

- 논문(Full paper) 12페이지 초과 시 탈락
- 영문과 국문 양식 중에 선택하여 작성
- 홈페이지 內 응모요강 > 작성가이드도 참조

# 미디어 빅 데이터 분석을 통한 청와대 국민청원의 여론 반영 분석

청와대 국민청원은 “국민이 물으면 정부가 답한다”는 국정 철학 하에 실시된 대국민 소통의 장이다. SNS로 로그인만 하면 누구나 청원을 등록할 수 있고, 청원 분야에는 제한이 없다. 여기에는 ‘국정 현안에 국민 여론을 반영한다’는 현 정부의 취지가 담겨있으며 누구나 청원을 올릴 수 있고, 청원 분야에는 제한이 없다. 등록된 청원이 30일 이내에 20만 명 이상의 동의를 얻을 시 정부 또는 청와대 관계자가 영상으로 답하는 방식으로 진행되고 있다. 국민이 정책을 제안하고 의견을 내는 직접민주주의의 의의에는 의심의 여지가 없으나, 표면적인 동의자수가 20만 명만 넘으면 청원이 채택된다는 점에서 여론 반영에 대한 객관적인 연구가 필요함을 느꼈다. 이에 현재 청원 시스템과 미디어 사이의 관계를 파악하였다. 미디어와의 관계를 통해 청와대 청원이 여론을 반영 할 수 있는지 분석하였다. 연구 결과, 청와대 청원은 저널리즘적 성격을 지니고 있어 여론을 형성하여 모든 청원이 여론을 반영한다고는 할 수 없다는 것을 확인하였다.

## 1. 서론

### 1.1. 연구 동기

청와대 국민청원은 “국민이 물으면 정부가 답한다.”는 국정 철학 하에 실시된 대국민 소통의 장이다. 2017년 8월 17일부터 시행되고 있으며 전체적인 시스템은 백악관의 국민청원 사이트인 ‘We The People’과 유사하다. SNS로 로그인만 하면 누구나 청원을 등록할 수 있고, 청원 분야에는 제한이 없다. 등록된 청원이 30일 이내에 20만 명 이상의 추천을 얻을 시 정부 및 청와대 관계자가 영상으로 답하게 된다. 현재까지 53개 청원에 대한 답변이 청와대 국민청원 홈페이지에 올라와있고, 3개의 청원이 답변 대기 중에 있다 (2018년 10월 23일 기준).

2008년부터 운영되고 있는 국민신문고에 비해 청원 처리가 더 빠르게 이루어지고 있어 활발히 이용되고 있다. 국민이 정책을 제안하고 의견을 내는 직접 민주주의의 새로운 가치를 창출해내고 있다는 점에서 주목할 만하다.

본 연구는 청원의 동의자수가 여론을 얼마나 반영하는지에 대한 물음에서 출발하였다. 현재 청원 시스템에서는 오로지 동의한 사람의 수만 표시되며, 청원 반대 여론의 규모와는 상관없이 표면적인 동의자수가 20만 명만 넘으면 청원이 채택된다. 다시 말해, 아무리 청원 반대 여론의 규모가 크다고 할지라도 이들은 청원 채택 여부에 영향을 주지 못한다. 현재 청원 시스템으로는 청원이 잠재 여론을 포함한 모든 여론을 정확하게 대변한다고 할 수 없다. 국내외의 선행 연구를 조사해본 결과, 청와대 국민청원, We The People 등 청원 사이트의 여론 반영을 분석한 연구는 없었다. ‘국정 현안에 국민 여론을 반영한다.’는 청와대 국민청원의 취지 달성을 위해서 여론 반영에 대한 객관적인 연구가 필요하다고 생각하였고, 본 연구를 진행하게 되었다.

본 연구의 목적은 청원 시스템이 여론을 반영할 수 있는가를 확인하는 것이다. 미디어 및 소셜 빅 데이터 분석에 관한 많은 선행 연구에 따르면, 인터넷 내 특정 키워드 언급량과 기사에 달린 댓글, SNS 게시물 등의 데이터를 분석한 결과가 실질적인 여론과 유사한 경향을 보인다. [7] 따라서 잠재 여론 분석을 위해서는 먼저 청원 시스템과 미디어의 관계 파악이 필요하다라고 생각하였다. 이에 Python을 통해 인터넷 기사와 관련 트윗을 스크레이핑하여 특정 키워드 언급량 데이터를 얻었다. 또한, 댓글을 감성 분석하여 청원에 대한 국민의 감정을 객관적으로 파악하였다. 본 연구에서는 이를 청와대 청원의 동의자수 변화 추이와 비교하여 현재 청원 시스템과 미디어 사이의 관계를 파악하고, 최종적으로 청원 시스템이 여론을 반영할 수 있는가를 확인하고자 한다.

