

국가 정책에 대한 언론과 SNS 반응의 감성 분석 연구 -아동 수당, 출산 장려금 정책을 중심으로-

윤혜민¹, 최은정^{2*}

¹서울여자대학교 정보보호학과 학부생, ²서울여자대학교 정보보호학과 부교수

A Study on Sentiment Analysis of Media and SNS response to National Policy: focusing on policy of Child allowance, Childbirth grant

Hye Min Yun¹, Eun Jung Choi^{2*}

¹Undergraduate Student, Department of Information Security, Seoul Women's University

²Associate Professor, Department of Information Security, Seoul Women's University

요 약 스마트폰, 태블릿 등의 이동 통신 기기와 PC 이용이 확장됨에 따라 인터넷 상에서 데이터가 기하급수적으로 수집되고 있다. 또한 SNS의 발전으로 인해 이용자 간의 자유로운 의사소통과 여러 분야의 정보를 공유할 수 있어 다양한 다량의 의견들이 빅데이터 형태로 쌓이고 있다. 이에 따라 빅데이터 분석 기법을 사용하여 일반 사람들의 반응과 언론사의 뉴스 기사 반응의 차이를 알아보는 기법이 대두되고 있다. 본 논문에서는 아동 수당과 출산 장려금에 대해 SNS에서 나타난 대중들의 반응과 언론사의 반응을 분석하였다. 이를 위해 일정 기간 동안 트위터에 올라온 이용자들의 글을 수집하고 뉴스 기사를 크롤링하여 감성 분석을 진행하였다. 이를 통해 SNS에 나타나는 대중의 의견과 언론사 뉴스의 반응을 비교하여 대중과 언론이 국가 정책에 대한 반응의 차이를 비교 분석하였다.

주제어 : 감성분석, 빅데이터, 소셜네트워크분석, 웹크롤링, 미디어 분석

Abstract Nowadays as the use of mobile communication devices such as smart phones and tablets and the use of Computer is expanded, data is being collected exponentially on the Internet. In addition, due to the development of SNS, users can freely communicate with each other and share information in various fields, so various opinions are accumulated in the from of big data. Accordingly, big data analysis techniques are being used to find out the difference between the response of the general public and the response of the media.

In this paper, we analyzed the public response in SNS about child allowance and childbirth grant and analyzed the response of the media. Therefore we gathered articles and comments of users which were posted on Twitter for a certain period of time and crawling the news articles and applied sentiment analysis. From these data, we compared the opinion of the public posted on SNS with the response of the media expressed in news articles. As a result, we found that there is a different response to some national policy between the public and the media.

Key Words : Sentiment Analysis, Big Data, SNS Analysis, Web Crawling, Media Analysis

This work was supported by a research grant from Seoul Women's University(2018).

*Corresponding Author : Eun Jung Choi(choi@swu.ac.kr)

Received December 31, 2018

Accepted February 20, 2019

Revised January 24, 2019

Published February 28, 2019

1. 서론

1.1 연구 배경

최근 스마트 폰이 보급되고 인터넷이 발달함에 따라 사람들은 인터넷 상에서의 사회적 관계망을 형성하는데 큰 관심을 가지고 있다. 특히 SNS가 점차 발달함에 따라 사용자 수가 증가하는 추세이고 사용자 간 관계를 쌓아 하나의 커뮤니티를 생성하며 다양하게 나타나고 있다. 이러한 SNS는 상호작용을 증가시켜 여론을 형성하고 확산하는데 큰 영향을 미치고 있다. 하지만 이렇게 형성된 여론은 특정 시점에 형성된 여론이고 계속적으로 변화하기 때문에 연속성 파악이 힘들고 대표성을 보장하는 것이 어렵다[1]. SNS를 통해 여론을 형성한 케이스는 19대 대통령 선거를 꼽을 수 있다. 선거 중, 국정원에서 SNS에 특정 후보를 지지하거나 비판하는 글을 올리는 행위를 조직적으로 진행한 것으로 판단하여 재판을 진행하였다[2,3]. 또한 일간 베스트(일베)가 고 노무현 대통령을 비하하고 세월호 희생자들을 모욕적으로 표현하는 등 만행을 저질렀다. 이러한 사례를 통해 SNS에서 표현의 자유를 어느 선까지 허용해야 하는지 정해야 하는 필요성을 느낄 수 있다[4].

