

페이스북 그룹 게시물 분석을 통한 우울증 관련 주제에 대한 고찰

Investigating Major Topics Through the Analysis of Depression-related Facebook Group Posts

주 영 준 (Yongjun Zhu)*, 김 동 훈 (Donghun Kim)**
이 창 호 (Changho Lee)***, 이 용 정 (Yongjeong Lee)****

목 차

- | | |
|----------|------------|
| 1. 서론 | 4. 연구 결과 |
| 2. 선행연구 | 5. 논의 및 결론 |
| 3. 연구 방법 | |

초 록

본 연구는 소셜 네트워크 서비스인 페이스북에서 우울증 관련 게시물을 분석하여 그 안에서 주로 논의되는 주제를 파악하고자 한다. 구체적으로, 접근 용이성, 개방성 및 익명성 등의 특징을 지니는 페이스북이라는 온라인 커뮤니티에서 사용자들이 다소 민감한 정신적 질환인 우울증에 관하여 어떤 내용을 논의하는지 살펴보고자 한다. 본 연구를 위해 페이스북 데이터 수집에서부터 주제어 추출에 이르기까지의 전반적인 과정을 포함하는 자연어 처리 기반의 데이터 분석 프레임워크를 구현하였다. 구현한 프레임워크를 이용하여, 본 연구는 우울증을 논의하는 페이스북 최대 사용자 그룹에서 최근 1년간 작성한 885개의 게시물을 수집하여 분석하였다. 주제어 추출의 완성도와 정확도를 위해 자동화된 기법과 수동적인 접근법(불용어 제거, 주제어 개수 지정)을 결합하였으며, 이를 통해 주제를 다각도에서 분석하였다. 분석 결과, 사용자들은 우울증 일반, 인간관계, 기분 및 느낌, 우울증 증상, 자살, 의료 참고, 그리고 가족 등에 대한 논의를 주로 하는 것으로 파악되었다.

ABSTRACT

The study aims to analyze the posts of depression-related Facebook groups to understand major topics discussed by group users. Specifically, the purpose of the study is to identify the topics and keywords of the posts to understand what users discuss about depression. Depression is a mental disorder that is somewhat sensitive in the online community, which is characterized by accessibility, openness and anonymity. The researchers have implemented a natural language-based data analysis framework that includes components ranging from Facebook data collection to the automated extraction of topics. Using the framework, we collected and analyzed 885 posts created in the past one year from the largest Facebook depression group. To derive more complete and accurate topics, we combined both automated and manual (e.g., stop words removal, topic size determination) methods. Results indicate that users discuss a variety of topics including depression in general, human relations, mood and feeling, depression symptoms, suicide, medical references, family and etc.

키워드: 소셜미디어, 소셜 네트워크 서비스, 페이스북 그룹, 우울증, 자연어 처리, 토픽 모델링
Social Media, Social Network Services, Facebook, Depression, Natural Language Processing, Topic Modeling

- * 성균관대학교 문헌정보학과 조교수(yzhu@skku.edu / ISNI 0000 0004 7933 6219) (제1저자)
** 성균관대학교 문헌정보학과 석사과정(antanta@skku.edu) (공동저자)
*** 성균관대학교 문헌정보학과 석사과정(lch519@skku.edu) (공동저자)
**** 성균관대학교 문헌정보학과 조교수(redpapa01@skku.edu / ISNI 0000 0004 7411 8220) (교신저자)
논문접수일자: 2019년 10월 9일 최초심사일자: 2019년 11월 6일 게재확정일자: 2019년 11월 19일
한국문헌정보학회지, 53(4): 171-187, 2019. <http://dx.doi.org/10.4275/KSLIS.2019.53.4.171>

1. 서론

최근 이슈가 된 진주아파트 방화사건, 강서구 PC방 살인사건, 그리고 60대 가장의 가족 살인 사건(매일경제 2019년 7월10일)과 같은 흉악 범죄행위들은 조현병이나 우울증과 같은 정신질환과 관련이 있는 것으로 나타났다. 이러한 정신질환은 조기에 치료를 받는 것이 중요하나, 정신과 질환의 경우, 사회적 수치심의 정도가 높아 정신과 방문을 꺼려하여 그 치료를 제대로 받지 못하는가 하면, 심지어 그러한 질환이 있음에도 불구하고 그 질환을 인정하는 것조차 어려워하는 경향이 있다(차지영, 차미영 2012; Lachmar et al. 2017). 본 연구에서 살펴보고자 하는 우울증은 널리 퍼져 있는 정신 질환의 하나로 우리사회에서도 그 심각성은 오래전부터 인지되어왔다(고광운 외 2016; Zhu et al. 2018). 우울증 증세가 있는 사람들은 자신의 심리적 상황을 다른 사람들에게 표현하여 감정적 내지는 사회적 지원(social support)을 받고자 하는 경향이 있으며 이를 위해 소셜 미디어를 활용하는 것으로 나타났다(Bazarova et al. 2017). 트위터나 페이스북 또는 소셜 Q&A 사이트(예, 지식인, 야후! Answers)와 같은 소셜 미디어는 우울증을 관찰하기에 매우 적절한 플랫폼이다. 이는 이용자들이 자신의 익명성을 보장받을 수 있는 장점으로 인해 우울증과 같은 민감한 건강관련 주제에 대해 자신들의 생각이나 정보를 자유롭게 공유할 수 있고, 무엇보다도 소셜미디어에서 생성된 빅데이터를 수집할 수 있어 표본 수의 한계를 극복할 수 있기 때문이다(Choudhury, Counts and Horvitz 2013; Lachmar et al. 2017).

소셜 미디어는 그 사용목적이나 주요 콘텐츠 측면에서 상이한 성격을 가지고 있다(이재범 외 2012). 소셜 Q&A 사이트의 경우 이용자들은 자신들의 건강문제에 대한 구체적인 질문들을 하고 여러 사람들로부터 이 질문들에 대한 답변을 구하는 정보추구활동을 벌인다(Yi 2018). 한편, 페이스북이나 트위터는 어떤 질문에 대한 답을 제공하는 방식보다는 주로 이용자들이 서로 간의 관계형성을 통해 자신의 질병이나 건강문제에 대한 의견이나 상태를 일상담화의 형태로 표현하거나 글을 주고 받는 형식으로 정보공유가 이루어진다. 특히, 페이스북은 온라인 네트워크를 통해 많은 사람들과 “친구”라는 사회적 관계를 형성하고 자신의 생각과 경험을 공유할 뿐 아니라 시시각각 변하는 자신들의 감정과 상태를 업데이트하기 때문에 페이스북 이용자가 어떤 상태에 있는지를 파악하기 좋고, 이러한 환경은 우울증을 진단하거나 그 증세를 예측하는 데 효과적이다(Guntuku et al. 2017; Nadeem 2016). 소셜 미디어가 제공하는 정보공유의 장에서 이용자들은 자신의 정보요구를 만족시키기 위한 활동을 활발하게 하고 있으며, 이러한 정보 공유는 서로의 생각 뿐 아니라 감정, 그리고 더 나아가 건강관련 의사결정에도 영향을 미치는 것으로 나타났다(목양숙 2015; 황유선 2016; Bae and Yi 2019).

