

‘어린이’와 ‘키즈’를 주제로 하는 유튜브(YouTube) 콘텐츠 분석: LDA 토픽모델링을 활용하여*

이근영** · 정부민***

본 연구는 유튜브(YouTube)에서 ‘어린이’와 ‘키즈’를 주제로 하는 콘텐츠를 분석하여 중심이 되는 주제를 파악하고 유아와 연관된 디지털 사회문화적 맥락을 찾는 데 목적을 둔다. 이를 위해 NetMiner의 SNS Data Collector를 활용하여 유튜브 데이터를 수집하였고, Excel과 NetMiner 4.5.0를 사용하여 LDA 토픽모델링 분석을 하였다. 분석 결과 첫째, ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠는 총 5개 토픽으로 분석되었고, ‘예술과 유아’, ‘종교와 유아’, ‘교육과 유아’, ‘사회와 유아’, ‘놀이와 유아’로 나타났다. 둘째, 토픽별 키워드 네트워크에서 모든 토픽의 평균 거리가 유의하게 멀었고, 공출현빈도가 높은 키워드를 통해 각 토픽의 흐름을 파악하였다. 교육-예술-놀이 토픽들이 서로 연관성을 보였으며, 사회 토픽에서는 현재 사회적 이슈가 주를 이루고 있었고, 종교 토픽에서는 유튜브가 미치는 영향력이 오락과 교육을 넘어 전 범위로 확장되고 있음을 알 수 있었다. 이러한 연구 결과를 바탕으로 유아를 대상으로 하는 유튜브 콘텐츠 연구의 기초자료를 제공하고자 한다.

주제어(Key words) : 유튜브(YouTube)

어린이 · 키즈 키워드 콘텐츠(contents with the keyword ‘child’ and ‘kids’)

LDA(Latent Dirichlet Allocation)

토픽모델링(topic modeling)

네트워크 분석(network analysis)

※ 논문 접수 : 2023. 12. 1. / 수정본 접수 : 2023. 12. 29. / 게재 승인 : 2023. 12. 29.

* 이 논문은 2023년 열린유아교육학회 추계정기학술대회 구두발표 논문임.

** 포항대학교 유아교육과 시간강사

*** 경북대학교 아동가족학과 박사수료 / 교신저자, qnals09@gmail.com

I. 서론

디지털 전환이라는 사회적 변화는 생애 초기 경험에도 변화를 가져왔다. 오늘날 유아들은 어린 시절부터 스마트폰, 컴퓨터 등과 같은 디지털 기기를 자유자재로 사용하고 디지털 환경에서 성장하는 ‘디지털 원주민(digital native)’ 세대이다. 아이들은 디지털 환경과 디지털 기기들이 친숙하며 인터넷을 통해 다양한 정보들을 쉽게 경험한다(Marsh et al., 2016; Verenikina & Kervin, 2011). 궁금한 것을 해결하기 위해 책과 사전을 찾아보고 그림을 그리기 위해 미술도구가 필요했던 과거와 달리, 디지털 전환의 시대에 살고 있는 유아들은 인터넷으로 지식을 습득하고, 디지털 기기로 표현하고 상호작용한다.

우리가 경험하는 디지털 콘텐츠들은 주로 Google, NAVER, Daum, Instagram, YouTube 등과 같은 인터넷 플랫폼과 애플리케이션을 통해 실행된다. 그중에서 유튜브(YouTube)는 가장 빠르게 이용자의 수가 증가하고 있는 동영상 공유 플랫폼이다. 대표적인 모바일 시장분석 기관인 모바일인덱스(mobileindex)의 ‘유튜브 사용자 조사’ 결과 유튜브가 온라인 동영상 채널 검색에서 압도적인 사용자 비율(91.8%)을 차지하고 있으며(모바일인덱스, 2023), 와이즈앱·리테일·굿즈(wisapp·retail·goods)의 ‘앱순위 분석’ 결과 한 달 동안 유튜브 앱의 사용은 1,044억 분으로 3위인 네이버(222억 분)보다 약 4.5배 이상 많이 사용되고 있다(와이즈앱·리테일·굿즈, 2023).

그리고 최근 연구에 의하면 어린 연령의 유튜브 사용에 대한 보고가 증가하고 있다. 한국언론진흥재단(2020)의 ‘인터넷 사용 플랫폼 연구’에 의하면 유아들이 사용하는 온라인 플랫폼의 대부분(94.8%)을 유튜브가 차지한다고 보고하였으며, 0~7세 유아의 80%가 유튜브를 사용하고, 주중의 하루 평균 1.39시간을 유튜브 영상 시청에 쓰는 것으로 나타났다(Neumann & Herodotou, 2020). 이는 인터넷 유저들이 주로 사용하는 플랫폼이 포털 사이트에서 유튜브로 이동하고 있으며(김형진, 정효정, 권보람, 2022; 이정희, 2018), 특히 어린 인터넷 유저들의 유튜브 의존 경향이 크다는 것 잘 보여준다. 어린 유튜브 사용자가 기하급수적으로 증가하는 원인을 분석해 보면, 첫째, 유튜브는 글로 설명하기 어려운 부분들을 시청각 자료로 제시하여 이해에 효과적이다(Dehghani et al., 2016). 흥미와 집중을 유발하는 다양한 편집 기술이 들어간 동영상이 많아지면서(송민주, 2020), 영상에 익숙한 요즘 세대에게 가장 편리한 정보 획득 수단이 되었다. 둘째, 유튜브는 기타 플랫폼과 달리 계정 없이도 쉽게 동영상을 검색하고 시청할 수 있는 구조적인 강점을 갖고 있어(김민경, 2019) 유아들이 접속하기 쉽고, 참여하기도 어렵지 않기 때문에 그 영향력이 점차 확대되고 있다. 마지막으로, 유튜브는 콘텐츠를 업로드하거나 공유하기가 쉬워서 전문가가 아니라도 편리하게 콘텐츠를 생성할 수 있다(변현진, 2018; 최미경, 2020). 이러한 편리성과 강점들로 인해 유튜브는 이제 유아가 가장 선호하는 플랫폼인 동시에 많은 어린이 크리에이터가 활동하는 장(field)이 되고 있다.

유아가 유튜브 콘텐츠의 주요 소비자이자 생산자가 되는 사회적 변화에서 유아들이 경험하는 유튜브 콘텐츠와 환경에 대한 연구가 최근 많이 이루어지고 있다. 유아와 관련된 유튜브 연구들은 첫째, 유튜브 콘텐츠를 이용하는 유아들의 연령, 성별, 이용 시간 등을 분석한 연구로, 많은 유아가 유튜브를 이른 시기부터 경험하며, 동영상 시청 시간이 점차 늘고 있다고 보고한다(강미리, 2021; 객선영, 2019; 김창숙, 2021; 박미선, 2020; 서희, 2023; 양유진, 전희경, 2021; Neumann & Herodotou, 2020). 둘째, 유아들의 유튜브 이용에 영향을 미치는 여러 가지 요인들에 대한 연구이다. 관련 연구들에 의하면 아동의 개인적 특성인 기질, 발달과 아동을 둘러싼 사회적 요소들에 따라 유튜브 사용을 예측할 수 있으며(Taylor & Cingel, 2023), 부모의 태도에 의해 유아들의 유튜브 사용에 차이가 있으므로 부모가 적절하게 유튜브 이용 방식을 조절하여 유아들이 바람직하게 유튜브를 시청할 수 있도록 제안하였다(김민정, 최운정, 2023; 이창호, 장석준, 2021; 조수선, 김봉현, 2021; Jindal & Kanozia, 2019). 셋째, 유아 유튜브 콘텐츠의 내용과 윤리에 대한 연구이다. 관련 연구들은 유아들에게 부적절한 유튜브 콘텐츠 필터링을 메타 데이터와 커뮤니티 피드백을 활용한 알고리즘에만 의존하고 있음을 비판하고 추가적인 안전 필터가 필요하다고 제안하였다(Alghowinem, 2018; Tahir. et al., 2019). 그리고 유아를 대상으로 하는 유튜브 콘텐츠들이 시각적 자극에만 집중하고 상업적인 내용이 많아서 유아의 교육적 자극에 적합한 콘텐츠 인지 의문을 제기하였고(Neumann & Herodotou, 2020; Papadamou, et al., 2020; Yadav, et al., 2018), 아동의 성장과 발달에 저해되는 무분별한 광고에 대한 규제와 책임을 강조하였다(이영주, 유수정, 2020). 넷째, 유아가 진행자가 되는 유튜브의 채널에서 유아의 권리와 복지에 관한 연구이다. 관련 연구들은 어린이 유튜브 보호를 위한 방안을 더욱 강화해야 하며 그들의 노동과 가치에 관심을 가지고 지속적으로 관리해야 한다고 주장한다(강희주, 정익중, 2020; 김은진, 2021; 김희연, 김명희, 2012; 백민제, 2021; 이현정, 2018).