포털 뉴스의 경우 여러 언론사들이 자신들의 홈페이지가 아닌 네이버나 다음 등과 같은 플랫폼에 기사를 업로드 하는 것을 의미한다. 네이버는 하루에 약 3000 만 명의 이용자와 75%의 검색 점유율로 한국의 대부분의 뉴스를 올리며 여론 형성에 막대한 영향을 끼치고 있기 때문이다. 하지만 이처럼 소비자들은 포털을 언론으로 생각하는데 이에 대한 규제가 거의 없고 이에 따라 위험성이 높아질 것이다[5]. 또한 네이버와 다음 등의 포털 여론영향력이 KBS와 조선일보 등의 언론사들을 앞서는 것으로 나타났다. 2015년 여론 집중도 조사 결과에 따르면 TV·라디오·인터넷 등 전체 뉴스 미디어를 합산한 '뉴스 이용창구 기준 여론영향력 점유율'은 네이버가 18%로 가장 많았다. KBS 계열사들은 17.0%로 2위, 조선일보 계열사들은 8.9%로 3위를 기록했다. 다음으로 동아일보 계열(7.6%), MBC 계열과 포털 다음(7.3%), SBS 계열(6.1%) 등의 순이었다. 인터넷 사이트 이용 점유율은 네이버와 다음이 78%를 차지하였다. 반면 조선닷컴은 약 2%, 동아닷컴은 약 1%, 조인스닷컴, 연합뉴스와 매경닷컴은 약 0.7% 정도로 아주 낮은 비율을 차지하였다. 더불어 포털을 거치지 않고 언론사 사이트에 직접 접속하는 방문자

는 전체의 22.7%에 불과한 것으로 나타났고 네이버를 경유하는 방문자의 비율은 45%, 네이버 검색 서비스로 들어가는 방문자는 약 25%였다[6]. 이와 같은 조사로 우리는 언론사 홈페이지에 직접 올라온 것이나 방송사에 업로드된 것을 선택한 것이 아니라 포털 이용자들이 가장 많이 사용하는 네이버에 올라온 뉴스 기사를 이용하였다.

본 논문에서는 SNS와 언론이 국가의 특정 정책에 대해 어떤 반응을 보여주고 있는지 분석하고자 한다.

1.2 연구 목적

본 논문에서는 요즘 가장 떠오르는 이슈인 아동 수당과 출산 장려금에 대해 SNS 상에서 나타난 여론형성 케이스와 언론의 여론 형성 케이스를 감성 분석 결과에 따라 비교하는 것을 목적으로 한다. 구체적으로 첫째, SNS 중 특히 트위터에서 위 2가지 사안에 대해 어떤 여론을 형성하는지 확인한다. 둘째, 포털 뉴스의 제목을 통해 언론이 어떤 방향으로 여론을 형성하는지 알아본다. 마지막으로 두 가지 데이터를 감성 분석하여 어떤 결과가 나오는지 알아본다.

2. 분석방법

2.1 분석대상

먼저 SNS(Social Network Service)는 이용자 간 자유로운 의사소통, 다양한 정보 공유, 사람 사이 관계 확대 등을 통해 사회적으로 관계를 생성하고 강화하는 온라인 플랫폼으로 이를 통해 정보가 공유 및 유통 될 때 의미가 생긴다. SNS는 웹을 기반으로 한 서비스이고 이메일이나 메신저 등을 통해 이용자들 간에 연락할 수 있는 수단을 제공한다. 이는 블로그, UCC 등처럼 소셜 미디어의 한 유형으로 생각해야 한다. 현 시대에는 스마트폰 이용자의 기하급수적인 증가와 무선 네트워크 서비스의 활성화와 함께 SNS의 이용자도 역시 급증하고 있다. SNS 분야를 주도하고 있는 페이스북과 트위터 등의 이용자 수는 이미 2018년에 18억 명을 넘겼고, 증가 추세는 당분간 계속 진행될 것으로 보인다. 기존에 오프라인 상에서만 알고 있었던 인맥 관계를 더욱 돈독하게 해주는 역할을 충분히 수행하며 온라인 커뮤니티를 통해서 새로운 인맥을 쌓을 수 있다는 점에서 SNS는 큰 장점을 가지고 있다. 또한 광범위하고 특정 집단으로 분류될 수 있는 서비

