

토픽 모델링을 활용한 '수돗물 유충' 뉴스 빅데이터 분석

News Big Data Analysis of 'Tap Water Larvae' Using Topic Modeling Analysis

이수연*, 김태종**

한국환경교육연구소*, 한국청소년정책연구원 청소년정책분석평가센터**

Su Yeon Lee(risy90@naver.com)*, Tae-Jong Kim(k2boy3@naver.com)**

요약

본 연구는 '수돗물 유충' 사태의 뉴스 빅데이터를 분석함으로써, '수돗물 유충'과 관련된 주요 키워드와 토픽을 파악하여, 환경문제에 대한 위기 대응력 제고방안을 제안하기 위해 수행되었다. 2020년 7월 13일부터 8월 31일까지 보도된 1,975건의 '수돗물 유충' 뉴스를 토픽 모델링 기법으로 분석하였다. 그 결과 언론에서 나타난 '수돗물 유충' 사태가 발생기, 확산기, 수습기로 구분되며, 각 5개의 토픽을 선정하여, 환경문제의 발생과 추진 과정을 확인할 수 있었다. 분석 결과를 바탕으로 환경문제에 대한 위기 대응 방안을 다음과 같이 제안하였다. 첫째, '수돗물 유충' 사건을 중심으로 얽혀있는 다양한 맥락을 탐구하고 통합적인 안목을 형성하는 교육으로 환경문제에 대한 대응력을 기를 수 있도록 해야 한다. 둘째, 인터넷 커뮤니티를 활용한 시민참여의 환경정보 공유와 환경감시 역할 부여가 필요하다. 셋째, 신속하고 정확한 환경정보 제공과 소통을 담당하는 환경 커뮤니티의 양성 및 배치가 필요하다. 본 연구는 '수돗물 유충' 관련 뉴스 빅데이터를 기반으로 국내에서 처음으로 토픽 모델링 분석기법을 활용하여 분석한 연구로서, 비정형 데이터로 나타나는 환경 관련 이슈를 실증적이고 체계적으로 분석했다는 학술적 의의와 환경교육 및 커뮤니케이션 개선 방안을 제시했다는 정책적 의의를 지닌다.

■ 중심어 : | 수돗물 | 유충 | 뉴스 빅데이터 | 토픽 모델링 분석 | 사건 중심 환경 탐구

Abstract

This study was conducted to propose measures to improve crisis response to environmental issues by analyzing the news big data on the 'tap water larvae' situation and identifying related major keywords and topics. To accomplish this, 1,975 cases of 'tap water larvae' reported between July 13 to August 31, 2020 were divided into three periods and analyzed using topical modeling techniques. The analysis output 15 topics for each period.

According to the result, the 'tap water larvae' incident, as reported in the media, is divided into the occurrence, diffusion, and rectification stages. The government's response and civilian risk consciousness and reaction could also be seen. Based on the result, the following measures to respond to environment risk is proposed. First, it is necessary to explore the various intertwined context with the 'tap water larvae' incident at its core and develop responsiveness to environmental problems through education which forms integrated views. Second, a role to monitor the environment must be implemented and civilian-participated environmental information must be shared through the application of internet communities. Third, the cultivation and deployment of environmental communicators who provide and communicate fast and accurate environment information is required.

This study, as the first in Korea to use the topic modeling analysis method based on big data related to 'tap water larvae', has academic significance in that it has empirically and systematically analyzed environmental issues which appear as unstructured data. It also political significance as it suggests ways to improve environmental education and communication.

■ keyword : | Tap Water | Larva | News Big Data | Topic Modeling Analysis | Event-centered Environmental Study |

접수일자 : 2020년 09월 21일

수정일자 : 2020년 10월 28일

심사완료일 : 2020년 11월 04일

교신저자 : 김태종, e-mail : k2boy3@naver.com

I. 연구목적 및 필요성

본 연구의 목적은 토픽 모델링 분석기법을 활용하여 '수돗물 유충' 관련 국내 뉴스 빅데이터를 분석함으로써, 뉴스 빅데이터에 나타난 '수돗물 유충' 관련 주요 키워드를 도출하고, 이를 바탕으로 사회적으로 이슈화되고 있는 수돗물 유충 관련 의제를 파악하는 것이다.

2020년 7월 13일 인천 서구 일대에서 수돗물에 유충이 나왔다는 민원이 최초 보도된 이후[1], 16일이 경과한 7월 29일까지 전국에서 접수된 '수돗물 유충' 관련 민원 신고는 2,318건이며, 인천 외에도 서울, 부산, 경기도 등 전국에서 민원신고가 접수됐다[2].

인천 지역 수돗물에서 발견된 유충은 파리목의 깔따구과(Chironomidae sp.)로서, 1~2mm 정도의 철사모양 지렁이 또는 붉은 색을 띤 실지렁이와 유사한 형태이며, 저수지, 강, 바닷물 등에서 서식하고, 염소(Chlorine)에 대한 저항성이 강해 잔류염소 50mg/L에서 48시간 처리 후에도 생존하는 것으로 알려져 있다[3]. 아직 깔따구 유충의 경우 유해성에 대한 국내 임상 결과는 나타나지 않았다[4].

'수돗물 유충' 민원이 급증함에 따라, 환경부는 7월 17일 전국 정수장배수지 등에 대한 위생상태 긴급점검을 요청하고, 전국 지방자치단체는 전수 조사에 착수했다[3]. 7월 20일 환경부장관은 '수돗물 유충' 사태에 대해 "국민께 불편을 드리고 우려를 낳게 해 주무장관으로서 막중한 책임감을 느낀다"고 공식 사과했다[5].

이러한 정부와 지방자치단체의 조치에도 불구하고, 건강과 직결되는 수질개선에 대한 공감대가 빠르게 확산되었다. 그리고 충분하지 않은 환경정보 제공으로 인한 불안으로, 수돗물 유충에 대한 민원과 청원이 전국적으로 증가함은 물론, 생수 및 정수기·샤워기 필터 판매량도 급증한 상황이다[6].