위에서 언급한 유튜브 연구들을 종합해 보면 유아들의 유튜브 사용이 점차 늘어나고 있으며, 동시에 유튜브 콘텐츠의 내용적 문제점과 부정적 영향을 우려하여 유아용 콘텐츠에 대한 더욱 면밀한 분석과 적절한 사용을 지속적으로 요구하고 있다. 그러나 급증하고 있는 유튜브 유아 콘텐츠의 디지털 사회문화적 맥락을 파악하여 앞으로의 방향성을 제시하기에는 부족함이 있다. 유아들의 디지털 기술 경험과 선호하는 내용에 대한 연구는 앞으로 교육의 방향과 개선을 위해 기본적으로 요구된다(Bennett, 2012). 유튜브 콘텐츠가 유아들에게 미치는 영향력이 점점 커지고 있는 사회적 변화 속에서, 유튜브 콘텐츠의 문제점에 초점을 맞추기보다는 현재 유튜브에서 주로 업로드되고 검색되는 콘텐츠의 양상을 파악하는 것이 우선되어야 할 것이다. 따라서 본 연구는 유튜브 동영상의 제목과 내용에서 ‘어린이’와 ‘키즈’를 포함하고 있는 콘텐츠의 중심 주제들을 파악하고 그 주제들 간의 관계를 분석하여 유아 유튜브 콘텐츠의 교육적·실용적 방향성을 제공하는 기초자료를 제공하고자 한다. 이를 위해 설정한 연구문제는 다음과 같다.

- 연구문제 1. ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠의 주요 토픽은 무엇인가?
 연구문제 2. ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠의 각 토픽에서 주요 키워드의 관계는 어떠한가?

II. 연구방법

1. 분석방법

본 연구에서는 ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠의 중심 주제를 분석하기 위해 잠재 디리클레 할당(Latent Dirichlet Allocation, 이하 LDA) 방식의 토픽모델링을 사용하였다(Blei, Ng, & Jordan, 2003). LDA 모델의 핵심은 하나의 문서 내에 복수의 토픽이 존재할 수 있다는 가정하에 연관어들 사이의 동시 출현 확률을 자동으로 분석하여 잠재적인 토픽을 찾아내는 것이다(Blei, 2012). 확률적인 방법에 근거하여 토픽을 추출하기 때문에 연구자의 주관성에 의존하지 않고 객관적인 방법으로 내용을 분석할 수 있다는 장점이 있다. 이러한 장점으로 인해, 텍스트에서 유용한 의미를 찾아내고자 하는 연구에서 활발히 활용되고 있다(박건령, 2023; 송민, 2017; 정부민, 이효림, 2023; 조수민 외, 2021).

LDA 방식은 기본적으로 토픽의 개수를 특정 값으로 고정하여 분석하는데, 실제로 그 특정 값을 알 수 없으므로 여러 값을 반복하여 모델의 성능이 가장 높은 값을 토픽 개수로 결정하게 된다. Blei가 제안한 적절한 토픽 수를 정하는 방법은 주제의 복잡도(Perplexity)와 주제의 일관성(Coherence) 지표를 확인하는 방식으로 복잡도 수치가 낮고 일관성 수치가 높을수록 내용이 잘 분류되었다고 판단한다(윤성진, 이관용, 홍정기, 2021; Blei et al., 2003). 따라서 본 연구에서는 주제의 일관성(Coherence) 지표를 확인한 후 최적의 토픽 수를 선정하였다.

2. 분석대상

유튜브는 내부 알고리즘에 따라 검색어와 영상 간의 연관성이 계속 업데이트된다. 유튜브는 소프트웨어 개발자들을 위해 채널, 댓글 등 다양한 정보를 수집할 수 있는 Data API를 공개하고 있고, 이를 통해 실시간으로 변하는 데이터를 얻을 수 있다. NetMiner는 SNS Data Collector 옵션에서 개발자가 아닌 일반인도 유튜브의 데이터를 수집할 수 있도록 기능을 제공한다(Cyram, 2022). 따라서 본 연구는 SNS Data Collector를 활용하여 ‘어린이’와 ‘키즈’를 포함하는 영상의 정보를 수집하였다.

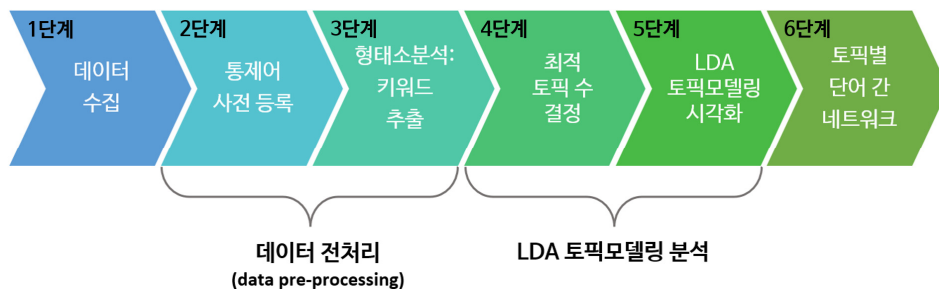
데이터 수집은 ‘어린이’와 ‘키즈’를 검색어로, 두 개의 단어 중 하나만이라도 포함되어 있으면 모두 수집되도록 하였다. 기간은 총 2014년부터 2023년까지에서 매년 4월 5일부터 4월 15일까지 11일간으로 설정하였다. 검색 시점은 2023년 6월 16일이다. 실효성 있는 분석을 위해 조회수가 100 이하인 영상은 제거하여, 최종적으로 2,071개의 동영상이 선정되었다. 그리고 최종 선정된 동영상의 제목(Title)과 내용(Full Text)에 있는 텍스트를 중심으로 분석하였으며, 연도별 수집 영상 수와 수집 항목은 다음 표 1과 같다.

<표 1> 연도별로 수집된 유튜브 동영상 수와 수집 항목

연도	수집된 영상 수	최종 수집 영상	키워드 설정	수집 항목
2014년	202개	182개	어린이 ANDOR 키즈	Title, Created Time, Writer(ID), Writer(Name), View Count, Like Count, Dislike Count, Comment Count, Full Text
2015년	250개	192개		
2016년	233개	192개		
2017년	253개	209개		
2018년	201개	181개		
2019년	251개	226개		
2020년	293개	250개		
2021년	281개	222개		
2022년	237개	204개		
2023년	270개	213개		
합계		2,071개		

3. 절차 및 자료분석

본 연구의 분석과정은 총 6단계로 진행되었다. 수집된 모든 데이터는 Excel로 정리한 후, NetMiner 4.5.0 프로그램을 이용하여 분석하였고, 절차는 다음 그림 1과 같다.



[그림 1] 연구 분석 절차 6단계

첫째, 1단계 데이터 수집은 앞에서 설명하였고 총 2,071개의 동영상상이 수집되었다.

둘째, 2단계와 3단계는 데이터 전처리(data pre-processing)에 해당한다. 2단계에서 NetMiner에 미리 통제어 사전을 등록하여, 연구목적과 데이터의 성격에 맞게 단어 정제 기준을 정하였다. 통제어는 유사어, 지정어, 제외어 사전을 .csv파일로 구성하여 정제된 키워드가 추출될 수 있도록 하였다. 3단계에서는 형태소 분석 및 한 글자 단어를 제거하였다. 건, 외 등과 같은 단어도 명사로 추출되므로, 의미가 없는 단어인 한 글자 단어는 제거한다(정부민, 이효림, 2023; 조수민 외, 2021). 추출된 형태소 단어를 확인한 후 다시 통제어 사전을 구성하는 식으로 2단계와 3단계는 순환적으로 반복된다. 이 과정을 통해 총 8,661개의 단어가 정제되었고, 그중 한 글자 단어를 제거하여 8,098개의 단어를 최종 분석에 사용하였다. 등록한 통제어 사전은 표 2와 같다.

