



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위 논문

인터넷 매체가 가상화폐 가격에  
미치는 영향에 관한 연구

A Study on the Effect of Internet  
Media on Virtual Currency Price

2018년 6월

승실대학교 대학원

IT융합학과

정 성 엽



석사학위 논문

인터넷 매체가 가상화폐 가격에  
미치는 영향에 관한 연구

A Study on the Effect of Internet  
Media on Virtual Currency Price

2018년 6월

숭실대학교 대학원

IT융합학과

정 성 엽

석사학위 논문

인터넷 매체가 가상화폐 가격에  
미치는 영향에 관한 연구

지도교수 정 윤 원

이 논문을 석사학위 논문으로 제출함

2018년 6월

숭실대학교 대학원

IT융합학과

정 성 엽

정 성 엽 의 석 사 학 위 논 문 을 인 준 함

심 사 위 원 장 김 영 한 인

---

심 사 위 원 신 요 안 인

---

심 사 위 원 정 윤 원 인

---

2018년 6월

승실대학교 대학원

## 감사의 글

좋은 기회를 얻어 숭실대학교 대학원 IT융합학과에 입학하여 좋은 동기분들, 선후배분들과 함께 석사과정을 잘 마무리할 수 있어 정말 행복하고 감사했습니다.

부족한 논문내용을 보다 나은 방향으로 연구를 진행할 수 있도록 지도해주시고 피드백해주신 정운원 교수님, 넓고 깊은 시각으로 논문의 문제점들을 파악해주시고 심사해주신 김영한 교수님과 신요안 교수님, 작년까지 학과를 이끌어 주셨던 유명식 교수님, 올해부터 IT융합학과를 더욱 멋지게 이끌어주실 박민호 교수님, 재학중 저에게 많은 가르침을 주신 IT융합학과 교수님들, 학과사무실에서 열심히 일해주는 조교님들 등 IT융합학과의 구성원 모든 분들에게 진심으로 감사드립니다.

2년의 기간동안 학업적으로, 외적으로 많은 것들을 얻게 해준 숭실대학교에도 감사드리고, 항상 저에게 많은 사랑이 담긴 쓴소리를 해주시는 부모님에게도 언제나 감사드립니다.

# 목 차

국문초록 .....	iv
영문초록 .....	v
제 1 장 서론 .....	1
제 2 장 관련연구 및 배경 .....	3
2.1 가상화폐 .....	3
2.1.1 가상화폐와 전자화폐 .....	3
2.1.2 가상화폐의 특징 .....	4
2.2 블록체인 .....	5
2.2.1 비트코인 .....	6
2.2.2 이더리움 .....	6
2.3 주가예측분석 .....	7
2.4 감성분석 .....	8
제 3 장 연구 방안 .....	10
3.1 연구데이터 .....	10
3.2 연구 절차 .....	10
3.2.1 데이터 수집 .....	11



3.2.2 형태소 추출 .....	12
3.2.3 가상화폐 특화 감성사전 구축 .....	13
3.2.4 감성분석 .....	14
 <b>제 4 장 실험 및 결과</b> .....	 16
4.1 실험 설계 .....	16
4.2 데이터 수집 .....	16
4.3 감성사전 구축 .....	17
4.4 결과 분석 .....	18
 <b>제 5 장 결론</b> .....	 22
 참고문헌 .....	 24

## 표 목 차

[표 1-1] ECB 가상화폐 분류 정의 .....	4
[표 1-2] 양방향 가상화폐 종류 .....	4
[표 1-3] 각 화폐별 특성 .....	5
[표 4-1] 사전별 감성지수 예 .....	19

## 그 립 목 차

[그림 3-1] 연구 절차 .....	11
[그림 3-2] 데이터 수집과정 .....	12
[그림 3-3] 감성사전 구축 과정 및 방법 .....	13
[그림 3-4] 신규 데이터 감성분석 과정 .....	15
[그림 4-1] 데이터베이스에 저장된 뉴스기사 데이터 일부 .....	17
[그림 4-2] 구축된 감성사전 일부 .....	18
[그림 4-3] 맞춤형 감성사전과 범용 감성사전 예측률 비교(2월) .....	19
[그림 4-4] 맞춤형 감성사전과 범용 감성사전 예측률 비교(3월) .....	20

국문초록

## 인터넷 매체가 가상화폐 가격에 미치는 영향에 관한 연구

정성엽

IT융합학과

승실대학교 대학원

블록체인 기술의 발전으로 사회의 다양한 분야에 블록체인을 접목하려는 시도와 실제 적용한 사례들이 많아지고 있으며 블록체인 기술에 대한 관심, 대형 가상화폐 거래소들의 개장 등으로 남녀노소를 가리지 않고 많은 사람들이 가상화폐의 투자에 나서고 있다. 그러나 다양한 사회적 순기능을 수행할 수 있는 가상화폐임에도 불구하고 정말 실용 가치가 있는 가상화폐인지에 대한 정확한 정보를 투자자들이 명확히 얻지 못하는 문제들이 발생하고 있다. 그로 인해, 투자자들은 어떤 뉴스가 가상화폐의 가격등락에 어떻게 요인을 미치는지에 대해 명확히 알지 못하고 건전한 투자보다는 단순 투기적인 시장으로 변해가고 있는 상황이다. 이러한 무분별한 투자의 문제점과 가상화폐 시장에 영향을 끼치는 다양한 요인들을 알아보기 위해 본 논문에서는 가상화폐에 대한 인터넷 뉴스 데이터들을 수집하고 가상화폐 시장에 맞는 감성사전을 구축하여 가상화폐 등락에 영향을 미치는 요인들에 대해 알아본다.

## ABSTRACT

# A Study on the Effect of Internet Media on Virtual Currency Price

JEONG, SEONGYEOB

Department of IT Convergence

Graduate School of Soongsil University

With the development of Blockchain, many attempts have been made to apply Blockchain to various field of society, and cases have been applied in practice. With the interest in Blockchain technology and the opening of large virtual currency exchanges, many people are investing in virtual currency. However, despite the fact that it can fulfill various social roles, there are problems that investors can not clearly obtain worthwhile information and make irresponsible investment. As a result, investors doesn't know what news is causing price fluctuations in virtual currencies and it makes speculative market rather than a healthy investment. In order to investigate the problems of this indiscreet investment and various factors affecting the virtual currency market, this paper collected internet news data about virtual currency and constructed the sentimental dictionary suitable for the virtual currency market.

