Korean Journal of English Language and Linguistics, **Vol 25, March 2025, pp. 330-366**
DOI: 10.15738/kjell.25..202503.330

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

인공지능 디지털 영어교육의 현황과 전망: 자연어처리 기반 텍스트 분석

Youngkyo Oh (Chonnam National University)

This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: December 4, 2024
Revised: January 24, 2025
Accepted: March 8, 2025

Oh, Youngkyo
Lecturer, Dept. of Education, College of Education, Chonnam National University; 77, Yongbong-ro, Buk-gu, Gwangju, Korea, 61186
Email: 5young-kyo@hanmail.net

ABSTRACT

Oh, Youngkyo. 2025. Issues and prospects of AI digital English education: Text analysis Based on Natural Language Processing. *Korean Journal of English Language and Linguistics* 25, 330-366.

The study aimed to quantify the content and keyword relationships in AI-based English education texts to understand the current status of English education and gain insights into the future direction of English education. Data analysis involved a text analysis that utilized Natural Language Processing (NLP) techniques to extract meaningful content and information from large-scale text data, identifying new meanings and knowledge at the contextual level by considering the relationships between texts and words. The primary analysis methods used included keyword analysis, network text analysis, and topic modeling. The research findings are as follows. First, a frequency analysis of the collected texts revealed terms such as 'technology,' 'learning,' 'textbooks,' 'big data,' 'application,' 'mathematics,' 'information,' 'students,' and 'edtech' based on Term Frequency (TF) values. TF-IDF (Term Frequency-Inverse Document Frequency) values identified 'textbooks,' 'learning,' 'big data,' and 'edtech.' Additionally, N-gram analysis highlighted the term 'development and implementation of AI digital textbooks.' Next, through ego-network analysis for relationship exploration, words related to 'learning,' 'textbooks,' 'technology and future trends' were found to be connected around the term 'English.' Finally, topic modeling analysis using Latent Dirichlet Allocation (LDA) classified the texts into five topics: 'AI-Based Educational Innovation,' 'Transition to a Digital-Based Learning Environment,' 'Future-Oriented Changes in the Education System,' 'Learner-Centered Personalized English Education,' and 'Technological Innovation and Services in the Education Industry.' This study provides insights into the future direction of English education by providing educators and learners with AI-powered text analysis tools.

KEYWORDS

natural language processing, AI digital English education, network text analysis, topic modeling

© 2025 KASELL All rights reserved

1. 서론

국가 차원의 학령인구 감소로 인한 교원 임용 축소와 더불어 사교육의 온상으로서 지목되면서 수능 영어 절대 평가 도입과 고교 학점제 도입으로 인한 영어 선택과목 개설 감소 등의 교육정책과 인공지능 기술을 통한 외국어 학습 대체, 그리고 COVID-19로 인한 언택트 교육에 이르기까지 영어교육은 전방위적인 위기를 맞이하고 있다. 디지털 대변혁 시대에 영어교육의 위기는 더욱 심화될 것이라고 예측된다(Hwang 2020, Lee et al. 2021, Yeo 2020). 특히, 인공지능 기반 디지털 기술에 따른 영어교육의 교육과정 및 평가, 그리고 학습관리 및 교수법 전반의 변화가 가속화되고 있다(Hwang 2020). 이러한 상황에서 우리 영어교육의 현황을 면밀히 점검하고 미래 영어교육을 대비하려는 노력이 필요한 때이다.

미래 사회는 다양한 언어의 국가, 집단, 개인 간의 의사소통이 더욱 활발해질 것이고 효과적인 소통을 위해 영어의 기능이 더욱 강조될 것이다. 또한 영어는 디지털 세계에서 다양한 지식과 정보를 처리하고 공유하는 코딩과 프로그래밍, 그리고 인터넷에서 주로 사용되는 언어이다. 영어는 의사소통 도구나 지식의 생산이나 전파를 위한 기능어 역할 뿐만 아니라 학습자의 미래 핵심역량을 함양하는 언어가 된다. 즉, 미래 사회에서 영어학습은 단순한 영어학습 목표를 너머 미래사회에 필요한 핵심역량을 함양하는데 초점을 맞춰야 한다.

COVID-19 팬데믹 기간의 원격교육과 하이브리드 교육모델은 교육 방식의 급격한 전환을 야기했고, 이는 영어교육 방법론, 컨텐츠 개발, 교수법, 평가 방식, 교사와 학생 간의 상호작용 방법에 커다란 영향을 미쳤다(이은혜 외 2023). 포스트 코로나 시대에 인공지능 디지털 기술이 발전하면서 학습자의 영어 능력 개발과 교육 접근성에 새로운 도전과 기회를 제공하였다. 교육자들은 디지털 학습플랫폼을 통해 창의적이고 상호작용적인 교육 방법을 개발하고, 이 과정에서 다양한 디지털 도구와 애플리케이션을 통합하는 방법을 모색했다(James and Théiault 2020).

