

소셜빅데이터를 이용한 온라인 소비자감성지수(e-CCSI) 개발

Electronic-Composit Consumer Sentiment Index(CCSI) development by Social Bigdata Analysis

김 유 신¹ 홍 성 관² 강 희 주³ 정 승 렬^{2*}
Yoosin Kim Sung-Gwan Hong Hee-Joo Kang Seung-Ryul Jeong

요 약

인터넷, 소셜미디어, 모바일 등의 등장 이후 소비자들은 개인의 의견을 온라인을 통해 적극적으로 표명하기 시작했고 이는 확산 또한 실시간으로 이루어지고 있다. 인터넷 기반의 다양한 커뮤니케이션 활동들을 통해 생산되는 텍스트는 인터넷을 사용하는 사용자들이 공유하고 공감하는 자원으로서 단순한 소통의 도구를 넘어 분석의 가치가 있는 새로운 정보의 창고가 되고 있다. 세계 각국의 정부와 기업은 인터넷과 소셜미디어를 통해 생산되는 소셜 빅데이터를 활용하여 사회/경제적 문제의 해결과 정부의 정책을 효과적으로 추진하는데 적극 활용하고 있다. 특히 경제와 관련해서는 경기현황에 대한 경제주체들의 움직임을 보다 빠르고 정확하게 분석하고 예측하여 이에 알맞은 정책을 적기에 실시할 수 있도록 다각적인 활동을 펼치고 있다. 본 연구에서는 소셜 미디어에 내재된 소비자의 직접적이고 즉시성 있는 의견을 경제적 측면에서 활용할 수 있는 온라인 소비자감성지수 모형을 제시하고 구현하였다. 이를 위해 한국은행의 소비자자동향조사(CSI)와 소비자심리지수(CCSI)를 온라인으로 수행할 수 있는 어휘분류체계(온톨로지)와 감성사전을 구축하고 감성분석을 실시하여 생활형편, 경제상황, 소비와 수입 4가지 영역의 소셜감성지수를 도출하였다. 또한 이들을 결합한 온라인 소비자감성지수(e-CCSI)를 개발하고 소비자심리지수와 비교를 통해 유용성을 확인하였다.

☞ 주제어 : 온라인소비자심리지수, 텍스트마이닝, 온톨로지, 감성분석

ABSTRACT

With emergence of Internet, social media, and mobile service, the consumers have actively presented their opinions and sentiment, and then it is spreading out real time as well. The user-generated text data on the Internet and social media is not only the communication text among the users but also the valuable resource to be analyzed for knowing the users' intent and sentiment. In special, economic participants have strongly asked that the social big data and its' analytics supports to recognize and forecast the economic trend in future. In this regard, the governments and the businesses are trying to apply the social big data into making the social and economic solutions. Therefore, this study aims to reveal the capability of social big data analysis for the economic use. The research proposed a social big data analysis model and an online consumer sentiment index. To test the model and index, the researchers developed an economic survey ontology, defined a sentiment dictionary for sentiment analysis, conducted classification and sentiment analysis, and calculated the online consumer sentiment index. In addition, the online consumer sentiment index was compared and validated with the composite consumer survey index of the Bank of Korea.

☞ keyword : e-Consumer Sentiment Index, Text Mining, Ontology, Sentiment Analysis

1. 서 론

인터넷, 소셜미디어, 모바일 등의 등장 이후 소비자들

은 개인의 의견을 온라인을 통해 적극적으로 표명하기 시작했고 이는 확산 또한 실시간으로 이루어지고 있다. 인터넷 블로그, 카페 등의 온라인 커뮤니티를 비롯하여 트위터, 페이스북 등의 소셜네트워크서비스(SNS)까지 인터넷 기반의 다양한 커뮤니케이션 활동들을 통해 생산되는 텍스트는 인터넷을 사용하는 사용자들이 공유하고 공감하는 자원으로서 단순한 소통의 도구를 넘어 분석의 가치가 있는 새로운 정보의 창고가 되고 있다. 이러한 온라인 소셜미디어에서 생산되고 확산되는 데이터들에는

1 BigData Analytics Dept., Univ. of Seoul, Seoul,02504,Korea
2 Graduate School of Business IT, Kookmin Univ., Seoul, 02727, Korea
3 DataScience Center, Funnywork Corp., Seoul, 04165,Korea
* Corresponding author (srjeong@kookmin.ac.kr)
[Received 25 July 2017, Reviewed 26 July 2017(R2 14 August 2017), Accepted 17 August 2017]

소비자의 생각, 관심사, 라이프스타일, 선호 등 개인의 특정 기록들을 담고 있으며, 사용자들이 쌍방향 정보를 만들어 소통하는 형식이기 때문에, 소비자의 기호를 파악하여 마케팅에 적용할 수도 있고, 주식시장의 감성을 파악해 주가의 등락을 예측할 수도 있으며, 정치 선거에서 후보자의 당락을 예측할 수도 있다[1].

이러한 변화를 감지한 세계 각국의 정부와 기업은 인터넷과 소셜미디어를 통해 생산되는 소셜 빅데이터를 활용하여 사회/경제적 문제의 해결과 정부의 정책을 효과적으로 추진하는데 적극 활용하고 있다[2]. 특히 경제와 관련해서는 경기현황에 대한 경제주체들의 움직임을 보다 빠르고 정확하게 분석하고 예측하여 이에 알맞은 정책을 적기에 실시할 수 있도록 다각적인 활동을 펼치고 있다. 이러한 노력의 일환으로 소비자가 체감하는 경기현황을 설문조사와 인터뷰를 통해 조사하고 이를 지수화하는 소비자동향조사와 소비자심리지수 등이 대표적이다. 그러나 이러한 설문조사 방식은 소비자의 직접적인 생각을 파악할 수 있다는 장점이 있지만 한계점도 있다. 설문조사 항목에 대한 응답은 의식적인 판단이 개입되며, 무의식적인 마음에 대해서는 파악하기 어려운 면이 있으며, 조사에 소요되는 시간과 비용 문제도 간과할 수가 없다. 때문에 인터넷과 소셜미디어에서 소비자들이 직접 표출하는 개인의 의견은 소비자의 생각과 마음을 파악하는데 매우 유용한 데이터가 될 수 있으며, 이를 실시간으로 수집하고 분석한다면 설문조사와 분석에 소요되는 시간 또한 획기적으로 단축시킬 수 있을 것이다.

이에 본 연구는 기존의 설문조사 방식으로 이루어지던 소비자심리지수를 소셜미디어 빅데이터 분석을 통해 온라인 소비자감성지수를 개발하고자 한다. 소셜미디어 속에서 표출되는 소비자의 의견은 비정형 텍스트 데이터로 이들 가운데 의미있는 정보를 추출하고 수집하기 위해서는 분석 주제와 목표에 부합하는 텍스트분류체계와 어휘 자원이 필요하다. 이를 위해 사용되는 것이 온톨로지인데, 온톨로지는 해당 개념을 명시적으로 정의하고, 컴퓨터가 처리할 수 있는 형태로 표현하는 용어의 논리적인 집합 이면서 개념 간 관계를 명시한 사전의 역할을 의미한다. 그러므로 본 연구를 위해서 소비자들이 경제현황에 대해 언급하고 의견을 표출하는 데이터를 수집/분류/분석하기 위해 경기관련 온톨로지와 감성사전을 개발하고 이를 이용하여 소셜데이터의 분류와 감성분석을 실시하였다.