## 제 1 장 서 론

2008년 사토시 나카모토(Satoshi Nakamoto)의 비트코인을 통해 최초로 소개된 블록체인은 제 4차 산업혁명의 등장과 함께 주목을 받기 시작하였다. 2017년 대형 국내 거래소의 오픈, 비트코인 및 블록체인을 이용한 가상화폐 가격의 급등 등으로 가상화폐만이 아니라 가상화폐의 주를 이루는 기술인 블록체인에 대한 관심도 더더욱 늘어나고 있다. 그 관심과 블록체인의 활용성에 따라 의료보험금 자동청구 서비스, 공인인증서를 대체하는 전자서명, 은행 간 해외송금 등 실생활에 다양하게 적용되고 꾸준히 개발되고 있다 [1],[2]. 이러한 기술의 발전과 대중들의 관심과 함께 국내 대형 가상화폐 거래소들이 오픈하고, 오픈 당시 가상화폐 가격의 급등으로 인해 무분별한 투자가 투자자들의 손실을 유발하고 있다.

가상화폐 시장은 나이의 제한이 없고 주식시장과는 다르게 24시간동안 운영되고 상한가와 하한가가 존재하지 않기 때문에 더 위험성이 있음에도 불구하고 어떤 요인들이 가상화폐 가격의 등락에 요인이 되는지에 대한 연구가 활발히 이루어지지 않았다.

본 논문에서는 가상화폐와 관련된 인터넷 뉴스 정보들을 수집하고, 해당 뉴스들을 토대로 감성사전을 구축 후 가상화폐 가격의 등락과 범용 감성사전 대비 가상화폐 영역에 맞게 구축된 감성사전의 감성지수를 비교하여 어떤 요인들이 가상화폐의 등락에 영향을 미치는 지에 대해 연구하였다. 본 논문에서는 R을 이용하여 가상화폐와 관련된 인터넷 뉴스들을 수집하고 형태소를 분석하여 감성지수를 분석하고, 그 결과를 토대로 감성사전을 구축하여 어떤 요인들이 가상화폐의 가격등락에 영향을 미치는지 알아본다.

본 논문의 구성은 1장 서론을 시작으로 2장에서는 블록체인, 가상화폐 및 추가예측 분석, 감성분석에 관련연구 및 배경을 알아본다. 3장에서는 본고가 제안하는 인터넷 뉴스 데이터들을 활용하여 R을 이용한 감성사전 구축방법을 제시한다. 4장에서는 제안한 가상화폐 맞춤형 감성사전과 기존의 범용 감성사전의 성능평가를 통해 유용성을 검증한다. 마지막으로 5장에서는 결론 및 향후 연구 발전방향을 기술한다.

## 제 2 장 관련연구 및 배경

### 2.1 가상화폐

#### 2.1.1 가상화폐와 전자화폐

2012년 유럽중앙은행(ECB)은 가상화폐를 정부에 의해 통제받지 않는 디지털 화폐의 일종으로 가상공간에서 사용되고 민간이 발행·관리하는 화폐로 정의하고 전자화폐는 정부에서 법으로 허가받은 사업자가 화폐의 신용을 보증하고 중앙감독기구가 있는 화폐로 분류하며 가상화폐의 개념이 시작되었다[3]. ECB는 가상화폐를 아래의 [표 1-1]과 같이 ‘폐쇄형 가상화폐’, ‘단방향 가상화폐’, ‘양방향 가상화폐’로 분류하였다. 폐쇄형 가상화폐의 예는 게임머니로 가상의 공간에서만 사용이 가능한 특징으로 공식적으로는 법정화폐를 통해 폐쇄형 가상화폐를 구매하는 것은 인정하지 않는다. 단방향 가상화폐는 법정화폐를 통해 구매할 수 있고 해당 가상화폐를 이용하여 가상 또는 실재의 재화를 구매할 수 있다. 하지만, 구매한 가상화폐를 다시 법정화폐로의 환전은 불가하다. 그 예로는 라이엇 게임즈 리그오브레전드의 RP, 네이버 웹툰의 쿠키, 카카오톡의 초코 등이 있다. 양방향 가상화폐의 예는 비트코인, 이더리움, 리플 등으로 가상화폐와 법정화폐간의 자유로운 교환이 가능한 화폐이고 가상 또는 실재의 재화를 모두 구매할 수 있다.

[표 1-2]는 양방향 가상화폐 중 다양한 분야에 실제 적용, 활용되고 있는 가상화폐들이다. 비트코인은 영국, 호주, 말레이시아 등에서 부동산 거래나 다양한 쇼핑 등에 활용되고 있으며, 리플은 국내 각종 은행들이 해외송금에 적용하기위해 테스트 중에 있다. 마지막으로 이더리움도 비



트코인과 마찬가지로 다양한 실생활 거래에 적용되고 있고, 스마트계약이라는 새로운 기능을 토대로 다양한 거래의 투명성을 보장하고 있다.

[표 1-1] ECB 가상화폐 분류 정의[4]

구분	폐쇄적 가상화폐	단방향 가상화폐	양방향 가상화폐
특징	가상의 공간에서만 사용가능	법정화폐로 구매 가능하며 가상의 재화, 실재의 재화 모두 구매 가능	가상화폐와 법정화폐의 자유로운 환전이 가능하며 가상의 재화, 실재의 재화 모두 구매 가능
예	게임머니	리그오브레전드(RP), 네이버 웹툰(쿠키), 카카오툼(초코)	비트코인, 이더리움, 리플

[표 1-2] 양방향 가상화폐 종류

연도	명칭	창시자	단위	암호화 기술
2009	Bitcoin	Satoshi Nakamoto	BTC	SHA-256d
2013	Ripple	Chris Larsen & Jed McCaleb	XRP	ECDSA
2015	Ethereum	Vitalik Buterin	ETH	Dagger Hashimoto

### 2.1.2 가상화폐의 특징

가상화폐는 발행기관과 관리기관이 중앙정부가 아니며, 발행규모는 알고리즘으로 정해놓은 규모로 발행한다. 블록체인을 기반으로 하며 독자적인 화폐단위가 존재하고 그 화폐 단위로 법정화폐와의 교환이 가능하다. 허나, 시장의 논리인 수요와 공급에 의해 가격이 수시로 변동되는 특

정이 있다. [표 1-3]에서 볼 수 있듯이, 가상화폐는 전자화폐와 다르게 발행기관이 없고 법정화폐와의 교환이 가능한지 보장되지 않는다는 점에서 법적으로 인정되지 않은 화폐이기 때문에 정부의 가상화폐에 대한 투자 및 거래소 규제가 엄격해지고 있다.

