

소셜 Q&A 사이트의 질문 분석을 통한 청소년의 성 정체성(sexual identity) 고민에 대한 이해*

Understanding Sexual Identity-related Concerns through the Analysis of Questions on a Social Q&A Site

주 영 준 (Yongjun Zhu)**, 남 서 진 (Seojin Nam)***
이 다 정 (Dajeong Yi)****, 이 용 정 (Yong Jeong Yi)*****

< 목 차 >

I. 서론	IV. 연구 결과
II. 선행연구	V. 결론 및 논의
III. 연구 방법	

요약: 본 연구는 소셜 Q&A 사이트인 네이버 지식iN에서 성 정체성과 관련된 질문들을 분석하여 질문의 중심이 되는 주제를 추출하고 구체적인 고민들에 대해 이해하는 것이 목적이다. 이를 위해 네이버 지식iN에서 2010년부터 9년간 작성된 2,120개의 관련 질문들을 수집한 후, 자연어 처리와 정보 검색 방법을 사용하여 질문들의 주제와 고민들을 추출하고 분석하였다. 분석 결과, 주요 주제는 인간관계, 성 정체성 의심, 성적 지향, 감정 및 관계, 성 정체성 고민 등으로 나타났다. 나아가 사용자들은 주로 1) 일반적인 성 정체성, 2) 성적 지향, 3) 성 정체성에 대한 부정적 인식, 4) 감정 고백, 커밍아웃, 동성애, 5) 미래, 이성 관계, 군대에 대한 우려, 6) 성 정체성 혼란을 겪는 이유 등과 관련된 고민을 하고 있었다. 본 연구는 성 정체성에 대한 실제적 고민들을 수집하여 토픽 모델링 기법을 통해 정보 요구를 도출하였으며 exact match와 tf-idf 기반의 검색 방법을 비교함으로써 문헌정보학 분야의 방법론적 확장을 시사했다. 또한 특정 관심사에 대한 온라인 커뮤니티 이용자들의 정보요구 및 정보탐색행태를 관찰함으로써 정보행위분야의 학문적 성숙에도 기여했다.

주제어: 소셜 미디어, 소셜 Q&A, 성 정체성, 자연어 처리, 정보 검색, 토픽 모델링

ABSTRACT: The study aims to understand major topics and concerns of gender identity-related questions expressed by the users of the NAVER social Q&A site. To achieve this goal, we analyzed 2,120 questions created from 2010 to 2018 using natural language- and information retrieval-based methods. Results indicated that the major topics discussed by the users include interpersonal relationships, doubts about gender identity, sexual orientation, feelings and relationships, and concerns about gender identity. In addition, users mainly expressed concerns regarding general issues of gender identity: sexual orientation; negative cognition about gender identity; confession, coming-out, homosexuality; future, heterosexual relationships, military enlistment; and causes of gender identity confusion. The present study effectively derives information needs from real-world concerns about sexual identity by employing topic modeling techniques, and by comparing the advantages of exact match and tf-idf-based information retrieval methods extends methodology of Library and Information Science. Further, it has contributed to the academic maturity of the study of information behavior by observing the information needs or information-seeking behaviors of online community users with specific interests.

KEYWORDS: Social Media, Social Q&A, Gender Identity, Sexual Identity, Natural Language Processing, Information Retrieval, Topic Modeling

* 이 논문은 성균관대학교-삼성서울병원 미래융합연구 사업에 의하여 연구되었음. This study was supported by SKKU-SMC Future Convergence Research Program Grant.

이 논문 또는 저서는 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5A2A03045289).

** 성균관대학교 문헌정보학과 조교수(yzhu@skku.edu / ISNI 0000 0004 7933 6219) (제1저자)

*** 성균관대학교 문헌정보학과 석사과정(namsj0327@skku.edu) (공동저자)

**** 성균관대학교 문헌정보학과 석사과정(ekwjd3024@skku.edu / ISNI 0000 0004 9295 9250) (공동저자)

***** 성균관대학교 문헌정보학과 조교수(redpapa01@skku.edu / ISNI 0000 0004 7411 8220) (교신저자)

• 논문접수: 2020년 11월 16일 • 최종심사: 2020년 11월 26일 • 게재확정: 2020년 12월 8일

• 한국도서관·정보학회지, 51(4), 101-119, 2020. <http://dx.doi.org/10.16981/kliss.51.4.202012.101>

I. 서론

소셜 미디어의 발달은 건강정보의 생산자와 소비자의 경계를 허물었고, 익명성과 신속성이라는 혜택을 통해 누구나 편리하게 건강정보를 탐색하고 정보를 공유하게 되었으며, 다양한 성격의 플랫폼들이 발달했다. 이러한 소셜 미디어들은 그들이 지닌 특성에 따라 그 사용목적이 다르게 나타날 뿐 아니라 거기에서 다루는 주제들 또한 차이가 있다(이재범 외 2012). 사용목적 측면에서 볼 때, 카카오톡이나 페이스북으로 대표되는 소셜 네트워크 서비스는 관계형성이나 엔터테인먼트가 주요 목적인 반면 트위터와 같은 마이크로 블로그나 개인블로그 등은 정보공유의 목적이 강하다. 또한, 주제적인 측면에서 볼 때, 소셜 네트워크 서비스는 신변잡기적인 정보가 주요 콘텐츠인 한편 블로그 등은 특정 주제에 대한 주관적 생각이나 평가가 주를 이룬다. 따라서 대통령선거나 세월호 사건과 같은 커다란 사회적 이슈가 되는 주제들의 논평은 주로 트위터 등에서 이루어지는 반면 친구나 지인들과의 의사소통이나 일상생활에 대한 공유는 카카오톡이나 페이스북에서 활발하게 이루어진다(이서영, 권상집 2019; 유호선, 김현지, 오효정 2018; 동탁유, 유승관 2016; 손영화 2018).

그러나 이러한 일반적인 주제에 대한 논의와는 달리, 개인적으로 민감하고 심각한 건강 관련 문제들은 또 다른 특성, 즉 익명성과 신속성 외에도 협업에 의한 정보나 지식의 공유를 제공하는 소셜 Question & Answer(Q&A) 사이트에서 주로 논의된다(Bae and Yi 2019). 이렇게 여러 사람들이 정보나 지식을 제공하는 집단지성(Wisdom of crowds)의 정보탐색을 사회적 탐색(Social search)이라 하며 소셜 Q&A 사이트는 이러한 사회적 탐색이 활발하게 이루어지면서 전통적인 정보탐색의 틀을 확장시켰다. 즉, 질문자로서 정보 탐색자가 여러 사람들이 제공하는 답변 중에서 자신의 질문에 가장 적합한 정보를 제공하는 답변을 채택할 수 있다. 따라서 소셜 Q&A 사이트는 이용자의 정보요구에 적절한 맞춤형정보를 찾을 수 있는 기회의 장(Opportunistic places)이라 할 수 있다(Yi 2018). 이용자들은 이렇게 사회적 탐색을 통해 집단지성이 제공하는 정보를 공유하면서 인지적 만족 뿐 아니라 정서적·사회적 지원을 받으면서 자신의 건강에 관한 의사결정을 하는 것으로 나타났다(민희 2018; Bae and Yi 2019).

