

이슈 생존 주기 모형 기반 재난 아카이브 수집 방안*

Acquisition Methods for Disaster Archives Based on the Issue Life Cycle Model

유호선 (Ho-Suon Yoo)**

오효정 (Hyo-Jung Oh)***

초 록

재난 웹 기록물의 가치와 보존의 중요성으로 인해 전 세계적으로 재난 아카이브의 구축은 국가적인 과제가 되어가고 있다. 본 연구는 재난 아카이브의 웹 기록물 수집에 있어서 이슈 생존 주기 모형에 기반한 수집 방안을 제시하는 것을 목적으로 한다. 이를 위해 먼저 국내외 재난 아카이브의 웹 기록물 수집 현황, 수집 방법 및 기간에 대하여 분석하였다. 또한, 최근 10년간 발생한 재난 이슈를 수집 및 분석하여 이슈 생존 주기 모형을 도출하였다. 이슈 생존 주기 모형은 재난의 특성에 따라 돌발형 및 주기형으로 구분되며, 각 모형에 따른 웹 기록물 수집 방안을 제시하였고 적용 가능성을 검증하였다.

ABSTRACT

Due to the value and the importance of preservation of disaster web records, to build disaster archives is globally becoming a national challenge. This study proposes a acquisition methods based on the issue life cycle model for collecting disaster web records. We firstly analyzed web records acquisition status, methods and period of domestic and foreign disaster archives. In addition, the issue life cycle model was derived by collecting and analyzing the disaster issues in the last 10 years. As the results of the analysis, the issue life cycle model was divided into the sudden type and periodic type according to the characteristics of the disaster. In conclusion, this study propose a method to collect web records according to each model and verify its applicability.

키워드: 재난 아카이브, 수집 방안, 이슈 생존 주기, 웹 기록물
disaster archives, acquisition methods, issue life cycle, web records

* 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-재난안전플랫폼기술개발사업의 지원을 받아 수행된 연구임(No. NRF-2016M3D7A1912703).

이 논문은 2018년 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016R1A2B1008000).

** 전북대학교 기록관리학과 석사과정(rhsblossom@gmail.com) (제1저자)

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

■ 논문접수일자: 2018년 5월 20일 ■ 최초심사일자: 2018년 6월 8일 ■ 게재확정일자: 2018년 6월 8일
■ 정보관리학회지, 35(2), 115-139, 2018. [http://dx.doi.org/10.3743/KOSIM.2018.35.2.115]

1. 서론

1.1 연구배경 및 목적

도시의 인구 밀도 증가로 인한 생활공간의 고층화와 극단적인 양상을 보이는 기상조건으로 인해 재해 및 재난은 현대 인류가 마주한 대형 위험요인으로 자리 잡았다. 하지만 국내의 재난 관리는 사건의 물리적 복구에만 초점이 맞춰져 있고, 복구 이후 관련 이슈는 신속히 사그라지곤 한다. 전 세계적으로 많은 재난 사건이 발생했지만, 우리나라와 다른 점은 참사의 기억을 외면하기보다 오히려 수면 위로 끌어올리고자 노력한다는 점이다. 그리고 이들의 노력은 재난 기록의 꾸준한 수집 및 관리와 활용을 통해서 잘 드러난다. 재난 기록의 의미와 가치를 인지하고 재난 사건을 통해 생산된 여러 기록을 수집·보존하여 당대의 기억을 후세에 남기려는 노력을 하고 있다(이미옥, 2014).

나아가 '재난 기록 관리'란 재난 사건과 관련하여 생산되는 공공기록 또는 민간기록 중 영구 보존할 가치가 있는 기록을 수집 및 보존하여 활용될 수 있도록 관리하는 것으로 정의된다(이미영, 2009). 재난 기록은 증거와 정보로써 가치를 가지고, 공동의 기억을 담는 사회적 기억 장치로써 해당 지역의 공동체 복원 및 정체성 회복에 기여하는 중요 자원이다. 보다 구체적으로 재난 기록은 재난이 수습·복구되는 업무 과정에서 효율적인 업무 진행을 가능하게 도구이자, 이후 비슷한 재난 사건이 발생할 때 합리적인 의사 결정을 할 수 있도록 해주는 설명책임성을 가진다. 또한, 재난 연구의 원천자료로써 학습자료 및 학술연구의 자료로 활용되어 정보적 가치를 가진

다. 재난 연구를 통해 사고의 현황과 문제점 그리고 개선점이 드러나고, 이는 사건이 재발하지 않도록 물리적으로 예방 및 대비가 가능하게 하여 사건 발생 시 신속하게 대응할 수 있도록 기여한다. 또한 재난의 기억을 공유하고 학습함으로써 국민들의 사회적 기억을 공유할 수 있도록 하고 사회적 경험을 재현한다. 특히 재난 사건에 있어서 민간기록은 공공기록이 담지 못하는 다양한 사회상을 품고 있어서 재난 사건을 직접 겪지 못했던 사람들이 재난 경험을 공유할 수 있도록 돕는다(이미영, 2009).

따라서 재난 기록 관리에 있어서는 공공기록뿐 아니라 다양한 민간기록을 효율적으로 관리할 수 있는 방안이 중요하다. 정혜지(2016)는 재난 기록을 관리함에 있어서 재난 기록의 특수성인 기록 유형의 다양성을 고려해야한다고 지적하였다. 여러 영역에서 기록을 포함한 다양한 매체의 산물이 생산되므로 어떻게 수집 및 관리할지에 대해 논하는 것이 중요하다는 것이다. 특히, 현대 사회에서는 어떤 재난 사건이 발생했을 때 소셜네트워크 서비스(이하 SNS)를 통해 이슈와 여론이 활발하게 생산된다. 특히 소셜미디어에서는 특유의 확산성을 기반으로 재난 상황을 빠르게 전파하고, 근접한 시각에서 국민들의 의견이 실시간으로 생성된다. 일례로, 2015년 11월 13일에 발생한 프랑스 파리 테러 사건의 경우, 발생과 동시에 사건과 관련된 내용이 트윗(Tweet)을 통해 즉각적으로 알려졌다. 파리 테러 사건과 관련하여 생산된 기록물은 사건 발생 지역을 벗어나 전 세계로 퍼지게 되면서 해당 사건과 관련된 기록물이 여러 지역에서 생산되게 되었다. 이렇게 생산된 SNS 기록물은 재난 발생 당시의 상황을 생생

하게 재현해주는 기록물로, 임지훈, 오효정, 김수정(2017)은 해당 기록물들의 중요성을 강조하였다.

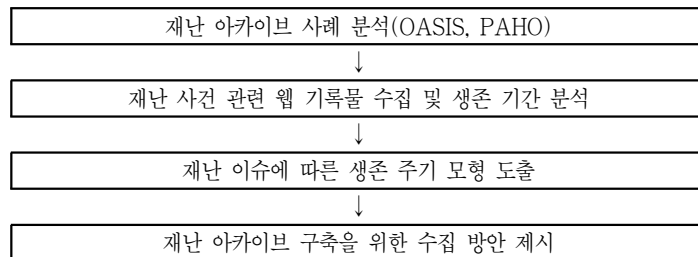
이와 같이 소셜미디어 및 온라인 뉴스미디어를 통해서 생산되는 기록물들은 공공기관에서 생산되는 보편적인 행정 기록물과 달리 다양한 주체 및 매체에 의해 생산되기 때문에 이를 관리하려면 기존의 공공기록물 관리에서 사용하던 수집·보존 및 활용 정책과는 또 다른 접근 방식이 필요하다. 해당 유형의 기록물들은 ‘웹’ 상에서 생산되고 제공되는 특성으로 인하여 정보의 활용성 측면에서 유용성 및 가치가 매우 높으나, 웹 기록물의 짧은 생명 주기 및 수집, 보존에 대한 방안이 적립되지 않아 소멸되기 쉽다(장인호, 황운영, 이규철, 2014). 또한, 웹 기록물의 방대한 생산량으로 인하여 수집에 있어서 자동 수집 방안에 대한 대책은 필수적인 요소가 되었다.

이렇듯 재난 웹 기록물의 가치와 보존의 중요성으로 인해 전 세계적으로 재난 아카이브의 구축은 국가적인 과제가 되어가고 있다. 실제로 다수의 국가가 온라인 아카이브를 구축함으로써 재난 관련 웹 자원을 보호하는 노력을 하고 있는데, 현재 국내에서는 OASIS 프로젝트의 일환으로 구축된 재난 아카이브가 그 역할

을 수행하고 있다. 한편, 이러한 재난 웹 기록물의 생산량은 국민들의 여론, 즉 사회적 관심도에 따라 큰 폭으로 좌우된다. 재난 이슈의 생산량은 이슈의 생성에서부터 소멸에 이르기까지 사회적 관심도에 따라 증가와 감소를 반복하는데, 이에 대한 특성을 파악할 수 있다면 효율적으로 이슈 수집 기간을 산정할 수 있다. 본 연구는 재난 이슈의 생존 특성 및 주기 모형을 분석한 유호선, 김현진, 오효정(2018)의 후속 연구로, 재난 이슈의 생존 주기 모형을 활용하여 이에 따른 재난 웹 기록물 자동 수집을 위한 수집 기간 및 주기를 제시하고자 한다. 구체적으로는 재난 이슈의 생존 주기를 분석함으로써 양질의 재난 정보가 생산되는 시점을 관측하고, 주기적 특성을 활용해 해당 이슈가 지속되는 기간을 미리 예측함으로써, 재난 아카이브 구축을 위한 수집 주기 및 기간의 기준을 마련하고자 한다.

1.2 연구내용 및 방법

본 연구는 재난 사건을 통해 생산되는 웹 기록물의 생산 주기 특성을 활용해 효율적인 수집 주기 및 기간을 제안하기 위하여 <그림 1>과 같은 과정으로 연구를 진행하였다.



<그림 1> 연구 방법

먼저 국내외 재난 아카이브 사례 분석을 통해 현재 재난 웹 기록물의 수집에 있어서 제한점을 파악하고자 하였다. 분석을 위해 국내 사례로는 OASIS 재난 아카이브를, 국외 사례로는 UN 산하 단체인 PAHO(The Pan American health Organization)에서 구축한 재난 아카이브를 선정, 해당 아카이브에서 다루고 있는 재난 유형과 수집 대상, 수집 기간 등의 현황을 분석하였다.