[표 1-3] 각 화폐별 특성[5]

	현금	전자화폐	가상화폐
발행기관	중앙은행	금융기관, 전자금융업자	없음
발행규모	중앙은행 재량	법정통화와 1:1 교환	알고리즘에 의해 사전 결정
거래기록 및 승인	불필요	발행기관 및 청산소	분산원장 기술
화폐단위	법정통화	법정통화와 동일	독자적인 화폐단위
법정화폐와 교환 여부	-	발행기관이 교환을 보장	교환 가능하나 보장되지 않음
법정통화와의 교환가격	-	고정	수요공급에 따라 변동
사용처	모든 거래	가맹점	네트워크 참가자

## 2.2 블록체인

비트코인에서 최초로 고안된 블록체인은 일종의 분산형 데이터베이스이다. 데이터들은 블록에 저장되며 각 블록들은 hash값으로 서로 연결되고 각 노드에 지속적으로 연결되며 쌓이게 된다. 블록체인에 저장되는 거래의 정보들은 분산된 노드에 저장되고 그 정보들은 모든 노드들로 전파되어 해당 정보들을 모든 거래의 참여자들이 가지게 된다. 그로 인해 하나의 거래에 대해서도 다수의 사람들이 해당 정보를 가지고 있기에 정보의 투명성과 무결성을 보장한다.

### 2.2.1 비트코인

비트코인은 사토시 나카토모의 'Bitcoin : A Peer-to-Peer Electronic Cash System' 의 논문으로 처음 알려지게 되었다[6]. 중앙 시스템으로부터 화폐의 발행이나 거래에 대한 통제를 받는 기존의 통화정책과는 다르게 분산화된 네트워크에 데이터를 저장하며 쌓인 데이터들은 임의의 조작이 사실상 불가능하도록 하여 데이터의 무결성을 보장한다. 또, 비트코인은 이중 지불 문제(Double-Spending Problem)를 작업증명(Proof of work)이라는 방식을 이용하여 해결하며 추후 많은 가상화폐 연구의 기초가 되었다.

### 2.2.2 이더리움

비트코인 프로토콜을 기반으로 새롭게 구현된 이더리움은 바이트코드를 거래 장부에 기록하여 코드를 EVM (Ethereum Virtual Machine)을 통해 해석하고 실행한다. 이렇게 올라간 코드들은 특정 조건을 만족했을 때 그 값을 얻어 계약을 실행하거나 소유권에 대한 변경 등을 실행할 수 있게 해준다.

이러한 이더리움의 코드를 스마트 컨트랙트(Smart Contract)라 하고, 이는 튜링 완전성(Turing Completeness), 가치 인지(Value Awareness), 블록체인 인지(Blockchain Awareness) 등의 특성을 포함하여 비트코인보다 조금 더 발전된 기능을 제공한다. 스마트 컨트랙트는 특정 조건에 충족할 경우 자판기와 같이 자동으로 계약이 실행되어 결과를 도출하게 되는데, 이를 이용하여 현실에서 사용하고 있는 다양한 계약들의 대체할 전망이다.

## 2.3 주가예측분석

주식시장은 가상화폐 시장과 유사한 예로 기업의 가치를 반영하는 주가를 다양한 관점으로 바라보며 그것을 분석하고 예측하는 연구들이 여러 분야에서 진행되어왔고 진행되고 있다. 주가는 회사의 경영 및 재무 정보인 재무제표에 포함된 기업의 실적이 반영되어 가치가 평가될 수도 있고, 기업의 실적뿐만 아니라 국내, 국제의 경제 상황 및 기업에서 발생한 다양한 공시, 뉴스들이 기업 가치평가에 영향을 미칠 수 있다[7],[8]. 위와 같은 이유로, 기업과 관련해 발생하는 온라인 뉴스들과 기업과 관련된 투자자들의 SNS 언급 등을 활용하여 주가를 예측하는 연구들이 꾸준히 수행되고 있다.

국내에서 감성분석을 활용한 주가 예측에 대한 연구로는 SNS상의 사용자 의견을 데이터화 후 구글 번역 API를 활용하여 한글 텍스트를 영문으로 번역 후 영문 감성사전으로 구축되어있는 SentiWordNet을 활용한 주가 예측 연구가 있다[9]. 그 외로는 OAR (Opinion Antonym Rule) 알고리즘을 활용하여 주식 관련 뉴스 데이터로 감성사전을 구축 후 주가의 상승과 하락을 분석한 연구, 주식시장에 특화된 감성사전을 구축하여 주가 지수의 방향성을 예측한 연구 등이 있다[10],[11].

해외에서 연구한 사례로는 2012년 SNS에서 기업명이 언급된 트위터 글을 수집하고 텍스트마이닝으로 개별 기업에 대한 주가 등락을 예측하는 모델을 제시한 연구가 있다[12]. 그 전으로는 투자자들에게 관심이 높은 기억들을 검색엔진 검색량과 기업가치의 유동성에 미치는 영향들을 활용하여 주식수익률과의 상관관계를 분석한 연구, 트위터 글들의 감성 분석을 통하여 주가의 변화를 예측하는 방안에 대한 연구 등이 있다 [13],[14].

## 2.4 감성분석

감성분석 (Sentiment Analysis) 기법은 사람들이 사용하는 자연어를 기계학습을 기반으로하여 문장에서 긍정이나 부정과 같은 문맥의 정보를 추출하거나 분류하는 방법을 의미한다. 감성분석의 방안으로는 감성사전을 구축하여 문서안의 내용 중 단어 빈도를 바탕으로 한 분석방법과 전체 문맥내의 정보를 해석하고 분석하는 방법들이 존재한다. 감성사전을 기반으로 한 문장에서 긍정·부정의 문맥정보에 대한 추출은 문장 내에 출현한 용어들의 종류나 관계 등에 따라 추출이 가능하며, 이러한 이유로 인해 해당 분야에 맞는 감성사전의 구축은 감성분석 성능에 많은 영향을 미치는 주요 요소 중 하나이다[15].

감성사전을 구축하는 방법은 크게 단어 또는 문장의 극성이 기존에 정의된 감성사전을 바탕으로 어휘들의 유사성과 관계 등의 비교를 통해 감성사전을 구축하는 방안이 있다. 또 다른 방안으로는 실제로 수집된 문장들의 형태소 분석을 기반으로 개별 단어의 극성을 새롭게 정의하여 감성사전을 구축하는 방법이 있다[16].