우리나라의 네이버 지식인은 가장 성공적인 소셜 Q&A 사이트 중의 하나로서 많은 연구들이 지식인에서 이루어지는 이용자의 정보탐색행태를 관찰하였고 타 문화와의 비교 연구를 위해 미국의 대표적인 소셜 Q&A 사이트인 Yahoo! Answers와도 자주 비교되어 논의되었다(Bae and Yi 2017; 2019; Liu and Jansen 2018). 다른 나라에서 운영되는 소셜 Q&A 사이트와 비교할 때, 네이버 지식인은 독특한 특성을 지닌다. 일반적으로 이러한 사이트들은 의학적 배경이 없는 일반인들이 건강정보를 서로 제공하여 공유하거나 전통의학이 답변을 제시하지 못하는 특수한 경우에 대해서도 개인들이 자신의 유사경험 바탕으로 답변을 제시하여 매우 긴요한 도움을 제공한다.

그러나 네이버 지식인은 일반 정보소비자뿐 아니라 의료 전문가가 자신이 속한 기관을 홍보할 목적으로 활발하게 활동하기 때문에 일반인이 제공하지 못하는 전문적 답변을 자주 제공하는 것을 발견할 수 있다.

건강정보 중에서도 성 관련 주제는 매우 개인적이며 민감한 정보이며, 우리나라는 그러한 주제를 타인과 자유롭게 논의하는 문화를 형성하고 있지 않다(Yi 2018). 특히 성 관련 주제 가운데서도 성 정체성에 관한 문제는 우리나라를 비롯한 여러 나라에서 제도적으로나 문화적으로 지속적인 논쟁이 이루어지고 있는데, 이 문제에 관해 보수적 성향이 강한 우리나라에서는 동성애와 같은 성적 성향을 커밍아웃 할 경우 사회적 오명의 위험을 부담해야 하며, 성소수자들의 경우 자신의 성 정체성과 관련된 혼란으로 인해 자살을 고민할 정도로 심각하며 사회적 합의가 제대로 이루어지지 않았다.

성 정체성에 관한 인지나 문제의식은 호르몬의 영향으로 인해 청소년기에 가장 많이 나타난다(이영선, 권보민, 이동훈 2012). 그러나 청소년을 비롯한 유사한 문제로 어려움을 겪는 성소수자들은 이러한 문제에 대한 수많은 질문들은 오프라인에서 자유롭게 표현하기 어려울 뿐 아니라 오히려 숨기게 되고 그러한 현상은 적절한 정보탐색을 방해하여 정보 빈곤을 초래하게 된다(Chatman 1996). 이러한 실정을 고려할 때, 소셜 Q&A 사이트는 성 정체성 문제를 논의하기에 가장 알맞은 장이라 할 수 있다(Yu 2019). 이는 이용자들이 자신의 익명성이 보장되는 혜택으로 인해 성 정체성과 같은 사회적 오명을 수반하는 주제에 대해 자유롭게 논의할 수 있기 때문이다.

이용자들이 소셜 미디어를 사용하면서 생산된 빅데이터를 분석하는 기법을 통해 그들이 주로 논의하는 특정 주제에 관해 이해하고자 하는 연구가 활발하게 진행되고 있다. 성 정체성과 관련된 연구는 지속되어 왔으나, 소셜 Q&A 사이트와 같은 소셜 미디어 플랫폼에서 사람들이 그 주제에 관해 무엇을 논의하고 있는지, 즉 그들이 어떠한 생각이나 감정들을 토로하는지를 살펴본 연구는 아직 부족하다. 따라서 본 연구는 소셜 Q&A 사이트에서 청소년들이 성 정체성과 관련된 어떠한 고민들을 논의하는 지를 살펴봄으로써 그들의 정보 요구와 정보 추구 행위를 파악하고자 한다. 이를 위한 구체적인 연구 문제는 다음과 같다.

- 소셜 Q&A 사이트에 나타난 청소년들의 성 정체성에 관한 정보 요구는 무엇이며, 어떠한 주제를 주로 논의하는가?

또한, 소셜 Q&A 사이트의 한글 자연어 텍스트를 분석하는 연구가 매우 초기 단계에 있다. 이용자들이 소셜 Q&A 사이트에 생성한 대용량의 텍스트를 수집하여 분석하는 연구는 전통적인 연구방법에서 나타나는 표본 수의 한계를 극복할 수 있어 연구의 신뢰성을 높이는데 기여한다(Yu 2019; Choudhury, Counts, and Horvitz 2013; Lachmar et al. 2017). 따라서 본 연구에서는

소셜 Q&A 사이트를 통해 생산되는 성 정체성에 관한 수많은 생각과 감정 등을 담은 한글 데이터를 분석함으로써 성 정체성에 대한 정보 요구에 대한 이해를 높이고 이를 통해 그러한 정보 요구를 충족시키기 위한 정보서비스 콘텐츠를 개발하는 시사점을 제고할 뿐 아니라 성 관련 건강정보 추구행태에 관한 학문적 함의를 제공하고자 한다.

II. 선행연구

1. 성 관련 건강정보 탐색행태

소셜 미디어가 제공하는 접근의 용이성과 익명성 등의 이점으로 인해 이용자들은 일반정보는 물론 건강정보를 탐색하는 데 있어서도 소셜 미디어를 자주 이용하고 있다(Song et al. 2016; Zhao and Zhang 2017). 소셜 미디어를 통해 건강정보를 탐색하는 이용자들은 자신의 정보요구를 빠르고 편리하게 만족시킬 수 있다는 긍정적인 인식을 가지고 있으면서도 건강정보를 탐색하면서 자신의 프라이버시가 노출될 수 있다는 위험과 소셜 미디어에서 제공하는 정보의 품질에 대해 회의적인 것으로 나타났다(Zhao and Zhang 2017). 특히 소셜 미디어는 이용자 간의 활발한 상호작용을 통해 효과적인 학습이 가능하다는 장점으로 인해 성 건강에 관한 커뮤니케이션을 촉진시킬 수 있는 플랫폼이다(Byron 2017).

반면에, 성 관련 정보는 낙인(stigma)이 수반되는 정보가 많기 때문에 페이스북 등의 개인정보가 노출되기 쉬운 소셜 미디어는 사생활 침해의 우려가 높아 이와 같은 건강정보를 공유하기에 적절하지 않은 것으로 나타났다(Evers et al. 2013). 즉 성병(Sexually transmitted diseases)이나 성소수자 문제를 비롯한 민감한 성 정보는 익명성이 보장된 텀블러(Tumblr)와 같은 플랫폼에서 활발하게 공유되며, 따라서 이와 같은 플랫폼은 성 정체성이나 성 교육에 관한 문제들에 대해 청소년들을 참여시키고 그들에게 신뢰할 만한 정보를 전달하는 데 효과적으로 이용될 수 있는 것으로 나타났다(Byron et al. 2019).

청소년들의 성 정보 탐색행태를 살펴본 연구들에 따르면, 온라인으로 성 정보를 탐색하거나 성적 표현을 많이 하는 청소년들의 경우 오프라인에서도 그와 유사한 활동들을 많이 하는 것으로 나타났다. 또한 온라인으로 위험한 성 관련 활동을 하는 청소년들의 경우, 부모와의 관계가 좋지 않거나, 낮은 자존감이나 자신에 대한 부정적인 이미지로 인해 인간관계에 어려움이 있는 것으로 나타났다(Livingstone and Mason 2015).

