

TF-IDF를 활용한 초·중등학교 저작권 교육 자료의 내용 분석

Analysis of Copyright Education Material Contents in Elementary and Secondary Schools Using TF-IDF

이경민[†] · 김자미^{††} · 이원규^{†††}

KyeongMin Lee[†] · JaMee Kim^{††} · WonGyu Lee^{†††}

요 약

교육 환경 내 디지털 자료와 기기 활용이 증대되면서, 디지털 저작권 교육에 대한 필요가 높아지고 있다. 본 연구는 초·중등학교 저작권 교육의 현황을 파악하고, 향후 디지털 저작권 교육의 방향성을 제시하는 데 목적을 두었다. 목적 달성을 위해 한국저작권위원회의 초·중등 표준 교재 및 강의자료에서 추출한 텍스트 데이터에 대하여 TF-IDF 및 동시출현 빈도를 분석하였다. 분석 결과, 현재 교재와 강의자료에서는 디지털 저작권 관련 단어가 높은 빈도로 출현하지 않았으며, 초·중·고등학교의 학교급별 교재 및 강의자료에서 서로 유사한 단어의 출현 빈도가 높았다. 이를 통해 본 연구는 다음과 같은 시사점을 도출하였다. 첫째, 생성형 AI를 포함하여 다양한 디지털 저작권이 강조되고 있으며, 2015 개정 교육과정과 2022 개정 교육과정에서 인공지능 및 디지털 관련 내용이 강화되고 있다. 저작권 교육에서도 이를 반영한 내용 체계, 교재, 강의자료를 개발할 필요가 있다. 둘째, 학교급별 교재 및 강의자료에서 사용하는 단어를 학습자의 발달 단계 및 수준에 맞게 적용할 필요가 있다. 본 연구는 기관의 저작권 교육과 국가 교육과정을 함께 고려하여, 현재 실제로 활용되는 교재 및 강의자료의 내용을 분석함으로써 향후 디지털 저작권 교육에 대한 방향성을 도출하기 위한 기초자료와 시사점을 제공했다는 데 의의가 있다.

주제어: 저작권, 디지털 저작권, 초중등학교 저작권 교육, 텍스트 마이닝, TF-IDF, 동시출현 빈도

ABSTRACT

With the increasing use of digital materials and devices in educational environments, the need for digital copyright education is growing. This study aimed to assess the current status of copyright education in elementary and secondary schools and to suggest directions for future digital copyright education. To achieve this goal, text data extracted from the Korea Copyright Commission's standard textbooks and lecture materials for elementary and secondary schools were analyzed using TF-IDF and co-occurrence frequency. The analysis revealed that digital copyright-related words did not appear frequently in the current textbooks and lecture materials, and similar words were often found in textbooks and lecture materials for elementary, middle, and high school levels. Based on this, the study drew the following implications. Firstly, amidst the emphasis on various digital copyrights including generative AI, artificial intelligence and digital-related content have also been strengthened in the revised curriculum of 2015 and 2022. Therefore, it is necessary to develop content systems, textbooks, and lecture materials for copyright education that reflect this. Secondly, there is a need to apply the words used in textbooks and lecture materials for each grade level according to the learners' developmental stages and levels. This study provides basic data and implications for the direction of future digital copyright education by analyzing the content of textbooks and lecture materials currently being used, taking into account institutional copyright education and national curriculum.

Keywords: Copyright, Digital Copyright, Copyright education in elementary and secondary schools, Text mining, TF-IDF, Co-occurrence frequency

[†]정 회 원: 고려대학교 컴퓨터학과 박사과정

^{††}중신회원: 고려대학교 교육대학원 컴퓨터교육전공 부교수

^{†††}중신회원: 고려대학교 대학원 컴퓨터학과 교수(교신저자)

논문투고: 2024년 01월 12일, 심사완료: 2024년 04월 03일, 게재확정: 2024년 04월 10일

1. 서론

디지털 대전환에 따라, 기술의 발전이 사회 문화적 인 변화와 더불어 학교 교육의 패러다임을 바뀌고 있다[1]. 2022 개정 교육과정에서는 미래를 살아가기 위한 기초 소양으로 디지털 리터러시를 강조하고 있다[2-4]. 디지털 역량 함양은 디지털 기술의 활용부터 창작에 이르는 다양한 능력을 포괄하는데, 디지털을 통한 창작은 큰 비용과 복잡한 작업 없이도 완벽한 복제물의 생산을 가능하게 한다. 디지털 환경에서는 저작권 침해의 가능성이 높으므로, 디지털 저작권에 대해서는 기존의 저작권과 다른 방식으로 접근할 필요성이 있다[5]. 타인의 창작물에 대한 청소년들의 콘텐츠 보호 및 의식 수준은 온라인과 모바일상에서의 콘텐츠 소비 수준과 속도에 비해 저조한 것으로 보고되었다[6]. 저작권 교육을 받은 학생은 그렇지 않은 학생보다 저작권에 대한 지식, 인지, 가치 점수가 높았고, 실천적 태도에서도 긍정적인 차이가 있다는 연구가 있으나, 디지털 사회의 변화를 반영한 것은 아니다[7].

실제 학교 현장에서 이루어지는 저작권 교육은 한국저작권위원회에서 각 학교를 찾아가서 진행하는 ‘찾아가는 저작권 교육’ 이 주를 이루고 있다. 한국저작권위원회는 매년 10,000회 이상의 ‘찾아가는 저작권 교육’ 프로그램을 진행하고 있으며, 이 중 70% 이상이 초·중등 학생 대상 교육으로 분석된다[8-9]. 한국저작권위원회의 초·중등학교 학생 대상 교육프로그램 역시 디지털 저작권 관련 내용이 미비한 편이지만, 일반적인 저작권 교육 현황을 분석하는 데는 용이할 것으로 판단되었다. 현재의 저작권 교육 내용을 파악하는 것은 향후 디지털 저작권 교육의 방향성을 제시하는 데도 의미가 있을 것으로 보인다.

따라서 본 연구는 한국저작권위원회에서 진행하는 ‘찾아가는 저작권 교육’ 중 초·중등학교 학생을 대상으로 하는 교육에서 활용되는 교재 및 강의자료를 분석함으로써, 현재의 저작권 교육 내용을 파악하고 향후 저작권 교육의 방향성을 제시하는 데 기초자료를 제공하고자 한다.