본 논문의 구성은 다음과 같다. 2장에서는 경기현황 조사와 관련한 선행연구들과 지표들을 살펴보고, 텍스트 데이터 분석을 위한 온톨로지와 감성분석의 개념을 정리

하였다. 이러한 선행연구를 기반으로 3장에서는 소셜미디어 데이터로부터 경기관련 온톨로지와 감성사전을 개발하고 분석하는 연구모형을 제시하였으며, 4장에서는 소비자심리지수(CCSI)와 대응되는 온라인 소비자감성지수(e-CCSI)의 개별항목 감성지수와 합성지수 편제과정을 기술하고 CCSI와 비교분석한다. 마지막 5장에서는 본 연구의 결론과 향후 연구방향을 제시한다.

2. 관련연구

2.1 소비자동향조사와 소비자심리지수(CCSI)

한국은행 및 통계청을 비롯한 정부기관과 경제단체들은 국가 경제의 현황을 파악하고 경기의 흐름을 빠르고 정확하게 포착하고자 다양한 지표들을 개발하여 활용하고 있다. 이러한 경제현황 파악에 사용되는 지표들은 개별 경제지표나 종합경기지표와 양적 통계를 기반으로 하거나, 경제심리지표와 같은 질적인 정보가 활용되기도 한다. 이 가운데 경제심리지표들은 전통적인 계수통계조사보다 경기순환변동에 더 민감하고 신속하게 포착할 수 있으며 전체 업황 전망, 소비지출계획 등 계수통계조사로는 조사가 어려운 정보를 획득할 수 있다는 장점으로 인해 경기 동향의 판단 및 예측의 주요지표로서 활용되고 있다.

소비자동향지수(CSI: Consumer Survey Index) 역시 이러한 경제심리지표의 하나로서 소비자의 경제에 대한 현재 인식과 향후 전망 내용 등을 설문조사하는 소비자동향조사 결과를 지수화한 것이다[3]. 소비자동향조사는 소비지출규모 등 소비관련 양적 통계를 보완하는 경기판단 및 전망지표로 폭넓게 활용되고 있는데, 고정된 조사 대상자에게 정기적으로 설문조사를 실시함으로써 응답거부율이 낮고 신속하게 조사가 이루어진다는 장점을 가지고 있다. 또한 기존의 계수 통계보다 넓은 범위의 정보를 제공하고 경기순환적인 측면을 쉽게 발견할 수 있어 경기순환에 대한 판단 및 예측에 적합한 지표로 활용되고 있다.

소비자동향조사는 1995년부터 실시되고 1998년부터 그 결과가 공표되었는데, 조사기관-대상 가구수-조사항목 등이 조정되어 몇 차례 이루어진 후 2008년 9월부터 한국은행이 전담하여 전국 시도 19개 지역의 도시 일반가구를 대표하는 표본가구 2,200가구를 대상으로 매월 15일을 전후하여 5가지 척도를 가진 20여개 항목으로 구성된 설문 조사를 실시하고 응답결과를 지수화 하여 매월 말일전후로 그 결과를 발표하고 있다.

소비자동향조사에서 조사되는 항목은 크게는 다섯 가지 영역의 20여 가지 항목으로 구성되어 있다. 먼저 경제 인식에 대한 조사에서는 현재 가계의 생활형편과 국내 경기에 대한 판단을 6개월 전과 비교한다. 다음으로 경제 전망은 향후 6개월 후의 생활형편과 국내경기, 취업기회, 금리수준을 묻는다. 소비지출전망 조사는 가계수입과 소비지출이 현재와 비교하여 어떻게 변화할지를 조사하며, 가계저축 및 부채에 대한 조사는 6개월 전과 비교하여 현재의 상황, 그리고 6개월 후의 예상치를 질의한다. 마지막으로 물가전망에 대한 조사는 현재와 비교하여 1년 후의 소비자물가 수준, 주택가격 수준, 근로임금 수준 전망과 물가상승의 1년 전후 상황을 조사한다.

소비자동향조사의 결과는 소비자동향지수(CSI)로 공표되는데, 이는 각 설문항목에 대한 소비자의 응답결과를 일정가중치를 부여한 지수형태로 산출한 것으로서 총 24개의 개별 소비자동향지수 항목이 있다. CSI 산출 방식은 5점 척도의 조사항목에 대해서 각 척도별로 가중치를 부여하여 응답수와 곱한 후 전체 응답소비자수로 나누고 100점으로 환산한다(그림 1 참조).

$$CSI = \frac{\text{매우긍정} \times 1.0 + \text{다소긍정} \times 0.5 + \text{비중} \times 0.0 - \text{다소부정} \times 0.5 - \text{매우부정} \times 1.0}{\text{전체 응답 소비자수}} \times 100 + 100$$

(그림 1) 소비자동향지수

(Figure 1) Consumer Survey Index

소비자동향지수는 산출식에서 알 수 있듯이 100을 기준으로 하며, 값이 0에서 200사이에서 계산되도록 되어 있다. 이는 100을 기준으로 그 이상일 때는 상승 또는 증가, 긍정의 의미를 가지며, 반대로 100 이하로 내려갈 때는 하락, 감소, 부정의 의미로 해석될 수 있다.

한편 CSI는 생활형편, 경제상황 등 개별 항목별로 소비자동향을 표현하는 지표이므로 각 항목에 대한 소비자들의 태도는 잘 나타내는데 반해 개별지수 간 결과가 상충되거나 또는 소비자의 심리를 대표하는 지수로 활용하기에는 어려움이 있다. 이에 조사기관인 한국은행에서는 2005년부터 소비상황, 경기인식 등에 대한 소비자 태도를 종합적으로 파악할 수 있는 대표지수로서 개별 항목 중 대표도가 높은 6개 항목을 합성하여 소비자심리지수(CCSI: Composite Consumer Sentiment Index)를 편제하여 발표하고 있다.

소비자심리지수(CCSI)에 합성되는 6개 개별 CSI는 경제지표와 상관성이 높고 선행성이 우수한 항목들로서 경제상황(현재경기판단지수, 향후경기전망지수), 생활형편

(현재생활형편지수, 생활형편전망지수), 소득 및 소비(가계수입전망지수, 소비지출전망지수)에서 선별되었다. 각 개별지수는 어느 하나의 지수가 과도한 영향을 미치지 않도록 표준화 과정을 거친 후 단순 합산하여 장기표준화구간 내에서 평균이 100, 표준편차 10이 되도록 재 표준화 후 최종적으로 100을 기준으로 지수화 된다. 결국 소비자심리지수는 현재의 장기 표준화구간인 2003년 1월부터 전년 12월까지의 소비자심리의 장기평균치 100과 대비하여 상승과 하락으로 해석될 수 있다. 즉 CCSI가 100보다 클 경우 평균적인 경기상황보다 긍정적임을 그 반대일 경우 부정적으로 해석할 수 있다.

