JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 21, No. 4, pp. 771-779, Apr. 2020

Check for updates

머신 러닝 기반 시각화를 통한 악성 댓글 문제 완화 연구

박 지 현[1] · 김 주 섭[2*]
[1]서강대학교 아트&테크놀로지학과 석사과정
[2]서강대학교 아트&테크놀로지학과 부교수

Solving the Abusive Comments Problem through ML-based Visualization

Jihyun Park[1] · Jusub Kim[2*]
[1]Master's course, Department of Art & Technology, Sogang University, Seoul 04107, Korea
[2]Associate Professor, Department of Art & Technology, Sogang University, Seoul 04107, Korea

[요 약]

인터넷과 온라인 미디어의 발달로 다양한 관심사나 사회 이슈들에 대해 온라인상에서 댓글을 통해 의견을 교류하는 댓글 문화가 생겨났다. 하지만 악성 댓글들이 심각한 사회 문제로 대두되면서 댓글 문화 개선에 대한 요구가 높아지고 있다. 본 논문에서는 댓글 문화를 개선할 수 있는 하나의 해결 방안으로써 뉴스 댓글 시각화 시스템을 제안한다. 제안하는 시스템은 시각화, 인터랙션 디자인, 머신 러닝 기술을 활용하여 악성 댓글의 작성과 노출이 모두 자연스럽게 줄어들도록 유도하는 것을 특징으로 한다. 댓글을 남기는 사용자들에게 자율적으로 악성 댓글에 대한 경각심을 갖게 도와 악성 댓글 작성 행위를 줄이도록 돕고, 작성된 악성 댓글은 뉴스 커뮤니티 사용자들에 의해 크라우드 소싱 방식으로 자율적으로 노출이 줄어들도록 유도하는 자정기능을 제공한다. 효과성을 알아보기 위하여 프로토타입을 구현하여 10대~30대 100명에 대해 사용자 평가를 진행하였다. 실험 결과, 기존의 대표적인 뉴스 댓글 시스템과 비교하여 제안하는 방법이 통계적으로 유의한 차이를 보이며 악성 댓글 문제를 완화하는 데 도움이 될 수 있음을 확인하였다.

[Abstract]

With the development of online news media, a culture where many people actively share their opinions through online comments has emerged. However, the severity of abusive comments problem has also risen, increasing the demand for improvements to online commenting culture. This paper proposes a novel news comments visualization system to improve the commenting culture. The system aims to reduce the writing and exposure of abusive comments through visualization and interaction design with the help of machine learning. It helps users alert themselves to the severity of hate comments, encouraging them to avoid making such comments, and it also features crowd-sourced abusive comment filtering where the exposure of abusive comments are reduced by the crowd participation. In the experiment conducted on 100 users in their 10s to 30s, the participants provided positive responses that have statistically significant differences in comparison to an existing popular news comment system. The system can be used as an alternative to current news comment systems for better online commenting culture.

색인어 : 악성 댓글, 댓글 시각화, 뉴스 시각화, 크라우드 소싱, 머신 러닝
Key word : Abusive Comments, Comments Visualization, News Visualization, Crowd-sourcing, Machine Learning

http://dx.doi.org/10.9728/dcs.2020.21.4.771

Received 14 February 2020; **Revised** 15 April 2020
Accepted 25 April 2020

***Corresponding Author; Jusub Kim**

Tel: +82-2-705-7976
E-mail: jusub@sogang.ac.kr

Ⅰ. 서 론

인터넷과 온라인 뉴스 미디어의 발달로 온라인상에서 많은 사람이 다양한 관심사나 이슈들에 대해 댓글을 통해 활발히 의견을 교류하는 댓글 문화가 생겨났다. 댓글은 누군가가 인터넷에 올린 원문에 대하여 짧게 답하여 올리는 글이다[1]. 댓글은 자유로운 토론의 장을 형성한다는 점에서 긍정적인 기능을 갖지만, 익명성을 악용해 남을 비방, 혐오, 조롱, 또는 허위 사실을 퍼뜨릴 수 있다는 점에서 부정적인 기능도 동시에 가진다[2].

2000년대 들어 유명 연예인들의 자살 사건 등 악성 댓글로 인한 피해가 발생하면서 댓글 문화의 부정적 이면인 악성 댓글의 심각성에 대한 관심이 대두되기 시작하였다. 악성 댓글은 온라인에서 상대방에게 정신적 상처를 줄 수 있는 글을 작성하는 행위를 가리킨다[3]. 악성 댓글은 익명성, 비대면성, 집단성을 가지는 인터넷 댓글의 특성으로 인해 활성화 된다. 온라인상에서 자신의 신분을 노출하지 않는 익명성은 욕설, 비방, 그리고 근거 없는 소문들을 보다 자유롭게 인터넷상에 유포시키게 되고, 상대방과 대면하지 않기 때문에 타인을 살아있는 인격체로 보지 않게 하는 비대면성은 폭력적인 성격을 소유하고 있는 사람의 경우, 인터넷상에서 더욱 폭력을 행사하는 경향을 높일 수 있다. 또한, 비슷한 악의적 내용을 쓰는 사람들이 많아지면 집단 속에 숨어 악성 댓글을 다는 것에 대해 책임감을 크게 느끼지 않게 만든다[4].

악성 댓글을 줄이기 위한 여러 시도들에도 불구하고 여전히 악성 댓글 문제는 해결되지 않고 확장되고 있는 온라인 사회에서 해결해야 할 중요한 사회적 문제 중 하나로 인식되고 있다[5][6]. 2010년 이후에는 인스타그램, 페이스북 등 소셜네트워크의 발달로 인해 악성 댓글의 피해 대상이 유명인 뿐 만 아니라 일반인으로 확장되면서 악성 댓글은 보다 큰 사회적 문제로 주목받기 시작했으며 최근에는 매년 1만 건 이상의 사이버폭력 피해사례가 접수되고 있다[7].

본 논문에서는 악성 댓글 문제를 완화시킬 수 있는 새로운 뉴스 댓글 시스템 *SG-Comment (Self-reGulating Comment)*를 제안한다. 제안하는 시스템은 머신 러닝 기반으로 댓글의 긍 • 부정 정도를 자동으로 인식하는 것을 기반으로 하며, 시각화와 인터랙션 디자인을 통해 댓글을 작성하는 과정과 타인의 댓글을 읽는 과정 모두에서 악성 댓글의 작성과 노출이 자연스럽게 줄어들도록 유도하는 것을 특징으로 한다. 댓글을 남기는 사용자들에게 자율적으로 악성 댓글에 대한 경각심을 갖게 도와 악성 댓글 작성 행위를 줄이도록 유도하고, 작성된 악성 댓글은 뉴스 커뮤니티 사용자들에 의해 자율적으로 노출이 줄어들도록 유도하는 자정기능을 제공한다.