〈표 2〉 Dictionary에 등록한 통제어 사전: 유사어, 지정어, 제외어

유사어(Thesaurus)	지정어(Defined Words)	제외어(Exception List)
의미가 유사한 단어들을 한 개의 단어로 통일	고유명사, 복합명사, 신조어, 줄임말 등을 등록	결과에 제외할 단어를 등록
노래 곡명 → 동요 송(song)	① 신조어 : 코로나, SNS (인스타 그램, 트위터 등)	① 유튜브에서 주로 사용 되는 단어 삭제 : 구 독, 좋아요, 클릭, 업로 드, 광고, 댓글 등
어린이 영양 영양실조 → 건강 건강식 및 영양 식생활	② 유튜브의 고유명사 : - 채널명(유니버, 깨비키즈, 헤이 지니, 보람튜브 등) - 어린이가 선호하는 캐릭터 (팬 더키키, 묘묘, 페파피그, 캐치 티니핑 등)	② 날짜 및 시간과 관련 된 단어 삭제 : 오전, 오후, 올해, 내년 등
키즈월드 키즈클럽 → 키즈카페 키즈랜드	③ 누리과정 관련 복합명사 : 역 할놀이, 쌓기놀이, 생활습관 등	③ 아이돌 관련 단어 삭 제: 스트레이키즈, 배 드키즈, 직캠, 아이돌 연구소
엄마 아빠 → 엄마아빠 어머니 아버지		

셋째, 4단계와 5단계는 연속해서 진행되고 LDA 토픽모델링 분석에 해당한다. 4단계에서 는 모델 성능이 가장 좋은 토픽 수와 하이퍼 파라미터(α , β)를 찾기 위해 Evaluation of Topic Models 옵션을 이용하였다. 모델 성능에 영향을 주는 α , β 와 토픽 수(K)의 범위를 부여한 후 반복 분석하고, 일관성(Coherence) 값이 1에 가장 가까운 경우를 찾는다. 그리고 5단계에서는

[Automatic Plugin] Topic Modeling 옵션에서 4단계에서 나온 최적 토픽 수(K)와 α , β 를 기준으로 토픽 분석하였다. TF-IDF¹⁾값을 조절하면 흔하게 사용된 단어를 제거하고 주제 예측 능력이 강한 키워드를 찾을 수 있으므로(이기창, 2019; Silge & Robinson, 2017), 본 연구에서 TF-IDF ≥ 0.2 로 필터링하여 워크파일을 새롭게 생성한 후 분석하였다. 그리고 분석 결과를 토픽-단어 네트워크로 시각화하여 토픽 간의 관계를 이미지로 살펴볼 수 있도록 하였다.

마지막으로, 6단계에서는 토픽별 키워드 네트워크 분석을 실시하였다. 토픽모델링 분석 결과를 토대로 ‘토픽별 중심 단어들의 네트워크’를 분석하였다. Window size²⁾=3, Link frequency threshold³⁾=1, Un-directed로 필터링하여 각 토픽의 키워드 공출현 네트워크를 형성하였고, 이를 통해 각 토픽 내에서 단어 간의 관계를 파악할 수 있었다.

III. 연구결과

1. ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠 분석: LDA 토픽모델링

1) 주요 키워드 빈도 분석

‘어린이’와 ‘키즈’를 주제로 하는 유튜브 영상의 제목과 내용에서 나타난 주요 키워드는 동요, 동화, 놀이 등이 상위에 나타났고, 교육, 영어, 초등, 수과학 같이 교육과 관련된 키워드, 엄마아빠, 가족, 친구, 사랑 등 사회적 관계와 관련된 키워드도 나타났다. 또한, 교회, 예배와 같이 종교와 관련된 키워드와 홈페이지, 인스타그램, 페이스북과 같이 인터넷과 연관된 키워드도 포함되었다. 상위 30개 주요 키워드는 표 3과 같다.

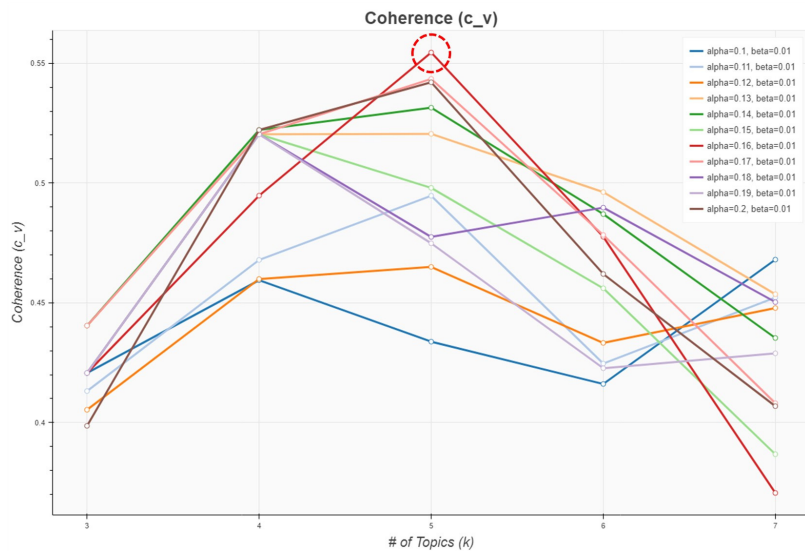
2) 최적 토픽 수 결정

LDA 토픽모델링의 성능이 가장 좋은 토픽 수(K), α , β 를 찾기 위해 반복분석을 110번 실시하였다. 그 결과 $K=5$, $\alpha=.16$, $\beta=.01$ 일 때 $coherence=0.554$ 로 가장 높았다. 따라서 토픽 수를 5개로 정하여 LDA 토픽모델링 분석하였고, 반복분석 결과를 그래프로 표현하면 그림 2와 같다.

- 1) TF-IDF는 ‘단어 빈도-역문서 빈도(Term Frequency-Inverse Document Frequency)’라고 하며, 단순한 빈도가 아닌 특정 단어가 문서 내에서 얼마나 중요한 위치를 차지하는지 그 정도를 나타내는 텍스트 분석 지표이다(정원준, 2022). 통계적으로, 모든 문서에서 사용될수록 0에 가깝고 소수의 문서에서 사용될수록 1에 가깝게 산출된다.
- 2) 단어 간 네트워크를 생성할 때 범위로, 최솟값인 2를 입력하면 바로 나란히 등장한 단어들끼리 네트워크가 형성된다(Cyram, 2022).
- 3) 단어 간 링크의 가중치(동시출현빈도)의 임계치로, 2를 입력하면 2회 이상 함께 사용된 단어 쌍만 추출된다(Cyram, 2022).

<표 3> 주요 키워드 상위 30개

	Keyword	Frequency		Keyword	Frequency		Keyword	Frequency
1	어린이	4,573	11	놀이	235	21	댄스	159
2	동요	1,359	12	안전	223	22	홈페이지	157
3	동화	577	13	영어	205	23	동물	155
4	친구	411	14	초등	183	24	놀이공원	143
5	교육	398	15	가족	171	25	수과학	140
6	장난감	369	16	찬양	167	26	인스타그램	137
7	엄마아빠	266	17	키즈카페	166	27	페이스북	133
8	만화	256	18	시간	162	28	음악	133
9	교회	250	19	깨비키즈	162	29	지니키즈	132
10	키즈댄스	239	20	예배	159	30	사랑	126

