

# 악성 집단 댓글 분석에 의한 SNS 여론 소셜데이터 분석

황윤찬\*, 고 찬\*\*

스토리조아 (디지털콘텐츠 제작회사) 대표\*, 서울과학기술대학교 교수\*\*

## Analysis of Opinion Social Data on the SNS (Social Network Service) by Analyzing of Collective Damage Reply

Hwang, Yun Chan\*, KOH, Chan\*\*

CEO, Storyjoa Co.\*, Professor, Seoul National University of Science & Engineering\*\*

**요약** 미디어를 통한 많은 소셜 데이터가 유통, 활용, 공개 되고 있다. 이 소셜 데이터를 이용한 미디어에 대한 즐거움과 정보의 효율적인 측면만 부각되고, 여기에서 발생하는 지나친 정보 노출과 사용자에 대한 인신 공격적 집단 댓글의 피해 문제는 소홀히 취급되고 있다. 본 연구에서는, 악성 집단 댓글 분석에 의한 SNS 여론 소셜 데이터 분석을 하였다. 소셜 네트워크가 가진 구조적 정보 이용을 통해 분석된 정보 분석 데이터의 양, 즉 SNS 언급 횟수 인 버즈량이 얼마나 많은 사람들에게 배포되고 악용되는가에 대한 문제를 다양한 측정 방법으로 분석하였다

**주제어** : SNS, 소셜데이터, 소셜커머스, 소셜컨슈머, 버즈량

**Abstract** A lots of social data are distributed, utilized and opened through the social media. They have characterized effectiveness and pleasure of information to the media by social data but it is ignored about excessive exposure of information and damage from collective reply of personal attack type. In this paper, we study about analysis of opinion social data on the SNS (Social Network Service) by analyzing of collective damage reply. It is analysed by diverse measurement method for distribution and disuse of the amount of Buzz data that is analysed data from structured social network.

**Key Words** : SNS, Social Data, SocialCommerce, Buzz Data

### 1. 서론

기업이 마케팅 활동으로 고객을 모으는 구조가 아닌 소비자의 기호에 맞는 맞춤형 시장이 인기를 더해가는 것은 스마트 미디어의 혁신 때문이다. 스마트 미디어란

스마트 IT 혁명에 맞는 스마트폰, 태블릿 PC, 클라우드 네트워킹 등 SNS로 대변되는 시간과 공간의 제약 없이 실시간으로 제품에 대한 정보를 생산해내고 다양한 경로로 공유하는 미디어를 말한다. 기업들은 소비자와 소통의 필요성이 더욱 커졌다. 소비자는 소비자대로 기업의

\* 이 연구는 서울과학기술대학교 교내 학술연구비 지원으로 수행되었습니다.

Received 27 March 2013, Revised 25 April 2013

Accepted 20 May 2013

Corresponding Author: Chan KOH (Professor, Seoul National University of Science & Engineering)

Email: chankoh@snu.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISSN: 1738-1916

상품 및 구매 후기를 언제 어디서나 전파하고 기업들은 상품의 홍보를 위해 자신의 브랜드 이미지를 향상시킬 수 있는 스토리텔링 및 소비 환경 개발에 변화를 모색하고 있다. 소비자의 요구에 따라 기업들은 SNS에서 거론되는 데이터들을 새로운 마케팅 수단으로 보고 활용하고 있다. 소셜 데이터는 크기, 속도, 형태를 분석할 때 그 의미가 있지만, 소셜 데이터의 성공적인 분석이 때론 역기능도 발생하고 있다. SNS에서 개인 정보가 무단으로 도용되어 범죄에 이용된다거나 거짓된 정보가 확산되어 새로운 사회문제로 변질될 수 있는 소지가 매우 많다. 소셜 데이터의 양이 저조하다고 방치해 두었다가 한꺼번에 대량의 정보가 진실여부를 망각하고 쏟아져 넘친다면 기업과 소비자의 신뢰는 무너지고 이러한 문제로 인해 발생하는 리즘<sup>1)</sup> 형태의 네트워크 구조는 또 다른 갈등과 경제 문화 발전에 안 좋은 영향을 미칠 것이다.

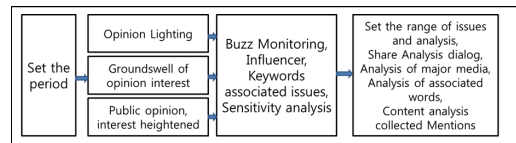
SNS를 통해 파급되는 여론들의 형태를 분석하여 국내 외 정치, 경제, 사회 문화현상 들의 소셜 분석을 통해 대응할 수 있는 정책을 수립하고 실행해야 SNS를 통한 스마트한 소통 환경 및 미디어 정책으로 인한 효율적인 문화 생산성에 도움을 줄 것이다.

## 2. 이론적 고찰

### 2.1 연구 분석의 틀

사회적인 사건, 사고에 따라 생성되는 많은 이슈들은 인터넷 공간을 떠돌아다닌다. 그중에서도 사람들의 먹거리에 대한 이슈는 우리의 건강에 직결되므로 매우 신중하게 처리되어야 하는데 그 전달은 이슈의 본질의 내용과 관계없이 기하급수적으로 확대 재생산된다. 요즘 홈페이지 게시판에는 뉴스의 정보를 빠르게 전달할 수 있는 SNS 버튼을 마련하여 현재 이슈가 될 수 있는 글을 버튼 클릭으로 쉽게 전달할 수 있기에 그 파급효과는 매우 크다. 본 연구에서는 2012년 1월부터 12월까지 소셜네트워크 서비스를 이용하여 이슈화된 데이터들중 코카콜

라사레, 채선당사레, 알츠하이머사레와 같은 트위터나 페이스북에 올라온 버즈광과 댓글의 양(수)를 바탕으로 만들어진 소셜 분석툴<sup>2)</sup>을 비교 분석하였다. 또한 SNS 상 이슈 확산과 매체별 분석 및 기업의 SNS를 통한 고객 대응 사례 분석을 통한 기업의 대응 방안 수립, 개선 전략 등을 파악해 보았다. [Fig. 1]은 소셜 서비스를 통해 얻어진 데이터들은 일단 기간 설정을 통해 만들어진 다양한 의견들을 모아 여론 접화 시점부터 여론의 관심이 증대되는 시기 및 여론의 관심이 고조되는 시기를 나누어 분석하여 분석 구간별 이슈 및 대상 미디어를 분석하게 된다. 이 연구를 통해 여론이 만들어진 후에 주요 미디어에 노출되었을 때 아주 빠른 속도로 여론의 관심이 높아지는 것을 알 수 있었다.