따라서, 최근에는 소셜 미디어에 표현된 대량의 텍스트를 분석하는 기법을 통해 우울증과 같은 특정 질환자들이 자주 사용하는 언어의 특징을 파악하고 그 특징에 따라 관련 질환에 대한 이해를 돕고자 하는 연구가 진행되고 있으며(Lachmar et al. 2017; Resnik et al. 2015), 더 나아가 그 텍스트의 분석을 활용하여 우울증의

정도를 진단하거나 우울증을 예측하는 모델을 개발하는데 주의를 기울이고 있다(고광운 외 2016; Nadeem 2016). 그러나 소셜 미디어 데이터로 우울증을 이해하는 연구는 아직 신생 분야일 뿐 아니라, 특히 국내에서는 한글 텍스트를 분석하는 연구가 매우 초기 단계에 있다. 따라서 본 연구에서는 소셜 미디어를 통해 생산되는 우울증에 관한 수많은 생각과 감정 등을 담은 한글 빅데이터를 분석함으로써 우울증에 대한 이해를 높이고 이를 통해 우울증을 예방하거나 인지하는데 기여하는 실용적 함의를 제고할 뿐 아니라 우울증에 관한 한글 빅데이터의 분석 기술을 정교화시키는 학문적 지평을 확장하고자 한다.

2. 선행연구

2.1 소셜 미디어와 건강정보추구행태

소셜미디어에서 논의되는 건강관련 정보탐색의 주제들은 구체적인 질병에서부터 공중보건 문제에 이르기 까지 다양하다(Zhao and Zhang 2017). 소셜 미디어를 통해 건강정보를 탐색하는 이용자들의 특징을 살펴 본 바에 따르면, 온라인 건강정보탐색에 대한 자기 효용이 높을 뿐 아니라 건강에 대한 우려가 많거나 자신의 건강 상태에 대한 자신감이 낮은 사람일수록 건강정보를 탐색하는 정보원으로서 소셜 미디어를 선호하는 것으로 나타났다(김수정 2012). 소셜 미디어는 수많은 사람들과 빠르게 소통할 수 있는 장점으로 인해 건강에 관한 문제를 지닌 이용자에게는 특정 질병에 관한 정보를 탐색하거나 공유하기에 매우 적합한 장이다. 건

강에 관한 질문은 일반 질문과 비교할 때, 매우 개인적이거나 민감한 내용(예, 정신질환, 성 관련 질병)이 많기 때문에 사람들은 익명성이 보장되는 소셜 미디어가 자신의 생각이나 질문을 자유롭게 표현할 수 적절한 곳으로 인식한다(Aldarwish and Ahmad, 2017). 이와 더불어 소셜 미디어를 제공하는 특별한 혜택 중의 하나는 이용자들 간의 상호작용을 통해 사회적 지원 또는 감정적 지원을 받을 수 있다는 점이다(Lachmar et al, 2017; Yi 2018). 정신질환이나 성 관련 질병으로 힘들어 하는 사람들의 경우, 사회적 수치심으로 인해 의료 전문가에게 조차 편안하게 대면하여 말 할 수 없는 내용들이 많고, 무엇보다도 감정적으로 예민한 경우가 많기 때문에, 질환에 관한 정확한 정보를 추구하기 보다는 오히려 사회적 내지는 감정적 지원, 즉 자신의 문제에 대한 공감을 받고자 소셜 미디어를 활용한다(Zhao and Zhang 2017).

또한, 명확한 치료방법이 없거나 충분한 임상 경험이 부족한 의학적 한계 내지는 현실적인 한계로 인해 의료 전문가에게서도 만족스러운 답을 찾기 힘들 때, 사람들은 집단 지성(Crowd's wisdom)에 의지거나 소셜 미디어를 방문하는 수많은 사람들 중에서 유사한 경험을 가진 이용자로부터 그 답을 찾고자 활발한 정보탐색활동을 한다(Bae and Yi 2017; Song et al. 2016). 이러한 활동은 전통적인 웹 검색과 달리 다른 사람들과의 상호작용을 통해 정보를 탐색한다고 하여 사회적 검색(social search)으로 정의되었다(Shah, Kitzie and Choi 2014). 이러한 사회적 검색이 소셜 미디어 사용자들에게 널리 이용되면서 그들이 어떤 기준을 가지고 정보를 신뢰하거나 선택하는지에 대한 연구들이 주목받

았다(Ha, Aikat and Jung 2015; Zhang 2013). 즉 전문가 뿐 아니라 많은 일반인들에 의해서 제공된 수 많은 정보들 중에는 잘못된 정보도 많고 이러한 정보는 소셜 미디어에 대한 불신감을 조장한다. 특히, 질병이나 치료에 관련된 정보들은 잘못된 정보일 경우 인체에 미치는 해를 클 수 있기 때문에 소셜 미디어 사용자가 정보의 품질을 어떻게 판단하는지 내지는 어떠한 정보를 선호하는지에 대한 이해가 필요했으며 이는 건강정보제공자들에게 효과적인 정보전달에 대한 함의를 제공했다. 즉, 소셜 미디어 이용자들은 건강정보를 평가하는 데 있어서 사회적 규범이나 손실 프레이밍(loss framing)과 같은 메시지의 특성 뿐 아니라 정보제공자의 전문성이나 참고정보(references)와 같은 정보원의 존재 유무를 중요하게 고려했으며(Bae and Yi 2017; Jin et al. 2016), 텍스트의 가독성(readability) 수준이나 위험 정보(risky information) 제공 여부를 포함하는 인지적 요인 이외에도 공감, 긍정 및 부정적 감정 내지는 낙관 및 비관적 정보와 같은 감정적 요인 또한 이용자들의 정보 채택에 유의미한 영향을 미치는 것으로 파악되었다(Chua and Banerjee 2015; Fong et al. 2014; Yi 2018).