### 1.2. 이론적 배경

#### 1.2.1. 오피니언 마이닝(Opinion Mining)

오피니언 마이닝은 소비자의 감성과 관련된 텍스트 마이닝(Text Mining)의 기술 중 하나로, 텍스트에서 사람들의 주관적인 성향과 의견 등을 추출해내는 기술이다.[6] 오피니언 마이닝에는 감성 분석(Sentiment Analysis)라는 기술이 사용되는데, 빅 데이터 기술의 발전에 따라 핵심 기술로 자리 잡으면서 현재 상품, 도서, 영화평 등의 다양한 의견 텍스트의 감성을 분석하는데 사용되고 있다. 이러한 감성 분석은 데이터 수집, 주관성 탐지, 극성 탐지의 순서로 이루어진다. 트위터와 페이스북 등의 소셜 미디어 매체에서 정보를 수집하고 수집한 텍스트에서 주관성이 없다고 판단되는 부분(성별, 나이, 텍스트 저자)을 제외시킨다.[4] 극성 탐지 단계에서는 각 단어가 나타나는 빈도 또는 긍정, 부정과 같은 감성 속성에 따라 점수를 부여하여 텍스트 전체의 감성을 분석한다. 감성 분석

의 대표적 기법으로 머신러닝 기반의 문서 단위의 극성 분석과 감성어 사전 기반의 속성 단위의 분석 방법이 있다. 머신러닝의 경우 수작업으로 감성이 판단된 데이터를 제작하는 과정이 필요하다. 이 때문에 본 연구에서는 연구자의 주관적 판단이 들어갈 것을 고려하여 감성어 사전 기반의 감성 분석을 진행한다. Naïve Bayes와 SVM(Support Vector Machine)등의 알고리즘을 사용하면 감성 사전을 통해 감성 분석이 가능하다. 감성 분석 전에 감성어 사전의 구축이 선행되어야 하지만 이미 KOSAC(Korean Sentiment Analysis Corpus), Korean Twitter Emotion Analysis 등의 한글 감성어 사전이 구축되어 있다. 이러한 감성어 사전에는 기본형 단어들만 존재하기 때문에 형태소 분석과 같은 전처리 작업이 필요하다. Korean Twitter Emotion Analysis는 요즘 인터넷 사회에서 등장하는 욕설이나 신조어까지 감성어 사전이 구축되어 있지만 KOSAC은 그렇지 않다. 따라서 본 연구에서는 인터넷 기사 댓글에서 등장하는 신조어와 욕설에 대한 감성 분석을 진행할 수 있는 Korean Twitter Emotion Analysis를 사용하였다. [12]

### 1.2.2. 여론(Public Opinion)

‘여론(Public Opinion)’이라는 단어는 1588년 미셸 드 몽테뉴에 의해 처음 사용되었다. [10] 그 이후 여론에 대한 여러 정의가 등장하였다. 다만, 여론은 추상적인 개념이고 주관적인 특성과 대상이 모호하다는 특성 때문에 정의하기 어려운 점이 있다. 본 연구에서는 브루머와 헤네시가 정의 내린 여론의 개념을 이용하였다. 브루머에 따르면 여론이란 “이슈를 가진 공중들에 의해 형성된 공중의 의견”이다. [9] 또한, 헤네시는 “중요성을 지닌 쟁점에 대해 상당히 많은 사람들이 표명한 여러 가지 선호 의견들”이라고 여론을 정의하였다. [11]

2012 12월 10일에 있었던 18대 대선 2차 TV 토론회가 방영되는 동안 생성된 트윗을 텍스트 마이닝 하여 분석한 결과 ‘고소득층’, ‘지하경제 활성화’, ‘자유토론’, ‘순환출자’ 등 국민의 관심이 반영된 트윗이 생성되었다고 한다. 또한, SNS 여론분석 전문회사인 ‘PR&Korea’에서 대선공식선거운동 기간인 2012년 11월 27일부터 12월 18일까지 리트윗을 제외하여 각 후보별 빅데이터 버즈량(SNS상 연관 키워드 언급량)을 분석한 결과를 보면 박근혜 후보와 관련된 버즈량은 164만 8264건, 문재인 후보와 관련된 버즈량은 164만 6717건이었다는 것을 알 수 있다. 이는 실제로 18대 대통령 대선 결과인 박근혜 후보가 51.6%의 지지율로 당선된 결과와 매우 유사함을 알 수 있다. 즉, 트위터를 통해 브루머와 헤네시가 정의한 여론을 파악할 수 있음을 알 수 있다. 따라서 본 연구에서는 여론을 파악하기 위해 트위터를 표본 집단으로 사용하였다. [8]

## 2. 연구 방법 및 절차

### 2.1. 연구 대상

2018년 8월 15일부터 9월 14일까지, 30일 동안 날마다 동의자수 기준 상위 1000개에 해당하는 청원을 대상으로 스크레이핑을 진행하였다. 스크레이핑을 통해 청원 번호, 청원 분류(카테고리), 청원 제목, 청원 기간(청원 시작일), 그리고 청원 동의자수의 정보를 수집하였다.

