

재난안전 기록정보자원 아카이빙 전략의 실증적 검증*

- 유관기관 웹 기록물 자동수집 결과를 통해 -

Empirical Study of Disaster and Safety Record Information Resources Archiving Strategy based on Automatic Acquired Web Records from Related Organizations

강 주 연 (Ju-Yeon Gang)**, 한 희 정 (Hui Jeong Han)***
김 용 (Yong Kim)****, 오 효 정 (Hyo-Jung Oh)*****

초 록

본 연구는 선행연구에서 제안된 재난안전 기록정보자원 아카이브 구축전략의 적용 가능성을 살펴보고, 실제 운용 시 문제점이 무엇인지를 파악하는데 그 목적이 있다. 이에 본 연구에서는 55개의 재난안전 유관기관을 선정하고 해당 기관 웹에 공개된 재난안전 기록정보자원을 수집하여 구축 전략의 세부 단계를 수행해 보았다. 구축 결과 검증을 통해 기존에 제안된 아카이브 구축 전략의 문제점을 분석하였으며, 보다 현실적인 전략이 될 수 있도록 정책적, 기술적, 관리적, 서비스적 측면에서 개선방안을 제안하였다.

ABSTRACT

This study explores the feasibility of our proposed disaster and safety record information archiving strategy and determines several problems encountered during actual operations. For practical experiments, 55 disaster and safety related organizations are selected and the detailed steps of our archiving strategy are implemented. Based on the real disaster and safety record information acquisition results, we analyzed the challenges of the proposed strategy. Furthermore, we suggested the improvements plans in terms of policy, technical, managerial, and service-oriented so as to become a more realistic strategy.

키워드: 재난기록, 재난안전정보, 재난아카이브, 실증적 검증, 적용 가능성

Disaster Records, Disaster and Safety Information, Disaster Archives, Empirical Verification, Feasibility

-
- * 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 - 재난안전플랫폼기술개발사업의 지원을 받아 수행된 연구임(No. NRF-2016M3D7A1912703).
이 논문은 2017년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(한국연구재단-2017-미래기초과학핵심리더양성사업/글로벌박사양성사업).
- ** 전북대학교 기록관리학과 박사과정(juyeongang@naver.com) (제1저자)
- *** 전북대학교 문화융복합아카이빙연구소 전임연구원(freebirdhhj@naver.com) (공동저자)
- **** 전북대학교 문헌정보학과 교수(yk9118@jbnu.ac.kr) (공동저자)
- ***** 전북대학교 기록관리학과 조교수, 문화융복합아카이빙연구소(ohj@jbnu.ac.kr) (교신저자)
- 논문접수일자 : 2017년 11월 27일 논문심사일자 : 2017년 12월 16일 게재확정일자 : 2017년 12월 17일
한국비블리아학회지, 28(4): 223-246, 2017. [<http://dx.doi.org/10.14699/kbiblia.2017.28.4.223>]

1. 서론

1.1 연구배경 및 필요성

최근 들어 '조류독감(AI)'과 같은 신종 전염병이나 '경주 지진', '포항 지진'과 같이 기존에 우리나라에 자주 나타나지 않았던 재난이 발생하고 있다. 뿐만 아니라, '후쿠시마 원전사고'처럼 복합적 재난이 발생함에 따라 재난과 안전 문제가 사회적 이슈로 떠오르고 있다. 특히 정부의 잘못된 대응과 신빙성 없는 정보의 공유 및 확산은 국민들을 더욱 불안하게 만들고 있는데, 이는 정부 기관 간 원활하지 않은 정보 공유와 신뢰성 없는 생산자에 의해 생산된 정보의 확산에서 기인한다. 실제로 한희정 외(2017a)가 진행한 재난 관련 업무 담당자들의 인터뷰를 보면 재난안전 유관기관 간 정보가 공유되지 않아 발생하는 문제점과 고충을 살펴볼 수 있다. 또한 그로 인한 업무 진행의 애로사항을 확인할 수 있다. 한편, 일반 국민들 역시 잘못된 정보의 공유와 확산으로 인하여 불안과 공포에 사로잡히는 경우가 종종 발생하곤 하였다. 예컨대, 지난 2015년 발생한 '메르스' 사태는 상기 두 양상, 즉 재난안전 유관기관 간 원활하지 않은 정보 공유에서 비롯된 늦은 대응과 잘못된 정보의 확산으로 공포에 떨었던 국민의 모습을 가장 잘 보여준다.

이러한 신빙성 없는 정보의 공유와 확산이나 기관 간 원활하지 않은 정보의 공유는 상기한 바와 같이 신뢰성 없는 생산자에 의해 생산된 정보의 확산과 신뢰성 있는 정보의 위치 공유에 대한 부재에서 비롯된 것이다. 따라서 신뢰성 있는 생산자에 의해 생산된 정보와 이에 대한

위치 정보가 공유되어야만 이러한 작금의 사태를 해결할 수 있다. 이러한 관점에서 한희정 외(2017a)는 재난안전 기록정보자원의 현황을 조사하고 재난에 대한 선제적 대응과 예방을 위한 체계적인 관리방안을 제안하였다. 한발 더 나아가 한희정 외(2017b)는 원활한 정보 공유를 위한 재난안전 기록정보자원 아카이브 구축 전략을 제안하였다.

재난안전 기록정보자원 아카이브는 재난안전 유관기관, 즉 신뢰성 있는 정보 생산자에 의해 생산된 기록정보자원을 수집, 보존, 관리 및 활용하고 공유할 수 있는 기반이다. 그러나 한편으로는 재난안전 유관기관으로 분류되는 수많은 공공기관들이 방대한 양의 기록을 생산하기 때문에 이러한 기록정보자원을 보존하기 위해서는 엄청난 비용과 시간이 소요될 것이다. 이에 한희정 외(2017b)는 기록정보자원의 위치 정보만을 보존, 서비스하는 전략을 제안하였는데, 본 연구는 이러한 필요성 및 문제점에 공감하여 그 전략의 적용가능성을 살펴보고자 한다. 즉, 한희정 외(2017b)가 제안한 4단계 전략, 재난안전 유관기관 기록정보자원의 전수조사, 자동화 가능성 분석, 아카이빙 대상 선별, 메타데이터 자동수집을 실증적으로 검증해보으로써 실제 운용 시 문제점을 살펴보고자 한다. 나아가 재난안전 기록정보자원 아카이브 구축을 위한 현실적인 전략에 대해 논하고자 한다.

1.2 연구목적 및 방법

본 연구는 한희정 외(2017b)의 후속연구로서 재난안전 기록정보자원 아카이브 구축 전략의 실증적 검증을 목적으로 한다. 이에 선행연구의

내용 및 범위를 그대로 적용하였으며, 55개의 재난안전 유관기관을 선정하여 해당 기관의 재난안전 기록정보자원을 실험 대상으로 사용한다. 실험 대상을 중심으로 선행연구가 제안한 4단계(전수조사-자동화 가능성분석-아카이빙 대상 선별-메타데이터 자동수집)의 아카이브 구축 전략의 세부 단계별 적용을 통해 재난안전 기록정보자원을 수집하고, 재난안전 기록정보자원 아카이브를 구축하고자 한다. 구축결과 검증을 통해 문제점을 분석하고 이에 기반해 선행연구에서 제안한 아카이브 구축 전략을 현실적인 전략으로 발전시키기 위한 시사점을 제안한다.

1.3 선행연구

본 연구와 맥락을 같이 하는 연구로서 아카이브 구축에 관한 연구는 이미 다수의 연구가 다양하게 진행되어 왔다. 이에 본 연구는 최근에 진행된 연구 위주로 살펴보았다(설문원 2012; 여진원, 장우권 2013; 설문원, 최이랑, 김슬기 2014; 현문수, 전보배, 이동현 2014; 한희정 외 2017b).

먼저 설문원(2012)은 로컬리티 기록화를 위한 참여형 아카이브 구축에 대해 연구하였는데, 사례조사를 통해 아카이브 구축 시 고려해야 할 요소에 대하여 논하였다. 그 요소로는 1) 추진 주체 및 방식, 2) 수집 기관 및 공동체의 협력 네트워크 구축, 3) 생산맥락의 보존과 재 맥락화, 4) 평가 선별, 5) 이용자 참여를 제안하였다. 이와 같은 맥락에서 설문원, 최이랑, 김슬기(2014)는 밀양 송전탑 건설 갈등사건의 기록화를 위한 디지털 아카이브 구축 방법론에 대해

연구하였다. 아카이브 구축 방향으로 1) 구축 대상 선정, 2) 사건 및 주민 기억의 기록화, 3) 오픈 소스(Omeka)를 활용한 구축, 4) 다양한 접근점 제공, 5) 맥락에 대한 풍부한 기술, 6) 망라적 수집을 제안하였다. 최종적으로 Omeka를 활용하여 디지털 컬렉션과 전시를 구현하였다. 현문수, 전보배, 이동현(2014) 역시 설문원(2012)과 맥락을 같이 하는데, 이들은 참여형 노동 아카이브 개발 모형을 디지털 노동 아카이브 구축에 적용하고자 하였다. 아카이브 구축 대상은 부산 영도의 조선 사업이며, 구축 도구로는 Omeka를 사용하였다. 구축 절차로는 1) 대상 범위 설정, 2) 참여주체 조사, 3) 맥락 및 기록 조사와 수집, 4) 기록 기술, 5) 맥락 기술 및 콘텐츠 구축을 제안하였다. 최종적으로 아카이브 구축 실험을 진행하였고, 결과를 통해 시사점을 제안하였다.

한편, 여진원, 장우권(2013)은 도시 아카이브 구축에 관한 연구를 진행하였다. 사례 조사를 바탕으로 기획, 수집, 분류, 정리, 활용에 이르는 5단계의 프로세스를 제안하였다. 한희정 외(2017b)는 재난안전 기록정보자원 아카이브를 위한 자동 수집 기반의 구축 전략을 제안하였다. 아카이브 구축 전략은 총 4단계로 구성되며, 단계별 진행사항에 대해 자세히 기술하였다. 이상의 연구들을 살펴보면 제언으로 그친 경우(설문원 2012; 여진원, 장우권 2013; 한희정 외 2017b)와 실제 구축까지 진행한 경우(설문원, 최이랑, 김슬기 2014; 현문수, 전보배, 이동현 2014)로 나눌 수 있는데, 전자는 실증적 검증을 진행하지 않았기 때문에 한계가 있으며, 후자는 특정 사례를 중심으로 구축하였기 때문에 일반화에 어려움이 있다는 한계가 있다.

