

SNS의 해시태그를 이용한 감정 단어 수집 시스템 개발

이중화* · 이운재** · 이현규***

〈 목 차 〉

I. 서론	III. 연구방법과 프레임워크
II. 선행연구	IV. 연구 알고리즘 실험과 결과
2.1 감정바퀴	V. 결론 및 향후 과제
2.2 해시태그	참고문헌
2.3 웹드라이버	<Abstract>

I. 서론

새로운 산업혁명의 시작은 데이터에서부터 시작되고 있다. 물론, 네트워크와 하드웨어, 소프트웨어 성장과 함께 ICT는 인류에 모바일 환경이 가능하게 만들어준 주요한 기폭제가 되고 있다. 기업 데이터가 정보시스템에서 사용되면서 디지털기업(digital firm)화 되었으며, 그 가치를 창출하기 위한 자동화와 인터넷의 등장인 3차 산업혁명이 지금의 새로운 파괴적 기술(disruptive technology)의 밑거름이 되고 있다. 4차 산업혁명은 모든 사물의 네트워크와 센싱이 가능한 사물인터넷(IoT), 인류에서 발생하는 모든 데이터 분석의 빅데이터(bigdata), 데이터들의 패턴을 학습한 인공지능(AI), 로봇산업,

자율주행 등 이러한 모든 ICT 환경의 개발을 돕기 위한 클라우드 기술까지 한 산업의 굴레가 아닌 다양한 산업의 융합으로 만들어진 형태이다(박성원, 2017; 서새남, 2017; 서병조, 나성욱, 2017).

데이터 기반 새로운 산업혁명은 인간이 보다는 삶을 지속할 수 있도록 새로운 원료로써 그 가치를 더해가고 있다. 수많은 네티즌들이 표현하고 있는 이야기에서 새롭고 유용한 정보나 패턴을 찾고 고객의 니즈를 얻기 위한 노력으로 웹서비스의 진화에 집중할 수 있다(이영재, 이성수, 2011; 유은지 등, 2012). 웹 1.0의 단순 카테고리 관점으로 분류하던 서비스에서 웹 2.0은 누구나 데이터의 생성과 접근이 가능한 개방 환경과 플랫폼 기반의 콘텐츠가 혼합

* 부경대학교 경영학부, newjwcom@daum.net(주저자)

** SK Telecom MNO사업부, yj6048@nate.com(공동저자)

*** 부경대학교 경영학부, hyunqlee@pknu.ac.kr(교신저자)

되었고 웹 3.0은 수많은 작은 단위의 개인 중심의 콘텐츠를 보다 가치 있게 제공하기 위한 큐레이션(curation)의 역할이 커지고 있다(Kouper, 2016; Wu et al., 2017).

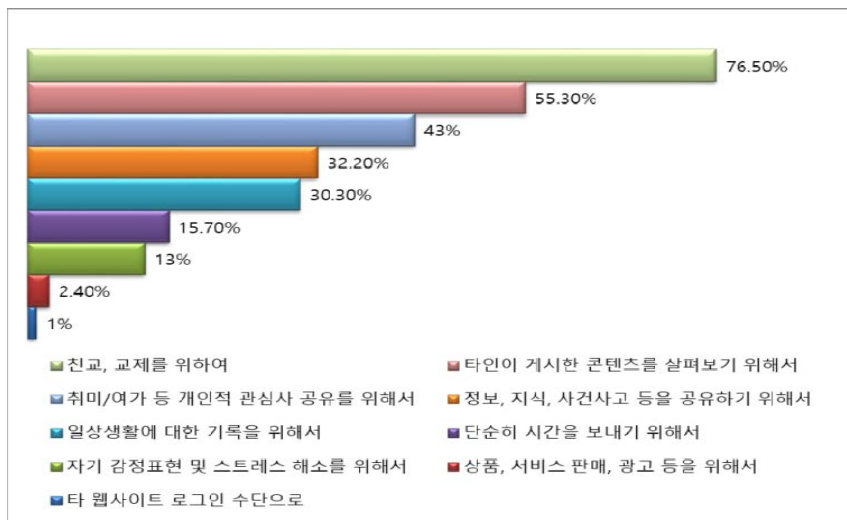
모든 산업은 풍선효과(balloon effect)처럼 글로벌 환경에 영향을 받고 있다. 한 부분이 성장하면 다른 한 부분은 축소되기 마련이며 빠른 변화를 감지하고 대응하는가에 따라 기업의 생존이 결정되며 의사결정에 다양한 데이터와 분석을 필요로 하고 있다. 무엇보다 글로벌 시장 환경에 적응하기 위하여 더 많은 경쟁자(competitor) 분석, 자사(company) 분석, 고객(customer) 분석의 데이터 처리가 필요하다.

특히 고객이 원하고 바라는 것을 기업이 알고 있다면 운영의 수월성을 확보할 수 있으며 그에 따른 경쟁환경에서의 경쟁우위를 선점할 수 있을 것이다. 이러한 고객의 니즈 분석에 주요한 SNS 데이터의 해시태그(hashtag)를 웹 3.0 서비스를 활용하여 연구하고자 한다.

SNS는 자신의 이야기를 알리거나 여러 사람

들과 쉽고 빠르게 대화할 수 있는 장점이 있다. 이런 정보 공유의 장인 SNS 이용은 모바일 디바이스 사용 증가와 인터넷 이용의 증가로 인터넷 이용자의 68.2%가 사용하는 것으로 집계되고 있다. 2017년 한국인터넷진흥원에 따르면 3세 이상 인구 중 인터넷 이용 인구는 90.3%이며 모바일 디바이스 사용은 94.1%로 현대인들의 필수품으로 조사되었다. <그림 1>과 같이 SNS를 이용하는 이유를 보여주고 있다. 2017년 전국 25,000가구 구성원들을 대상으로 조사한 결과 SNS를 사용하는 이유로 친교 및 교제 목적(76.5%), 타인이 게시한 콘텐츠를 보기 위한 목적, 개인적 관심사 공유, 정보 및 지식 공유 등으로 나타났다(한국인터넷진흥원, 2017).

해시(hash)는 사용자가 의미 있게 전달하고자 하는 함축적인 문장에 태그(tag)인 꼬리표를 연결하여 특정 주제에 관심과 지지를 나타내는 방식으로 검색의 편리함을 위한 SNS의 큐레이션이라 할 수 있다(안희정 등, 2017). 원하는 정보를 빠르게 검색하여 가치 있는 정보만을 제



<그림 1> SNS 이용 이유(한국인터넷진흥원, 2017)

공하기 위한 기능이라 할 수 있다. 이런 해시태그들은 사용자들의 감정을 고스란히 가지고 있다(남민지 등, 2015).

SNS 콘텐츠의 해시태그를 이용한 감정 단어 사전은 어떤 특정 제품이나 서비스, 인물, 정책, 사건, 사고 등 사회 이슈들에 대한 네티즌이 느끼는 여러 감정의 연결 고리를 파악할 수 있다. SNS에서 사용자들이 작성한 해시태그는 본 연구에 있어서 감정 씨앗(seed)의 하나이다. 또한, 기업관점에서 제품이나 서비스의 마케팅 전략 측면에서는 소비자에게 적절히 평가되고 있는가 아니면 특정 감정으로 치우쳐지는가에 대한 분석을 하기 위한 다차원 감정 사전의 재료가 되는 것이다.