제안한 방법의 효과성을 알아보기 위하여 프로토타입을 구현하여 10대~30대 100명에 대해 사용자 평가를 진행하였다. 실험 결과, 기존의 대표적인 네이버 뉴스 댓글시스템(Naver News Comments)과 비교하여 통계적으로 유의한 차이를 보이며 악성 댓글 완화에 보다 도움이 될 수 있음을 확인하였다. 본 연구는 사회적 문제가 되고 있는 댓글 문화 개선을 위한 하나의 대안적인 시스템을 제안하고 효과성을 검증했다는 점에서 의의를 지닌다.

Ⅱ. 관련 연구

2-1 악성 댓글 문제점 개선 사례

악성 댓글의 문제점을 개선하기 위해 국내에서 시도했던 대표적인 사례는 2003년에 시행되었던 인터넷 실명제이다. 인터넷 실명제는 인터넷 이용자의 실명과 주민등록번호가 확인될 때만 인터넷에 글을 올릴 수 있도록 하는 제도로써 인터넷 상 익명성이 존재하는 환경에서는 여론 왜곡이 불가피하며 공론의 장에서의 자유로운 발언에는 책임감이 따라야 한다는 관점에서 대안으로 제시되었다. 하지만 2012년 헌법이 보장하는 개인 표현의 자유를 침해할 뿐만 아니라 공익의 효과도 미미하다는 점을 근거로 위헌 판결을 받아 폐지되었다[8]. 인터넷 실명제 제도가 폐지된 후에 국내 주요 포털사이트인 네이버와 다음 카카오에서 악성 댓글의 문제를 해결하고자 다양한 시도를 해오고 있다. 대표적으로 2017년 6월 23일, 네이버에서는 자사 뉴스 댓글 서비스에 '접기 요청' 기능을 추가하여 개편하였다. '접기 요청' 기능은 사용자가 보고 싶지 않은 댓글을 선택해 접기 요청을 하면 댓글 창에서 해당 내용을 바로 사라지게 할 수 있다. 또한 다수의 사용자가 접기 요청을 누른 댓글은 누적 요청 건수에 따라 자동으로 숨김 처리가 된다. 이 기능은 댓글을 신고할 수 있는 기존 기능보다 사용자의 의견을 더 신속하게 반영할 수 있는 장점을 가졌으나 많은 사람들이 보고 싶지 않다는 이유만으로 댓글을 못 보게 한다면 소수의견은 배제되는 결과를 초래할 수 있고, 다양한 댓글을 볼 권리가 침해 된다는 등의 비판으로 인해 현재는 제공하지 않고 있다. 카카오는 악성 댓글의 심각성이 반복되자 2019년 10월 31일부터 다음 연예 뉴스 댓글에 대한 부작용을 최소화시키기 위해 댓글 서비스를 폐지하였다. 사회 구성원들의 다양한 목소리를 듣는 장으로써 댓글 서비스를 운영하였지만 건강한 소통과 공론의 장을 마련한다는 목적에도 불구하고 그에 따른 부작용이 계속 되었기 때문에 이를 개선하고자 서비스를 잠정 폐지하였다.

최근에는 악성 댓글을 줄이고자 인공지능 기술을 활용한 대안이 나오고 있다. 네이버에서는 댓글 '클린봇'이라는 불쾌한 욕설이 포함된 댓글을 AI 기술로 감지하여 자동으로 숨기는 기능을 제공한다. 이 클린봇을 활성화하면 악성 댓글이 자동으로 숨겨진다. 구글 직소(Google Jigsaw)는 인공지능을 이용하여 자동으로 악성 댓글을 찾아내 이를 줄이기 위해 '퍼스펙티브(Perspective)'API를 공개하였다[9]. 페이스북과 인스타그램에서는 페이스북의 인공지능 시스템인 '딥 텍스트(Deep Text)'를 이용하여 악의적이고 혐오스러운 게시물 또는 악성 댓글을 직접적으로 차단할 수 있는 기능을 제공하고 있다.

2-2 댓글 시각화 연구 사례

이윤정 등[1]은 블로그 게시물에 달린 많은 수의 댓글을 사용자 정의 사전을 이용하여 내용에 따라 분류하고 이를 크기가 다른 붉은 색과 주황색 원으로 시각화 하는 시스템인 TRIB(Telescope for Responding comments for Internet Blog)를 제안하였다. 또한, TRIB연구의 후속연구로써 댓글을 내용의 유사도에 따라 여러 클러스터로 분류하고, 분류의 결과물을 태양계와 유사한 구조로 배치하여 색상과 모양이 다른 원으로 직관적으로 시각화하였다[10].

BBC에서는 뉴스에 대한 댓글을 시각화하는 시스템인 Spectrum을 개발하였다. 감정, 지역, 성별 등에 따라 댓글을 클러스터링하고 이를 시각화하였다. 각 댓글은 감정별로 분류되어 서로 다른 색의 원으로 시각화되며, 감정과 지역, 성별 등과 같이 그룹화 할 기준을 선택할 수 있는 사용자 인터페이스를 제공하고 있어 원하는 기준으로 댓글을 필터링할 수 있고, 움직이는 입자를 클릭하면 토론에서 사용된 댓글을 볼 수 있다[11]. Tsuda, K. 등은 블로그 토론 댓글들을 시각화한 새로운 형태의 애플리케이션을 제안하였다. 한 번에 전체의 댓글을 시각화하기 보다는 특정 범위의 점수를 가지고 있는 여러 개의 댓글 뭉치들을 연속적으로 시각화 하였다[12].