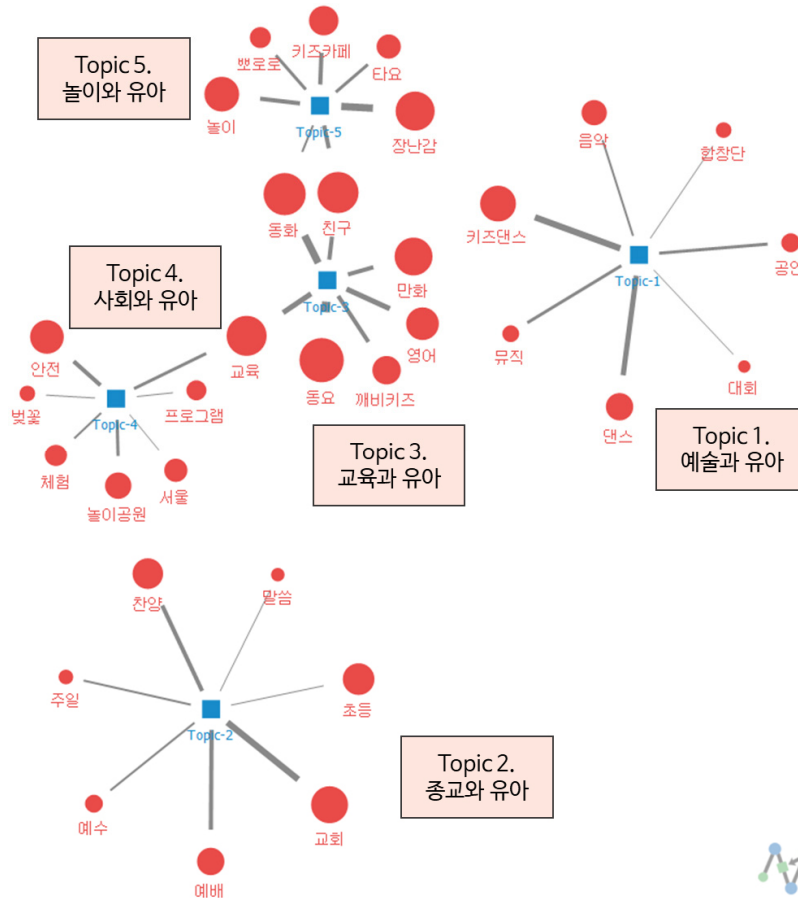


[그림 2] 최적 토픽 수 결정: 일관성(Coherence) 지표 확인

3) LDA 토픽모델링 분석

‘어린이’와 ‘키즈’를 주제어로 하는 유튜브 콘텐츠의 주요 주제 분석을 위해 LDA 토픽모델링을 사용하여 10년간 유튜브 콘텐츠의 제목과 내용을 분석하였다. 결과를 ‘토픽-단어 간 네트워크’로 시각화하면 빨간 원의 크기는 해당 단어의 출현빈도를 의미하며, 원이 클수록 출현빈도가 높음을 의미한다. 선의 굵기는 해당 토픽에 대한 할당 확률을 의미하고, 선이 두꺼울수록 확률이 높음을 의미한다(Cyram, 2022). 추출된 키워드의 군집을 대표할 수 있는

주제 영역의 명칭은 함께 연구를 진행한 연구자 2명이 각 토픽에 포함된 주요 키워드와 유튜브 제목을 검토하여 명명하였다. 토픽 분석 결과를 시각화하면 그림 3과 같다.



[그림 3] 유튜브 키즈 콘텐츠 토픽모델링 시각화

Topic 1은 ‘키즈댄스’, ‘댄스’, ‘공연’, ‘뮤직’, ‘음악’, ‘합창단’ 등의 키워드를 포함하는 군집으로, 어린이가 시연하는 음악 및 댄스 공연, 아이돌 음악과 댄스를 어린이가 따라 하는 영상 등으로 구성되어 있어 ‘예술과 유아’로 명명하였다. Topic 2는 ‘교회’, ‘찬양’, ‘예배’, ‘예수’, ‘주일’ 등의 키워드를 포함하는 군집으로, 어린이의 종교 활동과 관련된 영상으로 구성되어 있어 ‘종교와 유아’로 명명하였다. Topic 3은 ‘동요’, ‘동화’, ‘교육’, ‘영어’, ‘친구’, ‘만화’, ‘개비키즈’ 등의 키워드를 포함하는 군집으로, 교육용 동요 및 동화 영상으로 구성되어 있어 ‘교육과 유아’로 명명하였다. Topic 4는 ‘안전’, ‘교육’, ‘놀이공원’, ‘체험’, ‘서울’, ‘벚꽃’ 등의 키워드를 포함하는 군집으로, 어린이가 일상에서 참여할 수 있는 다양한 사회적 경험 등으로 구성되어 있어 ‘사회와 유아’로 명명하였다. Topic 5는 ‘장난감’, ‘친구’, ‘놀

이’, ‘키즈카페’, ‘타요’, ‘뽀로로’ 등의 키워드를 포함하는 군집으로, 장난감, 만화 캐릭터 영상 등으로 구성되어 있어 ‘놀이와 유아’로 명명하였다.

2. ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠의 토픽별 키워드 관계 분석

각 토픽에서 나타나는 키워드 간의 관계를 구체적으로 살펴보기 위해 키워드 네트워크 분석을 하였다. 순서대로 등장하는 3개 단어까지 동시 출현 단어로 간주하고, 1회 이상 함께 등장하는 단어를 선으로 연결하여 네트워크를 형성하였다. MCMC(Markov Chain Monte Carlo)⁴⁾검정에서 interaction 횟수를 196,000회로 지정하여 랜덤 네트워크를 생성한 후 유의성 검증을 시도하였다. 네트워크의 구조적 특성과 그 측정값의 통계적 유의성은 표 4와 같다.

<표 4> 토픽별 키워드 네트워크의 구조적 특성

	Topic 1. 예술과 유아		Topic 2. 종교와 유아		Topic 3. 교육과 유아		Topic 4. 사회와 유아		Topic 5. 놀이와 유아	
노드 수(of nodes)	95		97		98		100		98	
링크 수(of links)	254		472		459		471		360	
밀도(Network density)	.057		.101		.097		.095		.076	
Mean / SD	.057/.037		.101/.068		.097/.091		.095/.065		.076/.064	
중심성	Network degree centralization		18.097%		30.066%		54.338%		39.765%	
	37.521%									
관찰치(Obs)/기대치(Exp)	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
평균 연결 정도 (average degree)	5.347		9.732		9.367		9.42		7.347	
평균 거리 (mean distance)	2.985***	2.747	2.444***	2.226	2.23***	2.176	2.295***	2.224	2.656***	2.383
구성요소 (components)	2**	1.006	1***	1	1***	1	1***	1	1***	1
지름(diameter)	7		5		4		4*	4.017	5	

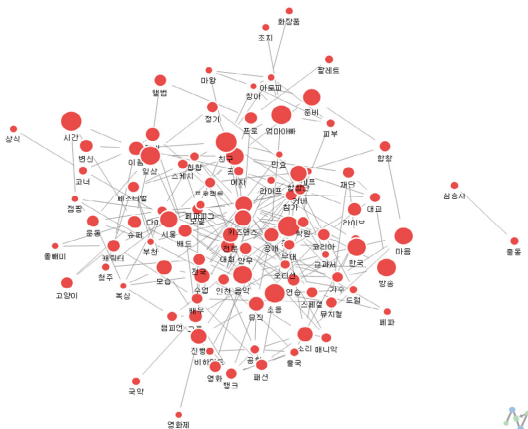
* $p < .05$, ** $p < .01$, *** $p < .001$

먼저 ‘평균연결정도’는 모두 유의하지 않게 나타났는데, 이는 각 토픽이 평균적으로 다수의 관계망을 가지는지 아닌지 판단할 수 없음을 의미한다(민기채, 최소영, 조성은, 2019). ‘평균 거리’는 모두 유의하게 나타나 평균 노드 간의 거리가 유의미하게 멀다는 것을 확인

4) 기존 네트워크에서 측정된 통계량(평균 거리, 네트워크의 지름 등)과 랜덤하게 연결된 네트워크에서 측정된 통계량을 비교하여 확률적으로 유의미한지 검증하는 방식이다(Roberts Jr, 2000).