언급량 분석 및 오피니언 마이닝을 위해 관련 인터넷 뉴스 기사와 트윗을 추가로 스크레이핑하였다. 한 사건이 매체를 통해 알려지고, 또 이것이 청원으로 연결되기까지의 기간을 감안하여 청원 시작일 5일 전부터 종료일까지 날마다 스크레이핑하였다. HTML을 통해 파싱하는 스크레이핑의 원리를 고려하여 네이버를 포맷으로 하는 인터넷 뉴스 기사를 대상으로 하였다. 인터넷 뉴스 기사에서는 기사 제목, 기사 게시일, 언론사, 기사 본문, 그리고 기사에 달린 댓글을, 트위터에서는 청원 관련 키워드가 포함된 트윗을 모두 스크레이핑하였다.

본 연구에서는 스크레이핑을 통해 얻은 자료 중 인터넷 뉴스 기사의 제목과 본문, 그리고 트윗으로부터 언급량의 데이터를 얻었고, 이를 통해 여론 형성의 정도를 파악하였다. 특히 인터넷 뉴스 기사 내 키워드 언급량 데이터를 통해 청원의 공론화 정도를 파악하였다. 또한, 각 청원에 대한 인터넷 뉴스 기사 댓글을 감성 분석하여  $\gamma$ 라는 척도를 얻었는데, 이에 대한 추가적인 설명과 활용 방안은 뒤에서 서술하도록 한다.

### 2.2. 연구 과정

#### 2.2.1. 청원 스크레이핑

먼저 정해진 기간(2018년 8월 15일부터 2018년 9월 14일까지)동안 날마다 동의자수 기준 상위 1000개에 해당하는 청원을 대상으로 여러 정보를 수집하였고, 이 중에서 분석을 진행할 청원을 추리는 작업을 진행하였다. 스크레이핑 기간 내에 만료되는 청원 중에서 동의자수가 10만 명 이상인 청원, 그리고 스크레이핑 종료일인 9월 14일에 동의자수가 1만 명 이상인 청원을 선정 기준으로 하였고, 이 과정에서 총 24개의 청원이 선정되었다. 이후 관련 인터넷 뉴스 기사와 트윗 스크레이핑을 진행하기 위해 청원의 제목 및 내용, 뉴스 자료 등을 참고하여 4명의 연구자 모두 이의가 없는 키워드를 추출하였다. 그 결과는 Table 1과 같다. Table 1은 수집한 청원의 제목, 청원 시작 날짜, 그리고 편의를 위해 각 청원에 부여한 분류코드와 키워드를 정리한 표이다.

**Table 1. Blue House Petition Case**

Case	Petition Title	Starting Date	Keyword
A	불법폭력조직 코마트레이드와 연루된 성남시장 은수미와 경기도지사 이재명 즉각 사퇴하	2018.07.22.	코마트레이드

