

# 자동수집 기반 재난안전 기록정보자원 아카이브 설계 및 구축전략\*

## Design and Construction Strategy for Disaster and Safety Record Information Resources Archives Based on Automatic Acquisition

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

### 목 차

- |              |                          |
|--------------|--------------------------|
| 1. 서론        | 4. 재난안전 기록정보자원 아카이브 구축전략 |
| 2. 이론적 배경    | 5. 활용방안                  |
| 3. 재난아카이브 사례 | 6. 결론                    |

### <초록>

최근 전 세계적으로 대형 복합 재난이 빈번하게 발생할 뿐 아니라 피해 반복되고 있다. 이에 따라 과거의 정리된 재난과 관련한 원시자료 및 가공정보에 대한 체계적인 관리와 활용에 대한 필요성이 증대되고 있다. 이에 본 연구는 범정부 차원에서 재난안전 기록정보자원을 종합적으로 제공하여 재난안전정보의 허브(Hub)로서 중추적 역할을 수행할 수 있는 자동수집 기반의 재난안전 기록정보자원 아카이브 구축 전략을 제안하고자 한다. 이를 위해 국내·외 재난아카이브 사례를 분석하여 재난안전 기록정보자원 아카이브 구축 시 고려할 사항들을 도출하였으며, 1) 재난안전 기록정보자원 전수조사, 2) 자동화 가능성 분석, 3) 아카이빙 대상 선별, 4) 메타데이터 자동수집에 이르는 4단계 구축전략을 제안하였다. 본 연구에서 제안하는 아카이브의 구축은 산발적으로 존재하는 정보의 통합관리와 정보의 공유 및 활용을 촉진하게 될 것이다.

주제어: 재난기록, 재난안전정보, 재난안전 기록정보자원, 재난아카이브, 자동수집

### <ABSTRACT>

Large-scale and complex disasters have frequently occurred all over the world recently, and they are repeated every year. Accordingly, the need for the systematic management and utilization of raw-data and processing information in the past has increased. For this purpose, this study proposed a construction strategy for disaster and safety record information resources archives based on automatic acquisition, which can be used as a hub for disaster and safety information. Based on local and foreign case studies, several consideration factors for building the archives are determined. Finally, this study proposed four steps for constructing the archives. These are as follows: 1) complete enumeration survey of the disaster and safety record information resources, 2) automatic acquisition possibility study, 3) selection of the resources for preservation, and 4) automatic acquisition of metadata. The construction of the archives proposed in this study will facilitate integrated management, sharing, and use of scattered information resources on disaster and safety.

Keywords: disaster records, disaster and safety information, disaster and safety record-information materials, disaster archives, automatic acquisition

\* 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 문화융복합아카이빙연구소 전임연구원(freebirdhj@naver.com) (제1저자)

\*\*\* 전북대학교 기록관리학과 박사과정(juyeongang@naver.com) (공동저자)

\*\*\*\* 전북대학교 문헌정보학과 교수(yk9118@jbnu.ac.kr) (공동저자)

\*\*\*\*\* 전북대학교 기록관리학과 조교수, 문화융복합아카이빙연구소(ohj@jbnu.ac.kr) (교신저자)

■ 접수일: 2017년 10월 29일 ■ 최초심사일: 2017년 11월 8일 ■ 게재확정일: 2017년 11월 13일

■ 한국기록관리학회지 17(4), 127-154, 2017. <<http://dx.doi.org/10.14404/JKSARM.2017.17.4.127>>

## 1. 서론

### 1.1 연구배경 및 필요성

재난의 발생은 그 자체로 현 사회 체계의 문제점을 극명하게 드러내기 때문에 모든 재난은 보다 나은 사회 구조를 설계하기 위한 계기가 된다(최형섭, 2014). 지금까지 재난이 발생하지 않았던 시대는 없었으며, 과학기술이 발전함에 따라 오히려 재난의 규모가 대형화·복합화 될 뿐 아니라 반복되고 있어 사회를 마비시키고 혼란을 초래하는 일이 끊이지 않고 있다. 이러한 상황에서 국가가 재난으로부터 시민을 보호하는 사회적, 조직적 안전장치나 의식, 시스템을 마련하지 않는다면 국민들의 사회적 불신과 자괴감, 불안감이 증폭되어 국가 발전을 저해하는 요소로 작용될 수 있다.

모든 재난을 더욱 위협스럽고 파괴적으로 만드는 원인은 대응책의 부재에 있다. '세월호' 참사 이후 안전이라는 화두를 두고 경제적 성장에 대한 집착에서 벗어나 우리 사회의 전반적인 변화를 꾀해야 한다는 인식이 점차 확산되고 있다(박성원, 2014). 즉, 이제는 모든 국민이 안전한 사회에 살고 있다는 믿음과 신뢰감을 회복시켜 주어야 하며, 이를 위해서는 재난에 대한 선제적 대응책 마련이 절실하게 필요하다. 재난을 겪고 복구하면서 재난의 피해가 반복되지 않고, 피해가 최소화 될 수 있도록 정책적, 과학기술적 차원에서 대응방안을 모색해야 한다. 특히 재난관리 분야에 있어 정부와 재난안전 유관기관 그리고 민간기관이 서로 협력하여 재난관리가 체계적이고 효과적으로 이루어질 수 있도록 해야 한다. 그리고 이러한 기관 간

협력이 제대로 이루어지기 위해서는 각 기관이 보유하고 있는 정보의 공유가 선행되어야 한다.

그러나 이들 각 기관들이 생산하여 보유하고 있는 재난안전 기록정보자원들이 공유되지 않아 제대로 활용되지 못하고 있다. 특히, 재난안전 유관기관들의 정보시스템에서 생산하고 있는 정보들은 내부적 차원에서 구성원 및 부서 간의 횡적·종적 차원의 정보공유는 물론이고, 부처 및 기관 간에도 서로 지원하는 수준에 그치고 있다. 특히, 재난안전 기록정보자원은 재난을 예방·대비·대응·복구하는 등의 재난 관련 전반적인 활동에서 생산되는 모든 정보자원들로서, 재난 관련 활동에 필요한 중요한 참고자료이자 증거 자료들이다. 그럼에도 불구하고 이러한 기록정보자원들이 어디에서 어떻게 생산되는지 제대로 공유되지 않아 활용되지 못한 채 사라지는 경우가 많다(한희정 외, 2017). 재난안전 기록정보 자원들을 제대로 수집·관리하여 적절하게 활용한다면 재난에 대한 선제적 대응과 예방이 제대로 이루어질 수 있으며, 국가 차원의 재난대응 체계의 수립이 가능해짐에 따라 미래 후손들에게 보다 안전하고 살기 좋은 사회를 물려줄 수 있는 토대를 제공할 수 있다. 따라서 이들 재난안전 유관기관이 생산하는 기록정보자원을 체계적으로 수집하고, 보존하여 관리 및 활용할 수 있는 기반인 재난안전 기록정보자원 아카이브를 구축할 필요가 있다. 특히 이러한 기록정보자원은 자동수집을 기반으로 이루어질 필요가 있다. 재난안전 유관기관이 생산하는 기록정보자원은 여기저기 흩어져 있으며, 생산되는 정보량 역시 지속적으로 증가하고 있다. 이러한 상황에서 수동으로 재난안전 기록정보자원을 수집하기에는 인적·재정적·물리적 한계가 있다.

이에 본 연구에서는 자동수집 기반의 재난안전 기록정보자원 아카이브 구축전략을 제안하고자 한다. 이렇게 구축된 아카이브는 해당 정보자원을 통합적으로 수집, 보존하고 나아가 재난안전 관련 연구 및 국가 정책 수행을 위한 중요한 정보원으로 활용될 수 있으며, 궁극적으로는 재난안전 기록정보자원의 체계적 관리에 기반한 정보의 공유를 위한 기틀이 될 수 있다.

### 1.2 연구목적 및 방법

본 연구의 최종 목적은 범정부 차원에서 재난안전 기록정보자원을 종합적으로 제공하여 재난안전 정보의 허브(hub)로서 중추적 역할을 수행할 수 있는 재난안전 기록정보자원 아카이브를 구축하는 데 있다. 이에 본 연구는 먼저 사례조사를 통해 해외 재난아카이브의 수집대상과 수집방식, 수집현황 등을 비교분석하였다. 그리고 국내 재난아카이브 현황 분석을 위해 국립중앙도서관 관계자를 대상으로 인터뷰를 진행하였다. 이를 통해 국립중앙도서관이 운영하는 OASIS 재난아카이브의 수집정책 및 현황을 파악하였다. 마지막으로 국내외 재난아카이브들의 비교분석을 통해 재난안전 기록정보자원 아카이브 구축 시 고려사항을 도출한 후 실제 재난안전 유관기관의 웹을 분석하여 적용 가능한 자동수집 기반의 재난안전 기록정보자원 아카이브 구축전략을 제안하였다.

### 1.3 선행연구

본 연구와 관련 있는 선행연구는 웹 아카이빙에 관한 연구와 재난안전 기록에 관한 연구

로 나눌 수 있다.

먼저 웹 아카이빙에 관한 연구는 김유승(2007, 2008), 이혁재, 김유승(2009), 차승준, 최윤정, 이규철(2009), 윤정옥(2010), 김광영 외(2011), 이연수, 남성운, 윤대현(2013) 등이 있다. 김유승(2007, 2008)은 웹 정보자원의 특징을 중심으로 웹 아카이빙의 범·제도적 문제를 고찰하였으며, 더 나아가 프랑스로부터의 사례를 중심으로 복합적 웹 아카이빙의 정책에 대하여 논하였다. 이를 통해 당시 웹 아카이빙의 위치를 살펴보고 나아갈 방향을 모색하였다. 이혁재, 김유승(2009)과 차승준, 최윤정, 이규철(2009)은 각각 정부기관과 공공기관의 웹 기록물에 대하여 연구하였다. 다만 이혁재, 김유승(2009)은 웹 기록물관리 정책에 초점을 맞추어 연구를 진행하였고, 차승준, 최윤정, 이규철(2009)은 공공기관의 웹 기록물 아카이빙을 위한 메타데이터 요소에 대한 연구를 진행하였다. 윤정옥(2010)은 국립중앙도서관이 중심이 되어 진행한 대표적인 웹 아카이빙 사례인 OASIS에서 '문학'과 '사회과학' 분야의 콘텐츠 및 서비스를 분석하여 개선사항 및 향후 방향을 제시하였다. 김광영 외(2011)는 웹 아카이빙에 필수적인 웹 크롤러에 대하여 연구하였다. 기존 크롤러들의 장단점을 분석하고 이를 통해 웹 아카이빙에 적합한 수집 도구 시스템을 개발하였다. 이와 유사하게 이연수, 남성운, 윤대현(2013)은 공공기관의 홈페이지에서 제공되는 웹 콘텐츠를 스냅샷 전략으로 수집하여 분석하였는데, 콘텐츠 중복 수집에 대한 문제가 발생함을 지적하고 이에 대한 해결방안을 제시하였다.

한편 국내 재난안전 기록과 관련된 연구는 그

리 많지 않은 편으로, 재난안전 기록정보자원의 현황분석, 분류, 서비스에 대한 연구(김태영 외, 2017; 박태연 외, 2017; 한희정 외, 2017), 재난안전 기록정보자원의 웹 아카이빙에 대한 연구(이수진 외, 2017; 김현정, 2017)가 있다. 이 중 본 연구와 맥락을 같이 하는 연구로는 김현정(2017), 이수진 외(2017), 한희정 외(2017)의 연구로, 먼저 김현정(2017)은 재난기록의 웹 아카이빙에 대한 연구를 진행하였는데, 특히 OASIS의 재난아카이브를 분석하여 문제점과 개선방향을 제시하였다. 윤정옥(2010)과는 OASIS의 컬렉션 중 재난아카이브만을 연구대상으로 하였다는 점에서 다르다. 한편 이수진 외(2017)는 재난안전정보가 분산되어 있으며 접근성에 문제가 있다는 점을 지적하면서 이를 해결하기 위한 방안으로 자동수집방안을 제시하였다. 이를 위해 알고리즘을 개발하고자 하였으며 국민재난안전포털을 중심으로 자동수집을 일부 수행하였다. 마지막으로 한희정 외(2017)는 온라인에서 확인이 가능한 재난안전 기록정보자원의 현황을 분석하고 실무자들과의 면담을 통해 요구사항을 수렴하여 재난안전 기록정보자원 아카이브 구축의 토대를 마련하고자 하였다.