스 이용자들을 데이터베이스에 의해 분류하고 관리할 수 있다는 점에서 마케팅으로 활용할 수 있는 장점이 있다. SNS를 사용하는 기업의 입장에서는 적은 비용으로 해당 집단에 효율적으로 타겟팅할 수 있는 맞춤형 마케팅을 진행할 수 있기 때문에 유용한 방법이다. SNS 업체도 페이지를 광고에 활용하거나 게임 등을 통해 수익을 얻을 수 있어 앞으로도 SNS 시장은 꾸준히 성장해 나갈 것이 다[7].

본 논문에서는 SNS 중 다양한 사용자들이 존재하는 트위터를 통해 아동 수당과 출산 장려금에 대해 분석 해 보았다. 트위터는 영어를 포함하여 다국어를 지원하며, 이용자가 어디에 있는 상관없이 다른 사용자와 함께 짧은 글로 대화를 주고받거나 친구를 맺을 수 있다. 이용자 수가 많고, 실시간 대화와 비슷한 방식으로 이야기를 나눌 수 있기 때문에, 한 사람이 작성한 글이나 루머가 빠르게 퍼져나가 세계적인 이슈가 되는 일도 있다[8].

장문으로 본인의 생각을 표현하는데 적합한 블로그와 달리, 트위터는 간단한 글을 손쉽게 쓸 수 있는 단문 전용 SNS이기 때문에, 휴대폰이나 태블릿 등을 이용한 글 등록도 편리하다. 또한 다른 이용자의 최근 활동을 알게 해주는 '팔로우'라는 기능이 있고, 메신저와 같은 신속성을 갖춘 점이 특징이다. 다른 SNS와는 달리 자신을 팔로우하지 않은 상대방이 허락하지 않아도 일방적으로 '팔로어'로 등록할 수 있다. 미국에서는 버락 오바마 전 대통령이 대통령 선거에서 승리하는 데 트위터를 이용한 홍보 효과를 많이 얻은 것으로 알려졌고 기업들도 역시 고객의 니즈나 불만 사항 등을 빠르게 파악하여 처리할 수 있다는 장점이 있다[9].

다음으로 언론의 경우 네이버에 올라온 뉴스를 이용하였다. 그 까닭은 국내 PC 사용자 방문자 수 1위를 차지한 네이버의 2016년 10월 검색 점유율은 67%로 카카오(21%) 보다 많고 라인의 월간 실 사용자 수가 2억1700만 명으로 4900만 명인 카카오 톡보다 4배 이상 많다. 또한 국내 검색 시장을 장악하면서 작년 광고 매출이 전국의 신문 인쇄 광고 매출 보다 2배가 넘게 육박한 것으로 추정되고 있어 네이버가 계속적으로 주도권을 잡고 있다 [10]. 하지만 네이버가 개편한 이후 언론사들이 구독자 확보 경쟁을 해야만 했다. 네이버가 시작 화면에서 뉴스를 제외하고 언론사 구독 서비스를 내세워, 누르면 언론사 웹사이트로 이동할 수 있게끔 하였고 채널 구독자 등을 바탕으로 수익을 배분하여 신경 쓸 수밖에 없도록 만

들었다. 이처럼 포털이 언론사를 조종하고 있다는 것을 알 수 있었다. 따라서 포털과 언론이 서로 공생할 수 있도록 정책을 만들어야 하며 포털이 뉴스를 노출하는 알고리즘이 각 언론사의 신뢰성, 보도 주제 범위 등을 고려해서 각 주제에 대해 높은 신뢰성을 가지고 있는 언론사의 기사를 선별하여 올려야 한다[11].