정치권에서도 "수돗물은 국민생활에 가장 필수적인 공공재로, 국민들이 믿고 마시며 사용할 수 있는 수돗물 공급과 관리는 정부의 기본 책무"라며, "국민들이 불편과 불안을 겪지 않도록 최선을 다하겠다"면서 정부의 신속한 대책을 촉구했다[7].

이처럼 환경 관련 뉴스는 사회적으로 이슈화되고 있는 환경문제와 관련된 다양한 기관과 사람들의 목소리

를 보도하면서, 사회문제를 제기하고 각자의 입장을 제시하며 논쟁의 과정을 거쳐 해결책을 마련하는 일종의 공론장(公論場) 역할을 수행한다.

특히, 뉴스는 '수돗물 유충'과 같이 국민의 건강과 밀접한 관련이 있는 사안을 중요한 의제로 설정함으로써, 문제를 부각시키고 국민의 어려움과 의견을 대변하며 전문가의 원인분석과 해결방안을 제시하여 정책에 반영시키는 기능을 한다.

따라서, 국내 최초로 발생한 '수돗물 유충'과 관련된 사안을 해결하기 위해서는 먼저, 국민, 정치인, 전문가, 정부 관계자 등의 목소리가 총체적으로 담겨있는 뉴스를 통해 사회적으로 구성된 의제가 무엇인지 파악하는 것이 필요하다. 이를 파악하는 데에는 대량의 뉴스 빅데이터를 대상으로, 텍스트 내에 잠재되어 있는 의제(토픽)를 파악하는 토픽 모델링 분석방법이 유용하게 활용될 수 있다.

현재까지 수돗물과 관련된 뉴스를 분석한 국내 연구(KCI 등재지 기준)로는 김성태·이창호(2007)의 '공적 이슈에 대한 미디어 보도 분석(수돗물 관련 기사를 중심으로)'[8]과 홍영식·이덕로(2019)의 '인천 상수도 사태의 이슈 특성과 향후 정책 변동 가능성 연구(매스미디어 분석을 중심으로)'[9] 외에는 발견되지 않았으며, 토픽 모델링 분석방법을 활용한 연구는 아직 진행되지 않았다.

수돗물 관련 연구동향을 분석한 연구에 따르면, 기존 선행연구 805건의 주요 연구주제로, 상수도 수질, 불소농도 조정사업, 잔류물 영향, 미생물 관리, 시스템 구축, 정수처리 방법, 영향평가, 시민들의 인식, 인체에 미치는 영향 등이 다루어졌으나[10], 수돗물에서 발생한 '유충'을 주제로 한 연구는 존재하지 않았다.

따라서, 본 연구에서는 국내에서 최초로 발생한 '수돗물 유충'을 주제로, 관련 뉴스 빅데이터를 토픽 모델링 기법을 활용하여 사회적으로 구성된 의제를 분석하기 위해 다음과 같이 연구문제를 설정하였다.

〈연구문제1〉 '수돗물 유충' 뉴스 빅데이터에 나타난 시기별 주요 키워드는 무엇인가?

〈연구문제2〉 '수돗물 유충' 뉴스 빅데이터에 나타난 시기별 주요 토픽은 무엇인가?

II. 선행연구 검토

1. 수돗물 관련 환경문제 선행연구 고찰

우리나라 상수 이용은 1950년대부터 시작하여 전국 평균 상수도 보급률이 '18년 기준 97%에 이르며[11], 물 사용의 편리성과 생활수준의 향상에 기여하고 있다. 그러나 1990년 금강광역상수도의 트리할로메탄(THM) 수질 기준 초과 사건을 시작으로, 1991년 낙동강 폐놀 유출 사건 등 수질 안전성에 대한 경각심이 촉발되었다. 이로 인해 건설부가 담당하던 상수도 업무와 보건사회부가 담당하던 음용수관리 업무를 환경처로 이관하였으며, 물관리 기능을 강화하고 일원화한다는 취지에서 1994년 환경처는 환경부로 격상되었다.

또한 국민의 환경에 대한 사회적 관심이 급속하게 확대되고 환경문제를 해결하기 위한 보다 체계적이고 근본적인 대책이 요구되자, 정부는 제6차 교육과정(1992년)에 환경을 독립과목으로 포함하였다. 이러한 영향으로 국내에 수돗물 관리(상수도 수질, 불소농도 조정사업, 정수처리 방법, 영향평가, 미생물 관리, 시스템 구축 등) 연구들이 다양한 주제로 활발히 진행되었다[10]. 최근 류재나 외(2018)는 '물환경 서비스와 물 인프라의 지속가능성 평가 연구'[12]에서 수질 서비스 품질의 형평성 확립이 필요함을 지적하고, 물 인프라(경제, 사회, 환경부문)의 지속가능성 평가 의무화를 통한 개선 및 이행체계 적립의 필요성을 제시하였다.

수돗물 환경 사건의 쟁점을 다룬 서이중(2001), 성하영(2007) 연구에서는 1997년부터 2002년까지 수돗물 바이러스 오염 여부에 대한 과학계와 정부[13], 환경시민단체 등 각 입장에 따른 논쟁 형성과정[14]을 다루었으나, 환경 위기 대응 논의는 다루지 않았다.

수돗물과 관련된 뉴스를 분석한 연구로는 김성태·이창호(2007)의 '공적 이슈에 대한 미디어 보도 분석(수돗물 관련 기사를 중심으로)'[8]과 홍영식·이덕로(2019)의 '인천 상수도 사태의 이슈 특성과 향후 정책 변동 가능성 연구(매스미디어 분석을 중심으로)'[9] 등이 진행되었으나, 그 밖의 수돗물 관련 뉴스 분석 연구는 발견되지 않았으며, 분석방법에서도 내용분석과 의미연결망 분석이 활용되었으며, 토픽 모델링 분석방법을 활용한 연구는 발견되지 않았다.