할 수 있었다. 본 연구에서 노드는 단어를, 링크는 문서(유튜브 제목과 내용)에서 같이 등장한 횟수를 나타내고 있으므로, 유튜브에서 사용된 단어들이 영상별로 구분되어 있다고 해석할 수 있다. 즉, A계열의 영상에서 사용되는 단어들과 B계열의 영상에서 사용되는 단어가 달라서, 다른 계열의 영상들 간에 중복되는 단어가 거의 없다는 것이다. 그리고 ‘구성요소’도 모든 토픽 네트워크에서 유의하게 나타났다. 이는 토픽 1 ‘예술과 유아’는 2개의 군집으로 이루어져 있고, 토픽 2, 3, 4, 5는 1개의 군집으로 이루어져 있음을 의미한다. 마지막으로, ‘지름’은 토픽 4 ‘사회와 유아’에서만 통계적으로 유의하게 나타났는데, 이는 네트워크를 관통하는 최단 거리가 유의하게 짧다는 것을 말한다. 즉, ‘사회와 유아’ 토픽의 영상에서 공통적으로 나타나는 단어가 있음을 유추할 수 있다.

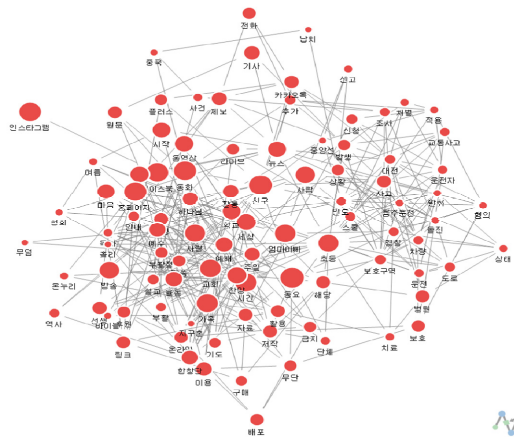
그리고 토픽별 키워드 네트워크를 시각화하였다. 키워드의 *TF-IDF*값이 클수록 원의 크기가 커지도록 표현하였고, 결과는 그림 4, 5, 6, 7, 8과 같다.



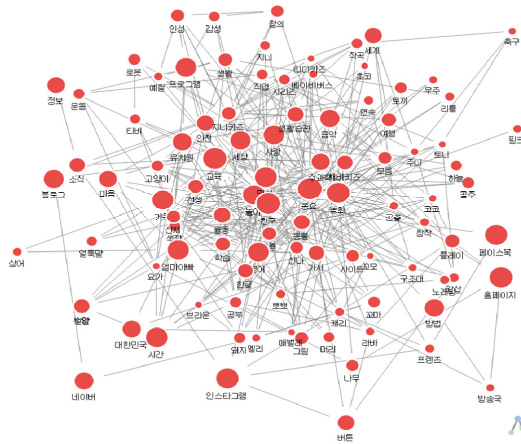
[그림 4] ‘예술과 유아’키워드 네트워크 시각화

토픽 1 ‘예술과 유아’의 키워드 네트워크에서 링크값이 높았던 키워드 간의 관계는 ‘뮤직-뱅크’, ‘댄스-키즈댄스’, ‘교과서-음악’, ‘음악-초등’, ‘교과서-초등’ 순으로 나타났다. 그 외에도 ‘공연-합창단’, ‘댄스-친구’, ‘대화-동요’ 등이 있었다. 이를 통해, 키즈댄스, 댄스, 친구가 서로 연관되어 있으며, 또한 초등 음악 교과와 관련된 영상도 많음을 알 수 있다. 그리고 뮤직과 뱅크는 다른 키워드와 연관성은 크지 않지만, 의미 있는 관계를 형성하고 있음을 알 수 있다.

토픽 2 ‘종교와 유아’의 키워드 네트워크에서 링크값이 높았던 키워드 관계는 ‘동화-성경’, ‘예배-주일’, ‘율동-찬양’, ‘주일-학교’, ‘교회-지구촌’, ‘교회-예배’ 순으로 나타났다. 이를 통해, 주로 교회에서 제공하는 유아를 위한 다양한 활동 영상이 서로 밀접하게 연관되어 있음을 알 수 있다.

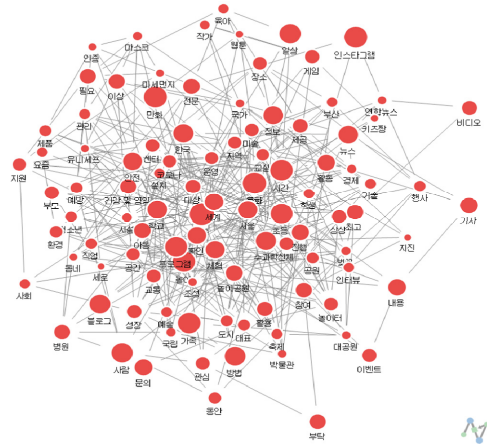


[그림 5] ‘종교와 유아’키워드 네트워크 시각화



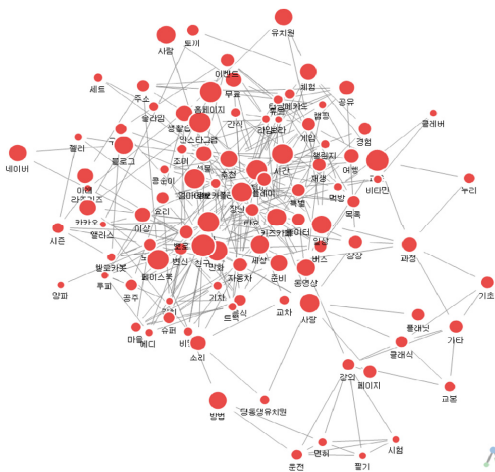
[그림 6] '교육과 유아' 키워드 네트워크 시각화

토픽 4 ‘사회와 유아’의 키워드 네트워크에서 링크값이 높았던 키워드 간의 관계는 ‘벚꽃-축제’, ‘놀이공원-서울’, ‘마스크-필요’, ‘교통-안전’, ‘직업-체험’, ‘교육-예방’, ‘안전-체험’ 순으로 나타났다. 이를 통해, 유아를 대상으로 한 안전교육 및 직업체험과 관련된 영상과, 현재 사회의 관심을 반영하는 키워드가 서로 연결되어 있으며, 또한 어린이가 참여할 수 있고 즐길 수 있는 행사 홍보가 연관되어 있음을 알 수 있다.



[그림 7] '사회와 유아' 키워드 네트워크 시각화

토픽 5 ‘놀이와 유아’의 키워드 네트워크에서 링크값이 높았던 키워드 관계는 ‘먼허-운전’, ‘강의-기타’, ‘놀이-장난감’, ‘라임-튜브’, ‘보라-조이’, ‘키즈카페-타요’, ‘누리-과정’ 순으로 나타났다. 그 외에도 ‘뽀로로-친구’, ‘버스-타요’, ‘슬라임-장난감’, ‘시즌-헬로카봇’ 등이 있었다. 이를 통해, 라임튜브와 보라조이 등의 채널에서 어린이들이 좋아하는 장난감(타요버스, 헬로카봇 등)과 게임영상을 많이 업로드하고 이들이 서로 연관되어 있음을 알 수 있다.



[그림 8] '놀이와 유아' 키워드 네트워크 시각화

IV. 결론 및 논의

본 연구에서는 ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠의 중심 토픽을 유형화하고 토픽별 내용 특성과 관계를 파악하고자 하였다. 이를 위해 10년간의 유튜브 콘텐츠 중 일정 기간 자료를 수집하여 LDA 토픽모델링을 활용해 분석한 연구 결과를 논의하면 다음과 같다.

첫째, ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠의 제목과 내용을 LDA 토픽모델링 분석한 결과 5개의 토픽으로 분류되었다. 토픽 1은 ‘예술과 유아’로 유아 음악과 댄스, 아이돌 음악과 댄스, 초등학교 음악 등으로 구성되어 있다. 토픽 2는 ‘종교와 유아’로 교회, 예배, 찬양 등과 같이 유아를 대상으로 하는 여러 종교 활동으로 구성되어 있다. 토픽 3은 ‘교육과 유아’로 유아를 위한 교육 동요 및 애니메이션으로 구성되어 있다. 토픽 4는 ‘사회와 유아’로 안전, 체험, 놀이공원 등과 같이 유아가 일상에서 참여할 수 있는 다양한 사회적 경험 등으로 구성되어 있다. 토픽 5는 ‘놀이와 유아’로 유아들이 선호하는 장난감 및 만화 캐릭터 영상 등으로 구성되어 있다. ‘어린이’와 ‘키즈’를 주제로 하는 유튜브 콘텐츠는 예술, 교육, 사회, 놀이, 그리고 종교와 연관된 단어들 이 주를 이루고 있었다.