디지털 대전환으로 인해 디지털 기술 이해 및 활용에 관한 교육적 요구가 증가하고, 미래 사회가 요구하는 창의적 인재양성이 필요하다는 사회적 인식에 따라 우리 정부는 디지털 인재 양성 종합방안을 마련하여 디지털 교육 전환 정책을 시행하였다(MoE 2022b). 다른 나라에서도 국가차원의 AI 인재 양성을 위해 다양한 디지털 교육정책을 이미 시행하고 있다(Chun et al. 2021, Im 2022). 우리나라에서는 교육부와 교육 개발원, 평가원 등 유관 정부 기관을 중심으로 AI 교육 및 종합방안을 수립하고, 2022 개정 교육과정에서도 디지털 소양을 포함하여 교수·학습에 AI를 비롯하여 에듀테크 활용을 강조하고 있다(MoE 2022a). 주요 목표로 디지털 분야 인재양성, 디지털 교육체제 전환, 에듀테크 활성화, 디지털 기반 교육혁신 역량강화 지원사업 등이 포함된다. 특히, 모두를 위한 맞춤형 교육 실현을 위한 AI 디지털 교과서 추진 방안 정책을 통해 개인 맞춤형 학습과 AI 기반 디지털 교과서 도입할 계획이다. 이를 통해 학교 현장에서 디지털 인프라를 구축하여 학생들의 디지털 역량을 강화하고 있다(MoE 2023).

영어교육 분야는 다른 분야에 비해 디지털 기술의 활용이 두드러지게 나타난다. 기존의 CD나 DVD 멀티미디어 활용을 거쳐 문법 검토, 영작문 자동 교정도구, 챗봇, 기계번역, 메타버스 등을 다양한 영어 교수·학습 보조도구가 활용되고 있다(Cha et al. 2021, Chun et al. 2021, Dizon 2020, Lee 2020). 최근에는 인공지능 기술 발달로 교육 현장에서 AI 활용이 학습관리, 학습지원,

© 2025 KASELL All rights reserved

교수·학습 방법측면에 도움을 줄 뿐만 아니라 연구방법에도 큰 영향을 미치고 있다(Chun et al. 2021). 디지털 기술의 발전은 영어 학습자들에게 자기주도적 학습이 가능한 학습플랫폼을 제공하고 여기에 AI 기술이 접목되어 실제적 영어능력을 함양할 수 있는 개별 맞춤형 상호작용 언어학습 환경이 마련되었다(Chun et al. 2021, Godwin-Jones 2019, Yun and Maeng 2017).

디지털 영어교육 분야의 연구 동향을 살펴본 선행연구에서 Lee(2020)의 연구는 2000년부터 2020년까지 발표된 연구를 연구 주제, 방법, 내용 중심으로 연도별 추이 및 현황을 살펴보았으며, Lee와 Kim(2022) 연구도 초등학교급에 대해 2017년부터 2021년까지 연구 시기, 내용, 방법을 살펴보았다. Chun 외 2인(2021)의 연구는 2016년부터 2020년에 해당되는 국내외 AI 활용 실험연구 24편에 대해 연구 주제, 대상, 방법 및 교육의 효과성을 살펴보았다. Cha 외 2인(2021)의 연구는 AI 챗봇에 관한 연구로 국한하여 국내 40편의 연구를 단순히 범주로 분류하고 그 특징을 비교 분석하였다.

그러나 영어교육 현황과 전망에 대한 선행연구는 주로 학술지에 게재된 논문에 국한하여 연구 동향을 분석하였기 때문에 실제 디지털 영어교육의 현실을 파악하여 미래 영어교육의 방향에 대한 시사점을 제시하기에는 한계가 있다. 이에 이 연구에서는 선행연구의 이러한 단점을 보완하여 빅데이터를 대상으로 체계적이고 객관적인 텍스트 분석 방법을 사용하여 인공지능 기반 디지털 영어교육 동향을 파악하고 미래 영어교육 방향을 설정하는데 시사점을 얻고자 하였다.

텍스트 분석은 언어학과 통계학을 접목한 기계학습을 기반으로 한 자연언어 처리기법을 활용하여 대규모 텍스트 데이터를 분석하는 방법이다. 텍스트 분석기법은 온라인 네트워크상에서 다양한 제품과 서비스에 대한 고객의 리뷰 또는 각종 사회 이슈에 대한 일반 대중의 댓글과 의견 및 여론 등을 분석하는 데 적용되어 왔다(Cho et al. 2017, Kim and Kang 2018). 이 기법의 핵심은 방대한 텍스트 내에 존재하는 키워드를 도출하여 키워드 간의 관계를 규명하는 데 목적이 있다(Kim et al. 2016).