2.2 웹 크롤링

우리는 뉴스 기사 제목을 이용하여 분석하기 위해 웹 크롤링 기법을 사용하였다. 웹 크롤러가 수행하는 일을 웹 크롤링(web crawling)이라 부르는데, 웹 크롤러는 조직화 및 자동화된 방법으로 웹을 탐색하는 컴퓨터 프로그램이다. 대부분의 사이트에서는 데이터의 상태를 최신으로 유지하기 위해 웹 크롤링을 하고, 웹 크롤러는 방문 사이트 모든 페이지의 복사본을 생성할 때 사용하고 검색 엔진은 이 복사본을 빠른 검색을 위해 인덱싱 한다.또한 크롤러는 링크 체크나 HTML 코드 검증과 같은 웹 사이트의 자동 유지 관리 작업을 위해 사용되기도 하며, 자동으로 이메일을 수집하는 등 웹 페이지의 특정 정보를 수집하는 것에도 이용된다[12].

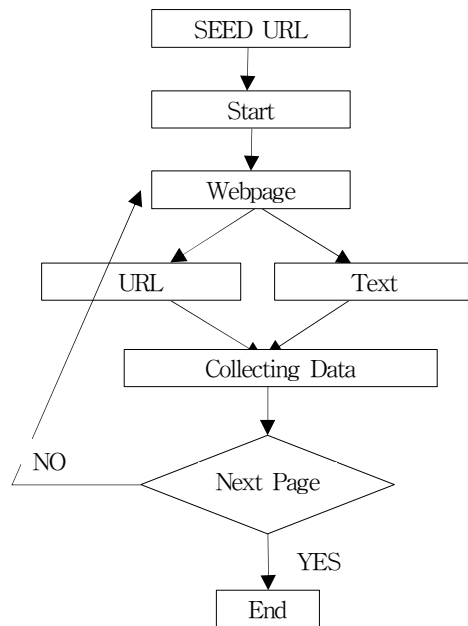


Fig. 1. Flow of Web Crawling

웹 크롤링은 웹 페이지를 보존하고 그대로 가져와서 데이터를 추출해 내는 행위이다. 예를 들어 구글은 굉장

히 짧은 시간에 60조개의 웹페이지를 훑어 사용자가 원하는 정보만을 보여준다. 먼저 검색 서비스를 제공하기 위해 웹페이지를 모으고, 잘 알려진 웹사이트부터 방문해 웹페이지를 수집한다. 네이버와 같은 사이트를 방문한 다음 사이트에서 볼 수 있는 모든 웹페이지 주소를 수집하는 방식이다[13].

Fig. 1은 본 논문에서 사용한 웹 크롤링을 하는 알고리즘이다. 사이트에 검색내용에 대한 URL을 추출하고 페이지의 텍스트를 추출하는 R기반 코드를 직접 개발하였다.

2.3 감성 분석

감성 분석이란 텍스트 정보를 추출하는 텍스트 마이닝과 다르게 한 주제에 대한 주관적인 인상, 감정, 태도, 의견들을 텍스트로부터 뽑아내는 분석을 말한다. 이 분석에서는 추출한 텍스트에서 긍정, 부정, 중립을 파악하고 정성적으로 표현된 언어에서 의미를 찾아 해석하고 여러 분야에서 활용하기 위해 정량적으로 분석한다. 사람들의 의사소통에서 나타난 주관적인 언어를 분석함으로써 일반적인 평가, 의견, 생각 등을 탐색한다. 탐색의 과정은 다량의 텍스트에 쓰인 감성표현 언어들을 추출하고 범주화하고 분류하여 해석하는 단계를 거쳐 정량화하고 도식화해서 의사결정에 도움이 되는 지표를 만드는 것이다[14].