두 번째로는 최근에 발생한 주요 재난 사건에 관련하여 생산된 기록물을 수집하여 그 생존 기간을 분석하였다. 본 연구에서는 구체적인 분석 대상을 웹 기록물 중 여론에 따른 주기적 생산 특성을 지닌 SNS 기록물과 인터넷 뉴스 기사로 정하였다. 분석 대상 재난 사건은 행정안전부(舊 국민안전처)에서 발행한 재난연감 및 재해연보를 참고하여 최근 10년간 발생한 대형 사회 재난 및 자연재해를 대상으로 선정, 해당 재난과 관련된 SNS와 뉴스 기사를 수집하였다. 재난 발생일 기준으로 매 주별로 간헐 기록의 생산량을 조사하여 해당 이슈의 생존 기간을 계산하였다.

세 번째로는 재난 이슈별 생존 주기 모형을 도출하였다. 이를 위해 재난 이슈 생존에 영향을 미치는 다양한 독립 변인들을 규명하였으며, 통계 분석 툴인 R을 활용하여 회귀 트리 모델을 구축하였다. 이를 토대로 재난 이슈의 생존 기간에 영향을 미치는 요인들을 탐색하고 영향력을 파악하였다.

마지막으로는 재난 사건들을 도출된 생존 주기 모형별로 범주화하였고, 생존 주기 유형에 따라 효율적인 아카이빙 기간 및 주기 산정 방법에 대하여 제안하였다. 나아가 제안된 모형에 기반

한 수집 기간과 기존 아카이브 수집 결과를 통해 적용 가능성을 검증하였다.

1.3 선행연구

본 연구에 관한 선행연구로는 크게 이슈 생존 주기에 관한 선행연구와 사건 및 재난 아카이브에 관한 선행연구 등이 있다.

우선 이슈 생존 주기에 관한 선행연구로, 이종혁, 신동호, 강성민(2013)은 운영함수 방법론을 사용해 언론의 이슈 주기 분석 및 언론의 넵비저널리즘을 현상을 연구하였다. 하지만 기존의 운영함수 방법론은 전조 징후에서부터 이슈의 소멸 및 재점화에 이르기까지 일정한 패턴을 보이는 언론 이슈의 분석에는 적합하나 사회 재난과 같은 재난 이슈와 같이 재난 발생과 동시에 전조 징후 없이 폭발적으로 이슈가 생산되고 재난 사건의 종결과 동시에 급격히 감소하는 해당 이슈의 분석에는 한계점이 있었다. 이에 김철민, 최충익(2015)은 재난 이슈의 주기 특성에 적합한 생존 기간 방법론을 적용하여 1960년대 이후 발생한 대형 재난사고 18건을 선정하여, 해당 재난 사고로 발생한 기사 건수를 조사하여 이슈별로 생존 주기를 분석하였다. 또한 유호선, 명현, 오효정(2017)은 재난 이슈의 생존 주기를 분석함에 있어서 뉴스보다 여론의 사회적 관심도를 직접적으로 확인할 수 있는 SNS 기록물로 그 연구범위를 넓혔고 상관분석을 활용하여 재난 이슈의 생존 주기에 직접적으로 영향을 미치는 요인들을 파악하였고, 언론 보도의 생산 양상과 비교하였다.

다음으로는 사건 및 재난 아카이브에 관한 선행연구로, 이미옥(2014)은 해외 재난 아카이

브의 사례 조사를 통해 “대구 지하철 화재 참사” 기록의 단계별 관리 방안을 제시하였다. 신동희, 김유승(2016)은 노근리 디지털 아카이브를 대상으로 “노근리 사건”과 노근리 디지털 아카이브의 성격과 특성을 분석하였고, 디지털 아카이브를 수집하는데 있어 수집 정책을 제안하였다. 다음으로 임지훈, 오효정, 김수정(2017)은 재난 아카이브를 포함하는 넓은 개념인 사건 아카이브에 관한 시론적 연구를 진행함으로써 사건 아카이브의 개념을 정의하고 사건 아카이브 구축 시 필요한 고려사항을 제시하였다. 이상의 연구들은 재난 아카이브를 구축하기 위한 전체적인 가이드라인을 제시함과 동시에 디지털 웹 자원의 수집 방안에 대하여 고찰하였다.

본 연구는 실질적인 자동 수집 과정에서 효율적인 수집 기간을 제안하는 것에 초점을 두었다. 이를 위해 웹 기록물의 생산량에 영향을 미치는 사회적 관심도를 수치화하여 제시하였고, 웹 기록물의 주기적 특성을 활용하여 수집 방안을 제시하였다는 점에서 선행연구들과 차별점이 있다. 또한 연구방법 측면에서 그동안 관련 분야에서는 시도되지 않았던 기계학습 기법을 활용하여 웹 기록물의 생존 모형을 도출하였는데, 수리 통계적 방법을 활용하여 그간 재난 아카이브에서 사용된 수집 기간 및 주기를 검토하였다는 점에서 의미가 있다.

2. 이론적 배경

2.1 재난 이슈 생존 주기

사회에서 나타나는 다양한 이슈들은 대중의

관심을 끌기 위해 서로 경쟁적으로 작용한다. 승리한 이슈는 ‘사회적 관심’을 끌게 되지만 일정 시간이 지나면 사라지는 일련의 과정을 겪는데, Hilgartner & Bosk(1988)은 이러한 ‘사회적 관심’을 언론이 다룬 기사의 건수, 글자 수, 게재면적 등을 활용하여 계량적으로 분석할 수 있다고 주장하였다. 이러한 사회적 문제로부터 탄생한 이슈는 일정한 단계를 거쳐 성장하고 소멸되어 가는 경향을 보이는데 Downs(1972)는 ‘이슈생존주기(Issue Attention Cycle)’ 모형을 다음과 같이 제시하였다.

생존 주기의 첫 번째 단계는 ‘이슈 잠복’ 단계로 특정한 사회적 상황이 나빠지고 있어 관련 분야 전문가를 중심으로 이슈가 부각되고 있으나 아직 일반 대중들은 인지하지 못하고 있는 상태를 나타낸다. 두 번째 단계는 ‘이슈발견 및 표면화’ 단계로 해당 문제가 사회적 관심을 받게 되면서 대중들도 관심을 보이는 시점을 말한다. 세 번째 단계는 ‘관심의 증가 및 비용인식’ 단계로 일반인들의 관심이 최대에 도달하고 해당 문제 해결을 위한 처리 비용 등을 고려하여 현실을 인지하는 단계이다. 네 번째 단계는 ‘점진적 관심 감소’ 단계로 해당 문제에 대한 사회적 관심이 최대점을 지나 천천히 식어가는 단계이고, 이후 ‘관심쇠퇴’ 및 ‘재점화’ 단계를 거친다.

한편 이러한 ‘이슈생존주기’ 모형은 지속 기간 동안 안정적으로 증가와 감소를 반복하는 이슈를 대상으로 해당 문제의 사회적 관심 등락을 패턴화하여 분석하는데 매우 유용한 방법이지만 재난 이슈와 같이 잠복기가 미미하고 사건 발생 이후에 폭발적으로 증가하는 이슈의 분석에는 그대로 적용하기가 어렵다(김철민, 최충익, 2015).

재난 이슈의 생존 주기는 재난의 특성에 따라 크게 돌발형 패턴과 주기적 패턴 두 가지로 나타난다. 대부분의 재난 사건은 재난 발생과 동시에 이슈 생산량이 폭발적으로 증가하는 돌발형 패턴을 보이는데 특히 뉴스 기사의 경우 어떤 이슈에 대하여 한순간 집중적으로 보도되다가 갑자기 잠잠해지는 '냄비저널리즘'의 형태를 보인다(이종혁, 신동호, 강성민, 2013). 하지만 이러한 패턴은 공공미디어에 해당하는 뉴스 기사와 소셜미디어인 SNS 기록물 간에 차이를 두고 나타난다. 재난 사건의 종결과 동시에 급격히 감소하는 패턴을 보이는 것을 일치하는 부분이지만, 여론에 영향을 받아 시간이 지난 뒤에도 비정형적으로 돌출 및 재점화되는 양상이 SNS 기록물에서는 더욱 자주 나타나게 된다. 한편 자연재해와 같이 주기적으로 발생하는 재해 사건의 경우 사건 종료 이후에도 매년 비슷한 시기에 관련 기록물이 다시 생산되는 일정한 재점화 양상을 확인할 수 있다. 효율적으로 재난 웹 기록물을 수집하기 위해서는 매체의 유형 및 재난 특성에 따라 다르게 나타나는 주기적 특징을 파악하고, 이에 맞는 수집 방법을 설계해야 한다.

2.2 이슈 생존율 및 생존 기간

김철민, 최충익(2015)은 운영함수 방법론의 한계를 극복하고 재난 이슈의 생존주기 특성을 분석하기 위해 이슈의 생존율과 감쇠율 개념을 정의하고 측정하였다. 이슈 생존율은 '1-(발생 시점에서 해당 기간까지의 기사 건수/전체 기간의 기사 건수)'를 백분율로 나타낸 것으로 식 (1)과 같다. 즉 재난 발생 시점부터 해당 주까

지 누적된 이슈 생산량을 전체 기간 동안 발생한 이슈 생산량으로 나눈 것이다. 이를 통해 사회적 관심이 어떤 지점에서 집중적으로 나타났는지, 해당 이슈의 평균적인 생존 기간이 어느 정도인지를 수치화하였다. 이를 위하여 생존율의 감소 추세를 '지수적 감쇠(Exponential Decay)'로 측정하였는데 이는 일정 시간이 경과된 이후 해당 이슈의 생존 강도를 측정하고 비교하기에 적합한 방법이다.

$$R_i = (1 - \frac{\sum_{k=1}^i a_k}{N_a}) \times 100(\%), i = 1 \dots n.. \quad (1)$$

또한 김철민, 최충익(2015)은 생존율을 활용하여 이슈 생존 기간을 계산하였는데 생존율이 50%가 되는 지점까지를 해당 이슈가 공론화되는 시기로 보았고, 쇠퇴하여 생존율이 10%가 되는 지점까지를 해당 이슈가 상실되어 가는 시점이라고 규정하였다. 이 시점까지의 기간이 '생존 기간'으로 정의된다. 한편, 재난 관련 신문기사의 생존 기간을 통해 드러난 생존 양상은 잠복기나 성숙기가 거의 존재하지 않다가 사건 발생 이후 폭발적으로 증가하여 4~5주 이후에는 급속하게 감소하는 지수 함수적 패턴을 보여주고 있었다.