따라서 이 연구에서는 자연어처리 기법에 기반을 둔 텍스트 분석 방법을 사용하여 인공지능 기반 디지털 영어교육에 대한 사회적 이슈와 동향 전반적인 맥락과 흐름을 파악하여 미래의 영어교육 방향과 비전을 보다 객관적이고 체계적으로 제시하고자 한다. 이 연구는 인공지능 기반 디지털 영어교육 관련 빅데이터를 데이터 분석가와 영어 교육자 입장에서 분석하고 이해하기 위해 NLP 기법을 활용하였다. 이 연구에서는 인공지능 기반 디지털 영어교육에 대한 사회적 인식과 이슈를 분석하고 주요 논의에 어떤 키워드와 주제가 포함되어 있는지 분석하여 향후 영어교육에서 인공지능 기반 디지털 기술을 어떻게 활용할지 방향성을 탐색하고자 한다. 이를 위해 다양한 온라인 검색 포털 채널을 통해 디지털 영어교육 관련 텍스트 데이터를 수집하고 전처리한 후 텍스트 마이닝 기법 중 키워드 분석, 네트워크 텍스트 분석, 토픽 모델링으로 분석하고자 한다. 이에 따른 연구 문제는 다음과 같다.

첫째, 인공지능 기반 디지털 영어교육 분야 관련 주요 키워드는 무엇인가?
둘째, 인공지능 기반 디지털 영어교육 분야 관련 주요 토픽은 무엇인가?
셋째, 인공지능 기반 디지털 영어교육 분야 관련 텍스트에 나타난 사회적 태도나 감정은 어떠한가?

© 2025 KASELL All rights reserved

2. 이론적 배경

2.1 디지털 교육정책 변화

교육부는 1996년 제1차 교육정보화 기본계획을 수립 이후, 매 5년 단위로 교육정보화 종합계획을 수립해오고 있다. 제6차 교육정보화 기본계획(2019~2023)도 이러한 연속성 선상에서 국가 수준의 교육정보화를 체계적으로 추진하고 있다. 2011년에는 미래 사회에 대비하기 위해 교육체제를 대폭 혁신하여 학습자의 미래핵심 역량 강화를 목표로 하는 2011 스마트교육 추진전략이 제안되었으며, 2016년에는 지능정보사회에서 우리나라 교육이 나아가야 할 미래 교육의 방향을 제시한 중장기종합계획이 발표되었다(MoE 2016). 이를 바탕으로 2019년에는 'AI 국가전략'을 발표하여 학생들이 어린 시절부터 쉽고 재미있게 SW와 AI를 접하고, 전 국민이 AI 기초역량을 배울 수 있는 교육체계를 구축하는 전략을 세웠다(MSICT 2019).

2020년 제1차 정보교육 종합계획(2020~2024)에서는 인공지능으로 대변되는 미래 지능정보사회의 발전을 주도하는 세계적 인재육성을 목표로 교과 사이의 연계와 융합, 인공지능, 메타버스, VR, AR 등의 에듀테크 기술 도입을 위한 목표를 세웠다(MoE 2020). 2022년 디지털 인재양성 종합방안으로 인해 AI 시대에 적극적으로 대응하기 위안 디지털 인재양성을 추진하고 AI 기반 디지털 교육 기반 마련과 교육정책을 수립하여 우수한 디지털 인재를 양성하고자 하였다. 2023년 교육부에서 디지털 기반 교육 혁신 방안을 발표하여 학생 맞춤형 교육 실현이라는 비전 아래 AI 디지털 교과서(AIDT) 개발, 디지털 리터러시 역량을 갖춘 교원양성을 위한 연수, 디지털 기술 활용 교수·학습방법 개발 등의 교육 정책을 추진하고 있다(MoE 2023). 이러한 AI 기반 디지털 교육정책들은 모두 교사와 학생의 AI 및 디지털 활용 역량 강화를 강조하고 있다. 특히 최근 들어 AI 교육, 코딩 교육, 소프트웨어 교육 등이 관심을 끌면서, 교사와 학생들의 디지털 리터러시 역량을 강화하고, 학교 현장에서 디지털 기술을 활용하는 것이 교육정책의 트랜드로 자리잡고 있다.

국내의 AI 및 디지털 교육 정책에 대한 연구는 먼저, 학교 현장에서 교사의 AI·디지털 활용 교육 실태 조사와 디지털 교육정책의 내실화 방안을 도출한 연구(Kang et al. 2023)와 Fairclough의 비판적 담론 분석을 통해 인공지능 교육정책의 핵심 담론과 질서, 담론이 정당화되기 위한 전략에 대한 연구를 수행하여 인공지능 교육 정책 변화의 필요성을 제언한 연구(Park 2024)가 있다. 또한, 디지털 인재 양성을 위해 교원의 디지털 리터러시 역량이 강조되는 점에 착안하여, 다차원 교육정책 분석 모형을 통해 디지털 기반 교원 연수 정책을 분석하고, 정책의 개선점과 방향을 제언한 연구도 있다(Jang et al. 2024).

교육부는 디지털 기반 학교 수업에서 학생들의 역량 차이에 따라 나타날 수 있는 학습 결손과 교육격차 문제를 예방하기 위해 디지털 튜터를 도입하였다. 이러한 인공지능 기반 디지털 튜터와 관련된 정책적 과제와 해결을 위한 방안 연구(Shim 2023)와 디지털 튜터 사업의 성공적인 안착을 위해 디지털 튜터가 갖춰야 할 기능과 역량을 도출한 연구(Hong et al. 2024)가 수행되었다.