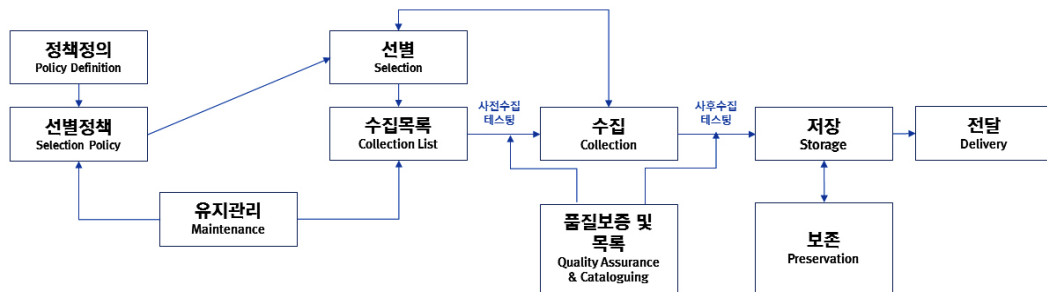
본 연구는 한희정 외(2017)의 후속연구로, 김

현정(2017)과 이수진 외(2017)의 연구와는 다음과 같은 차이가 있다. 첫째, 본 연구는 재난안전 유관기관들이 실제 생산한 재난안전 기록정보자원의 현황 분석 결과를 반영하여 아카이브 구축 전략을 제시한 반면, 김현정(2017)은 OASIS의 재난아카이브의 사례 분석을 통해 재난아카이브가 나아가야 할 방향을 제시하였다. 둘째, 본 연구는 재난안전 유관기관에서 생산한 모든 기록정보자원을 대상으로 하는 아카이브 구축 전략을 제시한 반면, 이수진 외(2017)는 국민재난안전포털만을 대상으로 하여 웹 크롤러를 개발, 구현하였다. 요컨대, 김현정(2017)과는 연구방법에서, 이수진 외(2017)와는 연구범위에서 차이가 있다.

## 2. 이론적 배경

### 2.1 웹 아카이빙

웹 아카이빙은 일반적으로 웹 자원의 선별, 수집, 저장, 전달의 과정으로 진행된다. 웹 아카이빙 절차를 도식화하면 <그림 1>과 같다(차승준, 이규철, 2008).



<그림 1> 웹 아카이빙 절차

웹 아카이빙을 위해서는 우선 선별정책이 개발되어야 한다. 정책의 정확한 특징은 개별조직 및 요구사항에 따라 매우 다양하지만 일반적으로 관련된 외부 선별정책과 적절한 상위기관의 정책 및 현존하는 내부 선별정책 내에서 결정되어야 한다. 선별정책이 결정되면 컬렉션 수집을 위한 특정 웹 소스에 대한 선별이 시작되는데 이는 컬렉션 리스트에 명시되고 이것이 선별과정의 최종 산출물이다. 컬렉션 리스트에는 모든 웹 자원의 영역 내지 범위가 분명히 명시되어야 하며, 이들 웹 자원의 적절한 수집 시간과 빈도 또한 명시되어 있어야 한다. 일단 적절한 선별 방법이 정해지면 다음으로 선별기준을 결정해야 한다. 선별기준을 결정하는 데 있어서 고려해야 할 사항은 내용·범위 및 시간/빈도 문제이다. 내용의 경우 지적 내용 측면에서 선별이 가능하도록 웹 자원의 특성을 정의하는 기준이 제시되어야 한다. 비선별 또는 주제적 접근은 최소한의 기준을 필요로 하는 반면 선별적 접근은 보다 정교한 선별기준이 필요하다. 다음으로 선별된 자원의 범위를 결정하기 위한 기준 또한 마련되어야 하는데, 특히 선별적 접근에 가장 세부적으로 적용되어야 한다. 세부적인 기술적 영역(경계)이 정의되어야 하는 경우 수집 리스트 내에 정교화 되어야 한다. 마지막으로 시간과 빈도를 결정하는 원칙이 정립되어야 한다. 어떤 웹 자원은 일회성 수집으로 끝날 수도 있지만, 웹의 역동적 특징상 대부분은 일정한 간격으로 수집된다. 웹 자원의 취약성과 일시성을 고려해 볼 때, 콘텐츠가 수집 가치가 있다고 보이는 시점에서 가능한 신속하게 수집이 되어야 한다. 또한 각 선별된 웹 자원의 수집 시점과 빈도는 수집 목록에

명시되어야 한다(한국기록관리협회, 2009).

수집대상의 범위 설정은 수집의 방법과 결합하여 더욱 다양한 수집 형태로 나타난다. 웹 아카이빙의 유형은 수집 범주를 기준으로 '선택적 아카이빙', '전체 도메인 아카이빙', '복합적 아카이빙'으로 구분된다. 선택적 아카이빙은 수집대상에 따라 다시 '주제별 웹 자원의 선택적 아카이빙', '정적 웹 자원의 선택적 아카이빙', '정적·동적 웹 자원의 선택적 아카이빙'으로 세분되며, 전체 도메인 아카이빙은 '단일 국가 전체 도메인을 대상으로 한 아카이빙'과 '전 세계 전체 도메인을 대상으로 한 아카이빙'으로 나뉜다(김유승, 2008). 웹상의 일정한 범위 안에 드는 웹사이트를 광범위하게 수집하는 경우 전량을 수집하는 것은 웹 크롤러에 전적으로 의존하며, 메타데이터 역시 자동으로 추출하는 기법을 사용한다. 실무자의 입장에서는 이러한 방식이 간단하고 공정을 요하므로 편리한 면이 있다. 반면 특정 이슈를 대상으로 선택적으로 수집하는 방식은 미리 정해진 선택기준의 범위에 드는 웹사이트들만을 대상으로 웹 아카이브를 구축하는 것으로 이 경우는 사전에 해당 웹사이트의 저작권자와 수집 및 이용조건에 대한 합의를 이루는 것이 전제가 된다. 이러한 방식은 전체 도메인을 아카이빙하는 방식보다 상대적으로 자원 집중적이며, 고비용이다. 일단 선택적으로 수집하기 위해서는 일정한 기준에 따른 평가를 수행해야 하기 때문에 상당한 시간과 노력이 필요하다(서혜란, 2004).

한편, 품질보증은 웹 아카이빙 절차에 필수적인 구성 요소로서 품질보증의 본질 및 등급은 요구사항과 수집하기 원하는 리소스에 많이 좌우된다. 품질보증은 크게 사전-수집 테

스팅(pre-collection testing)과 사후-수집 테스트(post-collection testing)이 있다. 목록화(cataloging)는 적절한 관리나 사용자들의 접근을 제공할 수 있게 하도록 아카이빙된 컬렉션에 메타데이터와 같은 설명 정보를 포함시키는 것이다. 그리고 보존의 목적은 대상의 가치가 유지되면서, 영속적인 접근을 보장하는 것이다. 즉, 사용자들의 접근을 보장하면서 본래의 가치도 그대로 보존하여 전달해야 한다. 웹 자원의 보존은 디지털 대상의 보존과 유사하여 같은 기술이 적용된다. 특히, 메타데이터는 웹 자원에 대해 이해하고 해석하는 방법을 제공하므로 장기 보존을 위해서는 설명정보의 메타데이터와 기술정보의 메타데이터의 저장이 필수 사항이다. 전달은 아카이빙된 웹 자원을 사용자가 사용할 수 있게 제공해주는 것으로 검색과 열람의 방법 등의 방법이 있다(차승준, 이규철, 2008).

## 2.2 웹 자원 수집방법

웹 자원을 수집하는 방법에는 직접 이관(direct transfer) 방식과 원격 수집(remote harvesting) 방식이 있다. 직접 이관은 가장 단순한 방식으로 원본 소스로부터 복제본을 직접 수집하는 방식으로 HTML 문서로만 구성된 정적 웹 사이트를 수집하는 데 가장 적합한 방법이다. 반면 원격 수집은 가장 널리 이용되는 방법으로 원격 웹서버로부터 콘텐츠를 수집하기 위해 웹 크롤러(crawler)를 이용하는 방식이다. 웹 크롤러는 리스트 상의 URL을 방문해서 웹 페이지를 수집하고, 페이지 내의 모든 하이퍼링크를 식별하며 그것들을 리스트 상에 추가하는

것을 반복적으로 수행한다. 크롤러는 문서를 분석, 인덱싱, 검색 및 마이닝할 수 있도록 데이터 소스에서 문서를 수집하는 프로그램이다. 크롤링하기 위한 시드(seed) URL이라고 부르는 URL 집합을 입력하면 크롤러는 지정된 정보와 관련된 웹 문서를 수집하고 다음 주소에 대한 정보를 분석 및 인덱싱(indexing) 한다. 크롤러에 모든 웹 주소를 입력할 수 없기 때문에 웹 크롤링 프로그램은 시드 URL에서 시작하여 방문한 웹페이지에 대한 파싱과정을 통해 앞으로 방문해야 할 링크 된 모든 URL을 추출한다. 그리고 추출된 URL중에서 방문하지 않은 URL에 대한 방문이 이루어지고 이와 같은 과정을 반복하여 시드 URL과 관련한 모든 웹 페이지를 방문한다.

한편, 웹 사이트는 정적 웹 사이트와 동적 웹 사이트로 구분할 수 있다. 정적 웹사이트는 보통 하나 이상의 페이지와 연계된 웹페이지들로 구성되어 있으며, 각 웹 페이지는 전형적으로 하나 이상의 개별적 요소로 구성되어 있다. 반면 동적 웹 사이트는 웹 페이지들이 보다 작은 콘텐츠 요소로부터 빠르게 생산되며, 요구사항이 접수된 후에 요구되는 요소들이 웹 페이지에 모여 전달된다. 즉, 검색을 통해서만 접근할 수 있는 동적 콘텐츠에 의해 이용가능하기 때문에 크롤러가 자동으로 수집하는 데 한계가 있다(한국기록관리협회, 2009). 최근 기술의 발전은 웹문서의 형태를 다양해지는 결과를 가져왔는데 기존 텍스트 위주의 웹문서에서 멀티미디어 형태의 웹문서가 만들어졌으며 웹은 정적인 상태에서 동적인 형태로 변화하고 있다. 즉, 웹 기술의 발전으로 웹 페이지에 이용자가 방문 하면 그 다음에 관련 정보를 데이터베이스에서

로드하여 페이지를 생성하는 방법을 많이 이용하고 있다. 이와 같은 페이지는 내용이 동적으로 생성되며 웹 크롤러가 방문한 순간에 웹 페이지에는 일부 정보만 존재한다. 이와 같은 경우 일반적인 크롤러로는 웹 페이지의 정보를 수집할 수 없다. 이처럼 페이지가 동적으로 생성되거나 또는 방문한 페이지에 검색폼을 이용해야만 페이지의 정보를 불러 올 수 있는 웹 페이지를 심층 웹(deep web)이라고 지칭한다.

카 대륙의 사람들의 건강과 삶의 표준을 증진시키기 위하여 일하고 있으며, 자신들이 관여한 재난 상황에 대하여 기록을 남기고 있다. PAHO Disasters Archives(이하 PAHO 재난아카이브)가 수집하는 기록물의 유형은 주로 사진, 동영상, 문서로 2017년 07월 기준으로 72개의 재난에 대한 기록물이 있다. 기록물은 연도별로 선택해서 볼 수 있으며, 원문 및 링크정보를 제공하고 있다. 주로 현지에서 있었던 재난 상황에 대한 개요와 사람에 대한 피해 현황을 기록한 기관 자체 생산 보고서이며 영어와 스페인어, 프랑스어 등으로 쓰여져 있다(〈그림 2〉 참조). 〈표 1〉은 PAHO가 제공하고 있는 재난아카이브 현황이며, 〈표 2〉는 PAHO 재난아카이브에 대한 내용을 정리한 표이다.