구축된 감성사전을 이용한 텍스트에 대한 감성분석의 방법으로는 단순히 문서에 포함되어 있는 단어의 발생 횟수만을 계산하는 방법뿐만 아니라 문서 내 단어들의 어의적 관계 등의 언어규칙을 Pointwise Mutual Information (PMI), NavieBayes, Support Vector Machine (SVM) 등의 통계적 분석 기법을 사용하여 분류하는 방안들이 존재한다. 이 외에도 언어학적 방법론들을 적용한 연구들이 다양한 형태로 수행되고 있다 [15]-[17].

해당 연구에서는 기존의 연구에서 진행되었던 감성분석 방안에서 나아가 개선한 내용을 적용하고 활용하여 가상화폐 시장에서의 등락을 인터

넷뉴스 데이터를 기반으로 감성분석을 이용한 예측에 대한 연구를 수행하고자 한다. 인터넷 뉴스를 기반으로 가상화폐에 특화된 감성사전을 구축하여 예측정확성을 높이는데 중요성을 두었다.

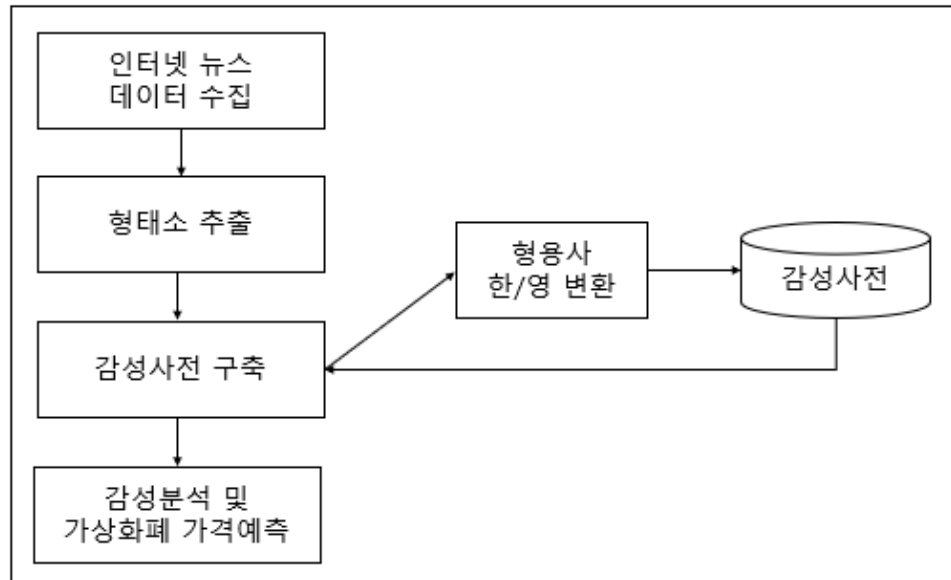
## 제 3 장 연구 방안

### 3.1 연구 데이터

본 연구에서 감성분석을 진행하기 위한 데이터는 네이버 뉴스를 크롤링하여 진행하였다. 2017년 2월부터 2018년 1월까지 ‘가상화폐’, ‘비트코인’으로 검색한 뉴스의 결과를 데이터베이스화 하였다. 두 가지의 검색어를 통해 추출한 데이터 중 중복데이터는 URL로 구분하여 제거하였다. 1년간의 데이터는 감성사전 구축의 용도로 사용되었으며, 2018년 2월부터 2018년 3월까지의 데이터를 검증의 데이터로 사용하였다.

### 3.2 연구 절차

본 논문에서는 [그림3-1]과 같은 절차를 통하여 수행되었다. 연구에 사용할 데이터는 네이버 뉴스를 크롤링으로 특정 검색어에 관련된 뉴스들을 수집하였다. 가상화폐와 관련된 감성사전을 구축하기 위한 훈련 데이터로 1년간의 뉴스를 기반으로 사용하였고, 사전 구축용 데이터와 검증용 데이터 분석은 통계와 자료분석에 널리 사용되는 R을 사용하였다. R의 패키지로 제공되는 ‘KoNLP’, ‘tm’ 등을 활용하여 데이터의 형태소 분석 및 텍스트 분석을 수행하였다.



[그림 3-1] 연구 절차

### 3.2.1 데이터 수집

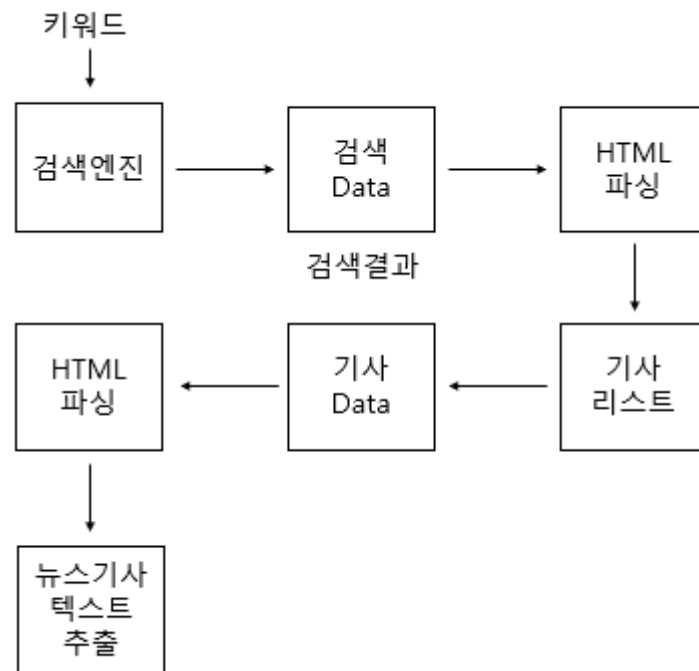
감성사전을 구축하기 위해 사용될 데이터인 인터넷 뉴스는 실시간으로 생산된다. 실시간으로 게재되는 뉴스들을 데이터로 활용하기 위해 뉴스를 일정한 시간마다 수집하여 데이터베이스에 저장하는 배치 프로그램을 구현하였다.

데이터의 수집 과정은 [그림 3-2]와 같다. 네이버 뉴스에서 해당 키워드로 검색한 내용의 HTML을 파싱하여 본문이 포함된 URL의 리스트를 추출한다. 추출된 URL의 리스트를 다시 요청하여 추출해낸 본문 HTML을 다시 파싱하고 HTML태그, HTML 문자 엔티티(nbsp, quot 등), 편집자주, 저작권자 등 감성사전에 등록하지 않아도 되는 의미없는 문자들을 제거하는 과정을 거친다.

