

# 웹 아카이빙을 위한 정보자원의 자동수집방법 개발\*

- 재난안전정보를 중심으로 -

## The Development of Automatic Collection Method to Collect Information Resources for Web Archiving: With Focus on Disaster Safety Information

이 수 진 (Su Jin Lee)\*\* , 한 희 영 (Hui Lyeong Han)\*\*\*  
심 민 정 (Min Jeong Sim)\*\*\*\* , 원 동 현 (Dong Hyun Won)\*\*\*\*\*  
김 용 (Yong Kim)\*\*\*\*\*

### 목 차

- |           |                  |
|-----------|------------------|
| 1. 서론     | 4. 재난안전정보 자동수집방안 |
| 2. 이론적 배경 | 5. 결론            |
| 3. 현황조사   |                  |

### <초록>

본 연구는 각 기관별로 산재되어 있는 재난안전정보의 효율적인 공유 및 활용과 일반 이용자들의 재난안전정보의 접근성 향상을 위하여 웹 크롤러를 활용한 자동수집방안을 제시하고자 한다. 그러나 현재 웹은 심층 웹이 대부분을 차지하고 있어 일반적인 아카이빙 전략에 사용되는 크롤러로 수집이 불가능하다. 이에 따라 본 연구에서는 심층 웹의 논리적 구조를 분석하여 정보 자원을 수집할 수 있는 알고리즘을 개발하고 실제 크롤러를 구현하였다. 이를 통해 재난안전정보의 자동수집을 수행하였고 이는 재난안전정보의 공유 및 활용으로 효율적인 재난 업무에 도움을 줄 것으로 예상된다. 나아가 본 연구에서 구현한 크롤러의 범용화를 통해 심층 웹 형태의 정보 자원을 자동수집할 수 있다는 점에 의의가 있다.

주제어: 재난안전정보, 웹 아카이빙, 자동수집, 정보공유, 심층 웹

### <ABSTRACT>

This study aims to provide the efficient sharing and utilization method of disasters scattered by each institution and develop automated collection algorithm using web crawler for disaster information in deep web accounts. To achieve these goals, this study analyzes the logical structure of the deep web and develops algorithms to collect the information. With the proposed automatic algorithm, it is expected that disaster management will be helped by sharing and utilizing disaster safety information.

Keywords: disaster safety information, web archiving, automatic collecting, information sharing, deep web

\* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016S1A5B8913575).

\*\* 전북대학교 기록관리학과 대학원(23635417@hanmail.net) (제1저자)

\*\*\* 전북대학교 기록관리학과 대학원(sharinny@naver.com) (공동저자)

\*\*\*\* 전북대학교 기록관리학과 대학원(smj810903@naver.com) (공동저자)

\*\*\*\*\* 전북대학교 재난정보표준화사업단 전임연구원(dkwon@jbnu.ac.kr) (공동저자)

\*\*\*\*\* 전북대학교 문헌정보학과 교수, 문화융복합 아카이빙연구소장(yk9118@jbnu.ac.kr) (교신저자)

■ 접수일: 2017년 7월 24일 ■ 초심사일: 2017년 11월 16일 ■ 게재확정일: 2017년 11월 20일

■ 한국기록관리학회지 17(4), 1-26, 2017. <<http://dx.doi.org/10.14404/JKSARM.2017.17.4.001>>

## 1. 서론

### 1.1 연구배경 및 필요성

지난 2014년 4월 발생한 세월호 참사는 전 국민적으로 커다란 충격과 혼란을 가져왔다. 세월호 침몰 당시 정부의 미흡한 대처능력은 국민들에게 국가의 재난대응시스템에 대한 커다란 불신을 초래했다. 세월호 참사 이후 재난에 대한 정부와 국민들의 경각심에도 불구하고 연달아 크고 작은 재난이 발생함에 따라 정부는 국가재난안전체계를 개선하였다. 세월호 참사라는 재난 상황에서 제 역할을 수행하지 못한 중앙행정부처를 해체(해양경찰청)하거나 축소(해양수산부, 안전행정부)하고 재난관리총괄기관으로 국민안전처를 설립하여 안전 관련 조직을 통합하였다. 이를 통해 정부는 실제 재난 상황 발생 시 드러났던 초기 대응의 미흡, 재난관리컨트롤타워 역할의 부재 등과 같은 국가재난대응시스템이 가진 전반적인 한계를 해결하고 국가재난관리컨트롤타워를 구축하고자 하였다(이상명, 2014). 또한 재난발생의 예방 및 재난발생 시 효율적인 초기 대응 그리고 재난의 확산을 방지하기 위해 재난 및 안전관리법을 개정하였다(김용섭, 2016).

정부의 국가재난안전체계의 개선에도 불구하고 1978년 대한민국 지진 관측을 시작한 이래 역대 최대 규모로 2016년 9월 12일 경주에서 발생한 5.8의 강한 지진에 대한 정부의 미흡한 대응과 재난대응시스템은 국민들에게 더 큰 불신을 안겨주었다. 경주 지진은 전국 대부분의 지역뿐만 아니라 일본, 중국 등에서도 감지될 정도로 강력했으며 대구, 경북, 구미, 부산 영남

권에는 통신장애 및 통화불능사태를 불러오면서 국민들에게 큰 혼란과 불안감을 불러일으켰다. 이는 우리 사회에 커다란 변화를 가져왔고 더불어 새로이 조직된 국가재난관리컨트롤타워로서의 국민안전처의 역할에 대한 의문을 가지게 하는 계기가 되었다. 이러한 상황을 기반으로 우리나라의 재난대응시스템의 전반적인 개선과 획기적인 개선방향을 요구하게 되었으며, 재난발생 시 대응 대책의 신속성과 일원화된 시스템 등이 개선점으로 떠올랐다. 또한 신속하게 국민들에게 상황을 알리고 대처할 수 있는 체계적인 재난대응시스템을 구축해야 한다는 여론이 높아지게 되었다.

『재난 및 안전관리 기본법(이하 재난 기본법)』에 따르면 재난 상황에 있어 재난관리업무의 효율성을 위해 각각의 재난 관련 기관은 수집·보유하고 있는 정보를 서로 공유해야 한다. 또한 국민안전처는 재난총괄기관으로 재난 정보의 효율적 관리 및 재난 시 신속 대응을 위해 재난 정보를 통합적으로 관리·운영해야 한다. 그러나 정부 교체에 따라 행정조직이 개편됨에 따른 잦은 조직과 인력의 변동 그리고 재난 관련 부서의 축소와 폐지는 효율적인 재난관리업무를 위한 정보 공유를 어렵게 한다. 뿐만 아니라 재난 전 과정의 정보화와 통합 운영을 위한 국가재난관리 정보시스템(NDMS; National Disaster Management System)이 있음에도 불구하고 재난 정보가 소관부처별로 분산되어 존재해 효율적인 정보 공유가 미흡한 실정이다(현석환, 2015).

재난은 매 상황마다 변동되는 정보를 생산한다. 또한 재난 정보는 국민의 안전과 관련된 정보로 일반 행정 정보와 비교해보았을 때 확연

히 그 가치가 높다고 할 수 있다. 대부분의 재난 정보는 웹 환경을 통해 실시간 상황정보, 기관별 재난관리정보 등 그 양이 방대하다. 그러나 각 기관별, 부서별로 재난 관련 정보들이 산재되어 있고, 그 위치정보를 파악하기 위해서는 각 재난주관기관의 고유 사이트에 들어가 직접 검색을 하거나 정보를 요청해야 하는 등과 같은 번거로움이 있다. 따라서 재난 관련 정보들이 재난 상황에 적절히 활용되고 각 기관 간의 원활한 정보 공유를 위해서는 재난 정보의 자동수집을 통한 통합적 관리·운영이 필요하다. 또한, 기관 간의 원활한 정보 공유뿐만 아니라 일반 국민들의 알 권리와 정보 욕구를 충족시켜주기 위해 국민안전처에서 통합 관리하고 있는 다양한 재난 관련 정보 및 기록물을 손쉽게 접근하기 위한 대책 마련이 필요하다. 대부분의 행정 정보와 같이 재난 정보 역시 웹 환경에서 정보를 생산하고 관리하고 있다. 이는 재난 정보가 웹 자원과 마찬가지로 휘발성, 역동성의 특성을 보여 중요한 가치를 지니는 재난 정보가 손실될 수 있는 위험성을 지니고 있다는 것을 의미한다. 따라서 방대한 양의 재난 정보의 수집을 위해 웹 크롤러와 같은 자동수집도구를 활용한 자동수집이 필요하다. 뿐만 아니라 웹의 성장과 함께 대부분의 웹 자원이 심층 웹에 존재하여 일반적인 크롤러를 통한 자동수집에 어려움이 존재해 이러한 기술적 한계를 극복할 수 있는 재난 정보의 자동수집 및 웹 아카이빙을 위한 대책이 요구된다.

따라서 본 연구는 심층 웹 환경에서 존재하고 있는 재난 정보의 효과적인 자동수집을 위해 웹 크롤러의 개발 및 구현을 통한 자동수집방안을 제시하고자 한다. 이는 국가재난관리 컨트

롤타워인 재난총괄기관이 재난의 예방, 대비, 대응, 복구 단계에서 신속한 대응과 효율적인 관리·운영을 하여 제 역할을 수행할 수 있도록 하며, 각 기관별로 산재되어 있고 기관의 고유 사이트에서 직접 검색을 통해야만 획득할 수 있는 재난 정보에 대해 누구나 손쉽게 접근할 수 있도록 하기 위한 재난 정보 웹 아카이빙의 기반 연구로 의의가 있다.

## 1.2 연구내용 및 방법

본 연구의 목적은 심층 웹 환경을 통해 생산되는 재난 정보를 수집해 기관 및 일반 이용자들이 손쉽게 접근하고 이를 공유·활용할 수 있도록 웹 크롤러를 활용한 재난 정보의 자동수집 방안을 제안하는 것을 목적으로 한다.

이러한 목적을 달성하기 위하여 본 연구의 범위를 다음과 같이 설정하였다. 『재난 기본법』은 재난 업무를 주관하는 중앙행정기관에 대해 재난관리주관기관이라 명명한다. 재난관리주관기관의 경우 각 기관들의 특성에 맞게 재난 유형이 분류되어 있으며 각 유형에 맞게 대응하도록 명시되어 있다. 이 중 국민안전처의 경우 재난관리총괄기관으로 재난 정보의 통합적 관리·운영에 목적을 가지기 때문에 국민안전처로 범위를 한정 지어 연구를 진행하였다. 특히 국민안전처에서 재난 정보를 관리·운영하고 있는 재난 정보 포털 중 기관은 물론 일반 이용자도 재난 정보를 획득할 수 있도록 대국민서비스를 제공하고 있는 국민재난안전포털을 중심으로 웹 사이트 및 정보 자원의 분석을 실시하였다.