최근 유호선, 김현진, 오효정(2018)의 연구에서는 기록물 생산 현황뿐 아니라 재난으로 발생한 피해 정도와 해당 재난에 대한 이용자 관심도 등, 생존 기간에 미치는 다양한 요인에 대한 분석을 수행하였다. 이를 통해 재난의 특성에 따라 각 재난 사건별로 생존 기간은 큰 폭으로 차이가 있었고, 매체의 유형에 따라 뉴스와 SNS 기록물의 생존 기간에도 차이가 있음

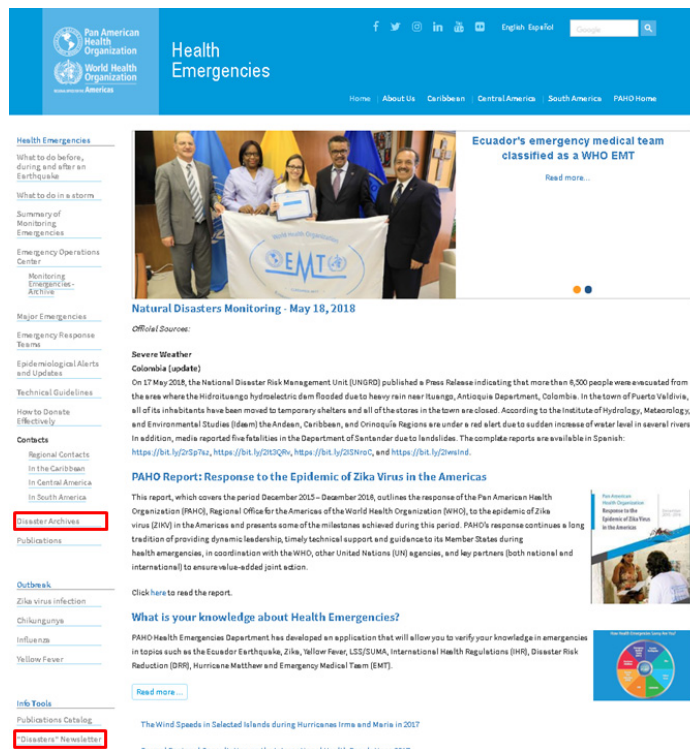
이 밝혀졌다.

3. 재난 아카이브 사례의 수집 주기 분석

본 장에서는 재난 웹 기록물의 자동 수집을 위한 수집 주기를 제안하기 위하여 현재 재난 아카이브를 운영하고 있는 국외 기관 PAHO와 국내 기관 OASIS의 사례를 분석해보고자 한다. 각 기관의 웹사이트 조사 및 담당자 인터뷰를 통하여 일반적인 서비스 현황과 수집 현황 및 수집 방법, 기간 등을 분석, 비교하였으며 한계점을 파악하였다.

3.1 PAHO

PAHO는 국제적인 공공 건강 단체로, 아메리카 대륙의 사람들의 건강과 삶의 표준을 증진시키기 위하여 1902년 12월에 설립되었다. PAHO는 세계보건기구(WHO)의 미국 사무소 역할을 하는 UN 시스템의 일부이다. PAHO는 재난 관련 기록물을 수집하고 웹 사이트를 통해 제공하고 있다. 제공하고 있는 방식은 크게 두 종류로 현지에서 발생한 재난에 대한 상황 및 인적 피해 현황을 기록한 보고서 중심의 기록물을 수집하는 [Disaster Archive] 탭과 재난 관련 뉴스 기록들을 수집하는 [Disaster Newsletter] 탭으로 나누어 서비스되고 있었다(〈그림 2〉 참조).



〈그림 2〉 PAHO 웹 사이트

[Disaster Archive] 탭에서는 기관이 자체적으로 생산한 기록물을 보존하기 위하여 주로 사진, 동영상, 문서를 수집하였다. <표 1>을 보면 2018년 05월 기준으로 78개의 재난에 대한 기록물이 있다. 기록물은 연도별로 선택해서 볼 수 있으며, 원문 및 링크 정보를 제공하고 있다. 주로 현지에서 있었던 재난 상황에 대한 개요와 사람에 대한 피해 현황을 기록한 기관 자체 생산 보고서이며 영어와 스페인어, 프랑스어 등으로 쓰여 있다.

다음으로 [Disaster Newsletter] 탭에서는 재난 관련 뉴스 기록물을 수집 및 제공하고 있다. <표 2>에서 확인할 수 있듯이 2001년을 시작으로 현재까지 33건의 재난 이슈에 대하여 언론의 관련 보도 및 인터뷰 자료 등을 수록하고 있다. 해당 기록물은 스페인어 및 영어로 제공되고 있고, 주제별 또는 연도별로 선택하여 이용할 수 있다. 또한 구독이 가능하며, PDF 형식으로 다운로드 가능하도록 지원하고 있다. 한편 [Archive와 Newsletter] 모두 SNS와의 연계를 통한 서비스를 활발히 제공하고 있었는

데 유튜브를 통해 관련 영상을 제공하고, 자신이 공유하고 싶은 기록물을 트위터에 '리트윗'하거나 페이스북 '좋아요'로 바로 연결되는 배너가 각 기록물마다 배치되어 있다.

한편 PAHO의 기록물 수집의 주기 및 기간은 비정기적으로 이루어진다. 재난이 발생한 이후부터 재난 사건이 완전히 종료될 때까지 아카이브가 주체적으로 내용을 선정하고 관련 기록을 수집하는 수동 방식으로 이루어지고 있기 때문이다.

3.2 OASIS

OASIS는 국립중앙도서관에서 현세대의 기록 문화유산인 인터넷 자원을 국가적인 차원에서 수집·축적하여 미래 세대에 연구 자료로 제공하기 위하여 2004년부터 추진하고 있는 온라인 디지털 자원 수집·보존 사업이다. 수집 대상은 국가기관, 공공기관, 교육·연구기관, 주요 상업기관, 민간단체 등에서 생산한 국가적 재난, 이슈, 주제 분야에 대한 웹 자원을 수집하는 것을 목표로 한다.

<표 1> PAHO 재난 아카이브 현황(2018년 05월 기준)

연도	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	계
기록 건수	2	58	23	18	43	28	14	34	7	15	2	0	3	39	10	6	302
재난 유형 (건수)	허리케인: 25, 태풍: 7, 화재: 2, 지진: 13, 홍수: 16, 화생방, (눈/산)사태: 1, 화산활동: 2, 전염병: 3, 호우, 한파, 기타: 3																

* 출처: 한희정 외(2017)에 2018년 자료 추가

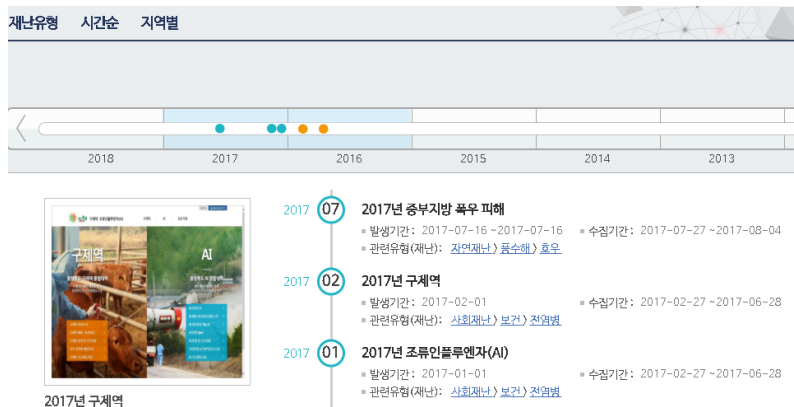
<표 2> Disaster Newsletter 현황(2018년 05월 기준)

연도	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	계
재난 이슈 (건수)	2	4	4	4	3	2	2	2	2	2	2			2	2		33

* 출처: PAHO, 연도별로 재구성

OASIS에서는 국가 재난 아카이브를 구축하여 관련 기록물을 수집하고 웹 사이트를 통해 시간순 또는 지역별로 제공하고 있다(〈그림 3〉 참조). OASIS 재난 아카이브는 소멸하기 쉬운 인터넷 자원을 수집하는 것이 주목적이고, 수

집 대상은 관련 기관·단체 웹 사이트, 뉴스 기사, SNS(개인 블로그), 웹 자료(문서, 이미지, 동영상) 등의 웹 자원에 초점이 맞춰져 있다(〈그림 4〉 참조). 따라서 PAHO가 보고서 및 일부 뉴스만을 수집 대상으로 국한했던 것에



〈그림 3〉 OASIS 재난 아카이브



〈그림 4〉 재난 기록물 유형

비해 다양한 웹 자원이 수집 대상으로 포함되어 있다. 그러나 현재 실제적으로 수집되고 있는 기록물은 주로 뉴스 기사와 일부 웹 자료에 국한되어 있는 것으로 파악되었다.

또한 OASIS 재난 아카이브의 수집 대상은 PAHO와 같이 UN이 관여한 모든 재난 사건에 대하여 수집되지는 않고, 우리나라에서 발생한 재난 사건 중 일부를 선정하여 해당 재난에 대한 기록물을 중심으로 수집하고 있다. 해당 기관 담당자 인터뷰 결과, 수집 대상의 재난 선정 기준은 가능한 많은 재난을 대상으로 해야 하지만, 예산 범위 내에서 구축할 수 있는 수량에 한계가 있기 때문에 특별재난선포 여부, 피해 규모, 사회적 영향, 수집 가능한 웹 자원 등을 고려하여 전문가 그룹의 자문을 참고하여 자의적으로 선정한다고 밝혔다.

그로 인해 OASIS 재난 아카이브의 수집 기간과 주기는 재난 사건별로 차이를 보였다. 그 중 수집 횟수는 특정 재난과 관련된 개별 신문 기사들과 특정 웹 페이지는 1번 수집(1-depth), 재난 관련 기관 웹사이트는 3번 수집(3-depth)

을 기본으로 하였다. 재난 아카이브에서 밝힌 수집 주기 즉, 수집 기간은 갱신 빈도 등을 고려하여 관련 이슈가 지속하는 동안 설정한다. 현재는 보통 1일~1년을 주기로 수집이 이루어진다는 점 이외에 구체적으로 어떤 유형의 재난들이 어떤 기준으로 수집 주기가 설정되어 수집되는지는 밝혀지지 않았다. 이후 이슈가 종료되어 더 이상 갱신이 없을 경우 수집주기를 1년(또는 중단)으로 변경한다.