### 3. 재난아카이브 사례

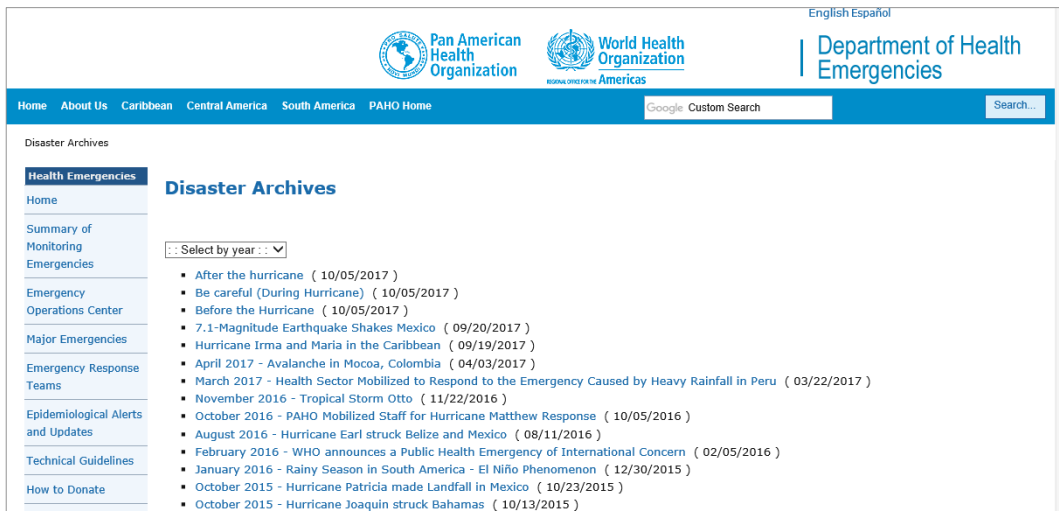
#### 3.1 국외 사례

##### 3.1.1 PAHO - Disaster Archives

The Pan American Health Organization (PAHO: originally the Pan-American Sanitary Bureau)는 국제적인 공공 건강 단체로, 아메리

##### 3.1.2 9·11 Memorial Museum

9·11 Memorial Museum(이하 9·11 아카이브)은 일종의 메뉴스크립트 보존소로서 기관



〈그림 2〉 PAHO 재난아카이브 화면

〈표 1〉 PAHO 재난아카이브 현황 (2017년 07월 기준)

| 연도                           | '03           | '04 | '05 | '06 | '07        | '08 | '09 | '10 | '11 | '12 | '13       | '14 | '15 | '16 | '17 | 계   |
|------------------------------|---------------|-----|-----|-----|------------|-----|-----|-----|-----|-----|-----------|-----|-----|-----|-----|-----|
| 기록건수                         | 2             | 58  | 23  | 18  | 43         | 28* | 14  | 34  | 7   | 15  | 2         | 0   | 3   | 39  | 10  | 296 |
| 재난유형<br>(건)                  | • 허리케인: 25    |     |     |     | • 태풍: 7건   |     |     |     |     |     | • 화재: 2건  |     |     |     |     |     |
|                              | • 지진: 13건     |     |     |     | • 홍수: 16건  |     |     |     |     |     | • 화생방: 1건 |     |     |     |     |     |
|                              | • (눈/산)사태: 1건 |     |     |     | • 화산활동: 2건 |     |     |     |     |     | • 전염병: 3건 |     |     |     |     |     |
|                              | • 호우: 5건      |     |     |     | • 한파: 1건   |     |     |     |     |     | • 기타: 3건  |     |     |     |     |     |
| ※ 복합재난의 경우 각각의 재난유형에 중복 포함시킴 |               |     |     |     |            |     |     |     |     |     |           |     |     |     |     |     |

\* 확인 가능한 최소 기록 건수임. 해당 년도에는 '터크스 케이커스 제도'로부터 수집된 재난 상황 사진 앨범이 있는데, 전체 수집 건수 확인이 어려움, 따라서 확인 가능한 최소 건수만 표시하였으며, 100장이 넘는 사진 기록이 있을 것으로 예상됨

〈표 2〉 PAHO 재난아카이브 (2017년 07월 기준)

|       |   |
|-------|---|
| 수집목적  | • 기관이 자체적으로 생산한 기록물을 보존하기 위함  |
| 수집대상  | • 자신들이 관여한 재난의 상황을 표현한 기록   |
| 수집유형  | • 동영상, 사진, 문서(보고서) 등  |
| 수집현황  | • 72개의 재난에 대한 기록물이 있음(296건 이상)  |
| 서비스내용 | • 원문 및 링크정보   |
| 특징    | <ul style="list-style-type: none"> <li>• 기록물은 재난이 발생한 연도별로 확인 가능</li> <li>• 기관에서 자체 생산 문서(보고서)들이 주를 이루며 다운로드 가능</li> <li>• 문서들은 영어와 스페인어, 프랑스어 등으로 쓰여져 있음</li> <li>• 현지에서 발생한 재난에 대한 상황 개요와 인적 피해 현황을 기록한 내용이 다수</li> <li>• 관련 정보로서 유튜브로의 링크를 제시하고 해당 재난에 대한 영상을 제공함</li> </ul> |
| 사이트주소 | <a href="http://www.paho.org/disasters/index.php?option=com_content&amp;view=article&amp;id=1933&amp;Itemid=915&amp;lang=en">http://www.paho.org/disasters/index.php?option=com_content&amp;view=article&amp;id=1933&amp;Itemid=915&amp;lang=en</a>                             |

\* 출처: PAHO 재난아카이브 홈페이지

미션에 맞는 컬렉션을 수집하고 있다. 수집대상은 2001년 9월 11일의 테러와 1993년 세계무역센터 폭탄 테러에 관한 기록으로 1) 다양한 매체 및 원본에서 파생되는 광범위한 자료로서 디지털 포맷으로 생산 또는 재생산된 자료, 2) 추모, 교육, 전시, 공공재, 연구에 있어서 가치를 부여할 할 수 있는 자료들이다. 수집유형은 음성자료, 동영상자료, 이미지(사진), 웹사이트, 문서, 희생자의 유품, 테러와 관련한 증거물, 추모기증품이나 추모기록물, 구술사, 9·11 관련 테러 정보, 테러와 관련된 예술작품, 유해 등이 있다. 2017년 07월을 기준으로 단명자료

(ephemera), 직물, 구술사, 책, 매뉴스크립트 등을 포함한 인공 유품 11,000건 이상이 수집되었으며, 동영상(moving images) 300편 이상과 40,000건 이상의 인화된 사진 및 디지털 사진자료가 수집되었다. 〈표 3〉은 9·11 아카이브에 대한 내용을 정리한 표이다.

### 3.1.3 Japan Disasters Digital Archive

Japan Disasters Digital Archive(이하 JDA)는 하버드 대학에서 진행한 프로젝트에서 시작된 디지털 아카이브로, 2011년 3월 발생한 지진과 쓰나미에 의해 직접적으로 영향을 받은 사람

〈표 3〉 9·11 아카이브 (2017년 07월 기준)

|       |  |
|-------|--|
| 수집목적  | • 1993년과 2001년 테러에 대한 증거 자료, 중요한 증언(testimonials), 역사적으로 중요한 기록들을 모으고, 보존하고, 기록화하고, 전시하고, 해석하고, 공공 교육 자료로 이용가능하게 하기 위함   |
| 수집대상  | • 다양한 매체 및 원본에서 파생되는 광범위한 자료로서 디지털 포맷으로 생산 또는 재생산된 자료<br>• 추모, 교육, 전시, 공공재, 연구에 있어서 가치를 부여할 할 수 있는 자료  |
| 수집유형  | • 음성자료, 동영상자료, 이미지(사진), 웹사이트, 문서, 희생자의 유품, 테러와 관련한 증거물, 추모기증품이나 추모기록물, 구술사, 9·11관련 정보, 테러와 관련된 예술작품, 유해 등  |
| 수집현황  | • 단명자료(ephemera), 직물, 구술사, 책, 매뉴스크립트 등을 포함한 인공 유품 11,000건 이상<br>• 동영상(moving images) 300편 이상과 40,000건 이상의 인화된 사진 및 디지털 사진자료  |
| 서비스내용 | • 원문   |
| 특징    | <ul style="list-style-type: none"> <li>• 컬렉션 관리 정책(Collection Management Policy)이라는 수집 정책 지침서가 있음</li> <li>• 매뉴스크립트 아카이브로서 기관 내 컬렉션 담당자가 수집 기록을 선별·평가하여 수집하고, 기록관리 전반의 프로세스를 따름</li> <li>• 오픈소스 컬렉션 관리 프로그램인 Collective Access를 통해 컬렉션의 전자적 관리를 지향하고 있음</li> <li>• 웹 아카이빙을 위하여 기념관 큐레이터가 사이트 목록을 선별하고 스냅샷을 위한 크롤러를 사용함</li> <li>• 웹 아카이빙 커뮤니티에서 국제적으로 인증된 표준 포맷(*.arc file)으로 압축</li> <li>• 수집된 파일들은 전용 서버에서 유지되며, 기념관은 수집한 모든 파일을 연 2회 전용 서버로부터 불러와 실행</li> <li>• 저작권은 웹사이트 제작에게 있으며 아카이빙 사실 공지를 통해 의의제기에 대하여 대처 하고 있음</li> </ul> |
| 사이트주소 | <a href="https://www.911memorial.org">https://www.911memorial.org</a>  |

\* 출처: 9·11 아카이브 홈페이지, Collection Management Policy(2016), 정혜지(2016)

들의 기억을 나누는 공간이다. 이를 위해 2011년 3월 일본에 대재앙을 불러온 지진과 쓰나미에 관련한 트윗, 증언(testimonials), 영어뉴스 기사 전문(full-text), 뉴스 헤드라인 등의 디지털 콘텐츠를 수집하고 있다. 주로 웹사이트, 사진, 비디오, 오디오 등과 같은 기록물들의 링크를 수집하여 서비스하고 있으며, 극히 일부에 대해서만 원문을 수집한다. JDA에 수집된 모든 기록물은 협력업체와 이용자가 제공하는 것이며, 그러한 정보들을 한 곳에 모아 정보검색 공간으로 활용하는 것이 JDA의 목표이다. 〈표 4〉는 JDA에 대한 내용을 정리한 것이다.

### 3.2 국내: 국립중앙도서관 - OASIS 재난아카이브

국립중앙도서관에서 구축한 OASIS 재난아카이브의 수집정책과 현황을 분석하기 위해 담당자와 면담 및 홈페이지 분석을 실시하였으며, 그 내용을 종합하면 다음과 같다.<sup>1)</sup>

국립중앙도서관은 2014년 처음 재난 관련 자료 수집을 시작하여, 2015년 OASIS 재난아카이브를 구축하였다. OASIS 재난아카이브의 대상이 되는 ‘주제(사건)’의 선정 기준은 특별 재난선포, 피해규모, 사회적 영향 등을 고려하여 구축 대상을 선정하며, 내·외부 전문가(16명)로 구성된 ‘웹사이트 아카이빙 전문가 그룹’은

1) 면담은 국립중앙도서관에서 2017년 2월에 진행되었으며, 면담대상자는 담당 사무관 1명, 사서 1명을 대상으로 진행하였다.

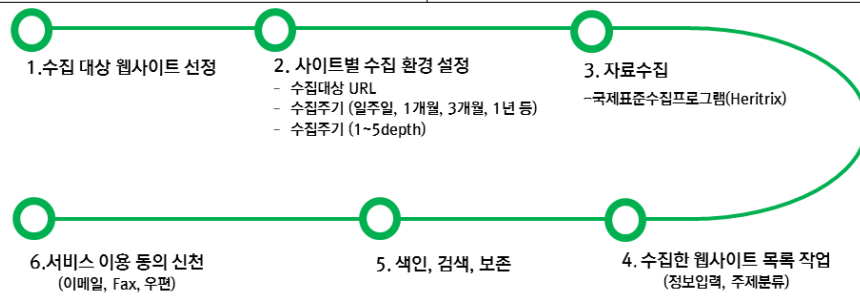
<표 4> JDA (2017년 07월 기준)

|       |  |
|-------|--|
| 수집목적  | <ul style="list-style-type: none"> <li>• 다른 재난뿐만 아니라 2011년 3월을 그 사건과 그 사후에 대한 디지털 자료를 접근 가능하게 하고 보존하고 찾을 수 있게 하기 위함</li> <li>• 시민과 연구자, 학생, 정책결정자에게 정보의 공유와 협업, 대화를 위한 공적인 장소를 제공하기 위함</li> <li>• 그 사건의 결과에 대해 가장 걱정하고 영향을 가장 많이 받은 사람들에게 대한 기억을 공유하기 위한 사이트로서의 임무를 다하기 위함</li> </ul>  |
| 수집대상  | • 2011년 3월 일본을 강타하여 핵 재앙을 유발한 지진과 쓰나미에 관련한 디지털 콘텐츠   |
| 수집유형  | • 웹사이트, 트윗, 비디오, 오디오, 뉴스 기사와 같은 디지털 콘텐츠  |
| 수집현황  | • 350개의 컬렉션  |
| 서비스내용 | • 일부 원문 및 링크   |
| 특징    | <ul style="list-style-type: none"> <li>• 사실상 모든 기록들은 협력업체와 이용자가 제공하는 것이며, 그러한 정보들을 한곳의 정보 검색 공간으로 가져오는 일을 하고 있음</li> <li>• 디지털 콘텐츠만 수집대상임</li> <li>• 트윗, 증언, 영어뉴스기사 전문, 뉴스 헤드라인 등을 포함한 일부는 직접 수집 및 보존</li> <li>• 수집된 데이터 셋을 가지고 프로젝트를 진행하고 있음                     <ul style="list-style-type: none"> <li>- Mapping the JD Archive(Fukushima, Twitter and the Politics of Disaster Communications): 당시 발생한 트위터를 가지고 다양한 분석을 시도하였음</li> <li>- 4-D Visualization of the Disaster-related data: 지도 위에 재난 당시 시공간에 대한 정보 분포도를 표현하고자 하는 시도</li> </ul> </li> </ul> |
| 사이트주소 | http://jdarchive.org/en  |

\* 출처: JDA 홈페이지

<표 5> 국립중앙도서관 OASIS 재난아카이브 구축 기간 (기준 - 주제 1건)

| 구분  | 기간                              |
|---|---------------------------------|
| 컬렉션 대상 선정 및 수집대상 조사                               | 2주일(컬렉션 규모 및 범위에 따라 1주 ~ 3주 소요) |
| 시스템 등록 및 Heritrix 수집                              | 1주일                             |
| 정상 수집여부 확인, 메타데이터 입력, 메인 화면 썸네일 생성, 점검, 품질검사, 서비스 | 2 ~ 4주일(구축 양에 따라 변동)            |



영'을 통해 자문을 받고 있다. 한 주제 당 아카이브 구축 소요 시간은 수집 규모에 따라 약 한

달에서 두 달 가량 소요되고 있으며, 구체적인 내용은 <표 5>와 같다.