2. 사회적 이슈 연구방법 고찰

사회적 이슈를 분석하는 데에는 연구목적에 따라 다양한 연구방법을 활용할 수 있다. 최근 사회적 이슈를 분석하는 데에 활용된 연구방법을 파악하기 위해, 2020년에 발행된 KCI 등재지를 대상으로 '이슈 분석' 관련 논문을 검색한 결과, 총 32편의 논문이 수집되었다.

사회적 이슈를 분석하는 데에 활용된 연구방법으로는 [표 1]과 같이 토픽모델링, 의미연결망, 문헌연구, 내용분석, 질적연구, 설문조사 등이 활용되었다. 특히, 토픽모델링과 의미연결망 분석은 뉴스, 보도자료, SNS, 인터넷 게시판 등 대량의 텍스트로 구성된 빅데이터를 분석하는 데에 유용하여 최근 이슈 분석에 많이 활용되고 있다.

표 1. 최근 주요 사회적 이슈별 연구방법

주요 사회적 이슈	연구방법	연구 수
미세먼지, 성범죄, 개인정보보호법, 마스터카드, 농촌 태양광, 4차산업혁명, 인공지능, 태권도	토픽모델링	8
공수처, 정신질환, 자사고 재지정, 사회적경제, 사회적가치, 약물투여, 블록체인, 스마트시티, 산림치유	의미연결망	9
데이터3법, 빅데이터, 사물인터넷 보안, 제로에너지, 스포츠경기	문헌연구	5
조국사태, 공유가치, 북한	내용분석	3
티베이크 포르노, 민주시민교육, 초등학습분석	질적연구	3
한일관계	설문조사	1
코로나19, 4차산업혁명, 사이버범죄	기타 (빈도,군집 등)	3
계		32

본 연구에서는, '수돗물 유출'과 관련된 뉴스에서 나타난 시기별 의제(토픽)를 파악하는 것이 연구목적므로 토픽 모델링 분석방법을 활용하였다. 특히, 토픽모델링 분석방법은 전통적인 뉴스 내용분석 방법의 한계인 코더 간 및 코더 내 신뢰도 문제, 대용량 데이터 분석 제한을 극복하는 데에 유용하다[15].

따라서, 본 연구에서는 뉴스에 잠재되어 있는 의제(토픽)를 도출하는 연구에 많이 사용되고 있는 토픽 모델링 기법인 LDA (Latent Dirichlet Allocation, 잠재 디리클레 할당) 기법을 적용하였다[16]. LDA기법은 텍스트 내 키워드들을 출현비율에 따라 토픽에 할당하는

분석기법이다[17][18].

III. 연구 방법

1. 자료 수집

분석 대상 언론매체는 한국언론진흥재단에 기사를 제공하는 총 54개 매체이며, 세부 현황은 [표 2]와 같다.

표 2. 분석 대상 언론매체 현황

유형	매체수	언론매체
중앙지	11	경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보
경제지	8	매일경제,머니투데이, 서울경제, 아시아경제, 아주경제, 파이낸셜뉴스, 한국경제, 헤럴드경제
지역지	28	강원도민일보, 강원일보, 경기일보, 경남도민일보, 경남신문, 경상일보, 경인일보, 광주매일신문, 광주일보, 국제신문, 대구일보, 대전일보, 매일신문, 무등일보, 부산일보, 영남일보, 울산매일, 전남일보, 전북도민일보, 전북일보, 제민일보, 충도일보, 중부매일, 중부일보, 충북일보, 충청일보, 충청투데이, 한라일보
전문지	2	디지털타임스, 전자신문
방송사	5	KBS, MBC, OBS, SBS, YTN

* 자료 출처: 빅카인즈 홈페이지(www.bigkinds.or.kr)에서 재구성

분석기간은 '수돗물 유출' 민원이 최초 보도된 2020년 7월 13일부터 관련 보도빈도가 크게 하락한 8월 31일까지 총 29일이다. 보도빈도가 [그림 1]과 같이 3개 시점을 기준으로 변화함에 따라 3개 시기로 구분하여 분석했다. 제1기는 7월 13일부터 18일까지, 제2기는 7월 19일부터 26일까지, 제3기는 7월 27일부터 8월 31일까지다.

뉴스 빅데이터 수집을 위한 검색어는 '수돗물 유출'으로 설정하였다. 수집 방법은 뉴스 빅데이터 분석 서비

스를 제공하고 있는 빅카인즈(BIGKinds)를 활용하였다.

분석 대상 뉴스 건수는 [표 3]과 같이 총 1,975건으로, 최초 수집된 2,326건에서 중복, 일정한내, 인물동정 등의 뉴스 351건을 제외하였다. 본 연구에서는 빅카인즈를 통해 비정형 데이터의 전처리와 형태소 추출이 완료된 데이터를 활용하여 분석을 수행했다.

표 3. 분석 대상 뉴스 건수(건)

구분	제1기 (7.13~7.18)	제2기 (7.19~7.26)	제3기 (7.27~8.31)	계
최초 수집 뉴스	378	1,342	606	2,326
제외 뉴스	37	233	81	351
분석 대상 뉴스	341	1,109	525	1,975

2. 분석 방법

본 연구에서 활용한 분석프로그램은 NetMiner 4이며, 뉴스 텍스트에서 추출된 키워드를 행렬 매트릭스로 구성하여 LDA분석($\alpha=0.1, \beta=0.01$)을 실시하였다[19].

토픽의 개수를 정하는 데에는 연구자가 사전에 정하는 방법과 데이터를 기반으로 정하는 방법이 있다[13]. 본 연구에서는 먼저 연구자가 사전에 토픽의 개수를 2개에서 30개로 정한 후, 샘플링 반복횟수를 1,000회로 설정하여 토픽 개수별로 중복되는 키워드의 수를 비교하였다. 이 데이터를 기반으로, 중복되는 키워드 수가 가장 적고 범주별로 해석이 용이하게 분류된 토픽의 개수를 판단한 결과, 적절한 토픽의 개수는 5개로 나타났다.