2-3 블로그 시각화 연구 사례

블로그 시각화에 대한 연구는 인터넷이 보급되면서 비교적 많이 진행되어 왔다. 대표적인 연구로는 Harris의 'We feel fine' 연구[13]를 들 수 있다. 이 연구에서는 일정 시간마다 전 세계에서 게시되는 블로그 게시물들을 수집하고 게시물에 포함된 감정 표현 문장들을 분석하여 행복, 슬픔, 우울과 같은 감정 상태로 분류한다. 각각의 감정 상태들은 색상이 다른 도형으로 표현되어 시각화 된다. 많은 블로그 게시물 혹은 기사들을 표현하고 있으나 어떤 블로그에서 게시되었는 지 혹은 게시물, 기사들의 앞, 뒤 연결을 알 수 없다는 단점이 있다. Indratmo 등은 블로그 시각화 도구인 iBlogVis 시스템을 제안하였다[14]. iBlogVis에서는 블로그 내의 게시물을 게시된 시간에 따라 수평인 시간 축의 위쪽에 배치하고 아래쪽에는 해당 게시물에 달린 댓글들의 개수나 글자 수를 고려하여 시간축의 아래쪽에 배치하였다. iBlogVis에서는 블로그 내의 게시물과 댓글의 전체적인 현황을 파악할 수 있으나 리스트로 제공 되는 것과 마찬가지로 댓글과 게시물과의 의미적 관계 등은 파악할 수 없다. Y. Takama 등은 블로그 공간에서 뉴스 기사와 블로그 게시물 그리고 블로그 사이트 분포를 시각화 하는 방법을 제안하였다[15]. 뉴스 기사의 중요도와 블로그 링크를 사용하여 블로그 사이트와 게시물들을 시각화 하였는데 게시물의 중요도는 조회 수로 계산하였다.

2-4 댓글 분석 및 분류 연구 사례

인터넷 게시물을 시각화 하는 것에 대한 연구에 비해 댓글의 분석 및 시각화에 대한 연구는 상대적으로 그 수가 적다.

Mishne 등은 블로그 게시물에 달린 댓글의 내용을 분석하여 논쟁 정도를 계산하는 방법을 제시하였다[16]. G. Mishne 등의 연구에서는 언어 모델을 이용하여 블로그의 글과 댓글, 댓글이 링크된 페이지 간의 유사도 비교를 통해 스팸 여부를 판단할 수 있는 방법을 제안하였다[17]. 또한, 클러스터링을 이용하여 스팸 필터링을 할 수 있는 연구도 진행되어왔다 [18][29]. 배민영 등은 토픽 시그너처(topic signature)를 이용하여 악성 댓글이 가지는 특징을 이용한 패턴 매칭 방법을 통해 악성 댓글을 분류하는 시스템을 제안하여 악성 댓글의 분류 성능을 개선할 수 있음을 보였다[20].

위와 같이 댓글과 블로그 게시물을 시각화한 다양한 연구에서의 목적은 수많은 정보를 한 눈에 볼 수 있게 하여 전체적인 현황을 파악할 수 있게 하는데 초점을 맞추고 있다. 본 연구에서의 시각화 목적은 악성댓글을 줄이는 데 있기에 기존 연구와 방향이 다르다고 할 수 있다. 또한, 악성 댓글을 필터링하여 자동으로 차단하는 강제적인 방법은 아직 그 기술이 완벽하지 않기 때문에 때로는 악성 댓글이 아닌 댓글을 차단하기도 하고 때로는 악성 댓글 차단에 실패하여 많은 피해자를 만들어 내고 있다. 본 연구에서는 사용자들이 보다 자율적인 판단으로 올바른 댓글 문화를 스스로 만들어 갈 수 있도록 유도하는 시각화 시스템을 제안한다는 점에서 차별점을 지닌다.

Ⅲ. SG-Comment 뉴스 댓글 시스템

본 논문에서 제안하는 새로운 댓글 시스템인 *SG-Comment*는 크게 네 가지 특징을 지닌다.

① 댓글의 긍 • 부정 시각화
② 색상 및 시간 지연 방식의 악성 댓글 필터링
③ 한 화면 내 전체 댓글 접근 기능
④ 댓글 중심의 뉴스 기사 시각화

3-1 댓글의 긍 · 부정 시각화

본 시스템은 머신러닝 기반 댓글 긍 • 부정 정도 인식 기능을 기본적으로 사용한다. 댓글 긍 • 부정 정도 인식 기능은 크게 두 가지 목적으로 사용된다.

우선, 사용자가 뉴스 기사에 대해 의견을 표하고자 댓글을 작성할 때 사용자가 입력하고 있는 댓글의 긍 • 부정 정도를 실시간으로 판단하여 즉각적인 피드백을 주는 데 사용된다. 사용자가 작성하고 있는 댓글이 긍정적이라 평가되면 화면에 어떠한 아이콘도 나타나지 않는다. 사용자가 작성하고 있는 댓글의 부정적 요소가 50%를 초과하면 댓글 창 위에 사이렌 아이콘 이미지를 나타내어 시각적으로 현재 작성하고 있는 댓글이 악성 댓글이 될 수도 있음을 1차적으로 경고 한다. 만약, 부정적 요소가 70%를 초과 하면 댓글 창 위에 나타났던 사이렌 아이콘 애

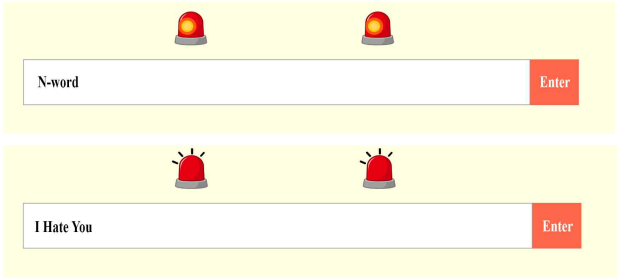


그림 1. 실시간 댓글 작성 피드백: 부정도 50% 초과 (위: 사이렌 아이콘), 부정도 70% 초과 (아래: 움직이는 사이렌 아이콘).
Fig. 1. Interactive feedback on the negativeness of comment: negativeness > 50% (top: siren icon), negativeness > 70% (bottom: animated siren icon)

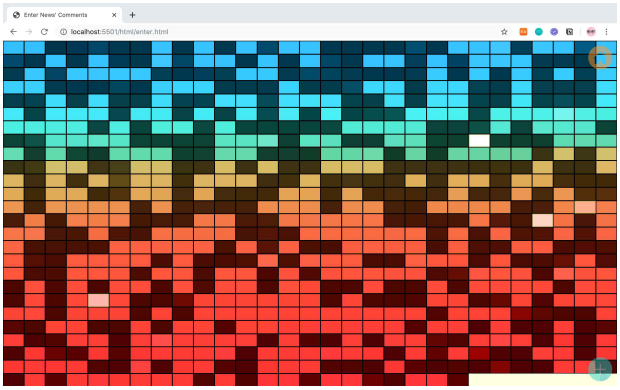


그림 2. 하나의 기사에 달린 모든 댓글들의 긍•부정 시각화
Fig 2. Visualizing negativeness of all the comments on one news article

니메이션이 활성화되면서 움직임으로 보다 강한 시각적 경고를 제공한다(그림 1). 일반 대중들에게 인식되는 사이렌의 시각적 의미는 경고(alert)로서 그것을 활용한 이러한 시각화 방법은 간단하지만 사용자들이 자신이 남기고 있는 댓글의 긍•부정 정도를 인지하고 자신의 말이 남길 영향에 대해 생각할 수 있는 기회를 주어 스스로 자정하게 하도록 고안되었다.