소셜 미디어를 통한 건강정보탐색에 대한 문화 간 차이를 살펴본 연구에 따르면, 한국인이나 홍콩인 같은 동양인의 경우 경험에 관한 정보를 제공하는 정보원, 즉 온라인 서포트 그룹(Online

support group), 소셜 네트워크 사이트, 그리고 블로그 등을 선호하는 반면, 미국인의 경우는 전문 지식을 제공하는 정보원, 즉 WebMD나 미국 질병관리본부 등을 선호하는 것으로 나타났다(Song et al. 2016). 또한, 집단지성이 참여하는 소셜 Q&A에서 이용자들이 민감한 건강정보, 예를 들면 성병에 대한 정보를 선택하는 데 영향을 미치는 요인들을 살펴본 연구에서도 문화 간 정보탐색행태의 차이가 발견되었다(Bae and Yi 2019). 미국의 Yahoo! Answers와 한국의 네이버 지식인에서의 성 정보 탐색행태를 비교한 결과, 네이버 지식인 이용자들은 Yahoo! Answers 이용자들에 비해 숫자로 표현한 정보, 위험에 대한 정보 내지는 낙관적인 내용의 정보를 선호하는 것으로 나타났다. 한편, Yahoo! Answers 이용자들의 경우는 더 높은 수준의 문장으로 쓰인 정보를 선호하는 것으로 나타났다. 선행연구에서 이용자들은 문장 수준이 높은 정보를 보다 과학적으로 인식하는 것으로 보고하였다(Yi, Stvilia, and Mon 2012).

선행연구를 종합해 볼 때, 소셜 미디어가 빠르게 확산됨에 따라 이용자들은 건강정보에 대한 탐색을 활발하게 할 수 있고 더 나아가 집단지성의 혜택으로 자신의 정보요구를 보다 효과적으로 충족시킬 수 있게 되었으나, 반면 전문가가 아닌 일반인들이 주로 정보를 제공하는 특성으로 인해 공유되는 건강정보의 신뢰성에 대한 우려와 함께 사생활 침해의 가능성이 지속적인 이슈로 논의되고 있음을 알 수 있다.

2. 성 정체성과 소셜 미디어

소셜 미디어가 그것이 제공하는 익명성과 접근성의 장점으로 인해 성 정체성(gender identity) 과 같이 개인적이고 민감한 주제에 관한 논의를 분석하기에 적합하다는 점은 여러 연구에서 관찰되었다(Bae and Yi 2017; 2019; Kitzie 2015; Yi 2018). 미국의 종합 성교육 가이드라인 (Sexuality Information and Education Council of the United States, 2004)은 성 관련 주제를 여섯 항목으로 분류하며, 이에는 1) 인간발달(human development), 2) 관계(relationships), 3) 개인적 스킬(personal skills), 4) 성 행태(sexual behavior), 5) 성 건강(sexual health), 그리고 6) 사회와 문화(society and culture)가 있다. 첫 번째 항목인 인간발달은 사춘기, 생식, 생식해부학과 병리학, 그리고 성 정체성 등의 네 가지 세부항목으로 구성된다. 이 가이드라인을 바탕으로 네이버 지식인에서의 전반적인 성 관련 정보요구를 분석한 연구에 따르면, 이용자들은 인간 발달 세부 항목 중 성 정체성에 관한 정보요구가 가장 큰 것으로 나타났으며(Yi 2018), 우리나라 청소년을 상대로 성 고민을 분석한 연구에서도 성 정체성이 가장 큰 비율을 차지하는 고민으로 파악되었다(유정옥 2019). 한편, Yahoo! Answers에 나타난 이용자들의 성 정체성에 대한 정보요구는 주로 다섯 가지 주제로 나타났다. 이에는 1) 성소수자(Lesbian, Gay, Bisexual, Transgender: LGBT) 정체성, 2)로맨스, 성관계, 그리고 관계성, 3)인간 소외, 4) 커밍아웃, 그리고 공동체에서

의 성소수자의 역할 등이 속하였다(Kitzie 2015).

청소년들이 소셜 미디어를 통해 성 건강 정보에 쉽게 접근할 수 있는 것은 그들에게 유익하게 작용할 수 있지만 반면 그러한 정보가 신뢰할 만한 정보가 아닐 경우, 잘못된 정보로 인해 위험에도 빠르게 노출된다는 것을 의미한다. 특히, 청소년 성소수자의 경우, 위험정보에 노출되기 쉬운 취약집단이기 때문에 신뢰할 만한 성 건강정보를 이용하느냐 아니면 잘못된 정보를 이용하느냐의 문제는 그들의 성 의식에 대한 교육뿐 아니라 그들의 건강과 생명에 크게 영향을 미칠 수 있다. 청소년 성소수자들의 사회·문화적 경험과 인식의 발달을 조사한 연구에 따르면(주재홍 2017), 그들은 사춘기에 타인과 다른 자신의 성적 지향을 깨달으면서 성 정체성에 대한 활발한 정보탐색을 하는 것으로 나타났다. 이는 유럽 청소년의 경우에서도 유사하게 나타난 행태로써 9-12세의 청소년들이 소셜 미디어와 같은 ICT 기술을 활용하여 위험성이 높은 성 정보를 탐색하기 시작하며, 청소년의 연령대가 높아질수록 온라인에서 위험성이 높은 성 관련 활동을 하는 것으로 나타났다(Livingstone and Mason 2015).

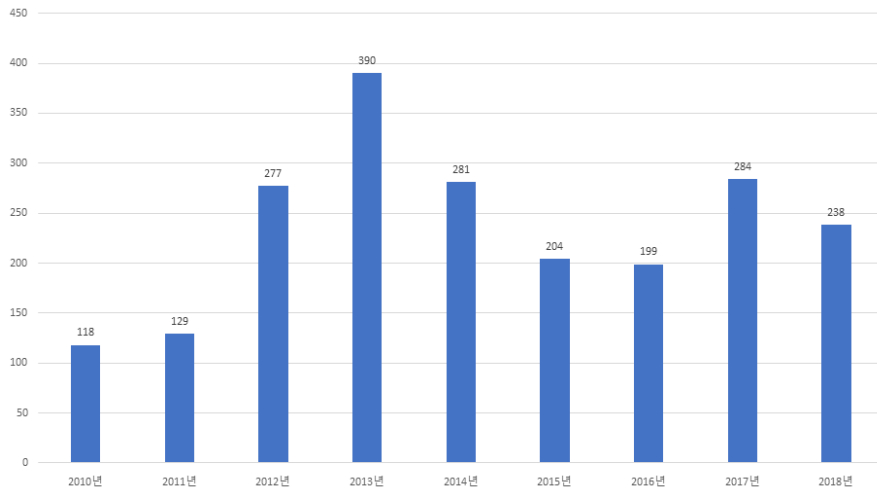
한편, 우리나라 청소년의 성의식과 동성애에 대한 태도를 사회·문화적 맥락에서 살펴본 선행 연구들은 성 정체성과 관련된 갈등 및 인권 문제들, 특히 성 정체성의 혼란으로 인한 우울증 및 자살 등과의 관계를 살펴봄으로써 우리 사회에서의 성 정체성 이슈를 이해하고자 노력했다(강병철, 하경희 2005; 김경호 2009; 김선호 2015; 신동열 2010; 이영선, 권보민, 이동훈 2012; 주재홍 2017). 성 관련 정보탐색을 통해 청소년 성소수자들은 자신의 성 정체성에 대해 긍정적으로 수용하는가 하면 자신을 사회 변화의 주체자로서 인식하기도 하였다. 그러나 대체로 우리 사회의 동성애에 대한 부정적 인식으로 인해 학교에서는 자신의 성 정체성을 숨기는 등 침묵과 금기의 문화를 경험하는가 하면 사회적 낙인이나 차별을 겪는 것으로 나타났다. 성 정체성과 관련된 문제들은 성 의식의 보수성 정도, 동성애의 경험유무, 그리고 동성애에 대한 혐오 정도는 상관관계가 있는 것으로 나타나 동성애 경험이 없는 경우 성에 대해 더 보수적이며 동성애에 대한 혐오 정도가 높았으며, 동성애 경험이 있는 경우 그렇지 않은 경우보다 우울증세가 더 높은 것으로 나타났다.

