상:

프랑스의 법률에 따르면, 다음과 같은 경우에 프랑스 국립도서관(Bibliothèque nationale de France)에 법정 납본 의무가 있다:

제1항 온라인 공중 통신 서비스: 2004년 6월 21일자 법률 제2004-575호의 제1조 IV항에 정의된 온라인 공중 통신 서비스 중 다음 조건을 충족하는 경우:

- ① 해당 서비스가 R. 132-34조에 언급된 서비스에 의해 편집되지 않았거나, 해당 서비스의 프로그램에 주로 집중되지 않은 경우
- ② .fr 도메인 또는 프랑스 도메인 관리 기관에 등록된 다른 도메인 이름으로 등록된 경우
- ③ 프랑스에 거주하는 개인에 의해 등록되었거나 프랑스 영토에서 제작된 경우

제2항 주문형 시청각 미디어 서비스: 1986년 9월 30일자 법률 제86-1067호의 제2조에 정의된 주문형 시청각 미디어 서비스 중 다음 조건을 충족하는 경우:

- ① 지상파를 통해 방송되지 않거나, R. 132-34조 1항에 언급된 서비스의 프로그램을 공중에 제공하지 않는 경우
- ② 해당 법률의 43-2조 및 43-3조에 따라 프랑스에 설립된 경우

○ 시행령 제R132-23-1조: 수집

제R132-23-1조

2011년 12월 19일 제정 법령 n°2011-1904 - 예술. 8

- I. - R.132-23조에 명시된 온라인 공공 커뮤니케이션 서비스와 주문형 시청각 미디어 서비스는 최소한 연 1회 수집되어야 한다.
- II. - R.132-23조에 언급된 해당 서비스가 자동 방식을 사용하여 전부 수집될 수 없는 경우 출판사는 프랑스 국립도서관의 요청에 따라 해당 서비스에 다음을 제공해야 한다. 수집에 필요한 보호 문서에 대한 비밀번호 및 액세스 키를 제공하거나 해당 문서의 복제본을 제공한다. 두 경우 모두 기탁자는 문서의 전달 및 장기 보존에 필요한 모든 기술 데이터도 제공해야 한다. 파일 추출 조건은 BnF와 기탁자가 공동으로 결정한다.
- III. - 프랑스 국립도서관의 온라인 공공 커뮤니케이션 서비스는 시행 중인 수집 절차와 사용 중인 자동 수집 도구의 기술 사양에 대해 대중에게 정보를 제공한다.
- IV. - 제L.131-1조 b에 규정된 온라인 공공 커뮤니케이션 서비스 및 주문형 시청각 미디어 서비스의 국가 장서 목록은 이러한 서비스에 대한 색인 형식으로 제공된다.

○ 시행령 제R132-23-2조: 현장 접근

제R132-23-2조

2011년 12월 19일 제정 법령 n°2011-1904 - 예술. 8

온라인 공공 커뮤니케이션 서비스와 주문형 시청각 미디어 서비스에 대한 현장 접근(on access on site)은 다음과 같이 장소에서 이루어져야 한다.

제1항 프랑스 국립도서관(Bibliothèque nationale de France) 및 문화부 장관이 승인한 기관

제2항 프랑스 국립도서관 또는 승인된 기관이 제공하는 접근, 검색 및 처리 인터페이스를 갖춘 개인용 단말기에서, 엄격하게 인증된 연구자들만 사용 가능

○ 참고자료

- 출처: <https://cdn.ifla.org/past-wlic/2011/193-stirling-en.pdf>



<그림 2-8> Peter Stirling, et.al. 프랑스 디지털 납본제도에 대한 소개 및 안내자료, IFLA 2011, June 1, 2011

<표 2-3> 프랑스 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거

	특징	법적 근거
프 랑 스	<ul style="list-style-type: none"> 2006년부터 국립시청각연구소(INA)와 프랑스 국립도서관(BnF)이 책임을 공유. 현장에서 액세스할 수 있으며 승인된 사용자로 제한 프랑스와 네덜란드에 본부를 둔 비영리단체인 인터넷 메모리 재단은 납본법의 적용을 받지 않음 	<ul style="list-style-type: none"> Code du patrimoine(프랑스어) 추가 정보 (<그림 2-8> IFLA 백서 참조)

4) 독일

○ 법적 근거

- 독일 국립도서관법

- ▶ 2006년 6월 22일자 독일 국립도서관법 제20조(BGBI. I p. 1338)와 1998년 10월 27일자 조직 법령(BGBI. I p. 3288)에 근거하여 연방 수상은 법령에 따라 독일 국립도서관장에게 수집 권한 부여



<그림 2-9> 독일 국립도서관법

▶ 저작권의 예외 형태로 허용되는 사용을 규정한 법률



<그림 2-10> 저작권의 예외 형태로 허용되는 사용을 규정한 온라인 법률 조희사이트

- 납본제법



<그림 2-11> 독일 납본제도 법률

<표 2-4> 독일 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거

	특징	법적 근거
독 일	<ul style="list-style-type: none"> 2006년부터 법적 납본. 2012년부터 수집. DNB와 웹 아카이빙에 협력하는 지역 도서관의 열람실에서 이용 가능 	<ul style="list-style-type: none"> 독일 국립도서관에 관한 법률 납본 규정 (<그림 2-11> 참조)

5) 일본

○ 법적 근거

- 국회도서관법

- ▶ ‘제11-2장 정부, 지방공공단체, 법인행정기관 등의 인터넷 자료에 관한 기록’에서 인터넷을 통해 제공되는 자료에 대한 수집 권한을 가짐

제11-2 장 정부, 지방공공단체, 법인행정기관 등의 인터넷 자료에 관한 기록

제25조의3. 관장은 제24조 또는 제24조의2에 규정된 기관이 대중에게 공개하거나 인터넷 서비스를 통해 제공하는 인터넷 자료를 국립국회도서관이 사용하는 매체에 기록하여 공식 사용을 위해 수집할 권한이 있다. “인터넷 자료”란 전자적, 자기적 또는 사람의 지각을 통해 인식할 수 없는 다른 수단으로 기록되어 인터넷을 통해 대중에게 공개되는 문자(텍스트), 이미지, 음향(사운드) 또는 프로그램을 말한다.

- 2 제24조 또는 제24조의2에 규정된 기관은 관장이 정하는 규칙에 따라 그러한 자가 대중에게 공개하거나 인터넷 서비스를 통해 제공하는 인터넷 자료를 관장이 적절하게 기록하도록 필요한 조치를 취할 의무가 있다. 이 항 및 다음 항의 경우 관장은 그 성질과 대중에게 공개하는 목적상 제1항에 규정된 목적을 달성하는 데 불필요한 특정 디지털 자료의 기록을 제외하는 규칙을 정할 수 있다.
3. 제24조 또는 제24조의2에 규정된 기관이 공중에게 제공하거나 인터넷 서비스를 통하여 제공하는 인터넷 자료 중에서 관장은 제1항에 규정된 목적을 달성하기 위하여 특히 필요한 것으로 관장이 지정한 디지털 자료를 국립국회도서관에 제공하도록 해당 기관에 요청할 수 있는 권한을 가진다. 이 경우 해당 기관은 정당한 거부 이유가 없는 한 요청에 응하여야 한다.

제11-3장 온라인 출판 기록

제25조의 4

- 1 제24조 또는 제24조의2에 규정된 자 외의 자가 인터넷 기타 첨단정보통신망을 통하여 온라인 간행물을 공중에 제공하거나 송신한 때에는 출판사는 문화유산의 축적과 이용에 이바지하기 위하여 전조의 규정에 해당하는 경우를 제외하고 관장이 정하는 규칙에 따라 국립국회도서관에 제공하여야 한다. "온라인 자료"이란 전자적, 자기적 또는 그 밖의 사람의 지각을 통해 인식할 수 없는 다른 수단에 의하여 기록된 문자(텍스트), 영상, 음향(사운드) 또는 프로그램으로서 인터넷 기타 첨단정보통신망을 통하여 공중에 제공하거나 송신한 것으로서 관장이 단행본 또는 연속간행물(비밀사항, 서식, 양식 및 그 밖의 간단한 자료 제외)에 준하는 것으로 정하는 것을 말한다.
- 2 전항의 규정은 다음 각 호에 규정된 경우에는 적용하지 아니한다.
 - (1) 제24조 또는 제24조의2에 규정된 자 외의 자가 국립국회도서관에서 사용하는 매체에 제1항의 제공 없이 공중에 제공하거나 송신하는 온라인 자료를 기록하여 줄 것을 관장에게 요청하고 관장이 이를 승인한 경우
 - (2) 온라인 자료의 내용이 이 조의 규정에 따라 이전에 수집한 온라인 자료의 양이나 내용에 차이가 없는 경우
 - (3) 관장이 온라인 자료가 그 성질 및 공중에 제공하거나 송신하는 목적상 제1항의 목적을 달성하는 데 불필요하다고 인정하는 경우
 - (4) 관장이 특별한 사유가 있다고 인정하는 경우
- 3 관장은 제1항의 규정에 따른 제공 또는 제1항의 승인에 해당하는 온라인 자료를 국립국회도서관에서 사용하는 매체에 기록하여 수집할 수 있는 권한을 가진다.
4. 제1항의 규정에 따라 온라인 자료를 제공한 자에게 관장은 같은 항에서 규정하는 제공에 통상적으로 소요되는 비용에 상당하는 금액을 지급하여야 한다. 다만, 해당자가 지급을 요구하지 아니한다는 의사를 표시한 경우에는 그러하지 아니하다.

<표 2-5> 일본 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거

	특징	법적 근거
일본	<ul style="list-style-type: none"> 2010년부터 정부 및 정부 유관기관 웹사이트 수집을 허용 (인터넷을 통한 접근을 위해서는 권한이 필요함) 2012년 법률이 개정되어 일본에서 출판된 전자책을 보관할 수 있는 권한을 국립국회도서관에 부여 (2013년 7월부터 시행) 	<ul style="list-style-type: none"> 국회도서관법 제25조의3 국립국회도서관 뉴스레터

6) 뉴질랜드

○ 법적 근거

- 수집 및 접근권한

- ▶ 뉴질랜드 국립도서관(Te Puna Mātauranga o Aotearoa) 법 2003은 2부 9. 국립도서관장의 기능 및 권한에 국립도서관장에게 정보 수집 및 접근 권한을 부여하는 것을 규정하고 있음

9 국립도서관장의 기능 및 권한

(1AA)

국립도서관장은 이 법에 따라 국립도서관장에게 부과된 기능과 의무를 수행하고 이행하는 데 필요한 모든 권한을 가진다.

(1) 국립도서관의 목적을 달성하기 위한 국립도서관장의 기능은 다음과 같다.

- 뉴질랜드와 뉴질랜드 국민과 관련된 포괄적인 문서 컬렉션을 포함하여 국가 문서 컬렉션을 개발하고 유지한다.
- 국립도서관 소장품 및 자료를 장관이 결정하는 방식과 조건에 따라 접근 가능하게 하여 해당 소장품 및 자료를 가장 유리하게 활용할 수 있다.
- 장관이 적합하다고 생각하는 정보 자원에 대한 접근, 서지 및 학교 서비스를 포함한 기타 서비스를 제공한다.
- 뉴질랜드 및 기타 지역의 당국 및 기타 사람들과 도서관 문제에 대한 협력을 촉진한다.
- 도서관 및 정보 문제와 관련된 사항에 대해 장관에게 조언하고 지원한다.
- 이 법에 따른 규정에 따라 정한 기타 기능을 수행한다.

- 납본제도

- ▶ 국립도서관 요구사항(전자문서) 공지 2006(SR 2006/118)은 8항에서, 뉴질랜드 국립도서관장은 뉴질랜드 국립도서관(Te Puna Mātauranga o Aotearoa) 법 2003년 제 31조에 따라, 국가 인터넷 문서에 대한 복사 요청을 통지하고 있음

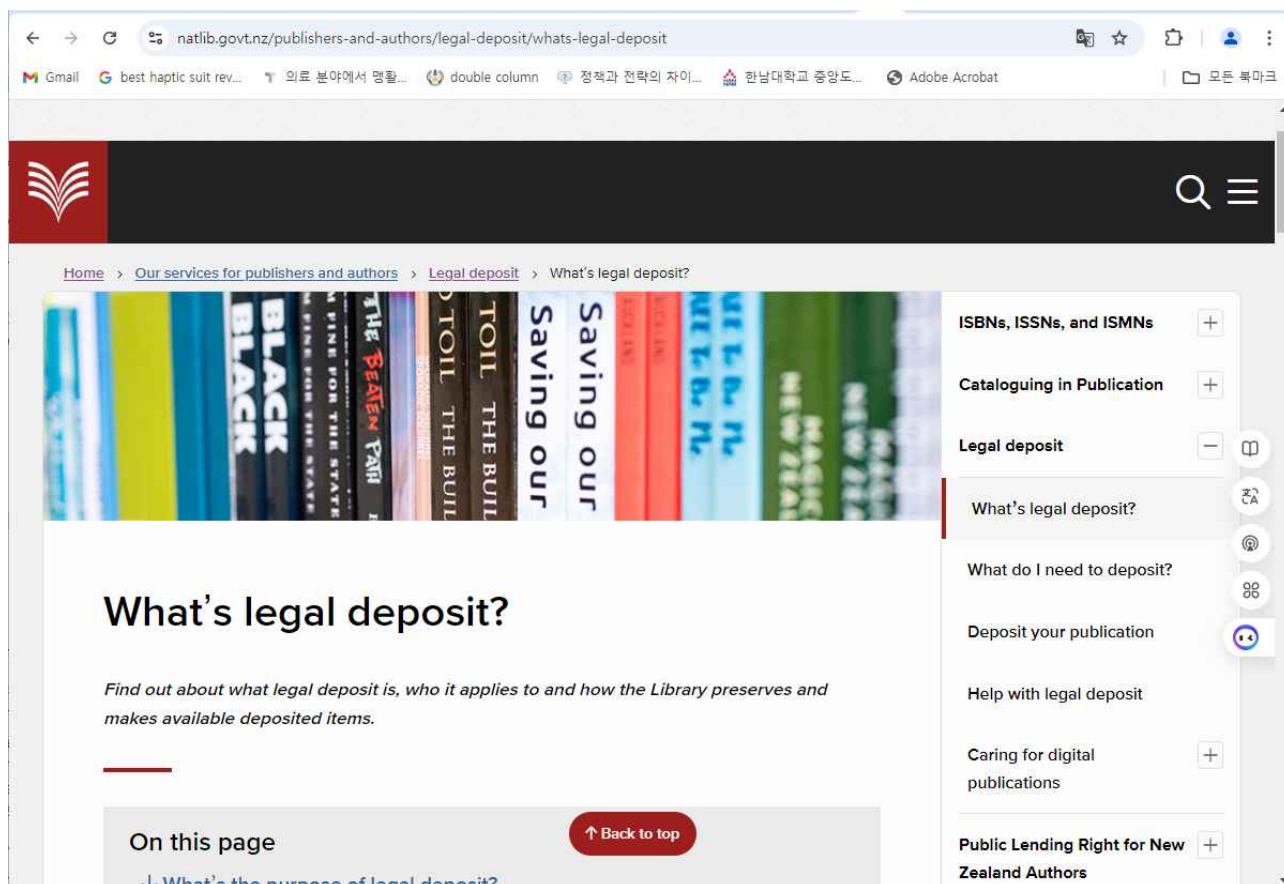
8 인터넷 문서 복제본 제작 권한

국가사서(National Librarian:국립도서관장)는 어떤 인터넷 문서든지 복사할 수 있는 권한을 가진다.

- 국립 사이트

▶ 뉴질랜드 국립도서관 웹사이트에 법적 근거를 제시하고 설명하고 있음

▶ 출처: <https://natlib.govt.nz/publishers-and-authors/legal-deposit/whats-legal-deposit>



<그림 2-12> 법적 납본에 대해 설명하는 뉴질랜드 국립도서관 웹 아카이브 웹사이트

<표 2-6> 뉴질랜드 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거

뉴 질 랜 드	<ul style="list-style-type: none"> 2006년 8월부터 웹 수집을 허용하고 전자 출판물에 대한 법적 기탁을 확대. 문서가 온라인에서 공개적으로 사용 가능하거나 소유자가 허가한 경우 보관된 복제본을 온라인에서 사용 가능. 문서에 대한 온라인 접근이 제한된 경우, 보관된 복제본은 웰링턴 국립도서관 캐서린 맨스필드 열람실 내 동시 사용자 3명으로 제한 	<ul style="list-style-type: none"> 2003년 뉴질랜드 국립도서관법 국립도서관 요건으로 전자문서에 대한 납본을 확대한 2006년 공지
------------------	--	---

7) 영국

○ 법적 근거

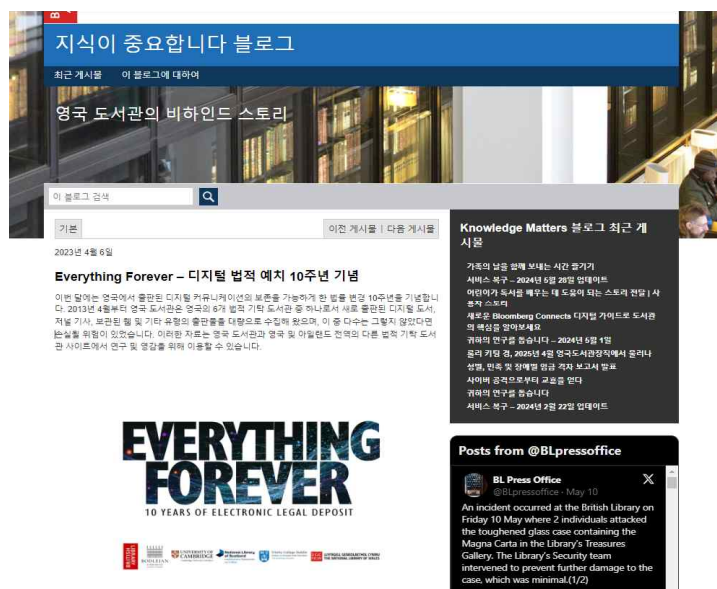
- 납본제도

- ▶ 2003년 납본 도서관법(Legal Deposit Libraries Act)를 통해 인쇄물뿐 아니라 비인쇄물에 대한 법적 납본 규정을 제정하였음

6 규정: 비인쇄 출판물의 납본

- (1) 국무장관은 인쇄 매체가 아닌 다른 매체에 출판된 저작물에 적용되는 1항 및 2항을 보완하는 규정을 만들 수 있다.
- (2) 이 항에 따른 규정은 특히 다음을 포함할 수 있다.
- (a) 납본 도서관이 1항에 따라 제출 받는 시점 또는 조건에 대한 조항을 규정한다.
 - (b) 1(1)항에 언급된 자에게 저작물 복제본과 함께 컴퓨터 프로그램과 저작물에 접근하는 데 필요한 정보 복제본, 그리고 저작물과 함께 대중에게 제공되는 매뉴얼 및 기타 자료 복제본을 전달하도록 요구한다.
 - (c) 출판 또는 다른 이벤트에 공개된 이후 규정된 시간 내에 제출해야 한다.
 - (d) 전자적 방법으로 제출을 허용하거나 요구한다.
 - (e) 저작물이 서로 다른 품질로 출판을 위해 제작된 경우 제출할 복제본의 품질을 지정한다.
 - (f) 저작물이 다양한 형식으로 공개되거나 대중에게 제공되는 경우, 납본 도서관 또는 그 중 하나가 지정한 요구 사항(일반적으로 또는 특정한 경우)에 따라 복제본을 전달하는 형식을 결정하도록 규정한다.
 - (g) 온라인에 출판된 자료가 영국에서 출판된 것으로 간주되거나 간주되지 않는 상황에 대한 규정을 제시한다.
 - (h) 온라인에 출판 저작물의 복제본을 전달하는 매체를 지정한다.

- ▶ 2013년 4월 6일 디지털 법적 납본제도를 시행하였음
- ▶ 2023년 4월 6일 디지털 법적 납본제도 10주년 기념 기사 게재
- ▶ 출처:
<https://blogs.bl.uk/living-knowledge/2023/04/everything-forever-marking-10-years-of-digital-legal-deposit.html>



<그림 2-13> 디지털 법적 납본 10주년 기념 블로그

<표 2-7> 영국 웹 아카이브 관련 납본제도의 특징 요약 및 법적 근거

	특징	법적 근거
영국	<p>IIPC 회원:</p> <ul style="list-style-type: none"> 영국 국립도서관 영국 국립기록보존소 스코틀랜드 국립도서관 한조(Hanzo) 아카이브(상용 웹 아카이빙 서비스) <p>법정 납본 콘텐츠에 대한 접근은 법정 납본도서관 내에서 제공</p>	<p>2013년 4월 6일부터 납본은 디지털 및 온라인으로 출판된 자료에도 적용되므로 납본 도서관은 웹사이트, 블로그, 전자저널 및 CD-ROM과 같은 영국의 비인쇄 출판 자료에 대한 국가 아카이브를 제공할 수 있음. 영국 법정 납본 도서관에는 다음이 포함됨.</p> <ul style="list-style-type: none"> 영국 국립도서관 옥스퍼드대학교 보들리안도서관 케임브리지대학 도서관 스코틀랜드 국립도서관 더블린 트리니티칼리지 도서관 웨일즈 국립도서관

8) 네덜란드

○ 법적 근거

- 2006년 라이덴대학교에서 네덜란드 관점, 즉, 네덜란드 법률 내에서 웹 보존의 법적 측면에 대한 연구를 수행하였음. 특히 저작권과 개인 정보 보호법(WBP)에 초점을 맞추었음. 예를 들어, 웹사이트 크롤링은 일종의 복사이므로 저작권의 적용을 받는 행위로 소유자의 허가가 필요함. 다양한 형식으로

여러 개의 새 복제본을 만드는 것은 때때로 여러 시간에 수집된 웹사이트의 모든 크롤러 복제본을 영구적으로 보존하는 데 필요함. 따라서 저작권이 다시 작용하지만 저작권법(Auteurswet)에는 보존 목적의 마이그레이션 복제본에 대한 예외가 포함되어 있음

- 출처: <https://www.kb.nl/en/over-ons/expertises/webarchivering/juridische-aspecten>

<표 2-8> 네덜란드 웹 아카이브 관련 납본제도의 특징 요약 및 참고자료

	특징	법적 근거
네 덜 란 드	<ul style="list-style-type: none"> ◆ 현장 접근: 앞으로는 인터넷에 접속하는 것을 목표로 설정 ◆ IIPC 회원: <ul style="list-style-type: none"> - 네덜란드 국립도서관 - 네덜란드 음향 및 시각 연구소 - Archiefweb.eu (상용 웹 아카이빙 서비스) 	<ul style="list-style-type: none"> ▪ 웹 아카이빙 법적 측면 연구 보고서

9) 미국

○ 법적 근거

- 저작권법

- ▶ 저작권법의 108조에 도서관 및 기록보존소에 의해 복제되는 것은 배타적 권리인 저작권을 침해하는 것이 아니라고 규정함

- 디지털 법적 납본

- ▶ Digital Legal Deposit - An IPA Special Report, July 24, 2014
- ▶ 전자 납본 자체가 의무사항은 아니고 인쇄물 의무 제출 시 전자버전 제출을 권고함

<표 2-9> 미국 웹 아카이브 관련 보존제도의 특징 요약 및 법적 근거

	특징	법적 근거
미국	IIPC 회원: • 컬럼비아대학교 도서관 • 하버드대학교 도서관 • 인터넷 아카이브 • 의회도서관 • 로스앨러모스 국립연구소도서관 • 올드도미니언대학교, 컴퓨터과학과 • 스탠포드대학교 도서관 • UCLA 연구 도서관 • 노스텍사스대학교 도서관 • 웹리코더 각각 보존 및 액세스 정책이 다름	예) 의회도서관(LC): 크롤링 의도를 사이트 저작권에게 통지. 경우에 따라 보존 및 외부 접근을 위해 권한이 필요함. 정부 웹사이트(미국 연방, 주, 지방)는 통지나 허가 없이 보존됨

2.1.2 요약 및 시사점

- IIPC 회원국의 법적 이슈는 웹 아카이브를 수집하기 위해 허가, 공정 사용 또는 거부 기반 시스템을 사용하는 회원국들(미국, 네덜란드, 한국 등)과 그리고, 해당 국가의 납본법 기반 웹 아카이브 수집을 하는 회원국(호주, 오스트리아, 캐나다, 독일, 일본, 프랑스, 영국 등)으로 크게 나눌 수 있음
- IIPC 회원국들은 기본 디지털 자원(ebook, digital audio, digital video)에 대한 의무 납본이나 허가, 공정 사용 또는 거부 기반의 두 경우 모두 웹사이트를 포함하여 시행하는 회원국이나 기관이 대다수이고 그렇지 못한 국가는 웹사이트를 포함하는 방향으로 확대되기를 노력하고 있음
- 또한, 현재 현장 접속이 주로 이루어지고 있는 상황에서 ‘현장’의 의미를 지점이나 파트너도서관으로 확대하여 인터넷 접속 이용이 가능한 방향으로 노력하고 있는 상황임
- 현재, 웹 아카이빙의 법적 이슈는 납본법과 저작권법과의 관계에서 발생하는 것으로, 상충되는 두 가치(보존을 위한 복제본 생산과 공익적 활용, 저작자 권리 보호)를 어떻게 조화롭게 유지하고 각 가치를 존중할 것인가를 풀어내는 것이 관건임

2.2 국제 웹 아카이빙 기술 동향

- IIPC 주요 회원기관들을 대상으로 2021년에 조사된 설문조사³⁾에서 언급된 각 기능별 도구들에 대하여 최신 동향을 조사 정리하였음

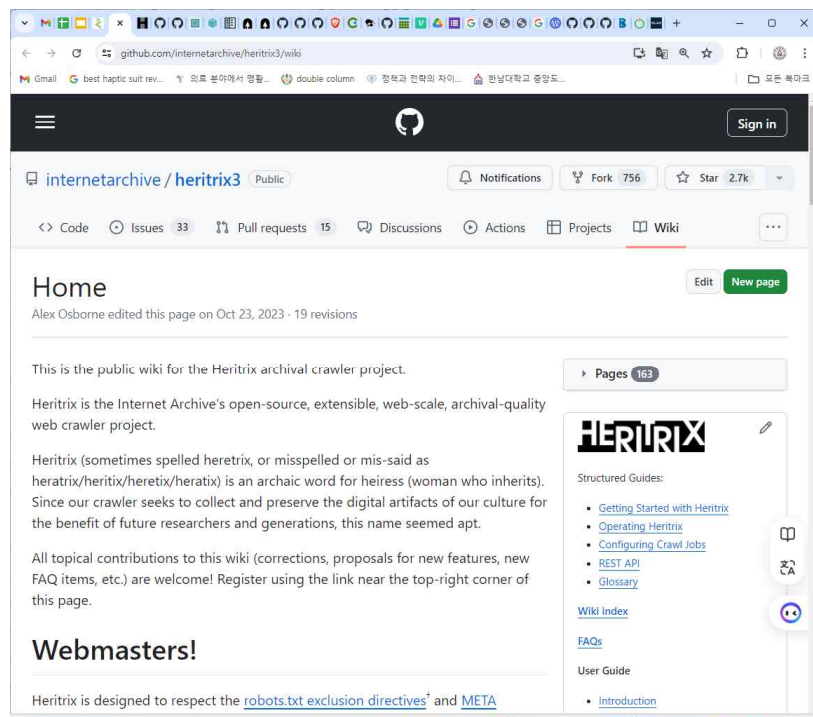
3) 국제인터넷보존컨소시엄(IIPC), 웹 아카이빙 소프트웨어 및 도구 관련 활용 현황 설문조사, 2021.

<https://docs.google.com/spreadsheets/d/1Ww5bhd7KbAH8QOI2qDhdakWh5aExN9JkeL431fhE0Vg/edit#gid=266509536>

2.2.1 수집(Capture)

1) Heritrix 3.x

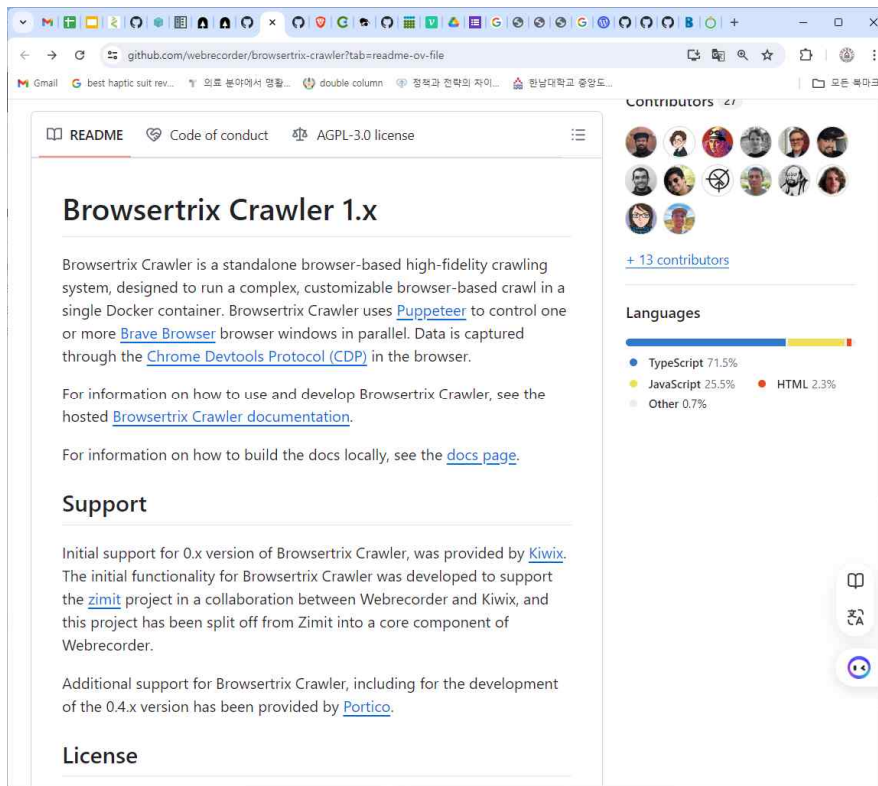
- Heritrix는 Internet Archive의 확장 가능한 오픈 소스, 웹 규모, 아카이브 품질의 웹 크롤러 프로젝트임
- 이 프로젝트는 미래 연구자와 세대의 이익을 위해 인류 문화의 디지털 유산을 수집하고 보존하는 프로젝트임



<그림 2-14> 오픈 소스 Heritrix3 Github 화면

2) Browsertrix

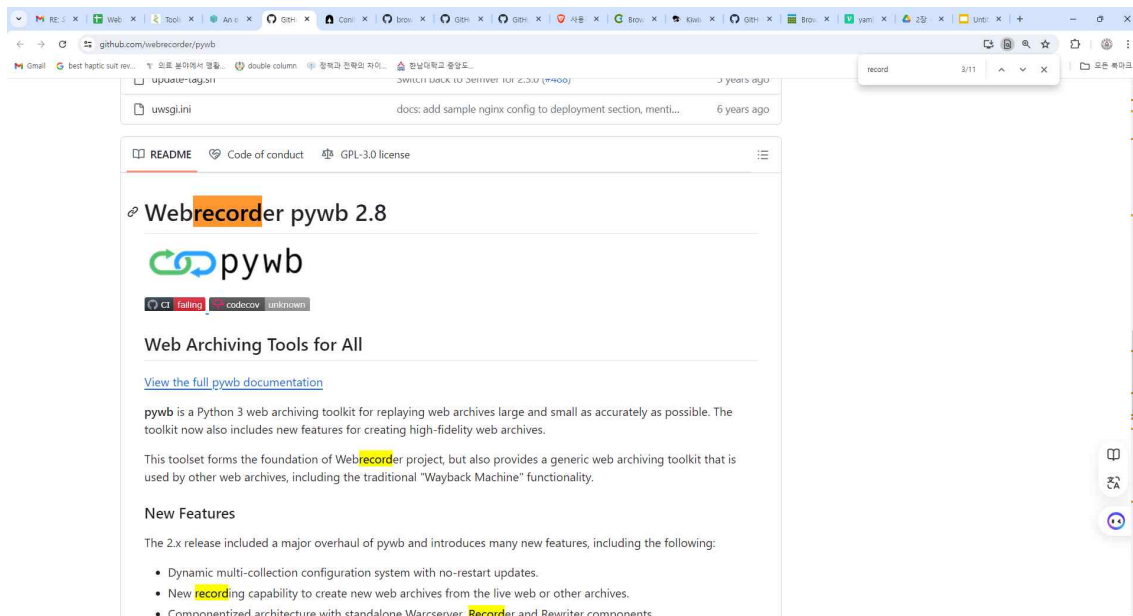
- 컨테이너에서 복잡하고 사용자 정의 가능한 브라우저 기반 크롤링을 실행하도록 설계된 독립형 브라우저 기반 고성능 크롤링 시스템임
- 이 크롤러는 Puppeteer를 사용하여 하나 이상의 BraveBrowser browser 창을 병렬로 제어함
 - Puppeteer는 DevTools Protocol 또는 WebDriver BiDi를 통해 Chrome 또는 Firefox를 제어하는 고급 API를 제공하는 Javascript 라이브러리임
 - Brave Software, Inc.에서 개발하는 크로미엄 웹 브라우저에 기반한 오픈 소스 웹 브라우저로 웹사이트 트래커(website tracker)와 광고를 차단함
- 데이터는 브라우저의 Chrome Devtools Protocol(CDP)을 통해 캡처됨



<그림 2-15> 오픈 소스 Browsertrix Crawler 메인 화면

3) Pywb

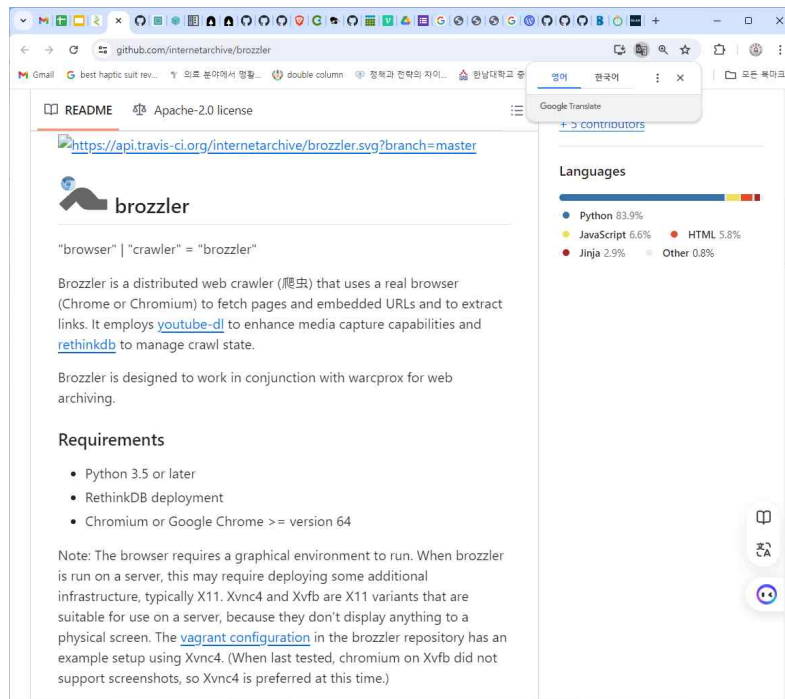
- Webrecorder.net 에서 개발한 파이썬(Python) 기반의 웹 아카이브 재생 도구임
- 2020년에 IIPC가 Pywb를 재생 솔루션으로 권장하고 IIPC 회원들이 Pywb를 채택하도록 지원함
- IIPC 회원들이 블로그 게시물과 기술 웨비나를 통해 공유해 온 공개적으로 사용 가능한 전환 가이드와 사용 사례가 있어 Pywb으로의 전환 시 참고자료가 풍부함
- Pywb와 함께 호주 국립도서관의 Alex Osborne 이 개발한 인덱싱 도구인 OutbackCDX를 사용하는 것을 권장함
- 재생을 위해 Pywb를 사용하는 웹 아카이브들로는 포르투갈(Arquivo.pt), 아이슬란드(Vefsafn.is), 영국(UK Web Archive) 및 미국 노스텍사스대학교(UNT의 Web Archive)을 포함함



<그림 2-16> 오픈 소스 Pywb 메인 화면

4) Brozzler

- 명칭인 Brozzler는 "browser" 와 "crawler"를 합쳐 만든 신조어임
- 실제 브라우저(Chrome 또는 Chromium)를 사용하여 페이지와 포함된 URL을 가져오고 링크를 추출하는 분산 웹 크롤러임
- youtube-dl을 사용하여 미디어 캡처 기능을 향상하고 크롤링 상태를 관리하기 위해 사용

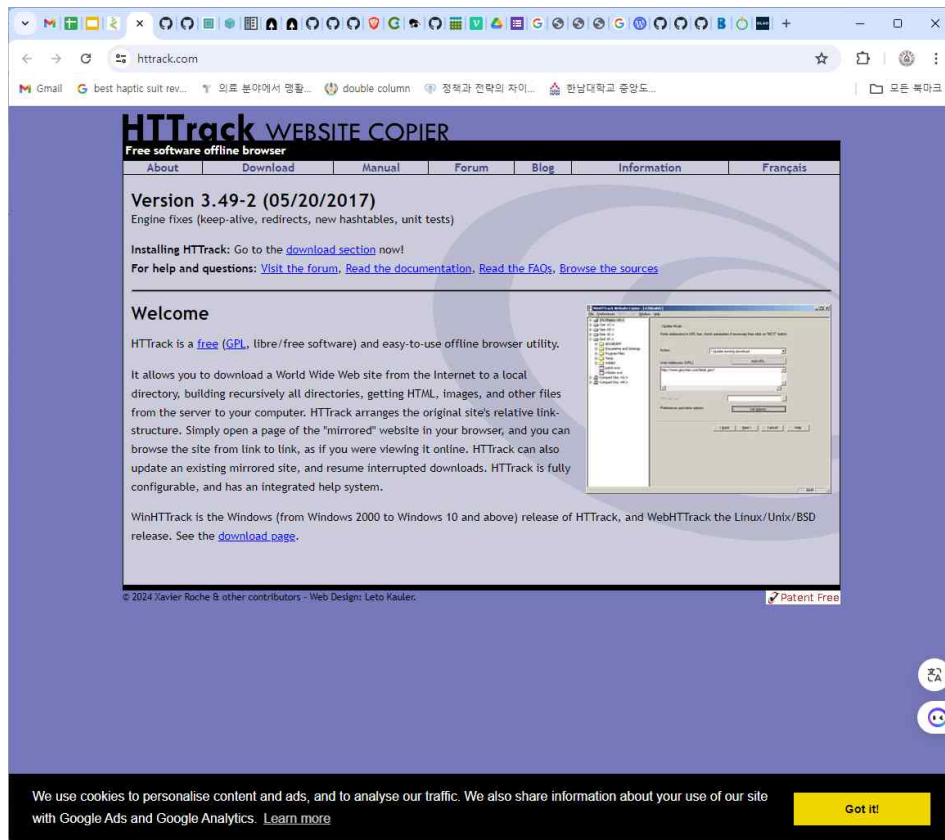


<그림 2-17> brozzler 메인 화면

5) HTTrack⁴⁾

- HTTrack은 무료(GPL, libre/free 소프트웨어)이며 사용하기 쉬운 오프라인 브라우저 유틸리티임
- 이를 통해 World Wide Web 사이트를 인터넷에서 로컬 디렉토리로 다운로드하고 모든 디렉토리를 재귀적으로 구축하고 HTML, 이미지 및 기타 파일을 서버에서 컴퓨터로 가져올 수 있음
- HTTrack의 특징으로는,
 - HTTrack은 원본 사이트의 상대 링크 구조를 정렬함
 - 브라우저에서 “미러링된” 웹사이트의 페이지를 열면 마치 온라인에서 보는 것처럼 링크에서 링크로 사이트를 탐색할 수 있음
 - HTTrack은 기존 미래 사이트를 업데이트하고 중단된 다운로드를 재개할 수도 있음
 - HTTrack은 완전히 구성 가능하며 통합 도움말 시스템이 있음

4) 출처: <https://www.httrack.com/>



<그림 2-18> HTTrack 프로젝트 메인 화면

6) Capture Tool 시사점

- 설문에 응한 22개의 IIPC 회원기관 중 17개의 기관이 현재 Heritrix 3.x를 사용하고 있음
- 11개의 IIPC 회원기관이 현재 Browsertrix(local deployment) 버전을 사용하고 있음
- 11개의 IIPC 회원기관은 앞으로 Browsertrix(local deployment) 버전을 사용할 계획으로 있다고 밝혔다
- Bozzler 수집기는 4개 기관(룩셈부르크 국립도서관, 헝가리 국립도서관, 콜롬비아대학교, 캐나다도서관 및 기록보존소)이 현재 사용 중임
- HTTrack 수집기관은 2개 기관(호주 국립도서관, 헝가리 국립도서관)이 현재 사용 중에 있음
- Pywb(record mode) 버전은 4개 기관(호주 국립도서관, 오스트리아 국립도서관, 노스텍사스대학교, 영국 국립도서관)이 현재 사용 중임
- (명령행을 통한) Browstrix는 단 한 곳(노스텍사스대학교)에서 사용 중이고 아이슬란드국립도서관에서 사용 예정임

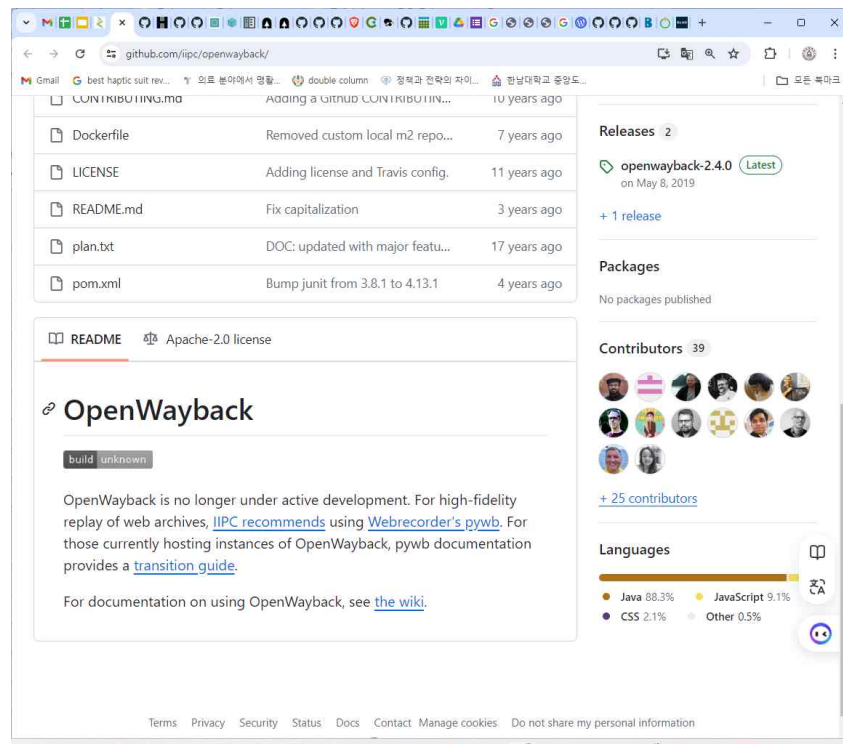
2.2.2 재생과 색인(playback+indexing)

1) Pywb

- 앞의 capture 내용과 중복되므로 앞의 절을 참조할 것

2) OpenWayback

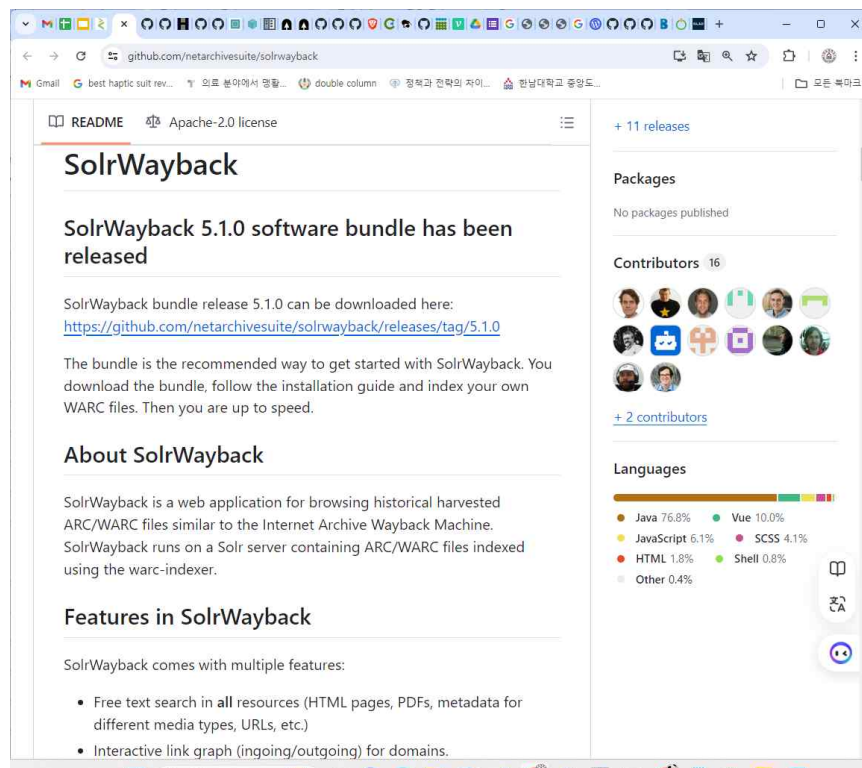
- OpenWayback은 더 이상 활발하게 개발되지 않음
- 웹 아카이브의 고품질 재생을 위해 IIPC는 Webrecorder의 Pywb를 사용하는 것을 권장함
- 현재 OpenWayback 인스턴스를 호스팅하는 경우 Pywb 설명서에서 전환 가이드를 제공함



<그림 2-19> 오픈 소스 OpenWayback 메인 화면

3) SolrWayback

- SolrWayback은 Internet Archive Wayback Machine과 유사하게 수집된 기록 ARC/WARC 파일을 검색하기 위한 웹 애플리케이션임
- SolrWayback은 warc-indexer를 사용하여 색인화된 ARC/WARC 파일이 포함된 Solr 서버에서 실행됨



<그림 2-20> 오픈소스 SolrWayback 메인 화면

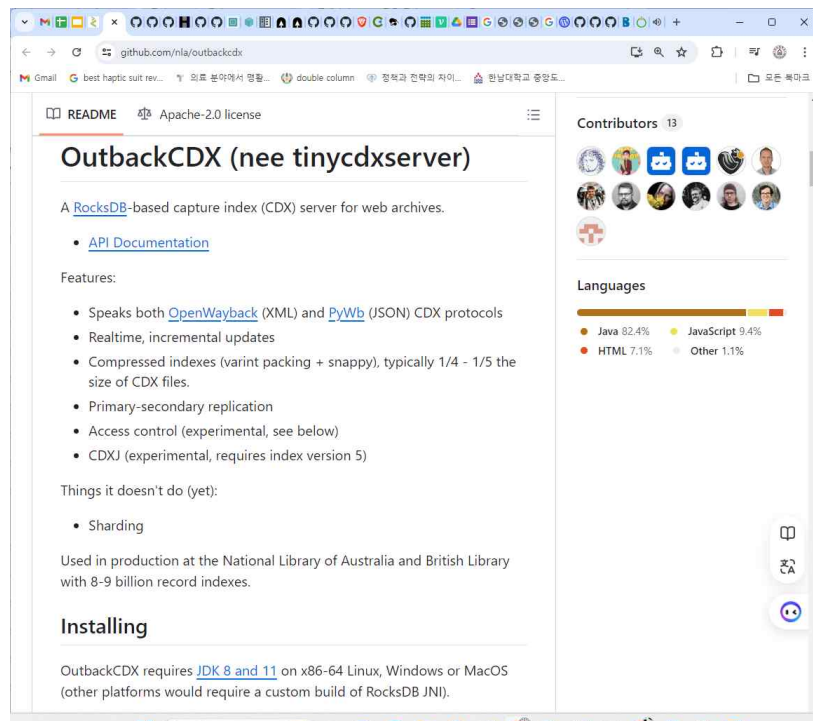
○ SolrWayback-Features는 다음과 같음

- 모든 리소스(HTML 페이지, PDF, 다양한 미디어 유형에 대한 메타데이터, URL 등)에서 무료 텍스트 검색
- 도메인에 대한 대화형 링크 그래프(수신/발신)
- 스트리밍 다운로드를 통해 검색 결과를 WARC 파일로 내보냄. 즉, 결과 집합의 크기에 제한이 없음
- 검색 결과에서 파일을 Zip으로 내보내기(예: HTML, 이미지 콘텐츠, 비디오 등을 기본 형식으로 내보내기)
- 사용자 정의 필드 선택을 통해 검색 결과를 CSV 텍스트로 내보낼 수 있음
- 도메인용 워드 클라우드 생성
- N-gram 검색 시각화
- 도메인별 검색 결과 시각화
- 크기, 들어오고 나가는 링크 수 등 시간 경과에 따른 다양한 도메인 통계를 시각화함
- Gephi 형식으로 링크 그래프를 대규모로 내보낼 수 있음(<https://labs.statsbiblioteket.dk/linkgraph/> 참조)
- 구글 이미지와 유사한 이미지 검색
- 이미지의 EXIF 메타데이터 정보를 사용하여 지도 위의 위치별 이미지 지리 검색

- 파일 업로드로 검색함(예: 이미지, PDF)을 사용하여 리소스가 수집되었는지 확인하고 이미지를 사용하는 HTML 페이지를 찾음
- 리소스에 대해 색인화된 모든 필드를 보고 레코드에 대한 warc-header를 표시함
- OpenWayback 또는 Pywb와 같은 재생 API를 사용하여 대체 재생 엔진을 모든 재생 엔진으로 구성함

4) OutbackCDX

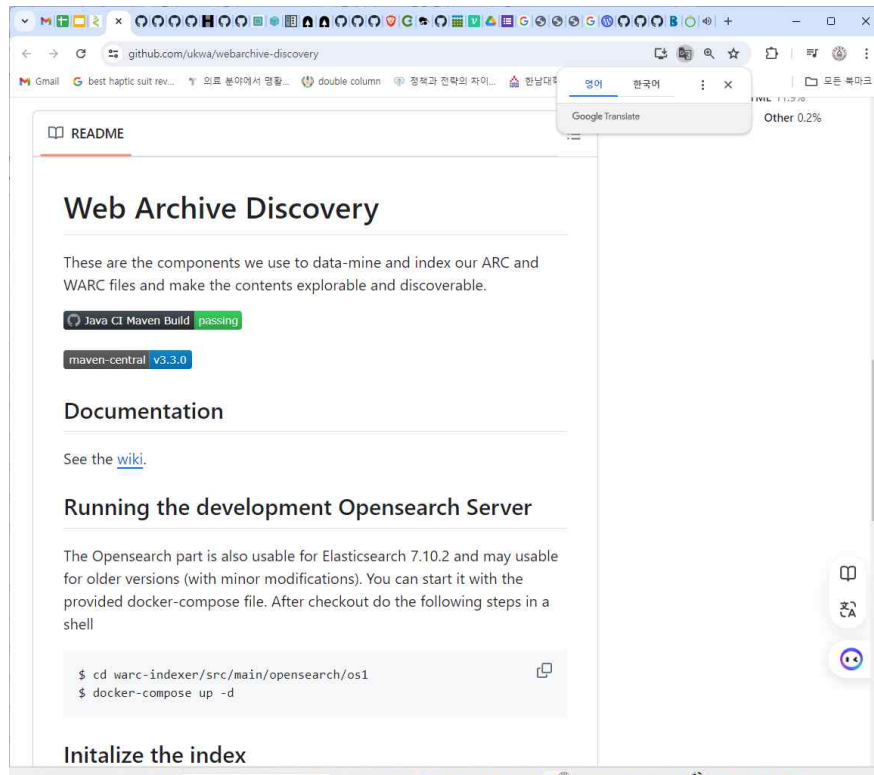
- 웹 아카이브용 RocksDB 기반 캡처 인덱스(CDX) 서버임
- 특징은 다음과 같음
 - OpenWayback(XML) 및 Pywb(JSON) CDX 프로토콜을 모두 사용함
 - 실시간, 증분 업데이트
 - 압축된 인덱스(varint 패킹 + snappy), 일반적으로 CDX 파일 크기의 1/4 - 1/5
 - 1차-2차 복제
 - 액세스 제어(실험적, 아래 참조)
 - CDXJ(실험적, 인덱스 버전 5 필요)



<그림 2-21> 오픈소스 OutbackCDX 메인 화면

5) Web Archive Discovery

- 이 프로젝트의 주요 목표는 웹 아카이브에 대한 전체 텍스트 검색을 제공하는 것임
- 이를 달성하기 위해 warc-indexer 구성 요소를 사용하여 (W)ARC 파일을 구문 분석하고 각 리소스에 대해 하나 이상의 Apache Solr 서버에 레코드를 게시함
- 그런 다음 이용자들이 Solr 색인을 쿼리하고 컬렉션을 탐색할 수 있는 클라이언트 도구를 사용함



<그림 2-22> 오픈소스 Web Archive Discovery의 메인 화면

6) Playback & indexing Tool 시사점

- 설문에 응한 22개의 IIPC 회원기관 중 12개의 기관이 현재 Pywb를 사용하고 있으며 4개 기관이 Pywb를 앞으로 사용할 계획이라고 밝혔음
- 12개의 IIPC 회원기관이 현재 OpenWayback 버전을 사용하고 있고 6개 기관은 과거에 사용했었다고 함
- 4개의 IIPC 회원기관은 현재 SolrWayback 버전을 사용하고 있고 추가로 5개 기관은 앞으로 SolrWayback을 사용할 계획임
- 9개 기관은 OutbackCDX 버전을 사용하고 있고, 2개 기관은 도입 중이거나 4개 기관은 OutbackCDX를 사용할 계획이 있다고 밝혔음
- Web Archive Discovery는 4개 기관이 사용하고 있고 현재 2개 기관이 앞으로 도입할 계획을 갖고

있음

- 전반적으로 OpenWayback이나 Pywb으로 이주하고 있음
- OutbackCDX도 도입하는 기관이 점점 늘어나는 추세임

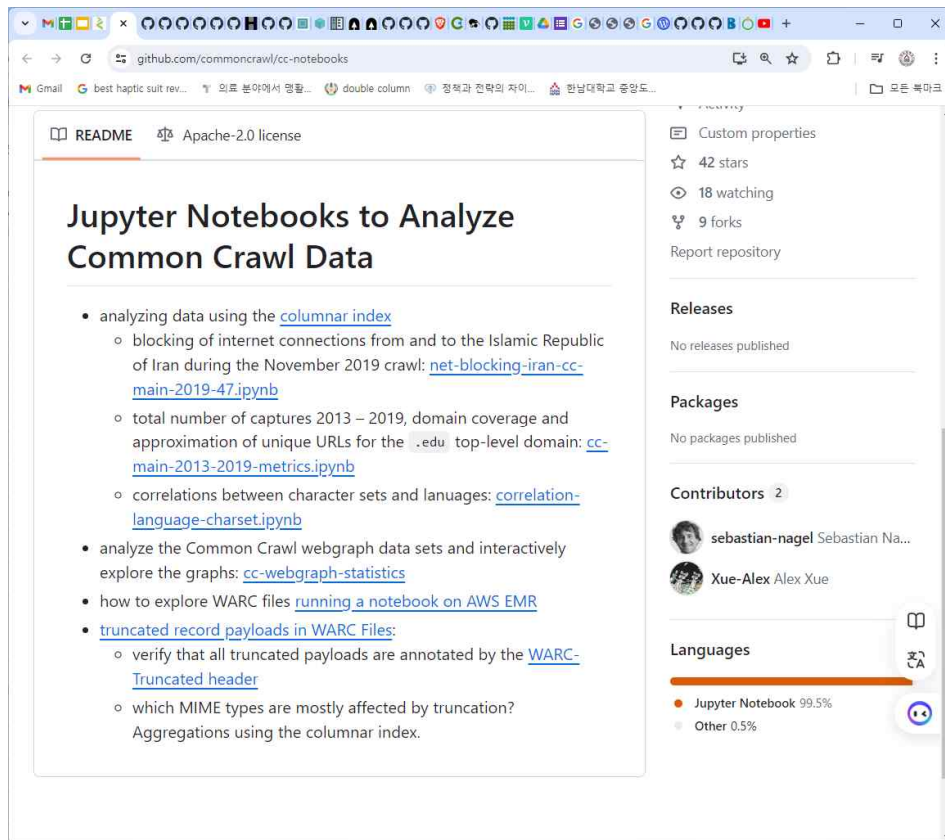
2.2.3 분석(Analysis)

1) SolrWayback

- 앞 절 2.1.2의 3) SolrWayback 내용과 동일하므로 앞의 절을 참조할 것

2) Jupyter Notebook

- Columnar Index(열형 인덱스)를 이용한 데이터 분석
 - 2019년 11월 크롤링 중 이란 이슬람 공화국과의 인터넷 연결 차단: net-blocking-iran-cc-main-2019-47.ipynb
 - 2013~2019년 총 캡처 수, 도메인 범위 및 .edu top-level 도메인에 대한 고유 URL의 근사치: cc-main-2013-2019-metrics.ipynb
 - 문자 집합과 언어 간의 상관 관계: correlation-language-charset.ipynb
- 공통 크롤링 웹 그래프 데이터 세트를 분석하고 그래프를 대화형으로 탐색: cc-webgraph-statistics
- AWS EMR에서 노트북을 실행하면서 WARC 파일을 탐색하는 방법
- WARC파일의 잘린 기록 페이로드:
 - 잘린 모든 페이로드가 WARC-Truncated header에 의해 주석 처리되었는지 확인
 - 잘림의 영향을 가장 많이 받는 MIME 유형은 무엇인지? 열형 인덱스를 사용한 집계



<그림 2-23> 오픈소스 Jupyter Notebooks 메인 화면

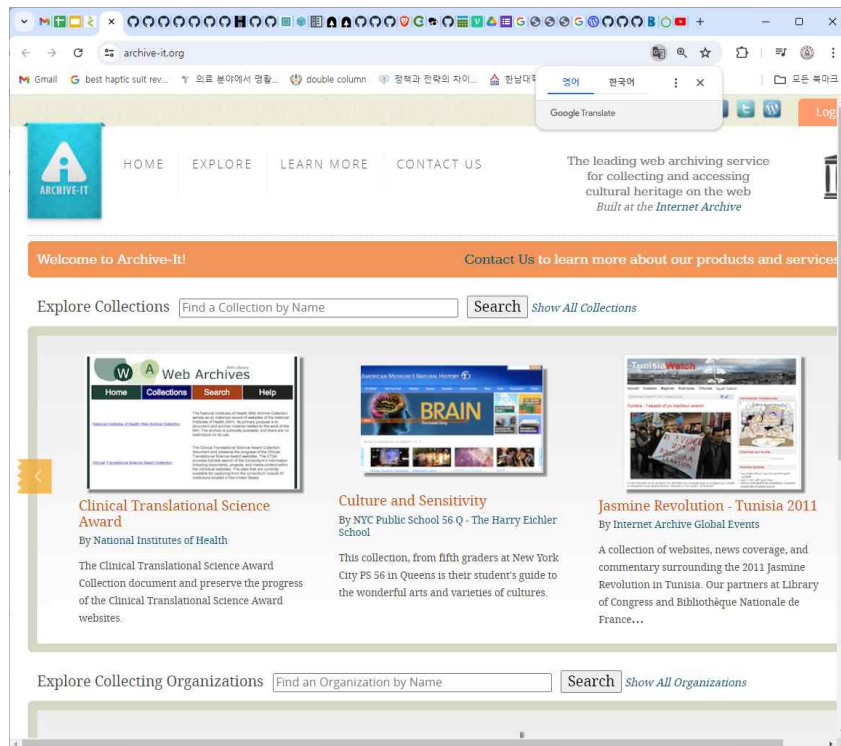
3) Analysis 도구 시사점

- 현재 4개 회원기관이 SolrWayback을 사용하고 있으며 3개 기관은 도입 중, 3개 기관은 도입 예정임
- Jupyter Notebook을 4개 기관이 사용하고 있고 한 기관(영국 국립도서관)은 과거에 사용하였다가 SolrWayback을 도입할 예정임

2.2.4 웹 아카이빙 서비스(Service)

1) Archive-It

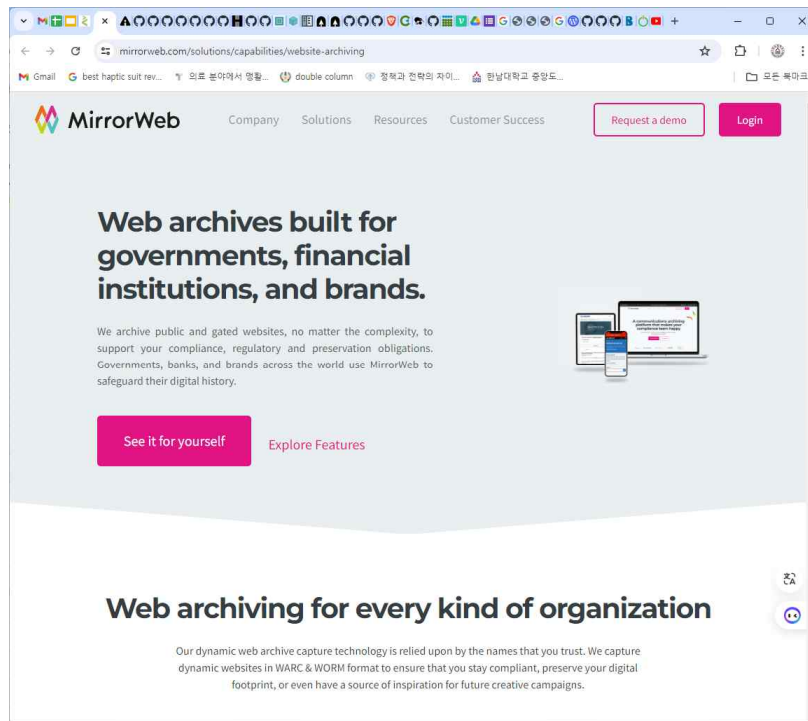
- 인터넷 아카이브(Internet Archive)에서 구축한 웹 상의 문화유산을 수집하고 이용할 수 있는 선도적인 웹 아카이빙 서비스



<그림 2-24> Internet Archive의 Archive-It 메인 화면

2) MirrorWeb

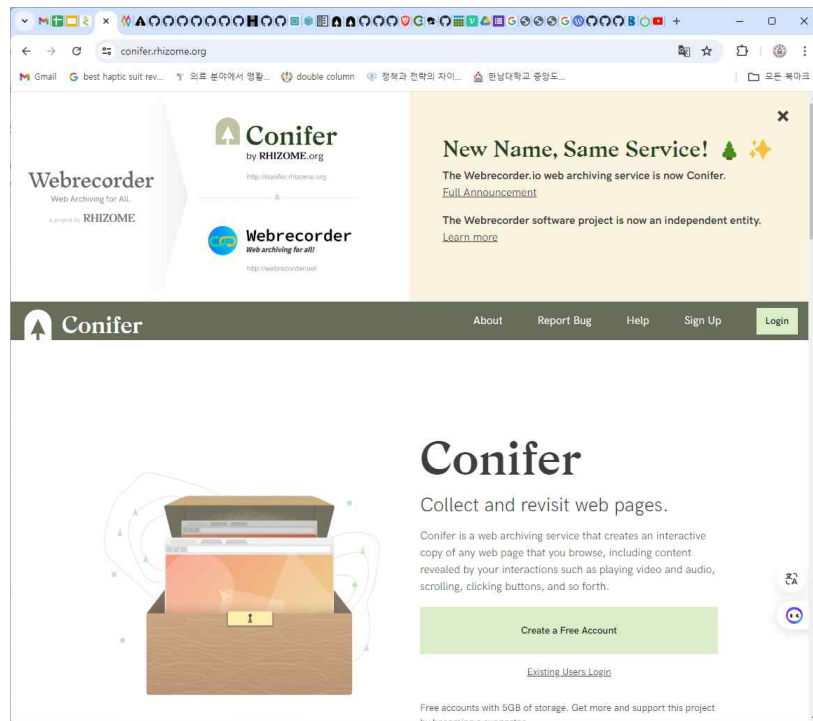
- 이용자의 규정 준수, 규제 및 보존 의무를 지원하기 위해 복잡성에 관계없이 공개 및 제한 웹사이트를 보존함. 전 세계의 정부, 은행 및 브랜드는 MirrorWeb을 사용하여 디지털 기록을 보호함
- 동적 웹 아카이브 캡처가 가능한 기술을 보유하고 있음. WARC 및 WORM 형식으로 동적 웹사이트를 캡처하여 규정을 준수하고 디지털 footprint를 보존하며 향후 창의적인 캠페인을 위한 영감의 원천이 되도록 보장함
- MirrorWeb은 정부, 금융 기관, 은행 및 브랜드와 협력하여 디지털 기록을 캡처하고 보존함. 지난 5년 동안 백만 개가 넘는 크롤링을 시작한 MirrorWeb은 모든 종류의 사용 사례와 규모에 맞게 동적 웹 아카이빙 기술을 운영함



<그림 2-25> 웹 아카이빙 서비스 MirrorWeb의 메인 화면

3) Conifer

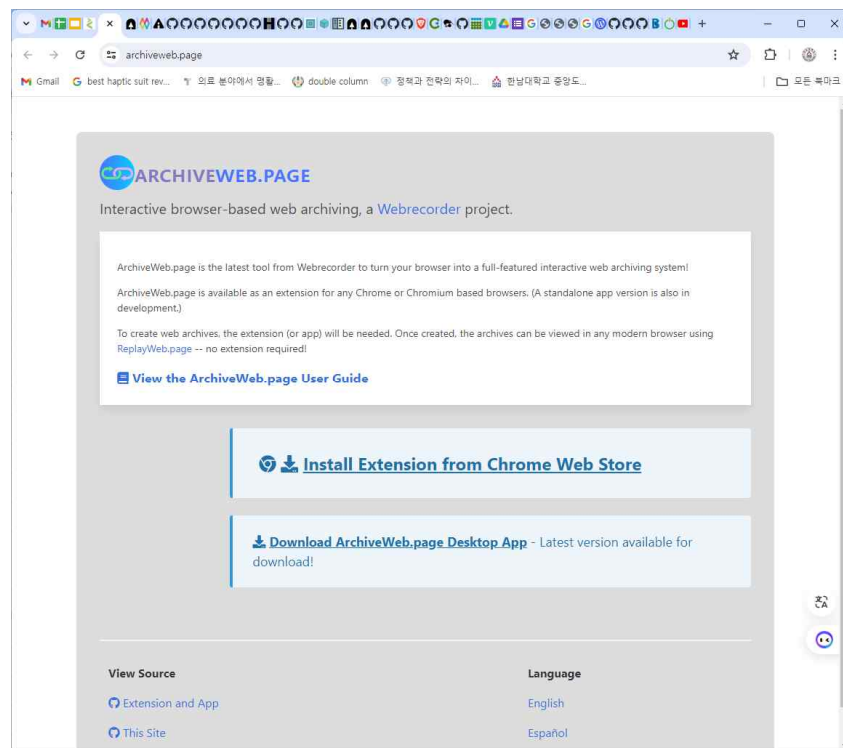
- 웹페이지를 수집하고 다시 방문하기 가능한 사용자 중심 온라인 서비스
- Conifer는 비디오 및 오디오 재생, 스크롤, 버튼 클릭 등과 같은 상호 작용을 통해 공개되는 콘텐츠를 포함하여 탐색하는 모든 웹 페이지의 대화형 복제본을 생성하는 웹 보존 서비스임
- 지금 온라인은 링크가 끊어지고 정보가 웹에서 지워지고 서비스가 사라지고 재설계가 발생함
- 기존의 크롤러 기반 웹 보존 방법과 달리 Conifer의 접근 방식을 사용하면 내장된 미디어, 복잡한 Javascript, 사용자별 콘텐츠 및 상호 작용, 기타 동적 요소가 포함된 복잡한 웹사이트도 캡처하고 다시 사용할 수 있음



<그림 2-26> Webrecorder의 Conifer 메인 화면

4) Archiveweb.page

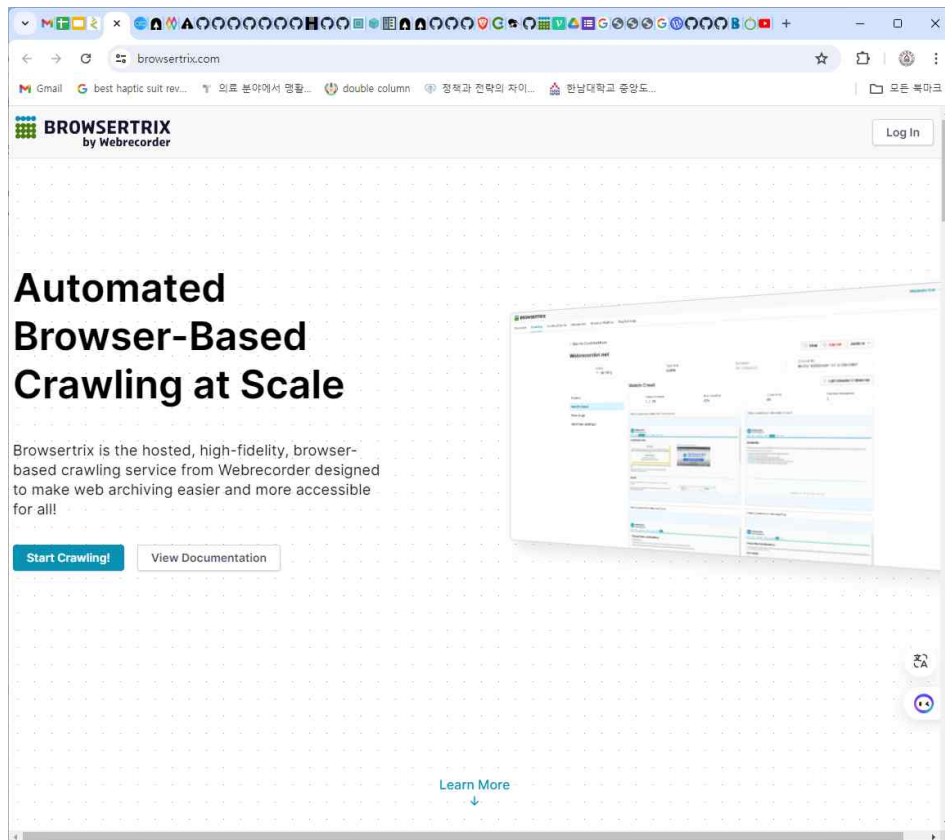
- ArchiveWeb.page는 브라우저를 모든 기능을 갖춘 대화형 웹 보존 시스템으로 바꿔주는 Webrecorder의 최신 도구임
- ArchiveWeb.page는 Chrome 또는 Chromium 기반 브라우저의 확장 프로그램으로 사용할 수 있음 (독립형 앱 버전도 개발 중)
- 웹 아카이브를 생성하려면 확장 프로그램(또는 앱)이 필요함. 일단 생성된 아카이브는 ReplayWeb.page를 사용하여 모든 최신 브라우저에서 볼 수 있어 확장이 필요하지 않음