머신 러닝 기반의 긍•부정 자동 인식 기능은 또한 전체 댓글의 긍•부정 경황을 파악하고 그에 따른 댓글 검색 기능을 제공하기 위해 사용된다. 기사에 달린 각각의 댓글을 긍•부정 정도에 따른 색조(hue)와 채도(saturation)가 다른 작은 사각형으로 시각화함으로써 하나의 기사에 달린 전체 댓글의 긍•부정 경향을 시각화 한다. 긍•부정도에 따른 색 매핑은 표 1에 따라 이루어진다. 즉, 가장 부정적인 댓글은 채도가 높은(100%) 빨간색(hue=0)으로 시각화 되며, 가장 긍정적인 댓글은 채도가 높은(100%) 파란색(hue=200)으로 시각화 된다. 긍정도 부정도 아닌 중립적인 댓글은 채도가 50%인 녹색(hue=100)으로 시각화되며 그 외의 경우들은 모두 선형 보간(linear interpolation)에 따라 색조와 채도가 결정된다(표 1).

3-2 색상과 시간 지연 기반 악성 댓글 필터링

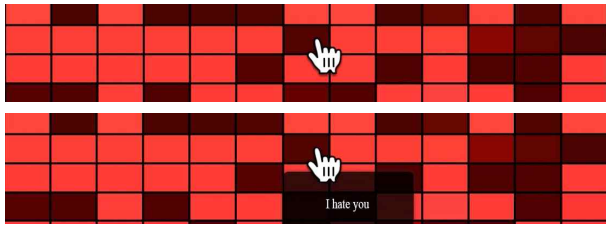


그림 3. 시간 지연 방식의 잠재적 악성 댓글 노출 차단. 부정도가 높은 (70% 초과) 댓글 중 비공감수가 공감수보다 많은 댓글은 사용자가 마우스 오버 한 후 수 초 후 (3초) 내용이 노출됨 (위: 마우스 오버시, 아래: 3초 후)
Fig. 3. Time-delayed exposure of a possible hate comment. Users have to wait for a few seconds (3 secs) to see the negative comment (negativeness > 70%) having more dislikes than likes

표1. 색조 및 채도를 이용한 댓글의 긍•부정도 시각화
Table 1. Visualizing positiveness/negativeness of a comment using hue and saturation

	Comment Negativeness: 100→0	Comment Positiveness: 0→100
Hue (0~359)	0(red)→100(green)	100(green)→200(blue)
Saturation (0~100%)	100%→50%	50%→100%

표 2. 2-단계 댓글 내용 노출 필터링
Table 2. 2-level comment exposure filtering design

	Negative comments having more dislikes than likes	All other comments
Level-I Filtering	Colored rectangles with low brightness	Colored rectangles with normal brightness
Level-II Filtering	Time delay (3 seconds)	-

본 시스템에서는 독자들에게 다른 사용자에 의해 작성된 댓글의 내용이 직접적으로 노출되지 않고 두 단계에 걸친 필터링을 거쳐 노출되도록 디자인되었다. 모든 댓글의 내용은 직접적으로 노출되지 않고 색상을 지닌 사각형으로만 표현된다. 긍정적인 댓글 또는 부정도가 높지 않은 댓글(부정도 70% 이하)은 사용자가 마우스 포인터를 해당 사각형에 올려 놓으면 바로 팝업 형태로 내용이 보여진다. 하지만 부정도가 높은 댓글(부정도 70% 초과) 중 다른 사용자들에 의해 공감(likes) 보다 비공감(dislikes)을 더 많이 받은 경우 해당 사각형은 낮은 명도(50%)로 표현되며 마우스 포인터를 해당 사각형에 위치하여도 바로 내용이 보여지지 않고 시간 지연(3초)을 주어 노출도를 더욱 제한한다. 즉, 일단 작성된 댓글은 대중의 참여에 의해 자율적으로 노출정도가 결정되도록 하였다. 이러한 크라우드소싱 방법은 악성 댓글을 프로그램으로 자동 검출하여 차단하는 방법에 비해 잘못된 차단이나 검출 실패 같은 오류를 줄일 수 있고, 사람들의 신고에 의해 사이트 관리자가 차단하는 방법에 비해서

는 수많은 댓글을 관리해야 하는 관리자의 부담을 줄이고 관리자의 개인적 판단이라는 변수를 제거하는 장점을 지닌다.

3-3 한 화면 내 전체 댓글 접근 기능

본 시스템에서는 일반적인 뉴스 플랫폼에서 사용하는 복수의 페이지 상에서의 댓글 열거(listing)와 화면 스크롤 방식으로 댓글을 보여주는 방법에서 탈피하여 한 페이지에서 모든 댓글을 접근할 수 있도록 디자인 하고 이러한 방식을 ACOP (All Comments on One Page) 이라 지칭한다.

ACOP 방식은 한 기사에 대한 독자들의 전체적인 반응과 생각의 경향을 파악할 수 있게 하고 그에 따라 독자들의 생각의 폭을 넓혀 일부 지엽적인 특정 악성 댓글로 인해 정신적 피해를 받는 경우를 방지하고자 디자인 되었다.