본 연구의 목적은 이런 해시태그를 활용하여 한글 기반 감정 사전을 구축하고자 한다. 이분법적인 긍정, 부정적 단어의 선별이 아니라 네티즌들의 자음, 모음의 다양한 표현을 다차원 감정 세트(emotional set)의 기반으로 연구하고자 한다. 본 논문의 2장은 감정 단어들에 대한 연구인 감정 바퀴와 해시태그, 단어 수집을 위한 가상 웹 드라이버에 관하여 기존 연구를 살펴본다. 3장에서는 본 연구의 핵심인 감정 단어를 추출하기 위하여 감정 세트별 감정 시드를 제시하였으며 본 연구의 프레임워크를 제시하였다. 4장은 크롬드라이버를 활용한 감정 단어 수집 알고리즘과 일곱 가지 감정 세트별 감정 단어 수집 통계를 제시하였으며 5장에서 결과와 시사점을 제시하고자 한다.

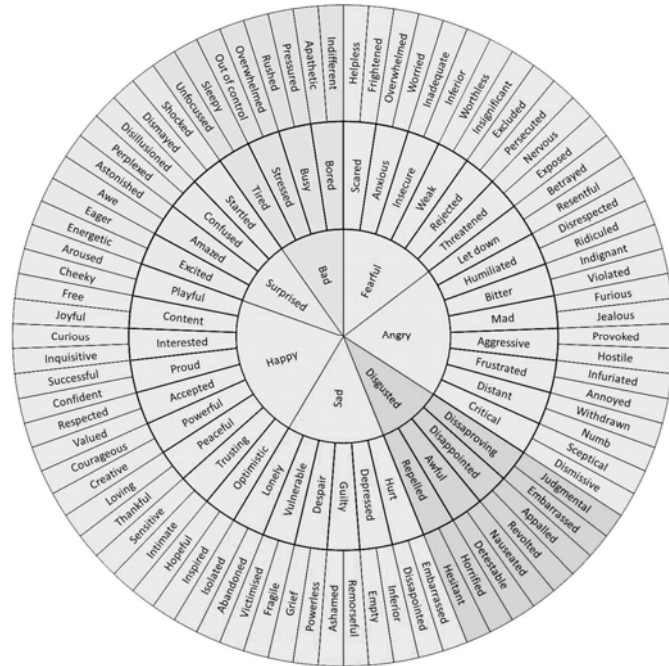
II. 선행 연구

2.1 감정바퀴(Wheel of Emotions)

영어는 소리를 나타내는 글자를 하나하나 늘어 놓는 방식의 나열식이며 WordNet을 활용한 연구가 활발히 진행되고 있다(Strapparava and Valitutti, 2004; Baccianella et al., 2010). 한글은 글자를 모아서 쓰는 방식으로 영어 표현과는 많은 차이를 보이고 있고 박인조와 민경환(2005), 이준웅 등(2009), 손선주 등(2012), 안정국과 김희웅(2015), 이종화 등(2016), 이종화와 이현규(2016), 이수상(2016) 등의 연구가 있었다. 최근 연구는 오픈 API를 활용한 긍정, 부정을 구분하는 한글 감성 연구(안정국, 김희웅, 2015)와 SNA 분석을 활용한 극성과 각성 차원의 한글 감성 연구(이상수, 2016), SNS의 비표준어 한글 연구(이종화 등, 2016)로 한글 감성 단어 연구가 진행되고 있다.

본 연구는 미국 심리학자 플러치(Plutchik, 1980)의 감정 바퀴(wheel of emotions) 연구를 바탕으로 기획하고자 한다. 플러치의 감정 바퀴는 Angry(화난, 성난), Bad(불쾌한, 나쁜), Disgusted(역겨운), Fearful(무서운, 두려운), Happy(행복한), Sad(슬픈), Surprised(놀란)의 7가지의 형용사로 구분하였으며 감정을 계층적으로 구분하여 조합하는 형태로 진행하였다.

이러한 플러치의 감정 바퀴를 활용하여 SNS 이용자의 감성 분석에 활용하고자 한다. SNS 콘텐츠는 텍스트, 동영상, 이미지 등 다양한 형태의 비정형적 자료들로 구성되어 있으며 그 중 텍스트 기반의 해시태그를 활용하여 플러치의 감정 단어와 연결하여 한글의 감정 단어 사



<그림 2> wheel of emotions(Plutchik, 1980)

전을 구축하고자 한다.

감성 모형은 각성과 극성을 이용한 테이어 (Thayer, 1989) 모형과 러셀(Russell, 1980) 감정 모형이 연구가 이루어져 있지만 단순 긍정, 부정의 측면 이전에 인간의 다양한 감정 전체를 텍스트 기반으로 정리하고자 플러치의 감정 바퀴를 활용하게 되었다.

본 연구는 특정 SNS의 해시태그를 이용한 감정 단어 추출을 통하여 감정 씨앗인 시드와 시드에서 확장된 단어들인 파생어를 함께 채굴하는 알고리즘을 활용하여 다양한 감정을 표현하고 있는 해시태그를 수집하고자 한다. 자음, 모음의 다양한 표현으로 네티즌의 신조어까지 함께 수집하여 고객 니즈 분석의 기준을 마련하고자 한다.

2.2 해시태그(hashtag)

웹 1.0인 월드 와이드 웹(www)은 사용자가 신문이나 방송처럼 일방적으로 정보를 받는 서비스인 반면 웹 2.0은 웹 사용자들이 참여할 수 있는 플랫폼 기반으로 정보를 함께 공유하고 제작하는 서비스로 발전되었다. 현재 웹 3.0은 인간이 사용하기 적합한 지능화된 웹 환경의 발달과 개인화 서비스 기반의 융합된 방향으로 개발되고 활용될 것이다(정소영, 2015). 이런 웹 환경에 맞게 SNS 또한 변화가 이루어진다. 1세대는 미니 홈페이지를 통한 기존 오프라인 인맥을 확장하는 형태이다. 2세대는 모바일 디바이스의 등장으로 단순한 내용을 실시간으로 전달하는 페이스북(facebook)이나 트위터(twitter)를 활용하는 형태이다. 3세대는 작은 단위의 소

셜 플랫폼을 연결하고 특정 주제를 중심으로 관심사 공유하는 큐레이션(curation)이다. 사진이나 영상 등 시각물을 중심으로 한 이용자 맞춤형 서비스로 링크트인, 인스타그램, 텀블러 등이 대표적이다(남민지 등, 2015).