페이스북, 트위터, 인스타그램 그리고 유튜브 등은 국제적으로 가장 활발하게 이용되는 소셜 미디어들일 뿐 아니라 그 각각의 특색이 다르기 때문에, 이용자들은 그 각각의 소셜 미디어를 사용하는 데 있어서 상이한 동기와 목적을 지니며, 따라서 그 특정 소셜 미디어의 성격에 맞게 정보를 생성한다(이재범 외 2012; Kneidinger 2010; Reavley and Pilkington 2014). 예를 들면, 페이스북과 트위터는 관계형성이라는 공통

된 사용목적은 지니지만 페이스북 이용자들이 신변잡기 정보를 주요콘텐츠로 공유하여 엔터테인먼트적 요소를 사용 동기 내지 목적으로 인식하는 것에 반해 트위터 이용자들은 자신의 현재의 상태나 개인적 감정을 주요 콘텐츠로 삼아 그러한 정보의 공유를 목적으로 한다(이재범 외 2012). 또한, 본 연구에서 살펴보고자 하는 페이스북의 경우, 다른 소셜 미디어와 비교할 때 그 이용자들은 사회관계에 필요한 정보를 추구하는 경향이 높으며, 다시 말해, 새로운 친구를 사귀거나 기존 친구들을 유지하기 위해서 페이스북을 사용하는 것으로 나타났다(Kneidinger 2010). 이는 페이스북이 다른 사람들과의 관계나 상호 작용을 활발히 하는데 자주 이용되는 것을 의미하며, 이를 뒷받침하는 연구에 따르면, 페이스북의 게시물은 그것을 읽은 사용자의 감정변화에 상당한 영향을 미치는 것으로 나타났다(목양숙 2015). 즉, 페이스북 게시물이 긍정적인 내용일 때에는 독자도 긍정적인 감정을 느끼며, 반대로 부정적일 때에는 독자도 우울함과 같은 부정적인 감정을 느낀다는 것이다. 이와 함께, 페이스북 게시물에 나타난 감정을 기쁨, 의문, 희망, 분노, 그리고 슬픔 등으로 분류하고 각 감정이 표현되는 경향을 분석한 연구(황유선 2016)를 통해서도 페이스북이 사람들과의 상호작용을 통해 이용자 자신의 생각뿐 아니라 감정을 표현함으로써 스트레스를 표출하는 수단으로 활용된다는 것을 파악할 수 있다.

2.2 우울증과 소셜 미디어

소셜 미디어가 우울증을 관찰하기에 적합하

다는 점은 여러 연구에서 관찰되었으며(Guntuku et al. 2017; Nadeem 2016), 우울증이 있는 사람과 그렇지 않은 사람들의 소셜 미디어에 대한 태도를 비교한 연구에서 그 밀접한 연관성을 찾을 수 있다. 즉, 우울증이 없는 사람들은 소셜 미디어를 단순히 정보를 공유하거나 소비하는 수단으로 여기는 반면 우울증이 있는 사람들은 소셜 미디어를 감정교류 내지는 사회적 인식(social awareness)이 이루어지는 수단으로 받아들인다는 것이다. 이는 우울증이 있는 사람들이 소셜 미디어를 자신들의 감정을 표출하고 사회적 지원(social support)을 구하는 장으로 인식하는 것으로 이해할 수 있다. 실제로 트위터 사용자들이 표현한 우울증에 대한 태도를 분석한 연구들을 살펴보면 상당수의 내용들이 우울증을 겪는 사람들에 대한 사회적 지원이나 도움을 제공하는 것으로 나타났다(Cavazos-Rehg et al. 2016; Reavley and Pilkington 2014).

소셜 미디어 이용자의 텍스트를 분석함으로써 우울증을 진단하거나 예측하는 연구들이 지속적으로 진행되었다. 우울증이 있는 이용자들이 자주 사용하는 언어를 분석하거나(Lachmar et al. 2017), 우울증 증세가 있는 사람들과 그렇지 않은 사람들이 언어를 사용하는 방식의 차이를 분석하는 연구들이 이에 속한다(Resnik et al. 2015). 우울증세가 있는 사용자들의 텍스트를 분석한 결과에 따르면 그것은 7가지 주제로 분류되며, 이는 역기능적 사고(예, 자기비하, 과장적 사고 등), 생활방식의 어려움, 사회생활의 어려움, 무감정과 슬픔, 자살에 관련된 생각과 행동, 심리적 안심의 추구, 그리고 가면 뒤에 숨기 등이 속한다(Lachmar et al. 2017).

소셜 미디어의 사용과 우울증과의 관계를 보

다 확장시켜 조사한 연구에 따르면, 그 두 요인의 상관관계는 다른 요인들, 즉 심리적, 사회적, 행태적, 그리고 개인적 요인들과도 복잡하게 연관된 것으로 나타났다(Baker and Algorta 2016; Yoon et al. 2019). 이를 뒷받침하는 연구에서도 우울증은 심리적인 스트레스는 물론 페이스북 사용과도 연관이 높은 것으로 나타났다. 즉, 스트레스가 높을수록 페이스북에서 우울증과 관련된 단어를 많이 사용하는 것으로 나타났을 뿐 아니라(Bazarova et al. 2017) 페이스북 사용 시간과 높은 신경질적인 증상 간에는 상관관계가 있는 것으로 파악되었다(Chow and Wan 2017).

페이스북 이용자를 대상으로 우울증을 분석한 연구에 따르면, 우울증의 경향이 높으면 부정적인 포스팅의 빈도가 늘어날 뿐 아니라(Scherr and Brunet 2017), 우울증 증상이 있는 사용자들의 경우, 페이스북을 이용하는 빈도가 높을수록 우울증에 대해 많이 언급하는 것으로 나타나 그들이 페이스북을 통해 자신의 증상이나 상태를 다른 사람들에게 표현한다는 것을 파악할 수 있다(Moreno et al. 2011). Yoon et al. (2019)의 연구에서도 이와 유사한 결과가 확인되었다. 즉, 소셜네트워크 서비스를 사용하는 시간이 길거나 이를 확인한 빈도가 많을수록 우울증의 정도가 높은 것으로 나타났으며, 특히, 사회적 비교(social comparison), 즉, 타인과 자신을 비교하는 내용의 텍스트를 많이 사용할수록 우울증과 연관이 높은 것으로 나타났으며, 이 사회적 비교는 Scherr, Toma and Schuster (2018)의 연구에서 언급하는 사회적 감시(social surveillance) - 예, 친구의 게시물을 브라우징하는 행위 - 와 유사한 개념으로 타인을 부러워하는 경향이 많을수록 우울증과 연관이 높은 것

을 시사한다.

우울증에 대한 이해를 높이고자 하는 연구들과 더불어 우울증을 진단하거나 치료하고자 하는 시도들이 다양하게 나타났다. 그 주류 중의 하나는 소셜 미디어에 이용자가 생성한 데이터를 기반으로 우울증 수준을 측정하여 분류하는 모델을 개발하거나(Aldarwish and Ahmad 2017), 페이스북 이용자의 데이터를 기반으로 우울증을 탐지하는 알고리즘을 개발하는 연구들이다(Katchapakirin et al. 2018). 이러한 도구들의 개발은 우울증 문제로 힘들어 하는 이용자들이 자신의 위험도를 스스로 진단하여 그 정신 질환의 문제를 예방적(proactive)으로 극복하는데 도움을 주는데 기여했다. 한편, 선행 연구들은 우울증과 다른 질환들과의 연관성에 대해서도 지속적으로 관찰해 왔다. 심장질환(Bar et al. 2004; Lichtman et al. 2008), 고혈압(Meng et al. 2012), 불면증(Irwin 2003), 알코올 중독(Boden and Fergusson 2011), 그리고 당뇨병(Katon et al. 2005; Knol et al. 2006) 등이 우울증과 관련이 높은 질환에 속하며, 이러한 질환들과 우울증, 그리고 그 치료제에 대한 선행 연구들을 분석함으로써 이러한 세 가지 요소들 간에 어떠한 상호작용이 발생하는지를 조사하고 이를 바탕으로 어떤 특정 질병이나 약이 우울증에 영향을 미치는지를 파악하고자 했다(Zhu et al. 2018).