라			
B	웹하드 카르텔과 디지털성범죄 산업에 대해 특별 수사를 요구한다.	2018.07.29	웹하드 카르텔
C	여성가족부의 예산 사용 실적을 공개를 촉구합니다.	2018.08.15.	여성가족부 예산
D	강릉 고양이 토막사건	2018.08.17	고양이 토막
E	"성상품화"로 가득찬 군대위문 공연을 폐지해주세요	2018.08.17.	성상품화 위문
F	540마리 사육곰 문제, 정부에 적극적인 대책 마련을 요구합니다.	2018.08.17	사육 곰
G	여성가족부폐지를 대한민국정부에게 강력히 촉구합니다!!	2018.08.18.	여성가족부 폐지
H	인천 여중생 자살 가해자 강력 처벌 희망 요망	2018.08.19.	인천 여중생 자살
I	엘리트과의 소송에 제출된 법무 답변서 관련 의혹에 대해 진상조사와 감사를 청원합니다	2018.08.22	엘리트 소송
J	전국 학교에 휴교령을 내려주세요.	2018.08.22.	학교 휴교령
K	국민연금 주식대여 금지 국민청원	2018.08.24	국민연금
L	산채로 냉동고에 가둬 죽음에 이르게 한 청주시반려동물보호소 유기견 학대 사건을 조사해주세요.	2018.08.25.	청주시반려동물센터
M	보건복지부의 낙태수술 의사 즉각처벌권을 철회해주세요	2018.08.28	보건복지부 낙태수술
N	학교 몰카사건에 대한 피해자 보호와 가해자 처벌 강화를 요청합니다.	2018.8.29.	학교 몰카
O	유은혜 의원의 교육부장관 후보 지명 철회해 주세요.	2018.08.30.	유은혜 의원 장관
P	김대용 축구국제심판 박탈하라	2018.09.01	김대용 축구국제심판
Q	문재인 대통령님, 약속을 지켜주세요..	2018.09.03	문재인 약속
R	어린이집 원장이 도둑년입니까?	2018.09.04	어린이집 원장
S	8만 치과위생사, 노동의 권리를 보장하라.	2018.09.05.	치과위생사 노동
T	제 남편의 억울함을 풀어주세요!	2018.09.06	성추행 누명 억울
U	보배드림 성추행 누명 사건 판사 징계를 청원합니다.	2018.09.08.	보배드림 성추행
V	성소수자 혐오 집회에서 시작된 성소수자 혐오 폭력 집회, 더 이상 가만히 있어서는 안됩니다.	2018.09.08	성소수자 집회
W	저는 성추행범이 아님에도, 법원 확정 판결이 났습니다. 정말 억울합니다.	2018.09.09	성추행범 판결
X	사람의 인권을 사고파는 반인륜적 행위인 성매매를 강력 처벌해주세요	2018.09.10	성매매

2.2.2. 관련 인터넷 뉴스 기사 및 트윗 스크레이핑  
 각 청원에서 추출한 키워드를 가지고 관련 인터넷 뉴스 기사와 트윗을 스크레이핑하였다. 기사 제목과 본문, 그리고 트윗에서 특정 단어가 얼마나 언급되었는지 계산하여 언급량 데이터를 얻었고, 이와 별개로 기사의 댓글을 수집하는 코드를 제작하여 오피니언 마이닝에 사용할 데이터를 얻었다.

### 2.2.3. 감성 분석

인간의 감정을 6가지(분노, 혐오, 두려움, 행복, 슬픔, 놀람)로 분류한 Korean Twitter Emotion Analysis의 Emotion Lexicons를 기반으로 하여 스크레이핑한 청원 관련 인터넷 뉴스 기사의 댓글을 감성 분석하였다. 오피니언 마이닝의 과정을 거침으로써 각 청원 별로 감성 점수 데이터를 얻었다.

얻어진 감성 점수 데이터를 분석에 활용하기 위하여 본 연구에서는  $\gamma$ 라는 척도를 정의하였다.  $\gamma$ 는 '부정 감정 지수'로 정의할 수 있는데, 댓글이 해당 청원에 얼마나 부정적인 감정을 가지고 있는지를 나타내는 척도이다. Korean Twitter Emotion Analysis의 감성 분류 체계 중에서 부정적인 감정에 해당하는 분노, 혐오, 두려움, 슬픔의 감성 점수는 더하고, 긍정적인 감정에 해당하는 행복의 감성 점수는 빼서  $\gamma$ 를 계산하였다. 놀람은 중립적인 감정으로 판단하여 계산에서 제외하였다. 계산식은 다음과 같다.

$$\gamma = Anger + Disgust + Fear + Sadness - Happiness$$

### 2.2.4. 수집 데이터를 바탕으로 한 여론 반영 분석

청원과 인터넷 뉴스 기사, 그리고 트윗을 스크레이핑하여 수집한 청원 동의자수 변화와 언급량 수치 자료를 Pyplot을 통해 그래프로 나타내었다. 청원 동의자수 그래프(이하 Agree)를 각각 관련 인터넷 뉴스 기사 내 키워드 언급량 그래프(이하 NKF, News Keyword Frequency), 트위터 내 키워드 언급량 그래프(이하 Tweets)와 겹침으로써 각각의 데이터 사이의 관계를 알아보았다.