<그림 2-27> Archiveweb.page 메인 화면

5) Browsertrix(subscription)

- Browsertrix는 모두가 웹 보존을 더 쉽고 더 쉽게 액세스할 수 있도록 설계된 Webrecorder의 호스팅 된 고성능 브라우저 기반 크롤링 서비스임



<그림 2-28> Browsertrix 메인 화면

6) Service 도구 시사점

- 현재 8개 회원기관이 Archive-It 툴을 사용하고 있고 한 기관(노르웨이 국립도서관)은 과거에 사용했었음
- 두 회원기관(영국 기록보존소, 미국 의회도서관)은 MirrorWeb 서비스를 이용하고 있음
- Conifer는 4개 기관이 이용하고 있고, 2개 기관이 사용했었음
- Archiveweb.page 서비스는 6개 기관이 이용하고 있으며 노스텍사스대학교는 사용할 계획임
- Browsertrix(subscription)는 한 기관(스탠포드대학교)이 이용하고 있으며 한 기관(노르웨이 국립도서관)이 도입 중에 있음

2.3 웹 아카이빙 우수 개선 사례

2.3.1 룩셈부르크 국립도서관 (BnL: Bibliotik National Luxemburg) 개선 사례⁵⁾

- 룩셈부르크 국립도서관(BnL)은 2016년부터 디지털 법적 납본으로 룩셈부르크 웹을 수집해 왔음
- 룩셈부르크 웹 아카이브는 전체 .lu 도메인의 광범위한 크롤링 외에도 특정 주제나 이벤트에 초점을 맞춘 타겟 크롤링을 수행함
- 법적 제한으로 인해 BnL의 웹 아카이브는 룩셈부르크에 있는 도서관 내에서만 볼 수 있음
- BnL은 2017년에 IIPC에 가입했고 2021년 온라인 총회 및 웹 아카이빙 컨퍼런스를 공동으로 조직했음



<그림 2-29> IIPC 블로그에 게시된 룩셈부르크 국립도서관 웹 아카이브 개선 사례

- 이전(migration) 계획

5) László Tóth, 룩셈부르크 국립도서관에서 Pywb로 마이그레이션, IIPC 블로그, 2024. 1. 22. <https://netpreserveblog.wordpress.com/2024/01/22/migrating-to-Pywb-at-the-national-library-of-luxembourg/>, 2024. 6. 22. 검색됨

- 2023년 동안 룩셈부르크 국립도서관(BnL)은 웹 아카이브를 새로운 인프라로 이전하는 작업을 수행
 - ▶ 하드웨어 : BnL은 인덱싱 및 재생과 관련된 애플리케이션을 호스팅하기 위해 4개의 새로운 고급 서버로 교체
 - ▶ 소프트웨어 : 오래된 OpenWayback 애플리케이션을 최신 Pywb/OutbackCDX 듀오로 업그레이드
 - ▶ 웹 아카이브 스토리지 : 339TB의 WARC 파일이 NetApp NFS에서 고성능 IBM S3 개체 스토리지로 마이그레이션

<표 2-10> 하드웨어 비교 - 이전 전 및 이전 후

비교항목	이전 전	이전 후
기술(S/W도구)	OpenWayback	SolrWayback + Pywb
전용 서버(H/W인프라)	1 대	4 대
메모리(GB)	24	3072
프로세서	8	192

○ 이전 전

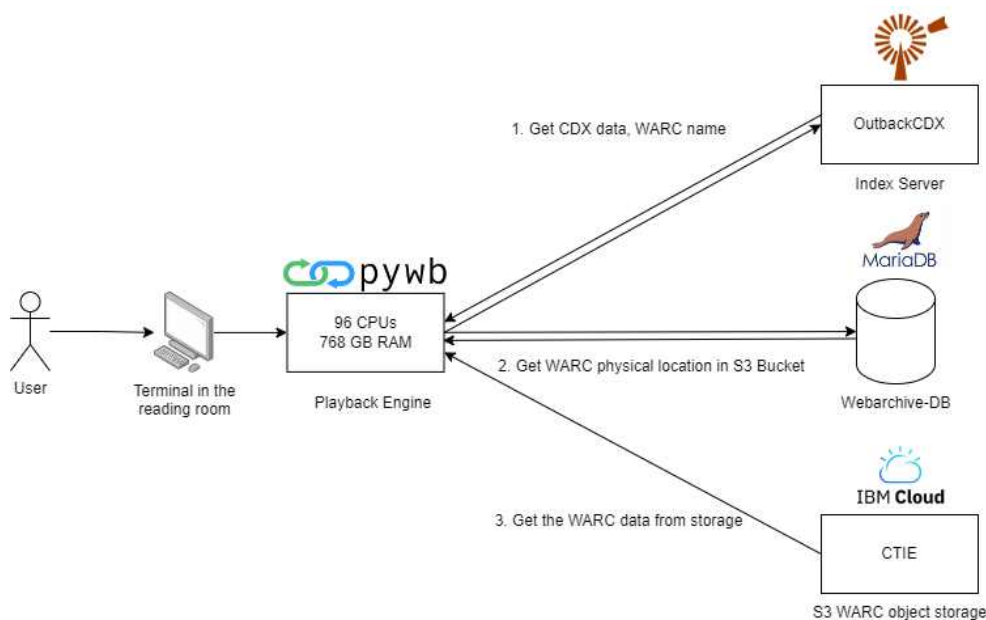
- 2023년 12월까지 BnL은 룩셈부르크 웹 아카이브에 접근하려는 사용자를 위한 재생 엔진(replay engine)으로 OpenWayback을 제공. OpenWayback은 간단하지만 느리고(그리고 사용하기 다소 번거로움) 인체공학적 사용자 경험과 아카이브의 효율적인 탐색에 필요한 여러 기능이 부족함
- 또한, WARC 파일은 로컬에 마운트된 NFS 공유 드라이브에 저장되었고, 보관 파일을 처리하고 클라이언트에 OpenWayback을 제공하는 전용 하드웨어 서버는 8개 코어와 24GB RAM이 있는 가상 머신으로 이 H/W 구성의 큰 단점 중 하나는 인덱스를 처리하는 방식임
- OpenWayback이 주어진 리소스에 액세스하려면 먼저 해당 리소스를 찾아야 했으며, 따라서 로딩 시간의 주요 원인은 CDX 파일 크기(총 3.1TB)와 해당 파일이 저장된 NFS 공유 드라이브의 느림에서 비롯됨
- 페이지 로딩 속도가 느리기 때문에 사용자는 웹 아카이브를 방문할 의향이 없음

○ 이전 과정

- 사용자에게 제공하는 서비스를 개선하기 위해, BnL은 Webrecorder가 개발한 재생 엔진인 Pywb와 호주 국립도서관이 개발한 고성능 CDX 인덱스 서버인 OutbackCDX로 전환하여 이러한 문제를 해결하기로 결정
- 이러한 애플리케이션을 호스팅하고 향후 SolrWayback 설정을 염두에 두고 BnL은 각각 96개의

CPU와 768GB RAM을 장착한 4대의 고급 서버를 구입

- 웹 아카이브가 스토리지 인프라의 상당 부분을 차지하기 때문에, BnL은 IBM S3로 이전하기로 공동으로 전략적 의사결정. 액세스 시스템을 마이그레이션하는 동안 스토리지 계층을 마이그레이션하는 것이 합리적이었기 때문에 먼저 이 작업을 수행



<그림 2-30> BnL의 웹 아카이빙 시스템 마이그레이션(Pywb + OutbackCDX 워크플로)

2.3.2 미국 의회도서관 (LC: Library of Congress) 개선 사례⁶⁾

- 미국 의회도서관(LC)의 웹 아카이빙 프로그램은 2023년에 23주년을 맞았음
- 초기에는 웹 아카이빙 워크플로를 시범 운영하다가, 보다 체계적으로 수집하는 방식으로 전환하여 컬렉션을 꾸준히 늘렸음
- 프로그램이 계속 성장하고 더 많은 콘텐츠가 온라인에 게시되고(예: 웹사이트, PDF 문서 및 보고서, 웹사이트 일부) LC와 협력하는 주제 및 언어 전문가에게 컬렉션 개발에 관심이 있는 콘텐츠가 늘어나면서, 더 많은 전문가가 참여했고 컬렉션이 더욱 커졌으며, 지난 팬데믹 시기에 보존하는 양이 더욱 증가함

6) Abbie Grotke, Improvements Ahead for the Web Archives, LC 블로그, <https://blogs.loc.gov/thesignal/2023/08/improvements-ahead-for-the-web-archives/>, 2024. 6. 10. 검색됨

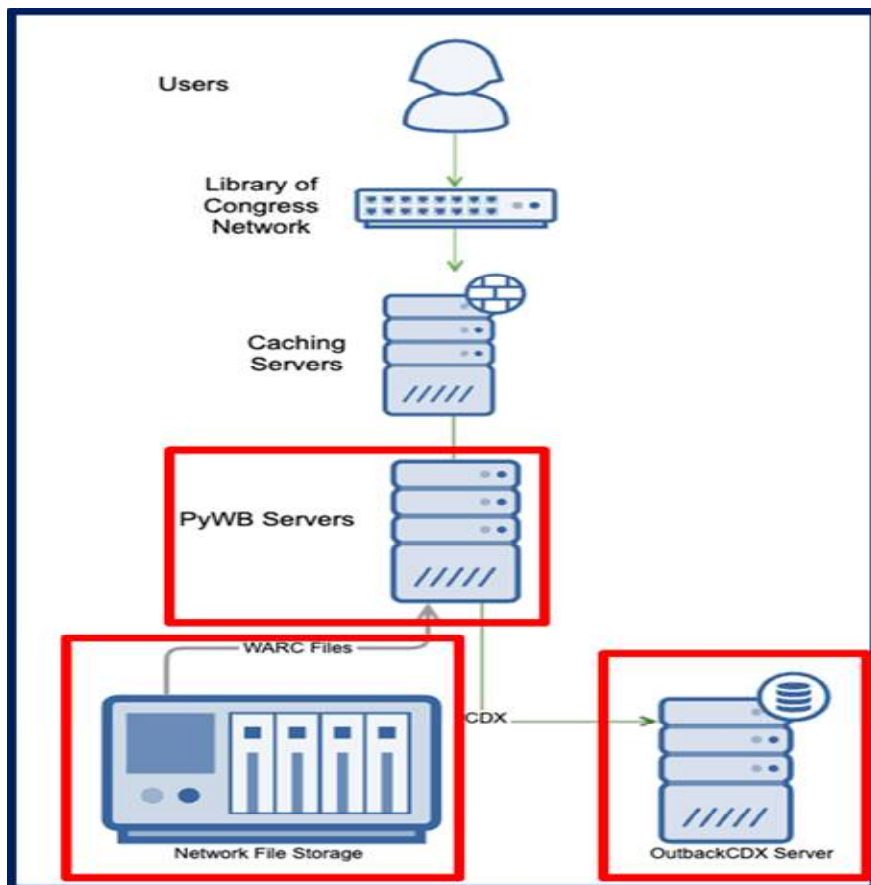
- 그러나 컬렉션에 대한 접근성에 영향을 미치는 규모가 되면서 문제 발생. 의회도서관 웹 아카이브는 현재 크기가 약 4페타바이트이며, 프로그램이 확장되고 보다 일상화된 수집 방법이 되면서 계속 문제가 점점 심각함



<그림 2-31> LC블로그에 게시된 LC 웹 아카이브 개선 사례

- 개선 과정 고려사항
 - 최근 몇 년 동안 웹 아카이브에 대한 접근성을 개선하기 위한 새로운 "차세대" 도구가 개발. Webrecorder에서 개발하고 웹 아카이빙 커뮤니티에서 지원하는 Pywb라는 오픈 소스 Python 기반 재생 시스템이 포함됨
 - IIPC는 OpenWayback에서 Pywb로의 잠재적으로 복잡한 전환에 대한 지원이 필요하다는 것을 알고 회원기관과 다른 사람들이 Pywb로 마이그레이션하는 데 도움이 되는 전환 가이드 개발에 자금을 지원함
 - 또한 재생 도구에서 렌더링하기 위해 웹 아카이브에도 CDX 파일이 필요하다는 점에 유의해야 함.

- CDX 파일은 연결된 메타데이터 줄로, 각 줄이 WARC 파일 내의 단일 객체를 나타냄
- 계획해야 할 또 다른 큰 변경 사항은 OpenWayback은 평평한 CDX 구조를 가지고 있으며, 규모와 Pywb로의 전환을 통해 평평한 구조를 OutbackCDX로 대체하여야 함. 그래야, 웹 아카이브 데이터를 보다 효율적으로 저장하고 조회할 수 있음
 - 4페타바이트의 웹 아카이브를 구성하는 수십억 개의 파일에 대한 모든 인덱스를 다시 수집하는 시간이 걸림



<그림 2-32> 미국 LC의 새로운 웹 아카이브 액세스 시스템의 구조(재생 SW(Pywb), 인덱스 서버(OutbackCDX) 및 WARC 저장소의 처리 흐름도)

- Pywb로의 전환을 위한 개발 작업이 우선 순위에 올랐으며, LC는 아직 설정해야 할 것이 몇 가지 있고, 23년간의 웹 아카이브에서 인덱스를 구축하는 데 시간이 걸릴 것으로 예상함
- 이러한 과도기에도 불구하고, LC는 의회도서관 웹 아카이브를 적극적으로 크롤링하고 추가하고 있고 최근 개발 중인 새로운 컬렉션에는 기후 변화 웹 아카이브, 대중 커뮤니케이션 웹 아카이브, Voices: Eastern and Central European Americans 웹 아카이브가 있고 이는 웹 콘텐츠를 계속 수집하고

보존하는 81개의 활성 이벤트 및 주제 컬렉션 중 일부임

2.3.3 우수 개선 사례 시사점

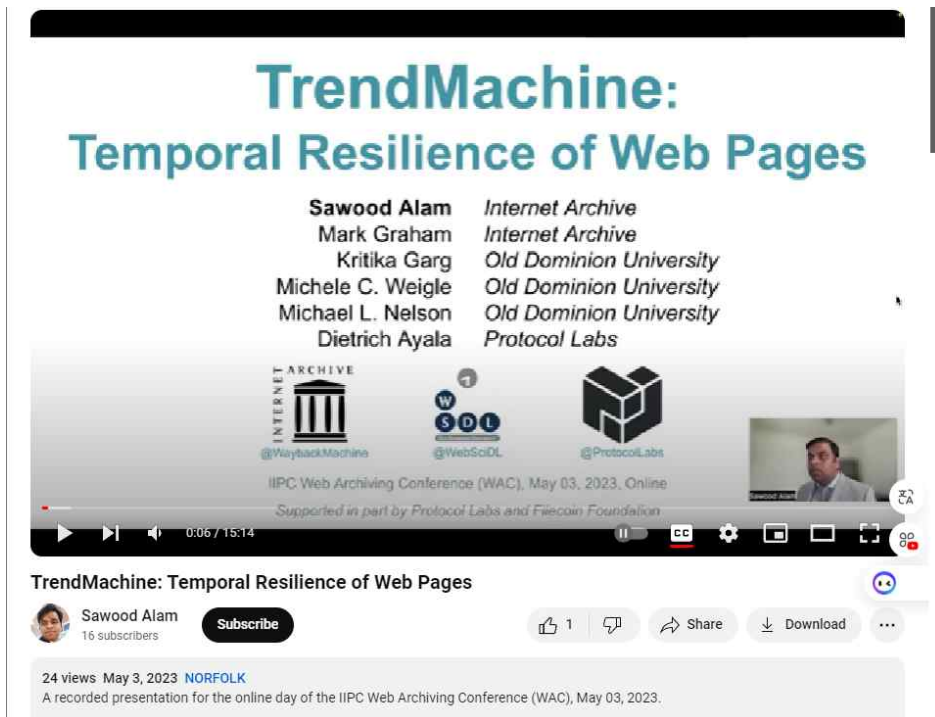
- 룩셈부르크 국립도서관(BnL)은 하드웨어와 소프트웨어에 대한 과감한 투자로 시스템 마이그레이션을 수행함
- 미국 의회도서관은 4페타바이트의 웹 아카이브를 마이그레이션(전체 인덱싱 포함)하는 매우 큰 규모의 프로젝트를 단계적으로 진행하는 것이 인상적임
- 두 기관 모두 공통적으로 재생SW로 Pywb를 사용하고 인덱싱 서버로 OutbackCDX 서버를 사용하여 최신의 SW 기술을 적용하려는 노력을 꾸준히 지속함
- 이는 20주년을 맞이하는 오아시스도 하드웨어와 소프트웨어에 대한 과감한 투자와 단계적인 마이그레이션 계획 수립이 필요함을 시사함

2.4 웹 아카이브 데이터 서비스 우수 개발 사례

2.4.1 Internet Archive

1) TrendMachine

- 출처: Sawood Alam, Kritika Garg, Michele C. Weigle, Michael L. Nelson, Mark Graham, and Dietrich Ayala. 2024. TrendMachine: A Temporal Webpage Resilience Portal. In Proceedings of the 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL '23). IEEE Press, 93-97. <https://doi.org/10.1109/JCDL57899.2023.00023>
- 전 세계 웹을 아카이빙하는 비영리기관인 Internet Archive 소속 연구원과 Old Dominion University 소속 교수 및 연구원들이 2023년 IIPC 총회의 온라인 세션에서 발표한 내용
- 또한, ACM/IEEE Joint Conference on Digital Library(JCDL) 2023(2023.6.26. - 30, Santa Fe, New Mexico, USA)의 Workshop인 WADL(Web Archive and Digital Library)에서 발표된 내용임
- 웹 아카이브에서 페이지에 대한 과거 관찰 결과를 기반으로 시계열 데이터로 웹페이지의 시간적 회복력을 정량화하기 위해 정규화된 점수를 계산하는 수학적 모델임



<그림 2-33> 2023년 IIPC 총회의 온라인 세션에서 저자가 TrendMachine에 대해 발표하는 장면

Sigmoid Function for Web Page Resilience

$$Resilience_t = \frac{Spread}{1 + e^{Shift - \frac{t}{Slope}}}$$

Spread: How far up or down the value can go from its starting position?

Shift: How soon any significant change in the value can begin?

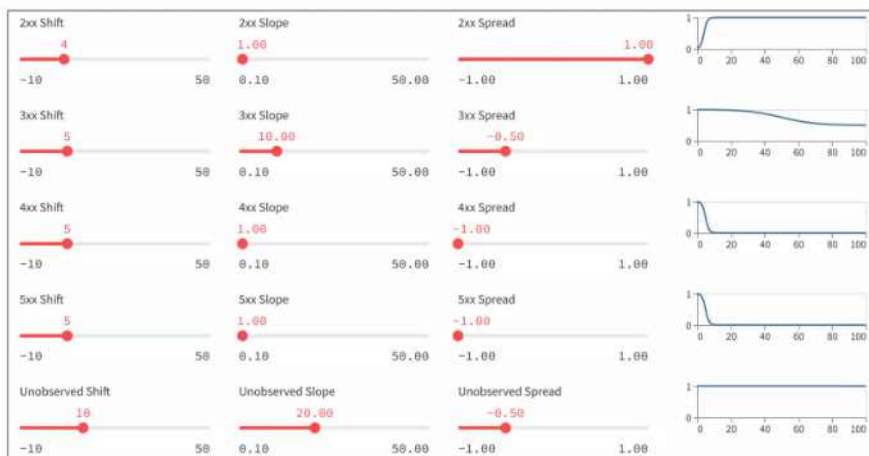
Slope: How quickly the value reaches close to the maximum change?



TrendMachine: Temporal Resilience of Web Pages | IIPC WAC 2023 | Sawood Alam <@ibnesayeed>

<그림 2-34> 웹페이지 회복력에 대한 수학적 모델(시그모이드 함수)

TrendMachine: Composite Sigmoid Parameters of Resilience



TrendMachine: Temporal Resilience of Web Pages | IIPC WAC 2023 | Sawood Alam <@ibnesayeed>

6

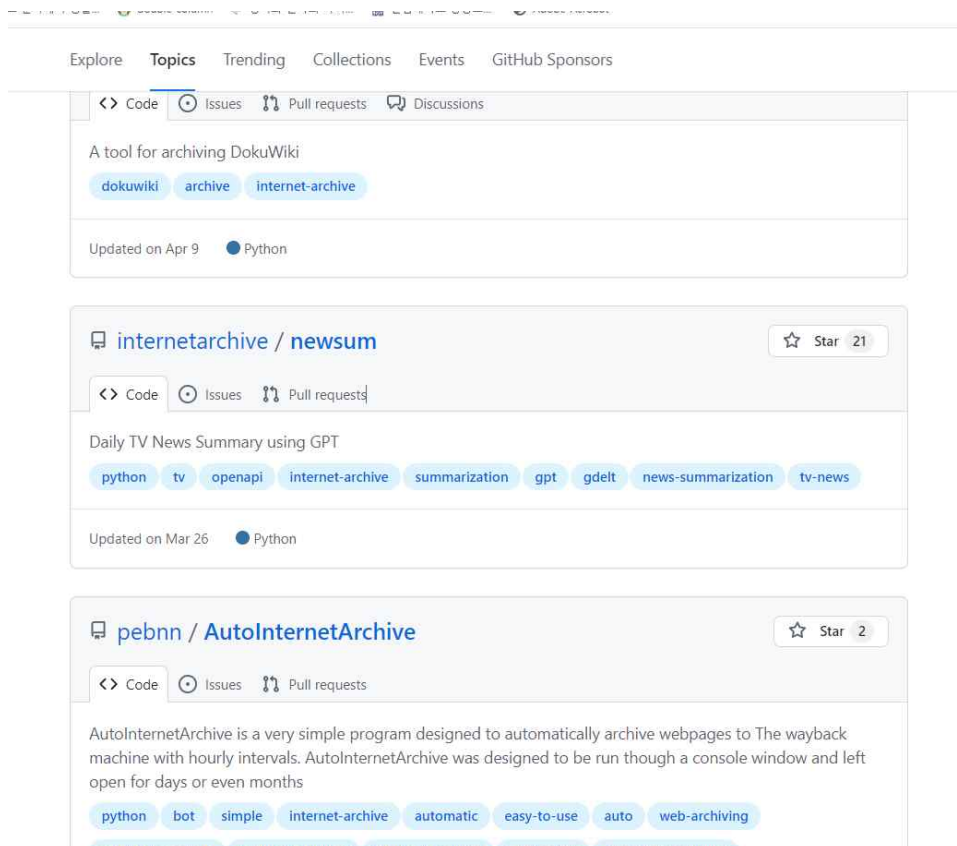
<그림 2-35> 웹페이지 회복력 가시화 예시 화면

2) NewSum

- 출처: Sawood Alam, Mark Graham, Roger Macdonald, and Kalev Leetaru, Utilizing Large Language Models for Semantic Search and Summarization of International Television

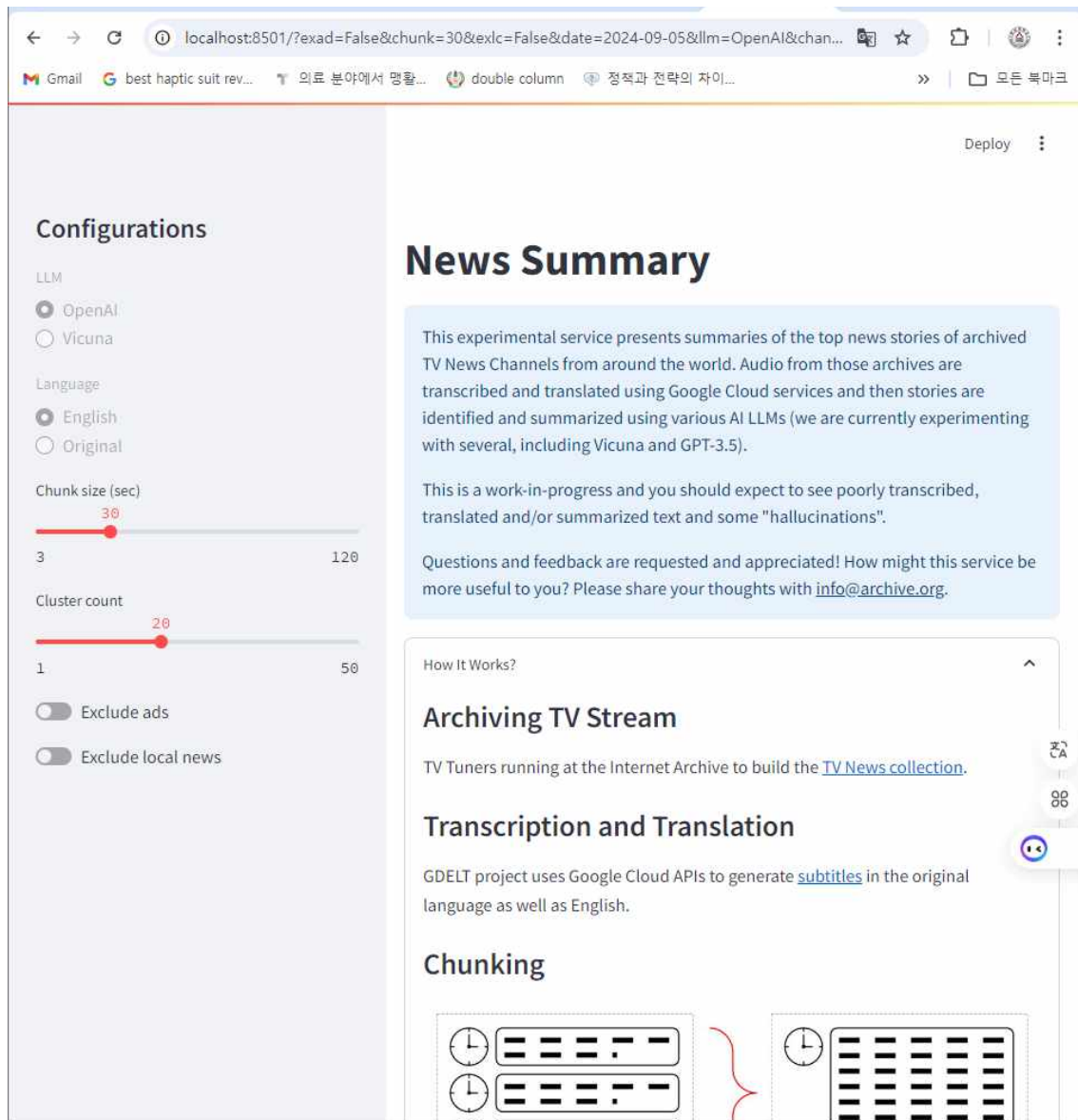
News Archives, IIPC General Assembly & Web Archiving Conference, 2024.4.24.-26.
National Library of France, <https://netpreserve.org/ga2024/abstracts/>

- 전 세계 웹을 아카이빙하는 비영리기관인 Internet Archive 소속 연구원과 GDELT⁷⁾ 프로젝트 연구원들이 2024년 IIPC 총회의 ARTIFICIAL INTELLIGENCE & MACHINE LEARNING 세션에서 발표한 내용임
- 이 애플리케이션은 GPT를 사용하여 인터넷 아카이브의 TV 뉴스 아카이브의 일일 뉴스를 요약함



<그림 2-36> GitHub에 오픈 소스로 공개된 Internet Archive의 NewSum

7) 사회적 글로벌 데이터베이스, Google Jigsaw의 지원을 받는 GDELT 프로젝트는 전 세계 거의 모든 국가의 100개 이상의 언어로 된 방송, 인쇄 및 웹 뉴스를 모니터링하고 매일 매초 우리의 글로벌 사회를 움직이는 인물, 위치, 조직, 주제, 출처, 감정, 횟수, 인용문, 이미지 및 이벤트를 식별하여 전 세계에서 컴퓨팅할 수 있는 무료 개방형 플랫폼을 구축하는 프로젝트임



<그림 2-37> TV 뉴스 요약 서비스 newsum docker image 실행 화면

2.4.2 데이터 서비스 우수 개발 사례 시사점

- Internet Archive에서는 지금까지 수집한 웹 아카이브 자료를 단순히 검색하여 재생하는 데 그치지 않고 특정 웹사이트의 페이지들이 얼마나 회복력 있게 지속되어 오는지에 대한 평가를 회복력(resilience)이라는 지표를 통해 측정하여 분석하려는 시도를 함. 수집하여 보존하고 있는 자료에 대한 세밀한 데이터 분석 기능을 제공하려는 노력을 지속적으로 기울임
- 마찬가지로 Internet Archive에서 제안한 또 다른 서비스로 최근 대규모 언어모델(Large Language Model)과 생성형 인공지능(ChatGPT or Vicuna)을 이용하여 TV뉴스를 요약하는 NewSum을 제공하여 최신 IT기술을 적극 수용하는 전략 방향을 취함

- 최근 전자저널 출판기업(Ebsco, Elsevier, Proquest)들은 데이터 분석기업으로 탈바꿈할 정도로 내부적으로 수집한 자료에 대하여 이용자들에게 좀 더 세밀한 분석을 할 수 있는 서비스 제공하려는 노력을 경주함

III

웹자원 아카이빙(OASIS) 시스템 데이터 현황 진단과 분석

3.

오아시스 시스템 데이터 현황 진단과 분석

- 본 장의 웹 아카이빙(오아시스) 시스템 진단 및 요구 조사·분석은 오아시스를 다루는 실무자, 전문가 의견과 오아시스 누리집 자체를 조사하고 분석함을 통해 오아시스의 요구사항을 다각도에서 파악하고 시사점을 제공하는 것을 목적으로 함
- 오아시스 현황 진단 및 요구 조사·분석은 실무자 및 전문가 심층 인터뷰를 진행하며 오아시스 누리집 사용성 평가를 통한 현황 진단 및 분석, 우수 사례 분석, 오아시스 성능 분석 순으로 기술함
 - 실무자(웹자원 수집/색인 및 검색) 심층 인터뷰를 통하여 표본이 너무 적어서 경향성 조사에서 추출할 수 없는 세부적인 내용을 심층 인터뷰 방법을 통하여 문제점 및 요구사항 도출
 - 대학생 예비 사서 47명으로부터 누리집을 이용하고 누리집의 전체 화면과 각 메뉴의 기능들을 사용하고 사용성 및 문제점, 개선점 조사를 진행하고 그 결과를 분석하는 방법으로 요구사항 도출
 - 우수 선진 웹 아카이브 누리집을 벤치마킹하여 오아시스 누리집에서 필요로 하는 요구사항을 도출함
 - 웹사이트 성능 분석 도구(Gmetrix)를 이용하여 오아시스 누리집의 성능을 분석하고 현황 진단 및 시사점을 도출함

3.1 오아시스 실무자 의견 조사

- 본 연구에서는 오아시스 개선 및 활용성 강화에 대한 다양한 의견을 수렴하기 위해 오아시스 실무자(6명)를 대상으로 FGI를 실시하였음

3.1.1 면담 조사 설계

- 조사 개요
 - 오아시스 포괄적 수집 실무자 6명을 대상으로 면담 조사를 실시함
 - 일시: 2024년 7월 22일 ~ 8월 14일(화)
 - 방식: 온라인 면담 질문지 작성
 - 조사내용 : 대한민국 웹 아카이빙(오아시스) 시스템에 대한 인식, 오아시스 시스템에 대한 인식 및 요구 파악(수집기, 수집 데이터, 색인기, 색인 데이터), 오아시스 시스템 개선 전략 및 방향성 등 제안

<표 3-1> 실무자 심층 면담 질문 및 내용

실무자	직무 분야	성별
실무자1	웹자원 수집 및 관리	남
실무자2	웹 데이터 수집 및 품질 관리	여
실무자3	전산 시스템 관리	남
실무자4	시스템 유지 운영	남
실무자5	시스템 유지 운영	남
실무자6	누리집 프로그램 개발 및 운영	남

○ 실무자 심층 면담 질문 구성 내용은 다음과 같음

<표 3-2> 오아시스 실무자 그룹: 심층 면담 질문 및 내용

면담 단계	질문 항목		질문 내용	No
시작 단계	대한민국 웹 아카이빙(오아시스) 시스템에 대한 인식		- 귀하가 중요하게 생각하는 우리나라 웹 아카이빙(오아시스) 서비스의 역할 및 위상은 무엇이라고 생각하십니까? (수집 목적, 수집 범위, 보존 품질, 활용도, 서비스 종류, 서비스 품질 등의 측면 중에서)	Q1
이행 및 핵심 단계	오아시스 시스템에 대한 인식 및 요구 파악	수집 과정의 특징과 문제점	- 현재 수집기(예.Heritrix)의 특징과 수집 과정에서 발생하는 문제점은 무엇이라고 생각하십니까?	Q2
		수집 데이터 특성 및 품질	- 현재 수집되고 있는 웹사이트와 WARC 파일의 특성 및 문제점은 무엇이라고 생각하십니까?	Q3
		색인기 특성과 문제점	- 색인기의 특징과 색인 과정에서 발생하는 문제점은 무엇이라고 생각하십니까?	Q4
		색인 데이터 특성 및 품질	- 귀하가 직무를 수행하시거나 경험하면서 평소에 인식하셨던 색인된 웹사이트의 검색은 잘 수행된다고 생각하십니까? 그 이유는 무엇인가요?	Q5
종결 단계	오아시스 시스템 개선 전략 및 방향성		- 귀하가 직무를 수행하시거나 경험하면서 평소에 가졌던 오아시스의 활용성 강화를 위한 개선점 및 방향성은? (시스템, 관리 및 운영 조직, 예산, 수집, 색인/검색, 웨이백 재생, 누리집 서비스 종류, 서비스 품질 등의 측면에서)	Q6

3.1.2 면담 조사 분석 결과

○ 실무자 심층 면담 주요 조사 결과는 아래 표와 같음

<표 3-3> 실무자 심층 면담 주요 조사 결과

Q1. 귀하가 중요하게 생각하는 우리나라 웹 아카이빙(오아시스) 서비스의 역할 및 위상은 무엇이라고 생각하십니까? (수집 목적, 수집 범위, 보존 품질, 활용도, 서비스 종류, 서비스 품질 등의 측면 중에서)	
인터뷰 내용	요약

- 1. 수집 목적은 국내 웹자원 정보를 수집하여 이를 보존하고 공유 활용 할 수 있도록 주제별 고품질 웹자원을 수집하여 구축 체계 마련 및 서비스 제공, 관리 및 서비스를 통하여 국가 웹자원 정보를 확산할 수 있는 기반 조성, 수집 구축된 웹자원 정보의 국제 자원 공유라고 생각함
- 2. 수집 범위는 국가 도메인 신규 웹사이트, 주기적으로 도메인 웹자원 재수집, 국가 재난, 행사, 주요 주제 및 이슈 수집하여 컬렉션 구축할 필요가 있음
- 3. 품질 보전은 수집되는 웹자원에 대하여 수집 공정 시부터 철저한 공정으로 품질을 확보하여 보존해야 함. 수집 대상 도메인 검증 (유효성 검사), 중복 조사, Heritrix 수집 -> 육안 전수 검수 (웹사이트와 WARC 파일), 레벨별(깊이) 수집 확인 및 링크 확인 (통상 3레벨), 악성코드 검사, 수집 서버에 업로드 등록, 이관 및 WARC파일 색인 확인

구축방법

- 준비단계
 - 도서관 : 수집대상 URL 제공
 - 현황분석 및 URL 오류 검증
- 구축단계 (Heritrix 수집기 운영)
 - 수집기 스케줄 환경설정 (컬렉션 신규 등록)
 - 수집기 운영 (정상 동작 점검)
 - 수집 및 웹자원 아카이브
 - 내용 확인 : 정상 수집 여부
 - 국가재난아카이브 및 주제·이슈 컬렉션 구축
- 검사단계(대상 : 수집 자료)
 - 담당자 : 육안 검사 및 링크확인
 - 입력정보 확인 및 검사 / 예외데이터 처리 및 별도 목록 관리

- 납품 및 관리
 - 담당자 : DB구축업체 > 품질관리 담당자 및 팀장
 - 검토완료 확인 및 납품처리
- 활용도는 소멸되기 쉬운 국가 웹자원을 수집 아카이빙하여 체계적으로 보존, 수집된 다양한 웹자원 서비스 (웹트렌드, 컬렉션 및 이슈별, 웹자원 메타), 국내외 웹자원 유통 및 상호 협력, 국가 디지털 장서로 활용함
- 서비스 종류는 웹자원 URL, 메타 검색 서비스 (사이트, 웹문서, 동영상, 이미지 등), 주제, 이슈 컬렉션 서비스, 재난 아카이브, 소셜미디어 아카이브, 웹트렌드임
- 서비스 품질은 서비스되는 웹자원의 품질은 양호하나 데이터 색인 처리 및 데이터 서비스 구현은 개선이 필요하다고 판단함 정확한 검색 요건 제시 및 구현 (메타, 키워드, 컬렉션 등), 검색 시 유사 사이트 웹자원 정보 제공 등, 검색 시 속도 저하 문제점 파악 및 개선할 필요가 있음 (실무자1)
- 인터넷을 통해 생성되고 사라지는 방대한 한국 관련 지식정보자원. 웹 데이터를 최대한 수집 및 보존하여 후대에 전승하는 국내 유일의 프로젝트로, 온라인 지식정보자원을 국가가 보존하고 영구 활용하고자 하는 시도 자체로 의의가 큼

- 국내 웹자원 수집하여 정보를 확산할 수 있는 기반을 마련하는 것이 가장 중요, 국제 자원 공유를 할 수 있어야 함
- 국가 재난, 행사, 주요 주제 및 이슈 수집하여 컬렉션 구축할 필요가 있음
- 품질 보증(QA)은 수집 공정 시부터 철저한 공정으로 품질을 확보하여 보존해야 함.
- 국내외 웹자원 유통 및 상호 협력, 국가 디지털 장서로 활용하고 있어 현재 기조를 유지하면서 추가적 확대 방안을 시도할 필요가 있음
- 웹자원의 품질은 양호하나 데이터 색인 처리 및 데이터 서비스 구현은 개선이 필요하다고 판단함. 검색 시 속도 저하 문제점이 있는 것으로 파악되는데 개선할 필요가 있음

<ul style="list-style-type: none"> - 현재 여건에서 최대한 수집하고 있지만, 수집 및 보존 품질 개선, 메타데이터 추출 및 제공 기능 강화, 검색 및 서비스 품질 개선, 데이터 활용성 강화 등 전반적인 개선, 후속 조치가 필요한 상황임 (실무자2) - 보존 품질과 활용도가 높아졌으면 함 (실무자3) - 관리자가 설정한 대상의 사이트를 보존하는 것에 주요 목적으로 판단됨 (실무자4) - 디지털 기록들은 휘발성이 강하고 분산되어 있어 관리와 보존이 어려운데, 웹 아카이빙은 디지털 기록물을 후대에 전해지도록 웹 콘텐츠를 수집 및 보존하는 것임 (실무자5) - 관리자가 설정한 수집 대상의 사이트를 크롤링하여 크롤링된 데이터를 보존하는데 목적을 두고 있다고 생각함 (실무자6) 	
--	--

Q2. 현재 사용 중인 수집기(예, Heritrix)의 특징과 수집 과정에서 발생하는 문제점은 무엇이라고 생각하십니까?

인터뷰 내용	요약
<ul style="list-style-type: none"> - 외부 웹자원 수집 특징으로 Heritrix 수집기는 웹 페이지의 구조와 콘텐츠를 최대한 원본 그대로 유지하면서 장기간 보존할 수 있는 형식으로 저장할 수 있음. 대규모 웹 크롤링 작업을 수행할 수 있도록 설계되었으며 오류 상황에 대해 재시도 기능을 제공함. Heritrix는 HTML의 링크 속성을 분석해서 수집함. 현대의 웹사이트는 동적인 페이지를 만들기 위해 Javascript를 사용하여 페이지를 생성함. Heritrix는 웹페이지의 Javascript를 실행할 수 없기 때문에 동적으로 생성되는 내용을 수집하지 못하여 불완전한 수집이 이루어짐 - 외부 웹자원 수집 과정 및 문제점으로 첫째, 웹자원 수집 검증 (사이트와 WARC 파일 비교), 둘째, 웹자원 수집 성공 WARC 파일 업로드, VPN 활용 HIWARE SFTP로 업로드 시 네트워크 및 서버 색인 전송 속도 저하(새로운 업로드 방식 및 색인 이관 프로세스 모듈 개선 필요). 전송 중 자주 장애 발생 (중단 및 수집 서버 상태 점검 어려움), 셋째, 국중 DCMS 시스템 등록 확인 (실무자1) - 헤리트릭스의 경우 텍스트 위주로 수집이 되며, 메인화면의 경우 어느 정도 수집이 가능함 - 다만, 동적 이미지 수집이 되지 않으며, 계층구조의 깊이있는 수집이 되지 않음 - 서브 메뉴 수집 안되며, 첨부파일도 수집이 되지 않음. - 영상, 이미지 등 대용량 동적인 웹사이트로 진화하는 추세임에 반해 수집 품질이 개선되지 않으며, SNS도 수집이 되지 않으므로 현재의 웹사이트를 수집하는 데 한계가 큼 (실무자2) - 악성코드까지 수집하는 경우가 있는데 이에 대한 검사가 누락되는 문제가 있음 (실무자3) - 국중 네트워크 환경 특성상 수집 속도 및 네트워크 트래픽 문제가 있다고 생각됨 (실무자4) - Heritrix는 웹크롤링 작업을 효율적으로 처리할 수 있음 	<ul style="list-style-type: none"> - 현재 웹 수집기로 Heritrix를 사용하고 있으며, 현대 웹사이트는 동적 웹페이지를 만들기 위해 Javascript를 사용하는 것이 특징인데, Heritrix는 Javascript를 실행할 수 없어 동적 웹페이지 수집에 있어 불완전한 수집을 보임 - 외부 웹자원 수집 과정은, 1) 웹자원 수집 검증 (사이트와 WARC 파일 비교), 2) 웹자원 수집 성공 WARC 파일 업로드, 3) 국중 DCMS 시스템 등록 확인이며, 수집 과정 문제점으로 1) VPN 활용 HIWARE SFTP로 업로드 시 네트워크 및 서버 색인 전송 속도 저하(새로운 업로드 방식 및 색인 이관 프로세스 모듈 개선 필요), 2) 전송 중 자주 장애(중단 및 수집 서버 상태 점검 어려움) 발생함

<p>나, 많은 시스템자원을 필요로 하며, 시스템이 복잡하여 유지운영에 어려움이 있고, 외부 웹사이트를 수집하는 부분이라 보완에 취약하다고 생각함 (실무자 5)</p> <p>- Heritrix는 WARC 파일 포맷을 사용하여 크롤링된 데이터를 저장함. 하지만 정적 콘텐츠를 크롤링하는 데 강점을 가지며, 동적인 콘텐츠 처리에는 제한적일 수밖에 없다는 문제점을 가지고 있음 (실무자 6)</p>	
Q3. 현재 수집되고 있는 우리나라 웹사이트와 WARC 파일의 특성 및 문제점은 무엇이라고 생각하십니까?	
인터뷰 내용	요약
<ul style="list-style-type: none"> - 도메인만 등록하고 실제 운영하지 않는 웹사이트가 많아서 수집 및 검증에 불필요한 시간을 초래 - 1. 국내 웹사이트는 이미지 활용 사이트가 많고 콘텐츠 서비스 구현이 복잡함 (Heritrix의 수집 문법 및 특성에 맞지 않게 구현된 사이트가 존재) - 2. 서비스 웹사이트 노출의 속도가 늦음. WARC 파일은 대규모 웹사이트를 저장하는 경우 개별 파일의 크기가 수백 기가바이트를 넘고 있으며 대용량 파일을 저장 및 이동하는데 어려움이 있음 (실무자1) - WARC 파일 특성상 Wayback 서비스 시, 압축 푸는 과정에서 로딩 시간이 오래 걸림 (실무자3) - Solr 색인으로 국가 도메인 웹사이트 수집 자료에 full-text 색인을 추출하고 있으나, 웹사이트 1건 색인에 2~3주 소요되는 경우 등 속도 지연이 심각하며, 검색 및 이용에 최적화되어 있지 않음 (실무자2) - 백 해외 웹사이트와 비교 시 우리나라 웹사이트는 용량이 큰 편이기에 수집되는 WARC 파일의 크기 역시 큰 사이즈이며, 이로 인해 WAYBACK 재생 속도가 저하되는 문제 등이 발생하게 됨 (실무자4) - 우 - 다양한 콘텐츠를 제공하고 있으나, 웹사이트가 디자인 요소가 많이 들어가 무겁고, 웹 접근성이 떨어지는 것 같음. 따라서 WARC파일 용량도 커지게 되고 웨이백 재생 시 속도가 걸리는 것 같음 (실무자5) - 감 - WARC 파일은 웹페이지와 관련된 데이터(이미지, 비디오 등 다양한 웹 리소스)를 포함하는 특징을 가지지만 웹페이지의 전체 콘텐츠를 저장하기 때문에 파일 크기가 매우 클 수 있으며, 저장 공간과 관리의 어려움을 초래할 수 있음 (실무자6) 	<ul style="list-style-type: none"> - 다수의 도메인 등록 후 미운영 웹사이트 수집 및 검증에 불필요한 시간 초래 - 이미지 활용도가 높은 국내 웹사이트의 특징 때문에 Heritrix 수집 문법 및 특성에 맞지 않은 사이트 다수 존재 - 전반적으로 수백 기가 바이트(gigabyte, GB)가 넘는 웹사이트의 대용량화 때문에 웹사이트를 저장하고 이동, 화면 제시하는 데 속도가 늦음
Q4. 색인기의 특징과 색인 과정에서 발생하는 문제점은 무엇이라고 생각하시나요?	
인터뷰 내용	요약
<ul style="list-style-type: none"> - 대용량 WARC 파일을 색인하고 검색하는 시간이 오래 걸리고 있으며 색인을 별도로 유지하는 것도 용량을 차지합니다. - 연속성 있는 Solr 검색엔진 구현으로 기수집된 데이터 및 수집 중 데이터를 원활하게 처리할 수 있게 재색인 및 방안 마련이 필요 (실무자1) - 오아시스는 오픈소스 Solr를 사용 중으로, 검색과 관련 	

<p>된 검색 사전이 개발되지 않았기에 기본적인 형태소 분석만 되는 상황(현재 조사 처리만 되는 중). 검색 기능을 향상 시키기 위해서는 사용 검색 엔진 도입이 필요한 듯 (실무자4)</p> <ul style="list-style-type: none"> - 여러 개의 파일을 병합하여 관리가 용이하나 파일 용량이 커져 성능이 떨어질 수 있음 (실무자5) - Solr 오픈소스를 사용 중으로 다양한 검색 요구를 충족할 수 있는 많은 기능을 갖추고 있는 특징이 있지만 대량의 데이터를 색인할 때 색인 지연이 발생할 수 있음 (실무자6) - 색인기를 통하지 않고 별도 사업을 통해 수집하는 경우, 문제가 있음 (실무자3) - Solr 색인으로 국가 도메인 웹사이트 수집자료에 full-text 색인을 추출하고 있으나, 웹사이트 1건 색인에 2~3주 소요되는 경우 등 속도 지연이 심각하며, 검색 및 이용에 최적화되어 있지 않음 (실무자2) 	<ul style="list-style-type: none"> - 대용량 WARC 파일을 색인하고 검색하는 시간이 오래 걸리고 있으며 색인을 별도로 유지하는 것도 용량을 차지함 - 색인 자료에 대하여 전문 검색(full-text)이 가능하도록 Solr 검색엔진 구현에 의한 재색인 방안 마련이 필요함
--	---

Q5. 귀하가 직무를 수행하시거나 경험하면서 평소에 인식하셨던 색인된 웹사이트의 검색은 잘 수행된다고 생각하십니까? 그 이유는 무엇인가요?

인터뷰 내용	요약
<ul style="list-style-type: none"> - 수집 WARC 파일 정보, 메타 등 색인 프로세스 강화 필요 - 검색 속도가 느리고 다양한 검색 조건이 필요 (실무자1) - 선택적 웹사이트의 경우 메타데이터 구축하고 있으나 오아시스 누리집에서 검색 시 검색 결과 표출이 누락되는 경우가 많으며, 표출 순서도 문제가 있음 - 정확성과 재현율이 모두 떨어짐 - 메타데이터 및 전문 검색(full-text)이 제대로 반영되지 않고 있음 - 웹사이트의 경우 표제, 저자, 발행자 등 일관된 형식이 아니라 표준화가 어려우며, 색인어가 다양한 형식으로 구축되어야 하며, 검색에 반영이 되어야 함 (실무자2) - 무난하게 수행되고 있다고 생각함 (실무자3) - 소장자료 검색은 잘 수행되는 것 같으며 검색엔진 한계로 형태소 검색 등은 안되는 듯 함 (실무자4) - 소장자료 검색은 잘 수행되는 것 같음 다만 웹사이트 본문 검색 부분은 시스템에 부하를 많이 주어 비효율적인 것 같음 (실무자5) - 형태소 분석 기능을 사용하여 검색이 되고 있으며 색인된 웹사이트 관련 단어로 검색하기 때문에 단순 검색 수행에는 문제없다고 생각됨 (실무자6) 	<ul style="list-style-type: none"> - 검색 속도가 느리고 다양한 검색 조건이 불가하므로 전문, WARC 파일 수집 정보, 메타 정보 등에 대하여 색인 프로세스 강화 필요

Q6. 귀하가 직무를 수행하거나 경험하면서 평소에 가졌던 오아시스의 활성화를 위한 개선점 및 방향성은? (시스템, 관리 및 운영 조직, 예산, 수집, 색인/검색, 웨이백 재생, 누리집 서비스 종류, 서비스 품질 등의 측면에서)

인터뷰 내용	요약
<ul style="list-style-type: none"> - 1. 오아시스 및 관련된 시스템(DCMS, 내부 수집기) 운영 및 관리 일원화하여 고품질 수집 및 등록, OASIS 	

<p>서비스 개선, 모니터링 효율성 증대. 시스템 장애 시 빠른 처리가 필요</p> <p>2. 현 시스템 운영 및 개선 방안 수립 (수집 대상 목표 재설정, 색인/검색 개선, 대용량 WARC 파일 처리 방안)</p> <p>3. 서비스 검색 및 운영 속도 개선 (검색부터 서비스까지)</p> <p>4. 웹자원 트렌드 분석, 데이터 서비스 등 신규 서비스 적용 및 개발이 필요 (실무자1)</p> <p>- 이용자를 위한 서비스 발굴이 필요해 보임 (현재는 자료 수집에만 집중하는 듯) (실무자4)</p> <p>- 웨이백 재생 속도 개선이 필요해 보이니, 수집되는 웹사이트 중량이 너무 무거워서 성능에 영향을 끼치는 것 같음 (실무자5)</p> <p>- 색인/검색은 오픈소스가 아닌 전문적인 검색엔진 솔루션을 도입하여 전문성 있게 관리 되어야 함 (실무자6)</p> <p>- 활용도 제고 방안이 강구됐으면 좋겠음</p> <p>- 또한 서비스 대비 수집 비율을 감안하여 수집량을 조절했으면 함 (실무자3)</p> <p>- 조직 및 기능 확대, 시스템 및 인프라 확충, 연구개발 및 서비스 개선 등 다각적인 지원 확대가 필요함</p> <p>- 사서+기술직 전담팀 확대 구성 및 유기적인 협업 체계, 오아시스 전용 시스템 구축, 수집기 및 색인/검색 고도화, 메타데이터 자동 추출 및 DB 분리, 네트워크 전용선 확충, 서버 및 스토리지 확충, 검증 자동화 도입 및 인공지능 기반 데이터 분석 기술 도입, 연구 개발 협업 체계, 누리집 및 모바일 서비스 개선, IIPC 국제활동 강화 등 전반적인 지원 확대 필요 (실무자2)</p>	<p>- 오아시스 및 관련된 시스템(DCMS, 내부수집기) 운영 및 관리 일원화하여 보다 체계적이고 효율적인 대처 필요</p> <p>- 현 시스템 운영의 문제점을 파악하여 서비스 속도 개선과 같은 개선 방안을 수립할 필요가 있음</p> <p>- 신규 서비스 적용 및 개발이 필요함</p>
---	--

3.1.3 요약 및 시사점

○ 우리나라 웹 아카이빙(오아시스) 서비스의 역할 및 위상에 대해

- 국내 웹자원 수집하여 정보를 확산할 수 있는 기반을 마련하는 것이 가장 중요하고 국제 자원 공유를 할 수 있어야 함
- 국가 재난, 행사, 주요 주제 및 이슈 수집하여 컬렉션 구축할 필요가 있음
- 품질 보증(QA: Quality Assurance)은 수집 공정 시부터 철저한 공정으로 품질을 확보하여 보존해야 함
- 국내외 웹자원 유통 및 상호 협력, 국가 디지털 장서로 활용하고 있어 현재 기조를 유지하면서 추가적 확대 방안을 시도할 필요가 있음
- 웹자원의 품질은 양호하나 데이터 색인 처리 및 데이터 서비스 구현은 개선이 필요하다고 판단함
- 검색 시 속도 저하 문제점이 있는 것으로 파악되는데 개선할 필요가 있음

○ 수집기(예, Heritrix)의 특징과 수집 과정에 대해

- 현재 웹 수집기로 Heritrix를 사용하고 있으며, 최신 웹사이트는 동적 웹페이지를 만들기 위해 Javascript를 사용하는 것이 특징인데, Heritrix는 Javascript를 실행할 수 없어 동적 웹페이지 수집

에 있어 불완전한 수집을 보임

- 외부 웹자원 수집 과정은, 1) 웹자원 수집 검증(사이트와 WARC 파일 비교), 2) 웹자원 수집 성공 (WARC 파일 업로드), 3) 국중 DCMS 시스템 등록 확인이며, 수집 과정 문제점으로 1)VPN 활용 HIWARE SFTP로 업로드 시 네트워크 및 서버 색인 전송 속도 저하(새로운 업로드 방식 및 색인 이관 프로세스 모듈 개선 필요). 2) 전송 중 자주 장애(중단 및 수집 서버 상태 점검 어려움) 발생함

○ 우리나라 웹사이트와 WARC 파일의 특성 및 문제점에 대해

- 다수의 도메인 등록 후 미운영 웹사이트 수집 및 검증에 불필요한 시간 초래
- 이미지 활용도가 높은 국내 웹사이트의 특징 때문에 Heritrix 수집 문법 및 특성에 맞지 않은 사이트 다수 존재
- 전반적으로 수백 기가 바이트(gigabyte, GB)가 넘는 웹사이트의 대용량화 때문에 웹사이트를 저장하고 이동, 화면 제시하는 데 속도가 늦음

○ 색인기의 특징과 색인 과정에서의 문제점에 대해

- 대용량 WARC 파일을 색인하고 검색하는 시간이 오래 걸림 있으며 색인을 별도로 유지하는 것도 용량을 차지함
- 색인자료에 대하여 전문 검색(full-text)이 가능하도록 Solr 검색엔진 구현에 의한 재색인 방안 마련이 필요함

○ 색인된 웹사이트의 검색에 대해

- 검색 속도가 느리고 다양한 검색 조건이 불가하므로 전문, WARC 파일 수집 정보, 메타 정보 등에 대하여 색인 프로세스 강화 필요

○ 오아시스의 활용성 강화를 위한 개선점 및 방향성에 대해

- 오아시스 및 관련된 시스템(DCMS, 내부수집기) 운영 및 관리 일원화하여 보다 체계적이고 효율적인 대처 필요
- 현 시스템 운영의 문제점을 파악하여 서비스 속도 개선과 같은 개선 방안을 수립할 필요가 있음
- 신규 서비스 적용 및 개발 필요

3.2 웹 아카이빙 전문가 의견 조사

- 본 연구에서는 오아시스 개선 및 활용성 강화 방안에 대한 다양한 검토 의견을 수렴하기 위해 웹 아카이빙 관련 학계, 연구계 전문가(3명)를 대상으로 면담 조사를 실시하였음

3.2.1 면담 조사 설계

○ 조사 개요

- 웹 아카이빙 관련 학계, 연구계 전문가 3명을 대상으로 면담 조사를 실시함
- 일시: 2024년 11월 7일(금) ~ 11월 20일(수)

- 방식: 온라인 면담 질문지 작성
- 조사내용 : 대한민국 웹 아카이빙(오아시스) 개선 및 활용성 강화 연구 절차 및 방향의 타당성 및 연구 결과 적절성, 오아시스 비전, 추진 전략, 개선 전략에 대한 타당성 및 적절성 파악, 오아시스 협력 체계 구축 및 홍보 전략의 방향성

<표 3-4> 전문가 심층 면담 대상자

전문가	직무 분야	성별
전문가1	재난 아카이브 연구	여
전문가2	웹 아카이빙 시스템 연구	남
전문가3	포괄적 수집 연구	남

○ 전문가 심층 면담 질문 구성 내용은 다음과 같음

<표 3-5> 웹 아카이빙 학계 연구계 전문가 그룹: 면담 질문 및 내용

면담 단계	검토 질문 항목		검토 질문 내용	No
시작 단계	대한민국 웹 아카이빙(오아시스) 개선 및 활용성 강화 연구 절차 및 방향의 타당성 및 연구 결과 적절성		- 대한민국 웹 아카이빙(오아시스) 개선 및 활용성 강화 연구의 최종보고 초안 검토를 통해 연구 방향, 연구 내용 및 연구 절차는 타당한가? (국제 웹 아카이빙 법적, 제도적, 기술적 동향 반영, 오아시스 시스템 현황 및 진단 반영, 이용자 및 관리자 요구 반영, 시사점 도출 반영, 비전, 개선 추진 전략 수립 등 연구 방향성 및 연구 내용, 보고서 내용에 대한 의견 제시)	Q1
이행 및 핵심 단계	오아시스 비전, 추진 전략, 개선 전략에 대한 타당성 및 적절성 파악	비전 및 추진 전략	- 제안된 오아시스의 비전 및 추진 전략에 대해 어떻게 생각하는가? (비전 및 추진 전략의 적절성 및 방향성에 대한 의견 제시)	Q2
		실행 과제	- 제안된 오아시스의 추진 전략에 따른 실행 과제에 대해 어떻게 생각하는가? (실행 과제에 대한 중복성, 실행 과제의 실현 가능성, 추진 전략과의 연계성 등에 대한 의견 제시)	Q3
		개선 및 활성화 방안	- 제안된 오아시스의 개선 및 활성화 방안에 대해 어떻게 생각하는가? (오아시스 단기 기술 고도화 방안, 중장기 개선 방안, 데이터 품질 관리 방안, 관리 및 운영 조직 개선, 누리집 서비스 품질 개선 방안의 적절성에 대한 의견 제시)	Q4
종결 단계	오아시스 협력 체계 구축 및 홍보 전략의 방향성		- 대한민국 웹 아카이빙(오아시스) 협력 체계 구축 및 홍보 전략에 대한 방향성 및 검토 의견은? (협력 체계 구축 방안, 협력 범위, 협력 내용의 적절성, 홍보 방안의 타당성 및 의견 제시)	Q5

3.2.2 면담 조사 분석 결과

○ 전문가 심층 면담 주요 조사 결과는 아래 표와 같음

<표 3-6> 전문가 심층 면담 주요 조사 결과

Q1. 대한민국 웹 아카이빙(오아시스) 개선 및 활성화 강화 연구의 최종보고 초안 검토를 통해 연구 방향, 연구 내용 및 연구 절차는 타당한가? (국제 웹 아카이빙 법적, 제도적, 기술적 동향 반영, 오아시스 시스템 현황 및 진단 반영, 이용자 및 관리자 요구 반영, 시사점 도출 반영, 비전, 개선 추진 전략 수립 등 연구 방향성 및 연구 내용, 보고서 내용에 대한 의견 제시)	
인터뷰 내용	요약
<p>전체적인 연구 방향 및 절차는 매우 타당하나 다음과 같은 관점에서의 보완을 고려해 주시기 바랍니다.</p> <p>1) 선정된 선진 사례 대부분이 도서관 주도의 웹 아카이빙 모델입니다. ‘아카이빙’이란 명칭에서도 나타난 바와 같이 기록관 주도의 웹 아카이빙, 재난 아카이브 등의 사례들도 포함되었으면 좋겠습니다. (TNA, NARA의 웹 아카이빙, Japan Disasters Digital Archive 등)</p> <p>2) 개선 및 활성화 방안 수립 시 시스템 담당 실무자의 의견은 반영되어 있으나 실제 ‘이용자’ 의견에 대한 검토 내용이 잘 보이지 않습니다. 새로 의견 수렴을 하기는 시기상 어려우시겠지만, 웹사이트를 통해 접수된, 혹은 기존 연구를 통해 수집된 의견에 대한 검토가 보완되면 좋겠습니다. (전문가1)</p> <p>첫째, 연구 방향 관련입니다. 최종보고서는 OASIS의 현재 상태를 진단하고, 향후 20년간 지속 가능한 웹 아카이빙 시스템을 구축하기 위한 방향성을 제시하고 있습니다. 이 과정에서 국제 동향과 국내·외 사례 분석을 통해 OASIS를 글로벌 수준의 서비스로 발전시키는 것을 목표로 하고 있습니다. 특히, 특히 서비스 품질 개선과 사용자 활용도를 높이는 데 초점을 두어 연구 관련 요구에 부합합니다. 목표를 명확히 하기 위해 현재 상태를 진단하고 장기 목표를 설정하였으며, 기존 시스템의 한계를 개선하고 미래를 대비할 수 있는 적절한 접근법을 보이고 있습니다. OASIS 자체는 대한민국 국립중앙도서관이 운영하는 프로젝트나 IIPC 등 국제적인 기구와의 협력 및 동향 파악이 핵심적입니다. 관련해서 국제 동향과 협력을 강조한 것은 국내 웹 아카이빙 시스템이 글로벌 맥락 속에서 발전할 수 있는 기반을 제공할 수 있습니다. 품질 개선과 활용성 강화라는 구체적인 목표를 설정한 점이 현실적이고 실행 가능성이 높다고 할 수 있습니다. 결론적으로 연구 방향은 현재와 미래의 과제를 균형 있게 다루고 있으며, 국내외 환경을 반영한 점에서 타당성이 높습니다.</p> <p>둘째, 연구 내용 관련입니다. 최종보고서에는 국제적 동향 조사 및 주요 국가 사례 분석(기술·법 제도 포함), OASIS 시스템 및 데이터 현황 진단(실무자 인터뷰, 사용성 평가 등), 중장기 기술 고도화와 안정적 운영을 위한 개선 방안 제안, 새로운 데이터 서비스 개발을 위한 프로토타입 연구와 시범 적용, 국제 협력 방안과 대국민 서비스 활성화를 위한 홍보 전략 수립을 수행하였습니다. 최종보고서는 국제 동향, 시스템 분석, 기술 개선, 홍보 및 협력 등 폭넓은 연구 내용과 핵심 문제를 다루</p>	<p>- 아카이빙’이란 명칭에서도 나타난 바와 같이 기록관 주도의 웹 아카이빙, 재난 아카이브 등의 사례들도 포함되었으면 좋겠음 (TNA, NARA의 웹 아카이빙, Japan Disasters Digital Archive 등)</p> <p>- 개선 및 활성화 방안 수립 시 시스템 담당 실무자의 의견은 반영되어 있으나 실제 ‘이용자’ 의견에 대한 검토 내용이 잘 보이지 않아, 웹사이트를 통해 접수된, 혹은 기존 연구를 통해 수집된 의견에 대한 검토가 보완되면 좋겠음</p> <p>- 연구 방향은 현재와 미래의 과제를 균형 있게 다루고 있으며, 국내외 환경을 반영한 점에서 타당성이 높음</p> <p>- 해당 연구 절차는 단계적이고 체계적으로 설계되어 있어 최종 결과물의 논리성과 실행 가능성을 높일 수 있도록 구성되었음</p> <p>- 연구 내용은 구체적이고 체계적이며, 요구사항을 충실히 반영하여 문제 해결 가능성을 높이고 있음</p> <p>- 연구 방법 및 접근은 동향 조사, 우수 사례 조사 및 사용성 평가와 같은 접근으로 적절했다고 보여짐</p>

<p>었습니다. 또 시스템 진단, 데이터 분석, 단계별 개선안 등 세부적이고 실현 가능한 내용이 포함하고 있습니다. 무엇보다 실무자와 이용자 의견을 수렴하여 실질적인 문제를 반영하고자 한 점이 돋보입니다. 결론적으로 연구 내용은 구체적이고 체계적이며, 요구사항을 충실히 반영하여 문제 해결 가능성을 높이고 있습니다.</p> <p>셋째 연구 절차 관련입니다. 최종보고서에서 제시한 연구 절차를 요약하면 다음과 같습니다.</p> <p>- 문헌 조사와 사례 분석을 통한 국내외 동향 파악 - 실무자와 이용자 인터뷰를 통한 의견 수렴 - 시스템 진단과 개선 방안 도출 및 프로토타입 개발 - 국제 협력과 홍보 방안 수립 - 중간 및 최종 보고를 통한 결과 공유</p> <p>본 보고서가 제안 및 준수하고 있는 연구 절차는 단계별로 체계적으로 구성되어 있으며, 목표 달성을 위한 논리 흐름이 벗어나지 않게 해줍니다. 실무자 의견과 전문가 자문을 반영하여 현실적인 접근을 강조한 점이 강점입니다. 연구 절차 중 중간 보고회와 최종보고회를 포함하고 있어 연구 타당성과 보완 사항을 검토하고 수정할 기회를 포함하고 있는 것도 적절하게 평가할 수 있습니다. 해당 연구 절차는 단계적이고 체계적으로 설계되어 있어 최종결과물의 논리성과 실행 가능성을 높일 수 있도록 구성되었습니다. (전문가2)</p> <p>- 연구 방향은 활용도 높은 새로운 서비스를 위한 품질 개선이라는 목표 달성을 위해 적절하게 설정되었다고 판단됨. 연구 절차 면에서 제도적 동향을 통한 환경 변화 분석과 기술적 동향 분석이 충실히 수행된 것으로 보임. 우수 사례를 선별하여 집중 분석한 것도 효과적인 접근이었음. 이후 기술적 동향과 현황 진단에서 면담 조사를 적절히 설계하여 개선점을 도출한 결과도 적절한 것으로 판단됨.</p> <p>- 다만 보고서 초안의 해외 사례 분석 중 일부 내용이 기계 번역 어투가 보이므로 내용 전달이 잘되도록 다듬을 필요가 있음. 일부 오타도 교정이 필요함.</p> <p>예: “귀하의 브라우저를”, “볼 수 있습니다. 확장이 필요하지 않음” “방대한 데이터를” (전문가3)</p>	
Q2. 제안된 오아시스의 비전 및 추진 전략에 대해 어떻게 생각하는가? (비전 및 추진 전략의 적절성 및 방향성에 대한 의견 제시)	
인터뷰 내용	요약
<p>질문 2~4에 대한 의견이 연결되어 통합 답변드립니다.</p> <p>전체적인 추진 전략 및 실행 과제, 이를 수행하기 위한 조직/예산 등의 계획 수립은 매우 체계적이고 현실적으로 작성되었습니다. 특히 ‘즉시 개선’, ‘차년도 개선’ 등 우선순위를 부여해 주관기관에서의 실행 가능성이 더 명확해진 점은 높이 평가합니다.</p>	

다만, 앞서 Q1의 3번 의견과 더불어, 국제 웹 아카이빙의 동향 분석은 매우 포괄적으로 내용적인 측면에서 진행된 반면, 3장의 시스템 분석은 사이트의 기능 점검 위주(링크 동작 여부, 속도 등)로 진행되다 보니 이후 추진 전략 역시 기능 보완 및 HW infra 구축 또는 확장 쪽으로 집중되어 있습니다. 현재 국중의 오아시스 서비스에 대한 개선이나 웹 아카이빙 구축 자체에 대한 개선 방향이 좀 더 보완되면 좋겠습니다.

예를 들면

1) 소셜미디어의 경우, 링크가 깨지는 등의 기능적 차원이 아닌 현행화(수집) 주기를 조절한다든지 없어진 계정에 대한 콘텐츠 보존 등의 문제를 고민해 주시면 좋겠습니다.

2) 재난 아카이브의 경우, 분류 체계가 일관되지 않고 해외 표준과도 맞지 않습니다. 일관된 기준의 분류 체계로 아카이빙 자료를 관리하고 이에 대한 내용이 반영된 색인, 검색 결과 제시 등의 내용적 개편에 대한 내용을 포함하면 좋겠습니다. (cf. TTA.KO-10.1047'재난안전정보 통합 메타데이터 구축을 위한 분류 체계' 표준)

3) 개인적으로, 이용자 차원에서 현 오아시스의 가장 큰 문제점은 검색 품질에 있다고 생각합니다. 검색 결과가 부족하거나, 결과가 제시되었어도 이해가 잘 되지 않는 경우가 자주 발생하는데, 이는 데이터 자체가 갖는 메타데이터 부족 문제도 있고, 이후 색인/검색 모듈 간의 불일치 문제 등에 대한 개선 방안이 필요합니다(6.1.2장 보완).