이와 같이 AI 기반 디지털 교육 정책에 관해 선행 연구가 이루어지고 있으나, 기존 연구는 주로 정책의 내용 검토를 통한 정책 제언에 초점을 둔 반면, 인공지능 기반 디지털 교육 정책과

© 2025 KASELL All rights reserved

이슈와 관련한 실제 다양한 온라인 웹 자료, 뉴스 기사, 보도자료, 블로그, 카페 등 전체 텍스트를 대상으로 한 텍스트 분석은 부족한 실정이다.

2.2 인공지능 기반 영어교육 디지털 플랫폼

인공지능 기반 디지털 교육의 등장은 영어교육에 큰 변화를 가져왔다. 선행연구에 따르면 전통적인 교실에서는 시간과 공간의 제약이 있지만, 디지털 교수·학습 플랫폼의 도입으로 학생들은 언제 어디서나 자신에게 적합한 학습 자료에 접근할 수 있게 되었다. 즉, 디지털 플랫폼을 통한 멀티미디어 자료와 상호작용이 가능한 학습 콘텐츠를 통해 학생들의 학습 경험은 다양하고 풍부해지고 있다(Cho et al. 2017, Jeong 2020, King et al. 2000, Yun and Maeng 2017).

특히, AI 기술의 발달로 인한 디지털 교육 환경에서 제공되는 학생 개별 맞춤형 학습은 학습의 효율성을 크게 향상시키며, 학생들의 동기부여와 자기주도적 학습능력을 기르는데 도움이 되었다(Godwin-Jones 2019, Kim 2018, Luckin et al. 2016). 특히, 영어교육과 관련하여 다양한 실제적인 언어자료를 활용하여 자연스럽게 소통할 수 있는 쌍방향 언어기능 연습 기회를 제공한다(Godwin-Jones 2019, Hubbard 2017, Kim et al. 2021).

최근 들어 AI 기반 학습플랫폼은 영어교육에서 다양한 형태로 교사 지원 및 학생 학습활동 지원 도구로 활용되고 있다. AI 도구는 검색, 기계번역, 음성인식, 이미지나 영상 생성 등을 통해 수업에서 학생들의 학습을 지원한다(Chun et al. 2021, Kim 2019). AI 활용 튜터와 같은 기능은 학습자의 인지능력, 동기, 학습유형 및 수준 등을 분석하고 이에 적합한 개별 맞춤형 학습을 제공하며 학습자의 진로 선택과 설계를 돕는 다양한 옵션을 제공한다(Lee 2020, Luckin et al. 2016, Zou 2017). 즉, AI는 디지털 튜터가 되어 학습자의 활동 이력을 추적하며 문제를 파악하고 해결법을 제안해 학습 효과를 향상시킨다.

AI는 교사처럼 학생에게 학습에 대한 다양한 선택과 대안을 제공하며 학습자가 필요로 하는 학습 자료를 제공하고, 학습활동을 모니터링하여 학습 문제점을 찾아 해결책을 제공한다(Godwin-Johones 2019, Kramsch et al. 2013, Zhang 2020). 더불어, 디지털 보조교사는 교사의 행정 업무를 지원하기 위해 과제 및 평가의 자동 채점, 오류 분석, 자동 피드백 등의 기능을 수행하여 교육의 효율성을 높이며 교사가 학생을 더 잘 지도할 수 있게 도와준다(Chun et al. 2021, Holmes et al. 2019, Kim et al. 2020, Park and Yang 2020).

영어교육 분야에서 인공지능을 활용한 영어 수업(신동광 2019, 이삭 2019, 최원경 2020), 메타버스, 가상현실을 활용한 영어 수업(이규영, 김희진 2020, 최원경 2021, 황요한 2021)처럼 영어 수업에서 디지털 리터러시를 적용한 사례들에 관한 연구가 진행되었다(정대연, 배지영 2021, Kim 2020). 교수·학습 활동 상황에서 디지털 기술의 가장 큰 장점은 개별화 수업이 가능하다는 것이다(Green et al. 2005). 기존교실 수업에는 교사 한 명이 다수의 학생들과 상호작용했다면, 학생들은 자신의 디지털 기기를 통해 AI 챗봇이나 튜터를 활용하여 각자의 수준과 관심사에 맞는 개별화 학습활동을 할 수 있다. 이러한 맞춤형 개별화 수업은 학생들이 학습에서의 자율성을 기를 수 있다는 점에서 2022 개정 교육과정이 추구하는 자기주도성과 연계된다(Savage and Barnett 2015).

또한, 학습자들은 디지털 소양의 함양을 통해 사회정서적 측면에서도 교육 효과를 경험할 수 있다. 영어 수업에서 AI 챗봇, 메타버스, 가상현실 등 디지털 도구를 활용한 다양한 형태의

© 2025 KASELL All rights reserved

수업이 학생들의 참여와 흥미를 개선시키고 결국 영어 실력을 높이는 데도 도움이 되었다(신동광 2022, 이삭 2019, 최원경 2021, 황요한 2021). 그러나 새로운 디지털 도구를 활용한 수업에 대한 참신성 효과가 사라지면 학생들의 흥미와 몰입도가 급격히 감소하기 때문에(최원경 2021), 이를 보완하려는 교사의 노력도 필요하다.