3. 연구 방법

3.1 데이터

페이스북 그룹은 공동 관심사를 가진 사용자들이 모여서 정보를 공유하고 의견을 나누는 온라인 커뮤니티이다. 소셜 네트워크 서비스의 특성상 사용자들은 익명성이 보장되는 환경에서 자유롭게 소통할 수 있다. “우울증”이라는 검색어로 페이스북 그룹을 검색하면 십여 개의 그룹이 검색결과로 나타나며, 본 연구의 연구 대상은 검색 결과에 포함된 그룹 중 사용자수가 가장 많은 그룹이다. 데이터 수집 당시에 해당 그룹에는 1,400명의 사용자가 있었고, 그들의 익명성 보장을 위하여 데이터 수집과정에서 사용자 식별 정보는 제거되었다. 본 연구는 해당 그룹에서 2018년 4월부터 2019년 4월까지 총 214명의 사용자가 작성한 885개의 게시물을 수집하여 분석하였다. 데이터 통계는 <표 1>과 같다.

3.2 연구 방법

본 연구의 목적은 페이스북의 우울증 그룹 사용자들의 게시물에 대한 분석을 통하여 페이스북에서 어떤 우울증 관련 주제들이 논의되고

<표 1> 데이터 통계

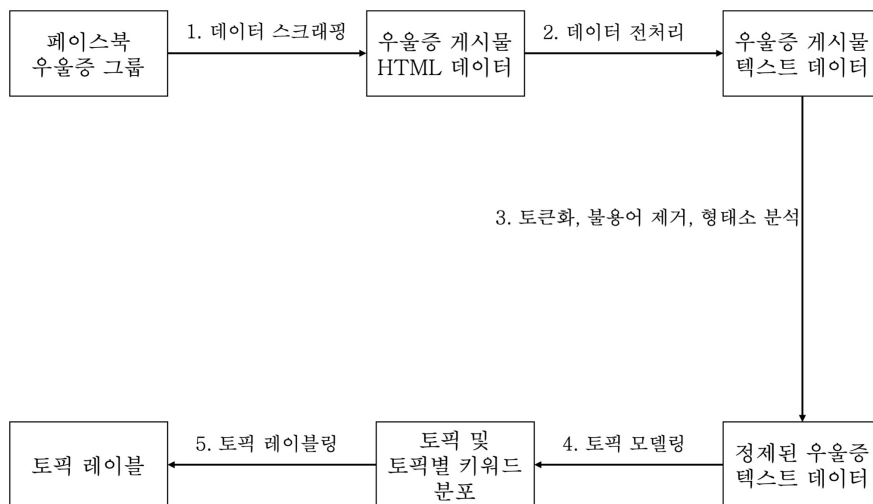
구분	최대	최소	평균
사용자 별 게시물 수	89	1	4
게시물 별 글자 수	2,398	5	89
게시물 별 댓글 수	90	1	3
게시물 별 좋아요 수	17	1	5

있는가를 이해하는 것이다. 따라서 자연어 처리 기법의 하나인 토픽 모델링을 이용하여 주제들을 추출하였다. 즉, 기존의 많은 연구에서 활용된 Latent Dirichlet Allocation(LDA) 기법(Blei, Ng and Jordan 2003)을 활용하여 각 주제를 대표할 수 있는 주제어들을 추출하고, 그 후 추출한 주제어들에 근거하여 주제를 정의하였다. 전반적인 데이터 분석 프레임워크는 <그림 1>과 같다.

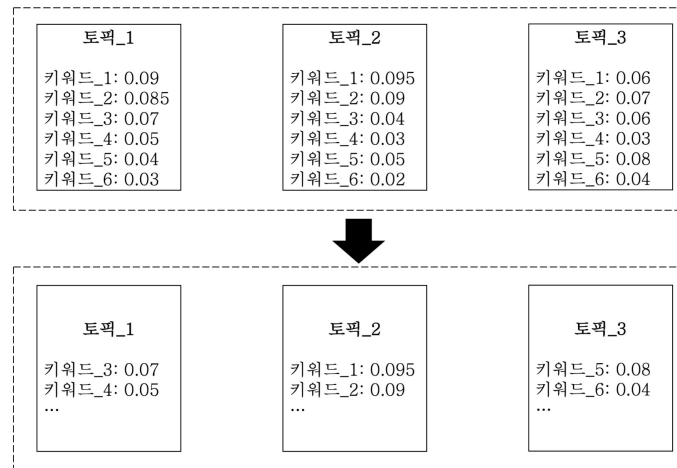
페이스북 우울증 그룹의 게시물은 파이썬 데이터 스크래핑 라이브러리를 이용하여 수집되었다. 수집된 HTML 데이터에서 데이터 전처리 과정을 통해 HTML 태그들이 제거가 되고 텍스트 데이터만 추출이 되었다. 한국어 형태소 분석기인 KoNLPy(박은정, 조성준 2014)의 Twitter Korean Text Class를 사용하여 토큰화 및 형태소 분석을 진행하였고 수동으로 불용어를 제거하였다. 토큰화된 텍스트 데이터는 Gensim(Řehůřek and Sojka 2010) 라이브

러리를 통하여 주제 및 주제별 주제어 분포가 추출되었다. 주제어 추출 시 Gensim 라이브러리의 기본 하이퍼 파라미터 조합이 사용되었다. 주제는 주제어들과 각 주제어들이 해당 주제에 속할 수 있는 확률로 구성 되어있다. 따라서 각 키워드들이 모든 주제에 각각 다른 확률값으로 존재한다. 본 연구는 여러 개의 주제에 중복되어 출현하는 키워드를 제거하기 위해 키워드들의 각 주제별 확률을 비교하고 제일 큰 확률값을 가지는 주제에만 해당 주제어를 포함시켰다 (<그림 2> 참조).

<그림 2>와 같이 키워드_1과 키워드_2는 토픽_2에서 가장 큰 확률값을 가지기 때문에 토픽_2에 배정되었다. 같은 방법으로 키워드_3과 키워드_4는 토픽_1에 키워드_5와 키워드_6은 토픽_3에 배정되었다. 이렇게 각 주제별로 유일하게 존재하는 키워드들을 각 주제를 대표할 수 있는 주제어로 간주하고 그 키워드들이 나타내는 주제에 대해 레이블링 하였다.