유튜브에서는 13세 미만 어린이를 위해 유튜브키즈(YouTubeKids)를 따로 구분하여 서비스하고 있다. 유튜브키즈는 ‘프로그램’, ‘음악’, ‘학습’, ‘탐색’의 4가지 주제로 나누어져 있다. 유튜브키즈 콘텐츠와 본 연구 결과를 비교해 보았을 때 유튜브키즈의 ‘음악’과 ‘학습’ 주제와 본 연구 결과인 ‘예술’, ‘교육’, ‘놀이’는 공통 주제이지만, ‘종교’와 ‘사회’는 본 연구에서 수집한 전체 유튜브 콘텐츠에서만 나타나는 주제였다. 이를 통해 유아들이 유튜브를 시청할 때 대부분 부모의 아이드로 접속하고, 자녀가 어릴수록 채널 선택에 대한 대부분의 권한을 부모가 가지고 있다는 연구결과(강소영, 2020; 양유진, 전희경, 2021; Neumann & Herodotou, 2020)를 다시 한번 확인할 수 있었다. 유튜브키즈가 어린이를 위해 구분되어 서비스되고 있지만 정작 유아들이 유튜브에 접속하고 동영상을 선택할 때에는 부모의 영향이 크게 작용하여, 유튜브키즈에서 나타나지 않는 종교적 내용이나 현재 사회에서 주로 관심받는 주제들이 유아와 연관된 일반 유튜브 콘텐츠의 다수를 차지하고 있다는 것을 알 수 있다.

더불어 유아들이 선호하는 유튜브 동영상 유형 및 분류에 대한 선행연구들을 살펴본 결과, 본 연구의 토픽 분석에서 나타난 ‘예술’, ‘교육’, ‘놀이’ 주제는 선행연구들과 동일하게 유아들이 많이 선택하는 주제로 확인되었다. 유아들이 주로 선택하는 유튜브 키워드가 놀이, 장난감, 동요라고 밝힌 연구(Neumann & Herodotou, 2020), 주로 이용하는 유튜브 장르가 교육/학습, 브이로그, 애니메이션, 언박싱 & 제품 리뷰, 게임, 음악이라고 보고한 연구(김창숙, 2021), 선호 유튜브 콘텐츠가 만화, 학습, 브이로그, 동요, 게임 등이라고 말한 연구(서희, 2023)가 이에 해당한다. 선행연구에서 나타난 동요, 음악 등은 ‘예술’ 주제와 연결되고,

장난감, 교육/학습, 학습, 애니메이션, 만화 등은 ‘교육’ 주제와 연결되며, 놀이, 브이로그, 게임 등은 ‘놀이’ 주제와 연결된다고 해석할 수 있다. 또한 유아들이 주로 만화, 애니메이션, 웹툰, 장난감 등의 콘텐츠를 선택하는 것으로 조사되어(한국언론진흥재단, 2020), 본 연구의 결과를 뒷받침해 준다. 이를 통해 유아들은 유튜브 콘텐츠 중 만화와 음악(동요)을 많이 시청하고 있으며, 유아들에게 제공되는 다수의 교육 콘텐츠들이 애니메이션의 형태임을 알 수 있다. 이는 디지털 네이티브 세대인 유아들이 미디어 환경에서 영상으로 경험하는 애니메이션, 게임 등이 새로운 놀이와 교구가 되고 있으며, 유아들이 선호하는 캐릭터를 활용한 영상물이 유아의 학습과 교육 그리고 놀이에서 많이 활용되고 있다고 해석할 수 있다. 그러므로 유아들에게 양질의 동영상 콘텐츠를 제공해야 함은 물론 어떻게 디지털 콘텐츠와 상호작용해야 하는지 교육하고 안내하는 것이 유아교육의 새로운 과제가 되고 있음을 시사한다.

둘째, 각 토픽에서 나타나는 키워드 간의 관계를 분석한 결과 모든 토픽이 유의미하게 평균 거리를 유지하고 있어, 토픽별 주요 단어들이 서로 구분되어 있음을 확인할 수 있었다. 그리고 토픽별로 키워드 관계를 구체적으로 살펴보면, 토픽 1 ‘예술과 유아’의 키워드 관계에서는 유아들은 어린이를 위한 음악을 많이 듣고 즐기지만 대중가요와 댄스도 선호하고 있음을 알 수 있었다. 이는 5세 유아가 좋아하는 대중가요와 이를 접하게 되는 경로를 분석한 연구에서 유아들은 유튜브를 통해 대중가요를 많이 접하고 이미 유아의 삶 깊숙이 대중가요가 자리 잡고 있다는 연구(김대옥, 2019)를 지지하는 결과이다. 유아들은 어린이 음악에만 흥미와 관심을 보이는 것이 아니라 대중적으로 인기 있는 음악과 댄스도 좋아한다. 그러므로 유아에게 어린이 음악만 강요하기보다는 긍정적으로 대중가요를 경험하고 소비할 수 있도록 대중가요 리터러시를 증진시킬 수 있는 교육과 프로그램 개발에 관심을 가져야 할 것이다.

토픽 2 ‘종교와 유아’의 키워드 간의 관계 분석에서 종교 분야에서 유아를 대상으로 하는 유튜브 콘텐츠 업로드가 활발하게 이루어지고 있음을 알 수 있었다. 이는 종교 활동에서 유튜브의 영향력을 파악하고 기존의 방식에서 벗어나 새로운 종교 활동의 가능성을 제안한 연구(김태룡, 안승범, 2021)에서 알 수 있듯이 유튜브가 미치는 영향력은 종교를 포함하여 오락, 만화, 음악, 교육을 넘어서 전 영역으로 확장되고 있다. 그러므로 유아 유튜브 콘텐츠의 범위를 교육이나 흥미 위주의 영상으로 한정할 것이 아니라 유아의 발달과 성장 전 분야에서 직·간접적으로 활용될 수 있는 콘텐츠 제작이 요구된다.

토픽 3 ‘교육과 유아’의 키워드 간의 관계 분석에서 유튜브 영상이 교수 자료로 다양하게 활용되고 있음을 알 수 있었다. 특히 유아 유튜브 안전교육 동영상 그래픽 표현 기법은 2D가 가장 많았다는 연구(진영주, 2023)에서 드러나듯이 유아를 위한 교육 콘텐츠는 대부분 동화나 만화와 같이 애니메이션으로 제시하고 있었다. 그리고 유튜브 생방송을 활용한 재난 안전교육이 기존의 PPT를 활용한 안전교육보다 참여도와 만족도를 더 높였다는 연구

결과(김승민, 2018)와 같이 음향과 움직임이 지원되는 동영상은 적극적으로 교수-학습 자료로 활용할 필요가 있다. 현재 교육 현장에서 유튜브 영상이 많이 활용되고 있지만 성교육, 영양교육, 다문화교육, 장애통합교육 등 주제가 있는 활동에서만 주로 사용되고 있기에(최효정, 이근영, 2022), 유아 과학교육과 같이 현장에서 실제로 구현하는 데 어려움이 있는 교육 활동에서 유튜브 콘텐츠를 활용하는 교수자의 열린 자세가 요구된다.

토픽 4 ‘사회와 유아’의 중심 키워드는 유아의 주변 환경에서 접할 수 있는 다양한 경험에 대한 내용이 서로 밀접한 연관성을 보였다. ‘벚꽃-축제’, ‘놀이공원-서울’, ‘마스크-필요’ 키워드의 연관성은 현재 사회의 관심이 어디에 있는지 잘 보여주는 결과이다. 특히 ‘벚꽃-축제’와 ‘놀이공원-서울’ 키워드의 연관성이 높게 나온 이유는 데이터 수집 기간이 4월이라는 시간적 특수성이 반영된 결과로 파악할 수 있다. 그리고 ‘마스크-필요’는 코로나19와 같은 감염병으로부터 유아들의 건강을 지키기 위해 가장 쉽게 실천할 수 있는 방법인 마스크 착용의 필요성을 잘 보여주는 결과이다. 특이하게 ‘면허-운전’의 관계가 상위에 나타난 것은 2020년 3월과 2022년 7월에 강화된 ‘어린이보호구역’ 규제와 관련하여 빈도가 많아졌을 것으로 유추할 수 있다. 유튜브는 비교적 간단하게 제작할 수 있을 뿐만 아니라 콘텐츠를 업로드하거나 공유하기가 쉬워서 전문가가 아니라도 편리하게 콘텐츠를 생성할 수 있기 때문에(변현진, 2018), 현재 사회에서 반짝이는 이슈나 빠르게 전달하고자 하는 내용인 경우 유튜브가 주로 사용되고 있음을 알 수 있다.