이를 바탕으로 토픽 모델링을 실시하여 토픽별로 출현비율이 높은 순으로 키워드를 도출한 후 다음과 같은 절차를 수행하였다. 첫째, 도출된 키워드와 뉴스 원문을 비교하여 어떤 맥락에서 키워드가 사용되었는지 파악

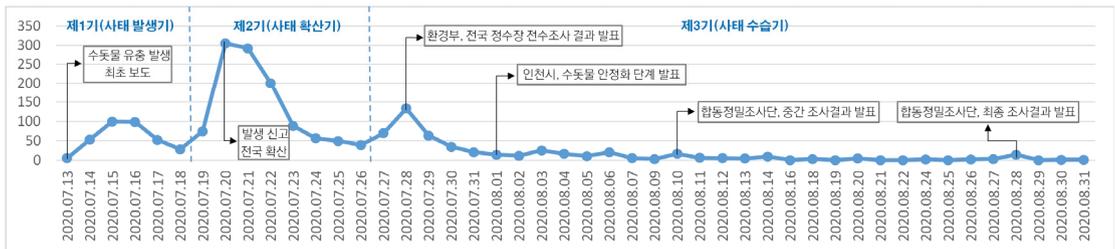


그림 1. 일자별 '수돗물 유출' 관련 보도 추이

* 자료 출처: 빅카인즈 홈페이지(www.bigkinds.or.kr)에서 재구성

정체 확인'으로 부여했으며, 관련 뉴스건수는 61건(17.9%)이다.

[토픽4]는 '신고', '불안', '서구', '벌레', '확산' 등의 키워드가 도출됨에 따라, 수돗물에서 유출을 발견한 시민들의 신고와 불안감이 확산되는 보도내용을 고려하여, 토픽명을 '수돗물 유출 불안 확산'으로 부여했으며, 관련 뉴스건수는 92건(27.1%)이다.

[토픽5]는 '청원', '임신', '아내', '깎따구', '유입' 등의 키워드가 도출됨에 따라, 수돗물 유출 관련 민원이 발생하는 보도내용을 고려하여, 토픽명을 '수돗물 유출 청원 발생'으로 부여했으며, 관련 뉴스건수는 56건(16.4%)이다.

[토픽1]은 '점검', '긴급', '정수장', '시설', '현장' 등의 키워드가 도출됨에 따라, 전국 지자체가 관내 정수장을 긴급 점검하는 보도내용을 고려하여, 토픽명을 '지자체 정수장 실태 점검'으로 부여했으며, 관련 뉴스건수는 196건(17.7%)이다.

[토픽2]는 '신고', '발견', '인천', '부산', '아파트' 등의 키워드가 도출됨에 따라, 전국에서 수돗물 유출 발견 신고가 접수되는 보도내용을 고려하여, 토픽명을 '수돗물 유출 신고 전국 확산'으로 부여했으며, 관련 뉴스건수는 316건(28.5%)이다.

[토픽3]은 '인천', '필터', '민원', '보상', '가정' 등의 키워

2. 제2기 '사태 확산기' 단계 분석 결과

제2기 2020년 7월 19일부터 26일까지 보도된 수돗물 유출 관련 뉴스 1,109건을 분석한 결과, 주요 키워드 간의 네트워크는 [그림 3], 토픽은 [표 5]와 같이 나타났다.

2기에서는 정수장-발견-신고-전국-확산-긴급 등의 키워드가 공동 출현하면서, 환경부를 중심으로 정수장 실태점검 노력과 국민들의 인식된 위험의 확산과 대응이 이슈로 부각되어, '사태 확산기'로 명명하였다.

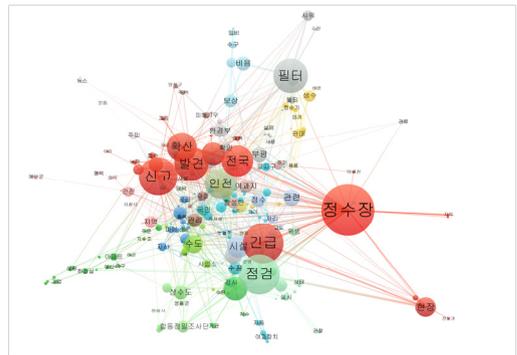


그림 3. 제2기 주요 키워드 네트워크

표 5. 제2기 토픽 및 주요 키워드

토픽명	주요 키워드	출현비율	주요 뉴스 제목	뉴스건수(건)	토픽 비중(%)
[토픽1] 지자체 정수장 실태 점검	점검	0.123	의왕시장, 수돗물 위생관련 청계정수장 [점검]	196	17.7
	긴급	0.081	충주시장, 수돗물 위생관련 단월정수장 [긴급] 점검		
	정수장	0.074	광주시, 광주 [정수장] 수도시설 긴급점검		
	시설	0.036	파주시장 문산정수장 등 수도공급 [시설] 긴급점검		
	현장	0.033	과천시, 관내 정수시설 [현장] 점검 실시		
[토픽2] 수돗물 유출 신고 전국 확산	신고	0.14	성남 운동등 한 아파트 단지서 '수돗물 유출' 의심 [신고]	316	28.5
	발견	0.088	청주서 [발견]된 유출 분석 의뢰 흥덕구 아파트서 신고 접수		
	인천	0.048	[인천] 수돗물 유출 10일동안 계속 발견 인천 자치단체에서도 의심신고		
	부산	0.035	[부산] 수돗물 유출 의심신고 총 77건 ...물탱크 등 유입 추정		
	아파트	0.035	파주 [아파트] 수돗물에서도 유출 발견 신고 접수 후 관계기관 조사 중		
[토픽3] 수돗물 유출 피해 보상 논란	인천	0.105	[인천] 수돗물 유출 '보상도' 논란 '유출' 확인 가구만 지원?	156	14.1
	필터	0.028	인천 수돗물 유출, 실제 벌레 확인된 가정만 [필터]비용 보상한다		
	민원	0.027	인천 수돗물 유출 [민원] 717건 21일에만 91건		
	보상	0.027	"수도요금이라도 깎아야" 인천서 '유출' [보상] 가이드라인 논란		
	가정	0.023	인천시, 유출 나온 [가정]만 필터비용 지원 생수구매하는 제외		
[토픽4] 전국 정수장 전수조사 진행	정수장	0.118	환경부 "인천 공촌외 6개 [정수장]서도 유출 발견" 여과지 문제	283	25.5
	인천	0.082	[인천] 울산 경남 정수장에서도 '수돗물 유출' 발견 환경부 이번주 전수조사 완료		
	발견	0.07	'수돗물 유출' 전국서 [발견] 이어져... 환경부 전국 정수처리장 전수조사		
	전국	0.063	환경부 [전국] 정수처리장 435개소 조사 이번 주 지나야 보완조치 마무리 될 듯		
	국무총리	0.048	수돗물 유출 사태 확산 정세균 [국무총리], 정수장 긴급점검 지시		
[토픽5] 수돗물 유출 불안 확산 및 주민 대응	필터	0.082	"수돗물 못 믿겠다" 불안감 확산...사위[필터] 생수 매출 급증	158	14.2
	불안	0.067	'수돗물 [불안]감' 확산에 사위기 필터 생수 판매 '몰타'		
	확산	0.043	울산 학부모 "급식실 수돗물은 어찌나" 유출 사태 [확산]에 불안 호소		
	생수	0.037	수돗물 유출 공포 일주일 [생수] 2000개 주문한 아파트		
	사위기	0.027	수돗물 유출에 놀란 사람들, 마트 달려가 [사위기] 필터 샀다		
계				1,109	100