〈표 3〉은 구체적으로 재난 아카이브에 수록되어 있는 재난 목록 및 수집 기간을 표로 정리한 것이다. 이슈 수집 기간은 OASIS 재난 아카이브에서 재난별로 수집된 뉴스 기사 및 SNS 기록물의 발행일, 즉 생산 시점을 기준으로 수집 기간을 조사한 것이다. 해당 유형의 기록물이 생산되지 않았을 시에는 ‘수집되지 않음’으로 표기하였다. 이슈의 수집은 대부분 사건 발생 당일 또는 이후부터 이루어졌으나 몇몇 재난 사건은 사건 발생 전의 기록물도 수집되었는데 이들은 주로 재난 전조 이슈에 해당하는 기록물이었다.

〈표 3〉 OASIS 재난 기록 수집 주기

재난 사건명	재난 구분	발생기간	이슈 수집 기간	수집 개시일
2017년 중부지방 폭우 피해	자연재난 > 풍수해 > 호우	2017-07-16 ~2017-07-16	2017-07-03 ~2017-08-02	사건 발생 전
2017년 구제역	사회재난 > 보건 > 전염병	2017-02-01	2014-06-02 ~2017-06-28	사건 발생 전
2017년 조류인플루엔자(AI)	사회재난 > 보건 > 전염병	2017-01-01	2006-11-28 ~2017-06-21	사건 발생 전
2016년 대구 서문시장화재 및 2017년 소래포구어시장화재	사회재난 > 화재 > 화재	2016-11-30 ~2017-03-18	2016-10-10 ~2017-05-31	사건 발생 전
2016년 경주지진	자연재해 > 지변 > 지진	2016-09-12	2016-09-12 ~2017-06-18	사건 당일
2015년 추자도 돌고래호 전복 사고	사회재난 > 해양사고 > 해난사고	2015-09-05	2015-09-06 ~2015-09-08	사건 발생 1일 후

재난 사건명	재난 구분	발생기간	이슈 수집 기간	수집 개시일
2015년 대한민국 중동호흡기증후군(메르스, MERS) 유행	사회재난 > 보건 > 의료	2015-05-20 ~2015-12-23	2014-05-27 ~2015-10-15	사건 발생 7일 후
2014년 판교 공연장 환풍구 붕괴 사고	사회재난 > 붕괴 > 건물붕괴	2014-10-17 ~2014-10-17	수집되지 않음	
2014년 장성 요양병원 화재 사고	사회재난 > 화재 > 건물화재	2014-05-28	2014-05-28 ~2014-06-02	사건 당일
2014년 세월호 침몰사고	사회재난 > 해양사고 > 해난사고	2014-04-16	2008-08-05 ~2017-06-13	사건 발생 전
2014년 경주 마우나오션리조트 체육관 붕괴 사고	사회재난	2014-02-17	2014-02-17 ~2014-05-16	사건 당일
우이산호 충돌 유류 오염사고	사회재난 > 해양사고 > 유조선사고	2014-01-31 ~2014-01-31	2014-02-03 ~2016-04-14	사건 발생 3일 후
2012년 구미 불산 누출 사고	사회재난 > 환경오염 > 유해화학물질 유출사고	2012-09-27 ~2012-09-27	2012-09-28 ~2015-05-07	사건 발생 1일 후
2012년 태풍 산바	자연재해 > 풍수해 > 태풍	2012-09-15 ~2012-09-17	2012-09-16 ~2015-09-10	사건 발생 1일 후
2011년 집중호우와 제9호 태풍 「무이파」	자연재해 > 풍수해 > 태풍	2011-08-06 ~2011-08-10	2011-08-02 ~2011-12-02	사건 발생 전
2011년 우면산 집중호우 산사태	자연재난 > 풍수해 > 산사태	2011-07-25 ~2011-07-28	2011-07-27 ~2013-02-22	사건 발생 2일 후
2008년 승례문 화재 사건	사회재난 > 화재 > 문화재재난	2008-02-10 ~2008-02-11	2003-12-31 ~2015-09-23	사건 발생 전
2007년 허베이스퍼리트호 원유 유출 사고	사회재난 > 해양사고 > 해난사고	2007-12-07 ~2007-12-07	2007-12-08 ~2015-08-28	사건 발생 1일 후
2007년 골든로즈호 충돌 침몰 사고	사회재난 > 해양사고 > 해난사고	2007-05-12 ~2007-05-12	2007-05-13 ~2007-05-30	사건 발생 1일 후
2007년 여수출입국관리사무소 화재 사고	사회재난 > 화재 > 건물화재	2007-02-11 ~2007-02-11	2007-02-11 ~2015-11-16	사건 당일
2005년 강원 양양군 화재	사회재난 > 화재 > 산불	2005-04-04 ~2005-04-06	2005-04-05 ~2005-05-17	사건 발생 1일 후
2003년 태풍 매미	자연재난 > 풍수해 > 태풍	2003-09-12 ~2003-09-13	2003-09-13 ~2004-09-13	사건 발생 1일 후
2003년 대구지하철 화재사고	사회재난 > 화재 > 화재	2003-02-18 ~2003-02-18	2003-02-19 ~2015-07-17	사건 발생 1일 후
2000년 동해안 산불	사회재난 > 화재 > 산불	2000-04-07 ~2000-04-15	2000-04-07 ~2000-05-10	사건 당일
1996년 씨랜드 청소년수련원 화재 사고	사회재난 > 화재 > 건물화재	1999-06-30 ~1999-06-30	1999-07-15 ~2015-10-07	사건 발생 15일 후
1997년 대한항공 801편 추락 사고	사회재난 > 교통 > 항공재난	1997-08-06 ~1997-08-06	수집되지 않음	
1995년 삼풍백화점 붕괴	사회재난 > 붕괴 > 건물붕괴	1995-06-29 ~1995-06-29	1995-07-01 ~2015-09-30	사건 발생 2일 후

* 출처: OASIS 재난 아카이브, 발생 시간별로 재구성

3.3 현황 비교 및 수집 한계점

앞선 두 사례를 살펴본 결과 재난 기록물을 수집 및 서비스하는 데 있어서 크게 양적인 측면과 수집 방법 적인 측면에서 두 기관은 차이를 드러내고 있었다.

우선 양적인 측면에서는 수집 대상으로 삼은 재난 건수가 OASIS 재난 아카이브가 31건, PAHO 재난 아카이브가 78건으로 약 2배 이상 차이가 남을 알 수 있었다. 이는 PAHO 재난 아카이브가 2003년 발생한 재난부터 수집한 데 비해 OASIS 재난 아카이브는 1995년 발생 재난부터 수집한 것에 비교하면 매우 큰 차이임을 확인할 수 있다. 이런 차이는 수집 대상의 선정에서 기인하는데 PAHO 재난 아카이브는 발생한 모든 재난 사건에 대하여 기록물을 수집하는 반면, OASIS 재난 아카이브는 일부 선정된 재난만을 수집하고 있기 때문이다. 따라서 OASIS 재난 아카이브의 수집 대상에는 “신종 인플루엔자”, “용흥지구 산불피해”, “상왕십리역 열차추돌사고” 등과 같이 인명 피해가 매우 크고 사회적으로 큰 이슈가 되었던 재난 사건들이 일부 포함되지 못하는 경우가 발생했다.

또한 수집 방식 측면에서는 PAHO와 OASIS 재난 아카이브 모두 수동 수집 방식을 택함으로써 해당 재난의 웹 기록물의 수집에 있어서 기록의 절대적인 양이 적고 유형 역시 다양하지 않았다. 실제로 PAHO는 공공기록물 역시 함께 수집하고 있었지만 민간기록물의 경우 뉴스 기사의 수집에 불과하였고, OASIS의 경우

에도 당초 수집 목표였던 SNS, 이미지 및 동영상 등의 웹 자원에 대한 수집은 거의 이루어지고 있지 못하는 상황이었다.

현재 OASIS 재난 아카이브는 수집 대상이 되는 재난의 선별 및 수집에 있어서 전문가의 자문 및 사이트의 특성을 분석하여 1회에서 3회까지 수집 횟수와 1일에서 1년까지 수집 주기를 설정하여 수집하고 있다. 전문가에 의해 선별 및 평가가 이루어지므로 품질은 보장되지만, 이로 인해 결국 수집되는 기록의 양이 적고 유형 역시 다양하지 않게 되므로 자동 수집 기술의 완전한 적용이 필수불가결한 요소가 되었다(강주연 외, 2017). 효율적인 자동 수집 지침이 마련되지 않으면 방대한 웹 자원의 특성상 인력 및 비용의 소모는 기하급수적으로 증가할 것이다.

4. 재난 이슈 생존 주기 모형

4.1 분석 대상

본 장에서는 앞서 지적한 수집 주기에 대한 문제점을 해결하기 위해서 재난 이슈 생존 주기 모형을 구축하고자 한다. 앞서 언급한 바와 같이, 본 연구에서는 웹 기록물 중 연구 범위를 SNS 기록물과 뉴스 기사로 정하였다. 데이터 수집은 SNS 매체 중에서는 트위터를 선정하여 Open API를 활용해 자동 수집하였고(유호선, 명현, 오효정, 2017), 신문 기사는 네이버 뉴스¹⁾를 활용하여 조선, 중앙 등 일간지와 지역신문사, 인터넷 신문사의 모든 보도 자료를 포

1) <https://news.naver.com/>

함하였다. SNS 매체 중에서 특히, 트위터를 선정한 이유는 여타 재난 이슈를 분석하는 연구에서 도구로 가장 활발히 사용되고 있기 때문이다(김현정, 2014). 트위터는 불특정 다수에게 신속 전파 가능한 특성으로 인해 과거 여러 재난 사례에서 재난 정보의 전파에 특히 많이 활용되어 왔다. 이러한 특성으로 인해 국립재난안전연구원 또한 재난 이슈에 관련한 여론을 파악하기 위해 매달 재난 트윗을 분석하고 있다. 또한, 우리나라 정부 부처 및 공공기관 대부분에서는 SNS를 이용하여 국민과의 대화 창구로 활용하고 있기 때문에 SNS에서 생산된 기록물은 공공기록의 측면도 일부 담당하고 있다(장인호, 황운영, 이규철, 2014).