4) 기존의 서비스 개선을 넘어 최근 전 국민이 범용적으로 쓰고 있는 생성형 AI 등의 기술이 적용된 새로운 서비스(예: 자연어 질문/답변, 어떤 사건에 대한 트렌드 분석, 자동 요약 등)에 대한 제언도 보완되면 좋겠습니다(p. 162 보완) (전문가1)

본 보고서는 오아시스의 비전과 추진 전략을 5장에 기술하였습니다. 현재 최종 보고서의 5장은 향후 추가 수정과 보완이 필요할 것으로 보입니다. 전체 내용이 아직 확정되지 않은 것으로 보입니다. 현재 사례로 조사한 내용은 룩셈부르크 웹 아카이브 서비스입니다. 국립중앙도서관은 2004년부터 웹 아카이빙 프로젝트를 수행했고, 그간 많은 연구개발이 이루어졌고, 다양한 사례 분석도 이루어졌습니다. 중복되지 않는 범위에서 사례 조사를 수행한 것으로 긍정적이며, 무엇보다 국립중앙도서관의 운영 현황에 맞춘 비전과 목표 설정이 현실성 있다는 점에서 현재의 접근에 대해서 문제점을 지적하는 것은 부적절하다고 생각합니다.

단, 비전은 조직이 이루어야 하는 이상적인 상태를 목표를 이룰 실현하기 위한 구체적인 활동과 성과입니다. 목표는 현황이나 사회적 변화에 따라 수준을 달리하거나 변화할 수 있습니다. 현재 보고서에 제시하고 있는 비전이 나쁘지는 않지만, 궁극적으로 웹 아카이빙이라는 활동을 대표한다고 보기에는 난해한 것 같습니다. 타임캡슐이라는 용어가 현재 OASIS 시스템에

- 무엇보다 국립중앙도서관의 운영 현황에 맞춘 비전과 목표 설정이 현실성 있기에 현재 시스템의 문제점 지적을 통한 비전 도출 접근보다 국립중앙도서관의 기관 비전과의 정합성을 검토할 필요가 있음
- 기능 보완 및 HW infra 구축 또는 확장 쪽으로 집중된 추진 전략에 현재 오아시스 서비스에 대한 개선이나 웹 아카이빙 구축 자체에 대한 개선 방향이 좀 더 보완되면 좋겠음
- 다른 나라의 웹 아카이빙 시스템과의 차별된 오아시스의 특징 중 하나인 트렌드 분석 등의 새로운 서비스를 더욱 부각하고 강화할 필요가 있으며 향후 데이터 활용 전략 강화를 목표로 인공지능, 데이터 분석 등 기술 활용 방안을 고려할 필요가 있어 보임
- 기존의 서비스 개선을 넘어 최근 전 국민이 범용적으로 쓰고 있는 생성형 AI 등의 기술이 적용된 새로운 서비스(예: 자연어 질문/답변, 어떤 사건에 대한 트렌드 분석, 자동 요약 등)에 대한 제언도 보완되었으면 좋겠음
- 대국민 홍보 전략을 다양화할 필요는 있다고 봄(디지털 미디어, 교육 프로그램, 시민 참여 프로젝트 등)
- 비전 도출 근거로 사례를 추가하면 좋겠음. 또한, 비전의 내용을 좀 더 구체적으로 설명할 필요가 있음

서도 활용하고 있어 기관 입장을 고려했을 때 적절한 사용이기도 합니다.

Wikipedia의 웹 아카이빙 기관 목록(https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives)은 2011년 국립중앙도서관 아카이빙 담당자도 참여하여 각국의 웹 아카이빙 프로젝트 현황을 정리한 것입니다. 비전과 목표 관련해서는 이 목록 중 국립도서관이 수행하고 있는 프로젝트들을 참고하면 좋지 않을까 합니다. 단, 현재 연구가 종료되는 시점으로 이미 중간 보고회 등을 통해 1차 검토가 완료된 시점인 점을 감안하여 추가 연구 수행이 어렵다는 점을 감안하여 추후 주관기관을 통해 추가 사례조사 등을 통한 비전 검토가 필요할 것으로 보입니다.

이후 비전 달성을 위한 전략과 단위과제는 잘 정리된 것으로 보입니다. 그런데, 5장에서 제시하고 있는 전략과 단위 과제와 6장과 내용이 이어지지 않습니다. 아직 정리가 더 필요할 것으로 보이며, 향후 5장에 비전, 전략, 단위 과제에 대한 상세 내용이 6장에서 그대로 이어져서 연결성을 갖도록 하는 것이 적절할 것으로 보입니다. 단, 연구진의 의도가 5장과 6장을 다른 내용을 분리할 의도라면 현재 상태를 유지하셔도 될 것 같습니다. 5장과 6장이 다른 내용이라는 언급 정도만 어딘가에 표기하시면 좋을 것 같습니다.

이 외 현재 OASIS 시스템이 다른 나라와 다르게 갖고 있는 특징은(적절한 것인지 여부와 상관없이) 트렌드 분석 등의 새로운 서비스를 시도하고 있다는 점에 있습니다. 이를 더욱 부각하고 강화할 필요가 있다면 이번 보고서가 아니더라도 향후 데이터 활용 전략 강화를 목표로 인공지능, 데이터 분석 등 기술 활용 방안을 고려할 필요가 있어 보입니다. 단 이는 현재 연구에 포함시키라는 필요성을 제안한 것은 아닙니다.

무엇보다 현재 보고서에도 언급하고 있지만, 대국민 홍보전략을 다양화할 필요는 있다고 봅니다(디지털 미디어, 교육 프로그램, 시민 참여 프로젝트 등). (전문가2)

- 비전 수립에 있어서 우수 운영기관인 룩셈부르크 웹 아카이브 서비스 사례는 충분히 모범으로 삼을 만 하지만, 한 사례만 제시한 것은 다소 아쉬움. 미국 의회도서관 등의 사례도 제시하면 더욱 도출된 결과에 대한 신뢰도가 확보될 것 같음. 예를 들어 미국 의회도서관의 비전 요소 중 하나로 변화에 대한 '인증(authentication)'을 언급하고 있는데(Potter, 2012), 웹 아카이브의 활용 방안을 생각하면 참고해 볼 만한 요소라고 여겨짐.

Potter, A. (2012). A Vision of the Role and Future of Web Archives: Conclusions and the Role of Archives. Library of Congress Blogs (May 24, 2012). Available: <https://blogs.loc.gov/thesignal/2012/05/a-vision-of-the-role-and-future-of-web-archives-conclusions-and-the-r>

<p>ole-of-archives/</p> <p>- 오아시스의 비전을 키워드만 제시하였는데, 조금 더 구체적으로 비전의 내용을 설명할 필요가 있음. 그 이후에 각 비전과 연계된 목표와 추진 전략 및 과제를 체계적으로 도식화하면서 제시하는 것이 더욱 전달이 잘 될 것으로 판단됨. (전문가3)</p>	
<p>Q3. 제안된 오아시스의 추진 전략에 따른 실행 과제에 대해 어떻게 생각하는가? (실행 과제에 대한 중복성, 실행 과제의 실현 가능성, 추진 전략과의 연계성 등에 대한 의견 제시)</p>	
인터뷰 내용	요약
<p>대략적인 답변은 Q2 답변으로 대신합니다. 이 외 검토 내용은 <표 5-2> 오아시스 중장기 계획 정책 과제·실행 과제를 중심으로 검토하였습니다.</p> <p>추진 전략 전략(정책) 과제 실행 과제</p> <p>1. 안심할 수 있는 데이터 보존과 품질 보장 (데이터 보존 및 무결)</p> <p>1.1 품질 보증(QA) 프로세스 자동화</p> <p>1.1.1 WARC 파일 보안 취약점 점검</p> <p>1.1.2 지능형 기술을 도입한 누락 웹페이지 자동 점검 프로세스 개발</p> <p>2. 이용자가 원하는 콘텐츠 (개인 맞춤)</p> <p>2.1 이용자 특성에 맞춘 콘텐츠 수집 및 제공</p> <p>2.1.1 이용자 페르소나 기법을 이용한 디지털 장서 정책 수립</p> <p>3. 시스템 기반 기술 강화 (기술 고도화)</p> <p>3.1 개별 시스템 기술 고도화</p> <p>3.1.1 수집기 고도화</p> <p>3.1.2 색인 검색기 고도화</p> <p>3.1.3 재생기(웨이백 재생) 고도화</p> <p>3.2 시스템 마이그레이션(장기)</p> <p>3.2.1 오아시스 시스템 정보화 전략 계획(ISP) 수립(단계적 운영 환경 개선)</p> <p>3.2.2 오아시스 시스템 고도화 사업</p> <p>4. 언제 어디서나 이용 가능한 오아시스(시스템 접근성)</p> <p>4.1 시스템 접근성 강화</p> <p>4.1.1 이용 접근성 강화된 누리집 개선</p> <p>5. 공유 개방화(콘텐츠 제공/전달/이용 체계 확대)</p> <p>5.1 공유 확산 강화</p> <p>5.1.1 웹 아카이빙 내 및 타자원과의 연계 추천 기술 개발</p> <p>6. 지역과 국제기관과의 긴밀한 동행</p> <p>6.1 지역 및 국제 협력 체계 구축</p> <p>6.1.1 국내 웹 아카이빙 협력 체계 구축 및 공유 확산</p> <p><표 5-2> 오아시스 중장기 계획 정책 과제·실행 과제</p> <p>해당 전략과 과제들은 다양한 사례 조사와 의견 수렴을 거치고 중복제거 등 과정을 거친 결과물로 본인의 검토 의견이 이를 대체할 수는 없습니다. 이런 이유로 실행 과제의 첫 번째 확인 내용인 중복성을 검토한 바 각 과제들</p>	<p>- 지능형 기술 도입 시 비용을 함께 고려할 필요가 있으며, 개발 방식, 구현 방식, 향후 유지 운영 방안을 함께 고려하여 단기로 수행하기보다는 장기 과제로 전환하는 것이 필요해 보임</p> <p>- 또한, 별도로 기능을 개발하여 운영하는 것이 IIPC SW와의 공동 운영에 미치는 영향이 없는지도 재검토가 필요해 보임</p> <p>- 시스템 마이그레이션의 경우도 중장기 과제로 보는 것이 타당해 보임</p> <p>- 지역 및 국제 협력 체계 구축의 실행 과제는 국내외 협력의 성과 지표나 구체적인 협력 방안이 부족해 보이며 이 항목은 직접 운영하는 기관의 현황(인력, 조직, 예산, 국내·외 협력 현황 등)에 따라 달라질 수 있다는 점을 고려하여 구체적으로 보완이 필요함</p> <p>- 실행 세부 방안에 대한 성과 지표를 설정할 수 있다면 향후 지속적인 사업 운영에 도움이 될 것으로 보임</p> <p>- 현재의 협력 방안은 '기관' 간의 수평적인 협력을 중시하고 있는데, 재난 안전 포털 등의 사례를 볼 때, 문화체육관광부(도서관기획단) 차원에서 행정안전부, 과기정통부 등 정부 행정 체계를 감안한 유관 부처와의 협력 체계를 설정하고 그 아래로 기관 간 단계적, 체계적 체계를 구축하는 것이 필요함</p>

이 고유한 성격을 갖고 있는 것으로 판단합니다.

두 번째는 실현 가능성입니다. 특별히 문제가 있을 것으로 보이는 부분은 없습니다. OASIS 역시 IIPC가 지원하는 다양한 도구들을 활용하고 있고, 대표적인 도구는 Heritrix, Wayback Machine, OpenWayback입니다. 이 도구들은 이 번 과제에서 제시한 기술적 도입 고려 사항인 누락된 웹페이지 자동 점검, WARC 파일의 보안 취약점 점검, 시스템 마이그레이션과 직접적으로 관련된 것은 없습니다. 따라서 이 기능 구현을 위해서는 추가적인 도구나 스크립트를 개발하거나, 다른 오픈 소스 커뮤니티에서 제공하는 도구를 활용하는 방안을 고려해야 합니다. 즉, 해당 기술 과제의 필요성은 정당하며, 이를 단위 과제로 제안한 것은 타당해 보입니다. 단, 지능형 기술을 도입한 누락 웹페이지 자동 점검 프로세스 개발의 경우는 기술 복잡성과 높은 초기 비용 등을 고려할 필요가 있을 것 같습니다. 기존에도 bad url 점검 등 기술은 일반적이거나 과제에서 제시한 것과 같은 지능형 기술의 도입은 비용을 함께 고려할 필요가 있으며, 개발 방식, 구현 방식, 향후 유지 운영 방안을 함께 고려하여 단기로 수행하기 보다는 장기 과제로 전환하는 것이 필요해 보입니다. 특히, 별도로 기능을 개발하여 운영하는 것이 IIPC SW와의 공동 운영에 미치는 영향이 없는지도 재검토가 필요해 보입니다. 마찬가지로 시스템 마이그레이션의 경우도 중장기 과제로 보는 것이 타당해 보입니다.

세 번째는 추진 전략과 실행 과제의 연결성입니다. 추진 전략과 실행 과제 간 연계성은 대체로 명확하며, 각 전략이 실행 과제를 통해 구체화 되고 있습니다. 다만, 지역 및 국제 협력 체계 구축의 실행 과제는 국내외 협력의 성과 지표나 구체적인 협력 방안이 부족해 보입니다. 물론 이 항목은 연구 개발 과제의 결과물로 나오기보다는 직접 운영하는 기관의 현황(인력, 조직, 예산, 국내·외 협력 현황 등)에 따라 달라질 수 있다는 점은 고려한 의견입니다.

전반적으로 큰 문제가 없으며, 다만 가능하다면 향후 실행 세부 방안에 대한 성과 지표를 설정할 수 있다면 향후 지속적인 사업 운영에 도움이 될 것으로 보입니다. (전문가2)

- 현재 제시된 이용자 요구 조사 등의 실행 과제는 추진 전략에 맞추어 충실히 고안된 것으로 판단됨. 특히 신규 사업으로 제시된 국내 웹 아카이빙 협력 체계 구축 및 공유 확산 사업은 기존 사업 내용의 한계를 극복할 수 있는 매우 절실한 내용이며 해외 사례와 비교해서도 국내에서 상당히 미흡했던 사업이므로 내용을 더욱 구체적으로 검토할 필요가 있음. 현재의 협력 방안은 ‘기관’ 간의 수평적인 협력을 중시하고 있는데, 국내 정부 행정 체계를 감안하고 재난 안전 포털 등의 사례를 고려하여 행정안전부, 과기정통부 등 유관 정부 부처와의 협력 체계를 문화체육관광부(도서관 기획단) 차원에서 설정하고 그 아래로 기관 간 협력 체계를 설정하는 등의 단계적, 체계적 기획이 필요함. (전문가3)

Q4. 제안된 오아시스의 개선 및 활성화 방안에 대해 어떻게 생각하는가? (오아시스 단기 기술 고도화 방안, 중장기 개선안, 데이터 품질 관리 방안, 관리 및 운영 조직 개선, 누리집 서비스 품질 개선 방안의 적절성에 대한 의견 제시)

인터뷰 내용	요약
<ul style="list-style-type: none"> - 개선 및 활성화 전략은 대체로 체계적이고 실현 가능성이 높은 방향으로 설계되었습니다. 특히, 기술 고도화, 전문성 강화, 사용자 중심 서비스 개선에 중점을 둔 점이 인상적입니다. 다만, 최신 기술 트렌드를 반영한 구체적인 실행 방안과 법적·윤리적 문제 해결 방안이 추가된다면 더욱 효과적이고 완성도 높은 전략이 될 것으로 보입니다. 8개 측면별로 구체적인 검토 내용은 다음과 같습니다. - 첫 번째, 시스템 측면입니다. 기술 고도화 및 시스템 마이그레이션 영역의 경우 기존 시스템의 한계를 극복하기 위해 경량 구조(마이크로서비스) 도입과 단계적 정보화 전략 계획(ISP)을 제안한 것은 적절하다고 생각합니다. 특히 웨이백 재생 기술 강화에서 다양한 브라우저 지원과 데이터 손실 복구 기능 강화는 사용자 경험 향상에 중요한 역할을 할 것으로 보입니다. 단, 클라우드 기반 아키텍처 도입이나 인공지능(AI) 기술 활용 방안이 구체적으로 제시할 필요성이 보입니다. - 두 번째 관리 및 운영 조직 측면입니다. 운영 조직 전문성 강화와 실무자 교육은 시스템의 지속 가능성을 높이는 데 핵심적인 역할을 할 것으로 보이며, 지역 및 국제 협력 체계를 강화하려는 방향성도 적절하다고 생각합니다. 단, 협력 체계 강화 방안이 구체적이지 않아, 국내외 전문가와의 네트워크 구축 및 공동 프로젝트 실행 등 실질적인 운영 방안 추가가 필요할 것으로 보입니다. 단, 이는 주관기관과 협의 하에 실행 가능성 여부를 논의 후 추가가 필요합니다. - 세 번째는 예산 측면입니다. 중장기적으로 안정적인 예산 확보를 위한 계획을 제안한 점이 긍정적이며, 단계별 예산 분배를 통해 실행 가능성을 높이고 있습니다. - 네 번째는 수집 측면입니다. 자동화 기술 도입을 통한 프로세스 개선은 효율성과 데이터 품질을 동시에 높일 수 있는 적절한 방안이라고 생각합니다. 다만, 데이터 수집 과정에서의 법적, 윤리적 문제(저작권, 개인정보 등)를 다루는 명확한 기준과 지침이 필요하며, 이는 주의할 필요가 있습니다. - 다섯 번째는 색인/검색 측면입니다. 고성능 검색엔진 도입과 사용자 맞춤형 검색 기능 강화는 데이터 접근성을 향상시킬 수 있는 핵심 요소입니다. 다만, 현재 OASIS 시스템의 구성을 고려했을 때 별도 검색 엔진을 도입하여 활용하는 것이 비용 대비 효과가 높을지는 의문입니다. OpenWayback이 포함하고 있는 검색으로도 충분하지 않을까 하는 생각입니다. - 여섯 번째는 웨이백 재생 측면입니다. OpenWayback은 이미 크로스 브라우징을 지원합니다. 다만 수집 대상이 되는 웹 콘텐츠가 특정 기능을 별도로 활용하고 지원하고 있는 경우에는 문제가 생길 수 있을 것으로 보이고 이를 일일이 대응하는 것은 불가능합니다. 표준을 벗어난 기능 구현으로 이를 지원하기 위한 별도 기 	<ul style="list-style-type: none"> - 최신 기술 트렌드를 반영한 구체적인 실행 방안과 법적·윤리적 문제 해결 방안 추가가 필요함 - 모바일 환경 최적화와 장애인을 위한 접근성 강화 기능의 추가적 고려 필요함 - 사용자 피드백 수집 및 반영 체계의 구체적 명시 필요 - 제안된 오아시스 개선 및 활성화 전략은 특히 시스템 측면에서 최신 동향에 맞춘 새로운 인프라 방향을 1, 2, 3안까지 대안을 제시하여 매우 우수하다고 판단됨

<p>술 개발은 더 큰 문제를 발생하게 한다고 생각합니다.</p> <ul style="list-style-type: none"> - 일곱 번째는 누리집 서비스 종류 측면입니다. 주제별 컬렉션 제공, 데이터 시각화 등 다양화된 서비스 제공은 대중 접근성을 높이는 데 효과적이라고 생각합니다. 굳이 기존 기능을 현재 상태에서 개선해야 한다면, 모바일 환경 최적화와 장애인을 위한 접근성 강화 기능이 추가적으로 고려되어야 할 것 같습니다. - 여덟 번째는 서비스 품질 측면입니다. 사용자 피드백 수집 및 반영 체계를 구체적으로 명시하면 더 좋을 것 같습니다. 현재도 큰 문제는 없다고 판단합니다. (전문가2) - 제안된 오아시스 개선 및 활성화 전략은 특히 시스템 측면에서 최신 동향에 맞춘 새로운 인프라 방향을 1, 2, 3안까지 대안을 제시하여 매우 우수하다고 판단됨. 절충적이라고도 볼 수도 있으나 가장 신뢰할 수 있을 만한 개선안은 이 중에서 2안으로 판단되며, 적극 추진할 필요가 있음. (전문가3) 	
Q5. 대한민국 웹 아카이빙(오아시스) 협력 체계 구축 및 홍보 전략에 대한 방향성 및 검토 의견은? (협력 체계 구축 방안, 협력 범위, 협력 내용의 적절성, 홍보 방안의 타당성 및 의견 제시)	
인터뷰 내용	요약
<ul style="list-style-type: none"> - 연구진이 제시하신 법적, 제도적 개선 방안이 기관 협력 체계에도 함께 반영되면 좋겠습니다. 특히 애초 정보가 생성된 기관으로부터 디지털 납본이 되는 형태가 가장 효과적이고 체계적이라는 점은 자명하니 이 부분에 대한 체계구축 방안이 제시되면 좋겠습니다. (전문가1) - 협력 체계 구축과 홍보 전략은 전반적으로 적절하며, 실행 가능성이 높습니다. 특히, 국내외 협력과 대중적 홍보를 강조한 점은 향후 웹 아카이빙의 지속 가능성을 보장하는 중요한 요소입니다. 첫 번째로 협력 체계 구축 방안과 관련하여 보고서에서 제안한 방향성은 국내 웹 아카이빙 관련 기관 간 협력 네트워크 구축, 지역 사회와의 협력을 통한 데이터 수집 확대, 국제 협력 체계 강화(IIPC 등 글로벌 네트워크 참여 확대)입니다. 협력 네트워크 구축은 중복 수집을 방지하고 효율적인 아카이빙을 가능하게 하며, 데이터의 보존 가치를 높이는 데 매우 적절합니다. 국제협력을 통해 글로벌 웹 아카이빙 표준 및 최신 기술을 수용하려는 접근은 장기적 관점에서 타당합니다. 가능하다면 더 적극적인 IIPC 참여와 활용이 필요하다고 봅니다. 또한 지역 사회와의 협력은 공공 기관뿐만 아니라 대학, 비영리단체, 민간 기업을 포함하여 협력 범위를 넓힐 필요가 있습니다(실제 실행하는데 존재하는 어려움은 잘 이해하고 있습니다.). 국제 협력의 경우, 단순 네트워크 참여를 넘어, 공동 연구 및 기술 개발 프로젝트를 추진하는 방식으로 구체화할 필요가 있습니다. - 두 번째는 협력 범위와 내용의 적절성 관련입니다. 협력 범위를 국내(공공 도서관, 학술 기관, 문화재 관련 	<ul style="list-style-type: none"> - 협력 체계 구축을 위한 법적, 제도적 개선 방안 포함되면 좋겠음 - 지역 사회와의 협력은 공공 기관뿐만 아니라 대학, 비영리단체, 민간 기업을 포함하여 협력 범위를 넓힐 필요가 있음 - 국제 협력의 경우, 단순 네트워크 참여를 넘어, 공동 연구 및 기술 개발 프로젝트를 추진하는 방식으로 구체화할 필요가 있음 - 협력의 범위와 내용에 있어서, 협력 범위를 국내(공공 도서관, 학술 기관, 문화재 관련 단체)와 국제(IIPC 등) 모두를 포괄하여 제안하며 지역 사회의 다양한 이해 관계자(대학, 비영리 단체, 민간 기업 등)를 포함하고, 협력 내용도 데이터 수집 및 보존 기술 공유, 공동 연구 및 정책 수립, 웹 아카이빙 이용 활성화를 위한 협력 서비스 개발 등으로 적절하며 향후 협력 내용을 단순 데이터 공유에서 넘어, 기술적·정책적 협업으로 확장하는 방안이 함께 제안될 필요가 있음 - 세부 추진 과제에 포함할 홍보 방안에 국립중앙도서관 측의 정기적 기획 분석, 결과 홍보를 수행할 필요가 있음. 도서관 북 큐레이션 서비스와 같이 사회적 이슈와 관련된 비정기적인 기획 분석 추진도 필요함 - 현재 제안된 국내 웹 아카이빙 협력 체계 구축 및 공유 협력 방안은 ‘기관’ 간의 수평적인 협력을 중시하고 있는데, 국내 정부 행정 체계를 감안하고 재난 안전 포털과 재난 아카이브 등의 과거 갈등 사례를 고려하여 행정안전부, 과기정통부 등 유관 정부 부처와의 협력 체계를 문화체육관광부(도서관 기획단) 차원에서 설정하고 그 아래로 기관 간 협력 체계를 설정하는 등의 단계적,

<p>단체)와 국제(IIPC 등) 모두를 포괄하여 제안하였으며, 지역 사회의 다양한 이해 관계자(대학, 비영리단체, 민간 기업 등)를 포함하고 있습니다. 협력 내용도 데이터 수집 및 보존 기술 공유, 공동 연구 및 정책 수립, 웹 아카이빙 이용 활성화를 위한 협력 서비스 개발 등으로 적절합니다. 향후 협력 내용을 단순 데이터 공유에서 넘어, 기술적·정책적 협업으로 확장하는 방안이 함께 제안되면 좋을 것 같습니다.</p> <ul style="list-style-type: none"> - 세 번째는 홍보 방안의 타당성입니다. 연구 결과로 대국민 홍보 캠페인 및 SNS를 통한 인식 확산, 교육 프로그램 개발 및 전문 강의 개최, 웹 아카이빙 데이터의 활용 사례 공유 등을 제안하였습니다. 대중에게 친숙한 방법(SNS, 미디어 활용)을 통해 웹 아카이빙의 중요성을 홍보하는 전략은 효과적이라고 생각합니다. 교육 및 강의 프로그램을 통해 관련 전문가를 육성하는 방안은 홍보와 실질적 역량 강화를 동시에 달성할 수 있습니다. 특별한 문제를 확인할 수 없으며, 향후 실제 수행 시 대국민 홍보 전략이 구체적인 타겟팅을 기반으로 세분화될 필요가 있을 것으로 보입니다. 사례 공유는 단순 정보 제공에서 벗어나, 웹 아카이빙 데이터를 활용한 혁신적인 응용 사례(연구, 산업적 응용 등)를 강조하면 더 큰 주목을 받을 수 있을 것으로 판단합니다. (전문가2) - 6장의 개선 및 활성화 추진 전략에는 ‘홍보 방안’이 제시되어 있는데, 7장의 과제별 세부 추진 방안에는 홍보 방안은 없음. 이는 ‘홍보’가 세부 추진을 위한 과제로 고려되지 않는다는 것인데, 세부 추진 과제에도 ‘홍보’를 포함할 필요가 매우 크다고 판단됨. - 세부 추진 과제에 포함할 홍보 방안에 국립중앙도서관 측의 정기적 기획 홍보를 구체적으로 포함하는 것도 좋을 것임. 현재는 홍보를 위한 활용 사례가 외부의 참여 위주로 제시되어 있는데, 초기에는 국립중앙도서관 측에서 정기적인 기획 분석과 결과 홍보를 수행할 필요가 있음. 도서관 북 큐레이션 서비스와 같이 사회적 이슈와 관련된 비정기적인 기획 분석 추진도 필요함. - 현재 제안된 국내 웹 아카이빙 협력 체계 구축 및 공유 협력 방안은 ‘기관’ 간의 수평적인 협력을 중시하고 있는데, 국내 정부 행정 체계를 감안하고 재난 안전 포털과 재난 아카이브 등의 과거 갈등 사례를 고려하여 행정안전부, 과기정통부 등 유관 정부 부처와의 협력 체계를 문화체육관광부(도서관 기획단) 차원에서 설정하고 그 아래로 기관 간 협력 체계를 설정하는 등의 단계적, 체계적 기획이 필요함. (전문가3) 	<p>체계적 기획이 필요함</p>
---	--------------------

3.2.3 요약 및 시사점

- 대한민국 웹 아카이빙(오아시스) 개선 및 활성화 연구 절차 및 방향의 타당성 및 연구 결과 적절성에 대해
 - ‘아카이빙’이란 명칭에서도 나타난 바와 같이, 기록관 주도의 웹 아카이빙, 재난 아카이브 등의 사례들도 포함되었으면 좋겠음 (TNA, NARA의 웹 아카이빙, Japan Disasters Digital Archive 등)

- 개선 및 활성화 방안 수립 시 시스템 담당 실무자의 의견은 반영되어 있으나 실제 ‘이용자’ 의견에 대한 검토 내용이 잘 보이지 않아, 웹사이트를 통해 접수된, 혹은 기존 연구를 통해 수집된 의견에 대한 검토가 보완되면 좋겠음
- 연구 방향은 현재와 미래의 과제를 균형 있게 다루고 있으며, 국내외 환경을 반영한 점에서 타당성이 높음
- 해당 연구 절차는 단계적이고 체계적으로 설계되어 있어 최종 결과물의 논리성과 실행 가능성을 높일 수 있도록 구성되었음
- 연구 내용은 구체적이고 체계적이며, 요구사항을 충실히 반영하여 문제 해결 가능성을 높이고 있음
- 연구 방법 및 접근은 동향 조사, 우수 사례조사 및 사용성 평가와 같은 접근으로 적절했다고 보여짐

○ 비전 및 추진 전략에 대해

- 무엇보다 국립중앙도서관의 운영 현황에 맞춘 비전과 목표 설정이 현실성 있기에 현재의 시스템의 문제점 지적을 통한 비전 도출 접근보다 국립중앙도서관의 기관 비전과의 정합성을 검토할 필요가 있음
- 기능 보완 및 HW infra 구축 또는 확장 쪽으로 집중된 추진 전략에 현재 오아시스 서비스에 대한 개선이나 웹 아카이빙 구축 자체에 대한 개선 방향이 좀 더 보완되면 좋겠음
- 다른 나라의 웹 아카이빙 시스템과의 차별된 오아시스의 특징 중 하나인 트렌드 분석 등의 새로운 서비스를 더욱 부각하고 강화할 필요가 있으며 향후 데이터 활용 전략 강화를 목표로 인공지능, 데이터 분석 등 기술 활용 방안을 고려할 필요가 있어 보임
- 기존의 서비스 개선을 넘어 최근 전 국민이 범용적으로 쓰고 있는 생성형 AI 등의 기술이 적용된 새로운 서비스(예: 자연어 질문/답변, 어떤 사건에 대한 트렌드 분석, 자동 요약 등)에 대한 제언도 보완 되었으면 좋겠음
- 대국민 홍보 전략을 다양화할 필요는 있다고 봄(디지털 미디어, 교육 프로그램, 시민 참여 프로젝트 등)
- 비전 도출 근거로 사례를 추가하면 좋겠음. 또한, 비전의 내용을 좀 더 구체적으로 설명할 필요가 있음

○ 실행 과제에 대해

- 지능형 기술 도입 시 비용을 함께 고려할 필요가 있으며, 개발 방식, 구현 방식, 향후 유지 운영 방안을 함께 고려하여 단기로 수행하기보다는 장기 과제로 전환하는 것이 필요해 보임
- 또한, 별도로 기능을 개발하여 운영하는 것이 IIPC SW와의 공동 운영에 미치는 영향이 없는지도 재검토가 필요해 보임
- 시스템 마이그레이션의 경우도 중장기 과제로 보는 것이 타당해 보임
- 지역 및 국제 협력 체계 구축의 실행 과제는 국내외 협력의 성과 지표나 구체적인 협력 방안이 부족해 보이며 이 항목은 직접 운영하는 기관의 현황(인력, 조직, 예산, 국내·외 협력 현황 등)에 따라 달

라질 수 있다는 점을 고려하여 구체적으로 보완이 필요함

- 실행 세부 방안에 대한 성과 지표를 설정할 수 있다면 향후 지속적인 사업 운영에 도움이 될 것으로 보임
- 현재의 협력 방안은 ‘기관’ 간의 수평적인 협력을 중시하고 있는데, 재난 안전 포털 등의 사례를 볼 때, 문화체육관광부(도서관 기획단) 차원에서 행정안전부, 과기정통부 등 정부 행정 체계를 감안한 유관 부처와의 협력 체계를 설정하고 그 아래로 기관 간 단계적, 체계적 체계를 구축하는 것이 필요함

○ 개선 및 활성화 방안에 대해

- 최신 기술 트렌드를 반영한 구체적인 실행 방안과 법적·윤리적 문제 해결 방안 추가가 필요함
- 모바일 환경 최적화와 장애인을 위한 접근성 강화 기능의 추가적 고려가 필요함
- 사용자 피드백 수집 및 반영 체계의 구체적 명시가 필요함
- 제안된 오아시스 개선 및 활성화 전략은 특히 시스템 측면에서 최신 동향에 맞춘 새로운 인프라 방향을 1, 2, 3안까지 대안을 제시하여 매우 우수하다고 판단됨

○ 오아시스 협력 체계 구축 및 홍보 전략의 방향성에 대해

- 협력 체계 구축을 위한 법적, 제도적 개선 방안에 포함되면 좋겠음
- 지역 사회와의 협력은 공공 기관뿐만 아니라 대학, 비영리단체, 민간 기업을 포함하여 협력 범위를 넓힐 필요가 있음
- 국제 협력의 경우, 단순 네트워크 참여를 넘어, 공동 연구 및 기술 개발 프로젝트를 추진하는 방식으로 구체화할 필요가 있음
- 협력의 범위와 내용에 있어서, 협력 범위를 국내(공공 도서관, 학술 기관, 문화재 관련 단체)와 국제(IIIPC 등) 모두를 포괄하여 제안하며 지역 사회의 다양한 이해 관계자(대학, 비영리 단체, 민간 기업 등)를 포함하고, 협력 내용도 데이터 수집 및 보존 기술 공유, 공동 연구 및 정책 수립, 웹 아카이빙 이용 활성화를 위한 협력 서비스 개발 등으로 적절하며 향후 협력 내용을 단순 데이터 공유에서 넘어, 기술적·정책적 협업으로 확장하는 방안이 함께 제안될 필요가 있음
- 세부 추진 과제에 포함할 홍보 방안에 국립중앙도서관 측의 정기적 기획 분석, 결과 홍보를 수행할 필요가 있음. 도서관 북 큐레이션 서비스와 같이 사회적 이슈와 관련된 비정기적인 기획 분석 추진도 필요함
- 현재 제안된 국내 웹 아카이빙 협력 체계 구축 및 공유 협력 방안은 ‘기관’ 간의 수평적인 협력을 중시하고 있는데, 국내 정부 행정 체계를 감안하고 재난 안전 포털과 재난 아카이브 등의 과거 갈등 사례를 고려하여 행정안전부, 과기정통부 등 유관 정부 부처와의 협력 체계를 문화체육관광부(도서관 기획단) 차원에서 설정하고 그 아래로 기관 간 협력 체계를 설정하는 등의 단계적, 체계적 기획이 필요함

3.3 오아시스 누리집 현황 진단

- 본 연구에서는 대한민국 웹자원 아카이빙(OASIS)의 개선 및 활용성 강화를 위해 현재 오아시스 누리집에 대한 다양한 의견을 수렴하기 위해 예비 사서(47명)를 대상으로 사용성 평가를 실시하였음

3.3.1 사용성 평가 설계

- 조사 개요
 - 한남대 문헌정보학과 예비사서 47명을 대상으로 오아시스 누리집 사용성 평가를 실시함
 - 일시: 2024년 6월 14일(금) ~ 2024년 6월 21일(금)
 - 평가 내용: 오아시스 제공 서비스, 오아시스 검색 사용성 평가, 컬렉션 사용성 평가, 재난 아카이브 사용성 평가, SNS 사용성 평가, 웹트렌드 사용성 평가, 소개 사용성 평가, 전반적인 사용 소감 및 개선 요구 사항
- 평가 결과는 평가자 주관에 반영된 내용이며 네트워크 환경 영향으로 인한 일시적인 장애일 수 있음

3.3.2 사용성 평가 결과

가. 오아시스 누리집 메뉴별 사용성 평가

1) 검색

- (검색-1) 웨이백 재생 무한 버퍼링
 - 예시) 검색 -> 2010년 고성군 보건소 자료, 웹사이트 보기 선택 → 버퍼링 화면 지속 (웹사이트 제공 X) (<그림 부록A-1>, <그림 부록A-2> 참조)
 - 개선점: 정기적으로 정보자원의 해당 웹사이트로의 이동이 원활한지 확인하고, 보완해야 함
- (검색-2) 미리보기 디폴트 이미지 상이함
 - 예시 1) 썸네일 영상이 없을 때 표시되는 디폴트 이미지로 OASIS 로고가 표시되는 경우 (<그림 부록 A-3> 참조)
 - 예시 2) 미리보기 디폴트 이미지로 문서 클립아트 이미지가 표시되는 경우 (<그림 부록A-4> 참조)
 - 예시 3) 미리보기 디폴트 썸네일 영상이 흰 바탕의 영상과 오아시스 영상으로 통일되지 않음 (<그림 부록A-5> 참조)
- (검색-3) 이미지 정보들이 미리보기가 제공되지 않음. 들어가서 확인해야 하는 번거로움이 존재
 - 예시) 이미지 검색 → '자연', 미리보기 제공되지 않음 (<그림 부록A-6> 참조)

- 개선점: 이미지나 동영상 등의 미리보기 이미지가 필수적인 정보자원들에 한해서는 미리보기 썸네일 이미지를 제공하여 원본에 대한 대강의 정보를 확인할 수 있는 방안에 대해 고려할 수 있음
- (검색-4) 사진이나 동영상, 문서 확인을 위해서는 별도의 뷰어 프로그램 설치가 요구되어 번거롭고, 해당 정보 획득을 위해 긴 시간이 소요됨 (<그림 부록A-7> 참조)
 - 설명: 저작권 보호를 받는 국립중앙도서관의 국립중앙도서관 모든 자료는 디지털저작권관리(DRM) 기능이 포함된 전용 뷰어를 통해 이용이 가능함.
 - 개선점: 이용자의 이해를 돕기 위해 저작권 보호정책 차원의 전용뷰어 사용 필요성에 대한 안내가 필요함
- (검색-5) 프로그램 설치 전 이용 범위 안내 필요
 - 일부 자료는 프로그램 설치가 필수적임. 이에 응하여 프로그램 설치를 완료했으나 ‘국립중앙도서관 지정 PC에서만 이용 가능’하다는 문구가 뜸. 이용 범위에 ‘국립중앙도서관’이 명시되어 있지만, 이를 보지 못하고 프로그램 설치를 먼저 할 이용자가 있을 것으로 예상됨. 이를 사전 예방하기 위해, 프로그램 설치 허용을 누르긴 전에 이용 범위를 안내해 주면 좋겠음
- (검색-6) 동영상 실행에 대해 ‘웹사이트 보기’ 버튼 제공 (<그림 부록A-8> 참조)
 - 설명: 기존 ‘온라인 보기’를 수정 요청한 상황으로 일시적인 적용 오류였고 개선 중임
 - 개선점: ‘웹사이트 보기’ 버튼보다는 ‘동영상 보기’ 혹은 전체적으로 ‘보기/이동’ 버튼으로 변경
- (검색-7) 국립중앙도서관에서만 자료 이용이 가능한 것 (<그림 부록A-9> 참조)
 - 개선점: 현재 디지털 자료 납본제도나 저작권법 제도하에서는 국립중앙도서관 내에서만 자료 이용이 가능하나 자유이용 허락 라이선스(Creative Commons License) 유도를 통해 이용 범위를 확대하는 방안 시도해 볼 수 있음
- (검색-8) 과거 자료의 변경 기록을 확인하고 과거의 보존된 웹사이트에 방문하면 이미지가 깨져 이미지 로딩에 실패한 재생 웹페이지가 제시됨 (<그림 부록A-10> 참조)
- (검색-9) 아이콘 확인 어려움. 아이콘에 마우스 커서를 올려야만 어떤 것이 문서이고, 웹사이트인지 알려주는 것이 조금 불편했음
 - 예시 1) 컬렉션 > 2022 카타르 월드컵 화면에서 수집 건수 정보를 보여줄 때, 수집 유형 아이콘과 숫자가 나오지만, 아이콘의 의미를 확인하기 위해 마우스 커서를 올려야 하는 어려움이 있음 (<그림 부록A-11> 참조)
 - 예시 2) 아이콘이 너무 작아 존재 여부를 자세히 들여다보아야만 알 수 있었음 (<그림 부록A-12> 참조)

- 예시 3) 아이콘을 보면, 세 번째인 동영상 아이콘과 네 번째인 사진 아이콘은 의미를 파악하기 쉽지만, 첫 번째, 두 번째 아이콘을 보고 각각 웹사이트와 문서를 뜻하는지는 쉽게 파악하기 어렵고, 어떤 의미를 가지고 있는지 알아내기 힘들 (<그림 부록A-12> 참조)
- 개선점: 조그맣게 글자를 써넣고 링크를 다는 것이 한눈에 확인하기도 쉬움. 아이콘의 크기를 키우고, 직관적인 아이콘 모형을 사용하여 이용자가 빠르게 아이콘을 인식하고, 아이콘을 눌러 자료를 탐색할 수 있도록 지원해야 함
- (검색-10) 12개씩 24개씩과 같이 나눠서 보는 것(<그림 부록A-13> 참조)이 아닌 전체 보기와 같은 시스템이 있으면 전체를 조망하면서 선택할 수 있을 것임
- 개선점: 윈도우즈 탐색기에서 자세히 보기와 같은 제목 위주의 리스트 자료 보기를 제시할 수 있을 것임
- (검색-11) 웹사이트 URL 링크 표시가 안 되어 있음 (<그림 부록A-14> 참조)
- 웹사이트 URL은 링크 표시(파란색 글씨에 밑줄)가 되는 것이 기본(default)인데 링크 표시가 되면 인지적 측면에서 링크가 있다는 것을 마우스로 가져다 대지 않아도 한눈에 파악할 수 있고 피로도 측면에서 마우스를 가져가는 마우스 움직임(action)을 줄임으로써 피로도를 줄일 수 있음
- (검색-12) 검색 기능의 불편함
- 오아시스 내에서는 오타를 자동 보정해 결과를 도출해 주거나 자동 완성과 같은 기능이 없어 키워드 입력 오타에 대해 수정하는 데에 불편함
- (검색-13) 복지 서비스 부재(장애인, 노인 등)
- 장애인과 노인 등 사회적 약자를 위한 복지 서비스가 부재함
- → 시각 장애인들을 위한 음성 지원 프로그램이나 노인을 위한 돋보기 기능 등 다양한 복지 서비스가 필요함
- (검색-14) 오아시스 이용 (빅)데이터 활용 이용자 맞춤 개인화 서비스의 필요성
- 자료를 탐색하여 이전에 보았던 자료를 안내해 주고, 다시 열람할 수 있는 기능이 없어 계속 이전으로 되돌아가는 과정에서 시간이 낭비되고 비효율적이라는 느낌이 듦
- → OASIS 웹사이트에 로그인한 이용자에게 최근 열람한 자료, 많이 검색한 키워드 등을 안내해 주는 서비스 등을 제공해야 함. 또한, 이용자가 열람한 자료를 바탕으로 한 알고리즘 기반 자료 추천 서비스, 자료 탐색을 끝내고 스크랩한 자료들을 한눈에 볼 수 있도록 정리하는 스크랩 기능을 만들어 이용자의 자료 활용을 용이하게 할 수 있음
- → 안내 서비스, 추천 서비스 등 개인화 분석 서비스를 제공하여 이용자의 OASIS에 대한 만족도를

향상할 필요가 있음

- (검색-15) 검색 항목에 <초성으로 검색하기>를 이용했을 때 선택한 초성이 제목을 말하는 것인지, 발행처인지, 사이트 이름이 정해지지 않아 불편함이 있었음 (<그림 부록A-15> 참조)
- (검색-16) 모바일 사용의 불편함
 - PC가 아닌 모바일 기기에서는 오아시스 사용 시 조작이 불편하기에 모바일에서도 편리하게 사용할 수 있는 방안이 필요함
- (검색-17) 일부 분류 주제에 있어 소수 수집자료
 - ‘KDC별 주제 검색’에서 일부 분야(잡지, 사전)에는 자료가 소수만 수집됨. (<그림 부록A-16> 참조)
 - 개선점: 국가도메인 전체 대상 수집 시도하나 기술적으로 수집이 유효한 사이트만 수집된 결과로 수집기 개선을 통해 자료를 점진적으로 채워 나가야 할 필요가 있음
- (검색-18) 앞선 정보테크. 한국의 링크가 작동하지 않음 (<그림 부록A-17>, <그림 부록A-18> 참조)
- (검색-19) 웹사이트 URL 이동 불가
 - ㈜태창(<https://www.needles.co.kr>, 한국형 발사체개발사업단(<http://www.kslv2.or.kr/>), 한국컴퓨터비전학회(<http://kcvs.kr/front/conference/2017>) 웹사이트도 마찬가지로 URL로 이동 불가 (<그림 부록A-19> 참조)
- (검색-20) 상세 검색 화면에서 ‘소장자료’ 탭과 ‘웹사이트 본문’ 탭의 차이가 명확하지 않음. 웹사이트 본문은 무엇의 구분인가? 필드 검색과 전문 검색의 차이인가?
 - ‘소장자료’ 탭은 메타(필드) 검색이 적용되고 ‘웹사이트 본문’ 탭은 전문(full-text) 검색이 적용됨
 - 도움말을 통해 설명이 일부 제시되고 있음
 - 개선점: 상세 검색의 메뉴 탭의 명칭을 검색 기법과 검색 대상을 함께 표현하는 방안(예, ‘소장자료 메타검색’, ‘웹사이트 본문 전문 검색’)을 고려해 볼 수 있겠음
 -
- (검색-21) 웹사이트 본문을 클릭했을 때, 검색 시간이 소장자료 검색에 비해 오래 걸린다(17초)
 - 전문 검색의 경우, 메타(필드) 검색에 비해 기본적으로 색인 및 검색 시간이 오래 걸림
 - 개선점: 색인기 및 검색기의 성능을 확인하여 기존 색인 검색기의 문제점이 파악되면 개선하거나 일반적으로 성능이 우수한 최신 기술을 적용 시도해 필요가 있음
- (검색-22) 자료유형에서 웹사이트, 누리집, 웹페이지, 뉴스, 웹사이트의 구분 및 분류가 직관적으로 이해되지 않음 (<그림 부록A-20> 참조)

- 자료유형은 모두 웹사이트이며, 2022년부터 장르 주제명 부여하여 누리집, 웹페이지, 블로그, 뉴스기사 등으로 세분화하고 있음
 - 개선점: 메뉴에 사용된 용어에 대한 분류 체계에 대한 설명 추가되면 이용자의 이해도를 제고할 수 있음
- (검색-23) '웹사이트 본문'을 선택할 경우, 자료유형이 full-text로 나타나지만, 전체를 선택하면 full-text 자료유형은 보이지 않음 (<그림 부록A-21> 참조)
- (검색-24) SINDSOFT의 경우, 검색에서는 검색 결과가 있으나 컬렉션, 재난아카이브, SNS에서는 검색결과가 없으며 웹트렌드에서는 검색이 됨 (<그림 부록A-22>, <그림 부록A-23>, <그림 부록A-24>, <그림 부록A-25> 참조)
- 수집한 웹사이트 전체는 일반 검색이 적용되며, 포괄적 수집한 웹사이트 대상으로 웹트렌드가 적용되었음.
 - 컬렉션, 재난아카이브, SNS는 큐레이션 개념으로 검색 대상이 다름
 - 개선점1: 웹트렌드 검색 대상과 다른 컬렉션들(컬렉션, 재난, SNS)에 대한 대상이 다름을 설명할 필요가 있음.
 - 개선점2: 현재 웹트렌드 분석 기술은 오아시스의 특화된 우수한 데이터 분석기술이며 최신 데이터 분석 기술을 추가하여 고도화할 필요가 있음
- (검색-25) 검색에서 마우스 조작 시 검색 아이콘의 박스는 작고, 분리되어 있지 않아 색이 다르고 큰 마지막 상세검색을 클릭하게 되어 불편함 (<그림 부록A-26>, <그림 부록A-27>, <그림 부록A-28> 참조)
- (검색-26) 상세검색에서 발행연도를 선택할 때 좌우로 한 달씩만 이동할 수 있고, 연 단위로 이동하거나 블록형으로 선택할 수 없어 불편함 (<그림 부록A-29>, <그림 부록A-30>, <그림 부록A-31> 참조)

○ 검색 종합시사점

순번	도출된 시사점
1	웨이백 재생 무한 버퍼링
2	디폴트 이미지 상이함
3	이미지 정보들이 미리보기가 제공되지 않음. 들어가서 확인해야 하는 번거로움이 존재
4	사진이나 동영상, 문서 확인을 위해서는 별도의 뷰어 프로그램 설치가 요구되어 번거롭고, 해당 정보 획득을 위해 긴 시간이 소요됨
5	프로그램 설치 전 이용 범위 안내 필요

순번	도출된 시사점
6	동영상 실행에 대해 '웹사이트 보기' 버튼 제공
7	국립중앙도서관에서만 자료 이용이 가능한 것
8	과거 자료의 변경 기록을 확인하고 과거의 웹사이트에 방문하면 이미지가 깨져 나옴
9	아이콘 확인 어려움. 아이콘에 마우스 커서를 올려야만 어떤 것이 문서이고, 웹사이트 인지 알려주는 것이 조금 불편했음
10	12개씩 24개씩과 같이 나눠서 보는 것이 아닌 전체 보기와 같은 시스템이 있으면 전체를 조망하면서 선택할 수 있을 것임
11	웹사이트 URL 링크 표시가 안 되어 있음
12	검색 기능의 불편함
13	복지 서비스 부재(장애인, 노인 등)
14	오아시스 이용 (빅)데이터 활용 이용자 맞춤 개인화 서비스의 필요성
15	검색 항목에 <초성으로 검색하기>를 이용했을 때 선택한 초성이 제목을 말하는 것인지, 발행처인지, 사이트 이름이 정해지지 않아 불편함이 있음
16	모바일 사용의 불편함
17	일부 분류 주제에 있어 소수 수집자료
18	앞선 정보테크. 한국의 링크가 작동하지 않음
19	웹사이트 URL 이동 불가
20	상세검색 화면에서 소장자료와 웹사이트 본문 탭의 차이가 명확하지 않음
21	웹사이트 본문을 클릭했을 때, 검색 시간이 소장자료 검색에 비해 오래 걸림(17초)
22	자료유형에서 웹사이트, 누리집, 웹페이지, 뉴스, 웹사이트의 구분 및 분류가 직관적으로 이해되지 않음
23	'웹사이트 본문'을 선택할 경우, 자료유형이 full-text로 나타나지만, 전체를 선택하면 full-text 자료유형은 보여지지 않음
24	SNDSOFT의 경우, 검색에서는 검색결과가 있으나 컬렉션, 재난아카이브, SNS에서는 검색결과가 없으나 웹트렌드에서는 검색이 됨
25	검색에서 마우스 조작 시 검색 아이콘의 박스는 작고, 분리되어 있지 않아 색이 다르고 큼지막한 상세검색을 클릭하게 되어 불편함
26	상세검색에서 발행연도를 선택할 때 좌우로 한 달씩만 이동할 수 있고, 연 단위로 이동하거나 블록형으로 선택할 수 없어 불편함

2) 컬렉션

○ (컬렉션-1) 보기 유형에서 12개 보기는 가능하나 24개 보기 이상은 동작하지 않음 (<그림 부록

- A-32>, <그림 부록A-33> 참조)
- 개선점: 버퍼링 시간을 단축하기 위해 여러 방안으로 노력해야 함
- (컬렉션-2) 컬렉션 부분에서 3.1운동 및 임시정부 100주년 기념 자료의 출처 링크가 깨졌음 (<그림 부록A-34>, <그림 부록A-35> 참조)
- 컬렉션 구축 당시 존재했으나 지금은 사라진 사이트임
 - 개선점: <https://www.together100.go.kr> 웹사이트에 대한 변경 기록 URL로 변경하는 방안
 - 전반적으로 링크 깨짐 현상에 대해 자동으로 추출하는 도구 개발하여 주기적으로 점검하는 방안
- (컬렉션-3) 미리보기 이미지가 제공되는 경우, 이미지가 모두 보이는데 2-4초 가량 소요됨 (<그림 부록A-36> 참조)
- 개선점: 미리보기를 위한 썸네일 이미지 생성 및 제시를 통한 웹페이지 로딩속도를 높이는 방안
- (컬렉션-4) 한국영화 100년 기념 사업 홈페이지 링크 깨짐- 입장 시 503 상태코드가 발생 (<그림 부록A-37>, <그림 부록A-38> 참조)
- 컬렉션 구축 당시 존재했으나 지금은 사라진 사이트임
 - 개선점: <https://www.together100.go.kr> 웹사이트에 대한 변경 기록 URL로 변경하는 방안
 - 전반적으로 링크 깨짐 현상에 대해 자동으로 추출하는 도구 개발하여 주기적으로 점검하는 방안
- (컬렉션-5) '컬렉션' 장서(collection)에서 검색할 때, 검색 키워드 입력 후 엔터를 치면 실행되나 검색 버튼(돋보기 모양 이미지) 클릭은 동작하지 않음 (<그림 부록A-39> 참조)
- 컬렉션 종합적 시사점

순번	도출된 시사점
1	보기유형에서 12개 보기는 가능하나 24개 보기 이상은 동작하지 않음
2	컬렉션 부분에서 3.1운동 및 임시정부 100주년 기념 자료의 출처 링크가 깨졌음
3	미리보기 이미지가 제공되는 경우, 이미지가 모두 보이는데 2~4초 가량 소요됨
4	한국영화 100년 기념사업 홈페이지 링크 깨짐
5	'컬렉션' 장서(collection)에서 검색할 때, 검색 키워드 입력 후 엔터를 치면 실행되나 검색버튼(돋보기 모양 이미지) 클릭은 동작하지 않음

3) 재난아카이브

- **(재난아카이브-1)** 자료유형에 따른 적절한 유형의 자료 제시 오류(이미지 선택 시 동영상 제시, 동영상 선택 시 이미지 제시) (<그림 부록A-40>, <그림 부록A-41> 참조)
- **(재난아카이브-2)** 해외재난 분류 체계가 상세하지 않음 (<그림 부록A-42> 참조)
 - 개선점: 해외재난도 우리나라의 재난을 분류해 둔 것과 마찬가지로 분류를 해두면 가독성이 더욱 높아질 것임
- **(재난아카이브-3)** 컬렉션 가장 최신 기사가 2022년 기사라 현재 기사가 궁금하면 따로 검색해 보는 방법밖에 없어서 불편했음 (<그림 부록A-43> 참조)
 - 매년 주요 이슈 컬렉션을 적시 수집하고 있으나 메타데이터 구축 및 컬렉션 구축하는 데 시간이 소요됨
 - 개선점: 2년 엠바고에 대한 설명과 함께 오아시스의 최대 목적이 보존이라는 점과 최신 기사에 대한 접근이 제한적일 수 있다는 정보를 이용자가 쉽게 알 수 있는 방법으로 제시할 필요가 있음.
- **(재난아카이브-4)** 재난아카이브에서 코로나바이러스에 대한 출처인 질병관리본부(<http://ncov.mohw.go.kr/>) 링크를 들어가자 접근 금지된 자료였음 (<그림 부록A-44> 참조)
 - 그 당시 공개 웹사이트를 수집한 것이었으나 지금은 접근 금지되어 수집시기와 접근시기의 접근정책 변화에 따른 사항임
 - 오아시스 자체의 문제라기보다는 수집대상 사이트의 공개정책에 따른 문제점으로 이에 대한 안내가 이루어지면 이용자의 이해를 도울 수 있음
- **(재난아카이브-5)** 재난아카이브에서 자료를 선택한 후 첫 번째 로딩 후에 재로딩하는 현상 (<그림 부록A-45> 참조)
 - 예상원인: 처음 로딩 후 스타일 적용을 위해 다시 메타DB 재로딩하는 것으로 예상됨. 스타일 적용 시 메타DB 재로딩이 있는지 확인할 필요가 있음
- **(재난아카이브-6)** 표기 오류
 - ‘재난 아카이브’에서 ‘재난유형별’의 ‘해외’에는 세 개의 자료가 표기되어 있지만 ‘지역별’의 ‘해외’에는 자료가 전혀 표기되어 있지 않음. 각 키워드에 적합한 문헌임에도 동시적으로 표기되지 않는 경우, 수정이 필요해 보임

○ 재난 아카이브 종합적 시사점

순번	도출된 시사점
1	자료유형에 따른 적절한 유형의 자료 제시 오류(이미지 선택 시 동영상 제시, 동영상 선택 시 이미지 제시)
2	해외재난 분류 체계가 상세하지 않음
3	컬렉션 가장 최신 기사가 2022년 기사라 현재 기사가 궁금하면 따로 검색해 보는 방법밖에 없어서 불편했음
4	재난아카이브에서 코로나바이러스에 대한 출처인 질병관리본부 (http://ncov.mohw.go.kr/) 링크를 들어가자 접근 금지된 자료였음
5	재난아카이브에서 자료를 선택한 후 첫 번째 로딩 후에 재로딩하는 현상
6	표기 오류: ‘재난 아카이브’에서 ‘재난유형별’의 ‘해외’에는 세 개의 자료가 표기되어 있지만 ‘지역별’의 ‘해외’에는 자료가 전혀 표기되어 있지 않다. 각 키워드에 적합한 문헌임에도 동시에 표기되지 않는 경우, 수정이 필요해 보임

4) 소셜미디어

- 소셜미디어 컬렉션은 2018년 소스 분석 방법으로 시범 수집하여 구축한 아카이브임
- 현재는 해당 소셜미디어(twitter)의 지속성 가능성이 불확실하여 추가 수집 및 내용 확장이 불가한 상황임
- 소셜미디어 메뉴에 대한 사용성 평가 결과는 다음과 같음
- (소셜미디어-1) 하이퍼링크화 되지 않아 불편함 (<그림 부록A-48> 참조)
- (소셜미디어-2) 제일 처음으로 올라온 트윗은 2010년인데 가입일은 2018년도로 뜨는 등 가입일과 트윗이 맞지 않는 경우가 있음 실제로 홈페이지에 들어가 보면 가입일은 2010년도로 뜸. 이런 경우가 많음 (<그림 부록A-49>, <그림 부록A-50> 참조)
 - 개선점: 확인 후 수정이 필요함
- (소셜미디어-3) 제한된 소셜미디어
 - 소셜미디어 창에 들어갔을 때 여러 기관의 소셜미디어가 정리되어 있음. 하지만 현재 트위터 계정들만 정리되어 있고 여러 기관이 유튜브, 인스타그램 등의 다양한 소셜미디어를 가지고 있는 만큼 다른 소셜미디어 계정들도 정리되어 있어야 한다고 생각함

○ (소셜미디어-4) 각 기관 문양 이미지 깨짐

- 고용노동부의 공식 트위터를 소개하는 기관 문양의 사진은 선명한 반면, 수집된 트윗에서의 고용노동부의 문양 사진은 선명하지 않고 이미지가 깨져있음. 고용노동부뿐만 아니라 다른 기관 모두 수집된 트윗에서의 기관 문양 사진은 깨져있음 (<그림 부록A-51> 참조)

○ (소셜미디어-5) 환경부 트위터에 자세히 보기 링크를 들어가도 페이지를 볼 수 없음 (<그림 부록 A-52>, <그림 부록A-53>, <그림 부록A-54>, <그림 부록A-55> 참조)

- 한 트위터 자세히 보기 URL(<https://bit.ly/2GY9iue>)이 링크화가 되어 있지 않고 페이지 링크 접속이 불가함
- 다른 트위터 자세히 보기 URL(http://bitly.kr/P3kIV_pic.twitter.com/lXjlliFAjV)이 링크화가 되어 있지 않고 페이지 링크 접속이 불가함
- 추후 링크 접속 시도 시 링크 접속이 가능함을 확인하였으며 이를 통해 일시적 접속 오류임을 확인함
- 본문에 URL이 제시될 경우 하이퍼링크를 제공하면 사용자가 편리하게 접속시도해 볼 수 있음

○ (소셜미디어-6) 소셜미디어 설명 부족

- 소셜미디어 관련하여 해당 기능이 있는 이유나 활용법에 대해 안내해 주면 좋을 것 같음

○ 소셜미디어 종합적 시사점

순번	도출된 시사점
1	하이퍼링크화 되지 않아 불편함
2	제일 처음으로 올라온 트윗은 2010년인데 가입일은 2018년도로 뜨는 등 가입일과 트윗이 맞지 않는 경우가 있음 실제로 홈페이지에 들어가 보면 가입일은 2010년도로 뜸 이런 경우가 많음
3	제한된 소셜미디어: 소셜미디어 창에 들어갔을 때 여러 기관의 소셜미디어가 정리되어 있음. 하지만 현재 트위터 계정들만 정리되어 있고 여러 기관이 유튜브, 인스타 등의 다양한 소셜미디어를 가지고 있는 만큼 다른 소셜미디어 계정들도 정리되어 있어야 한다고 생각함
4	각 기관 문양 이미지 깨짐: 고용노동부의 공식 트위터를 소개하는 기관 문양의 사진은 선명한 반면, 수집된 트윗에서의 고용노동부의 문양 사진은 선명하지 않고 이미지가 깨져있음. 고용노동부뿐만 아니라 다른 기관 모두 수집된 트윗에서의 기관 문양 사진은 깨져있음
5	환경부 트위터에 자세히 보기 링크를 들어가도 페이지를 볼 수 없음
6	소셜미디어 설명 부족

5) 웹트렌드⁸⁾

- 웹트렌드 분석 및 검색 서비스는 오아시스에서 자체 개발 적용한 데이터 분석 서비스로 풍부한 웹 데이터 분석 기능을 제공함
- 2022년 이후 업데이트 안되고 있는 상황으로 고도화하여 이용자의 데이터 접근성을 높일 수 있는 서비스임
- 웹트렌드 메뉴에 대한 사용성 평가 결과는 다음과 같음
- (웹트렌드-1) 안내 메시지가 화면에 가려지는 불편함 (<그림 부록A-56> 참조)
 - 원인: 여백 부족
 - 개선점: 이용자가 정보를 확인하는 데 불편함이 없는 크기의 여백을 제공함
- (웹트렌드-2) 그래프 보기에서 중복 라디오버튼 (<그림 부록A-57> 참조)
- (웹트렌드-3) 웹트렌드 분석 도움말이 기능 위주의 설명임. 의미와 해석 방법에 대해서 좀 더 상세한 설명이 있으면 좋겠음. 예를 들어, 레이블에 한국, 미국 등과 같은 나라명이 제시되어 있는데, 이 레이블의 의미가 무엇이고 그래프는 어떻게 해석할 수 있는지 제시되면 이용자가 웹트렌드 기능을 이해하는 데 도움이 될 것임 (<그림 부록A-58> 참조)
 - 개선점: 자세한 사항은 웹트렌드 활용예제를 참고하라는 안내메시지 포함
- (웹트렌드-4) 키워드 비중 도움말이 너무 간단함 (<그림 부록A-59> 참조)
 - 개선점: 키워드 비중이 키워드(용어) 빈도(term frequency)의 비중을 표현한 것인지 키워드가 출현한 문서(document frequency)의 비중을 표현한 것인지 설명을 추가하면 도움이 될 것임
- (웹트렌드-5) 웹트렌드의 대부분 링크를 클릭했을 때 사이트에 연결할 수 없음. 메시지가 결과로 나오음 (<그림 부록A-60>, <그림 부록A-61> 참조)
 - 개선점: 웹트렌드의 링크를 전반적으로 점검할 필요가 있음. 자동화 툴을 개발 적용할 것을 추천함
- (웹트렌드-6) "웹트렌드"라는 하나의 메뉴에서 갈라진 하위 메뉴임에도 불구하고 웹트렌드, 웹트렌드 활용예제, 태그클라우드 세 페이지의 UI 배치와 구성이 통일되어 있지 않아 이용이 어려움
 - 웹트렌드와 태그클라우드에서는 왼쪽의 바로가기가 없어 활용예제 혹은 태그클라우드로 이동할 수 없지만, 웹트렌드 활용예제에서는 둘 다 이동할 수 있음 (<그림 부록A-62>, <그림 부록A-63>, <그림 부록A-64> 참조)

8) 웹트렌드 검색 및 분석 서비스는 여건 부족으로 2022년 이후 업데이트 안 되고 있음

- 개선점: 웹트렌드와 태그클라우드에도 왼쪽의 바로가기 메뉴 제공

○ 웹트렌드 종합적 시사점

순번	도출된 시사점
1	안내메세지가 화면에 가려지는 불편함
2	그래프 보기에서 중복 라디오버튼
3	웹트렌드 분석 도움말이 기능 위주의 설명임. 의미와 해석 방법에 대해서 좀 더 상세한 설명이 있으면 좋겠음. 예를 들어, 레이블에 한국, 미국 등과 같은 나라명이 제시되어 있는데, 이 레이블의 의미가 무엇이고 그래프는 어떻게 해석할 수 있는지 제시되면 이용자가 웹트렌드 기능을 이해하는 데 도움이 될 것임
4	키워드 비중 도움말이 너무 간단함
5	웹트렌드의 대부분 링크를 클릭했을 때 사이트에 연결할 수 없음. 메시지가 결과로 나옴
6	"웹트렌드"라는 하나의 메뉴에서 갈라진 하위 메뉴임에도 불구하고 웹트렌드, 웹트렌드 활용예제, 태그클라우드 세 페이지의 UI 배치와 구성이 통일되어 있지 않아 이용이 어려움

6) 소개

- (소개-1) 자주 묻는 질문은 있었지만, 직접적인 질의응답 창구가 없음. 사이트를 사용하다가 궁금한 점이나 모르는 점이 생겼을 경우 바로 질문을 주고받을 수 있는 질의응답 창이 필요

○ 소개 종합적 시사점

순번	도출된 시사점
1	자주 묻는 질문은 있었지만, 직접적인 질의응답 창구가 없음. 사이트를 사용하다가 궁금한 점이나 모르는 점이 생겼을 경우 바로 질문을 주고받을 수 있는 질의응답 창이 필요

3.3.3 전반적 사용 소감 요약 및 시사점

① 검색 기능 관련 문제

- 불충분한 검색 결과의 관련성 : 일부 검색어에 대한 결과가 사용자의 기대와 다르게 나타나는 경우가 있음
- 필터링 옵션의 제한성 : 필터링 옵션이 제한적이어서 원하는 자료를 세밀하게 검색하기 어려운 문제가 발생함

② 다국어 지원 부족

- 한국어 외 사용자를 위한 정보 부족 : 비한국어 사용자가 이용할 수 있는 콘텐츠나 정보가 부족하여 사용 편의성이 떨어짐

③ 모바일 호환성 문제

- 모바일 기기에서의 불편한 접근성 : 스마트폰이나 태블릿과 같은 모바일 기기에서 웹사이트가 제대로 렌더링 되지 않아 사용자 경험이 저하됨

④ 검색 기록 관리 이슈

- 검색 기록 저장 및 삭제의 불편함 : 검색 기록이 자동으로 저장되거나 삭제되지 않아 사용자가 개인 정보 관리를 어려워함

⑤ 페이지 레이아웃 문제

- 화면 해상도 및 디바이스 호환성 문제 : 특정 화면 해상도나 디바이스에서 페이지 레이아웃이 깨지는 현상이 발생하여 사용자 경험이 저하됨

⑥ 검색 필터 오작동

- 특정 검색 필터 오류 : 특정 검색 필터를 적용했을 때 예상과 다른 결과가 나타나는 경우가 발생하여 검색 정확성에 문제가 생김

⑦ 링크 오류

- 링크 연결 실패 : 클릭 시 연결되지 않거나 잘못된 페이지로 연결되는 링크가 존재하여 사용자가 원하는 정보에 접근하기 어려움

⑧ 접근성 문제

- 장애인을 위한 접근성 부족 : 장애인 및 다양한 능력을 가진 사람들이 콘텐츠에 접근하기 어려운 경우가 있어 포용성이 부족함

⑨ 문서화 부족

- 서비스 이용 방법에 대한 부족한 설명 : 사용자가 웹사이트의 기능을 제대로 활용할 수 있도록 한정된 정보 제공으로 어려움을 겪을 수 있음

⑩ 자료 접근의 한계

- 국립중앙도서관 외 자료 접근의 제한 : 다양한 자료에 접근이 어려워 사용자가 필요한 자료를 찾기 어려움

⑪ 유관 웹사이트의 링크 제공 협약을 통한 적극적 홍보 필요

- 국립중앙도서관의 경우에는 국립어린이청소년도서관 등 분관이 있으며, 연관되어 있는 도서관인 경우 오아시스의 링크를 눈에 띄는 곳에 두는 것이 좋다고 생각함. 국립어린이 청소년도서관은 어린이와 청소년을 대상으로 한 특화 도서관이어서 그런지, 오아시스와 같은 웹사이트가 많이 발전되어 있지 않은 느낌이었음. 그렇기에 오아시스의 링크를 눈에 띄는 곳에 두면 사이트를 더 풍성하게 만들 수 있을 거라고 생각함

⑫ 이용자와의 원활한 상호 작용을 위한 챗봇 제공 고려 필요

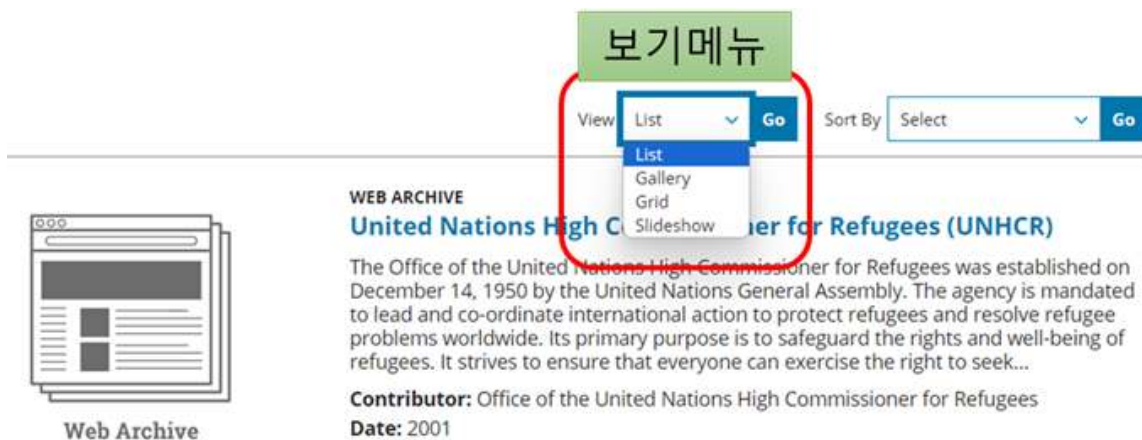
- 자주하는 질문에서는 해결되지 않는 부분이 있을 수 있음. 이용자와의 상호 작용성을 높이기 위해 소개 부분에서 자주 하는 질문 말고도 질문을 추가할 수 있는 기능이나 인공지능 챗봇을 추가하는 방안이 필요함