## 3. 연구 결과

### 3.1. Agree와 NKF의 비교분석

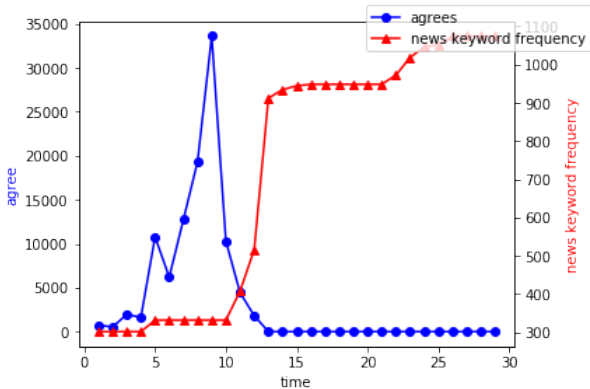
그래프 분석 결과를 살펴본 결과 그래프 상의 수치 변화가 두 가지 경향으로 나타난다는 것을 알 수 있다. 따라서 위의 분석 결과를 그래프 변화의 경향성을 기준으로 하여 청원을 두 가지 유형으로 분류하였고, 분류 결과는 Table 2와 같다. Table 2의 Type N1은 NKF의 피크가 Agree의 피크 이후에 따라오는 유형이다. 반면 Type N2는 관련 인터넷 뉴스 기사 내 특정 키워드 언급이 Agree의 피크 이전에도 꾸준히 발견되다가 Agree의 피크 이후에 NKF가 피크를 찍는 유형이다.

**Table 2. Two Types of Results (Relation between Agree and News Keyword Frequency)**

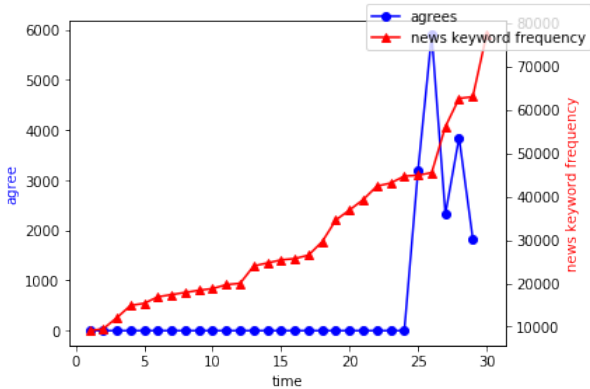
Type	Case
N1	A, B, D, F, H, J, S, U
N2	C, G, I, K, L, M, N, O, P, Q, R, T, V, W, X

분류 결과, 총 8개의 청원이 Type N1에, 15개의 청원이 Type N2에 해당하였다.

Fig 1은 Type N1으로 분류된 Case B(‘웹하드 디지털 성범죄’ 관련 청원), Fig 2는 Type N2로 분류된 Case X(‘성매매 처벌’ 관련 청원)가 보이는 Agree와 NKF의 그래프를 비교한 것이다. Fig 1, Fig 2에서 Agree는 파란색(원형 마커)으로, NKF는 빨간색(삼각형 마커)으로 표시하였다. 그래프의 가로축은 시간(일), 세로축은 키워드 언급량(개)을 의미한다.



**Fig 1. Graph of Agree and NKF in Type N1**



**Fig 2. Graph of Agree and NKF in Type N2**

### 3.2. Agree와 Tweets의 비교분석

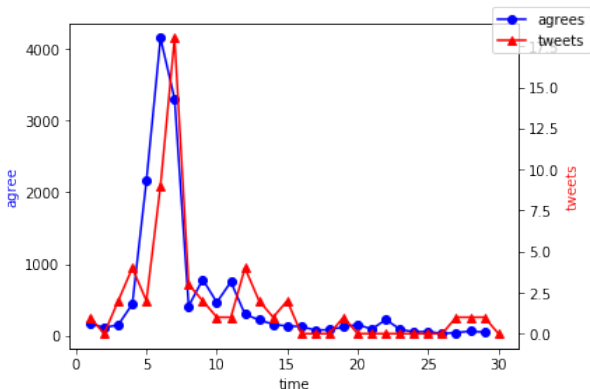
그래프 분석 결과 그래프 상의 수치 변화가 두 가지 경향으로 나타난다는 것을 알 수 있다. 위의 분석 결과를 그래프 변화의 경향성을 기준으로 하여 청원을 두 가지 유형으로 분류하였고, 분류 결과는 Table 3와 같다. Table 3의 Type T1은 Tweets의 피크가 Agree의 피크 이후에 따라오는 유형이다. 반면 Type T2는 트윗 내 키워드 언급이 Agree의 피크 이전에도 꾸준히 발견되다가 Agree 그래프의 피크 이후에 Tweets가 피크를 찍는 유형이다.