<그림 1> 데이터 분석 프레임워크



〈그림 2〉 주제 및 주제별 주제어 추출 방법

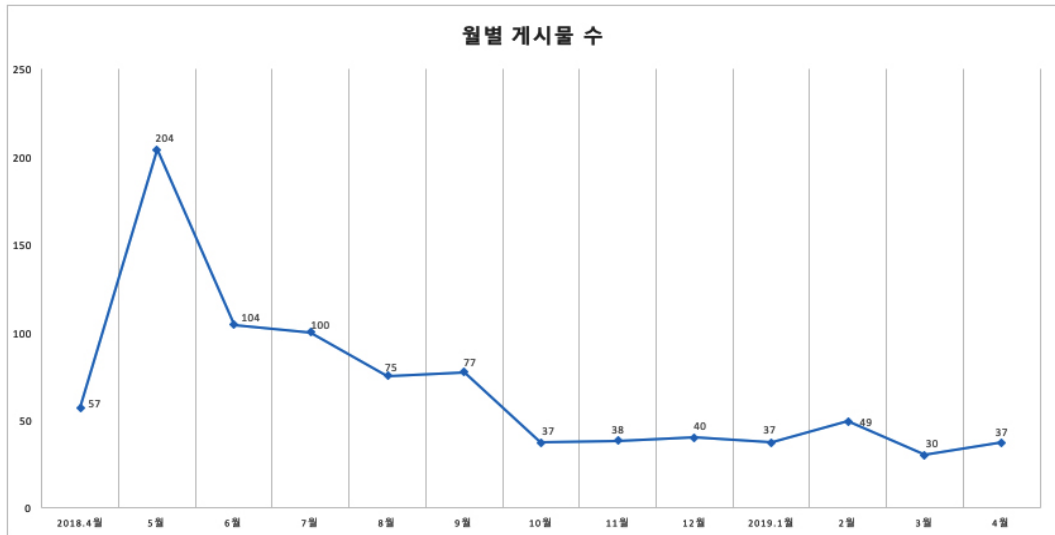
4. 연구 결과

4.1 게시물 분포

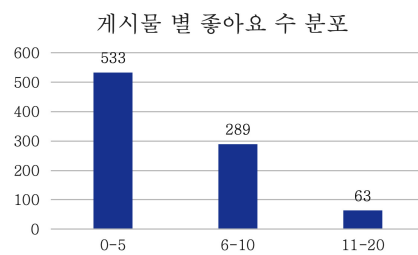
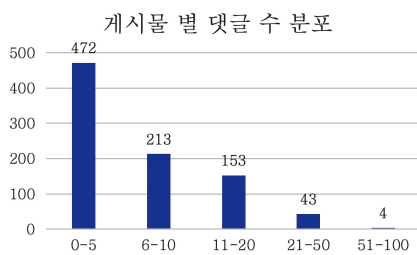
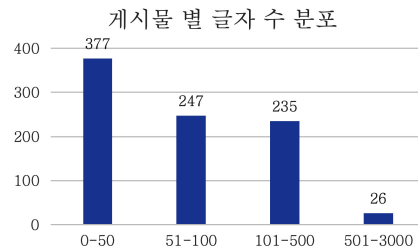
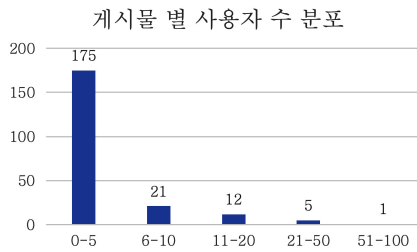
그룹 내의 게시물의 월 별 분포를 보면, 특정 달의 게시물 수가 현저히 많은 것을 볼 수 있다. 〈그림 3〉에서 제시된 바와 같이, 5월, 6월, 7월의 월별 게시물 수는 100개가 넘는 반면 8월, 9월의 월별 게시물 수는 80개를 넘지 않으며, 나머지 7개월의 월별 게시물 수는 60개 미만이다. 5월, 6월, 7월이 늦봄부터 초여름 사이라는 것을 고려할 때 계절의 변화가 우울증에 영향을 미칠 수도 있다고 추측할 수 있다.

〈그림 4〉는 사용자 별 게시물 수, 게시물 별 글자 수, 게시물 별 댓글 수, 게시물 별 좋아요 수의 분포를 보여준다. 총 214명의 사용자 중 대부분의 사용자(175명, 약 82%)는 6개 미만의 게시물을 작성하였다. 사용자들 가운데 20개 이상의 많은 게시물을 작성한 사용자는 6명에 불과하다. 따라서 본 연구는 많은 게시물을

작성한 소수의 사용자의 글을 대상으로 분석하기보다는 그러한 사용자를 포함한 다양한 사용자들의 게시물을 폭넓게 분석하여 데이터의 다양성을 보장하고자 하였다. 게시물 별 글자 수를 보면, 총 885개의 게시물 중 377개의 글자 수는 50자 미만이다. 글자 수가 500자 이상의 게시물은 26개이고 약 절반 가량의 게시물의 글자 수는 51자에서 500자 사이다. 게시물 별 글자 수로 미루어 볼 때, 많은 사용자들이 게시물 내용을 비교적 자세하게 작성한다는 것을 알 수 있으며, 이는 페이스북 그룹에 가입하는 사람들이 대체로 해당 그룹에 관심이 있는 사용자들이기 때문이다. 사용자들의 게시물에 대한 공감 정도는 게시물에 대한 댓글과 좋아요 개수로 일부 파악할 수 있다. 절반 이상의 게시물의 댓글과 좋아요 개수가 5개 이하지만 많은 댓글과 좋아요를 기록한 게시물의 양도 적지 않은 것을 발견할 수 있다. 이를 통해 그룹의 사용자들은 혼자 고립되기보다는 어느 정도 소통하고 있다는 것을 알 수 있다.



〈그림 3〉 월별 게시물 수



〈그림 4〉 사용자 및 게시물 분포

4.2 빈도 기반 키워드 추출

〈표 2〉는 전체 게시물에서 가장 많이 언급된

20개의 키워드를 보여준다. 빈도는 해당 키워드가 전체 게시물에서 언급된 개수이다. 불용어는 표에서 제외되었다.

