

Web of Science 빅데이터를 활용한 텍스트 마이닝 기반의 정보윤리 이슈 탐색

김한성[†]

요 약

본 연구의 목적은 Web of Science(WoS)에서 제공하는 학술 빅데이터를 활용하여 정보윤리 이슈를 탐색하고 향후 정보과 정보윤리 교육을 위한 시사점을 제공하는 것에 있다. 이를 위해 WoS에서 제공하는 학술 논문 중 정보윤리와 관련해 출판된 318편의 논문을 텍스트 마이닝 하였다. 구체적으로는 R을 활용해 주요 키워드에 대한 빈도 분석(TF, DF, TF-IDF), 토픽 모델링 기반의 정보윤리 이슈 분석, 그리고 각 이슈에 대한 연도별 출연 빈도를 분석하여 정보윤리 연구의 경향성을 탐색하였다. 주요 결과를 살펴보면 다음과 같다. 첫째, TF-IDF를 통해 'digital', 'student', 'software', 'privacy' 등의 단어가 주요 키워드임을 확인하였다. 둘째, 토픽 모델링 분석 결과, 'Professional value', 'Cyber-bullying', 'AI and Social Impact' 등을 포함한 총 8개 이슈로 분석되었고, 그 중, 'Professional value'와 'Cyber-bullying' 이슈가 상대적으로 높은 비율을 차지하고 있었다. 본 연구는 이러한 분석 결과를 기초로 우리나라 정보윤리 교육을 시사점을 논의하였다.

주제어 : 정보윤리, 정보교육, 텍스트 마이닝, 빅데이터, Web of Science

Exploring Information Ethics Issues based on Text Mining using Big Data from Web of Science

Han Sung Kim[†]

ABSTRACT

The purpose of this study is to explore information ethics issues based on academic big data from Web of Science (WoS) and to provide implications for information ethics education in informatics subject. To this end, 318 published papers from WoS related to information ethics were text mined. Specifically, this paper analyzed the frequency of key-words(TF, DF, TF-IDF), information ethics issues using topic modeling, and frequency of appearances by year for each issue. This paper used 'tm', 'topicmodel' package of R for text mining. The main results are as follows. First, this paper confirmed that the words 'digital', 'student', 'software', and 'privacy' were the main key-words through TF-IDF. Second, the topic modeling analysis showed 8 issues such as 'Professional value', 'Cyber-bullying', 'AI and Social Impact' et al., and the proportion of 'Professional value' and 'Cyber-bullying' was relatively high. This study discussed the implications for information ethics education in Korea based on the results of this analysis.

Keywords : Information Ethics, Informatics, Text Mining, Big Data, Web of Science

[†]정 회 원: 한국교육학술정보원 선임연구원
논문접수: 2019년 5월 18일, 심사완료: 2019년 5월 28일, 게재확정: 2019년 5월 30일

1. 서론

컴퓨터의 등장과 ICT의 빠른 발전은 생활의 편리함과 다양한 혜택을 가능하게 해주었다. 특히 참여, 공개, 공유를 표방했던 Web 2.0 시대를 빠르게 지나 Big data와 Artificial Intelligence(AI)의 등장은 생활 문화 전반에 큰 영향을 미치며 삶의 경험을 질적으로 변화 시키고 있다[1].

하지만 이렇게 빠르게 변화하는 사회문화 이면에는 프라이버시, 정보 접근성, 정보 신뢰성, 소프트웨어 안정성 등과 같은 윤리적 이슈들이 지속적으로 등장하고 있다[2][3][4][5]. 이러한 윤리적 이슈 중에는 기존의 갈등이 심화된 경우도, 새롭게 갈등이 되는 경우도 존재한다.

컴퓨터 기술로 인해 발생하는 윤리적 이슈로 인해 사회적 문제가 심각한 수준에 이르자 이를 예방하고 대응하기 위한 대책으로서 교육적 접근의 필요성이 대두 되고 있다[6][7]. 우리나라 초·중등학교에서는 도덕과, 정보과 등에서 크게는 교육과정의 한 영역으로서 작게는 교과서의 소단원으로 정보윤리 관련 내용을 편성하여 시행하고 하고 있다[8][9]

정보과는 2015 개정 교육과정에 정보문화 영역을 대영역으로 구성하고 핵심 개념 중 하나로 정보윤리를 가르치고자 하였다[8]. 하지만 교육과정에 제시된 정보윤리 내용이 교육과정의 목적에 명시된 것처럼 ‘청소년들의 정보문화 소양과 윤리적 의식 함양을 도울 수 있도록 적합한 내용으로 구성되어 있는가?’에 대한 객관적 차원의 검토는 교육과정의 발전을 위해 지속적으로 필요한 과제이다[10].

교육내용 선정은 교과 목표의 부합성, 개인적·사회적 요구, 학문적 특성 및 교수·학습 가능성 등을 만족해야한다[10]. 특히, 정보윤리의 학문적 역사가 길지 않았음을 고려하였을 때 교육 내용요소 선정을 위해 그 동안 시행된 학문적 속의 과정에 관한 재논의는 반드시 필요하며, 이를 보완하기 위한 보다 체계적인 내용요소 선정 과정이 필요하다.

이러한 문제의식에서 출발한 본 연구의 목적은 정보윤리 내용요소 선정을 위해 고려해야할 정보윤리 이슈가 무엇인지 탐색하고, 이를 토대로 정보윤리 교육을 위한 시사점을 도출하는 것에 있다.

본 연구의 목적 달성을 위해 구체적인 연구문제를

를 설정해 수행하고자 한다. 각 연구문제를 살펴보면 다음과 같다. 첫째, 정보윤리 관련 연구에서 사용되고 있는 주요 키워드는 무엇인가? 둘째, 정보윤리와 관련된 주요 이슈들은 무엇이라 할 수 있는가? 셋째, 정보윤리 관련 주요 이슈별 연구경향은 어떠한가?

연구문제를 해결하기 위해 Web of Science (WoS)에서 제공하는 학술 빅데이터를 활용해 키워드 빈도 분석, 토픽모델링 분석, 연도별 토픽 출현 빈도 분석을 실시하였다.

2. 이론적 배경

본 절에서는 정보윤리와 관련한 여러 학자들의 이론을 분석하여 정보윤리의 개념과 범위를 정의하고 정보과 및 관련 교과들에서 제시하고 있는 정보윤리 교육 내용요소를 살펴본다.