2. 관련 연구

2.1 초·중등학교의 저작권 교육 관련 연구 및 현황

2000년 이후 진행된 초·중등 저작권 교육 관련 연구는 교사 혹은 학생들의 저작권 관련 인식 및 지식수준을 조사하는 내용의 연구가 주를 이루고 있다. 저작권에 대한 초등학생 인지 및 태도 연구[10], 초등 교사의 저작권 지식수준 분석을 통한 교육정보 이용 활성화 방안 연구[11], 정보 영재 지도교사 저작권 인식 조사 및 저작권 교육프로그램 개발 연구[12] 등이 그것이다.

관련 연구를 분석한 바에 따르면, 저작권 교육이 교과 내용에 반영된 비중은 미미한 것으로 나타났다. 2007 개정 교육과정 내 초등 도덕과, 중학교 정보에서는 저작권 관련 내용이 충분히 다루어지지 않았고, 이를 더 확대해야 할 것으로 보고되었다[13-14]. 하지만 2015 개정 교육과정에서 성취기준 및 학습요소, 교과서에 저작권 내용이 반영된 정도를 살펴보았을 때, 그 비중이 미미한 것으로 보고되었다[6][15]. 국가 수준의 교육과정 차원에서 저작권 교육 내용을 포함하지 않거나 그 수준이 미흡하였으며, 교과별 성취기준 내 저작권 교육 반영의 필요성이 제시되었다[15]. 2022 개정 교육과정 문서의 교과별 저작권 내용 분석 연구에서도 대부분 교과의 성취기준에서 직접적인 교육 내용으로 제시되기보다는 교사가 학생들의 활동이나 과제를 평가할 때 타인의 저작권을 침해하지 않도록 안내하는 간접적 방식으로 저작권 교육이 구성되어 있다[16].

국내 저작권 전문 기관인 한국저작권위원회는 저작권 교육 활성화를 목표로 다양한 사업과 교육프로그램을 지속적으로 개발 및 운영하고 있다[6]. 한국저작권위원회는 다양한 저작권 강의에서 실용적으로 활용될 수 있도록 표준 교재 및 강의자료를 제작하였다[17-24]. 저작권 교육 경험이 많은 교사가 강의자료를 집필하였고, 저작권 전문가 다수의 검토를 통해 자료를 수정 및 보완하였다. 현재 이 자료는 ‘저작권 표준 교안 4종’ 이라는 이름으로 배포되고 있으며 저작권 강사들이 활발히 활용하고 있다. 또한 중등용 표준 교재 개발 연구 및 초등용 표준 교재 개발 연구를 통해 저작권 교육 내용 체계를 구안하여, 이를 바탕으로 한 표준 교재를 2종 개발하였다[17-18]. 학교 현장으로 찾아가 진행하는 초·중등학생 대상 ‘찾아가는 저작권 교육’ 은 2023년 기준 3년간 매년 7,000회 이상 실시되었다[11-12]. 즉, 초·중등학교 현장에서 이루어지는 저작권 교육은 교과 교육 내에 포함된 경우보다 기관에서 운영하는 교육이 큰 비중을 이루고 있으므로, 기관의 저작권 교육 내용을 분석할 필요가 있

다.

따라서 본 연구에서는 현재 초·중등학교의 저작권 교육 내용 및 방법 등에 대한 논의를 구체화하기 위해서 실질적으로 현장에서 활발히 운영되고 있는 한국 저작권위원회의 교육 내용을 분석하였다.

2.2 TF-IDF 분석

텍스트 마이닝은 비정형 텍스트 데이터에서 의미 있는 정보를 목적에 맞게 추출하는 기법이다[1, 25, 26]. 여러 텍스트 마이닝 기법 중에서도 가장 널리 사용되는 빈도분석 기법을 활용하면, 문서에 출현한 용어의 빈도를 다양한 관점으로 분석할 수 있다[27]. 문서와 단어의 연관성을 분석하는 키워드 빈도분석으로 TF(Term Frequency, 단어 빈도) 기법과 TF-IDF(Term Frequency-Inverse Document Frequency, 단어-역문서 빈도) 기법이 많이 활용된다.

TF 기법은 어떤 문서에서 특정 단어가 얼마나 출현했는지 산출하여, 문서 내 단어의 비중을 파악하는 기법이다[28]. t 는 특정 단어, d 는 특정 문서라고 할 때, TF 값을 도출하는 식은 다음과 같다.

$$TF(t,d) = \frac{t \text{가 } d \text{ 내에서 나타난 횟수}}{d \text{ 내의 모든 단어의 수}}$$

그러나 TF 분석의 경우, 일반적으로 자주 쓰이는 단어는 어느 문서에서든 높은 빈도를 보이기 때문에, 특정 문서에서 중요한 의미가 있는 주제어가 이에 비해 낮은 빈도를 보이는 한계가 있다[27]. 즉, TF 분석의 결과는 단어의 문서 내 단순 출현 빈도이므로, 특정 문서와 주제어의 연관성을 명확히 나타내기 어렵다. 연관성이 낮음에도 불구하고 자주 출현하는 특정 단어들에 제한을 주는 방식으로, TF 분석이 가지는 한계를 해결할 수 있다. 특정 단어가 출현한 문서의 개수로 전체 문서의 개수를 나누어주면, 역문서 빈도인 IDF 값을 산출할 수 있다. 즉, IDF 값은 여러 문서에 공통적으로 자주 출현하는 단어에는 특정 문서에서 특별히 자주 출현하는 단어보다 적은 가중치를 부여하여, 상대적 연관성을 산출한다[28]. 전체 문서의 집합을 D 라고 할 때, IDF 값의 산출식은 다음과 같다.

$$IDF(t,D) = \log \frac{\text{전체 문서의 수}(D)}{\text{단어 } t \text{를 포함하는 문서의 수}}$$

여러 문서에서 특정 단어가 많이 출현할 때 TF 값은 커지지만, 특정 문서 대비 전체 문서에서 특정 단

어가 많이 출현할수록 IDF 값은 낮아지기 때문에, TF 값과 IDF 값을 모두 고려하여야 특정 문서와 특정 단어 간의 연관성을 명확히 파악할 수 있다. TF-IDF 값은 TF와 IDF 값을 곱한 것으로, 그 식은 다음과 같다.

$$TF-IDF(t,d,D) = TF(t,d) \times IDF(t,D)$$

본 연구에서는 TF-IDF 기법을 활용하여, 한국저작권위원회 표준 교재 2종[19-20]과 강의자료 4종[21-24]을 분석하고, 각 문서가 어떤 단어와 연관성이 높은지 분석하였다.