[Fig. 1] Analysis structure of research

[Fig. 1]에서 첫째 버즈 모니터링이란 특정 키워드가 소셜 미디어상에서 얼마나 많이 언급이 되는지를 확인하는 것이다. 특정 키워드에 대한 언급이 일시적으로 증폭되면 평소와 다른 일이 일어났음을 유추할 수 있다. 둘째 소셜 분석툴의 수집 대상은 소셜 미디어이며 자연어 처리 기술을 이용해 메시지가 담고 있는 호감도와 동시에 언급되고 있는 연관 키워드를 분석하여 검색엔진과 차별화된 소셜 분석을 하려 한다. 메시지에 대한 자연어 처리를 통해 언급된 성향이 긍정적인지 부정적인지를 파악하는 것이 감성분석이다. 이는 펄스K를 통해 대규모 텍스트에서 대상에 대한 호감도를 즉시 분석해 여론의 경향성을 파악하는 방법이다. 셋째 연관이슈 키워드 분석은 특정 키워드와 동시에 언급되고 있는 또 하나의 키워드를 찾아내는 것이다. 연관 키워드 분석은 키워드와 관련된 호감도의 변화를 감지하며 라디안6(Radian6)라는

1) '리즘'은 들뢰즈와 가타리(Gilles Deleuze et Félix Guattari)가 그들의 명저 『천의 고원(Mille Plateaux)』(1980)의 입문적 표제어로서 사용하면서 널리 알려지게 이르렀다. 이들은 이 말을 '수목형'(樹木型)과 대비적으로 사용한다. 리즘형과 수목형은 '관계 맺기'의 두 방식을 가리킨다. 더 정확히 말해, 리즘형과 수목형이 따로 존재하는 것이 아니라 리즘형에 좀더 많은 규정들이 들어갈 경우 수목형으로 화하고 반대의 경우(규정성을 줄어들 경우) 리즘형으로 화한다. 즉 리즘형과 수목형은 상관적 정도(correlative degree)를 형성한다. 리즘은 관계를 맺는 방식이 보다 자유로운 쪽으로 갈 때 성립하고, 수목형은 관계 맺는 방식이 이항대립적(binary) 방식으로 화할 때 성립한다.

2) 소셜 분석을 활용해 얻을 수 있는 버즈모니터링 (Buzz Monitoring), 인플루언서(Influencer), 연관 이슈 키워드, 감성 분석 (Sentiment Analysis) 들을 가지고 소셜 데이터를 분석하는 틀.

소셜 분석 툴로 해당 키워드와 동시에 언급되고 있는 단어 무치를 제공한다. 넷째 인플루언서 분석은 소셜미디어에서 영향력을 행사하는 사람을 말한다. 소셜 영향력 측정 지수를 제공하는 클라우트(Klout)로 소셜 미디어 사용자의 참여도를 분석해 점수를 부여한다. 이 때 영향력 평가 기준은 트위터의 팔로워나 업데이트 빈도, 페이스북 친구 수, 댓글 수, 좋아요 등이다. 클라우트를 활용하면 어떤 사람이 소셜 미디어에서 얼마만큼의 영향력을 갖고 있는지 어떤 성향의 영향력자 인지도 가능할 수 있다. 이러한 4가지 단계를 이용하여 이슈 및 분석범위를 설정하고 대화 점유율, 주요 미디어, 연관어, 수집 멘션 내용들을 분석 사이트에 들어오는 검색어들을 분석하였다. 소셜미디어 여론 분석 및 소셜 미디어의 메시지를 분석해 새로운 의미를 찾는 소셜 분석을 통해 각 사례들의 사회적 관심이 나타난 시기, 증폭된 시기, 사회적 관심도의 변화가 일어나는 시기의 위험요인들의 의미를 분석하여 파악하였다.

## 2.2 소셜 데이터의 개념

소셜 데이터의 개념은 세계적 컨설팅 기관인 맥킨지 앤드 컴퍼니 (McKinsey & Company) 의 5월 보고서에서 ‘소셜 데이터의 정의는 기존 데이터베이스 관리 도구의 데이터 수집, 저장, 관리, 분석하는 역량을 넘어서는 데이터 셋 규모로, 그 정의는 주관적이며 앞으로도 계속 변화될 것이다. 또한 데이터량 기준에 대해 산업분야에 따라 상대적이며 현재 기준에서는 몇 십 테라이트에서 수 페타 바이트까지가 그 범위이다’라고 설명한다. 아직까지 소셜 데이터의 개념에 대해 많은 연구가 진행중이고 분석 사이트들이 등장하고 있지만 이를 어떠한 형태로 가공되어질 것인지에 대한 자료 조사는 많은 시행착오가 있어야 할 것이다. 국내 기업들은 미디어 시장의 세계화 전략에 맞춘 적극적인 마케팅을 펼치지 못하였다. 싸이월드, 블로그, 카페, 지식인등 국내 환경에 맞는 데이터만을 가지고 국내 시장만을 타겟으로 세계화 전략에 앞서가지 못한 이유도 있다. 혈연, 지연, 학연으로 구성된 SNS 보다는 관계 중심이나 뉴스중심의 SNS가 해외 시장의 글로벌 전략에는 더욱 확실한 커뮤니케이션 모델이었던 것이다.

소셜 데이터는 좋아하는 음식, 문자메시지, 구매 기록, ID, 위치정보, SNS 기록, 웹사이트 방문 기록, 검색 통계 등의 모든 데이터들을 분석하고 이를 계량화하여 소비자의 요구 및 만족할 수 있는 상품 등을 측정하는데 사용하

고 있다. 소셜 데이터의 특성을 보면 세가지로 분류할 수 있다. 정형(Structured), 준정형(Semi-Structured), 비정형(Unstructured) 데이터로 구분된다. 정형(Structured)은 고정된 필드에 저장된 데이터, 관계형 데이터베이스 및 스프레드시트 등이다. 준정형(Semi-Structured)은 고정된 필드에 저장되어 있지는 않지만, 메타데이터나 스키마 등을 포함하는 데이터, xml이나 HTML, 텍스트 등을 예로 들 수 있다. 비정형(Unstructured)은 고정된 필드에 저장되어 있지 않은 데이터, 텍스트 분석이 가능한 텍스트 문서 및 이미지/동영상/음성 데이터 등을 예로 들 수 있다. 이들의 특성은 <Table 1>과 같다.