2.2 빅데이터 기반 경제동향지표의 등장

PC와 인터넷, 스마트폰, 모바일, 소셜미디어 등의 출현 이후 인류의 생활방식은 급격한 변화를 맞고 있다. 이들은 사람들로 하여금 정보를 생산하고 유통하고 취득하는 방식뿐만 아니라 상호 소통하는 방식에서도 근본적 변화를 가져왔다. 사람들은 언론매체의 뉴스를 포털웹사이트(네이버, 다음) 또는 소셜네트워크서비스(페이스북, 트위터)에서 접하고 이에 대한 의견을 개진하고 또 누군가와 온라인으로 즉시 공유하는 등 정보의 생성과 전파가 지구촌 어디로든지 실시간으로 이루어지고 있다.

이러한 변화는 경제 환경에도 지대한 영향을 미치게 되어, 지구 건너편에서 발생한 정치 스캔들이 국내 주식 시장에 바로 영향을 미치고, 경제 환경의 변화들이 다시 또 실시간으로 온라인에서 확산 전파되는 실정이다. 실시간 정보 생산/유통의 시대에서는 생산되는 데이터의 양 또한 방대하여 이제는 바야흐로 빅데이터 시대가 분명 도래하였고, 그간의 정형데이터와 설문조사 위주의 경제동향분석에서도 빅데이터를 이용하여 경기의 판단근거를 찾고자 하는 노력들이 생겨나고 있다[4-6]. 이러한 연구들은 특히나 새로운 기술시대에서 급변하는 경제환경 변화를 기존의 경제지표에서는 즉시 반영할 수 없으므로, 그 대안으로서 온라인 검색서비스에서의 경제 키워드 검색[7], 소셜미디어에서의 경제상황 메시지와 감성분석[4,5]을 통해 보완 또는 대체할 수 있음을 제시하였다.

먼저, Daas와 Puts의 연구에 따르면[4], 네덜란드의 소비자컨피던스(Consumer confidence)와 소셜미디어 감성(Sentiment in Dutch Social Media) 사이의 상관관계를 분석하였다. 네덜란드의 소비자컨피던스는 한국의 소비자심리지수와 매우 유사한 경제심리지표로서 매월 1000가구를 대상으로 경제현황과 재무상황에 대한 5가지의 설문

에 대해 긍정-부정-중립으로 응답하는 설문조사 방식의 소비자동향조사결과이다. 연구자들은 트위터, 페이스북, 링크드인과 같은 SNS와 웹사이트, 포럼, 블로그 등에서 생성되는 네덜란드 메시지들을 모두 포함한 소셜미디어 빅데이터를 수집하고 감성분석을 통해 지수를 도출하였다. 분석결과에서는 네덜란드 소비자자만피던스와 소셜미디어 감성지수사이에 분명한 연관성이 있음이 확인되었으며, 특히 페이스북 메시지는 0.85의 상관계수로 가장 통찰력있는 소셜미디어였으며, 트위터가 그 뒤를 잇고 있음을 보여주었다. 그 외의 다양한 실험결과와 함께 결과적으로 소셜미디어 감성분석결과가 매월 공식 발표되는 소비자자만피던스보다 소비자의 심리변화를 선형하여 보여줄 수 있으며 즉시성 측면에서도 더욱 유용함을 주장하고 있다.

위와 유사한 연구가 한국 통계청에서도 실시되었다[5]. 본 연구에서는 소셜미디어에서 발생하는 경제상황 메시지를 이용하여 산출한 소셜미디어 감성지수와 한국은행에서 발표하고 있는 소비자 심리지수와와의 비교분석을 진행하였다. 소셜미디어 감성지수를 산출하기 위하여 국내 포털사이트(네이버/다음)의 블로그, 게시판, 뉴스, 카페에서 소비자 심리지수에 대응하는 검색어 경제, 가게, 수입, 소비, 지출, 살림살이, 생활형편의 키워드를 이용하여 데이터를 수집하고 이를 이용하여 감성분석을 진행하여 소셜미디어 감성지수를 도출하였다. 분석결과에서는 소셜미디어 감성지수와 소비자심리지수의 비교결과 역시 상관관계수가 0.734로 강한 상관관계를 나타냈다. 그리고 특정 기간 월별 집계된 소셜미디어 감성지수와 소비자 심리지수는 추세가 유사하게 나타났으며 주별 소셜미디어 감성지수에서 세월호 사건 발생, 메르스 사망자 발생 주를 기점으로 감성지수의 하락이 크게 나타나는 것을 볼 수 있었다. 본 연구 결과 주 단위 소셜미디어 감성지수는 월 단위로 발표되는 소비자심리지수에 비해 소비자심리지수가 포함하는 경제상황, 생활형편, 수입 및 소비지출에 대하여 더 빠른 정보를 제공할 수 있을 것으로 보고 있다.

또 다른 연구에서는 네이버에서 검색된 조회건수를 이용한 연구가 있다[6]. 본 연구는 국내 포털 점유율이 가장 높은 네이버의 네이버 트렌드를 이용하여 네이버 검색 경기지수를 만들고 경제 심리지수, 경기 실사지수와 같은 경기판단 지표들과의 분석을 진행하였다. 네이버 트렌드는 네이버에서 제공하고 있는 통합검색에서 발생하는 검색어 통계를 제공하는 서비스로써 PC와 모바일을 이용하여 네이버에서 검색한 검색어들의 통계를 검색어의 빈도

의 합이 아니라 검색키워드가 검색에서 가장 많이 검색된 주간 평균값을 100으로 보고 나머지 기간의 평균 검색횟수를 상대적으로 표현하여 0~100사이의 상대값으로 환산하여 그래프로 보여주게 된다. 이를 이용하여 호황과 불황관련 검색어군을 이용하여 네이버 트렌드의 검색통계를 수집하고 이를 이용하여 네이버 검색 경기지수를 도출하였다. 분석결과에서는 네이버 검색 경기지수와 기존의 경제지표들과의 상관관계 분석을 진행하였으며 이때 네이버 검색 경기지수는 경기 실사지수와 상관관계가 높게 나타났으며 경기실사지수와 소비자심리지수를 혼합한 경제심리지수와 상관관계가 가장 높게 나온 것을 알 수 있었다. 또한 교차상관분석을 진행한 결과 네이버 검색 경기지수는 경제심리지수와 동행하면서 상관관계수가 0.96으로 매우 높게 나타나 경제심리지수를 대체할 수 있을 것으로 보이고 있으며 예측력 분석에서는 경제성장률과 민간소비증감률이 네이버 검색 경기지수와 어느 정도 밀접한 관계를 보이고 있어 네이버 검색 경기지수가 경제성장률과 민간소비의 당기 예측에 유용하게 이용될 수 있음을 알 수 있다. 본 연구 결과 경제심리지수와 매우 밀접하게 움직이는 네이버 검색 경기지수는 주별로 작성되며 빠르게 데이터를 구할 수 있어 경제심리지수를 보완할 수 있는 지표로 활용될 수 있는 것으로 판단된다.