3. 연구 방법

3.1 연구 절차

텍스트 분석은 주로 정성 분석 방법, 내용 분석 방법, 언어분석 방법, 언어 네트워크 분석 방법으로 구분할 수 있다. 전통적 정성 분석 방법을 통해서 연구자의 관점에 따라 결과가 좌우될 수 있지만, 내용 분석은 이러한 오류에 대한 대안적 방법으로 텍스트를 일정 분류 범주에 따라 코딩하고 그 결과를 계량적으로 취합하는 방법이다. 언어 네트워크 분석은 기존 언어 텍스트의 분석 방법들을 통합하면서, 개념 간의 의미적 관계에 나타나는 새로운 특성을 파악하는 방법으로 볼 수 있다[29].

본 연구의 구체적인 연구 절차는 Figure 1과 같다.

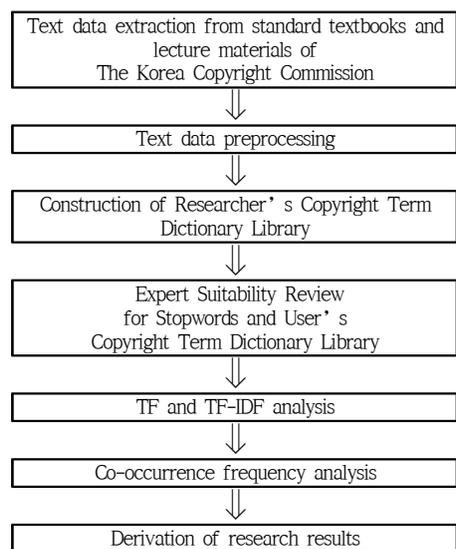


Figure 1. Research Procedure

본 연구는 한국저작권위원회에서 운영하는 저작권 교육의 표준 교재와 강의자료를 분석함으로써, 실제 현장에서 운영되고 있는 초·중등학교 저작권 교육의 내용을 분석하는 것에 목적을 두었다. 교재와 강의자료 내의 텍스트 데이터 분석을 위하여, TF-IDF를 통한 내용 분석과 주요 단어 간 동시출현 빈도 기반의 언어 네트워크 분석을 진행하였다.

첫째, 한국저작권위원회에서 개발한 저작권 교육 표준 교재 및 강의자료의 텍스트를 추출하였다. 둘째, 해당 텍스트 데이터에 대한 전처리를 진행하였다. 셋째, 한국저작권위원회에서 제공하는 저작권 용어 사전을 기준으로 저작권 용어 라이브러리를 구성하였다. 넷째, 한국어 형태소 분석기 komoran과 더불어, 추가로 구성된 저작권 용어 라이브러리를 기준으로 표준 교재와 강의자료의 텍스트 데이터에 대한 TF 분석 및 TF-IDF 분석을 진행하였다. 다섯째, TF-IDF 분석을 통한 형태소 출현 빈도를 기반으로, 한 문장 내 두 단어 간 동시출현 빈도 및 언어 네트워크를 분석하여 연구 결과를 도출하였다.

3.2 분석 방법

본 연구에서는 Google사의 Colaboratory를 통해 브라우저 내에서 Python 스크립트를 작성하여, 한국저작권위원회 표준 교재 2종과 강의자료 4종의 TF-IDF 값을 산출하고 그 결과값을 기반으로 동시출현 빈도를 분석하였다. 상세한 분석 과정은 다음과 같다.

첫째, 한국저작권위원회에서 개발한 표준 교재 2종과 강의자료 4종의 각각의 문서에서 텍스트 데이터를 추출하여 수집하였다.

둘째, 수집한 텍스트 데이터에 대한 전처리 과정을 수행하였다. 본 연구의 분석 대상은 저작권 교육자료 내 텍스트 데이터이므로, 저작권과 관련된 전문 용어의 의미를 오류 없이 반영하여 분석하는 데 중점을 두었다. 구체적인 데이터 전처리 과정은 다음과 같다. Konlpy 패키지를 설치하여, 형태소 분석기 komoran을 통해 텍스트 데이터를 형태소로 분리하고, 토큰화하였다. 토큰화된 결과에 대하여 두 음절 이상의 일반 명사에 해당하는 단어들만 추출하고, 구두점이나 불필요한 특수 문자 및 불용어를 제거하였다. 이때 학계 전문가 3인의 의견을 수렴하여 저작권 교육 내용 분석을 위해 필요하다고 보기 어려운 단어들을 불용어에 포함하였다. 불용어로 추가로 포함된 단어들은 총 16개로, ‘관련’, ‘해당’, ‘동안’, ‘별도’와

같은 단어들에 이에 해당한다.

셋째, 저작권 용어 사전 라이브러리를 구성하였다. 두 어절의 단어인 경우, 저작권과 관련하여 자주 사용되는 용어임에도 komoran 형태소 분석기 라이브러리에 단어로 등록되지 않아, 다른 단어로 분리되는 경우가 발견되었다. 예를 들면 “공정 이용”, “간접 인용”, “저작권 침해”, “저작권 보호” 등이 이에 해당한다. 또한 komoran 형태소 분석기는 한 어절 안에 있더라도 이를 구분하여 각각의 형태소를 분리하기 때문에, 저작권 전문 용어의 의미를 제대로 분석하지 못하는 경우가 발견되었다. 예를 들어 komoran 형태소 분석기는 “저작인접권”을 “저작”, “인접”, “권”이라는 각각의 형태소를 분리하는데, 이는 하나의 용어인 “저작인접권”으로 처리하여야 한다. 현재 다른 영역과 달리 저작권은 관련 용어를 하나의 코퍼스(corpus)로 처리할 수 있도록 정리한 패키지가 없기 때문에 이러한 문제가 발생하였다. 이에 따라 한국저작권위원회 저작권 용어 사전을 기준으로 447개의 단어를 ‘사용자 저작권 용어 사전’ 라이브러리로 만들어 포함하고, 그 외에도 본 연구에서 필요하다고 판단된 ‘저작권 보호’, ‘저작권 등록’, ‘저작권 침해’와 같은 단어 9개를 라이브러리에 추가하여 분석하였다[30]. 이때 사용자 사전에 의한 분석을 먼저 적용하지 않을 경우, komoran 형태소 분석기에 의해 단어가 분리되는 것을 방지할 수 없으므로, 사용자 사전에 의한 분석이 선행될 수 있도록 조치하였다. 사용자 사전에 등록된 456개의 단어 중 텍스트 데이터에서 출현한 단어들은 총 48개이며, 그 예시로는 ‘2차적저작물작성권’, ‘결합저작물’, ‘공공누리’, ‘공공저작물’, ‘저작인접권’ 등이 있다.