상기 언급한 연구 외에도 본 연구와 맥락을 같이 하는 연구는 공공기관의 웹 아카이빙에 관한 연구가 있다. 먼저, 차승준, 최윤정, 이규철(2009)은 공공기관의 웹사이트가 구조, 내용, 맥락, 즉 기록의 속성을 통해 증거적 능력을 확보할 수 있게 하는 특징을 가졌음을 언급하면서 공공기관 웹사이트 기록물을 수집, 보존해야 함을 주장하였다. 이를 위해 국가기록원과 행정안전부 웹사이트를 분석하였고 심층 웹에 대하여 정의하였다. 최종적으로 심층 웹 기록물의 아카이빙을 위한 메타데이터 요소를 제안하였다. 이연수, 남성운, 박성배(2010)는 웹사이트가 온라인 환경에서 대국민 업무와 관련해 중요한 역할을 담당하고 있기 때문에 웹 사이트에서 제공되는 웹 기록물의 중요성을 강조하고 있다. 이에 국가기록원 웹사이트 기록물을 중심으로 수집 테스트를 진행하였으며 결과 분석을 통해 원격 수집 시 고려해야 할 사항들을 제안하였다. 이연수, 남성운, 윤대현(2013)은 이연수, 남성운, 박성배(2010)의 연구에서 웹 기록물 수집 시 고려사항으로 제시된 중복 수집 문제에 관하여 집중적으로 논하였다. 국가기록원에서 개발한 웹 아카이빙 테스트 베드를 사용하여 공공기관 웹사이트를 수집한 결과를 바탕으로 중복 수집 문제를 해결하기 위한 방안을 제안하였다. 이상의 연구들은 공공기관의 웹 기록물을 아카이빙하고자 하였다는 측면에서 본 연구와 맥락을 같이하지만 연구 범위 및 대상이 협소하다는 점에서 한계가 있다.

이상의 내용을 정리해보면, 본 연구는 선행연구들과 달리 1) 기존에 제안한 아카이브 구축 전략을 실증적으로 검증하였다는 점, 2) 다양한 실험 대상에 제안 전략을 적용, 대량의 기록을

자동 수집함으로써 실제 운용 가능성을 검증하였다는 점에서 차이가 있다.

2. 이론적 배경

2.1 웹 기록물 수집도구

웹 아카이빙은 웹 전체 또는 그 일부를 수집하고 일반 대중을 위해 웹 아카이브에서 영구 보존하는 것이다. 웹 사이트에서 제공되는 웹 기록물은 웹의 역동성과 기술변화에 많은 영향을 받는다. 이러한 영향으로 인해 웹 기록물은 기록의 가치와는 관계없이 주기적 또는 비주기적으로 갱신되거나 소멸된다. 잦은 수정과 변경, 삭제되는 특성을 가진 웹 기록물의 수집 및 보존에 대한 중요성이 강조되면서, 관련 연구가 증가하고 있다. 또한 웹 기록물을 수집하고 보존하기 위한 수집도구 개발이 필요함에 따라 수집 관련 프로젝트를 통해 수집을 위한 도구가 개발되었다(김광영 외 2011).

웹 기록물을 수집하기 위한 도구를 웹 크롤러(web crawler) 또는 하베스터(harvester)라고 한다. 웹 크롤러는 웹 서버를 순환하며 각 홈페이지에서 제공하고 있는 수많은 양의 정보를 수집할 수 있게 하여, 사람이 홈페이지에 접근해 링크를 따라 수동으로 정보를 수집하는 작업을 대신 해주는 프로그램이다(조완섭, 이정은, 최치환 2013). 웹 크롤러는 일반적으로 사이트 방문이나 크롤러 실행 방식에 따라 집중 크롤러(focused crawler), 증분 크롤러(incremental crawler), 분산 크롤러(distributed crawler), 병렬 크롤러(parallel crawler)로 구분할 수 있다.

집중 크롤러(focused crawler)는 각 페이지들을 수집하기 위해서 특히 중요하다고 선정된 페이지나 주제 등을 수집하기 위한 방법이다. 증분 크롤러는 크롤링 중인 콘텐츠 원본에 지정되어 있는 웹 문서를 크롤링하여 마지막 크롤링 이후 수집되지 않은 정보를 크롤링하는 방식으로 기존 크롤링 정보에 변경 내용만 추가하는 방법이다. 분산 크롤러는 웹 문서를 수집하는 과정에 시간이 많이 소요되는 문제를 해결하기 위한 방안으로 크롤러가 동작하는 다수의 서버가 동시에 웹 수집을 수행하는 것이다. 즉, 중심 역할을 하는 서버를 두고 각 크롤러 서버를 관리하는 방식으로 웹 문서를 수집한다. 병렬 크롤러는 웹의 규모가 커짐에 따라 하나의 프로세스 또는 스레드만으로 전체 웹 페이지를 수집하기에는 어려움이 존재한다는 문제를 해결하기 위해 여러 개의 프로세스 및 스레드를 이용하여 대량의 웹 페이지들을 빠르게 수집하기 위한 크롤러이다(Desai et al. 2017).

이러한 일반적인 크롤러는 심층 웹의 정보를 수집하는데 어려움이 있는데, 이를 극복하기 위하여 심층 웹을 위한 크롤러에 대한 연구가 활발히 진행되고 있다. 예컨대, Deepbot, HiWE, Incremental Web Crawler 등이 이에 해당한다. Deepbot은 미니 웹 브라우저라는 클라이언트 스크립트 실행 도구를 내장하여 서버와의 세션 유지 및 클라이언트의 스크립트 실행이 가능하도록 하여 데이터를 수집하는 방식으로 동작한다(Álvarez et al. 2007). HiWE는 웹 질의어 인터페이스에 숨겨진 데이터를 추출하기 위해 입력 폼을 분석하고 입력 폼에 값을 전달하는 방식으로 심층 웹을 수집한다(Raghavan and Hector 2001). Incremental Web Crawler

는 심층 웹이 제공하는 정보의 변화에 즉시 반영된 결과를 저장하기 위한 크롤러이다. 이를 위해 크롤러가 웹페이지를 방문하는 방문주기를 확률적으로 정하고 웹페이지 변화 주기를 계산하여 재방문에 대한 최적의 값을 적용하는 방식이다(Madaan et al. 2010).

하지만 이 같은 심층 웹을 위한 크롤러들도 관련 정보 수집을 위해서는 많은 질의어를 필요로 한다는 단점이 존재한다(Pablo and Luis 2017). 질의어는 관련 주제나 범위에 따라서 관리되어야 하기 때문에 정확한 결과를 위해서는 사람의 노력이 필요하다. 이에 본 연구에서는 폼에 값을 입력하여 실행된 결과를 추출하는 기존 키워드 중심의 심층 웹 수집방안의 대안으로 클라이언트 스크립트를 일반 URL 처럼 활용이 가능한 크롤러를 적용하였다.

2.2 재난안전 기록정보자원 아카이브 구축 전략

본 연구는 한희정 외(2017b)가 제안한 재난안전 기록정보자원 아카이브 구축 전략을 실증적으로 검증하는데 그 목적이 있다. 이에 선행 연구에서 설정한 연구 내용 및 범위 즉, 재난안전 기록정보자원 아카이브의 목적, 수집 범위, 자동수집 기반의 구축 전략을 그대로 적용하였다. 그 내용을 정리하면 <표 1>과 같다.

한희정 외(2017b)가 제안한 재난안전 기록정보자원 아카이브는 “재난과 관련된 실질적 기록정보 및 현장정보, 학술정보, 실무정보 및 실생활에 유용한 재난관련 전반적인 정보를 통합적으로 제공”하는데 그 목적이 있다. 따라서 재난안전 유관기관이 재난 관련 업무를 수행하면서 생산하는 기록정보자원을 수집 대상으로

〈표 1〉 재난안전 기록정보자원 아카이브 구축 전략

목적	<ul style="list-style-type: none"> • 신뢰성 있는 재난안전 기록정보자원을 통합적으로 제공 • 관련 기관 및 단체에게 실질적 업무에 필요한 정보 제공 • 관련 분야의 연구자에게 학술정보 제공 • 일반 이용자에게 실생활에 유용한 재난관련 정보 제공 			
수집범위	대상	<ul style="list-style-type: none"> • 재난안전 기록정보자원의 목록 및 위치 정보 • 재난안전 기록정보자원: 재난안전 유관기관이 재난을 예방·대비·대응·복구하는 등의 재난 관련 전반적인 활동을 하면서 생산하는 모든 정보자원 		
	유형	<ul style="list-style-type: none"> • 데이터 세트를 제외한 모든 유형의 재난안전 기록정보자원 중 온라인에 공개되어 있는 기록정보 자원 		
구축 전략	1단계	2단계	3단계	4단계
	전수조사	자동화 가능성 분석	아카이빙 대상 선별	메타데이터 자동수집

*출처: 한희정 외(2017b)

하였다. 다만, 원문 수집에 대한 저작권 및 법적 문제로 인하여 온라인에 공개되어 있는 재난안전 기록정보자원의 목록과 위치정보로 수집 대상 및 유형을 한정하였다.

재난안전 기록정보자원(이하 재난기록정보)은 1) 최근 발생하는 재난의 특성으로 인하여 장기적으로 기록화 될 필요가 있으며, 2) 생산 기관이 다양하고, 3) 기록물의 양이 방대하기 때문에 사람이 직접 수집하는데 어려움이 있다. 따라서 사전에 생산기관과 수집 대상 및 범위를 정하여 자동수집 기반의 구축 전략을 세워야 할 필요가 있다. 이에 한희정 외(2017b)는 1) 재난안전 기록정보자원의 전수조사, 2) 자동화 가능성 분석, 3) 아카이빙 대상 선별, 4) 메타데이터 자동수집, 총 4단계의 구축 전략을 제안하였다.

본 연구에서는 상기 기술한 내용을 전제로 하여 4단계의 재난안전 기록정보자원 아카이브 구축 전략을 실제 재난기록정보를 통해 검증하고자 하였다. 검증 방법 및 결과에 대한 자세한 내용은 3장에 기술하였다.

3. 재난안전 기록정보자원 아카이브 구축 전략 검증

3.1 검증 방법

재난안전 기록정보자원 아카이브 구축 전략의 적용 가능성을 검증하기 위하여 재난안전 유관기관을 선정하고 해당 기관들의 재난기록정보 현황을 분석하였다. 이를 위해 총 55개의 재난안전 유관기관을 선정하였으며 이 중 5개 기관을 선정하여 실제 자동수집을 검증하였다. 아카이브 구축 절차는 한희정 외(2017b)의 연구를 바탕으로 하였으며, 구축 전략의 단계별 내용과 검증 방법, 수행 기간은 〈표 2〉와 같다.