“#” C언어에서 헤더 파일을 호출하기 위한 기호로 C/C++언어가 우선적으로 처리해야 하는 헤더파일 선언문에 사용되었던 기호이다. SNS에서 해시태그는 게시물에 달린 일종의 꼬리표이며 검색어에 링크를 설정하는 기호로 사용되고 있다. SNS 게시물 앞에 해시태그인 “#”을 표시하면 같은 해시태그를 사용한 게시물은 함께 찾아볼 수 있다는 것이다. 사진과 글을 입력하고 “#”를 하는 순간 사람들 속으로 들어간다. 나이, 직업도 모르는 누군가의 세상으로 그들과 나를 연결하는 것이 해시태그이다(윤여경, 2016). 인스타그램, 페이스북, 트위터 등 소셜 미디어에서 특정 단어들을 빠르게 검색할 수 있도록 하는 의미적인 메타데이터 체계를 한 키워드 “#오늘은뭘먹지”, “#공유합니다”, #가볼을 경우 다른 네티즌에게 연결의 의미로 공유되는 비율이 증가한다고 한다. 2014년 미국에서 18세 흑인 소년이 백인 경관이 쏜 총에 사망하자 누군가의 SNS에서 “#BlackLivesMatter”, “#흑인의생명도소중하다” 해시태그의 등장과 함께 공유가 시작되었고 공감하는 시민들이 급속히 생겨나게 되었다. 미국 법 집행 기관의 인종차별적인 행태에 분노한 수많은 사람들이 이런 마음을 널리 알리고 공유하기를 원했다는 것을 의미한다고 볼 수 있다. 이런 해시태그의 영향으로 자신이 좋아하는 분야에 관련된 것을 수집하거나 찾아보는 “덕질”, “덕후” 등이 있고 어떤 단어나 문장에 “#”를 하는 순간

확장되는 가능성과 몇 분 지나지 않아 “#”된 사진들은 다양한 지구촌의 반응을 이끌어낸다. 본 연구는 네티즌들의 감정 표현 해시태그를 다양한 차원으로 구분하여 수집하고자 하는 연구를 진행하고자 한다.

2.3 웹드라이버(web driver)

단순 검은 액체 물질에서 인류의 산업 혁명을 주도 했던 석유의 가치는 혁명 그 자체였다. 하지만 현재 또 다른 산업혁명을 주도하는 것은 바로 미래의 석유인 데이터이다. 이러한 데이터의 관심은 새로운 직업군을 만들어내고 있다. 웹 환경에서 필요한 데이터 패턴 분석과 해당 웹 페이지 데이터 수집을 담당하는 크롤러(crawler)에 대한 직업군 인기도가 높아지고 있다. 이러한 크롤링 과정에서 오픈 소스 소프트웨어 API를 개발하기 위한 언어로 파이썬(python)을 사용한다. 다소 문법이 간결하고 표현 구조가 인간의 사고 체계와 닮았다는 파이썬은 사용자 수도 계속 증가하고 있다(<http://github.info/>).

파이썬에서 데이터 크롤링에 필요한 모듈은 Requests, BeautifulSoup, Selenium 등이 있다. 보통은 웹페이지를 요청하는 Requests 모듈과 요청 결과를 파이썬에서 사용할 수 있는 타입으로 변환하는 BeautifulSoup 모듈을 조합해서 사용하거나, 동적 웹페이지의 데이터를 가져와야 할 때 웹 테스트 모듈인 Selenium을 사용한다(Kaur and Gupta, 2013; Gojare et al., 2015).

Selenium을 사용하기 위해서는 웹을 테스트하기 위한 드라이버를 설치해야 하는데 드라이버는 크롬, 익스플로러, 파이어폭스 등 브라우

저에 맞는 드라이버를 사용하게 되며 Selenium에서는 “WebDriver” 명을 명시한다. Selenium을 구동할 수 있는 언어들에는 Java, C#, 파이썬, 루비, 펄, PHP, JavaScript 등에서 웹드라이버 API를 사용할 수 있다.

Selenium의 WebDriver는 브라우저에 내장된 자동화 지원 기능을 사용하여 브라우저를 직접 호출해서 데이터를 가져오는 기능을 한다. 특히 Selenium-WebDriver는 페이지 자체가 다시 로드되지 않고 페이지 요소가 변경될 수 있는 동적 웹 페이지를 보다 잘 지원하도록 개발되어 있다. 웹 드라이버의 종류는 Firefox Driver, Internet Explorer Driver, Chrome Driver, Opera Driver, IOS Driver, Android Driver 등이 사용되고 있다(<https://www.seleniumhq.org/>).

본 연구에서는 인터넷 뉴스 기사, 누리꾼들의 기사 댓글, 인스타그램, 페이스북, 트위터 등과 같은 SNS에 방대한 텍스트기반 한글 해시태그 처리를 목적으로 진행되고 있다. 누리꾼들이 자연스럽게 표현하는 이야기를 여러 가지 감정 세트로 구분하기 위하여 선행되어야 하는 감정별 단어를 수집할 수 있는 시스템을 구축하고자 한다. 플러치(Plutchik, 1980)의 감정 바퀴를 기반으로 감정 세트를 구성하고 게시물에 꼬리표 역할을 담당하며 검색어에 대한 링크로 사용되고 있는 해시태그에서 감정 단어를 추출하고자 한다. “#기분좋아”, “#기분 좋음”, “#기분 좋은”, “#기분 좋다”, “#기분 최고”, “#기분 업”, “#기분 좋게” 등으로 “기분”을 표현하는 해시태그를 보듯 다양한 표현으로 인터넷 사용자들의 감정을 나타내고 있다. 웹 드라이버 크롬을 이용한 가상 웹 브라우저의 환경을 사용하

여 웹 브라우저 검색창에 감정 시드를 넣으면 나타나는 자동 완성(autofill) 단어를 수집하기 위한 시스템 개발을 하고자 한다.

Ⅲ. 연구방법과 프레임워크

사용자의 감정을 읽을 수 있는 감정 단어의 수집을 SNS의 해시태그를 통하여 확인하고자 한다. 플러치(1980)의 감정 바퀴를 이용하여 영문으로 된 감정단어를 한글로 번역하고 SNS의 한글 해시태그에서 감정 단어를 찾고자 한다. 플러치는 긍정, 부정의 이원화된 감성 사전이 아니라 심리학에서 인간의 희노애락(喜怒哀樂)을 표현하듯 일곱 가지의 감성을 나타내고 있다(장문수, 2012). 이러한 감성 바퀴를 이용하여 감성 세트를 구축하였으며 다양한 네티즌들의 감정 표현을 한글 해시태그를 통하여 정리하고자 한다.

본 연구에서 가장 중요한 것은 감정 단어 수집이다. 표준어로 표현한 단어와 함축적으로 적용한 신조어, 인터넷어 등 비표준어이지만 사용자가 감정의 의미를 갖고 있는 해시태그라면 수집 대상이 된다. 이러한 감정 단어를 수집하는 대상 SNS는 페이스북([facebook.com](https://www.facebook.com))의 자회사인 인스타그램([instagram.com](https://www.instagram.com))으로 설정하였다. 인스타그램은 2010년 서비스를 시작으로 같은 해 100만 명, 2011년 1,000만 명으로 사용자가 늘어나면서 2016년 6월에 5억 명 이상이 친구나 가족과 사진, 동영상, 텍스트 메시지 전달 등의 콘텐츠를 공유하며 사용하고 있다([wikipedia.org](https://www.wikipedia.org)).