넷째, TF-IDF 결과값을 벡터화하고 분석하였다. 위 과정에서 형태소 추출, 불용어 제거, 적합성 검토를 여러 차례 반복하여 텍스트 데이터의 정규화 완결성을 높인 후, 출력된 데이터를 TF-IDF 행렬로 변환하였다. TF-IDF 행렬은 각 문서의 형태소별 빈도를 반영한 벡터값으로 표현하였고, 이에 따라 산출된 TF 값, TF-IDF 값을 내림차순으로 정리하였다.

다섯째, 문장 내 특정 단어 간 동시출현 빈도를 분석하였다. 동시출현 빈도 분석법은 단어 간의 관계를 확인하기 위해, 자주 함께 출현하는 단어의 쌍을 살펴보는 내용 분석 기술의 하나로, 텍스트의 주제를 파악하거나 개념의 유사성 및 근접성을 확인할 수 있다 [31]. 이를 통해 텍스트 내용에 포함된 단어들을 그대로 분석에 활용하는 정성적 분석, 내용 분석 과정에서

발생하는 코딩 오류, 단어들의 단순 빈도 분석에 그치는 오류를 예방할 수 있다. 한 문장 안에서 둘 이상의 키워드들이 함께 사용되고 있다는 것은 이 키워드들이 해당 문장 내에서 주제적으로 서로 밀접한 관계를 가진다는 것을 의미한다[32]. 본 연구에서는 한국어 특성상, 단어 하나만으로 그 용어가 사용된 맥락과 의미를 명확히 파악하기 어렵다는 점을 고려하여 문장을 기준으로 동시출현 빈도를 분석하고 연구결과를 도출하는 데 반영하였다.

4. 연구 결과

표준 교재 2종 간의 TF-IDF 분석 결과와 강의자료 4종 간의 TF-IDF 분석 결과는 Table 1과 같다. Table 1은 TF-IDF 결과값 중 상위 20개를 발췌한 것이다.

Table 1. TF-IDF Results of Textbooks and Lecture Materials

Rank	Between two types of standard textbooks		Between four types of lecture materials			
	Elementary	Secondary	Elementary Lower grade	Elementary Upper grade	Middle school	High school
1	copyrighted work (0.64)	copyrighted work (0.58)	copyright (0.57)	copyrighted work (0.85)	copyrighted work (0.45)	copyrighted work (0.46)
2	copyright (0.40)	copyright (0.46)	copyrighted work (0.40)	copyright (0.28)	copyright (0.44)	copyright (0.31)
3	use (0.26)	use (0.27)	use (0.31)	commonality (0.15)	use (0.33)	use (0.29)
4	person (0.24)	right (0.17)	method (0.17)	author (0.14)	source (0.29)	right (0.22)
5	protection (0.21)	author (0.15)	quiz (0.16)	expression (0.14)	Sharing Square (0.18)	album (0.17)
6	creation (0.17)	protection (0.14)	creation (0.15)	name (0.11)	indication (0.18)	performance(0.14)
7	author (0.13)	person (0.12)	moment (0.14)	illustration (0.09)	permission (0.18)	protection (0.14)
8	right (0.12)	copyright owner(0.11)	illustration (0.14)	originality (0.08)	sharing (0.14)	industry (0.13)
9	method (0.09)	creation (0.09)	friend (0.11)	architecture (0.08)	copyright infringement (0.11)	profit (0.13)
10	self (0.09)	infringement (0.08)	permission (0.11)	art (0.08)	hint (0.11)	purpose (0.12)
11	piece of work(0.09)	copyright infringement (0.08)	person (0.11)	photograph (0.08)	author (0.10)	creation (0.11)
12	expression (0.09)	permission (0.08)	card (0.11)	music (0.08)	protection (0.10)	indication (0.11)
13	quote (0.08)	performance (0.08)	Sharing Square (0.09)	creation (0.08)	photograph (0.09)	lesson (0.10)
14	type (0.08)	neighboring rights (0.08)	imitation (0.09)	person (0.06)	creation (0.08)	profit (0.10)
15	content (0.07)	expression (0.07)	source (0.08)	emotion (0.06)	illustration (0.08)	permission (0.10)
16	writing (0.07)	self (0.07)	box (0.09)	method (0.06)	person (0.08)	person (0.10)

Rank	Between two types of standard textbooks		Between four types of lecture materials			
	Elementary	Secondary	Elementary Lower grade	Elementary Upper grade	Middle school	High school
17	explanation (0.07)	copyright protection (0.07)	game (0.08)	photographer(0.06)	square (0.07)	author (0.10)
18	source (0.07)	music (0.07)	Photograph (0.08)	creator (0.06)	newspaper (0.07)	disclosure (0.09)
19	indication (0.07)	content (0.07)	content (0.08)	thought (0.06)	utilizing (0.07)	broadcast (0.09)
20	permission (0.07)	industry (0.07)	promise (0.08)	classroom (0.05)	choreography(0.06)	production (0.09)

표준 교재 2종과 강의자료 4종 모두에서 디지털 저작권과 관련된 단어는 상위 20개의 값에서 나타나지 않았고, ‘저작물’, ‘저작권’이 가장 높은 TF-IDF 값을 보였다. 그 외 단어들은 상호 간 편차는 크지 않은 데 반해, ‘저작물’, ‘저작권’과 비교해서는 크게 낮은 TF-IDF 값을 나타내었다. 또한 표준 교재 2종 간에 상위 3개의 값은 완전히 일치하였고, 4위부터 10위까지도 거의 유사하였다. 8번째의 ‘저작권자’, 10번째의 ‘침해’라는 두 단어는 초등 교재에서는 상위 결과값으로 나타나지 않았지만, 중등 교재에서 높은 순위의 TF-IDF 값으로 나타났다.