연구의 내용은 자동화된 재난 정보의 수집 및 웹 아카이빙을 위하여 먼저, 재난 주관 기관

의 웹 사이트 분석을 통해서 기관의 업무 분석과 기관에서 산출되는 재난 정보의 유형과 특성을 살펴보았다. 기관 별로 생산되는 정보의 유형과 특성이 표준화되어 있지 않고 통합적으로 관리되고 있지 않다는 점과 웹 사이트 별로 각각의 특성이 다르다는 점에 주목해 재난 정보의 자동수집방안 필요성에 대해 고찰하였다. 둘째, 분석된 웹 사이트의 구조, 특히 기존의 자동수집기능에서는 접근 및 수집이 어려운 심층 웹의 특성과 웹 상의 재난 정보의 유형 및 특성에 대한 분석을 수행하였다. 또한 문헌연구를 통해 웹과 웹 아카이빙에 대해 살펴보고 국내 재난관리시스템의 현황과 실태에 대해 살펴보았다. 셋째, 심층 웹을 포함한 재난 정보의 웹 아카이빙을 위한 자동수집 알고리즘을 개발하였으며, 마지막으로 제안된 알고리즘을 기반으로 웹 크롤러의 설계 및 구현을 수행하였다.

이러한 연구목적과 내용을 달성하기 위하여 본 연구에서는 아래의 방법론을 적용하였다. 먼저, 기관별 웹 사이트에 대한 현황분석을 통하여 재난 정보의 유형을 조사하였다. 지금까지의 재난관리주관기관 간의 재난발생 시 업무 프로세스를 살펴보고자 기관의 웹 사이트 조사를 실시하였고, 자동수집을 위해 국민안전처의 국민재난안전포털 웹 사이트 특성 및 재난 관련 기록물을 분석하였다. 또한 자동수집 및 웹 아카이빙과 관련된 문헌조사를 실시하였다. 그리고 재난 업무 상에서 발생하는 재난 정보의 유형 및 형태를 살펴보고 현재 재난 관리 시스템 상에서 이루어지고 있는 업무 환경의 현황과 문제점을 파악하기 위한 기관 현장조사를 실시하였다.

이상의 연구 방법을 통해 재난관리주관기관

에서 생산하는 재난 정보의 현황과 기관 간 업무 프로세스 분석을 통해 재난발생 시 이루어지는 정보 소통에 대해 파악하여 이를 바탕으로 재난 정보의 자동수집방안 마련의 기초 자료를 획득하였다. 그리고 재난관리주관기관 간의 정보 공유 및 운영의 효율성을 위해서 심층 웹에 존재하고 있는 재난 정보 아카이빙을 위한 자동수집방안 마련에 중점을 두고 연구를 진행하였다.

### 1.3 선행연구

재난에 대한 관심이 높아짐에 따라 재난과 관련된 많은 연구가 이루어지고 있다. 본 연구에서는 이러한 연구들을 재난 정보의 수집방안 연구를 위해 재난 상황에서 정보의 소통을 파악하기 위한 재난관리시스템에 대한 연구와 생산되는 재난 정보 및 기록물을 파악하기 위한 연구, 그리고 웹 환경에서 존재하는 재난 정보의 효과적인 자동수집을 위해 웹 아카이빙에 관한 연구로 크게 세 가지로 구분해 살펴보고자 하였다.

먼저, 재난기록물에 대한 연구를 보면, 정혜지(2016)는 재난기록이 단순히 참사를 기억하고 추모하는 것이 아니라 이전과 다른 변화를 가져오기 위해 필요한 것임을 강조하였다. 또한 9·11테러가 어떤 유형의 재난에 속해 있는지 정리하고 재난기록물의 수집방안을 정리하였다. 이 연구는 다수의 희생자가 발생한 재난 사건과 관련된 기관을 설립하는 과정에서 협력적 거버넌스(governance)가 이루어져야 하며 재난기록을 잘 관리하기 위해서는 시민여론과 정부와의 공정하고 합리적인 의사결정이 내려

져야 한다고 주장하였다. 이미옥(2015)은 '재난 사건'으로 인해 생산되는 공적이거나 사적인 재난기록을 단계별로 관리하여 효과적으로 활용될 수 있는 방안을 연구하는데 목적을 두고 재난기록관리의 선진사례를 분석하여 대구 지하철화재참사 기록관리 현황을 분석하여 기록의 단계별 관리 방안을 제안하였다. 이와 같이 기록학계는 대부분 매뉴스크립트 수집을 다루고 있으며, 기록물에 한정해 수집방안을 제안하고 있다. 이는 행정기관에서 생산하는 기록물뿐만 아니라 개개인으로부터 획득한 기록과 재난의 전반에 걸친 기록을 수집한다는 점에서 의의가 있다. 그러나 재난이라는 긴박한 상황 속에서 발생하는 다양한 정보는 수집 대상에서 제외되었다는 한계를 가진다.

다음으로 재난관리체계에 대한 연구를 살펴보면, 정부와 민간부문의 협력적 네트워크 구축의 중요성을 강조한 박병식(2008)은 재난관리를 효율화하기 위한 방안을 제시하기 위하여 민간재난관리기구인 부산재난안전네트워크의 운영 실태를 분석하여 재난안전네트워크 회원단체 간 역할분담의 명확화, 회원단체의 재난 대응 조직체계와 행동매뉴얼의 구축 등 발전방안을 제시하였다. 재난대응체계 구축을 위한 네트워크체계를 구축하고자 일본과 한국의 재난 대응 사례를 비교 분석한 류상일, 안혜원(2007)은 일본의 경우 중앙정부와 지방정부 등의 공공부문과 민간부문 및 NGO간 상호 네트워크에 의해서 협력적으로 재난에 대응하고 있으나, 한국의 경우 각 부문 간 네트워크 협력체계 구축이 미약하여 더 많은 재난피해를 불러일으키고 있다고 지적하며 중앙정부와 지방정부 등의 공공부문의 노력만으로는 재난관리를 효과적으로

수행하기에는 한계가 있다고 보고 각 주체들을 연결할 수 있는 매개조직이 필요하고 재난 대응 네트워크에 참여하는데 있어 각 기관들의 역할과 기능 분담을 위한 재난 대응 교육과 훈련 프로그램이 요구된다고 하였다. 이들 연구에서는 재난관리체계에 대한 근본적 문제에 대한 해결책 마련을 위해 해외 사례와 비교하는 정책적 연구이다. 하지만 재난 상황에서 정보의 소통 네트워크와 한계를 확인할 수 있었다.

웹 의존도의 향상과 웹 아카이빙 사례 증가에 따라 웹 아카이빙의 중요성이 강조되고 있는 만큼 웹 아카이빙과 관련된 연구들이 활발하게 진행되고 있다. 그중 이연수, 남성운, 박성배(2010)는 공공기관의 웹 사이트 기록물의 본격적인 수집에 대비하여 대표적인 몇몇 기관을 대상으로 실시한 수집 테스트 및 분석 결과를 분석하여 향후 원격수집기의 기능 개선과 원격수집 시 요구되는 수집 기준 자료로 제공하였다. 성시애(2009)는 공공기관의 웹 기록물의 중요성을 인식하여 공공기록을 남기기 위해서는 공공기관의 체계적인 웹 아카이빙의 시행이 시급하다고 하였다. 김유승(2008)은 웹 정보자원이 우리 삶의 전 영역에서 지니는 가치와 중요성에 비해 이를 지키고 보존하려는 제도적 환경과 노력이 미흡하다는 인식 아래 웹 정보 자원의 수집·보존정책의 발전 방향을 언급하고 각국의 웹 아카이빙 사례를 7가지 유형으로 나누어 분석한 결과 최적의 웹 아카이빙 접근법으로서의 복합적 아카이빙 정책을 채택하고 있는 프랑스 국립도서관의 사례를 들어 우리나라의 웹 아카이빙 정책발전을 위한 두 가지 방안을 제시하였다. 이성숙(2005)은 웹 아카이빙의 활성화를 위한 기초 자료를 제공하기 위해 웹 아카이빙 관

런 프로젝트에서 사용한 도구와 프로젝트 주체와 접근방법에 의해 범주화하고 사용된 S/W를 살펴보고 이미 진행된 프로젝트에서 남겨두었던 문제들과 새로운 웹 기반 기술, 인증이 필요한 웹에 대한 연구의 필요성을 주장하였다. 이들 연구는 웹 환경의 발달로 웹 정보 자원의 수집에 대한 필요성을 언급하고 있으나 공공기록을 대상으로 하고 있으며 최근 웹 환경에서 주목받고 있는 심층 웹 아카이빙에 대해서는 한계로 지적하고 있다. 이에 대해 이규희(2009)는 기존 웹 아카이빙의 논의에서 제대로 다뤄지지 않았던 심층 웹의 개념 및 유형을 정의하고, 실질적으로 심층 웹의 아카이빙을 위한 개선방안을 제안하고 있다. 차승준, 최윤정, 이규철(2009)은 웹을 통해 생산, 관리, 활용되는 웹 기록물에 대해 정의하며, 웹 기록물의 한 분류인 심층 웹에 대해 국내외 공공기관에서 제공하는 심층 웹 기록물을 토대로 정의하고 있다. 또한 심층 웹 기록물 아카이빙 기반기술로 더블링크어와 전자기록물 장기보존포맷을 활용해 심층 웹 기록물 메타데이터를 설계하였다. 그러나 우리나라 웹 환경 자체에 대한 분석보다는 외국 사례와의 비교를 통해 분석하고 있으며 심층 웹의 아카이빙을 위한 구체적인 방안을 제안하고 있지 않다는 한계를 지닌다.

## 2. 이론적 배경

### 2.1 웹 아카이빙

#### 2.1.1 웹 아카이빙의 정의

웹에서 생성되는 중요한 정보 자원이 웹의 성

장과 함께 폭발적으로 증가하고 있지만 웹 자원에 대한 의존도가 증폭될수록 웹이 가지고 있는 역동성, 휘발성, 기술의존성 등과 같은 여러 가지 특성으로 인하여 중요한 정보 자원으로서의 가치가 있는 웹 기록물들이 보존되지 않고 수정되고 바뀌고 사라져버리는 문제점이 발생하고 있다(차승준, 이규철, 2008). 이러한 이유로 전 세계적으로 웹 정보 자원들을 보존하기 위한 웹 아카이빙(web archiving)의 중요성과 필요성에 대한 논의와 관심이 점차 커지고 있다. 특히 아카이빙 대상이 되는 웹 자원의 방대함으로 인하여 웹 크롤러(web crawlers)를 활용한 다양한 수집전략이 지속적으로 개발되고 있다(김희정, 2011).