2.1 정보윤리의 개념과 범위

정보윤리에 대한 학문적 관심은 컴퓨터 기술의 윤리적 문제를 다루기 위해 태동하였던 컴퓨터윤리학(Computer Ethics)에 토대를 두고 있다고 할 수 있다[11]. 컴퓨터윤리학은 1950년 Wiener(1950)의 ‘Human and Human Use of Human’이라는 저서에서 컴퓨터와 정보 기술이 만들어 낼 수 있는 수많은 윤리적 문제를 제기하며 출발하였다[12]. 이후, Maner(1980)는 컴퓨터기술이 아니라면 발생하지 않았을 윤리적 이슈를 제시하면서 완전히 새로운 윤리 분야로 컴퓨터윤리학을 정의하고, 컴퓨터윤리를 위한 ‘Starter Kit’을 발표하였다. 여기에는 컴퓨터윤리의 이론적 근거, 정보보호 및 보안, 컴퓨터 범죄, 컴퓨터 기반 의사결정, 기술 의존과 같은 주제와 교육과정, 교수법 등을 제시하였다[13]. 반면, 존슨은 컴퓨터 기술로 인해 기존에 존재하던 관례적인 도덕적 문제들과 딜레마들이 복잡해지며 (give them a new twist), 이를 새롭게 해석하는 연구 분야로 컴퓨터윤리학을 정의하였다. 존슨은 이를 통해 컴퓨터윤리학을 컴퓨터로 인해 새롭게 발생하는 윤리적 이슈뿐만 아니라, 기존의 이슈를 복잡하게 만드는 기술과 윤리적 문제를 다루는 분야로 확장하였다[14]. 이 후, Moor(1985)가 ‘What

is computer ethics?’를 통해 컴퓨터윤리학을 ‘컴퓨터 기술로 인해 발생하는 정책적 공백(policy vacuum)을 해결하기 위한 학문’으로 정의하였으며, 이는 현재까지 가장 보편적인 정의 중 하나로 활용되고 있다[15].

이 후, 컴퓨터윤리학은 ‘정보윤리’, ‘사이버윤리’, ‘인터넷윤리’, 등과 같이 명명되며 그 연구 범위가 다양하고 확대되어졌다. 임상수(2003)는 정보윤리의 교육의 범위를 명확히 하기 위해 정보윤리의 범위를 ‘협회의 정보윤리’, ‘사이버윤리’, ‘컴퓨터윤리’로 구체화하여 그 내용을 살필 필요가 있다고 하였다[16]. 그는 협회의 정보윤리를 정보의 일생사에 관련된 윤리적 쟁점을 다루는 규범과 이론 체계로 규정하고, 사이버 윤리는 사이버 공간과 관련된 윤리적 쟁점을 다루는 규범과 이론으로 정의하였다. 그리고 컴퓨터윤리는 컴퓨터 기술과 관련된 윤리적 쟁점을 다루는 규범과 이론체제로 한정할 필요가 있다고 하였다. 반면, 김한성(2010)은 정보교과와 같은 경우 임상수(2003)가 정의한 광의의 정보윤리 중 컴퓨터윤리에 보다 초점을 맞추어 내용을 구성할 필요가 있음을 주장하였다[17]. 이는 Maker, 프로슈머와 같이 프로그래밍을 통한 개발경험이 더 이상 전문가들 소유물이 아니라 초·중등학생들에게도 매우 보편화되고 있는 시점에서 정보과 또한 IT의 올바른 사용뿐만 아니라 생산과정에서 발생할 수 있는 윤리적 이슈들에 대해 대응할 수 있는 소양이 필요함을 강조한 것이다.

2.2 정보윤리 교육 관련 내용 요소

정보윤리 교육 내용요소를 살펴보기 위해 2009, 2015 개정 교육과정에서 관련 내용을 살펴보면, <표 1>과 같다[8][9][18][19].

우선, 현재 적용되고 있는 2015 개정 교육과정에서는 도덕과와 정보과에서 공통적으로 정보윤리를 다루고 있다. 도덕과의 경우, 초등학교에서는 사이버 예절, 중학교에서는 사이버 공간에서의 소통 그리고 고등학교에서는 정보 기술 발달로 인해 발생하는 윤리적 문제들을 알아보는 과정으로 구성되어 있다. 정보과의 경우 초·중·고 모두 개인정보와 저작권을 다루고 있으며, 중·고등학교에 사이버 윤리(예절)가 추가적으로 구성되어 있다.

<표 1> 정보과와 도덕과의 정보윤리 내용요소

구분	2009 개정 교육과정		2015 개정 교육과정	
		내용요소		내용요소
정보과	초 (실과)	• 정보기기와 사이버 공간	• 개인 정보와 지식 재산 보호	
	중 (정보)	• 정보과학과 정보사회 • 정보의 윤리적 활용 • 정보 사회의 역기능과 대처	• 개인정보와 저작권 보호 • 사이버 윤리	
	고 (정보)	• 정보과학과 정보사회 • 정보의 윤리적 활용 • 정보사회의 역기능과 대처	• 정보보호와 보안 • 저작권 활용 • 사이버 윤리	
도덕과	초 (도덕)	• 정보사회에서의 올바른 생활	• 사이버 공간에서 지켜야 할 것은 무엇일까? (사이버 예절, 준법)	
	중 (도덕)	• 사이버 윤리와 예절	• 정보화 시대에 우리는 어떻게 소통해야하는가? (정보통신윤리)	
	고 (생활과 윤리)	• 정보 통신 기술과 윤리 • 사이버 공간과 인간의 자아 정체성	• 정보 기술 발달과 정보윤리 • 정보 사회에서의 매체윤리	

이를 통해 도덕과의 경우, 초·중학교에서는 사이버 예절을 중심으로 가르치며, 고등학교에서 다양한 윤리적 이슈들을 살펴보고 해결 방안을 탐구하는 내용으로 확대해 구성하였음을 알 수 있다. 반면, 정보과의 경우 개인정보와 저작권에 초점을 맞추어 학교급이 올라갈수록 내용을 심화되는 형태로 구성되어 있음을 알 수 있다.

이렇게 구성된 두 교과의 내용을 2009 개정 교육과정과 비교해 살펴보면, 2009 개정 교육과정 개발 당시 제기 되었던 두 교과간의 중복성의 문제는 다소 해결된 것으로 보인다[17], 정보과의 경우 그 내용의 범위 및 세부 주제가 2009년에 비해 다소 줄어든 것을 알 수 있다. 이는 수업 시수 및 교과내 타 영역과의 비중을 고려한 구성이라 할지라도 선정된 교육내용의 타당성 및 학습 가능성 등에 대해서는 지속적인 논의가 필요할 것이다.

3. 연구 방법

3.1 데이터

3.1.1 데이터 수집

본 연구에서 분석 대상으로 하는 학술 데이터는 WoS에서 제공하는 SSCI, SCI, SCIE, A&HCI급 학술 논문이다. 구체적인 분석 대상은 각 논문의

제목, 저자 키워드, 학술지 지정 키워드, 그리고 요약이다. 학술지 저자 키워드와 함께 학술지 지정 키워드까지 분석대상으로 삼은 이유는 각 키워드들이 핵심적인 정보를 담고 있을 뿐만 아니라, 각 학술지에서 사용하는 공통적인 키워드까지 고려한 분석이 가능하기 때문이다.