소셜 빅데이터를 이용한 새로운 유형의 연구들은 비단 경기현황과 경제환경에 대한 분석 및 예측에 한정되지 않고 뉴스콘텐츠를 분석한 주가예측, 영화의 리뷰를 분석해서 흥행실적을 예측하고, 여행/숙박 고객이 남긴 의견들을 분석하여 서비스 개선을 도출하는 연구, 흡연과 같은 사회갈등 이슈에 대한 분석 등 다양한 분야에서 적용되고 있다. 그리고 이러한 빅데이터 분석연구에서는 다양한 분석 방법들이 혼용되어 사용되고 있지만, 공통적으로 소셜미디어 데이터가 가지고 있는 데이터의 속성, 즉 비정형 텍스트 데이터라는 원본 데이터의 특성을 감안할 때 자연어를 처리하는 텍스트 마이닝이 매우 빈번히 활용되고 있다.

2.3 텍스트마이닝과 온톨로지

소셜 빅데이터를 이용하여 소비자의 경제인식과 향후 전망, 소비와 지출 현황 등을 도출하기 위해서는 여러 가지 절차와 활동들이 필요하다. 먼저 경기 관련 데이터를 수집하는 단계에서부터 불필요한 어휘 자료를 제거하는 전처리 단계, 정제된 데이터를 구성 항목에 따라 선별하는 분류 단계, 각 문서의 감성 극성을 부여하는 감성분석

단계, 그리고 마지막으로 이러한 텍스트 마이닝 결과를 수치화 하고 e-CCSI로 수치화하는 단계가 필요하다.

그러나 한편으론 이러한 텍스트 마이닝 전단계에 걸쳐 텍스트 콘텐츠에 적용하는 온톨로지 개발이 선행되어야 한다. 온톨로지는 비정형텍스트 데이터를 분석하기 위해 마이닝 과정에서 해당분야에 대한 전문적 지식을 컴퓨터가 이해할 수 있게 논리적이고 명시적으로 표현한다[8]. 의미적 모호성을 해결하고 용어의 의미를 파악함으로써 텍스트 마이닝의 품질을 향상시킬 수 있다. 하나의 용어가 여러 개념에 대응되는 문제를 해결하기에 적합한 체계라는 측면에서 텍스트 내 용어의 의미를 파악하는 장치로 사용하고 있다. 각 개념이 지닌 고유한 속성, 개념들 간의 관계 및 이들 사이의 제약조건, 지식추론을 위한 개념을 정의해야 한다. 결국 경기 분야의 지식체계를 컴퓨터가 해석하고 이해하여 처리할 수 있도록 형식화한 일종의 표준명세서이다.

이러한 온톨로지는 해당 도메인의 지식과 노하우가 반영되어야 함으로, 경기현황을 분석하기 위해서는 경제분야를 표현하는 개념들의 의미를 정의하는 온톨로지 개발 작업이 필요하다. 최근 시장 경기영역 에서도 SNS 또는 소셜미디어상의 데이터를 분석하여 전통적인 분석방법에서 얻을 수 없었던 가치 있는 정보와 새로운 통찰력을 발견하려는 시도들이 늘어나고 있다. 이러한 과정에서 온톨로지는 해당 영역의 데이터와 지식을 결합함으로써 데이터를 의미 있게 해석하는 분석틀으로써 중요한 역할을 하고 있다. 소셜미디어에서 표현되는 비정형 데이터들 가운데 의미 있는 키워드를 추출하고 수집하기 위해서는 경기동향 개념을 추출하고 해당 개념들 간의 관계를 나타내는 온톨로지가 필요하다. 온톨로지 구축 방법은 경험적 개발을 통해 유용한 것을 발견해 나가는 일반적인 방법으로 명확한 기존의 방법론을 개선해 가면서 적용해야 한다.

2.4 감성분석

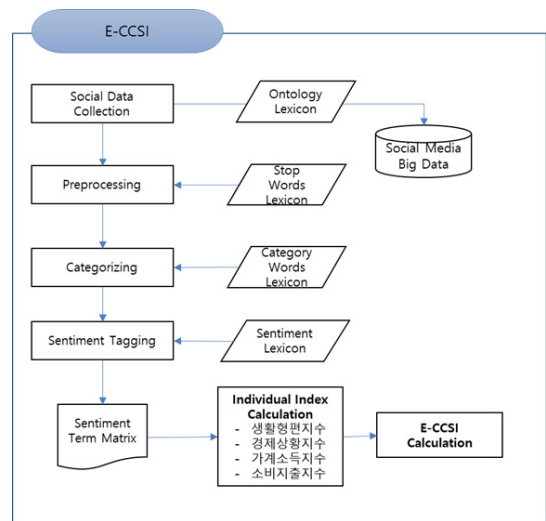
감성분석은 정형화되지 않은 텍스트로부터 사람들의 태도, 의견 혹은 성향과 같은 정보를 알아내는 기법이다. 소비자의 감정과 관련된 텍스트 정보를 자동으로 추출하는 텍스트마이닝 기술의 한영역이기도한데, 감성어휘 자원을 활용하는 방법과 기계학습을 이용한 감성분류 기법이 있다. 이 중에서 감성어휘사전을 활용하는 감성분석은 비교적 간단하면서도 정확한 방법이다. 감성사전이란 긍정적이거나 부정적인 감정을 나타내는 단어들을 모아놓

은 사전으로 국내에서도 분야별로 연구가 진행되고 있다. 감성분석은 각문서의 최소단위인 단어의 감성극성(sentiment polarity)에 따라 결정되며 단어의 감성극성이 미리 정의된 감성사전을 구축한 후, 새로 주어진 문서에 출현한 단어의 감성극성에 따라 문서 전체의 감성을 분류하게 된다[7]. 따라서 감성분석은 단어의 감성극성을 정확하게 반영한 감성사전을 사용하는 것이 중요하다. 감성사전 구축에 관한 연구로는 단어별 극성 편별에 집단 지성을 활용한 연구[9], 단어의 출현 빈도수를 활용해 주제별 감성사전을 구축한 연구[10], 주시관련 콘텐츠에 OAR(Opinion Antonym Rule) 알고리즘을 사용하여 감성사전을 구축한 연구[11], LP(Label Propagation) 방법을 이용해 단어 간의 인접도를 통해 감성사전을 구축한 연구[12] 등이 있으며 이밖에도 TF-IDF(Term Frequency-Inverse Document Frequency)등의 수치에 기반하거나 PMI 등의 통계적 수치를 활용하여 단어 간의 극성을 정의하는 방법도 있다.