OASIS 재난아카이브 홈페이지에는 수집대상 자료에 대하여 “재난 관련 기관(정부, 공공기관, 지방자치단체, 민간단체)의 웹사이트, 발행문서(보도자료, 발표문, 정책자료, 백서, 연보, 조사·분석·연구보고서 등) 및 동영상, 사진, 개인이 작성·공개하는 온라인 자료”로 명시되어 있다. 여기에서 개인이 작성·공개하는 온라인 자료의 경우 자료의 희귀성에 따라 선별하여 수집하는데, 예를 들면 ‘세월호’에 관한 유일한 동영상 정보를 포함하는 포스팅의 경우는 보존할 만한 가치가 있다고 여겨 수집·보존하고 있다. 개인 블로그는 수집대상이 맞지만 공개 및 서비스는 아직 시작하지 않고 있다. 이는 자료의 희귀성이나 유일성 측면에서 수집·보존의 대상이 되긴 하지만 공개할 경우 저작권 문제 및 사적인 정보에 대한 내용 공개 등의 문제가 발생할 소지가 많아서 개인 블로그 자료는 공개 대상 정보로 하고 있지는 않다. 현재 국립중앙도서관 OASIS 재난아카이브가 수집한 현황은 <표 6>과 같다.

OASIS 재난아카이브 구축을 위한 기록물 수집은 대부분 수작업으로 진행되고 있다. 크롤링 작업의 경우 Heritrix를 사용하여 내용 추출

이 아닌 형상자체를 가져와서 재현하는 방식이다. 재난은 시작과 끝이 명확하지 않기 때문에 정기적인 크롤 작업을 실시하고 있다. 뉴스 같은 경우는 저작권 때문에 관내에서만 볼 수 있으며, 관내에서는 대다수의 자료에 대해 접근이 가능하다. 전체 공개는 저작권 동의를 받은 경우만 가능하다.

국립중앙도서관 재난아카이브의 특징을 살펴보면 재난관련 정보자원에 대한 모든 수집 활동이 전문가에 의해 수동으로 이루어진다. 따라서 정보의 양은 적으나 질적인 측면에서 우수한 편이다. 다만 주제 선정에 대한 기준이 명확하지 않으며, 원문 저장에 따른 스토리지 비용이 지속적으로 상승할 것으로 예상된다. <표 7>은 국립중앙도서관 OASIS의 재난아카이브에 대한 내용을 정리한 표이다.

### 3.3 재난아카이브 비교분석

본 절에서는 상기에서 살펴본 국내·외 재난아카이브 사례들을 비교분석을 통해 재난안전 기록정보자원 아카이브 구축을 위한 고려사항을 도출하고자 한다.

<표 6> 국립중앙도서관 OASIS 재난아카이브 수집현황 - 기록 유형별 (2017년 10월 기준)

| 재난유형(건) |         | 웹사이트  | 이미지 | 동영상   | 웹문서   | 총계     |
|---------|---------|-------|-----|-------|-------|--------|
| 사회재난    | 화재(7)   | 922   | 136 | 75    | 135   | 1,268  |
|         | 해양사고(5) | 7,633 | 163 | 684   | 895   | 9,375  |
|         | 붕괴(3)   | 223   | 4   | 80    | 307   | 614    |
|         | 환경오염(1) | 55    | 139 | 119   | 101   | 414    |
|         | 보건(1)   | 728   | 25  | 7     | 697   | 1,457  |
|         | 교통(1)   | 0     | 1   | 18    | 0     | 19     |
| 자연재난    | 풍수해(4)  | 153   | 9   | 85    | 13    | 260    |
|         | 지변(1)   | 138   | 0   | 26    | 173   | 337    |
| 총계(23)  |         | 9,852 | 477 | 1,094 | 2,321 | 13,744 |

〈표 7〉 국립중앙도서관 OASIS 재난아카이브

|       |   |
|-------|---|
| 수집목표  | • 해당 시기에 집중 생산되고 쉽게 소멸되는 주요 재난에 대한 온라인 자료로 현재뿐 아니라 미래를 위해 컬렉션 단위로 수집하여 보존   |
| 수집대상  | • 국가기관, 공공기관, 교육 및 연구기관, 주요 상업기관, 민간단체 등에서 제공하는 국가적 재난, 이슈, 주제 분야에 대한 웹 자원  |
| 수집유형  | • 관련 기관 및 단체의 웹 사이트, 뉴스기사(댓글포함), 개인블로그, 웹 자료(문서, 이미지, 동영상)  |
| 수집현황* | • 웹 사이트: 225,470(2016년 기준)<br>• 웹 자료: 897,877(2016년 기준)<br>* 재난안전 기록정보자원을 포함한 OASIS 전체컬렉션 통계임                       |
| 서비스   | • 원문 및 링크정보   |
| 특징    | • 모든 수집 활동이 전문가에 의해 수동으로 이루어짐<br>• 정보의 양은 적으나 질적인 측면에서 우수함<br>• 주제 선정에 대한 기준이 명확하지 않음<br>• 원문 저장에 따른 스토리지 비용 지속적 상승 |
| 수집절차  |   |
| 사이트주소 | http://www.oasis.go.kr  |

\* 출처: OASIS 재난아카이브 홈페이지

앞서 살펴본 재난아카이브 사례들은 모두 재난을 주제로 하고 있으며 범위에 따라 성격이 상이하다. 여기에서 '범위'란 내용적인 측면에서 어떤 재난을 다루고 있는지를 말하는데, 상기 사례의 경우에는 한 가지의 재난 사례(case)에 대한 기록물을 수집하는 경우와 모든 재난 유형(type)에 대한 기록물을 수집하는 경우로 나눌 수 있다.

먼저 단일 재난 사례에 대한 기록물을 수집하고 있는 9·11 아카이브와 JDA는 기록이 양

적으로 많고 유형 역시 다양하다. 이는 상기 두 아카이브가 특정 재난 사례에 대한 기억의 보존을 목표로 다양한 생산자(출처)를 갖는 기록을 수집, 보존하고 있기 때문이다. 또한 관심있는 이용자의 자발적인 기증과 기탁이 주된 수집 방법이라는 점이 기록의 양과 유형의 다양성에 큰 역할을 하고 있다. 예컨대, 9·11 아카이브는 현장에 남아있던 최초 구조 차량과 강철 구조물 같은 대형 박물부터 개인 소지품과 수집품, 웹 기록까지 다양한 생산자로부터 생산되고

다양한 매체를 통해 표현되는 기록을 다수 수집, 보존하고 있다(정혜지, 2016).

한편, 다양한 재난 유형에 대한 기록물을 수집하는 경우로는 OASIS 재난아카이브와 PAHO 재난아카이브가 있다. 이 두 아카이브는 내용적인 측면에서 다양한 유형의 재난을 포함하고 있지만 그 범위가 너무 넓기 때문에 각 재난 유형별로 수집된 기록의 양이 적고 기록의 유형 역시 문서로 제한적이다. 예컨대, PAHO 재난아카이브의 경우 하나의 보고서(문서 유형)에 재난을 집약적으로 정리, 기록화하고 있으며, 기관 자체적으로 해당 기록을 생산하여 수집, 보존하기 때문에 더욱 이러한 양상을 보인다. 또한 모든 재난 유형에 대한 기록물을 수집하는 아카이브의 경우에는 민간기록보다는 수집이 용이한 공공기관이나 언론기관에서 생산한 기록들을 수집, 보존하는 경향이 있다. 이러한

특징은 국립중앙도서관 주도로 구축된 OASIS 재난아카이브에서 확인할 수 있었다.

이상과 같이 재난아카이브는 아카이브가 수집하고자 하는 기록물의 내용적인 측면의 범위에 따라 그 특성이 달라진다. 하지만 상기 모든 사례들은 주제 아카이브로서 자료의 회귀성이나 유일성의 측면을 중요시여기고 있기 때문에 일반적인 재난안전에 관련한 기록정보자원은 수집대상에서 제외될 가능성이 있다. 또한 자동수집을 지향하고 있으나 1회성 수집으로 끝나는 경우가 많다. 물론 OASIS 재난아카이브의 경우 일부 기록물에 대하여 주기적으로 크롤러를 작동시키고 있지만 주제 선정에 상당한 시간이 걸리고 크롤러를 보낼 곳을 전문가가 직접 선정해야 한다는 점에서 자동수집 기술을 완전히 적용했다고 보기에는 어려움이 있다. 이상을 정리하면 <표 8>과 같다.

<표 8> 국내·외 재난아카이브 비교분석

| 구분   | 단일 재난 사례 아카이브  |                                     | 다양한 재난 유형 대상 아카이브                      |                                 |
|------|--|-------------------------------------|--|---------------------------------|
|      | 9·11 아카이브(국외)  | JDA(국외)                             | PAHO 재난아카이브(국외)                        | OASIS 재난아카이브(국내)                |
| 주제   | • 9·11 테러  | • 2011년 3월 대지진                      | • PAHO가 구호활동을 벌이는 모든 재난                | • 전문가에 의해 선별된 유형의 재난            |
| 수집대상 | • 9·11 테러와 관련된 모든 기록   | • 2011년 3월 지진과 관련된 디지털 콘텐츠          | • 기관 자체 생산 기록                          | • 재난에 대한 웹 자원                   |
| 수집방식 | • 수동(일부 자동)  | • 수동(일부 자동)                         | • 수동                                   | • 수동(일부 자동)                     |
| 서비스  | • 원문   | • 일부 원문 및 링크정보                      | • 원문 및 링크정보                            | • 원문 및 링크정보                     |
| 기록출처 | • 주로 민간  | • 주로 민간                             | • 기관 자체                                | • 주로 공공기관 및 언론                  |
| 특징   | • 다양한 생산자의 기록을 수집하여 두터운 기억을 보존   | • 디지털 콘텐츠를 수집하기 때문에 기록의 활용 및 분석이 용이 | • 기록물(보고서) 한 건에서 재난 전체에 대한 상황 정보 확인 가능 | • 전문가에 의해 선별·평가가 이루어지므로 품질이 보장됨 |
| 시사점  | <ul style="list-style-type: none"> <li>• 이벤트(단일 재난 사례)에 대한 기억을 보존하는데 최종 목표가 있기 때문에 다양한 출처의 기록물을 수집</li> <li>• 기록의 양이 많고 유형 역시 다양함</li> <li>• 관심 있는 이용자의 기증 및 기탁이 주된 수집 방법</li> </ul> |                                     |  |                                 |
|      | <ul style="list-style-type: none"> <li>• 자료의 회귀성이나 유일성 측면이 강조되며, 일반적인 재난안전에 관한 정보는 수집대상에서 제외될 가능성이 있음</li> <li>• 크롤러를 통한 자동수집을 지향하지만 자동수집 기술을 완전히 적용했다고 보기는 어려움</li> </ul>           |                                     |  |                                 |

## 4. 재난안전 기록정보자원 아카이브 구축전략

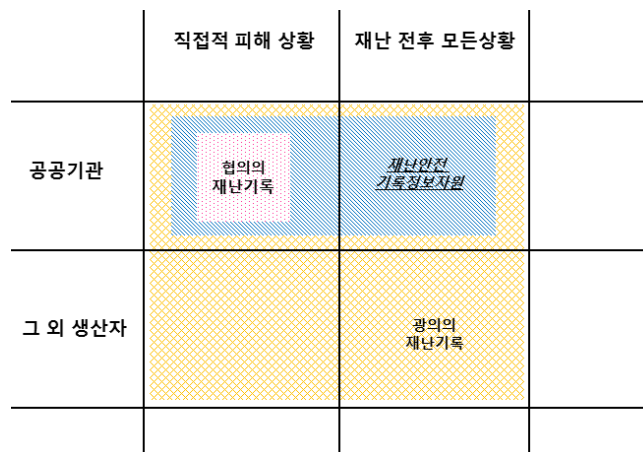
### 4.1 설계 시 고려사항

앞서 살펴본 재난아카이브 현황분석 통해 도출된 재난안전 기록정보자원 아카이브 설계 시 고려사항은 다음과 같다.