아카이브 구축 전략은 총 4단계로 구성되는데, 먼저 1단계에서는 재난안전 유관기관별 재난기록정보 현황을 파악하고 이에 대한 세부 정보를 작성한다. 이를 통해 향후 재난안전 기록정보자원 아카이브를 위한 수집 대상 선별 기준 및 자동수집을 위한 기초자료를 마련할 수 있다. 전수조사 단계에서는 기록물의 세부정보

〈표 2〉 자동수집 기반의 재난안전 기록정보자원 아카이브 구축 전략

분석 단계	내용	방법	수행 기간
1단계 전수조사	재난안전 유관기관별로 수집해야 할 재난기록정보를 파악하고 이에 대한 세부정보를 작성	<ul style="list-style-type: none"> 재난안전 유관기관 선정(55개)¹⁾ 재난기록정보 전수조사 및 세부정보 작성 	2016년 11월 ~ 2017년 01월
2단계 자동화 가능성 분석	1단계에서 파악된 재난기록정보에 대하여 자동수집 가능 여부를 판단하고 기술적인 부분에 대한 세부정보를 파악	<ul style="list-style-type: none"> 재난기록정보 게시 형식 분석 자동화 가능성 분석 기준 마련 자동화 가능성 분석 	2016년 12월 ~ 2017년 02월
3단계 아카이빙 대상 선별	자동수집된 재난기록정보의 아카이빙 여부를 결정하기 위한 기준 및 세부지침 마련	<ul style="list-style-type: none"> 대표 기관 5개 선정²⁾ 아카이빙 대상 선별 기준 마련 아카이빙 대상 선별 	2017년 01월 ~ 2017년 05월
4단계 메타데이터 자동수집	아카이빙 대상으로 선별된 재난기록정보에 대한 목록 및 위치정보의 메타데이터 자동수집 가능성 분석	<ul style="list-style-type: none"> 수집할 메타데이터 요소 추출 메타데이터 수집 수집 가능성 분석 	2017년 01월 ~ 2017년 05월

를 작성해야 하는데, 여기서 세부정보란 생산 기관, 기록물 제목, URL 등의 메타정보를 의미하며 차후 기록물의 식별 및 자동수집을 위한 기초 정보를 의미한다. 1단계의 검증을 위하여 본 연구에서는 총 55개의 재난안전 유관기관이 온라인에 공개한 재난기록정보에 대해 전수조사를 실시하고, 세부정보를 작성하였다. 실험 대상 재난안전 유관기관은 한희정 외(2017a)에서 제시한 36개의 기관을 바탕으로 국민안전처(現 행정안전부)와의 협의를 통해 확대, 선정하였다.

구축 전략 2단계에서는 1단계에서 전수조사한 결과를 바탕으로 자동수집 가능성을 분석한다. 자동수집 가능성 분석은 크롤러 설계에 필요한 기초자료를 마련하기 위한 것으로, 표면 웹과 심층 웹을 구분하여 심층 웹을 수집할 수 있는 방안을 모색하는데 필요하다. 즉, 자동수집 가능성 분석을 통해 심층 웹에서 수집 가능

한 기록정보자원의 유형을 분석하고 수집 대상의 물리적인 범주를 정의하여 수집 가능한 기록물 건수를 예측한다. 2단계의 검증을 위하여 55개 재난안전 유관기관의 재난기록정보 전수조사 목록을 대상으로 자동수집 가능성을 분석하였다. 특히 표면 웹과 심층 웹의 구분, 크롤링 가능 여부, 웹에서 재난기록정보가 게시된 형식 등을 전체 목록에 대하여 검토하였다. 이 과정에서 홈페이지에 재난기록정보가 제공되는 형식은 게시판과 페이지로 구분되며 그 유형이 다양하다는 점을 확인하였다(〈표 3〉, 〈표 4〉 참조). 이에 〈표 5〉와 같이 각 유형별로 자동화 가능성 분석 기준을 마련하였다. 〈표 5〉는 앞서 분석한 내용을 바탕으로 실시한 샘플링 작업을 통해 작성되었다.

3단계는 아카이빙 대상을 선별하기 위하여 기준 및 세부지침을 마련하는 단계이다. 이에 본 연구에서는 재난기록정보의 내용적 특징 및

1) 재난안전 유관기관(총 55개기관) 목록은 〈표 7〉 참조
 2) 국민안전처(現 행정안전부), 국립재난안전연구원, 한국원자력안전위원회, 한강홍수통제소, 농림축산검역본부