〈표 2〉 게시물에서 가장 많이 언급된 20개 키워드

순위	키워드	빈도	순위	키워드	빈도
1	사람	261	11	자해	58
2	생각	180	12	기분	53
3	오늘	149	13	이제	52
4	우울증	132	14	요즘	52
5	병원	118	15	하루	52
6	마음	70	16	자살	51
7	친구	69	17	자신	45
8	시간	62	18	상처	44
9	혼자	61	19	이유	41
10	다시	58	20	자기	37

〈표 2〉에서 제시된 바와 같이, 우울증이라는 키워드 외에 사람 또는 자신을 나타내는 키워드들(사람, 친구, 혼자, 자신, 자기), 심리 또는 감정을 나타내는 키워드들(생각, 마음, 기분, 이유), 시간을 나타내는 단어들(오늘, 시간, 요즘, 하루), 그리고 아픔 및 자해를 나타내는 키워드들(자해, 자살, 상처)이 다양하게 분포되어 있다. 가장 많이 언급된 키워드들은 사용자들의 주된 관심사를 표현한다.

4.3 토픽 모델링 기반 주제 추출

빈도 기반의 키워드 추출이 단어의 단순 출현빈도를 분석한 관심사에 대한 추출인 점과 비교할 때, 토픽 모델링 기반의 주제 추출은 게시물에 나타난 전반적인 키워드들 간의 관계 내지는 맥락을 고려하여 관련성이 높은 것으로 나타나는 주제들을 추출하는 것이다. 주제 추출 과정은 다음과 같은 순서로 진행 된다. 우선, 주제 개수를 정하고, 다음에는 주제별 주제어를 추출한 후, 마지막으로 주제별 주제어를 고려하여 어떤 주제에 관해 논의하고 있는지에

대해서 레이블링을 한다. 본 연구에서는 보다 정확한 주제어의 개수를 정하기 위하여 토픽 개수를 미리 정하지 않고, 자동화 기법을 통해 나타나는 주제 개수 별 분석결과를 확인한 후 적절한 주제 개수를 정하고 주제어를 추출하였다. 추출한 여러 개의 주제 개수에 대해 연구자 세 명이 각각 독자적으로 레이블링을 진행하였으며, 레이블링 결과에 대해 충분한 토론을 거쳐 합의를 이룬 후, 최종 주제어 개수 및 주제 추출을 진행하였다.

본 연구를 위해 주제 개수를 3개에서 10개까지 설정하고 3명의 저자가 레이블링을 진행하여 레이블링 결과에 대해 토론을 한 결과, 주제 개수가 3개일 때 완전한 일치를 보았고, 5개일 때는 최소 2명이 레이블링의 결과의 일치를 보였다. 따라서 본 연구에서는 주제 개수가 3개와 5개일 때 전체 게시물의 주제를 비교적 정확하게 추출한다고 가정하고 데이터 분석을 실시하였다. 〈표 3〉은 주제 개수가 3개일 때의 확률 기반 최상위 20개의 주제어와 그에 따른 레이블을 보여준다.

〈표 3〉의 주제 1의 레이블은 ‘친구’, ‘남’, ‘엄

〈표 3〉 주제, 레이블, 및 주제어(3개 주제)

주제 번호	주제 1	주제 2	주제 3
레이블	인간관계	우울증 일반	기분 및 느낌
주제어 1	병원	사람	마음
주제어 2	생각	우울증	혼자
주제어 3	오늘	시간	하루
주제어 4	친구	이유	집
주제어 5	이제	다시	자해
주제어 6	상처	기분	술
주제어 7	같이	자살	요즘
주제어 8	힘드네요	자신	우울
주제어 9	남	다른	잠
주제어 10	엄마	소리	머리
주제어 11	자기	세상	우리
주제어 12	여자	돈	사랑
주제어 13	아침	감정	모든
주제어 14	클럽	힘	정신과
주제어 15	인생	보고	한번
주제어 16	삶	먹고	병
주제어 17	아이	상담	느낌
주제어 18	모두	노력	남편
주제어 19	방법	무기	언제
주제어 20	부모님	위로	힘들어요

마, '여자', '아이', 그리고 '부모님' 등으로 대표되는 '인간관계'로 정했으며, 주제 2의 레이블은 '우울증', '자살', '세상', '감정', 그리고 '위로' 등이 우울증에 관련된 일반 단어들이므로 '우울증 일반'으로 정했고, 주제 3은 '마음', '우울', '사랑', '병', '느낌', 그리고 '힘들어'로 대표되는 '기분 및 느낌'으로 세 연구자가 동일하게 주제의 명칭을 정하였다. 이를 통해 페이스북 그룹 사용자들이 전반적으로 우울증 일반에 관한 주제를 비롯하여 인간관계와 우울증의 관계 그리고 우울증과 관련된 기분이거나 느낌에 대해 논의하고 있다는 것을 알 수 있다. 〈표 4〉는 주제 개수가 5개일 때의 주제, 레이블 및 주제어를 나타낸다.

주제 개수가 5개일 때 각 레이블링 결과에 대해 최소 2명의 연구자가 일치하였다. 주제 1의 레이블은 우울증에 관련된 일반 단어들로 '우울증 일반', 주제 2의 레이블은 '힘들어', '우울하고', '힘들어요', '힘들고', '이상', '무기(력)', '자다' 등으로 대표되는 '우울증 증상', 주제 3은 '자살', '시도' 등으로 대표되는 '자살', 주제 4는 '병원', '입원', '변호사' 등으로 대표되는 의료 참고, 주제 5는 '우리', '엄마', '아이', '남편' 등으로 대표되는 '가족'으로 각각 레이블링 하였다. 〈표 3〉과 비교할 때 주제의 개수가 증가함에 따라 더 상세한 주제를 얻을 수 있음을 알 수 있다.

〈표 4〉 주제, 레이블, 및 주제어(5개 주제)

주제 번호	주제 1	주제 2	주제 3	주제 4	주제 5
레이블	우울증 일반	우울증 증상	자살	의료 참고	가족
주제어 1	우울증	마음	사람	병원	시간
주제어 2	혼자	자해	오늘	다시	우리
주제어 3	자신	우울	생각	상처	이제
주제어 4	남	힘들어	친구	술	하루
주제어 5	감정	힘드네요	요즘	사랑	삶
주제어 6	부모님	세상	자살	아침	보고
주제어 7	잠	우울하고	기분	내일	엄마
주제어 8	힘	몸	집	입원	모든
주제어 9	아니	힘들어요	이유	고민	먹고
주제어 10	상담	돈	자기	저녁	병
주제어 11	같이서	어제	같이	어떤	아이
주제어 12	모두	힘들고	연락	눈	한번
주제어 13	손	낚시	다른	담배	좋은
주제어 14	먹으면	이상	시도	클럽	살기
주제어 15	넌	먹었는데	여자	우울하다	하루하루
주제어 16	일자리	남자	머리	떠진다	많은
주제어 17	사는게	위로	인생	슬퍼요	스트레스
주제어 18	아니고	무기	나이	식욕	소리
주제어 19	가족	버스	정신	변호사	느낌
주제어 20	그룹	자다	가장	관련	남편