<Table 1> Characteristic and classification of social data

| Group           | Characteristic   |
|-----------------|--|
| Structured      | Data which is stored in fixed field.<br>ex: relational database and spread sheet |
| Semi-Structured | Data contains meta data and schema,<br>ex: XML, HTML, text                       |
| Unstructured    | Data which is stored in non fixed field.<br>ex: image, video, voice data         |

<Table 1>은 소셜 데이터의 구분과 특성에 관련된 내용으로 소셜 데이터를 구분할 때 정형, 준정형, 비정형 데이터로 크게 구분되며 이러한 분류에 대한 다양한 해석 방식에 대해 그 특성을 알아 볼 수 있다.

## 3. 사례 분석

### 3.1 코카콜라 사례 연구

본 연구는 경험적 연구와 경험적 지식, 귀납적 추리를 강조하여 잠정적인 결론에 도달하는 방법인 경험주의적 방법으로 온라인상에 있는 ‘코카콜라’의 여론에 대한 소셜 데이터가 어떤 형식으로 분석되는지에 대한 사례연구 분석을 진행해 보았다. 코카콜라는 SNS 여론이 안 좋는데 이에 대한 위기대응능력이 뛰어나 연구 조사하게 되었다. 코카콜라 데이터를 분석하기 위해 분석된 사이트는 네이버랩, 구글트렌드, 펄스K, Socialmetrics 사이트이다. 코카콜라의 전체 평가를 분석한 결과 소셜 미디어 위기 대응 및 방법에 대한 그 이미지가 긍정효과를 미치는 반면 부정적인 이미지도 함께 나타나는 것을 볼 수 있는

것으로 조사되었다.코카콜라는 몸에 나쁘다는 인식을 가지고 있지만 그에 따른 예방이나 조치가 없는 상태에서 코카콜라의 다양한 마케팅과 SNS 홍보를 통해 단점을 보완 극복하고 있는 사례를 분석해 본다.

### 3.1.1 자료의 수집과 과정

결정된 연구문제를 기술, 분석하는데 도움이 되는 모든 자료들을 SNS 분석사이트를 기준으로 조사하였다. 코카콜라에 모든 SNS 상에 있는 자료들을 모으기에는 상당히 많은 양의 데이터가 있었다. 데이터의 질도 중요하지만 데이터의 양에 따른 분석을 위주로 기간을 정하여 트위터, 블로그, 페이스북, 티무데이 등의 소셜 데이터를 위주로 데이터를 모으는 분석 사이트를 기준으로 버즈량 분석, 인플루언서 분석, 연관어 분석, 감성분석이 제대로 이루어지는지에 대해 비교 분석해 보았다. 소셜 분석을 활용해 얻을 수 있는 버즈모니터링 (Buzz Monitoring), 인플루언서(Influencer), 연관 이슈 키워드, 감성 분석 (Sentiment Analysis)들을 가지고 소셜 데이터를 분석할 수 있는 소재를 찾아 보았다. 데이터에 대한 연구가 초창기인 지금은 버즈량에 대한 분석만으로 데이터를 분석하여 보여 주었다. 하지만 이런 데이터의 문제점은 양에 따라 긍정이나 부정을 단정 짓거나 한사람이 많은 양의 데이터를 반복적으로 올리면 의미가 점점 조각되어 진다는 것이다. 이러한 점을 보완하고자 그 버즈에 대해 영향력 분석 및 연관어 분석 및 감성 분석을 통하여 좀 더 가치 있는 정보를 추출하고 이를 통해 예측할 수 있는 다양한 사실들을 활용하고자 한다. 분석에 대한 자세한 내용은 <Table 2>와 같다.

<Table 2>에서 첫째 버즈 모니터링이란 특정 키워드가 소셜 미디어상에서 얼마나 많이 언급이 되는지를 확인하는 것이다. 특정 키워드에 대한 언급이 일시적으로 증폭되면 평소와 다른 일이 일어났음을 유추할 수 있다. 둘째 소셜 분석툴의 수집 대상은 소셜 미디어이며 자연어 처리 기술을 이용해 메시지가 담고 있는 호감도와 동시에 언급되고 있는 연관 키워드를 분석하여 검색엔진과 차별화 된 소셜 분석을 하려 한다. 메시지에 대한 자연어 처리를 통해 언급된 성향이 긍정적인지 부정적인지를 파악하는 것이 감성분석이다. 이는 펄스K를 통해 대규모 텍스트에서 대상에 대한 호감도를 즉시 분석해 여론의 경향성을 파악하는 방법이다. 셋째 연관 키워드 분석은 특정 키워드와 동시에 언급되고 있는 또 하나의 키워드

를 찾아내는 것이다. 연관 키워드 분석은 키워드와 관련된 호감도의 변화를 감지하며 라디안6(Radian6)라는 소셜 분석 툴로 해당 키워드와 동시에 언급되고 있는 단어 문치를 제공한다. 넷째 인플루언서 분석은 소셜미디어에서 영향력을 행사하는 사람을 말한다. 소셜영향력 측정 지수를 제공하는 클라우트(Klout)로 소셜 미디어 사용자의 참여도를 분석해 점수를 부여한다. 이 때 영향력 평가 기준은 트위터의 팔로워나 업데이트 빈도,페이스북 친구 수, 댓글 수, 좋아요 등이다.클라우트를 활용하면 어떤 사람이 소셜 미디어에서 얼마 만큼의 영향력을 갖고 있는지 어떤 성향의 영향력자인지도 가능할 수 있다. 이러한 4가지 단계를 이용하여 분석사이트에 들어오는 검색어들을 분석하였다.