이슈 생존 주기 분석 대상으로 선정한 재난 및 재해 사건은 행정안전부에서 발행한 2016 재난연감과 2016 재해연보를 활용하여 특별재난지역으로 선포된 재난을 중심으로 사회재난 12건, 자연재해 6건을 선정하였다. 또한 수집 기간은 사회재난의 경우 각 재난 사건에 대해서 재난안전대책본부가 운영된 시점부터 2016년 12월 31일까지로 설정하였고, 자연재해의 경우 전조 징후를 파악하기 위하여 대책본부가 운영되기 한 달 전부터 수집하였다. 또한 분석의 신뢰성을 유지하기 위하여, 관련 트윗 생산량이 500건 이하인 재난은 제외하였다. 한편 “세월호 침몰사고”는 재난 자체의 특성 보다는 정치·사회적인 영향으로 국민의 관심이 유지되고 있는 현안으로 특수한 경우에 해당되므로 분석 대상에서 제외하였다. 또한 자연재해의 경우에는 트윗을 수집하기 위한 키워드 선정이 용이한 태풍을 중심으로 수집하였으며 호우, 대설 등은 제외하였다.

〈표 4〉와 〈표 5〉는 각각 사회재난과 자연재해로 나누어 재난 이슈에 따른 SNS 기록물 및 뉴스 기사의 실제 생존 기간을 계산한 결과를 정리한 것으로, 생존 기간에 영향을 미치는 주요 요인 지표를 함께 제시하였다. 지표 중 사망자 수는 해양사고의 경우 실종자 수를 포함하고 있으며, 부상자 수는 감염병 유형의 경우 감염자 수를 포함하고 있다. 또한 인명 피해 규모 이외에도 1주 차 생존율과 2주 차 트윗 생산량은 재난 이슈의 생존 기간에 큰 영향을 미친다. 이중 1주 차 생존율은 사건 발생 1주일 동안 해당 이슈가 어느 정도 살아남았는지를 보여주는 것으로, 예를 들어 “신종 인플루엔자” 사건의 경우 생존율이 0.87이므로 1주 차에 전체 생산량의 13%가량이 생산되었다는 의미인 반면 “용흥지구 산불피해”의 생존율은 0.02로 첫 주에 전체 생산량의 98%가 생산되었음을 의미한다.

생존율 및 생존 주기를 활용한 재난 이슈 생존 특성 분석은 트위터와 뉴스, 두 유형의 매체 중에서 트위터가 국민의 관심도를 더 직접적으로 나타낸다는 점에 주목하여(김현정, 2014), 트위터를 중심으로 진행하였다. 우선 〈그림 5〉는 재난 이슈의 생성부터 소멸에 이르기까지 전체적인 생산 추이를 살펴보기 위하여 분석 대상이 되는 전체 재난 사건들의 1년간 이슈 생존율을 주별로 나타낸 것이다. 살펴보면 재난 사건별로 이슈의 생존율은 시간이 가면서 점차 줄어드는 추세를 공통적으로 보이지만 그 감소 속도와 폭은 각 재난 사건별로 재난 특성에 따라 매우 다르게 나타남을 알 수 있다. 여기서 생존율의 감소폭이 완만하게 이어진 재난 사건의 경우 꾸준히 계속해서 생산되었다는 의미가

〈표 4〉 기록물 생존 기간(사회재난)

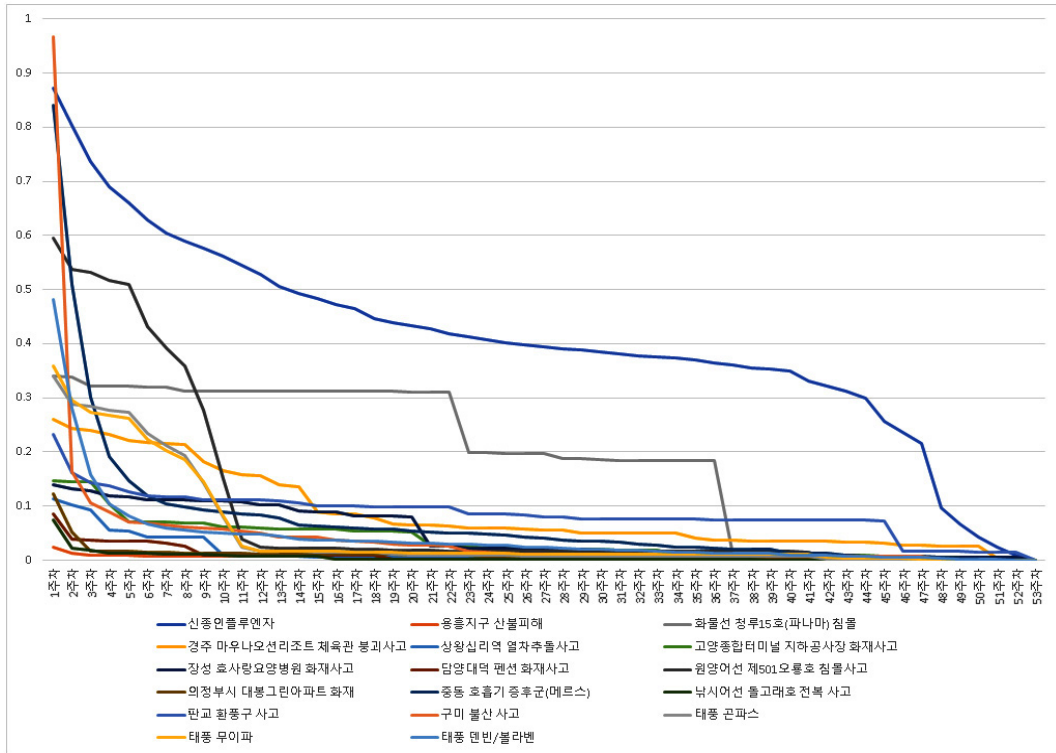
재난 사건명	기간	생존 기간 (일)		인명 피해 규모(명)		1주 차 생존율	2주 차 트윗 생산량
		트윗	뉴스	사망	부상		
신종인플루엔자	2009-11-04 ~2009-12-10	292	291	270	763,761	0.87	1913
용흥지구 산불피해	2013-3-9 ~2013-3-10	1	156	1	29	0.02	47
화물선 청루15호, 파나마침몰	2013-10-15 ~2013-10-30	22	67	10	8	0.33	2
경주 마우나오션리조트 체육관 붕괴사고	2014-02-17 ~2014-02-21	24	34	10	105	0.26	139
상왕십리역 열차추돌사고	2014-05-02	4	5	0	477	0.11	49
고양종합터미널 지하공사장 화재사고	2014-05-26	16	52	8	116	0.15	164
장성 효사랑요양병원 화재사고	2014-05-28	11	52	21	8	0.14	44
담양대덕 펜션 화재사고	2014-11-15	1	14	4	6	0.09	26
원양어선 제501오룡호 침몰사고	2014-12-01	9	6	27	7	0.59	164
의정부시 대봉그린아파트 화재	2015-01-10	2	39	5	125	0.12	44
메르스	2015-05-28 ~2015-07-28	15	32	38	186	0.84	88644
낙서어선 돌고래호 전복 사고	2015-09-05 ~2015-09-30	1	3	18	3	0.08	134

〈표 5〉 기록물 생존 기간(자연재해)

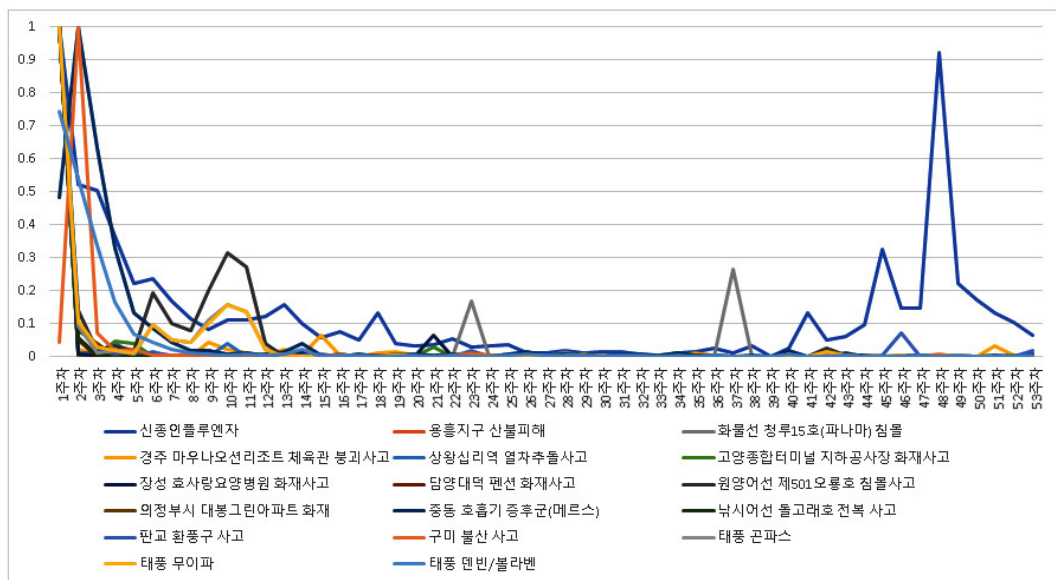
재난 사건명	기간	생존기간	
		트윗	뉴스
태풍 곤파스	2010-09-01~2010-09-03	67	65
태풍 무이파	2011-08-06~2011-08-10	2	4
태풍 덴빈, 볼라벤	2012-08-25~2012-08-30	3	3
태풍 산바	2012-09-16~2012-09-17	2	3
태풍 차바	2016-09-12~2016-10-06	3	2

고, 해당 재난의 생존 기간 또한 길게 나타나게 된다. 〈그림 6〉은 재난 사건들의 생존율을 상대적으로 비교하기 위해 각 재난 사건 별로 가장 많은 생산량을 보인 주의 생산량을 1로 기준하여, 정규화한 그래프이다. 이를 보면 최초의 재

난 이슈 발생 및 감소는 비슷한 양상으로 진행되지만 시간이 지남에 따라, 특정 재난 사건들은 이슈의 재점화가 지속적으로 일어나서 전체 생존 기간의 길이에 영향을 미치는 것으로 해석된다.