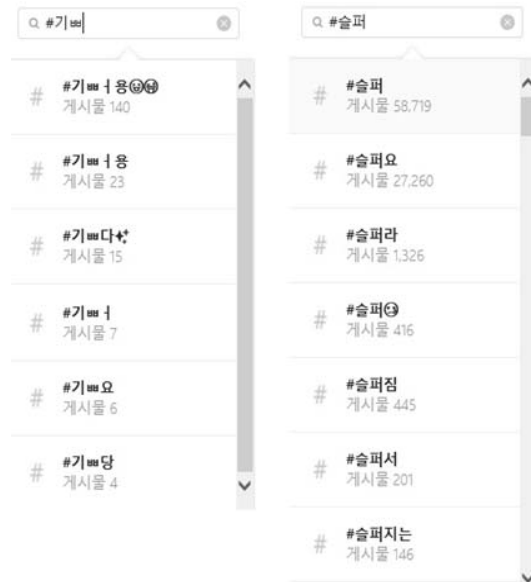
연구에 필요한 해시태그 수집은 인스타그램

웹 페이지에서 인기 해시태그를 사용한 계정 수집과 수집된 계정마다 접속하여 해당 계정의 모든 콘텐츠내의 해시태그를 수집하고자 한다. 실제 데이터 수집 범위 및 기간을 따로 설정하지 않고 있으며 콘텐츠마다 해시태그만을 수집하여 연구에 활용하고자 한다.

포털사이트나 SNS에서 제공하는 기능으로 검색어 첫 음절만 입력하면 이어지는 문자열이 자동으로 나타나 마지막 음절까지 입력하지 않아도 단어를 완성시켜주는 자동완성(autofill) 기능이 있다. 이러한 자동완성기능을 활용하여 해당 감정 단어를 크롤링하고자 한다.

<그림 3>은 연구 대상 SNS 인스타그램의 해시태그를 확인하기 위한 화면을 보여주고 있다. 두 음절을 넣으면 입력한 글자에서 시작되는 단어를 자동완성 기능을 통해 확인할 수 있는 것이다. “#기쁨”과 “#슬퍼”를 확인한 결과 해당 단어로 시작되는 모든 해시태그와 사용한 빈도까지 제공되는 것을 확인할 수 있었다. 해시태그의 검색어에 따라 나타나는 자동완성 단어 빈도는 자동 완성 창 의 스크롤바를 비교해 보아도 그 차이가 크다는 것을 알 수 있다. “#기쁨”보다는 “#슬퍼”로 시작되는 해시태그가 훨씬 많은 것을 보여주고 있다. 또한 자동 완성 단어로 나타난 단어를 해시태그로 사용하는 게시물 빈도와 빈도순으로 정렬된 상태에서 자동완성 단어들을 보여주고 있다. “#기쁨”로 시작되는 해시태그의 최대 빈도는 140개이며 “#슬퍼”로 시작되는 해시태그는 무료 6만 번 사용하였다는 뜻이다. <그림 3>의 “#슬퍼”를 살펴보면 “슬퍼요”, “슬퍼라”, “슬퍼😊”, “슬퍼짐”, “슬퍼서”, “슬퍼지는” 등 인기 있는 해시태그인 경우 100개까지 자동완성 키워드를 제공하는

것으로 확인되었다.



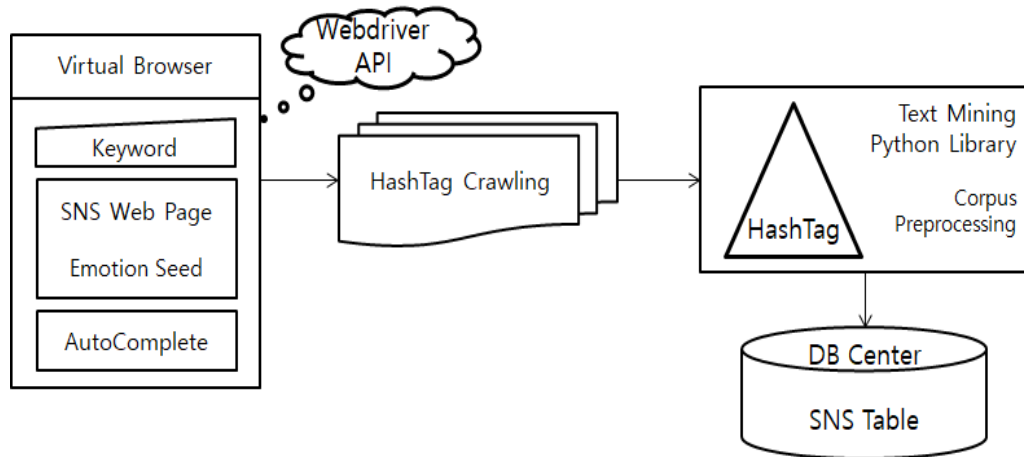
<그림 3> 자동완성기능을 통한 감정 단어

플러칙(Plutchik, 1980)의 감정 바퀴 연구를 바탕으로 한글 감정 단어 세트를 구성하였다. 플러칙의 감정 바퀴는 Angry, Bad, Disgusted, Fearful, Happy, Sad, Surprised 의 7가지의 형용사로 구분하였으며 감정을 계층적으로 구분하여 조합하는 형태로 진행하였다. 이러한 플러칙의 감정 단어와 연결하여 한글의 감정 단어 사전을 구축하였다. 특히 네티즌이 표현하는 감정이 담겨 있는 해시태그를 사용하였으며 한글의 자음, 모음의 이모티콘화 된 음절도 함께 감정 단어에 포함시켰다.

각각의 자극의 변화에 따라 느끼는 여러 가지 감정 단어를 수집하기 위하여 감성 시드를 <표 1>과 같이 지정하였다. 시드는 감정을 표현하기 위한 기본 단위인 두 음절을 기초로 하여 작성하였다. 첫 음절과 자음을 넣은 경우, 두