3.4 우수 운영사례 분석

3.4.1 미국 의회도서관(LC)

○ (LC-1) 보기 유형으로 슬라이드쇼 제공

- 보기 형식 메뉴 제공(List, Gallery, Grid, Slideshow). 미 의회도서관 웹 아카이브 보기 메뉴는 반응형 웹을 적용하였으며 하나의 보기형식을 정하면 반응형 웹에서도 그대로 유지됨
- 참고로 오아시스즈의 경우, 반응형 웹 디자인을 통해서, 윈도우 크기가 큰 화면에서 List 보기 형식에서 윈도우 크기가 작아지면서 Gallery 방식으로 변화하는 방식임
- 미국 의회도서관은 다양한 보기 메뉴를 제공함으로써 개인에 맞춘 보기 기능을 제공함

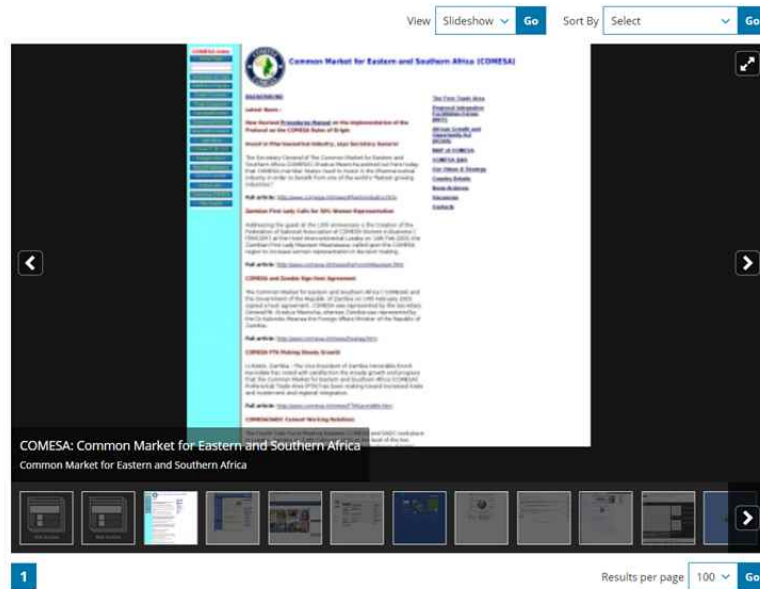


<그림 3-1> 미국 의회도서관 웹 아카이브가 제공하는 보기 메뉴(List, Gallery, Grid, Slideshow)

- 출처:

<https://www.loc.gov/collections/intergovernmental-organizations-web-archive/?st=list>

- 슬라이드쇼 보기 메뉴는 다른 보기 메뉴(list, gallery, grid)와 비교하여 화면의 틀은 바뀌지 않고 콘텐츠가 바뀌는 방식이므로 눈이 고정되어 피로도가 덜한 인터페이스임
- PC 환경에서뿐만 아니라 손가락으로 밀어서 조작하는 모바일 환경에 적합한 인터페이스임

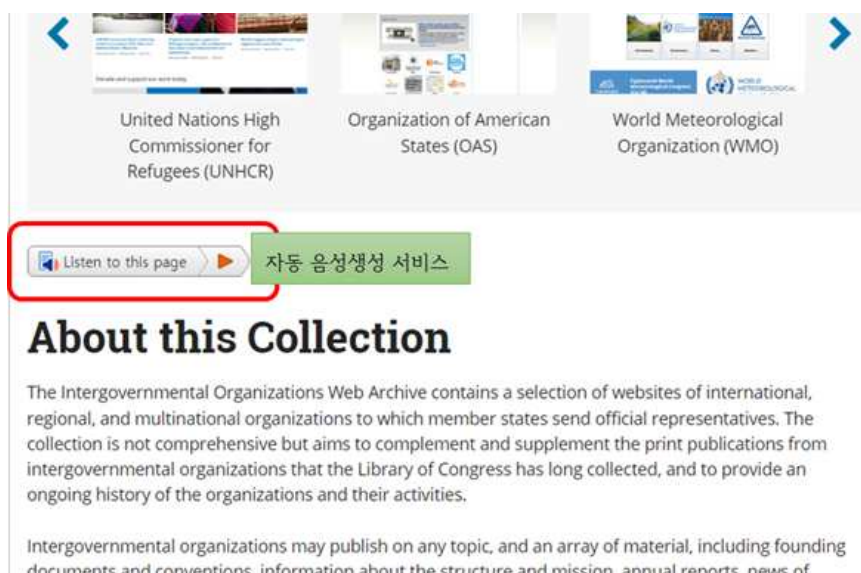


<그림 3-2> 미국 의회도서관 웹 아카이브 슬라이드쇼 보기 화면

- 출처: <https://www.loc.gov/collections/intergovernmental-organizations-web-archive/?st=slideshow#slide-3>

○ (LC-2) 자동 음성 생성 서비스 제공

- 자동 음성 생성 서비스를 제공하여 일반인들이나 시각장애인들에게 텍스트가 아닌 음성으로 정보를 제공할 수 있음. 본 기능은 컬렉션의 소개 페이지에 제공되고 있음



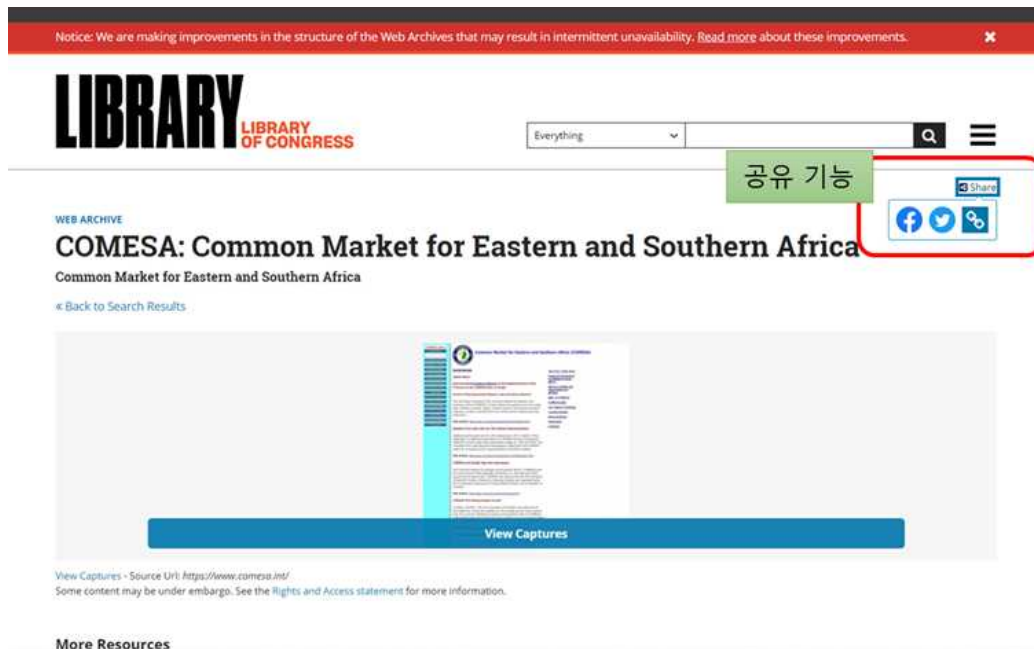
<그림 3-3> 미국 의회도서관 웹 아카이브 자동음성생성 서비스

- 출처: <https://www.loc.gov/collections/intergovernmental-organizations-web-archive/about>

t-this-collection/

○ (LC-3) 공유(share) 기능 제공

- 아래 그림의 웹페이지에서 우측 상단에, SNS를 통해 다른 미디어들과 공유할 수 있는 share 기능을 제공함
- 이러한 기능은 콘텐츠의 활용성을 높일 수 있는 좋은 도구임

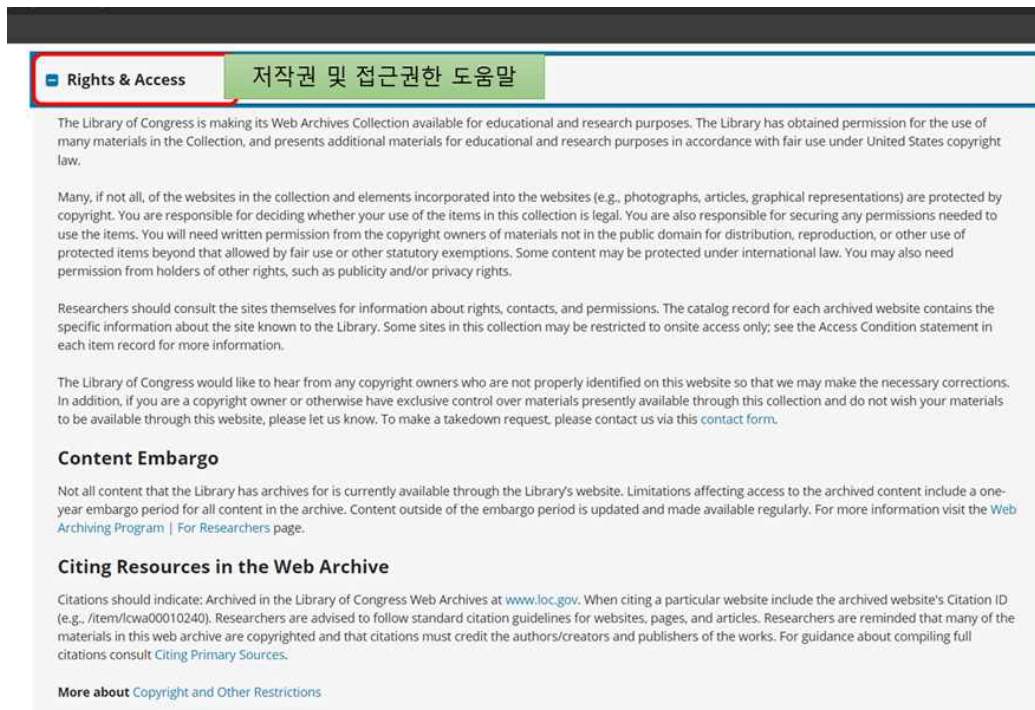


<그림 3-4> 미국 의회도서관 웹 아카이브 공유 기능

- 출처: <https://www.loc.gov/item/lcwaN0003866/>

○ (LC-4) 저작권 및 접근 권한 안내 메시지 제공

- 저작권 관련 안내 메시지는 이용자가 콘텐츠를 적법하게 활용할 수 있게 하는 매우 중요 요소로, 명확한 관련 정보가 제시될 수 있어야 함
- 이는 웹자원 이용을 적법하게 활용할 수 있도록 함을 통해 신뢰 가운데 웹자원을 공유 및 개방할 수 있도록 유도하는 방안임



<그림 3-5> 미국 의회도서관 웹 아카이브 저작권 및 접근권한 도움말 제공

- 출처: <https://www.loc.gov/item/lcwaN0003866/>

○ (LC-5) 인용(cite) 기능 제공

- 인용 기능은 논문(작성 시 참고문헌(references)에서 인용할 수 있도록, 주요 인용스타일(Chicago, APA, MLA)을 제공하고 있음
- 인용 기능은 연구자들로 하여금 아카이브에서 제공하는 웹자원을 공식적인 학술자료로 활용할 수 있도록 하는 데 편의성을 제공함
- 이 기능은 학술 연구자들이 공유하는 형식으로 앞의 일반인을 포함하는 공유(share) 기능과 대상만 다를 뿐 그 역할은 유사한 기능임
- 이 인용 기능을 이용해 각 웹자원 이용을 학술적으로 활용할 수 있도록 하기 위해서는, 유일하게 식별할 수 있는 유일 식별자를 제공하는 것이 필요함

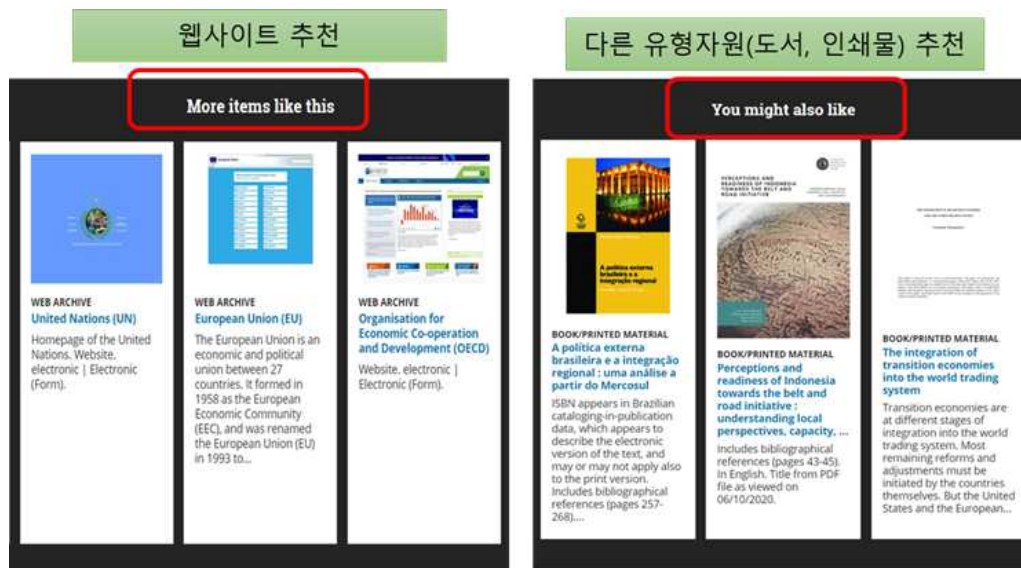


<그림 3-6> 미국 의회도서관 웹 아카이브 인용 기능

- 출처: <https://www.loc.gov/item/lcwaN0003866/>

○ (LC-6) 추천(recommend) 기능 제공

- 추천 기능은 두 파트를 제공하는데 하나는 동일 유형의 자료인 웹사이트를 추천('More items like this')하는 기능과 다른 유형의 자료인 도서, 인쇄물을 추천(You might also like)하는 기능을 제공함
- 이러한 기능은 이용자로 하여금 관련 자원들을 손쉽게 이용할 수 있는 접근점을 제공하는 것으로 오아시스 자원 이용 활성화에 도움을 줄 수 있는 좋은 기능임
- 다른 유형의 정보자원까지 추천함으로써 웹사이트의 특수성에 머무르지 않고 이용자들이 익숙한 기존 유형의 자원(도서 및 인쇄자료)으로 연관 지음으로써 웹자원에 대한 인식 장벽을 낮추는 효과를 제공하는 기능이라 할 수 있음

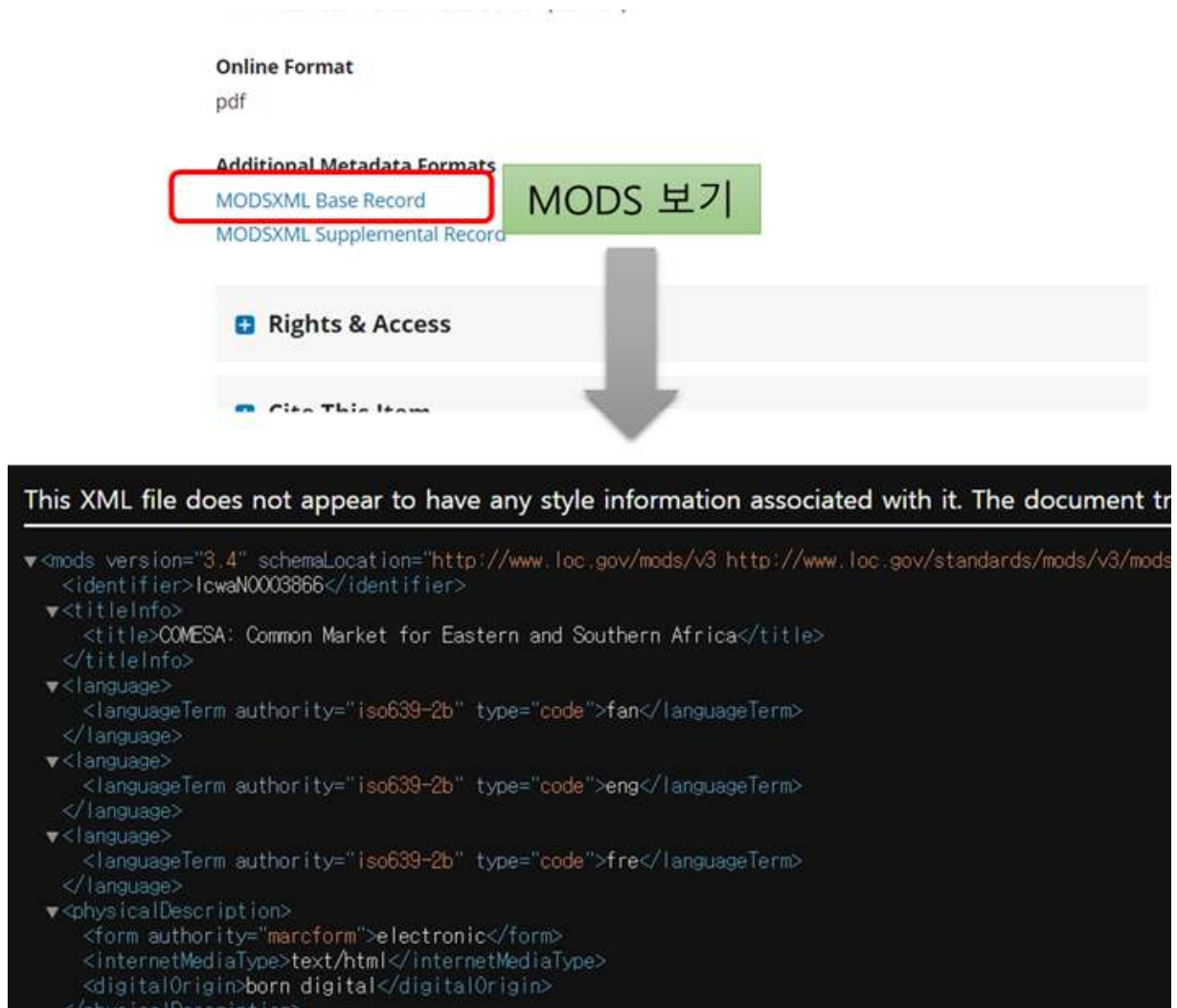


<그림 3-7> 미국 의회도서관 웹 아카이브 추천 기능

- 출처: <https://www.loc.gov/item/lcwaN0003866/>

○ (LC-7) 공유 및 개방 강화

- 국제인터넷보존컨소시엄(IIPC)은 공유 및 개방 강화를 전략으로 삼고 있음
- MODS 보기 기능을 통해 메타데이터에 대한 공개 및 공유 기능을 제공하고 있음
- 이 경우, 이용자가 읽거나 다운로드 받아서 활용하게 되겠으나, 해당 웹자원의 접근점을 자동화를 통해 확대하고 그 웹자원의 메타데이터의 활용성을 높이기 위해서는 웹로봇이 수집하거나 혹은 다른 시스템에나 틀에서 해당 정보를 활용할 수 있도록 API 형태로 제공될 필요가 있음



<그림 3-8> 미국 의회도서관 웹 아카이브 메타데이터 보기 기능

- 출처: <https://www.loc.gov/item/lcwaN0003866/>
- MODS 출처: https://tile.loc.gov/storage-services/service/webcapture/project_1/mods/general-web-archives/lcwaN0003866.xml

3.4.2 우수 운영기관(미국 의회도서관) 사례 분석의 시사점

순번	도출된 시사점
1	미국 의회도서관(LC) 우수 사례 조사 시사점→ 보기유형으로 슬라이드쇼 제공
2	미국 의회도서관(LC) 우수 사례 조사 시사점→ 자동 음성 생성 서비스 제공
3	미국 의회도서관(LC) 우수 사례 조사 시사점→ 공유(share) 기능 제공
4	미국 의회도서관(LC) 우수 사례 조사 시사점→ 저작권 및 접근권한 도움말 제공
5	미국 의회도서관(LC) 우수 사례 조사 시사점→ 인용 기능 제공
6	미국 의회도서관(LC) 우수 사례 조사 시사점→ 추천 기능 제공
7	미국 의회도서관(LC) 우수 사례 조사 시사점→ 메타데이터(MODS) 제공

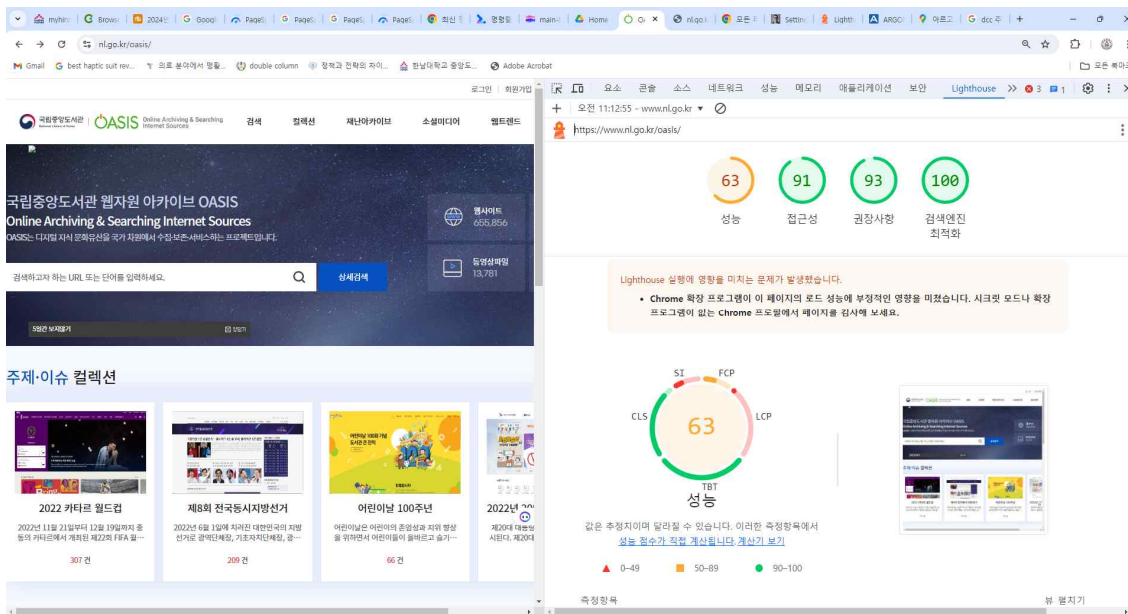
3.5 오아시스 누리집 성능 측정

3.5.1 성능 측정도구

- 웹사이트 성능 측정 주요 도구에는 1) Google 개발자 도구 Lighthouse, 2) PageSpeed Insights, 3) GTmetrix가 있음

1) 구글 개발자 도구 Lighthouse

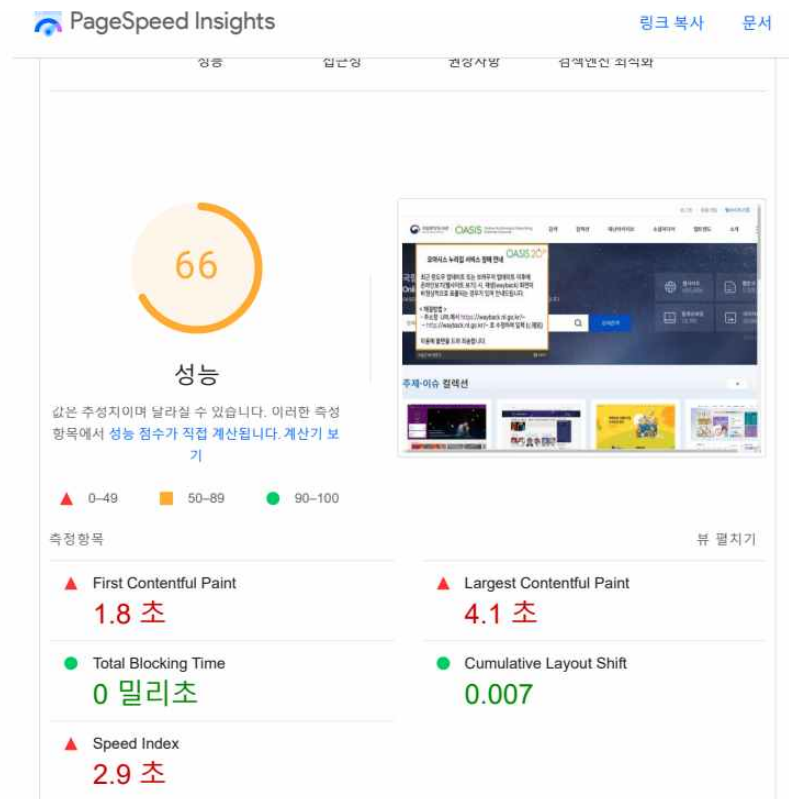
- Google 개발자 도구는 구글 크롬 브라우저에서 가장 손쉽게 사용할 수 있는 웹페이지 성능 측정 도구임
- 구글 크롬 브라우저에서 F12키를 누르고 전체 메뉴에서 Lighthouse 메뉴를 선택하고 분석하기 (Analyze)를 선택하면 아래 그림에서 볼 수 있듯이 성능, 접근성, 권장 사항, 검색엔진 최적화에 대한 분석을 제시해 줌



<그림 3-9> 웹브라우저 개발자 도구(F12) > 성능평가도구(Lighthouse)를 통해 오아시스 메인 웹페이지(<https://www.nl.go.kr/oasis/>) 분석

2) PageSpeed Insights

- Google PageSpeed Insights는 구글 크롬 개발자 도구 lighthouse와 마찬가지로 구글에서 제공하는 데스크톱 및 모바일 모두에서 사이트 페이지 성능을 테스트하는 도구임
- 이 도구는 Chrome 사용자 경험(CrUX) 보고서를 사용해 실제 방문자 성과를 분석하고 Google Lighthouse를 사용해 페이지에 성능 점수를 제공함



<그림 3-10> PageSpeed Insight를 통해 오아시스 메인 웹페이지(<https://www.nl.go.kr/oasis/>) 분석

3) GTmetrix

- GTmetrix는 사이트 성능에 대한 자세한 보고서를 얻기 위해 점점 더 많이 사용하고 있는 도구 중 하나임
- Google PageSpeed와 YSlow를 사용하여 페이지의 속도 성능을 분석하는 무료 도구임
- GTmetrix의 경우는 초기에 정해진 test server를 사용해야 하고, 원하는 test server region을 설정하거나 더 많은 웹사이트를 분석하기 위해 유료 회원 가입이 필요

3.5.2 GTmetrix를 이용한 오아시스 누리집 성능 분석

- 본 절에서는 최신 웹사이트 성능 평가 도구로 많이 사용되는 GTmetrix를 이용하여 오아시스 누리집 성능 분석을 수행함

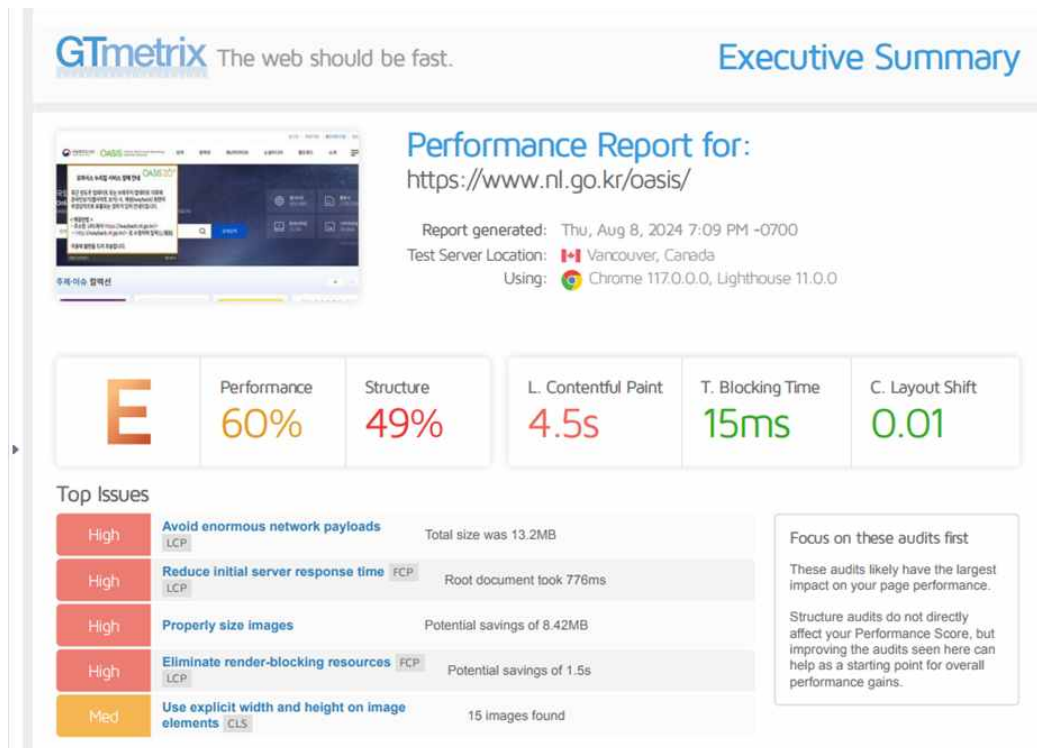
1) 오아시스 메인페이지 성능 분석

- 본 페이지는 <https://www.nl.go.kr/oasis/>에 대해서 성능 분석을 수행한 결과임
- 성과 등급⁹⁾은 E(50~59%)이고, 성과(performance)는 60%, 구조(structure)는 49%임. 성과(주황색)

9) 성과 등급(GTmetrix Grade)은 두 가지 백분율 기반 점수의 가중 평균임

는 중간이나 구조(빨간색)는 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간이 4.5초 걸려 문제가 심각함

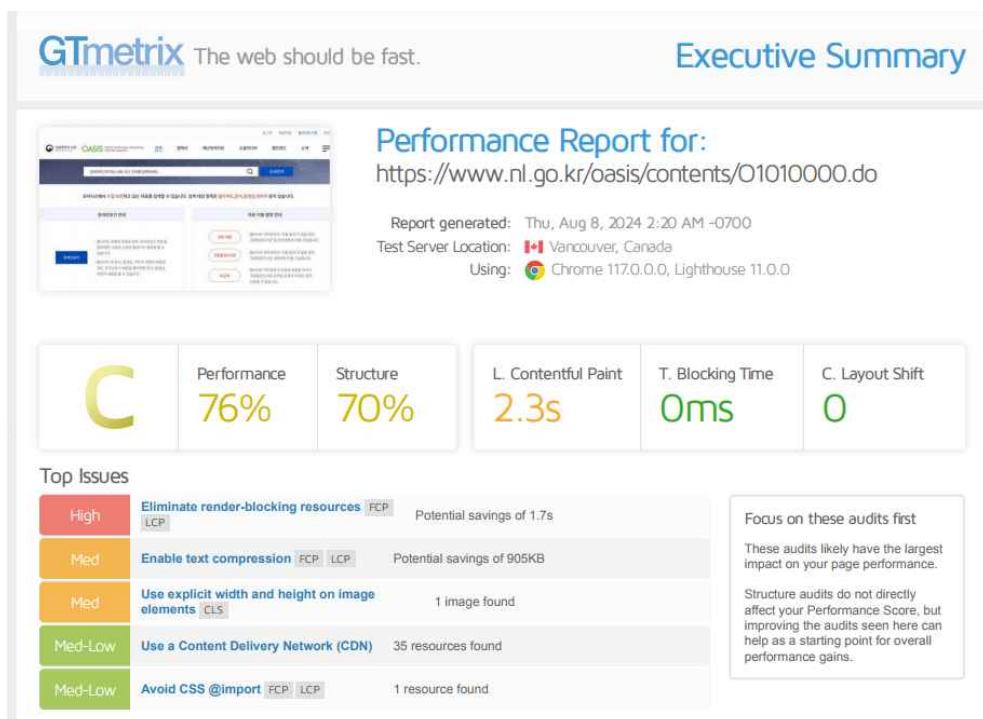
- 그 외, 심각한 네트워크 부하 회피, 초기 서버 응답 시간 감소, 영상 크기를 적절히 사용하기, 렌더 방해 요소 제거를 통해 웹페이지를 개선할 필요가 있음



<그림 3-11> 오아시스 메인페이지 성능 분석 화면

2) 검색 > 통합검색 성능 분석

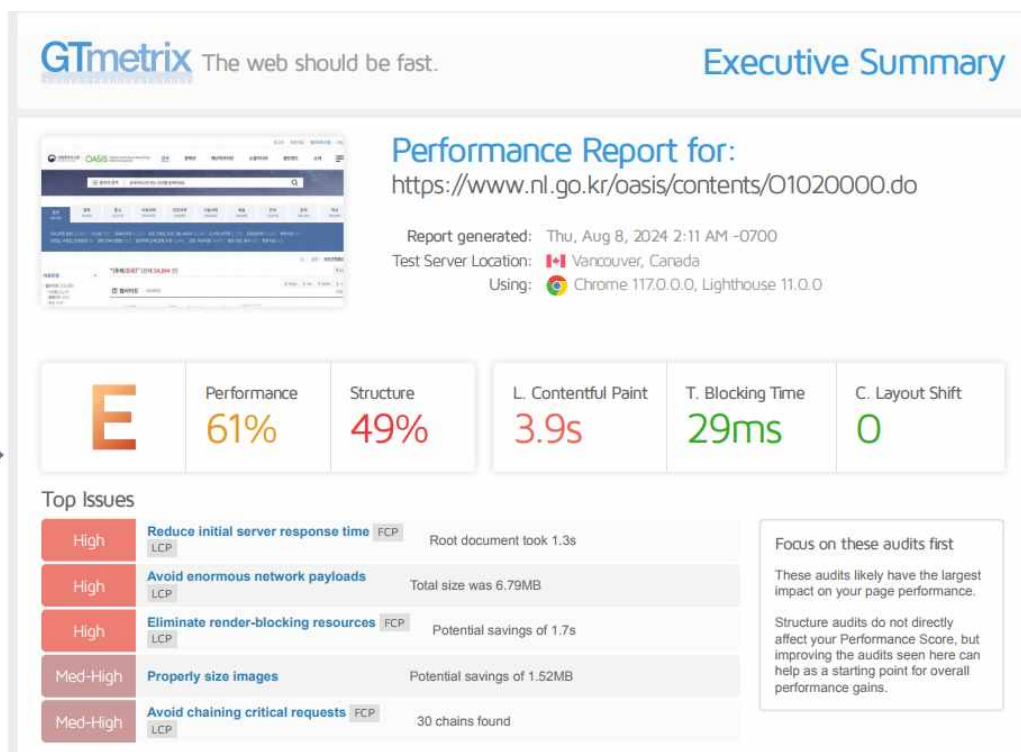
- 본 페이지는 <https://www.nl.go.kr/oasis/contents/O1010000.do>에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 C(70~79%)이고, 성과(performance)는 76%, 구조(structure)는 70%임. 성과(노란색)는 중간이나 구조(노란색)는 중급의 성능을 보임. 가장 큰 콘텐츠 로딩 시간이 2.3초 걸려 좀 더 줄일 필요가 있음
- 오아시스 메인 웹페이지와 마찬가지로, 심각한 네트워크 부하 회피, 초기 서버 응답시간 감소, 영상 크기를 적절히 사용하기, 렌더 방해 요소 제거를 통해 웹페이지를 개선할 필요가 있음



<그림 3-12> 오아시스 검색메뉴페이지 성능 분석 화면

3) 검색 > KDC 주제별 웹페이지 성능 분석

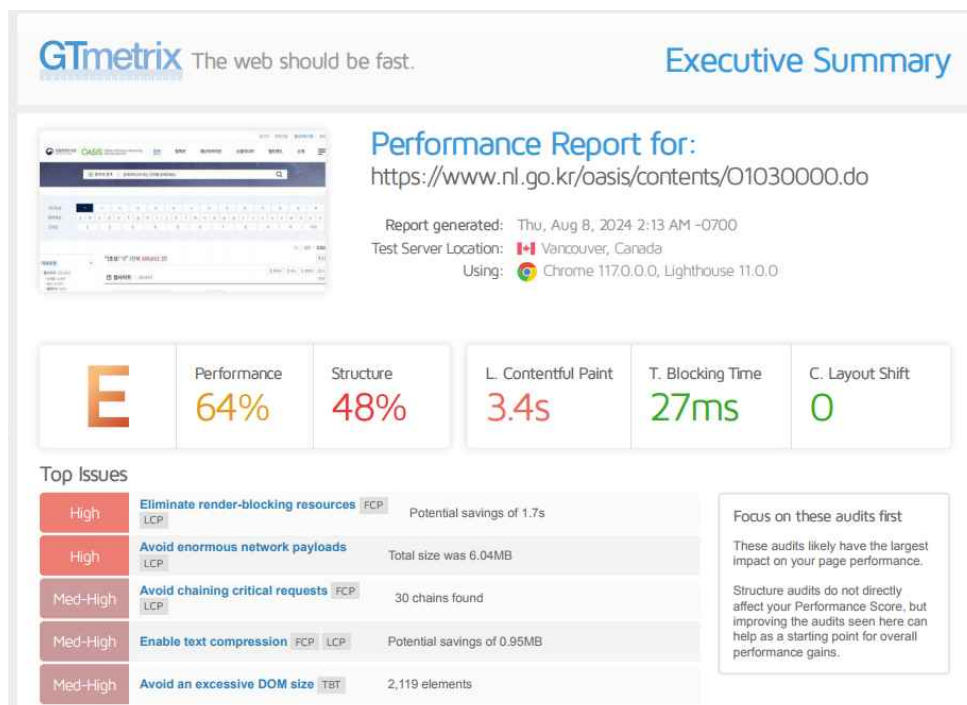
- 본 페이지는 <https://www.nl.go.kr/oasis/contents/O1020000.do>에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 E(50~59%)이고, 성과(performance)는 61%, 구조(structure)는 49%임. 성과(주황색)는 중간이나 구조(빨간색)는 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간이 3.9초 걸려 문제가 심각함
- 그 외, 심각한 네트워크 부하 회피, 초기 서버 응답시간 감소, 렌더 방해 요소 제거를 통해 웹페이지를 개선할 필요가 있음. 영상 크기를 적절히 사용하기는 중상 레벨로 개선 및 시급성을 보여줌



<그림 3-13> 오아시스 검색 메뉴 중 KDC 주제별 웹페이지 성능 분석 화면

4) 검색 > 초성 웹페이지 성능 분석

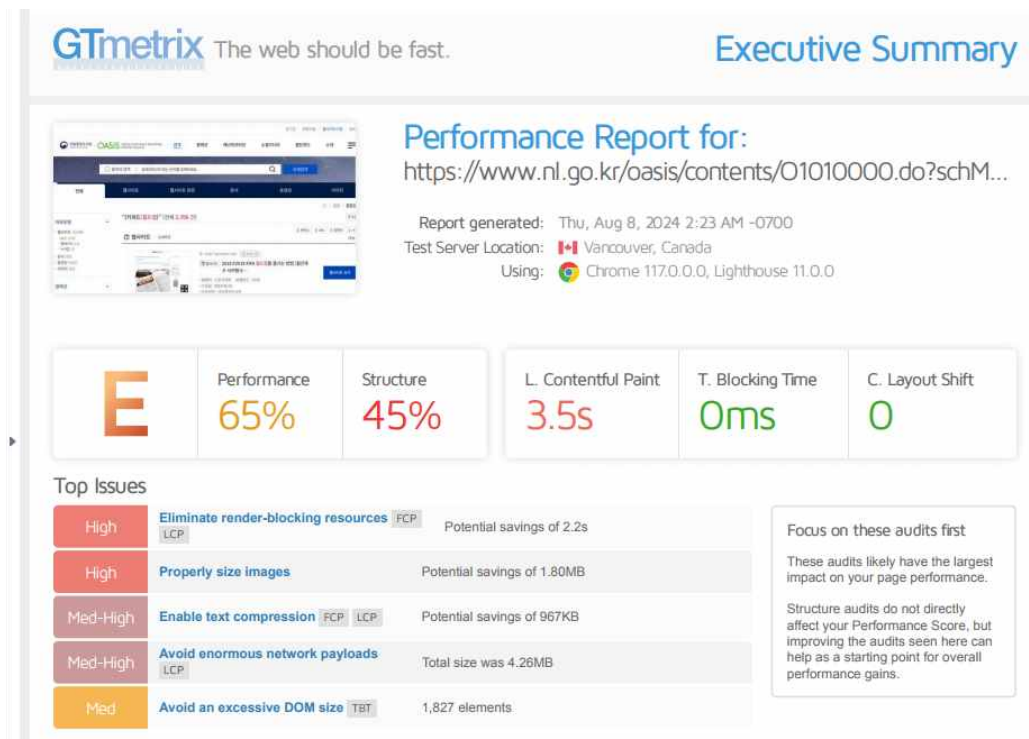
- 본 페이지는 검색 > 초성 웹페이지(<https://www.nl.go.kr/oasis/contents/O1030000.do>)에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 E(50~59%)이고, 성과(performance)는 64%, 구조(structure)는 48%임. 성과(주황색)는 중간이나 구조(빨간색)는 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간(LCP)이 3.4초 걸려 문제가 심각함
- 그 외, 렌더 방해 요소 제거, 심각한 네트워크 부하 회피를 우선적으로 해결해야 함



<그림 3-14> 오아시스 검색 > 초성 웹페이지 성능 분석 화면

5) 검색 > 통합검색 > 월드컵

- 본 페이지는 검색 > 통합검색 > 월드컵¹⁰⁾에 대해서 성능 분석을 수행한 결과로 통합검색 기능 및 결과 웹페이지 성능을 확인하기 위한 평가임
- 성과 등급은 E(50~59%)이고, 성과(performance)는 65%, 구조(structure)는 45%임. 성과(주황색)는 중간이나 구조(빨간색)는 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간(LCP)이 3.5초 걸려 문제가 심각함
- 그 외, 렌더 방해 요소 제거, 영상 크기를 적절히 사용하기가 우선적으로 개선해야 할 사항임

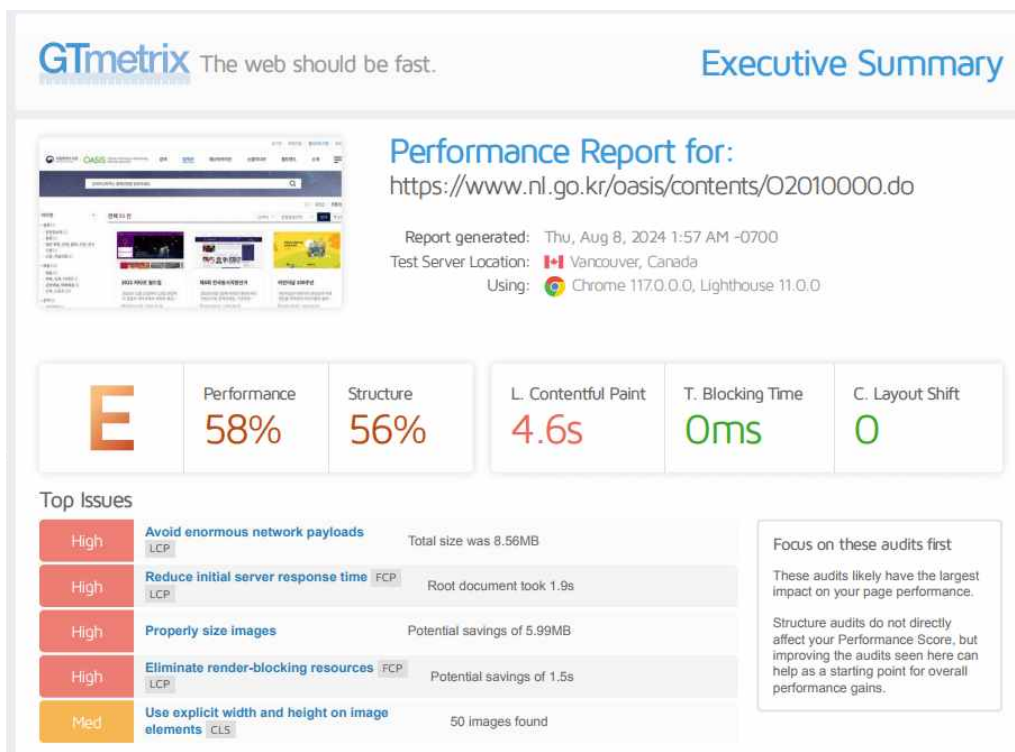


<그림 3-15> 오아시스 검색 > ‘월드컵’ 검색 성능 분석 화면

10) 오아시스 홈페이지에 접속 후 통합검색에서 키워드 “월드컵”으로 검색한 결과 웹페이지, <https://www.nl.go.kr/oasis/contents/O1010000.do?schM=list&schType=simple&historyYn=Y&schSiteNm=%EC%9B%94%EB%93%9C%EC%BB%B5>

6) 컬렉션 메뉴 웹페이지

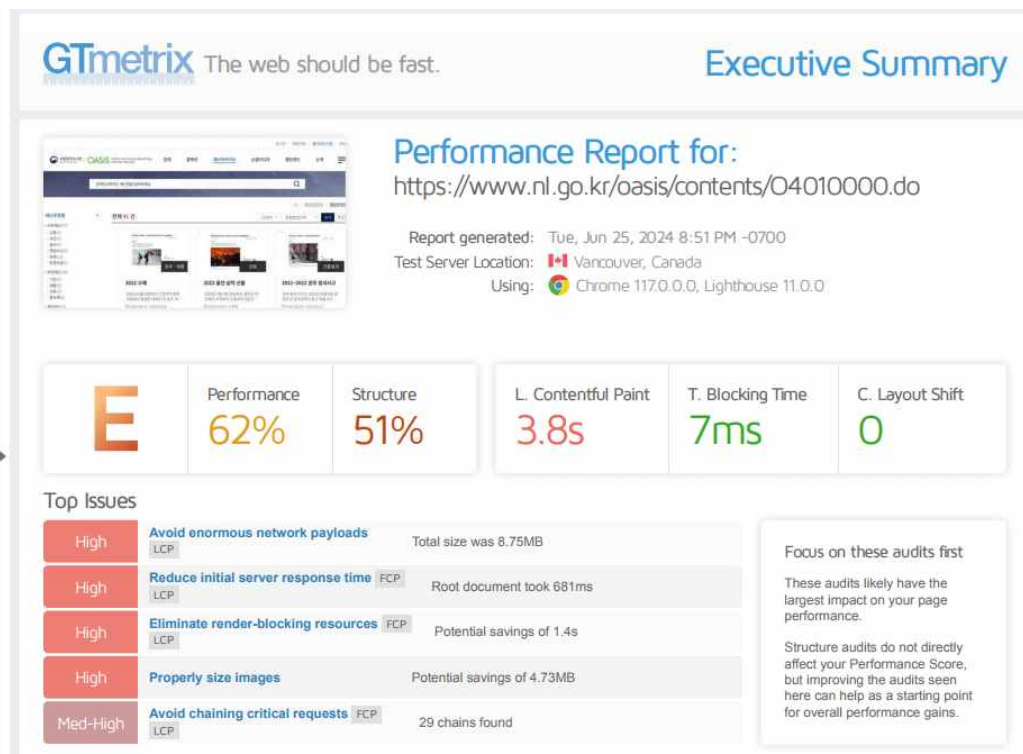
- 본 페이지는 컬렉션 메뉴 웹페이지(<https://www.nl.go.kr/oasis/contents/O2010000.do>)에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 E(50~59%)이고, 성과(performance)는 58%, 구조(structure)는 56%임. 성과(빨간색)와 구조(빨간색)는 모두 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간이 4.6초 걸려 문제가 심각함
- 그 외, 심각한 네트워크 부하 회피, 초기 서버 응답시간 감소, 영상 크기를 적절히 사용하기, 렌더 방해 요소 제거를 통해 웹페이지를 개선할 필요가 있음
- 성능이나 증상은 오아시스 메인 웹페이지 성능과 매우 유사함



<그림 3-16> 오아시스 컬렉션 메뉴 웹페이지 성능 분석 화면

7) 재난아카이브

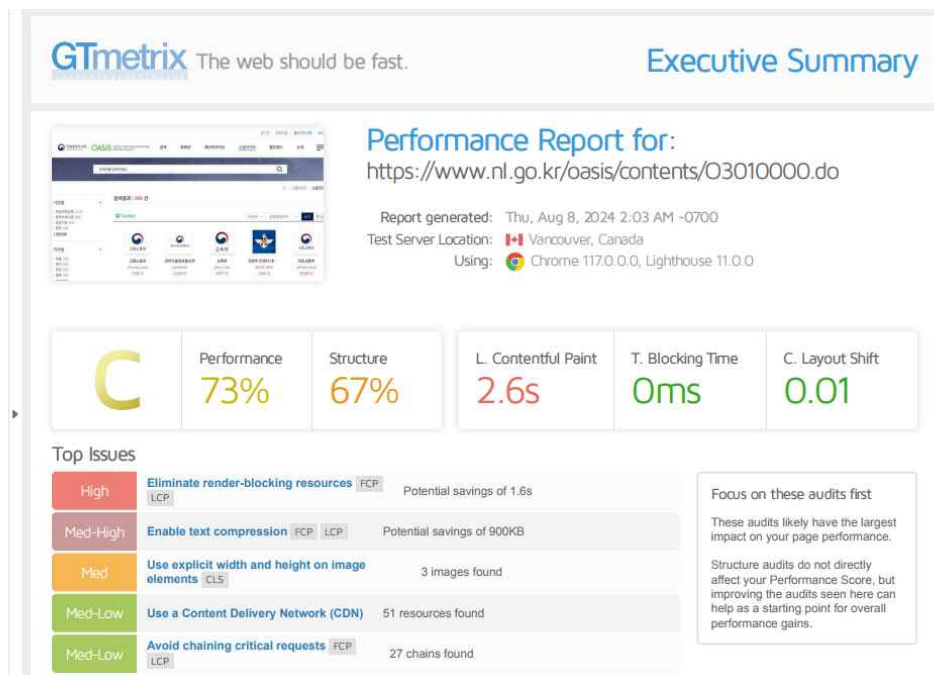
- 본 페이지는 재난아카이브 메뉴 웹페이지 (<https://www.nl.go.kr/oasis/contents/O4010000.do>)에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 E(50~59%)이고, 성과(performance)는 62%, 구조(structure)는 51%임. 성과(주황색)는 중간이나 구조(빨간색)는 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간(LCP)이 3.8초 걸려 문제가 심각함
- 그 외, 심각한 네트워크 부하 회피, 초기 서버 응답시간 감소, 영상 크기를 적절히 사용하기, 렌더 방해 요소 제거를 통해 웹페이지를 개선할 필요가 있음
- 오아시스 메인 웹페이지, 컬렉션 메뉴와 성능이 매우 유사함



<그림 3-17> 오아시스 재난 아카이브 메뉴 웹페이지 성능 분석 화면

8) 소셜미디어 메뉴 웹페이지

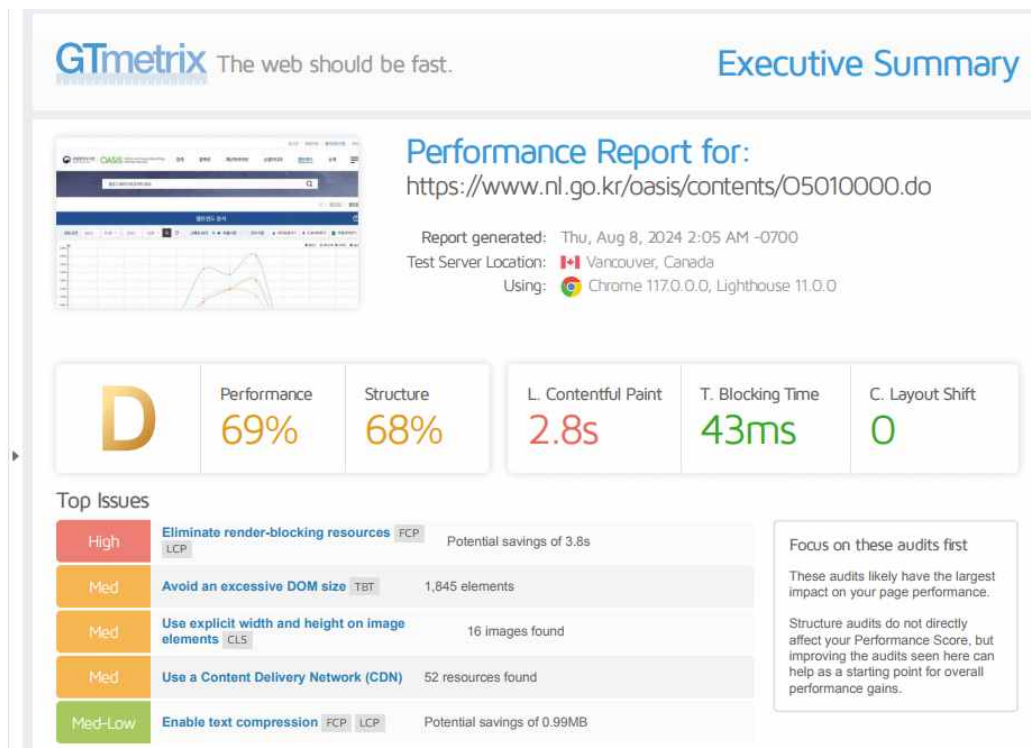
- 본 페이지는 소셜미디어 메뉴 웹페이지(<https://www.nl.go.kr/oasis/contents/O3010000.do>)에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 C(70~79%)이고, 성과(performance)는 73%, 구조(structure)는 67%임. 성과(노란색)는 양호이지만 구조(주황색)는 문제가 중급도를 시사함. 가장 큰 콘텐츠 로딩 시간(LCP)이 2.6초 걸려 문제가 심각함.
- 그 외, 렌더 방해 요소 제거를 통해 웹페이지를 우선적으로 개선할 필요가 있음
- 이미지 요소가 주로 정부 부처 로고로 구성되어 다른 웹페이지에 비해 비교적 작은 사이즈임을 확인할 수 있음



<그림 3-18> 소셜미디어 메뉴 웹페이지 성능 분석 화면

9) 웹트렌드 메뉴 웹페이지

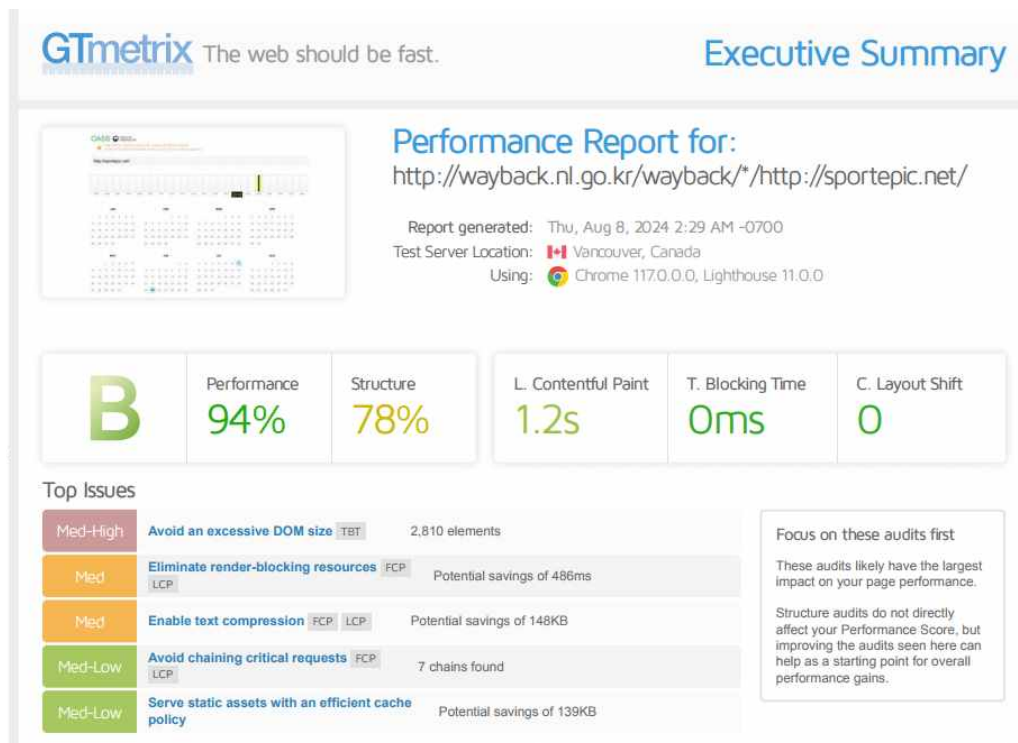
- 본 페이지는 웹트렌드 메뉴 웹페이지(<https://www.nl.go.kr/oasis/contents/O5010000.do>)에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 D(60~69%)이고, 성과(performance)는 69%, 구조(structure)는 68%임. 성과(주황색)는 중간이나 구조(주황색)는 문제가 보통 수준임을 시사함. 가장 큰 콘텐츠 로딩 시간(LCP)이 2.8초 걸려 문제가 심각함.
- 그 외, 렌더 방해 요소 제거를 통해 웹페이지를 우선적으로 개선할 필요가 있음. 다음으로 과도한 DOM 크기 줄이기, 이미지 요소의 명시적 폭, 너비 지정하기, 콘텐츠 배포망(Content Delivery Network) 사용하기를 수행하여 웹페이지를 개선할 수 있음



<그림 3-19> 웹트렌드 메뉴 웹페이지 성능 분석 화면

10) 컬렉션-웨이백 웹페이지

- 본 페이지는 컬렉션의 한 웹사이트에 대해서 웨이백 재생¹¹⁾의 캘린더 페이지에 대한 성능 분석을 수행한 결과임
- 성과 등급은 B(80~89%)이고, 성과(performance)는 94%, 구조(structure)는 48%임. 성과(초록색)는 매우 양호하고 구조(노란색)는 양호함을 시사함. 가장 큰 콘텐츠 로딩 시간이 1.2초 걸려 1초 내외로 콘텐츠를 로딩하는 것을 볼 때 양호함
- 즉, 본 페이지는 웨이백 재생의 캘린더 페이지로 용량이 큰 이미지 요소를 포함하지 않고 그래픽 요소를 포함하여 성능이 오아시스 누리집 웹페이지 중 가장 좋은 것(B등급)으로 나타남
- 그 외, 당장 시급한 문제는 없으나 과도한 DOM 크기 회피, 렌더 방해 요소 제거 및 텍스트 압축과 같은 방식으로 웹페이지를 여유 있게 개선할 필요가 있음

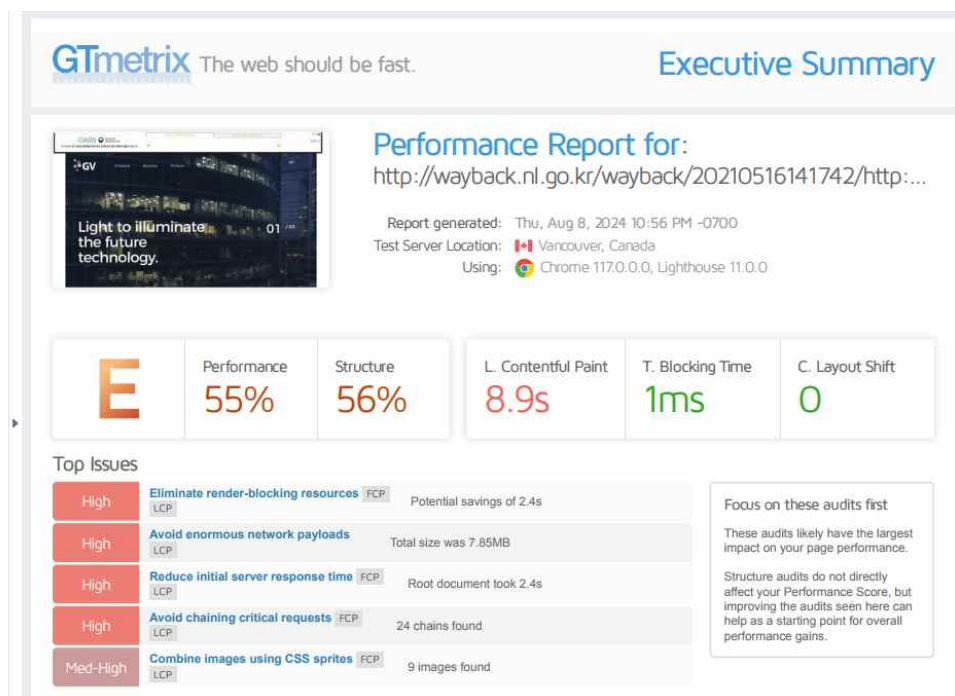


<그림 3-20> 컬렉션-웨이백 웹페이지 성능 분석 화면

11) http://wayback.nl.go.kr/wayback/*/http://sportepic.net/

11) 웨이백 재생-성공-금빛

- 본 페이지는 웨이백 재생을 통한 아카이브 데이터((주)금빛) 렌더링 페이지¹²⁾에 대한 성능 분석을 수행한 결과임
- 성과 등급은 E(50~59%)이고, 성과(performance)는 60%, 구조(structure)는 49%임. 성과(주황색)는 중간이나 구조(빨간색)는 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간이 8.9초 걸려 문제가 심각함
- 그 외, 렌더 방해 요소 제거, 심각한 네트워크 부하 회피, 초기 서버 응답시간 경감, 결정적 사슬 요구 회피를 통해 웹페이지를 우선적으로 개선할 필요가 있음

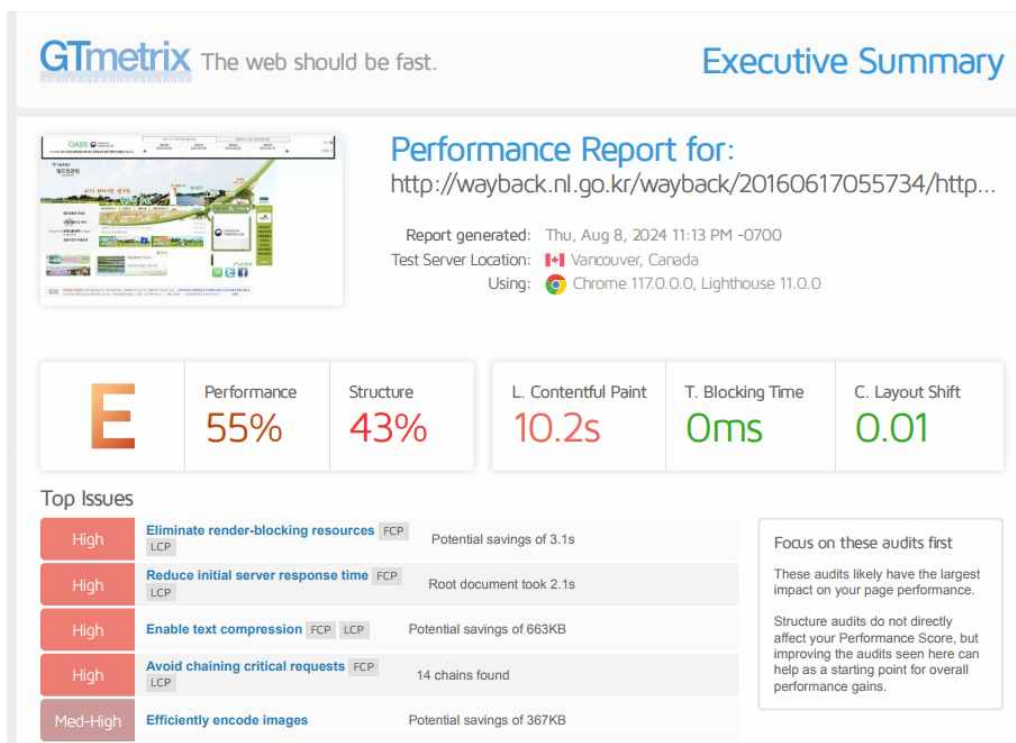


<그림 3-21> 웨이백 재생-성공-금빛 웹사이트 성능 분석 화면

12) 성공-금빛, <http://wayback.nl.go.kr/wayback/20210516141742/http://dblightec.co.kr/#>

12) 웨이백 재생-월드컵파크

- 본 페이지는 웨이백 재생을 통한 월드컵파크 웹사이트¹³⁾에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 E(50~59%)이고, 성과(performance)는 55%, 구조(structure)는 43%임. 성과(빨간색)와 구조(빨간색)는 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간이 8.9초 걸려 매우 오랜 시간이 걸리는 것으로 파악되어 문제가 심각함
- 그 외, 심각한 네트워크 부하 회피, 초기 서버 응답시간 감소, 텍스트 압축 시행, 결정적 요구 사슬 회피를 통해 웹페이지를 우선적으로 개선할 필요가 있음

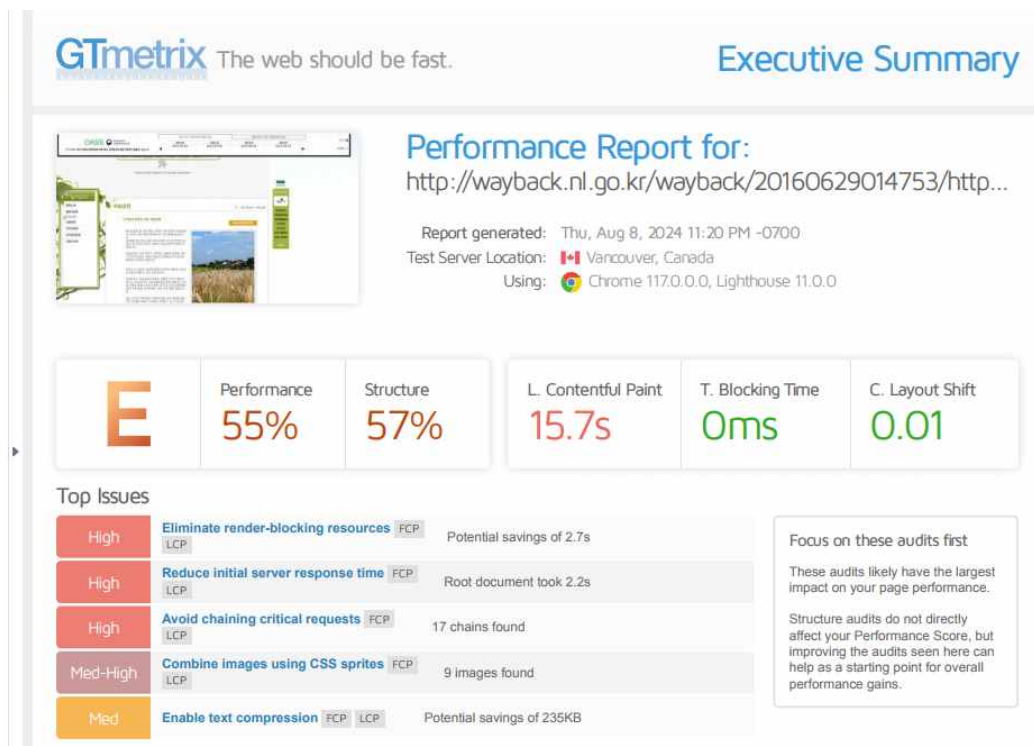


<그림 3-22> 웨이백 재생-월드컵파크 웹사이트 성능 분석 화면

13) 월드컵파크, <http://wayback.nl.go.kr/wayback/20160617055734/http://worldcuppark.seoul.go.kr/#>, 2024.8.8. 검색됨

13) 웨이백 재생_월드컵파크_20160629_서브 메뉴_하늘공원이용

- 본 페이지는 웨이백 재생을 통해 로드된 월드컵파크 웹사이트의 서브 메뉴 웹페이지¹⁴⁾에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 E(50~59%)이고, 성과(performance)는 55%, 구조(structure)는 57%임. 성과(빨간색)는 중간이나 구조(빨간색)는 문제가 심각함을 시사함. 가장 큰 콘텐츠 로딩 시간이 15.7초 걸려 문제가 심각함
- 그 외, 렌더 방해 요소 제거, 심각한 네트워크 부하 회피, 결정적 요구 사슬 회피를 통해 웹페이지를 개선할 필요가 있음
- 이 웹페이지는 아카이브 된 데이터를 로딩해 오는 과정으로 해당 웹페이지의 성능에 따라 차이가 많이 남

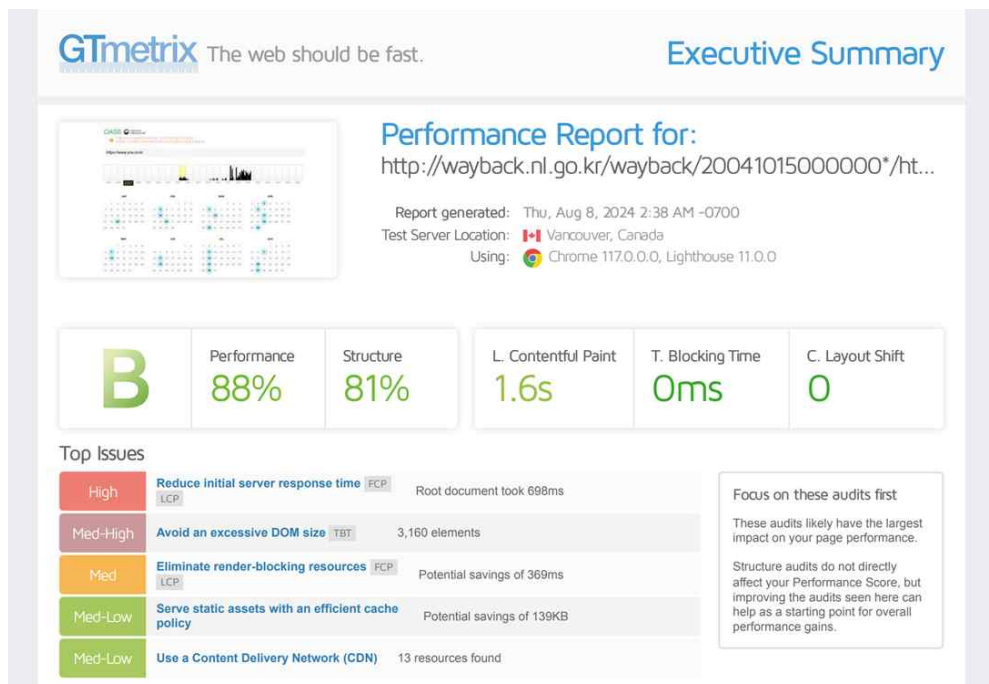


<그림 3-23> 웨이백 재생_월드컵파크_20160629_서브 메뉴_하늘공원이용 웹페이지 성능 분석 화면

14) 월드컵파크_20160629_서브 메뉴_하늘공원이용,
http://wayback.nl.go.kr/wayback/20160629014753/http://worldcuppark.seoul.go.kr/parkinfo/parkinfo3_1.html,
2024.8.8. 검색

14) 연합뉴스(yna) 현재 웹페이지 혹은 웨이백 재생 캘린더 화면

- 본 페이지는 웨이백 재생에서 연합뉴스(yna) 웹사이트의 웨이백 캘린더 화면에 대해서 성능 분석을 수행한 결과임
- 성과 등급은 B(80~89%)이고, 성과(performance)는 88%, 구조(structure)는 81%임. 성과(초록색)는 중간이나 구조(초록색)는 매우 양호함을 시사함. 가장 큰 콘텐츠 로딩 시간이 1.6초 걸려 로딩 성능이 양호함
- 그 외, 초기 서버 응답시간 감소를 통해 웹페이지를 우선적으로 개선할 필요가 있음
- 앞선 웨이백 재생의 캘린더 화면을 로딩할 때와 마찬가지로, 유사한 사유(이미지 요소 부재)로 가장 좋은 성능을 보여주고 있음



<그림 3-24> 연합뉴스(yna) 웨이백 재생 캘린더 화면 성능 분석 화면

3.5.3 오아시스 누리집 GTmetrix 성능 지표 요약 및 비교

○ GTmetrix를 이용한 오아시스 누리집 성능 지표 비교 평가

<표 3-7> GTmetrix를 이용한 오아시스 누리집 성능 지표 비교 평가

메뉴	Level	Performance	Structure	Total page size	Total Page Requests	Fully Loaded Time	FCP	SI	LCP	TTI	TBT	CLS
메인	E	60	49%	13.2MB	85	9.0s	3.0s	3.9s	4.5s	3.7s	15ms	0.01
검색	C	76%	70%	2.00MB	41	3.1s	2.2s	2.3s	2.3s	2.7s	0ms	0
검색-KDC 주제별	E	61%	49%	6.79MB	64	7.2s	3.6s	4.3s	3.9s	7.1s	29ms	0
검색-초성	E	64%	48%	6.04MB	64	7.0s	3.3s	3.4s	3.4s	6.8s	27ms	0
검색-통합월드컵	E	65%	45%	4.26MB	62	5.9s	2.7s	3.1s	3.5s	3.1s	0ms	0
컬렉션	E	58%	56%	8.56MB	66	8.5s	3.9s	4.7s	4.6s	4.4s	0ms	0
재난아카이브	E	62%	51%	8.75MB	67	6.4s	2.9s	3.9s	3.8s	6.2s	7ms	0
소셜미디어	C	73%	67%	2.41MB	58	3.4s	2.2s	2.4s	2.6s	2.7s	0ms	0.01
웹트렌드	D	69%	68%	2.25MB	64	3.7s	2.4s	2.8s	2.8s	3.6s	43ms	0
컬렉션-웨이백	B	94%	78%	241KB	18	3.7s	1.1s	1.2s	1.2s	1.1s	0ms	0
웨이백 재생-성공-금빛	E	55%	56%	7.83MB	115	25.2s	8.8s	25.5s	8.9s	8.9s	1ms	0
웨이백 재생-월드컵파크	E	55%	43%	2.31MB	273	51.2s	7.6s	19.0s	10.2s	9.8s	0ms	0.01
웨이백 재생_월드컵파크_20160629_서브 메뉴_하늘공원이용	E	55%	57%	933KB	142	31.4s	6.7s	14.3s	15.7s	25.3s	0ms	0.01
yna 현재 웹페이지	B	88%	81%	774KB	17	3.2s	1.5s	1.5s	1.6s	1.5s	0ms	0

○ 시사점

- 전체적으로 단순한 웹페이지의 경우 B-D까지 걸쳐 나오지만, 웹 아카이브 컬렉션들을 포함한 페이지의 경우 E와 같이 성능이 낮을 것을 볼 수 있음
- 그리고 성능이 B인 페이지를 제외한 대부분의 웹페이지(C~E) 경우 FCP와 SI, LCF가 매우 느린 것을 확인할 수 있음
- 웹페이지의 이미지 크기를 조정함으로써 전체적인 로딩 시간을 줄일 수 있음을 시사하고 있음
- 하지만, 이것은 단순히 누리집을 수정하는 것으로 해결될 수 있는 문제와 수집한 웹 아카이브 데이터를 수정해야 하는 문제를 구분하여 접근해야 오아시스 누리집 성능을 향상시킬 수 있음

IV

시스템 개선 방안 및 프로토타입

4.