첫째, 아카이빙 대상이 되는 기록물의 정의 및 범위를 명확히 해야 한다. 본 연구에 앞서 한희정 외(2017)가 정의하였듯이, 재난기록은 ‘공공기관이 “단순히 재난피해와 인명피해에 초점을 맞춰 재난의 행정관리의 목적상” 생산하는 기록’이라는 협의의 기록과 ‘생산자와 관계없이 “재난으로 인한 그 지역사회에 충격과 혼란 상태까지”를 기술한 기록’이라는 광의의 기록으로 나눌 수 있다. 그리고 협의와 광의의 재난기록 사이에는 「재난 및 안전관리 기본법」을 기준으로 하여 “국가 및 공공기관이 재난관리 및 안전관리와 관련된 활동을 수행하면서 생산한

기록자료 및 정보자원”으로 정의될 수 있는 재난안전 기록정보자원이 있다(〈그림 3〉 참조). 본 연구에서는 이 중 재난안전 기록정보자원을 아카이빙 대상으로 삼고자 한다. 그 이유는 재난안전 기록정보자원 아카이브 구축의 목적이 재난과 관련된 실질적 기록정보 및 현장정보, 학술정보, 실무정보 및 실생활에 유용한 재난관련 전반적인 정보를 통합적으로 제공하는 데 있으며, 이는 다시말해 본 연구에서 제안하는 아카이브 구축 목적과 재난안전 기록정보자원의 정의 및 범위가 일치함을 의미한다. 따라서 재난안전 기록정보자원 아카이브 구축을 위해서는 공공기관, 특히 재난안전 유관기관이 생산하면서 재난을 예방·대비·대응·복구하는 등의 재난 관련 전반적인 활동에서 생산되는 모든 정보자원들을 수집, 보존해야 한다.

둘째, 지속적인 수집 및 관리·활용을 위해 자동수집 전략을 구축해야 하며, 자동수집 전략에는 필수로 수집해야 할 메타데이터가 포함되어야 한다. 기존의 주제 아카이브는 기관의 미



〈그림 3〉 재난안전 기록정보자원의 범위  
\*한희정 외, 2017, p. 194 도식화

선에 부합하는 기록을 수집, 보존하기 위하여 별도의 수집 전략을 기반으로 개별적인 기록을 직접 수집했다. 그러나 재난안전 기록정보자원 아카이브의 주제가 되는 재난은 갈수록 발생의 빈도가 잦아지고 발생 양상이 복합적으로 진화한다는 점에서(이관형, 이원호, 양원직, 2016) 장기적인 기록화가 필요한 상황이므로 단발적 수집이 아닌 지속적인 수집이 필요하다. 지속적인 수집은 수집하고자하는 기록의 양에 따라 그 방법이 수동과 자동으로 달라질 수 있는데, 본 연구에서는 수집할 기록의 범위를 볼 때 방대한 양의 수집이 예상되므로 자동수집이 반드시 필요하다. 따라서 자동수집을 위한 전략이 구축되어야 하며, 전략에는 기록을 식별하기 위한 최소한의 메타데이터와 자동수집이 가능한 정도를 파악하여 사전에 설정하여야 한다.

셋째, 아카이빙 대상의 선별·평가를 위한 전략이 수립되어야 한다. 아카이빙 대상이 되는 기록물의 정의 및 범위가 정해지면, 대상 중에서도 우선적으로 수집해야 할 대상을 선별하기 위한 선별·평가 전략이 수립되어야 한다. 본 연구에서는 자동수집 기반의 재난안전 기록정보자원 아카이브를 구축하고자 하므로, 재난안전 기록정보자원을 생산하는 기관에서 수집 가능한 기록정보자원 목록을 확인한 후 이에 대한 자동수집 가능성을 파악해야 한다. 자동수집 가능성 파악 후에는 다시 아카이빙을 우선적으로 진행할 목록을 선정한다. 이때 아카이브 구축 목적에 부합하는 정보를 우선적으로 아카이빙하기로 한다. 즉, 재난과 직접적 관련이 있는 기록정보 및 현장정보, 학술정보, 실무자에게 필요한 정보, 실생활에 유용한 재난관련 기록정보자원을 우선적으로 수집하도록 한다.

넷째, 이용자에게 공유의 편의성 및 공개성을 보장해야 한다. 재난안전 기록정보자원은 재난 관련 전반적인 활동에서 생산되는 모든 기록정보자원들로서, 재난 관련 활동에 필요한 참고자료이자 증거 자료들이다. 그러나 이러한 기록정보자원들의 생산기관이나 보존장소 같은 위치 정보의 공유는 미흡한 실정이며, 이로 인하여 활용이 제대로 이루어지지 않고 있다. 이러한 문제는 한희정 외(2017)의 연구에서 실시한 실무자 인터뷰를 통해서 확인된 바 있다. 따라서 재난안전 기록정보자원 아카이브 구축을 위해서는 수집하는 기록정보자원의 공유의 편의성과 공개성을 보장해야 한다. 이에 본 연구에서는 재난안전 기록정보자원의 목록 및 위치 정보 등을 제공함으로써 이용자들이 재난안전 기록정보자원에 보다 손쉽게 접근할 수 있도록 하고자 한다. 목록 및 위치정보의 제공은 원문 저장에 따른 저작권 문제를 해소함으로써 재난안전 기록정보자원의 공유의 편의성을 높이고 공개성을 보장한다.

## 4.2 구축전략

본 절에서는 자동수집 기반의 재난안전 기록정보자원 아카이브 구축전략에 대하여 기술하였다. 구축전략은 4단계로 구성되며(〈그림 4〉 참조) 상기의 고려사항을 반영하였다.

### 4.2.1 1단계 - 재난안전 기록정보자원 전수 조사

재난안전 기록정보자원 아카이브 구축을 위해서는 먼저 재난안전 유관기관이 온라인에서 제공하고 있는 재난안전 기록정보자원에 대한



〈그림 4〉 재난안전 기록정보자원 아카이브 구축 프로세스

전수조사를 실시해야 한다. 이를 통해 기관별로 수집해야 할 재난안전 기록정보자원을 파악하고 이에 대한 세부정보를 작성해야 한다. 재난안전 기록정보자원 전수조사는 실제 재난안전 기록정보자원의 위치를 파악하고, 자동수집을 위한 크롤링 알고리즘을 설계하기 위한 기초조사이다. 즉, 표면웹과 심층웹을 구분하고 이에 맞는 크롤링 알고리즘을 설계하기 위해 아카이브 대상이 되는 재난안전 유관기관의 웹 페이지를 전수 조사 해야 한다. 본 연구에서는 재난안전 기록정보자원 전수조사는 자동수집을 위한 사전조사로서 다음과 같이 진행할 것을 제안한다.

첫째, 필드명에 대한 정의를 내려야 한다. 전수조사 대상에 대한 필드명을 정의해야 하는 이유는 크롤링을 통한 자동수집의 대상을 결정하기 위해서이다. 재난안전 유관기관이 운영하는 홈페이지에는 자체 서버 시스템과 외부시스템 등이 혼재되어 있으므로 이를 사전에 구분

하여 중복 수집되지 않도록 해야 한다. 재난안전 기록정보자원 전수조사 대상에 대한 필드명 정의에 대한 예시는 〈표 9〉에 나와 있으며, 수집대상의 여부는 수집정책에 따라 달라질 수 있다.

둘째, 웹페이지 세부주소를 작성해야 한다. 세부주소의 확인을 통해 재난안전 기록정보자원 자동수집 가능성을 분석할 수 있으며, 나아가 크롤링 알고리즘에 대한 설계 방향을 세울 수 있다. 〈표 10〉은 세부주소 작성 방법의 예시이다.

셋째, 링크가 표현하는 정보에 대한 사이트 경로를 작성해야 한다. 이 경우 사이트맵을 참고하여 작성하되, 사이트맵에서 표현되지 않는 최하위 경로는 작성자가 부가적으로 작성한다. 그 외 서브시스템에서 외부링크로 이동하는 경우 첫 단계를 홈(home)이 아닌 서브시스템을 기입한다. 예를 들어, 원자력안전정보공개센터에서 유사정보기

〈표 9〉 재난안전 기록정보자원 전수조사 대상에 대한 필드명 정의 예시

| 입력 값          | 정의  |
|---------------|---|
| 메인            | <ul style="list-style-type: none"> <li>• 기관의 메인 홈페이지</li> </ul>   |
| 서브시스템명        | <ul style="list-style-type: none"> <li>• 기관에서 운영하는 서브시스템</li> <li>• 외부 링크의 유형으로 갑자기 등장할 수도 있음</li> <li>• 여기서 <u>서브시스템이란</u>,                         <ol style="list-style-type: none"> <li>1. 기관에서 운영하는 시스템으로 공지된 경우</li> <li>2. (기관 관계자 인터뷰가 진행되었을 때) 사전 인터뷰자료에서 확인되는 시스템</li> <li>3. 서브시스템 페이지에서 기관 로고가 발견되는 경우로 한정한다.                                 <ul style="list-style-type: none"> <li>ex. 화학물질안전원의 경우 “화학물질안전관리정보시스템”이 상기 3번 경우에 해당하여 서브시스템에 해당하므로 “화학물질안전관리정보시스템”을 입력값으로 기입한다.</li> </ul> </li> </ol> </li> </ul> |
| 타기관 (시스템명)    | <ul style="list-style-type: none"> <li>• 외부 링크 유형 중 하나</li> <li>• 페이지를 제공하고 있는 기관의 성격이 현재 조사 대상인 기관의 성격과 완전히 상이하지만 해당 정보가 기본 원칙 2번에 해당한다면 수집대상으로 한다.                         <ul style="list-style-type: none"> <li>ex. 한국가스안전공사의 경우 “home &gt; 공사소개 &gt; 경영공시 &gt; 정보센터 &gt; 연구보고서”를 클릭했을 때 공공 기관 경영정보 공개시스템인 “ALIO”로 이동하게 되는데, 이 기관은 한국가스안전공사와 성격이 완전히 상이하다. 그러나 기본 원칙 2번에 해당한다고 볼 수 있으므로 클릭으로 생성되는 첫 페이지를 “링크” 필드에 입력하고 “시스템(메인, 서브)”에는 “타기관(ALIO)”을 입력 값으로 기입한다.</li> </ul> </li> </ul>                                   |
| 유사정보기관 (시스템명) | <ul style="list-style-type: none"> <li>• 외부 링크 유형 중 하나</li> <li>• 타 기관에서 운영하는 페이지이지만 조사 기관과 유사한 재난 정보를 제공하고 있다면 수집대상으로 한다.                         <ul style="list-style-type: none"> <li>ex. 한국원자력안전위원회의 경우 “HOME &gt; 알림마당 &gt; 일본원전 방사능 정보방 &gt; 국가환경방사선자동감시망”을 클릭했을 때 “국가환경방사선자동감시망”으로 이동하게 되는데, 이 페이지는 한국원자력안전위원회에서 제공하는 정보와 유사한 정보를 제공하고 있으므로 유사정보기관으로 판단할 수 있다. 따라서 “유사정보기관(국가환경방사선자동감시망)”을 입력한다.</li> </ul> </li> </ul>   |
| 확인불가 (시스템명)   | <ul style="list-style-type: none"> <li>• 외부 링크 유형 중 하나</li> <li>• 상기의 유형 중에 포함되지 않는 유형</li> <li>• 작업자가 판단하기 불가능한 유형</li> </ul>  |