성 정체성에 관한 선행 연구들을 종합해 보면, 청소년 성소수자들은 이성애만을 정상적인 것으로 수용하고 동성애를 혐오하는 문화로 인해 정신적·육체적 폭력과 억압을 경험할 뿐 아니라 성소수자에 대한 사회적 낙인으로 많은 고통을 겪게 되면서 위험한 약물을 사용하거나 우울증 또는 자살 위험행태를 더 많이 보였다(강병철, 하경희 2007; Johns et al. 2018; Clements-Nolle et al. 2018). 따라서 많은 연구들이 청소년 성소수자들을 위한 사회적 지지와 상담이 절실하다고 시사했으며(김경호 2009; Byron et al. 2019; Livingstone and Mason 2015), 이를 위해 소셜 미디어는 성 정체성에 관한 정보 공유 및 전달을 위한 효과적인 플랫폼으로 평가되었다(Clements-Nolle et al. 2018; Livingstone and Mason 2015).

Ⅲ. 연구 방법

1. 데이터

네이버 지식iN은 네이버가 제공하는 온라인으로 질문하고 답변을 받는 서비스이다. 사용자들은 사회, 정치, 경제 등 다양한 분야와 관련하여 자유롭게 질문을 할 수 있고 일반인 또는 전문가들의 답변을 받을 수 있다. 본 연구는 2010년 1월부터 2018년 12월까지 작성된 지식iN 질문 중 “성 정체성”이라는 키워드가 들어간 네이버 질문 5,156개를 웹 스크래핑에 유용한 Python 라이브러리인 BeautifulSoup를 사용하여 수집한 후, 질문자가 청소년임을 나타내는 질문을 선별하는 과정을 거쳤다. 이러한 필터링 작업은 연구자들이 질문을 리뷰하며 질문자가 본인의 나이 또는 신분을 중학생, 고등학생 등 청소년이라고 직접적으로 드러내거나 학생, 학년, 학교, 사춘기 등 관련 단어로 설명을 한 경우 연구에 적합하다고 판단하였다. 결과적으로 전체 공개된 성 정체성 고민에 대한 질문 2,120개를 분석하였으며, 데이터 수집 시 익명성 보장을 위하여 사용자를 식별할 수 있는 정보는 제거되었다. 연도별 질문 수는 <그림 1>과 같다.



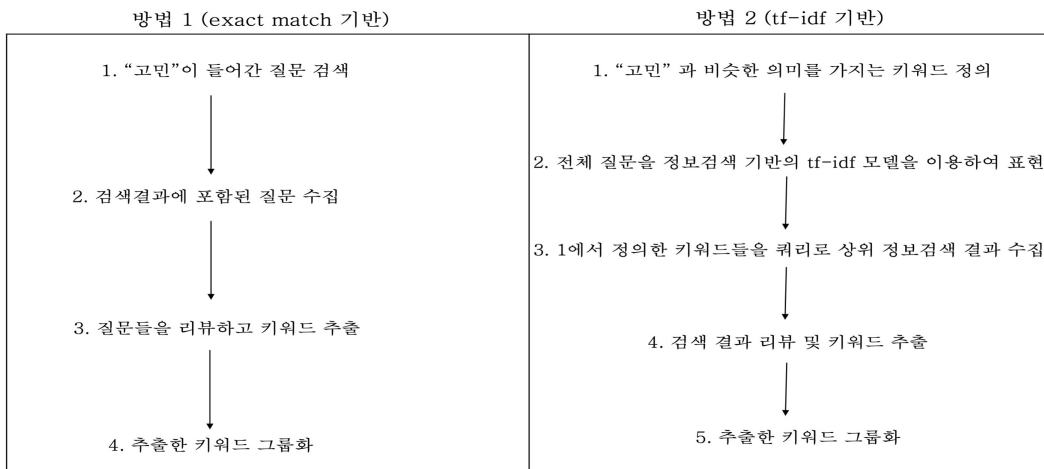
<그림 1> 연도별 성 정체성 질문 수

2. 연구 방법

본 연구는 두 가지 목적이 있다. 첫 번째로 성 정체성 관련 모든 질문들에 대한 분석을 통하여 전반적으로 어떤 주제에 대해 논의되고 있는지를 파악하고자 하였다. 토픽 모델링 기법인 Latent

Dirichlet Allocation(LDA) (Blei, Ng, and Jordan 2003)을 이용하여 질문에 존재하는 주제를 추출하고 주제를 설명할 수 있는 주제어들을 살펴보았다. 수집한 한글 데이터는 한국어 형태소 분석기인 KoNLPy(박은정, 조성준 2014)에서 지원해주는 패키지들 중 Kkma를 사용하여 전처리하였고 이후 불용어를 제거하였다. 이러한 불용어에는 무의미한 접미사나 부사, 성 정체성과 무관한 답변 채택 등에 대한 단어들을 비롯하여, 단순 자음 또는 모음의 나열, 숫자, 문장부호 등이 포함되었으며, 1음절의 단어라도 그 의미를 파악하는 것이 어렵지 않고 주제와 무관하지 않다고 판단될 경우 불용어로 처리하지 않았다. 전처리된 데이터는 gensim(Řehůřek and Sojka 2010) 라이브러리를 이용하여 토픽 모델링을 진행하였다.

본 연구의 두 번째 목적은 질문들 중 고민과 연관된 질문만을 분석하여 성 정체성에 대해 구체적으로 어떠한 고민들이 있는지 이해하는 것이다. 성 정체성 질문을 하는 가장 큰 이유 중의 하나가 고민을 해결하기 위한 것이기 때문에 고민과 연관된 키워드들을 찾아내는 것은 고민 해결을 위한 중요한 과제이다. <그림 2>는 고민과 관련된 키워드들을 추출하고 그룹화 하는 두 가지 방법을 설명하고 있다.



<그림 2> 고민 관련 키워드 추출 방법

방법 1은 exact match 기반의 정보검색 방법이다. 전체 질문에서 “고민”이라는 키워드를 포함한 질문을 검색하고 검색결과에 포함된 모든 질문을 수집한 뒤 연구자가 질문들을 차례로 리뷰하면서 키워드를 추출하고 그룹화를 진행한다. 이 방법의 장점은 “고민”과 직접적으로 연관된 질문들을 모두 추출할 수 있다는 점이다. 단점은 “고민”이라는 키워드가 포함이 되어있지 않지만 실제로 고민에 대해 다루는 질문들을 놓칠 수 있다는 점이다. 해결 방법은 “고민”과 비슷한 의미를

가지는 키워드들을 추가로 정의하고 그 키워드들을 하나라도 포함하는 질문들을 모두 수집하여 리뷰 하는 것이다. 하지만 추가 키워드들로 검색했을 때 검색 결과에 포함되는 질문의 양을 고려하면 이를 모두 리뷰 하는 것은 현실적으로 불가능하며 exact match로 검색된 결과는 연관성에 근거하여 순위를 매길 수 없다는 문제점이 있다.