<Table 2> Four steps of Framework of social data analysis

| Framework of social data analysis | Contents   |
|-----------------------------------|--|
| Buzz Monitoring                   | how many references are to be in social media on a specific keyword.   |
| Influencer                        | It means a person to exercise the influence of social media. To grant score by analyzing the participation of users of social media cloud client that provides the social influence measurement index and (Klout).   |
| Keywords associated issues        | It is intended to find a keyword another are mentioned at the same time as a particular keyword  |
| Sensitivity analysis              | It is a social media, by analyzing the related keywords that are referred to at the same time as the favorability message is on using natural language processing techniques, collected social analysis tools, social it is differentiated and search engine |

### 3.1.2 '코카콜라'에 대한 입력 데이터

코카콜라에 대한 많은 여론들은 블로그와 트위터 페이스북 등 다양한 소셜 네트워크 서비스를 통해 올려지고 재생산되며 재해석되어진다. 그것이 진실이든 거짓이든 간에 입소문이 나는데 걸리는 시간은 짧아졌고 비용은 없어졌다. 누구나 사용할 수 있고 쉽고 빠르다는 점과 편집되지 않은 데이터들과 정보들은 여과 과정 없이만 들어지고 이들을 분석하여 만들어진 분석내용들도 여과 없이 모든 사람들에게 노출된다.

<Table 3> Example of social data input of affirmation and negation in Coca Cola case

| Classification                                 | Information about the case.   |
|--|---|
| Input example of circumstances of the incident | <ol style="list-style-type: none"> <li>1. State high-speed police many of the United States are returnable by piling in the car Coca-Cola of about two gallons.</li> <li>2. In addition, emotion becomes unstable intelligence dropped the children eat too much cola, which accounts for 13% of the sugar.</li> </ol>  |
| Positive example                               | <ol style="list-style-type: none"> <li>1. I've never missed one of 20 years Best brand Coca-Cola</li> <li>2. I will go to the World Cup along with Coca-Cola.</li> <li>3. Campaign of "Open the Happiness" of Coca-Cola</li> </ol>  |
| Negative examples                              | <ol style="list-style-type: none"> <li>1. Intended use of the Coca-Cola, is to turn off the blood stains adhering to the road when the traffic accident occurred.</li> <li>2. During two days, the piece of meat you would multi Sakua and if I put the hospitality that was filled with Coca-Cola of beef steak.</li> <li>3. Coca-Cola is good to get rid of the time of the toilet bowl. Citric acid in the Coca-Cola will be cleanly removed was attached to the toilet bowl, when you haggard.</li> <li>4. Pour the Coca-Cola to battery cable you want the rust to car battery, no more rust while out bubbles.</li> </ol> |
| Examples of add-on products                    | <ol style="list-style-type: none"> <li>1. Rust is no longer clean To clean the rust chrome bumpers attached to the front and rear of the car, and wipe with paper soaked in Coca-Cola.</li> <li>2. To wash the clothes smeared with 2 oil, after having poured the Coca-Cola where oil is attached, you can wash off cleanly and washing by the addition of detergent.</li> <li>3. It is clean and wipe the paint with Coca-Cola when the windshield of the car was cloudy.</li> </ol>  |

<Table 3>은 코카콜라에 대한 소셜 데이터 긍, 부정 및 기능에 관한 입력 사례에 대한 내용으로 사건 경위에 대한 입력 예, 긍정과 부정적인 예, 제품의 부가 기능에 관한 예를 인터넷상에서 올라온 글들의 내용을 정리해 본 표이다. 이 내용은 무작위로 추출되었으며 내용에 관련된 내용은 코카콜라에 대한 다양한 관점에서의 생각이므로 옳고 그름의 판단이 아닌 여론성 글임을 밝힌다.

### 3.1.3 소셜데이터 분석에 대한 사례

코카콜라가 지닌 특성이나 문제를 종합적으로 진단

분석하기 위하여 심층적인 분석을 통하여 근본적인 특성이나 문제를 발견하고자 했다. 이를 분석하기 위해 국내 가장 많은 사람들이 찾고 있는 블로그나 트위터, 페이스북에 있는 글을 분석함과 동시에 소셜 분석 서비스가 갖고 있는 툴을 이용하여 그 데이터의 진위 여부와 데이터와 연관되어 있는 단어와 감성의 흐름 및 영향력 지수를 측정하였다. 데이터의 분석 기간을 정하고 그 기간 동안 발생하는 버즈량을 분석하였다. 네이버 소셜 네트워크 검색은 로그인한 상태에서만 서비스가 가능하다. 검색범위는 블로그 이웃 글들과, 가입한 카페 내 글, 미투데이 글들을 검색해 준다. 특히 네이버의 실시간 급상승 검색어는 단위 시간동안 네이버 검색창으로 입력되는 검색어를 분석해 입력 횟수의 증가 비율이 가장 큰 검색어 순서대로 보여주는 서비스로, 이용자들이 보여주는 관심의 흐름을 반영하고 있다. 실시간 급상승 검색어는 이용자들이 보여주는 관심의 흐름을 반영하기 때문에 짧은 시간에도 검색어 순위가 변경될 수 있다. 또한 연관 검색어는 검색 사용자의 검색의도를 파악하여 적합한 검색어 제공을 통해 이용자의 정보 탐색을 할 수 있도록 지원하는 서비스이다. '코카콜라'의 연관검색어를 보면 코카콜라, 크리스마스, 코카콜라의 비밀, 코카콜라 마케팅, 코카콜라 광고, 코카콜라 캔, 코카콜라 채용, 코카콜라 북극곰, 코카콜라 한정판등의 내용이 나온다. 네이버 소셜 네트워크 전체 내용 중 대부분 긍정이라기 보다는 기업 소개 또는 홍보성 글의 연관 검색어가 많이 나오는 것을 알 수 있다. <Table 4>은 버즈데이터 분석툴에 대한 종류를 나타낸다.

<Table 4>에서 제시한 버즈데이터 분석툴에 대한 설명을 자세하게 조사해 보았다. 다섯가지 툴에 대한 조사를 통해 이 툴들이 가지고 있는 분석 기능 및 내용을 살펴 보았다.

[Fig. 2]의 구글 트렌드에서 보여주는 검색은 0부터 100 사이를 기준으로 검색되다가 갑자기 증폭되는 현상을 볼 수 있다. 이에 따른 이슈를 분석된 화면으로 처리하면 A부터 M까지 어떤 이슈로 인해 코카콜라에 대한 이슈가 증폭되었는지를 볼 수 있다.

<Table 5>에서 소셜 인사이트에 올라온 탐색어 맵의 추이를 비교 분석하면 전체 연관어는 펍시를 시작으로 음료, 음료, ceo, 광고, 브랜드, 기업, 평창, 올림픽, 정당, 가치, 미국, 그림, 게임, 펍시콜라, 손가락으로 분류되는 것으로 조사되었다.