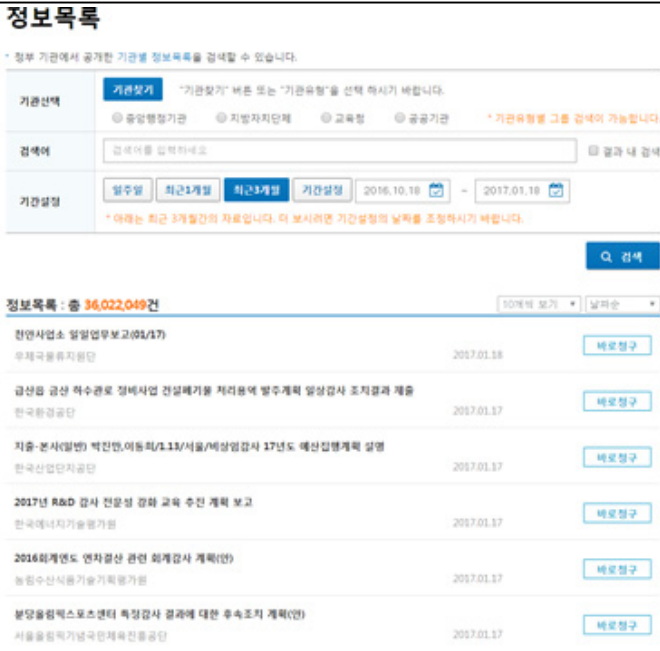
〈표 3〉 게시판의 유형과 예시

유형	예시
<p>하단에 페이지 번호가 표기된 경우</p>	
<p>번호, 제목, 작성, 등록일 등의 게시판 형식을 따르는 경우</p>	

〈표 4〉 페이지의 유형과 예시

유형	예시
<p>게시판 형식이 아닌 단순 자료의 나열</p>	

유형	예시																																																																																																																									
<p>통계수치를 단순히 나열</p>	<p>* 총 228건 (1/23 페이지) * 검색년도 : 2014년 * 검색지역 : 전체지역 * 검색업종명 : 전체업종 * 검색종류 : 전체종류 * 검색업종 : 전체업종</p> <p>* 다음 있는 항목명을 마우스로 클릭하면, 검색가능을 사용할 수 있습니다.</p> <table border="1"> <thead> <tr> <th>CAS No.</th> <th>화학물질명</th> <th>백분 비율%</th> <th>대기중농도 Omg/L</th> <th>수중중농도 Omg/L</th> <th>토양중농도 Omg/L</th> <th>백분비 Omg/L</th> <th>잔류농도 Omg/L</th> <th>폐수미포함 Omg/L</th> <th>배기중농도 Omg/L</th> <th>외포함 Omg/L</th> </tr> </thead> <tbody> <tr> <td>001330-20-7</td> <td>자일렌(C₈H₁₀, p-이성질체 혼합물)</td> <td>517</td> <td>17,660,860</td> <td>7</td> <td>0</td> <td>17,660,867</td> <td>0</td> <td>115,629</td> <td>12,896,596</td> <td>13,002,225</td> </tr> <tr> <td>000108-88-3</td> <td>톨루엔</td> <td>633</td> <td>8,537,509</td> <td>309</td> <td>0</td> <td>8,537,818</td> <td>0</td> <td>233,067</td> <td>15,273,220</td> <td>15,506,286</td> </tr> <tr> <td>000141-78-6</td> <td>아세트산 에틸</td> <td>268</td> <td>4,221,517</td> <td>182</td> <td>0</td> <td>4,221,699</td> <td>0</td> <td>342,938</td> <td>6,826,507</td> <td>7,169,445</td> </tr> <tr> <td>000079-93-3</td> <td>에틸 에틸 케톤</td> <td>297</td> <td>3,421,560</td> <td>0</td> <td>0</td> <td>3,421,560</td> <td>0</td> <td>456,968</td> <td>3,964,443</td> <td>3,921,411</td> </tr> <tr> <td>000100-41-4</td> <td>에틸벤젠</td> <td>124</td> <td>2,821,071</td> <td>1</td> <td>0</td> <td>2,821,072</td> <td>0</td> <td>12,353</td> <td>1,082,843</td> <td>1,095,196</td> </tr> <tr> <td>000067-56-1</td> <td>에틸 알코올</td> <td>523</td> <td>2,658,950</td> <td>57,192</td> <td>0</td> <td>2,716,142</td> <td>0</td> <td>1,507,668</td> <td>15,928,511</td> <td>17,436,379</td> </tr> <tr> <td>000075-09-2</td> <td>디클로로메탄</td> <td>127</td> <td>2,587,899</td> <td>1</td> <td>0</td> <td>2,587,900</td> <td>0</td> <td>431,198</td> <td>6,190,835</td> <td>8,622,033</td> </tr> <tr> <td>000067-63-0</td> <td>n-프로판올</td> <td>334</td> <td>2,295,614</td> <td>17,509</td> <td>0</td> <td>2,313,123</td> <td>0</td> <td>14,734,393</td> <td>15,476,808</td> <td>30,211,202</td> </tr> <tr> <td>000068-12-2</td> <td>N,N-디에틸아세트아미드</td> <td>141</td> <td>1,124,120</td> <td>0</td> <td>0</td> <td>1,124,120</td> <td>0</td> <td>3,264,991</td> <td>12,976,394</td> <td>16,241,386</td> </tr> <tr> <td>000106-97-9</td> <td>부탄</td> <td>43</td> <td>741,844</td> <td>14</td> <td>0</td> <td>741,858</td> <td>0</td> <td>43</td> <td>22,762</td> <td>22,805</td> </tr> </tbody> </table> <p>< 처음 < 이전 1 2 3 4 5 6 7 8 9 10 다음 > 끝 ></p>	CAS No.	화학물질명	백분 비율%	대기중농도 Omg/L	수중중농도 Omg/L	토양중농도 Omg/L	백분비 Omg/L	잔류농도 Omg/L	폐수미포함 Omg/L	배기중농도 Omg/L	외포함 Omg/L	001330-20-7	자일렌(C ₈ H ₁₀ , p-이성질체 혼합물)	517	17,660,860	7	0	17,660,867	0	115,629	12,896,596	13,002,225	000108-88-3	톨루엔	633	8,537,509	309	0	8,537,818	0	233,067	15,273,220	15,506,286	000141-78-6	아세트산 에틸	268	4,221,517	182	0	4,221,699	0	342,938	6,826,507	7,169,445	000079-93-3	에틸 에틸 케톤	297	3,421,560	0	0	3,421,560	0	456,968	3,964,443	3,921,411	000100-41-4	에틸벤젠	124	2,821,071	1	0	2,821,072	0	12,353	1,082,843	1,095,196	000067-56-1	에틸 알코올	523	2,658,950	57,192	0	2,716,142	0	1,507,668	15,928,511	17,436,379	000075-09-2	디클로로메탄	127	2,587,899	1	0	2,587,900	0	431,198	6,190,835	8,622,033	000067-63-0	n-프로판올	334	2,295,614	17,509	0	2,313,123	0	14,734,393	15,476,808	30,211,202	000068-12-2	N,N-디에틸아세트아미드	141	1,124,120	0	0	1,124,120	0	3,264,991	12,976,394	16,241,386	000106-97-9	부탄	43	741,844	14	0	741,858	0	43	22,762	22,805
CAS No.	화학물질명	백분 비율%	대기중농도 Omg/L	수중중농도 Omg/L	토양중농도 Omg/L	백분비 Omg/L	잔류농도 Omg/L	폐수미포함 Omg/L	배기중농도 Omg/L	외포함 Omg/L																																																																																																																
001330-20-7	자일렌(C ₈ H ₁₀ , p-이성질체 혼합물)	517	17,660,860	7	0	17,660,867	0	115,629	12,896,596	13,002,225																																																																																																																
000108-88-3	톨루엔	633	8,537,509	309	0	8,537,818	0	233,067	15,273,220	15,506,286																																																																																																																
000141-78-6	아세트산 에틸	268	4,221,517	182	0	4,221,699	0	342,938	6,826,507	7,169,445																																																																																																																
000079-93-3	에틸 에틸 케톤	297	3,421,560	0	0	3,421,560	0	456,968	3,964,443	3,921,411																																																																																																																
000100-41-4	에틸벤젠	124	2,821,071	1	0	2,821,072	0	12,353	1,082,843	1,095,196																																																																																																																
000067-56-1	에틸 알코올	523	2,658,950	57,192	0	2,716,142	0	1,507,668	15,928,511	17,436,379																																																																																																																
000075-09-2	디클로로메탄	127	2,587,899	1	0	2,587,900	0	431,198	6,190,835	8,622,033																																																																																																																
000067-63-0	n-프로판올	334	2,295,614	17,509	0	2,313,123	0	14,734,393	15,476,808	30,211,202																																																																																																																
000068-12-2	N,N-디에틸아세트아미드	141	1,124,120	0	0	1,124,120	0	3,264,991	12,976,394	16,241,386																																																																																																																
000106-97-9	부탄	43	741,844	14	0	741,858	0	43	22,762	22,805																																																																																																																
<p>조직도</p>	<p>전체 국문이름 검색내용 조회</p> <table border="1"> <thead> <tr> <th>이름</th> <th>소속</th> <th>전화</th> <th>이메일</th> <th>담당업무</th> </tr> </thead> <tbody> <tr> <td>김성희</td> <td>측량과</td> <td>051-638-9815</td> <td>측량</td> <td></td> </tr> <tr> <td>임관장</td> <td>영해측지</td> <td>051-400-4280</td> <td>영해측지 담당 업무 총괄</td> <td></td> </tr> <tr> <td>노경식</td> <td>인사서무</td> <td>051-400-4110</td> <td>인사서무 담당</td> <td></td> </tr> <tr> <td>김효원</td> <td>기획분석</td> <td>051-400-4353</td> <td>연구지원</td> <td></td> </tr> <tr> <td>도승민</td> <td>역측연구</td> <td>051-400-4364</td> <td>.</td> <td></td> </tr> <tr> <td>김기미</td> <td>기획예산</td> <td>051-400-4123</td> <td>홍보</td> <td></td> </tr> <tr> <td>진형태</td> <td>통해호호</td> <td>033-535-1716</td> <td>조사선 관리운영</td> <td></td> </tr> <tr> <td>김이지</td> <td>영해측지</td> <td>051-400-4284</td> <td>영해</td> <td></td> </tr> <tr> <td>박현규</td> <td>해양2000호</td> <td>051-633-1902</td> <td>선박</td> <td></td> </tr> <tr> <td>김정섭</td> <td>바다로1호</td> <td>051-633-1902</td> <td>선박</td> <td></td> </tr> </tbody> </table> <p>< 처음 < 이전 1 2 3 4 5 6 7 8 9 10 다음 > 끝 ></p>	이름	소속	전화	이메일	담당업무	김성희	측량과	051-638-9815	측량		임관장	영해측지	051-400-4280	영해측지 담당 업무 총괄		노경식	인사서무	051-400-4110	인사서무 담당		김효원	기획분석	051-400-4353	연구지원		도승민	역측연구	051-400-4364	.		김기미	기획예산	051-400-4123	홍보		진형태	통해호호	033-535-1716	조사선 관리운영		김이지	영해측지	051-400-4284	영해		박현규	해양2000호	051-633-1902	선박		김정섭	바다로1호	051-633-1902	선박																																																																			
이름	소속	전화	이메일	담당업무																																																																																																																						
김성희	측량과	051-638-9815	측량																																																																																																																							
임관장	영해측지	051-400-4280	영해측지 담당 업무 총괄																																																																																																																							
노경식	인사서무	051-400-4110	인사서무 담당																																																																																																																							
김효원	기획분석	051-400-4353	연구지원																																																																																																																							
도승민	역측연구	051-400-4364	.																																																																																																																							
김기미	기획예산	051-400-4123	홍보																																																																																																																							
진형태	통해호호	033-535-1716	조사선 관리운영																																																																																																																							
김이지	영해측지	051-400-4284	영해																																																																																																																							
박현규	해양2000호	051-633-1902	선박																																																																																																																							
김정섭	바다로1호	051-633-1902	선박																																																																																																																							
<p>용어사전</p>	<p>해양조사 용어사전</p> <p>가 나 다 순 [전체] 가 나 다 라 마 바 사 아 자 차 카 타 파 하 A B C 순 [전체] A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 용어검색 검색내용 조회</p> <p>* 총 1166개의 검색결과가 있습니다. (1/117페이지)</p> <p>1일1회조 (single day tide) 하루에 한 번 고조와 저조가 나타나는 조석을 말한다. 일주조라고도 하며, 1일2회조에 대비되는 용어다. 일조부들의 크기가 매우 클 때에는 저고조 및 고저조가 거의 소멸되어 1일1회조가 된다. 주기는 약 24시간 50분이며, 고리도 지방의 바다에서 1일1회조가 주로 나타난다.</p> <p>1일2회조 (double day tide) 하루에 두 번 고조와 저조가 나타나는 조석을 말한다. 반일주조라고도 하며, 1일 1회조에 대비되는 용어다. 주기는 약 12시간 25분이다. 지구의 적도면이 달의 공전 궤도면과 약 28° 기울어져 있기 때문에 위도에 따라 고조와 저조가 일어나는 횟수는 달라진다. 반일주조는 적도지역에서 우세하게 일어나고 극지방에서는 일주조가 우세하다.</p>																																																																																																																									

유형	예시
정보공개포털의 정보목록	 <p>The screenshot shows the '정보목록' (Information List) page of a government portal. It includes search filters for '기관선택' (Agency Selection) with options like '중앙행정기관', '지방자치단체', '교육청', and '공공기관'. There are also date range filters for '기간설정' (Date Range) with options like '일주일', '최근3개월', '최근3개월', and '기간설정'. Below the filters, a search button '검색' is visible. The main content area displays a table of records with columns for document title, date, and a '바로보기' (View) button. The first record is '친연사업소 설립업무보고(09/17)' dated 2017.01.18. Other records include '급산급산 특수교육 장애학생 진로교육 지원사업 실적조사 조치결과 제출' (2017.01.17), '지출-본사(일본) 적진현,이동희(1.13/서울/예산)임금사 17년도 예산집행계획 설명' (2017.01.17), '2017년 R&D 감사 전문성 강화 교육 추진 계획 보고' (2017.01.17), '2016회계연도 연차결산 관련 회계감사 계획(안)' (2017.01.17), and '분당울림텍스트센터 특정감사 결과에 대한 후속조치 계획(안)' (2017.01.17).</p>

<표 5> 재난안전 유관기관의 기록정보자원 자동수집 가능성 분석 기준

필드명	분석기준
유형 (게시판, 페이지)	<ul style="list-style-type: none"> • 게시판: 웹 페이지 유형이 게시판 형식일 경우(<표 3> 참조) <ul style="list-style-type: none"> - 하단에 페이지 번호가 표기된 유형 - 번호, 제목, 작성자, 등록일 등의 게시판 형식을 따르는 경우 • 페이지: 단일 페이지일 경우 <ul style="list-style-type: none"> ※ 다음의 경우 게시판의 형식으로 보여도 페이지로 구분함(<표 4> 참조) <ul style="list-style-type: none"> - 게시판 형식이 아닌 단순 자료의 나열은 페이지로 구분 - 통계 수치를 나열한 경우 페이지로 구분 - 조직도는 페이지로 구분 - 용어사전은 페이지로 구분 - 정보공개포털의 정보목록은 페이지로 구분 • 게시판과 페이지 내용이 복합된 웹페이지일 경우에 게시판과 페이지 단위로 나누어 각각 작성한다.
예상건수	<ul style="list-style-type: none"> • 링크가 연결되는 정보 페이지에서 크롤링이 가능한 예상 게시물 건수를 작성 • 작성 요령: <ul style="list-style-type: none"> - "유형(게시판, 페이지)" 필드의 입력값이 "페이지"인 경우 : 1 - "유형(게시판, 페이지)" 필드의 입력값이 "게시판"인 경우 : 총 게시물 수 ex. 마지막 게시물의 목록 번호가 456인 경우 "456"으로 작성
수집건수	<ul style="list-style-type: none"> • 실제로 크롤링을 통해 수집한 게시물 건수를 작성

구조적 특징을 직접 분석하기 위하여 재난안전
유관기관 55개 중 국민안전처, 국립재난안전연
구원, 원자력안전위원회, 농림축산검역본부, 한
강홍수통제소, 총 5개 기관을 선정하였다. 기관
선정 기준은 다음과 같다.