한국어 형태소 특성상 단어 하나만으로는 그 의미를 명확히 파악하기 어렵기 때문에, 교재별로 한 문장 내 두 단어의 동시출현 빈도를 분석하였다. 초등 교재와 중등 교재의 단어 간 동시출현 빈도분석 결과는 Table 2와 같다. Table 2는 각 교재별 동시출현 빈도 상위 결과값 20개를 발췌한 것이다.

Table 2. Co-occurrence Frequency of Standard Textbooks

Elementary Textbook	Frequency	Secondary Textbook	Frequency
Copyrighted Work - Utilization	28	Copyright - Infringement	105
Copyright - Protection	23	Copyright - Protection	102
Utilization - Permission	11	Copyrighted Work - Utilization	75
Copyrighted Work - Type	11	Utilization - Permission	39
Expression - Creative Work	11	Copyright - Personal Intent	37
Legal - Protection	10	Copyrighted Work - Utilization	25
Copyright - Registration	10	Copyrighted Work - Authoring	23
Copyrighted Work - Protection	8	Language - Copyrighted Work	20
Free - Use	6	Broadcasting - Business Owner	19
Computer - Program	6	Public - Transmission	18
Permission - Indication	6	Copyright - Industry	18
Agreement - Joining	6	Computer - Program	18
Protection	5	Copyrighted Work -	17

Elementary Textbook	Frequency	Secondary Textbook	Frequency
- Duration		Creation	
Copyright - Personal Intent	5	Art - Copyrighted Work	16
Copyright - Restriction	5	Video - Copyrighted Work	16
Copyright - Infringement	5	Copyrighted Work - Copyrighted Work	16
Copyrighted Work - Creation	5	Expression - Creative Work	16
Copyrighted Work - Author	5	Thought - Emotion	15
Author - Copyrighted Work	5	Human - Thought	14
Source - Indication	5	Copyright - Registration	14

분석 결과, 두 교재에 공통적으로 높은 빈도로 나타난 단어인 ‘이용’, ‘보호’는 ‘저작물 이용’, ‘저작권 보호’, ‘저작물 보호’, ‘자유 이용’ 등과 관련된 키워드로 나타났다. ‘침해’라는 단어는 ‘저작권 침해’로 자주 출현했음을 알 수 있다. 이는 한국저작권위원회의 ‘청소년 저작권 교육내용 가이드’에서 제시하고 있는 학교급별 중점 내용과도 일치하는 결과다[33]. 청소년 저작권 교육내용 가이드에서는 저작권 침해의 유형 및 사례, 표절에 대해 초등학생보다는 중학생, 고등학생 대상 교육에서 더욱 강조해야 한다고 표시하고 있다.

또한 저작권 산업, 진로, 직업 연관성과 관련된 내용은 다른 학교급에 비해 고등학생을 대상으로 할 때 가장 강조해야 함을 나타내고 있다. 이는 고등학생 대상 강의자료 TF-IDF 분석 결과값과도 일치하는 결과다. Table 1에서 나타난 것처럼 고등학생 대상 강의자료는 다른 학교급의 강의자료에 비해 상대적으로 “산업”, “수익”, “영리” 등과 같은 단어들의 TF-IDF 값이 더 높은 것으로 나타났는데, 해당 단어들어 어떤 맥락에서 출현한 것인지 분석하기 위해 동시 출현 빈도를 분석해 본 결과는 Table 3와 같다.

Table 3. Co-occurrence Frequency of 'Industry' and 'Revenue'

With Industry	Frequency	With Revenue	Frequency
Property Rights	5	Allocation	4
Copyright	2	Character	2
Development	2	Naver	1
Design	1	Advertisement	1
Economy	1	Copyright Fee	1
...		Author	1
		Copyright	1
		Sound Source	1
		Distribution	1
		Business Owner	1
		Rights Holder	1
		...	

‘산업’과 관련하여서는 ‘산업재산권’뿐 아니라 ‘저작권 산업의 발전’에 관련된 단어들도 함께 출현한 것을 볼 수 있고, ‘수익’ 역시 ‘작가’, ‘광고’, ‘상품’, ‘저작권료’, ‘음원’과 같이 실제 저작권을 기반으로 한 영리활동에 관련된 단어들과 동시 출현하고 있음을 알 수 있다. 다른 학교급 학생들을 대상으로 한 강의자료에 비해 ‘산업’, ‘수익’, ‘영리’ 등의 단어들도 비교적 많이 출현하였다. 이는 저작권이라는 개념이 저작물이나 콘텐츠를 이용하는 데 있어서 규제와 제한의 역할만을 하지 않으며, 저작권을 통해 저작자의 권리를 지켜주고 이를 기반으로 경제적 수익 창출이나 관련 산업 발전에 기여할 수 있음을 보여주고자 한 것으로 해석할 수 있다.

5. 결론 및 논의

본 연구는 국내에서 활발히 운영되고 있는 한국저작권위원회에서 개발한 교재를 대상으로, 교육 내용의 구성을 분석하였다. 연구 결과를 바탕으로 도출한 결론은 다음과 같다.

첫째, 디지털 기술의 발전에 따라 변화하는 사회와 교육 환경에 맞추어, 저작권 교육에 디지털 관련 내용을 중점적으로 반영할 필요가 있다. 교재 및 강의자료를 대상으로 TF 및 TF-IDF를 분석한 결과에서는 디지털 저작권과 관련된 단어가 높은 비중으로 나타나지 않았다. 컴퓨터프로그램이나 데이터베이스와 같이 저작권법에 명시된 내용 외에도 디지털과 관련된 저작권 문제는 계속해서 발생하고 있다. 특히 생성형 AI에 의해 생성되는 저작물에 대한 저작권 이슈는 계속해서 논의되고 있다[34]. 기술의 발전과 사회 변화에 맞추어 교육 환경도 변화하고 있으며, 학생들이 학교 안팎으로 접하는 디지털 환경과 디지털 저작물이 증가하고 있음을 고려하여, 디지털 저작권 관련 내용을 반영하여 저작권 교육 내용을 구성할 필요가 있다.