웹 아카이빙(web archiving)의 개념에 대해 기록학 용어 사전에 따르면 웹 자원에 대한 장기적 접근을 보장하기 위한 보존활동으로 정의하고 있다(한국기록학회, 2008). 김정심(2008)은 웹 아카이빙이란 시간의 경과에 의해 변화되거나 그 일부 또는 전체가 없어질 우려가 있기 때문에 웹 상에 여러 가지(학술적, 보존적, 증거적 등)를 갖고 있는 기록물을 아카이빙 함으로써 콘텐츠의 유용성을 보장하고 항구적으로 해당 기록물을 기록, 보존하는 총체적인 과정이라고 정의하였다.

#### 2.1.2 심층 웹의 정의

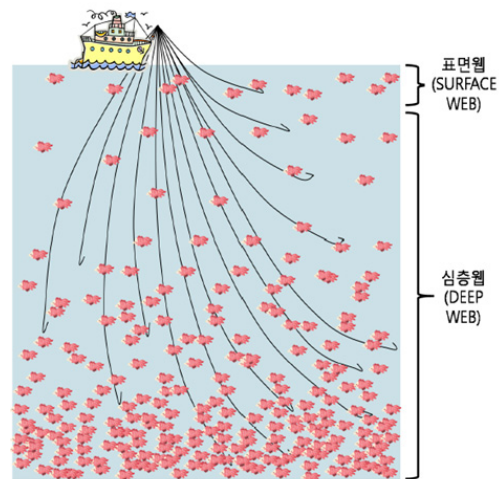
웹 공간에는 검색엔진을 통해서 자유롭게 접근할 수 있는 표면 웹(surface Web)과 일반적 검색에 의해 색인되지 않는 심층 웹(deep Web)이 함께 존재하는데 웹 아카이빙은 웹의 형태에 따라 크게 표면 웹 아카이빙과 심층 웹 아카이빙으로 나눌 수 있다(서혜란, 2004). 표층 웹

은 웹 브라우저로 표현되는 표면적인 웹을 의미하고, 심층 웹은 웹 브라우저로 표현되지 않지만 내부적인 동작과 연결을 담당하거나 가공되지 않는 데이터를 말한다(손영진, 2014). 이규희(2009)는 심층 웹을 웹 크롤러를 통해 수집되지 않는 양질의 정보를 지닌 웹 사이트 또는 웹 자원으로, 시각에 따라 작게는 관계형 데이터베이스 기반의 웹 데이터베이스로부터 크게는 크롤링이 불가능한 모든 부분까지로 보기도 한다. 영국도서관에서는 심층 웹을 보이지 않는 웹(invisible Web)이라고 부르며 온라인 데이터베이스와 같이 큰 저장소에 있는 콘텐츠로 웹 크롤러들이 접근할 수 없는 것으로 기존 검색 엔진의 제약으로 사용되지 않고 있다.

심층 웹은 크게 두 가지 측면으로 다뤄지고 있다. 이는 웹 크롤러를 통해 크롤링이 불가능한 영역과 기술적으로 크롤링이 되지 않는 기술적 측면을 일컫는다. 첫 번째 측면은 사적 웹과 소유 웹처럼 의도적으로 크롤러의 접근을 제한한 웹 페이지의 경우이다. 이 경우 기술적으로 크롤링이 가능하지만 암호화, 로봇 접근 금지 등 웹 마스터의 의도적 방해로 인해 발생한다. 두 번째로 기술적 측면은 기술적으로 크롤러가 웹 페이지에 접근해 수집하는 것이 어렵게 프로그래밍된 웹 페이지로 앞서 언급한 관계형 데이터베이스 기반의 웹을 포함해 고정된 URL이 아닌 동적 URL로 데이터베이스에서 일시적으로 생성되는 등의 동적 페이지에 의한 것이다.

그러나 <그림 1>에서 볼 수 있는 것과 같이 웹 환경의 일부분만 표면 웹으로 구현되고 있으며 그 외는 존재하고 있지만 웹 크롤러로 접근하기 어려운 심층 웹으로 구현됨을 알 수 있다.

심층 웹으로 구현된 웹 환경에 존재하는 웹 자원은 일반 이용자가 접근하기 어려우며 비가시적이라고 할 수 있다(이규희, 2009). 일반적인 검색 엔진으로 접근 가능한 표면 웹은 현재 15억 개의 웹 페이지 정도인 반면, 중요한 내용을 담고 있지만 접근이 불가능한 심층 웹은 무려 5500억 개에 달한다는 통계도 있다. 물론 집계에 따라 데이터가 바뀌기는 하지만 심층 웹의 규모가 표면 웹의 400배에 달한다는 통계는 웹 아카이빙에 있어 심층 웹의 수집이 얼마나 중요한지를 말해주고 있다(차승준, 최윤정, 이규철, 2009).



<그림 1> 표면 웹과 심층 웹

출처: 차승준, 최윤정, 이규철(2009), p. 183

## 2.2 재난안전정보

### 2.2.1 재난의 개념 및 유형

재난은 시대와 상황에 따라 차이를 가지는 상대적인 개념으로 인식되고 있다. 이미옥(2015)은 재난에 대해 현대사회 이전과 이후로 나누

어 이전에는 태풍, 홍수, 지진과 같은 천재지변을 재난으로 보았으나 최근인 현대사회 이후에는 인위적 요인으로 발생하는 화재, 붕괴, 폭발과 같은 대형사고, 테러 및 전쟁 등의 사고까지도 포함하는 개념으로 본다고 말하고 있다. 더불어 재난과 유사한 개념으로 위기, 위협, 재해, 사건, 사고 등과 같은 용어가 혼용되어 사용된다고 보았다.

『재난 기본법』에 따르면 재난에 대해 국민이나 국가에 피해를 주거나 줄 수 있는 것으로 자연현상 등과 같은 자연적인 위험요인에 의해 발생하는 태풍, 홍수, 호우, 대설, 가뭄, 지진 등과 같은 자연재난과 사회적, 인적 위험요인으로 발생하게 되는 화재, 폭발, 교통사고 등의 사회재난으로 분류하고 있다. 사회재난의 범주에는 감염병, 가축 전염병 등으로 발생하는 피해 역시 포함된다. 더불어 재난관리총괄기관인 국민안전처는 재난을 자연재난, 사회재난 그리고 생활안전으로 세분화하여 분류하고 있다. 생활안전에는 수난사고, 산악사고, 전기사고, 가스사고 등과 같은 실생활과 밀접한 관련성이 있는 재난이 포함된다.

이상으로 살펴본 재난의 개념 및 유형을 토대로 재난은 대규모의 피해를 가져오며 단일적인 하나의 사건이 아닌 다방면에서 발생하는 국민적·국가적 손실을 가져오는 사건으로 정의할 수 있다. 최근에는 법령과 국민안전처에서 분류하여 정의하고 있는 재난 유형인 자연재난과 사회재난이 복합적으로 발생하며, 이상 기후 현상으로 인한 새로운 자연재난의 발생, 그리고 이전에는 경험하지 못한 새로운 형태의 재난 발생이 증가하고 있다. 즉, 재난이 가지고 있는 불확실성(uncertainty), 상호

작용성(interaction), 복잡성(complexity)와 같은 특성이 두드러지고 있는 추세이다(차세영, 임도빈, 2014).

### 2.2.2 재난안전정보의 개념 및 특성

한국기록관리학회(2013)에서는 기록관리에 관한 국제 표준 'ISO 15489'를 통하여 기록(records)에 대해서 법적 의무와 업무의 증거나 정보로 생산, 접수, 유지되는 정보라 정의하고 있다. 재난 관련 정보 및 기록물 역시 재난의 증거이자 추후 발생할 수 있는 유사 재난에 대한 후대의 가치 있는 연구 자료로 활용될 수 있는 중요한 정보이기 때문에 기록관리적 측면에서 중요하게 다루어야 한다. 또한 국민과 국가의 안전과 밀접한 연관성이 있는 정보로 재난 정보는 일반 행정 업무에서 생산되는 정보와 비교했을 때 보다 중요한 가치를 지니기 때문에 재난 정보의 체계적인 관리가 요구된다.

『재난 기본법』은 재난관리단계에 대하여 재난의 예방·대비·대응 및 복구 단계로 구분하여 정의하고 있다. 이에 따라 법령에서 명시되어져 있는 재난관리책임기관은 재난 상황에 대하여 각각의 단계에 맞게 재난관리업무를 수행하여야 한다. 재난관리업무를 수행하는 과정에서 발생하게 되는 재난 관련 정보에 대하여 동법에 따르면 재난관리에 필요한 재난상황정보, 시설물정보, 지리정보 등과 같은 “재난관리정보”와 재난 및 사고에 관한 통계, 지리정보, 안전정책, 재난원인조사 결과 등과 같은 “안전정보”에 대하여 명시하고 있다. 뿐만 아니라 재난관리책임기관은 위기대응표준매뉴얼, 위기대응실무매뉴얼, 현장조치매뉴얼과 같은 재난 상황에 적합한 매뉴얼을 생산해야 할 의무를 가진다.



이러한 법령을 토대로 매뉴얼을 비롯한 재난 관련 정보가 기록화 되고 있으며 재난관리책임 기관을 비롯한 재난 업무를 수행하는 다양한 기관에 의하여 계속적으로 생산되어짐을 알 수 있다. 또한 재난관리단계에 따라 정보의 생산이 다르게 나타남을 예상할 수 있다. 재난 업무의 효율적인 운영을 위해서는 이러한 재난 정보가 체계적이고 통합적으로 관리되고 공유되어야 한다. 뿐만 아니라 추후에 발생 가능성이 농후한 재난의 예방 및 신속한 대응을 통하여 피해를 최소화하기 위해서는 재난 정보의 관리·공유는 매우 중요하다. 재난 정보의 체계적인 수집 및 활용을 위하여 동법 제70조(재난상황의 기록관리)에서는 재난관리책임기관의 장이 재난 정보를 기록하고 보관해야 하며 기록사항에 대하여 시장·군수·구청장에게 통보해야 함을 의무화하고 있다는 점에서 재난 정보 및 기록물의 기록관리의 필요성에 대해 살펴 볼 수 있다. 특히 최근 개정된 「재난기본법」에 따르면 재해연보 또는 재난연감뿐만 아니라 시설·재산 등의 피해 상황과 같은 재난 상황, 재난에 관한 조치 결과, 재난원인조사 결과 등과 같은 보다 포괄적이고 현장성을 가진 재난 정보의 기록화 및 보관이 강조되고 있는 추세이다.