오아시스 시스템 개선 방안 및 프로토타입

- 본 장의 앞 장에서 시스템 현황 및 진단을 통해 파악된 시사점들을 우리나라 웹 아카이빙 시스템(오아시스) 개선에 적용하기 위하여 시사점을 유형화하고 개선 방안을 도출함
- 유형화된 개선 방안들 중에 현 오아시스 시스템 개선 방안의 효과성을 파악하기 위하여 프로토타입을 정의하고 시범 적용할 항목을 선정함을 목적으로 함

4.1 오아시스 시스템 시사점 유형화 및 개선 방안

4.1.1 오아시스 누리집 시사점 유형화

1) 시사점 유형화

- 앞 장에서 현황 분석을 통해 제시된 오아시스 시스템, 누리집의 개별 시사점들에 대하여 시사점이 시스템의 어느 요소(시스템(하드웨어, 소프트웨어 포함), 데이터, 누리집(홈페이지), 정책 및 제도)에 속하는지 분류하고 각 범주의 오류를 유형화하였음. 이 유형화 분류 체계는 상향식(bottom-up)으로 도출되었음. 따라서, 새로운 시사점이 도출되면, 동일한 유형화 프로세스를 거쳐 오류 유형 분류 체계는 수정 및 확장될 수 있음
- 오류 및 시사점 유형화:
 - 1. 시스템
 - ▶ 1.1 웨이백 재생 프로그램
 - ▶ 1.2 문서(동영상) 전용 뷰어(재생) 프로그램
 - ▶ 1.3 불편한 검색 기능
 - 2. 데이터
 - ▶ 2.1 누리집 데이터
 - ▶ 2.2 웹 데이터 품질(이미지 깨짐)
 - ▶ 2.3 웹 데이터 품질 유형(링크 깨짐)
 - ▶ 2.4 자료 분류 체계 모호성(유형, 주제)

- ▶ 2.5 오류를 포함하는 메타데이터
- 3. 누리집
 - ▶ 3.1 누리집 디자인
 - ▶ 3.2 누리집 실행 순서(전용뷰어 설치 전 이용 범위 안내)
 - ▶ 3.3 누리집 장애인 지원 기능(음성 안내, 돋보기 기능)
 - ▶ 3.4 개인 맞춤형 서비스 부재
 - ▶ 3.5 누리집 도움말 미비
 - ▶ 3.6 모바일 비친화적 누리집
 - ▶ 3.7 불편하고 모호한 검색 인터페이스
 - ▶ 3.8 부적절한 미리보기 이미지 사이즈
 - ▶ 3.9 누리집 기능 오동작
 - ▶ 3.10 누리집 웹페이지 이중 디스플레이(재로딩)
 - ▶ 3.11 이용자와 상호 작용이 부족한 메뉴
- 4. 정책 및 제도
 - ▶ 4.1 제한된 자료 이용 범위
 - ▶ 4.2 수집 웹사이트 주제 범위의 편중화 -> 균일화 및 다양화
 - ▶ 4.3 트윗에 제한된 소셜미디어 수집 범위

2) 도출된 시사점 유형화 및 개선 방안 제안

○ 검색(Search) 시사점 유형화 및 개선 방안

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
SRC -1	웨이백 재생 무한 버퍼링	1.1 웨이백 재생 프로그램	정기적으로 정보자원의 해당 웹사이트로의 이동이 원활한지 확인하고, 보완해야 함
SRC -2	디폴트 이미지 상이함	2.1 누리집 데이터	자동 음성 생성 서비스 제공
SRC -3	이미지 정보들이 미리보기가 제공되지 않음. 들어가서 확인해야 하는 번거로움 존재	3.1 누리집 디자인	이미지나 동영상 등의 미리보기 이미지가 필수적인 정보자원들에 한해서는 미리보기 썸네일 이미지를 제공하여 원본

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
			에 대한 대강의 정보를 확인할 수 있는 방안에 대해 고려할 수 있음
SRC-4	사진이나 동영상, 문서 확인을 위해서는 별도의 뷰어 프로그램 설치가 요구되어 번거롭고, 해당 정보 획득을 위해 긴 시간이 소요됨	1.2 문서(동영상) 전용 뷰어(재생)	전용뷰어 설치가 아닌 브라우저 플러그인을 설치 활용하는 방안
SRC-5	프로그램 설치 전 이용 범위 안내 필요	3.2 누리집 실행 순서(전용뷰어 설치 전 이용 범위 안내)	이를 사전 예방하기 위해, 프로그램 설치 허용을 누르긴 전에 이용 범위를 안내해 주면 좋겠음
SRC-6	동영상 실행에 대해 ‘웹사이트 보기’ 버튼 제공	3.1 누리집 디자인	‘웹사이트 보기’ 버튼보다는 ‘동영상 보기’ 혹은 전체적으로 ‘보기/이동’ 버튼
SRC-7	국립중앙도서관에서만 자료 이용가능한 것이 불편함	4.1 제한된 자료 이용 범위	현재 디지털 자료 납본제도나 저작권법 제도하에서는 국립중앙도서관 내에서만 자료이용이 가능하나 자유이용 허락 라이선스(Creative Commons License) 유도를 통해 이용 범위를 확대하는 방안 시도해 볼 수 있음
SRC-8	과거 자료의 변경 기록을 확인하고 과거의 웹사이트에 방문하면 이미지가 깨져 나옴	2.2 웹 데이터 품질(이미지 깨짐)	링크가 있는 곳과 그렇지 않은 곳을 한눈에 구분하는 화면
SRC-9	아이콘 확인 어려움. 아이콘에 마우스 커서를 올려야만 어떤 것이 문서이고, 웹사이트인지 알려주는 것이 조금 불편함	3.1 누리집 디자인	조그맣게 글자를 써넣고 링크를 다는 것이 한눈에 확인하기도 쉬움. 아이콘의 크기를 키우고, 직관적인 아이콘 모형을 사용하여 이용자가 빠르게 아이콘을 인식하고, 아이콘을 눌러 자료를 탐색할 수 있도록 지원해야 함
SRC-10	12개, 24개씩 같이 나눠서 보는 것이 아닌 전체 보기와 같은 시스템이 있으면 전체를 조망하면서 선택할 수 있을 것임	3.1 누리집 디자인	윈도우즈 탐색기에서 자세히 보기와 같은 제목 위주의 리스트 자료 보기를 제시할 수 있을 것임
SRC-11	웹사이트 URL 링크 표시가 안 되어 있음	3.1 누리집 디자인	웹사이트 URL은 링크 표시(파란색 글씨에 밑줄)
SRC	검색 기능의 불편함	1.3 불편한 모호한	검색 기능 고도화

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
-12		검색 기능	
SRC -13	복지 서비스 부재(장애인, 노인 등)	3.3 누리집 장애인 지원 기능(음성 안내, 돋보기 기능) 부재	시각 장애인들을 위한 음성 지 원 프로그램이나 노인을 위한 돋보기 기능 등 다양한 복지 서비스가 필요함
SRC -14	오아시스 이용 (빅)데이터 활용 이용자 맞춤 개인화 서비스의 필요성	3.4 개인 맞춤형 서비스 부재	안내 서비스, 추천 서비스 등 개인화 분석 서비스를 제공하 여 이용자의 OASIS에 대한 만 족도를 향상할 필요가 있음
SRC -15	검색 항목에 <초성으로 검색하기>를 이 용했을 때 선택한 초성이 제목을 말하 는 것인지, 발행처인지, 사이트 이름이 정해지지 않아 불편함이 있었음	3.5 누리집 도움말 미비	상세한 도움말 제공
SRC -16	모바일 사용의 불편함	3.6 모바일 비친화적 누리집	모바일 지향의 누리집 개선 과 제 수행
SRC -17	일부 분류 주제에 있어 소수 수집자료	4.2 수집 주제범위의 편중화	다양한 주제 분류가 수집될 수 있도록 지속적이고 포괄적 수 집 도메인 확대
SRC -18	앞선 정보테크. 한국의 링크가 작동하 지 않음	2.3 웹 데이터 품질 유형(링크 깨짐)	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도메인을 가려내는 QA 프로세 스 보완
SRC -19	웹사이트 URL 이동 불가	2.3 웹 데이터 품질 유형(링크 깨짐)	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도메인을 가려내는 QA 프로세 스 보완
SRC -20	상세검색 화면에서 소장자료와 웹사이 트 본문 탭의 차이가 명확하지 않음	3.7 불편하고 모호한 검색 인터페이스	상세검색 메뉴 재설계함
SRC -21	웹사이트 본문을 클릭했을 때, 검색시 간이 소장자료 검색에 비해 오래 걸림 (17초).	1.3 불편하고 모호한 검색 기능	색인 및 검색을 다시 살펴볼 필요가 있음
SRC -22	자료유형에서 웹사이트, 누리집, 웹페이 지, 뉴스, 웹사이트의 구분 및 분류가 직관적으로 이해되지 않음	3.5 누리집 도움말 미비	2년 엠바고에 대한 설명을 이 용자가 쉽게 알 수 있는 방법 으로 제시할 필요가 있음
SRC -23	‘웹사이트 본문’을 선택할 경우, 자료유 형이 full-text로 나타나지만, 전체를 선택하면 full-text 자료유형은 보여지 지 않음	2.4 자료 분류 체계 모호성(유형, 주제)	누리집의 유형 분류 체계 개선

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
SRC-24	SNDSOFT의 경우, 검색에서는 검색 결과가 있으나 컬렉션, 재난아카이브, SNS에서는 검색 결과가 없으나 웹트렌드에서는 검색이 됨	1.3 불편한 검색 기능	웹트렌드 검색 대상과 다른 컬렉션들(컬렉션, 재난, SNS)에 대한 검색 대상이 다를 수 있음을 설명할 필요가 있음
SRC-25	검색에서 마우스 조작 시 검색 아이콘의 박스는 작고, 분리되어 있지 않아 색이 다르고 큼지막한 상세검색을 클릭하게 되어 불편함	3.1 누리집 디자인	누리집 해당 페이지 디자인 분석 후 개선
SRC-26	상세검색에서 발행연도를 선택할 때 좌우로 한 달씩만 이동할 수 있고, 연 단위로 이동하거나 블록형으로 선택할 수 없어 불편함	3.1 누리집 디자인	누리집 해당 페이지 기능 분석 후 개선

○ 컬렉션(collection) 시사점 유형화 및 개선 방안

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
CLT-1	보기 유형에서 12개 보기는 가능하나 24개 보기 이상은 동작하지 않음	3.1 누리집 디자인 3.6 모바일 비친화적 누리집	버퍼링 시간을 단축하기 위해 여러 방안으로 노력해야 함.
CLT-2	컬렉션 부분에서 3.1운동 및 임시정부 100주년 기념 자료의 출처 링크가 깨졌음	2.3 웹 데이터 품질 유형(링크 깨짐)	https://www.together100.go.kr 웹사이트에 대한 변경 기록 URL로 변경하는 방안 전반적으로 링크 깨짐 현상에 대해 자동으로 추출하는 도구 개발하여 주기적으로 점검하는 방안
CLT-3	미리보기 이미지가 제공되는 경우, 이미지가 모두 보이는 데 2-4초 가량 소요됨	3.8 부적절한 미리보기 이미지 사이즈	미리보기를 위한 썸네일 이미지 생성 및 제시를 통한 웹페이지 로딩속도를 높이는 방안
CLT-4	한국영화 100년 기념사업 홈페이지 링크 깨짐	2.3 웹 데이터 품질 유형(링크 깨짐)	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도메인을 가려내는 QA 프로세스 보완 필요
CLT-5	‘컬렉션’ 장서(collection)에서 검색할 때, 검색 키워드 입력 후 엔터를 치면 실행되나 검색 버튼(돋보기 모양 이미지) 클릭은 동작하지 않음	3.9 누리집 스크립트 기능 오작동	누리집 해당 페이지 기능 분석 후 개선

○ 재난 아카이브(Disaster Archive) 시사점 유형화 및 개선 방안

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
DS T-1	자료유형에 따른 적절한 유형의 자료 제시 오류(이미지 선택 시 동영상 제 시, 동영상 선택 시 이미지 제시)	3.9 누리집 기능 오동작	누리집 해당 페이지 기능 분석 후 개선 버퍼링 시간을 단축하기 위해 여러 방안으로 노력해야 함
DS T-2	해외재난 분류 체계가 상세하지 않음	2.4 자료 분류 체계 모호성(유형, 주제)	해외재난도 우리나라의 재난을 분류해 둔 것과 마찬가지로 분 류를 해두면 가독성이 더욱 높 아질 것임
DS T-3	컬렉션 가장 최신 기사가 2022년 기사 라 현재 기사가 궁금하면 따로 검색해 보는 방법밖에 없어서 불편했음	3.5 누리집 도움말 미비	2년 엠바고에 대한 설명을 이용 자가 쉽게 알 수 있는 방법으로 제시할 필요가 있음
DS T-4	재난아카이브에서 코로나바이러스에 대한 출처인 질병관리본부 (http://ncov.mohw.go.kr/)링크를 들 어가자 접근 금지된 자료였음	2.3 웹 데이터 품질 유형(링크 깨짐)	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도 메인을 가려내는 QA 프로세스 보완
DS T-5	재난아카이브에서 자료를 선택한 후 첫 번째 로딩 후에 재로딩하는 현상	2.10 누리집 웹페이지 이중 디스플레이(재로딩)	처음 로딩 후 스타일 적용을 위 해 다시 메타DB 재로딩하는 것 으로 예상됨. 스타일 적용시 메 타DB 재로딩이 있는지 확인할 필요가 있음
DS T-6	표기 오류: ‘재난 아카이브’에서 ‘재난 유형별’의 ‘해외’에는 세 개의 자료가 표기되어 있지만 ‘지역별’의 ‘해외’에는 자료가 전혀 표기되어 있지 않음. 각 키워드에 적합한 문헌임에도 동시에 로 표기되지 않는 경우, 수정이 필요해 보임	2.4 자료 분류 체계 모호성(유형, 주제) 3.9 누리집 기능 오동작	자료분류 체계 재설정 후 데이 터 품질 확보(QA) 방안 수립 누리집 해당페이지 기능 분석 후 개선

○ 소셜미디어(SNS) 시사점 유형화 및 개선 방안

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
SN S-1	하이퍼링크화 되지 않아 불편함	3.1 누리집 디자인	누리집 해당 페이지 기능 분석 후 개선
SN S-2	제일 처음으로 올라온 트윗은 2010년 인데 가입일은 2018년도로 뜨는 등 가입일과 트윗이 맞지 않는 경우가 있	2.5 오류를 포함하는 메타데이터 3.9 누리집 기능 오동작	해당 메타데이터 확인 후 메타 데이터 품질 확보 방안 수립 누리집 해당 페이지 확인 후 수

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
	음 실제로 홈페이지에 들어가 보면 가 입일은 2010년도로 뜸. 이런 경우가 많음		정 필요
SN S-3	제한된 소셜미디어: 소셜미디어 창에 들어갔을 때 여러 기관의 소셜미디어 가 정리되어 있음. 하지만 현재 트위 터 계정들만 정리되어 있고 여러 기관 이 유튜브, 인스타 등의 다양한 소셜 미디어를 가지고 있는 만큼 다른 소셜 미디어 계정들도 정리되어 있어야 한 다고 생각함	4.3 트윗에 제한된 소셜미디어 수집 범위	이미 제시된 트윗자료는 유지하 고 추가로 소셜미디어 자료를 수집할 수 있을지 확대 전략 논 의 필요
SN S-4	각 기관 문양 이미지 깨짐: 고용노동부 의 공식 트위터를 소개하는 기관 문양 의 사진은 선명한 반면, 수집된 트윗에 서의 고용노동부의 문양 사진은 선명 하지 않고 이미지가 깨져있음. 고용노 동부뿐만 아니라 다른 기관 모두 수집 된 트윗에서의 기관 문양 사진은 깨져 있음	3.8 부적절한 미리보기 이미지 사이즈	확인 결과 이미지 사이즈가 현 저히 작음. 트윗 문양에 맞게 사 이즈를 조금 크게 할 필요가 있 음
SN S-5	환경부 트위터에 자세히 보기 링크를 들어가도 페이지를 볼 수 없음	2.3 웹 데이터 품질 유형(링크 깨짐)	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도 메인을 가려내는 QA 프로세스 보완
SN S-6	소셜미디어 설명 부족	3.5 누리집 도움말 미비	상세한 도움말 마련 및 제공

○ 웹트렌드(Trend) 시사점 유형화 및 개선 방안

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
TR N-1	안내 메시지가 화면에 가려지는 불편 함	3.1 누리집 디자인	이용자가 정보를 확인하는 데 불편함이 없는 크기의 여백을 제공함
TR N-2	그래프 보기에서 중복 라디오버튼	3.1 누리집 디자인	중복 제거
TR N-3	웹트렌드 분석 도움말이 기능 위주의 설명임. 의미와 해석 방법에 대해서 좀더 상세한 설명이 있으면 좋겠음. 예를 들어, 레이블에 한국, 미국 등과	3.5 누리집 도움말 미비	자세한 사항은 웹트렌드 활용 예제를 참고하라는 안내메시지 포함

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
	같은 나라명이 제시되어 있는데, 이 레이블의 의미가 무엇이고 그래프는 어떻게 해석할 수 있는지 제시되면 이 사용자가 웹트렌드 기능을 이해하는 데 도움이 될 것임		
TR N-4	키워드 비중 도움말이 너무 간단함	3.5 누리집 도움말 미비	키워드 비중이 키워드(용어)빈도 (term frequency)의 비중을 표현한 것인지 키워드가 출현한 문서(document frequency)의 비중을 표현한 것인지 설명을 추가하면 도움이 될 것임
TR N-5	웹트렌드의 대부분 링크를 클릭했을 때 사이트에 연결할 수 없음. 메시지가 결과로 나옴	2.3 웹 데이터 품질 유형(링크 깨짐)	웹트렌드의 링크를 전반적으로 점검할 필요가 있음. 자동화 툴을 개발 적용할 것을 추천함
TR N-6	"웹트렌드"라는 하나의 메뉴에서 갈라진 하위 메뉴임에도 불구하고 웹트렌드, 웹트렌드 활용예제, 태그클라우드 세 페이지의 UI 배치와 구성이 통일되어 있지 않아 이용이 어려움	3.1 누리집 디자인	웹트렌드와 태그클라우드에도 왼쪽의 바로가기 메뉴 제공

○ 소개(About) 시사점 유형화 및 개선 방안

순번 코드	도출된 시사점	시사점 유형	제안된 개선 방안
AB T-1	자주 묻는 질문은 있었지만, 직접적인 질의응답 창구가 없음. 사이트를 사용하다가 궁금한 점이나 모르는 점이 생겼을 경우 바로 질문을 주고 받을 수 있는 질의응답 창이 필요함	3.11 이용자와 상호 작용이 부족한 메뉴	상호 작용을 높일 수 있는 질문 게시판 기능 제공 인공지능 챗봇과 같은 신기술 도입 전략 방안을 고려해 볼 수 있음

○ 미국 의회도서관 우수 사례 시사점 및 개선 방안

순번코드	도출된 시사점	제안된 개선 방안
LC-1	미국 의회도서관(LC) 우수 사례 조사 시사점→ 보기 유형으로 슬라이드쇼 제공	모바일에 초점 맞춘 슬라이드쇼 (slideshow) 인터페이스 기능 적용
LC-2	미국 의회도서관(LC) 우수 사례 조사 시사점→ 자동 음성 생성 서비스 제공	자동 음성 생성 서비스 제공
LC-3	미국 의회도서관(LC) 우수 사례 조사 시사점→ share 기능 제공	Share 기능 추가
LC-4	미국 의회도서관(LC) 우수 사례 조사 시사점→ 저작권 및 접근권한 도움말 제공	도움말 제공
LC-5	미국 의회도서관(LC) 우수 사례 조사 시사점→ 인용 기능 제공	Cite this term 기능
LC-6	미국 의회도서관(LC) 우수 사례 조사 시사점→ 추천기능 제공	추천 사이트 기능
LC-7	미국 의회도서관(LC) 우수 사례 조사 시사점→ 메타데이터(MODS) 제공	메타데이터(MODS) 목록을 다른 웹 서버에 제공하는 API 기능

4.2 오아시스 개선 프로토타입 시범 적용

4.2.1 오아시스 누리집 프로토타입 시범 적용 논의 결과

○ 시스템 관련 개선점 리스트

시사점 유형		시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
1. 시스템	1.1 웨이백 재생 프로그램	검색-1	웨이백 재생 무한 버퍼링	정기적으로 정보 자원의 해당 웹사이트로의 이동이 원활한지 확인하고, 보완해야 함	오아시스 시스템의 중장기 계획에 잘 반영될 수 있도록 해외 우수 사례 분석 및 시사점 도출 요망
	1.2 문서 (동영상) 전용 뷰어(재생)	검색-4	사진이나 동영상, 문서 확인을 위해서는 별도의 뷰어 프로그램 설치가 요구되어 번거롭고, 해당 정보 획득을 위해 긴 시간이 소요됨	전용뷰어 설치가 아닌 브라우저 플러그인을 설치 활용하는 방안	차년도 개선
	1.3 불편한 검색 기능	검색-24	SNDSOFT의 경우, 검색에서는 검색 결과가 있으나 컬렉션,	웹트렌드 검색 대상과 다른 컬렉션들(컬렉션, 재난, SNS)에 대한 검색	차년도 개선

시사점 유형		시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
			재난아카이브, SNS에서는 검색결과가 없으나 웹트렌드에서는 검색이 됨	대상이 다름을 설명할 필요가 있음	
	1.3 불편한 모호한 검색기능	검색-12	검색 기능의 불편함	검색 기능 고도화	차년도 개선
	1.3 불편한 모호한 검색기능	검색-21	웹사이트 본문을 클릭했을 때, 검색 시간이 소장자료 검색에 비해 오래 걸림(17초)	색인 및 검색을 다시 살펴볼 필요가 있음	차년도 개선
	1.4 다양한 웹 아카이브 분석 기능 부족	우수 사례-1	선진 기관 운영 사례 분석 → 웹페이지 회복력 분석 기능(trend machine) 제공	장기적 시스템 개선 방안에 웹 아카이브 회복력 분석 기능 포함	적용 가능 여부 검토 필요
	1.5 이용자 인지력 향상 기능 제공 부족	우수 사례-2	선진 기관 운영 사례 분석 → 뉴스 웹 아카이브 인공지능 활용 요약 기능(newsum) 제공	장기적 시스템 개선 방안에 웹 아카이브 인공지능 활용 (요약 및 추천) 기능 포함	생성형 인공지능의 유료 API를 사용하는 것은 내부 논의가 필요함

○ 웹 데이터 관련 개선점 리스트

시사점 유형		시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
2. 웹 데이터	2.1 누리집 데이터	검색-2	썸네일 디폴트 이미지 상이함	누리집 디폴트 썸네일 이미지 확인 후 통일해야 함	일부 썸네일 디폴트 이미지는 의도적으로 달리하였으므로 문제없음
	2.10 누리집 웹페이지 이중 디스플레이 (재로딩)	재난아카이브-5	재난아카이브에서 자료를 선택한 후 첫 번째 로딩 후에 재로딩하는 현상	처음 로딩 후 스타일 적용을 위해 다시 메타DB 재로딩하는 것으로 예상됨. 스타일 적용시 메타DB 재로딩이 있는지 확인할 필요가 있음	문제없다면 문제점 항목에서 제외해도 무방
	2.2 웹 데이터 품질(이미지 깨짐)	검색-8	과거 자료의 변경 기록을 확인하고 과거의 웹 사이트에 방문하면 이미지가 깨져 나옴	링크가 있는 곳과 그렇지 않은 곳을 한눈에 구분하는 화면	차년도 개선
	2.3 웹 데이터 품질 유형	검색-18	앞선 정보테크. 한국의 링크가 작동하지 않음	링크 깨짐(broken link)이나 더 이상 존재하지 않는	차년도 개선

시사점 유형	시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
(링크 깨짐)			비존재 도메인을 가려내는 QA 프로세스 보완	
2.3 웹 데이터 품질 유형 (링크 깨짐)	검색-19	웹사이트 URL 이동 불가	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도메인을 가려내는 QA 프로세스 보완	차년도 개선
2.3 웹 데이터 품질 유형 (링크 깨짐)	컬렉션-2	컬렉션 부분에서 3.1운동 및 임시정부 100주년 기념 자료의 출처 링크가 깨졌음	https://www.together100.go.kr 웹사이트에 대한 변경기록 URL로 변경하는 방안	즉시 개선
2.3 웹 데이터 품질 유형 (링크 깨짐)	컬렉션-2	컬렉션 부분에서 3.1운동 및 임시정부 100주년 기념 자료의 출처 링크가 깨졌음	전반적으로 링크 깨짐 현상에 대해 자동으로 추출하는 도구 개발하여 주기적으로 점검하는 방안	차년도 개선
2.3 웹 데이터 품질 유형 (링크 깨짐)	컬렉션-4	한국영화 100년 기념사업 홈페이지 링크 깨짐	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도메인을 가려내는 QA 프로세스 보완	차년도 개선
2.3 웹 데이터 품질 유형 (링크 깨짐)	재난아카이브-4	재난아카이브에서 코로나바이러스에 대한 출처인 질병관리본부(http://ncov.mohw.go.kr/) 링크를 들어가자 접근 금지된 자료였음	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도메인을 가려내는 QA 프로세스 보완	차년도 개선
2.3 웹 데이터 품질 유형 (링크 깨짐)	소셜미디어-5	환경부 트위터에 자세히 보기 링크를 들어가도 페이지를 볼 수 없다.	링크 깨짐(broken link)이나 더 이상 존재하지 않는 비존재 도메인을 가려내는 QA 프로세스 보완	즉시 개선
2.3 웹 데이터 품질 유형 (링크 깨짐)	웹트렌드-5	웹트렌드의 대부분 링크를 클릭했을 때 사이트에 연결할 수 없음 메시지가 결과로 나옴	웹트렌드의 링크를 전반적으로 점검할 필요가 있음. 자동화 툴을 개발 적용할 것을 추천함	차년도 개선
2.4 자료 분류 체계 모호성(유형, 주제)	검색-23	‘웹사이트 본문’을 선택할 경우, 자료유형이 full-text로 나타나지만, 전체를 선택하면 full-text 자료유형은 보여지지 않음	누리집의 유형 분류 체계 개선	차년도 개선
2.4 자료 분류 체계 모호성(유형, 주제)	재난아카이브-2	해외재난 분류 체계가 상세하지 않음	해외재난도 우리나라의 재난을 분류해 둔 것과 마찬가지로 분류를 해두면 가독성이 더욱 높아질 것임	차년도 개선
2.4 자료 분류 체계 모호성(유형, 주제)	재난아카이브-6	표기 오류: ‘재난 아카이브’에서 ‘재난유형별’의 ‘해외’에는 세 개의 자료가 표기되어	자료 분류 체계 재설정 후 데이터 품질 확보(QA) 방안 수립	차년도 개선

시사점 유형	시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
		있지만 ‘지역별’의 ‘해외’에는 자료가 전혀 표기되어 있지 않음. 각 키워드에 적합한 문헌임에도 동시에 표기되지 않는 경우, 수정이 필요해 보임		
2.5 오류를 포함하는 메타데이터	소셜미디어-2	제일 처음으로 올라온 트윗은 2010년인데 가입일은 2018년도로 뜨는 등 가입일과 트윗이 맞지 않는 경우가 있음 실제로 홈페이지에 들어가 보면 가입일은 2010년도로 뜸 이런 경우가 많음	해당 메타데이터 확인 후 메타데이터 품질 확보 방안 수립	차년도 개선

○ 오아시스 누리집 관련 개선 방안

시사점 유형	시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
3. 누리집	3.1 누리집 디자인	이미지 정보들이 미리보기가 제공되지 않음. 들어가서 확인해야 하는 번거로움이 존재	이미지나 동영상 등의 미리보기 이미지가 필수적인 정보 자원들에 한해서는 미리보기 썸네일 이미지를 제공하여 원본에 대한 대강의 정보를 확인할 수 있는 방안에 대해 고려할 수 있음	차년도 개선
	3.1 누리집 디자인	동영상 실행에 대해 ‘웹사이트 보기’ 버튼 제공	‘웹사이트 보기’ 버튼보다는 ‘동영상 보기’ 혹은 전체적으로 ‘보기/이동’ 버튼	즉시 개선
	3.1 누리집 디자인	아이콘 확인 어려움. 아이콘에 마우스 커서를 올려야만 어떤 것이 문서이고, 웹사이트인지 알려주는 것이 조금 불편하였음	조그맣게 글자를 써 넣고 링크를 다는 것이 한눈에 확인하기도 쉬움. 아이콘의 크기를 키우고, 직관적인 아이콘 모형을 사용하여 이용자가 빠르게 아이콘을 인식하고, 아이콘을 눌러 자료를 탐색할 수 있도록 지원해야 함	즉시 개선
	3.1 누리집 디자인	12개씩, 24개씩 같이 나눠서 보는 것이 아닌 전체 보기와 같은 시스템이 있으면 전체를 조망하면서 선택할 수 있을 것임	윈도우즈 탐색기에서 자세히 보기와 같은 제목 위주의 리스트 자료보기를 제시할 수 있을 것임	즉시 개선
	3.1 누리집 디자인	웹사이트 URL 링크 표시가 안 되어 있음	웹사이트 URL은 링크 표시(파란색 글씨에 밑줄)	즉시 개선

시사점 유형		시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
3.1 누리집 디자인	3.1 누리집 디자인	검색-25	검색에서 마우스 조작 시 검색 아이콘의 박스는 작고, 분리되어 있지 않아 색이 다르고 큼지막한 상세검색을 클릭하게 되어 불편함	누리집 해당 페이지 디자인 분석 후 개선	즉시 개선
	3.1 누리집 디자인	검색-26	상세검색에서 발행연도를 선택할 때 좌우로 한 달씩만 이동할 수 있고, 연 단위로 이동하거나 블록형으로 선택할 수 없어 불편함	누리집 해당페이지 기능 분석 후 개선	즉시 개선
	3.1 누리집 디자인	컬렉션-1	보기 유형에서 12개 보기는 가능하나 24개 보기 이상은 동작하지 않음	누리집 해당 페이지 기능 분석 후 개선	즉시 개선
	3.1 누리집 디자인	소셜미디어-1	하이퍼링크화 되지 않아 불편함	누리집 해당 페이지 기능 분석 후 개선	즉시 개선
	3.1 누리집 디자인	웹트렌드-1	안내 메시지가 화면에 가려지는 불편함	이용자가 정보를 확인하는 데 불편함이 없는 크기의 여백을 제공함	즉시 개선
	3.1 누리집 디자인	웹트렌드-2	그래프 보기에서 중복 라디오버튼	중복 제거	즉시 개선
	3.1 누리집 디자인	웹트렌드-6	"웹트렌드"라는 하나의 메뉴에서 갈라진 하위 메뉴임에도 불구하고 웹트렌드, 웹트렌드 활용예제, 태그클라우드 세 페이지의 UI 배치와 구성이 통일되어 있지 않아 이용이 어려움	웹트렌드와 태그클라우드에도 왼쪽의 바로가기 메뉴 제공	즉시 개선
	3.11 이용자와 상호 작용이 부족한 메뉴	소개-1	자주 묻는 질문은 있었지만, 직접적인 질의응답 창구가 없음. 사이트를 사용하다가 궁금한 점이나 모르는 점이 생겼을 경우 바로 질문을 주고받을 수 있는 질의응답 창이 필요	상호 작용을 높일 수 있는 질문 게시판 기능 제공.	현재 전체 누리집에서 대응하고 있음
	3.11 이용자와 상호 작용이 부족한 메뉴	소개-1	자주 묻는 질문은 있었지만 직접적인 질의응답 창구가 없음. 사이트를 사용하다가 궁금한 점이나 모르는 점이 생겼을 경우 바로 질문을 주고받을 수 있는 질의응답 창이 필요	인공지능 챗봇과 같은 신기술 도입 전략 방안을 고려해 볼 수 있음	차년도 개선
	3.12 공유 및 인용 기능 미제공	우수 사례-5	미국 의회도서관(LC) 우수 사례 조사 시사점 → Share 기능 제공	Share 기능 추가	즉시 개선 적용
	3.12 공유 및 인용 기능	우수 사례-7	미국 의회도서관(LC) 우수 사례 조사 시사점 → cite	Cite this term 기능	즉시 개선 적용

시사점 유형	시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
미제공		this term 기능 제공		
3.12 공유 기능 미제공	우수 사례-6	IIPC 우수 운영 사례 조사 시사점→공유 및 개방 강화	메타데이터(MODS) 목록을 다른 웹 서버에 제공하는 API 기능	차년도 개선(저작권법과 관련성은 없는지 점검 후 진행)
3.12 누리집 디자인	검색-27	이미지 클릭에 의한 웹사이트 링크 기능 미제공	온라인 보기 버튼과 병행하여 이미지 클릭 기능 추가	즉시 개선 적용
3.13 추천 콘텐츠 미제공	우수 사례-8	미국 의회도서관(LC) 우수 사례 조사 시사점 → 추천 사이트 기능 제공	추천 사이트 기능	현재 KOLIS에서 제공하는 관련 자료 제시 기능 활용하여 시범 적용 가능할 것으로 보임
3.2 누리집 실행 순서(전용뷰어 설치 전 이용 범위 안내)	검색-5	프로그램 설치 전 이용 범위 안내 필요	이를 사전 예방하기 위해, 프로그램 설치 허용을 누르기 전에 이용 범위를 안내해 주면 좋을 듯함	즉시 개선
3.3 누리집 장애인 지원 기능(음성 안내, 돋보기 기능)	우수 사례-4	미국 의회도서관(LC) 우수 사례 조사 시사점 → 자동 음성 생성 서비스 제공	자동 음성 생성 제공 서비스	차년도 개선(운영부서와 논의 후진행)
3.3 누리집 장애인 지원 기능(음성 안내, 돋보기 기능) 부재	검색-13	복지 서비스 부재(장애인, 노인 등)	시각장애인들을 위한 음성 지원 프로그램이나 노인을 위한 돋보기 기능 등 다양한 복지 서비스가 필요함	차년도 개선
3.4 개인 맞춤형 서비스 부재	검색-14	오아시스 이용 (빅)데이터 활용 이용자 맞춤 개인화 서비스의 필요성	안내 서비스, 추천 서비스 등 개인화 분석 서비스를 제공하여 이용자의 OASIS에 대한 만족도를 향상할 필요가 있음	차년도 개선
3.5 누리집 도움말 미비	검색-15	검색 항목에 <초성으로 검색하기>를 이용했을 때 선택한 초성이 제목을 말하는 것인지, 발행처인지, 사이트 이름이 정해지지 않아 불편함이 있었음	상세한 도움말 제공	즉시 개선
3.5 누리집 도움말 미비	검색-22	자료유형에서 웹사이트, 누리집, 웹페이지, 뉴스, 웹사이트의 구분 및 분류가 직관적으로 이해되지 않음	누리집 해당 페이지 확인 후 수정 필요	즉시 개선
3.5 누리집	재난아카	컬렉션 가장 최신 기사가	2년 엠바고에 대한 설명을	즉시 개선

시사점 유형	시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
도움말 미비	이브-3	2022년 기사라 현재 기사가 궁금하면 따로 검색해 보는 방법밖에 없어서 불편했음	이용자가 쉽게 알 수 있는 방법으로 제시할 필요가 있음	
3.5 누리집 도움말 미비	소셜미디어-6	소셜미디어 설명 부족	상세한 도움말 마련 및 제공	즉시 개선
3.5 누리집 도움말 미비	웹트렌드-3	웹트렌드 분석 도움말이 기능 위주의 설명임. 의미와 해석 방법에 대해서 좀더 상세한 설명이 있으면 좋겠음. 예를 들어, 레이블에 한국, 미국 등과 같은 나라명이 제시되어 있는데, 이 레이블의 의미가 무엇이고 그래프는 어떻게 해석할 수 있는지 제시되면 이용자가 웹트렌드 기능을 이해하는 데 도움이 될 것임	자세한 사항은 웹트렌드 활용 예제를 참고하라는 안내메시지 포함	즉시 개선
3.5 누리집 도움말 미비	웹트렌드-4	키워드 비중 도움말이 너무 간단함	키워드 비중이 키워드(용어) 빈도(term frequency)의 비중을 표현한 것인지 키워드가 출현한 문서(document frequency)의 비중을 표현한 것인지 설명을 추가하면 도움이 될 것임	즉시 개선
3.6 모바일 비친화적 누리집	우수 사례-3	미국 의회도서관(LC) 우수 사례 조사 시사점 → 보기 유형으로 슬라이드쇼 제공	모바일에 초점 맞춘 인터페이스 기능 적용	기요청사항
3.6 모바일 비친화적 누리집	검색-16	모바일 사용의 불편함	모바일 지향의 누리집 개선 과제 수행	차년도 개선
3.6 모바일 비친화적 누리집	컬렉션-1	보기 유형에서 12개 보기는 가능하나 24개 보기 이상은 동작하지 않음	누리집 해당 페이지 기능 분석 후 개선	즉시 개선
3.7 불편하고 모호한 검색 인터페이스	검색-20	상세검색 화면에서 소장자료와 웹사이트 본문 탭의 차이가 명확하지 않음	상세검색 메뉴 재설계함	차년도 개선
3.8 부적절한 미리보기 이미지 사이즈	컬렉션-3	미리보기 이미지가 제공 되는 경우, 이미지가 모두 보이는데 2-4초 가량 소요됨	미리보기를 위한 썸네일 이미지 생성 및 제시를 통한 웹페이지 로딩 속도를 높이는 방안	즉시 개선
3.8 부적절한 미리보기 이미지 사이즈	소셜미디어-4	각 기관 문양 이미지 깨짐: 고용노동부의 공식 트위터를 소개하는 기관 문양의 사진은 선명한 반면, 수집된 트위터에서의 고용노동부의 문양 사진은 선명하지 않고	확인 결과 이미지 사이즈가 현저히 작음. 트윗 문양에 맞게 사이즈를 조금 크게 할 필요가 있음	차년도 개선

시사점 유형		시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
			이미지가 깨져있음. 고용노동부뿐만 아니라 다른 기관 모두 수집된 트윗에서의 기관 문양 사진은 깨져있음		
	3.9 누리집 기능 오동작	재난아카이브-1	자료유형에 따른 적절한 유형의 자료 제시 오류(이미지 선택 시 동영상 제시, 동영상 선택 시 이미지 제시)	누리집 해당 페이지 기능 분석 후 개선	즉시 개선
	3.9 누리집 기능 오동작	재난아카이브-6	표기 오류: '재난아카이브'에서 '재난유형별'의 '해외'에는 세계의 자료가 표기되어 있지만 '지역별'의 '해외'에는 자료가 전혀 표기되어 있지 않음. 각 키워드에 적합한 문헌임에도 동시에 표기되지 않는 경우, 수정이 필요해 보임	누리집 해당 페이지 기능 분석 후 개선	즉시 개선
	3.9 누리집 기능 오동작	소셜미디어-2	제일 처음으로 올라온 트윗은 2010년인데 가입일은 2018년도로 뜨는 등 가입일과 트윗이 맞지 않는 경우가 있음 실제로 홈페이지에 들어가 보면 가입일은 2010년도로 뜸 이런 경우가 많음	누리집 해당 페이지 확인 후 수정 필요	즉시 개선
	3.9 누리집 스크립트 기능 오작동	컬렉션-5	'컬렉션' 장서(collection)에서 검색할 때, 검색 키워드 입력 후 엔터를 치면 실행되나 검색 버튼(돋보기 모양 이미지) 클릭은 동작하지 않음	누리집 해당 페이지 기능 분석 후 개선	즉시 개선

○ 정책 및 제도 관련 개선 방안

시사점 유형		시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
4. 정책 및 제도	4.1 제한된 자료 이용 범위	우수 사례-6	IIPC 우수 운영 사례 조사 시사점→공유 및 개방 강화	메타데이터(MODS) 목록을 다른 웹 서버에 제공하는 API 기능	차년도 개선(저작권법과 관련성은 없는지 점검 후 진행)
	4.1 제한된 자료이용범위	검색-7	국립중앙도서관에서만 자료 이용 가능한 것이 불편함	현재 디지털 자료 납본제도나 저작권법 제도하에서는	차년도 개선(국중과

시사점 유형		시사점 코드	도출된 시사점	제안된 개선 방안	논의결과
				국립중앙도서관 내에서만 자료 이용이 가능하나 자유이용 허락 라이선스(Creative Commons License) 유도를 통해 이용 범위를 확대하는 방안 시도해 볼 수 있음	논의 후 진행)
	4.2 수집 주제 범위의 편중화	검색-17	일부 분류 주제에 있어 소수 수집 자료	다양한 주제 분류가 수집될 수 있도록 지속적이고 포괄적 수집 도메인 확대	차년도 개선(국중과 논의 후 진행)
	4.3 트윗에 제한된 소셜미디어 수집 범위	소셜미디어-3	제한된 소셜미디어: 소셜미디어 창에 들어갔을 때 여러 기관의 소셜미디어가 정리되어 있음. 하지만 현재 트위터 계정들만 정리되어 있고 여러 기관이 유튜브, 인스타 등의 다양한 소셜미디어를 가지고 있는 만큼 다른 소셜미디어 계정들도 정리되어 있어야 한다고 생각함	이미 제시된 트윗 자료는 유지하고 추가로 소셜미디어 자료를 수집할 수 있을지 확대 전략 논의 필요	현재 트윗만 수집 가능하고 트윗도 경영권자에 의해 점점 수집가능 범위가 축소하고 있음



V

**오아시스 개선 및
활용성 강화
추진 전략**

5.

오아시스 개선 및 활용성 강화 추진 전략

5.1 오아시스 단기 기술 고도화 전략

- 시스템 현황과 진단 수행하여 오류 유형화 및 개선점을 도출하였음
- 오류 유형화를 통해 오아시스 기술 고도화 전략은 수집기, 색인 검색기, 재생기, 웹 데이터 품질 보증 기술의 개선으로 세분화하여 접근할 수 있음
- 본 연구에서는 각 세부 기술에 대하여 비교 분석하고 고도화 방향을 제시하고자 함

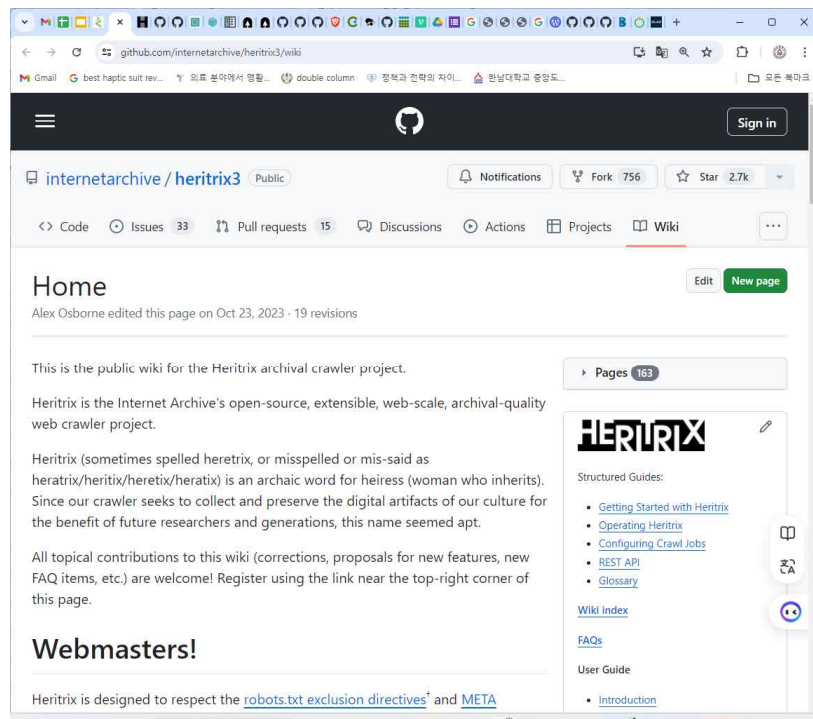
5.1.1 수집 기술 고도화 방안

- 필요성 및 목적
 - 동적 웹 기술 발전으로 인한 웹 환경의 변화로 기존 수집기의 현재 웹 환경 반영이 어려움
 - 수집기 고도화를 통하여 효과적이고 효율적인 웹자원 수집 및 웹 데이터 품질 향상이 필요함
- 추진 방안
 1. 현재 웹자원 수집 오류 현황 파악 및 웹 데이터 품질 측정
 - 웹자원 수집 오류 현황 파악
 - ▶ 웹자원 수집 오류 추출
 - ▶ 웹자원 수집 오류 유형화
 - 웹 데이터 품질 자동 측정
 - ▶ 수집된 웹 데이터 품질 지표 선정
 - ▶ 수집된 웹 데이터 품질 측정
 - 수집 오류 해결을 위해 조사된 후보 웹 수집 기술 비교 분석
 - ▶ 오류를 포함한 테스트 웹페이지 링크 표본 추출
 - ▶ 추출된 웹페이지 링크 표본에 대해서 최신 웹 수집 기술 적용
 - ▶ 각 후보 웹 수집 기술에 대한 품질 측정
 2. 최신 웹 수집 기술 동향 조사 및 비교 분석 및 적용
 - 1안: Heritrix, Browsertrix-crawler, Pywb-crawler 등 IIPC 회원국에서 사용하는 기존 웹 수집기 활용 가능성 파악
 - 2안: 최신 동적 웹 수집 기술인 selenium을 이용한 웹 수집 기술 조사

○ 후보 기술 개요

1. 1안: Heritrix 3.x

- Heritrix는 현재 많은 IIPC 회원들이 사용하고 있는 수집기임
- 국내 웹페이지들의 수집 결과를 보면서 오픈 소스를 커스터마이징하는 방법을 적용
- 오아시스를 위한 조직 정비가 필요하며 수집기를 보수 및 개선을 위해서는 전담할 기술직을 조직 내에 포함할 것을 제안
- Heritrix는 Internet Archive의 확장 가능한 오픈 소스, 웹 규모, 아카이브 품질의 웹 크롤러 프로젝트임
- 이 프로젝트는 미래 연구자와 세대의 이익을 위해 인류 문화의 디지털 유물을 수집하고 보존하는 프로젝트임

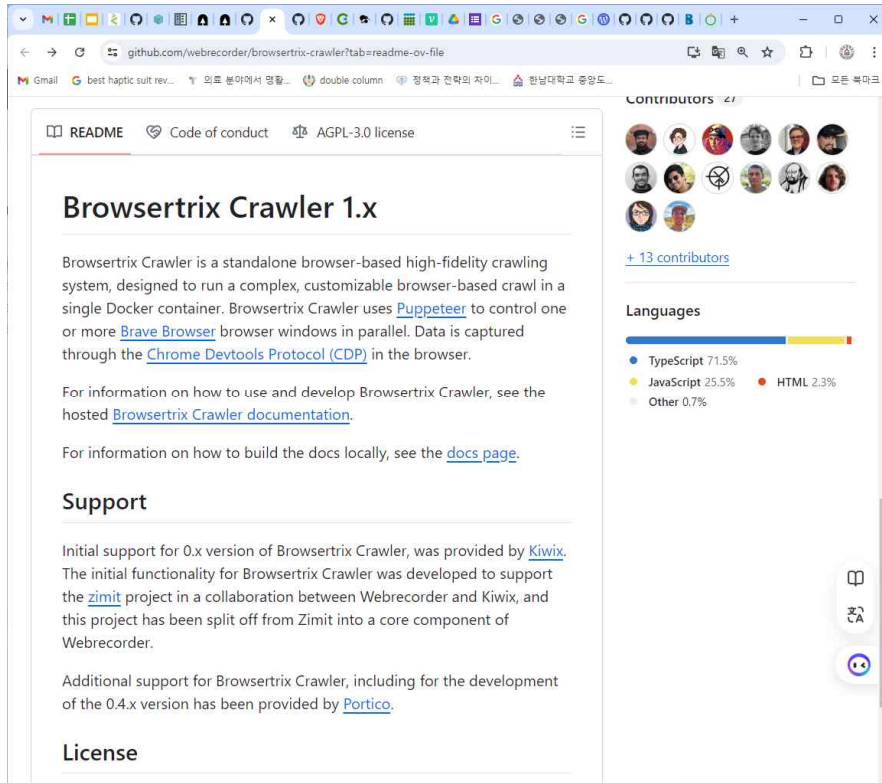


<그림 5-1> 오픈 소스 Heritrix3 Github 화면

2. 2안: Browsertrix-crawler

- Browsertrix는 단일 Docker 컨테이너에서 복잡하고 사용자 정의 가능한 브라우저 기반 크롤링을 실행하도록 설계된 독립형 브라우저 기반 고성능 크롤링 시스템임
- 이 크롤러는 Puppeteer를 사용하여 하나 이상의 BraveBrowser browser 창을 병렬로 제어함

- 데이터는 브라우저의 Chrome Devtools Protocol(CDP)을 통해 캡처됨
- 1) Browsertrix-crawler 설치 및 실행¹⁵⁾



<그림 5-2> 오픈 소스 Browsertrix Crawler 메인 화면

- ▶ Browsertrix Crawler를 사용하려면 크롤링을 실행하는 머신에 Docker를 설치해야 함
- ▶ Docker가 설치되어 있다고 가정하면, 다음 단계에 따라 크롤링을 실행하고 아카이브를 테스트할 수 있음
- ▶ Browsertrix Crawler 저장소를 복제할 필요도 없음
- ▶ 크롤링 데이터를 배치할 디렉터리를 선택한 후 다음 명령을 실행하면 됨([URL]을 크롤링하려는 웹사이트로 바꿀 것)

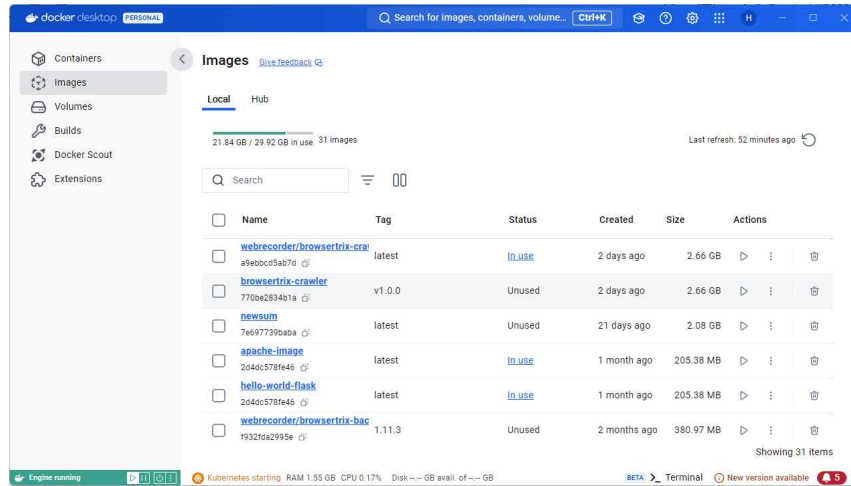
docker pull webrecorder/browsertrix-crawler 실행

docker run -v \$PWD/crawls:/crawls/ -it webrecorder/browsertrix-crawler creep --url [URL] --generateWACZ --text --collection 테스트

- ▶ 이제 크롤링이 실행되고 JSON Lines 형식의 로그가 콘솔에 출력됨
- ▶ 크롤링이 완료되면, 크롤링을 실행한 디렉터리의 crawls/collection/test/test.wacz에 WACZ 파

15) <https://crawler.docs.browsertrix.com/user-guide/> 문서참조

일이 생성



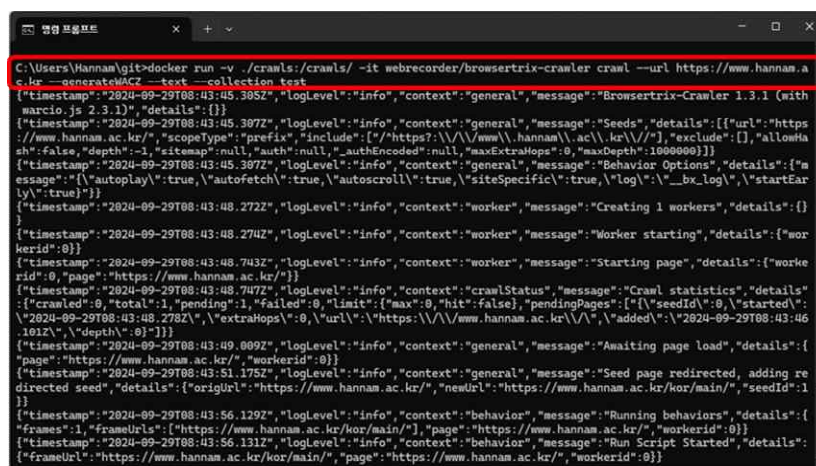
<그림 5-3> Browsertrix-crawler docker image가 docker desktop에서 사용되고 있는 설치 화면

- 2) 웹페이지 수집(크롤링)

- ▶ Cmd 창 명령행에서 아래 명령을 실행

```
docker run -v ./crawls:/crawls/ -it webrecorder/browsertrix-crawler crawl --url https://www.hannam.ac.kr --generateWACZ --text --collection test
```

- ▶ https://www.hannam.ac.kr을 수집하도록 명령



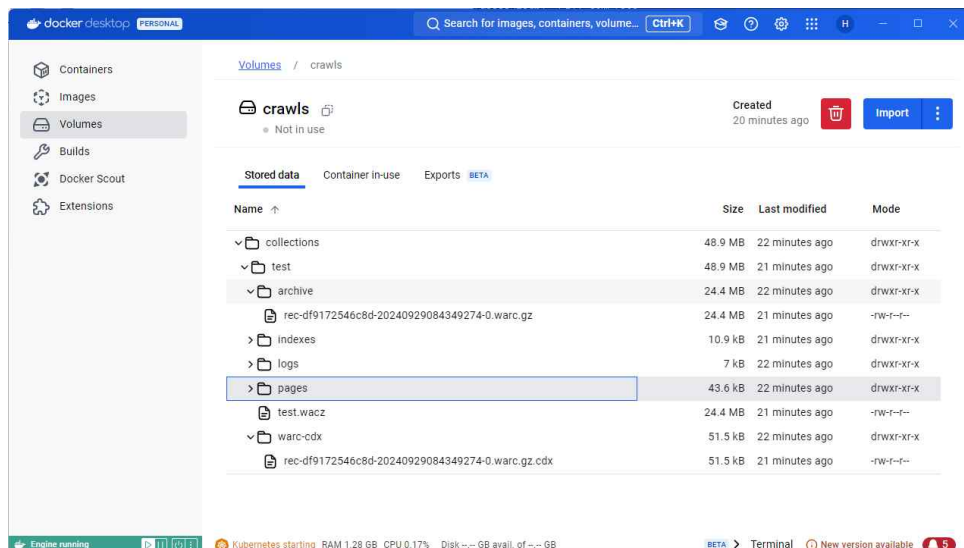
<그림 5-4> CLI 창에서 한남대학교 홈페이지를 수집하도록 명령 입력

- 3) 실행결과 log

- ▶ {"timestamp": "2024-09-29T08:43:45.305Z", "logLevel": "info", "context": "general", "message": "Browsertrix-Crawler 1.3.1 (with warcio.js 2.3.1)", "details": {}}
- ▶ {"timestamp": "2024-09-29T08:43:45.307Z", "logLevel": "info", "context": "general", "message": "Seeds", "details": [{"url": "https://www.hannam.ac.kr/", "scopeType": "prefix", "include": ["/^https?:\\W\\W\\W\\.hannam\\.ac\\.kr\\W/"], "exclude": [], "allowHash": false, "depth": -1, "sitemap": null, "auth": null, "_authEncoded": null, "maxExtraHops": 0, "maxDepth": 1000000}]}
- ▶ {"timestamp": "2024-09-29T08:43:45.307Z", "logLevel": "info", "context": "general", "message": "Behavior Options", "details": {"message": {"autoplay": true, "autofetch": true, "autoscroll": true, "siteSpecific": true, "log": {"__bx_log": true, "startEarly": true}}}}
- ▶ {"timestamp": "2024-09-29T08:43:48.272Z", "logLevel": "info", "context": "worker", "message": "Creating 1 workers", "details": {}}

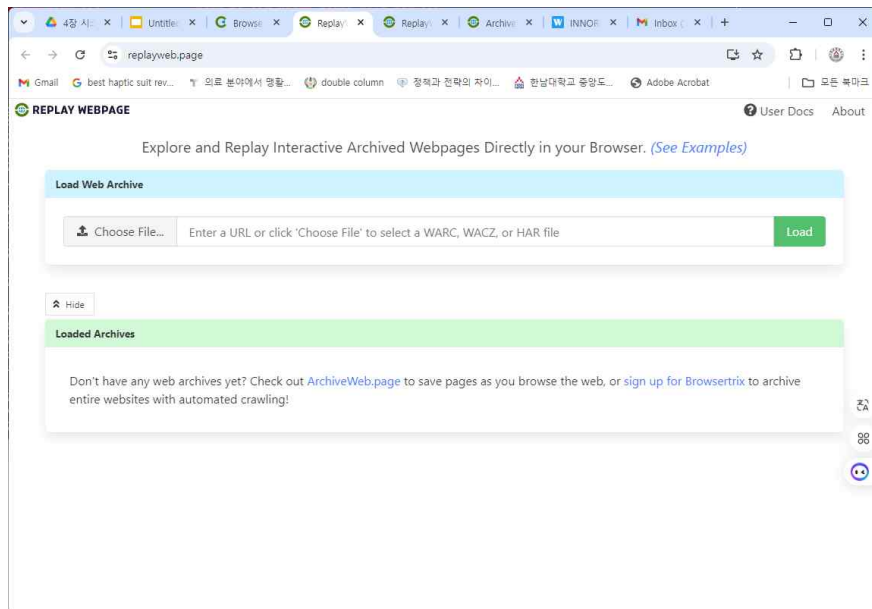
- 4) 수집된 웹페이지

- ▶ Docker desktop속에 volume으로 crawls 폴더 밑에 저장되어 있음

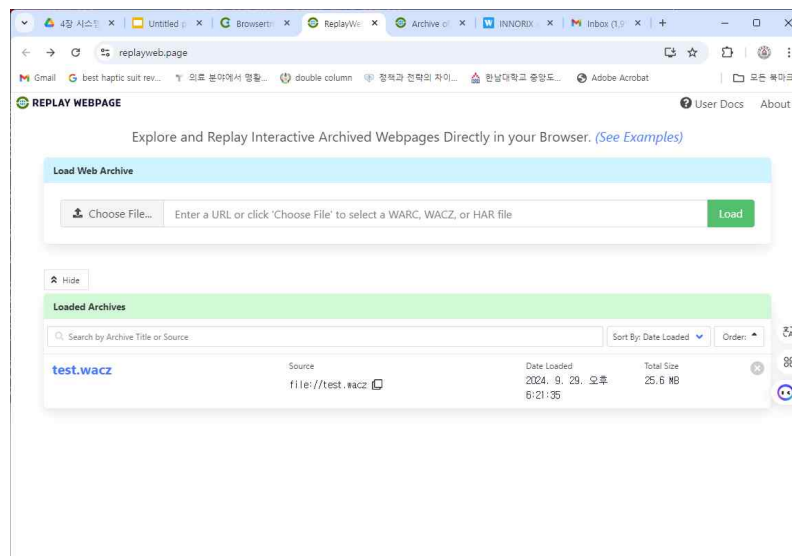


<그림 5-5> Docker desktop속에 volume으로 crawls 폴더 밑에 수집된 웹페이지

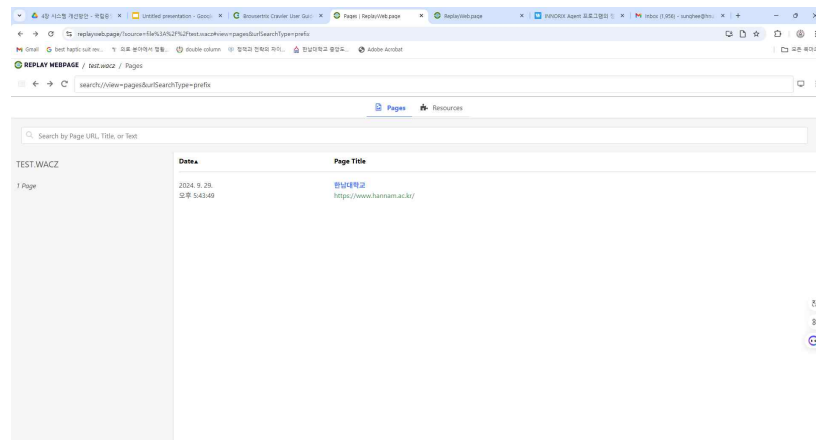
- 5) WARC 파일뷰어를 통한 웹페이지 확인하기



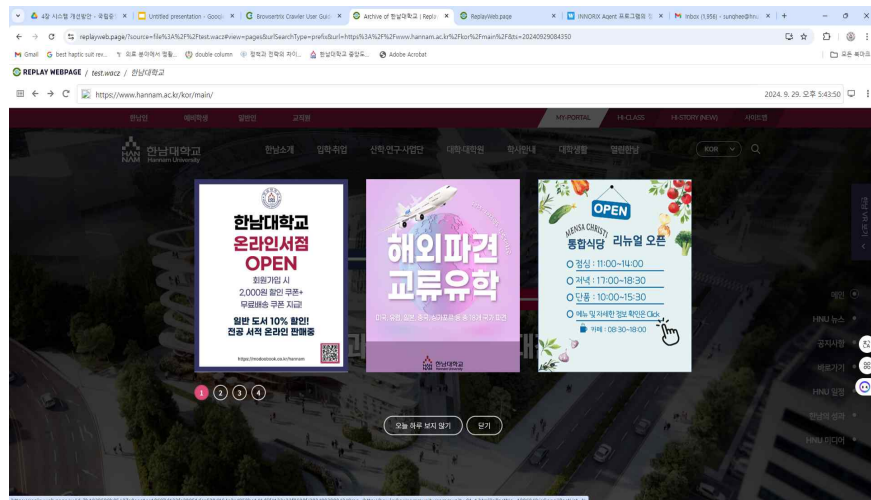
<그림 5-6> WARC 파일 뷰어인 Replay Webpage 실행 화면 (WARC 파일 로드 전)



<그림 5-7> WARC 파일 뷰어인 Replay Webpage 실행 화면 (WARC 파일(test.wacz) 로드 후)



<그림 5-8> TEST.WACZ 파일 상세정보 제시 화면 (WACZ 파일명, 수집 일시, 웹페이지 제목 및 URL)

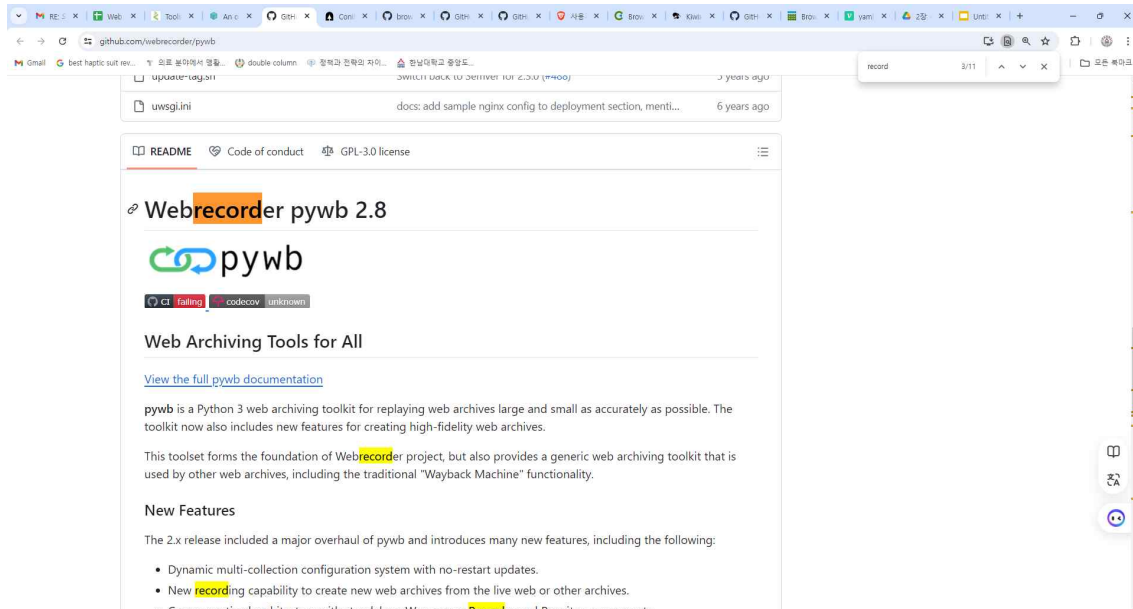


<그림 5-9> 수집된 WACZ 파일 재생 화면 (한남대학교 홈페이지: <https://www.hannam.ac.kr>) 화면

3. 3안: Pywb-crawler

- Webrecorder.net에서 개발하였고 Python 기반의 웹 아카이브 재생 도구임
- 2020년에 IIPC가 Pywb를 재생 솔루션으로 권장하고 IIPC 회원들이 Pywb를 채택하도록 지원함
- IIPC 회원들이 블로그 게시물과 기술 웨비나를 통해 공유해 온 공개적으로 사용이 가능한 전환 가이드와 사용 사례가 있어 Pywb으로의 전환 시 참고자료가 풍부함
- Pywb와 함께 호주 국립도서관의 Alex Osborne이 개발한 인덱싱 도구인 OutbackCDX를 사용하는 것을 권장함
- 재생을 위해 Pywb를 사용하는 웹 아카이브들로는 포르투갈(Arquivo.pt), 아이슬란드(Vefsafn.is),