드가 도출됨에 따라, 수도물에서 유충이 확인된 가구만 피해를 보상한다는 인천시 지침과 논란에 대한 보도내용을 고려하여, 토픽명을 '수도물 유충 피해 보상 논란'으로 부여했으며, 관련 뉴스건수는 156건(14.1%)이다.

[토픽4]는 '정수장', '인천', '발견', '전국', '국무총리' 등의 키워드가 도출됨에 따라, 국무총리 지시로 환경부가 전국 정수장을 점검하는 보도내용을 고려하여, 토픽명을 '전국 정수장 전수조사 진행'으로 부여했으며, 관련 뉴스건수는 283건(25.5%)이다.

[토픽5]는 '필터', '불안', '확산', '생수', '샤워기' 등의 키워드가 도출됨에 따라, 시민들의 불안이 확산되고 생수나 샤워기 필터를 구매하는 보도내용을 고려하여, 토픽명을 '수도물 유충 불안 확산 및 주민 대응'으로 부여했으며, 관련 뉴스건수는 158건(14.2%)이다.

3. 제3기 '사태 수습기' 단계 분석 결과

제3기 2020년 7월 27일부터 8월 31일까지 보도된 수도물 유충 관련 뉴스 525건을 분석한 결과, 주요 키워드 간의 네트워크는 [그림 4], 토픽은 [표 6]과 같이 나타났다.

3기에서는 상수도-시설-점검-결과-관리-부실 등의 키워드가 공동 출현하면서, 수도물 유충 원인규명과 환

경부 전국 정수장 전수조사 결과가 이상 없음에도 유충이 추가 발견되고, 국민들의 알 권리를 위한 환경정보 제공 및 교육과 수도물에 대한 신뢰 회복이 필요함을 보여주는 이슈가 부각되어 '사태 수습기'로 명명하였다.

[토픽1]은 '발견', '외부', '조사', '전국', '신고' 등의 키워드가 도출됨에 따라, 환경부가 전국 정수장을 대상으로 조사한 결과를 발표한 보도내용을 고려하여, 토픽명을 '전국 정수장 전수조사 결과 발표'로 부여했으며, 관련 뉴스건수는 121건(23.0%)이다.