<표 1> 자극의 변화에 따른 감정 시드

Emotion Set	Seed Word
Angry	경멸, 고함, 괴씸, 괴핌, 굴욕, 귀차, 귀차, 귀찬, 귀찮, 귀차, 끄짜, 끄짜, 난폭, 날카, 더러, 더럽, 더럽, 따분, 떨떠, 모욕, 무례, 미운, 미움, 미워, 미칠, 밋다, 밋다, 밋당, 반대, 반항, 배신, 보복, 복수, 부담, 분노, 분노, 분하, 분한, 분함, 불만, 불안, 불쾌, 불쾌, 불편, 비난, 비명, 비열, 비판, 성급, 성나, 성난, 성질, 속상, 시끄, 실망, 싫, 심술, 씹쓸, 앙심, 약오, 알미, 알미, 억울, 언짢, 언짢, 여겨, 여겨, 열받, 욕나, 욕하, 욕한, 욕함, 욕해, 욕했, 원망, 적대, 좌절, 지겨, 지겹, 지루, 질책, 질투, 짜, 짜, 짜, 짜, 짜, 짜, 통제, 풍자, 찢대, 하찮, 혐오, 혼나, 혼난, 혼난, 혼나, 화나, 화남, 화남, 화났, 화났, 후덥
Bad	나빠, 나빠, 나빴, 나빠, 나쁘, 나쁜, 나쁨, 냉담, 바빠, 바빠, 바빴, 바빠, 바빠, 바빴, 바빠, 바빠, 바빠, 바빠, 바빠, 바빠, 졸라, 졸라, 졸려, 졸림, 졸리, 졸린, 졸림, 졸림, 지루, 통제, 피고, 피고, 피곤
Disgusted	겹에, 격되, 궁색, 망서, 망설, 무서, 무서, 무섭, 불쾌, 비판, 실망, 역기, 역겨, 역겹, 혐오
Fearful	거부, 거절, 걱정, 놀라, 놀란, 놀람, 놀랍, 놀랐, 놀래, 놀랜, 놀렘, 놀랬, 두려, 두려, 불안, 약하, 약한, 약함, 약해, 위협, 제외, 학대, 협박
Happy	기쁘, 감사, 감사, 팬찮, 팬찮, 궁금, 귀중, 그림, 기대, 기뻐, 기뻐, 기뻐, 기뻐, 기뻐, 기쁨, 기쁨, 기쁨, 기쁨, 기쁨, 기쁨, 낙관, 느긋, 마냥, 만족, 무료, 믿는, 믿어, 믿었, 믿었, 반가, 발랄, 벅차, 벅찬, 벅참, 벅참, 사랑, 살맛, 상쾌, 상큼, 새심, 세심, 시원, 씩씩, 아늑, 안전, 온화, 용감, 인정, 자랑, 짜릿, 친하, 친한, 친함, 친함, 친해, 쾌활, 쾌활, 포근, 행복, 화사, 화창, 황홀, 후련, 흐뭇, 흥미, 흥분, 희망
Sad	고독, 고립, 고통, 공허, 궁색, 낙담, 낙심, 낭패, 단절, 더러, 더럽, 뜸하, 뜸한, 뜸함, 뜸해, 뜸했, 막막, 맥빠, 무력, 몽클, 민감, 부끄, 복받, 불행, 비참, 상처, 상처, 서글, 서운, 슬고, 슬퍼, 슬프, 슬픈, 슬픔, 슬픔, 신음, 쓸쓸, 암담, 애꿎, 애석, 애절, 애뜻, 외로, 우울, 울머, 울었, 울음, 울적, 위축, 적적, 절망, 좌절, 주눅, 주눅, 참담, 처량, 충격, 취약, 침울, 풀죽, 하위, 허약, 허진, 허탈, 황량, 후회
Surprised	갈망, 깜짝, 깜짝, 난처, 놀르, 놀라, 놀란, 놀람, 놀랍, 놀랐, 놀래, 놀랜, 놀렘, 놀랬, 당황, 두려, 두려, 열망, 충기, 충겨, 충격, 혼란, 활기, 활기, 흥너, 흥분



<그림 4> 본 연구의 프레임워크

음절을 사용하는 경우가 자동완성 기능을 통하여 확인되었다.

<표 1>의 자극의 변화에 따른 감정 시드는

플러칙의 일곱 가지 감정 세트를 활용하였으며 각 감정 세트의 시드는 감정을 계층적으로 구분하여 조합한 단어들을 재구성하여 수집하였

다. 시드는 두 음절 기반으로 먼저 수집하였으며 웹 브라우저의 자동완성 기능을 활용하여 수작업을 진행한 결과이다.

<그림 4>는 본 연구에서 중점적인 감정 단어 수집에 관한 프레임워크이다. 웹드라이버(webdriver)는 가상 웹 브라우저를 사용하여 사람이 마치 각각의 해시태그를 입력하여 자동완성단어를 확인하듯이 컴퓨터 시스템에서 대신 해주는 API를 개발할 수 있는 개발 환경이다. 이런 웹 자동 테스트 프레임워크인 웹드라이버를 사용하여 <표 1>의 감정 시드를 활용한 감정 세트 일곱 가지에 해당되는 해시태그를 세트별로 수집하고자 설계하였다. 연구 대상인 인스타그램은 각 해시태그 시드에 최대 100개의 해당단어를 제공하였고 추가적으로 나타나지 않은 해시태그를 더 확보하기 위하여 시드를 더 확장하는 연구도 함께 기획하였다.

“Surprised” 감정 세트의 감정 단어인 “#놀란”의 1차 시드에서 더 확장하기 위하여 음절 단위를 증가시켜서 자동완성 기능을 활용하였

다. “놀란”의 시드에서 나타난 단어는 “놀란표정”, “놀란토끼눈”, “놀란기슴”, “놀란척”, “놀란눈” 등의 1차 시드에서 못 찾아낸 감정 단어를 추가적으로 확인할 수 있었다. “Happy” 감정 세트의 감정 단어인 “#행복”이란 시드는 6,546,060개의 게시물에 사용되고 있으며, 100개의 자동 완성 단어를 빈도순으로 나열하면 최소로 사용된 게시물 수가 3천 개 이상이 있는 것을 확인할 수 있었다. 즉, 자동완성 기능을 통해서 확인한 100개 이외의 단어가 더 등장한다고 볼 수 있다. 추가 확보를 위하여 시드의 음절 수를 세 개로 확대하여 확대된 세 음절 시드 또한 100개의 자동완성 단어를 제공하기 때문에 “#행복”이란 시드에 최대 10,000개의 감정 단어를 추출할 수 있으며 중복된 단어는 별도 확인이 필요한 것으로 나타났다.

<표 2>는 “Angry” 감정 세트에서 “경멸” 시드를 이용한 감정 단어 추가 추출 예이다. 두 음절의 ‘#경멸’에서 1차 추출된 감정 단어는 “경멸하는”에서부터 “경멸하다”까지 8개의 단어

<표 2> 감정 시드를 이용한 감정 단어 추출

시드	1차 추출	2차 추출	최종(시드 + 1차 + 2차)
경멸	경멸하는 경멸하는거같다 경멸하는거아님 경멸하는내친구 경멸하는눈빛으로보개 경멸하는눈빛으로보지말아줘 경멸하는사람이니께 경멸하다	경멸하는 경멸하는거같다 경멸하는거아님 경멸하는내친구 경멸하는눈빛으로보개 경멸하는눈빛으로보지말아줘 경멸하는사람이니께 경멸하는인간 경멸하는자 경멸하는중 경멸하는표정 경멸하다 경멸하던시선 경멸하리라	경멸 경멸하는 경멸하는거같다 경멸하는거아님 경멸하는내친구 경멸하는눈빛으로보개 경멸하는눈빛으로보지말아줘 경멸하는사람이니께 경멸하다 경멸하는인간 경멸하는자 경멸하는중 경멸하는표정 경멸하던시선 경멸하리라

가 수집되었다. 즉 “경멸”로 시작되는 해시태그의 모임이라 할 수 있다. 하지만 한글의 특수성 때문에 미처 단어 완성 기능에 나타나지 않은 단어를 위하여 2차 배양을 실행하였다. 2차 배양은 1차 배양 단어에서 추출된 단어를 기준으로 하였으며 세 음절까지 확장하여 수집하였다. <표 2>의 예시는 “#경멸하”를 이용하여 “경멸하는”부터 “경멸하리라”까지 14개의 단어가 수집됨을 확인하였다. “경멸하는인간”, “경멸하는자”, “경멸하는중”, “경멸하는표정”, “경멸하던시선”, “경멸하리라”가 추가된 것을 확인할 수 있다. 감정 단어 추출의 시드와 1, 2차 추출 단어를 중복 제거한 결과로 15개의 ‘경멸’ 감정 단어를 추출하였다. 이러한 알고리즘을 이용하여 <표 1>에서 제시된 일곱 가지 감정 세트의 시드를 본 연구의 감정 단어 수집 시스템으로 수집하였다.