3. 연구방법

3.1 온라인 소비자감성지수(e-CCSI) 개발 모형

소셜미디어 데이터를 이용하여 소비자의 경제인식과 생활형편, 수입과 부채, 소비 지출의 인식을 도출하기 위해서는 여러 가지 절차와 활동들이 필요하다. 먼저 경기



(그림 2) e-CCSI 개발 모형
(Figure 2) e-CCSI Development Model

관련 데이터를 수집하는 단계에서부터 불필요한 어휘 자료를 제거하는 전처리 단계, 정제된 데이터를 구성 항목에 따라 선별하는 분류 단계, 각 문서의 감성 극성을 부여하는 감성분석 단계, 그리고 마지막으로 이러한 감성분석 결과를 이용하여 개별 구성지수를 산출 하고 e-CCSI를 산출하는 단계가 필요하다. 그러나 한편으로 이러한 텍스트 마이닝 전단계에 걸쳐 텍스트 콘텐츠에 적용하는 온톨로지 개발이 선행되어야 한다.

경기 감성분석을 위해서는 첫째 경기와 관련된 영역으로 수집 대상과 수집 범위를 설정한 후 대상 채널(네이버 뉴스, 블로그 등)에서 수집엔진을 이용하여 수집한다. 경기와 관련 연관 키워드 그룹을 지정·분류하기 위하여 해당 토픽에 대한 이론적 배경을 분석하여 온톨로지를 개발한 후 온톨로지 키워드를 적용한다. 둘째 수집된 데이터들은 바로 사용할 수 없기 때문에 비정형데이터는 텍스트 마이닝, 오피니언 마이닝을 통하여 분류하고 정제하는 작업이 필요하다. 정제된 비정형 데이터 분석은 버즈 분석, 키워드 분석, 감성분석 등을 진행한다.

소셜 미디어에 표현된 언어들 은 일반인 들이 사용하는 대화이기 때문에 분석을 위해서는 분석틀이 필요하다. 분석틀의 내용은 관련 주제가 어떤 개념 영역으로 구성되어 있는지와 각 개념간의 관계에 대한 정의가 필요하다. 이를 반영한 온톨로지 개발이 필요하다. 수집되는 소셜 데이터들은 비정형적으로 다양하게 표현됨으로서 온톨로지를 구성하는 개념을 설명하는 용어와 그 유의어를 정의하여 용어체계를 만드는 것이 필요하다. 이렇게 개발된 온톨로지와 감성사전을 이용하여 도출된 e-CCSI는 기존의 설문조사 방식을 통해 개발된 CCSI보다 여러 가지 측면에서 장점이 있다.

(표 1) CCSI vs. E-CCSI
(Table 1) CCSI vs. E-CCSI

	CSI	E-CCSI
대상	전국 56개도시의 2,200가구	소셜 미디어를 사용하는 모든 사용자
조사 방법	우편조사에 의한 자기기입방식 및 미회수 가구에 대한 전화인터뷰 조사 병행	뉴스, 블로그, 트위터와 같은 비정형 데이터
구성 지수	현재생활형편, 생활형편전망, 가계수입전망, 소비지출전망, 현재 경기판단, 향후경기전망	생활형편, 경제상황, 가계수입 및 부채, 소비지출
공표 및 분석 기간	최소 월단위 조사표 발송은 매월 15일경 공표는 매월 25-28일경	최소 일단위 분석과 결과 발표 가능 주 또는 월단위 집계도 매우 쉽게 산출할 수 있음

3.2 경기 관련 온톨로지 개발

본 연구는 온라인상의 경기 주제에 대해 수집된 빅데이터 자료를 식별하고 활용하기 위한 분석틀로서, 경기관련 주제를 분류하고, 온톨로지와 용어체계를 개발하였다. 온톨로지는 *Ontology Development 101* 방법론[13]에 근거하여 개발하였으며 온톨로지를 구성하는 용어는 선행연구 조사에 따른 관련 문헌 및 한국은행에서 조사하는 지표를 기반으로 수집하였다. 선행연구 결과를 통하여 경기 지수 개발을 위해서는 생활형편, 경제상황, 소득 및 소비에 대한 내용을 포함하고 있음을 확인하였으며 이를 포함하는 개념의 온톨로지를 개발하였다.

온톨로지는 기존에 사용하고 있는 개념들을 재사용하여 적용하는 방법을 사용한다. 경기관련 온톨로지를 사전에 조사하여 경기 관련 영역과 일치하면 재사용할 수 있는지를 고려하여 추가적인 정보와 분류를 얻을 수 있는지 검토하여 적용한다. 본 연구에서는 7개의 선행연구를 참고하여 생활형편, 경제상황, 소득 및 소비관련 대응 어휘자원 45개를 추출하고 소득 및 소비와 관련한 전반적인 경제 용어 45개중 중복어휘2개를 선정하여 메타어휘로서의 재사용 여부를 고려하였다. 경기 상황을 대변할 추가적인 키워드를 도출하기 위해 소셜 데이터인 네이버 지식인과 뉴스를 수집하여 분석하였다.

데이터 수집을 위하여 소셜 데이터인 네이버에서 경제상황, 경기, 생활형편 등의 키워드를 선정하여 네이버 지식인에서 2,700건을 수집하였으며 약 23,000건의 키워드들의 빈도수를 기준하여 추출하였다. 경기관련 주요 어휘 선정방법은 주제별 수집 어휘 중에서 빈도분석, 연관성분석을 통해 상위 200개 어휘를 선정하여 각 주제에 맞는 어휘를 연구자들이 상호 점검을 통해 최종 어휘를 선정하였다.

클래스는 일반적으로 우리가 사용하는 사물이나 개념등에 붙이는 이름으로 경기와 관련된 용어를 열거하고 용어의 속성을 정의하는 것을 말한다. 용어를 설명하기 위해 필요한 것을 열거하여 클래스들을 선별하고 개념 정의하는데 커뮤니티 및 그 구성원은 물론 용어의 의미에 대해 합의를 이루는 것이다. 경기와 관련된 클래스들은 생활형편, 주거환경, 근로환경, 사회계층, 경제상황, 경기상황, 체감경기, 고용상황, 금리수준, 소비지출, 생활비, 공공요금, 내구재, 의류비, 외식비, 여행비, 교육비, 의료·보건비, 교양·오락·문화생활비, 교통·통신비, 주거비, 가계수입 및 부채, 가계소득, 가계지출, 가계부채로 선정하여 각 어휘에 대한 개념을 정의하였다. 클래스 계층 정의는 클래스

간의 존재하는 관계(포함관계, 상하관계, 부분관계 등)들을 정리하여 관계가 있는 클래스들은 taxonomic hierarch에 구조화하였다. 경기관련 온톨로지의 클래스는 대분류-중분류-소분류로 구분하고 각 클래스별 개념에 해당하는 용어를 영역 수준별로 정의하였다. 또한 각 용어별 동의어와 유사어는 인터넷 검색과 선행 문헌 검색 등의 방법을 이용하여 추가하였다.