5. 논의 및 결론

본 연구에서는 소셜 네트워크 서비스인 페이스북의 우울증 관련 그룹 게시물을 분석하여 우울증 관련 주제를 분석하였다. 정신질환의 하나로서 다소 민감한 주제인 우울증에 대하여 사용자가 어떤 내용을 논의하는지 살펴보았다. 본 연구를 위해 페이스북 데이터 수집부터 주제어 추출에 이르기까지의 전반 과정을 포함하는 자연어 처리 기반의 데이터 분석 프레임워크를 구현하였다. 구현한 프레임워크를 이용하여 페이스북의 우울증 관련 최대 그룹(사용자 기준)에서 최근 1년간 작성한 885개의 게시

물을 수집하여 분석을 실시하였다. 분석 결과 사용자들이 우울증 일반, 인간관계, 기분 및 느낌, 우울증 증상, 자살, 의료 참고, 그리고 가족 등에 대해 논의하고 있는 것을 발견하였다. 전반적으로 페이스북 우울증 관련 그룹에서는 사용자들이 우울증에 대하여 다양한 각도에서 논의하고 있다는 것을 알 수 있다. 페이스북은 개방된 플랫폼이라는 특성으로 인해 사용자의 접근이 용이할 뿐 아니라 익명성이 보장된다는 장점으로 인해 우울증과 같이 사회적 낙인의 영향을 받는 개인적인 정신건강문제를 논의하는 것으로 파악된다.

또한 본 연구에서 발견한 주제어들은 인간관

계로 인한 갈등이나 삶의 어려움으로 인한 부정적인 감정, 그리고 자신의 우울증 문제와 관련된 고민 등을 다른 사람과 공유 내지는 상담함으로써 해결하고자 하는 노력을 반영한다고 해석할 수 있다. 따라서 본 연구는 사용자들이 쉽게 접근해서 자신의 문제를 논의하기 쉬운

페이스북과 같은 소셜 미디어 플랫폼을 통해 우울증치료 전문가들이 온라인 상담을 하거나 심리치료서비스를 제공하는 것이 이러한 정신 질환의 치료 뿐 아니라 예방에 효과적일 수 있다는 실질적인 함의를 시사한다.

참 고 문 헌

- [1] 고광운 외. 2016. Social Attributes 를 이용한 트위터 사용자의 감정 상태 예측. 『EXTENDED ABSTRACTS OF HCI KOREA 2016 학술대회 발표 초록집』, 2016년 1월 27일, 정선: 하이원리조트: 639-641.
- [2] 김수정. 2012. 소셜 미디어 환경에서 대학생들의 건강정보 요구와 추구행태에 관한 탐험적 연구. 『한국비블리아학회지』, 23(4): 239-260.
- [3] 매일경제. 2019. '우울증' 60대 가장, 환각상태서 아내와 딸 흥기로 잔혹 살해. [online] [cited 2019. 7. 23.] <<https://www.mk.co.kr/news/economy/view/2019/07/505355/>>
- [4] 목양숙. 2015. 페이스북 콘텐츠 감성이 사용자의 감정 변화 반응에 미치는 영향. 『정보디자인학연구』, 24: 91-99.
- [5] 박은정, 조성준. 2014. KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지. 『제26회 한글 및 한국어 정보처리 학술대회 논문집』, 2014년 10월 10일, 춘천: 강원대학교 춘천캠퍼스: 133-136.
- [6] 이재범 외. 2012. 블로그-트위터 매체 간 특성 차이 및 사용자 제품정보 처리와 평가차이 비교에 관한 연구. 『정보시스템연구』, 21(1): 69-91.
- [7] 차지영, 차미영. 2012. 트위터에서 우울의 담론. 『2012년 한국간호과학회 추계학술대회 논문집』, 2012년 10월 26일, 서울: 한국과학기술회관: 276-276.
- [8] 황유선. 2016. 페이스북 일상담화의 감정 탐색. 『한국콘텐츠학회논문지』, 16(1): 1-13.
- [9] Aldarwish, M. M. and Ahmad, H. F. 2017. "Predicting Depression Levels Using Social Media Posts." In *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*, 22-24 March, 2017: Bangkok: 277-280.
- [10] Bae, B. J. and Yi, Y. Y. 2017. "What Answers Do Questioners Want on Social Q&A? User Preferences of Answers about STDs." *Internet Research*, 27(5): 1104-1121.
- [11] Bae, B. J. and Yi, Y. Y. 2019. "Identification and Comparison of the Persuasive Elements

- Present in “Best Answers” to STD-Related Questions on Social Q&A Sites: Yahoo! Answers (United States) Versus Knowledge-iN (South Korea).” *International Journal of Communication*, 13: 2516-2534.
- [12] Baker, D. A. and Algorta, G. P. 2016. “The Relationship Between Online Social Networking and Depression: A Systematic Review of Quantitative Studies.” *CyberPsychology, Behavior & Social Networking*, 19(11): 638-648.
- [13] Bar, K. J. et al. 2004. “The Influence of Major Depression and Its Treatment on Heart Rate Variability and Pupillary Light Reflex Parameters.” *Journal of Affective Disorders*, 82(2): 245-252.
- [14] Bazarova, N. N. et al. 2017. “Psychological Distress and Emotional Expression on Facebook.” *CyberPsychology, Behavior & Social Networking*, 20(3): 157-163.
- [15] Boden, J. M. and Fergusson, D. M. “Alcohol and Depression.” *Addiction*, 106(5): 906-914.
- [16] Blei, D. M., Ng, A. Y. and Jordan, M. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research*, 3(4/5): 993-1022.
- [17] Cavazos-Rehg, P. A. et al. 2016. “A Content Analysis of Depression-related Tweets.” *Computers in human behavior*, 54, 351-357.
- [18] Choudhury, M., Counts, S. and Horvitz, E. 2013. “Social Media as a Measurement Tool of Depression in Populations.” In *Proceedings of the 5th Annual ACM Web Science Conference*, 2-4 May, 2013, Paris: 47-56.
- [19] Chow, T. S. and Wan, H. Y. 2017. “Is there any ‘Facebook Depression’? Exploring the moderating roles of neuroticism, Facebook social comparison and envy.” *Personality and Individual Differences*, 119: 277-282.
- [20] Chua, A. Y. and Banerjee, S. 2015. “Measuring the Effectiveness of Answers in Yahoo! Answers.” *Online Information Review*, 39(1): 104-118.
- [21] Fong, P. et al. 2014. “Quality of Online Information About Sexually Transmitted Diseases: Which Websites Should Patients Read?” *Online Information Review*, 38(5): 650-660.
- [22] Guntuku, S. C. et al. 2017. “Detecting Depression and Mental Illness on Social Media: An Integrative Review.” *Current Opinion in Behavioral Sciences*, 18, 43-49.
- [23] Ha, J. H., Aikat, D. D. and Jung, E. H. 2015. “Theories and Messages in South Korean Antismoking Advertising.” *Health Communication*, 30(10): 1022-1031.
- [24] Irwin, M. et al. 2003. “Nocturnal Catecholamines and Immune Function in Insomniacs, Depressed Patients, and Control Subjects.” *Brain Behavior and Immunity*, 17(5): 365-372.
- [25] Jin, J. et al. 2016. “How Users Adopt Healthcare Information: An Empirical Study of an