먼저 국민안전처는 국민의 안전과 국가적 재
난관리를 위한 재난안전 총괄기관으로서, 모든
재난안전유형에 관한 정보를 생산, 접수, 가공,
서비스하고 있으며, 유관기관 중 가장 많은 양
의 재난기록정보를 다루고 있다. 국립재난안전
연구원은 국가 재난 및 안전관리 총괄 연구기
관으로서 모든 재난안전유형에 관한 실용적 재
난관리기술을 연구하고, 연구 결과물인 재난기
록정보를 생산, 발행, 서비스하고 있어 선정하
였다. 원자력안전위원회의 경우 일본 지진에서
유발된 후쿠시마 원전사고와 최근 경주 지진으
로 인하여 방사능물질사고에 대한 우려가 커짐
에 따라 해당기관의 재난기록정보가 중요하
다고 판단하였다. 농림축산검역본부의 경우 최근
AI와 구제역의 빈번한 발생으로 인하여 해당
기관에서 제공하는 정보가 중요하다고 판단하
여 선정하였다. 마지막으로 한강홍수통제소는
최근 10년간 우리나라에서 가장 많이 발생한 대
형 자연재난은 대부분 물과 관련된 유형이었기
때문에 해당기관의 재난기록정보의 사용 빈도
가 높을 것으로 예상하여 선정하였다.

상기 기관에서 생산되는 재난기록정보의 내
용적, 구조적 특징을 분석한 결과와 재난안전
기록정보자원 아카이브 구축을 위해 수집하고
자 하는 기록물의 정의 및 범위를 기준으로 본
연구에서는 다음의 경우를 아카이빙 대상에서
제외하였다. 아카이빙 대상에서 제외하는 경우
는 첫째, 재난기록정보와 내용적으로 직접적인

성이 없는 기관의 일반적인 행정업무내용으로
서 계약, 입찰, 채용 정보는 아카이빙에서 제외
한다. 둘째, 실시간 정보와 데이터는 제외한다.
기록은 단순한 데이터의 집합이 아니며 업무활
동과 연계된 특정 사안의 결과나 산출물이다.
따라서 정보시스템의 데이터가 누적된 결과물
로서의 데이터베이스 자체가 기록이 될 수 없으
며, 고정된 형태로 존재해야 기록으로 볼 수 있
다. 즉, 재난기록정보 중 데이터 유형은 내용이
고정된 형태로 존재하지 않고, 데이터가 누적된
결과물로 볼 수 있어 아카이빙 대상에서 제외하
다. 셋째, 정보를 포함하는 웹페이지가 아닌 메
인페이지로 연결되는 경우는 아카이빙 대상에
서 제외한다. 즉, 특정한 정보가 포함되어 있지
않는 단순한 메인홈페이지 화면은 아카이빙 하
지 않는다. 마지막으로 자주 묻는 질문과 일반
민원 게시관 역시 아카이빙 대상에서 제외한다.
재난안전 기록정보자원 아카이브는 신뢰성 있는
기록정보자원의 제공을 목표로 하기 때문이다.
따라서 민원게시관의 글은 신뢰성 있는 기관에
의해 작성된 내용이 아니므로 제외한다. <표 6>
은 재난안전 기록정보자원 아카이브 구축에서
제외하는 정보의 내용이다.

<표 6> 재난안전 기록정보자원 아카이브
구축 제외 대상

No.	정보의 내용
1	계약정보, 입찰, 채용
2	실시간 정보, 데이터
3	정보를 포함하는 웹페이지가 아닌 메인페이지 (ex. 서브시스템 메인페이지)
4	자주 묻는 질문, 일반 민원 게시관

마지막 4단계는 아카이빙 대상으로 선정된

재난기록정보에 대한 메타데이터의 자동수집 가능성을 분석하는 단계이다. 즉, 3단계까지의 결과물에 대하여 웹 크롤러를 구현하여 재난기록정보의 목록 및 위치정보에 대한 메타데이터를 수집한다. 이에 본 연구에서는 상기 5개 기관의 재난기록정보 목록 및 위치정보 수집을 위해 직접 크롤러를 개발하여 메타데이터를 자동수집 하였다. 자동수집을 위한 메타데이터 요소는 1단계에서 작성한 세부정보를 바탕으로 재난기록정보의 식별이 가능한 최소한의 요소로 선정하였으며, 제목, 원출처, URL, 연도, 첨부파일명, 첨부파일링크, 기술내용, 사이트경로 등이 있다.

3.2 검증 결과

3.2.1 재난기록정보 전수조사 및 자동수집 가능성 분석

본 연구에서는 국민안전처 및 재난안전 유관기관 총 55개 기관을 대상으로 이들 기관이 보유한 재난기록정보의 현황을 파악하고, 이들 기록물에 대한 자동수집 가능성을 분석하였다. 전수조사 결과는 실제 재난안전 유관기관이 보유한 기록정보자원의 현황이며, 자동수집 가능성 분석 결과는 이러한 기록정보자원 중 크롤링이 가능한 경우로써 그 결과는 <표 7>과 같다.

<표 7> 재난안전 유관기관 전수조사 및 자동수집 가능성 분석 결과

no.	기관명	전수조사 결과				자동수집 가능성 분석 결과			
		게시판	게시물수	개별 페이지 수	예상수집 건수	게시판	게시물수	개별 페이지 수	예상수집 건수
1	국립재난안전연구원	31	2,215	44	2,259	31	2,215	44	2,259
2	원자력안전위원회	34	3,338	109	3,447	34	3,338	108	3,446
3	한국가스안전공사	78	126,954	269	127,223	78	126,954	268	127,222
4	화학물질안전원	29	31,755	58	31,813	29	31,755	57	31,812
5	부산시설공단	10	1,616	123	1,739	10	1,616	122	1,738
6	한국국토정보공사	18	1,548	21	1,569	18	1,548	21	1,569
7	국토교통부	33	162,433	137	162,570	33	162,433	136	162,569
8	산림청	27	577,468	80	577,548	27	577,468	75	577,543
9	국립해양조사원	22	364,392	127	364,519	22	364,392	89	364,481
10	경찰청	39	13,931	204	14,135	39	13,931	201	14,132
11	농림축산검역본부	116	42,444	225	42,669	116	42,444	217	42,661
12	한국수력원자력	109	7,502	132	7,634	109	7,502	132	7,634
13	국립공원관리공단	19	39,636	93	39,729	19	39,636	90	39,726
14	국립산림과학원	40	36,407	86	36,493	40	36,407	86	36,493
15	인천국제공항공사	22	4,475	51	4,526	22	4,475	51	4,526
16	국립수산과학원	34	60,926	277	61,203	34	60,926	274	61,200
17	한국소방안전협회	33	43,053	99	43,152	33	43,053	91	43,144
18	한국정보화진흥원	29	6,443	8	6,451	29	6,443	8	6,451
19	한국수자원공사	71	54,583	275	54,858	71	54,583	246	54,829

no.	기관명	전수조사 결과				자동수집 가능성 분석 결과			
		게시판	게시물수	개별 페이지 수	예상수집 건수	게시판	게시물수	개별 페이지 수	예상수집 건수
20	한국시설안전공단	67	8,457	264	8,721	67	8,457	255	8,712
21	식품의약품안전처	155	39,272	113	39,385	155	39,272	104	39,376
22	질병관리본부	112	41,125	452	41,577	112	41,125	452	41,577
23	한강홍수통제소	17	10,727	410	11,137	17	10,727	363	11,090
24	한국전기안전공사	73	33,161	132	33,293	73	33,161	93	33,254
25	미래창조과학부	41	36,688	11	36,699	41	36,688	11	36,699
26	한국철도공사	16	49,710	71	49,781	16	49,710	71	49,781
27	한국원자력안전기술원	63	11,728	326	12,054	63	11,728	320	12,048
28	행정자치부	74	1,070,626	33	1,070,659	74	1,070,626	33	1,070,659
29	산림항공본부	6	14,350	10	14,360	6	14,350	10	14,360
30	해양수산부	55	20,590	96	20,686	55	20,590	91	20,681
31	환경부	150	771,227	231	771,458	150	771,227	230	771,457
32	산업통상자원부	33	24,924	28	24,952	33	24,924	28	24,952
33	기상청	79	59,265	529	59,794	79	59,265	451	59,716
34	한국산업안전보건공단	226	30,172	308	30,480	226	30,172	304	30,476
35	한국항공공사	20	3,805	22	3,827	20	3,805	22	3,827
36	한국토지주택공사	108	30,680	296	30,976	108	30,680	296	30,976
37	한국승강기안전공단	39	3,547	125	3,672	39	3,547	110	3,657
38	고용노동부	28	15,935	48	15,983	28	15,935	48	15,983
39	한국철도시설공단	23	6,113	83	6,196	23	6,113	83	6,196
40	국민안전처	295	413,914	573	414,487	295	413,914	560	414,474
41	교통안전공단	100	10,672	133	10,805	100	10,672	133	10,805
42	항공철도사고조사위원회	12	10,350	17	10,367	12	10,350	17	10,367
43	해양환경관리공단	61	5,624	121	5,745	61	5,624	121	5,745
44	대한건설기계협회	4	535	10	545	4	535	10	545
45	조달청	1	2,110	4	2,114	1	2,110	4	2,114
46	한국가스공사	2	187	10	197	2	187	10	197
47	대한적십자사	34	30,270	147	30,417	34	30,270	147	30,417
48	건강보험심사평가원	18	7,473	87	7,560	18	7,473	87	7,560
49	국가교통정보센터	4	2,228	70	2,298	4	2,228	70	2,298
50	한국농어촌공사	38	257,580	288	257,868	38	257,580	288	257,868
51	한국환경공단	196	89,475	794	90,269	196	89,475	794	90,269
52	한국전력거래소	23	2,636	23	2,659	23	2,636	23	2,659
53	한국도로공사	48	6,334	372	6,706	48	6,334	372	6,706
54	한국전력공사	26	3,401	82	3,483	26	3,401	82	3,483
55	문화재청	31	19,471	49	19,520	31	19,471	49	19,520
총 계		3,072	4,725,481	8,786	4,734,267	3,072	4,725,481	8,458	4,733,939

※ 예상수집건수: 게시물 수 + 페이지 수

재난안전 유관기관의 재난기록정보에 대한 전수조사를 실시한 결과 예상 수집 건수는 4,734,267건이었으며, 이 중 자동수집이 가능한 건수는 4,733,939건이었다. 자동수집 가능성을 분석한 결과 약 99%가 크롤링이 가능하다는 결론을 도출하였다(〈표 8〉 참조).

3.2.2 아카이빙 대상 선별

기준에 1, 2단계에서 수집된 재난기록정보의 수작업 검수를 위해 55개의 기관 중 5개의 기관

을 선정하여 실제 자동수집을 검증하고자 하였다. 이에 앞서 〈표 6〉을 기반으로 전수조사 목록 중 아카이빙 대상 선별을 진행하였다. 5개의 기관 중 먼저 재난안전 유관기관 4개 기관(국립재난안전연구원, 한국원자력안전위원회, 한강홍수통제소, 농림축산검역본부)을 대상으로 하여 재난기록정보 아카이빙 선별 기준을 적용하여 분석해 본 결과 약 81.5%가 아카이빙 대상인 것으로 나타났다(〈표 9〉 참조).