영국(UK Web Archive) 및 미국 노스텍사스대학교(UNT의 Web Archive)를 포함함

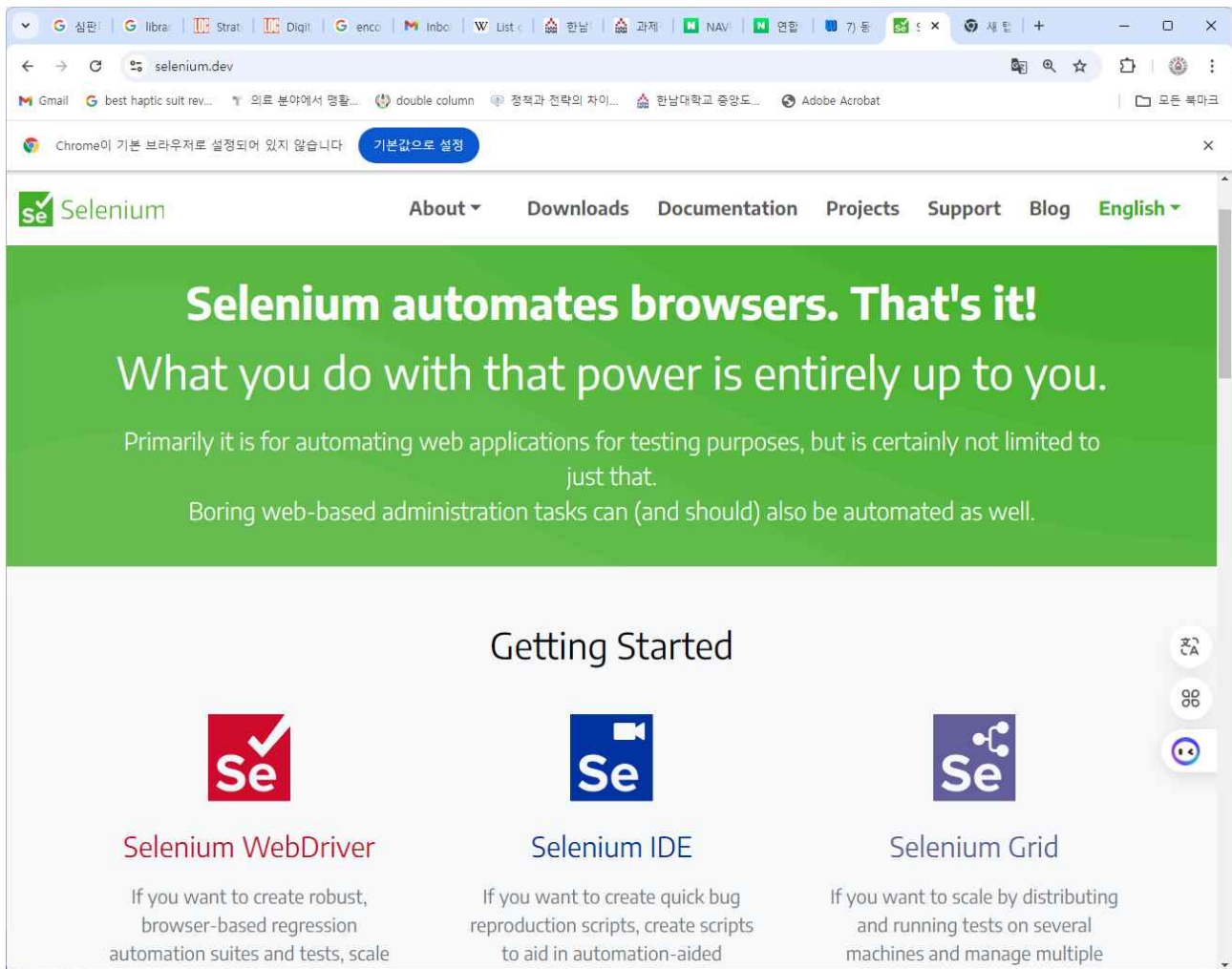


<그림 5-10> 오픈 소스 Pywb 메인 화면

4. 4안: Selenium¹⁶⁾

- ▶ Selenium은 웹 브라우저의 자동화를 가능하게 하고, 지원하는 다양한 도구와 라이브러리를 위한 포괄적인 프로젝트임
- ▶ 브라우저와 사용자 상호 작용을 에뮬레이트하는 확장 기능, 브라우저 할당을 확장하는 분산 서버, 모든 주요 웹 브라우저에서 상호 교환이 가능한 코드를 작성할 수 있는 W3C WebDriver 사양 구현을 위한 인프라를 제공함
- ▶ 다양한 언어(Java, Python, C#, Ruby, Javascript, Kotlin) 기반의 동적 웹 수집기 (Dynamic Web Crawler)
- ▶ 최신 웹페이지의 특징은 동적 웹으로, 기존 웹 수집기들을 사용하더라도 특정 컨트롤(예, 키워드 입력, 버튼 클릭)을 조작하여 결과를 크롤링하는 것은 불가능한데 이러한 부분을 추출하는 것이 가능(웹 스크래핑)

16) Selenium 프로젝트 공식 홈페이지: <https://www.selenium.dev/>



<그림 5-11> Selenium 프로젝트 공식 홈페이지 화면

5.1.2 색인 및 검색 기술 고도화 방안

○ 필요성 및 목적

- 사용자 요구에 맞는 웹자원 검색의 정확도를 높이기 위해서 필요
- 검색 기술의 고도화를 통해 이용자들에게 웹자원에 대한 빠르고 효과적인 접근성 제공
- 방대한 데이터를 효과적으로 분석 관리하여 활용도를 높이려는 목적

○ 추진 방안

1. 현재 검색 오류 현황 파악 및 색인 및 검색 성능 측정
 - 웹자원 검색 오류 현황 파악
 - ▶ 웹자원 검색 오류 추출

- ▶ 웹자원 검색 오류 유형화
- 현 색인 검색기 성능 측정
 - ▶ 색인 검색기에 대한 성능 지표 선정
 - ▶ 현재 사용 중인 색인 검색기 성능 측정
- 검색 성능 향상을 위해 조사된 후보 색인 검색 기술 비교 분석
 - ▶ 색인 검색 테스트 데이터 표본 추출
 - ▶ 추출된 테스트 데이터 대해서 최신 웹 색인 검색 기술 적용
 - ▶ 각 후보 웹 색인 검색 기술에 대한 색인 검색 성능 측정

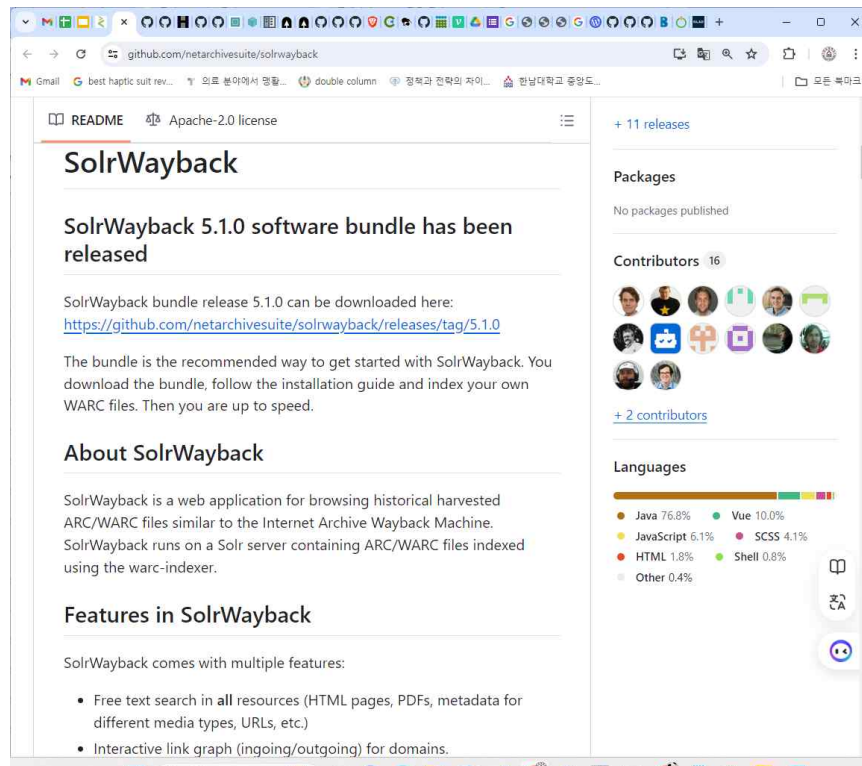
2. 최신 웹 검색 기술 동향 조사 및 비교 분석 및 적용

- 최신 웹 검색 기술 동향 조사 및 후보 색인 검색 기술 선정
 - ▶ 1안 : SolrWayback 등 IIPC 회원국에서 사용하는 기존 웹 수집기 활용 가능성 파악
 - ▶ 2안 : 최신 색인 검색 기술인 Elastic Search를 이용한 웹 색인 및 검색 기술 조사

○ 후보 기술 개요

1. 1안: SolrWayback

- SolrWayback은 Internet Archive의 Wayback Machine과 유사하게 수집된 기록 ARC/WARC 파일을 검색하기 위한 웹 애플리케이션임
- SolrWayback은 warc-indexer를 사용하여 색인화된 ARC/WARC 파일이 포함된 Solr 서버에서 실행됨



<그림 5-12> 오픈소스 SolrWayback 메인 화면

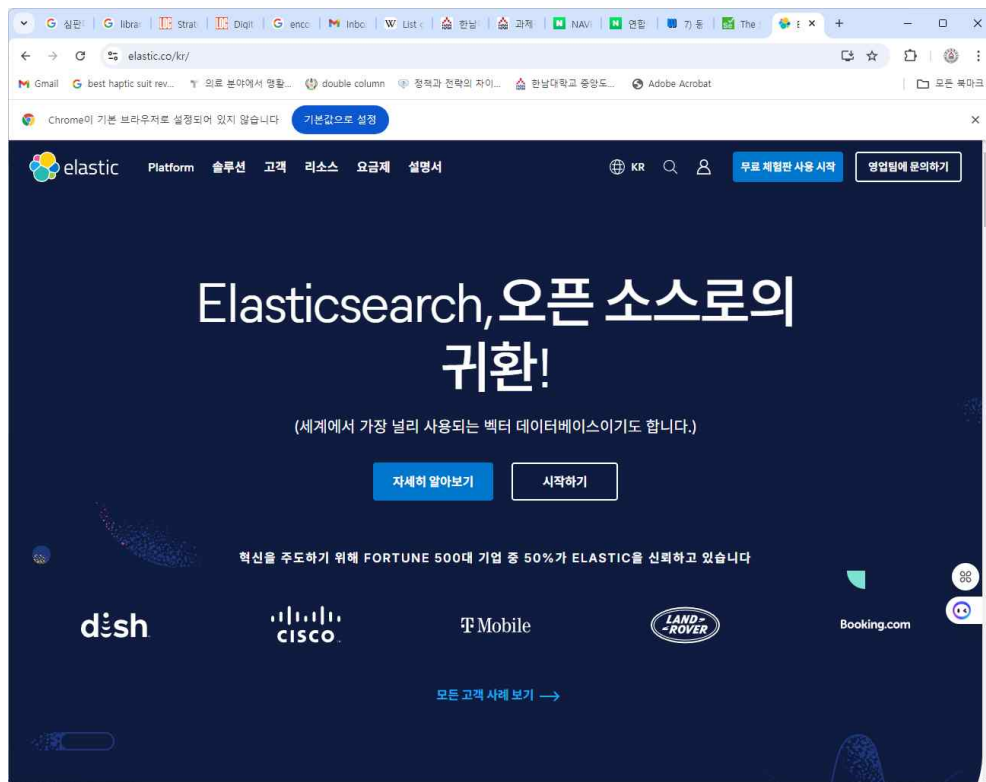
- SolrWayback-Features는 다음과 같음

- ▶ ① 모든 리소스(HTML 페이지, PDF, 다양한 미디어 유형에 대한 메타데이터, URL 등)에서 무료 텍스트 검색
- ▶ ② 도메인에 대한 대화형 링크 그래프(수신/발신)
- ▶ ③ 스트리밍 다운로드를 통해 검색 결과를 WARC 파일로 내보냄. 즉, 결과 집합의 크기에 제한이 없음
- ▶ ④ 검색 결과에서 파일을 Zip으로 내보내기(예: HTML, 이미지 콘텐츠, 비디오 등을 기본 형식으로 내보내기)
- ▶ ⑤ 사용자 정의 필드 선택을 통해 검색 결과를 CSV 텍스트로 내보낼 수 있음
- ▶ ⑥ 도메인용 워드 클라우드 생성
- ▶ ⑦ N-gram 검색 시각화
- ▶ ⑧ 도메인별 검색 결과 시각화
- ▶ ⑨ 크기, 들어오고 나가는 링크 수 등 시간 경과에 따른 다양한 도메인 통계를 시각화함
- ▶ ⑩ Gephi 형식으로 링크 그래프를 대규모로 내보낼 수 있음 (<https://labs.statsbiblioteket.dk/linkgraph/> 참조)
- ▶ ⑪ 구글 이미지와 유사한 이미지 검색

- ▶ ⑫ 이미지의 EXIF 메타데이터 정보를 사용하여 지도 위의 위치별 이미지 지리 검색
- ▶ ⑬ 파일 업로드로 검색함. (예: 이미지, PDF를 사용하여 리소스가 수집되었는지 확인하고 이미지를 사용하는 HTML 페이지를 찾음)
- ▶ ⑭ 리소스에 대해 색인화된 모든 필드를 보고 레코드에 대한 warc-header를 표시함
- ▶ ⑮ OpenWayback 또는 Pywb와 같은 재생 API를 사용하여 대체 재생 엔진을 모든 재생 엔진으로 구성함

2. 2안: Elastic Search

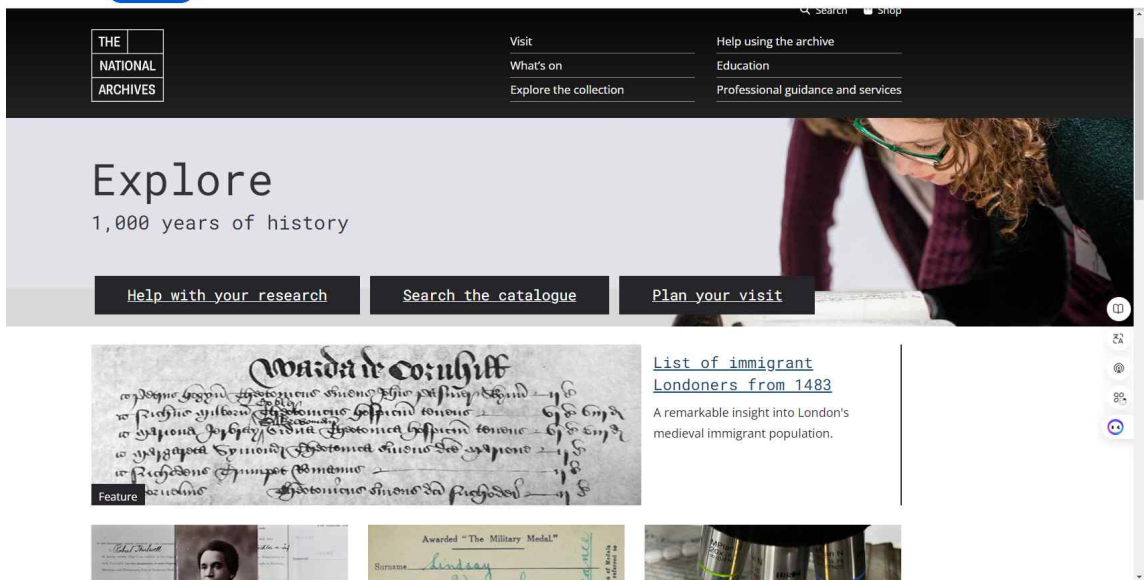
- ▶ Shay Banon은 2010년 2월에 Elastic search의 첫 번째 버전을 출시함
- ▶ Elastic search는 Apache Lucene을 기반으로 하는 검색엔진
- ▶ IIPC 회원인 영국 국립기록보존소에서 운영 중인 웹 아카이브도 검색 기능을 구현하는 데 Elastic search를 사용하고 있음¹⁷⁾
- ▶ 현재, Elastic Search는 DB-Engine에서 제공하는 검색엔진 랭킹에서 1위를 차지하고 있음¹⁸⁾



<그림 5-13> Elastic Search 프로젝트 공식 홈페이지 화면

17) Claire Newing, The National Archives (UK) and Phil Clegg, MirrorWeb. 영국 정부 소셜미디어 아카이브를 더욱 개선하기, IIPC blog, Tue, 09 February 2021, <https://netpreserveblog.wordpress.com/tag/elasticsearch/>, 2024. 11. 20. 검색됨

18) DB-Engine, DB-Engines Ranking of Search Engines, <https://db-engines.com/en/ranking/search+engine>, 2024. 11. 20. 검색됨



<그림 5-14> 검색 기능에 Elastic Search를 사용하고 있는 인덱스 서버(The National Archive)

☐ include secondary database models

26 systems in ranking, November 2024

Rank			DBMS	Database Model	Score		
Nov 2024	Oct 2024	Nov 2023			Nov 2024	Oct 2024	Nov 2023
1.	1.	1.	Elasticsearch	Multi-model ⓘ	131.64	-0.20	-7.98
2.	2.	2.	Splunk	Search engine	88.47	-2.81	-8.86
3.	3.	3.	Apache Solr	Search engine, Multi-model ⓘ	32.42	-0.56	-12.20
4.	4.	4.	OpenSearch +	Search engine, Multi-model ⓘ	17.60	+0.29	+4.29
5.	5.	↑ 6.	Algolia	Search engine	5.99	-0.16	-0.85
6.	6.	↑ 8.	Sphinx	Search engine	5.92	+0.00	-0.08
7.	7.	7.	Microsoft Azure AI Search	Search engine, Multi-model ⓘ	5.51	-0.20	-0.92
8.	8.	↓ 5.	MarkLogic	Multi-model ⓘ	4.28	-0.10	-3.89
9.	9.	9.	Virtuoso	Multi-model ⓘ	3.87	-0.04	-1.75
10.	10.	10.	ArangoDB	Multi-model ⓘ	3.09	-0.35	-1.45
11.	11.	11.	Coveo	Search engine	2.32	+0.08	-0.34
12.	12.	12.	Amazon CloudSearch	Search engine	1.74	-0.08	-0.57
13.	13.	13.	Meilisearch	Search engine	1.10	+0.03	-0.25
14.	14.	↑ 15.	Typesense	Search engine	0.80	-0.18	-0.12
15.	15.	↓ 14.	Xapian	Search engine	0.70	-0.01	-0.39
16.	16.	16.	CrateDB	Multi-model ⓘ	0.64	-0.03	-0.26
17.	17.	17.	Vespa	Multi-model ⓘ	0.61	0.00	-0.04
18.	18.	↑ 19.	Alibaba Cloud Log Service +	Search engine	0.42	-0.01	-0.04

<그림 5-15> DB-Engine에서 제공하는 검색엔진 대중성 랭킹에서 1위인 Elastic Search

5.1.3 재생 기술 고도화

○ 필요성 및 목적

- 중요한 웹 콘텐츠 보존 및 접근성을 높여 역사적 데이터의 활용성 증대
- 웹사이트의 다양한 요소(미디어, 스크립트 등)를 정확히 재현하여 신뢰성 있는 데이터 제공
- 디지털 문화유산으로서의 과거 웹페이지의 정확한 재생을 통해 사용자 경험을 향상하고 오류를 최소화

○ 추진내용

1. 수집된 웹 데이터 재생 오류 현황 파악 및 재생 웹 데이터 품질 측정

- 웹 데이터 재생 오류 현황 파악
 - ▶ 웹 데이터 재생 오류 추출
 - ▶ 웹 데이터 재생 오류 유형화
- 현 재생(웨이백 재생) 데이터 품질 측정
 - ▶ 재생 웹 데이터에 대한 품질 지표 선정
 - ▶ 현재 사용 중인 재생 웹 데이터 품질 측정
- 재생 성능 향상을 위해 조사된 후보 재생 기술 비교분석
 - ▶ 재생 테스트 웹 데이터 표본 추출
 - ▶ 추출된 테스트 웹 데이터 대해서 최신 웹 재생 기술 적용
 - ▶ 각 후보 웹 재생 기술에 대한 재생 성능 및 재생 웹 데이터 품질 측정

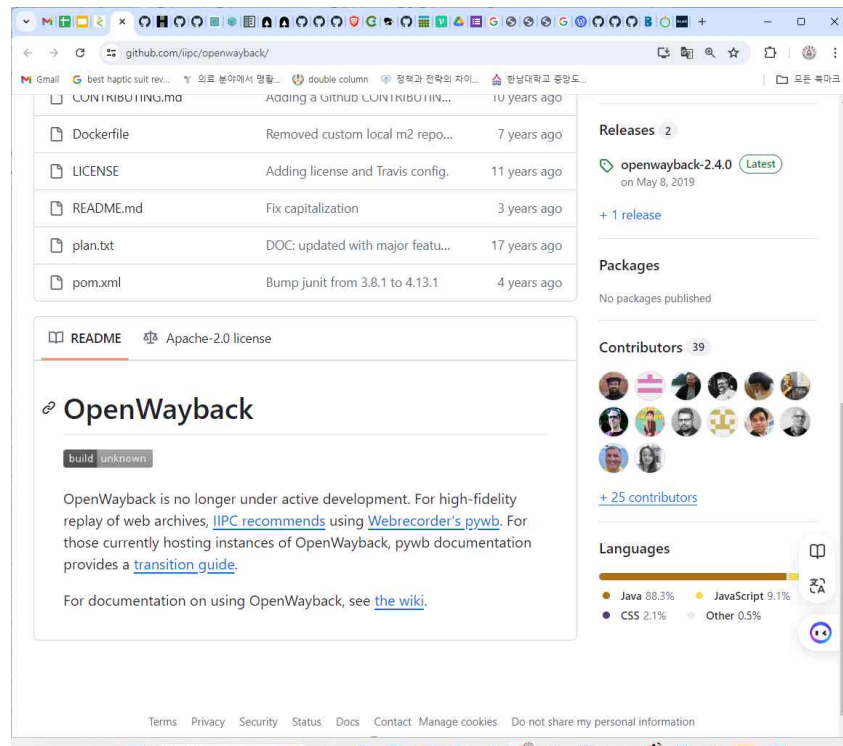
2. 최신 웹 재생 기술 동향 조사 및 비교 분석 및 적용

- 최신 웹 재생 기술 동향 조사 및 후보 색인 검색 기술 선정
 - ▶ 1안: OpenWayback 등 IIPC 회원국에서 사용하는 기존 웹 재생기 지속 활용 가능성 파악
 - ▶ 2안: 최신 재생 기술인 Pywb, 인덱싱 서버 OutbackCDX를 이용한 웹 색인 및 검색 기술 적용 가능성 파악

○ 후보 기술 개요

1. 1안: OpenWayback

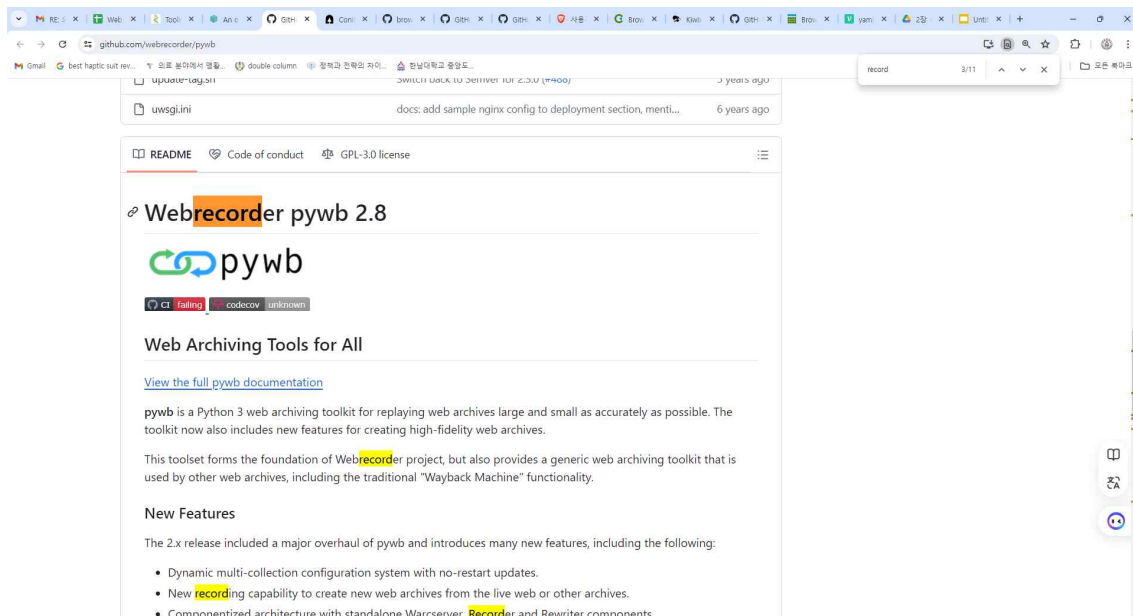
- OpenWayback은 더 이상 활발하게 개발되지 않음
- 웹 아카이브의 고품질 재생을 위해 IIPC는 Webrecorder의 Pywb를 사용하는 것을 권장함
- 현재 OpenWayback 인스턴스를 호스팅하는 경우 Pywb 설명서에서 전환 가이드를 제공함



<그림 5-16> 오픈 소스 OpenWayback 메인 화면

2. 2안: Pywb방안

- Webrecorder.net 에서 개발하였고 Python 기반의 웹 아카이브 재생 도구임
- 2020년에 IIPC가 Pywb를 재생 솔루션으로 권장하고 IIPC 회원들이 Pywb를 채택하도록 지원함
- IIPC 회원들이 블로그 게시물과 기술 웨비나를 통해 공유해 온 공개적으로 사용이 가능한 전환 가이드와 사용 사례가 있어 Pywb으로의 전환 시 참고자료가 풍부함
- Pywb와 함께 호주 국립도서관의 Alex Osborne이 개발한 인덱싱 도구인 OutbackCDX를 사용하는 것을 권장함
- 재생을 위해 Pywb를 사용하는 웹 아카이브들로는 포르투갈(Arquivo.pt), 아이슬란드(Vefsafn.is), 영국(UK Web Archive) 및 미국 노스텍사스대학교(UNT의 Web Archive)를 포함함



<그림 5-17> 오픈 소스 Pywb 메인 화면

5.1.4 데이터 품질 보증(QA: Quality Assurance) 기술

○ 필요성 및 목적

- 현재 웹자원 수집 시 진행되는 백신 점검에 더하여 기존 WARC 파일 보안 취약점 점검을 통한 수집된 웹자원의 보안성 강화에 기여
- 웹자료의 보안 취약점을 점검 시 개인 정보 보호 및 유해 콘텐츠 차단과 같은 웹자원 관리에 인공지능과 같은 최신 기법을 적용하여 “안전한” 웹자원 보존 기대

○ 추진내용

- 품질 보증 프로세스 수립
 - ▶ WARC 파일 파서 개발 및 보안 취약점 점검
 - ▶ ① 웹페이지 보안 취약 점검 도구를 WARC 파일에 적용이 가능하도록 파서를 개발하고 기존 취약점 리스트와 추가되는 리스트를 종합적으로 관리할 수 있는 파서 개발이 목적임
 - ▶ ② WARC 파일 파서 요구 사항 도출
 - ▶ ③ 보안 취약점 리스트 적용
- 인공지능 기반 개인 정보 보호 및 유해 콘텐츠 식별 기술 개발
 - ▶ ① 수집된 웹페이지에 개인정보 포함 시 개인 정보를 식별하여 익명 처리를 수행하여 개인 정보 보호 성능을 향상시켜야 하고, 유해 콘텐츠가 포함되었는지 확인하여 “안전한” 웹페이지를 보존

할 수 있는 기능을 확보해야 함

- 국외 웹 아카이빙 우수 사례에서 요구 사항으로 나타난 사항임

- ▶ ② 수집 시 수집자 육안으로 점검하기에는 무리가 있는 사항으로 인공지능 기법을 적용한 자동화가 필요한 상황임

- 1안: ChatGPT와 같은 기존 생성형 인공지능 적용 방안

- 2안: 자체 인공지능 기법 기술 개발 방안

- ▶ ③ 1안의 경우, 바로 적용할 수 있는 장점이 있으나 최소한의 커스터마이징이 필요함
- ▶ ④ 2안의 경우, 기술력 확보와 진화하는 웹페이지에 따라 기술 수정이 가능한 장점이 있으나 예산 및 기술 인력 확보가 전제되어야 함

- 지능형 기술을 이용한 누락 웹페이지 자동 점검 프로세스 개발

- ▶ ① 누락 웹페이지 자동 점검을 위한 지능형 기술 동향 조사

- ChatGPT

- OpenCV

- ▶ ② 누락 웹페이지 점검 프로세스 업무 흐름도 작성

- 웹페이지 수집(crawling)

- 웹페이지 URI 웹 접속

- 웹페이지 저장 폴더로 이동

- 원본 웹페이지와 수집 웹페이지 스크롤링을 통한 비교

- 유사도 측정 (누락 여부 확인, 404 코드 및 빈 웹페이지)

- ▶ ③ 업무 자동화(RPA: Robotic Process Automation)기법을 활용한 누락 웹페이지 자동 점검 프로세스 개발

- 웹페이지 수집(crawling)

- 웹페이지 URI 웹 접속

- 웹페이지 저장 폴더로 이동

- 지능형 기술 적용한 원본 웹페이지와 수집 웹페이지 비교(RPA)

- 자동 유사도 측정 (누락 여부 확인, 404 코드 및 빈 웹페이지) (RPA)

5.1.5 데이터 서비스 고도화

○ 필요성 및 목적

- 웹 아카이빙 내 연관 자원과의 추천을 통한 접근성과 활용도를 높일 필요성 대두
- 타 자원(단행본, 전자책, 정기간행물 등)과의 추천을 통한 공유 확산 강화 방안으로 고려
- 이용자로 하여금 웹 아카이빙 자료와 타 자원을 연결하여 더욱 연관이 있는 자원을 추천받는 사용자 경험을 가질 수 있도록 서비스 제공

○ 추진 방안

1. 웹 아카이빙 자료 분석 및 구조화

- 자료의 메타데이터 표준화 작업 수행
 - ▶ 웹 아카이빙 자료에 대한 메타데이터 구성 요소를 정의하고 표준화하여 정보의 일관성을 확보
 - ▶ 다양한 자료유형에 맞는 메타데이터 모델을 설계하여 데이터 검색과 관리의 효율성을 높임
- 웹 아카이빙 자료의 분류 및 태깅 시스템 구축
 - ▶ 자료의 주제, 형식, 생성일자 등을 기준으로 분류 체계를 마련
 - ▶ 사용자가 쉽게 접근할 수 있도록 태그 시스템을 도입하여 관련 자료 검색을 용이

2. 타 자원과의 연계 모델 설계

- 국립중앙도서관 자료관리시스템과의 연동
 - ▶ 다른 데이터베이스와의 연계를 위해 RESTful API를 설계하고 구현하여 상호 작용 가능
 - ▶ 데이터 호출과 전송 방식을 표준화하여 다양한 시스템 간의 원활한 연계
- 연관 자료 추천 알고리즘 개발 및 최적화
 - ▶ 사용자의 행동 패턴과 선호도를 분석하여 개인화된 추천 알고리즘을 개발
 - ▶ 추천 결과의 정확도를 높이기 위해 기계 학습 기법을 적용하여 알고리즘을 지속적으로 개선
- 지역 웹 아카이브 자료들과의 연계 추진
 - ▶ 교육 및 기술 제공을 통한 지역 웹 아카이브 구축 독려
 - ▶ 표준 메타데이터 구축을 통한 메타데이터 수준에서의 통합검색 제공 및 상호 운용성 향상

3. 추천 기술 구현 및 테스트

- 사용자 맞춤형 추천 시스템 인터페이스 개발
 - ▶ 직관적인 사용자 인터페이스(UI)를 설계하여 사용자가 쉽게 추천 기능을 활용
 - ▶ 사용자 경험(UX)을 고려한 피드백 수집 시스템을 도입하여 지속적으로 개선할 수 있는 기반을 마련
- 테스트와 피드백을 통해 추천 정확도 개선

- ▶ 프로토타입을 사용자 그룹에 배포하여 피드백을 수집하고 추천 알고리즘을 테스트
- ▶ 테스트 결과를 분석하여 알고리즘을 조정하고 개선하여 정확도를 높임

4. 사용자 피드백 기반 시스템 고도화

- 사용자 피드백을 수집하여 추천 정확도 조정
 - ▶ 사용자가 제공하는 피드백을 기반으로 추천 결과를 정기적으로 업데이트하여 만족도를 높임
 - ▶ 피드백 분석 결과를 바탕으로 사용자 맞춤형 설정을 강화
- 시스템 성능 모니터링 및 주기적 업데이트 수행
 - ▶ 시스템의 성능 지표를 지속적으로 모니터링하여 문제가 발생할 경우 즉시 대응
 - ▶ 정기적인 시스템 점검 및 업데이트를 통해 최신 기술과 트렌드를 반영

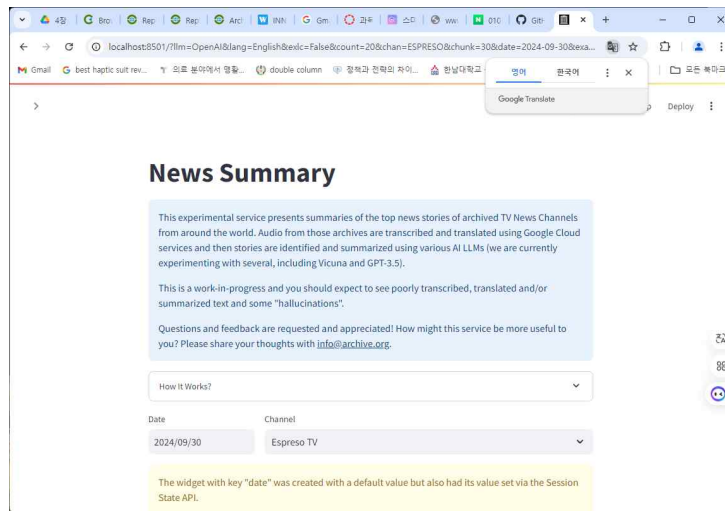
5. 데이터 후보 서비스(안)

- 이용자 빅데이터 수집
- 이용자 이용행 태 분석
- 이용자 요구사항 수집 및 빅데이터 분석 서비스 적용 제안
- 생성형 인공지능 기능 활용 요약 및 추천 서비스 제공
- 기존 웹 아카이브를 활용한 LLM 구축 제공 및 이에 기반한 fact checker 서비스 제공

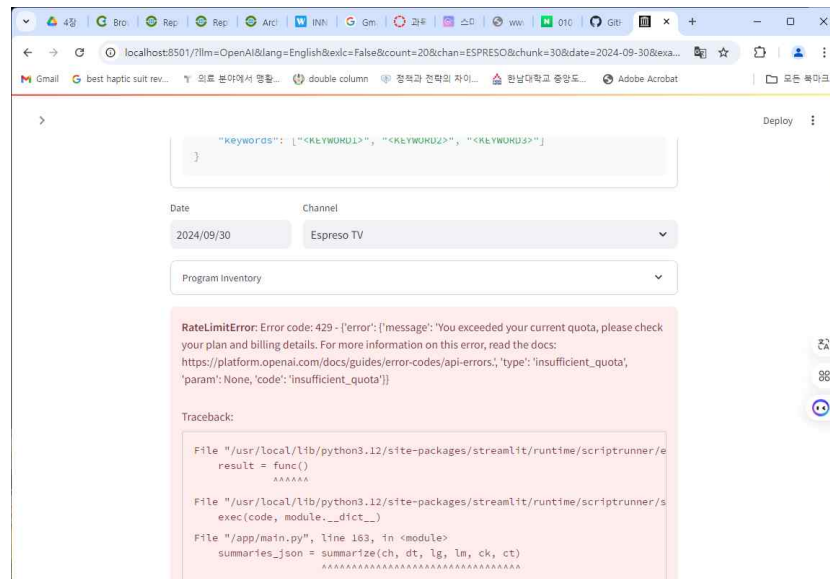
6. 후보 데이터 분석 서비스 예시

① News Summary

- Internet Archives와 Old Dominion University 연구진에 의해 개발된 오픈 소스
- 수집되고 보존되고 있는 뉴스 웹 아카이브 데이터에 대한 이용자 접근성을 높이기 위한 분석 요약 서비스
- 최신 생성형 인공지능 서비스를 웹 아카이브 데이터에 적용하여 웹 아카이브 활용성을 높이는 방안으로 파악되어 오아시스 신규 데이터 서비스로 추천



<그림 5-18> Internet Archive의 ChatGPT를 활용한 뉴스 요약 서비스 화면



<그림 5-19> Internet Archive의 ChatGPT를 활용한 뉴스 요약 서비스 화면(ChatGPT OpenAPI 허용 이용량 초과 에러 화면)

② Trendmachine

- 전 세계 웹을 아카이빙하는 비영리기관인 Internet Archive 소속 연구원과 Old Dominion University 소속 교수 및 연구원들이 2023년 IIPC 총회의 온라인 세션에서 발표한 내용

- 웹 아카이브에서 페이지에 대한 과거 관찰 결과를 기반으로 시계열 데이터로 웹페이지의 시간적 회복력을 정량화하기 위해 정규화된 점수를 계산하는 수학적 모델임
- Github에 오픈 소스로 공개하고 있기 때문에 오아시스에 적용하여 보존되고 있는 웹사이트의 회복력 가시화 서비스 제공을 이용자들에게 제안함
- 본 서비스를 적용하기 위해 충분한 양의 시계열 자료가 필요하므로 오아시스에 축적된 웹 아카이브 자료가 웹페이지 회복력 분석에 적합한지 시범 적용 후에 정식 서비스를 진행할 필요가 있음



<그림 5-20> 웹페이지 회복력 가시화 예시 화면

5.2 시스템 중장기 개선(migration) 전략

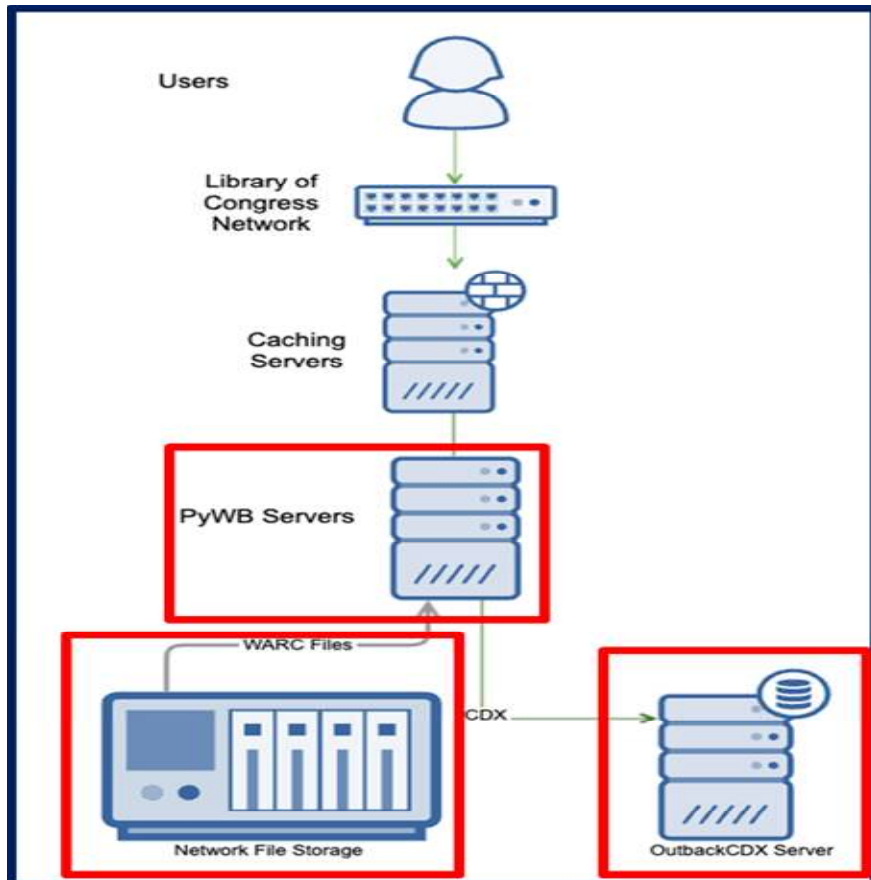
5.2.1 기존 시스템 개선 사례

○ 미국 의회도서관(Library of Congress)

- 최근 몇 년 동안 웹 아카이브에 대한 접근성을 개선하기 위한 새로운 "차세대" 도구가 개발. Webrecorder에서 개발하고 웹 아카이빙 커뮤니티에서 지원하는 Pywb라는 오픈 소스 Python 기반 재생 시스템이 포함됨
- IIPC는 OpenWayback에서 Pywb로의 잠재적으로 복잡한 전환에 대한 지원이 필요하다는 것을 알고 회원 기관과 다른 사람들이 Pywb로 마이그레이션하는 데 도움이 되는 전환 가이드 개발에 자금을 지원
- 또한 재생 도구에서 렌더링하기 위해 웹 아카이브에도 CDX 파일이 필요하다는 점에 유의해야 함. CDX 파일은 연결된 메타데이터 줄로, 각 줄이 WARC 파일 내의 단일 객체를 나타냄
- 계획해야 할 또 다른 큰 변경 사항은 OpenWayback은 평평한 CDX 구조를 가지고 있으며, 규모와

Pywb로의 전환을 통해 평평한 구조를 OutbackCDX 로 대체하여야 함. 그래야, 웹 아카이브 데이터를 보다 효율적으로 저장하고 조회할 수 있음

- 4페타바이트의 웹 아카이브를 구성하는 수십억 개의 파일에 대한 모든 인덱스를 다시 수집하는 시간이 걸림



<그림 5-21> 미국 LC의 새로운 웹 아카이브 액세스 시스템의 구조(재생SW(Pywb), 인덱스 서버(OutbackCDX) 및 WARC 저장소의 처리 흐름도)

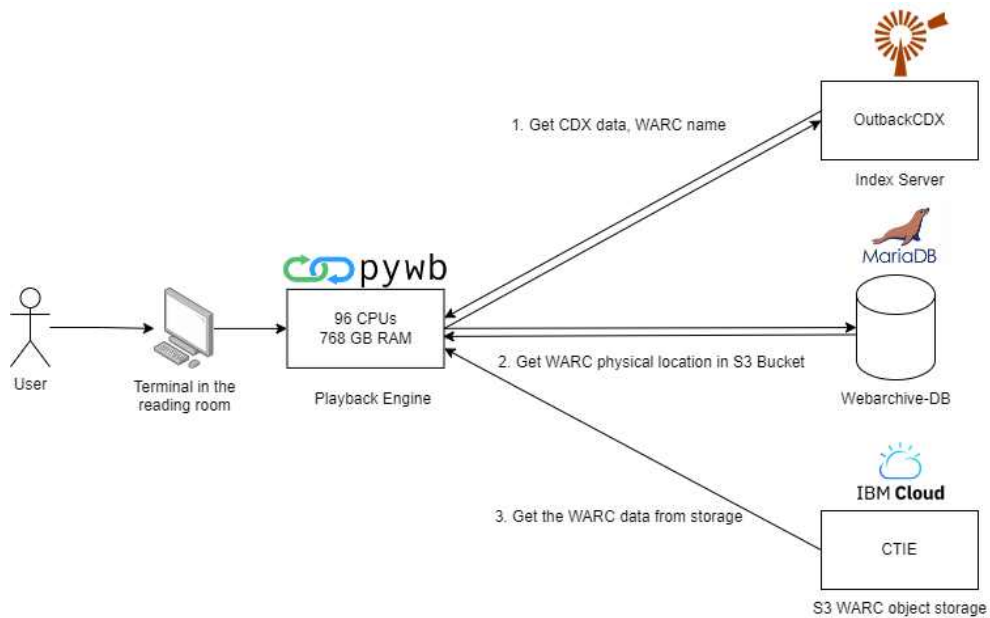
○ 룩셈부르크 국립도서관(BnL: National Library of Luxembourg)

- 2023년 동안 룩셈부르크 국립도서관(BnL)은 웹 아카이브를 새로운 인프라로 이전하는 작업을 수행
 - ▶ 하드웨어 : BnL은 인덱싱 및 재생과 관련된 애플리케이션을 호스팅하기 위해 4개의 새로운 고급 서버로 교체
 - ▶ 소프트웨어 : 오래된 OpenWayback 애플리케이션을 최신 Pywb/OutbackCDX 듀오로 업그레이드
 - ▶ 웹 아카이브 스토리지 : 339TB의 WARC 파일이 NetApp NFS에서 고성능 IBM S3 개체 스토리지로 마이그레이션

<표 5-1> 하드웨어 비교 - 이전 전 및 이전 후

비교 항목	이전 전	이전 후
기술(S/W 도구)	OpenWayback	SolrWayback + Pywb
전용 서버(H/W 인프라)	1 대	4 대
메모리(GB)	24	3072
프로세서	8	192

- 사용자에게 제공하는 서비스를 개선하기 위해, BnL은 Webrecorder가 개발한 재생 엔진인 Pywb 와 호주 국립도서관이 개발한 고성능 CDX 인덱스 서버인 OutbackCDX로 전환하여 이러한 문제를 해결하기로 결정
- 이러한 애플리케이션을 호스팅하고 향후 SolrWayback 설정을 염두에 두고 BnL은 각각 96개의 CPU와 768GB RAM을 장착한 4대의 고급 서버를 구입
- 웹 아카이브가 스토리지 인프라의 상당 부분을 차지하기 때문에, BnL은 IBM S3로 이전하기로 공동으로 전략적 의사결정. 액세스 시스템을 마이그레이션하는 동안 스토리지 계층을 마이그레이션하는 것이 합리적이었기 때문에 먼저 이 작업을 수행

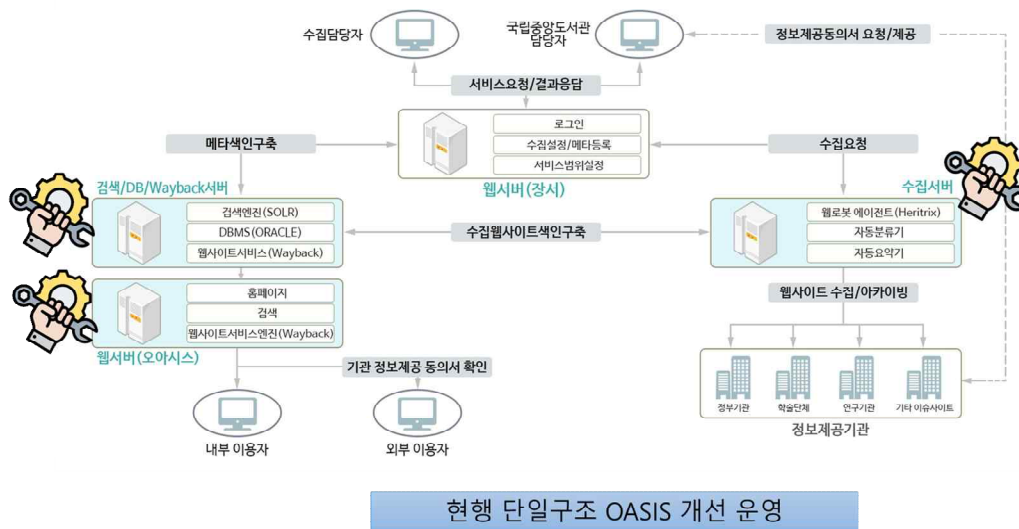


<그림 5-22> BnL의 웹 아카이빙 시스템 마이그레이션(Pywb + OutbackCDX 워크플로)

5.2.2 개선 방안(1안): 웹 데이터 수집·보존 개선을 위한 단기 기술 고도화

○ 방안 1:

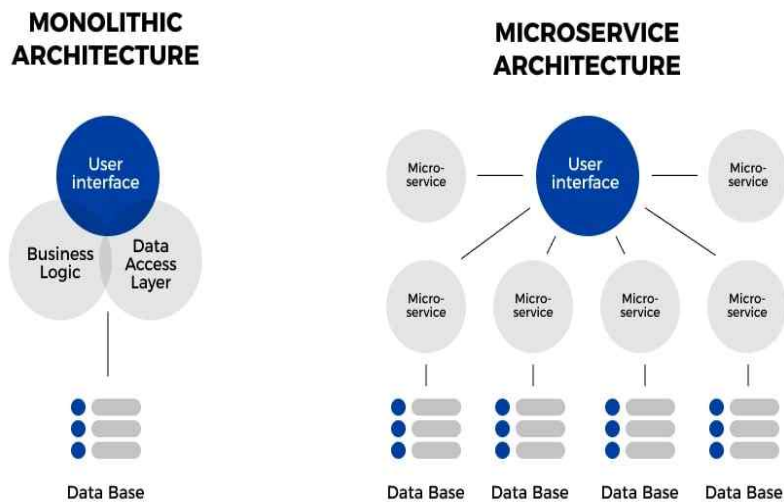
- 현행 시스템 평가 결과에 따라 현행 시스템을 개선 운영하는 방안을 제안함
- 기존 수집 보존 활용 프로세스를 그대로 유지하면서 즉시 개선이 가능한 개선점 위주로 개선하고 지속적으로 유지 보수하는 접근임
- 앞에서 언급한 단기 기술 고도화를 통해 실현이 가능함. 하지만, 다른 해외의 선진 웹 아카이빙 사례를 고려할 때, 장기적인 측면에서는 한계가 있는 방안임
- 선진 웹 아카이빙 기관 개선 사례(5.2.1 참조)에서 볼 수 있듯이, 장기적인 측면에서, 개선 웹 아카이브 이전(migration)을 함께 진행해야 함



<그림 5-23> 현행 단일구조 오아시스 개선 운영 개념도(방안1)

5.2.3 개선 방안(2안): 새로운 경량 구조(마이크로서비스) 도입

- 방안 2: **현행 시스템 운영과 함께 새 경량 시스템 시범 운영 병행(안) (2-way)**
 - 새로운 시스템 구조(예, 클라우드, 마이크로서비스) 도입 제안

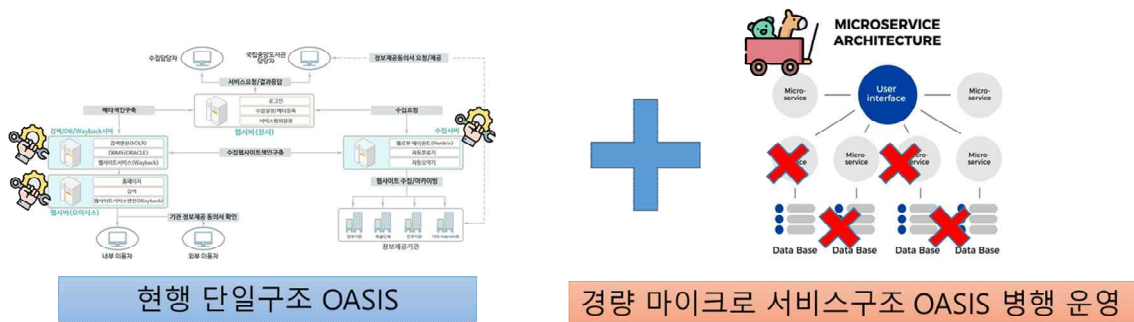


<그림 5-24> 단일 아키텍처 대 마이크로서비스 아키텍처 비교

<표 5-2> 단일 아키텍처와 마이크로서비스 아키텍처 비교(장단점)

	단일 아키텍처 (Monolithic Architecture)	마이크로서비스 (MicroService)
장점	<ul style="list-style-type: none"> 어떤 기능(서비스)이든지 개발되어 있는 환경이 같아서 복잡하지 않음 쉽게 고가용성 서버 환경을 만들 수 있음(같은 어플리케이션으로 하나 더 만들면 됨) End-to-End 테스트가 용이함 (MSA의 경우 테스트에 필요한 서비스들을 모두 동작시켜야 함) 	<ul style="list-style-type: none"> 기능별로 마이크로서비스를 개발하고, 작업 할당을 서비스 단위로 하면 개발자가 해당 부분을 온전히 이해할 수 있음 새로 추가되거나 수정 사항이 있는 마이크로서비스만 빠르게 빌드, 배포가 가능함. 해당 기능에 맞는 기술, 언어 등을 선택하여 사용할 수 있음 일부분의 오류가 있으면 해당 기능에만 오류가 발생하고 그 부분만 빠르게 고쳐서 정상화가 가능함
단점	<ul style="list-style-type: none"> 한 프로젝트의 덩치가 너무 커져서 어플리케이션 구동 시간이 늘어나고 빌드, 배포 시간도 길어짐 조그마한 수정 사항이 있어도 전체를 다시 빌드하고 배포를 해야 함 많은 양의 코드가 몰려있어 개발자가 모듈을 이해할 수 없고 유지 보수도 힘들 일부분의 오류가 전체에 영향을 미침 기능별로 알맞은 기술, 언어, 프레임워크를 선택하기가 까다로움 	<ul style="list-style-type: none"> 무엇보다 관리가 힘들. 작은 여러 서비스들이 분산되어 있기 때문에 모니터링이 힘들 서로를 호출하여 전체 서비스가 이루어지기 때문에 무조건 다른 서비스를 호출하는 코드가 추가되는데 이 부분이 모놀리식 아키텍처의 개발보다 조금 까다로움 통신 관련 오류가 잦을 수 있음. 마이크로서비스들끼리 계속 서로 통신을 하다보니, 단일 아키텍처에 비해 통신 관련 오류가 잦았음 테스트가 불편함. 예로 End-to-End 테스트를 위해 UI, Gateway 등등 여러 개의 마이크로서비스를 구동시켜야 했었음

- 본 안은 시스템 장기 개선 방안의 다른 하나로 기존 웹 아카이브 액세스 시스템과 새로운 경량 구조 마이크로서비스 병행하는 시스템 구조를 제안



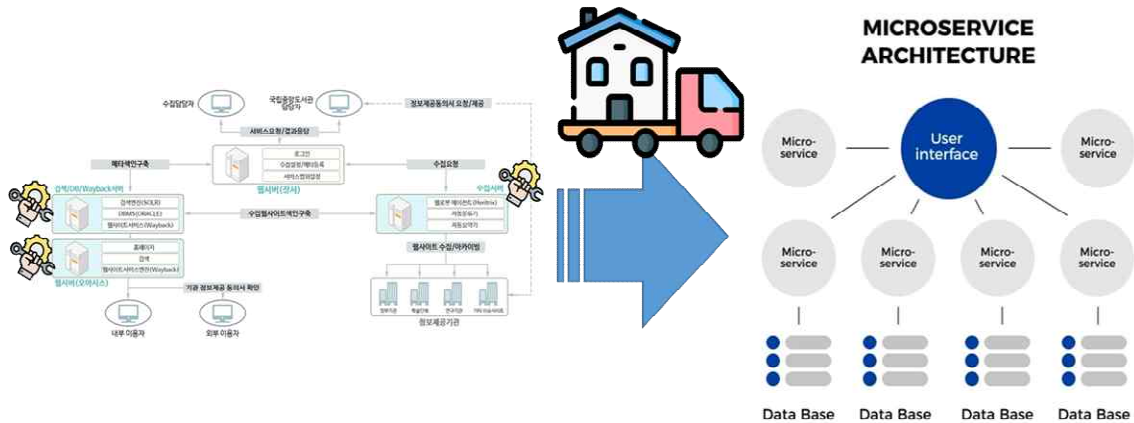
<그림 5-25> 현행 단일구조(OASIS)와 경량 마이크로서비스 구조병행 운영을 통한

점진적 마이그레이션(방안2)

5.2.4 개선 방안(3안): 새로운 웹 아카이브 시스템 구조로의 전면적인 마이그레이션

- 방안 3: 새로운 웹 아카이브 시스템 구조로의 전면적인 마이그레이션(안)

- 본 안은 시스템 장기 개선 방안 중 또 다른 하나로 새로운 마이크로 서비스 시스템 구조로의 전면적인 마이그레이션을 제안
- 본 방안 채택 시, ISP, ISMP 진행을 통해 개선 시스템으로의 면밀한 이전계획 수립이 추가로 필요함



현행 시스템에서 마이그레이션을 통한 새 시스템구조의 OASIS 개발 및 운영

<그림 5-26> 새로운 웹 아카이브 시스템 구조로의 전면적인 마이그레이션(방안3)

5.3 협력 및 홍보 방안

5.3.1 우리나라 관련 웹자원의 수집·보존 및 공유 강화를 위한 국제 협력 방안

○ 오아시스 및 국내 웹 아카이빙 체계 구축

- 전담 조직 구성: 전문 조직 구성으로 운영 지원
 - ▶ 조직 내 역할 분담: 데이터 수집, 메타데이터 관리, 기술 지원 분담
 - ▶ 협력 네트워크 형성: 관련 기관과의 네트워크를 통한 정보 공유
- 기술 인프라 개발: 웹 아카이빙을 위한 인프라 구축
 - ▶ 자동화 시스템 구축: 효율성을 위한 자료 수집 및 처리 자동화
 - ▶ 메타데이터 표준화: 일관성을 위해 국제적 표준 적용
- 국내 웹 아카이빙 협력체 조직 및 운영
 - ▶ 국내 광역대표도서관 및 디지털 자원 보존 의무기관을 중심으로 국내 웹 아카이빙 협력체 조직
 - ▶ 지역 사회의 '디지털 기억'으로서의 지역 디지털 웹자료 보존 독려
 - ▶ 디지털 문화유산으로서의 국내 웹자원 아카이빙 관련 공통 관심사 논의 및 문제 해결

○ IIPC의 회원국 간의 긴밀한 협력

- K 웹 콘텐츠를 바탕으로 수집 리스트 선별 및 국제 공동 아카이빙 사업 발굴
 - ▶ 정기적인 데이터 업데이트: 최신 정보 반영
 - ▶ 보안 관리 체계 수립: 데이터 보안 강화 및 점검
 - ▶ 사용자 피드백 반영: 서비스 품질 개선

○ 지역 및 민간 웹 아카이빙 교육 및 훈련 프로그램 운영

- 전문 인력 양성: 교육 프로그램 운영
 - ▶ 커리큘럼 개발: 아카이빙 기술 및 데이터 관리 커리큘럼 개발
 - ▶ 실습 기회 제공: 이론과 실무를 결합한 교육 실시

5.3.2 대국민 이용 활성화를 위한 오아시스 홍보 방안 제안

○ SNS나 유튜브 채널과 같은 디지털 마케팅 채널을 통한 오아시스 소개 및 활용 콘텐츠 제공

○ 대학 및 학회와 연계하여 OASIS 활용 강좌 및 워크숍 진행

- 워크숍 및 세미나 개최: 정보 교류와 네트워킹 촉진
 - ▶ 주제별 전문가 초청: 최신 동향 및 사례 공유
 - ▶ 참여자 피드백 수집: 지속적인 프로그램 개선

○ OASIS 전시회 개최: 전시회 개최를 통한 OASIS 홍보 및 체험 행사 개최

○ 이용 후기 공모전: OASIS 이용 후기를 공유하는 공모전 개최

VI

오아시스 비전 및 목표 수립

6. 오아시스 비전 및 목표 수립

6.1 오아시스 비전 및 목표 수립

○ 대한민국 웹 아카이빙(오아시스) 비전 및 목표 수립을 위해 다음과 같은 방법으로 수행함

6.2 비전 수립 근거

6.2.1 우수 운영 기관의 비전과 전략 분석

1) 룩셈부르크 웹 아카이브 서비스

○ 비전:

- 1. 완전성(품질 보증: 무결성)
 - ▶ 룩셈부르크 웹의 모든 콘텐츠는 LWA에 캡처되므로 라이브 웹의 콘텐츠가 손실되지 않음. 웹사이트
트의 보관된 버전은 사용자 경험의 모든 측면에서 원본에 충실함
- 2. 보존(보존: 시간적 완전, 기술 고도화 포함)
 - ▶ 룩셈부르크 웹은 미래 세대를 위해 장기적으로 보존됨
- 3. 접근(이용 가능)
 - ▶ LWA에 대한 접근은 법적으로 가능한 한 많은 청중에게 열려 있음
- 4. 연구용(핵심 이용자)
 - ▶ LWA는 국내 및 국제 연구에 적극적으로 사용되는 필수 정보 소스임
- 5. 인식(이해 관계자)
 - ▶ 웹사이트에 대한 납본과 웹 아카이빙의 필요성에 대한 인식은 웹사이트 소유자, 사용자, 콘텐츠
제작자에게 널리 알려져 있음

○ 전략:

- 1. 전략은 비전에 설명된 목표를 달성하기 위해 정의해야 하는 룩셈부르크 웹 아카이브의 다양한 측면을 다루고 있음
- 2. 룩셈부르크 법률
 - ▶ 납본법은 룩셈부르크 국립도서관의 사명과 임무를 정의함. 그러나 아카이브 해야 하는 항목에 대한 정의는 상당히 광범위함. 따라서 BnL은 룩셈부르크 웹의 최적 범위를 확보하기 위해 수단의 우선 순위를 정해야 함

- 3. 자원
 - ▶ 직원과 예산 모두 BnL이 목표를 얼마나 잘 달성할 수 있는지, 웹 아카이브 사용자에게 어떤 수준의 서비스를 제공할 수 있는지에 대한 중요한 요소
- 4. 기술 개발
 - ▶ 웹 아카이빙은 아직 기술적인 장벽이 많은 비교적 초기 단계의 분야임. 개발자와 웹 기록 보관자는 인터넷의 발전과 변화하는 과제를 따라잡기 위해 열심히 노력하고 있음. 크롤링 및 보존 기술의 새로운 발전은 웹 보존의 모습을 완전히 바꿀 수 있음
- 5. 사용자와 기여자의 요구
 - ▶ 현재의 웹사이트 소유자는 내일, 또는 10년, 20년 후의 웹 아카이브 사용자임. 게시자 및 콘텐츠 제작자와 연락을 유지하면 록셈부르크 웹 공간과 현재 보존해야 할 사항을 더 잘 이해할 수 있음
- 6. 지역 및 국제 협력
 - ▶ 납본과 디지털 보존은 국립도서관만의 영역이 아님. 공통의 과제와 책임을 다하기 위해 다른 아카이빙 기관, 연구원 및 웹 아카이빙 실무자 네트워크와 협력해야 하는 영역이 많이 있음

2) 미국 의회도서관 전략 구조(Library of Congress strategic framework)

○ 사명:

- 보편적이고 지속적인 지식과 창의성의 원천을 통해 의회와 미국 국민의 참여를 유도하고 영감을 주며 정보를 제공함(Engage, inspire, and inform Congress and the American people with a universal and enduring source of knowledge and creativity)

○ 비전:

- 모든 미국인은 미의회에 연결됨(All Americans are connected to the Library of Congress)

○ 전략목표: 1) 접근성 확장, 2) 서비스 개선, 3) 역량 강화, 4) 혁신 장려

- 1) 접근성 확장
 - ▶ 검색 가능성, 가용성 및 접근성 향상
 - ▶ 의회와 미국 지역 사회에 봉사하기 위해 도서관 컬렉션과 콘텐츠를 구축하고 강화
 - ▶ 지식과 창의성에 대한 지속적인 접근
 - ▶ 매력적이고 영감을 주는 디지털 공간과 대면 공간으로 방문객의 참여를 유도
- 2) 서비스 개선
 - ▶ 커뮤니티를 이해하고 참여
 - ▶ 지역 사회의 요구를 충족하기 위해 디지털 서비스를 강화
 - ▶ 신뢰할 수 있고 권위 있는 데이터 제공

- ▶ 영향력 확대를 위한 협력
- 3) 역량 강화
 - ▶ 유연하고 다양하며 응집력 있는 인력 지원
 - ▶ 디지털 기술에 중점을 두고 직원 역량 개발
 - ▶ 안전하고 공정하며 포용적이고 협력적인 환경을 조성
 - ▶ 성능 향상을 위해 운영 및 시스템 최적화
- 4) 혁신 장려
 - ▶ 진화하는 과제에 대처하기 위해 호기심과 실험을 유도
 - ▶ 상상력이 풍부한 접근 방식과 최신 기술로 혁신
 - ▶ 데이터를 활용하여 새로운 경로와 기회 발견

3) 국립중앙도서관 비전(2024-2028)

○ 사명:

- 국가지식문화유산을 전승하고, 국민의 지적인 삶을 풍요롭게 하며, 국가 및 세계의 문화 발전에 이바지함

○ 비전:

- 지성과 문화로 세계를 잇기

○ 5대 핵심 가치:

- 1) 활용성, 2) 신뢰성, 3) 포용성, 4) 전문성, 5) 지속가능성

○ 추진 전략

목표	세부 목표
1. 국민의 지적·문화적 삶을 풍요롭게 하는 도서관	① 수요 기반 지식정보 제공 ② 국가지식자원에 대한 차별 없는 접근 제공과 디지털 격차 해소 ③ 국민 모두를 위한 국가대표도서관
2. 국가지식자원의 보고로서의 도서관	④ 국가지식자원의 수집과 보존의 최상위 책임기관 ⑤ 현재와 미래에 접근성을 보장하는 지식정보의 수집과 보존 ⑥ 신뢰도 높은 도서관 데이터의 생산과 제공
3. 국내외 도서관계를 선도하는 도서관	⑦ 도서관을 중심으로 한 참여와 협력의 리더십 강화 ⑧ 우리나라 도서관계 영향력 강화 ⑨ 도서관 가치 확산과 사서 전문성 강화

4. 유연한 조직 운영과 효율적인 자원 분배	⑩ 변화에 능동적으로 대비하는 조직 ⑪ 영향력 측정을 위한 평가 시스템 마련
--------------------------	---

6.2.2 국내외 웹 아카이빙 서비스 비전의 시사점

- 국내외 웹 아카이빙 서비스 비전의 핵심 키워드를 정리해보면, ‘디지털 보존(digital preservation)’, ‘공개 개방(open access)’, ‘지식 유산(intellectual asset)’, ‘전 국가적(national-wide)’, ‘혁신강화(fostering innovation)’ 등이 주로 나타나는 것을 알 수 있음

6.3 오아시스 비전 및 목표, 추진 전략

- 오아시스의 비전, 목표, 추진 전략 및 과제는 다음과 같음
 - 앞서 도출된 시사점을 바탕으로 오아시스의 비전을 다음과 같이 정의함

국민과 함께하는 디지털 세상으로의 타임캡슐

- 핵심 추진 전략
 - 비전 도출에서 나타났던 서비스 비전의 핵심 키워드 (보존, 개인 맞춤, 지능화, 모바일, 공유 개방, 협력)를 반영하여 핵심 추진 전략을 도출함
 - 1. 문화유산 안심 보존(데이터 수집 확대 및 QA 강화)
 - 2. 핵심 이용자 집중(개인화)
 - 3. 신기술 융합을 통한 시스템 기반 기술 강화(기술 고도화 및 지능화)
 - 4. 사용 편의성 강화(모바일화)
 - 5. 다양한 이용자 확대(공유 개방)
 - 6. 지역과 국제기관의 긴밀한 동행(지역 및 국제 협력)

- 추진 전략:

<표 6-1> 오아시스 중장기 계획 추진 전략

1. 안심할 수 있는 품질 (데이터 보존 및 무결)	2. 이용자가 원하는 콘텐츠 (개인 맞춤)	3. 시스템 기반 기술 강화 (기술 고도화)
4. 모바일화 (시스템 접근성)	5. 공유 개방화 (콘텐츠 확대)	6. 지역과 국제기관의 긴밀한 동행 (지역 및 국제 협력)

○ 전략 (정책)과제:

<표 6-2> 오아시스 중장기 계획 정책 과제·실행 과제

추진 전략	전략(정책) 과제	실행 과제
1. 안심할 수 있는 데이터 보존과 품질 보장 (데이 터 보존 및 무결)	1.1 품질 보증(QA) 프로세스 자동화	1.1.1 WARC 파일 보안 취약점 점검
		1.1.2 지능형 기술을 도입한 누락 웹페이지 자동 점검 프로세스 개발
2. 이용자가 원하는 콘텐츠 (개인 맞춤)	2.1 이용자 특성에 맞춘 콘텐 츠 수집 및 제공	2.1.1 이용자 페르소나 기법을 이용한 디지털 장서 정책 수립
3. 시스템 기반 기술 강화 (기술 고도화)	3.1 개별 시스템 기술 고도화	3.1.1 수집기 고도화
		3.1.2 색인 검색기 고도화
		3.1.3 재생기(웨이백 재생) 고도화
	3.2 시스템 마이그레이션(장 기)	3.2.1 오아시스 시스템 정보화 전략 계획(ISP) 수립(단계적 운영 환경 개선)
		3.2.2 오아시스 시스템 고도화 사업
4. 언제 어디서나 이용 가 능한 오아시스(시스템 접근성)	4.1 시스템 접근성 강화	4.1.1 이용 접근성 강화된 누리집 개선
5. 공유 개방화(콘텐츠 제공 /전달/이용 체계 확대)	5.1 공유 확산 강화	5.1.1 웹 아카이빙 내 및 타 자원과의 연계 추천 기술 개발
6. 지역과 국제기관의 긴밀 한 동행	6.1 지역 및 국제 협력 체계 구축	6.1.1 국내 웹 아카이빙 협력 체계 구축 및 공유 확산