마지막으로, 교사는 학생들의 디지털 소양 함양을 위해서 다양한 디지털 도구를 적절하게 활용할 수 있도록 관련 디지털 리터러시를 길러야 한다. Koehler와 Mishra(2009)는 기존의 내용지식, 교육학적 지식과 더불어 오늘날 교사가 반드시 갖추어야 할 지식으로 기술적 지식을 제안하여 TPACK(technological pedagogical and content knowledge)이라는 개념 체계를 제시하였다. 최근 인공지능 기반 디지털 기술이 교육 분야에 활용되는 에듀테크라는 용어가 생겨났고(박지수, 길준민 2020), 2022 개정 영어과 교육과정에도 디지털 소양이 처음 등장하는 등, 영어과에서 디지털 기술 활용을 적극적으로 권장하고 있다. 영어 교수·학업과 관련된 다양한 인공지능 기술 지식을 활용할 수 있는 교사와 그렇지 않은 교사 사이의 교육격차가 확대될 가능성이 크다. 따라서 영어 교사는 영어 교과에 대한 내용지식, 교수법, 학습자 이해 등 교육학적 지식과 함께 인공지능 기반 디지털 지식을 통합하여 수업에 적용할 수 있는 디지털 리터러시 역량을 길러야 할 것이다.

2.3 인공지능 기반 디지털 영어교육 선행연구

포스트 코로나 영어교육 전문가와 다양한 이해관계자들의 주된 관심은 디지털 기술이 영어 교수·학습에 어떻게 접목되며 앞으로 교수·학습 패러다임이 어떻게 변화될 것인가이다. 인공지능 기반 디지털 기술의 적용은 코로나19와 함께 전면적으로 적용된 실시간 화상 강의, 인공지능 기반 교수학습 프로그램, 메타버스 등의 가상공간을 활용한 외국어 교육 플랫폼 등은 디지털 신기술에 대한 이해와 활용을 요구하며 기존 수업의 패러다임을 바꾸고 있다(Persson and Nouri 2018).

코로나19 기간 동안 국내 영어교육에 대한 웹 데이터를 텍스트 네트워크 분석한 선행연구에서 영어교육 네트워크의 주요 키워드, 참여 주체 및 학습 공간별 특성, 그리고 가장 큰 영향력을 미치는 주체를 확인하였다. 키워드 네트워크 분석과 웹문서 4,497건에 대한 에고 네트워크 분석 결과, 학습, 학교, 학생, 수업, 온라인, 교사, 학원 등이 상위에 포함되었다. 특히, 학생, 학교, 온라인, 학부모의 네트워크 크기가 컸으며, 특히 학부모와 온라인 노드의 영향력이 두드러져 네트워크에서 중요한 위치를 차지하였다(권선희 2023).

또 다른 선행연구에서는 한국멀티미디어 언어교육학회 학회지(MALL) 2002년부터 2021년 사이에 게재된 총 567편의 논문을 대상으로 멀티미디어와 테크놀로지와 관련된 언어 교육 연구의 양적 추이, 네트워크 분석, 클러스터 분석을 실시하였다(Nam 2022). 그 결과 2010년도에 영어교육 분야에서 멀티미디어나 테크놀로지를 활용한 논문이 늘어났으며, 2020년도에는 COVID-19 팬데믹으로 인해 원격교육에 대한 논문의 수가 증가하였다. 분석 방법으로 키워드, 논문 제목, 초록에서의 명사구 형태의 용어를 추출하여 출현 빈도에 따라 네트워크 시각화 분석을 한 결과 20년간 '블렌디드 러닝', '초등', '한국 대학'이라는 단어와 관련된 연구가 주를 이루고 있었다. 또한, MALL에 게재된 논문의 키워드 단어를 군집 분석한 결과 '블렌디드 러닝', '글쓰기 능력', '초등학교 학습자', '컴퓨터기반 언어(교육)', '영어교육'이라는 연구 분야와 주제로

© 2025 KASELL All rights reserved

분류하였다.

최근 디지털·AI 기반 영어교육 연구 동향을 텍스트 네트워크 분석 및 토픽모델링으로 분석한 선행연구 결과에 따르면 토픽 1 '교수·학습 모형', 토픽 2 'AI 학습 어플리케이션', 토픽 3 '읽기 전략 및 평가', 토픽 4 '온라인 학습 및 콘텐츠', 토픽 5 '디지털·AI 도구를 활용한 쓰기 피드백', 마지막 토픽 6은 '개별화 학습요인 및 효과'가 보고되었다(맹은경 외 2023). 또 다른 선행연구에서는 AI 디지털 교과서 도입에 대한 우려와 기대가 공존하는 상황에서 인식 조사를 통해 맞춤형 학습경험, 실시간 피드백, 평가의 자동화 등 긍정적인 면과 디지털 과의존과 문해력 저하, 디지털 격차 확대 등 부정적인 면을 확인하고 AI PengTalk과 같은 AI 기반 도구를 활용한 챗봇 영어교육이 학습과 정의적 요인을 개선시킨다는 연구 결과를 보고하였다(심창용 2024).