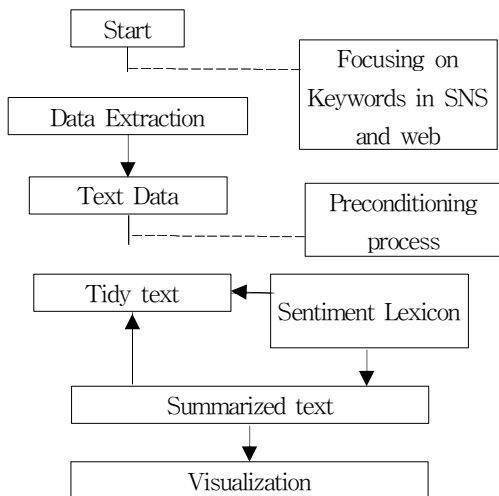


Fig. 2. Flow of Sentiment Analysis

예를 들어, 한 웹 페이지 또는 문서에 대한 전반적인 의견 등의 여부를 측정하거나 정치, 선거 등에 대한 유권

자들의 태도를 파악하는 데 감성 분석을 이용할 수 있다. 감성 데이터는 종종 페이스북 북, 인스타그램, 트위터 등의 SNS와 유사한 사용자 생성 콘텐츠에서 파생된다. 따라서 추출한 감성 데이터는 빅 데이터라고 불리울 만큼 대규모로 수집되는 경향이 있다[15]. Fig. 2는 본 논문에서 감성분석을 진행한 과정을 흐름으로 Python을 이용하여 개발한 코드로 분석하였다.

3. 연구 결과

3.1 분석 결과 통계

먼저, 트위터에서 아동수당과 출산 장려금과 관련된 데이터를 수집하여 감성분석을 진행하였다. 2018년 11월 20일부터 30일까지의 데이터를 수집하였다. 1일 최대 1만5천개의 트윗에 대한 크롤링을 수행한 결과 아동수당 665개, 출산 장려금 2,231개로 총 2,896개의 트윗을 추출할 수 있었으며 광고 등을 제거하는 전처리 과정 진행을 완료한 후 총 2,735개로 나타났다.

다음으로, 언론의 의견을 알아보기 위해 네이버에 올라온 각 언론사의 뉴스 기사 제목을 크롤링 하여 추출하였고 이를 사용해 감성 분석을 진행하였다. 데이터 수집 일자는 2018년 11월 20일부터 12월 1일까지이며 수집 건수는 아동 수당 602개, 출산 장려금 230개로 총 832개였고 관련 없는 기사를 제거하는 전처리 과정 진행 후 데이터의 총 량은 801개로 나왔다.

3.2 아동수당 정책 반응 비교

트위터를 대상으로 분석한 아동 수당과 관련된 감성 분석 결과는 다음과 같다.

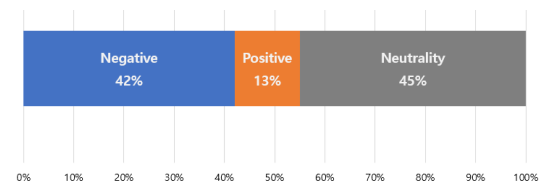


Fig. 3. Result I of Sentiment Analysis

그래프를 중심으로 비율을 계산해 보면, 총 639개 중 부정으로 판단된 비율은 약 42%, 긍정은 13%, 리트윗이나 단순 정책 소개를 보여주는 중립은 45%를 차지하였다. 즉, 트위터에서는 아동 수당에 대해 대체적으로 부정

적 의견이 강하였고 연관분석을 통해서 주로 언급된 키워드는 “형편”, “불분명” 등의 정책의 형평성에 대한 반응이 주요하였다.

다음으로 언론을 대상으로 분석한 아동 수당에 대한 감성분석 결과는 다음과 같다.

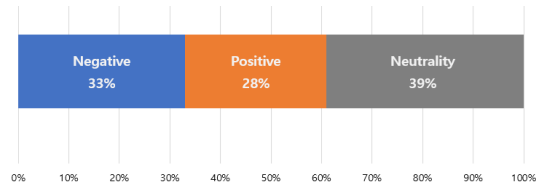


Fig. 4. Result II of Sentiment Analysis

아동 수당에 대한 뉴스 기사 제목 데이터는 총 573개로 부정에 대한 비율은 약 33%, 긍정적 입장은 약 28%, 중립적 입장은 39%이다. 이 결과, 언론사의 입장에서는 보도 형태의 중립적 기사와 부정적 견해가 많이 작성한 것을 알 수 있다.