VII

과제별 세부 추진 방안

7. 과제별 세부 추진 방안

7.1 정책 과제 및 실행 과제별 세부 추진 방안

- 오아시스 개선 및 활성화 방안을 수행하기 위한 정책 과제 및 실행 과제는 6개의 추진 전략, 7개의 정책 과제, 11개의 실행 과제로 구성되어 있음

추진 전략	전략(정책)과제	실행 과제
1. 안심할 수 있는 데이터 보존과 품질 보장 (데이터 보존 및 무결)	1.1 품질 보증(QA) 프로세스 자동화	1.1.1 WARC 파일 보안 취약점 점검
		1.1.2 지능형 기술을 도입한 누락 웹페이지 자동 점검 프로세스 개발
2. 이용자가 원하는 콘텐츠 (개인 맞춤)	2.1 이용자 특성에 맞춘 콘텐츠 수집 및 제공	2.1.1 이용자 페르소나 기법을 이용한 디지털 장서 정책 수립
3. 시스템 기반 기술 강화 (기술 고도화)	3.1 개별 시스템 기술 고도화	3.1.1 수집기 고도화
		3.1.2 색인 검색기 고도화
		3.1.3 재생기(웨이백 재생) 고도화
	3.2 시스템 마이그레이션(장기)	3.2.1 오아시스 시스템 정보화 전략 계획(ISP) 수립(단계적 운영 환경 개선)
		3.2.2 오아시스 시스템 고도화 사업
4. 언제 어디서나 이용 가능한 오아시스(시스템 접근성)	4.1 시스템 접근성 강화	4.1.1 이용 접근성 강화된 누리집 개선
5. 공유 개방화(콘텐츠 제공/전달/이용 체계 확대)	5.1 공유 확산 강화	5.1.1 웹 아카이빙 내 및 타 자원과의 연계 추천 기술 개발
6. 지역과 국제기관의 긴밀한 동행	6.1 지역 및 국제 협력 체계 구축	6.1.1 국내 웹 아카이빙 협력 체계 구축 및 공유 확산

1.1.1

WARC 파일 보안 취약점 점검 [신규사업]

○ 필요성 및 목적

- 현재 웹자원 수집 시 진행하는 백신 점검에 더하여 기존 WARC 파일 보안 취약점 점검을 통한 수집된 웹자원의 보안성 강화에 기여

- 웹자료의 보안 취약점을 점검 시 개인 정보 보호 및 유해 콘텐츠 차단과 같은 웹자원 관리에 인공지능과 같은 최신 기법을 적용하여 “안전한” 웹자원 보존 기대

○ 추진내용

- 1. WARC 파일 파서 개발 및 보안 취약점 점검
 - ▶ 웹페이지 보안 취약 점검 도구를 WARC 파일에 적용이 가능하도록 파서를 개발하고 기존 취약점 리스트와 추가되는 리스트를 종합적으로 관리할 수 있는 파서 개발이 목적임
 - ▶ WARC 파일 파서 요구사항 도출
 - ▶ 보안 취약점 리스트 적용
- 2. 인공지능 기반 개인 정보 보호 및 유해 콘텐츠 식별기술개발
 - ▶ 수집된 웹페이지에 개인 정보 포함 시 개인 정보를 식별하여 익명 처리를 수행하여 개인 정보 보호 성능을 향상시켜야 하고, 유해 콘텐츠가 포함되었는지 확인하여 “안전한” 웹페이지를 보존할 수 있는 기능을 확보해야 함
 - 국외 웹 아카이빙 우수 사례에서 요구 사항으로 나타난 사항임
 - ▶ 수집 시 수집자 유관으로 점검하기에는 무리가 있는 사항으로 인공지능 기법을 적용한 자동화가 필요한 상황임
 - 1안: ChatGPT와 같은 기존 생성형 인공지능 적용 방안
 - 2안: 자체 인공지능 기법 기술 개발 방안
 - ▶ 1안의 경우, 바로 적용할 수 있는 장점이 있으나 최소한의 커스터마이징이 필요함
 - ▶ 2안의 경우, 기술력 확보와 진화하는 웹페이지에 따라 기술 수정이 가능한 장점이 있으나 예산 및 기술 인력 확보가 전제되어야 함

○ 추진 일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정		○	○	○			
소요 예산	구독/자체개발	15	15	15			45

○ 기대 효과

- WARC 파일 보안 취약점 점검 및 웹페이지 정보 보안을 통해 안전한 웹자원 보존 및 서비스 제공

- ▶ WARC 파일 보안 관리 기술을 확보함으로써 추후 WARC 파일 처리에 있어 기술력을 바탕으로 한 활용성 강화에도 도움이 될 것으로 예상

1.1.2

지능형 기술을 도입한 누락 웹페이지 자동 점검 프로세스 개발 [신규사업]

○ 필요성 및 목적

- 웹자원 수집 시 수집된 웹페이지가 제대로 수집되었는지 수집기 성능 및 수집된 웹 데이터 품질 점검을 해야 함
- 현재 이 품질관리 작업을 매 페이지 마다 육안 확인에 의한 수작업으로 진행 중임
 - ▶ 이는 비용 증가로 인한 웹자원 수집에 있어서 제한적인 웹자원을 수집하도록 하는 제약으로 작용하고 있는 실정임
 - ▶ 이에 수집이 제대로 이루어지지 않은 누락된 웹페이지를 (인공)지능형 기술을 도입하여 자동으로 점검하는 자동화 점검 프로세스 개발이 필요함

○ 추진내용

- 1. 지능형 기술을 이용한 누락 웹페이지 자동점검 프로세스 개발
 - ▶ 누락 웹페이지 자동 점검을 위한 지능형 기술 동향 조사
 - ChatGPT
 - OpenCV
 - ▶ 누락 웹페이지 점검 프로세스 업무 흐름도 작성
 - 웹페이지 수집(crawling)
 - 웹페이지 URI 웹 접속
 - 웹페이지 저장 폴더로 이동
 - 원본 웹페이지와 수집 웹페이지 스크롤링을 통한 비교
 - 유사도 측정 (누락 여부 확인, 404 코드 및 빈 웹페이지)
 - ▶ 업무 자동화(RPA) 기법을 활용한 누락 웹페이지 자동점검 프로세스 개발
 - 웹페이지 수집(crawling)
 - 웹페이지 URI 웹 접속
 - 웹페이지 저장 폴더로 이동
 - 지능형 기술 적용한 원본 웹페이지와 수집 웹페이지 비교(RPA)

- 자동 유사도 측정 (누락 여부 확인, 404 코드 및 빈 웹페이지) (RPA)

○ 추진 일정 및 소요 예산

단위 : 백만원

구분	2025	2026	2027	2028	2029	계
추진 일정	0	0	0			
소요 예산	15	15	15	0	0	45

○ 기대 효과

- 웹자원 수집 시 효율적인 웹 데이터 품질 보증 프로세스 확립
- 효율적인 자동화된 품질 보증(QA) 프로세스 개발을 통해 수집 자원 품질 점검 시간 절약 및 수집 자원의 양적 증대

2.1.1

이용자 페르소나 기법을 이용한 디지털 장서(웹자원) 개발
정책 수립 [\[계속사업\]](#)

○ 필요성 및 목적

- 현재 개발된 웹자원들의 활용도를 측정하여 다음 디지털 장서(웹자원) 개발 정책에 반영할 필요가 있음
- 이용자 페르소나 기법을 적용하여 연구자, 정책전문가, 교육생들이 필요로 하는 자원들에 대한 요구 조사를 할 필요가 있음
- 이를 통해 활용도 높은 디지털 자원(웹자원) 장서를 개발할 목적임

○ 추진내용

- 1. 이용자 페르소나를 통한 요구 조사
 - ▶ 연구자
 - 웹자원을 활용한 논문 조사
 - 인용 네트워크 기법을 활용하여 인용, 피인용 논문 조사
 - 논문 저자 및 공저자 논문 조사
 - ▶ 정책 전문가
 - 각 부처의 웹페이지 수집
 - 각 부처에 특화된 2차 웹자원 수요조사

- 각 부처에 수집 보존이 필요한 3차 웹자원 수요 조사
 - ▶ 교육생
 - 각 분야별 교육에 필요한 현용 웹자원 수요 조사
 - 각 분야별 교육에 필요한 과거 웹자원 수요 조사 및 웨이백 머신을 통한 접근성 제공
- 2. 디지털 자원(웹자원) 장서 개발 정책 수립
 - ▶ 포괄적 수집
 - 장서의 현재 이용 및 후대의 접근성 향상을 위해 가능한 모든 자료의 보존 관리
 - 가능한 수집에 의한 자원 확보를 원칙으로 하되 필요에 따라 자료 이관을 통한 자원 확보 병행 보존도 실시
 - 국가기록원, 지자체 등 유관 기관과의 보존 협력 강화
 - ▶ 선택적 수집
 - 장서의 수집 및 개발 기준에 따라 진행하되 가능한 매체 변환, 시스템 전환 등의 방식을 통한 자원 보존에 노력
 - 디지털 보존 공간의 확보 및 이용자 접근 개선, 관리 비용의 절감 등을 위한 중장기적 장서 수집 및 보존 정책 수립

○ 추진 일정 및 소요 예산

단위 : 백만원

구분	2025	2026	2027	2028	2029	계
추진 일정	0	0	0	0	0	
소요 예산	120	140	140	140	140	680

○ 기대 효과

- 체계적인 웹자원 수집을 통해 문화유산으로서의 웹자원 보존
- 시민을 위한 사회적 기억으로서의 웹자원 접근성 제공으로 시민의 만족도 제고
- 연도별 계획을 통한 웹자원 수집에의 효율적인 예산 집행

3.1.1

수집기 고도화 [계속사업]

○ 필요성 및 목적

- 동적 웹 기술 발전으로 인한 웹 환경의 변화로 기존 수집기의 현재 웹 환경 반영의 어려움
- 수집기 고도화를 통하여 효과적이고 효율적인 웹자원 수집 및 웹 데이터 품질 향상이 필요함

○ 추진내용

- 1. 현재 웹자원 수집 오류 현황 파악 및 웹 데이터 품질 측정
 - ▶ 웹자원 수집 오류 현황 파악
 - 웹자원 수집 오류 추출
 - 웹자원 수집 오류 유형화
 - ▶ 웹 데이터 품질 자동 측정
 - 수집된 웹 데이터 품질 지표 선정
 - 수집된 웹 데이터 품질 측정
- 2. 최신 웹 수집 기술 동향 조사 및 비교 분석 및 적용
 - ▶ 최신 웹 수집 기술 동향 조사 및 후보 수집 기술 선정
 - 1안: Heritrix, Browsertrix-crawler, Pywb-crawler 등 IIPC 회원국에서 사용하는 기존 웹 수집기 활용 가능성 파악
 - 2안: 최신 동적 웹 수집 기술인 selenium을 이용한 웹 수집 기술 조사
 - ▶ 수집 오류 해결을 위해 조사된 후보 웹 수집 기술 비교 분석
 - 오류를 포함한 테스트 웹페이지 링크 표본 추출
 - 추출된 웹페이지 링크 표본에 대해서 최신 웹 수집 기술 적용
 - 각 후보 웹 수집 기술에 대한 품질 측정

○ 추진 일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정	수집기 고도화	0	0				
	소요 예산	20	20	0	0	0	40

○ 기대 효과

- 최근 웹 기술의 발전에 따른 웹 환경의 변화를 반영한 웹자원 수집이 가능
- 기존 웹자원 수집기의 한계로 인한 웹 데이터 품질 저하의 개선을 기대할 수 있음

3.1.2

색인 검색기 고도화 [계속사업]

○ 필요성 및 목적

- 사용자 요구에 맞는 웹자원 검색의 정확도를 높이기 위해 필요
- 검색 기술의 고도화를 통해 이용자들에게 웹자원에 대한 빠르고 효과적인 접근성 제공
- 방대한 데이터를 효과적으로 분석 관리하여 활용도를 높이려는 목적

○ 추진내용

- 1. 현재 검색 오류 현황 파악 및 색인 및 검색 성능 측정
 - ▶ 웹자원 검색 오류 현황 파악
 - 웹자원 검색 오류 추출
 - 웹자원 검색 오류 유형화
 - ▶ 현 색인 검색기 성능 측정
 - 색인 검색기에 대한 성능 지표 선정
 - 현재 사용 중인 색인 검색기 성능 측정
- 2. 최신 웹 검색 기술 동향 조사 및 비교 분석 및 적용
 - ▶ 최신 웹 검색 기술 동향 조사 및 후보 색인 검색 기술 선정
 - 1안: SolrWayback 등 IIPC 회원국에서 사용하는 기존 웹 수집기 활용 가능성 파악
 - 2안: 최신 색인 검색 기술인 Elastic Search를 이용한 웹 색인 및 검색 기술 조사
 - ▶ 검색 성능 향상을 위해 조사된 후보 색인 검색 기술 비교 분석
 - 색인 검색 테스트 데이터 표본 추출
 - 추출된 테스트 데이터 대해서 최신 웹 색인 검색 기술 적용
 - 각 후보 웹 색인 검색 기술에 대한 색인 검색 성능 측정

○ 추진 일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정	색인 검색기 고도화		0	0			
	소요 예산	0	20	20	0	0	40

○ 기대 효과

- 방대한 웹자원으로부터 원하는 자원들을 효과적으로 검색할 수 있는 검색 정확도의 향상
- 검색기 술의 고도화를 통해 이용자들에게 웹자원에 대한 효과적이고 빠른 접근성 제공
- 개선된 접근성과 정확도를 통해 사용자 만족도를 제고할 수 있음

3.1.3

재생기(웨이백 재생) 고도화 [계속사업]

○ 필요성 및 목적

- 중요한 웹 콘텐츠 보존 및 접근성을 높여 역사적 데이터의 활용성 증대
- 웹사이트의 다양한 요소(미디어, 스크립트 등)를 정확히 재현하여 신뢰성 있는 데이터 제공
- 디지털 문화유산으로서의 과거 웹페이지의 정확한 재생을 통해 사용자 경험을 향상하고 오류를 최소화

○ 추진내용

- 1. 수집된 웹 데이터 재생 오류 현황 파악 및 재생 웹 데이터 품질 측정
 - ▶ 웹 데이터 재생 오류 현황 파악
 - 웹 데이터 재생 오류 추출
 - 웹 데이터 재생 오류 유형화
 - ▶ 현 재생(웨이백 재생) 데이터 품질 측정
 - 재생 웹 데이터에 대한 품질 지표 선정
 - 현재 사용 중인 재생 웹 데이터 품질 측정
- 2. 최신 웹 재생 기술 동향 조사 및 비교 분석 및 적용
 - ▶ 최신 웹 재생 기술 동향 조사 및 후보 색인 검색 기술 선정

- 1안: OpenWayback 등 IIPC 회원국에서 사용하는 기존 웹 재생기 지속 활용 가능성 파악
- 2안: 최신 재생기술인 Pywb, 인덱싱서버 OutbackCDX를 이용한 웹 색인 및 검색 기술 적용 가능성 파악
- ▶ 재생 성능 향상을 위해 조사된 후보 재생 기술 비교 분석
 - 재생 테스트 웹 데이터 표본 추출
 - 추출된 테스트 웹 데이터 대해서 최신 웹 재생 기술 적용
 - 각 후보 웹 재생 기술에 대한 재생 성능 및 재생 웹 데이터 품질 측정

○ 추진 일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정	(웨이백) 재생기 고도화			0	0		
소요 예산		0	0	20	20	0	40

○ 기대 효과

- 웹 데이터 복원 재생 품질 개선으로 사용자들이 정확한 과거 웹 콘텐츠를 접할 수 있음
- 다양한 기기와 환경에서도 동일한 품질의 콘텐츠 제공이 가능해짐
- 정확한 웹페이지 복원이 가능하여 학술 연구, 미디어 분석 등에서의 사용자 경험 만족도 증가

3.2.1

오아시스 시스템 정보화 전략 계획(ISP) 수립 (단계적 운영 환경 개선) [신규사업]

○ 필요성 및 목적

- 급변하는 웹 환경에 대응하여 안정적인 데이터 아카이빙 체계 구축
- 국내 주요 웹사이트와 콘텐츠의 장기적 보존을 통해 지속 가능한 국가 디지털 문화유산 보존
- 웹 아카이브 데이터의 공공 이용성을 확대하여 연구 및 교육에 기여

○ 추진내용

- 1. 운영 인프라 업그레이드
 - ▶ 최신 서버와 스토리지 적용으로 시스템 안정성 및 처리 속도 개선

- 2. 데이터 보안 및 접근성 강화
 - ▶ 고도화된 보안 프로토콜 적용 및 이용자 맞춤형 접근성 기능 추가
 - ▶ 지능형 기술 도입을 통한 개인 정보 보호 및 보안 취약점 상시 점검
- 3. 자동화 및 효율성 향상
 - ▶ 수집 자동화 도입 및 관리 시스템 개선을 통해 운영 효율 증대
 - 수집 자동화 기술 도입
 - 관리 시스템 개선

○ 추진 일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정	정보화 전략 계획(ISP) 수립	○					
	소요 예산	30	0	0	0	0	30

○ 기대효과

- 공공 자료와 중요 웹 콘텐츠의 지속적 보존을 통해 사회적 기억으로서의 웹 데이터 유실 방지
- 편리한 검색과 접근 방식 제공으로 공공 활용도 증대
- 웹 아카이브 시스템을 통한 국가 디지털 자산 보존 및 체계적 관리 강화

3.2.2

오아시스 시스템 고도화 사업 [신규사업]

○ 필요성 및 목적

- 급변하는 웹 환경에 대응하여 안정적인 데이터 아카이빙 체계 구축
- 국내 주요 웹사이트와 콘텐츠의 장기적 보존을 통해 지속 가능한 국가 디지털 문화유산 보존
- 웹 아카이브 데이터의 공공 이용성을 확대하여 연구 및 교육에 기여

○ 추진내용

- 1. 운영 인프라 업그레이드
 - ▶ 최신 서버와 스토리지 적용으로 시스템 안정성 및 처리 속도 개선

- 2. 데이터 보안 및 접근성 강화
 - ▶ 고도화된 보안 프로토콜 적용 및 이용자 맞춤형 접근성 기능 추가
 - ▶ 지능형 기술 도입을 통한 개인 정보 보호 및 보안 취약점 상시 점검
- 3. 자동화 및 효율성 향상
 - ▶ 수집 자동화 도입 및 관리 시스템 개선을 통해 운영 효율 증대
 - 수집 자동화 기술도입
 - 관리 시스템 개선

○ 추진일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정	시스템 고도화		0	0			
소요 예산		0	300	300	0	0	600

○ 기대 효과

- 공공 자료와 중요 웹 콘텐츠의 지속적 보존을 통해 사회적 기억으로서의 웹 데이터 유실 방지
- 편리한 검색과 접근 방식 제공으로 공공 활용도 증대
- 웹 아카이브 시스템을 통한 국가 디지털 자산 보존 및 체계적 관리 강화

4.1.1

이용 접근성 강화된 누리집 개선 [계속사업]

○ 필요성 및 목적

- 다양한 사용자층이 웹 아카이브 콘텐츠에 쉽게 접근할 수 있도록 접근성 향상
- 정보 취약 계층(노인, 장애인 등)의 정보 접근성 개선을 통해 디지털 격차 해소
- 법적 웹 접근성 준수로 공공기관의 법적 책임을 다하고 신뢰도 강화

○ 추진내용

- 1. 웹사이트 성능 향상
 - ▶ 이미지 깨짐 현상 개선

- 이미지 깨짐 현상 자동 인식 기술 적용
- 링크된 이미지 경로의 업데이트 여부를 자동 검수하는 이미지 자동 업데이트 기능 적용
- ▶ 접속 불가능한 링크 및 누락 페이지 관리
 - 정기적이고 자동화된 링크의 유효성을 점검 통해 접속 불가능한 링크 즉시 보고 및 관리
- ▶ 페이지 로딩 시간 최적화
 - 캐싱 및 로딩 최적화를 통한 웹사이트 응답 속도 향상
 - 서버 응답 시간을 줄이기 위해 성능이 좋은 서버와 네트워크 인프라를 구축
- 2. 모바일 최적화
 - ▶ 반응형 웹 디자인 적용
 - 다양한 화면 크기에 따라 제공되는 보기 메뉴가 자동 조정되어 데스크톱, 스마트폰, 태블릿에서 최적화된 사용자 경험을 제공
 - ▶ 터치 친화적 디자인 적용
 - 터치 기반 UI 요소인 슬라이드 보기 메뉴 제공해 모바일 사용자의 편의성을 높임
- 3. 접근성 및 활성화를 강화한 메뉴 제공
 - ▶ Share 기능 추가
 - ▶ 도움말 제공
 - ▶ 인용 기능
 - ▶ 추천 컬렉션 기능
- 4. 정보 취약 계층 접근성 강화를 위한 보조 기술 지원
 - ▶ 스크린 리더 및 자동 음성 생성 서비스 제공

○ 추진 일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정	누리집 개선		0	0			
소요 예산		0	20	20	20	20	20

○ 기대효과

- 접근성 강화로 다양한 사용자들의 사용자 경험 및 이용 만족도 증대
- 정보 취약 계층의 접근성 확대를 통해 포용적 디지털 환경 조성

- 웹 접근성 준수로 법적 요구사항 충족 및 이미지 제고

5.1.1

웹 아카이빙 내 및 타 자원과의 연계 추천 기술 개발

[계속사업]

○ 필요성 및 목적

- 웹 아카이빙 내 연관 자원과의 추천을 통한 접근성과 활용도를 높일 필요성
- 타 자원(단행본, 전자책, 정기간행물 등)과의 추천을 통한 공유 확산 강화 방안으로 고려
- 이용자로 하여금 웹 아카이빙 자료와 타 자원을 연결하여 더욱 연관 있는 자원을 추천받는 사용자 경험을 가질 수 있도록 서비스 제공

○ 추진 내용

- 1. 웹 아카이빙 자료 분석 및 구조화
 - ▶ 자료의 메타데이터 표준화 작업 수행
 - 웹 아카이빙 자료에 대한 메타데이터 구성 요소를 정의하고 표준화하여 정보의 일관성을 확보
 - 다양한 자료유형에 맞는 메타데이터 모델을 설계하여 데이터 검색과 관리의 효율성을 높임
 - ▶ 웹 아카이빙 자료의 분류 및 태깅 시스템 구축
 - 자료의 주제, 형식, 생성 일자 등을 기준으로 분류 체계를 마련
 - 사용자가 쉽게 접근할 수 있도록 태그 시스템을 도입하여 관련 자료 검색을 용이
- 2. 타 자원과의 연계 모델 설계
 - ▶ 국립중앙도서관 자료관리시스템과의 연동
 - 다른 데이터베이스와의 연계를 위해 RESTful API를 설계하고 구현하여 상호 작용 가능
 - 데이터 호출과 전송 방식을 표준화하여 다양한 시스템 간의 원활한 연계
 - ▶ 연관 자료 추천 알고리즘 개발 및 최적화
 - 사용자의 행동 패턴과 선호도를 분석하여 개인화된 추천 알고리즘을 개발
 - 추천 결과의 정확도를 높이기 위해 기계 학습 기법을 적용하여 알고리즘을 지속적으로 개선
- 3. 추천 기술 구현 및 테스트
 - ▶ 사용자 맞춤형 추천 시스템 인터페이스 개발
 - 직관적인 사용자 인터페이스(UI)를 설계하여 사용자가 쉽게 추천 기능을 활용
 - 사용자 경험(UX)을 고려한 피드백 수집 시스템을 도입하여 지속적으로 개선할 수 있는 기

반을 마련

- ▶ 테스트와 피드백을 통해 추천 정확도 개선
 - 프로토타입을 사용자 그룹에 배포하여 피드백을 수집하고 추천 알고리즘을 테스트
 - 테스트 결과를 분석하여 알고리즘을 조정하고 개선하여 정확도를 높임
- 4. 사용자 피드백 기반 시스템 고도화
 - ▶ 사용자 피드백을 수집하여 추천 정확도 조정
 - 사용자가 제공하는 피드백을 기반으로 추천 결과를 정기적으로 업데이트하여 만족도를 높임
 - 피드백 분석 결과를 바탕으로 사용자 맞춤형 설정을 강화
 - ▶ 시스템 성능 모니터링 및 주기적 업데이트 수행
 - 시스템의 성능 지표를 지속적으로 모니터링하여 문제가 발생할 경우 즉시 대응
 - 정기적인 시스템 점검 및 업데이트를 통해 최신 기술과 트렌드를 반영

○ 추진 일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정	타 자원 연계	○	○	○	○	○	
소요 예산		15	15	15	15	15	75

○ 기대 효과

- 웹 아카이빙 자료 활용도 증대 및 사용자 만족도 향상
- 통합된 정보 접근성 제공으로 연구 및 학습 효율성 제고
- 데이터 자원 간의 시너지 효과로 정보 가치 향상

6.1.1

국내 웹 아카이빙 협력 체계 구축 및 공유 확산 [신규사업]

○ 필요성 및 목적

- 국내 웹사이트의 다양한 웹페이지와 웹 콘텐츠를 보존하여 디지털 문화유산의 손실을 방지하고, 보존하는 데 용이
- 사용자와 연구자들이 필요로 하는 웹 자료에 대한 접근성을 높이고, 정보 활용을 촉진하기 위해 협력

체계 구축의 필요성이 대두

- 기관 간의 협력을 통해 웹 아카이빙의 효율성을 높이고, 데이터 공유를 통해 보다 풍부한 정보 생태계 조성을 목적으로 함

○ 추진내용

- 1. 웹 아카이빙 체계 구축
 - ▶ 전담 조직 구성: 전문 조직 구성으로 운영 지원
 - 조직 내 역할 분담: 데이터 수집, 메타데이터 관리, 기술 지원 분담
 - 협력 네트워크 형성: 관련 기관과의 네트워크를 통한 정보 공유
 - ▶ 기술 인프라 개발: 웹 아카이빙을 위한 인프라 구축
 - 자동화 시스템 구축: 효율성을 위한 자료 수집 및 처리 자동화
 - 메타데이터 표준화: 일관성을 위해 국제적 표준 적용
- 2. 교육 및 훈련 프로그램 운영
 - ▶ 전문 인력 양성: 교육 프로그램 운영
 - 커리큘럼 개발: 아카이빙 기술 및 데이터 관리 커리큘럼 개발
 - 실습 기회 제공: 이론과 실무를 결합한 교육 실시
 - ▶ 워크숍 및 세미나 개최: 정보 교류와 네트워킹 촉진
 - 주제별 전문가 초청: 최신 동향 및 사례 공유
 - 참여자 피드백 수집: 지속적인 프로그램 개선
- 3. 지속적인 데이터 유지 및 관리
 - ▶ 정기적인 데이터 업데이트: 최신 정보 반영
 - 보안 관리 체계 수립: 데이터 보안 강화 및 점검
 - 사용자 피드백 반영: 서비스 품질 개선

○ 추진 일정 및 소요 예산

단위 : 백만원

구분		2025	2026	2027	2028	2029	계
추진 일정	협력 체계 구축	0	0	0	0	0	
소요 예산		15	15	15	15	15	75

○ 기대 효과

- 사용자들의 아카이빙된 정보 접근성 용이
- 협력 체계 구축으로 자원 및 정보 공유 확대
- 아카이빙을 통한 지식 축적 및 활용 가능성 증대

7.2 정책 과제 및 실행 과제의 연차별 예산 소요 내역

7.2.1 추진 전략별 소요 예산

○ 오아시스 개선 및 활용성 강화 추진 전략별 소요 예산은 다음과 같음

<표 7-1> 오아시스 개선 및 활용성 강화 추진 전략별 소요 예산

추진 전략		소요 예산(백만원)					
		계	2025	2026	2027	2028	2029
추진 전략 1	안심할 수 있는 데이터 보존과 품질보장(데이터 보존 및 무결)	90	30	30	30	0	0
추진 전략 2	이용자가 원하는 콘텐츠(개인맞춤)	680	120	140	140	140	140
추진 전략 3	시스템 기반 기술 강화 (기술 고도화)	750	50	340	340	20	0
추진 전략 4	언제 어디서나 이용 가능한 오아시스(시스템 접근성)	100	20	20	20	20	20
추진 전략 5	공유 개방화(콘텐츠 제공/전달/이용 체계 확대)	75	15	15	15	15	15
추진 전략 6	지역과 국제기관의 긴밀한 동행	75	15	15	15	15	15
계		1,770	250	560	560	210	190

○ 오아시스 개선 및 활용성 강화 추진 전략별 정책 과제·실행 과제의 소요 예산은 다음과 같음

<표 7-2> 오아시스 개선 및 활용성 강화 추진 전략별 정책 과제·실행 과제의 소요 예산

추진 전략	정책 과제	실행 과제	소요 예산(백만원)					
			계	2025	2026	2027	2028	2029
1. 안심할 수 있는 데이터 보존과 품질 보장 (데이터 보존 및 무결)	정책 과제	1.1 품질 보증(QA) 프로세스 자동화	90	30	30	30	0	0
	실행 과제	1.1.1 WARC 파일 보안 취약점 점검	45	15	15	15	0	0
		1.1.2 지능형 기술을 도입한 누락 웹페이지 자동 점검 프로세스 개발	45	15	15	15	0	0
	계		90	30	30	30	0	0
2. 이용자가 원하는 콘텐츠 (개인 맞춤)	정책 과제	2.1 이용자 특성에 맞춘 콘텐츠 수집 및 제공	680	120	140	140	140	140
	실행 과제	2.1.1 이용자 페르소나 기법을 이용한 디지털 장서 수집	680	120	140	140	140	140
	계		680	120	140	140	140	140
3. 시스템 기반 기술 강화 (기술 고도화)	정책 과제	3.1 개별 시스템 기술 고도화	120	20	40	40	20	0
	실행 과제	3.1.1 수집기 고도화	40	20	20	0	0	0
		3.1.2 색인검색기 고도화	40	0	20	20	0	0
		3.1.3 재생기(웨이백 재생) 고도화	40	0	0	20	20	0
	정책 과제	3.2 시스템 마이그레이션(장기)	630	30	300	300	0	0
	실행 과제	3.2.1 오아시스 시스템 정보화 전략 계획(ISP) 수립(단계적 운영환경 개선)	30	30	0	0	0	0
		3.2.2 오아시스 시스템 고도화 사업	600	0	300	300	0	0
	계		750	50	340	340	20	0
4. 언제 어디서나 이용 가능한 오아시스(시 스템 접근성)	정책 과제	4.1 시스템 접근성 강화	100	20	20	20	20	20
	실행 과제	4.1.1 이용 접근성 강화된 누리집 개선	100	20	20	20	20	20
5. 공유 개방화 (콘텐츠 제공/전달/이 용 체계 확대)	정책 과제	5.1 공유 확산 강화	75	15	15	15	15	15
	실행 과제	5.1.1 웹 아카이빙 내 및 타 자원과의 연계 추천 기술 개발	75	15	15	15	15	15
6. 지역과 국제기관의 긴밀한 동행	정책 과제	6.1 지역 및 국제 협력 체계 구축	75	15	15	15	15	15
	실행 과제	6.1.1 국내 웹 아카이빙 협력 체계 구축 및 공 유 확산	75	15	15	15	15	15
	계		75	15	15	15	15	15
계			1,770	250	560	560	210	190

VIII

결론

8.

결론

8.1 제언

- 지금까지의 오아시스 시스템 개선 및 활용성 강화 방안 연구를 수행하여 다음과 같이 제언함
- 향후 대한민국 웹 아카이빙 오아시스는 ‘국민과 함께하는 디지털 세상으로의 타임캡슐’이라는 비전하에 ‘안심할 수 있는 품질 (데이터 보존 및 무결)’, ‘이용자가 원하는 콘텐츠(개인 맞춤)’, ‘시스템 기반 기술 강화 (기술 고도화)’, ‘모바일화 (시스템 접근성)’, ‘공유 개방화 (콘텐츠 확대)’, ‘지역과 국제기관 의 긴밀한 동행(지역 및 국제 협력)’ 전략을 추구해 나갈 것을 제언함
- 전략을 추진하기 위한 방법론으로 본 연구에서 국제 웹 아카이빙 동향 분석, 시스템 현황을 파악하기 위하여 실무자 의견 조사, 전문가 의견 조사, 시스템 사용성 평가, 성능 진단 등 조사, 분석한 내용을 바탕으로 단기 기술 고도화, 중장기 시스템 개선 방안 및 활용성 강화 방안을 제안하였으며 세부 전략 과제를 도출하였음
- 국제 웹 아카이빙 법적 기술적 동향 및 이용자 요구 반영
 - 국제 우수 웹 아카이빙 운영 국가의 법적 제도적 동향을 고려할 때, 디지털 자원을 적극적 수집하고 있어 우리나라 웹 아카이빙 자원도 적극적 수집이 필수적이고 활용 측면에서 여러 브랜치 도서관 협약에 의한 웹자원의 상호 대차를 통해 현장 열람의 범위를 확대할 필요가 있음
 - 웹 아카이빙 기술적 동향을 감안하고 변화하는 웹 기술에 의한 웹자원을 안정적으로 수집, 보존하고 지속 가능한 웹 수집 도구의 활용을 위해서는 최신 버전의 Heritrix나 Pywb와 같은 최신 수집기를 사용할 필요가 있음
 - 우수 기관의 운영 사례는 웹자원의 활용을 위해 이용자 접근성을 확대하는 방향으로 데이터 서비스를 제공하고 지속적으로 개선하고 있으며, 모바일 기기에 맞춘 UI/UX 제공과 Share 기능, 추천 기능을 제공하고 있으므로 현재 오아시스 누리집도 이러한 이용자의 웹자원 접근성과 활용을 확대하는 기능을 포함할 필요가 있음
- 기술 고도화 방안(단기)
 - 수집기 관련하여서는, 현재 수집기의 최신 버전을 사용하는 방안(1안)과 웹 아카이빙 선진 운영 기관에서 사용하는 수집기로의 전환(2안)과 한국 웹페이지의 특성과 동적 웹 기술을 고려하여 다른 웹 아카이빙 기관에서는 사용하지 않으나 최신 동적 웹페이지 수집 기술을 도입하는 방안(3안)을 제시하였음

- ▶ 1안: 현재 오아시스가 사용 중인 기존 웹 수집기 Heritrix의 최신 버전 활용 및 개선 가능성 파악
- ▶ 2안: 현재 오아시스가 사용 중인 Heritrix 외에 Browsertrix-crawler, Pywb-crawler 등 IIPC 회원국에서 사용하는 기존 웹 수집기 활용 가능성 파악
- ▶ 3안: 최신 동적 웹 수집 기술인 selenium을 이용한 웹 수집 기술 조사
- 색인 검색기 관련하여서는 현재 사용 중인 SolrWayback 도구를 그대로 사용하는 방안과 최신 기술 중 Elastic Search 도구를 도입하는 방안을 제안함
 - ▶ SolrWayback 도구의 경우 현재 색인 검색기의 문제점을 유형화하고 해결 방안을 찾아서 개선하는 과정을 거쳐야 함
 - ▶ Elastic Search 도구는 현재 많은 텍스트 색인 검색에 사용되는 도구로서 빠른 삽입 제거를 할 수 있으며 분석 처리도 가능하여 확장성이 뛰어나
- 재생기 관련하여서는 수집기와 유사하여 현재 사용 중인 전통적 웨이백 재생 도구 OpenWayback을 Pywb으로 변경할 것을 제안함. 또한, 선진 운영 기관들이 이전에서 채택하는 도구 셋인 OutbackCDX 서버를 연동할 것을 제안하였음
- 데이터 서비스 기술 관련하여서는, 웹 아카이빙 자료 분석 및 구조화하고 타 자원과의 연계 모델 설계, 추천 기술 구현 및 테스트 및 사용자 피드백 기반 시스템 고도화를 통하여 데이터 후보 서비스를 도출할 것을 제안하였고 후보 데이터 분석 서비스의 예시를 제공하였음
- 데이터 품질 보증(QA: Quality Assurance) 기술 관련하여서는, 현재 지능형 자동화 기술(RPA)을 활용한 품질 보증 프로세스를 수립하기 위하여, 인공지능 기반 개인 정보 보호 및 유해 콘텐츠 식별 기술 개발과 지능형 기술을 이용한 누락 웹페이지 자동 점검 프로세스를 개발할 것을 제안하였음

○ 시스템 개선 방안(중장기)

- 개선 방안(1안): 웹 데이터 수집·보존 개선을 위한 단기 기술 고도화
 - ▶ 현행 시스템 평가 결과에 따라 현행 시스템을 개선 운영하는 방안을 제안함
 - ▶ 기존 수집·보존 활용 프로세스를 그대로 유지하면서 즉시 개선 가능한 개선점 위주로 개선하고 지속적으로 유지·보수하는 접근임
 - ▶ 앞에서 언급한 단기 기술 고도화를 통해 실현 가능함. 하지만, 다른 해외의 선진 웹 아카이빙 사례를 고려할 때, 장기적인 측면에서는 한계가 있는 방안임
 - ▶ 선진 웹 아카이빙 기관 개선 사례(5.2.1 참조)에서 볼 수 있듯이, 장기적인 측면에서, 웹 아카이브 이전(migration)을 함께 진행해야 함
- 개선 방안(2안): 새로운 경량 구조(마이크로 서비스) 도입 제안
 - ▶ 새로운 시스템 구조(예, 클라우드, 마이크로 서비스) 도입 제안
 - ▶ 본 안은 시스템 장기 개선 방안의 다른 하나로 기존 웹 아카이브 액세스 시스템과 새로운 경량 구조 마이크로 서비스 병행하는 시스템 구조를 제안

- 개선 방안(3안): 새로운 웹 아카이브 시스템 구조로의 전면적인 마이그레이션(안)
 - ▶ 본 안은 시스템 장기 개선 방안 중 또 다른 하나로 새로운 마이크로 서비스 시스템 구조로의 전면적인 마이그레이션을 제안
 - ▶ 본 방안 채택 시, ISP, ISMP 진행을 통해 개선 시스템으로의 면밀한 이전 계획 수립이 추가로 필요함

○ 협력 및 홍보 방안

- 우리나라 관련 웹자원의 수집·보존 및 공유 강화를 위한 국제 협력 방안과 대국민 이용 활성화를 위한 오아시스 홍보 방안 제안을 각각 제안하였음
- 우리나라 관련 웹자원의 수집·보존 및 공유 강화를 위한 국제 협력 방안
 - ▶ 오아시스 및 국내 웹 아카이브 체계 구축
 - ▶ IIPC의 회원국 간의 긴밀한 협력
 - ▶ 지역 및 민간 웹 아카이빙 교육 및 훈련 프로그램 운영
- 대국민 이용 활성화를 위한 오아시스 홍보 방안 제안
 - ▶ SNS나 유튜브 채널과 같은 디지털 마케팅 채널을 통한 오아시스 소개 및 활용 콘텐츠 제공
 - ▶ 대학 및 학회와 연계하여 OASIS 활용 강좌 및 워크숍 진행
 - ▶ OASIS 전시회 개최: 전시회 개최를 통한 OASIS 홍보 및 체험 행사 개최
 - ▶ 이용 후기 공모전: OASIS 이용 후기를 공유하는 공모전 개최

참고 문헌

- 국립중앙도서관 도서관연구소 (2008). 주요국 도서관법령집(14. 캐나다도서관·기록관법), 도서관연구소 자료집; 4, pp. 187-204
<https://oak.go.kr/nl-ir/bitstream/2020.oak/394/1/%EC%A3%BC%EC%9A%94%EA%B5%AD%EB%8F%84%EC%84%9C%EA%B4%80%EB%B2%95%EB%A0%B9%EC%A7%91-1.pdf>
- Alam, S., Garg, K., Weigle, Michele C., Nelson, Michael L., Graham, M. & Ayala D. (2024). TrendMachine: A Temporal Webpage Resilience Portal. 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL), 93-97p. <https://doi.org/10.1109/JCDL57899.2023.00023>
- Alam, S., Graham, M., Macdonald, R. & Leetaru, K. (24.04.24-26). Utilizing Large Language Models for Semantic Search and Summarization of International Television News Archives. IIPC GENERAL ASSEMBLY & WEB ARCHIVING CONFERENCE. NATIONAL LIBRARY OF FRANCE. <https://netpreserve.org/ga2024/abstracts/>
- Peter Stirling, Gildas Illien, Pascal Sanz, Sophie Sepetjan. (2011). The state of e-legal deposit in France: looking back at five years of putting new legislation into practice and envisioning the future. <http://conference.ifla.org/past/ifla77/193-stirling-en.pdf>
- Potter, A. (2012). A Vision of the Role and Future of Web Archives: Conclusions and the Role of Archives. Library of Congress Blogs.
<https://blogs.loc.gov/thesignal/2012/05/a-vision-of-the-role-and-future-of-web-archives-conclusions-and-the-role-of-archives/>

참고 웹사이트

- 국립중앙도서관 OASIS. [online] [cited 2024.06.12] <https://www.nl.go.kr/oasis>
- 국립중앙도서관 OASIS. 3.1운동 및 임시정부100주년 기념. [online] [cited 2024.12.02.]
https://www.nl.go.kr/oasis/contents/O2010000.do?page=1&pageUnit=12&schM=search_list&schType=theme&schStr=3.1&schIsFa=ISU-000000000361
- 국립중앙도서관 OASIS. 웹트렌드 분석. [online] [cited 2024.12.02.]
<https://www.nl.go.kr/oasis/contents/O5010000.do>
- 국립중앙도서관 OASIS. 재난아카이브. [online] [cited 2024.12.02.]
<https://www.nl.go.kr/oasis/contents/O4010000.do>
- 국립중앙도서관 OASIS. 초성보기. [online] [cited 2024.12.02.]
<https://www.nl.go.kr/oasis/contents/O1030000.do>
- 국립중앙도서관 OASIS. 컬렉션-주제·이슈. [online] [cited 2024.12.02.]
<https://www.nl.go.kr/oasis/contents/O2010000.do>
- 국립중앙도서관 OASIS. 통합검색. [online] [cited 2024.12.02.]

<https://www.nl.go.kr/oasis/contents/O1010000.do>

국립중앙도서관 OASIS. 통합검색-월드컵 검색. [online] [cited 2024.12.02.]

<https://www.nl.go.kr/oasis/contents/O1010000.do?schM=list&schType=simple&historyYn=Y&schSiteNm=%EC%9B%94%EB%93%9C%EC%BB%B5>

국립중앙도서관 OASIS. KDC 주제별 보기. [online] [cited 2024.12.02.]

<https://www.nl.go.kr/oasis/contents/O1020000.do>

국제인터넷보존컨소시엄(IIPC). 웹 아카이빙 소프트웨어 및 도구 관련 활용 현황 설문조사 (2021). [online] [cited 2024.12.02.]

<https://docs.google.com/spreadsheets/d/1Ww5bhd7KbAH8QOI2qDhdakWh5aExN9JkeL431fhE0Vg/edit#gid=266509536>

㈜태창산업. [online] [cited 2024.12.02] <https://www.needles.co.kr>

한남대학교. [online] [cited 2024.12.02] <https://www.hannam.ac.kr>

British Library. Everything Forever-marking 10 years of digital legal deposit. (2023.04.06.) [online] [cited 2024.12.02.]

<https://blogs.bl.uk/living-knowledge/2023/04/everything-forever-marking-10-years-of-digital-legal-deposit.html>

Browsertrix. [online] [cited 2024.12.02] <https://crawler.docs.browsertrix.com/user-guide/>

Claire Newing. The National Archives (UK) and Phil Clegg, MirrorWeb. 영국 정부 소셜미디어 아카이브를 더욱 개선하기. [online] [cited 2024.11.20.]

<https://netpreserveblog.wordpress.com/tag/elasticsearch/>

DB-Engine. DB-Engines Ranking of Search Engines. [online] [cited 2024.11.20.]

<https://db-engines.com/en/ranking/search+engine>

Graphs [online] [cited 2024.12.02] <https://labs.statsbiblioteket.dk/linkgraph/>

Government of Canada. Justice Laws Website. 지적재산권법(R.S.C., 1985, c. C-42) [online] [cited 2024.06.22.] <https://laws-lois.justice.gc.ca/eng/acts/L-7.7/index.html>

Government of Canada. Justice Laws Website. 캐나다도서관 · 기록관법(S.C. 2004. c.11) [online] [cited 2024.06.22.] <https://laws-lois.justice.gc.ca/eng/acts/L-7.7/index.html>

GTmetrix. Everything You Need to Know About GTmetrix Reports. GTmetrix Performance Blog. [online] [cited 2024.06.22.]

<https://gtmetrix.com/blog/everything-you-need-to-know-about-the-new-gtmetrix-report-powered-by-lighthouse/>

HTTrack [online] [cited 2024.12.02] <https://www.httrack.com/>

KB national library of the netherlands. Legal aspects in web archiving. [online] [cited 2024.12.02] <https://www.kb.nl/en/over-ons/expertises/webarchivering/juridische-aspecten>

László Tóth. Migrating to pywb at the National Library of Luxembourg. [online] [cited 2024.06.22.]

<https://netpreserveblog.wordpress.com/2024/01/22/migrating-to-pywb-at-the-national-library-of-luxembourg/>

Library of Congress. Collection Intergovernmental Organizations Web Archive-About this

Collection. [online] [cited 2024.12.02.]

<https://www.loc.gov/collections/intergovernmental-organizations-web-archive/about-this-collection/>

Library of Congress. Collection Intergovernmental Organizations Web Archive-Collection Items List. [online] [cited 2024.12.02.]

<https://www.loc.gov/collections/intergovernmental-organizations-web-archive/?st=list>

Library of Congress. Collection Intergovernmental Organizations Web Archive-Collection Items Siledshow. [online] [cited 2024.12.02.]

<https://www.loc.gov/collections/intergovernmental-organizations-web-archive/?st=slideshow#slide-3>

Library of Congress. COMESA: Common Market for Eastern and Southern Africa. [online] [cited 2024.12.02] <https://www.loc.gov/item/lcwaN0003866/>

Library of Congress. MODS general web archives. Library of Congress Web Archive. [online] [cited 2024.12.02.]

https://tile.loc.gov/storage-services/service/webcapture/project_1/mods/general-web-archives/lcwaN0003866.xml

National edeposit. [online] [cited 2024.12.02] <https://ned.gov.au/ned/>

NATIONAL LIBRARY OF AUSTRALIA. Legal deposit. [online] [cited 2024.12.02.]

<https://www.nla.gov.au/using-library/services-publishers/legal-deposit>

NATIONAL LIBRARY OF AUSTRALIA. Legal deposit frequently asked questions. [online] [cited 2024.12.02.]

<https://www.nla.gov.au/using-library/services-publishers/legal-deposit/legal-deposit-frequently-asked-questions#websites>

NATIONAL LIBRARY OF NEW ZEALAND. What's legal deposit? [online] [cited 2024.06.22.]

<https://natlib.govt.nz/publishers-and-authors/legal-deposit/whats-legal-deposit>

République française. Légifrance. TITRE III : DÉPÔT LÉGAL (Articles R131-1 à R133-1-1) [online] [cited 2024. 12. 23.]

https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006074236/LEGISCTA000024240034/#LEGISCTA000024240034

Selenium. [online] [cited 2024.12.02.] <https://www.selenium.dev/>

Tracee Haupt. Improvements Ahead for the Web Archives. [online] [cited 2024.6.10.]

<https://blogs.loc.gov/thesignal/2023/08/improvements-ahead-for-the-web-archives/>

Wikipedia. List of Web archiving initiatives. [online] [cited 2024.12.02]

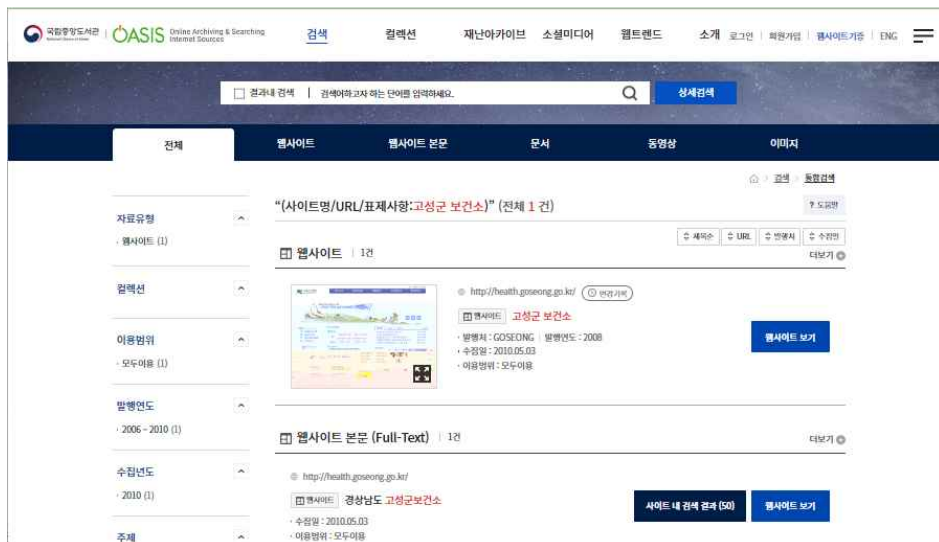
https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives

WORLD TRADE ORGANIZATION. TRIPS - Trade-Related Aspects of Intellectual Property Rights. [online] [cited 2024.12.02] https://www.wto.org/english/tratop_e/trips_e/trips_e.htm

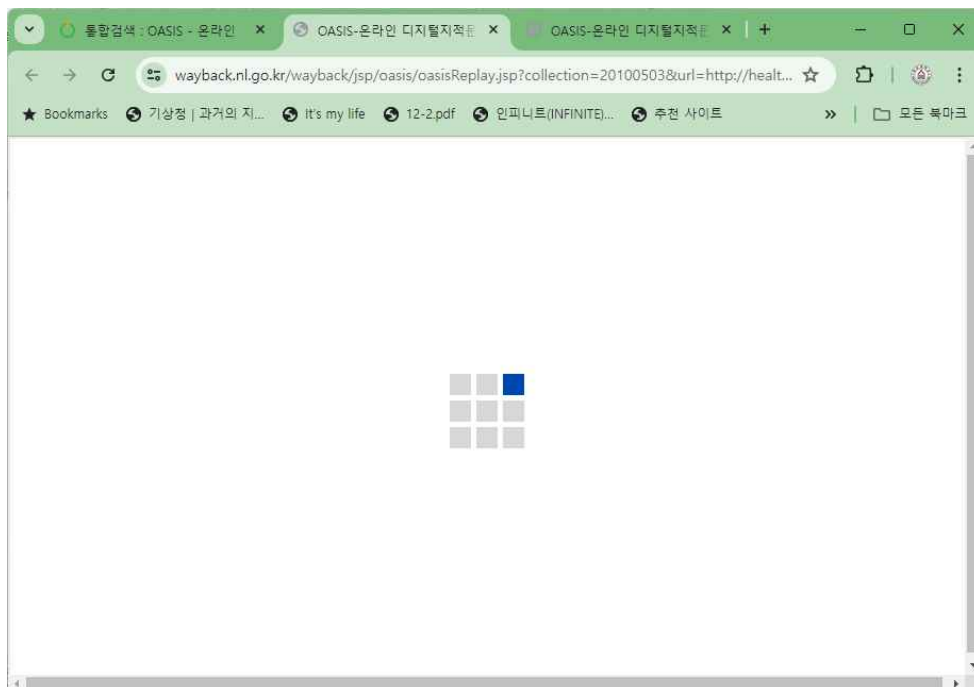
부록A

오아시스 누리집 사용성 평가 결과 화면 모음

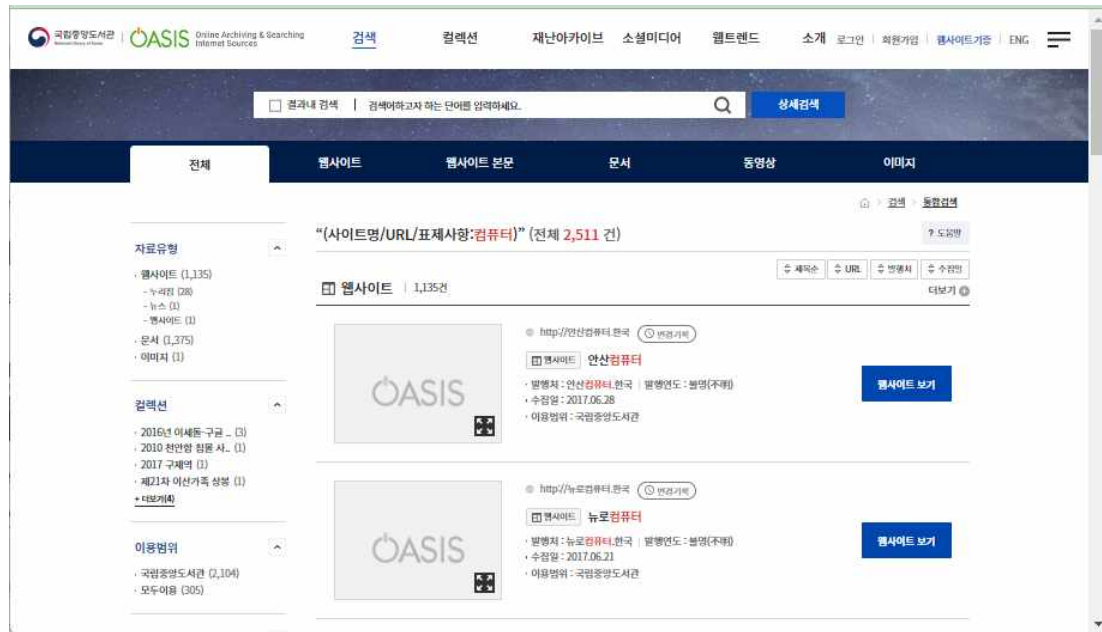
- 본 부록은 본 보고서 3.4 오아시스 누리집 사용성 평가 결과 제시된 문제점에 대하여 실제 오아시스 누리집에서 그 문제점에 대해 확인하고 캡처한 결과 화면임
- 이는 오아시스 누리집 사용성 평가에 대한 신뢰성을 제공하고자 근거로 제시하였음



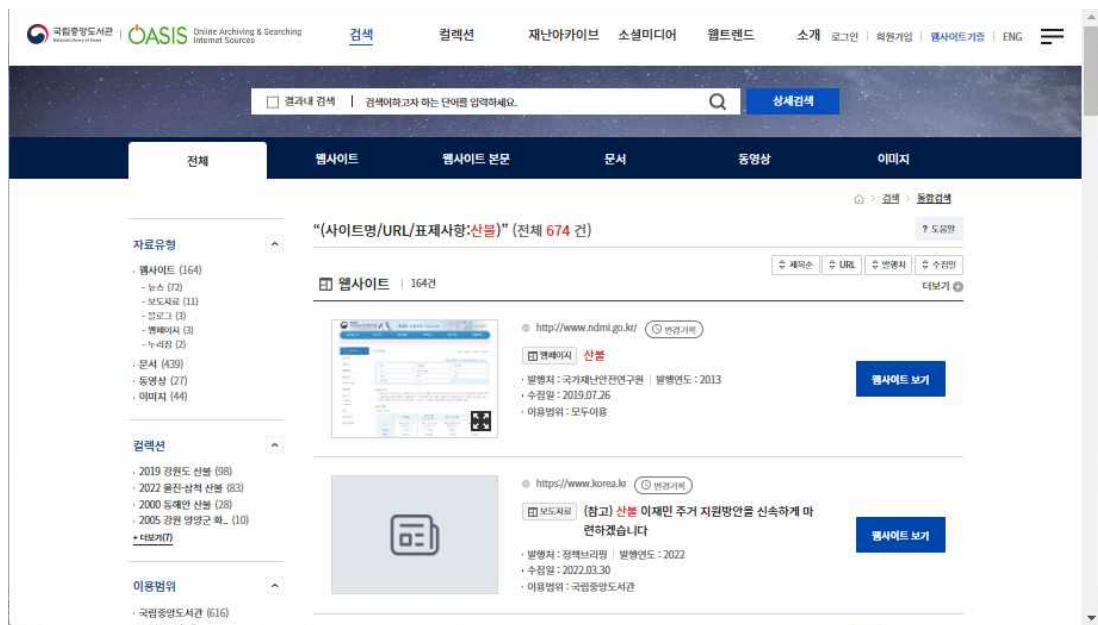
<그림 부록A-1> 웹사이트 검색 화면



<그림 부록A-2> 웨이백 재생 무한 로딩



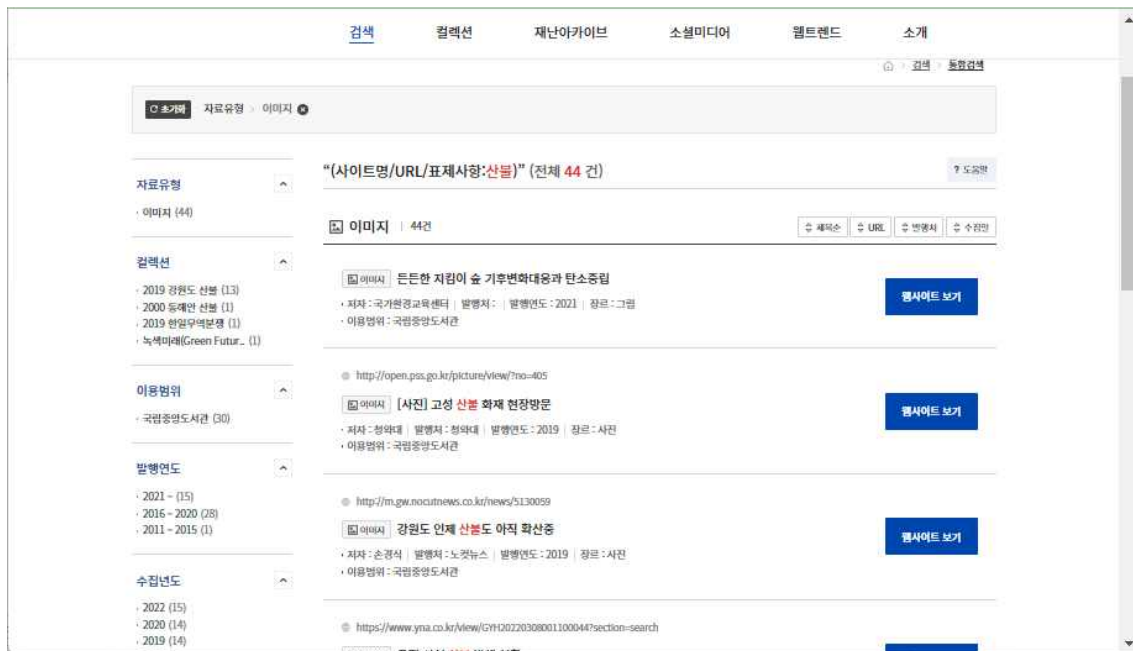
<그림 부록A-3> 썸네일 영상이 없을 때 표시되는 디폴트 이미지(OASIS 로고)



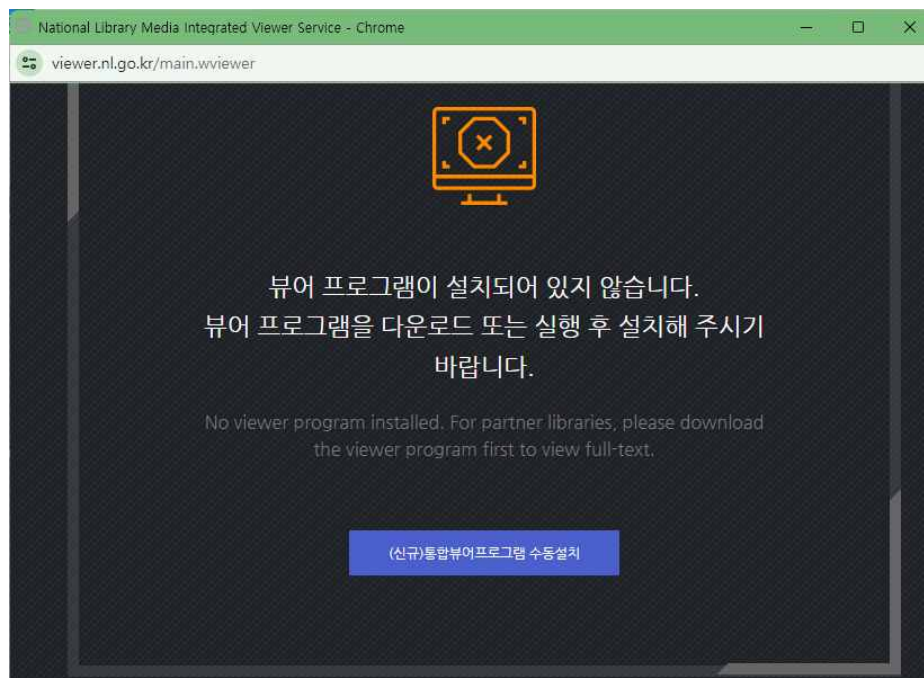
<그림 부록A-4> 썸네일 영상이 없을 때 표시되는 디폴트 이미지(문서 이미지)



<그림 부록A-5> 미리보기 영상이 흰 바탕의 영상과 오아시스 영상이 통일되지 않음



<그림 부록A-6> 미리보기가 제시되지 않은 이미지 결과 제시 화면



<그림 부록A-7> 전용 뷰어 프로그램 설치 안내 화면



<그림 부록A-8> 동영상에 대해 ‘웹사이트 보기’ 버튼



<그림 부록A-9> 국립중앙도서관 내 자료 이용 가능 안내 메시지



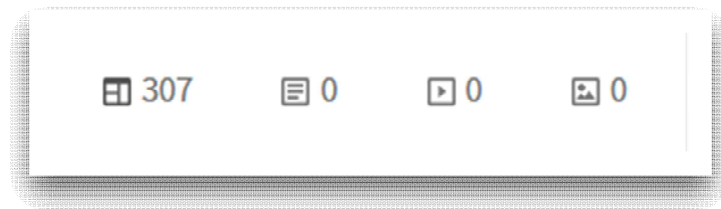
<그림 부록A-10> 보존된 웹페이지의 이미지 로딩을 실패한 웨이백 재생 화면

2022 카타르 월드컵

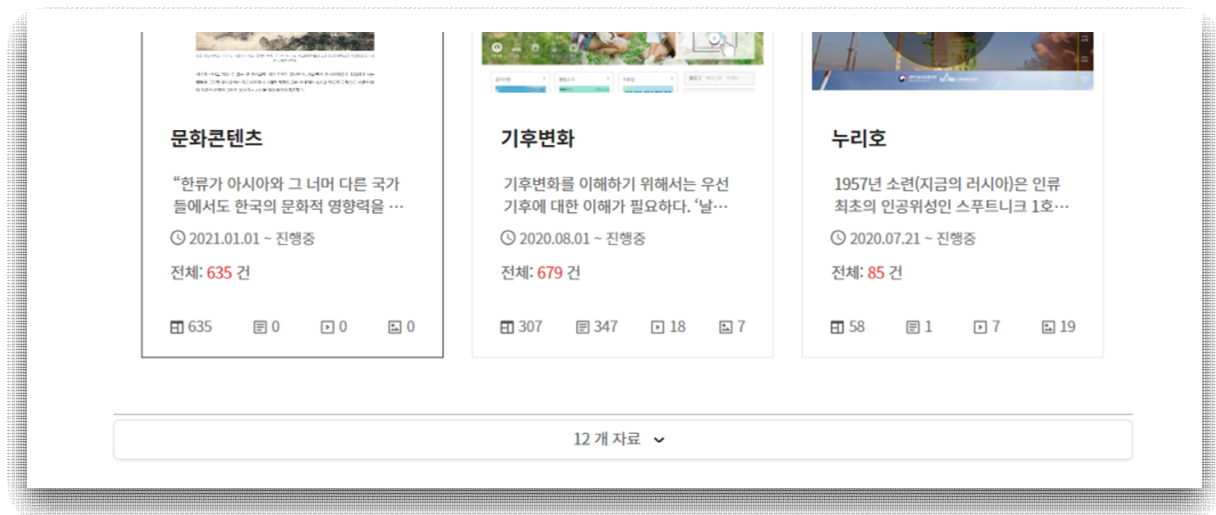


- 발생기간 2022년 11월 20일 ~ 2022년 12월 18일
- 수집기간 2022년 11월 15일 ~ 2022년 12월 29일
- 관련주제 예술 > 오락, 스포츠
- 수집건수 307 0 0 0

<그림 부록A-11> 컬렉션 > 2022 카타르월드컵



<그림 부록A-12> 수집 건수 (확대영상) 자료유형별
아이콘이 너무 작아 구분이 잘 안 됨



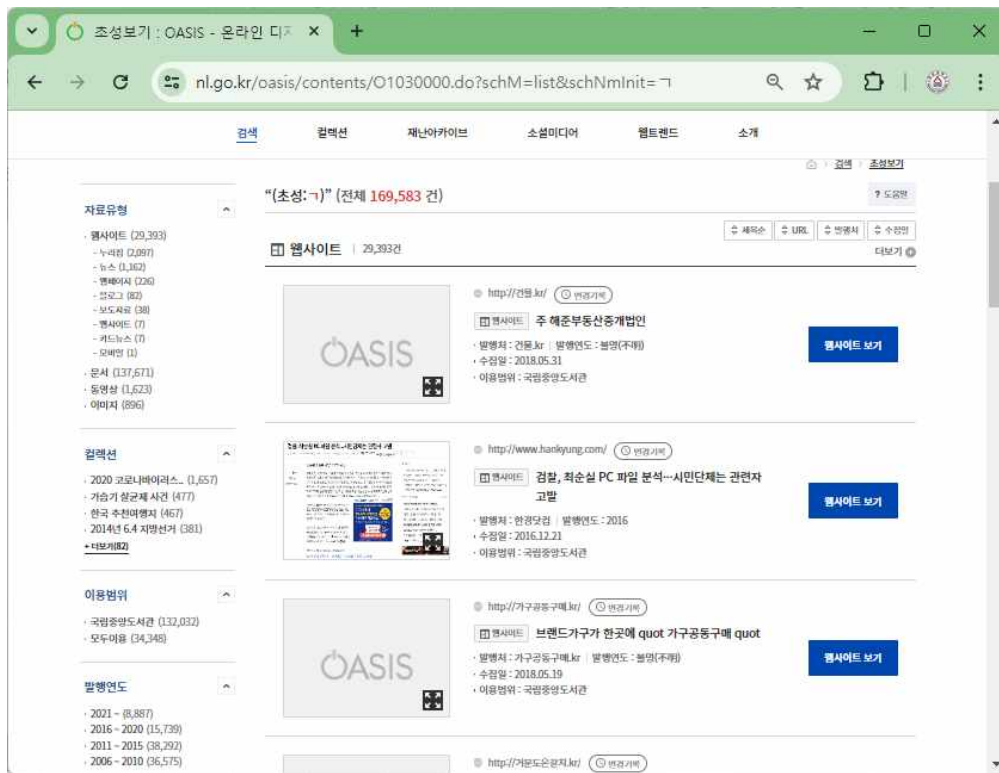
<그림 부록A-13> 보기 형식 중 12개 자료 보기 화면

<http://www.pcj2003.com/> 변경기록

웹사이트 제17대 국회의원선거 박찬종

- 발행처 : 박찬종 | 발행연도 : 불명(不明)
- 수집일 : 2004.04.20
- 이용범위 : 국립중앙도서관

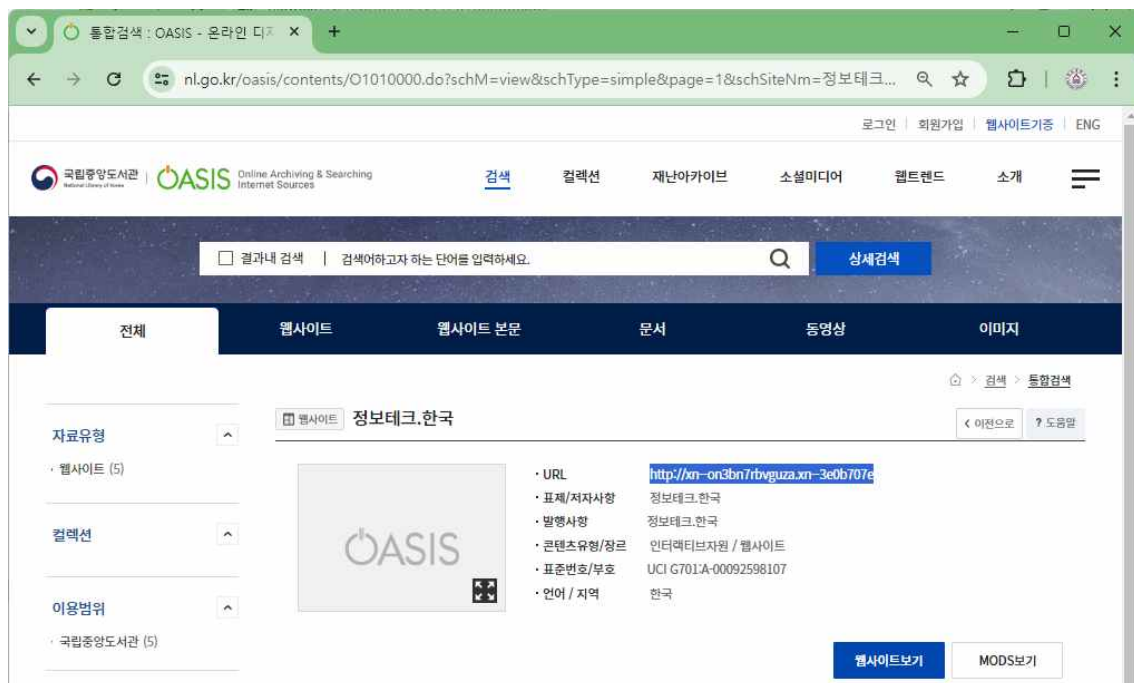
<그림 부록A-14> 링크 표시가 안 되어있는 웹사이트
URL



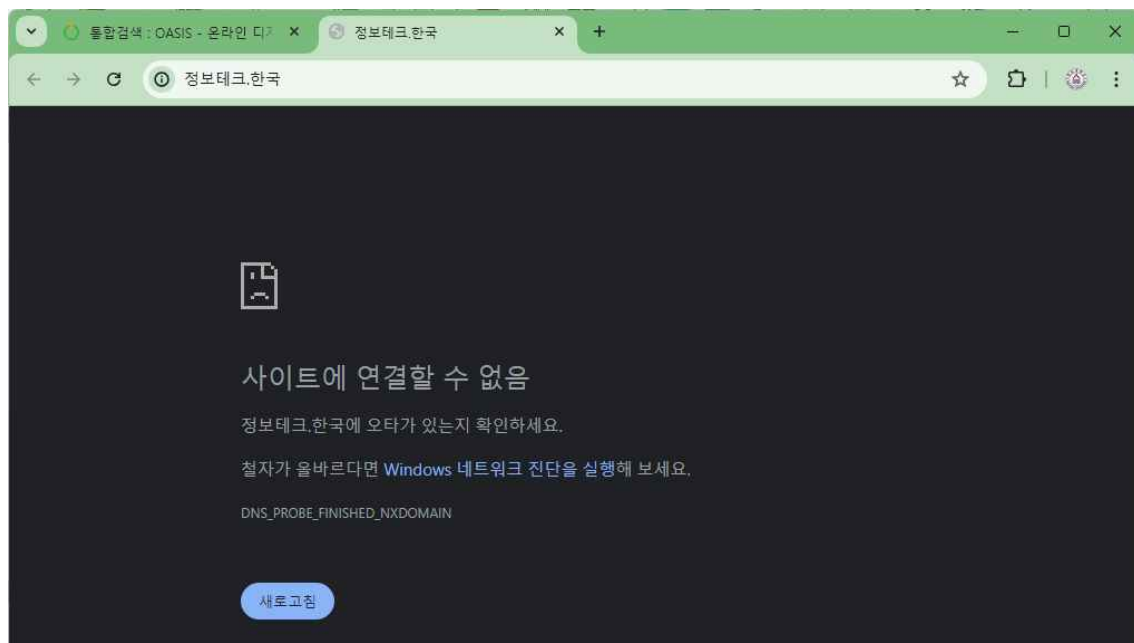
<그림 부록A-15> 검색 > 초성보기 > ‘ㄱ’으로 브라우징한 결과 화면