ACOP 방식의 또 다른 장점은 직관적인 검색이 가능하다는 점이다. 사용자들은 색조(긍 • 부정), 명도(대중의 순공감수: 공감수 - 비공감수), 마지막으로 깜박임(대중의 전체 반응수: 공감수 + 비공감수)으로 한 페이지 내에서 댓글을 특정하여 그곳에 마우스를 이동하는 방식으로 모든 댓글에 직관적으로 접근할 수 있다. 예를 들면, 특정 댓글이 사람들에게 반응을 많이 받으면 해당 사각형은 깜박임(blinking) 효과를 갖게 되어 독자들은 어떤 댓글이 가장 관심을 많이 받고 있는지 직관적으로 검색이 가능하다. 또한 그 댓글의 색조와 명도를 통해 그 댓글이 긍정적인 댓글인지 부정적인 댓글인지와 공감을 많이 받고 있는지 비공감을 더 많이 받고 있는지를 알 수 있으며, 깜박이면서 명도가 낮은 댓글의 경우 의견 대립이 심한 글이라는 것을 직관적으로 알 수 있다. 시간 지연 대상 댓글의 경우는 독자의 관심을 받지 않게 하기 위해 깜박임 효과에서 제외된다.

ACOP 방식의 마지막 장점은 독자들의 의견 개진 참여를 독려한다는 점이다. 순위에 밀릴 경우 대부분의 댓글들이 숨겨지고 접근이 어려워지는 리스트 방식과 달리 모든 댓글이 한 페이지에 보이기에 댓글 작성자는 자신의 댓글이 모든 다른 댓글과 함께 노출되는 것을 인식하고 성취감을 느낄 수 있으며 그에 따라 더 많은 참여를 유도하도록 디자인 되었다.

3-4 댓글 중심의 뉴스 기사 시각화

SG-Comment 시스템에서는 위에 서술한 댓글 시각화 방법을 뉴스 기사 시각화에도 확장하여 적용한다. 일반적인 온라인 뉴스 사이트에서 사용하는 기사 열거 방식에서 탈피해서 특정 기간 (예, 지난 3일) 동안의 모든 뉴스를 한 페이지에서 보여 주는 방식을 따른다 (ANOP; All News on One Page). ANOP 방식에서는 댓글 시각화와 같은 방식으로 각 뉴스 기사는 색상을 지니는 사각형으로 시각화 된다. 각 뉴스 기사에 달린 댓글들의 긍 • 부정도의 평균을 계산하여 색조와 채도로 나타낸다. 댓글이 많은 기사의 경우에는 깜박이는 효과를 주어 독자들이 직관적으로 사람들이 많은 반응을 보이고 있는 기사를 특정할 수 있도록 하였다. 각 기사를 대표하는 사각형에 마우스를 위치하면 기사 요약이 팝업 방식으로 나타나며 클릭할 경우 기사로 이동하게 된다.

ANOP 방식은 독자들이 최근 발생한 뉴스 기사들에 대해 어떤 반응을 보이고 있는지 경향을 직관적으로 파악할 수 있는 장점을 지닌다. 예를 들어, 대부분의 시각화된 기사들의 색이 붉은 색 계통이라면 독자들이 뉴스 기사들을 읽고 부정적인 의견을 많이 남겼다는 것을 의미하며 이는 현재 부정적인 사건들이 더 많이 뉴스화 되고 있다는 것을 암시한다.

ANOP 방식은 또한 기사들이 긍정에서 부정으로 댓글 시각화와 같은 방식으로 한 화면에서 정렬되어 나열되기 때문에 독자들은 그 기준으로 기사를 정하여 읽을 수 있는 장점을 지닌다. 즉, 어떤 독자들의 경우 긍정적인 기사 위주로 읽고 싶을 수 있는 데 그것을 가능하게 해준다.

마지막으로 ANOP 방식은 독자들에게 보다 많은 뉴스에 대한 접근성을 높여 생각의 폭을 넓힐 수 있는 기회를 제공한다. 일반적으로 뉴스 홈은 사람들이 많이 읽은 기사, 최신 기사, 또는 사이트에서 정해서 올리는 헤드라인 뉴스 기사 등 소수의 기사들만 독자들에게 노출이 된다. ANOP 방식에서는 특정 기간 동안의 모든 뉴스가 동일한 화면 크기를 가지며 노출이 되기 때문에 여러 이유로 특정 기사가 묻히는 일을 방지하게 된다.

Ⅳ. 실험 방법

4-1 실험 데이터

본 논문에서는 네이버 뉴스의 정치, 사회, 연예 부문에서 각 하나의 뉴스를 선정하여 댓글 데이터를 수집하였다.

댓글이 달렸다는 사실은 다른 사람들도 그 뉴스에 주목하였다는 점을 가시적으로 알 수 있게 해주는 일종의 신호(signal) 역할을 한다[10]. 따라서 정치, 사회 부문에서는 2019년 10월 14일, 정오~1시까지 집계한 조회 수가 가장 많은 뉴스 기사를 선정하였고, 연예 부문은 2019년 10월 14일, TV 연예뉴스에서 일간 많이 본 뉴스 중 랭킹 1위 뉴스 기사를 선정하여 각 뉴스기사에 달린 댓글의 내용, 공감수, 비공감수 데이터를 수집한 후에 각각의 댓글의 순공감수(공감수 − 비공감수)와 전체반응수(공감수 + 비공감수)를 계산하여 이를 추가한 새로운 데이터 셋(Set)을 만들었다.

4-2 긍 • 부정 머신 러닝 모델

수집한 댓글 데이터의 긍정, 부정 정도를 분류하기 위해서 네이버 영화의 평점 및 리뷰 데이터 셋을 사용하였다. 이 데이터 셋은 영화당 100개의 리뷰를 모아 총 20만개의 리뷰로 이루어져 있고, 1~10까지의 평점 중에서 중립적인 평점(5~8)은 제외하고 1~4점을 긍정으로, 9~10점을 부정으로 동일한 비율로

샘플링 되어있다[21].

본 논문에서는 이 데이터 셋을 기반으로 Python으로 구현된 머신 러닝 라이브러리 중 하나인 케라스(Keras)를 활용하여 댓글의 긍 • 부정도를 추측할 수 있는 모델을 만들었다. 가장 빈도수가 높은 1만개의 단어에 대해 Count Vectorization 기법으로 각 댓글 문장을 1만 차원의 벡터로 매핑하여 학습 데이터를 생성하였다. 학습을 위해 15만개의 네이버 영화 리뷰 데이터를 사용하였고 성능 평가를 위해 3만개의 데이터를 사용한 결과 85% 이상의 정확도를 보였다. 긍정 댓글 예로, '힘내세요~ 세상은 넓고 아름답습니다' 라는 댓글은 99.79%로 긍정 댓글로 분류가 되었고, '재수없으니~나오지마라..퉤퉤퉤' 라는 부정 댓글은 96.66%의 확률로 부정 댓글로 분류되었다.