재난기록에 대해 정혜지(2016)는 기록의 기억성에 초점을 맞춰 재난의 원인, 과정, 결과에서 생산된 현대의 기억으로 좁게는 사건의 진상규명 활동, 적절한 배·보상 및 처벌을, 넓게는 교육적·학문적 활용가치를 지니고 있는 도구로 보았다. 이미옥(2015)은 사고의 대응과 복구 과정에서 발생하는 모든 행정기록과 재난을 경험한 개인 및 단체의 활동과 기억을 표상하는 기록, 역사적·사회적·지역적으로 보존 가

치가 있는 기록 등의 민간기록까지 포함해 정의하고 있다. 이러한 재난기록의 정의는 업무상 발생하는 행정기록뿐만 아니라 언론 및 민간기록까지 포괄적으로 살펴보고 있다. 그 외 재난기록물에 대해 다루고 있는 연구가 대부분 민간기록 및 추모기록과 같은 매뉴스크립트적인 성격을 띠고 있다.

앞선 연구에서 다루고 있는 재난기록물은 재난 종료 후 기록물을 보존하고자 하기 때문에 본 연구에서 다루고 있는 재난의 상황에서 발생하는 정보의 효율적인 공유를 위해 기록물을 수집하고자 하는 목적과는 차이가 존재한다. 하지만 재난의 예방, 대응 그리고 기록으로써 역사적·증거적 가치를 지닌 재난기록이 가지는 특성과 의미를 찾는 관점은 동일하다.

재난은 다양성, 복합성, 불가측성과 같은 특성(정영철, 최익수, 배용근, 2016)을 가지고 있어 생산되는 정보 역시 재난의 특성을 담고 있다. 더불어 재난 정보가 웹 환경 속에서 생산, 관리되어진다는 점에 주목한다면 웹 자원적인 성격도 가진다고 볼 수 있다. 따라서 앞서 살펴본 재난기록의 정의와 의미, 재난의 특성, 웹 자원의 특성 등을 종합해 다양한 형태로 생산, 관리가 이루어지고 있는 재난 정보에 대해 본 연구에서는 재난안전정보라 명명하고자 한다. 재난안전정보에는 공공기록물 관리법에서 명시한 기록물 유형과 더불어 재난 상황에서 발생하는 전자기록물로 웹 사이트의 기록물 및 현황, 통계 데이터 등 모두를 포함하는 포괄적 개념으로 정의하고자 한다.

즉, 재난안전정보는 「공공기록물 관리에 관한 법률(이하 공공기록물 관리법)」 제2조(정의)에서 명시하고 있는 전자기록물인 전자문서, 웹

기록물 및 행정정보 데이터세트 등의 “기록정보자료”와 함께 「재난 기본법」 제3조 제10항의 “재난관리정보”, 제66조 제9항의 “안전정보”를 모두 포함한 것을 말한다.

재난관리주관기관은 법에 따라 재난 관련 업무에 따른 매뉴얼 및 재난안전정보들을 포털 및 시스템을 통해 생산 및 관리하고 있다. 재난관리주관기관은 각 기관의 특성에 따라 담당하는 재난 및 사고의 유형이 다르게 나타난다.

〈표 1〉은 「재난기본법」과 같이 자연재난, 사회재난으로 재난관리주관기관을 분류한 것으로 자연재난과 사회재난을 포괄적으로 관리하는 기관은 공통으로 분류하였다. 각 기관이 담당하고 있는 재난 및 사고의 유형이 다르기 때문에 발생하는 재난에 따라 재난관리업무에 참여도도 다르게 나타나며 생산되는 정보의 유형

### 3. 현황조사

#### 3.1 재난안전정보 유형

현재 「재난기본법」에서는 재난관리주관기관을 재난 및 사고 유형별로 지정해놓고 있다. 각

〈표 1〉 재난관리주관기관 별 재난 및 사고의 유형

재난분류	재난관리주관기관	재난 및 사고의 유형
자연재난	산림청	산불, 산사태
	교육부	학교 및 학교 시설 사고
사회재난	미래창조과학부	우주전파재난, 정보통신사고, 위성항법장치(GPS) 전파혼신
	법무부	교정시설사고
	국방부	국방시설사고
	행정자치부	정부주요시설사고
	문화체육관광부	경기장, 공연장 사고
	농림축산식품부	가축질병, 저수지사고
	산업통산자원부	가스사고, 원유수급사고, 원자력안전사고, 전력사고, 전력생산용 댐사고
	보건복지부	감염병, 보건의료사고
	고용노동부	사업장 대규모 인적 사고
	국토교통부	교통사고, 댐사고, 붕괴사고
	금융위원회	금융 전산, 시설 사고
	원자력안전위원회	원자력안전사고, 인접국가 방사능 누출 사고
	문화재청	문화재 시설사고
	공통	외교부
환경부		환경오염, 조류 대발생, 황사식용수사고, 화학물질사고
해양수산부		조류 대발생, 조수, 해양 환경오염, 해양 선박사고
국민안전처		풍수해, 지진, 화산 등 자연재난, 수난사고, 화재, 위험물사고 등 타 재난관리주관기관에 속하지 않는 재난·사고

출처: 「재난기본법」 시행령 [별표 1의 3], 재정리함.

도 다르게 나타난다. 대부분의 기관은 재난 시 발생한 정보에 대해 기관 내에서 관리하거나 국민안전처로 보내지나 실제 정보가 어떻게 관리되는지 알기 어렵다. 또한 정보의 공유를 위해서는 기관 간의 협조를 요청하거나 실제 업무 과정에서 검색을 통하는 등의 번거로움이 존재한다.

이렇게 각 기관별, 부처별로 흩어져 있는 재난안전정보들을 국민안전처는 신속하고 정확한 재난상황판단과 현장대응력을 높이기 위해 통합 재난안전정보체계 구축사업을 시작하여 개별적으로 분산 운영되던 재난 관련 27개 시스템을 3개의 포털(재난관리업무-NDMS, 모바일, 국민재난안전포털)로 통합하여 단일화했다. 재난관리업무는 NDMS라는 국가재난관리 정보시스템으로서 예방·대비·대응·복구 등 재난관리 단계별 업무를 지원하며 23개 중앙부처 234개 지방자치단체 등 국가재난관리에 관련되는 모든 기관이 동시에 접속·활용할 수 있다. 하지만 NDMS는 접근 권한이 주어진 기관만 접근해 정보를 활용할 수 있다. 이는 재

난 상황에서 행정기관을 망라하고 여러 유관 기관, 유관단체와 협업이 필요한 재난 업무를 수행하는데 원활한 소통을 어렵게 만든다고 볼 수 있다.

국민재난안전포털은 각 기관별, 부서별로 흩어져 있는 재난안전정보를 통합·연계하여 다양한 재난안전정보를 종합적으로 제공함으로써 국민들이 쉽게 접근할 수 있도록 인터넷 포털형태로 구축한 시스템이다. 국민재난안전포털은 일반 국민들이 재난발생 시 신속한 대처를 할 수 있도록 <표 2>와 같이 실시간 상황정보 등 예방부터 복구까지 재난에 대한 각종 정보를 올려놓고 관리하고 있다.

### 3.2 재난안전정보 노출도

웹 기록물을 수집하기 위해서는 수집하고자 하는 웹 사이트에서 제공하고 있는 웹 기록물에 대한 사전조사가 필요하며 수집 목적에 웹 기록물을 선별하는 작업이 우선시되어야 한다.

<표 2> 국민재난안전포털에서 제공하는 정보 유형

분류	세부내용
재난예방대비	- 자연재난이나 사회재난 등 비상시 국민행동요령(텍스트, 동영상) - 비상연락망(자치단체, 소방서, 약국, 응급의료센터 등)
안전현황	- 재난사고발생 시 재난유형, 재난명, 피해발생지역에 대한 정보 - 국민안전관리 일일상황, 재난문자, 안전뉴스 - 지역별 위험지수, 안전사고일지, 예방요령 - 기상정보, 강우량과 수위, 재난시설과 물자 의료기관정보, 무더위쉼터, 급경사지, 도로교통상황, 가뭄 예·경보, 재해정보지도 - 년도, 유형별 재난피해현황, 이재민조회 - 재해로 인해 피해를 입은 도로, 교량, 하천/제방 철도 등의 국가기반 공공시설의 응급복구현황
국민참여와 통계	- 피해신고 - 월별 재난안전종합상황 분석 및 전망, 재해연보, 재난연감
재난심리상담	- 재난심리회복지원
풍수해보험	- 풍수해보험에 관련된 모든 정보

재난안전정보의 공유 및 활용이라는 수집 목적에 부합하는 웹 기록물의 선별을 위해 먼저 재난안전정보에 대하여 분석하였다. 「재난기본법」 제3조(정의)에서 명시하고 있는 재난의 유형인 '자연재난'과 '사회재난'으로 나눈 후, 법령에서 지정한 재난관리단계를 예방, 대비, 대응, 복구의 4단계를 예방 및 대비, 대응, 복구로 3단계로 분류하였다. 이러한 분류 기준을 토대로 국민재난안전포털에서 제공하고 있는 재난

안전정보의 현황을 파악하고자 하였다.

일반 이용자로 하여금 재난안전정보를 손쉽게 접근할 수 있고 이를 활용할 수 있는지 살펴보기 위해 「재난기본법」에 따라 재난관리단계에 따라 국민재난안전포털의 게시판을 분류하고 게시판에 속해 있는 재난안전정보가 담긴 게시물을 웹 포털에 검색해 도출한 결과를 분석하였다. 그 결과는 <표 3>과 같다.

재난상황 시 잦은 업데이트와 변경이 이루어

<표 3> 재난안전정보 웹 검색 및 노출 현황

재난 유형	분류		게시판	웹 검색 가능 여부	노출 페이지	노출 값
	재난관리 단계	하위 분류				
자연 재난	예방·대비	현황	재난사고발생	△	3 페이지	0
			기상정보	△	2 페이지	0.2
			강우량과 수위	○	1 페이지	1
			무더위습터	○	1 페이지	1
			급경사지	○	3 페이지	1
			가뭄예·경보	X	-	0
	대응	행동요령	자연재난행동요령	△	1/2/3/4/12 페이지	0.36
	복구	보험	보험금지급사례	△	1 페이지	0.4
			풍수해보험자료	X	-	-
			풍수해보험공지	X	-	-
사회 재난	예방·대비	현황	재난사고발생	X	-	-
			안전사고경보	△	1 페이지	0.3
			도로교통상황	X	-	0
	대응	행동요령	사회재난 행동요령	△	1/2/4 페이지	0.32
			방호행동요령	X	-	0
복구	-	-	-	-	-	
공통	예방·대비	예방요령	예방요령	X	-	0
		현황	지역별 위험지수	○	1 페이지	1
			재난 시설과 물자	○	1/2 페이지	1
	대응	보고	일일재난상황	X	-	0
			재난문자	X	-	0
	복구	정보제공	의료기관정보	○	4 페이지	1
			피해현황	△	1/3/8 페이지	0.8
복구현황			○	6 페이지	1	