<Table 4> Analysis tools of Buzz data

| 버즈데이터 분석 툴  |   |   |  |
|---|---|---|--|
| (1)   | (2)   | (3)   | (4)  |
| NaverLab  | pulseK  | Google Trend  | SocialMetrics  |
| It can you search for the positive, the negative of the data contained in the blog. | It can you analyze data through the social value of the brand and company through social score, social awareness and favorability social. | When you search for the subject of interest of web search, indicate the statistical survey how people around the world if it were a lot of searching in the search. | It is shown by the ranking keyword sensibility week surge and keywords related keywords Gurudo~uru posted in blog and Twitter. |



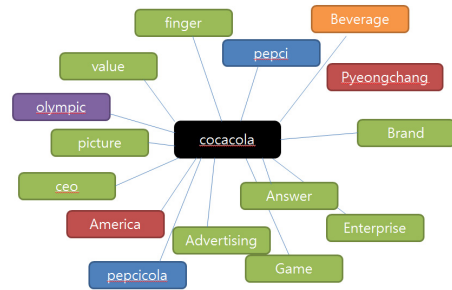
[Fig. 2] Interest trend of Coca Cola case by Google Trend. Ref: Googletrend, 2012

<Table 5> Social keyword analysis of Coca Cola case (Rearrange of social matrix 2012.12)

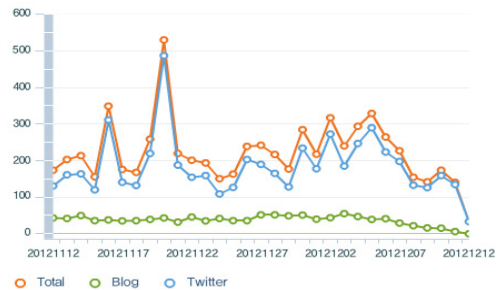
| Rank | Associated keyword ranking | Emotional keyword rankings | Weekly surge keyword ranking |
|------|----------------------------|----------------------------|------------------------------|
| 1    | pepsi                      | love                       | Company                      |
| 2    | Beverage                   | Fake                       | mother                       |
| 3    | ceo                        | Prosperity                 | took                         |
| 4    | Advertising                | Opposition                 | Squeeze                      |
| 5    | Brand                      | Red                        | Schedule                     |

[Fig. 3]에서 코카콜라와 연관 지어진 단어를 보면 속성에서 9개, 장소 2개, 브랜드 2개, 취미 또는 여가 1개로 분류되는 것으로 조사되었다. 자세히 분석해 보면 이 개수 또한 중복되는 글을 하나로 합쳐야 하는 분석이 필요성이 과제인 것으로 조사되었다.

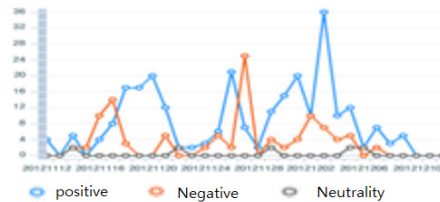
[Fig. 4]에서는 블로그와 트위터의 코카콜라에 대한 탐색어 추이에 대해 날짜별 추이를 보여준다. 코카콜라와 블로그에 올라온 글들이 비슷하게 그래프를 그리는 것을 볼 수 있다. 이것은 트위터상에 있는 글들이 블로그에 링크되어 나타내주고 있음으로 분석될 수 있다.



[Fig. 3] Relation word search by retrieval in Coca Cola case. Ref: Social Matrix 2012



[Fig. 4] Trend of search word in Coca Cola case. Ref: Social Matrix 2012



[Fig. 5] Trend of affirmation and negation. Ref: Social Matrix 2012

[Fig. 5]에서는 코카콜라의 긍정과 부정에 관한 추이를 보여준다. 소셜메트릭스에서는 긍정과 부정이 서로 맞대응 하면서 나타나는 것을 볼 수 있다. 이것은 코카콜라의 부정에 대한 반응을 댓글 피해 방지를위해 모니터링하며 대응하는 것으로 조사되었다.

### 3.1.4 코카콜라 소셜 분석결과

소셜데이터 분석 툴은 버즈모니터링, 인플루언서, 연관 이슈키워드, 감성 분석을 중심으로 4개의 사이트인 네이버랩, 펄스K, 구글트렌드, 소셜메트릭스가 소셜데이터를 어떻게 분석하고 있으며 그에 따른 영향력 및 내용들

을 분석해 보았다. 코카콜라를 검색어로 설정하여 데이터를 분석한 결과 소셜분석 사이트들에 대한 데이터의 내용과 실제 SNS 사이트에 올라오는 데이터와는 다소간 차이가 있는 것으로 분석되었다.

**<Table 6> Result of social data by analysis tools (N: Naver lab, K: Pulse K, G: Google trend, S: Social Matrix)**

| Framework of social data analysis | Analysis |   |
|-----------------------------------|----------|---|
| Buzz Monitoring                   | N        | Positive data if you'll continue to the specific data that people, the reliability of the negative is reduced.  |
|                                   | K        | Materials twitter whereas 80%, the data on Facebook is significantly lower 0.3%, the reliability is lowered. Also, I publish only a certain level data. |
|                                   | G        | A clear interpretation of the data to determine to have tens of thousands of search is not possible.  |
|                                   | S        | Standards come to the source of the data is limited to the blog and Twitter.  |
| Influencer                        | N        | Influence of power bloggers, which operates the blog, so are better than bloggers general, analysis of public opinion is even difficult.                |
|                                   | K        | Criteria for evaluating the social awareness is unknown.  |
|                                   | G        | The search has tens of thousands of impact can not be evaluated.  |
|                                   | S        | Classification tab of the associated control for the impact analysis is easy.   |
| Keywords associated issues        | N        | It is difficult to analyze the associated control.  |
|                                   | K        | It is difficult to analyze the associated control.  |
|                                   | G        | The keywords on issues associated between the naohjiman keyword connectivity drops.   |
|                                   | S        | Related searches is easy by us are shown in descending order of search order issues related keywords.   |
| Sensitivity analysis              | N        | Their because spindle document, the blogger who runs the blog, sensitivity analysis is easy.  |
|                                   | K        | Because I show positive, negative by analyzing the apartment, sensitivity analysis is easy.   |
|                                   | G        | The sensitivity analysis is difficult because it consists of the yearly event.  |
|                                   | S        | The sensitivity analysis presented for positive and negative keywords, content and technology placed.   |

<Table 6>을 토대로 결과를 정리하면, 소셜 데이터 분석 틀인 버즈모니터링 유형의 네이버랩으로 분석한 결과, 한 사람이 특정자료를 계속해서 올릴 경우 자료의 긍정, 부정에 대한 신뢰도가 떨어진다는 결과가 나왔다. 펄스K 분석틀에서 보면, 트위터의 자료는 80%인데 비해 페이스북의 자료는 현저히 낮은 0.3%이기 때문에 신뢰도가 떨어진다는 단점을 가지고 있는 것으로 조사되었다. 또한 자료를 일정수준 까지만 공개한다고 분석되었다. 구글트렌드는 검색수만을 가지고 판단하기 때문에 명확한 자료해석이 불가능하였다. 소셜메트릭스는 자료를 출처해오는 기준이 트위터와 블로그에 한정되어 있다는 단점을 가지고 있는 것으로 조사되었다.