가상화폐 등락에 대한 데이터도 크롤링을 통하여 Raw data를 추출 후 데이터베이스를 구축하고 일정 간격의 시간마다 수집하여 데이터베이스



에 저장되도록 돌아가는 배치 프로그램을 구현하였다.



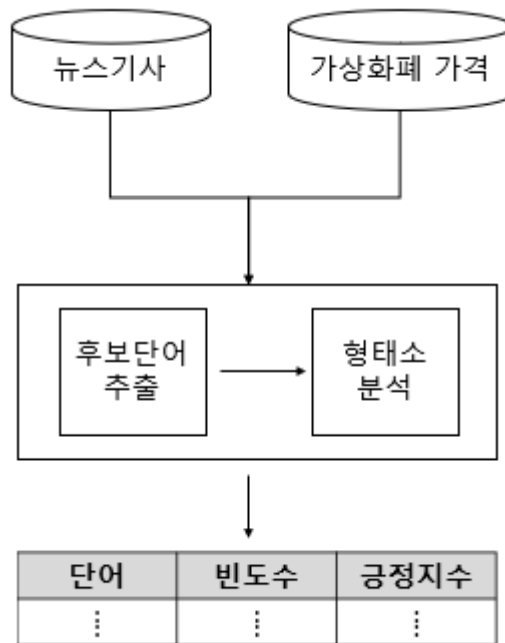
[그림 3-2] 데이터 수집과정

### 3.2.2 형태소 추출

파싱한 데이터를 이용하여 감성분석을 진행하기 위해서는 텍스트들을 형태소 단위로 분리하여 각 형태소의 극성을 분류한다. 본 연구에서는 명사를 추출하기 위해 R의 패키지인 'KoNLP'를 이용하여 추출하였고, 형용사는 원활한 추출을 위해 1차로 구글번역을 통해 영문으로 번역 후 'OpenNLP'를 이용하여 추출하였다.

### 3.2.3 가상화폐 특화 감성사전 구축

감성사전의 구축과정 및 방법은 [그림 3-3]과 같다. 감성사전은 감성 분석의 과정에서 해당 형태소들의 극성을 분류하는 용도로 사용된다. 분야별로 특화된 감성사전들은 기존 범용 감성사전에 비해 분석을 용이하게 하고 정확도를 높이는데 큰 비중을 차지한다.



[그림 3-3] 감성사전 구축 과정 및 방법

데이터 수집 및 형태소 추출 과정을 거쳐 나온 데이터들을 토대로 감성사전에 포함될 후보 단어들을 추출하게 된다. 크롤링을 통해 데이터베이스화된 일별 가상화폐 등락 데이터로 전날에 비하여 가격이 상승한 경우, 해당 날짜의 뉴스들을 긍정적 영향을 갖는 뉴스들로 가정하여 분류

한다. 반대로 전날에 비하여 가격이 하락한 경우는 부정적 영향을 주는 뉴스들로 가정하여 분류한다.

상승과 하락의 요인으로 분류된 뉴스들에서 가장 많이 발생한 단어들의 순위를 매겨 0에서 1사이의 값으로 맵핑하여 감성사전에 등록한다. 추출한 형태소의 순위 중 발생빈도가 현저하게 낮은 단어들은 가상화폐 가격에 미치는 영향도가 적다고 판단하고 순위에서 제거한다.

상승과 하락으로 구분된 데이터들에서도 동일하게 출현빈도가 높은 중복어휘들이 발생하게 되는데 이 부분은 식(1)과 같이 긍정 / 부정 시점의 어휘 발생 횟수와 각 점수를 곱하여 총 발생 횟수로 나누어 감성점수를 등록한다.

$$(1) \text{ OverlapSentScore} = \frac{\sum_{i=1}^{n-pos} \text{NegScore} + \sum_{i=1}^{n-neg} \text{PosScore}}{n}$$

### 3.2.4 감성분석

구축된 감성사전을 이용하여 새롭게 발생하는 인터넷 뉴스기사들은 [그림 3-4]의 과정으로 진행된다. 감성사전 구축과 마찬가지로 인터넷 뉴스를 크롤링을 통해 수집하고 데이터를 형태소 분석을 하여 명사와 형용사를 추출한다. 추출한 형태소들을 감성사전 점수에 대입하여 아래의 식(2)를 이용하여 해당일 뉴스의 감성 점수를 계산한다.

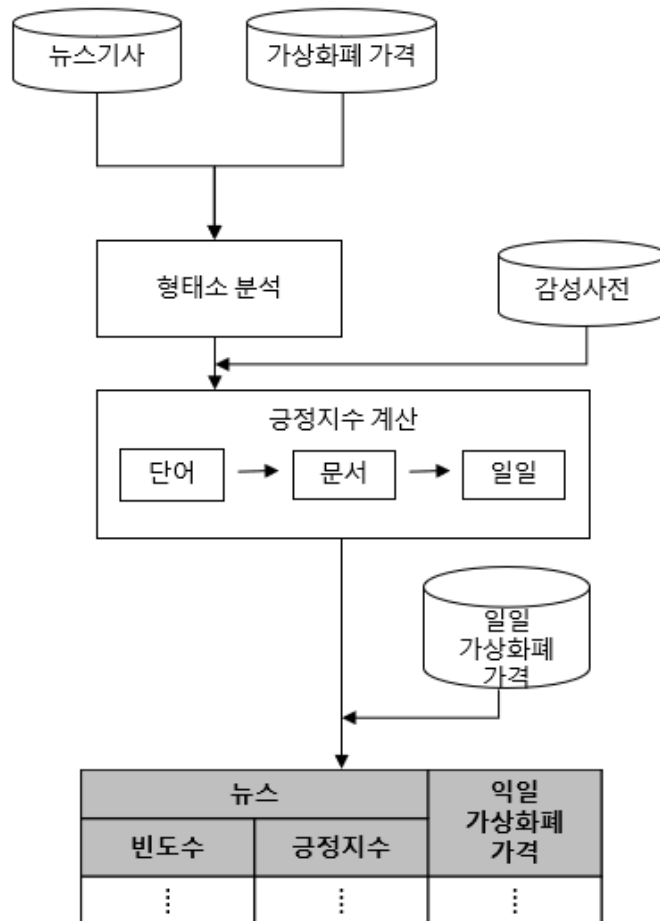
$$(2) \text{ NewsScore}(s) = \frac{\sum_{i=1}^n (\text{Num}(s_i) \times \text{SentScore}(s))}{\sum_{i=1}^n \text{Num}(s_i)}$$

분석하고자 하는 일자에 발생한 모든 뉴스에서 발생한 감성사전에 있

는 어휘 s의 발생횟수와 해당 어휘의 감성지수 SentScore(s)를 곱하여 전체 어휘의 발생 횟수로 나누어 해당 어휘 s의 감성점수를 산정한다.