디지털 환경에서 자란 '디지털 세대'(Prensky 2001)는 서책형 학습 자료보다는 디지털 자료에 더 친숙하고, 흥미와 학습 동기를 보인다. 다양하고 풍부한 학습 자료를 가지고 개별화 학습 맞춤형 학습이 가능한 학습 환경을 제공하기 위한 인공지능 디지털 교과서 개발이나 효과적 교수-학습 활용 방안에 관한 연구가 필요한 실정이다(김정렬 2015, 임희정 2014). 선행연구에서 학교 영어교육에서 서책형 교과서보다 디지털 교과서를 적용했을 때 학습 효과가 나타나는 연구 결과는 영어 인공지능 디지털 교과서를 활용한 영어교육의 가능성을 시사한다(김인옥 2019, 한종임 2014).

그러나 최근 중등 영어교육 분야의 AI 디지털 교과서 도입 쟁점에 대한 연구 결과 AI 디지털 교과서의 제작만으로는 성공이 보장되지 않으며 영어 교사가 AI 교과서의 다양한 기능과 분석 역량을 얼마나 잘 활용하는지에 달려 있음이 드러났다(박문영 2023). 즉, 다양한 최신 AI 자원을 교육적으로 통합하는 교사의 역량이 강조되었다. 또한, 초등 영어교육에서 AI 디지털 교과서 전망을 연구한 결과 AI 기반 학습도구와 디지털 교과서가 초등학생 영어학습에 일부 긍정적인 영향을 미치지만, AI 음성인식 기술의 부정확성, 과도한 디지털 의존, 급격한 교사의 역할 변화, 초등학생의 신체적·정신적 문제, 학습 적응과정에서의 어려움, 다양한 디지털 기기의 활용 문제, 빠른 개발 일정과 높은 비용, 개인정보 및 보안 문제 등의 이슈를 제기하였다(이정화 2023).

2.4 텍스트 분석기법

텍스트 분석기법으로 Feldman과 Dagan(1995)에 의해 처음 언급된 텍스트 마이닝은 데이터베이스 상의 지식발견이다. 즉, 정보 검색과 정보 추출을 위해 자연어처리(NLP) 기술을 사용하여 텍스트 속의 알려지지 않은 유용한 패턴과 새로운 지식을 발견하는 것을 말한다. 즉, 비정형 텍스트 데이터를 구조화하여 의미 있는 지식, 패턴, 추이를 도출해내며, 이를 통해 사용자가 원하는 형태로 데이터의 내재적인 정보를 추출하는 것이다(송민 2017). 텍스트 마이닝은 통계적 설문조사 기법과 유사하지만, 분석 대상을 모집단인 빅데이터를 대상으로 한다는 점이 특징이다(Lau et al. 2005).

텍스트 마이닝은 데이터 전처리 과정과 데이터 분석 과정으로 나뉘며, 전처리 과정은 비정형화된 데이터를 분석이 가능하도록 가공 및 정제하는 단계이고, 데이터 분석은 머신러닝 알고리즘을 이용하여 텍스트에서 의미 있는 정보를 추출하는 단계이다(Hotho et al. 2005). 연구의 주제나 연구자의 의도, 데이터의 성격에 따라 다양한 분석 알고리즘을 적용할 수 있으며, 연구의 목적에 적합한 분석기법을 선정해야 한다).

© 2025 KASELL All rights reserved

구체적으로 텍스트 마이닝의 절차는 데이터 수집, 데이터 전처리 및 정제, 텍스트 분석, 결과 해석 및 논의의 4단계로 이루어진다(Kim and Park 2023). 먼저 데이터 수집 단계에서는 뉴스 기사, 책, SNS, 논문, 웹문서 등 인터넷의 모든 텍스트 데이터를 대상으로 문서를 수집한다. 이때 연구하려는 텍스트 데이터의 기준과 범위를 명확히 세워야 다음 단계에서의 혼란을 줄일 수 있다. 데이터 전처리 단계는 자연어 상태인 텍스트 데이터에서 분석에 불필요한 요소를 제거 또는 대체하고 분석할 수 있는 정형화된 형태로 바꿔주는 구조화 과정이다. 데이터 정제를 위해 특수문자, 공백 등 불용어를 삭제하고 텍스트를 기본 단위로 나누는 토큰화 작업을 거친다. 이후 형태소를 추출해 품사를 분류한 후 분석의 대상이 되는 유의미한 품사만을 따로 추출한다(송민 2017).