**Table 3. Two Types of Results (Relation between Agree and Tweets)**

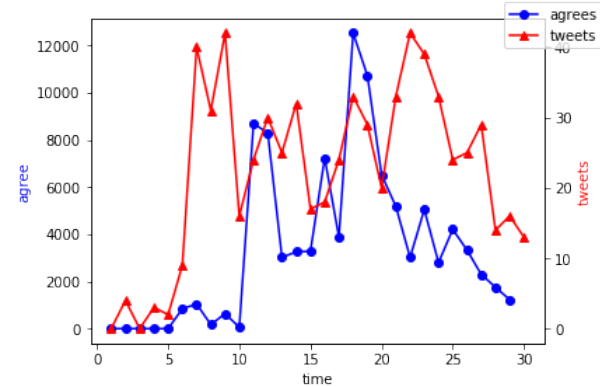
Type	Case
T1	B, D, E, F, H, J, L, O, P, S, T, U, V, W
T2	A, C, G, I, K, M, N, Q, R, X

분류 결과, 총 14개의 청원이 Type T1에, 10개의 청원이 Type T2에 해당하였다.

Fig 3은 Type T1으로 분류된 Case D(‘강릉고양이토막사건’ 관련 청원), Fig 4는 Type T2로 분류된 Case I(‘엘리엇 법무부 답변서’ 관련 청원)가 보이는 Agree와 Tweets를 비교한 것이다. Fig 3, Fig 4에서 Agree는 파란색(원형 마커)으로, NKF는 빨간색(삼각형 마커)으로 표시하였다. 그래프의 가로축은 시간(일), 세로축은 트윗수(개)을 의미한다.



**Fig 3. Graph of Agree and Tweets in Type T1**



**Fig 4. Graph of Agree and Tweets in Type T2**

### 3.3. 청원 관련 기사 댓글 감성 분석

각 청원 별로 6가지 감정에 대해 감성 분석을 진행하였고, 그 결과는 Table 4와 같다.

**Table 4. Relation between Petition Case and Emotion Score (Top Three and Bottom Three Values are Highlighted in Bold.)**

Case	Anger	Disgust	Fear	Happiness	Sadness	Surprise
A	2.661241	3.08568	1.180943	2.112604	2.234615	1.512811
B	3.595928	3.96724	1.65058	2.918553	3.302004	2.20732
C	3.490000	3.898500	1.281600	2.606700	2.796600	1.879100
D	2.681900	2.812600	1.083800	2.433300	2.568900	1.673200

E	3.287900	3.412500	1.533900	2.770300	2.971000	<b>2.395700</b>
F	4.548266	3.506015	1.498487	3.383937	<b>4.260066</b>	2.009714
G	3.310630	3.780302	1.224048	2.388109	2.517987	1.710839
H	3.167200	3.694000	1.231000	2.410000	2.897600	1.571500
I	3.503800	4.470600	1.425700	2.641400	2.763200	1.966000
J	2.122955	2.085251	0.883998	1.663268	1.986814	1.117968
K	3.986444	4.764295	1.388749	2.809165	2.849549	2.050786
L	2.189280	2.281794	1.000110	2.008651	2.048374	1.320458
M	3.391519	3.672687	1.349910	2.832540	3.166632	1.989006
N	2.960129	3.227061	1.476860	2.417974	2.801733	1.932903
O	3.135242	3.659654	1.211545	2.158868	2.252164	1.581725
P	2.574348	2.873426	1.407509	1.723200	2.082024	1.696652
Q	3.307675	3.979046	1.203120	2.444616	2.512916	1.812147
R	3.233408	3.264042	1.349408	2.637906	2.978165	1.887235
S	<b>4.616024</b>	<b>4.834544</b>	<b>1.833649</b>	<b>3.798045</b>	4.223539	2.090840
T	2.950920	3.261900	1.160143	2.336133	2.528942	1.668981
U	3.307391	3.614120	1.529735	2.552690	2.853595	2.023185
V	3.681620	3.866936	1.602447	3.090309	3.095191	2.038000
W	3.369267	3.685397	1.678812	2.614833	2.897774	2.008866
X	3.072668	3.416291	1.433892	2.484098	2.783315	1.894843

**Table 5. Relation between Petition Case and  $\gamma$**   
(Top Three and Bottom Three Values are Highlighted in Bold.)

Case	$\gamma$	Case	$\gamma$
A	7.049875	M	8.748208
B	9.597197	N	8.047809
C	8.860243	O	8.099737
D	<b>6.713995</b>	P	7.214107
E	8.435239	Q	8.558141
F	<b>10.428897</b>	R	8.187117
G	8.444858	S	<b>11.709711</b>
H	8.579953	T	7.565772
I	9.521993	U	8.752151
J	<b>5.415750</b>	V	9.155885
K	<b>10.179872</b>	W	9.016417
L	<b>5.510907</b>	X	8.222068

Table 4를 보면 Case S에서 가장 높은 Anger, Disgust, Fear, 그리고 Happiness를 나타냈다. Case F에서 가장 높은 Sadness를 나타냈고, Case E에서 가장 높은 Surprise를 나타냈다.