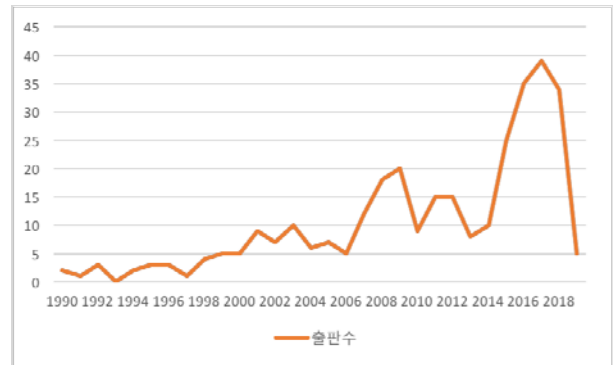
분석 대상 논문을 수집하기 위하여 다음과 같은 절차를 거쳤다. 첫째, 2019년 3월 31일을 기준으로 WoS에 접속하여 정보윤리와 관련된 논문을 검색하였다. 자료 검색 시 정확하면서도 가급적 많은 저널이 포함될 수 있도록 검색 기간은 제한하지 않았으며, 정보윤리와 관련한 다양한 키워드를 활용하였다. 키워드는 선행연구에서 도출된 ‘정보윤리(Information ethics)’, ‘컴퓨터 윤리(Computer ethics)’, ‘인터넷 윤리(Internet ethics)’, ‘사이버 윤리(Cyber ethics)’, ‘디지털 윤리(Digital ethics)’를 활용하였으며, ‘ethics’과 유사어로 활용되고 있는 ‘moral’을 ‘ethics’과 대체하여 검색 구문을 추가해 활용해 검색하였다. 검색 구문 예시를 살펴보면 다음과 같다.

‘Computer ethics’ & ‘Information ethics*’
& ‘Internet ethics*’ & ‘Cyber ethics*’ &
‘Digital ethics*’ ... (생략)*

이렇게 수집된 논문 중 요약 정보를 제공하지 않거나, 정보량이 극히 제한적인 논문들을 제외하고 총 318편의 논문을 분석 대상으로 선정하였다.

각 연도별 논문 수를 살펴보면 [그림 1] 과 같다. 연도별 논문 수를 살펴보면 1990년 2편으로부터 시작하여 꾸준히 출판 논문 수가 증가하고 있으며, 2015년에는 25편, 2017년에는 약 40편이 출판되며, 정보윤리 관련 출판 논문 수가 가파르게 증가하고 있는 것을 알 수 있다. 단, 2019년의 경우 3월 기준으로 5편이 출판되었다.

정보윤리가 출판된 주요 저널지를 살펴보면, SCIENCE AND ENGINEERING ETHICS(19편), JOURNAL OF BUSINESS ETHICS(19편), ETHICS AND INFORMATION TECHNOLOGY(14편), COMPUTERS IN HUMAN BEHAVIOR(8편), COMPUTERS & EDUCATION(7편) 순으로 나타났다.



[그림 1] 연도별 출판 논문 수

3.1.2 데이터 전처리

텍스트 마이닝 과정에서는 수집된 비정형 데이터를 연구 목적을 달성하기에 적합한 형태로 전처리하는 과정이 무엇보다 중요하다. 본 연구에서는 이를 위해 다음과 같은 절차를 통해 전처리를 수행하였다.

첫째, 분석 데이터를 말뭉치(Corpus)로 변환한 후, 숫자, 특수 문자, 문장 부호를 제거하였다. 둘째, 이렇게 처리된 데이터를 1차로 확인한 후, 대문자로 나타난 텍스트를 소문자로 전환하고 tm 패키지의 en 함수를 활용해 불용문자를 제거하였다. 여기서 en 함수는 ‘i’, ‘my’, ‘me’ 등과 같이 일반적으로 분석의 의미를 갖지 못하는 최소한의 단어들을 포함하는 함수이다. 셋째, 어근 동일화 함수를 이용해 일반적으로 같은 의미에서 파생한 단어의 다양한 형태(시제, 단복수 등)를 하나의 단어로 치환하였다. 넷째, 끝으로 이렇게 정리된 단어들의 주요 빈도 수를 살펴보고 본 연구 목적에 맞지 않는 단어를 추가적으로 삭제하는 과정을 반복하였다. 특히, 거의 모든 문서에 공통적으로 나타나면서도 큰 의미를 갖지 못하는 ‘ethics’, ‘moral’, ‘research’, ‘study’ 등과 같은 단어들을 추가적으로 삭제하였다.

이렇게 전처리 과정을 거쳐 본 연구에 활용된 최종 단어 수는 36,763개이며, 이 중 고유 단어 수는 3,837개이다.

3.2 분석 절차 및 도구

전처리 과정 후 다음과 같이 각 연구 내용을 중

심으로 분석하였다.

첫째, 연구문제1: 주요 키워드 빈도 분석을 위해 단어빈도(Term Frequency, TF), 문서빈도(Document Frequency, DF), 그리고 TF-IDF(Term Frequency - Inverse Document Frequency) 가중치를 이용한 빈도 분석을 하였다.

TF는 단어 빈도를 산출하는 가장 간단한 방법으로 전체 문서 내에 나타나는 해당 단어의 총 빈도를 사용하는 것이다. 이를 통해 가장 단순하면서도 간결한 방법으로 정보윤리와 관련된 주요 단어들 살펴볼 수 있다. DF의 경우 각 단어가 등장하는 문서의 수로서 널리 사용되고 있는 단어가 무엇인지 참고할 수 있다. TF-IDF는 각 단어가 문서에서 차지하는 중요도를 고려해 가중치를 부여하는 통계적 수치로 일반적으로 관용어처럼 쓰이는 단어들의 중요도를 낮게 고려하는 방식이다.

둘째, 연구문제2: 정보윤리 이슈를 도출하기 위해 토픽 모델링을 실시하였다. 토픽 모델링은 문서와 단어로 구성된 데이터를 기반으로 각 문서에 잠재되어 있다고 가정되는 토픽의 등장 확률을 추정하는 통계적 텍스트 처리기법이다[20]. 본 연구에서는 토픽 모델링을 위해 상관토픽모형(Correlated Topic Model, CTM)을 사용하였다. CTM은 토픽 모델링을 위해 일반적으로 사용되는 잠재적 할당 모형(Latent Dirichlet Allocation, LDA)을 보완한 것으로, 잠재토픽 사이의 상관관계 발생을 고려해 토픽을 도출하는 보다 정교한 모형이다.