다음으로 텍스트 분석 단계에서는 전처리 된 자료를 대상으로 범주화, 분류, 문서 요약 등의 정보 추출기법을 사용하여 텍스트 분석을 수행한다. 마지막 결과 해석 및 논의 단계에서는 분석 결과를 바탕으로 의미 있는 정보와 지식을 도출한다(Punitha et al. 2014). 이 연구에서는 자연어처리 기반 텍스트 마이닝 분석 방법인 키워드 분석, 네트워크 텍스트 분석, 토픽 모델링을 적용하였다(김성근 외 2016).

3. 연구방법

3.1 데이터 수집 및 전처리

연구를 위해 국내외 주요 온라인 포털 사이트 데이터를 웹크롤링 검색한 후 텍스트 마이닝의 대표적인 분석 방법인 자연어처리 기법으로 분석하였다. 자연어처리는 텍스트 데이터에서 정보를 추출하고 의미 있는 지식으로 전환하는 기법으로 문서 분류, 군집화, 정보 추출, 요약 등 다양한 분야에 활용된다. 분석과정은 우선 수집된 데이터 전처리 후 정제된 텍스트를 대상으로 키워드 분석 및 시각화, N-gram 분석, 네트워크 텍스트 분석(에고 네트워크), 토픽 모델링을 진행하여 인공지능 기반 디지털 영어교육 분야에 관한 최근 논점과 현황 그리고 미래 전망을 하고자 하였다. 주요 분석과 모델링은 파이썬을 기반으로 한 텍스톰(TEXTOM)을 활용하였다. 구체적인 연구 절차는 [그림 1]과 같다.

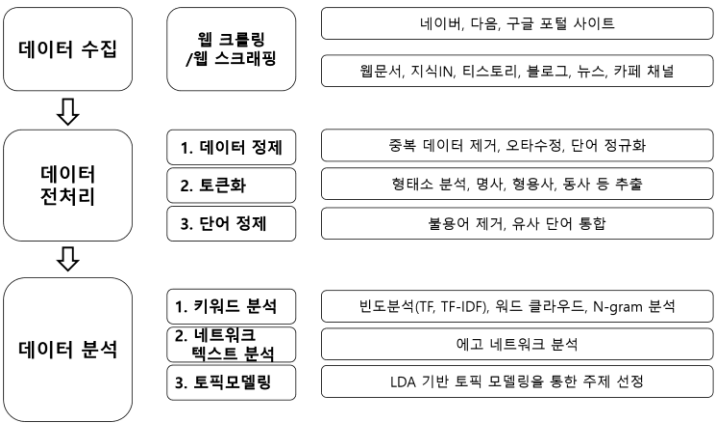

그림 1. 연구절차

© 2025 KASELL All rights reserved

3.1.1 데이터 수집

웹 크롤러/스크래퍼는 인터넷 사이트에서 원하는 웹 정보를 자동으로 수집하는 도구로, 웹사이트의 특정 영역과 내용을 지정하고 조건을 설정하면 자동으로 정보를 수집하고 저장하는 기능을 수행한다. 데이터 수집 채널로는 대표적인 포털 사이트인 네이버(블로그, 뉴스, 카페, 지식IN, 웹문서), 다음(블로그, 뉴스, 카페, 웹문서), 그리고 구글(페이스북, 뉴스, 웹문서)을 통해 인공지능 기반 디지털 영어교육 관련 정보를 수집하였다. 다음으로 수집된 텍스트 데이터를 분석하여 키워드를 도출하고, 이를 매트릭스 모델로 구성하여 기초분석 자료를 구성하였다. 자료 수집기간은 최근 포스트 코로나 시기 중 디지털 교육이 활성화되는 기간을 중심으로 2021년 11월 7일부터 2024년 11월 6일까지 3년으로 설정하여 월 단위로 수집하였다. 데이터 수집을 위한 키워드는 “인공지능 + 디지털 + 영어 + 교육”이다.

3.1.2 데이터 전처리

자연어 데이터인 텍스트를 컴퓨터가 분석할 수 있도록 구조화하는 데에 인덱스가 사용된다. 이 처리 과정에서 텍스트의 핵심 내용을 반영하는 키워드를 추출하고, 단순 문법 기능을 수행하는 불필요한 단어는 불용어로 제거한다. 이후 형태소 분석기를 활용하여 문장을 품사별로 나누고, 이 중에서 명사나 동사, 형용사와 같은 중요한 요소를 추출하는 라이브러리를 적용한다. 데이터 전처리는 텍스트를 분해, 정제, 여과하는 과정으로 자연어처리의 가장 기초적이면서 중요한 과정이다.

데이터의 전처리 과정은 1단계 데이터 정제, 2단계 토큰화, 3단계 단어정제의 3가지 단계로 진행되고, 이 연구에서는 정제/형태소 분석기를 사용하여 데이터의 전처리 과정을 진행하였다. 1단계 데이터 정제방법으로는 직접 선택과 자동정제 중 직접 선택을 통해 텍스트 제목과 본문이 포함된 열을 대상으로 데이터를 정제하였다. 데이터 정제 단계 중 분리정제는 URL, 제목과 본문 전체를 선택하여 제목과 내용을 통합하여 분석을 진행하였다. 중복 제거는 본문을 기준으로 동일한 데이터를 삭제하였다. 또한, 이미 데이터 수집 단계에서 주제와 관련 있는 뉴스나 웹문서만 수집하였기 때문에 특정 단어를 포함하거나 제외한 데이터만 추출하는 필터링 과정은 생략하였다.