그 대안으로 방법 2를 사용할 수 있다. 방법 2의 절차는 다음과 같다. 전체 질문을 문장으로 나누어 하나의 문장을 하나의 문서로 간주하고, tf-idf를 단순히 문장을 구성하는 토큰들의 tf-idf 값을 구하는 것에 그치지 않고 정보검색이라는 큰 틀에서 각 문서를 표현하는데 이용한다. 즉, 각 문장과 키워드들의 연관성을 문장 안의 토큰들의 tf-idf 값을 고려하여 구하고 그 값에 근거하여 문장들을 연관성으로 순위를 매기는 것이다. 본 연구에서는 “고민”과 비슷한 의미를 가지는 키워드들로 “혼란”, “걱정”, “궁금”을 정의하고 이 키워드들을 쿼리로 tf-idf 모델 기반 검색을 진행하였다. 검색결과 상위에 랭크 된 top n개의 문장들만 수집하여 직접 리뷰하며 키워드를 추출하고 그룹화 하였다. <표 1>은 이러한 고민 관련 키워드 추출의 두 가지 방법을 비교 정리한 것으로써 결론적으로 두 방법은 키워드와 검색결과의 범위 면에서 서로를 보완하기 때문에 본 연구에서 함께 사용되었다.

<표 1> 고민 관련 키워드 추출의 방법 비교

	exact match 기반	tf-idf 기반
사용한 키워드	고민	고민, 혼란, 걱정, 궁금
검색방법	키워드 포함 여부	키워드와의 연관성
랭킹	없음	있음
리뷰한 검색결과	전부	상위 n개

IV. 연구 결과

1. 토픽 모델링 기반 주제 추출

토픽 모델링 기법은 사전에 정의한 주제 개수에 따라 주제를 추출하고 각 주제를 나타내는 주제어를 나열한다. 각 주제는 주제어의 조합으로 설명이 되고 주제어를 전반적으로 고려하여 주제에 대해 레이블링을 진행하는 작업이 필요하다. 따라서 주제 개수 선정에 따라 주제 추출 결과가 상이하다. 주제 개수에 따라 토픽 모델링 결과가 다르기 때문에 본 연구에서는 최대한 많은 주제 개수를 고려하여 정확한 주제를 찾으려고 하였다. 주제 개수를 3에서 10까지 설정하고 3명의 저자가 각자 레이블링을 진행하였고 레이블링의 일치성을 고려하여 최종 주제 개수를 선정

하였다. 결과 주제 개수가 3과 4일 때 토론을 거쳐 일치율을 보았다. 주제 개수가 3개 일 때의 결과는 <표 2>와 같다. 주제어는 확률을 고려하여 top 20개의 주제어만 나열하였다.

<표 2> 주제, 레이블, 및 주제어 (3개 주제)

주제 번호	주제 1		주제 2		주제 3	
레이블	인간관계		성 정체성 의심		성적 지향	
주제어 1	친구	.0364	지금	.0093	여자	.0453
주제어 2	좋아하	.0280	여자애	.0092	남자	.0363
주제어 3	좋	.0184	모르	.0091	생각	.0281
주제어 4	아이	.0079	사귀	.0078	케이	.0144
주제어 5	감정	.0071	정체성	.0074	사람	.0105
주제어 6	친하	.0068	동성애자	.0067	싫	.0076
주제어 7	같이	.0066	이상	.0063	사랑	.0069
주제어 8	어떻	.0064	느끼	.0058	몸	.0053
주제어 9	마음	.0062	혼란	.0058	보다	.0049
주제어 10	다른	.0057	이쁘	.0056	언니	.0049
주제어 11	말하	.0051	생기	.0053	여성	.0042
주제어 12	계속	.0048	남	.0051	흥분	.0041
주제어 13	학교	.0046	사춘기	.0043	고민	.0040
주제어 14	만나	.0040	남자애	.0043	그때	.0034
주제어 15	못하	.0040	미치	.0041	만지	.0033
주제어 16	장난	.0038	입	.0041	보고	.0031
주제어 17	년	.0037	요즘	.0037	동성애	.0030
주제어 18	지내	.0036	눈	.0037	옷	.0029
주제어 19	놀	.0034	관심	.0036	레즈비언	.0027
주제어 20	한번	.0033	귀엽	.0035	질문	.0026

<표 2>와 같이, 주제 1의 레이블은 ‘친구’, ‘친하’, ‘같이’, ‘만나’, ‘장난’, ‘지내’ 등으로 대표되는 ‘인간관계’로 명명될 수 있다. 주제 2는 ‘모르’, ‘정체성’, ‘이상’, ‘느끼’, ‘혼란’, ‘미치’, ‘관심’ 등으로 대표되는 ‘성 정체성 의심’으로 레이블링 되었다. 특히, 주제 2의 ‘여자애’, ‘남자애’, ‘사춘기’, 그리고 ‘부모님’ 등은, 여학생이 동성인 여학생에게 또는 남학생이 동성인 남학생에게 성적으로 끌리게 되어 본인이 사춘기라서 느끼게 되는 혼란스런 감정들이 표현되었으며, 이러한 감정에 대해 부모님이 걱정하실 것을 우려하는 표현들이 많이 나타났다. 주제 3은 ‘케이’, ‘사랑’, ‘몸’, ‘동성애’, ‘레즈비언’ 등으로 대표되는 ‘성적 지향’이다. 전반적으로 인간관계에 대한 질문으로부터 시작하여 성 정체성 의심 및 성적 지향에 대한 심도 있는 질문도 포함하고 있다는 것을 알 수 있다. <표 3>은 주제 개수가 4일 때의 주제, 레이블, 및 주제어를 보여준다.

〈표 3〉 주제, 레이블, 및 주제어 (4개 주제)

주제 번호	주제 1		주제 2		주제 3		주제 4	
레이블	성적 지향		감정 및 관계		성 정체성 고민		성 정체성 의심	
주제어 1	사람	.0147	친구	.0406	여자	.0609	좋아하	.0277
주제어 2	언니	.0105	생각	.0292	남자	.0442	지금	.0108
주제어 3	느끼	.0070	좋	.0168	싫	.0098	여자애	.0076
주제어 4	동성애자	.0063	게이	.0134	남	.0080	이상	.0064
주제어 5	어떻	.0059	모르	.0086	말하	.0062	남자애	.0056
주제어 6	정체성	.0058	사랑	.0079	못하	.0057	미치	.0053
주제어 7	선배	.0056	아이	.0075	몸	.0048	학교	.0050
주제어 8	생기	.0049	사귀	.0070	고민	.0047	관심	.0048
주제어 9	보다	.0048	마음	.0069	여성	.0046	혼란	.0048
주제어 10	만나	.0046	감정	.0067	성격	.0042	답변	.0046
주제어 11	성적	.0043	친하	.0061	입	.0042	귀엽	.0042
주제어 12	동성	.0036	다른	.0059	이쁘	.0041	발기	.0040
주제어 13	요즘	.0036	같이	.0056	옷	.0040	장난	.0038
주제어 14	고백	.0033	계속	.0046	힘들	.0036	그때	.0037
주제어 15	눈	.0032	한번	.0037	태어나	.0032	만지	.0036
주제어 16	손	.0031	년	.0036	부모님	.0031	우리	.0036
주제어 17	동성애	.0031	느낌	.0035	사춘기	.0031	가슴	.0036
주제어 18	질문	.0031	학년	.0035	머리	.0031	연예인	.0034
주제어 19	이성애	.0031	양성	.0033	나이	.0031	신고	.0033
주제어 20	애들	.0024	키스	.0031	인정	.0028	가끔	.0032

〈표 3〉의 주제어를 전반적으로 고려하여 주제 1은 ‘동성애자’, ‘동성’, ‘동성애’, ‘이성애’로 대표되는 ‘성적 지향’, 주제 2는 ‘친구’, ‘모르’, ‘사귀’, ‘친하’, ‘같이’, ‘사랑’, ‘마음’, ‘감정’, ‘느낌’으로 대표되는 ‘감정 및 관계’, 주제 3은 ‘말하’, ‘못하’, ‘고민’, ‘힘들’, ‘태어나’, ‘부모님’, ‘사춘기’로 대표되는 ‘성 정체성 고민’, 주제 4는 ‘이상’, ‘관심’, ‘혼란’ 등으로 대표되는 ‘성 정체성 의심’으로 레이블링을 진행하였다. 〈표 2〉와 비교할 때 ‘성적 지향’과 ‘성 정체성 의심’, ‘인간관계’ 등 주제 외에 ‘성 정체성 고민’이라는 새로운 주제가 출현한 것을 알 수 있다.