토픽 모델링을 통해 도출된 정보윤리 이슈의 적정 수는 Perplexity Value(PV)를 평가를 통해 확인하였다. PV는 그 값이 적게 나올수록 적절한 모델로 평가되지만, 일반적으로 도출하는 토픽 수가 많을수록 값이 낮아지는 경향이 있다. 이에, 본 연구에서는 특정 토픽 수 구간을 정해 CTM을 여러 번 구현한 뒤, 각 토픽을 대표하는 논문을 직접 확인한 후 적절히 분석이 되었는지 판단하는 과정을 반복하였다. 이를 위해 PV를 기준으로 7개에서 12개 사이의 토픽 구간을 정한 뒤, 각 이슈를 대표하는 15개 키워드를 도출하였다. 이를 통해 각 키워드간의 상관성을 살펴본 후, 1차로 각 토픽의 이름을 명명하였다. 이 후, posterior 함수를 활용해 각 토픽을 대표하는 논문을 확인하는 사후 분석을

실시하였다. 사후 분석 과정을 통해 임시 명명되었던 이슈와 실제 대표 논문간의 일치도를 확인한 후, 최종적으로 8개 토픽과 토픽명을 선정하였다.

끝으로 연구문제3: 각 이슈별 연구동향을 살펴보기 위해 각 연도별 저널에 해당 이슈가 등장하는 확률 값을 계산하였으며, 이를 전체 이슈에 대한 상대적인 분포와 함께 각 이슈별 세부적인 경향으로 나누어 분석하였다.

본 연구는 텍스트 마이닝을 위해 RStudio version 1.1.463(for mac)을 활용하였으며, R에서 제공하는 자연어 처리 패키지인 tm과 토픽 모델링 패키지 topicmodel를 사용하였다.

4. 연구 결과

4.1 키워드 분석

<표 2> 주요 키워드 빈도 분석

순위	단어빈도(TF)		문서빈도(DF)		TF-IDF
	단어	빈도	단어	빈도	단어
1	information	1042	information	236	computer
2	computer	463	use	164	digital
3	use	373	computer	135	student
4	technology	264	issue	129	internet
5	internet	237	technology	122	information
6	issue	229	develop	118	data
7	student	229	social	106	software
8	social	217	result	97	privacy
9	develop	216	relate	95	technology
10	digital	216	approach	87	use
11	data	188	practice	83	behavior
12	education	169	provide	81	social
13	system	168	data	79	education
14	behavior	162	make	79	professional
15	privacy	160	system	79	system
16	relation	158	value	79	model
17	value	152	education	78	security
18	software	147	design	76	develop
19	professional	146	new	75	responsibility
20	approach	140	base	74	value

<표 3> 토픽 모델링 기반의 주요 이슈 분석

토픽 번호	이슈명	주요 키워드 및 내용		출연 문서 확률 총합*
1	Professional value	키워드	professional, value, development, code, problem	64.296
		주요내용	소프트웨어 개발 과정에서 경험할 수 있는 윤리적 딜레마, 지켜야 할 윤리적 가치, 그리고 윤리적 소양에 대한 논의	
2	Cyber-bullying	키워드	cyber-bullying, unethical, student, behavior, internet	46.144
		주요내용	온라인에서 발생하는 폭력, 비윤리적 행위 등에 대한 실태와 영향요인 등에 대한 탐색에 대한 논의	
3	Media and Reliability	키워드	algorithm, web, bias, media, use, design	41.278
		주요내용	인터넷이 가지고 온 사회적 변화, 인터넷에서 공유되는 정보의 공정함, 정보 제공 알고리즘의 투명도, 정치적 영향력, 인종차별 등과 같은 사회적 영향력에 대한 논의	
4	Security and Privacy	키워드	security, cyber, risk, privacy, protect	34.077
		주요내용	소프트웨어의 취약성, 보안, 그로 인한 사회적 영향과 개인 프라이버시에 미치는 영향력 등에 대한 논의	
5	IT and Social Impact	키워드	social, culture, intellectual, property, perspective, right	34.022
		주요내용	IT가 가지고온 사회적 변화에 대한 전반적인 주제와 사회적 영향력, 글로벌 윤리, 문화적 차이 등에 대한 논의	
6	Teaching and Learning	키워드	education, use, discussion, learning, student	33.046
		주요내용	정보윤리 교육을 위한 교육 내용, 교수 학습 방법, 적용 그리고 효과성에 대한 논의	
7	Property and Behavior	키워드	piracy, software, behavior, attitude, intent, cognition	32.952
		주요내용	소프트웨어 불법복제, 저작권, 그리고 이에 영향을 미치는 요인들에 대한 논의	
8	AI and Social Impact	키워드	artificial intelligence, agent, distribution ethics, network, system	32.187
		주요내용	인공지능이 삶에 미치는 영향과 인공지능 객체에 대한 도덕적/윤리적 가치 그리고 새롭게 등장할 윤리적 문제에 대한 논의	

*출연 문서 확률이 1인 경우 특정 연구 논문 1편에 해당 이슈가 등장할 확률이 100%에 수렴한다는 의미로, 총합은 전체 문서에서 각 이슈가 등장하는 문서의 빈도를 추측할 수 있는 정보를 제공함

주요 키워드 중 상위 20개의 분석 결과를 살펴보면 <표 2>와 같다. 우선 TF 분석 결과를 살펴보면 ‘information’, ‘computer’, ‘use’, ‘technology’, ‘internet’ 순으로 나왔다. 이는 전체 문서에 걸쳐 각 해당 단어가 가장 많이 등장하는 것을 의미한다. 다음으로 각 단어가 등장하는 문서 수를 의미하는 DF 분석 결과를 살펴보면 ‘information’, ‘use’, ‘computer’, ‘issue’, ‘technology’ 순으로 나타났다. TF와 유사한 경향을 보였으나 ‘issue’의 순위가 상위권에 있는 것을 알 수 있다. 반면, TF-IDF를 보면 ‘computer’,

‘digital’, ‘student’, ‘internet’, ‘information’ 순으로 나타났다. TF/DF 분석과는 다르게 ‘computer’가 1순위로 나타났고, ‘digital’, ‘student’, ‘data’가 상위권에 등장한 것을 알 수 있다. 그 외, TF/DF와 비교했을 때 ‘software’, ‘privacy’가 순위가 향상되었다.

4.2 토픽 모델 분석

토픽 모델링을 활용한 주요 이슈 분석 결과를 살펴보면 <표 3>과 같다. 각 이슈별 주요 키워드

는 TF-IDF를 고려해 각 이슈를 대표할 수 있는 주요 키워드들 중심으로 표기하였다. 그리고 사후 분석을 통해 각 이슈를 대표하는 논문과 각 논문에서 다루고 있는 주요 주제를 살펴보면 다음과 같다.