[토픽2]는 '인천', '사태', '관리', '활성탄', '부실' 등의 키워드가 도출됨에 따라, 인천시 수도물 유충 사태의

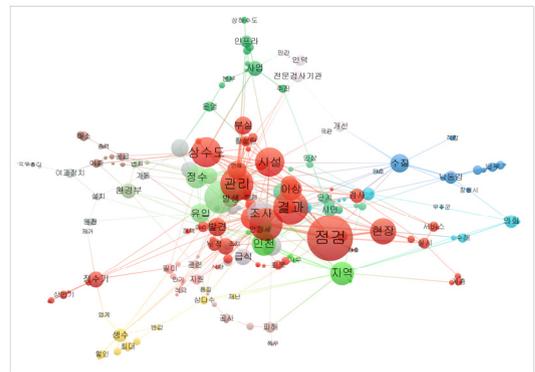


그림 4. 제3기 주요 키워드 네트워크

표 6. 제3기 토픽 및 주요 키워드

토픽명	주요 키워드	출현비율	주요 뉴스 제목	뉴스건수(건)	토픽 비중(%)
[토픽1] 전국 정수장 전국 전수조사 결과 발표	발견	0.065	환경부 "전국 일반정수장 435곳 전수 조사 유충 [발견] 못 해"	121	23.0
	외부	0.056	서울 수도물은 이상 無 유충 발견 신고는 [외부] 요인에 의한 것		
	조사	0.045	환경부 "전국 정수장 전수 [조사] 결과, 유충 발견되지 않아"		
	전국	0.042	[전국] 정수장 435곳 중 3곳만 여과지에 유충 발견 수도물엔 없어		
	신고	0.042	"수도물 유충, 정수 과정서 발생된 것 아냐" 서울시, [신고]된 73건 조사 결과 발표		
[토픽2] 인천 물 관리 부실	인천	0.116	[인천]시 수도물 유충사태의 원인은 '정수장 관리부실'로 최종 결론	92	17.5
	사태	0.054	인천 수도물 유충 [사태] 원인 "유충신란 쉬운데 밀폐 부실"		
	관리	0.046	인천시 물[관리]가 문제다, 붉은 수도물 이후 개선 약속 안 지켜 유충 사태도 정수장 부실 원인		
	활성탄	0.038	인천 수도물 유충 발생 원인은 '인재' 정수장 [활성탄]지 관리부실		
	부실	0.029	합동조사단 "인천 유충 수도물은 정수기 관리[부실] 탓"		
[토픽3] 지자체별 상수도 안전 발표	점검	0.09	부여군, 상수도 주요 시설물 [점검] 안전	141	26.9
	긴급	0.043	동두천시 정수장 배수지 [긴급]점검 결과 '양호'		
	인천	0.042	오산시 수도물 [인천] 시 배수지 5곳 긴급점검		
	상수도	0.041	"세종[상수도] 유충논란 이상무" 정수장 4곳 긴급점검		
	현장	0.032	옥천군, 정수장 3개소 [현장] 특별점검 수도물 유충 '안전'		
[토픽4] 수도물 유충 부대 효과	정수기	0.03	코로나에 수도물 유충까지 [정수기] 관련 제품 판매 급증	77	14.7
	낙동강	0.023	창원시장 "낙동강" 수질개선 안되면 물값 납부 거부"		
	생수	0.023	수도물 사태에 [생수] 수요 급증 편의점 업계, 생수 판매 박차		
	필터	0.023	수도물 유충 사태에 '소나기 주문' 샤워기 [필터] 공장 풀가동		
	수질	0.017	낙동강 [수질], 경남지역 의제로 급부상		
[토픽5] 수도물 유충 추가 발견	인천	0.078	[인천] 수도물 유충 2건 추가 누적 254건	94	17.9
	발견	0.054	강릉 연곡정수장 여과지서 유충 [발견]		
	추가	0.038	인천 '수도물 유충' 5건 [추가] 발견		
	시민	0.034	인천 수도물 비상, 적수 이어 유충까지... 지자체는 [시민]들		
	확인	0.021	인천 수도물 갈따구 유충 2건 추가 [확인]		
계				525	100

원인을 규명한 결과 정수장 관리 부실로 나타났다는 보도내용을 고려하여, 토픽명을 '인천 물 관리 부실'로 부여했으며, 관련 뉴스건수는 92건(17.5%)이다.

[토픽3]은 '점검', '긴급', '안전', '상수도', '현장' 등의 키워드가 도출됨에 따라, 각 지자체의 긴급 점검 보도내용을 고려하여, 토픽명을 '지자체별 상수도 안전 발표'로 부여했으며, 관련 뉴스건수는 141건(26.9%)이다.

[토픽4]는 '정수기', '낙동강', '생수', '필터', '수질' 등의 키워드가 도출됨에 따라, 수돗물 유출 사태로 인한 정수기필터·생수 판매량 증가, 낙동강 수질 개선 요구 등의 보도내용을 고려하여, 토픽명을 '수돗물 유출 부대 효과'로 부여했으며, 관련 뉴스건수는 77건(14.7%)이다.

[토픽5]는 '인천', '발견', '추가', '시민', '확인' 등의 키워드가 도출됨에 따라, 인천 등 각 지역에서 수돗물 유출이 지속적으로 발견된다는 보도내용을 고려하여, 토픽명을 '수돗물 유출 추가 발견'으로 부여했으며, 관련 뉴스건수는 94건(17.9%)이다.

V. 요약 및 제언

본 연구는 '수돗물 유출' 사태의 뉴스 빅데이터를 분석함으로써, '수돗물 유출'과 관련된 주요 키워드와 토픽을 파악하여, 환경 문제에 대한 위기 대응력 제고 방안을 제안하기 위해 수행되었다. 이를 위해 2020년 7월 13일부터 8월 31일까지 보도된 1,975건의 뉴스를 3개 시기로 구분하여 토픽 모델링 기법으로 분석하였으며, 시기별로 다음과 같은 토픽이 나타났다.

제1기(사태 발생기)에는 ①수돗물 유출 문제 감지 ②인천 수돗물 유출 논란 ③수돗물 유출 정체 확인 ④수돗물 유출 불안 확산 ⑤수돗물 유출 청원 발생, 제2기(사태 확산기)에는 ①지자체 정수장 실태 점검 ②수돗물 유출 신고 전국 확산 ③수돗물 유출 피해 보상 논란 ④전국 정수장 전수조사 진행 ⑤수돗물 유출 불안 확산 및 주민 대응, 제3기(사태 수습기)에는 ①전국 정수장 전수조사 결과 발표 ②인천 물 관리 부실 ③지자체별 상수도 안전 발표 ④수돗물 유출 부대 효과 ⑤수돗물 유출 추가 발견 토픽이 도출되었다.

이상과 같이 국내 최초로 발생한 '수돗물 유출' 사태

토픽들을 통해 다음과 같은 시사점을 발견할 수 있다.

첫째, 수돗물 유출 사건을 중심으로 얽혀있는 다양한 맥락을 탐구하고 통합적인 안목을 형성하는 교육으로 환경문제에 대한 대응력을 기를 수 있도록 해야 한다. 수돗물 유출 사태의 전 시기에서 나타난 토픽들은 수질 환경 문제를 둘러싼 정부, 지자체, 국민들의 위기 상황을 경험하는 과정으로 나타난다. 수돗물 유출사태는 우리나라 수자원 관리의 축적된 지식에서 벗어난 전례없는 사건이다. 국민들은 일상생활 속에서 경험하거나, 건강과 직결되는 내용에 대한 관심과 앎의 욕구가 높다. 따라서 환경교육으로 지속가능한 삶의 토대를 만드는 수자원 이용에 대한 '알 권리'가 충족되어야 하고, 환경적인 실천을 수반해야 한다. 제1기의 '논란', '불안'과 관련된 키워드와 뉴스를 통해 볼 수 있듯이, 시민들은 과거 발생한 인천 붉은 물, 낙동강 폐놀 유출사건 등으로 수질 안정성에 대한 불신감이 누적되어 있다. 제한된 정보로 인한 대중의 공포를 줄이기 위해 생활 환경교육 내용으로 정수장으로부터 집으로 공급되는 물의 정수 과정, 정수장의 운영과 정수처리 방법 등을 다룰 수 있다. 제2기와 제3기의 '실태점검', '관리 부실' 등의 토픽은 우리나라 수질 관리의 시스템의 보완 필요성을 나타내며, 중앙과 지자체의 체계적인 관리와 기후위기 등 변화하는 환경에서 지속가능한 수자원 관리를 위한 예방 차원의 역할에 대한 반성적 성찰이 필요하다. 이를 종합하여 환경에 대한 통합적 안목과 문제해결력을 기르기 위해 이재영(2016)이 제안한 '사건 중심 환경 탐구'를 환경교육 방법으로 활용할 수 있다. '사건 중심 환경 탐구'는 사건을 사유의 단위로, 사건에 얽혀있는 다양한 관계를 조망하여 환경문제에 대한 통합적인 탐구를 제안한다[20]. 수돗물 유출의 발생과 확산, 정부와 시민들의 대응 과정에 대한 심층 탐구로 통합적인 안목을 형성하여 향후 환경문제로 야기된 위기에 대응력을 기를 수 있다.