해당 일자의 감성지수를 계산하기 위한 수식은 식(3)과 같이 전체 발생한 뉴스들을 식(2)로 계산하여 발생한 뉴스의 개수 n으로 나누어 구한다.

$$(3) \text{DailyScore}(d) = \frac{\sum_{i=1}^n \text{NewsScore}(i)}{n}$$



[그림 3-4] 신규 데이터 감성분석 과정

## 제 4 장 실험 및 결과

### 4.1 실험 설계

실험은 2017년 2월 1일부터 2018년 1월 31일까지의 가상화폐, 비트코인과 관련된 데이터를 실험에 수집하여 가상화폐에 특화된 감성사전을 구축하는데 사용하고, 2018년 2월 1일부터 2018년 3월 31일까지의 데이터는 구축된 감성사전을 실험을 통해 검증하는데 사용하였다.

가상화폐 시장에서의 등락 데이터를 수집하기 위해 가상화폐의 기축통화의 역할을 하는 비트코인 가격등락 데이터를 위의 뉴스 데이터와 마찬가지로 2017년 2월부터 2018년 1월까지의 데이터는 감성사전을 구축하는데 사용, 2018년 2월부터 3월까지의 데이터는 감성사전을 이용하여 실제 결과를 도출하는데 사용하였다.

실험은 3장에서 제안한 순서에 따라 데이터 수집, 형태소 추출, 감성사전 구축 및 감성 분석의 순으로 진행하였다.

### 4.2 데이터 수집

연구에 필요한 데이터는 파이썬으로 제작한 배치프로그램을 이용하여 매일 00시에 인터넷 뉴스와 가상화폐 가격 데이터를 데이터베이스에 저장하였다. 인터넷 뉴스 기사의 경우 비트코인, 가상화폐의 키워드가 제목과 본문에 존재하는 기사들을 대상으로 데이터를 추출하여 총 67,851건의 기사를 수집하였고 그 중 50,654건은 감성사전 구축에, 나머지는 감성사전을 이용한 결과 도출에 사용하였다. 아래 [그림 4-1]은 저장된 데이터의 예시이다.

date	article
2017-03-26 07:02:00	[4차 산업혁명 도래]‘나만의 보험’ 가시화...관련법 개정 속제
2017-03-26 07:01:00	[4차 산업혁명 도래]은행권, 블록체인·핀테크 등 4차 산업혁명에 ‘성큼’
2017-03-28 08:07:00	<u>비트코인</u> 헤지 나선 투자자들...매도 or 다른 가상화폐
2017-03-30 10:04:00	1만조 가상화폐시장, <u>비트코인</u> 보다 <u>알트코인</u> 가치 더 커
2017-03-31 11:20:00	반년에 3배 <u>쾅</u> 니다에 <u>슬깃</u> ...日가상통화 피해분쟁 ‘비상’
2017-03-31 11:19:00	‘가상화폐의 쿠데타’ 블록체인 비즈니스, 한국도 다가온다
2017-03-31 11:05:00	미쓰비시UFJ, 내년부터 ‘블록체인’ 기반 국제송금

[그림 4-1] 데이터베이스에 저장된 뉴스기사 데이터 일부

### 4.3 감성사전 구축

수집한 인터넷 뉴스기사와 가상화폐 가격등락 데이터를 이용하여 가상화폐에 특화된 감성사전을 구축한다. 우선 인터넷 뉴스 기사를 형태소 분석을 통해 명사와 형용사를 추출한 뒤 빈도수를 기준으로 후보단어들을 도출한다. 이 중 비트코인, 가상화폐와 같이 검색한 키워드 등 불필요한 단어들과 평균 빈도수 이하인 단어들을 제거한다. 필터링된 단어들의 빈도수와 해당 뉴스가 발생한 일일의 가상화폐 가격의 추이를 이용하여 긍정 지수를 계산하여 감성사전의 구축을 완료한다. 구축된 감성사전의 일부를 [그림 4-2]에 나타내었다.

word	<u>positive index</u>	word	<u>positive index</u>
가시화	0.657193	비상	0.294437
가속화	0.561075	분쟁	0.276881
규제	0.138252	유혹	0.227915
배당	0.457136	절감	0.548169
<u>에어드랍</u>	0.336591	ICO	0.661027
스냅샷	0.508753	하드포크	0.520371
<u>알트코인</u>	0.680123	해킹	0.118340

[그림 4-2] 구축된 감성사전 일부

#### 4.4 결과분석

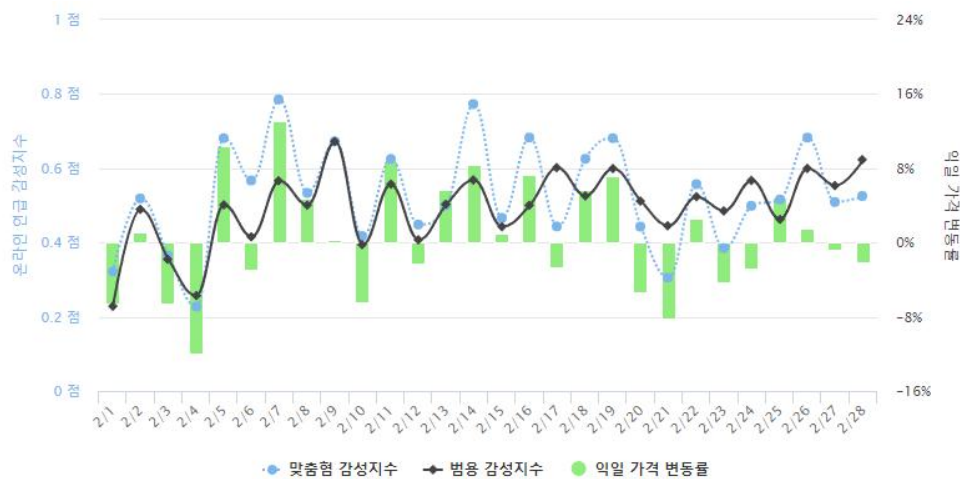
본 연구에서 구축한 가상화폐 특화 감성사전과 이미 구축된 범용 감성사전을 이용하여 2018년 2월 1일부터 2018년 3월 31일까지 저장한 데이터에 대해 성능을 분석한다. 본 연구에서는 두 사전의 성능을 비교하기 위해 일자별 뉴스를 각 감성사전에 대입하여 일자별 감성지수를 비교하였고, 식(4)와 같이 향상도를 범용감성사전을 사용하여 추출한 감성지수 대비 가상화폐 맞춤 감성사전을 사용한 경우 얼마만큼 등락의 예측률이 향상되었는지를 보여준다.