첫째, 출연 문서 총합을 보았을 때 가장 높게 나타난 이슈는 'Professional value'이다. Professional value는 professional, value, development, code, problem 등의 주요 키워드와 이론적 배경에서 분석한 문헌을 고려해 명명하였다. 사후 분석을 통해 'Professional value'가 등장하는 주요 논문을 살펴보면 Gotterbarn(2001), Gorniak(2007), Ferguson, Thornley, Gibb(2007)의 연구가 이를 대표하는 것으로 나타났다 [5][21][22]. 그리고 각 논문에서 다루고 있는 주요 주제는 소프트웨어 개발 과정에서 경험할 수 있는 윤리적 딜레마, 지켜야 할 윤리적 가치, 그리고 윤리적 소양에 대한 논의임을 확인할 수 있다.

둘째, 'Cyber-bullying'이다. 이는 주요 키워드인 cyber-bullying, unethical, student, behavior, internet 등과 이론적 배경에서 분석한 문헌을 기초로 명명하였다. 사후 분석을 통해 'Cyber-bullying'이 등장하는 주요 논문을 살펴보면 Namly, Odabasi & Hatice(2007), Byssey, Fitzpatrick & Raman(2015), Wang, Yang & Yang(2017)의 연구가 이를 대표하는 것으로 나타났다 [23][24][25]. 그리고 각 논문에서 다루고 있는 주요 주제는 온라인에서 발생하는 폭력, 비윤리적 행위 등에 대한 실태와 영향요인 등에 대한 논의임을 확인할 수 있다.

셋째, 'Media and Reliability'이다. 이는 주요 키워드인 algorithm, web, bias, media, use, design 등과 이론적 배경에서 분석한 문헌을 기초로 명명하였다. 사후 분석을 통해 'Media and Reliability'이 등장하는 주요 논문을 살펴보면 Flores, Nadia, Vidales, Plemons(2011), Mittelstadt(2016), Seyvanizad, Kashkuli(2017)의 연구가 이를 대표하는 것으로 나타났다 [26][27][28]. 그리고 각 논문에서 다루고 있는 주요 주제는 인터넷이 가지고 온 사회적 변화, 인터넷에서 공유되는 정보의 공정함, 정보 제공 알고리즘의 투명도, 정치적 영향력, 인종차별 등과 같은

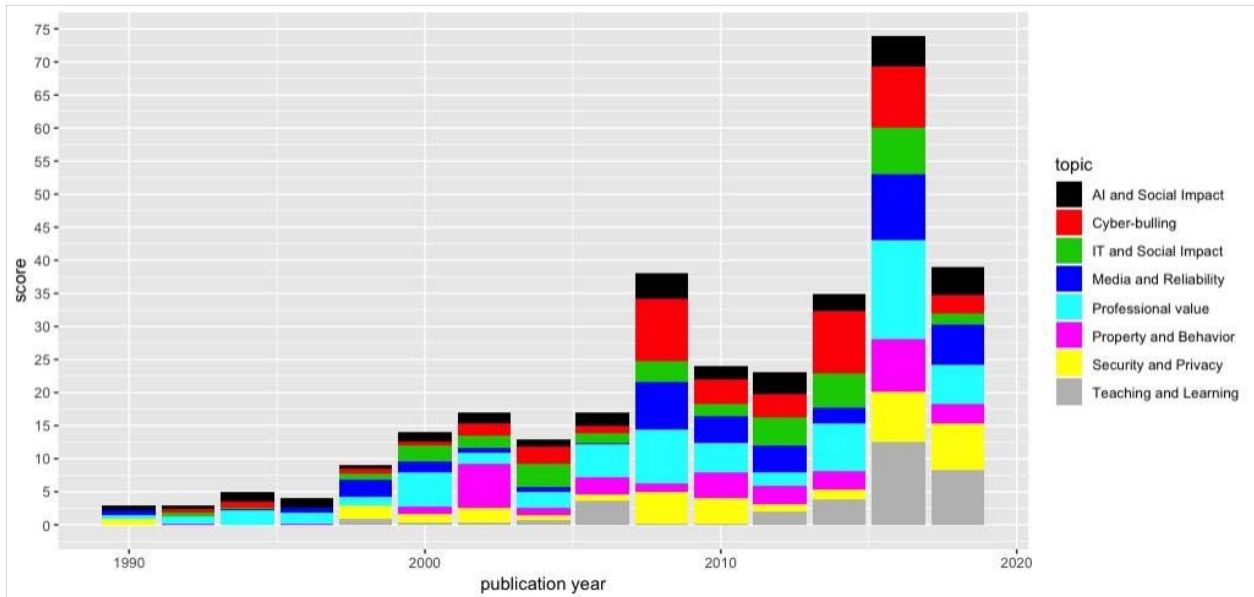
사회적 영향력 등에 대한 윤리적 논의임을 확인할 수 있다.

넷째, 'Security and Privacy'이다. 이는 주요 키워드인 security, cyber, risk, privacy, protect 등과 이론적 배경에서 분석한 문헌을 기초로 명명하였다. 사후 분석을 통해 'Security and Privacy'가 등장하는 주요 논문을 살펴보면 Sung, Kim(2016), Wolf(2016), Wachter(2017)의 연구가 이를 대표하는 것으로 나타났다 [2][3][29]. 그리고 각 논문에서 다루고 있는 주요 주제는 소프트웨어의 취약성, 보안, 그로 인한 사회적 영향과 개인 프라이버시에 미치는 영향력 등에 대한 논의임을 확인할 수 있다.