3.3 감성사전 개발

감성분석에 사용되는 감성사전이 잘 개발되어있을수록 감성분석의 정확도와 신뢰도는 향상될 수 있다. 본 연구에서는 경기관련 신뢰도가 높은 감성사전을 만들기 위해서 기존의 경기관련 연구에서 기본어휘를 찾고, 이를 활용하여 감성어휘를 확장하는 방식을 이용하였다. 확장형 감성어휘를 얻는 기본적인 방법은 사전의 해설 부분을 이용하여 기본형 감성어휘의 유의어를 찾는 것으로, 사전은 오랜 시간에 걸쳐 단어의 의미를 쉽고 자세하게 기술해 놓은 것이므로 내용의 정확도에 신뢰도를 줄 수 있다. 감성사전을 개발하기 위해 감성분석 주제영역을 정의하고 선행연구에서의 재사용가능한 어휘들을 살펴보고, 이를 기준어휘로 자연어처리를 통한 빈도분석과 연관분석, 연구자 교차 검증 등을 거쳐 최종 감성사전을 개발하였다.

감성사전을 개발하기 위해 먼저 기존연구에서 재사용이 가능한지를 검토하였다. 선행연구[5]와 [6]에서 경기와 관련된 감성 어휘 149개(긍정어휘 40개, 부정어휘 109개)를 확인하고, 이를 기준어휘로 활용하였다.

추가적인 감성어휘 선정을 위해 수집데이터의 빈도분석, 연관성 분석 및 형태소 분석을 실시하였다. 빈도분석은 어휘별 출현 빈도를 계산하여 상위 노출어휘를 감성분석어휘로 활용하는 것인데, 실제 출현빈도가 높더라도 감성어휘로서 활용되지 못하는 것들이 많기 때문에 1차적인 어휘군 선별로서만 활용되었다. 연관성 분석은 소비자 감성지수의 주제영역인 생활형편, 가계수입, 소비지출, 경제상황의 기준어휘와 동시출현확률이 높은 어휘들을 선별해내는 방식이다. 예를 들어 생활형편을 키워드로 추출된 단어들의 연관성 분석을 진행하여 연관성이 0.1 이상의 모든 어휘를 도출하고, 각 어휘들이 감성어휘로 적절한지 연구자들이 크로스 체크 방식으로 선별한다. 이러한 감성어휘 추출 과정을 거쳐, 생활형편 감성어휘는 선행조사에는 긍정3개, 부정 19개가 도출되고 수집데이터의 빈도분석과 연관성분석을 이용해 추가로 선별한 결

과 긍정 73개, 부정 211개가 정의되었다.

앞서 빈도분석과 연관성분석은 R 프로그래밍의 자연어처리 KoNLP 패키지의 내장함수인 extractNoun을 이용하여 선별된 어휘를 이용하였기에 명사형태가 주를 이루었다. 한국어의 언어적 속성상 하나의 어휘라도 어미·접두사·접미사·품사변형 등을 통해 다양한 형태로 변형됨으로 이를 감성사전에 반영하기 위한 추가적으로 형태소 분석을 통한 동사·형용사 어휘 추출 작업을 실시하였다. 형태소 분석은 SimplePos09를 이용하여 9개 품사를 태그하여 동사나 형용사 어휘를 선별한 후 선행연구에서 동사, 형용사 어휘들을 비교하여 감성어휘를 선정하였다.

감성사전은 4개 구성지수 영역인 생활형편, 경제상황, 가계수입, 소비지출 부문으로 작성되었다. 각 영역별로 생활형편 425개, 경제상황 534개, 가계수입 371개, 소비지출 410개로 구축하였으며 이후 매뉴얼 점검을 통해 중복어휘와 주제 영역에 해당하지 않은 어휘를 제거하는 등의 최종 선별을 거쳐, 긍정 어휘 537개 부정어휘 1,161개로 구성된 경기관련 감성사전을 정의하였다.

(표 2) 감성사전 어휘

(Table 2) Emotion dictionary vocabulary

분류	대응어휘		
	긍정	부정	합계
생활형편	126	286	412
경제상황	161	365	526
소득	108	253	361
소비	142	257	399
합계	537	1,161	1,698

3.4 온라인 소비자감성지수(e-CCSI) 개발

온라인 소비자감성지수(e-CCSI)는 한국은행의 소비자심리지수(CCSI)와 대응하는 소셜경기지수이다. 그러므로 e-CCSI 역시 CCSI와 동일하게 생활형편, 경제상황, 소비지출 및 부채, 가계수입 4가지 구성영역으로 정의되었으며, 개별 구성지수를 산출 후 4개의 영역의 합성지수인 e-CCSI로 산출하였다. 각 구성지수를 산출하는 방식은 선행연구 [13]에서 제시된 빅데이터를 이용한 경기판단지표 개발의 감성지수를 산출식을 참고로 하였다.

$$\text{감성지수} = \frac{(\text{긍정문서수}) - (\text{부정문서수})}{\text{감성이 태깅된 문서수}} \times 100 + 100$$

(그림 3) 소셜 감성 지수

(Figure 3) Social Sentiment Index

개별 소셜감성지수 산출 후 이를 합성한 온라인 소비자감성지수(e-CCSI)를 편제하는 방식은 기존 한국은행의 CCSI 산출 방식을 동일하게 적용하였다. 합성지수 산출 방식은 개별 구성지수 평균이 0, 표준편차가 1이 되도록 감성지수를 표준화하고 이후 표준화된 개별 구성지수를 단순 합산하고 이 단순합산된 지수를 다시한번 평균이 0, 표준편차가 1이 되도록 표준화한다. 표준화된 합성지수를 장기평균 100, 표준편차가 10이 되도록 값을 재조정한다. 재조정된 지수는 기존의 지수의 범위인 0~200범위를 가지는 지수값으로 산출되며 이 값이 100보다 클 경우 긍정이며, 100보다 작은 경우는 부정을 나타내게 된다.

4. 온라인 소비자감성지수 유용성 검토

4.1 경기관련 감성분석 및 지수화

경기 관련 감성지수를 산출하기위해 수집된 문서 각각에 감성어휘 사전을 매칭하여 문서별 감성극성을 태깅한다. 이후 일자별로 감성이 태깅된 긍정, 부정문서의 수를 집계하여 개별 감성지수를 산출한다.

이렇게 태깅된 감성값은 감성지수 산출식에 따라 개별 소셜감성지수로 표준화된다.

4.2 온라인 소비자감성지수 산출

개별 구성지수는 기본적으로 일별지수이지만, 필요에 따라 주별지수나 월별지수로 변환하여 사용될 수 있다. 여기서는 CCSI와 비교를 위해서 일별지수를 월평균으로 변환작업을 실시하였다. 월평균으로 변환할 때 비교대상인 소비자심리지수와 동일하게 표준화를 하기위해서는 전년동월의 평균과 표준편차를 이용하여 표준화를 진행

(표 3) 일자별 매칭된 긍정, 부정 문서 수
(Table 3) Daily sentiment tagging results

DATE	긍정	부정	중립
2011-01-01	25	7	29
2011-01-02	58	24	92
2011-01-03	382	153	1078
2011-01-04	321	107	976
2011-01-05	278	133	960
2011-01-06	266	141	863
2011-01-07	233	120	798
2011-01-08	26	10	25
2011-01-09	67	10	111
...