$$(4) \text{Improvement} = 1 - \frac{\text{PredictScore}_{\text{SpecializedDictionary}}}{\text{PredictScore}_{\text{OriginalDictionary}}}$$

[표 4-1]은 일자별 감성지수를 범용 감성사전과 가상화폐 맞춤형 감성 사전에 대한 예이고, [그림 4-3]은 검증기간동안의 비트코인 가격 추이와 각 감성사전을 비교하여 성능을 분석한 2월 1일부터 3월 31일까지의 예측률을 비교한 그래프이다.

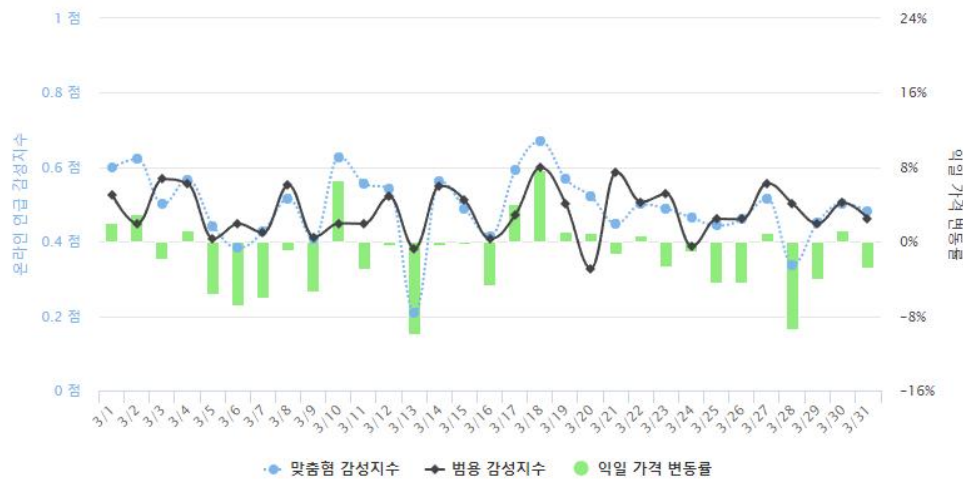
[표 4-1] 사전별 감성지수 예

날짜	맞춤형 감성사전	범용 감성사전
2월 21일	0.305271	0.445286
2월 22일	0.557160	0.523189
2월 23일	0.384973	0.485217
2월 24일	0.498712	0.567139



[그림 4-3] 맞춤형 감성사전과 범용 감성사전 예측률 비교 (2월)





[그림 4-4] 맞춤형 감성사건과 범용 감성사건 예측률 비교 (3월)

위 [그림 4-3], [그림 4-4]과 같이 감성지수 0.5 이상을 상승으로 예상할 때, 범용 감성사건으로 예측한 가상화폐 가격에 비해 약 8% 가량의 성능향상을 보였다. 또, 감성지수가 높을수록 상승률이 높고, 낮을수록 하락률이 높다고 가정했을 때, 등락을 예측하여 성공한 날짜의 경우 상승률이나 하락률을 감성지수에 대입하였을 때, 가상화폐 맞춤형 감성사건이 범용 감성사건에 비해 상승률이나 하락률에 더 근접한 감성지수를 얻었음을 확인할 수 있다.

뉴스기사별 분석으로는 2월 28일에 조선비즈에서 게재한 ‘업비트, 가상통화 배당 안하나 못하나’ 이라는 뉴스는 범용감성사건과 가상화폐 특화 감성사건 모두 0.5 이하로 부정적인 감정을 추출하였지만, 가상화폐 시장에서만 사용되는 ‘에어드랍’이라는 단어의 특성에 의해 가상화폐 특화 감성사건으로 추출한 감성지수가 약 0.15정도 낮은 점수를 기록했다. ‘에어드랍’은 특정 가상화폐를 보유하고 있을 때 기존 가상화폐의 블록체인을 하드포크하여 분리할 때 정해진 가치만큼의 새로운 가상화폐를 보유자들에게 지급하는 주식시장에서의 배당과 비슷한 의미이다. ‘에어드랍’ 외에

도, ‘하드포크’, ‘핫월렛’, ‘콜드월렛’, ‘ICO’, ‘스캠’ 등의 가상화폐시장에서만 통용되는 단어들의 감성지수들이 각 뉴스들의 감성점수를 계산할 때 범용 감성사전에 비해 높은 예측률에 영향을 끼치는 것을 확인할 수 있었다.

또, 특화된 감성사전을 비트코인 외의 가상화폐 시장에 범용적으로 적용할 수 있는지 확인하기 위해 비트코인 외의 가상화폐 거래소에 상장된 주요 가상화폐 가격등락과 비교해보았다. 리플(XRP)의 경우 범용감성사전을 이용하여 비교하였을 때는 2월은 42.3%, 3월은 45.7%의 예측률을 보인 반면, 본 연구로 구축된 가상화폐 특화 감성사전을 이용하였을 때는, 2월은 49.1%, 3월은 50.8%로 각각 약 6%와 5% 상승된 예측률을 보였다. 이더리움에 사용한 결과로는 범용사전을 사용하였을 때는 각각 2월에는 45.7%, 3월에는 50.8%의 예측률을 보였고, 본 연구에서 구축한 감성사전을 이용했을 때는 2월 54.2%, 3월 55.9%로 각각 5% 정도 상승된 예측률을 보였다.

## 제 5 장 결론

본 논문에서는 인터넷에 발생하는 가상화폐 뉴스들이 실제 가상화폐에 어떤 영향을 끼치는지 알아보기 위해 관련 인터넷뉴스의 텍스트들의 감성사전을 구축하여 범용감성사전과의 비교를 진행하였다. 데이터는 가상화폐, 비트코인의 키워드가 포함된 포털의 기사들을 수집하는 프로그램을 제작하였으며, 수집한 데이터를 이용하여 감성사전을 구축하였다.