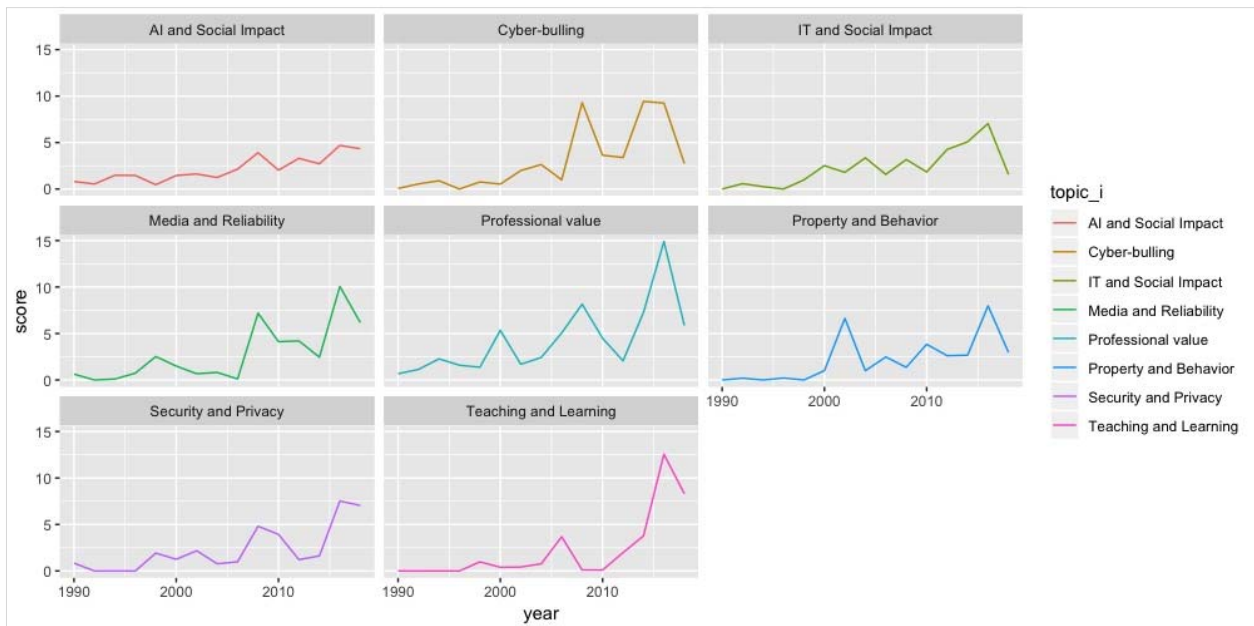
다섯째, 'IT and Social Impact'이다. 이는 주요 키워드인 social, culture, intellectual, property, perspective, right 등과 이론적 배경에서 분석한 문헌을 기초로 명명하였다. 사후 분석을 통해 'IT and Social Impact'가 등장하는 주요 논문을 살펴보면 Vanacker, Heider(2012), Rehg(2015), Healey, Stephens (2017)의 연구가 이를 대표하는 것으로 나타났다 [30][31][32]. 그리고 각 논문에서 다루고 있는 주요 주제는 IT가 가지고 온 사회적 변화에 대한 전반적인 주제와 사회적 영향력, 글로벌 윤리, 문화적 차이 등에 대한 논의임을 확인할 수 있다.

여섯째, 'Teaching and Learning'이다. 이는 주요 키워드인 education, use, discussion, learning, student 등과 이론적 배경에서 분석한 문헌을 기초로 명명하였다. 사후 분석을 통해 'Teaching and Learning'이 등장하는 주요 논문을 살펴보면 Liu, Yang(2012), Choi(2016), Huang, Huang, Wu(2016)의 연구가 이를 대표하는 것으로 나타났다 [6][7][33]. 그리고 각 논문에서 다루고 있는 주요 주제는 정보윤리 교육을 위한 교육 내용, 교수 학습 방법, 적용 그리고 효과성에 대한 논의임을 확인할 수 있다.

일곱째, 'Property and Behavior'이다. 이는 주요 키워드인 piracy, software, behavior, attitude, intent, cognition 등과 이론적 배경에서 분석한 문헌을 기초로 명명하였다. 사후 분석을 통해 'Property and Behavior'가 등장하는 주요 논문을 살펴보면 Tang, Fam(2005), Hashim, Kannan,



[그림 2] 전체 이슈에 대한 연도별 연구 동향



[그림 3] 각 이슈에 대한 연도별 연구 동향

Wegener(2018), Akbulut, Donmez(2018)의 연구가 이를 대표하는 것으로 나타났다[4][34][35]. 그리고 각 논문에서 다루고 있는 주요 주제는 소프트웨어 불법복제, 저작권, 그리고 이에 영향을 미치는 요인들에 대한 논의임을 확인할 수 있다.

마지막은, ‘AI and Social Impact’이다. 이는 주요 키워드인 artificial intelligence, agent, distribution ethics, network, system 등과 이론

적 배경에서 분석한 문헌을 기초로 명명하였다. 사후 분석을 통해 ‘AI and Social Impact’가 등장하는 주요 논문을 살펴보면 Powers(2013), Vassilis(2016), Floridi(2016)의 연구가 이를 대표하는 것으로 나타났다[36][37][38]. 그리고 각 논문에서 다루고 있는 주요 주제는 인공지능이 삶에 미치는 영향과 인공지능 객체에 대한 도덕적·윤리적 가치 그리고 새롭게 등장할 윤리적 문제에 대한 논

의임을 확인할 수 있다.

4.3 주요 이슈별 연구 동향

첫째, 전체 이슈에 대한 연도별 등장 빈도를 살펴보면 [그림 2]와 같다. 전체 이슈에 대한 연도별 등장 빈도는 전체 이슈별 연구 트렌드의 상대적 비교를 파악하기에 수월한 정보를 제공한다. 전체 8개 이슈는 1990년부터 일정 비율을 유지하며 등장하고 있다. 다만, 2010년 전후로 'Professional value'와 'Cyber-bullying'이 상대적으로 높은 비율을 차지하고 있는 것을 볼 수 있다. 이와 함께, 'Teaching and Learning'과 관련된 연구가 타 주제와 비교해 상대적으로 많은 등장 비율을 보여 주고 있는 것을 확인할 수 있다.

둘째, 각 이슈에 대한 연도별 등장 비율을 살펴보면 [그림 3]과 같다. 각 이슈에 대한 연도별 등장 비율은 각 이슈의 연구 트렌드를 파악하기에 수월한 정보를 제공한다. 대부분의 이슈가 2010년 전후로 등장 비율이 상승하는 것을 파악할 수 있다. 특히, 'Professional value', 'Cyber-bullying', 'Media and Reliability', 'Teaching and Learning' 등의 이슈가 높은 상승 폭을 보이는 것을 확인할 수 있다.

5. 논의 및 결론

본 연구는 정보윤리 연구의 주요 이슈와 연구 경향을 살펴보고 이를 토대로 우리나라 정보윤리 교육을 위한 시사점을 제공하기 위해 수행하였다.

이를 위해 Wos의 학술 빅데이터를 활용해 텍스트 마이닝 기반의 키워드 분석, 토픽 모델링 분석 그리고 연구 경향 분석을 하였다. 먼저, 본 연구에서 설정한 세 가지 연구문제를 중심으로 주요 결과를 간략히 살펴보면 다음과 같다.

첫째, TF/DF 모두 'information', 'computer', 'use', 'technology', 'internet' 등의 단어가 높은 빈도를 차지하고 하고 있었다. 다만, TF-IDF 분석 결과를 살펴보면, 'digital', 'student', 'software', 'privacy' 등의 단어가 TF/DF에 비해 순위가 상승된 것을 확인할 수 있었다.

둘째, 토픽 모델링 분석에 기초한 주요 이슈들을

살펴보면, 'Professional value', 'Cyber-bullying', 'Media and Reliability', 'Security and Privacy', 'IT and Social Impact', 'Teaching and Learning', 'Property and Behavior', 'AI and Social Impact' 으로 분석되었다.