Table 5에서 부정적 감정의 포함 정도를 나타내는  $\gamma$ 의 경우, Case F, K, S에서 가장 높은  $\gamma$  수치를 보였고, Case D, J, L에서 가장 낮은  $\gamma$  수치를 보였다.

## 4. 결론

4.1. 청원과 미디어의 상관관계 및 여론 형성에 끼치는 영향

본 연구에서는 청원과 미디어 간의 상관관계를 알아보기 위해 그래프 상의 피크를 기준으로 Type N1과 Type N2, 그리고 Type T1과 Type T2의 형태로 분류하였다. 모든 유형에서 Agree의 피크 이후에 NKF와 Tweets의 증가 추세가 나타나는 것으로 보아 청원이 여론 형성에 영향을 준다는 점을 확인할 수 있었

다. 즉, 대부분의 청원에서는 청원이 시작된 뒤 이를 바탕으로 여론이 형성됨을 알 수 있다.

다만, Type N2와 Type T2를 보면 청원 전에 존재했던 여론이 청원을 발생시킨 반면에 Type N1과 Type T1의 경우 청원이 발생하면서 관련 여론이 형성된 것에서 차이가 있다. 이를 통해 미디어가 청원을 유도하고, 다시 청원이 미디어를 유도하는 상호 유도 관계에 있다는 것을 알 수 있다. 청원에는 사회적인 문제를 제기하는 유형도 있지만 개인적인 불만 제기로부터 시작되는 청원도 있다. Type N1을 보면, 청원이 시작된 이후로 개인의 문제가 공론화되고, 이것이 뉴스를 통해 확산되면서 결과적으로 여론이 형성된다고 볼 수 있다.

다만 Type N2의 사례와 같이 뉴스에서의 언급이 지속되고, 이러한 언급이 일정 수준에 도달한 후 청원이 시작되는 경우는 특별하게 다뤄야 할 필요가 있다. 이때는 이미 인터넷 뉴스 기사를 통해 공론화된 문제가 청원으로 발전하여 큰 규모의 여론이 형성된다고 설명하는 것이 더 타당하다. 하지만 이 경우에도 청원이 중요한 요인으로 작용한다. 인터넷 뉴스 기사를 통해 어느 정도 여론이 형성되었는데, 그 규모가 청원 이후에 더욱 증가하였다. 이러한 점에서 청원이 여론을 증폭하는 효과를 지니고 있다고 할 수 있다.

### 4.2. 청원 관련 댓글 감성 분석에 대한 고찰

감성 분석 결과, Table 5에서  $\gamma$ 를 기준으로 상위 3개와 하위 3개의 Agree, NKF, Tweets의 개형을 비교하였다. Case F, K, S는 가장 높은  $\gamma$  수치를 보였고 Case D, J, L은 가장 낮은  $\gamma$  수치를 보였다.  $\gamma$ 는 댓글이 얼마나 부정적인 감정을 가지고 있는지 상대적으로 나타내는 값이다. 이에  $\gamma$ 가 높을수록 격한 감정으로 인해 Agree, NKF, Tweets의 개형에 차이점이 있을 것이라고 예상했다. 연구 결과, Tweets의 경우,  $\gamma$ 에 따른 그래프 개형의 차이가 나타났다. Tweets의 유동은  $\gamma$ 가 높은 Case F, K에서  $\gamma$ 가 낮은 Case D, J, L보다 큰 폭으로 나타났다. 그러나 NKF의 경우 개형의 차이는 없는 것으로 확인되었다.

아직 감성 분석이 좋은 성능을 내지 못하고 본 연구에서는 정확도가 높지 않은 Bag of Words 방식을 이용했기 때문에 감성 분석에 대한 결과는 신중하게 재검토할 필요가 있다. 본 연구에서 이용한 Korean Twitter Emotion Analysis의 Emotion Lexicons의 경우 대략 60%의 성능을 보인다. [12] 따라서 댓글의 감성 분석이 정확하게 되지 않은 가능성도 배제할 수는 없다.