한편, 국민안전처의 경우 자동수집 가능한

〈표 8〉 재난기록정보 전수조사 결과 및 자동수집 가능성 비교

(2017.04.17. 기준)

	재난기록정보 전수조사	자동수집 가능 재난기록정보	
게시판 수	3,072	3,072	<p>재난기록정보 자동수집 가능성 99%</p> <p>예상수집: 4,734,267(건) 자동수집: 4,733,939(건)</p>
게시물 수	4,725,481	4,725,481	
개별페이지 수	8,786	8,458	
예상수집 건수 (게시물 수+페이지 수)	4,734,267	4,733,939	

〈표 9〉 재난안전 유관기관 - 재난기록정보 자동화 및 아카이빙 선별 예상 수집

(2017.04.17. 기준)

	재난기록정보 전수조사	자동수집 가능 재난기록정보	선별된 아카이빙 대상	
게시판 수	198	198	182	<p>재난기록정보 아카이빙 대상 81.5%</p> <p>예상수집: 59,512(건) 자동수집: 59,456(건) 아카이빙 대상: 48,486(건)</p>
게시물 수	58,724	58,724	47,811	
개별페이지 수	788	732	675	
예상수집 건수 (게시물 수 + 페이지 수)	59,512	59,456	48,486	

※ 분석대상: 국립재난안전연구원, 한국원자력안전위원회, 한강홍수통제소, 농림축산검역본부

재난기록정보를 대상으로 아카이빙 선별기준을 적용하여 분석한 결과 약 53.8%가 아카이빙 대상인 것으로 나타났다. 이는 국민안전처의 재난기록정보에는 실시간 정보가 많이 포함되어 있어 아카이빙 대상에서 상당수 제외되었기 때문이다(〈표 10〉 참조).

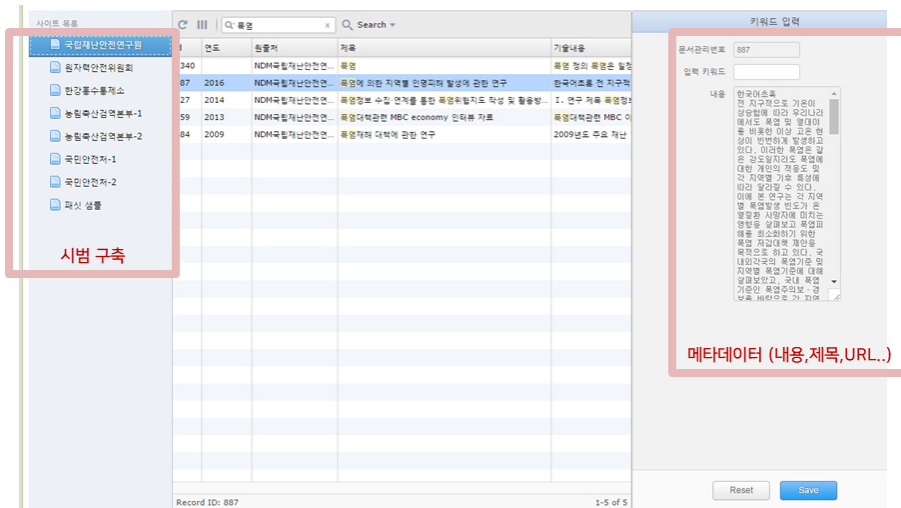
3.2.3 재난기록정보 목록 및 메타데이터 자동 수집 가능성 분석
3단계까지는 자동수집 가능성을 예측한 결과

로서 실제 자동수집 여부에 대해서는 크롤러를 직접 구현하여 그 결과를 비교해 볼 필요가 있다. 이에 본 연구에서는 직접 크롤러를 구현하여 5개 기관의 재난기록정보 목록 및 위치 정보에 대한 메타데이터를 자동수집 하였다. 〈그림 1〉은 실제로 시범 구축한 재난안전 기록정보자원 아카이브이며, 여기에 자동수집된 메타데이터 요소는 제목, 원출처, URL, 연도, 첨부파일명, 첨부파일링크, 기술내용, 목록번호, 조회수, 사이트경로, InsertDBTime이다.

〈표 10〉 국민안전처 - 재난기록정보 자동화 및 아카이빙 선별 예상 수집

(2017.04.17. 기준)

	재난기록정보 전수조사	자동수집 가능 재난기록정보	선별된 아카이빙 대상
게시판 수	295	295	242
게시물 수	413,914	413,914	222,645
개별페이지 수	573	560	431
예상수집 건 수 (게시물 수+페이지 수)	414,487	414,474	223,076



〈그림 1〉 재난안전 기록정보자원 아카이브 시범구축 화면

4개의 재난안전 유관기관 기관에서 생산된 재난기록정보 중 아카이빙 대상으로 선별된 기록정보의 목록 및 위치정보 메타데이터 자동수집 가능성을 분석한 결과 약 96.5%가 실제 수집이 가능한 것으로 나타났다. <표 11>은 재난안전 유관기관 4개 기관을 대상으로 재난기록정보의 목록 및 위치정보 메타데이터를 자동수집한 결과표이다.

한편, 국민안전처의 재난기록정보 중 아카이

빙 대상으로 선별된 기록정보의 목록 및 위치정보 메타데이터 자동수집 가능성을 분석한 결과 기술내용은 약 98% 수집이 가능한 반면, 첨부파일명은 약 6%만 수집 가능하였다. 이는 국민안전처가 제공하는 재난기록정보 대부분이 기술내용을 포함되어 있음을 의미하므로, 다른 재난안전 유관기관의 기록정보보다 정확한 정보 추출 및 분류가 가능할 것으로 예상된다 (<표 12> 참조).

<표 11> 재난안전 유관기관 - 재난기록정보 목록 및 위치정보에 대한 메타데이터 자동수집 (2017.06.28. 기준)

요소	건수	%
제목/원출처/URL	46,812	100
연도	40,637	86.8
첨부파일명	27,039	57.8
첨부파일링크	27,036	57.8
기술내용	19,403	41.4
목록번호	44,635	95.3
조회수	29,507	63.0
사이트경로	46,801	100
InsertDBTime	46,812	100

※ 분석대상: 국립재난안전연구원, 한국원자력안전위원회, 한강홍수통제소, 농림축산검역본부

<표 12> 국민안전처 - 재난기록정보 목록 및 위치정보에 대한 메타데이터 자동수집

(2017.06.28. 기준)

	제목/원출처/URL	연도	첨부파일명	첨부파일링크	기술내용	목록번호	조회수	사이트경로	InsertDBTime
건수	226,284	97,325	15,035	849	222,211	136,091	58,729	226,284	226,284
%	100	43.0	6.6	0.4	98.2	60.1	26.0	100	100

4. 실증적 검증 결과 및 개선 방안

4.1 검증 결과 및 시사점

상기의 실증적 검증을 통해 발견한 문제점은 다음과 같다.

첫째, 재난기록정보를 생산하는 기관(재난안전 유관기관)의 범위를 선정하는 데 어려움이 있었다. 본 연구에서는 실험 대상 재난안전 유관기관의 범위를 선행연구(한희정 외 2017a)에서 제시한 36개 기관을 중심으로 국민안전처와의 협의를 통해 55개로 확대하였다. 그러나 전수조사를 진행하는 과정에서 55개 기관 목록에는 포함되어 있지 않았으나, 재난기록정보를 생산하는 산하 기관 및 서브시스템이 빈번히 등장하였다. 이에 재난기록정보 생산기관의 범위를 어떻게 정할 것이며, 만약 확장시킨다면 무엇에 근거할 것인가에 대한 논의가 필요하였다.

둘째, 재난안전 유관기관에서 생산하고 있는 정보의 위치를 파악하는데 어려움이 있었다. 본 연구는 연구 범위를 온라인에서 제공하고 있는 재난기록정보로 한정하였음에도 불구하고 이에 대한 전수조사에는 많은 시간과 노력이 소요되었다. 이는 재난안전 유관기관이 해당 기관에서 제공하고 있는 온라인 재난기록정보에 대한 내용을 모르고 있을 뿐만 아니라, 각 기관별로 동일한 정보를 동시에 제공함에도 서로 공유가 되지 않았다는 점에서 기인한다. 즉, 재난안전 유관기관은 해당 기관의 온라인 재난기록정보 목록을 관리하고 있지 않으며, 이를 서로 공유하고 있지 않았다.

셋째, 브로큰 링크(broken links)로 인하여

재난기록정보의 식별이 불가능한 경우가 발생하였다. 본 연구는 55개 재난안전 유관기관의 홈페이지를 조사하였기 때문에 상당한 시간이 소요되었는데, 그 기간 동안 기관의 홈페이지 개편이나 게시물의 등록과 삭제가 빈번하게 발생하였다. 이로써 URL로 식별하였던 재난기록정보의 위치정보를 모두 잃어버린 경우가 발생하였으며, 각 단계를 진행하면서도 이러한 경우가 자주 발생하였다.

넷째, 적절한 수집 주기를 정하는데 어려움이 있었다. 본 연구는 5개의 대표기관을 선정하여 실제 크롤링을 수행하였는데, 각 기관별 업데이트 주기가 상이하야 수집 주기를 정하는데 어려움이 있었다.

다섯째, 중복 수집이 발생하였다. 본 연구는 실험 대상 유관기관의 홈페이지 분석을 통하여 재난기록정보를 조사하였는데, 이 때 하나의 기록정보를 다수의 게시판에서 확인할 수 있는 경우가 있었다. 이에 이러한 경우의 중복 수집을 방지하기 위하여 중복 제거 과정을 적용했음에도 불구하고 일부 중복 수집이 발생하였다. 또한 크롤러가 일정한 주기를 가지고 수집한 목록에 대해서도 중복이 확인되었다. 이런 경우는 동일한 재난기록정보임에도 불구하고 덮어쓰기가 되지 않았거나 내용은 거의 일치하지만 텍스트에서 완전일치가 일어나지 않은 경우였다.

여섯째, 메타데이터 요소 선정에 어려움이 있었다. 본 연구는 한희정 외(2017b)에서 제안한 아카이브 구축 전략의 실증적 검증에 목적을 두고 있기 때문에 기록 식별을 위한 최소한의 메타데이터 요소만을 선정하였다. 하지만 10개 남짓한 메타데이터 요소마저도 자동 수집이 어려

운 경우가 있었기 때문에 자동 수집을 위한 메타데이터 요소에 대해 다시 고민해볼 필요가 있다.