셋째, 각 이슈별 연구 경향을 살펴보면, 'Professional value'와 'Cyber-bullying'이 상대적으로 높은 비율을 차지하고 있으며, 특히 2010년을 전후로 'Professional value', 'Cyber-bullying', 'Media and Reliability', 'Teaching and Learning' 이슈가 가파르게 상승하는 것을 알 수 있었다.

이상의 주요 결과가 정보과 정보윤리 교육을 위해 시사하고 있는 사항을 논의하면 다음과 같다.

첫째, 'Professional value'가 강조 되고 있다는 것이다. 'Professional value'는 직역하였을 때 전문가들이 지켜야할 윤리로 볼 수 있으나, 다수의 논문은 특정 직업인으로서의 전문가가 아니라, 소프트웨어 개발 과정을 경험하는 누구라도 갖추어야 하는 윤리적 소양을 다루고 있다[21][22]. 즉, 그동안 우리나라 교육과정에서 다루고 있는 사용자 관점에서 정보윤리가 아니라, 개발자 관점에서 경험할 수 있는 정보윤리 교육에 대한 고민이 필요하다는 시사점을 제시한다.

둘째, 'AI and Social Impact'가 독립된 이슈로 분석되었다는 것이다. 자율주행 자동차, 인공지능 스피커, 워드스미스(WordSmith) 등의 로봇 저널리즘과 같이 AI 기반의 서비스는 더 이상 먼 미래의 이야기가 아니라 우리의 삶속에 깊이 들어왔다고 볼 수 있다. 하지만, AI 발달로 인해 생기는 정책적, 윤리적 공백은 아직 심도 있게 논의 되지 못하고 있는 것이 사실이다. 최근 IEEE, ACM 등이 앞다투어 AI 서비스 개발에 앞서 지켜야할 윤리적 가치를 지침을 만들었으나[39][40], 아직 초·중등 교육을 위한 내용과 교육 방법에 대한 연구는 부족한 상태이다. 이에 향후 AI의 발달로 인해 발생할 수 있는 윤리적 이슈를 보다 면밀히 분석하고, 초·중등 교육에서 다루어야 할 교육내용 및 방법에 대한 연구가 이루어져야 할 것이다.

셋째, 다양한 정보윤리 이슈가 도출되었다는 것이다. 토픽 모델링 결과 8개의 이슈로 유형화될 수 있었다. 이는 2015 개정 정보과 교육과정에 저작

권, 개인정보, 그리고 사이버우리로 매우 간단히 구성되어 있는 것과 대조적임을 알 수 있다[8]. 현재 교육과정의 정보윤리 교육 내용요소가 교육과정의 전체 시수와 타 영역과의 비중을 고려해 불가피하게 구성된 내용이라 할지라도 향후 미래 사회를 살아가야하는 학생들에게 요구되는 정보윤리 소양이 무엇인지에 대해서는 보다 심도 있는 논의와 관련 교육내용의 구성을 위한 노력이 필요하다.

이상의 주요 시사점을 토대로 향후 체계적인 정보과 정보윤리 교육을 위한 두 가지 제언을 하면 다음과 같다.

첫째, 정보 교육의 목표가 '정보윤리 의식을 갖추고, 컴퓨터 과학의 기본 개념과 원리를 토대로 실생활을 문제를 창의적이고 효율적으로 해결하는 것'이라면[8], 정보윤리의 목표 또한, 문제를 해결하는데 있어 요구되는 윤리적 경험에 관한 것이어야 한다. 즉, Maker 또는 개발자 관점에서 경험할 수 있는 윤리적 이슈를 토대로 초·중등 수준에 맞는 교육내용으로 구성해야 할 것이다.

둘째, 정보 윤리의 특성 상 새로운 기술의 등장으로 인해 발생하는 윤리적 공백을 최대한 적절한 시점에 다룰 수 있어야 한다. 본 연구의 결과에서도 2010년 전후로 'Professional value', 'Cyber-bullying' 관련 연구가 대폭 증가하는 것으로 나타났으며, 이는 2007년 스마트폰의 등장과 함께, 전 세계적으로 Computational thinking 기반의 정보교육의 중요성이 대두되는 시점과 무관하다고 할 수 없을 것이다. 이를 고려했을 때 현 시점에서는 AI, BigData, Cloud, IoT 등과 같이 새롭게 등장하는 기술과 이를 통해 발생하는 윤리적 이슈를 보다 적절한 시점에 가르칠 수 있도록 교육과정 개정의 주기를 보다 빠르게 가져갈 필요가 있다.

그럼에도 불구하고, 여러 현실적 여건을 고려해 물리적으로 빠른 호흡의 교육과정 개정이 어려울 수 있다. 이럴 경우, 정보윤리 교육이 단순히 올바른 것과 나쁜 것을 다루는 역기능에 대한 교육이 아니라, 새로운 기술이 사회에 어떠한 영향을 미칠 것인지에 대한 토론과 논의를 토대로 합리적이고 윤리적인 사고가 이루어질 수 있는 교육 중심으로 구성되어야 할 것이다.

본 연구는 정보윤리 교육을 위한 이슈를 탐색하기

위해 방대한 양의 논문과 텍스트 마이닝을 활용한 과학적 접근을 시도하였다는 것에 그 의의가 있다. 하지만, 그 대상을 WoS에서 제공하는 전문 학술 저널만을 대상으로 하였다는 한계점이 존재한다. 이를 보완하기 위해서는 향후 소셜 빅데이터 등의 분석을 통해 보다 다각적인 방식으로 접근할 필요가 있다. 또한 체계적인 정보과 정보윤리 교육을 위해서는 정보윤리 교육의 목적, 방법, 평가 등에 대한 연구와 논의가 지속적으로 이루어져야 할 것이다.

참 고 문 헌

- [1] WEF (2016). *World Economic Forum Annual Meeting 2016 : Mastering the Fourth Industrial Revolution*. Davos-Klosters, Switzerland 20-23 January.
- [2] Sandra, W. (2017). Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR. *Computer Law & Security Review*, 34(3), 436-449.
- [3] Sung, W., Kim, D. (2016). Effects of Internet Users' Perception Regarding the Risks and Benefits of the Internet on Cyberspace Trust. *International Journal of Security and Its Applications*, 10(7), 71-81.
- [4] Hashim, M. G., Kannan, K. N., Wegener, D. T. (2018). *Central Role of Moral Obligations in Determining Intentions to Engage in Digital Piracy*. *Journal of Management Information Systems*, 35(3), 934-963.
- [5] Gorniak, K., (2007). From computer ethics to the ethics of global ICT society. *Library HI Tech*, 25(1), 47-57.
- [6] Choi, M. (2016). A Concept Analysis of Digital Citizenship for Democratic Citizenship Education in the Internet Age. *Theory and Research in Social Education*, 44(4), 565-607.