IV. 연구 알고리즘 실험과 결과

본 연구는 SNS의 해시태그를 이용한 감정 단어 수집 시스템을 개발하였다. 심리학자 플러치(Plutchik, 1980)의 감정 바퀴 연구를 바탕으

로 감정 세트를 구분하였다. 웹 브라우저의 자동 완성 기능을 활용하여 두 음절의 해시태그를 사용하여 감정 시드를 각 감정 세트를 기준으로 확보하였다. 두 음절로 구성된 시드 단어를 기준으로 1차 추출하고 세 음절 기준으로 추가 2차 추출하여 시드 기준 감정 단어를 확장시켜 진행하였다.

연구에 필요한 감정 단어 수집 시스템은 풍부한 라이브러리를 제공하며 웹프로그래밍과 데이터 분석에 많이 사용되는 파이썬을 사용하여 개발하였다. 브라우저에 내장된 자동화 지원 기능을 사용하여 브라우저를 직접 호출해서 데이터를 가져오는 기능이 가능한 Selenium의 Webdriver를 사용하였다. 특히 크롬드라이버는 페이지 자체가 다시 로드되지 않고 페이지 요소가 변경될 수 있는 동적 웹 페이지 기능을 활용하여 자동 완성 단어를 감정 단어 연구 재료로 사용하였다.

<그림 5>는 Selenium의 Webdriver를 사용한 파이썬으로 개발한 자동 완성 기능의 단어를 추출할 수 있는 알고리즘이다. 연구자가 지정한 시드에서 파생된 단어를 함께 채굴하는 알고리즘을 활용하여 한글 자연어 처리(KoNLP)를 수행하였다.

```

from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.action_chains import ActionChains
import time
import urllib.parse

path = " 크롬드라이버 설치 경로 "
driver = webdriver.Chrome( path )

```

```

#emosional.txt파일의 Seed(씨앗)를 읽어올 준비
f = open("emosional.txt", "r")
# sad_hashes.txt에 자동완성 단어를 저장하기 위해서 파일 지움.
f2 = open("emosional_hashes.txt", "w")
f2.write("")

while True:
    #sad.txt 한 줄씩(한 단어) 읽어오기
    line = f.readline()
    #단어 앞에 #붙임(해시태그를 검색하기 위함)
    keyword = '#'+line

    if not line:
        break

    #인스타 그램 접속
    driver.get("https://www.SNS web page.com/")
    #검색창 찾을
    inputbox = driver.find_element_by_css_selector("검색창")
    #검색창 비움
    inputbox.clear()
    #검색창에 검색할 해시태그 입력
    inputbox.send_keys(keyword)

    #3초 기다림
    time.sleep(3)

    #자동완성이 나타나는 위치 찾기
    elems = driver.find_element_by_css_selector("오토필")
    #그 단어들 중, 해시태그인 a 태그 가져오기
    aelems = elems.find_elements_by_css_selector("오토필내의 태그")

    #해시태그 전체 리스트
    for i in aelems:
        #해시태그의 링크 주소(href)만 가져옴
        hash_tags = i.get_attribute("href")
        #링크를 /로 분할, n 번째가 해시태그 텍스트임.
        a = hash_tags.split("/")

        # 자동완성해시태그 - url디코딩
        txt2 = urllib.parse.unquote(a[n], encoding='utf-8')

        try:
            # sad_hashes.txt에 자동완성 단어를 저장
            f2.write(txt2)
            f2.write("\n")
        except:
            pass

```

<그림 5> 감정 단어 수집 알고리즘(Python Script)

<그림 5>의 알고리즘을 살펴보면 가상 웹 브라우저인 크롬드라이버를 사용하기 위한 웹 드라이버 환경설정을 진행하였다. 감정 세트 단위 두 음절 시드를 하나씩 로딩하며 연구 재료가 되는 SNS에 접속하여 로딩된 시드를 검색 창에 입력하게 된다. 자동 완성 단어가 웹에 들어오는 시간을 3초를 설정하였으며 자동완성 단어가 나타난 해시태그의 Elements를 확보한다. 확보된 Elements만큼 반복을 통하여 해당 웹 클래스 태그의 텍스트 값을 가져온다. 이런 알고리즘을 이용하여 또 다른 두 음절 시드의 감정 단어를 반복 수집하게 된다.

<그림 6>은 파이썬 라이브러리를 활용하여 SNS 웹 페이지의 콘텐츠를 크롤링 후 문장내의 해시태그만을 추출하기 위한 알고리즘이다. “#” 기호를 찾고 이후 한 음절씩 비교하여 공백이나 특정 특수문자가 들어오면 이전까지의 음절이 하나의 해시태그로 구성된 것을 확인 후 저장하고 같은 방법으로 반복하는 구조이다.

사용 빈도가 높은 시드인 경우 추가 확보를 위하여 세 음절 시드 또한 같은 방법으로 크롤링이 가능하며 같은 DB내 전처리를 통하여 최종 감정 단어를 수집하였다. 물론 범용적인 SNS DB 접속 계정이 있다면 가상 웹 드라이버

```
# python 라이브러리를 활용하여 웹 페이지 크롤링 내용 활용
hashtags <- "SNS crawling web data"

# 콘텐츠 문장 길이
s_len<-nchar(hashtags)

# 콘텐츠 전체 문장 길이-1 까지 반복
for(j in seq(nchar(hashtags)-1))
{
  # 해시태그 “#”을 찾게되면
  if(substr(hashtags, j, j) == "#"){

    # 해당 문장의 길이만큼 한 글자씩 확인 작업
    for(i in j+1:s_len){
      one_char <- substr(hashtags, i, 1)

      # 단, 공백, “#”, “\n” 등을 만나면 이전까지 음절을 해시태그로 확인
      if(one_char==" " || one_char=="#" || one_char=="\n" ||
        one_char=="\" || one_char==",")
      {
        a <- substr(hashtags, j+1, i-1)
        h <- paste(h, a);
        break
      }
    }
  }
}
```

<그림 6> 크롤링 후 해시 태그 추출 알고리즘(R Script)

<표 3> 감정 세트별 감정단어 통계

감정세트	시드	확장 시드	배양된 감정단어 수
Happy	65	6,893	41,166
Angry	100	8,245	39,740
Sad	66	6,072	29,158
Bad	32	2,050	18,063
Fearful	23	2,207	13,414
Surprised	26	3,249	8,831
Disgusted	15	1,554	4,970
합계	327	30,270	155,342

를 사용하지 않고 원하는 데이터를 가져올 수 있을 것이다. 대부분의 연구자들은 그렇지 않기 때문에 마치 사람이 하나 하나 필요한 자료를 가져 오듯 웹드라이버를 통하여 반복적인 작업을 수행하는 시스템이 필요한 것이다. 수집 로봇이 웹에 접속하여 데이터가 웹 브라우저에 로딩되는 시간까지 고려하여 크롤링 작업이 수행되다보니 3초라는 지연 시간이 알고리즘에 포함되었다.