<그림 부록A-16> 검색 > KDC주제별보기> 철학에서 잡지, 사전은 아카이브 자료가 소수(0건)임



<그림 부록A-17> 아카이빙된 정보테크 웹사이트 정보



<그림 부록A-18> URL(정보테크.한국) 링크가 깨어진 웹사이트



needles.co.kr  변경기록

 웹사이트 Error

- 발행처 : Tae chang | 발행연도 : 2017
- 수집일 : 2024.04.24
- 이용범위 : 국립중앙도서관

[웹사이트 보기](#)

<그림 부록A-19> 웹사이트가 Error로 나오는 needles.co.kr 웹사이트

전체

웹사이트

웹사이트 본문

문서

동영상

자료유형

웹사이트 (1,446)

누리집 (72)

웹페이지 (3)

뉴스 (1)

웹사이트 (1)

컬렉션

국립중앙도서관 (1,158)

모두이용 (83)

이용범위

국립중앙도서관 (1,158)

모두이용 (83)

발행연도

“(사이트명/URL/표제사항:월드컵)” (전체 1,446 건)

웹사이트

1,446건

http://joomal.com



변경기록

웹사이트

주말

발행처 : JOOMAL.com | 발행연도 : 불명(不明)

수집일 : 2017.05.29

이용범위 : 국립중앙도서관

http://www.seoulworldcupst.or.kr/



변경기록

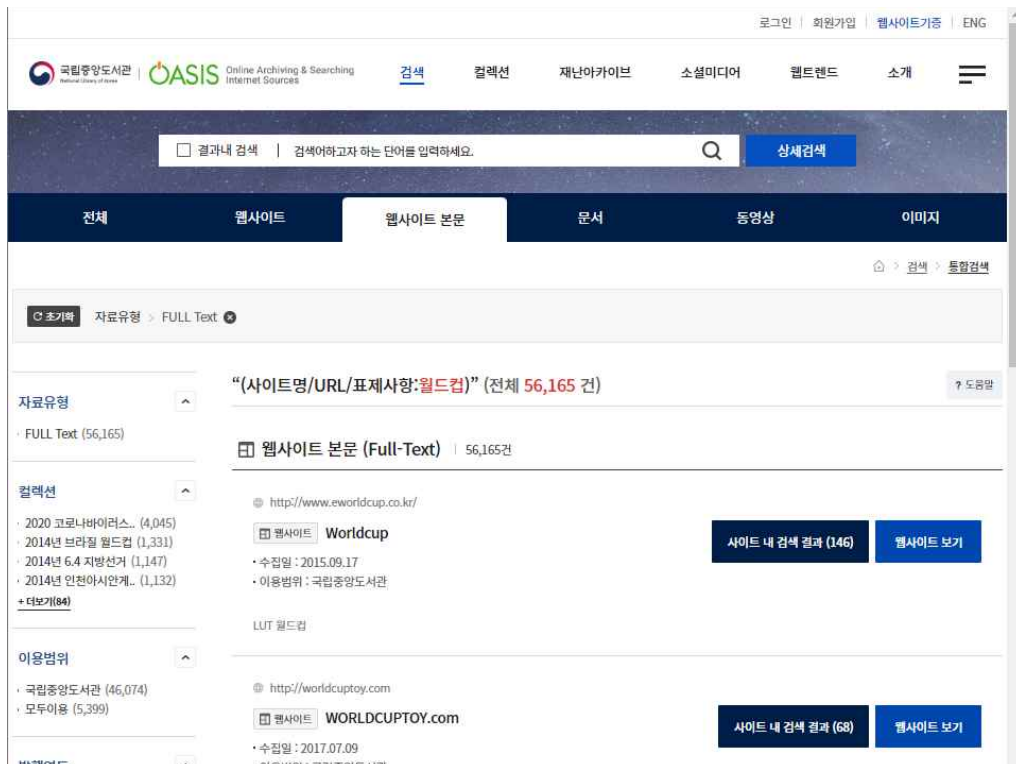
웹사이트

서울월드컵경기장

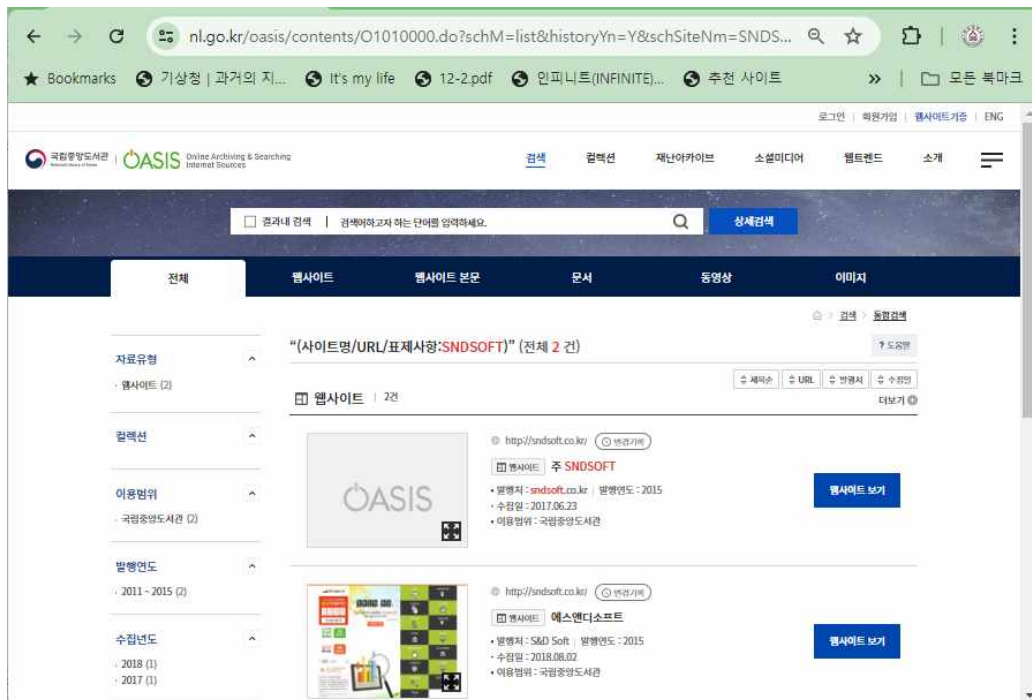
발행처 : seoulworldcupst.or.kr | 발행연도 : 2001

<그림 부록A-20> ‘월드컵’ 검색 결과 자료유형으로 웹사이트, 누리집, 웹페이지, 뉴스, 웹사이트로 구분되어 있음

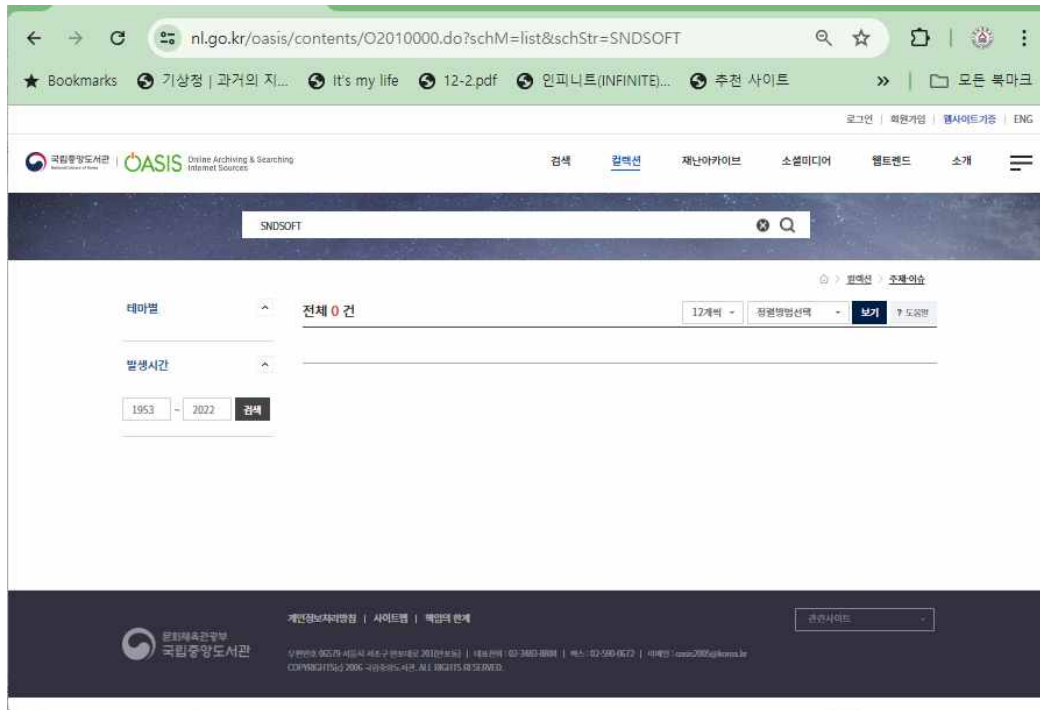
- 215 -



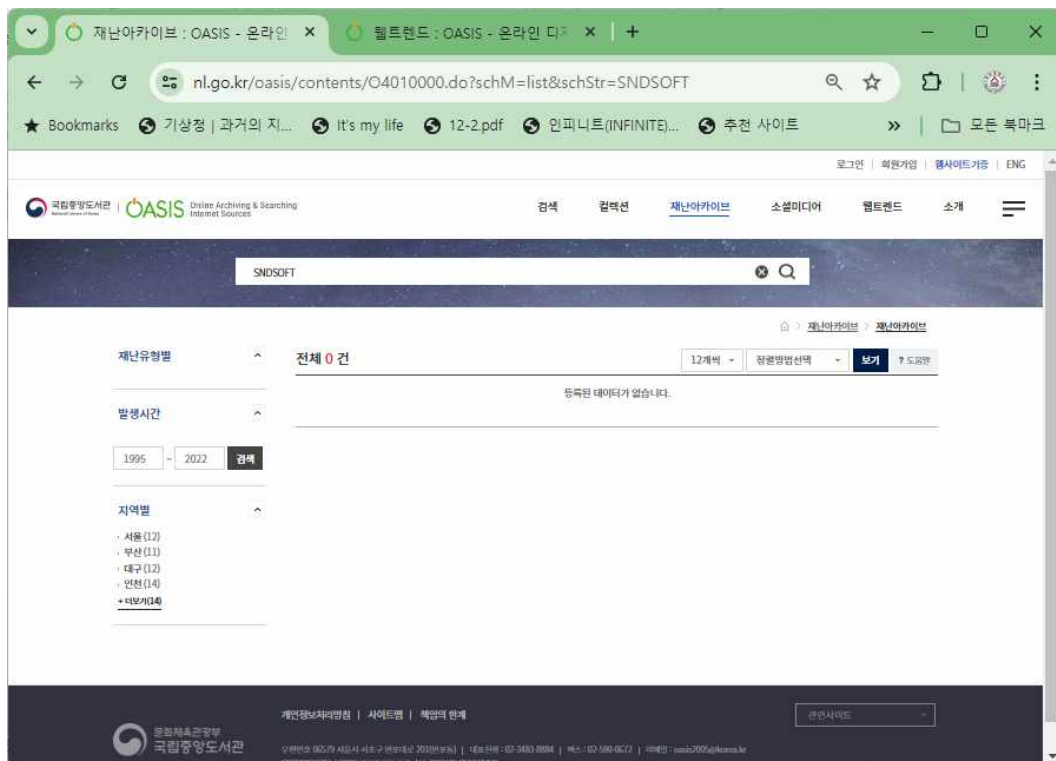
<그림 부록A-21> ‘월드컵’ 검색 결과 자료유형은 웹사이트, 누리집, 웹페이지, 뉴스, 웹사이트로 구분되지 않고 full-text로만 제시되고 있음



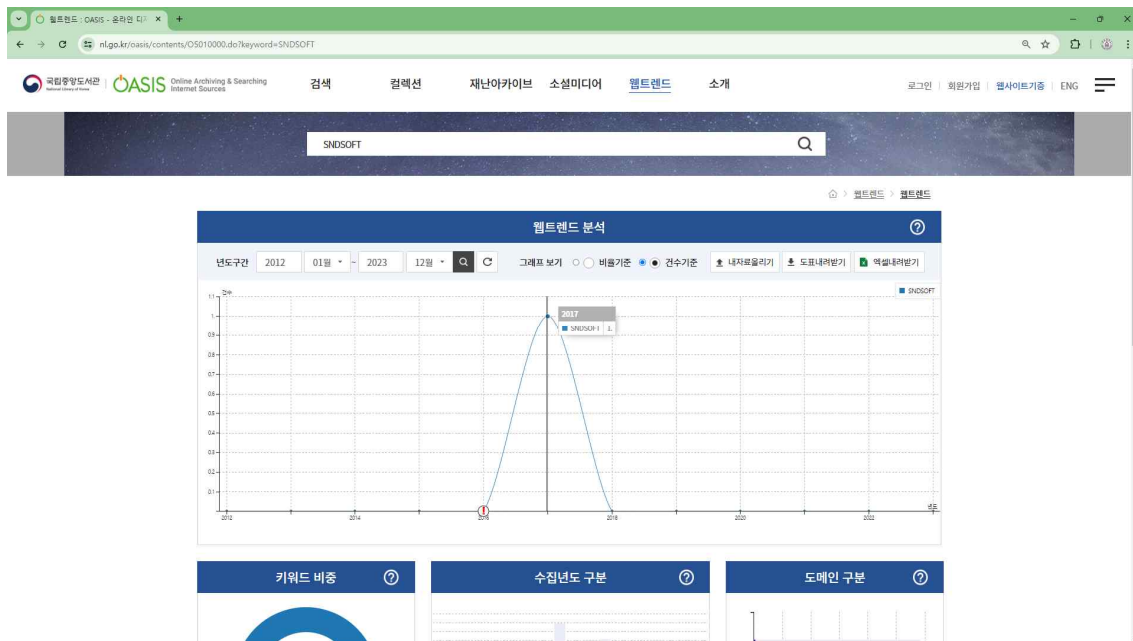
<그림 부록A-22> ‘SNDISOFT’ 검색 결과 자료유형은 웹사이트(2)로 검색된 결과



<그림 부록A-23> 컬렉션에 대해 ‘SNSOFT’ 검색 실행, 전체 0건 검색된 결과



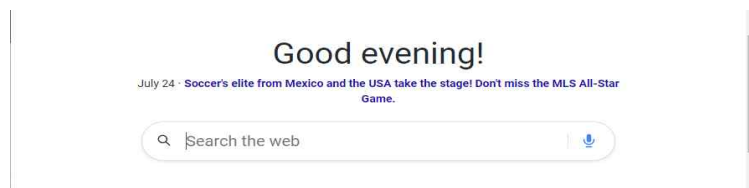
<그림 부록A-24> 재난아카이브에 대해 ‘SNSOFT’ 검색 실행, 전체 0건 검색된 결과



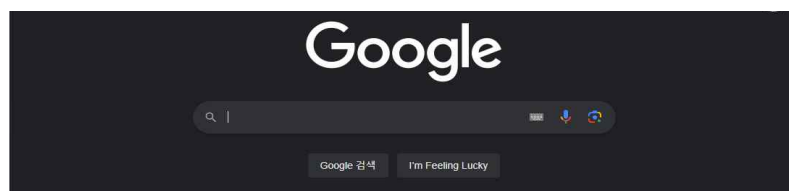
<그림 부록A-25> 웹트렌드 분석에 대해 ‘SNSD’ 검색 실행, 전체 1건 검색된 결과



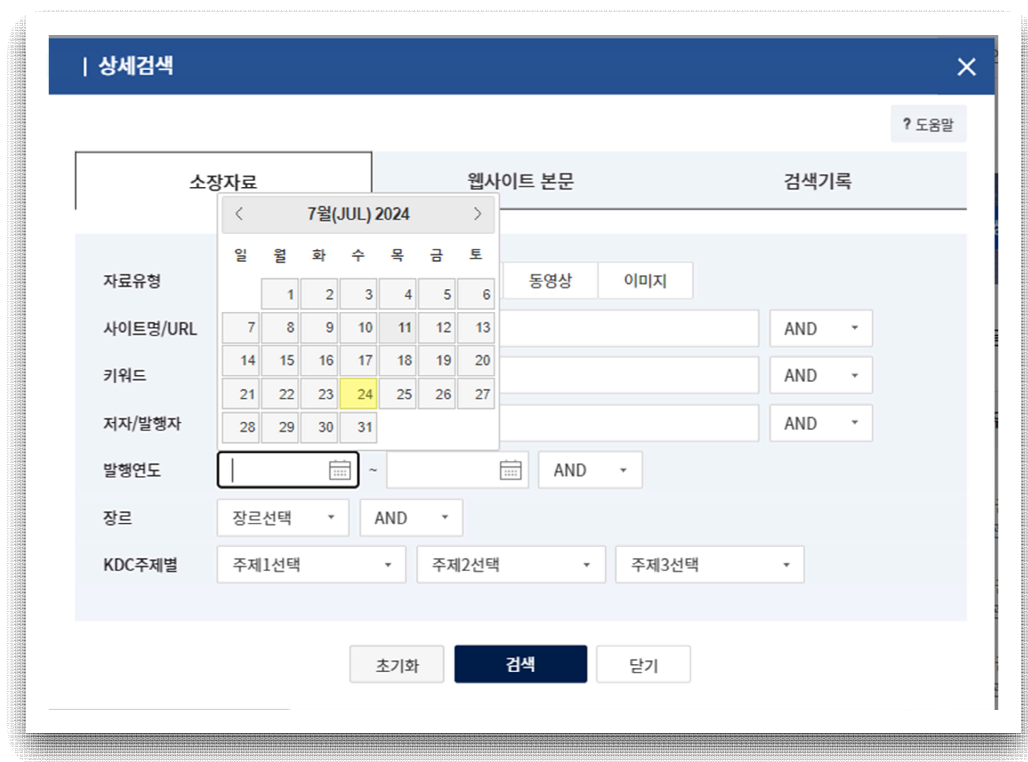
<그림 부록A-26> ‘검색’ 메뉴에서 작은 돋보기 아이콘과 비교되게 넓은 면적의 상세검색 버튼



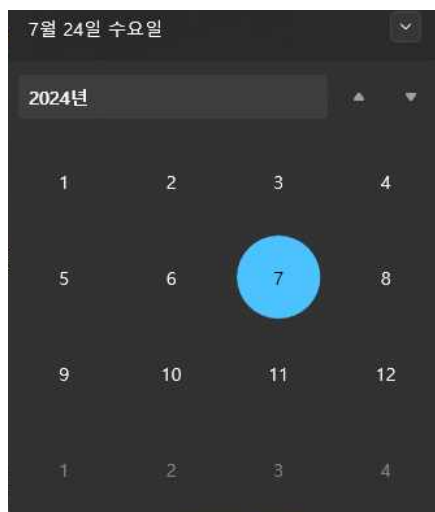
<그림 부록A-27> 야후 검색 창의 경우 돋보기는 검색을 의미하는 아이콘이고 버튼 없이 엔터로 검색 실행



<그림 부록A-28> 야후와 유사하게 돋보기 아이콘은 앞쪽에 검색 레이블 대신 사용되었고, 검색 실행은 엔터(enter)나 아래쪽 ‘Google 검색’ 버튼을 눌러 실행



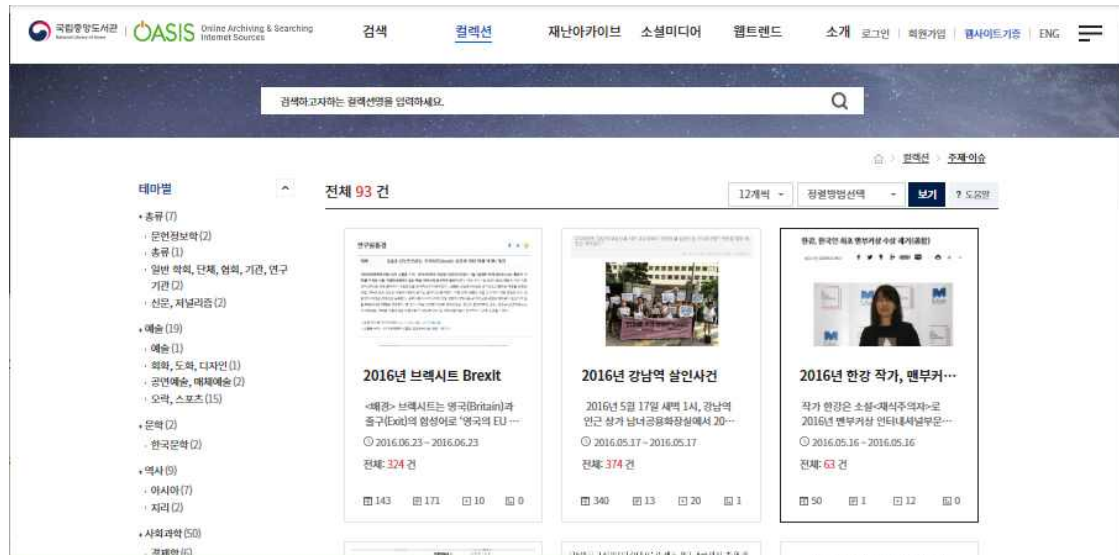
<그림 부록A-29> 상세검색에서 발행연도 날짜 선택 컨트롤



<그림 부록A-30> Windows11 날짜 선택 컨트롤(월 선택)



<그림 부록A-31> Windows11 날짜 선택 컨트롤(연도 선택)



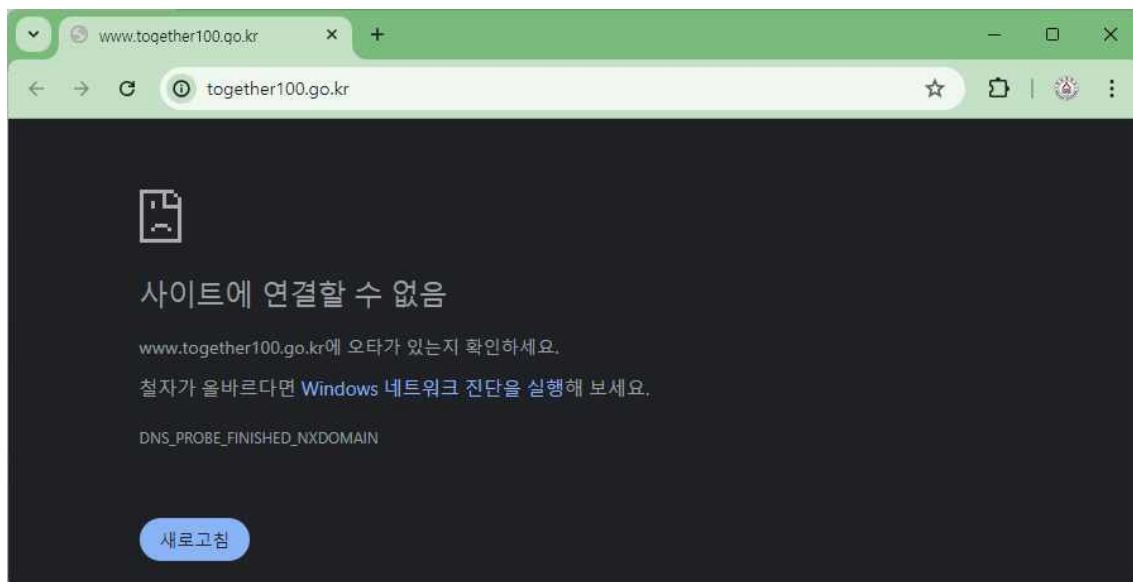
<그림 부록A-32> 12개씩 보기 화면



<그림 부록A-33> 24개씩 보기 화면



<그림 부록A-34> 컬렉션 > 3.1운동 및 임시정부 100주년 기념 전체보기 화면 및 출처 (<https://www.together100.go.kr>)



<그림 부록A-35> <https://www.together100.go.kr> 접속 오류 화면



어린이날 100주년

어린이날은 어린이의 존엄성과 지위 향상을 위하여 어린이들이 올바...

🕒 2022.05.05 ~ 2022.05.05

전체: 66 건

어린이날 100주년

어린이날은 어린이의 존엄성과 지위 향상을 위하여 어린이들이 올바...

🕒 2022.05.05 ~ 2022.05.05

전체: 66 건

<그림 부록A-36> 주제별 컬렉션 선택 시 원본 이미지 로딩 시간과 같은 미리보기 이미지 로딩 시간

테마별

예수 (1)

공연예술, 매체예술 (1)

발생시간

2019

~

2021

검색

한국영화

발생기간

2019년 05월 21일

수집기간

2019년 05월 29일 ~ 2019년 08월 26일

관련주제

예술 > 공연예술, 매체예술

수집건수

230 0 8 66

1919년 10월 27일, 우리나라 최초의 상설 영화관인 단성사에서 우리 자본으로 제작된 최초의 연쇄극 <의리적 구토> 1962년 10월 27일이 '영화의 날'로 제정되었으며, 2019년은 그로부터 100년을 맞이하는 뜻깊은 해이다. 한편 이를 기념하여 2019년 5월 21일 72회 칸 영화제에서 한국영화 최초로 황금종려상을 수상하였다. 이렇게 뜻깊은 해를 두 체들을 재조명해본다.

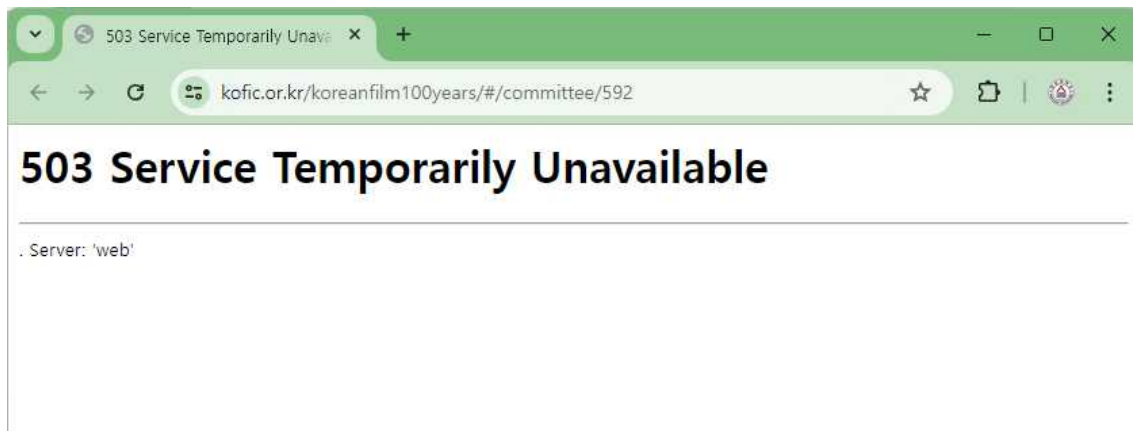
<출처>

한국영화진흥위원회 홈페이지 (<https://www.kofic.or.kr/kofic/>)

한국영화 100년 기념사업 홈페이지(<https://www.kofic.or.kr/koreanfilm100years/#/committee/592>)

<그림 부록A-37> 한국영화 <출처> 한국영화 100년 기념사업 홈페이지 정보

- 222 -



<그림 부록A-38> 한국영화 100년 기념사업 홈페이지 링크 접속 결과(503 Service Temporarily Unavailable)



<그림 부록A-39> 컬렉션 메뉴에서 검색 버튼(돋보기 모양)이 동작하지 않음



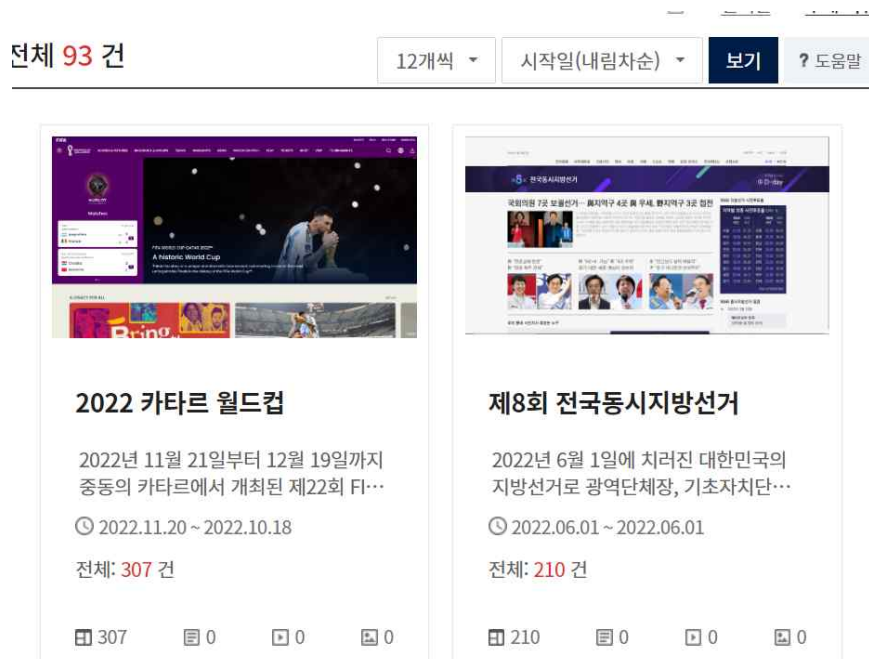
<그림 부록A-40> 재난아카이브 > 콘텐츠목록 > 이미지 메뉴 선택 시 동영상 자료 결과 제시



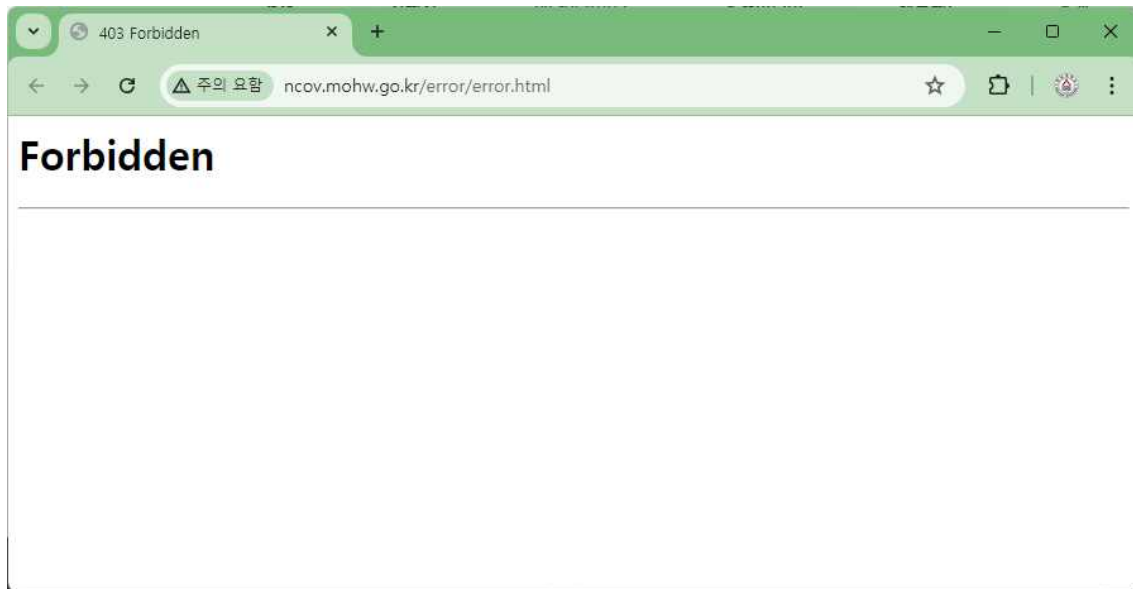
<그림 부록A-41> 재난아카이브 > 콘텐츠목록 > 동영상 메뉴 선택 시 이미지 자료 결과 제시

+ 해외재난(3)
· 해외(3)

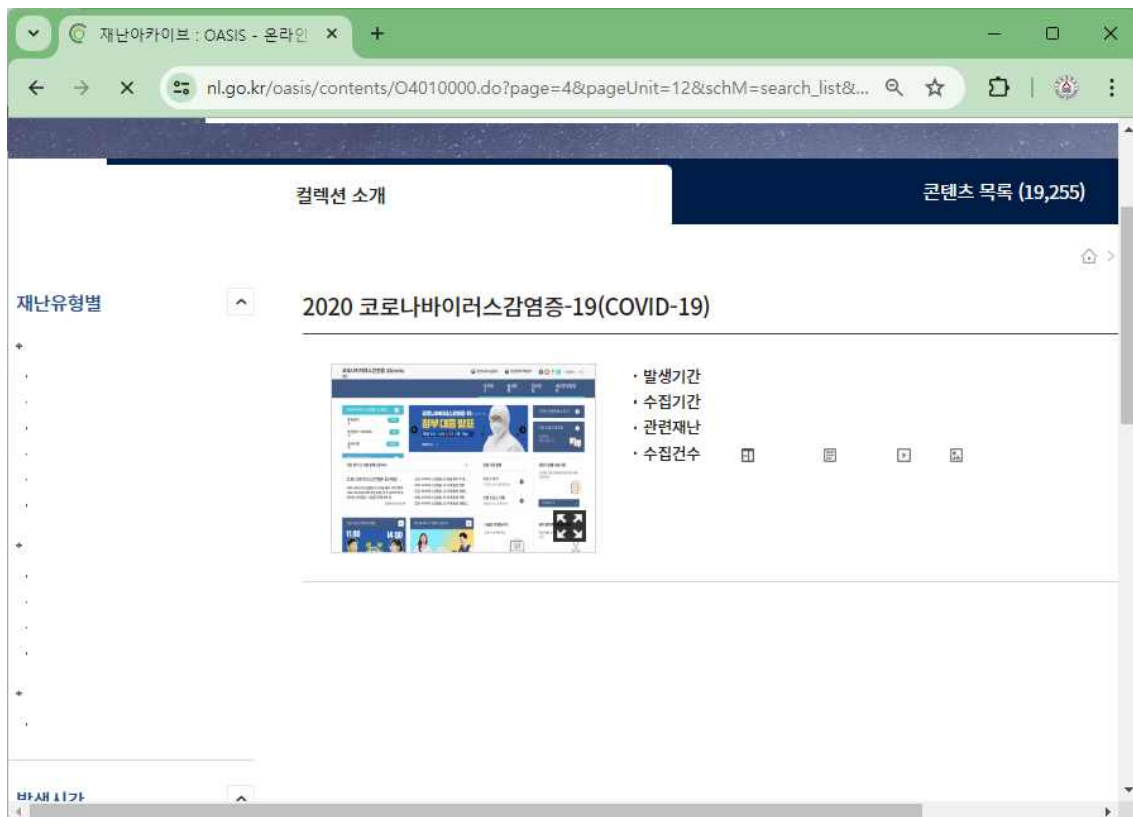
<그림 부록A-42> 재난 분류 중 직관적으로 이해가 어려운 해외재난 분류 체계(해외재난>해외)



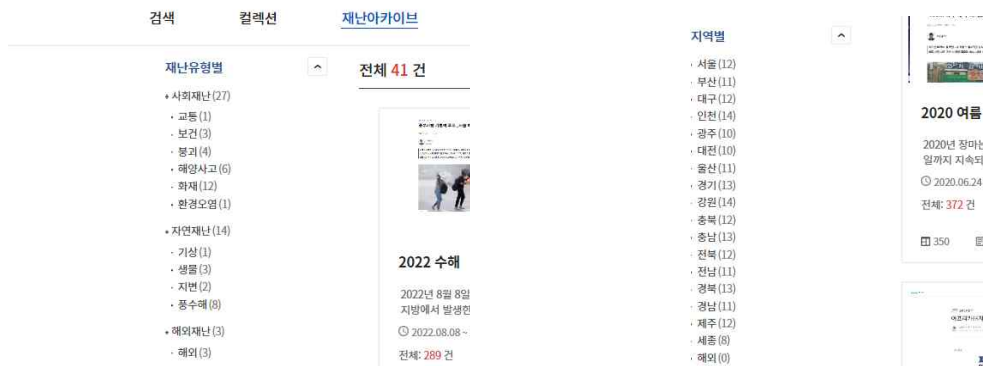
<그림 부록A-43> 2022년이 최신이며 엠바고에 대한 부가 설명이 없는 화면



<그림 부록A-44> 접근 금지된 질병관리본부 URL



<그림 부록A-45> 재난아카이브 자료 선택 시 화면이 두 번 그려지면서 깜빡이는 현상



<그림 부록A-46> 재난아카이브
분류(재난유형별) 해외 3건

<그림 부록A-47> 재난아카이브
분류(지역별) 해외 90건



<그림 부록A-48> 소셜미디어 정부기관 설명에서 URL이 링크화가
되어 있지 않음

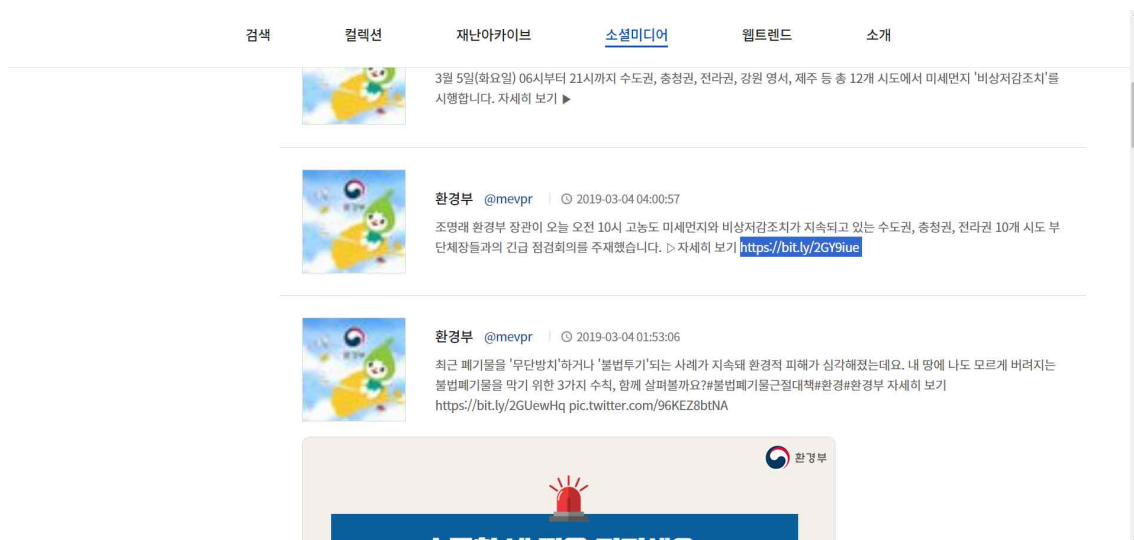


<그림 부록A-49> 가입일보다 최초
트윗일이 빠른 사례 1(합천군)

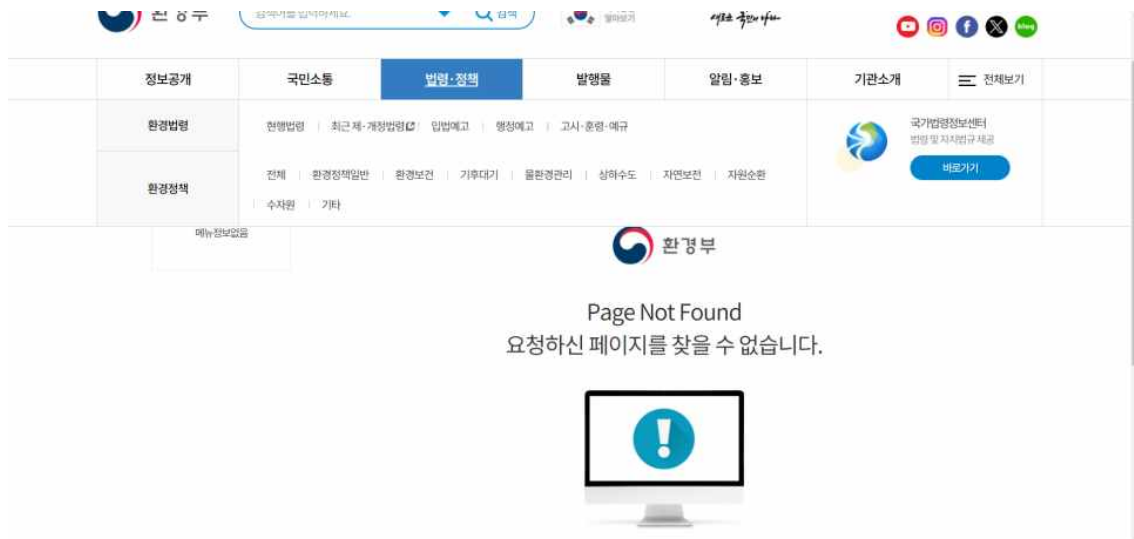
<그림 부록A-50> 가입일보다 최초
트윗일이 빠른 사례 2(노사발전재단)



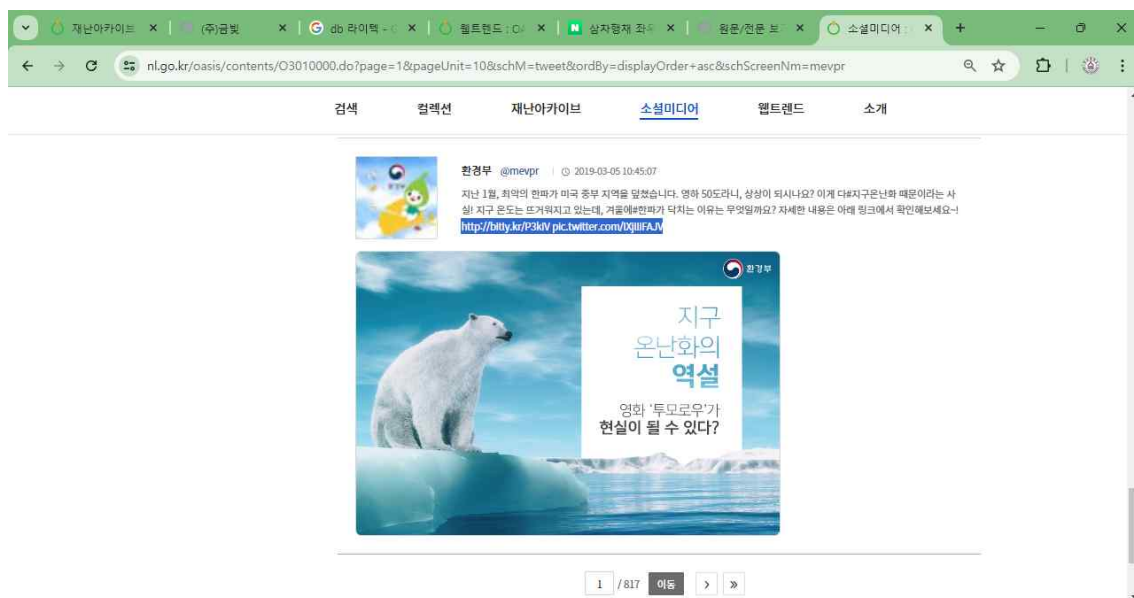
<그림 부록A-51> 고용노동부의 공식 트위터를 소개하는 기관 문양의 사진은 선명한 반면, 수집된 트윗에서의 고용노동부의 문양 사진은 선명하지 않고 이미지가 깨져있음



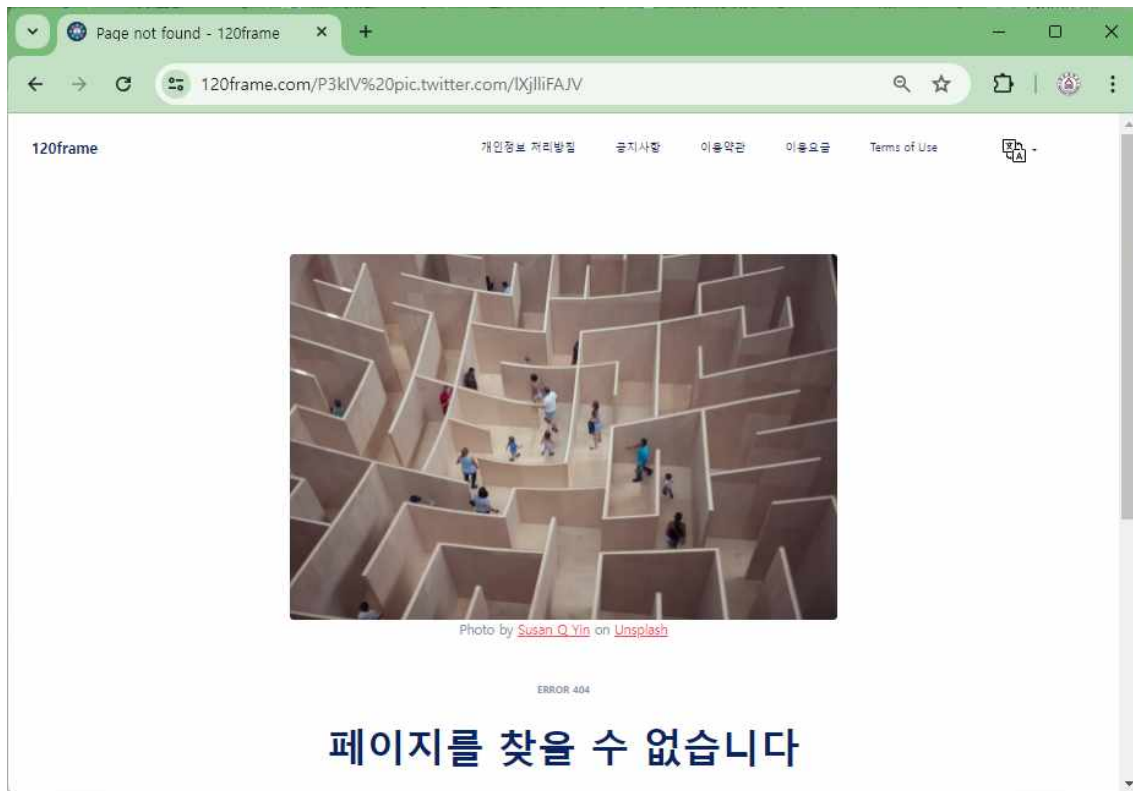
<그림 부록A-52> 소셜미디어 > 환경부 > 자세히 보기 URL (링크화되어 있지 않음)



<그림 부록A-53> 앞 그림의 환경부 자세히 보기 URL(<https://bit.ly/2GY9iue>) 복사 웹 브라우저 주소창에 붙여넣기



<그림 부록A-54> 소셜미디어 > 환경부 트윗 중 하나(1/817 페이지 중 제일 마지막 트윗)로 자세히 보기 링크가 링크화가 안 되어 있음



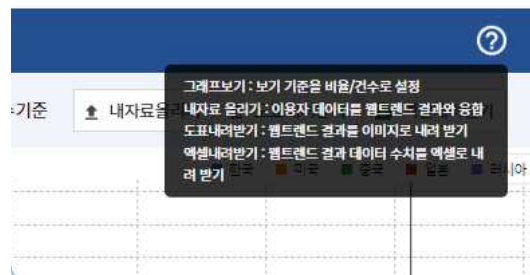
<그림 부록A-55> 앞 그림 URL([http://bitly.kr/P3kIV pic.twitter.com/lXjlliFAJV](http://bitly.kr/P3kIV_pic.twitter.com/lXjlliFAJV))을 복사 붙여넣기 한 결과



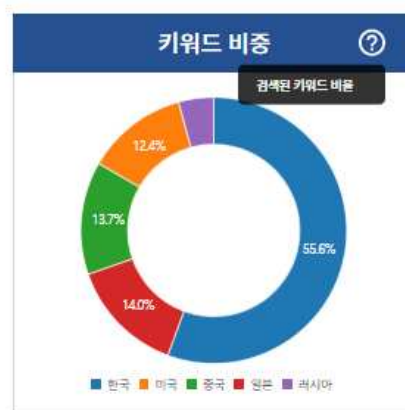
<그림 부록A-56> ‘키워드 비중 및 ’수집년도 구분’ 화면에 가려진 안내 메시지(‘2016년부터 ...’)



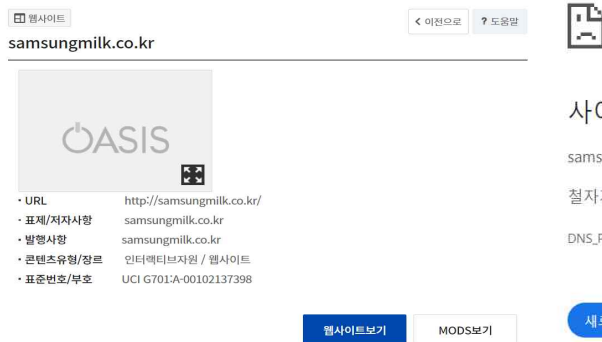
<그림 부록A-57> 중복된 라디오 버튼



<그림 부록A-58> 웹트렌드 도움말 설명



<그림 부록A-59> 단편적인 정보를 포함한 키워드 비중 도움말



사이트에 연결할 수 없음

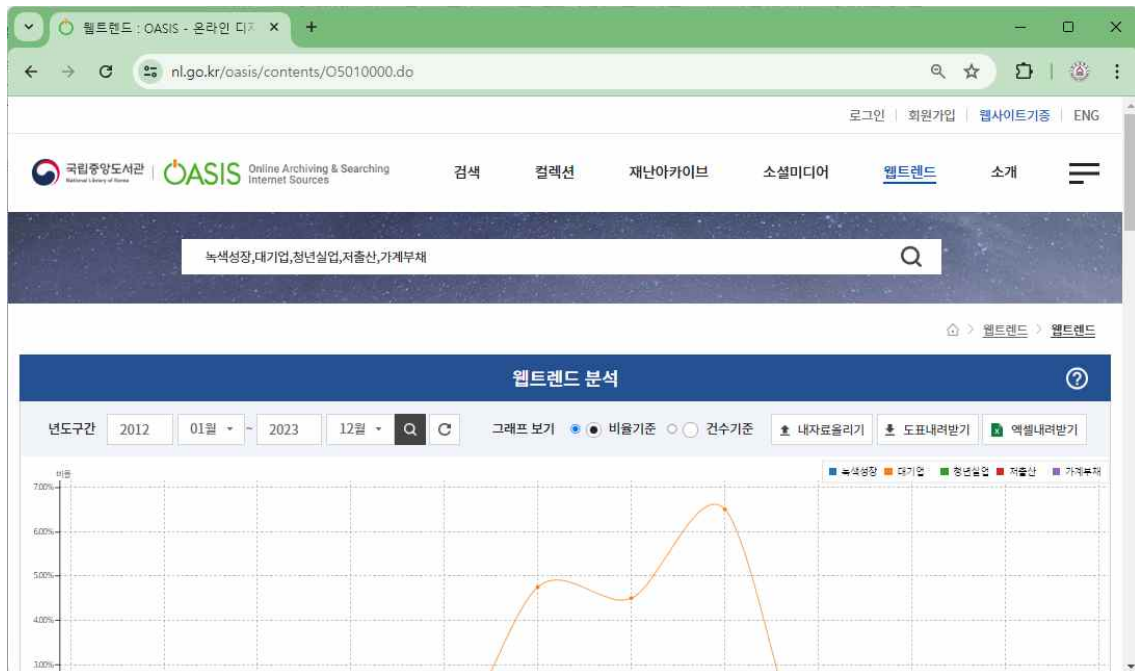
samsungmilk.co.kr에 오타가 있는지 확인하세요.

철자가 올바르게되면 Windows 네트워크 진단을 실행해 보세요.

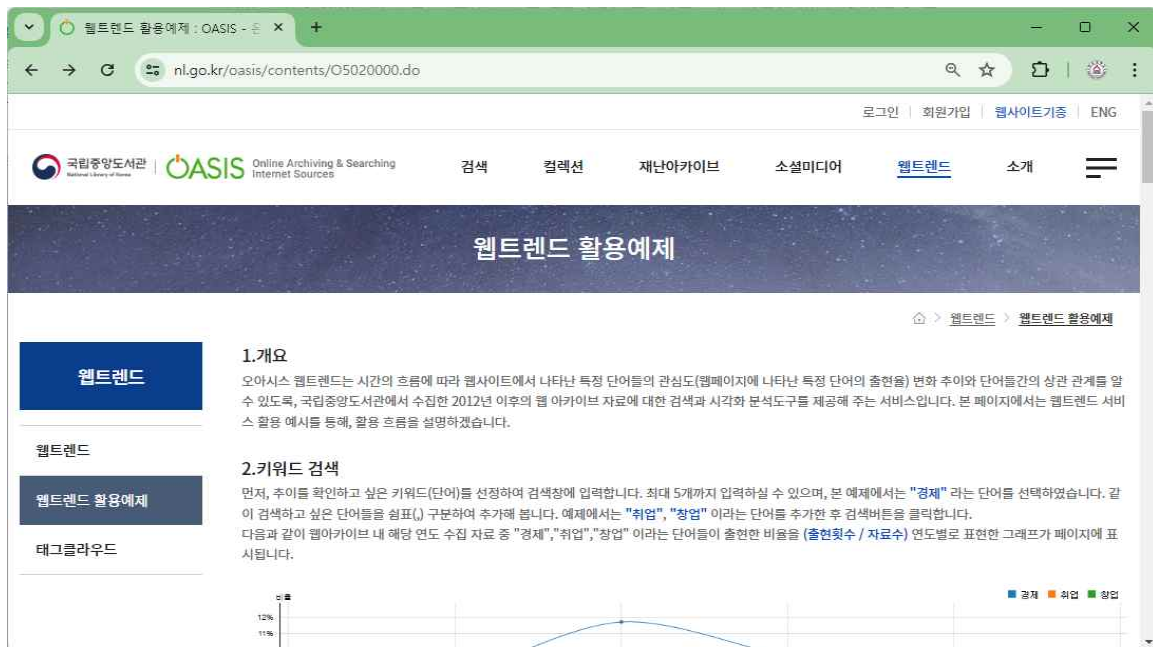
DNS_PROBE_FINISHED_NXDOMAIN

<그림 부록A-60> 웹트렌드 키워드 검색
결과 화면

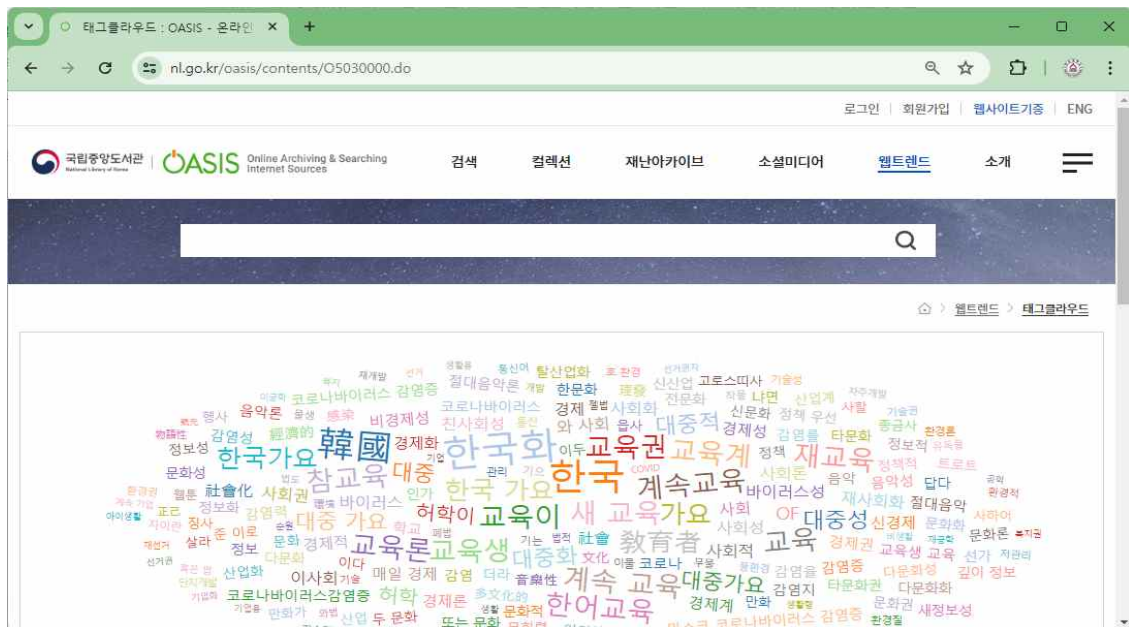
<그림 부록A-61> URL링크 접속 실패
화면



<그림 부록A-62> 웹트렌드 > 웹트렌드 분석 메뉴 선택 시 화면(왼쪽 사이드 메뉴 미제공에
주목)



<그림 부록A-63> 웹트렌드 > 웹트렌드 활용예제 메뉴 선택 시 화면(왼쪽 사이드 메뉴 제공에
주목)



<그림 부록A-64> 웹트렌드 > 태그 클라우드 메뉴 선택 시 화면(왼쪽 사이드메뉴 미제공에
주목)

부록B**GTmetrix 성능 지표 해설**

- 본 부록은 본 보고서의 3.5 오아시스 누리집 성능 측정에서 사용된 GTmetrix 도구의 성능 결과 보고서를 이해하는 데 도움을 주기 위한 내용임
- 아래 내용은 다음의 웹페이지¹⁹⁾에서 번역한 내용임
- GTmetrix 등급은 성능 및 구조 점수에서 파생된 가중치가 적용된 등급임

○ GTmetrix 등급

- GTmetrix Grade는 페이지의 성능(로딩, 상호 작용성, 시각적 안정성 측면)과 최적의 성능을 위해 얼마나 잘 구축되었는지를 결합하여 평가함
- 이 등급은 프론트엔드 구조와 사용자가 실제로 경험한 성능을 모두 고려하여 전반적인 웹페이지 경험을 반영함
- GTmetrix 등급은 방문자의 페이지 로드 경험을 종합적으로 나타낸 것으로, 페이지가 얼마나 잘 구성되었는지와 실제 성능을 모두 고려한 것임
- GTmetrix Grade는 두 가지 백분율 기반 점수의 가중 평균임. 즉,
 - ▶ 성과 점수 (60%)
 - ▶ 구조 점수 (40%)
- 페이지의 가중 평균은 GTmetrix 보고서에 표시되는 A에서 F 사이의 문자 등급에 해당함
- 각 등급별 % 범위는 다음과 같음

GTmetrix 등급(%)	GTmetrix 등급 문자 등급
90 - 100	A
80 - 89	B
70 - 79	C
60 - 69	D
50 - 59	E
0 - 49	F

- 이러한 가중치는 지속적으로 모니터링되고 조정됨. 점점 더 많은 성능 데이터를 수신하여 분석하고 핵심 Lighthouse 플랫폼 내부에서 변경이 발생함에 따라 가중치는 시간이 지남에 따라 진화함

○ 성과 점수

- 성능 점수는 기본적으로 GTmetrix 테스트를 통해 얻은 Lighthouse 성능 점수이며, 브라우저, 하드웨어 사양 및 지정된 분석 옵션(AdBlock, 연결 속도 등)을 적용한 결과임
- 성과 점수는 사용자 관점에서 페이지의 성능이 얼마나 좋은지 알려줌

19) GTmetrix, GTmetrix 보고서에 대해 알아야 할 모든 것, GTmetrix 블로그, <https://gtmetrix.com/blog/everything-you-need-to-know-about-the-new-gtmetrix-report-powered-by-lighthouse/> 2024. 6. 22. 검색

- 이 점수는 다음과 같은 가중치를 갖는 6가지 핵심 지표로 구성됨
- 로딩 성능(45%)
 - ▶ 첫 번째 콘텐츠 로딩(FCP: First Contentful Paint)(10%)
 - ▶ 속도지수(10%)
 - ▶ 가장 큰 콘텐츠 로딩(LCP: Largest Contentful Paint)(25%)
- 상호 작용성(30%)
 - ▶ 총 차단 시간(30%)
- 시각적 안정성(25%)
 - ▶ 누적 레이아웃 이동(25%)
- 성과 점수 위에 마우스를 올리면 각 측정 항목이 전체 성과 점수에 얼마나 기여했는 지에 대한 페이지별 세부 정보를 볼 수 있음
- 성과 점수 위에 마우스를 올려놓으면 각 측정 항목이 페이지의 전체 성과 점수에 얼마나 기여했는지를 확인할 수 있음
- 각 측정 항목은 점수로 측정 및 계산된 후 임계값과 비교되고, 집계된 결과(적절한 가중치 적용)가 최종 성과 점수를 구성함
- 성과 점수에 대한 자세한 내용은 GTmetrix 보고서의 성과 탭에 나와 있음
- 이 점수는 Google PageSpeed Insights(PSI)에서 확인하는 성과 점수와 비교하면 어떻습니까?
 - ▶ PSI 역시 Lighthouse에 의해 구동되지만, 성능 점수를 생성하기 위해 자체 하드웨어와 테스트 방법론을 사용함
- 점수가 달라질 수 있는 몇 가지 주요 차이점은 다음과 같음
- 지리적 차이
 - ▶ Google이 지리적 지역에 기반한 서버에서 테스트하는 지는 알 수 없음. GTmetrix를 사용하면 페이지를 분석할 테스트 지역을 정의할 수 있음
- 하드웨어 차이
 - ▶ GTmetrix와 Google은 테스트를 위한 CPU/메모리 지정이 다를 수 있으며, 이는 메트릭에 영향을 미칠 수 있음
- 네트워크 차이
 - ▶ 네트워크 연결과 분산의 차이는 개별 지표의 변화에 영향을 미칠 수 있음
 - ▶ 다른 Google 마법의 Google은 성능 점수를 계산할 때 추가 고려 사항이 있을 수 있음(예: CPU/네트워크 제한을 위한 Lantern)

- GTmetrix 성능 점수와 Google에서 생성된 성능 점수는 직접 비교할 수는 없지만 비슷할 것임

○ 구조 점수

- 구조 점수는 Lighthouse Opportunity 및 Diagnostic 감사에 대한 독창적 평가임
- 구조 점수는 페이지가 최적의 성능을 위해 얼마나 잘 구축되었는지 알려줌
- 이러한 평가를 실시하고 GTmetrix 팀이 인식한 잠재적 절감 및 중요도를 포함한 다양한 요소에 따라 포인트 값을 할당했음
- 이러한 구조 감사의 전체 목록은 여기에서 확인하기
- 구조 점수가 얼마나 좋은지는 사이트가 최적의 성능을 발휘하도록 얼마나 잘 구축되었는지를 반영함
- 당사의 맞춤형 GTmetrix 감사
 - ▶ 표준 Lighthouse 감사 외에도 웹 성능과 관련이 있다고 생각되는 자체 맞춤 감사도 포함했음. 이러한 맞춤형 GTmetrix 감사에는 다음이 포함됨
- Keep-Alive 활성화
 - ▶ 이 문제는 Keep-Alive ON 지시문이 감지되지 않은 HTTP/1.1을 사용하는 페이지에서만 발생함
 - ▶ 페이지가 HTTP/2를 사용하는 것으로 감지되면 이 감사는 영향을 미치지 않음
- CSS 스프라이트를 사용하여 이미지 결합
 - ▶ 페이지에 여러 개의 작고 스프라이트 이미지가 제공되는 경우 이 감사가 트리거 됨
 - ▶ 이 감사에는 페이지가 HTTP/1.1을 사용하는지 HTTP/2를 사용하는지에 따라 다른 임계값이 적용됨
- 콘텐츠 전송 네트워크(CDN)를 사용하기
 - ▶ 저희는 지난 기사에서 CDN의 이점에 대해 논의했으며, CDN이 전 세계적으로 꾸준히 빠른 웹사이트를 제공하는 데 중요한 역할을 한다고 여전히 생각함
 - ▶ GTmetrix는 이전 보고서에서처럼 처벌적이지는 않지만, 글로벌 전송을 개선하기 위해 CDN에서 제공될 수 있는 모든 정적 리소스에 대한 플래그를 계속 지정할 것임
- CSS @import를 피하기
 - ▶ 가능하다면 CSS 사용을 피하는 것이 @import 여전히 권장되는 관행임. 브라우저에서 해당 defer 속성이 없는 모든 import 지시문을 접하면 즉시 다운로드, 구문 분석, 실행되기 때문임
 - ▶ 이렇게 하면 나머지 페이지의 렌더링이 차단될 수 있음
- 구조 점수에 대한 전체 세부 정보는 GTmetrix 보고서의 구조 탭에 나열되어 있음
- 구조 점수가 성과 점수에 영향을 미치는가?
 - ▶ 구조 점수는 성과 점수에 직접적인 영향을 미치지 않음
 - ▶ 그러나 구조 점수를 개선하면 일반적으로 페이지 성능이 전반적으로 향상될 수 있음. 이는 더 나

은 성능 점수와 확장하여 더 나은 GTmetrix 등급으로 이어질 것임

- ▶ 새로운 GTmetrix 보고서의 구조 감사를 더 나은 웹 성능을 위한 지침서로 생각하고, 성능 점수 측정 항목을 실제 웹 성능의 스톱워치로 생각해 볼 필요가 있음
- ▶ 총 차단 시간이 긴 것은 메인 스레드 작업과 Javascript 실행이 긴 데 기인할 가능성이 높음
- 구조 감사를 다루는 것은 전반적인 페이지 로드 시간을 개선하기 위한 좋은 시작점이 될 수 있음. 또한, 일부 감사는 상관관계가 있으므로 한 감사를 수정하면 다른 감사에 영향을 미칠 수 있음

○ 웹 바이탈

- Google은 빠른 웹사이트 경험을 제공하기 위해 집중해야 할 핵심 지표로 웹 바이탈을 설정했음
- 웹 바이탈은 페이지 성능에 가장 큰 영향을 미치는 주요 지표임
- 이러한 측정 항목에는 다음이 포함됨
- 가장 큰 콘텐츠 로딩(LCP):
 - ▶ 가장 큰 콘텐츠 요소가 표시되는 데 걸린 시간임
- 총 차단 시간(TBT):
 - ▶ 페이지 로딩 과정 중 스크립트에 의해 차단된 시간임
- 누적 레이아웃 변화(CLS)
 - ▶ 사용자가 페이지를 로드하는 동안 경험한 레이아웃 변화량임
- 이런 측정 항목에서 좋은 성적을 거두는 것은 방문자에게 빠르고 (Google에서 말하는) 즐거운 경험을 제공하고 있다는 좋은 지표임
- 웹 바이탈은 성과에 대한 인식에 가장 큰 영향을 미치는 핵심 지표 몇 가지에 초점을 맞추므로 페이지를 벤치마킹하는 좋은 방법임. 새로운 GTmetrix 보고서에서 GTmetrix 등급과 성과/구조 점수와 함께 웹 바이탈이 중심에 표시됨

대한민국 웹자원 아카이빙(OASIS) 개선 및 활용성 강화 방안 연구

주소: (06579) 서울특별시 서초구 반포대로 201 국립중앙도서관

발행일: 2024년 12월 30일

발행인: 김희섭

발행처: 문화체육관광부 국립중앙도서관

연락처: 02-590-0500

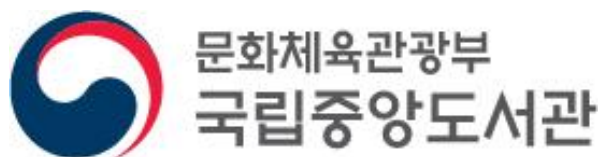
홈페이지: <https://www.nl.go.kr>

연구수행기관: 한남대학교 산학협력단

ISBN: 979-11-6513-473-0 <비매품>



문화체육관광부의 창작 저작물인 대한민국 웹자원 아카이빙 (OASIS) 개선 및 활용성 강화 방안 연구의 경우 [공공누리 제4 유형] 출처표시 + 상업적 이용금지 + 변경금지 조건에 따라 이용하실 수 있습니다.



대한민국 웹자원 아카이빙(OASIS)
개선 및 활용성 강화 방안 연구