〈표 10〉 재난안전 기록정보자원 웹 기록물 세부주소 작성 방법 예시

|          |   |
|----------|---|
| 작성 방법(안) | <ul style="list-style-type: none"> <li>• 정보를 확인할 수 있는 가장 최하위 단위의 탭, 페이지까지의 링크를 기입한다.</li> <li>• 하위 카테고리에서 다수의 탭이 확인되는 경우, 모든 탭의 링크를 기입한다. 단, 다수의 탭 중 하나의 탭에서 수집대상이 되는 모든 정보를 포함하고 있다면 그 탭의 링크만을, 그렇지 않다면 해당하는 탭의 링크를 기입한다.</li> <li>• 검색창으로 구성된 페이지의 경우, “검색” 버튼이나 검색어를 입력하지 않은 첫 페이지 링크를 기입한다.</li> <li>• 다수의 카테고리에서 중복된 정보가 확인되는 경우 해당되는 모든 페이지 및 탭에 대한 정보를 기입한다. “비고” 필드에는 중복된 정보 페이지 링크가 있는 행 번호를 기입하거나, 중복 내용에 대하여 글로 작성할 수 있다.                         <ul style="list-style-type: none"> <li>ex. 국립재난안전연구원의 경우 “HOME &gt; 부서안내 &gt; 연구실적”과 “HOME &gt; 연구실적 &gt; 연구보고서”의 내용이 90% 이상 일치하는 것으로 보이므로 “비고” 필드에 이러한 내용에 대하여 기술하도록 한다.</li> </ul> </li> <li>• 소프트웨어의 설치가 필요한 경우에는 기입할 수 있는 정보를 모두 기입하고, “비고” 필드에 소프트웨어 설치 여부에 대하여 기입한다.</li> <li>• 로그인을 해야만 정보를 확인할 수 있는 경우는 수집대상에서 제외한다.</li> <li>• 게시판에 게시된 글이 0건일지라도 게시판의 성격 상 재난안전 기록정보자원에 포함되는 게시물이 미래에 생산될 수 있다면 수집대상에 포함한다.</li> <li>• “정부 3.0” 정책의 일환으로 게시된 게시물에 대하여 다음을 따르도록 한다.                         <ol style="list-style-type: none"> <li>① “공공데이터포털”과 같은 외부 링크로 연결되는 경우: 외부 링크로 간주하여 작성한다.</li> <li>② 기관에 속한 게시판이나 페이지 형태로 확인되는 경우: 일반 게시물로 보고 작성한다.</li> </ol> <ul style="list-style-type: none"> <li>※ 이외에 발견되는 “비고” 필드에 상세히 기입하도록 한다.</li> </ul> </li> </ul> |
|----------|---|

관 “방사성폐기물안전관리통합정보시스템”으로 이동하게 된다. 이 때 사이트 경로는 “홈 > 핵연료/폐기물안전 > 방사성폐기물 > 중저준위폐기물 > 방사성폐기물안전관리통합정보시스템”이 아니라 “원자력안전정보공개센터 > 핵연료/폐기물안전 > 방사성폐기물 > 중저준위폐기물 > 방사성폐기물안전관리통합정보시스템”으로 기입한다.

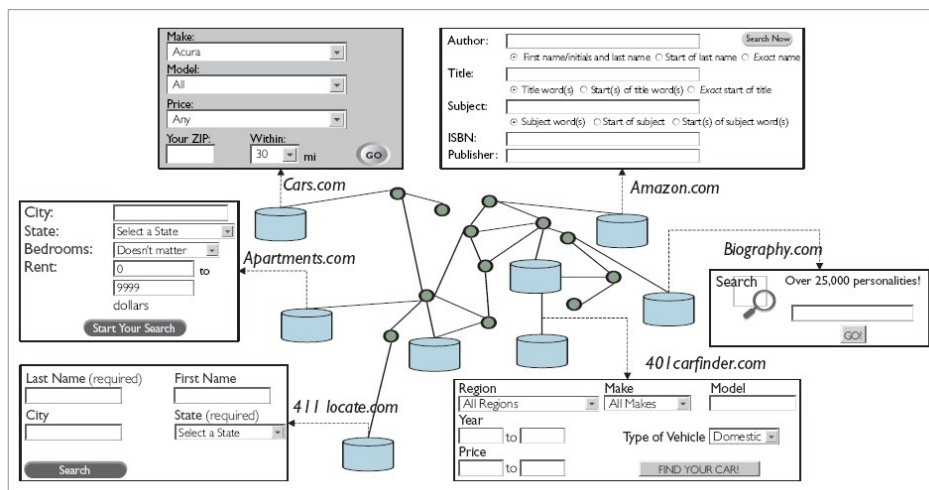
#### 4.2.2 2단계 - 자동화 가능성 분석

자동화 가능성 분석은 1단계에서 전수 조사한 결과를 바탕으로 자동화여부를 판단하고, 기술적인 부분에 대하여 세부 정보를 보완하는 단계이다. 본 단계에서는 전수조사에서 누락된 페이지와 정보를 재확인한 후 보완하도록 한다. 또한 자동화가능성 분석 단계의 핵심은 표면웹과 심층웹을 구분하고, 크롤링이 가능 여부에 대해서도 분석하는 일이다. 표면웹과 심층웹을 구분하여 자동수집 알고리즘을 설계할 수 있어야 한다.

본 연구에서 제안하는 웹 크롤러는 웹페이지를 읽고 색인을 생성하는 프로그램으로 대부분의 검색 엔진은 이러한 기능을 하는 크롤러를 통해 웹 기록물을 수집하고 서비스를 제공한다. 하지만 이러한 조건을 만족하기 위해서는 페이지가 정적, 즉 웹브라우저가 페이지를 방문함과 동시에 모든 정보를 가지고 있어야 한다. 하지만 <그림 5>와 같이 특정 검색결과를 통해 동적으로 페이지가 만들어 지는 경우에는 페이지가 완성되기 전 까지 웹페이지는 존재하지 않는다. 이러한 경우 크롤러는 웹 기록물을 수집하지 못한다. 이처럼 웹사이트에 특정 질의를 통해 동적으로 문서를 생성하고 가져올 수 있는 형태의 심층웹 문서와 일반적인 표층웹 문서를 구별하여 개발하여야 한다.

#### 4.2.3 3단계 - 아카이빙 대상 선별

전수 조사한 결과를 바탕으로 아카이빙 대상이 되는 정보를 선별한다. 즉, 본 단계에서는 재



<그림 5> 심층 웹 크롤링

\* 출처: Bin He, Mitesh Patel, Zhen Zhang, Kevin Chen-Chuan Chang (2007)

난안전 기록정보자원에 대한 수집 기준을 좀 더 구체화시키는 단계로서 전수 조사된 내용을 분석하여 아카이빙 대상을 결정해야 한다. 선별기준은 재난안전 기록정보자원 아카이브 구축의 목적에 따라 정한다. 본 연구에서 제안하는 재난안전 기록정보자원 아카이브 구축의 목적은 첫째, 아카이브 대상과 관련 실질적 기록 정보 및 현장정보를 제공하고 둘째, 관련 분야 연구자에게 필요한 학술정보를 제공하며 셋째, 관련기관 및 단체에게 실무에 필요한 정보를 제공하고 넷째, 일반이용자에게 실생활에 유용한 재난관련 정보를 제공하는 데 있다. 이러한 목적을 달성할 수 있는 정보는 수집하고, 그 외 정보는 제외하도록 한다.

또한, 본 단계에서는 링크가 연결되는 페이지의 정보에 대하여 아카이빙 여부를 판단한 후 크

롤링을 통해 자동수집을 할지 결정한다. 그리고 자동으로 수집되어야 할 메타데이터 요소에 대해 결정한 후 요소명을 정의한다(〈표 11〉 참조).

#### 4.2.4 4단계 - 재난안전 기록정보자원 목록 및 메타데이터 자동수집

3단계에서 아카이빙 대상으로 결정된 재난안전 기록정보자원의 목록과 메타데이터 요소를 자동수집할 수 있도록 크롤링 알고리즘을 설계한다. 먼저 웹사이트에 특정 질의를 통해야만 문서를 생성하고 가져올 수 있는 형태의 웹 문서를 심층웹 문서라고 하며 이러한 문서를 수집하기 위해서는 다음과 같은 사항을 고려해야 한다. 먼저 웹 문서 주소이다. 일반적인 정적인 형태의 웹 문서는 하나의 주소에 하나의 문서가 존재하지만 동적인 웹 기록물은 하나의 웹 페이

〈표 11〉 재난안전 기록정보자원 메타데이터 자동수집 요소 정의

| 요소명     | 필수여부      | 요소 정의  |
|---------|-----------|--|
| 제목      | 필수        | <ul style="list-style-type: none"> <li>• 다음을 참조하여 제목으로 기입한다.                             <ul style="list-style-type: none"> <li>- 전자계시물의 경우: 전자계시물의 제목을 기입한다.</li> <li>- 웹페이지의 경우: 웹페이지 최상단에서 확인되는 텍스트를 제목으로 간주한다.</li> <li>- 웹페이지 내 하위 탭으로 다시 페이지 구분이 가능한 경우: 하위 탭을 웹페이지로 간주하고 하위 탭명을 제목으로 기입한다.</li> </ul> </li> </ul> |
| 원출처     | 필수        | <ul style="list-style-type: none"> <li>• 정보를 생산한 기관명을 원출처로 한다.</li> </ul>  |
| 연도      | 필수        | <ul style="list-style-type: none"> <li>• 작성일, 등록일자, 기간 등의 메타데이터 필드에서 확인되는 정보로, 재난안전 기록정보 자원이 생산, 게시, 등록된 날짜를 의미하며 “년도”만 기입한다.</li> </ul>   |
| 첨부파일명   | 해당시<br>필수 | <ul style="list-style-type: none"> <li>• 전자계시물에서 첨부파일이 확인되는 경우 첨부파일 명을 기입한다.</li> </ul>  |
| 첨부파일 링크 | 해당시<br>필수 | <ul style="list-style-type: none"> <li>• 전자계시물에서 확인되는 첨부파일의 링크를 기입한다.</li> </ul>   |
| 기술내용    | 필수        | <ul style="list-style-type: none"> <li>• 내용(콘텐츠)을 기입한다.</li> </ul>   |
| 번호      | 해당시<br>필수 | <ul style="list-style-type: none"> <li>• 전자계시물의 경우 게시물 번호를 기입한다.</li> </ul>  |
| 조회수     | 해당시<br>필수 | <ul style="list-style-type: none"> <li>• 전자계시물의 경우 조회수를 기입한다.</li> </ul>   |
| 사이트경로   | 필수        | <ul style="list-style-type: none"> <li>• 전자계시물 혹은 웹페이지까지 도달하기 위한 사이트경로를 기입한다.</li> </ul>   |
| URL     | 필수        | <ul style="list-style-type: none"> <li>• 웹페이지 세부 주소를 작성한다.</li> </ul>  |

지 안에서 스크립트 실행을 통해 여러 페이지의 정보가 제공되기 때문에 각각의 명령어를 웹페이지의 주소처럼 수집할 필요가 있다. 다음은 실행 시간이다. 방문 즉시 모든 정보를 제공하는 정적인 웹 기록물과 달리 심층 웹 문서의 경우 스크립트가 실행되어 페이지를 완전히 완성하기 전까지 대기한 후 웹페이지 정보를 읽어야 한다. 네트워크의 상태 및 서버의 상황에 따라 응답시간이 달라질 수 있기 때문에 특정 시간동안 대기한다고 해도 모든 문서를 크롤링 하기는 쉽지 않다.

재난안전 관련기관의 홈페이지는 크게 웹페이지와 게시판으로 구성되어 있으며, 이들에 대한 메타데이터를 수집하기 위한 방법은 다음과 같다. 먼저, 웹페이지를 수집하기 위해서는

웹 페이지와 수집 정보가 어떠한 형태로 저장되어 있는지 분석하고 수집된 정보를 데이터베이스에 저장하는 과정을 거치게 된다. 예를 들어 중앙응급의료센터 응급상황 시 대처요령 웹페이지<sup>2)</sup>의 수집 과정은 다음과 같이 진행된다. 먼저 <그림 6>과 같이 해당 웹 페이지를 분석해보면 콘텐츠와 관련된 정보는 <div class="layout-content" id="layoutContents"> 안에서 제공되고 있음을 알 수 있다. 즉 DIV 태그의 ID 속성 값이 layoutContents인 정보를 수집해야 원하는 정보들을 저장할 수 있다. 개발된 크롤러에 관련 속성 정보들을 입력하게 되면 해당 정보를 수집하여 데이터베이스에 저장하게 된다.