마지막으로 토픽 5 ‘놀이와 유아’의 중심 키워드에서 ‘장난감’, 유아들이 좋아하는 ‘만화 캐릭터’, ‘게임 영상’, ‘키즈카페’가 서로 연결되어 있다. 이러한 결과는 포스트 휴먼 시대에 디지털 미디어는 새로운 놀이문화와 배움, 학습이 되고 있으며(동폴잎, 2022), 디지털 세상을 살아가고 있는 유아들은 인터넷으로 정보를 찾고, 상상놀이를 하며 상호작용한다(Marsh, et al., 2016; Verenikina & Kervin, 2011)는 연구와 맥을 같이 한다. 오늘날 유아의 놀이는 디지털과 아날로그가 융합되어 그 경계를 구분 짓기 어렵다. 손으로 점토를 만들며 상상놀이를 하고 동시에 놀이를 디지털로 저장하여 기억하고 전달하면서 즐거움을 나눈다. 디지털, 미디어, 인터넷은 발전된 기술이기도 하지만 유아들에게는 놀이의 도구가 되기도 한다.

한편, 유아들의 유튜브 사용을 우려하는 걱정과 함께 이를 변화하는 과정으로 받아들여야 한다는 의견도 있다. 유튜브를 최초로 시청하는 시기가 점점 빨라지고 있으며, 부모들은 유아의 흥미와 재미 때문에 유튜브 시청을 허용하는 경우가 많아졌다(김은영, 임신일, 2021). 더 나아가 유튜브를 유아들의 놀이, 창의성, 호기심을 자극하는 정보자원으로 인식하는 부모들은 양육에서 유튜브를 적극적으로 활용한다(Jindal & Kanozia, 2019). 그러므로 어린 시기부터 유튜브 영상을 놀이로 즐기는 유아와 미디어를 긍정적으로 생각하는 부모의 인식에 발맞춰, 유아교육도 유튜브 영상을 활용한 놀이에 관심을 가지고 유아교육 현장에 도입하려는 자세가 요구된다. 그리고 선행연구에서 제시한 유아용 뉴미디어 놀이 콘텐츠를 선택하는 기준(정지용, 2018)을 참고하여, 유아들의 발달 수준과 흥미, 재미, 유해성 등을 고려

한 교구로서의 동영상 도입해 보는 시도도 필요하다. 더 나아가, 교실에서 유아들과 함께 교실 촬영하기, 영상 속 주인공 되기 등 직접 영상을 촬영하는 것을 놀이로 시도해 봄으로써, ‘유튜브 영상’ 그 자체가 하나의 놀이 방식이 될 수도 있을 것이다.

그러나 본 연구는 여러 가지 한계점들도 지니고 있다. 끊임없이 만들어지고 업로드되는 유튜브 콘텐츠를 분석하는 것은 매우 어려운 작업이다. 본 연구에서 사용한 SNS Data Collector는 유튜브에서 제공하는 내부 알고리즘을 이용하기 때문에, 수집 시기에 따라서 데이터 결과가 다르게 도출될 수 있어 이를 고려한 수집 방법을 모색할 필요가 있다. 또한, 콘텐츠를 일정 시간으로 한정하여 수집하여 분석한 것은 변화무쌍한 유튜브 콘텐츠를 총체적으로 분석하기에 미흡함이 있다. 후속 연구에는 콘텐츠 수집 기간을 더욱 장기화하거나 연중 여러 기간을 설정하여 자료를 수집하면 더욱 신뢰성 있는 결과를 도출할 수 있을 것이다. 그리고 영상 제작자들이 자신의 동영상을 유튜브 알고리즘 상위에 랭크시키기 위해 제목에 최대한 많은 키워드를 넣거나 영상과 상관없는 현재 유행하는 키워드를 넣기도 하였다. 본 연구에서는 이러한 오염 요인을 정제하는 작업에 많은 시간과 노력이 소모되었다. 이를 보완하기 위해 키워드 분석 단계를 더욱 세분화하거나, 전문가 집단의 타당성 검증을 거쳐 더욱 구체적 키워드를 설정할 것을 제안한다.

참고문헌

- 강미리 (2021). 유아의 유튜브 시청 실태와 또래 유능성 간의 관계. 광주교육대학교 대학원 석사학위논문.
- 강소영 (2020). 유아의 유튜브 시청 실태 및 유튜브 시청현황에 따른 자기조절력의 차이. 명지대학교 교육대학원 석사학위논문.
- 강희주, 정의중 (2020). 아동 출연 유튜브에 나타난 아동학대 현황. **아동과 권리**, 24(4), 585-613.
- 곽선영 (2019). 유튜브 키즈콘텐츠의 구성적 특성과 적절성 탐색. 가천대학교 대학원 석사학위논문.
- 김대욱 (2019). 유아가 좋아하는 대중가요의 종류와 대중가요를 접하게 되는 경로 분석. **문화기술의 융합**, 5(2), 157-161.
- 김민경 (2019). 유튜브(YouTube)의 한국어 교육 채널 분석. **학습자중심교과교육연구**, 19(8), 941-964.
- 김민정, 최윤정 (2023). 부모의 시청 중재가 어린이들의 유튜브 시청 시간·시청 콘텐츠를 거쳐 정서적 불안정과 과잉 행동에 미치는 영향. **방송문화연구**, 35(1), 177-223.
- 김승민 (2018). 유튜브 생방송 활용 미디어 교육의 효과 탐색: 초등학생 대상 재난안전 프로

- 그램을 중심으로. 충남대학교 대학원 석사학위논문.
- 김은영, 임신일 (2021). 부모가 인식하는 유아의 유튜브 시청 실태 연구. **한국아동교육학회**, 30(1), 79-101.
- 김은진 (2021). 키즈 크리에이터 콘텐츠 수용 행태 연구: 보람튜브를 중심으로. **한국언론학보**, 65(2), 114-161.
- 김창숙 (2021). 어린이 동영상 콘텐츠 장르 분류: YouTube 어린이 동영상 콘텐츠의 내용, 형식, 제작자 특성을 중심으로. **사회과학연구논총**, 37(2), 183-213.
- 김태룡, 안승범 (2021). 뉴 노멀 시대 기독교 유튜브 콘텐츠의 기능과 가능성-국내 크리스천 1인 미디어 채널을 중심으로. **신앙과 학문**, 26(1), 165-193.
- 김형진, 정효정, 권보람 (2022). 디지털 네이티브 세대를 위한 유튜브 리터러시 체크리스트 개발 연구. **한국콘텐츠학회논문지**, 22(3), 273-284.
- 김희연, 김명희 (2012). TV 애니메이션 ‘뽀롱뽀롱 뽀로로’에 등장하는 캐릭터의 성차 분석. **열린유아교육연구**, 17(5), 611-634.
- 동폴잎 (2022). 포스트휴먼 시대의 새로운 유아들의 놀이와 학습에 대한 탐구: 디지털 놀이. **유아교육연구**, 42(6), 356-383.
- 모바일인덱스 (2023). **유튜브 사용자 조사**. <https://www.mobileindex.com/app/search>에서 2023년 10월 19일 인출.
- 민기채, 최소영, 조성은 (2019). 북한 영유아·아동 지원사업 네트워크 분석. **입법과 정책**, 11(3), 233-258.
- 박건령 (2023). LDA 토픽 모델링 기법을 활용한 국내 유·초 연계 관련 연구의 토픽 유형 및 동향 분석: 국가 수준 유치원 교육과정 개정 시기별 비교. **유아교육연구**, 43(1), 399-421.
- 박미선 (2020). 영유아 유튜브 시청 현황과 시청 현황에 따른 과몰입 차이. 중앙대학교 교육대학원 석사학위논문.
- 백민제 (2021). 어린이가 출연하는 유튜브 콘텐츠 보호에 대한 연구: 유튜브 키즈 콘텐츠와 키즈 유튜버를 중심으로. **연기예술연구**, 23(3), 247-264.
- 변현진 (2018). 유튜브 콘텐츠의 제작, 이용환경 특성과 인기 채널 분석 및 함의점 고찰. **조형미디어학**, 21(4), 227-239.
- 서희 (2023). 유아의 유튜브 콘텐츠 이용 실태와 문제행동의 관계. 한양대학교 대학원 석사학위논문.
- 송민 (2017). **텍스트 마이닝(Text Mining)**. 서울: 도서출판 청람.
- 송민주 (2020). 초등학생의 유튜브 활용에 대한 교육적 함의. **부산교육이슈페이퍼**, 44(9), 1-30.
- 양유진, 전희경 (2021). 유튜브 어린이 콘텐츠 현황 분석 및 적절성 평가. **어린이미디어연구**,