둘째, 국가 교육과정의 변화를 저작권 교육 내용 체계 및 내용 구성에 반영할 필요가 있다. 2015 개정 교육과정에서는 SW 교육 활성화를 위해 초등학교에서 17시간, 중학교는 34시간 동안 SW 교육을 필수화하였다[35]. 이는 다양한 문제를 창의적이고 효율적으로 해결하는 컴퓨팅 사고력을 갖춘 융합 인재를 양성하기 위한 것으로, 국가 교육과정의 방향성을 확인할 수 있는 부분이다[35-37]. 하지만 2021년도에 발표된 한

국저작권위원회의 청소년 저작권 교육내용 가이드에서는 디지털 혹은 SW 관련 내용이 비중 있게 반영되었다고 보기 어렵다[33]. 2022 개정 교육과정에서도 미래 사회를 준비하기 위한 역량 함양과 디지털 및 인공지능 기술 발달로 인한 사회 변화에 적극적으로 대처하기 위해 디지털 기초 소양을 강화할 필요가 있음을 제시하고 있다[37]. 국가 교육과정의 변화를 고려하여, 저작권 교육 내용 체계를 개발할 필요가 있다.

셋째, 한국저작권위원회에서 개발한 표준 교재와 강의자료는 학교급에 관계 없이 유사한 단어들이 출현하였다. 청소년 저작권 교육내용 가이드에서 제시된 것과 일치하는 방향에서 학교급별로 차별화된 내용 요소도 일부 있었으나, TF-IDF 분석 결과로 나타난 20개의 상위값에 해당하는 대부분 단어가 서로 비슷하게 나타났다. 초등학교 저학년과 고등학생은 학습 수준과 발달 단계가 명확히 다르므로, 자료에서 사용하는 용어와 내용 구성이 분명히 차별화될 필요가 있다. 저작권이라는 방대한 개념에 대해 초·중등학교 학생들이 알아야 하는 내용이 학교급별로 크게 다르지 않고, 청소년 저작권 교육내용 가이드에 포함된 대부분의 중점 내용 또한 모든 학교급에서 중복되어 구성되어 있지만, 같은 내용을 다루더라도 학습자의 발달 단계 및 수준에 따라서 다른 용어를 사용할 수 있을 것이다. 예를 들어, 초등학교 저학년 대상 강의자료에서도 다른 학년을 대상으로 한 강의자료와 마찬가지로 ‘이용’, ‘창작’, ‘출처’, ‘모방’ 등의 단어가 빈출하고 있는데, 초등학교 저학년 수준에 어려울 것으로 판단되는 용어는 수준에 맞게 순화할 필요가 있을 것이다.

본 연구는 한국저작권위원회에서 ‘청소년 저작권 교육내용 가이드’의 내용체계에서 제시하고 있는 학교급별 중점 내용을 기준으로 초·중등학교 저작권 교육 교재 및 강의자료가 어떻게 구성되어 있는지 분석하였다. 그러나 청소년 저작권 교육내용 가이드의 중점 내용이 해당 학교급의 학생 수준을 고려하고 있는 것으로 판단하기는 어렵다. 저작권 교육 내용 구성에 관한 연구가 지속되기 위해서는 한국저작권위원회의 교육 내용 체계 및 교육내용 가이드를 구성할 때 특정 영역 혹은 중점 내용이 왜 해당 학교급에서 중요도를 갖는지에 대한 타당성을 확보할 필요가 있다. 저작권 교육 내용 체계 구체화를 통해 향후 교재 개발 및 저작권 교육 내용 구성의 객관적 기준점이 될 수 있을 것으로 판단된다.

본 연구는 ‘기관의 저작권 교육’ 뿐 아니라 향후

‘국가 수준의 교육과정을 통한 학교 현장의 저작권 교육’ 방향성 도출을 위한 기초자료를 제시하였다는 점에서 의의가 있다. 디지털·AI 기술 발전과 사회 변화로 인한 저작권 교육의 중요성이 지속적으로 강조되고 있으나, 이와 관련된 저작권 교육 내용 연구는 미비한 상태이다. 본 연구가 ‘방대한 저작권의 내용을 어떻게 구성하여, 어떤 방식으로 학생들에게 교육할 것인지’에 대한 향후 저작권 교육 후속 연구의 초석이 될 것을 기대한다.

참고문헌

- [1] Yang, H., Kim, J., & Lee, W. (2023). Analyzing the Alignment between AI Curriculum and AI Textbooks through Text Mining. *Applied Sciences*, 13(18), 10011. DOI : 10.3390/app131810011
- [2] Shin, S. (2021). A Study to Design the Instructional Contents for National Curriculum of Computer Education in Elementary School. *JOURNAL OF The Korean Association of information Education*, 23(1), 13-31. DOI : 10.14352/jkaie.2021.25.1.13
- [3] Seo, K., & Han, S. (2021). Keyword Analysis of Artificial Intelligence Textbooks for Elementary and Secondary Schools: Focusing on Language Network Analysis. *Journal of The Korean Association of Artificial Intelligence Education*, 2(3), 7-23. DOI : 10.52618/aied.2021.2.3.2
- [4] Ministry of Education (2022). *The national guidelines for the 2022 Revised Elementary and Secondary School Curriculum*.
- [5] Choi, H., & Lee, J. (2005). Study on Implementation of Rights Information System and its Use-process for Libraries' Digital Rights Management. *Journal of the Korean Society for Information Management*, 22(3), 163-181. DOI : 10.3743/KOSIM.2005.22.3.163
- [6] Kim, H., & Park, G. (2022). Development and Application of Educational Program for Cultivation of Copyright Basic Knowledge: Focusing on Second-year Commercial High School Class case in Seoul. *The Journal of Learner-Centered Curriculum and Instruction*, 22(24), 235-260. DOI : 10.22251/jlcci.2022.22.24.235
- [7] Kim, H. (2011). *A comparative study on youth copyright training effectiveness*. Master's thesis, The Graduate School of Education, Ewha Womans