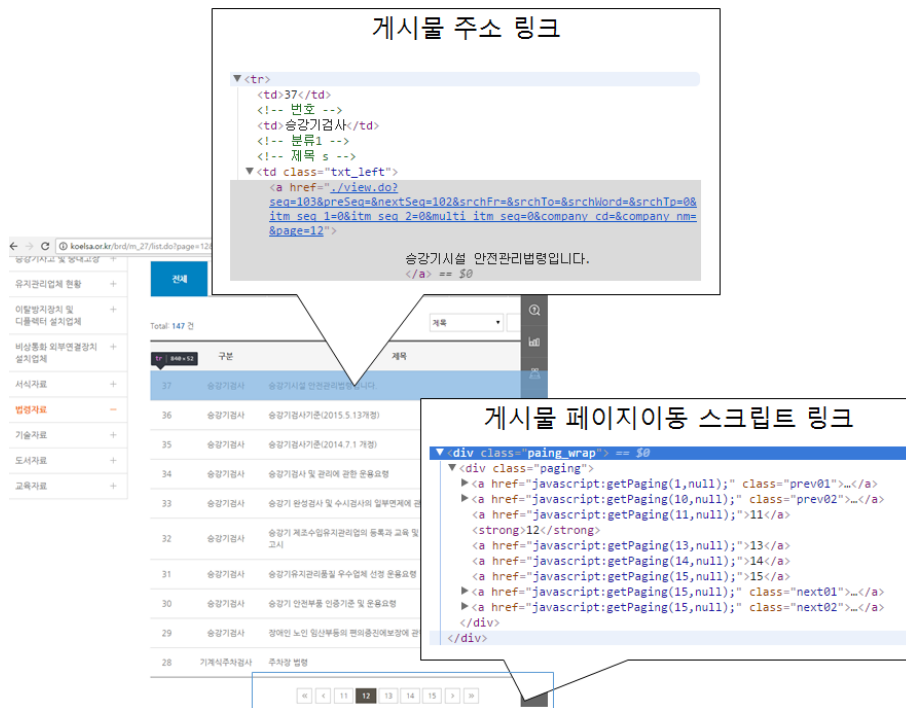
<그림 6> 중앙응급의료센터 응급상황 시 대처요령 웹 페이지 분석

2) 중앙응급의료센터. [http://www.e-gen.or.kr/egen/emergency\\_treat.do](http://www.e-gen.or.kr/egen/emergency_treat.do)

게시판 수집의 경우 웹 페이지 수집 과정에 게시물의 주소 수집과정이 추가된다. 웹 페이지의 경우 하나의 웹 페이지에 하나의 콘텐츠가 존재하지만 게시판의 경우 하나의 웹 페이지에 여러 게시물이 존재하고 각 게시물마다 고유한 주소가 존재하기 때문이다. 그렇기 때문에 크롤러의 게시판 수집 과정은 게시물 주소를 수집하고 수집한 주소를 웹 페이지를 수집하는 것과 동일한 방법으로 실행된다. 예를 들어 한국승강기안전공단의 승강기정보 게시판<sup>3)</sup>의 정보를 수집하기 위해서는 게시판 주소 및 게시물의 주소가 있어야만 관련 정보들을 수집할 수 있다. <그림 7>과 같이 게시물 주소

링크를 확인해보면 TR태그 내 3번째 TD태그에서 게시물 링크를 확인할 수 있다. 링크의 주소가 스크립트 형태가 아니기 때문에 크롤러는 이 주소를 직접 방문하여 게시물을 가져 올 수 있다. 하지만 게시물들이 여러 페이지에 나뉘어서 제공되기 때문에 게시판의 경우 첫 페이지를 읽고 나서 다음 페이지로 이동 후 링크 정보를 수집해야 한다.

한편, 게시물이 고유한 주소를 가지고 있지 않은 동적인 게시판의 경우, 예를 들어 Ajax기술과 같이 특정 스크립트 실행을 통해서 서버로부터 필요시마다 정보를 전달받아 동적으로 게시물을 생성하는 경우, 게시물은 고유한 주소를 가지고



<그림 7> 한국승강기안전공단의 승강기정보 게시판 분석

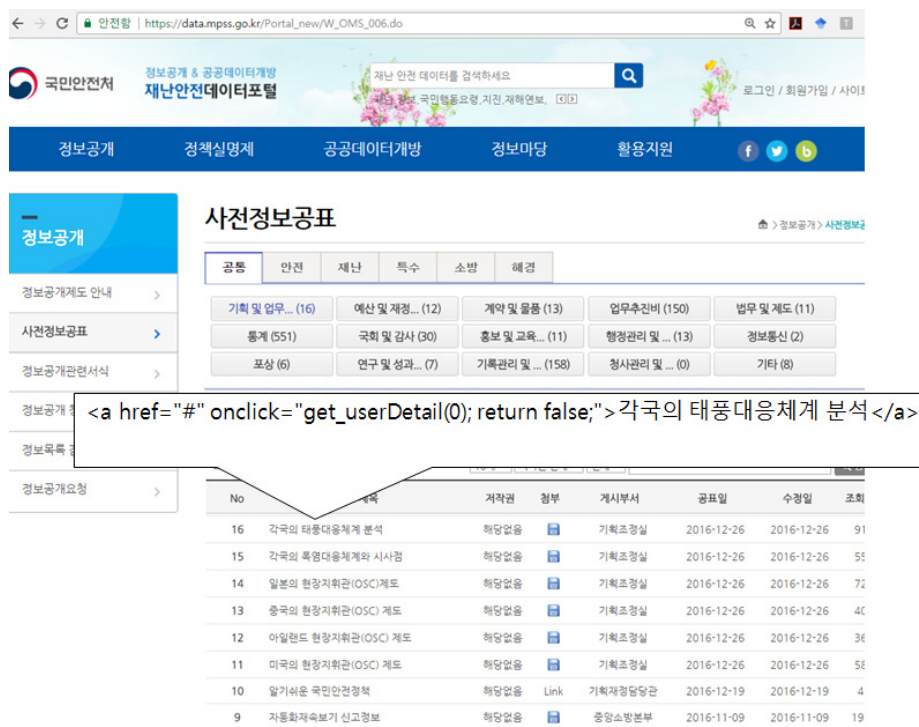
3) 한국승강기안전공단. [http://koelsa.or.kr/brd/m\\_27/list.do?page=12](http://koelsa.or.kr/brd/m_27/list.do?page=12)

있지 않다. 이 경우 주소가 존재하지 않기 때문에 주소 대신 웹 페이지를 호출하는 명령어를 수집해야 한다. 국민안전처<sup>4)</sup> 사전정보공표 게시판<sup>5)</sup>의 경우가 이에 해당한다. <그림 8>과 같이 국민안전처 사전정보공표 게시판을 분석해보면 게시물에 접근하기 위해서 get\_userDetail 메소드를 웹브라우저의 인터프리터가 실행하고 실행 결과를 서버로부터 전달 받아 웹브라우저가 웹 페이지를 갱신한 결과를 보여줘야 게시물 내용을 확인할 수 있다. 국민안전처 사전정보공표 게시판을 분석한 결과 <그림 8>의 get\_userDetail 메소드는 게시물 순서에 따라 숫자를 입력으로 받고

실행 결과를 국민안전처 서버로부터 전달 받아 게시물 내용을 보여준다. 이와 같은 경우 게시물 관련 명령어를 수집하고 실행한 다음 실행결과를 크롤링 해야 한다.

### 5. 활용방안

재난안전 기록정보자원 아카이브는 재난안전 기록정보자원을 체계적으로 수집, 정리, 보존, 검색, 제공하는 재난안전 기록정보자원에 대한 지식의 보고이자 정보교류의 장으로서 역할 수




<그림 8> 국민안전처 사전정보공표 게시판 분석

4) 現 행정안전부 재난안전관리본부, 본 연구는 진행된 당시(2017년 7월까지) 기관명을 기입함  
 5) 국민안전처, https://data.mpss.go.kr/Portal\_new/W\_OMS\_006.do

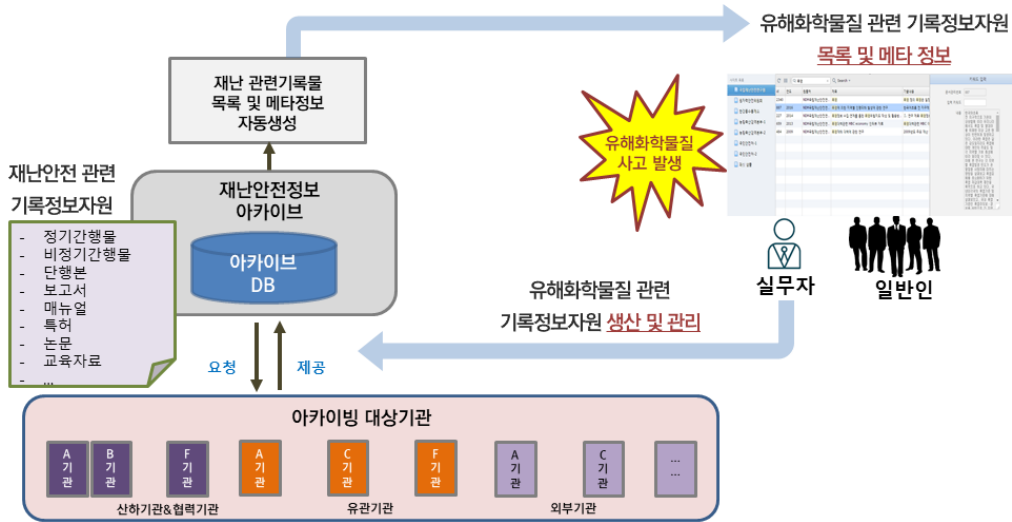
행이 가능하다. 이를 통해 쉽게 사라져 버리는 웹 상의 재난안전 기록정보자원을 보존하는 한편, 이를 활용하고자 하는 이용자들의 다양한 요구사항을 만족시킬 수 있다. 이와 관련하여 먼저, 재난안전 기록정보자원 아카이브를 통해 제공된 재난안전 기록정보자원의 활용도를 높이기 위한 웹 기반 아카이빙 서비스를 개발할 필요가 있다. 예컨대, 실시간 검색어 1위를 키워드로 가지는 기록을 가장 많이 보유한 기관을 순서대로 나열하여 보여주거나 기관을 클릭하게 되면 기관에서 보유하고 있는 해당 키워드의 기록목록 제공이 가능하다. 또한 기관기록목록을 팝업 형태로 제시하여 언제든지 첫 화면으로 돌아갈 수 있게 하는 편의성을 제공하거나, 실시간 검색어 1위를 키워드로 가지는 기록목록을 제공하는 등 이용자의 재난안전 기록정보자원 검색의 편의성과 효율성을 높여줄 수 있는 인터페이스가 가능하다. <그림 9>는 웹 기반 아

카이빙 서비스를 통해 이용자에게 제공할 수 있는 재난안전 기록정보자원의 상세내용이다.

보다 근본적으로는 재난안전 기록정보자원 아카이브는 이용자 집단에 따라 실무자에게는 재난 발생 시 신속한 관련 기록정보자원의 목록 및 메타정보를 제공을 통한 재난관련 업무 지원이, 일반 이용자를 위해서는 재난관련 기록정보 제공을 통한 대국민서비스 지원이 가능하다. 예컨대, <그림 10>과 같이 재난안전 유관 기관에서 수집한 다양한 재난안전 관련 기록정보자원을 재난안전 기록정보자원 아카이브에 저장한 후 재난관련기록물 목록 및 메타데이터를 자동 생성이 가능하다. 이를 통해 재난이 발생했을 때 이용자와 실무자들이 해당 재난과 관련된 기록정보자원을 한 곳에서 통합하여 검색하고 보다 빠르게 재난에 대처할 수 있도록 관련 정보를 제공함으로써 효율적인 업무 지원이 가능하다.

| 상세보기        |                     |   |   |        |        |  |
|-------------|---------------------|---|---|--------|--------|--|
| 제목          | 가뭄관련 저수지 현장조사 결과 보고 |   |   |        |        |  .pdf |
| 패킷분류        | 재난유형                | 자연재난)가뭄   | 재난관리단계  | 예방     | 피해대상   | 물적피해   |
|             | 주관기관                | 중앙행정기관)5처   | 정보유형  | 비정형)문서 | 재난안전기술 | 재난안전체계/분석  |
| 위치정보1_원출처   |                     |   | 국립재난안전연구원 ( <a href="http://www.ndmi.go.kr/index.jsp">http://www.ndmi.go.kr/index.jsp</a> )               |        |        |  |
| 위치정보2_사이트경로 |                     |   | HOME > 연구실적 > 현장조사보고  |        |        |  |
| 위치정보3_URL   |                     |   | <a href="http://www.ndmi.go.kr/research/place/list.jsp">http://www.ndmi.go.kr/research/place/list.jsp</a> |        |        |  |
| 위치정보4_목록번호  |                     |   | 64  |        |        |  |
| 기술내용        |                     | <input type="checkbox"/> 개요<br>○ 조사목적 : 평년대비 많은 강수에도 불구하고 지속적으로 저수량 확보에 어려움이 발생함에 따라 저수현황 및 문제점 분석을 위한 현장 출장<br>○ 일 시 : '16.6.15. ~ '16.6.17.<br>○ 장 소 : 인천시 강화군 삼산면 (삼산저수지, 상하저수지)<br>○ 방문기관 : 삼산·상하 양수장, 한국농어촌공사 강화지사 삼산지소<br>※ 세부내용은 첨부파일 참조 |   |        |        |  |
| 관계정보        |                     | -   |   |        |        |  |
| 관련기관        |                     | 국민안전처: <a href="http://www.mpss.go.kr/home/index.html">http://www.mpss.go.kr/home/index.html</a>  |   |        |        |  |

<그림 9> 재난안전 기록정보자원 아카이브 UI 상세보기 예시



〈그림 10〉 재난안전 기록정보자원 아카이브 활용방안 예시

## 6. 결론

최근 전 세계적으로 재난에 대한 관심이 높아지고 있으며, 우리나라 역시 국민안전을 위한 대응책으로써 재난관련 기관을 신설하고, 관련 시스템을 구축하는 등 재난대응 및 관리를 위한 활동 등이 활발하게 진행되고 있다. 이에 따라 재난과 관련하여 생산되는 원시자료 및 가공정보 등도 급속하게 증가하고 있지만, 이들 정보의 체계적인 관리와 활용을 위한 통합관리 체계는 여전히 미흡한 상황이다. 따라서 국가적 차원에서 재난재해에 대한 관심의 확대와 국민안전을 위한 재난안전 기록정보자원의 체계적인 관리와 활용을 위한 통합관리 체계가 필요하다. 특히, 국가안전체의 통합재난안전체계에서 관리·연동되는 재난안전 기록정보자원을 기반으로 신규 재난안전 기록정보자원, 타 부처 정보 및 비정형데이터의 통합관리 및 민간 활용을 위해 향후 통합 플랫폼 구축이 필요하다.