- 20(3), 27-57.
- 와이즈앱·리테일·굿즈 (2023). **앱순위 분석**. <https://www.wiseapp.co.kr/report/>에서 2023년 11월 16일 인출.
- 윤성진, 이관용, 홍정기 (2021). 토픽 모델링을 활용한 부동산 대책 및 사회적 이슈 분석. **不動産政策研究**, 22(1), 15-37.
- 이기창 (2019). **한국어 임베딩: 자연어 처리 모델의 성능을 높이는 핵심 비결**. 서울: 에이콘 출판.
- 이영주, 유수정 (2020). 유튜브 아동용 콘텐츠 정책 변화 이후 동영상 광고의 상품 유형 및 유효성에 관한 연구. **애니메이션연구**, 15(3), 114-136.
- 이정희 (2018). 유튜브 교육/강화 콘텐츠의 만족도와 지속 이용 의도에 관한 연구: 이용 동기, 신뢰도, 매체 특성을 중심으로. 동국대학교 대학원 석사학위논문.
- 이창호, 장석준 (2021). 어린이의 유튜브 및 게임이용시간에 영향을 미치는 부모 중재요인 분석: 어린이집, 유치원, 초등학교 저학년 비교. **미래사회의 삶과 교실: 유아를 위한 미디어의 과거, 현재, 미래**, 한국어린이미디어학회 추계학술대회 자료집, pp.102-110. 11월 20일. 나주시: 동신대학교(온라인 화상).
- 이현정 (2018). 크리에이터의 성별에 따른 유튜브 키즈 채널 놀이 콘텐츠 및 댓글 분석. 성균관대학교 대학원 석사학위논문.
- 정부민, 이효림 (2023). COVID-19시기의 영유아 건강 관련 뉴스 토픽모델링 분석 - 휴교를 중심으로. **열린유아교육연구**, 28(6), 269-291.
- 정원준 (2022). **NetMiner를 활용한 빅데이터 텍스트 분석 기법과 활용**. 서울: 학지사.
- 정지용 (2018). 유아용 뉴미디어 콘텐츠의 재미 요소. **한국콘텐츠학회논문지**, 18(10), 40-52.
- 조수민, 김민주, 안은미, 이운경, 이정애, 정익중 (2021). 토픽모델링을 활용한 아동 행복관련 기사분석. **한국아동복지학**, 70(2), 223-260.
- 조수선, 김봉현 (2021). 어린이의 유튜브 이용량에 미치는 영향 요인 연구. **한국엔터테인먼트산업학회논문지**, 15(7), 45-57.
- 진영주 (2023). 유아안전교육 유튜브 영상콘텐츠 현황 및 적절성 분석. 이화여자대학교 대학원 석사학위논문.
- 최미경 (2020). 유튜브 어린이 콘텐츠 연구. **스토리앤이미지텔링**, 19(-), 325-354.
- 최효정, 이근영 (2022). 유아교사들이 경험한 교수-학습 매체로서 유튜브(YouTube)에 대한 요구도 분석. **유아교육연구**, 42(4), 249-272.
- 한국언론진흥재단 (2020). **2020 어린이 미디어 이용 조사**. 서울: 한국언론진흥재단.
- Alghowinem, S. (2019). A safer youtube kids: An extra layer of content filtering using automated multimodal analysis. In Janusz Kacprzyk (Eds.), *Advances in Intelligent Systems and Computing* (pp.294-308). Switzerland: Springer International Publishing.

- Bennett, S. (2012). Digital natives. In Fischer, W. K., Katz, E. J., Rainie, L., Swan, K., Wellman, B., Young, K., & Zhao, Y. (Eds.), *Encyclopedia of cyber behavior* (pp.212-219). PA: IGI-Global.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(-), 993-1022.
- Cyram (2022). *NetMiner 4.5*. Seoul: Cyram Inc.
- Dehghani, M., Niaki, M. K., Ramezani, I., and Sali, R. (2016). Evaluating the influence of YouTube advertising for attraction of young customers. *Computers in Human Behavior*, 59(-), 165-172.
- Jindal, R., & Kanozia, R. (2019). *Do YouTube based children channels impact parenting? An exploratory study*. December 24, 2023 from https://www.researchgate.net/publication/333951387_Review_of_Research_DO_YOUTUBE_BASED_CHILDREN_CHANNELS_IMPACT_PARENTINGAN_EXPLORATORY_STUDY.
- Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J., & Scott, F. (2016). Digital play: A new classification. *Early Years*, 36(3), 242-253.
- Neumann, M. M., & Herodotou, C. (2020). Young children and youtube: A global phenomenon. *Childhood Education*, 96(4), 72-77.
- Papadamou, K., Papasavva, A., Zannettou, S., Blackburn, J., Kourtellis, N., Leontiadis, I., Stringhini, G., & Sirivianos, M. (2020). *Disturbed YouTube for kids: Characterizing and detecting inappropriate videos targeting young children*. Proceedings of the fourteenth international AAAI conference on web and social media. December 24, 2023 from <https://ojs.aaai.org/index.php/ICWSM/article/view/7320>
- Roberts Jr, J. M. (2000). Simple methods for simulating sociomatrices with given marginal totals. *Social Networks*, 22(3), 273-283.
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. CA: O'Reilly Media, Inc.
- Tahir, R., Ahmed, F., Saeed, H., Ali, S., Zaffar, F., & Wilson, C. (2019). *Bringing the kid back into YouTube kids: detecting inappropriate content on video streaming platforms*, Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. November 16, 2023 from <https://dl.acm.org/doi/pdf/10.1145/3341161.3342913>.
- Taylor, L. B., & Cingel, D. P. (2023). Predicting the use of YouTube and content exposure among 10 - 12-year-old children: Dispositional, developmental, and social factors. *Psychology of Popular Media*, 12(1), 20-29.
- Verenikina, I., & Kervin, L. (2011). iPads, digital play and pre-schoolers. *He Kupu*, 2(5), 4-19.
- Yadav, S., Chakraborty, P., Mittal, P., & Arora, U. (2018). Children aged 6-24 months like to watch YouTube videos but could not learn anything from them. *Acta Paediatrica Nurturing the Child*, 107(8), 1461-1466.

Abstract

LDA topic modeling analysis of YouTube content with the keywords “child” & “kids”

Lee, Keunyoung · Jung, Boomin

This study aimed to analyze YouTube content using the keywords “child” and “kids” to identify central themes and explore the digital sociocultural context related to young children. YouTube data were collected using NetMiner’s SNS Data Collector. Additionally, topic modeling based on Latent Dirichlet Allocation (LDA) was conducted using Excel and NetMiner 4.5.0. YouTube content was classified into five topics with the keywords “child” and “kids”: “Arts and Children,” “Religion and Children,” “Education and Children,” “Society and Children,” and “Play and Children.” In the keyword network of each topic, the mean distance between the nodes for all topics was statistically significant. The thematic flow of each topic was elucidated through keywords with a high co-occurrence frequency. We found that “Education and Children,” “Arts and Children,” and “Play and Children” were correlated to each other, while “Society and Children” predominantly focused on current societal issues. “Religion and Children” showed that YouTube’s influence extends beyond children entertainment and education. We hope this study provided basis for further research on YouTube content for young children.