- University.
- [8] The Korea Copyright Commission. (2022). *Copyright Statistics. 1X12*, 178-179. The Korea Copyright Commission.
- [9] The Korea Copyright Commission. (2023). *Copyright Education Training Basic Plan of 2023*. The Korea Copyright Commission.
- [10] Koo, J. (2007). The perceptions of 5th-grader on the Copyright. *The Journal of Education, 27*(2), 41-59.
- [11] Kim, J., Park, J., & Lee, W. (2011). A Study of the Copyright Knowledge of Elementary School Teachers in Association with the Utilization of Education Information. *JOURNAL OF The Korean Association of information Education, 15*(2), 241-251.
- [12] Han, S., & Kim, S. (2011). The Information Gifted Teacher's Recognition and Developing the Copyright Education Program. *JOURNAL OF The Korean Association of information Education, 15*(3), 327-334.
- [13] Kim, I., Han, K., Lee, S., & Lee, C. (2012). Analysis of Copyright Education in Informatics Curriculum and Textbook for Middle School. *The Journal of Korean association of computer education, 15*(4), 1-12. DOI : 10.32431/kace.2012.15.4.001
- [14] Yu, B. (2009). The improvement of information ethics and copyright education in elementary moral education. *Journal of Human Rights & Law-related Education, 2*(3), 1-30.
- [15] The Korea Copyright Commission. (2020). *A study on the reflection of copyright content in textbooks and the development of educational programs*. The Korea Copyright Commission.
- [16] Lee, K., Kim, J., & Lee, W. (2023, August). Analysis of Copyright-Related Content by Subject in the 2022 Revised Curriculum Documentation. *Proceeding of the Korean Association of Computer Education, 27*(2), 51-54.
- [17] The Korea Copyright Commission. (2022). *Study on the Development of Secondary School Level Copyright Standard Textbooks*. The Korea Copyright Commission.
- [18] The Korea Copyright Commission. (2023). *Study on the Development of Elementary School Level Copyright Standard Textbooks*. The Korea Copyright Commission.
- [19] The Korea Copyright Commission. (2022). *Easy-to-Understand Copyright Textbook for Adolescents : Standard Textbook for Secondary School Students*. The Korea Copyright Commission.
- [20] The Korea Copyright Commission. (2023). *Easy-to-Understand Copyright Textbook for Elementary School Students : Standard Textbook for Elementary Students*. The Korea Copyright Commission.
- [21] The Korea Copyright Commission. (2022). *Copyright Journey with "Creators" & "Sharers" : For Lower Elementary Grades*. The Korea Copyright Commission.
- [22] The Korea Copyright Commission. (2022). *Let's Embark on a Copyright Journey with Easy Learning Copyright Classroom : For Upper Elementary Grades*. The Korea Copyright Commission.
- [23] The Korea Copyright Commission. (2022). *Useful Copyright Stories to Know: For Middle School Students*. The Korea Copyright Commission.
- [24] The Korea Copyright Commission. (2022). *Understanding Copyright Through Examples: For High School Students*. The Korea Copyright Commission.
- [25] Tan, A. H. (1999). Text mining: The state of the art and the challenges. *Proceeding of the pakdd 1999 workshop on knowledge discovery from advanced databases, 8*, pp. 65-70.
- [26] Jockers, M. L., & Thalken, R. (2020). *Text analysis with R*. Springer International Publishing. DOI : 10.1007/978-3-030-39643-5
- [27] Park, J., Park, H, Jung, S., & Lim, K. (2022). An Analysis of Research Trends on Personalized Learning : Using TF-IDF, LDA based Text Mining. *The Korean Journal of Educational Methodology Studies, 34*(4), 711-739.
- [28] Lee, J., Lee, M., & Kim, J. (2019). A study on Korean language processing using TF-IDF. *Journal of information systems, 28*(3), 105-121. DOI : 10.5859/KAIS.2019.28.3.105
- [29] Lee, S. (2014). A Content Analysis of Journal Articles Using the Language Network Analysis Methods. *Journal of the Korean Society for Information Management, 31*(4), 49-68. DOI : 10.3743/KOSIM.2014.31.4.049
- [30] The Korea Copyright Commission. (2024). *Copyright Term Dictionary*. Accessed: Jan. 08, 2024. [Online]. Available: <https://www.copyright.or.kr/information-materials/dictionary/list.do>
- [31] Kim, H., Lee, Do., & Kang, B. (2011). A Network of

Co-occurring nouns of Event Nouns. *Language and Linguistics*, 0(50), 81-106.

- [32] Kang, M. (2000). A network analysis of political power structure. *Journal of Communication Research*, 37, 93-130.
- [33] The Korea Copyright Commission. (2021). *Adolescent Copyright Education Content Guide*. The Korea Copyright Commission.
- [34] The Korea Copyright Commission. (2023). *U.S. Copyright Office Imposes Restrictions on Copyright Protection for Creations Using AI*. Copyright Trends of 2023, Vol 3. The Korea Copyright Commission.
- [35] Kim, J. (2019). Analyzing contents of software education area shown in the 2015 revised elementary Practical Art textbook. *JOURNAL OF The Korean Association of information Education*, 23(1), 9-18. DOI : 10.14352/jkaie.2019.23.1.9
- [36] Ministry of Education (2015). *The national guidelines for the 2015 Revised Elementary and Secondary School Curriculum*.
- [37] Kim, J. (2023). A Critical Discourse Analysis on Government Policy Documents Related to the 2022 Revised Curriculum. *The Journal of Curriculum Studies*, 41(3), 153-178. DOI : 10.15708/KSCS.41.3.7



이 경 민

2018년 부산대학교
일반사회교육과 (문학사)
2023년 고려대학교
교육정보전공 (교육학석사)
2024년~ 고려대학교
컴퓨터학과 박사과정

관심분야: 정보교육, 교육과정, 교육평가, 인공지능 교육
E-Mail : leekm.inc@gmail.com



김 자 미

1992년 이화여자대학교
교육학과 (문학사)
1995년 이화여자대학교
교육학과 (문학석사)
2011년 고려대학교
컴퓨터교육학과 (이학박사)

2011 ~ 2015년 고려대학교 컴퓨터학과 연구교수
2015 ~ 현재 고려대학교 교육대학원 컴퓨터교육전공 부교수
관심분야: 정보교육, 교육과정평가, 에듀테크
E-Mail : celine@korea.ac.kr



이 원 규

1985년 고려대학교
영어영문학과(문학사)
1989년 츠쿠바대학
이공학연구과 (공학석사)
1993년 츠쿠바대학 공학연구과
전자정보공학 전공(공학박사)
1993 ~ 1995년 한국문화예술진흥원
문화정보본부 책임연구원

1996 ~ 2014년 고려대학교 사범대학 컴퓨터교육과 교수
2014 ~ 2020년 4월 고려대학교 정보대학 컴퓨터학과 교수
2020년 5월 ~ 현재 고려대학교 대학원 컴퓨터학과 교수
관심분야: 정보교육, 정보표현, 정보관리, 교육정책
E-Mail: lee@inc.korea.ac.kr

부 록

<표 1> 교재 및 강의자료 TF 결과값(상위 20개)