이에 본 연구는 범정부 차원에서 재난안전 기록정보자원을 제공하여 특히 실무자를 위한 재난안전 기록정보자원의 공유 틀로서의 역할을 수행할 수 있는 자동수집 기반의 재난안전 기록정보자원 아카이브 구축전략을 제안하였다. 재난안전 기록정보자원 아카이브 구축을 위한 고려사항은 첫째, 아카이빙 대상이 되는 기록물의 정의 및 범위를 명확히 해야 하며, 둘째, 지속적인 수집을 위해 자동수집 방안을 개발해야 하고, 셋째, 아카이빙 대상의 선별·평가를 위한 전략이 수립되어야 한다. 마지막으로 공유의 편의성 및 공개성을 보장하여 재난안전 기록정보자원의 목록 및 위치정보 등을 제공함으로써 이용자들이 재난안전 기록정보자원에 보다 손쉽게 접근할 수 있도록 해야 한다. 이를 고려하여 본 연구에서 제안하는 자동수집 기반의 재난안전 기록정보자원 구축전략은 총 4단계로서 다음과 같다. 1단계에서는 자동수집을 위한 사전조사로써 재난안전 유관기관이 보유한 정보자원에 대한 전

수조사를 실시하여 수집하여 할 재난안전 기록 정보자원을 파악하고, 이에 대한 세부정보를 작성해야 한다. 2단계는 1단계에서 수집된 정보에 대하여 자동수집 가능 여부를 판단하고, 기술적인 부분에 대한 세부정보를 면밀히 파악해야 한다. 3단계에서는 2단계 분석결과를 바탕으로 아카이빙 대상 선별을 위한 수집기준 및 세부지침 등을 마련하는 등 아카이빙 수집정책을 마련해야 하며, 마지막 4단계에서는 아카이빙 대상으로 선별된 재난안전 기록정보자원에 대한 목록 및 메타데이터를 자동수집 할 수 있는 방안을 마련해야 한다.

재난안전 기록정보자원 아카이브는 재난안전 기록정보자원을 체계적으로 수집, 정리, 보존, 검색, 제공하는 역할을 수행해야 한다. 이를 통해 휘발성을 가진 웹상의 재난안전 기록정보자원을 보존하는 동시에 이용자들의 다양한 요구사항을 만족시킬 수 있어야 한다. 본 연구에

서 제안하는 자동수집 기반의 재난안전 기록정보자원 아카이브를 구축하게 되면 여기저기 산재해 있는 재난안전 기록정보자원을 통합적으로 자동수집하여 제공할 수 있어 재난안전 관련 연구 및 국가 정책 수행에 따른 중요한 정보원으로 활용할 수 있다. 또한 재난안전 기록정보자원의 통합관리를 위한 정보의 공유 및 활용 기반을 구축하게 되어 재난안전 기록정보자원의 상호연계성을 통한 정보의 시너지 효과를 창출하고, 나아가 재난안전 기록정보자원의 가치를 극대화할 수 있을 것으로 기대된다.

향후 연구 방향은 본 연구에서 제시한 구축전략에 대한 검증으로, 제안된 전략을 통해 재난안전 기록정보자원을 실제로 수집하고, 수집된 기록정보자원의 현황 및 특징을 분석하여 재난안전 기록정보자원 아카이브 구축 가능성을 확인하고자 한다.

## 참 고 문 헌

- 김광영, 이원구, 윤희목, 신성호, 이민호 (2011). 웹 자원 아카이빙을 위한 웹 크롤러 연구 개발. 한국콘텐츠학회논문지, 11(9), 9-16.
- 김유승 (2007). 웹 아카이빙의 법·제도적 문제에 대한 고찰. 한국문헌정보학회지, 41(3), 5-24.
- 김유승 (2008). 복합적 웹 아카이빙 정책에 관한 고찰: 프랑스국립도서관의 사례를 중심으로. 한국문헌정보학회지, 42(4), 159-179.
- 김태영, 강주연, 김혜영, 김용 (2017). LOD기반의 재난안전 정보서비스 확장에 관한 연구. 한국문헌정보학회지, 51(3), 163-188.
- 김현정 (2017). 재난기록의 웹 아카이빙에 대한 연구. 한국기록관리학회지, 17(2), 27-56.
- 박성원 (2014). Introduction: 우리가 살면서 만날 재난과 그 대응. 과학기술정책연구원, Future Horizon: Summer, 21, 4-7.

- 박태연, 한희정, 김용, 김수정 (2017). 재난안전정보의 통합 관리를 위한 분류체계 현황분석 및 개선방안에 관한 연구. 한국비블리아학회지, 28(3), 125-150.
- 서혜란 (2004). 웹 아카이빙의 성과와 미래 전망. 한국비블리아발표논문집, 10, 7-25.
- 유호선, 명현, 오효정 (2017). 소셜미디어에서의 사회재난 이슈 생존 주기 분석. 한국정보기술학회논문지, 15(7), 139-147.
- 윤정옥 (2010). 웹 아카이브 OASIS에 관한 고찰. 한국문헌정보학회지, 44(2), 5-27.
- 이관형, 이원호, 양원직 (2016). 사회재난의 재난위해분석에 관한 연구. 한국방재안전학회 학술대회 논문집, 84-92.
- 이수진, 한희영, 심민정, 원동현, 김용 (2017). 웹 아카이빙을 위한 정보자원의 자동수집 알고리즘 개발 및 구현. 한국도서관정보학회 하계 학술발표회, 127-134.
- 이연수, 남성운, 윤대현 (2013). 대규모 웹 기록물의 원격수집을 위한 콘텐츠 중복 필터링 개선 연구. 기록학연구, 35, 133-160.
- 이혁재, 김유승 (2009). 정부기관 웹 기록물관리 정책에 관한 연구. 한국비블리아학회지, 20(3), 33-46.
- 정지나, 정힘찬, 김용 (2017). 언어네트워크 분석을 통한 재난안전정보와 관련한 국내 연구동향 분석. 한국비블리아학회지, 28(3), 67-93.
- 정혜지 (2016). 미국 9·11 메모리얼의 재난기록물 수집에 관한 연구. 석사학위논문. 명지대학교 기록정보과학대학원, 기록관리전공.
- 정힘찬, 김태영, 김용, 오효정 (2017). 재난안전정보 관리를 위한 어휘자원 현황분석 및 활용방안. 정보관리학회지, 34(2), 137-158.
- 차승준, 이규철 (2008). 웹기록물 아카이빙 기반기술 연구 개발. 한국전자거래학회 심포지움 및 기타 간행물, 359-368.
- 차승준, 최윤정, 이규철 (2009). 공공기관 심층 웹기록물 아카이빙을 위한 메타데이터 설계. 한국전자거래학회지, 14(4), 181-193.
- 최형섭 (2014). 재난의 기록: 재난보고서의 사회적 기능. 과학기술정책연구원, Future Horizon: Summer, 21, 24-27.
- 한국기록관리협회 (2009). 전자기록물의 이해. 경기: 조은글터.
- 한희정, 박태연, 오효정, 김용. (2017). 재난안전정보 아카이브 구축을 위한 온라인 기록정보 현황분석 및 개선방안 연구. 한국도서관·정보학회지, 48(2), 187-213.

• 국문 참고자료의 영어 표기

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

Cha, Seung-Jun & Lee, Kyu-Chul (2008). Research and development of base technology for archiving

- web records. Federation of Knowledge Information Societies Conference, 359-368.
- Cha, Seung-Jun, Choi, Yun-Jeong, & Lee, Kyu-Chul (2009). Metadata design for archiving public deep web records. *The Journal of Society for e-Business Studies*, 14(4), 181-193.
- Choi, Hyung Seob (2014). Record of Disaster: the social role of disaster reports. *Future Horizon*, 21, 24-27.
- Chung, Hye-Ji (2016). A study of acquisition for disaster collections on The National September 11 Memorial & Museum. Master's Thesis, Major of Records and Archival Information Management, Myongji University. Seoul.
- Han, Hui-Jeong, Park, Tae-Yeon, Oh, Hyo-Jung, & Kim, Yong (2017). 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.
- Jeong, Him-Chan, Kim, Tae-Young, Kim, Yong, & Oh, Hyo-Jung (2017). A Study on the Utilization Plan of Lexical Resources for Disaster and Safety Information Management Based on Current Status Analysis. *Journal of the Korean Society for Information Management*, 34(2), 137-158.
- Jeong, Ji-Na, Jeong, Him-Chan, & Kim, Yong (2017). Analysis of Trends on Disaster Safety Information based on Language Network Analysis Methods. *Journal of the Korean Biblia Society for Library and Information Science*, 28(3), 67-93.
- Kim, Hyunjung (2017). A Study on the Web Archiving of Disaster Records - Focusing on the Roles and Functions of the OASIS Disaster Archive -. *Journal of Korean Society of Archives and Records Management*, 17(2), 27-56.
- Kim, Kwang-Young, Lee, Won-Goo, Yoon, Hwa-Mook, Shin, Sung-Ho, & Lee, Min-Ho (2011). Development of Web Crawler for Archiving Web Resources. *Journal of the Korea Contents Association*, 11(9), 9-16.
- Kim, Tae-Young, Gang, Ju-Yeon, Kim, Hye-Young, & Kim, Yong (2017). A Study on the Extension of Disaster Safety Information Service based on Linked Open Data. *Journal of the Korean Society for Library and Information Science*, 51(3), 163-188.
- Kim, You-Seung (2007). A study of legal issues for web archiving. *Journal of the Korean Society for Library and Information Science*, 41(3), 5-24.
- Kim, You-Seung (2008). A study of combined web archiving policy: BnF's three layers web archiving strategy. *Journal of the Korean Society for Library and Information Science*, 42(4), 159-179.
- Korea Society of Archives and Records Management (2009). *Understanding Electronic Records*.

Gyeonggi: Joeungeulteo.

- Lee, Hyuk-Jae & Kim, You-Seung (2009). A Study on the Web Based Records Management Policy for Government Agencies. *Journal of the Korean Biblia Society for Library and Information Science*, 20(3), 33-46.
- Lee, Kwan Hyoung, Yi, Waon Ho, & Yang, Won Jik (2016). A Study on Risk Analysis of Social Disaster. *Korean Society of Disaster & Security Conference Proceedings*, 84-92.
- Lee, Su-Jin, Han, Hui-Lyeong, Sim, Min-Jeong, Won, Dong-Hyun, & Kim, Yong (2017). The Development of Automatic Collection Algorithm to Collect Information Resources for Web Archiving: Focused on Disaster safety Information. *Journal of Korean Library and Information Science Society Summer Conference Proceedings*, 127-134.
- Lee, Yeon-Soo, Nam, Sung-un, & Yoon, Dai-Hyun (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.
- Park, Seong-won (2014). Introduction: Disasters we will encounter while living and their response. *Future Horizon: Summer*, 21, 4-7.
- Park, Tae-Yeon, Han, Hui-Jeong, Kim, Yong, & Kim, Soojung (2017). A Study on the Analysis and Improvement of Classifications for Integrated Management of Disaster and Safety Information. *Journal of the Korean Biblia Society for Library and Information Science*, 28(3), 125-150.
- Suh, Hye-Ran (2004). Web archiving: What we have done and what we will do. *The Korean Biblia Society for Library and Information Science Conference Proceedings*, 10, 7-25.
- Yoo, Ho-Suon, Myung, Hyun, & Oh, Hyo-Jung (2017). Issue Survival Cycle Analysis of Social Disasters in Social Media. *The Journal of Korean Institute of Information Technology*, 15(7), 139-147.
- Yoon, Cheong-Ok (2010). Research on the OASIS, a web archive in Korea. *Journal of the Korean Society for Library and Information Science*, 44(2), 5-27.