- Online Q&A Community.” *International Journal of Medical Informatics*, 86: 91-103.
- [26] Katchapakirin K. et al. 2018. “Facebook Social Media for Depression Detection in the Thai community.” In Proceedings of 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), 11-13 July, 2018, Nakhonpathom.
- [27] Katon, W. J. et al. 2005. “The Association of Comorbid Depression with Mortality in Patients with Type 2 Diabetes.” *Diabetes Care*, 28(11): 2668-2672.
- [28] Kneidinger, B. 2010. Facebook und Co. Eine soziologische Analyse von Interaktionsformen in Online Social Networks. Wiesbaden, Germany: Springer. Quoted in Scherr, S. and Brunet, A. 2017. “Differential Influences of Depression and Personality Traits on the Use of Facebook.” *Social Media and Society*, 3(1): 1-14.
- [29] Knol, M. J. et al. 2006. “Depression as a Risk Factor for the Onset of Type 2 Diabetes Mellitus. A Meta-analysis.” *Diabetologia*, 49(5): 837-845.
- [30] Lachmar, E. M. et al. 2017. “# MyDepressionLooksLike: Examining Public Discourse About Depression on Twitter.” *JMIR Mental Health*, 4(4): 1-11.
- [31] Lichtman, J. H. et al. 2008. “Depression and Coronary Heart Disease: Recommendations for Screening, Referral, and Treatment.” *Circulation*, 118(17): 1768-1775.
- [32] Meng, L. et al. 2012. “Depression Increases the Risk of Hypertension Incidence: A Meta-analysis of Prospective Cohort Studies.” *Journal of Hypertension*, 30(5): 842-851.
- [33] Moreno, M. et al. 2011. “Feeling Bad on Facebook: Depression Disclosures by College Students on a Social Networking Site.” *Depression & Anxiety*, 28(6): 447-455.
- [34] Nadeem, M. 2016. *Identifying Depression on Twitter*. [online] [cited 2019. 7. 7.] <<https://arxiv.org/abs/1607.07384>>
- [35] Reavley, N. J. and Pilkington, P. D. 2014. “Use of Twitter to Monitor Attitudes Toward Depression and Schizophrenia: An Exploratory Study.” *PeerJ*, 2: 1-15.
- [36] Řehůřek, R. and Sojka, P. 2010. “Software Framework for Topic Modelling with Large Corpora.” In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, May 22, 2010, Valletta: Mediterranean Conference Centre: 46-50.
- [37] Resnik, P. et al. 2015. “Beyond LDA: Exploring Supervised Topic Modeling for Depression-related Language in Twitter.” In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology*, June 5, 2015, Denver: 99-107.
- [38] Scherr, S. and Brunet, A. 2017. “Differential Influences of Depression and Personality Traits on the Use of Facebook.” *Social Media and Society*, 3(1): 1-14.
- [39] Scherr, S., Toma, C. L. and Schuster, B. 2018. “Depression as a Predictor of Facebook

- Surveillance and Envy: Longitudinal Evidence from a Cross-lagged Panel Study in Germany.” *Journal of Media Psychology: Theories, Methods, and Applications*, [online] [cited 2019. 7. 5.] <<https://psycnet.apa.org/record/2018-50882-001>>
- [40] Shar, C., Kitzie, V. and Choi, E. 2014. “Modalities, Motivations, and Materials - Investigating Traditional and Social Online Q&A Services.” *Journal of Information Science*, 40(5): 669-687.
- [41] Song, H. et al. 2016. “Trusting Social Media as a Source of Health Information: Online Surveys Comparing the United States, Korea, and Hong Kong.” *Journal of Medical Internet Research*, 18(3): e25.
- [42] Yi, Y. Y. 2018. “Sexual Health Information-seeking Behavior on a Social Media Site: Predictors of Best Answer Selection.” *Online Information Review*, 42(6): 880-897.
- [43] Yoon, S. et al. 2019. “Is Social Network Site Usage Related to Depression? A Meta-analysis of Facebook-depression Relations.” *Journal of Affective Disorders*, 248(1): 65-72.
- [44] Zhang, Y. 2013. “Toward a Layered Model of Context for Health Information Searching: An Analysis of Consumer-generated Questions.” *Journal of the American Society for Information Science and Technology*, 64(6): 1158-1172.
- [45] Zhao, Y. and Zhang, J. 2017. “Consumer Health Information Seeking in Social Media: A Literature Review.” *Health Information & Libraries Journal*, 34(4): 268-283.
- [46] Zhu, Y. et al. 2018. “Understanding the Research Landscape of Major Depressive Disorder via Literature Mining: An Entity-level Analysis of PubMed Data from 1948-2017.” *JAMIA Open*, 1(1), 115-121.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Go, Gwangwoon et al. 2016. “Mood State Prediction of Twitter Users from Social Attributes.” *Extended Abstracts of HCI Korea 2016*, January 27, 2016, Jeongseon: High 1 Resort: 639-641.
- [2] Kim, Soojung. 2012. “An Exploratory Study of Undergraduate Students’ Health Information Needs and Seeking Behaviors in Social Media.” *Journal of the Korean Biblia Society for Library and Information Science*, 23(4): 239-260.
- [3] Maeil Business News Korea. 2019. ‘Depression’ in his 60’s, brutally murdered with his wife and daughter in hallucinations. [online] [cited 2019. 7. 23.] <<https://www.mk.co.kr/news/economy/view/2019/07/505355/>>
- [4] Mock, Yang Suk. 2015. “A Study on the Effect of Facebook Contents Emotion on Users’

- Emotional Change Reaction.” *Journal of Korean Society of Communication Design*, 24: 91-99.
- [5] Park, Eunjeong L. and Cho, Sungzoon. 2014. “KoNLPy: Korean Natural Language Processing in Python.” In *Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology*, October 10, 2014, Chuncheon: Kangwon National University Chuncheon Campus: 133-136.
- [6] Lee, Jae-Beom et al. 2012. “A Comparative Study on Different Characteristics of Social Media and Product Information Processing and Evaluation.” *Journal of Information System*, 21(1): 69-91.
- [7] Cha, Chiyong and Cha, Meeyoung. 2012. “Discourse of Depression on Twitter.” In *Proceedings of 2012 Korean Society of Nursing Science*, October 26, 2012, Seoul: Korean Science and Technology Hall: 276-276.
- [8] Hwang, Yoosun. 2016. “Exploration of the Emotion for Daily Conversation on Facebook.” *Journal of the Korean Contents Association*, 16(1): 1-13.