일곱째, 재난안전 기록정보 아카이브 구축 이후 이용자들의 재난기록정보에 대한 접근과 실질적 이용을 위한 전략의 부재이다. 4단계의 구축전략에는 아카이브 구축 후 사용을 위한 요소에 대해서는 설계되어 있지 않다. 예컨대, 인터페이스의 구성, 서비스 대상 이용자층의 분석이나 그룹화, 서비스 제공 방식과 같은 전략이 부재한 것이다. 비록 한희정 외(2017b)에서는 활용방안 측면에서 인터페이스 일부를 제안하고 있으나, 이것만으로는 자동 수집이 필요한 정보들을 선정하는데 어려움이 있다. 앞서 언급한 메타데이터 요소 선정의 경우도 보다 명확한 서비스 방향이 제공되었다더라면 근거에 기반한 요소 선정이 이루어졌을 것이다.

상기의 문제점은 정책적, 기술적, 관리적, 서비스적 측면으로 나누어 질수 있으며 이에 대한 시사점을 정리하면 <표 13>과 같으며 이를 위한 개선방향을 4.2에서 자세히 기술하였다.

4.2 개선 방안

4.2.1 정책적 측면

재난안전 기록정보자원 아카이브 구축을 위해 기 생산된 재난기록정보에 대해 보다 구체적인 수집정책을 마련해야 한다. 현재로서는 「재난 및 안전관리 기본법 74조의 2(재난관리정보의 공동이용)」와 「재난 및 안전관리 기본법 시행령 83조(재난관리정보의 공동이용절차 등)」에서 공동이용 재난관리정보의 범위와 의무조항이 명시되어 있지만 상시적 공유는 현실과 거리가 멀다. 따라서 재난기록정보에 대한 포괄적, 상시적인 정보공유를 위한 근거를 마련하는 한편 재난기록정보 보유기관장의 정보제공 의무의 실효성을 확보할 수 있는 방안을 모색해야 한다.

이를 위해서는 다음과 같은 방안을 고려해 볼 수 있다.

첫째, 재난기록정보의 디지털 납본 제도를 통한 수집 방안이다. 재난기록정보 디지털 납본 제도와 같은 법 제도를 만들어 재난기록정

<표 13> 실증적 검증 결과 및 시사점

구분	문제점	시사점
정책적 측면	<ul style="list-style-type: none"> 재난안전 유관기관의 범위 선정 문제 재난기록정보의 목록 및 위치정보의 독점 	<ul style="list-style-type: none"> 재난기록정보의 디지털 납본 제도 의무화 재난안전 유관기관의 연계 및 제휴 저작권 문제에 대한 제도적 장치 마련
기술적 측면	<ul style="list-style-type: none"> 브로큰 링크로 인한 재난기록정보의 식별 문제 수집 주기 선정 문제 중복 수집 발생 	<ul style="list-style-type: none"> 심층 웹 정보자원의 수집 방안 마련 실험적인 수집 과정을 거쳐 적절한 주기 선택
관리적 측면	<ul style="list-style-type: none"> 메타데이터 요소 선정 문제 	<ul style="list-style-type: none"> 표준 메타데이터 스키마 개발
서비스적 측면	<ul style="list-style-type: none"> 접근과 이용을 위한 전략 부재 	<ul style="list-style-type: none"> 이용자 세분화 재난안전 유관기관 간 협업기능 강화 서비스 개발 온라인과 모바일 서비스 전략 수립

보가 통합적으로 수집·관리될 수 있는 방안을 마련할 필요가 있다. 또한 이를 통해 재난기록 정보 원문에 대한 접근점이 마련될 수 있어야 한다. 둘째, 재난안전 유관기관들과 연계 및 제휴를 통한 재난기록정보 자동수집 방안이다. 즉, 재난안전 유관기관별 기록정보 생산 목록의 하베스팅 또는 해당 기관에서의 자동 업로드를 위한 규정을 제정할 필요가 있다. 특히, 거버넌스 관점에서 협업기능이 요구되는 업무를 중심으로 재난기록정보의 목록정보가 공유될 수 있도록 제도적 장치를 마련하고, 공개가능 여부에 대한 권한 관리와 정보접근단계 설정 등 제도적인 보안 장치를 마련해야 한다. 셋째, 온라인에 공개된 재난기록정보의 원문 공유를 위해서는 저작권 문제를 해결할 수 있는 제도적 방안을 마련해야 한다. 이를 통해 별도의 저작권 협의를 거치지 않아도 공유가 가능한 방안을 모색하여 행정적 낭비를 최소화 할 수 있도록 해야 한다. 예를 들어, 정보공유를 위한 자유이용 라이선스의 일종인 CCL³⁾이나 원본 콘텐츠를 보호하기 위한 저작권관리시스템(이하 “DRM”)⁴⁾의 적용을 고려해 볼 수 있다. DRM으로 디지털 콘텐츠를 보호하고, 대상 콘텐츠 사용 권한 정보를 관리할 수 있다(한희정 외 2017a).

4.2.2 기술적 측면

재난안전 기록정보자원 아카이브 구축을 위해서는 재난기록정보 아카이빙 대상에 대한 명

확한 범위 선정과 정보의 유형 등에 대한 현황이 파악되어야 한다. 이를 통해 재난안전 기록정보자원 아카이브 구축대상에 대하여 범위를 선정하고, 현황을 분석하여 구축 목표치를 명확히 한 후 재난안전 기록정보자원 아카이브 구축을 진행하여야 한다. 나아가 웹 크롤링을 이용해 산재되어 있는 재난기록정보를 수집 분석하여 아카이브를 구축하되, 표면 웹뿐만 아니라 심층 웹도 수집할 수 있는 방안을 마련해야 한다. 웹은 수시로 변하기 때문에 웹 크롤러의 웹사이트 방문주기에 따라 웹의 변화된 정보를 수집하지 못할 가능성이 있다. 그리고 일반적인 웹 크롤러는 정적으로 연결된 페이지만 방문 가능하다는 점이다. 정적인 웹페이지는 방문하는 순간 이미 모든 페이지가 완성된 형태로 제공 되어 웹 크롤러가 다음 링크를 참조하여 페이지를 이동하거나 정보를 저장하는 것이 가능하였으나, 최근에는 웹 기술의 발전으로 웹 페이지에 이용자가 방문 한 후 관련 정보를 데이터베이스에서 로드하여 페이지를 생성하는 방법을 많이 이용하고 있다. 이 같은 페이지는 그 내용이 동적으로 생성되며 크롤러가 방문한 순간에 웹 페이지에는 일부 정보만 존재한다. 이와 같은 경우 일반적인 크롤러로는 웹 페이지의 정보를 수집할 수 없다. 즉, 일반적인 웹 크롤러는 정적으로 연결된 페이지만 방문이 가능하기 때문에 일반적인 크롤러로는 심층 웹의 정보를 수집하기 어렵다. 그러나 심층 웹이 가진 정보는 정적으로 구성되는 표면 웹

3) CCL(Creative Commons License)는 원칙적으로 모든 이의 자유이용을 허용하되 몇 가지 이용방법 및 조건을 추가하는 방식의 개방적인 자유이용 라이선스이다.

4) DRM(Digital Rights Management)은 전자책(e-Book), 음악, 비디오, 게임, 이미지 등 온라인에서 유통되는 디지털 콘텐츠의 보호 및 저작권 관리 기술을 일컫는 것으로, 원본으로 배포되는 디지털 콘텐츠에 대한 불법적인 접근 및 배포를 차단하기 위한 기술을 말한다.

보다 약 450~550배 이상의 정보를 가지고 있을 것으로 추산하고 있으며, 따라서 이들 심층 웹에 있는 정보들을 크롤러를 이용해 자동으로 수집할 수 있는 방안이 필요하다(Bergman 2001; Pandya et al. 2017). 그 외 재난기록정보에 대한 적절한 수집주기를 설정할 필요가 있다. 웹 기록물의 경우 기록의 가치와는 관계없이 주기적 또는 비주기적으로 갱신되거나 소멸된다. 잦은 수정과 변경, 삭제되는 특성을 가진 웹 기록물의 특성 상 크롤러를 통한 수집이 제대로 이루어지지 않을 수 있다. 따라서 적절한 수집주기를 설정하여 재난기록정보가 안정적으로 수집될 수 있도록 해야 한다.

4.2.3 관리적 측면

먼저, 재난기록정보 관리를 위해서는 표준 메타데이터 스키마 개발이 필요하다. 즉, 재난기록정보 관리를 위한 메타데이터 스키마를 개발하여 재난기록정보 공유의 편의성이 확보될 수 있어야 한다. 이를 위해서는 우선적으로 기관별 관리 아카이브 메타데이터 현황을 파악해야 한다. 그리고 기존 메타데이터 표준들과 매핑하여 재난기록정보 관리를 위한 표준 메타데이터 스키마를 개발해야 한다. 다음으로 정보 표준화 작업이 이루어져야 한다. 재난안전 기록정보자원 아카이브 시스템에 대한 정보항목 및 용어를 표준화 대상으로 하여 현행 산재된 정보항목 요소에 대한 명명규칙, 구조 등을 정의하고, 지속적인 표준 적용과 변화관리를 수행할 수 있도록 정보항목 표준을 적용해야 한다. 기본적으로 정보항목 요소에 대해 정의 시 “행정정보데이터베이스 표준화 지침”을 준수하도록 한다. 코드 역시 행정안전부의 “행정기관의 코드

표준화 추진지침”에서 제시하는 표준화 지침을 따르고, 정보항목과 서식 가운데 표준화가 필요한 항목을 선별하여 표준화한다. 또한 행정표준이 제공되는 코드는 행정표준코드를 활용하며, 제공되지 않는 코드는 관리체계의 총괄적인 업무관점에서 타 업무 및 시스템과 정보를 공동 활용할 수 있도록 표준화하여 적용한다. 그 외에 웹 접근성 및 호환성 역시 확보해야 한다. 웹 접근성을 확보하기 위해서는 모든 콘텐츠는 사용자가 인식할 수 있고, 이해할 수 있도록 만들어져야 한다. 그리고 사용자 인터페이스 구성요소는 조작 가능하고 네비게이션할 수 있어야 하며, 콘텐츠는 미래의 기술로도 접근할 수 있도록 견고하게 만들어질 필요가 있다. 또한 어떠한 브라우저로 웹페이지에 접속하더라도 동일한 결과를 볼 수 있도록 웹 표준을 준수해야 한다.