지는 재난의 특성을 담고 있는 재난안전정보의 특수성에 따라 게시판에서 제공하고 있는 재난 안전정보는 최신 정보를 중심으로 게시판 첫 페이지에 등록되어 있는 게시물을 위주로 검색에 활용하였다. 웹 검색 가능 여부의 경우 모든 게시물이 웹 포털에서 검색될 경우 'O', 부분 검색의 경우 '△', 검색되지 않을 경우 'X'로 표시하였고, 노출 페이지는 포털의 상위 순위로 2페이지 이상의 경우 노출이 되지 않는다고 간주하였다. 이러한 기준으로 노출 값을 계산하였고 소수점 아래 2자리까지 반올림해 노출 값을 도출하였다. 그 결과 자연재난 유형에 속해 있는 게시판은 10개 중 3개의 게시판만이 포털에서 검색했을 때 완전히 노출이 되었고, 사회재난 유형에 속해 있는 게시판은 5개 중 2개의 게시판이, 공통에 속해 있는 게시판은 8개 중 4개의 게시판이 포털에서 검색했을 시 완전히 노출됨을 알 수 있었다. 따라서 재난안전정보에 일반 이용자의 접근성이 높지 않음을 확인하였다. 즉, 국민재난안전포털의 재난 정보는 검색엔진으로 검색이 불가능한 심층 웹 환경에서 존재한다고 볼 수 있다. 심층 웹 환경의 정보 자원은 포털 사이트에서 운영하고 있는 크롤러로 수집하는데 제약이 있다. 그렇기 때문에 재난안전정보를 획득하기 위해서는 국민재난안전포털에 이용자가 직접 접근해야만 한다. 이 경우 이용자의 정보탐색능력 수준에 따라 정보 획득의 여부에 차이가 있으며, 웹 환경이 지니는 잦은 변동성으로 인해 이용자가 원하는 정보가 이미 갱신과 삭제를 통해 정보 획득이 불가능할 수 있다. 그리고 국민재난안전포털에서 원하는 정보를 찾지 못할 경우 타 기관의 홈페이지 등에 접근해야 하는 번거로움이 존재한다.

## 4. 재난안전정보 자동수집방안

### 4.1 고려사항

효율적인 재난 업무가 이루어지기 위해서는 재난안전정보의 체계적인 수집 및 관리 그리고 공유가 필수적이다. 재난안전정보는 재난이 가지는 불확실성, 상호작용성, 복잡성 등과 같은 고유의 특성이 담겨져 있다. 본 연구에서 다루고 있는 재난안전정보는 국민재난안전포털이라는 웹 상에서 제공되고 있기 때문에 웹 기록물의 휘발성, 다양성, 증식성 등과 같은 특성을 지니고 있다(성시에, 2009). 따라서 웹 상에서 제공되고 있는 재난안전정보가 일반적인 행정업무 상에서 생산되는 기록물이나 특정 주제에 따라 생산되는 메뉴스크립트 기록물과 다르기 때문에 웹 기록물 고유의 특징과 속성을 고려한 재난 안전정보의 자동수집방안을 마련해야 한다.

재난안전정보의 자동수집방안 마련을 위한 고려사항으로 첫째, 웹 기록물의 형태를 가지고 있는 재난안전정보의 자료 유형 및 특성을 파악하여야 한다. 웹은 일반적으로 웹 공간에서 검색엔진을 통하여 접근이 자유로운 '표면 웹(surface web)'과 이와 달리 검색만으로는 접근이 어려운 '심층 웹(deep web)'으로 나눌 수 있다(김광영 외, 2011). 심층 웹의 웹 기록물은 일반적인 웹 크롤러로는 수집에 어려움이 있다. 특히, 본 연구의 대상인 국민재난안전포털은 심층 웹과 표면 웹이 복합적으로 구성되어 있으며 대부분의 게시판에서 재난안전정보는 반정형 형태나 비정형 형태로 제공되는 심층 웹 형태로 구성되어 있다. 따라서 일반적인 크롤러로는 심층 웹의 웹 기록물을 완벽하게 아카이빙하

기에는 어려움이 존재하기 때문에 국민재난안전포털의 재난안전정보를 아카이빙하기 위해서는 국민재난안전포털의 웹 사이트의 자료 특성을 고려한 자동수집방안이 필요하다. 둘째, 재난안전정보의 수집 기준 및 범위에 대한 체계적인 수집정책이 마련되어야 한다. 재난안전정보를 수집하기 위해서는 국민재난안전포털에서 제공하고 있는 웹 기록물에 대한 사전조사를 바탕으로 수집정책이 세워져야 한다. 재난 상황에 대한 효율적인 업무 및 효과적인 재난관리라는 목적에 부합하는 기록물을 선별하고 그에 맞는 수집 범위를 설정해야 한다. 재난과의 관련성에 따라 수집 대상의 중요도에 따라 우선적으로 수집해야 할 재난안전정보에 대한 판단이 요구된다. 체계적인 수집정책에 따라 자동수집 시 반영하여야 한다. 셋째, 웹 크롤러가 웹 서버에 접근하여 재난안전정보를 수집할 때 적절한 주기를 설정해야 한다. 웹 크롤러는 관심 주소 리스트가 등록되면 반복적으로 같은 링크에 접근하여 사용자가 원하는 정보를 수집하기 때문에 적절한 주기를 설정하여 데이터베이스 시스템의 불필요한 자원 낭비를 막을 수 있다. 탐색 주기는 검색 시스템의 도메인의 특성을 반영하여 하루에 한번 또는 2-3일에 한 번씩 탐색하도록 설정해야 한다(장진욱, 2014). 넷째, 웹 기록물인 재난안전정보의 중복 수집을 막기 위한 중복 필터링 작업이 필요하다. 웹 크롤러는 웹 기록물의 URL을 기준으로 웹 기록물을 수집한다. 따라서 한 웹 사이트에 대하여 주기적으로 수집할 경우 동일한 내용으로 변경되지 않은 링크 또는 링크가 같으나 내용이 다른 경우, 수집 목적에 부합하지 않는 의미 없는 링크가 반복적으로 수집될 수 있다(이연수, 남성운, 윤대현,

2013). 이러한 문제점을 해결하기 위하여 자동으로 수집된 재난안전정보의 중복 유형을 파악하고 유형에 맞는 중복 필터링에 대하여 고려하여야 한다. 마지막으로 웹 기록물의 특성인 휘발성에 대한 고려사항으로 홈페이지 및 자료 등의 수정, 갱신, 삭제 여부를 고려해야 한다. 웹 기록물은 온라인 환경에서 생산되기 때문에 웹 사이트의 특성을 담고 있다. 즉 웹 사이트가 폐쇄되거나 자료의 수정, 갱신, 삭제가 발생했을 경우 웹 기록물의 수집 및 보존, 활용에 어려움이 발생한다(이연수, 남성운, 박성배, 2010). 수집 데이터베이스 시스템의 자원 낭비를 방지하고 갱신된 웹 기록물을 판단하기 위하여 일반적으로 "HTML header" 파일을 읽어 수정 날짜와 시간을 비교해 변화여부를 판단하거나 소스를 통하여 파일의 크기 변화를 통하여 판단할 수 있다. 하지만 파일의 내용 변화에 대하여 파악하지 못하는 문제가 발생한다(장진욱, 2014). 따라서 국민재난안전포털의 웹 페이지 및 웹 기록물의 분석을 통하여 적합한 갱신 판단 방법이 필요하다.

그 외 웹 크롤러가 수집 작업을 진행함에 있어 네트워크 과부하를 방지하기 위하여 사용자 트래픽의 고려, 프록시 또는 방화벽 등의 문제에 대한 해결 방안, 정보 제공 및 검색 방법을 위한 색인 및 문서요약과 같은 사항도 효과적이고 체계적인 자동수집방안 마련을 위하여 고려하여야 한다.

#### 4.2 자동수집절차

웹 크롤러는 시작 URL로 지정된 인터넷 상에서의 웹 서버에 접근하여 HTML 문서를 읽

어와 참조되지 않은 하이퍼링크(HyperLink)를 자동으로 추적하여 원하는 정보를 수집하고 데이터베이스에 수집된 정보를 저장하는 것을 목적으로 한다(장문수, 정준영, 2007).

일반적인 웹 크롤러는 먼저 다운로드할 URL 목록을 저장한 후 HTTP 프로토콜을 사용하여 문서를 다운로드한다. 다음으로 HTML 문서에서 링크를 추출하고 이전에 추출한 링크와의 중복되는 링크가 있는지 확인하는 과정으로 진행되며 <그림 2>와 같이 이 과정이 계속적으로 반복된다.

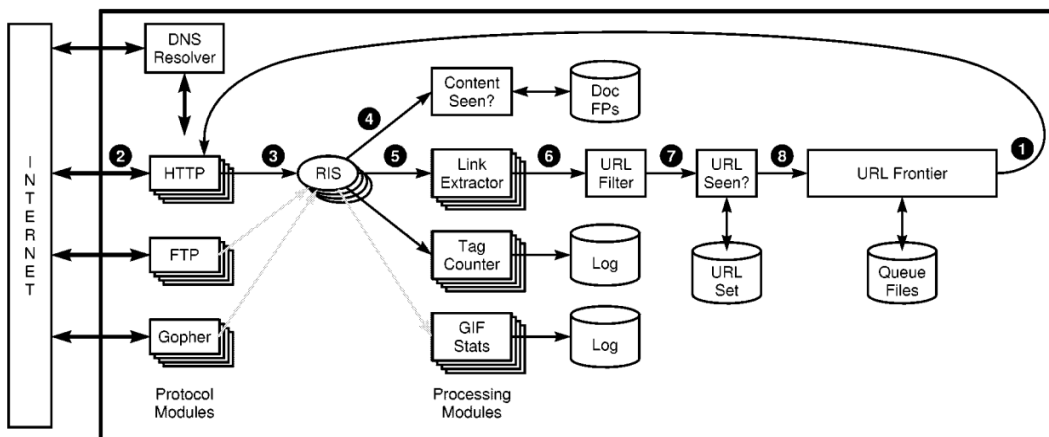
그러나 심층 웹 형태로 구성되어 있는 국민재난안전포털의 재난안전정보를 자동수집하기 위해서는 일반적인 크롤러로는 수집이 불가능하기 때문에 앞선 장에서 언급한 고려사항을 반영하여 자동수집절차를 마련해야 한다. 심층 웹은 일반적으로 게시판 형태를 가지고 있다. <그림 3>은 국민재난안전포털의 게시판 형식을 보여주고 있다. <그림 3> 상단에 표시되어 있는 네모는 게시글의 리스트이며 하단의 네모는 게시판

의 페이지 수를 나타낸다. 즉, 해당 페이지는 검색을 위한 값을 입력하고 검색 버튼을 클릭해야 하는 과정을 통해 결과를 보여주는 심층 웹이므로 직접적인 링크를 참조하는 일반적인 크롤러는 해당 게시판을 수집할 수 없다. 따라서 위와 같은 유형의 게시판을 수집하기 위해서 <그림 4>와 같이 페이지를 분석하였다.