(표 4) 생활형편의 일별지수 예시

(Table 4) Daily social sentiment index of living condition

DATE	INDEX
2011-01-01	156.25
2011-01-02	141.46
2011-01-03	142.80
2011-01-04	150.00
2011-01-05	135.28
2011-01-06	130.71
2011-01-07	132.01
2011-01-08	144.44
2011-01-09	174.03
2011-01-10	131.51
...	...

(표 5) 일별지수의 월평균

(Table 5) Monthly average of daily index

날짜	구성지수 (월평균)				
	소비자 심리지수	경제상황 구성지수	생활형편 구성지수	가계수입 구성지수	소비지출 구성지수
2011. 01	111	123.26	133.81	162.72	103.91
2011. 02	109	112.56	116.56	153.65	102.08
2011. 03	99	99.85	112.96	156.59	95.40
2011. 04	101	112.00	127.77	158.42	87.44
2011. 05	104	104.53	113.54	149.44	90.53
2011. 06	102	92.22	107.65	146.77	81.89
2011. 07	102	105.55	116.07	144.26	86.82
2011. 08	98	81.27	86.26	125.02	65.08
2011. 09	97	84.83	93.43	131.45	55.33
2011. 10	100	87.91	100.74	131.79	54.71
2011. 11	104	86.88	105.14	136.41	51.75
2011. 12	98	90.35	108.03	145.96	42.35
평균		98.43	110.16	145.21	76.44
표준편차		12.64	12.72	11.40	20.51

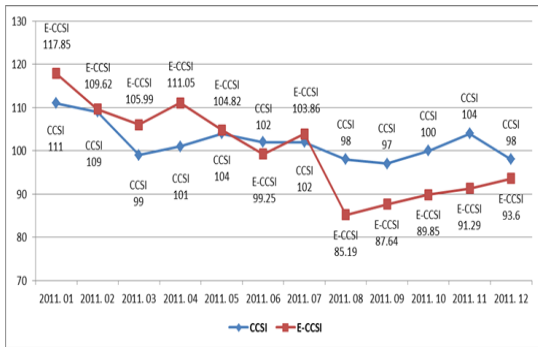
해야하지만 본 연구에서 사용한 수집 데이터가 2011년 1년간 데이터에 한정되므로, 표준화구간을 2011년 1월부터 12월까지로 한정하였다.

월평균으로 변환한 개별 구성지수를 평균이 0이고 표준편차가 1이 되도록 표준화를 만든다. 표준화식은 합성지수 산식에 의하여 계산되며 표준화 기간은 1년을 단위로 한다. 개별 표준화 지수를 산출하기 위하여 해당 지수값에 평균지수 값을 뺀 후 표준편차로 나눈다. 표준화된 합성지수를 0~200범위를 가지는 지수값으로 바꾸기 위하여 장기평균이 100, 표준편차가 0이 되도록 재조정하여 소셜 경기지수를 산출한다. 이렇게 산출된 소셜 경기지수는 100보다 작은값이 나오면 부정, 100보다 큰 값을 가지면 긍정을 의미한다.

(표 6) CCSI와 e-CCSI 비교
(Table 6) Social economic index

날짜	CCSI	E-CCSI
2011. 01	111	117.85
2011. 02	109	109.62
2011. 03	99	105.99
2011. 04	101	111.05
2011. 05	104	104.82
2011. 06	102	99.25
2011. 07	102	103.86
2011. 08	98	85.19
2011. 09	97	87.64
2011. 10	100	89.85
2011. 11	104	91.29
2011. 12	98	93.60

표 6은 소비자심리지수(CCSI)와 온라인 소비자감성지수의 상관관계 분석을 위하여 월 기준으로 지수를 표준화하여 비교한 표이다. 소비자심리지수와 온라인 소비자감성지수와 상관관계계수를 도출해본 결과 **Correlation Coefficient = 0.7153**로 비교적 강한 상관관계를 나타내었지만, 두 지표를 상호 비교하기에 12개월은 기간 상 충분치 않다는 한계가 있다.



(그림 4) 월별 e-CCSI와 CCSI 추이 그래프
(Figure 4) E-CCSI and CCSI on time series

2011년 1월부터 12월까지의 월별 집계된 e-CCSI와 CCSI의 추세가 유사한 것을 발견할 수 있다. 따라서 일별, 주별, 월별 등 즉시성있게 분석되고 공표될 수 있는 e-CCSI가 설문조사를 통해 발표하는 소비자심리지수를 보완하여 신속한 경기판단에 도움을 줄 수 있는 지표로서 유용성이 있음을 확인할 수 있다.

5. 결 론

정부기관과 경제단체들은 국가 경제의 현황을 파악하고 경기의 흐름을 빠르고 정확하게 포착하고자 다양한 지표들을 개발하여 활용하고 있다. 이러한 경제현황 파악에 사용되는 지표들 중 하나인 소비자동향지수(CSI)는 소비자의 경제에 대한 현재 인식과 향후 전망 내용 등을 설문조사하고 그 결과 지수화한 것이며, 소비자심리지수(CCSI)는 CSI의 여러 구성지표 중 생활형편, 경제상황, 소비지출 및 부채, 가계수입 4가지역의 6가지 구성 지표를 합성한 합성지표로서 경기 동향의 판단 및 예측의 주요 지표로서 활용되고 있다. 그러나 CSI와 CCSI는 한달에 한번 실시되는 설문조사 방식의 경제심리지표로 지금과 같이 시시각각 변화되는 경제 환경을 반영하기에는 시간의 간격이 존재한다. 또한 일반 소비자들이 자신들의 의견과 감정을 온라인으로 표출하고 실시간으로 소통하는 모바일·소셜미디어 시대에도 과연 2,200가구만을 조사 대상으로 한정하는 것으로 충분한지, 이를 대체 또는 보완할 수 있는 다른 방법은 없는지 그 가능성에 대한 검토가 필요한 상황이다.

SNS를 비롯하여 블로그, 카페, 인터넷 쇼핑 등에서 생성되는 텍스트 데이터에는 소비자의 생각, 관심사, 라이프스타일, 선호 등 개인의 특정 기록들이 담겨 있으므로, 이를 실시간으로 수집하고 분석한다면 설문조사에서 드러나지 않는 개인의 직접적인 의견과 감정, 선호도 등을 알 수 있을 뿐만 아니라 조사와 분석에 소요되는 시간 또한 획기적으로 단축시킬 수 있을 것이다. 이에 본 연구는 기존의 설문조사 방식으로 이루어지던 소비자심리지수를 소셜미디어 빅데이터 분석을 통해 온라인 소비자감성지수를 개발하고자 하였다.