4-3 실험 대상 선정 및 과정

본 연구에서는 제안하는 SG-Comment 시스템이 악성 댓글을 줄이고 댓글 문화를 개선하는데 효과적인지 알아보기 위해 온라인을 통해 모집한 10대~30대, 총 100명을 대상으로 네이버 뉴스 댓글 시스템과 비교 실험을 진행 후 설문조사를 진행하였다. 실험 집단은 네이버 뉴스 기사를 일주일에 한 건 이상 보는 사람들을 대상으로 선정하였다.

실험자들은 20분 동안 네이버 뉴스 사이트와 본 논문에서 제안하는 SG-Comment 시스템에서 각각 정치, 사회, 연예 3가지 분야의 기사를 읽고 기사에 달린 댓글들을 확인하고 댓글을 직접 작성해 보도록 안내 받았다. 시스템 체험 순서에 따른 영향을 최소화 하기 위해 역균형화(counterbalancing) 방식으로 실험을 진행하였다. 즉, 100명을 A, B 두 그룹으로 각각 50명 씩 나눈 뒤 A그룹은 네이버 뉴스 댓글 시스템을 먼저 경험한 후 제안한 방식을 사용하도록 하였고, B그룹은 본 논문에서 제안하는 시스템을 먼저 경험하고 다른 방식을 사용하도록 지시 받았다. 두 그룹 간 나이 및 성별 분포에 대한 동질성 검증 결과 통계적으로 유의한 차이를 보이지 않아 두 그룹은 동질한 그룹으로 간주하였다(표 3).

실험 참가자는 각 시스템을 체험한 후 설문조사에 응답하였으며 모든 설문은 7점 척도의 리커트 스케일로 측정되었다. 참가자들은 두 시스템 간 사용자 경험 비교 평가를 위한 설문 문항 5가지에 응답하였으며 제안하는 SG-Comment 시스템에 대해서는 추가로 5가지 문항에 응답하였다.

두 시스템 간 비교를 위한 설문 문항은 표4와 같이 구성하였고 참가자들은 매우 그렇지 않다(1)에서 매우 그렇다(7) 사이로 응답하였다.

제안하는 SG-Comment 시스템에만 관련된 추가적인 설문 문항은 "B-1. 한 페이지에서 세 가지 기준(전체공감순(깜박임), 순공감순(밝기), 긍부정순(색깔))으로 댓글을 검색하는 것이 도움이 되었다. B-2. 뉴스 홈에서 특정 시간에 사람들이 어떤 기사에 관심을 가지는 지, 각 기사에 달린 전체 댓글 수와 연동된 깜박이는 시각화와 아울러 그 댓글들의 평균 긍 • 부정도를 색깔로 시각화한 것이 기사를 검색하는 데 도움이 되었다. B-3. 뉴스 홈 시스템을 통해 전체 기사에 대한 독자들의 긍 • 부정 반응 경향을 쉽게 파악할 수 있었다. B-4. 이 댓글 시스템은 댓글 문화를 개선하는데 도움이 될 것 같다. B-5. 이 뉴스 시스템을 다시 사용하고 싶다"로 구성되었으며 역시 참가자들은 매우 그렇지 않다(1)에서 매우 그렇다(7) 사이로 응답하였다.

설문 후, 추가적으로 실험 참가자 중 5명을 임의로 선정하여 심층 인터뷰를 진행하였다.

표 3. 역균형화 실험을 위한 A, B 그룹에 대한 나이 및 성별에 대한 동질성 검증 결과.
Table 3. Homogeneity testing for counterbalancing group A and B in age and gender.

		Group A (N=50)	Group B (N=50)	p
Age		24.34	25.12	.560
Gender	Male	6 (12.0)	6 (12.0)	1
	Female	44 (88.0)	44 (88.0)	

표 4. 네이버 뉴스 댓글 시스템과 SG-Comment 시스템 간 비교를 위한 설문 문항.
Table 4. Questionnaire for comparison between Naver News Comments system and SG-Comment system.

Questions	
A-1	Was it easy to see what many readers thought about the content of the article through the comment system?
A-2	Was it easy to grasp the overall trend of positive / negative reactions of the readers about the content of the article through the comment system?
A-3	When writing a comment, did you think about the effect of the comments I leave on others?
A-4	Do you think abusive comments are easily exposed in the comment system?
A-5	Was it easy to access many news articles through the news home system?

표 5. 5 가지 사용자 경험 질문 항목에 대한 네이버 뉴스 댓글 시스템 vs SG-Comment 대응표본 t-검정 결과.
Table 5. Paired t-test results between Naver News Comments system vs SG-Comment system on UX questions

Questions	Naver News Comments	SG-Comment	t	p	dF
A-1	4.92±1.28	5.82±1.01	-6.25	$<0.001^{***}$	99
A-2	4.88±1.62	5.82±1.30	-5.10	$<0.001^{***}$	99
A-3	5.32±1.57	5.92±1.09	-4.24	$<0.001^{***}$	99
A-4	5.89±1.12	5.49±1.54	2.28	$.025^{*}$	99
A-5	5.52±1.11	5.66±1.15	-1.04	.302	99

Ⅴ. 실험 결과

5-1 정량 평가 결과

네이버 뉴스 댓글 시스템과 본 논문에서 제안하는 새로운 댓글 시스템 간 통계적으로 유의한 차이가 있는 지 알아보기 위해 대응표본 t-검증을 실시하였다. 분석 결과, 참가자들은 A-1, A-2, A-3 의 항목에서 네이버 뉴스 플랫폼보다 제안하는 SG-Comment 시스템에서 통계적으로 매우 유의한 차이를 보이며 더 높은 점수를 응답하였다 ($p<0.001$). 즉, 참가자들은 제안하는 SG-Comment 시스템에서 전반적인 독자들의 생각을 긍•부정 경향 정보와 함께 쉽게 파악할 수 있었고, 댓글 작성 시 타인에게 미칠 수 있는 영향에 대해 보다 더 생각한 것으로 나타났다. 참가자들은 A-4 항목에서도 통계적으로 유의한 차이를 보이며 ($p<0.05$) 제안한 SG-Comment 시스템에서 악성 댓글이 덜 노출 될 것 같다고 응답하였다. 하지만, 뉴스 홈 시스템을 통해 많은 뉴스 기사에 쉽게 접근 할 수 있었는 지를 물어 본 A-5 항목에서는 두 시스템 간 통계적으로 유의한 차이가 없었다(표 5).