소셜 데이터 분석 틀인 인플루언서 유형의 네이버랩의 경우는 블로그를 운영하는 파워 블로거들의 영향력이 일반 블로거들보다 월등하기 때문에 균등한 여론분석이 힘들다고 조사되었다. 펄스K 경우는 소셜 인지도를 평가하는 기준이 불분명했다. 구글트렌드의 경우에는 검색수만 가지고는 영향력을 평가할 수 없다고 조사되었다. 특히 소셜메트릭스 연관어 중 분류탭이 있어 영향력 분석이 용이했다. 연관어키워드 유형에서는 네이버랩과 펄스K의 경우 연관어 분석이 어려웠다. 구글트렌드는 연관어에 관한 키워드는 나오지만 키워드간의 연계성이 떨어지는 것으로 조사되었다. 소셜메트릭스는 연관어 키워드를 검색 순위에 따라 차례로 나타나 줌으로써 연관 검색을 용이하게 했다. 감성 분석의 경우 네이버랩은 블로그를 운영하는 블로거들이 주축이 된 문서이기 때문에 감성 분석이 쉬운 점이 있었다. 펄스K는 멘션을 분석하여 긍정,부정을 나타내어 주기 때문에 감성분석이 용이함으로 조사되었다. 구글트렌드는 년도별 사건을 중심으로 구성되어 있기 때문에 감성분석이 어렵다는 단점이 있었다. 소셜 메트릭스는 감성분석에 대해 긍정적, 부정적 키워드를 제시하고 그에 대한 내용 또한 기술해두었다는 점에서 소셜 메트릭스는 4가지 요건에 가장 적합한 소셜 분석 사이트로 평가 되었다. 최근들어 SNS 버즈량을 활용한 여론 조사가 활발해졌다. 그러나 아직도 여론 조사의 신뢰도와 정확성은 여전히 요구된다고 하겠다. 보다 정확한 데이터와 분석이 가능한 분석틀이 개발되기 전까지 단순 버즈량 데이터만 가지고 평가하는 틀에 절대적 신뢰를 보내는 것이 무의미하거나 위험한 일이 되는 이유이다.

3.1.5 채선당 소셜 분석결과

채선당을 검색어로 설정하여 데이터를 분석한 결과 소셜분석 사이트들에 대한 데이터의 내용과 실제 SNS 사이트에 올라오는 데이터와는 다소간 차이가 있는 것으로 분석되었다.

|   |  |
|---|--|
| G | Because it is organized around the events of each year, sensitivity analysis is difficult.     |
| S | Presents the keywords positive, negative for sensitivity analysis, it was also described that. |

<Table 7> Result of social data by analysis tools (N: Naver lab, K: Pulse K, G: Google trend, S: Social Matrix)

| Framework of social data analysis | Analysis |  |
|-----------------------------------|----------|--|
| Buzz Monitoring                   | N        | Documentation for pregnant women is greater than the material of Cheson hit when you search the Cheson per pregnant woman.             |
|                                   | K        | Similar to the case of Coca-Cola, fairness SNS search is reduced.  |
|                                   | G        | A clear interpretation of the data to determine to have tens of thousands of search is not possible.                                   |
|                                   | S        | Standards come to the source of the data is limited to the blog and Twitter.   |
| Influencer                        | N        | Influence of power bloggers, which operates the blog, so are better than bloggers common Analysis of public opinion is even difficult. |
|                                   | K        | Criteria for evaluating the social awareness is unknown.   |
|                                   | G        | It is not possible to evaluate the impact it will have several tens of thousand search.  |
|                                   | S        | There is a tab classification of related words, analysis of influence is easy.   |
| Keywords associated issues        | N        | Related term analysis is difficult.  |
|                                   | K        | Related term analysis is difficult.  |
|                                   | G        | Cooperation with the keyword coming out Keywords related to related issues fall.   |
|                                   | S        | Related searches is easy by us are shown in descending order of search order issues related keywords.                                  |
| Sensitivity analysis              | N        | Because it is a document that bloggers who run the blog has become the main shaft, sensitivity analysis is easy.                       |
|                                   | K        | Because I show positive, negative by analyzing the apartment, sensitivity analysis is easy.  |

<Table 7>을 토대로 결과를 정리하면, 소셜 데이터 분석 틀인 버즈모니터링 유형의 네이버랩으로 분석한 결과, 네이버 검색 시스템을 그대로 적용하여 채선당 입산 부를 찾을 경우 채선당 임신부에 관련된 글이 나오는 것이 아니라 채선당 따로 임신부 따로 관련글이 나오기 때문에 효과적인 긍정, 부정을 검색하기가 어렵다. 펄스k 분석틀에서 보면, 트위터의 자료는 80%인데 비해 페이스북의 자료는 현저히 낮은 0.3%이기 때문에 신뢰도가 떨어진다. 또한 자료를 일정수준 까지만 공개한다고 분석되었다. 구글트렌드는 검색수만을 가지고 판단하기 때문에 명확한 자료해석이 불가능하였다. 소셜메트릭스는 자료를 출처해오는 기준이 트위터와 블로그에 한정되어 있다는 단점을 가지고 있는 것으로 조사되었다.