<표 3>은 본 연구에서 수집된 일곱 가지의 감정세트별 두 음절 시드 통계와 사용 빈도가 높은 감정 단어의 추가 수집을 위한 세 음절 확장 시드 통계이다. “Surprised” 감정 세트가 두 음절보다 125배 증가 하였으며 “Happy” 감정 세트도 106배 증가한 것으로 나타났다. 일곱 가지 감정 세트가 두 음절 시드와 세 음절 시드로 확장 했을 때 평균 96배 증가한 것으로 나타났다. 최종 수집된 감정단어 수를 기준으로 초기 두 음절 시드를 비교해보면 “Happy” 감정 세트가 633배 많은 감정 단어가 수집된 것으로 나타났다. “Fearful” 감정 세트가 583배로 그 뒤를

이었으며, “Disgusted” 감정 세트가 331배로 가장 적게 수집된 것으로 나타났다. 감정 시드 100개로 실험한 “Angry” 감정세트는 4만개의 감정단어가 수집되어 초기 감정 시드에서 400배 증가되었으며, “Sad”, “Bad” 감정 세트 역시 각각 440배, 560배 이상의 감정 단어가 수집되었다. <표 1>의 자극의 변화에 따른 감정 시드의 해시태그에서 600가지 이상 파생된 단어를 수집하는데 활용되었다고 볼 수 있다. 이렇듯 네티즌들의 감정 표현이 그만큼 다양하며 본 연구에서 개발된 수집 시스템의 성능을 검증한다고 볼 수 있다.

V. 결론 및 향후 과제

본 연구는 감정 세트에 필요한 감정단어를 네티즌들이 자유롭게 사용하는 해시태그를 활용하여 연구하였다. 2017년 한국인터넷진흥원의 발표에 따르면 SNS 이용자는 일주일 평균 23회 이상 SNS를 사용하는 것으로 조사되었다

(한국인터넷진흥원, 2017). 이처럼 모바일 환경과 함께 살아가는 현대인들의 SNS 생활은 밀접한 관계가 있으며 그들이 사용하는 언어들에 대한 연구가 지속적으로 이루어져야 한다. SNS 생활 속에는 다양한 감정 표현은 물론 원하고 바라는 그들의 이야기들이 묻어져 있다.

본 연구 과정과 의미를 종합적으로 정리하면 다음과 같다.

첫째, 본 연구에서는 네티즌들의 일상을 표현하고 있는 SNS에서 텍스트 데이터를 중심으로 한글 감정 연구를 진행하였다. 영어 기반의 플러치(1980)의 감정 기준을 한글로 재해석하고 SNS상의 함축적 의미가 있는 해시태그를 활용하였다. SNS 사용자간 의미 있는 해시태그의 감정 단어를 이용하여 ‘행복’, ‘슬픈’, ‘억울’, ‘화난’, ‘두려운’, ‘나쁜’, ‘놀란’ 등 일곱 가지의 감정 세트로 구분하여 각 세트별 감정 시드를 선정하였다. 한글 자연어 처리(KoNLP) 기반 감정 사전에 필요한 단어를 수집하기 위한 시스템을 개발한 것이다. 본 연구의 결과물은 수집 시스템을 이용하여 <표 3>의 결과를 집계하였고 일곱 감정 세트 내 327개의 시드를 활용하여 155,342개의 감정 단어를 배양하여 초기 시드에서 470배 이상의 단어를 검출하여 정리하였다. 이러한 150만개의 감정 단어를 활용하여 네티즌들의 다양한 감정 분석이 가능할 것으로 기대된다.

둘째, 본 연구의 수집 시스템은 웹 브라우저의 자동 완성 기능과 한글의 자음과 모음을 분리하는 한글 자연어 처리 알고리즘을 융합한 결과물이다. 연구에서 감정 세트에 필요한 감정 단어 수집 알고리즘을 살펴보면 마치 연구자가 직접 웹 서핑을 하듯 가상 웹브라우저인 크롬

웹드라이버를 활용하여 특정 SNS 접속 후 검색창의 자동완성 텍스트를 수집하였다. 준비된 감정 시드를 감정 세트 별로 구분하였으며 일곱 가지 세트를 순차적으로 시스템에 로딩하여 감정 단어를 수집하도록 하였다. 검색창에 입력된 감정 시드의 자동완성기능을 이용하여 자동완성 단어에 설정되어 있는 ‘a’태그의 ‘href’ 속성의 값에서 해시태그 텍스트만을 추출하였다.

셋째, 본 연구의 진행과정은 오픈 소스 소프트웨어를 활용하여 구현하였다. 리눅스 서버에 크롬 웹드라이버를 활용하고 SNS 웹 페이지 소스에서 데이터 패턴을 찾고 수집 가능한 API를 개발하기 위한 도구로 파이썬(V. 3.6) 개발 환경을 이용하였다. 마이닝처리된 데이터에서 해시태그만을 추출하기위한 작업은 오픈 소스 소프트웨어인 R 프로그램을 활용하였으며 크롤링 내용의 공백 및 특수문자 등을 활용하여 태그 분리하는 작업과 정리하는 처리 과정을 함께 진행하였다.

넷째, SNS 콘텐츠의 해시태그를 이용한 감정 단어 사전은 어떤 특정 제품에 대해서 느끼는 감정들의 연결 고리를 파악할 수 있다. SNS에서 사용자들이 작성한 해시태그는 본 연구에 있어서 감정 씨앗 중 하나이다. 기업관점에서 제품의 마케팅 전략 측면에서는 소비자의 다차원 감정 사전에서 적절히 평가되고 있는가 아니면 특정 감정으로 치우쳐지는가에 대한 분석이 가능하다. 온라인 사용자들이 많이 사용하는 비정형적 데이터에 감정적 판단을 할 수 있는 도구가 개발되었다고 볼 수 있다.

본 연구는 감정을 표현하는 단어들을 일곱 가지 카테고리로 선별하는 작업을 진행하였다. 하지만 인터넷상의 고객 감정 데이터를 데이터

베이스화만 진행한 상태이다. 빅데이터 분석 기술을 활용하여 고객의 감정에 어필하는 제품과 서비스를 만들 수 있는 마케팅 전략에 사용가능한 솔루션으로 추가적이고 지속적인 연구가 필요하다. 또한 현시대를 반영하여 인터넷 사용자들이 표현하는 감정이 지속적으로 변화되므로 본 연구의 감정 사전을 지속적으로 업데이트 할 수 있는 수집 시스템이 연구되어야 할 것이다. 특히, 네티즌들의 신조어 사용으로 “금사빠”, “깜놀”과 같은 글자 줄임을 활용하여 감정을 표현하는 해시태그 연구도 지속되어야 할 것이다.