2. 고민과 관련된 키워드 및 그룹화

Exact match와 tf-idf 기반의 정보검색 결과에서 참여 연구자들 간의 합의를 통해 각각의 고민 관련 키워드를 추출하고 이를 소분류로 그룹화한 후 대분류로 통합하였다. 추출한 키워드들의 대분류와 소분류는 〈표 4〉와 같다.

〈표 4〉 고민 관련 키워드의 대분류 및 소분류

대분류	소분류	검색방법
일반적인 성 정체성 관련	성 정체성 혼란	exact match, tf-idf
	성 정체성의 성향 불일치	exact match, tf-idf
	성 정체성의 성적 측면 불일치	tf-idf
	성전환수술	exact match, tf-idf
성적 지향 관련	성적 지향 혼란	exact match, tf-idf
	특정 동성/양성에 대한 감정	exact match, tf-idf
	동성에 대한 성적인 생각	exact match, tf-idf
	성소수자에 대한 호기심 및 추구	tf-idf
성 정체성에 대한 부정적 인식	이성에 대한 무경험	exact match, tf-idf
	스스로에 대한 부정적 인식	exact match, tf-idf
	자살	exact match, tf-idf
감정 고백, 커밍아웃, 동성에	치료 및 상담	tf-idf
	감정 고백	exact match, tf-idf
	커밍아웃	exact match, tf-idf
	동성 간의 연애	exact match, tf-idf
미래, 이성 관계, 군대에 대한 우려	미래에 대한 우려	tf-idf
	이성과의 관계에 대한 우려	tf-idf
	군대에 대한 우려	tf-idf
성 정체성 혼란을 겪는 이유	성 정체성 혼란을 겪는 이유	tf-idf
	성 정체성 혼란 경험의 일반성	tf-idf

〈표 4〉는 성 정체성 고민 키워드들을 6개 대분류와 20개의 소분류를 나누어 보여준다. 소분류는 〈표 5〉에 있는 구체적 질문들을 요약한 키워드들을 기반으로 분류되었다.

〈표 5〉 고민 관련 키워드의 소분류에 대한 구체적 키워드

소분류	키워드
성 정체성 혼란	성 정체성에 대한 혼란, 어디에 속하는지 모르겠네, 판단 방법, 여자답다의 의미
성 정체성의 성향 불일치	생물학적 성과 불일치, 주어진 성이 싫어, 반대 성이 되고 싶어, 반대 성으로 인정 욕구, 숨기고 살아야 하나, 여성/남성적 성향(e.g., 체형, 행동, 성격)
성 정체성의 성적 측면 불일치	성관계시 여성의 역할을 원함, 여자로서의 성적 측면(생식기, 성관계) 호기심
성전환수술	성전환수술을 고민, 한국병원의 신뢰성, 수술 가능한 곳, 구체적 비용, 이후에 공격(기록, 자격증)으로 바뀌는지
성적 지향 혼란	동성애자/케이/레즈비언/양성애자/이성애자인가, 동성/양성을 좋아함, 어느 성을 좋아하는지 몰라, 성적 지향 판단 확률, 인지했던 지향성에 혼란
특정 동성/양성에 대한 감정	동성/양성의 친구/언니/형/선생님/연예인에게 감정(e.g. 좋아함, 설렘, 질투, 우정 이상), 감정의 정체(e.g. 동경, 착각 아닌가), 감정에 대한 대응
동성에 대한 성적인 생각	동성친구에게 성적 흥분, 게이아동으로 받기, 아한 생각에 동성이 먼저 생각남, 동성과 성관계 호기심

소셜 Q&A 사이트의 질문 분석을 통한 청소년의 성 정체성(sexual identity) 고민에 대한 이해

소분류	키워드
성소수자에 대한 호기심 및 추구	성적 소수자의 삶에 대해, 이반으로서 활동/친목 방법
이성에 대한 무경험	이성과 아무것도 안 해본, 사귀 적 없고
스스로에 대한 부정적 인식	자기혐오(e.g. 이상, 무섭다, 더러움, 싫고 안 좋게 느껴), 남들과 다름, 머리가 어떻게 돼버린
자살	자살 충동까지, 자살하고 싶다는 생각, 어떤 방식을 살까
치료 및 상담	어떻게 고치나, 성 정체성을 의학적으로 바꿀 가능성, 누구한테 상담 좋을지, 혼란 벗어나기 위한 진단/관리
감정 고백	마음을 전해야 할지, 좋아하는 상대의 마음, 좋아하는 걸 알게 된 상대의 반응, 그 애랑 불편해질까, 고백 어려움
커밍아웃	커밍아웃을 할지, 친구/부모님에게 말해야 할지, 부모님의 반응, 사회적으로 받아들여지지 않는, 소문 걱정
동성 간의 연애	관계 지속을 고민, 성관계에서 리드, 사귀어야 할지 말지, 연애에 대한 시선 걱정, 교제 좋았지만 두려운 기분
미래에 대한 우려	어른이 되어 어떻게 살아야, 게이 인정했지만 미래 걱정
이성과의 관계에 대한 우려	결혼/이성교제 어려울 텐데, 결혼 후 이성과 성관계 좋지 않을
군대에 대한 우려	군대 견디 낼 수 있을지, 군대 뺄 수 있을 거라지만 무지해서
성 정체성 혼란을 겪는 이유	사춘기적 아니면 정신적 문제인지, 후천적으로도 게이가 되는지
성 정체성 혼란 경험의 일반성	보통의 사람들이 같은 생각, 나이 때 원래 감정, 살면서 하는 경험 남들보다 늦게

첫 번째 대분류인 일반적인 성 정체성 관련 그룹은 스스로를 규정하는 성에 관한 고민들로 이를 판단하는 것이 혼란스럽거나, 생물학적 성과 본인의 성향 및 성관계에서의 역할 등 성적인 측면이 일치하지 않는 것, 그리고 그에 따른 성전환수술과 관련된 키워드들을 포함한다. 두 번째 성적 지향 관련 그룹은 스스로가 감정적·성적으로 끌리는 성에 관한 고민들로 이를 분류하는 동성애자, 양성애자, 이성애자로서 자신을 판단하는 것이 혼란스럽거나, 동성 또는 양성의 특정 대상에게 감정을 느끼는 것, 나아가 동성에 대한 성적인 호감이 드는 것과 관련된 키워드들을 포함한다. 이는 실제 성소수자의 삶에 대한 호기심 및 이를 추구하고자 하는 데서 오는 고민들로 표현되는가 하면, 이성 관계의 경험이 부재한 데서 비롯된 성적 지향에 대한 고민들로 나타났다.

세 번째 대분류인 성 정체성에 대한 부정적 인식은 성 정체성 혼란으로 인해 겪게 되는 자기혐오를 비롯한 스스로에 대한 부정적 인식과, 이에 따른 극단적 선택으로써 자살, 또는 벗어나기 위한 방법으로써 치료나 상담과 관련된 키워드들의 그룹이다. 네 번째 그룹에서는 좋아하는 동성에게 감정을 고백하거나 자신의 주변사람들에게 커밍아웃하는 것과 관련하여 고민하는 키워드들, 나아가 동성 간의 실제적인 연애에 있어서 생기는 고민들을 묶어서 분류하였다. 다섯 번째 그룹은 성소수자로서 앞으로 살아갈 미래에 대한 우려, 구체적으로는 결혼 등 이성과의 관계와 군대에 대한 우려와 관련된 키워드들을 포함하고 있다. 마지막으로 이러한 성 정체성 혼란을 겪는 이유와 경험의 일반성에 대한 고민들, 예를 들면 사춘기적 문제인지 아니면 정신적 문제인지, 이 나이

때 원래 경험하는 감정인지 등과 관련된 키워드들을 하나의 그룹으로 분류하였다.