둘째, 인터넷 커뮤니티를 활용한 시민참여의 환경정보 공유와 환경감시 역할 부여가 필요하다. 수돗물 유출 사태의 '청원 발생', '주민 대응' 등의 토픽에서 볼 수 있듯이, 시민들은 환경 이슈를 직접적으로 경험하면서 국가를 상대로 문제를 제기하고, 지역 커뮤니티에서 의견을 공유하며, 주체적으로 문제에 대응하는 모습을 보

었다. 특히 수돗물 유출 발생 후 일부 지자체와 관계자들의 느장 대응으로 인터넷 맘카페 등 지역 커뮤니티를 활용해 유출 발생 관련 현황을 알리는 글이 빠르게 확산되었다. 언론에서도 인터넷 맘카페 정보를 활용한 기사가 총 120건으로 전체 뉴스의 약 6% 비중을 차지하였으며, 유튜브에는 수돗물 유출과 관련하여 시민들이 제작한 콘텐츠가 132개 등록되었다. 주로 수질 측정, 정수기능 실험, 정수장 현장, 샤워기 필터 제작 또는 제품 비교 등 시민들의 수돗물 유출 관련 유의미한 정보 생산과 공유, 대응 노력 등을 확인 할 수 있었다. 그러므로 환경문제가 발생했을 때 더 적극적으로 시민 환경감시 역할을 부여할 수 있다. 이를 위해 수돗물 유출 발견, 신고 등의 구체적인 가이드를 제공하여 신속한 현황을 진단하고 대책을 수립할 수 있으며, 오인 신고를 줄이는데 기여할 수 있다.

셋째, 환경문제 발생 초기에 신속하고 정확한 정보 제공과 소통을 담당하는 환경 커뮤니케이터의 양성 및 배치가 필요하다. 수돗물 유출 사태의 '논란', '불안 확산' 등의 토픽에서 볼 수 있듯이, 환경문제 발생시 신속하고 정확한 정보 제공과 국민과의 긴밀한 소통은 논란과 불안을 최소화하고 위험을 정확하게 인식시켜 효과적인 대응을 할 수 있다. 즉, 수돗물 유출 제1기인 사태 발생기를 살펴보면, 유출 발견 신고 후 2~3일만에 청와대 국민청원이 접수되고, 1만 1천명 이상이 동의하며 국민들의 불안과 분노에 대한 내용이 초기에 집중적으로 나타난다. 이는 수돗물 유출(깔따구)의 객관적인 유해성이 밝혀지지 않은 상황과 비교하여 국민들에게 '인식된 위험'의 영향에 주목할 필요가 있다. 뉴스 미디어에서 얻어지는 정보의 신뢰도가 타 정보원보다 높은 점[21]을 고려할 때, 국민들이 언론에서 제공하는 타당한 정보를 바탕으로 위험에 대한 객관적인 인식을 형성하는 것이 중요함을 알 수 있다. 따라서, 사태 발생 초기부터 환경 정보를 쉽게 전달하고, 진행상황에 대한 투명한 공유를 담당하는 환경 커뮤니케이션 관련 전문가의 양성 및 배치로, 심리적 공포감을 낮추고 사태를 직면하여 적절한 대응방안을 모색하는데 기여할 수 있다.

본 연구는 '수돗물 유출' 관련 뉴스 빅데이터를 기반으로 국내에서 처음으로 토픽 모델링 분석기법을 활용

하여 분석한 연구로써, 비정형 데이터로 나타나는 환경 관련 이슈를 실증적이고 체계적으로 분석했다는 학술적 의의가 있다. 또한, 사건을 중심으로 한 환경교육, 온라인 커뮤니티를 활용한 시민 감시 역할, 전문적인 환경 커뮤니케이터의 양성 및 배치의 필요성을 제시했다는 정책적 의의를 지닌다.

본 연구의 제한 및 제언사항은 다음과 같다.

첫째, 언론에서 보도하는 뉴스만을 분석 대상으로 삼아, 정부-미디어-시민으로 이어지는 환경 커뮤니케이션의 전체적인 순환 과정을 담아내는 데에는 제한이 있었다. 추후 정부의 보도자료, 시민들의 SNS 콘텐츠에 담겨 있는 환경 관련 비정형 데이터를 포함하여 분석을 수행한다면 더욱 의미있는 함의를 도출할 수 있을 것이다.

둘째, 본 연구는 수돗물 유출이 최초 신고된 2020년 7월 13일부터 합동정밀조사단의 조사결과 발표 후 보도빈도가 감소한 8월 31일까지를 분석기간으로 삼았다. 그러나 2020년 10월 23일 현재까지도 제주도에서 수돗물 유출 사태가 발생해 국정감사에서 논의되는 등 아직도 수돗물 유출사태가 진행 중임을 고려할 때, 분석기간을 더욱 확대하여 분석한다면 수돗물 유출 사태에서 나타난 의제를 더욱 단계적이고 심층적으로 분석하여 정책적 함의를 제시할 수 있을 것이다.