4.2.4 서비스적 측면

재난안전 기록정보자원 아카이브의 이용대상은 재난 관련 업무를 수행하는 현장실무자들과 연구자를 포함한 일반이용자들로 예상된다. 이들의 이용목적은 각 이용자층별로 상이하기 때문에 이용자들을 세분화하여 아카이브 서비스를 제공해야 한다. 이를 위해 다음의 사항을 고려하여 재난안전 기록정보자원 아카이브를 구축하여야 한다. 이를 위해서는 먼저, 다양한 검색 조건을 갖춘 메타검색 기능과 정보연계 기능을 갖춘 재난안전 기록정보자원 아카이브를 구축해야 한다. 그리고 재난안전업무의 협업 기능을 강화할 수 있는 서비스를 개발해야 한다.

일반이용자를 위해서는 이용자서비스 방향을 정립하여, 분류체계 정립 및 각 기준별 조회

편의를 제공해야 하며, 활용 관점에서 이용자 맞춤 정보를 제공할 수 있어야 한다. 이와 관련하여 온라인 기반의 서비스 제공전략 수립하고, 대국민 서비스를 위한 모바일연계 및 앱 개발, 최신형의 반응형 웹 방식을 도입하여 필요한 재난기록정보들을 쉽고 편리하게 찾아볼 수 있도록 지원할 수 있어야 한다. 그 외에도 검색기능 및 디렉토리 서비스 제공을 위한 체계화된 분류체계를 개발해야 한다.

5. 결론

본 연구의 목적은 한희정 외(2017b)가 제안한 재난안전 기록정보자원 아카이브 구축 전략을 실증적으로 검증해보고 문제점 분석을 통해 시사점을 도출하는데 있다. 이를 위하여 재난안전 유관기관 55개를 선정하였고 해당 기관의 재난기록정보를 수집하여 실험 대상으로 삼았다. 전수조사-자동화 가능성분석-아카이빙 대상 선별-메타데이터 자동수집, 총 4단계로 구성된 아카이브 구축 전략을 기반으로 각 세부 단계별 적용을 통해 적용 가능성을 검증하였으며, 이를 분석하여 문제점과 시사점을 도출하였다. 먼저 문제점으로는 1) 재난기록정보 생산기관 범위 선정 문제, 2) 재난기록정보의 목록 및 위치정보의 독점, 3) 브로큰 링크로 인한 재난기록정보의 식별 문제, 4) 수집 주기 선정 문제, 5) 중복 수집 문제, 6) 메타데이터 요소 선정의

문제, 7) 접근과 이용을 위한 전략의 부재가 있었다. 상기의 문제점을 통하여 도출한 개선 방안으로는 1) 재난기록정보의 디지털 납본제도 의무화, 2) 재난안전 유관기관이 연계 및 제휴, 3) 저작권 문제에 대한 제도적 장치 마련, 4) 심층 웹 정보자원이 수집 방안 마련, 5) 실험적인 수집 과정을 통한 적절한 주기 선택, 6) 재난기록정보를 위한 표준 메타데이터 스키마 개발, 7) 이용자 세분화, 8) 재난안전 유관기관 간 협업기능 강화 서비스 개발, 9) 온라인과 모바일 서비스 전략 수립이 있다.

본 연구는 전략 제언으로만 그쳤던 기존 연구들과 달리 선행연구를 통해 제안된 전략을 실험대상을 통해 검증하고 활용가능성을 검토하였다는데 의의가 있다. 또한 이를 통하여 문제점을 분석하고 개선방향을 제언하여 보다 현실적인 적용이 가능한 아카이브 구축전략을 제안하였다는데 의의가 있다.

그러나 재난안전 기록정보자원 아카이브는 목록 및 위치정보를 수집하여 서비스한다는 점에서 “인간이 개인 또는 조직적 차원에서 활동을 수행하는 과정에서 생산하거나 입수한 데이터나 정보를 특정매체에 고정시킨” ‘기록’을 수집, 보존한다는 ‘아카이브’의 기본적 개념과는 거리가 있다는 한계가 있다. 하지만 이러한 한계는 법적·제도적 개선과 합의를 통해 충분히 극복해나갈 수 있으며, 현재 이 부분에 대한 논의가 진행되고 있는 만큼 빠른 시일 내에 극복이 가능할 것으로 보인다.

참 고 문 헌

- 김광영, 이원구, 윤화목, 신성호, 이민호. 2011. 웹 자원 아카이빙을 위한 웹 크롤러 연구 개발. 『한국콘텐츠학회논문지』, 11(9): 9-16.
- 설문원. 2012. 로컬리티 기록화를 위한 참여형 아카이브 구축에 관한 연구. 『기록학연구』, 32: 3-44.
- 설문원, 최이랑, 김슬기. 2014. 오픈소스를 활용한 사건 아카이브 구축에 관한 연구. 『한국기록관리학회지』, 14(4): 7-36.
- 여진원, 장우권. 2013. 도시아카이브 구축 방향에 관한 연구. 『한국문헌정보학회지』, 47(2): 315-335.
- 이연수, 남성운, 박성배. 2010. 공공기관 웹 사이트 기록물 수집 사례분석. 『한국멀티미디어학회 학술 발표논문집』, 644-647.
- 이연수, 남성운, 윤대현. 2013. 대규모 웹 기록물의 원격수집을 위한 콘텐츠 중복 필터링 개선 연구. 『기록학연구』, 35: 133-160.
- 조완섭, 이정은, 최치환. 2013. 웹크롤러의 수집주기 최적화. 『한국콘텐츠학회논문지』, 13(6): 30-39.
- 차승준, 최윤정, 이규철. 2009. 공공기관 심층 웹기록물 아카이빙을 위한 메타데이터 설계. 『한국전자거래학회지』, 14(4): 181-193.
- 한국기록학회 (편). 2008. 『기록학용어사전』. 서울: 역사비평사.
- 한희정, 강주연, 김용, 오효정. 2017b. 자동수집 기반 재난안전 기록정보자원 아카이브 설계 및 구축 전략. 『한국기록관리학회지』, 17(4): 127-154.
- 한희정, 박태연, 오효정, 김용. 2017a. 재난안전정보 아카이브 구축을 위한 온라인 기록정보 현황분석 및 개선방안 연구. 『한국도서관·정보학회지』, 48(2): 187-213.
- 현문수, 전보배, 이동현. 2014. 참여형 디지털 아카이브 구축 실행 방안. 『기록학연구』, 42: 245-285.
- Álvarez, Manuel, Juan Raposo, Alberto Pan, Fidel CACHEDA, Fernando Bellas, and VictorCarneiro. 2007. "DeepBot: A Focused Crawler for Accessing Hidden Web Content." In *Proceedings of DEECS2007*, San Diego CA, 18-25.
- Bergman, M. K. 2001. "White Paper: The Deep Web: Surfacing Hidden Value." *Journal of Electronic Publishing*, 7(1). <<http://www.press.umich.edu/jep/07-01/bergman.html>>.
- Desai, Keyur, Virala Devulapalli, Smita Agrawal, Preeti Kathiria, and Atul Patel. 2017. "Web Crawler: Review of Different Types of Web Crawler, Its Issues, Applications and Research Opportunities." *International Journal of Advanced Research in Computer Science*, 8(3): 1199-1202.
- Madaan, Rosy, Ashutosh Dixit, A.K. Sharma, and Komal Kumar Bhatia. 2010. "A Framework for Incremental Hidden Web Crawler." *International Journal on Computer Science and*

Engineering, 2(3): 753-758.

Pandya, Ishan, Hitanshu Joshi, Biren Patel, and Harshil Joshi. 2017. "Threats that Deep Web Possess to Modern World." *International Journal for Innovative Research in Science & Technology*, 140-148.

Raghavan, Sriram and Hector Garcia-Molina. 2001. "Crawling the Hidden Web. VLDB Conference." In: *Proceedings of VLDB2001*, Rome Italy, 129-138.

• 국문 참고자료의 영어 표기

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

Cha, Seung-Ju, Yun-Jeong Choi, and Kyu-Chul Lee. 2009. "Metadata Design for Archiving Public Deep Web Records." *The Journal of Society for e-Business Studies*, 14(4): 181-193.

Cho, Wan-Sup, Jeong-Eun Lee, and Chi-Hwan Choi. 2013. "Refresh Cycle Optimization for Web Crawlers." *Journal of the Korea Contents Association*, 13(6): 30-39.

Han, Hui Jeong, Ju-Yeon Gang, Yong Kim, and Hyo-Jung Oh. 2017b. "Design and Construction Strategy for Disaster and Safety Record Information Resources Archives Based on Automatic Acquisition." *Journal of Korean Society of Archives and Records Management*, 17(4): 127-154.

Han, Hui Jeong, Tae-Yeon Park, Hyo-Jung Oh, and Yong Kim. 2017a. "A Study on Improvement and Analysis of Records Management Status for Disaster Safety Archives in Online Environment." *Journal of Korean Library and Information Science Society*, 48(2): 187-213.

Hyun, Moonsoo, Bobae Jeon, and Dong-Hyun Lee. 2014. "Implementing a Model for Developing Participatory Labor Archives for Shipbuilding Labor Digital Archives in Young-do, Busan Metropolitan City." *The Korean Journal of Archival Studies*, 42: 245-285.

Kim, Kwang-Young, Won-Goo Lee, Hwa-Mook Yoon, Sung-Ho Shin, and Min-Ho Lee. 2011. "Development of Web Crawler for Archiving Web Resources." *Journal of the Korea Contents Association*, 11(9): 9-16.

Korean Society of Archival Studies. 2008. *Dictionary of records and archival terminology*. Seoul: Critical Review of History.

Lee, Yeon-Soo, Sung-un Nam, and Dai-hyun Yoon. 2013. "A Study on the Enhanced Filtering Method of the Deduplication for Bulk Harvest of Web Records." *The Korean Journal of Archival Studies*, 35: 133-160.

Lee, Yeon-Soo, Sung-Un Nam, and Sung-Bae Park. 2010. "A Study on the Analysis of the

- Web Archiving Case of Public Institutions.” *Journal of Korean Multi-media Conference Proceedings*, 644-647.
- Seol, Moon-Won. 2012. “Building Participatory Digital Archives for Documenting Localities.” *The Korean Journal of Archival Studies*, 32: 3-44.
- Seol, Moon-Won, Yi-Lang Choi, and Seul-Gi Kim. 2014. “Development of Digital Archives Using Open Source Software to Document Mi-ryang Transmission Towers Construction Conflicts.” *Journal of Korean Society of Archives and Records Management*, 14(4): 7-36.
- Yeo, Jin-Won and Woo-Kwon Chang. 2013. “A Study on the Urban Archives of Building Direction.” *Journal of the Korean Society for Library and Information Science*, 47(2): 315-335.