### 4.3. 연구의 시사점

국내의 다른 연구[2]에서는 인터넷 게시판 여론의 형성과정을 기준으로 여론의 형태를 분류했을 때 다음과 같이 3가지로 나눌 수 있다고 밝혔다. 첫 번째는 순수 인터넷 기반 여론으로 기존 언론에서 사건이 전혀 보도되지 않고 오로지 인터넷 자유 게시판에서



개인적 의견이나 호소, 고발 등이 사이버상의 의제로 설정되어 수많은 네티즌의 의견과 토론이 이루어지면서 여론화된 경우이다. 두 번째는 증폭형 인터넷 여론으로, 기존 언론에서 사건이 보도된 후 네티즌들이 인터넷 관련 사이트에 의견을 제시하고 토론하는 형태의 인터넷 여론이다. 세 번째는 시차형 인터넷 여론으로, 기존 언론에 사건이 보도되었으나 다른 사회적 이슈 등으로 여론이 매몰되었다가 상당한 기간이 지난 후 다시 인터넷 네티즌들을 중심으로 여론화되는 경우이다. [2]

지금까지 많은 국내 연구자들[1][3]에게 제기되었던 인터넷의 저널리즘적 특성들은 다음과 같다. 감정적이고 불확실한 정보전달과 이론화, 사회자가 없는 의제 설정, 의제설정의 탈공간성과 탈시간성, 의제설정의 무한성, 쌍방향 커뮤니케이션 체제, 사적 영역과 공적 영역이 불분명한 커뮤니케이션 공간, 익명적 커뮤니케이션 공간, 강력한 응집력, 정보의 무한 확장성과 무한 복제성 등이 있다.

본 연구의 결과를 이에 비교해 보았을 때, Type N1과 T1는 순수 인터넷 기반 여론에, Type N2와 T2는 증폭형 인터넷 여론에 대응된다. 즉, 청원이 순수 인터넷 기반 여론과 증폭형 인터넷 여론이라는 새로운 여론의 형성을 유도한다. 더 나아가, 청와대 청원은 인터넷의 저널리즘의 특성 중 쌍방향 커뮤니케이션의 특성을 제외한 나머지 특성들을 모두 갖추고 있다. 이것으로 미루어 보았을 때, 청와대 청원은 뉴스를 생산하고 전달하는 저널리즘적 성격을 가지고 있다고 할 수 있다. 따라서 청원은 여론 형성에 영향을 미치지만 모든 청원이 여론을 반영한다고 할 수 없다.

최근 ‘이수영 폭행 사건’이라는 청와대 국민청원 글에 35만명이 동의하면서(2018년 11월 18일 기준) 젠더 갈등으로 번지고 있다. 그러나 국민청원 글이 사실과 다른 부분이 있다는 점이 경찰 발표로 알려지면서 청와대 청원의 저널리즘적 성격이 불필요한 사회적 갈등을 불러 일으킬 수 있다는 점이 확인되었다. 현재 이런 청와대 청원의 저널리즘적 성격의 악영향을 차단할 제도가 존재하지 않는다. 따라서 청와대 청원은 잘못된 정보로 인한 불필요한 사회적 갈등을 막기 위한 제도적 보완이 필요하다.

## 참고문헌

- [1] 강미은, 인터넷 저널리즘과 여론, 커뮤니케이션북스, 2001
- [2] 박성호, 여론형성공간으로서 인터넷 자유 게시판의 저널리즘적 특성과 사회적 영향에 관한 연구, 언론과학연구, 5(3), 191-226, 2005
- [3] 박성호, 인터넷 미디어의 이해와 활용, 서울: 커뮤니케이션 북스, 2002
- [4] 신수정, “‘글에서 감정을 읽다’ 감성 분석의 이해”, IDG Tech Report, 2014
- [5] 오미영, 정인숙, 커뮤니케이션 핵심이론, 서울: 커뮤니케이션북스, 222-239, 2005

- [6] “오피니언 마이닝(Opinion Mining)”. 정보통신용어사전. 2012.
- [7] 장성복, 성열홍, "소셜 빅데이터를 통한 공연예술의 온라인 여론 분석", 한국디자인포럼, 57, 79-90, 2017
- [8] 한창진, 김경수, "TV 토론회에서 트위터가 선거에 미치는 영향", 한국디지털콘텐츠학회 논문지, 14(2), 207-214, 2013
- [9] Blumer. H, "The mass, the public, and public opinion. Reader in public opinion and communication", 2, 43-50, 1966
- [10] Braatz, Kurt, "Friedrich Nietzsche: eine Studie zur Theorie der Öffentlichen Meinung", Walter de Gruyter, 1988
- [11] Hennessy B. C, Public opinion, Wadsworth Publishing Company, 1970
- [12] Hyo Jin Do, Ho-Jin Choi, "Korean Twitter Emotion Classification Using Automatically Built Emotion Lexicons and Fine-Grained Features", 29th Pacific Asia Conference on Language, Information and Computation, 142-150, 2015
- [13] James Manyika, Michael Chui, "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, 2011
- [14] Laney, Douglas, "3D Data Management: Controlling Data Volume, Velocity and Variety", Gartner, 2001
- [15] Snijders. C, Matzat. U, Reips U. D, "Big Data.: Big gaps of knowledge in the field of Internet", International Journal of Internet Science, 1-5. 2012.