<그림 4>의 실제 소스 코드 중 다음의 게시물 리스트의 소스 코드를 통하여 일반적인 크롤러로 수집할 수 없는 심층 웹 형태임을 확인할 수 있다.

<그림 5>에서 “(시스템 재훈련) 재해구호물자 지급훈련(2017.1.17.)” 게시물에 접근하기 위해서는 javascript:apiView("2017","1002","1")을 실행해야 하는데 해당 명령어는 단순히 주소창에 입력한다고 해서 게시물로 이동하거나 해당 명령어를 실행할 수 없다. 또한 게시판에서 페이지를 이동시키는 링크도 역시 같은 문제를 가지고 있다.

<그림 6>은 페이지 이동 버튼의 소스 코드로



<그림 2> 일반적인 크롤러 방식

출처: Heydon & Najork (1999). Mercator: A scalable, extensible Web crawler. p. 2. 그림 1 인용

대설	(혼련)대설	20170125 ~ 20170125	경기도
대설	(시스템 재훈련) 재해구호물자 지급훈련(2017.1...	20170117 ~ 20170117	서울특별시,부산광역시,인천광역시,경기도, 강원도,충청북도,충청남도,전라북도,전라남도, 경상북도,경상남도
대설	(시스템훈련) 재해구호물자 지급 훈련	20161227 ~ 20161227	서울특별시,부산광역시,대구광역시,인천광역시, 광주광역시,대전광역시,울산광역시,경기도, 강원도,충청북도,충청남도,전라북도,전라남도, 경상북도,경상남도,제주특별자치도
대설	12.26-대설	20161226 ~ 20161227	피해자료 없음
호우	12.26 호우~	20161226 ~ 20161227	부산광역시
풍랑,강풍	12.21-풍랑	20161221 ~ 20161224	부산광역시,울산광역시,강원도
강풍	12.21 ~ (강풍)	20161221 ~ 20161224	부산광역시,충청남도
호우	12.21 ~ (호우)	20161221 ~ 20161222	부산광역시,경상남도
대설	(혼련)12.21대설풍랑	20161221 ~ 20161221	서울특별시,부산광역시,인천광역시,광주광역시, 경기도,전라북도,경상북도,경상남도
태풍	(혼련)전시중합상항실 NDMS 재난상황 조지훈련	20161213 ~ 20161213	부산광역시

<그림 3> 국민재난안전포털 자료 유형

The image shows a screenshot of the National Disaster Safety Portal (www.safekorea.kr) search results page. The page displays a list of disaster events with columns for search period, disaster type, and occurrence period. One event is highlighted: '태풍 (혼련)전시중합상항실 NDMS 재난상황 조지훈련' occurring on 20161213. Below the search results, the browser's developer console is open, showing the JavaScript code for the search results list. The code uses jQuery to iterate over search results and display them in a table format. The console also shows a warning message: 'Synchronous XMLHttpRequest on the main thread is deprecated because of its detrimental effects to the end user's experience. For more help, check https://xhr.spec.whatwg.org/'.

<그림 4> 국민재난안전포털 소스코드

`<a id="api_bt_r_1_dstr_nm" class="w2anchor2 ellipsis winter_link" href='javascript:apiView("2017","1002","1");'>(시스템 재훈련) 재해구호물자 지급훈련(2017.1.17.)</a>`

<그림 5> 게시물 리스트 소스 코드



```
<a id="wq_uuid_145" class="w2anchor2_go_btn" href="javascript:void(null);">이동</a>
```

〈그림 6〉 페이지 이동 버튼 소스 코드

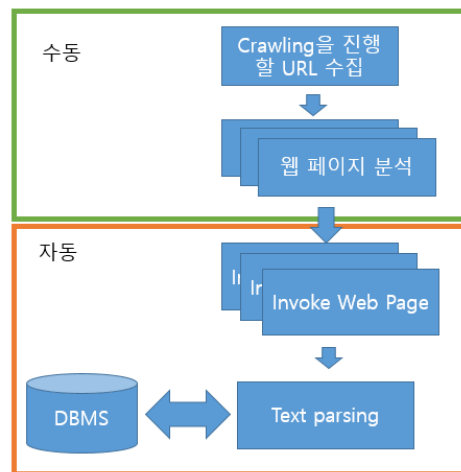
해당 버튼을 클릭하는 경우 javascript:void(null) 링크를 클릭하도록 되어 있다. 이러한 명령어 역시 웹 브라우저 주소창에서 실행이 불가능하다. 이러한 이유 때문에 게시판의 수집하고자 하는 페이지로 이동하거나, 수집하고자 하는 게시물을 불러오는 것이 일반적인 크롤러로는 어려움이 따른다. 따라서 직접 데이터베이스에 접근하거나 또는 마우스로 해당 링크를 클릭하여 웹 페이지에서 명령어를 실행하는 직접 이관과 같은 아카이빙 방법으로 정보를 수집할 수 있다. 하지만 직접 데이터베이스에 접근하는 방법은 사이트의 보안 문제, 저작권과 같은 문제가 발생한다. 또한 마우스로 각 링크를 클릭하여 데이터를 수집하는 것 역시 많은 시간이 소요되고 인력적인 문제와 같은 현실적인 어려움이 나타난다.

같으며 웹 기록물을 수집하기 위한 방법으로 크게 수동과 자동으로 자동수집절차를 나눴다. 우선적으로 수집목적에 부합하는 재난안전정보의 수집 기준을 선정하고 수집할 해당 URL을 선정하는 작업이 수동으로 이루어진다. 심층 웹은 실제 데이터가 데이터베이스에 저장되어 있기 때문에 수집하고자 하는 링크에 접근하는 것만으로는 수집이 불가능하며 소스 코드로 구성된 명령어나 함수를 실행시켜야 서버로부터 데이터를 수집할 수 있다. 따라서 웹 페이지 분석을 통하여 페이지 내의 소스 코드의 게시물 관련성과 페이지 이동 관련성을 분석해야 한다. 다음으로 선정된 수집 대상을 자동으로 수집하는 과정에서 앞서 언급한 것과 같이 심층 웹은 수집하고자 하는 대상의 링크를 직접적으로 참조하는 일반적인 크롤러로 수집이 불가능하다.

### 4.3 적용 알고리즘

수집하고자 하는 정보의 직접적인 링크를 따라가는 일반적인 크롤러를 통한 자동수집절차는 국민재난안전포털의 재난안전정보의 수집에 어려움이 존재한다. 따라서 이러한 문제를 해결하기 위해 본 연구에서는 심층 웹의 형태를 가지고 있으며 재난이라는 특수성을 가진 정보 자원을 생산하는 국민재난안전포털의 웹 특성을 고려하여 알고리즘으로 동작하는 크롤러를 구현하였다.

재난안전정보의 자동수집절차는 〈그림 7〉과



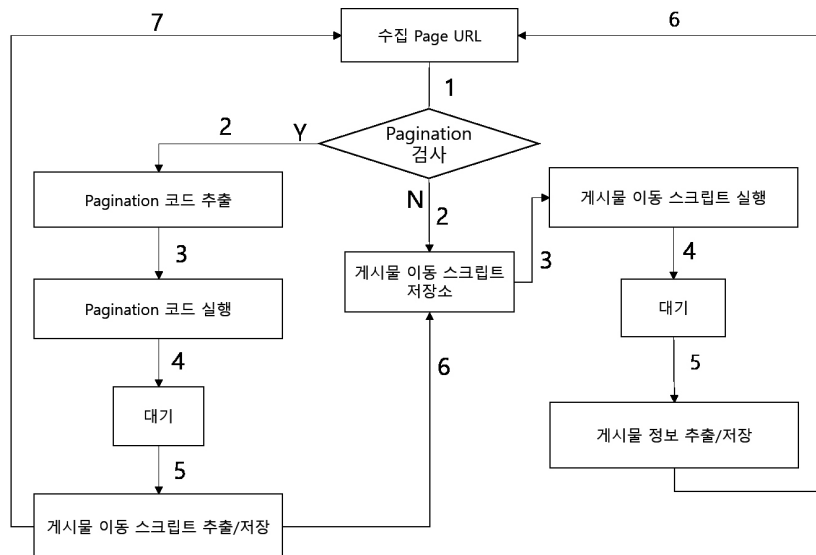
〈그림 7〉 재난안전정보 자동수집절차

따라서 크롤러를 C#으로 구현하였으며 소스 코드 실행을 위한 웹브라우저 객체는 Microsoft Web Browser 객체<sup>1)</sup>를 이용하였다. 크롤러를 통해 수집된 데이터는 데이터베이스에 저장되는데, 본 연구에서는 Linux 기반의 서버에 MySQL 5.7 버전을 설치하고 국민재난안전포털 DB를 구축한 후 Text 형식으로 수집된 데이터를 저장하였다. 전체적인 재난안전정보의 자동수집에 대한 알고리즘은 <그림 8>과 같다.

수동 수집 절차를 통하여 웹 페이지를 분석한 결과 페이지 이동 관련 소스코드와 게시물 관련 소스 코드를 확인하였다면, 먼저 수집하려는 페이지로 이동한 다음 Pagination 검사를 수행한다. Pagination 검사는 해당 게시판이 하나 이상의 페이지로 구성 되어 있는 경우 페이지를 이동하기 위해 수행한다. Pagination 검사의 세

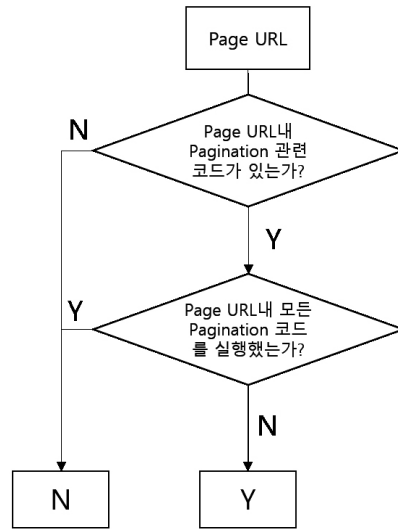
부 동작은 <그림 9>와 같다.