최종적으로 아동 수당에 대한 트위터와 뉴스 기사의 긍정과 부정 정도를 비교하면 아래 Fig. 5와 같다. 부정적인 견해가 모두 앞서 있다는 것을 보여주고 있다.

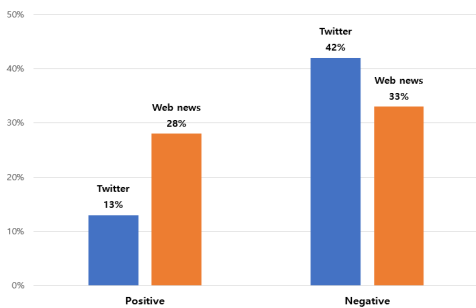


Fig. 5. Result I of Comparative Analysis

3.3 출산장려금 정책 반응 비교

트위터 상에서 출산장려금과 관련된 감성분석 결과는 다음과 같다.

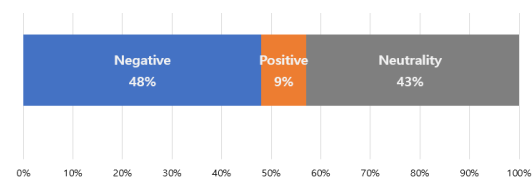


Fig. 6. Result III of Sentiment Analysis

그래프에 따르면 총 2096개 중 부정으로 판단된 비율은 약 48%, 긍정은 9%, 중립은 43%로 나타났다. 즉, 트위터에서는 출산 장려금에 대해 부정적인 반응이 크다는 것을 알 수 있다. 이러한 주된 원인 파악을 위해 연관분석을 하면 “예산”, “복지”, “혜택” 등의 키워드가 주로 나타나고 있어 예산의 과지출이나, 복지 및 혜택의 지나친 남발에 대한 부정적 반응이 나타나는 것을 알 수 있다.

다음은 언론을 대상으로 하는 출산 장려금에 대한 데이터 그래프 Fig. 7이다.

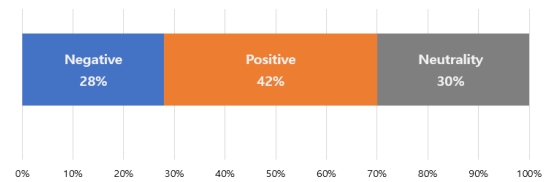


Fig. 7. Result IV of Sentiment Analysis

출산 장려금 관련 데이터는 총 228개이고, 부정적 입장은 28%, 긍정적 입장은 42%, 중립적 입장은 30%로 나타났다. 따라서 언론사는 출산 장려금에 대해 비교적 긍정적인 견해의 기사가 많다는 것을 알 수 있다.

출산장려금에 대한 트위터와 뉴스기사의 긍정과 부정의 비율을 비교하면 다음 Fig. 8과 같다. 아동수당과는 달리 트위터의 반응과 언론의 반응이 다른 것을 알 수 있다.

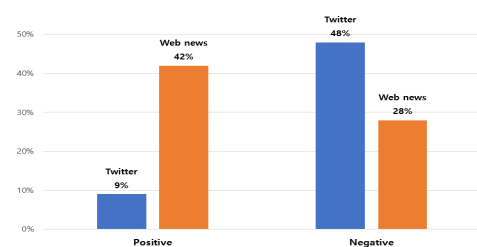


Fig. 8. Result II of Comparative Analysis

4. 결론

분석 결과에 따라 국가의 주요한 정책에 대한 SNS의 반응이 언론에 비해 부정적 견해가 훨씬 더 강하다는 것을 알 수 있었고 오히려 특정 정책에 대해서는 SNS에서 부정적 견해가 많이 나타남을 알 수 있었다. 더불어 긍정, 부정의 강도 역시 언론에 비해 SNS가 훨씬 큰 차이를 보였다. 국가 정책에 대한 반응은 언론보다 대중들이 더 극

과 극으로 나뉘는 현상을 보여주는 것을 알 수 있다.