제안한 SG-Comment 시스템에 대해서만 추가로 물어본 B 설문 항목에서 참가자들은 5개 항목 모두에서 평균 점수 5점 이상을 응답하여 (B-1(M=5.5), B-2(M=5.8), B-3(M=6.1), B-4(M=5.5), B-5(M=5.5)) 제안하는 기능들에 대체로 만족하고 재사용 의도를 가지며 댓글 문화 개선에도 도움이 될 것으로 생각하는 것으로 응답하였다.

5-2 정성 평가 결과

실험 참가자 중 임의로 선정한 5명을 대상으로 심층 인터뷰를 진행하여 전체적인 만족도와 함께 시각화 시스템에 대한 긍정적/부정적 피드백을 알아보았다. 긍정적인 피드백 중 가장 많이 언급된 사항은 댓글의 긍•부정도를 시각화한 점과 작성하는 댓글의 긍•부정을 판단하여 사이렌 아이콘으로 시각화 한 점이었다.

"댓글을 보는 이유가 전체적으로 사람들이 해당 기사에 대해서 어떤 생각을 가지는지, 어떠한 감정을 나타내고 있는지 살펴보기 위해 댓글을 읽는데 색과 채도, 그리고 명도로 댓글을 직관적으로 볼 수 있게 하는 부분에서 도움이 되었다. 또한, 서로 다른 의견이 충돌할 때, 상대방의 입장도 볼 수 있어서 완충재 역할을 하여 댓글 문화를 개선하는데 도움이 될 것 같다." (23세, 여성)

"색으로 긍정적인 댓글과 부정적인 댓글로 나눠져 있어 보고 싶은 성향의 댓글만 볼 수 있어서 보기 편했다. 악성 댓글을 볼 때 느꼈던 피로도를 줄이는데 도움이 되었다"(20세, 여성)

"부정적인 댓글을 작성할 때 사이렌 아이콘이 나타나서 중간에 작성한 댓글을 다시 보게 되었다. 그 이후로 댓글을 쓸 때마다 신경을 써서 댓글을 작성하게 되었다." (28세, 남성)

한편, 부정적인 피드백으로는 뉴스 메인 화면의 시각화 방식이 불편했다는 의견이 가장 많이 언급되었다.

"댓글은 색깔로 시각화해서 보여주는게 도움이 된다고 생각하는데 뉴스 홈 화면에서는 내용 위주로 노출이 되면 좋을 것 같다. 사람들의 반응 보다는 기사에 대한 정보를 바탕으로 시각화하면 좋을 것 같다."(26세, 여성)

이외에 댓글의 긍•부정 판단 정확도에 대해서 몇 차례 언급되었다. 댓글의 긍•부정 판단이 정확하다고 느껴지지 않을 때가 있어서 시스템에 대한 신뢰도가 떨어진다고 응답하였다.

Ⅵ. 논의

실험결과에서 가장 주목해야 할 부분은 실험 참가자들이 기존 포털 사이트 뉴스 댓글 시스템(Naver News Comments)과 비교하여 제안한 SG-Comment 시스템에서 댓글을 작성할 때 내가 남기는 댓글이 타인에게 미칠 수 있는 영향에 대해 더 생각하였는지(A-3), 댓글 시스템에서 악성 댓글이 쉽게 노출될 것 같은지(A-4), 그리고 댓글 시스템이 댓글 문화를 개선하는데 도움이 될 것 같은지(B-4) 라고 할 수 있다. 긍정적인 B-4 결과 (M=5.5)와 함께, A-3과 A-4 모두에서 통계적으로 유의한 차이를 보이며 긍정적인 결과가 나온 것은 주목할 만한 결과라고 할 수 있다. 특히 내가 남기는 댓글이 타인에게 미칠 수 있는 영향에 대해서 더 생각하게 되었다는 부분(A-3)에서 통계적으로 매우 유의한 차이($p<0.001$)를 보이며 긍정적인 결과가 나오고 심층 인터뷰에서도 가장 많이 언급되었다는 점은 그러한 인터랙티브한 피드백이 간단하면서도 큰 실효성이 있을 수 있음을 알려준다고 할 수 있다. 다만, 그러한 피드백이 반복될 경우 효과가 경감될 수 있기에 장기적인 사용에 따른 효과성 검증이 필요하다.

제안하는 기능 중 뉴스 메인 홈 화면을 댓글과 유사한 방식으로 시각화 한 부분은 긍정적인 피드백도 있었지만 부정적인 피드백도 많았는데 심층 인터뷰에서 보고된 것처럼 댓글과 기사는 서로 다른 특성을 지니고 사용자가 원하는 것이 다르기 때문에 댓글에 적용된 시각화를 확장하는 것은 효과적이지 않았다고 추측할 수 있다. 뉴스 메인 화면의 UX 디자인 연구에 대한 보다 심층적인 탐구가 필요하다.

본 실험의 결과는 몇 가지 제한점을 가진다. 첫째, 댓글의 긍•부정을 판단하는 모델이 실제 댓글 데이터를 학습시킨 것이 아니라 네이버의 영화 평점 댓글 데이터를 학습시켰기 때문에 모델의 정확도가 실험 결과에 영향을 미쳤음을 배제할 수 없다. 이것은 심층 인터뷰에서도 '신뢰도가 떨어진다'고 부정적인 의견으로 보고되었는데, 이에 따라 추후 실험에서는 실제 댓글 데이터를 학습시킨 모델을 활용하여 실험을 진행할 필요가 있으며 아울러 머신 러닝 모델도 현재의 기본적인 모델이 아닌 좀 더 정교한 모델의 사용이 필요하다. 둘째, 이 실험은 10대~30대를 대상으로 진행되었기 때문에 뉴스 플랫폼을 사용하는 50대~60대 등의 다른 높은 연령대로 결과를 일반화시키기는 어

렵다. 마지막으로, 제안하는 방법은 데스크 탑 컴퓨터 사용에 최적화 되어 있어 최근 많은 뉴스 접근이 이루어지는 모바일 기기에서의 사용은 별도의 연구가 필요하다.