- [7] Huang, S., Huang, Y., Wu, T. (2016). Problem-based learning effectiveness on micro-blog and blog for students: a case study. *Interactive Learning Environments*, 24(6). 1334-1354.
- [8] 교육부 (2015). 실과(기술·가정)/정보과 교육과정. 교육부 고시 제2015-74호 별책 10. 교육부.
- [9] 교육부 (2015). 도덕과 교육과정. 교육부 고시 제2015-74호 별책 6. 교육부.
- [10] 민용성 (2011). 교과 내용의 적합성 준거 개발. *학습자중심교과교육연구*, 11(3), 19-41.
- [11] Bynum, T. (2018). *Computer and Information Ethics. The Stanford Encyclopedia of Philosophy (Summer 2018 Edition)*. Edward N. Zalta (ed.).
- [12] Wiener, N. (1950). *The Human Use of Human Beings: Cybernetics and Society*. Boston: Houghton Mifflin; Second Edition Revised. NY: Doubleday Anchor 1954.
- [13] Maner, W. (1980). *Starter Kit in Computer Ethics, Hyde Park*. NY: Helvetia Press and the National Information and Resource Center for Teaching Philosophy.
- [14] Johnson, D. (1985). *Computer Ethics*. First Edition, Englewood Cliffs. NJ: Prentice-Hall.
- [15] Moor, J. (1985). What Is Computer Ethics?. *Metaphilosophy*, 16(4), 266-75.
- [16] 임상수 (2003). 정보윤리교육의 방법론 모색. *윤리연구*, 54, 261-291.
- [17] 김한성 · 정혜진 · 이원규 (2010). 정보교과의 정보윤리 교육내용에 대한 고찰. *학습자중심교과교육연구*, 10(2), 95-119.
- [18] 교육과학기술부 (2011). 실과(기술·가정) 교육과정. 교육과학기술부 고시 제 2011-361 호 별책10. 교육과학기술부.
- [19] 교육과학기술부 (2011). 도덕과 교육과정. 교육과학기술부 고시 제 2011-361호 [별책 6]. 교육과학기술부.
- [20] 백영민 (2017). R를 이용한 텍스트 마이닝. 경기도 : 한올아카데미.
- [21] Gotterbarn, D. (2001). Informatics and Professional Responsibility. *Science and Engineering Ethics*. 7(2). 221-230.
- [22] Ferguson, S., Thornley, C., Gibb, F. (2016). Beyond Codes of Ethics: How Library and Information Professionals Navigate Ethical Dilemmas in A Complex and Dynamic Information Environment. *International Journal of Information Management*. 36(4), 543-556.
- [23] Namlu, A. G., Odabasi, H. F. (2007). Unethical Computer Using Behavior Scale: A Study of Reliability and Validity on Turkish University Students. *Computers & Education*, 48(2), 205-215.
- [24] Bussey, K., Fitzpatric, S., Raman, A. (2015). The Role of Moral Disengagement and Self-Efficacy in Cyberbullying. *Journal of School Violence*, 14(1), 30-46.
- [25] Wang, X., Yang, L., Yang, J. (2017). Trait Anger and Cyberbullying among Young Adults: A Moderated Mediation Model of Moral Disengagement and Moral Identity. *Computers in Human Behavior*, 73, 519-526.
- [26] Flores, Y., Nadia, Y., Vidales, G., Plemons, A. (2011). The Latino Cyber-Moral Panic Process in the United States. *Information Communication & Society*, 14(4), 568-589.
- [27] Mittelstadt, B. (2016). Auditing for Transparency in Content Personalization Systems. *International Journal of Communication*, 10, 4991-5002.
- [28] Seyvanizad, J., Kashkuli, M. H. (2017). General Principles of International Law concerning Law of Media. *Helix*, 7(4), 1606-1613.
- [29] Wolf, M., Fresco, N. (2016). Ethics of the Software Vulnerabilities and Exploits Market. *Information Society*, 32(4), 269-279.
- [30] Vanacker, B., Heider, D. (2012). Ethical Harm in Virtual Communities

convergence. *The International Journal of Research into New Media Technologies*, 18(1), 71-84.

- [31] Rehg, W. (2015). Discourse Ethics for Computer Ethics: a Heuristic for Engaged Dialogical Reflection. *Ethics and Information Technology*, 17(1), 27-39.
- [32] Healey, K., Stephens, N. (2017). Augmenting justice: Google glass, body cameras, and the Politics of Wearable Technology. *Journal of Information Communication & Ethics in Society*, 15(4), 370-384.
- [33] Liu, C. J., Yang, S. C. (2012). Applying the Practical Inquiry Model to Investigate the Quality of Students' Online Discourse in an Information Ethics Course Based on Bloom's Teaching Goal and Bird's 3C model. *Computers & Education*, 59(2), 466-480.
- [34] Tang, J., Fam, C. (2005). The Effect of Interpersonal Influence on Softlifting Intention and Behaviour. *Journal of Business Ethics*, 56(2), 149-161.
- [35] Akbulut, Y., Donmez, O. (2018). Predictors of Digital Piracy among Turkish Undergraduate Students. *Telematics and Informatics*, 35(5), 1324-1334.
- [36] Powers, T. M. (2013). On the Moral Agency of Computers. *An International Review of Philosophy*, 32(2), 227-236.
- [37] Vassilis, G. (2016). Singularitarianism and Schizophrenia. *AI & Society*, 32(4), 573-590.
- [38] Floridi, L. (2016). Faultless Responsibility: on the Nature and Allocation of Moral Responsibility for Distributed Moral Actions. *Philosophical Transactions of the Royal Society : Mathematical, Physical and Engineering Sciences*, 374(2083), 20160112.
- [39] IEEE (2019). *Ethically Aligned Design : A Vision for Prioritizing Human Well-being*

with Autonomous and Intelligent Systems, First Edition. IEEE.

- [40] ACM (2018), *ACM Code of Ethics and Professional Conduct.* ACM.



김 한 성

2005 공주대학교 사범대학
컴퓨터교육과(이학사)
2007 공주대학교 일반대학원
컴퓨터과학과(이학석사)

2014 고려대학교 일반대학원
컴퓨터교육학과(이학박사)

2009 ~ 2010 홋카이도 대학 방문연구원

2011 ~ 2014 고려대학교, 덕성여대, 춘천교대 강사

2013 ~ 현재 한국교육학술정보원 선임연구원

관심분야: 정보교육, 정보윤리, 교육빅데이터,
디지털리터러시

E-Mail: hansung.kim@keris.or.kr