〈그림 5〉 재난 사건별 생존율 그래프



〈그림 6〉 재난 사건별 생존율 정규화 그래프

4.2 이슈 모형

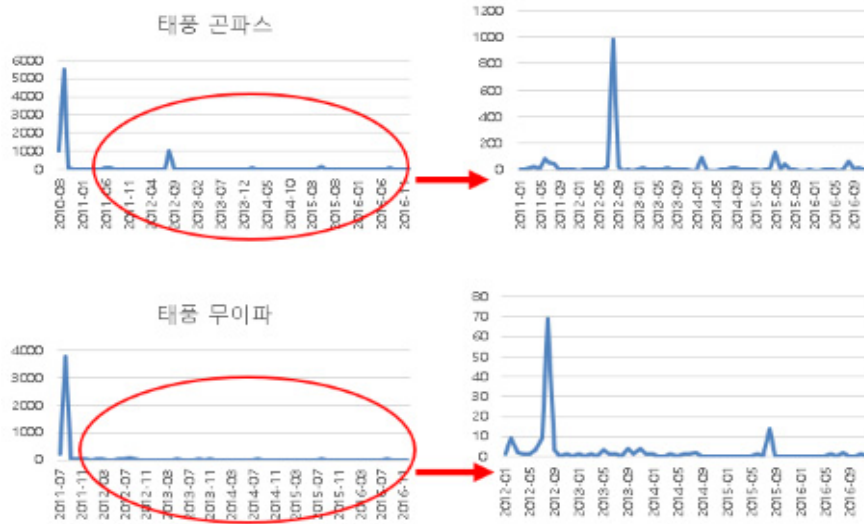
앞서 생존율 그래프를 통해서 사건 발생 후 1년만에 걸쳐 재난 이슈들의 폭발적 생산과 신속한 감소 패턴을 직접 확인할 수 있었다. 하지만 생존율 그래프는 재난 이슈의 대략으로 생존율과 생존 기간의 관계를 살펴볼 수 있지만, 세부적인 생존 주기의 유형을 살펴보는 데는 제한적이다. 따라서 본 장에서는 구체적인 재난 사건별로 실제 이슈의 생산량의 추이를 살펴보고, 공통적인 양상을 보이는 재난을 묶어 유형별로 구분하여 생존 주기 이슈 모형을 설계하고자 한다. 전체 재난 이슈의 발생 및 소멸은 공통적으로 사건 발생과 동시에 생산량이 폭등하고, 이후 감소하는 추세를 보이지만 세부적으로는 사건별로 이슈 생산 주기가 구분된

다. 재난 이슈의 이슈 모형은 크게 돌발형과 주기형으로 구분된다. 돌발형 모형은 대부분의 사회재난에서 나타나는 이슈 생산 주기 패턴이다. 특정한 전조 징후가 없이 사건 발생과 동시에 최대치로 생산된 이슈 생산량은 이후 매우 빠르고 일정한 감소폭을 보이게 된다. <그림 7>은 돌발형 유형을 가지는 재난 사건의 이슈 생산 그래프이고, 해당 재난들은 재난 발생과 동시에 급격한 이슈의 생산 이후에 일정한 비율로 그 생산 폭이 감소함을 알 수 있었다.

한편 주기형 유형은 태풍과 같은 자연재해 또는 조류 독감과 같은 감염병 유형의 사회재난처럼 특정하고도 일정한 기간에 걸쳐 주기적으로 일어나는 사건들이 주를 이룬다. 태풍은 주기형 유형을 뚜렷하게 보여주는 대표적인 재해 유형으로 <그림 8>은 “태풍 곤파스”와 “태



<그림 7> 돌발형 이슈 생산 그래프



〈그림 8〉 주기형 이슈 생산 그래프

풍 무이파”의 사건 발생 이후 몇 년간에 걸친 트윗 생산 빈도를 그래프로 나타낸 것이다. 그래프의 앞부분을 살펴보면 돌발형 재난 사건과는 다르게 사건 최초 발생 이전부터 전조 징후로 이슈가 점차 증가하는 단계가 확실하게 드러난다. 또한 빨간색 원으로 표시된 부분은 사건 이후 향후 몇 년간의 트윗 생산 빈도를 나타낸 것인데 주기적으로 매년 여름 일정한 기간 동안에 이슈의 재점화가 일어나는 것을 확인할 수 있다. 즉 주기 모형을 보이는 재해 및 재난 사건의 경우 매년 일정한 기간 동안 이슈가 주기적으로 생산되고 있었고, 재난 유형별로 주로 사건이 일어나는 기간을 분석하여 이를 자동 수집을 위한 수집 기간의 설정에 용이한 지표로 활용될 수 있을 것이다.

4.3 생존 주기 모형

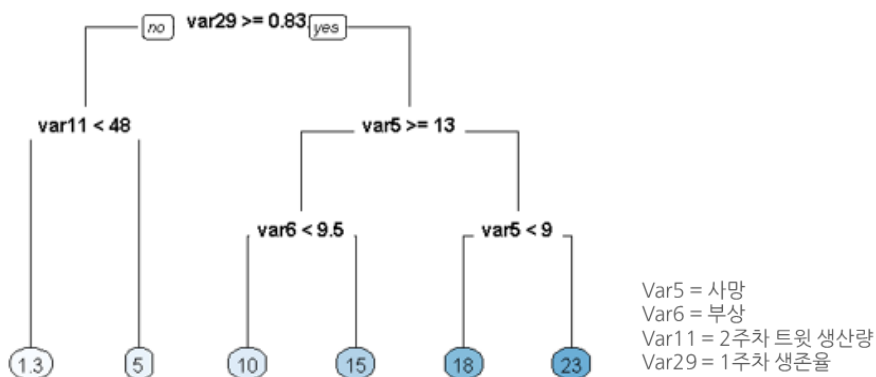
앞서 주요 재난 이슈들을 분석한 특성을 정

리해보면, 대부분 사회재난의 경우에는 돌발형 생산 유형을 보이고, 자연재해의 경우에는 주기적 생산 유형을 보였다. 주기적 생산 유형의 경우 매년 일정한 시기에 관련 기록물이 주기적으로 생산되므로 자동 수집을 위한 생존 기간 및 주기의 설정이 용이하다. 반면 돌발형 생산 유형의 경우 대부분 매년 반복적으로 생산되는 것은 아니기 때문에 재난 사건별로 수집 기간 및 주기가 산정되어야 한다. 그러나 주기형 재난과는 다른 형태로 돌발형 재난과 관련하여 생산된 기록물은 생성에서부터 소멸에 이르기까지 일련의 규칙성을 확인할 수 있었다. 본 장에서는 돌발형 재난의 생존 주기를 확인하기 위하여 통계 분석 툴인 R을 이용한 기계학습 방법을 사용하였다. 이를 통해 돌발형 재난의 지속적인 생존에 영향을 미치는 방대한 생존 요인을 대상으로 그중 적합한 생존 요인 및 그 영향력을 자동으로 계산하고자 한다. 구체적으로는 유호선, 명현, 오효정(2017)이 밝

힌 주요 생존 요인인 피해규모와 트윗 생산량을 중심으로, 인명 피해 및 재산피해, 각 주별 트윗, 리트윗, 좋아요 생산량 등을 독립 변인으로 설정하여 회귀 트리를 도출하여 분석하였다.

〈그림 9〉는 “신종 인플루엔자”, “메르스”와 같은 주기적 생산 양상을 보이는 감염병 유형의 재난을 제외한 돌발형 사회재난을 대상으로 제작한 생존 기간 회귀 트리의 알고리즘을 나타낸 것으로, 돌발형 재난의 생존 기간이 크게 사망 및 부상자 수와 1주 및 2주 차에 생산된 트윗 생산 빈도에 영향을 받는다는 것을 알 수 있다. 이 중 인명 피해 규모 이외에 재난 이슈의 전체 생존 기간이 처음 2주간의 트윗 생산 빈도가 가장 큰 영향을 미치는 것은 재난 발생 직후의 생산 빈도가 결국 전체 이슈 생존 기간으로 이어진다는 뜻이다. 즉, 최초 재난 발생 이후 1주 차에 가장 많은 기록물이 생산되는데 이후 2주 차로 들어서면서 그 생산량의 감소 비율을 통해 전체적인 생존 기간의 운곽이 드러난다는 의미이다. 이와 같이 최초 발생 당시의 생존율과 인명 피해 규모 간의 상호 연관성을 분석함으로써 생존 기간 알고리즘을 도출하였다.

〈표 6〉은 도출된 생존 주기 모형의 정확도를 측정하기 위하여 생존 기간 회귀 트리를 통해 계산한 예상 생존 기간과 실제 생존 기간을 비교하여 정리한 것으로, 재난별 평균 오차(MAE: Mean Absolute Error)와 평균 백분율 오차(MAPE: Mean Absolute Percent Error)를 계산하였다. 평균 오차는 생존 기간과 예상 생존 기간 사이의 차의 평균이고, 평균 백분율 오차는 실제 생존 기간에 대한 평균 오차의 비율이다. 돌발형 재난 생존 주기 모형의 정확도는 평균 오차 1.07, 평균 백분율 오차 51%로 나타났다. 오차가 가장 크게 나타난 “돌고래호 전복 사고”는 피해자 수가 매우 큰 사건이었음에도 트윗과 뉴스 보도 상으로도 크게 이슈가 되지 않은 특이한 사건이었다. 이 사건의 사망자 수는 18명으로 “청루 15호 침몰 사고” 사망자 수인 10명보다 많지만 생존 기간은 큰 차이를 보였다. 이는 이 사건의 전체 트윗 발생량 자체가 적어서 생겨난 오차로 보인다. 이 사건을 제외하면 각 재난별 평균 백분율 오차는 14%로 나타났으며, 1주 미만인 0.79주의 평균 오차를 보이는 것으로 평가되었다. 이는 새로운 재난 사



〈그림 9〉 돌발형 재난 생존 기간 회귀 트리

〈표 6〉 돌발형 재난 예상 생존 기간

재난 사건명	생존 기간(주)	예상 생존 기간(주)	오차 비율(%)
용흥지구 산불피해	1	1.3	30
화물선 청루15호, 파나마침몰	22	23	5
경주 마우나오션리조트 체육관 붕괴사고	24	23	4
상왕십리역 열차추돌사고	4	5	25
고양종합터미널 지하공사장 화재사고	16	17.5	9
장성 효사랑요양병원 화재사고	11	10	9
담양대덕 펜션 화재사고	1	1.3	30
원양어선 제501오룡호 침몰사고	9	10	11
의정부시 대봉그린아파트 화재	2	1.3	35
메르스	15	15	0
남시어선 돌고래호 전복 사고	1	5	400
평균 오차(MAE)		1.07	
평균 백분율 오차(MAPE)			51

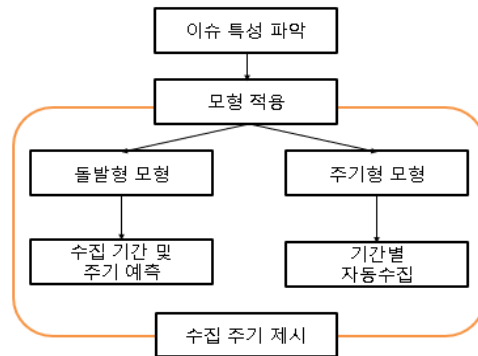
* 출처: 유호선, 김현진, 오효정(2018) 인용 후 재구성

건이 발생했을 때 해당 이슈의 생존 기간을 조사하기 위하여 오랜 기간을 들여서 이슈 생산량을 조사하는 것보다는, 최초 2주 동안 발생한 생존을 및 인명 피해 규모를 통해 예측하는 것이 더 효율적이라는 의미로 해석할 수 있다.