고민과 관련된 키워드를 추출하는 분석 과정에서 사용되는 tf-idf의 장점은 exact match와 비교할 때 다음과 같은 장점이 두드러진다. 즉, 성 정체성 고민과 밀접한 관련이 있는 키워드로만 이루어진 간결한 문장들이 검색 결과의 상위에 표시되어 고민의 핵심을 파악하는 것이 용이하다는 것이다. 또한, tf-idf 기반의 검색을 통해 고민과 관련된 보다 다양하고 구체적인 주제의 키워드들을 추출할 수 있었다. 예를 들어 성 정체성의 성적 측면 불일치, 치료 및 상담, 그리고 미래, 이성과의 관계, 군대에 대한 우려와 같이 정확한 매치(exact match)를 통해서 나타났던 새로운 그룹의 키워드들의 추출이 가능한 한편, 성전환수술과 같이 공통적으로 나타난 그룹에 대해서는 그 수술에 관련되는 병원의 신뢰성이나 비용, 수술 이후 공적인 문서에서의 성별 표시 변화 등의 구체적인 키워드들을 추출할 수 있었다.

V. 결론 및 논의

성 정체성 관련 질문은 민감한 주제이기 때문에 오프라인에서 질문하고 상담 받기 어려운 점이 많다. 네이버 지식iN은 익명성이 보장되는 온라인 Q&A 사이트로 성 정체성을 고민하는 사용자들이 질문을 하고 답변을 받는 커뮤니티로 활발히 활용되고 있다. 이를 통해 해당 플랫폼은 사용자의 고민을 해결하는 면에서 긍정적인 역할을 하는 것으로 파악된다. 따라서 네이버 지식iN의 성 정체성 관련 질문을 분석하는 것은 성 정체성과 관련된 국내 집단지성들의 온라인 토론의 현황을 이해함으로써 그와 관련된 사회적 인식 및 규범 등을 하는데 매우 중요하다. 본 연구에서는 Q&A 사이트인 네이버 지식iN에서 성 정체성과 관련된 질문들을 분석하여 질문의 중심이 되는 주제와 구체적인 고민들을 추출하였다. 이를 위해 네이버 지식iN에서 2010년 1월부터 2018년 12월까지 작성된 2120개의 질문을 수집하였다. 자연어 처리와 정보 검색 방법론을 사용하여 질문들에 존재하는 주제와 고민들을 추출하고 분석하였다. 분석 결과 질문들의 주요 주제는 인간관계, 성 정체성 의심, 성적 지향, 감정 및 관계, 성 정체성 고민 등으로 나타났다. 사용자들은 주로 1) 일반적인 성 정체성, 2) 성적 지향, 3) 성 정체성에 대한 부정적 인식, 4) 감정 고백, 커밍아웃, 동성애, 5) 미래, 이성 관계, 군대에 대한 우려, 6) 성 정체성 혼란을 겪는 이유 등과 관련된 고민을 하고 있는 것으로 나타났다. 하지만, 본 연구는 “성 정체성”이란 키워드 외에 “성 소수자”, “게이”, “레즈비언” 등의 관련 키워드를 최대한 포함하여 데이터를 수집·분석하지 못한 한계가 있어 후속 연구에서는 이를 보강할 필요가 있다.

본 연구의 주요 목적 중의 하나인 고민과 관련된 키워드를 추출하여 분석하는 데 사용된 tf-idf 기법은 연구 맥락에 적절한 구체적인 주제들을 파악하는 데 효과적이었다. 이는 성 정체성의 혼란

으로 인해 힘들어 하는 사람들에게 그 고민을 덜어주거나 해결해 주기 위해 집단 지성의 커뮤니티가 도울 수 있는 시사점을 제시한다. 예를 들면, 성 정체성 혼란에 대한 지식의 부재로 인해 고민을 하는 청소년들에게 그러한 혼란이 청소년기나 사춘기에 오는 것이 어떤 개인에게만 해당하는 특수한 문제가 아니라 신체 발달과 함께 누구에게나 오는 정상적인 과정임을 알려줌으로써 그러한 고민의 해결을 도와줄 수 있다. 무엇보다도, 동성애에 대한 사회적 낙인이나 부정적 인식으로 인해 누구에게도 자유롭게 질문하지 못하거나 고통스러워하는 청소년들에게 정서적·사회적 지원을 제공하는 것은 이러한 고민을 극복하는 데 효과적일 수 있다.

따라서, 본 연구는 소셜 Q&A에 나타난 성 정체성이라는 민감한 주제에 대한 실제적 고민들을 수집하여 자연어 텍스트 분석과 토픽 모델링 기법을 통해 정보 요구를 효과적으로 도출하였다는 점에서 건강정보행위분야의 연구방법론적 확장에 기여했다. 본 연구는 exact match와 tf-idf 기반의 검색 방법의 장점을 비교함으로써 이후 다른 주제의 건강정보요구 및 행위를 분석하는 연구에도 적용할 수 있는 잠재성을 제시하였다는 학문적 의의가 있다. 더 나아가, 특정 관심을 가지고 있는 이용자들이 온라인 커뮤니티에서 실행하는 정보탐색 및 정보공유행태를 관찰함으로써 정보행위분야의 학문적 성숙에도 기여하였다.

참 고 문 헌

- 강병철, 하경희. 2005. 청소년 동성애자의 동성애 관련 특성이 자살 위험성에 미치는 영향. 『청소년학연구』, 12(3): 267-289.
- 강병철, 하경희. 2007. 청소년 동성애자의 성적체성 드러내기가 자살위험성에 영향을 미치는 과정에 관한 연구. 『사회복지연구』, 35: 167-189.
- 김경호. 2009. 청소년 동성애와 상담방안에 관한 연구. 『아시아교육연구』, 10(2): 135-168.
- 김선호. 2015. LGBTQ의 인권과 도서관서비스. 『한국도서관·정보학회지』, 46(4): 21-44.
- 등탁유, 유승관. 2016. 모바일 인스턴트 메신저 이용 동기, 만족도 및 충성도 차이가 정보공유에 미치는 영향에 관한 비교연구: 카카오톡과 중국의 위챗을 중심으로. 『커뮤니케이션학연구』, 24(2): 227-250.
- 민희. 2018. 북한이탈주민의 사회적응. 『정보화정책저널』, 25(2): 67-83.
- 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.
- 손영화. 2018. 페이스북 이용자의 자기노출수준, 자아표현 동기 및 사회교류 동기가 주관적 행복