Ⅶ. 결론

본 논문에서는 악성 댓글을 줄이고 댓글 문화를 개선할 수 있는 방법으로써 새로운 뉴스 댓글 플랫폼을 제안했다. 본 논문에서 제안하는 SG-Comment 시스템은 머신 러닝 기반으로 댓글 작성 중 댓글의 긍 • 부정도를 실시간으로 판단하여 그에 따른 시각화를 제공함으로써 자기 검열의 기회를 제공하고, 색상과 시간 지연 방식으로 악성 댓글의 노출도를 줄이며, 한 페이지에서 모든 댓글을 직관적으로 검색할 수 있는 기능을 특징으로 한다. 10대~30대 100명을 대상으로 진행한 예비 실험 결과 참가자들은 통계적으로 유의한 차이를 보이며 제안한 SG-Comment 시스템에서 기존의 일반적인 댓글 방식보다 댓글 작성 시 타인에게 미칠 수 있는 영향에 대해 더 생각하였으며, 댓글 문화를 개선하는 데 도움이 될 수 있다는 긍정적인 평가를 얻었다. 본 연구는 악성 댓글의 심각성이 앞으로도 지속적으로 대두될 것으로 예상되는 온라인 문화에서 악성 댓글을 줄이고 건전한 댓글 문화를 만드는데 대안적인 시스템으로 사용될 수 있을 것이다. 향후, 댓글 시각화와 연계된 보다 나은 뉴스 홈 화면 디자인 연구와 모바일 디바이스 등 다른 플랫폼에서의 UX를 고려한 후속 연구가 필요하다.

참고문헌

[1] Lee, Y., Ji, J., Woo, G. and Cho, H. TRIB : A Clustering and Visualization System for Responding Comments on Blogs. The KIPS Transactions:PartD, 16D(5), pp.817-824, 2009.

[2] lee, y. (2015). [online] Bulmanzero.com. Available at: http://www.bulmanzero.com/news/articleView.html?idxno=14337 [Accessed 14 Oct. 2019].

[3] Ybarra, M. and Mitchell, K. Online aggressor/targets, aggressors, and targets: a comparison of associated youth characteristics. *The Journal of Child Psychology and Psychiatry*, 45(7), pp.1308-1316, 2004.

[4] Changhwan Kim, “Malicious Comments and Countermeasures”, ITFIND 1437, 2010.3.17

[5] Koo, H. and Seo, E. A Study on Injurious Comment Spam－Its Typology and Suggestions for Improvement. Korean Language Research, null(30), pp.5-32, 2012.

[6] H. G. Park. "Paralinguistic Expressions and ‘Publicness’ of Language of On-line Messages". *The Journal of Speech, Media & Communication Research,* Vol. 10, pp.7-37, Dec 2008.

[7] Police.go.kr. (n.d.). [online] Available at: https://www.police.go.kr/www/open/publice/publice03_2018.jsp [Accessed 13 Jun. 2019].

[8] Constitutional Court of Korea, 2012.08.23. 2010 Heon-Ma 47

[9] Hosseini, Hossein, Kannan, Sreeram, Zhang, Baosen and Poovendran, Radha. Deceiving Google's Perspective API Built for Detecting Toxic Comments, 2017.

[10] Lee, Y., Ji, J., Woo, G. and Cho, H. Analysis and Visualization for Comment Messages of Internet Posts. *The Journal of the Korea Contents Association*, 9(7), pp.45-56, 2009.

[11] Anon, (n.d.). [online] Available at: http://www.bbc.co.uk/white/spectrum.shtml. [Accessed 6 Jun. 2019].

[12] K. Tsuda and R. Thawonmas, “Visualization of Discussions in Comments of a Blog Entry Using KeyGraph and Comment Scores,” *Proc. Of 4th WSEAS International Conference on E-ACTIVITIES, Florida, USA*, Vol. 5, pp. 21-26, 2005.

[13] Jonathan Harris, We Feel Fine, 2006, [online] Available at: www.wefeelfine.org [Accessed 11 Mar. 2020].

[14] Indratmo, Julita Vassileva, and Carl Gutwin, “Exploring blog archives with interactive visualization,” In Proceedings of the Working Conference on Advanced Visual Interfaces, pp.:39-46, 2008.

[15] Takama, Y., Matsumura, A. and Kajinami, T. Interactive Visualization of News Distribution in Blog Space. *New Generation Computing*, 26(1), pp.23-38, 2007.

[16] G. Mishne and N. Glance, "Leave a reply: An analysis of weblog comments," In Third annual workshop on the weblogging ecosystem, 2006.

[17] G. Mishne and D. Carmel, "Blocking Blog Spam with Language Model Disagreement," 1st International Workshop on Adversarial Information Retrieval on the Web. pp.1-6, 2005.

[18] L. Xiao-bing and N. Zhang, "Incremental Immune-Inspired Clustering Approach to Behavior-Based Anti-Spam Technology," *International Journal of Information Technology,* Vol.12, No.3, pp.111-120, 2006.

[19] W.-F. Hsiao, T.-M. Chang, and G.-H. Hu, "A cluster-based approach to filtering spam under skewed class distributions," In HICSS, pp.53-59, 2007.

[20] Min-Young Bae, Jeong-Won Cha, “Comments Classification System using Topic Signature”, *The Journal of KIISE : Software and Applications* (SA), Vol.35, No.12, pp.774-779, 2008.

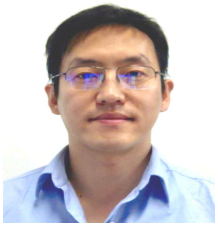
[21] Cyc1am3n's Blog. (n.d.). [online] Available at:

https://cyc1am3n.github.io/2018/11/10/classifying_korean
_movie_review.html [Accessed 11 Aug. 2019].

박지현(Jihyun Park)

2017년 : 국민대학교 (공학사-컴퓨터공학)

※관심분야 : 데이터 시각화, 딥러닝



김주섭(Jusub Kim)

2000년 : 연세대학교 (공학사-전자공학)
2002년 : 연세대학교 대학원 (공학석사-전자공학)
2008년 : 미국 메릴랜드대 대학원(공학박사-컴퓨터공학)
2008년~2012년 : 미국 Rhythm & Hues Studios
2012년~현재 : 서강대학교 아트&테크놀로지 학과 교수
※관심분야 : Creative Computing, HCI, New Media 등