간과 예측 수집 기간을 비교하였다. 이를 통해 기존 수집 기간을 통해 누락된 기록물은 없는지 또는 수집 기간의 선정이 비효율적이지는 않았는지 확인하였다.

5. 이슈 생존 주기 모형 기반 재난 아카이브 수집 방안

이번 장에서는 본 연구의 최종 목표인 재난 아카이브 구축을 위한 수집 방안을 제안하고자 한다. 보다 구체적으로는 〈그림 10〉과 같은 순서로 앞에서 제시한 이슈 생존 주기 모형을 활용하여 그간 수작업으로 산정되었던 재난 웹 기록물의 수집 기간 및 수집 주기를 자동으로 설정하는 방안을 제안하고자 한다. 또한 제안한 방법의 적용 및 검증을 위해, OASIS 재난 아카이브의 특정 사건을 대상으로 실제 수집 기



〈그림 10〉 수집 주기 설정 순서도

5.1 재난 아카이브 수집 주기 제안

새로운 재난 사건이 발생한 경우, 재난 기록 수집을 위한 첫 번째 과정은 해당 이슈의 특성

을 파악하는 것이다. 4장에서 언급한 바와 같이, 재난 이슈 생존 기간을 예측하기 위한 가장 주요 요인은 해당 재난 발생 후 1~2주 내 발생한 기록의 양과 피해 규모이다. 따라서 최초 2주 동안 새로 발생한 사건과 관련된 뉴스 및 SNS 기록물의 생산 현황과 피해 상황을 집계하여 이슈 생존 모형을 예측하기 위한 입력치로 활용한다.

구체적인 수집 기간의 설정은 이슈 생존 주기 모형 적용 결과에 따라 두 가지로 분류하여 제시하고자 한다. 먼저 돌발형 모형에 따른 수집 기간 제안은 다음과 같다. 우선 대상 재난으로는 대부분의 사회재난에 적용이 가능하다. 특히 돌발형 모형의 경우에는 사건 발생과 동시에 폭발적으로 증가한 이슈의 생산량이 이후 일정하게 감소하기 때문에 비교적 정확한 범위에서 생존 기간의 예상이 가능하다. 예상 생존 기간을 바탕으로 재난 사건 발생 이후 수집 기간 및 수집 주기를 낭비 없이 효율적으로 설정할 수 있다. 예를 들면 인명 피해 규모가 10명, 1주 트윗 발생량이 500건 정도 되는 돌발형 재난이 새롭게 일어났다고 했을 때 모형을 통해 생존 기간이 2주 내외일 것으로 예상 가능하기 때문에 사건 이후 빠르고 효율적으로 수집 계획을 세울 수 있다.

주기 모형의 경우에는 다음과 같은 방법으로 수집 방안을 제안한다. 우선 해당 모형에 적용 가능한 재해들은 태풍, 호우, 대설과 같은 자연 재해와 AI, 구제역과 같이 주기 모형을 갖는 감염병 형태의 재난에 적용이 가능하다. 해당 유형의 재난의 경우에는 예상 생존 기간을 정확히 계산하는 것이 불가능하므로 주기 특성에 따른 수집 계획을 수립해야 한다. 일단 주기형

재해 사건이 발생한 경우에는 특정 기간 동안에만 집중적으로 이슈가 생산되므로 최초 수집 기간을 길게 유지할 필요성은 없다. 그러나 해당 유형의 재난의 경우에는 매년 관련 이슈가 반복 재생산되므로 같은 유형의 재난 별로 장기적인 수집 계획을 수립할 필요성이 있다. 예를 들면 태풍의 경우 매년 8~9월을 중심으로 재해 사건이 발생하기 때문에 해당 주기별로 자동 수집 계획을 설계할 필요성이 있다. 또한 발생 전조 징후를 포착하기 위해 발생 1개월 전을 수집 시작 기간으로 설정해야 한다.

5.2 수집 설계시 고려사항

실제적으로 이슈 생존 주기를 활용한 수집을 위해서는 수집 기록의 선정 정확도를 높이기 위해서 자동 수집 방안 설계 단계에서 다음과 같은 과정이 포함되어야 한다.

첫 번째 과정은 중복 기록물을 제거할 수 있는 방안이 마련되어야 한다. SNS 기록물에서는 '리트윗'과 같은 '떠가기'를 통해 본질적으로 같은 기록물이 많이 생산되고, 뉴스 기사에서도 역시 같은 내용 및 맥락의 기사들의 재생산이 많이 이루어지고 있기 때문이다.

두 번째 과정은 생존 기간을 통해 수집된 기록물의 품질을 확인하여야 한다. 본 연구에서는 생존 기간을 생존율이 10%가 되는 지점으로 정의하여 수집 기간을 제시하였지만, 해당 기간 동안에 불필요한 기록물들이 많이 생산된다면 그 시점을 20% 또는 50%와 같이 유동적으로 조절하여 수집 품질을 유지하여야 한다.

세 번째 과정은 재난 사건과 관련된 관련 정치 사회적 이슈의 흐름을 살펴볼 필요가 있다.

재난 이슈는 최초에 예측된 생존 기간보다도 더욱 오랜 시간 뒤에 정치 사회적 이슈에 의해 재점화될 수 있는데 이러한 기록물까지 수집 범위에 포함시켜야 할지 고찰하고, 이슈의 흐름을 관찰할 필요가 있다. 특히 돌발형 재난의 경우 이러한 예외 요인이 크게 작용하므로 수집이 끝난 이후에도 1년 주기로 정치·사회적 이슈에 의한 예외적인 돌발 요인이 발생하지 않았는지 확인이 필요하다. 수집 이후에는 수집 기록물에 대한 생존 기간 및 수집 기간의 검증을 통해 해당 재난 관련 웹 기록물이 수집 기간의 부족으로 인해 수집 누락이 되지 않도록 사후 검증이 수반되어야 할 것이다.

이와 같은 수집 기간의 자동 산정은 웹 기록물의 자동 수집 단계에 포함됨으로써 그 효용성이 증가한다. 더욱이 이와 같은 웹 기록물은 휘발성으로 인해 수집 시기가 늦어질 경우 주요 기록물들이 소멸될 가능성이 있어서 수집 시기가 중요하게 평가되고 있다. 유형별 재난 기록물 수집 기간 선정의 자동화를 통해 절약되는 예산 및 인력을 활용해 현재까지는 재난

아카이브의 수집 대상에 포함되지 못했던 다른 대형 재난 사건들까지 폭넓게 수집되어야 함을 제언한다.

5.3 제안 방법의 검증

제안 방법에 의한 수집 기간의 설정이 실제 재난 웹 기록물의 수집에 적합한지 살펴보기 위하여 <표 7>과 같이 OASIS 재난 아카이브의 재난별 이슈 수집 기간과 본 연구에서 도출한 제안 수집 기간을 비교하였다. 비교 대상의 선정은 본 연구의 연구 범위 안에 해당되던 사회재난과 자연재해 중에서 OASIS의 수집 대상 재난과 일치하는 것들을 선정하였다.

<표 7>의 검증 결과를 요약하면 다음과 같다. 첫 번째로 사회 재난의 경우 “메르스”를 제외한 재난 사건들에서 OASIS의 수집 기간이 제안 수집 기간보다 짧은 것으로 파악되었다. 이는 또한 이슈의 종결 이전 즉, 이슈가 계속 생산되는 동안에도 OASIS 재난 아카이브의 수집이 종료되었다는 의미이기도 하다. 현재와 같

<표 7> 이슈 수집 기간 비교

재난 사건명	발생 기간	OASIS 이슈 수집 기간	제안 수집 기간
경주 마우나오션리조트 체육관 붕괴사고	2014-02-17 ~2014-02-21	2014-02-17 ~2014-05-16	2014-02-17~2014-08-10
장성 효사랑요양병원 화재사고	2014-05-28	2014-05-28 ~2014-06-02	2014-05-28~2014-08-19
메르스	2015-05-28 ~2015-07-28	2014-05-27 ~2015-10-15	2015-05-28~2015-07-29
낙서어선 돌고래호 전복 사고	2015-09-05 ~2015-09-30	2015-09-06 ~2015-09-08	2015-09-05~2015-09-12
태풍 무이파	2011-08-06 ~2011-08-10	2011-08-02 ~2011-12-02	2011-07-06~2011-08-18 (매년 8, 9월 발생 시 주기적으로)
태풍 산바	2012-09-16 ~2012-09-17	2012-09-16 ~2015-09-10	2012-08-16~2012-09-28 (매년 8, 9월 발생 시 주기적으로)

이 재난 사건별로 해당 이슈가 완전 종결될 때까지 계속해서 기록물의 생산을 확인하여 수집하려면 비용이 가중되고, 또한 수집이 일찍 종료될 때 수집 대상에 포함되어야 할 여러 유형의 기록물들이 실제로 수집되지 못할 가능성이 있다. 따라서 본 연구에서 제시한 수집 기간에 의해 아카이빙 전략을 수립할 시 기존에 수집되던 기록물들은 누락 없이 수집될 수 있을 것이다.

두 번째로 자연 재해의 경우, 전조 징후로 인하여 수집 시작 기간이 달라질 수 있다. 본 연구에서는 자연 재해의 경우 사건 발생 이전 한 달간을 전조 징후를 포착하기 위한 기간으로 설정하였다. 태풍 무이파의 경우, 실제 재난은 2011년 8월 6일에 발생했으므로 제안 수집 기간은 전월인 7월 6일부터 8월 18일로 제안하였다. 실제 OASIS 재난 아카이브의 '태풍 산바'의 이슈 수집 결과를 보면 재난 발생 한 달보다 더 이전에 이미 전조 징후로 관련 이슈들이 생산되었다는 것을 확인할 수 있다. 반면 실제 발생한 기간 내의 기록물은 수집되지 않았다. 따라서 자연재해의 경우 더 넓은 범위에서 전조 징후를 포착하기 위한 노력이 요구된다. 한편, 자연재해의 경우 매년 해당 사건들이 회자되며 관련 기록물이 재생산되는데 현재 OASIS 재난 아카이브에서는 사건 당해 연도의 기록물만 수집 대상으로 하고 있었다.