데이터의 수집 시기가 거의 동일했기 때문에 언론이 SNS로 또는 SNS가 언론으로의 영향을 미치는 연관에 대한 연구는 한계가 있었다. 또한 대중의 반응이 특정 정책에 대한 반응일 수도 있지만 정책 내용보다 정권에 대한 부정적 반응으로 나타날 수 있어 이러한 부분에 대한 추가적인 분석도 필요하다.

추후에는 더 많은 정책과 견해에 대한 반응을 비교해 보고 각각의 영향력 또는 연관성에 대한 관계도 연구 할 수 있을 것이다. 이를 통해 올바른 여론, 대중의 반응을 판단할 수 있는 근거를 제시하고자 한다.

REFERENCES

- [1] S. M. Ko, B. H. Hwang & Y. G. Ji. (2010). A Study on Social Network Service and Online Social Capital : Focusing on a Korean and Chinese Case, *The Journal of Society for e-Business Studies*, 15(1), 103-118.
- [2] Y. N. Lee, E. J. Choi & M. J. Kim. (2018). Analysis of the Influence of Presidential Candidate's SNS Reputation on Election Result: focusing on 19th Presidential Election, *Journal of Digital Convergence*, 16(2), 195-201.
- [3] E. J. Choi, S. W. Choi, S. Y. Lee & M. J. Kim. (2017). Analysis of the effect of the mention in SNS on the result of election, *Journal of Digital Convergence*, 15(2), 191-197.
- [4] G. H. Kim. (2014). An analysis of Courts' Decisions on Freedom of Expression on Social Network Services: Focusing on Twitter and Facebook Cases, *Press and Law*, 13(2), 165-190.
- [5] S. Iyengar & R. Reeves. (1997). Do the Media Govern? Politicians, Voters and Reporters in America, *Electoral Studies*.
- [6] S. H. Kim & J. W. Lee. (2016). *Naver's influence on the public opinion, ahead of KBS · Chosun*, Journalists Association of Korea, <http://www.journalist.or.kr/news/article.html?no=38363>
- [7] Gerbaudo, Paolo. (2018). *Tweets and the streets: Social media and contemporary activism*. Pluto Press.
- [8] A. Lavalée. (2007). Friends swap twitters, and frustration new real-time messaging services overwhelm some users with mundane updates from friends. *Wall Street Journal*.
- [9] Barack Obama, <https://twitter.com/barackobama>
- [10] S. I. Lee. 'Advertising' Naver vs 'Content' Cacao ...

Naver is in the lead, Energy Economic News. <http://www.ekn.kr/news/article.html?no=266715>

- [11] J. J. Lee & Y. M. Sang. (2008). A Study on the Remedies of Infringement of Personal Rights by Portal News Service. *Korean Journal of International Communication*, 37-83.
- [12] M. Kobayashi & K. Takeda. (2000). "Information retrieval on the web", *《ACM Computing Surveys》 (ACM Press)*, 32(2), 144-173.
- [13] Google, *Privacy and terms-Privacy policy*, <https://policies.google.com/privacy?hl=en>.
- [14] B. Pang & Lillian Lee. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information, Retrieval*, 21(1), 1-135.
- [15] Liu, Bing & Lei Zhang. (2012). A survey of opinion mining and sentiment analysis. *Mining text data*. Springer, Boston, MA, 415-463.

윤 혜 민(Yun, Hye Min)

[학생회원]



- 2015년 3월 ~ 현재 : 서울여자대학교 정보보호학과
- 관심분야 : 정보보호, 암호, 인공지능
- E-Mail : hyemin33078@gmail.com

최 은 정(Choi, Eun Jung)

[정회원]



- 2000년 2월 : 서울여자대학교 대학원 컴퓨터학과(이학석사)
- 2005년 8월 : 서울여자대학교 대학원 컴퓨터학과(이학박사)
- 2006년 3월 ~ 현재 : 서울여자대학교 정보보호학과 교수
- 관심분야 : 보안, 빅데이터, 인공지능
- E-Mail : chej@swu.ac.kr