1개월간의 가상화폐 가격의 추이를 각각 가상화폐에 특화된 감성사전과 범용감성사전으로 가격을 예측해본 결과, 범용 감성사전을 토대로 예측한 결과에 비해 약 8% 향상된 예측정확도의 연구결과를 얻었다.

본 연구는 기존 연구되지 않았던 가상화폐의 가격등락에 미치는 영향을 가상화폐에 특화된 감성사전 구축을 통한 감성분석을 진행함으로써 어떤 요소들이 가상화폐 시장에 영향을 미치는지를 알게 되고, 가상화폐 관련 감성사전 구축이 범용 감성사전을 통한 예측에 비해 성능 향상이 있음을 보인 점이 이 논문의 주된 학문적 기여라고 볼 수 있다. 그러나 가상화폐 분야의 감성분석에 사용할 맞춤형 감성사전 구축을 위해서는 해당 연구가 갖는 한계점은 아래와 같으며 이후 추가적인 연구가 필요하다.

본 연구에서는 ‘비트코인’의 키워드로만 뉴스를 검색하여 감성사전을 구축하였는데, 추후 연구에서는 다양한 가상화폐들의 데이터들도 감성사전을 구축할 데이터로 삼아 각 가상화폐별로 특화된 감성사전으로의 개선, 가상화폐 범용 감성사전의 구축에 대한 연구가 필요하다.

본 연구에서 사용한 가격의 등락지표는 일별 종가를 기준으로 삼아 진행하였다. 허나, 가상화폐 시장은 24시간동안 지속적인 흐름이 발생하고 세계에서 다양한 이슈들이 발생하기 때문에 시간별, 국가별로 가상화폐

의 등락데이터와 뉴스데이터를 실시간으로 수집하여 보다 능동적인 연구를 진행하는 연구가 필요하다.

또, 본 논문에서 사용한 데이터는 국내 최대의 포털 사이트에서 데이터를 수집하였는데, 인터넷 뉴스 외에도 각 가상화폐의 공식 트위터, 대형 가상화폐 커뮤니티 등 보다 다양한 각도에서 데이터를 수집하여 보다 폭넓은 감성사전을 구축하여 나은 결과를 얻을 수 있을 것으로 예상된다.

마지막으로, 가상화폐 시장도 시장경제의 한 영역이기 때문에 시장에 영향을 끼치는 다양한 요인들도 추후 연구에 가상화폐 가격등락의 분석요인에 추가하면 단편적인 요소들을 통해서만 예측한 결과보다 나은 결과를 얻을 수 있을 것으로 예상된다.

## 참고 문헌

- [1] 장운형, “[4차 산업혁명 시작 ‘블록체인’]<7>의료정보 공유부터 실손 보험 청구까지 ‘메디컬 체인 시대’”, 전자신문, 2018.03.26.  
(<http://www.etnews.com/20180326000179>)
- [2] 백만호, “은행권, 블록체인 ‘공인인증’ 도입”, 내일신문, 2018.02.07.  
([http://www.naeil.com/news\\_view/?id\\_art=265902](http://www.naeil.com/news_view/?id_art=265902))
- [3] 정경민, “[경제 view &] 비트코인은 가상화폐인가 암호화폐인가”, 중앙일보, 2017.11.09. (<http://news.joins.com/article/22097491>)
- [4] 신상화, 강성훈, “가상화폐 이용 증대에 따른 과세상 쟁점 분석 및 대응 방안 연구 : 비트코인을 중심으로”, 재정포럼, 2015
- [5] 김동섭, “분산원장 기술과 디지털통화의 현황 및 시사점”, 한국은행 지급결제조사자료, 2016
- [6] Nakamoto, Satoshi, “Bitcoin: A Peer-to-Peer Electronic Cash System”, “<http://bitcoin.org/bitcoin.pdf>”, 2008
- [7] 김미형, “재무제표분석을 이용한 장부시장가비율이 낮은 기업의 주가 수익률 예측력분석”, 국제회계연구 제 30집, pp.1-18, 2010
- [8] 전세미, 정여진, 이동엽, “개별 기업에 대한 인터넷 검색량과 주가 변동성의 관계 - 국내 코스닥시장에서의 산업별 실증분석”, 한국지능정보 시스템학회 지능정보연구, 제22권, 제2호, pp.81-96, 2016
- [9] 김영민, 정석재, 이석준, “소셜 미디어 감성분석을 통한 주가 등락 예측에 관한 연구”, Entrue Journal of Information Technology, 제13권, 제3호, pp.50-69, 2014
- [10] 조혜진, 서지훈, 최진탁, “주식 뉴스 콘텐츠를 활용한 오피니언마ining 기반의 OAR 감성사전 알고리즘 기법”, 한국정보기술학회논문지, 제

13권, 제3호, pp.111-119, 2015

[11] 김유신, 김남규, 정승렬, “뉴스와 주가: 빅데이터 감성분석을 통한지  
능형 투자의사결정 모형”, 한국지능정보시스템학회 지능정보연구, 제18  
권, 제2호, pp.143-156, 2012

[12] Evangelopoulos, N., Magro, M. J., and Sidorova, A., “The Dual  
Micro/Macro Informing Role of Social Network Sites: Can Twitter  
Macro Messages Help Predict Stock Prices?,” Informing Science: the  
International Journal of an Emerging Transdiscipline, Vol. 15, pp.  
247-268, 2012

[13] Bank, M., M. Larch, and G. Peter, “Google Search Volume and  
Its Influence on Liquidity and Returns of German Stocks,” Financial  
Markets and Portfolio Management, Vol.25, No.3, pp.239-264, 2011

[14] Bollen, J., Mao, H., and Zeng, X., “Twitter mood predicts the  
stock market,” Journal of Computational Science, Vol. 2, No. 1, pp.  
1-8, 2011

[15] 김승우, 김남규, “오피니언 분류의 감성사전 활용효과에 대한 연구”,  
한국지능정보시스템학회 지능정보연구, 제20권, 제1호, pp.133-148, 2014

[16] 조은정, “감성 분석 연구의 현황과 말뭉치에 기반한 사례 분석”, 연  
어과학회 언어과학연구, 제61권, pp.259-282, 2012

[17] 송상일, 이동주, 이상구, “PMI 를 이용한 우리말 어휘의 의미 극성  
판단”, 한국정보과학회 학술발표논문집 , pp.260-265, 2010