- 감에 미치는 영향. 『한국사회과학연구』, 37(2): 115-150.
- 신동열. 2010. 성소수자의 성인식·성태도·성행동에 관한 연구. 『에니어그램연구』, 7(2): 111-141.
- 유정옥. 2019. 텍스트마이닝을 이용한 청소년의 성고민 분석. 『보건정보통계학회지』, 44(2): 181-188.
- 유호선, 김현진, 오효정. 2018. 재난 사건별 이슈 생존 주기 유형 분석. 『한국정보기술학회논문지』, 16(3): 125-134.
- 이서영, 권상집. 2019. 트위터 메시지 분석을 통한 선거 결과 예측 고찰. 『한국콘텐츠학회논문지』, 19(4): 174-186.
- 이영선, 권보민, 이동훈. 2012. 청소년의 동성애 경험, 성의식, 동성애 혐오, 우울의 관계. 『재활심리연구』, 19(1): 129-147.
- 이재범 외. 2012. 블로그-트위터 매체 간 특성 차이 및 사용자 제품정보 처리와 평가차이 비교에 관한 연구. 『정보시스템연구』, 21(1): 69-91.
- 주재홍. 2017. 한국의 청소년 성소수자들로부터 알게 된 그들의 삶의 이야기들: 질적 사례 연구. 『교육문화연구』, 23(1): 175-215.
- Bae, Beom Jun and Yong Jeong Yi. 2017. "What Answers Do Questioners Want on Social Q&A? User Preferences of Answers About STDs." *Internet Research*, 27(5): 1104-1121.
- Bae, Beom Jun and Yong Jeong Yi. 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.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, 3(4/5): 993-1022.
- Byron, Paul. 2017. "Friendship, Sexual Intimacy and Young People's Negotiations of Sexual Health." *Culture, Health & Sexuality*, 19(4): 486-500.
- Byron, Paul et al. 2019. "Hey, I'm Having These Experiences": Tumblr Use and Young People's Queer (Dis)connections." *International Journal of Communication*, 13(2019): 2239-2259.
- Chatman, Elfreda A. 1996. "The Impoverished Life-World of Outsiders." *Journal of the American Society for Information Science*, 47(3): 193-206.
- Clements-Nolle, Kristen et al. 2018. "Sexual Identity, Adverse Childhood Experiences, and Suicidal Behaviors." *Journal of Adolescent Health*, 62(2): 198-204.

- De Choudhury, Munmun, Scott Counts and Eric Horvitz. 2013. "Social Media as a Measurement Tool of Depression in Populations." *Proceedings of the 5th Annual ACM Web Science Conference*, 2013: 47-56.
- Evers, Clifton W. et al. 2013. "Young People, Social Media, Social Network Sites and Sexual Health Communication in Australia: This is Funny, You Should Watch It." *International Journal of Communication*, 7(1): 263-280.
- Johns, Michelle M. et al. 2018. "Violence Victimization, Substance Use, and Suicide Risk Among Sexual Minority High School Students-United States, 2015-2017." *Morbidity and Mortality Weekly Report*, 67(43): 1211-1215.
- Kitzie, Vanessa. 2015. "Labels are for Clothing: Negotiating LGBT Identities Within Social Question-Answering Sites." *iConference 2015 Proceedings*, 2015: 1-13.
- Lachmar, E. Megan et al. 2017. "#MyDepressionLooksLike: Examining Public Discourse About Depression on Twitter." *JMIR Mental Health*, 4(4): e43.
- Liu, Zhe and Bernard J. Jansen. 2018. "Questioner or Question: Predicting the Response Rate in Social Question and Answering on Sina Weibo." *Information Processing & Management*, 54(2): 159-174.
- Livingstone, Sonia and Jessica Mason. 2015. *Sexual Rights and Sexual Risks Among Youth Online: A Review of Existing Knowledge Regarding Children and Young People's Developing Sexuality in Relation to New Media Environments*. European NGO Alliance for Child Safety Online(ENACSO).
- Řehůřek, Radim and Petr Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 2010: 46-50.
- Sexuality Information and Education Council of the United States. 2004. *Guidelines for Comprehensive Sexuality Education: Kindergarten-12th Grade*.
<www.siecus.org/_data/global/images/guidelines.pdf> [cited 2016. 9. 17].
- Song, Hayeon 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.
- Yi, Yong Jeong. 2018. "Sexual Health Information-Seeking Behavior on a Social Media Site: Predictors of Best Answer Selection." *Online Information Review*, 42(6): 880-897.

- Yong, J. Yi, Besiki Stvilia and Lorri Mon. 2012. "Cultural Influences on Seeking Quality Health Information: An Exploratory Study of the Korean Community." *Library & Information Science Research*, 34(1): 45-51.
- Zhao, Yuehua and Jin Zhang. 2017. "Consumer Health Information Seeking in Social Media: A Literature Review." *Health Information & Libraries Journal*, 34(4): 268-283.

• 국한문 참고문헌의 영문 표기

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

- Deng, Zhuo Wei and Seung Kwan Ryu. 2016. "A Comparative Study on the Relationships Among Mobile Instant Message Users' Motivation, Satisfaction, and Loyalty Factors on Information Sharing and Diffusion: Focusing on Kakao Talk and Chinese WeChat." *Korean Journal of Communication Studies*, 24(2): 227-250.
- Joo, Jae-Hong. 2017. "Hidden Lives of Korean LGBT Adolescents: A Qualitative Case Study." *Journal of Education & Culture*, 23(1): 175-215.
- Kang, Byung Chul and Kyung Hee Ha. 2005. "A Study on Homosexual Adolescents' Sexual Identity and Suicidal Risk." *Korean Journal of Youth Study*, 12(3): 267-289.
- Kang, Byung Chul and Kyung Hee Ha. 2007. "A Path Model Analysis on the Influences of Disclosure on the Suicidal Risk of Gay and Lesbian Youth." *Korean Journal of Social Welfare Studies*, 35: 167-189.
- Kim, Kyounggho. 2009. "Gay and Lesbian Adolescents and Counseling Intervention: Focusing on One-to-One, Group, Family, and School Counseling." *Asian Journal of Education*, 10(2): 135-168.
- Kim, Seon-Ho. 2015. "LGBTQ's Human Rights and Library Services." *Journal of Korean Library and Information Science Society*, 46(4): 21-44.
- Lee, Jae Beom et al. 2012. "A Comparative Study on Different Characteristics of Social Media and Product Information Processing and Evaluation." *The Journal of Information Systems*, 21(1): 69-91.
- Lee, Seo Young and Sang Jib Kwon. 2019. "A Study on Predicting Presidential Election Results by Analyzing Twitter Message Contents: A Focus on the 18th Presidential Election in Korea." *Journal of the Korea Contents Association*, 19(4): 174-186.

- Lee, Young Sun, Bo-min Kwon and Donghoon Lee. 2012. "The Relationships Among Homosexual Experiences, Sexual Attitude, Sexual Homophobic, and Depression in Adolescents." *Journal of Rehabilitation Psychology*, 19(1): 129-147.
- Min, Hee. 2018. "North Korean Defectors' Adaptation to South Korean Society: Their Emotional States, Social Media Use, and Life Satisfaction." *Informatization Policy*, 25(2): 67-83.
- Park, Eunjeong L. and Sungzoon Cho. 2014. "KoNLPy: Korean Natural Language Processing in Python." *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology*, 2014: 133-136.
- Shin, Dong Ryeol. 2010. "Research on the Sexual Cognizance, Attitude and Behaviour of Sexual Minority." *Journal of Enneagram Studies*, 7(2): 111-141.
- Son, Younghwa. 2018. "The Effect of Facebook Users Levels of Self-disclosure, Motives for Self-expression and Social Interaction on Subjective Happiness." *Korean Journal of Social Science*, 37(2): 115-150.
- Yoo, Hosuon, Hyeon-Jin Kim and Hyo-Jung Oh. 2018. "Analysis of Issue Life Cycle Types of Disaster Events." *Journal of Korean Institute of Information Technology*, 16(3): 125-134.
- Yu, Jungok. 2019. "Text Mining for Identifying Topics in Internet Q&A About Adolescents' Sexual Concerns." *Journal of Health Informatics and Statistics*, 44(2): 181-188.