소셜 데이터 분석 틀인 인플루언서 유형의 네이버랩의 경우는 블로그를 운영하는 파워 블로거들의 영향력이 일반 블로거들보다 월등하기 때문에 균등한 여론분석이 힘들다고 조사되었다. 펄스k 경우는 소셜 인지도를 평가하는 기준이 불분명했다. 구글트렌드의 경우에는 검색수만 가지고는 영향력을 평가할 수 없다고 조사되었다. 특히 소셜메트릭스 연관어 중 분류탭이 있어 영향력 분석이 용이했다. 연관어슈키워드 유형에서는 네이버랩과 펄스K의 경우 연관어 분석이 어려웠다. 구글트렌드는 연관어에 관한 키워드는 나오지만 키워드간의 연계성이 떨어지는 것으로 조사되었다. 소셜메트릭스는 연관어 키워드를 검색 순위에 따라 차례로 나타내 줌으로써 연관 검색을 용이하게 했다. 감성 분석의 경우 네이버랩은 블로그를 운영하는 블로거들이 주축이 된 문서이기 때문에 감성 분석이 쉬운 점이 있었다. 펄스k는 멘션을 분석하여 긍정,부정을 나타내어 주기 때문에 감성분석이 용이함으로 조사되었다. 구글트렌드는 년도별 사건을 중심으로 구성되어 있기 때문에 감성분석이 어렵다는 단점이 있었다. 소셜 메트릭스는 감성분석에 대해 긍정적, 부정적 키워드를 제시하고 그에 대한 내용 또한 기술해두었다는 점에서 소셜 메트릭스는 4가지 요건에 가장 적합한 소셜 분석 사이트로 평가 되었다. 최근들어 SNS 버즈량을 활용한 여론 조사가 활발해졌다. 그러나 아직도 여론



조사의 신뢰도와 정확성은 여전히 요구된다고 하겠다. 보다 정확한 데이터와 분석이 가능한 분석틀이 개발되기 전까지 단순 버즈량 데이터만 가지고 평가하는 틀에 절대적 신뢰를 보내는 것이 무의미하거나 위험한 일이 되는 이유이다.

### 3.1.6 알츠하이머 소셜 분석 결과

알츠하이머를 검색어로 설정하여 데이터를 분석한 결과 소셜분석 사이트들에 대한 데이터의 내용과 실제 SNS 사이트에 올라오는 데이터와는 다소간 차이가 있는 것으로 분석되었다.

**<Table 8> Result of social data by analysis tools (N: Naver lab, K: Pulse K, G: Google trend, S: Social Matrix)**

| Framework of social data analysis | Analysis |  |
|-----------------------------------|----------|--|
| Buzz Monitoring                   | N        | To survey data there is no relationship is coming up more and search data is difficult.  |
|                                   | K        | There is a limit to investigate the incident and diseases such as Alzheimer's disease.   |
|                                   | G        | Specific data interpretation for determining to have to only be able to search is impossible.  |
|                                   | S        | Standards come to the source of the data is limited to the blog and Twitter.   |
| Influencer                        | N        | Influence of power bloggers, which operates the blog, so are better than bloggers general, analysis of public opinion is even difficult. |
|                                   | K        | Criteria for evaluating the social awareness is unknown.   |
|                                   | G        | It is not possible to evaluate the impact it will have several tens of thousand search.  |
|                                   | S        | There is a tab classification of related words, analysis of influence is easy.   |
| Keywords associated issues        | N        | Related term analysis is difficult.  |
|                                   | K        | Related term analysis is difficult.  |
|                                   | G        | Cooperation with the keyword coming out Keywords related to related issues fall.   |
|                                   | S        | Cooperation with the keyword coming out Keywords related to related issues fall.   |
| Sensitivity analysis              | N        | Because it is a document that bloggers who run the blog has become the main shaft, sensitivity analysis is easy.                         |
|                                   | K        | Because I show positive, negative by analyzing the apartment, sensitivity analysis is easy.  |

|   |  |
|---|--|
| G | Because it is organized around the events of each year, sensitivity analysis is difficult.               |
| S | Presents the keywords positive, negative for sensitivity analysis, it should be also describes that was. |

<Table 8>을 토대로 결과를 정리하면, 소셜 데이터 분석 틀인 버즈모니터링 유형의 네이버랩으로 분석한 결과, 한 사람이 특정자료를 계속해서 올릴 경우 자료의 긍정, 부정에 대한 신뢰도가 떨어진다는 결과가 나왔다. 펄스k 분석틀에서 보면, 트위터의 자료는 80%인데 비해 페이스북의 자료는 현저히 낮은 0.3%이기 때문에 신뢰도가 떨어진다는 단점을 가지고 있는 것으로 조사되었다. 또한 자료를 일정수준 까지만 공개한다고 분석되었다. 구글트렌드는 검색수만을 가지고 판단하기 때문에 명확한 자료해석이 불가능하였다. 소셜메트릭스는 자료를 출처해오는 기준이 트위터와 블로그에 한정되어 있다는 단점을 가지고 있는 것으로 조사되었다.

소셜 데이터 분석 틀인 인플루언서 유형의 네이버랩의 경우는 블로그를 운영하는 파워 블로거들의 영향력이 일반 블로거들보다 월등하기 때문에 균등한 여론분석이 힘들다고 조사되었다. 펄스k 경우는 소셜 인지도를 평가하는 기준이 불분명했다. 구글트렌드의 경우에는 검색수만 가지고는 영향력을 평가할 수 없다고 조사되었다. 특히 소셜메트릭스 연관이 중 분류탭이 있어 영향력 분석이 용이했다. 연관이슈키워드 유형에서는 네이버랩과 펄스k의 경우 연관어 분석이 어려웠다. 구글트렌드는 연관 이슈에 관한 키워드는 나오지만 키워드간의 연계성이 떨어지는 것으로 조사되었다. 소셜메트릭스는 연관 이슈 키워드를 검색 순위에 따라 차례로 나타내 줌으로써 연관 검색을 용이하게 했다. 감성 분석의 경우 네이버랩은 블로그를 운영하는 블로거들이 주축이 된 문서이기 때문에 감성 분석이 쉬운 점이 있었다. 펄스k는 멘션을 분석하여 긍정,부정을 나타내어 주기 때문에 감성분석이 용이함으로 조사되었다. 구글트렌드는 년도별 사건을 중심으로 구성되어 있기 때문에 감성분석이 어렵다는 단점이 있었다. 소셜 메트릭스는 감성분석에 대해 긍정적, 부정적 키워드를 제시하고 그에 대한 내용 또한 기술해두었다는 점에서 소셜 메트릭스는 4가지 요건에 가장 적합한 소셜 분석 사이트로 평가 되었다. 최근들어 SNS 버즈량을 활용한 여론 조사가 활발해졌다. 그러나 아직도 여론 조사의 신뢰도와 정확성은 여전히 요구된다고 하겠다.