<그림 9>와 같이 수집하려는 페이지 내에 페이지 이동 관련 소스 코드가 있는지 검사하고 만약 있으면 해당 페이지를 방문 했는지 검사해야 한다. 그래서 만약 방문하지 않은 페이지가 있다면 해당 페이지로 이동 하도록 Y 또는 N을 리턴하는 과정을 반복한다. <그림 9>의 Pagination 검사를 통해 이동하려는 페이지가 있는 경우 Y로 이동하여 Pagination 검사를 수행한다. Pagination 검사 과정은 Pagination 소스 코드를 수집하는 과정이다. 그 후 추출된 소스 코드를 실행하면, 실행된 소스 코드가 동작하는 시간동안 대기한다. 다음으로 페이지 로드가 완료되면 페이지 내 게시물 링크를 저장하고 다시 수집 URL로 이동한다. Pagination 검사를 통해 더 이상 이동하려는 페이지가 없는



<그림 8> 자동수집 적용 알고리즘

1) [https://msdn.microsoft.com/en-us/library/aa752040\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/aa752040(v=vs.85).aspx)



〈그림 9〉 Pagination 검사 세부 동작

경우는 N으로 이동하여 게시물 이동 소스 코드 저장소로 이동한다. 게시물 이동 소스 코드 저장소에 저장된 게시물 이동 소스 코드를 가져와서 게시물 이동 소스 코드를 실행한다. 그 다음으로 소스 코드가 실행되어 웹 페이지가 로드되는 동안 대기하였다가 로드된 정보를 추출하여 저장한다. 마지막으로 페이지 로드가 끝나면 수집 Page URL로 이동한다.

#### 4.4 구현 및 결과

앞선 장에서의 수집 절차를 수행하기 위하여 모델링 알고리즘을 작성하였으며 이를 통하여 크롤러는 모든 페이지를 방문할 때까지 반복적으로 실행된다. 먼저 크롤링을 하기 위한 URL을 불러오고, 페이지가 로드될 때까지 대기한다. 다음으로 로드된 페이지의 소스 코드를 실행한다. 소스 코드가 실행되는 일정 동안 대기하고 소스 코드가 실행되었다면 실행된 웹 페이지를

임시 저장한다. 저장된 페이지를 분석하고 분석된 데이터를 저장한다. 그 후 입력된 웹 문서 형태별로 저장하고 특정 형태의 문서인 경우 즉, 수집하고자 하는 속성의 정보를 읽어와 데이터베이스에 저장하는 이와 같은 과정으로 크롤링이 이루어진다.

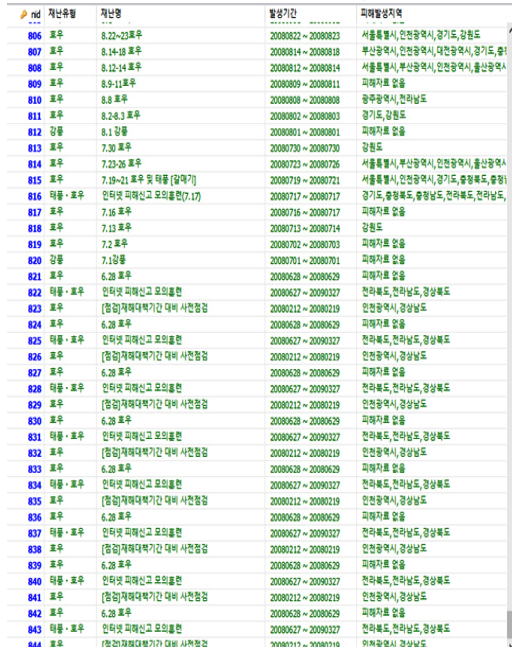
국민재난안전포털에서 제공하고 있는 재난안전정보를 자동수집하기 위해 심층 웹으로 구성되어 있는 웹 페이지의 특성을 고려하였으며 이상의 과정을 통하여 자동수집절차를 마련하였다. 그리고 실제로 마련한 자동수집절차를 통하여 게시판에 적용하여 재난안전정보의 자동 수집을 실시하였다. 실행된 결과는 〈그림 10〉, 〈그림 11〉과 같다.

〈그림 10〉은 수집하고자 하는 게시판에서 크롤러가 웹 페이지를 로드한 것을 보여준 것으로 크롤러 시작 버튼으로 크롤러를 실행시키게 되면 프로그램이 지정한 소스 코드가 실행되게 된다. 다음으로 소스 코드 실행에 따라 페



〈그림 10〉 크롤러 실행 중인 화면

이지의 이동이 이루어지게 된다. 웹 페이지가 이동하면서 〈그림 10〉의 하단과 같이 웹 페이지 내에서 텍스트를 읽어 오며, 읽어 온 텍스트 중 의미 있는 정보를 추출하고 추출한 정보를 〈그림 11〉과 같이 사전에 구축한 국민재난안전포털 DB에 저장하게 된다. 이는 대부분의 게시판이 반정형 형태 또는 비정형 형태인 심층 웹으로 이루어져 일반적인 웹 크롤링이 불가능한 국민재난안전포털의 게시판에서 제공하고 있는 재난안전정보를 수집함으로써 기존의 웹 크롤러가 수행하지 못한 심층 웹의 수집이 가능함을 보여주고 있다. 자동수집된 재난안전정보는 국민재난안전포털 DB에 국민재난안전포털에서 재난안전정보를 분류한 형식을 토대로 Text 형식으로 저장되기 때문에 추후 카테고리 분류와 같은 작업을 통해 재난 유형, 발생기간, 피해발



〈그림 11〉 자동 수집 결과

생지역 등으로 검색해 이용이 가능하다. 재난안전법에 따르면 재난관리업무의 예방, 대비, 대응, 복구 4단계로 나눠 각 단계별로 생산되어야 하는 재난안전정보가 존재한다. 따라서 본 연구에서 각 단계와 관련된 기록물을 국민재난안전포털 DB에 통합적으로 관리하고 있기 때문에 이용자의 요구에 맞게 필요로 하는 각 단계의 기록물을 이용할 수 있을 것이다.

일반적인 웹 아카이빙 전략에서 가장 광범위하게 사용되고 있는 원격하베스팅의 다음과 같은 한계점을 가지고 있다. 첫째, 플래시, 소스 코드 등과 같이 직접적인 링크 참조로 크롤링이 불가능한 기술 영역이 존재한다는 점이다. 특히 심층 웹은 실제 데이터를 수집하기 위하여 명령어를 실행시키는 등과 같은 논리적 단계를 거쳐야 한다는 기술적 어려움이 존재한다. 둘째, 심

층 웹을 비롯한 웹 사이트의 완벽한 아카이빙이 불가능하다는 것이다(이규희, 2009). 셋째, 심층 웹이 대부분인 웹 자원 특성에 따라 자동수집만으로 수집이 이루어지지 않는 경우가 발생하기도 한다. 스크립트가 브라우저의 호환성으로 인해 실행이 안 되는 경우와 프로그램의 규칙성이 없어서 페이지와 리스트의 구분이 명확하지 않은 경우로 일반적인 크롤러를 통한 자동수집으로 수집이 불가능한 경우가 발생한다. 따라서 심층 웹의 논리적 단계를 거쳐 크롤링을 하기 위하여 수동으로 웹 페이지의 분석을 통하여 명령어의 역할 및 문서 연결 관계를 파악해야 한다. 이를 위해 소스 코드 타입을 2개로 나누어 게시물 관련 소스 코드, 페이지 이동 관련 소스 코드를 확인한 다음에 크롤러를 통하여 자동으로 정보 자원을 수집할 수 있도록 알고리즘을 구현하였다. 즉, 일반적인 웹 크롤러로 수집이 불가능한 심층 웹의 자동 수집이 가능하도록 수동과 자동 작업이 병행되는 반자동 크롤러 알고리즘을 개발 및 구현하였다.

#### 4.5 기대효과

이러한 결과를 토대로 실제 국민재난안전포털에 적용한다면 다음과 같은 기대효과가 예상된다. 첫째, 심층 웹으로 검색 엔진에서 검색이 원활하게 이루어지지 않았던 국민재난안전포털에서 제공하고 있는 재난안전정보를 검색했을 시 일반 이용자가 재난안전정보에 대한 접근성을 향상시킬 수 있다. 이는 재난의 예방, 대비, 대응, 복구와 관련해 일반 이용자가 원하는 정보를 공유하고 활용할 수 있게 함으로써 재난정보의 대국민 서비스라는 국민재난안전포털의

역할을 수행하는데 도움을 줄 수 있다. 둘째, 자동수집을 통하여 웹 포털에 노출시켜 재난안전정보에 대한 이용자들의 접근성을 향상시킬 수 있기 때문에 재난 상황이 발생했을 때, 적재적소에서 재난안전정보를 공유하고 활용하여 신속한 대응 및 효율적인 재난 업무를 수행할 수 있게 한다. 이는 현재 재난발생 시에 재난정보의 공유가 제대로 이루어지지 않아 관련기관 및 국민들의 미흡한 대응으로 인하여 대규모 재난으로 이어지는 현실적인 문제 해결에 도움을 줄 수 있다. 셋째, 웹 아카이빙 측면에서 볼 때, 크롤러 구현을 위한 알고리즘이 작성되고 수집하고자 하는 대상이 증가할수록 웹 페이지 사이의 규칙성을 발견할 수 있기 때문에 추후에는 대상의 양에 구애받지 않고 주기적으로 크롤링이 가능하다. 따라서 현재 재난안전정보라는 한정적인 웹 정보 자원의 아카이빙에서 나아가 일반 웹 정보 자원으로 확대 가능성이 있으며 크롤러의 범용화가 가능할 것으로 예상된다.

## 5. 결 론

재난은 언제 어디에서든 발생하며 현대사회에 들어서 그 유형 및 피해규모가 점점 다양해지며 커져가고 있는 추세이다. 따라서 재난 상황에 대처하기 위한 종합적이고 체계적인 관리를 위하여 재난총괄기관이 마련되었다. 재난총괄기관은 재난에 관련된 정보를 통합적으로 관리하고 각 기관 및 국민들을 위한 공유 및 활용 서비스를 제공하고 있다. 또한 재난에 대한 예방, 대비, 대응, 복구 단계에 대한 정보 제공을 위하여 국민재난안전포털을 통하여 대국민 서

비스를 제공하고 있다.

그러나 국민재난안전포털에서 제공하고 있는 재난안전정보는 웹 검색엔진으로 검색했을 시 일반 이용자의 접근성이 떨어진다는 문제점을 가지고 있다. 따라서 본 연구에서는 이러한 문제점을 해결하기 위한 방안의 하나로 자동수집방안을 제시하였다. 이를 위하여 먼저, 국민재난안전포털의 각 게시판 별 특성에 대하여 살펴본 후 게시판에서 제공하고 있는 재난안전정보의 유형을 파악하였다. 대부분의 게시판이 일반적인 웹 크롤러로 수집이 어려운 심층 웹 형태로 제공되고 있었기 때문에 본 연구에서는 국민재난안전포털의 특성을 반영한 수집방안의 필요성에 대해 제시하였다.

심층 웹 형태로 구성되어 있는 국민재난안전포털의 특성을 반영한 수집방안을 위하여 먼저 수집절차 모델링을 수행하였으며, 모델링을 기반으로 한 수집 웹 크롤러를 활용하여 반정형 유형의 게시판을 선정하였다. 이 과정에서 반정형의 데이터를 정형으로 변환하는 과정을 추가하여 실제로 자동수집을 실시해 결과를 도출할 수 있었다.