셋째, 본 연구에서는 토픽 모델링 분석 방법을 통해 토픽을 구성하는 주요 키워드를 출현비율에 따라 군집화하여 의제를 파악했으나, 토픽별 주요 키워드 간에 연결된 의미관계를 파악하는 데에는 한계가 존재한다. 추후 연구에서는 주요 키워드 간의 네트워크를 파악하는 의미연결망 분석을 병행하여 활용한다면 토픽별 의미를 더욱 정교하게 도출할 것으로 예상된다.

넷째, 본 연구에서는 토픽의 수를 정할 때 토픽이 얼마나 정확하게 분류되는 지를 파악하는 복잡도(perplexity) 등의 지수를 활용하지 않고, 연구자가 복수의 토픽 수를 기준으로 도출된 결과를 바탕으로 해석가능성과 토픽 분류의 적합성을 고려하여 토픽의 수를 정하였다. 추후 연구에서는 복잡도 지수를 적용하여 분석을 수행한다면 좀 더 실증적인 토픽 분석 결과가 나타날 것이다.

참고 문헌

[1] 김주영, “‘붉은 수돗물’ 터진 인천, 이번엔 ‘유출 보인다.’” 세계일보, 2020.7.13.

[2] 환경부, “인천 외 지역 유출 발생 민원 312건은 외부 유입사례,” 환경부 보도자료, 2020.7.30.

[3] 환경부, “전국 정수장배수지 청결 등 위생상태 긴급 점검,” 환경부 보도자료, 2020.7.17.

[4] 백순영, “‘수돗물 유출’ 사태 전국으로… 원인과 유해성은?,” YTN, 2020.7.21.

[5] 천권필, “환경부 장관 ‘수돗물 유출, 국민 눈높이에서 수질기준 초과’ 사과,” 중앙일보, 2020.7.22.

[6] 배재성, “수돗물 유출’ 불안 전국 확산, 물 관리 혁신으로 신뢰 높여야,” KBS, 2020.7.22.

[7] 김미경, “김종인 ‘수돗물 유출, 솔직히 창피스러운 일’, 이해찬 ‘수돗물 정상화 시급,’” 디지털타임스, 2020.7.22.

[8] 김성태, 이창호, “공적 이슈에 대한 미디어 보도 분석 (수돗물 관련 기사를 중심으로),” 한국언론정보학보, 제39호, pp.40-69, 2007.

[9] 홍영식, 이덕로, “인천 상수도 사태의 이슈 특성과 향후 정책 변동 가능성 연구(매스미디어 분석을 중심으로),” 한국지방공기업학회보, 제15권, 제2호, pp.23-48, 2019.

[10] 김지운, 도윤호, 주기재, 김은희, 박은영, 이상협, 백명수, “수돗물 이용에 대한 국내 연구동향과 사회적 인식,” 생태와 환경, 제49권, 제3호, pp.208-214, 2016.

[11] 환경부, 상수도 통계, 2018.

[12] 류재나, 황상일, 김호정, 이병국, 이영경, 정아영, *물 환경 서비스와 물 인프라의 지속가능성 평가(V)*, 환경정책평가연구원, 2018.

[13] 성하영, “수돗물 바이러스 논쟁(바이러스의 실재와 위험 인식의 구성),” 과학기술연구, 제7권, 제2호, pp.125-154, 2007.

[14] 서이중, “환경 문제의 ‘과학기술과 사회’적 형성과정 (수돗물 바이러스 논쟁을 중심으로),” 환경사회학연구 ECO, 통권, 제1호, pp.64-91, 2001.

[15] 황용석, 이준용, 손동영, 김상민, 유경한, 최세정, 김익현, 김대원, 박령주, 백영민, 이상엽, 성운택, *데이터 시대의 언론학 연구*, 커뮤니케이션북스, 2017.

[16] 김태중, “뉴스 빅데이터를 활용한 코로나19 언론보도 분석,” 한국콘텐츠학회논문지, 제20권, 제5호, pp.457-466, 2020.

[17] 정원준, 김대욱, 윤호영, 이형민, 박진우, 김동성, 손영곤, 전홍식, 천용석, 정유미, 박종구, *빅데이터의 분*

석방법과 활용, 학지사, 2020.

[18] D. M. Blei and J. D. Lafferty, “Topic Models,” Text Mining: Classification, clustering, and application, Vol.10, No.71, pp.1-34, 2009.

[19] W. Zhao, J. Chen, and W. Zen, “Best Practices in Building Topic Models with LDA for Mining Regulatory Textual Documents,” CDER 9Th November, 2015.

[20] 이재영, *사건 중심 환경 탐구*, 공주대학교 출판부, 2016.

[21] 양성관, 양성운, “뉴스 미디어 신뢰도가 뉴스 수용자의 위기지각에 미치는 영향(위기에 관한 수용자의 관여도 조절효과를 중심으로),” 한국언론학보, 제47권, 제6호, pp.279-305, 2003.

저자 소개

이수연(Su Yeon Lee)

정회원



- 2013년 2월 : 공주대학교 환경교육과(이학사), 교육학과(교육학사)
- 2016년 8월 : 공주대학교 교육학과(교육학석사)
- 2013년 3월 ~ 2017년 4월 : 한국환경교육연구소 연구원
- 2019년 5월 ~ 2020년 9월 : 한국청소년정책연구원 전문연구원
- 2020년 10월 ~ 현재 : 한국환경교육연구소 객원연구원 (관심분야) : 환경교육, 생태전환교육, 환경 프로젝트 학습, 사건 중심 환경 탐구, 청소년교육, 평생교육

김태종(Tae-Jong Kim)

정회원



- 1999년 2월 : 경북대학교 신문방송학과(문학사)
- 2012년 2월 : 연세대학교 언론학과(언론학석사)
- 2019년 8월 : 공주대학교 교육학과(교육학박사)
- 2020년 4월 ~ 현재 : 한국청소년정책연구원 부연구위원 (관심분야) : 환경정책, 보건정책, 청소년정책, 미디어정책, 평생교육, 트렌드 분석, 뉴스 빅데이터 분석, 토픽 모델링 분석, 텍스트 네트워크 분석