소셜미디어 속에서 표출되는 소비자의 의견이 비정형 텍스트 데이터에 담겨있으므로, 이를 추출하고 수집하기 위한 텍스트분류체계로서 경기관련 온톨로지와 감성사전을 개발하였고, 이를 이용하여 소셜미디어 감성분석을 실시하였다. 소셜감성지수는 소비자심리지수와 동일하게 생활형편, 경제상황, 소비지출 및 부채, 가계수입 4가지 구성지수로 개발하였으며 이 4개의 영역의 구성지수를 합성하여 합성지수인 온라인 소비자감성지수를 산출하였다. 그리고 마지막으로 온라인 소비자감성지수와 한국은행의 소비자심리지수를 비교분석하였다.

결과적으로 온라인 소비자감성지수는 소셜데이터의 실시간 수집과 분석을 통해 즉시성있는 경제심리지표를 생성하고 공표할 수 있다는 장점을 가지고 있었으며, 비

록 1년으로 한정된 비교였지만, 소비자심리지수와의 높은 상관관계를 보임으로서, 소비자심리지수를 보완할 수 있는 지표로 활용될 수 있음을 확인하였다.

이러한 긍정적 결과와 함께 향후 연구에서는 보다 향상된 온라인 소비자감성지수의 개발이 요구되는데, 이를 위해서 적어도 수년 이상의 데이터를 분석하고 지수화하여 소비자심리지수 뿐만 아니라 다른 경제지표들과도 통계적인 비교 검증을 거칠 필요가 있다. 다른 한편으로는 감성사전을 개발함에 있어서 유니그램을 중심으로한 빈도분석, 연관성분석, 매뉴얼 검증을 보완할 수 있는 기술적 접근과 멀티그램 어휘자원의 검토도 이루어질 필요가 있다.

참고문헌(Reference)

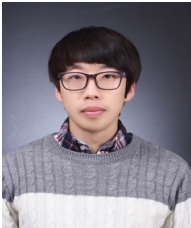
- [1] K. Lerman, "Social Information Processing in News Aggregation," *Internet Computing, IEEE*, Vol. 11, No. 6, pp.16~28, 2007.
<http://dx.doi.org/10.1109/MIC.2007.136>
- [2] H. S. Ryu, "A Study on Risk Perception and Communication via Social Media," *The Korea Institute of Public Administration, Research Report* Vol. 25, No. 3, 2013.
<https://www.kipa.re.kr>
- [3] "Comprehension of Consumer Survey Index ", *The Bank of Korea*, 2014. <http://www.bok.or.kr>
- [4] P. J. H. Daas and M. J. H. Puts, "Social Media Sentiment and Consumer Confidence," *European Central Bank, Statistics Paper Series* No 5, 2014.
<https://www.ecb.europa.eu>
- [5] Y. J. Hwang, "Consumer Survey using SNS Data", *Statistics Korea*, 2015 Research Paper, 2015.
<http://kostat.go.kr>
- [6] G. H. Lee and S. P. Hwang, "Business Cycle Indicator Using Big Data: Compilation of the Naver Search Business Index", *Economy Analysis of The Bank of Korea*, Vol. 20, No. 4, 2014. 12,
<http://kiss.kstudy.com/journal>
- [7] S. W. Kim and N. G. Kim, "A Study on the Effect of Using Sentiment Lexicon in Opinion Classification," *Journal of Intelligent Information Systems*, Vol. 20, No.1, pp.133~148, 2014.
<http://dx.doi.org/10.13088/jiis.2014.20.1.133>
- [8] S. G. No. "Framework about ontology development," *Korea Intell Inf Syst Soc*, Vol. 11, No. 1, pp.141 - 148, 2009.
<https://www.dbpia.co.kr/Journal/>
- [9] J. K. An and H. W. Kim, "Building a Korean Sentiment Lexicon Using Collective Intelligence," *Journal of Intelligent Information Systems*, Vol. 21, No. 2, pp.49~67, 2015.
<http://dx.doi.org/10.13088/jiis.2015.21.2.49>
- [10] E. J. Yu, Y. Kim, N. Y. Kim and S. R. Jeong, "Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary," *Journal of Intelligent Information Systems*, Vol. 19, No.1, pp.95~102, 2013.
<http://dx.doi.org/10.13088/jiis.2013.19.1.095>
- [11] H. J. Jo, J. H. Seo and J. T. Choi, "OAR Algorithm Technology Based on Opinion Mining Utilizing Stock News Contents," *Journal of Korean Institute of Information Technology*, Vol. 13, No. 2, pp. 111~119, 2015.
<https://www.kci.go.kr/>
- [12] J. H. Kim, Y. J. Oh and S. H. Chae, "The Construction of a Domain-Specific Sentiment Dictionary Using Graph-based Semi-supervised Learning Method," *Korean Journal of the Science of Emotion and Sensibility*, Vol. 18, No. 4, pp.97~104, 2015.
<http://kiss.kstudy.com/journal>
- [13] "Ontology Development 101: A Guide to Creating Your First Ontology," *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05*, March 2001. <http://www.ksl.stanford.edu/>

● 저 자 소 개 ●



김 유 신(Yoosin Kim)

2000년 국민대학교 정보관리학부 졸업(학사)
2009년 국민대학교 비즈니스IT전문대학원 졸업(석사)
2013년 국민대학교 비즈니스IT전문대학원 졸업(박사)
2016년~현재 서울시립대학교 융합전공학부 빅데이터분석학 객원교수
관심분야 : 텍스트 마이닝, AI 빅데이터 분석 및 활용, 시스템 구현 etc
E-mail : yoosin25@uos.ac.kr



홍 성 관(Sung-Gwan Hong)

2015년 강원대학교 컴퓨터정보통신공학전공 졸업(학사)
2016년~국민대학교 비즈니스IT전문대학원 재학(석사)
관심분야 : 텍스트 마이닝, 빅데이터 분석 및 활용 etc
E-mail : play_w7@kookmin.ac.kr



강 희 주(Hee-Joo Kang)

1996년 서경대학교 전산통계학과 졸업(학사)
2005년 국민대학교 전자계산교육 졸업(석사)
2011년 국민대학교 비즈니스IT전문대학원 졸업(박사)
2005년~현재 명지전문대학교 보건의료정보과 외래교수
2017년~(주)퍼니웍 수석 연구원
관심분야 : 프로젝트관리, 예측모델, 텍스트 마이닝, AI 빅데이터 분석 및 활용 etc
E-mail : joyriver316@fuunywork.co.kr



정 승 렬(Seung-Ryul Jeong)

1985년 서강대학교 경제학과 졸업(학사)
1989년 미국 위스컨신 대학교 대학원 경영정보학과 졸업(석사)
1995년 미국 사우스캐롤라이나 대학교 대학원 경영정보학과 졸업(박사)
1997년~현재 국민대학교 비즈니스IT전문대학원 교수
관심분야 : 정보시스템 구현, 프로세스 관리, 프로젝트 관리, 정보자원 관리 etc.
E-mail : srjeong@kookmin.ac.kr