이를 통하여 심층 웹 형태로 이루어진 국민재난안전포털의 재난안전정보를 자동수집할 수 있음을 확인할 수 있었고, 일반 이용자로 하여

금 재난안전정보를 웹 포털의 검색을 통하여 공유 및 활용할 수 있는 기반을 마련했다는 것에 의의가 있다. 또한 심층 웹의 정보 자원을 수집을 위하여 새로운 알고리즘을 개발하고 크롤러를 구현하는데 기초 연구로써 의의가 있다. 특히 수집하고자 하는 웹 자원의 양이 증가함에 따라 규칙성을 발견하여 범용화의 가능성이 있다는 점에서 가치를 찾을 수 있다. 그러나 일반적인 아카이빙 전략에서 사용되고 있는 크롤러에 비하여 실행 속도가 느리다는 점, 대기 시간을 예상하기 어렵다는 점, 웹 페이지를 분석할 전문가가 필요하다는 한계를 지니고 있다. 또한 본 연구는 재난안전정보의 자동수집 알고리즘 개발에 초점을 두고 있기 때문에 수집된 재난안전정보를 효과적으로 이용하기 위해서는 이용을 위한 별도의 프로그램 제공이 요구된다. 이에 따라 심층 웹의 특성을 고려해 효과적인 탐색과 브라우징 검색이 가능한 인터페이스를 갖추고 있으며 재난안전정보를 활용하고자 하는 이용자의 요구를 반영한 컬렉션 개발에 관한 향후 연구가 필요하다. 그럼에도 불구하고 본 연구는 대부분의 웹 자원이 심층 웹으로 제공되고 있는 현실을 고려해 볼 때, 다양한 정보 자원을 자동수집하기 위한 기초 연구로 활용될 수 있을 것이라 기대한다.

## 참 고 문 헌

- 김광영, 이원구, 이민호, 윤화목, 신성호 (2011). 웹 자원 아카이빙을 위한 웹 크롤러 연구 개발. 한국콘텐츠학회논문지, 11(9), 9-16.
- 김용섭 (2016). 재난 및 안전 관리 법제의 현황과 법정정책 과제. 행정법연구, 45, 51-79.

- 김유승 (2008). 복합적 웹 아카이빙 정책에 관한 고찰: 프랑스국립도서관의 사례를 중심으로. 한국문헌정보학회지, 42(4), 159-179.
- 김정심 (2008). 행정부처 웹 사이트의 아카이빙 방안에 관한 연구. 석사학위논문. 중앙대학교 대학원, 기록물관리학과.
- 김희정 (2011). 웹 아카이빙 도구 비교분석 연구. 제 18회 한국정보관리학회 학술대회 논문집, 95-98.
- 류상일, 안혜원 (2007). 네트워크를 통한 효율적 재난대응체계 구축. 한국콘텐츠학회논문지, 7(2), 170-179.
- 박병식 (2008). 민간부문 재난관리 대응체제의 실태와 구축방안. 한국정책과학학회보, 12(4), 203-226.
- 서혜란 (2004). 웹 아카이빙의 성과와 과제. 한국비블리아, 15(1), 5-22.
- 성시에 (2009). 중앙행정기관의 웹 기록물 수집방안에 관한 연구. 석사학위논문. 중앙대학교 대학원, 기록관리학과.
- 손영진 (2014). 웹 표준을 이용한 공공기관 웹 아카이빙 개선방안 연구. 석사학위논문. 전남대학교 대학원, 기록관리협동과정.
- 이규희 (2009). 심층웹 아카이빙 방안 연구. 석사학위논문. 명지대학교 기록정보과학전문대학원, 기록관리학과.
- 이미옥 (2015). 대구지하철화재참사 기록의 단계별 관리 방안. 석사학위논문. 경북대학교 대학원, 기록학과.
- 이상명 (2014). 재난안전관리체계의 개선에 관한 법적 고찰. 한양법학, 25(4), 397-417.
- 이성숙 (2005). 웹 아카이빙 도구에 관한 연구. 한국정보관리학회 학술대회 논문집, 185-193.
- 이연수, 남성운, 박성배 (2010). 공공기관 웹 사이트 기록물 수집 사례분석. 한국멀티미디어 학회 학술발표논문집, 644-647.
- 이연수, 남성운, 윤대현 (2013). 대규모 웹 기록물의 원격수집을 위한 콘텐츠 중복 필터링 개선 연구. 기록학연구, 35, 133-160.
- 장문수, 정준영 (2007). URL 패턴 스크립트를 이용한 효율적인 웹문서 수집 방안. 퍼지 및 지능시스템 학회논문지, 17(6), 849-854.
- 장진욱 (2014). 웹 페이지 비교통합 기반의 정보 수집 시스템 설계 및 개발에 대한 연구. 한국IT서비스학회지, 13(1), 147-159.
- 정영철, 최익수, 배용근 (2016). 사회안전을 위한 빅데이터 활용의 재난대응 정책. 한국정보통신학회논문지, 20(4), 683-690.
- 정혜지 (2016). 미국 9·11 메로리얼의 재난기록물 수집에 관한 연구. 석사학위논문. 명지대학교 기록정보과학전문대학원, 기록관리학과.
- 차세영, 임도빈 (2014). 재난관리 단계별 대응 체계에 대한 연구 - 구미 불산 누출 사고 사례의 네트워크 방법론적 분석. 지방정부연구, 18(1), 727-754.

- 차승준, 이규철 (2008). 웹 기록물 아카이빙 기반기술 연구 개발. 한국전자거래학회 심포지움 및 기타간행물, 359-368.
- 차승준, 최윤정, 이규철 (2009). 공공기관 심층 웹 기록물 아카이빙을 위한 메타데이터설계. 한국전자거래학회, 14(4), 181-193.
- 한국기록관리학회 (2013). 기록관리론: 증거와 기억의 과학. 성남: 아세아문화사.
- 한국기록학회 (2008). 기록학 용어 사전. 서울: (주)역사비평사.
- 현석환 (2015). 한국의 재난관리 시스템의 문제점과 효율화 방안. 한국사회안전학회지, 10(1), 39-56.
- Heydon, Allan & Najork, Marc (1999). Mercator: A scalable, extensible Web crawler. World Wide Web, 2, 219-229.

[ 관련법령 ]

- 「공공기록물 관리에 관한 법률」 (2014.11.19.)
- 「재난 및 안전관리 기본법」 (2017.1.17.)

[ 웹사이트 ]

- 국가법령정보센터. 검색일자: 2017. 1. 17. <http://www.law.go.kr/>
- 국민안전처. 검색일자: 2016. 12. 15. <http://www.mps.go.kr/home/index.html>
- 국민재난안전포털. 검색일자: 2016. 12. 15. <http://www.safekorea.go.kr/idsiSFK/index.jsp>

• 국문 참고자료의 영어 표기

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

- 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.
- 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, Se-Yeong & Im, Tobin (2014). Study on Disaster Management System: Case Study Using Social Network Analysis on Hydrofluoric Acid Leakage Accident in Gumi, Korea. The Korean Journal of Local Government Studies, 18(1), 727-754.
- Chang, Moon-Soo & Jung, June-Young (2007). A Method of Efficient Web Crawling Using URL Pattern Scripts. Journal of Korean Institute of Intelligent Systems, 17(6), 849-854.
- Choung, Young-chul, Choy, Ik-su, & Bae, Youg-guen (2016). Social security aimed disaster response policy based on Big Data application. Journal of the Korea Institute of Information and



- Communication Engineering, 20(4), 683-690.
- 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.
- Hyun Suk-Hwan (2015). Problems of Korean disaster management system and methods for its efficiency. The Journal of Korean Association of Security and Safety, 10(1), 39-56.
- Jang, Jin-Wook (2014). A Study on Design and Development of Web Information Collection System Based Compare and Merge Method. Journal of information technology services, 13(1), 147-159.
- Kim, Hee-jung (2011). Comparative Analysis of Web Archiving Tools. Korea Society for Information Management Conference Proceeding, 95-98.
- Kim, Jung-Sim (2008). The Plan Research of Administration Ministries and Offices Website Archiving. Master's Thesis. Major of Records and Archives Management, Chung-Ang University. Seoul.
- Kim, Kwang-Young, Lee, Won-Goo, Lee, Min-Ho, Yoon, Hwa-Mook, & Shin, Sung-Ho (2011). Development of web crawler for archiving web resources. Journal of the Korea Contents Association, 11(9), 9-16.
- Kim, Yong-Sup (2016). Status of Disasters and Safety Management Legislation and its Legal and Policy Issues. Administrative Law Journal, 45, 51-79.
- Kim, You-Seung (2008). A Study of Combined Web Archiving Policy: BnF's Three Layers Web Archiving Strategy. Journal of the Korea Social for Library and Information Science, 42(4), 159-179.
- Korean Society of Archival and Records Management(ed) (2013). Theory of Records Management: Science of Evidence and Memory (3rd). Seongnam: AsiaMunHwaSa.
- Korean Society of Archival Studies (2008). Dictionary of Records and Archival Terminology. Seoul: Yuksabipyongsa.
- Lee, Kyu-He (2009). A Study on Deep web Archiving. Master's Thesis. Major of Records and Archival Information Management, Myongji University. Seoul.
- Lee, Me-Ok (2015). A Study on Phased Management Plan of 'Daegu Subway Disaster' Record. Master's Thesis. Department of Archival Science, Kyungpook National University. Daegu.
- Lee, Sang-Myeong (2014). Legal Considerations on Improvement of the Disaster Safety Management System: Focused on 'Laws Related to Safe Management of Vessels'. Hanyang Law Review, 25(4), 397-417.

- Lee, Sung-Sook (2005). A Study on Web Archiving Tools. Korea Society for Information Management Conference Proceeding, 185-193.
- Lee, Yeon-Soo, Nam, Sung-Un, & Park, Sung-Bae (2010). A study on the analysis of the web archiving case of public institutions. Korea Multimedia Society Conference Proceeding, 644-647.
- 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, Byoung Silk (2008). The Reality of Private Sector for Disaster Management and the Response System Construction: In the Focus of Busan Disaster Safe Network. Korean Policy Science Review, 12(4), 203-226.
- Ryu, Sang-Il & Ahn, Hye-Won (2007). Improving Disaster Response System Using Network: Focused on Korea and Japan's Disaster Cases. Journal of the Korea Contents Association, 7(2), 170-179.
- Seong, Si-Ae (2009). A research on how to collect web records of state administrative agencies. Master's Thesis. Major of Records and Archives Management, Chung-Ang University. Seoul.
- Son, Young Jin (2014). A Study on web archiving improvement using web standardization of public Institutions. Master's Thesis. Interdisciplinary Program of Archival Studies, Chonnam National University. Gwangju.
- Suh, Hye-Ran (2004). Web archiving: What we have done and what we should do. Journal of the Korean Biblia Society for Library and Information Science, 15(1), 5-22.