참고문헌

- 남민지, 이은지, 신주현, “인스타그램 해시태그를 이용한 사용자 감정 분류 방법,” *Journal of Korea Multimedia Society* 제18권, 제11호, 2015, pp. 1391-1399.
- 박성원, “새로운 과학기술이 일하는 방식에 미치는 영향,” *과학기술정책*, 제27권, 제7호, 2017, pp. 26-31.
- 박인조, 민경환, “한국어 감정단어의 목록 작성과 차원 탐색,” *한국심리학회지: 사회 및 성격*, 제19권, 제1호, 2005, pp. 109-129.
- 서병조, 나성욱, “지능정보화 시대에 대비한 네트워크 발전전략 연구,” *한국통신학회지(정보와통신)*, 제34권, 제8호, 2017, pp. 30-37.
- 서새남, “4 차 산업혁명 주요기술에 대한 법적 고찰-한국 및 중국을 중심으로,” *문화·미디어·엔터테인먼트 법*, 제11권, 제1호, 2017, pp. 141-172.
- 손선주, 박미숙, 박지은, 손진훈, “한국어 감정 표현단어의 추출과 범주화,” *감성과학*, 제15권, 제1호, 2012, pp. 105-120.
- 안정국, 김희웅, “Building a Korean Sentiment Lexicon Using Collective Intelligence,” *지능정보연구*, 제21권, 제2호, 2015, pp. 49-67.
- 안희정, 김기원, 김승훈, “도서 정보 및 본문 텍스트 통합 마이닝 기반 사용자 맞춤형 도서 큐레이션 시스템,” *Journal of Information Technology Applications & Management*, 제24권, 제1호, 2017, pp. 33-43.
- 유은지, 김정철, 이춘열, 김남규, “시맨틱 텍스트 마이닝을 위한 온톨로지 활용 방안,” *정보시스템연구*, 제21권, 제3호, 2012, pp. 137-161.
- 윤여경, “SNS 해시태그 방식에 대한 비교평가 UI 디자인 연구,” *한국상품문화디자인학회 논문집*, 제46권, 2016, pp. 103-113.
- 이수상, “한글 감정단어의 의미적 관계와 범주 분석에 관한 연구,” *한국도서관·정보학회지* 제47권, 제2호, 2016, pp. 51-70.
- 이영재, 이성수, “텍스트마이닝 기반의 인적재난사고사례 신뢰도 측정연구,” *정보시스템연구*, 제20권, 제3호, 2011, pp. 63-79.
- 이종화, 레환수, 이현규, “소셜네트워크서비스에 활용할 비표준어 한글 처리 방법 연구,” *한국산업정보학회논문지*, 제21권,

- 제3호, 2016, pp. 35-46.
- 이중화, 이현규, “오픈소스 소프트웨어를 활용한 자연어 처리 패키지 제작에 관한 연구,” 정보시스템연구, 제25권, 제4호, 2016, pp. 121-139.
- 이준웅, 송현주, 나은경, 김현석, “정서 단어 분류를 통한 정서의 구성 차원 및 위계적 범주에 관한 연구,” 한국언론학보, 제52권, 제1호, 2009, pp. 85-116.
- 장문수, “심리학적 감정과 소셜 웹 자료를 이용한 감성의 실증적 분류,” 한국지능시스템학회논문지, 제22권, 제5호, 2012, pp. 563-569.
- 정소영, “빅데이터 시대에 카드뉴스의 유형과 활용 현황 분석,” 한국디자인문화학회지, 제21권, 제4호, 2015, pp. 609-620.
- Baccianella, S., Esuli, A. and Sebastiani, F., “Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining,” *In LREC*, Vol. 10, No. 2010, pp. 2200-2204.
- Gojare, S., Joshi, R. and Gaigaware, D. “Analysis and Design of Selenium WebDriver Automation Testing Framework,” *Procedia Computer Science*, Vol. 50, 2015, pp. 341-346.
- Kaur, H. and Gupta, G. “Comparative study of automated testing tools: selenium, quick test professional and testcomplete,” *Journal of Engineering Research and Applications ISSN*, 0132 pp. 2248-9622.
- Kouper, I., “Professional participation in digital curation,” *Library & Information Science Research*, Vol. 38, No. 3, 2016, pp. 212-223.
- Plutchik, R., “*Emotion: A psychoevolutionary synthesis*,” Harpercollins College Division, 1980.
- Russell, J. A., “A circumplex model of affect,” *Journal of Personality and Social Psychology*, Vol. 39, No. 6, 1980, pp. 1161-1178.
- Strapparava, C. and Valitutti, A., “Wordnet affect: an affective extension of wordnet,” *In Lrec*, Vol. 4, 2004, pp. 1083-1086.
- Thayer, R. E., “*The Biopsychology of Mood and Arousal*,” New York: Oxford University Press, 1989.
- Wu, Y., Wang, N., Kropczynski, J. and Carroll, J. M., “The appropriation of GitHub for curation,” *PeerJ Computer Science*, Vol. 3, 2017, p. 134.

이 종 화 (Lee, Jong-Hwa)



부경대학교 경영학 박사과정을 취득하고, 현재 부경대학교 경영학부 겸임강사로 재직 중이다. 주요 관심분야는 BigData, Mining, Content Analysis 등이다.

이 윤 재 (Lee, Yun-Jae)



국립공주대학교 산업공학 박사학위를 취득하고, 현재 SK Telecom MNO사업부 재직 중이다. 주요 관심분야는 정보전략, 전략마케팅, 경영 전략, BigData 등이다.

이 현 규 (Lee, Hyun-Kyu)



연세대학교에서 경영학 박사학위를 취득하고, 현재 부경대학교 경영학부 교수로 재직하고 있으며, 주요 관심분야는 정보시스템전략, Data-Mining & Analysis 등이다.

<Abstract>

Development of Emotional Word Collection System using Hash Tag of SNS

Lee, Jong-Hwa · Lee, Yun-Jae · Lee, Hyun-Kyu

Purpose

As the amount of data became enormous, it became a time when more efforts were needed to find the necessary information. Curation is a new term similarly to the museum curator, which is a service that helps people to collect, share, and value the contents of the Internet.

In SNS, hash tag is used for emotional vocabulary to be transmitted between users by using (#) tag.

Design/methodology/approach

As the amount of data became enormous, it became a time when more efforts were needed to find the necessary information. Curation is a new term similarly to the museum curator, which is a service that helps people to collect, share, and value the contents of the Internet.

In SNS, hash tag is used for emotional vocabulary to be transmitted between users by using (#) tag.

Findings

This study base on seven emotional sets such as 'Happy', 'Angry', 'Sad', 'Bad', 'Fearful', 'Surprised', 'Disgusted' to construct 327 emotional seeds and utilize the autofill function of web browser to collect 1.5 million emotional words from emotional seeds. The emotional dictionary of this study is considered to be meaningful as a tool to make emotional judgment from unstructured data.

Keyword: Emotional Word, HashTag, Text Mining, NLP, Bigdata

* 이 논문은 2018년 5월 31일 접수, 2018년 6월 20일 1차 심사, 2018년 6월 28일 게재 확정되었습니다.