세 번째로 메르스 사건의 경우 제안 수집 기간보다 OASIS 이슈 수집 기간이 긴데, 이것은 예외적인 재점화가 일어났기 때문이다. 생존 기간에 의한 "메르스" 이슈의 소멸은 그 전에 7월 29일에 일어났지만 10월 13일에 기존에 완치되었다고 했던 마지막 메르스 환자가 바이러스 재검출 되는 사건이 발생하였다. 이에 따라

이슈가 소폭 재점화 되었지만, 생존 기간에는 큰 영향을 미치지 못했다. 본 연구에서 제안한 모형의 경우 이 같은 예외상황에 취약한 반면, OASIS 재난 아카이브에서는 손으로 직접 재난 기록의 선별 및 수집 대상을 고려하므로 이러한 기록물도 포함될 수가 있었다. 이처럼 생존 기간 이후에도 일부 중요한 의미를 가지는 기록물이 생산될 수 있으므로 설계시 고려사항에서 지적한 것처럼 특정 이슈들에 대한 재점화 여부의 확인이 요구된다.

6. 결론 및 제언

재난 사건을 통해 생산된 기록을 수집하여 아카이브를 구축하고 관리하는 것은 재난을 관리하기 위함 뿐 아니라 새로운 재난의 대비에도 매우 중요한 일이다. 본 연구는 향후 자동 수집으로 나아가기 위한 수집 방안을 제시하는 것을 목표로 재난 웹 기록물의 자동 수집 기반 수집 기간 및 수집 주기를 제안하는 것을 목표로 하였다.

연구결과를 요약하면 다음과 같다. 우선 국내에서 주로 발생하는 재난 이슈들은 그 생존 특성에 따라 돌발형 유형 및 주기형 유형으로 분류되었다. 대부분의 사회재난이 해당되는 돌발형 유형은 전조 징후 없이 사건 발생과 동시에 이슈가 폭발적으로 생산되었지만, 이후 재난 특성에 따라 규칙성을 가지고 점차 감소하였다. 또한 이 때문에 돌발형 재난 이슈들은 피해 규모를 파악한 뒤 최초 1~2주의 생산 추이를 분석함으로써 향후 이슈의 생존 기간을 예측할 수 있었다. 이러한 특성을 반영한 이슈 생

존 주기 모형을 통해 이후 새로운 재난 이슈가 발생하였을 때 신속하고도 효율적으로 수집 기간을 정하여 SNS 기록물 및 뉴스 기사를 수집하는 방안을 제안하였다. 또한 주기형 재난 이슈들은 대부분의 자연재해들이 포함되었는데, 국내에서 발생하는 자연재해는 지리적 기후적 특성으로 인하여 종류가 제한되고 같은 유형의 재해가 반복하여 일어나고 있었다. 또한 해당 이슈들은 돌발형 재난 이슈들과는 다르게 사건이 종결된 이후에도 매년 같은 시기에 이슈의 재점화가 반복적으로 일어나고 있었다. 따라서 주기형 재난 이슈들의 수집에 있어서 사건별 수집이 아닌 재난 유형별로 수집 기간을 산정해 자동 수집하는 방안을 제안하였다.

한편 본 연구는 다음과 같은 한계점들은 남겼다. 첫 번째로는 분석 방법 측면에서 기계학습을 통한 모형 도출시 분석 대상의 종류가 클

수록 정확도가 높아진다. 그러나 사회적 이슈가 큰 대형 재난을 그 대상으로 함으로써 분석 대상의 수가 제한적이었다. 두 번째로는 모형의 검증에 있어서 OASIS 재난 아카이브의 재난 수집 기간과 실제 생존 기간을 비교하였는데, OASIS에서 수집한 재난이 대형 재난 이슈를 모두 포함하지 못함에 따라 비교 대상이 많지 않았다. 한편 아직 웹 기록물이 적절하게 수집 및 보존되기 위해서는 메타데이터 프레임워크 개발 및 일관된 장기보존포맷 유지 등 해결되어야 할 과제가 많다(장인호, 황윤영, 이규철, 2014). 이 시점에도 수집되지 못한 웹 기록물들은 언제라도 소멸할 위기에 처해있는바 실제 자동 수집에 있어서 이러한 한계를 고려할 수 있는 방법 및 적절한 수집 기간의 산정뿐만 아니라 수집의 질을 높이기 위한 설별 방안에 관한 연구가 촉구된다.

참 고 문 헌

- 강주연, 한희정, 김용, 오효정 (2017). 재난안전 기록정보자원 아카이빙 전략의 실증적 검증. 한국비블리학회지, 28(4), 223-246.
- 김철민, 최충익 (2015). 대형 재난사고 이슈의 생존주기 분석. 서울도시연구, 16(4), 147-162.
- 김현정 (2014). 재난 위기 관련 정부 기관 소셜미디어의 위기 관리 커뮤니케이션 도구로서의 역할에 대한 연구. 한국광고PR실학회, 7(4), 60-98. <https://doi.org/10.21331/jprapr.2014.7.4.003>
- 신동희, 김유승 (2016). 노근리 디지털 아카이브 수집정책 연구. 한국기록관리학회지, 16(3), 1-30. <https://doi.org/10.14404/jksarm.2016.16.3.001>
- 유호선, 김현진, 오효정 (2018). 재난 사건별 이슈 생존 주기 유형 분석. 한국정보기술학회논문지, 16(3), 125-134. <https://doi.org/10.14801/jkiit.2018.16.3.125>
- 유호선, 명현, 오효정 (2017). 소셜미디어에서의 사회재난 이슈 생존 주기 분석. 한국정보기술학회논문지, 15(7), 139-147. <https://doi.org/10.14801/jkiit.2018.16.3.126>
- 이미영 (2009). 재해 지역의 기록화 사례 연구. 기록학연구, (21), 85-116.

- 이미옥 (2014). 대구지하철화재참사 기록의 단계별 관리 방안. 경북대학교 기록학석사학위논문.
- 이종혁, 신동호, 강성민 (2013). 이슈 보도 주기로 관찰된 '냄비저널리즘' 현상. 한국방송학보, 27(4), 206-250.
- 임지훈, 오효정, 김수정 (2017). 사건 아카이브의 시론적 연구. 기록학연구, (51), 175-208.
- 장인호, 황윤영, 이규철 (2014). 정부부처의 페이스북 페이지 기록물 수집·보존·재현 방법. 한국기록관리학회지, 14(2), 117-128. <https://doi.org/10.14404/jksarm.2014.14.2.117>
- 정혜지 (2016). 9/11 컬렉션을 통해 본 재난 기록의 관리. 한국비블리아학회 발표논집, 157-161.
- 한희정, 강주연, 김용, 오효정 (2017). 자동수집 기반 재난안전 기록정보자원 아카이브 설계 및 구축전략. 한국기록관리학회지, 17(4), 127-154.
- Downs, A. (1972). Up and down with ecology-the issue attention cycle. *PublicInterests*, 28, 38-50.
- Hilgartner, S., & Bosk, C. L. (1988). The rise and fall of social problems: A public arenas model. *American Journal of Sociology*, 94(1), 53-78. <https://doi.org/10.1086/228951>

[웹 사이트]

- OASIS 재난 아카이브 (2018, May). Retrieved from <http://www.oasis.go.kr/collection/mainDisa.do>
- PAHO 재난 아카이브 (2018, May). Retrieved from https://www.paho.org/disasters/index.php?option=com_content&view=article&id=1933:disaster-archives&Itemid=915&lang=en

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Chung, Hye-Ji (2016). 9/11 as seen through the collection of disaster management of records. Korean Biblia Society for Library and Information Science and Korean Society of Archives and Records Management Joint Conference Proceeding, 157-161.
- Gang, Ju-Yeon, Han, Hui-Jeong, Kim, Yong, & Oh, Hyo-Jung (2017). Empirical study of disaster and safety record information resources archiving strategy based on automatic acquired Web records from related organizations. *Journal of the Korean Biblia Society for Library and Information Science*, 28(4), 223-246.
- Han, Hui-Jeong, Gang, Ju-Yeon, Yong Kim, & Oh, Hyo-Jung (2017). 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.

- Jang, In-Ho, Hwang, Yun-Young, & Lee, Kyu-Chul (2014). The methods of collecting, preservation, reproduction for records of public sector's Facebook pages. *Journal of Korean Society of Archives and Records Management*, 14(2), 117-128.
<https://doi.org/10.14404/jksarm.2014.14.2.117>
- Kim, Chul-Min & Choi, Choong-Ik (2015). An exploration of issue attention cycle of great disasters. *Seoul Studies*, 16(4), 147-162.
- Kim, Hyun-Jung (2014). A study on the role of the disaster and crisis-related government's social media as the crisis management communication tool. *The Korean Advertising and PR Practitioners Society*, 7(4), 60-98. <https://doi.org/10.21331/jprapr.2014.7.4.003>
- Lee, Jong-Hyuk, Shin, Dong-Ho, & Kang, Sung-Min (2013). The examination of 'Pot Journalism' through issue-cycle analysis - an application of spline function analysis. *Korean Journal of Broadcasting and Telecommunication Studies*, 27(4), 206-250.
- Lee, Me-Ok (2014). A study on phased management plan of 'Daegu Subway Disaster' record. Department of Archival Science Graduate School, kyungpook National University.
- Lee, Mi-Young (2009). A case study on the documentation in a disaster area - on the basis of great Hanshin-Awaji earthquake -. *The Korean Journal of Archival Studies*, (21), 85-116.
- Lim, Ji-hoon, Oh, Hyo-jung, & Kim, Soojung (2017). A preliminary study on event archives. *The Korean Journal of Archival Studies*, (51), 175-208.
- Sinn, Dong-Hee, & Kim, You-seung (2016). Collection development policy for the No Gun Ri digital archive. *Journal of Korean Society of Archives and Records Management*, 16(3), 1-30. <https://doi.org/10.14404/jksarm.2016.16.3.001>
- Yoo, Ho-Suon, Kim, Hyeon-Jin, & Oh, Hyo-Jung (2018). Analysis of issue life cycle types of disaster events. *The Journal of Korean Institute of Information Technology*, 16(3), 125-134.
<https://doi.org/10.14801/jkiit.2018.16.3.125>
- 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. <https://doi.org/10.14801/jkiit.2018.16.3.126>