순위	표준 교재 2종		강의자료 4종			
	초등용	중등용	초등 저	초등 고	중학교	고등학교
1	이용(59)	이용(252)	이용(33)	표현(11)	이용(35)	이용(42)
2	사람(56)	보호(129)	창작(16)	공통점(10)	출처(25)	보호(21)
3	보호(48)	사람(108)	그림(15)	이름(9)	표사(16)	창작(16)
4	창작물(40)	권리(107)	방법(15)	그림(7)	허락(16)	사람(14)
5	권리(21)	창작(85)	사람(12)	음악(6)	보호(10)	표사(13)
6	자신(21)	집해(76)	친구(10)	창작(6)	사진(9)	허락(11)
7	방법(21)	허락(71)	허락(10)	독창(5)	창작(9)	목적(11)
8	표현(20)	표현(68)	출처(8)	사람(5)	그림(8)	산업(10)
9	작품(20)	자신(67)	내용(7)	사진(5)	사람(8)	수익(10)
10	종류(18)	내용(61)	사진(7)	감정(4)	사용(5)	영리(10)
11	내용(17)	작품(55)	권리(6)	방식(4)	영리(5)	표현(10)
12	작성(16)	인정(47)	카드(6)	생각(4)	이름(5)	방법(9)
13	허락(15)	음악(47)	표현(6)	보호(3)	자유(5)	수업(9)
14	표사(15)	공정(44)	게임(5)	사건(4)	조건(5)	자신(9)
15	출처(15)	표사(42)	상황(5)	여행(3)	공유(4)	제작(9)
16	설명(15)	산업(43)	종류(5)	창작자(3)	목적(4)	공표(8)
17	생각(14)	영해(41)	표사(5)	교실(2)	방식(4)	금지(7)
18	자유(13)	의미(41)	생각(4)	미술(2)	자료(4)	조건(7)
19	인정(12)	제작(40)	약속(4)	시간(2)	자신(4)	학교(7)
20	노력(11)	사진(40)	음악(4)	어문(2)	정품(4)	공연(6)

<표 2> 강의자료별 동시출현 빈도수(상위 10개값)

초등 저학년 강의자료	빈도수	초등 고학년 강의자료	빈도수
저작물 - 이용	17	저작물 - 공통점	9
저작권 - 이점	14	공통점 - 이름	9
공유 - 마당	7	저작물 - 저작물	9
다른 - 사람	7	독창 - 표현	5
관련 - 저작물	7	저작물 - 사람	4
저작권 - 답	6	사람 - 생각	4
답 - 퀴즈	6	생각 - 감정	4
창작 - 나눔	5	감정 - 독창	4
모방 - 창작	5	표현 - 저작물	4
허락 - 이용	4	저작물 - 창작	4
중학교 강의자료	빈도수	고등학교 강의자료	빈도수
공유 - 마당	15	저작물 - 이용	15
저작물 - 이용	13	저작물 - 작성	7
저작권 - 집해	12	저작권 - 보호	6
출처 - 표시	8	출처 - 표시	6
공유 - 저작물	5	동인성 - 유지	5
이용 - 허락	5	저작물 - 보호	4
저작물 - 저작자	4	공유 - 마당	4
공유 - 저작권	4	성명 - 표시	4
저작물 - 창작	4	영리 - 목적	4
이용 - 단계	4	수익 - 배분	4

<표 3> 저작권 용어 사용자 사전 단어의 초등 및 중등 교재 내 출현 빈도

No.	단어	교재		강의자료			
		초등용	중등용	초등 저학년	초등 고학년	중학생	고등학생
1	2차적저작물작성권	2	7	-	-	1	5
2	결합저작물	-	9	-	-	-	-
3	공공누리	-	2	-	-	-	4
4	공공저작물	-	4	-	-	-	3
5	공동저작물	-	14	-	-	-	-
6	공연권	2	15	-	-	1	4
7	공유마당	1	8	8	-	16	5

*0은 -로 표기

No.	단어	교재		강의자료			
		초등용	중등용	초등 저학년	초등 고학년	중학생	고등학생
8	공중송신권	2	14	-	-	1	2
9	대여권	2	11	-	-	1	4
10	동시 중계 방송권	-	4	-	-	-	1
11	동일성유지권	2	10	-	-	1	4
12	디지털 음성 송신	-	6	-	-	-	1
13	리이선스	-	5	-	-	1	2
14	링크	-	17	-	-	-	-
15	방송권	-	4	-	-	-	1
16	배타적발행권	-	2	-	-	-	-
17	배포권	2	11	-	-	1	4
18	법정허락	-	4	-	-	-	-
19	복제권	3	14	-	-	1	5
20	불법복제물	3	9	-	-	-	-
21	빅데이터	-	3	-	-	-	-
22	성명표시권	4	9	-	-	1	4
23	세어웨어	-	2	-	-	-	-
24	업무상저작물	-	8	-	-	-	-
25	영상저작물	1	15	-	-	-	1
26	웹 캐스팅	-	2	-	-	-	-
27	이용 허락	8	25	-	-	2	3
28	자유 이용	-	-	-	-	-	1
29	저작권대리중개업	-	5	-	-	1	1
30	저작권 등록	12	14	-	-	-	-
31	저작권 보호	13	66	-	-	1	-
32	저작권 보호 기간	1	24	-	-	-	-
33	저작권신탁관리업	-	5	-	-	-	-
34	저작권자	10	101	2	-	6	11
35	저작권 침해	4	75	-	-	-	-
36	저작물 이용	6	9	-	-	-	-
37	저작인격권	7	21	-	-	2	4
38	저작인접권	-	50	-	-	-	5
39	저작재산권	9	41	-	-	3	5
40	전송	-	18	-	-	-	3
41	전시권	3	6	-	-	1	2
42	지식재산권	-	10	-	-	-	6
43	창작성	4	30	-	-	-	4
44	출처 표시	-	1	-	-	-	4
45	카피레프트	-	4	-	-	-	-
46	패러디	-	24	-	-	-	-
47	표절	-	10	-	-	2	-
48	프리웨어	-	2	-	-	-	-

<표 4> 전문가 검토에 의해 불용어 처리된 단어

No.	단어	No.	단어
1	단답	9	이중
2	왼쪽	10	구체
3	관련	11	동안
4	원래	12	단계
5	순서	13	별도
6	지급	14	해당
7	이번	15	경우
8	부분	16	다음