보다 정확한 데이터와 분석이 가능한 분석틀이 개발되기 전까지 단순 버즈량 데이터만 가지고 평가하는 틀에 절대적 신뢰를 보내는 것이 무의미하거나 위험한 일이 되는 이유이다.

#### 4. 연구의 함의와 한계

본 연구는 하루 수 백 만 건의 트윗을 어떠한 형식으로 분석할 것인가에 대한 소셜 분석 분류 방법을 연구하여 그 피해 사례 및 대응 방법을 찾기 위해 조사되었다. 연구 범위는 댓글 등으로 확산 전파되는 내용들을 버즈 모니터링 (Buzz Monitoring), 인플루언서(Influencer), 연관 이슈 키워드, 감성 분석(Sentiment Analysis) 등을 주요 소셜 분석 사이트를 중심으로 여론이 어떻게 분류되는지를 조사 분석하였다. 코카콜라, 채선당, 알츠하이머 사례중에서 감성분석의 관점을 예로 들면 펄스k는 멘션을 분석하여 긍정, 부정을 나타내어 주기 때문에 감성분석이 용이함으로 조사되었다. 구글트렌드는 년도별 사건을 중심으로 구성되어 있기 때문에 감성분석이 어렵다는 단점이 있었다. 소셜 매트릭스는 감성분석에 대해 긍정적, 부정적 키워드를 제시하고 그에 대한 내용 또한 기술해두었다는 점에서 소셜 매트릭스는 4가지 요건에 가장 적합한 소셜 분석 사이트로 평가 되었다. 각 사례별 연구를 보면 버즈모니터링, 인플루언서, 연관이슈키워드, 감성분석을 통해 이슈 및 분석범위를 설정하고 대화점유율 및 주요미디어나 연관어 분석 수집 멘션등의 내용을 분석하는 웹사이트나 분석틀들이 다소 많은 한계점을 가지고 있는 것으로 연구되었다. 인터넷에서 허구와 진실을 가려내는 문제는 순간적으로 밝혀질 수 있는 문제가 아니기에 시간이 지날수록 재 가공되며 이 문제가 새로운 알 수 없는 문제들로 급격하게 재생산 된다는데 그 위험성은 더욱 커진다고 했다. 정보는 받아들이는 개인의 성향과 직업 나이 연령별로 각각 다르므로 그에 맞는 정책을 마련한다는 것은 대단히 힘든 일이다. 이러한 점에서 볼 때 미디어리터러시를 통한 소셜 네트워크 서비스의 자율적 규제가 가능할 수 있도록 정부와 국민 모두가 사회문화적인 현상들에 대해 정직과 신뢰가 바탕이 되는 이야기 문화가 필요하며 SNS에 대한 교육과 기술적인 미디어의 해석과 접근보다 사이버 윤리에 대한 의식과 미디어 해독력을 통해 보다 올바른 소통 문화를 위한 연

구가 지속적으로 진행되길 기대해 본다.

#### ACKNOWLEDGMENTS

This study was supported by Seoul National University of Science and Technology.

#### References

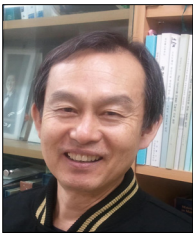
- [1] Hansuwan (2011.2). The impact of Twitter motivations and negative experience satisfaction. Master's thesis, Hanyang University, Graduate School of Public Policy, pp1-101
- [2] Information Society Agency Korea (2012.1.31.), Enhancing Literacy Research SNS users, pp3
- [3] Ju Anna (2011.8), Master Thesis: SNS research use online word-of-mouth information about the impact on consumer purchase intention, Inha University, pp10-50
- [4] Jim Stern(2011), Subdue social media, MulByungJaRi, pp50-100
- [5] Kim Ji Hye, Lee Hyung Hyo(2011.2). Language privacy policies for social networking services, and privacy protection module implementation, the Institute of Information Security and Cryptology, Vol. 21, No. 1, 2011.2, pp53-63
- [6] Kim Yong Hwan, Park Ji Hong(2009), Explore the behavior of SNS users who do not know (stranger) and social networks, 16th Proceedings of the Korean Society for Information Management, pp17-35
- [7] Kim Sang Hyun(2010.3), Social media and privacy, Journal of Information Science 제28권 제3호, pp41-46

### 황 윤 찬 (Yun Chan Hwang)



- 2013년 2월 : 서울과학기술대학교 정책학석사 (IT정책전문대학원 디지털문화정책학)
- 1998년 2월 : 숭실대학교 이학사 (정보통계학)
- 2005년 8월 ~ 2013년 2월 : 경희사이버대학교 미디어콘텐츠학 외래교수
- 2011년 12월 ~ 2013년 현재 : 디지털콘텐츠 제작 회사 스토리조아 대표
- 2001년 8월 ~ 2005년 7월 : 중앙일보 ITEA 웹마스터 과정 교육 강사
- 2009년 3월 ~ 2013년 현재 : 중앙공무원교육원 외 지자체, 관공서 소셜미디어 & 스마트폰 활용 교육
- 관심분야 : 디지털문화정책, 앱개발, 소셜미디어분석
- E-Mail : korea56@gmail.com

### 고 찬 (Chan KOH)



- 2008년 2월 : 서울대학교 경제학박사(기술정책학)
- 1991년 2월 : 경희대학교 공학박사(전자공학)
- 1983년 8월 : 연세대학교 공학석사(전산학)
- 1974년 2월 : 경희대학교 공학사(기계공학)
- 1992년 12월 ~ 1993년 12월 : 미국,NCSU대학 Post Doc.
- 1974년 10월 ~ 1978년 2월 : 해군장교 복무 (초계함승조 및 해군대학)
- 1987년 12월 : 정보처리 기술사 (한국산업인력공단)
- 2007년 4월 : 정보시스템 수석감리원 (정통부/ 서울체신청)
- 2006년 4월 ~ 2009년 3월 : 한국게임산업진흥원 이사
- 2005년 12월 ~ 현재 : 핀란드HMU대학교 강의겸임교수
- 2004년 12월 ~ 현재 : 한국디지털정책학회 부회장
- 1983년 10월 ~ 현재 : 서울과학기술대학교 교수
- 관심분야 : 기술정책, 경영정보, 컴퓨터그래픽스, 게임디자인
- E-Mail : chankoh@seoultech.ac.kr