2단계 토큰화 및 품사추출을 위한 형태소 분석을 수행하였다. 고유명사, 복합명사를 그대로 결과값에 반영하여 분석하는 Expresso K 분석 모듈과 한국어 형태소를 분석하기 위해 만들어진 소프트웨어인 Mecab-ko 중에서 Mecab-ko 분석 모듈을 활용하였으며, 분석 품사로는 명사, 형용사, 동사, 영어를 대상으로 하였고, 명사를 세분화하면 일반명사, 고유명사, 의존명사, 단위명사, 수사, 대명사로 이루어졌다.

3단계 단어정제에서는 데이터 정제 및 형태소 분석을 통해 추출된 단어는 오타가 포함된 단어나 조사가 제거 또한 불용어와 동의어들이 많기 때문에, 데이터를 편집하는 과정에서 직접 조사를 추가하거나 수정하고 불용어를 제거하였다. 품사 추출 후 키워드 정제 기능으로 불용어를 제거하고 유사 단어를 통합하였지만 한 글자로 이루어진 단어나 무의미한 단어를 2차로 정제하는 과정을 거쳤다. 불용어 제거 및 유사 단어 통합의 예는 다음 [표 1]과 같다.

© 2025 KASELL All rights reserved

표 1. 불용어 제거 및 유사 단어 통합 예시

과정	예시
불용어 제거 (한국어/영어)	등, 있, 것, 이, 주, 되, 외, 같, 만, 받, 년, 수, 일, 중, 월, 하, 개, 명, 차, 말, 초, 전, 많, 억, 시, 나, 보, 때, 읽, 사, 맞, 좋, 위, 내, 데, 간, 분, 부, 후, 높, 알, 오, 앞, 쓰, 번, 회, 장, 속, 들, 고, 그, 곳, 넘, 조, 가, 천, 크, 안, 세, 거, 바, 트, 기, 호, 미, 급, 관, 금, 달, 씨, 목, 과, 저, 열, 배, 문, 영, 백, 선, 살, 공, 풀, 의, 남, 얻, 업, 티, 제, 발, 방, 플, 건, 등등 b, D, A, C, B, G, X, T, of, and, I, S, to, c, in, E, for, e, Q, R, the, s, The, a, M, n, p, N, vs, x, Y, m, Z, W,
유사 단어 통합	챗 GPT→챗GPT, 플랫 폼→플랫폼, 메타 버스→메타버스, 리터 러시→리터러시, 튜터→튜터, 자율 주행→자율주행, 조 회수→조회수, 상호 작용→상호작용, 펭 톡→펭톡, 자바 스트립트→자바스크립트, 첨부 파일→첨부파일, 트랜스 포메이션→트랜스포메이션, →설 명회→설명회, 뤼 이드→뤼이드, NE 능률→NE능률, 부총리 겸→부총리겸, 미드저니→미드저니, youtu be→youtube, 헬 스케어→헬스케어, 인텔 리전→인텔리전, 템플릿→템플릿, 사이언 티스→사이언티스, 베리 타스→베리타스, Pre trainded→Pretrained, 빅→빅데이터, 데이터→빅데이터, 인공→인공지능, 지능→인공지능, 듀→에듀, 챗→챗봇, 봇→챗봇, AI→인공지능, 딥→딥러닝 등등

3.2 분석절차

3.2.1 키워드 분석

이 연구에서는 인공지능 기반 디지털 영어교육 관련 인터넷 웹문서나 뉴스의 제목과 본문에서 자주 언급되는 주요 키워드를 추출하고 빈도에 따른 중요도를 분석하기 위해 키워드 분석을 실시하였다. 텍스트 마이닝의 가장 기초적인 분석단계로 텍스트 데이터에서 명사, 형용사, 동사 같은 단어들을 추출하여 그 빈도를 분석하면 자주 등장하는 단어들이 텍스트의 주제를 나타낸다는 가정을 바탕으로 한다(김수현 외 2019, 백영민 2020, 윤태일, 이수안 2018).

키워드 분석 방법에는 TF(Term Frequency) 분석과 TF-IDF(Term Frequency-Inversed Document Frequency) 분석이 있다. 빈도분석을 위해서는 특정 문서에 등장하는 특정 단어의 출현 빈도를 행렬로 나타낸 DTM(Document-Term Matrix)을 구축해야 한다. DTM을 통해 전체 말뭉치에서 특정 단어의 등장 횟수를 분석하여 단어별 TF 값을 구할 수 있으며, 그 값이 높은 단어를 핵심 단어로 판단한다(황은희 외 2019).

그러나 단순히 출현 빈도수가 높다고 해서 무조건 주제를 드러내는 키워드라고 보기 어렵다. 오히려 모든 문서에서 자주 출현하는 일반적인 상투어인 경우가 많기 때문에 이를 보완한 TF-IDF를 사용한다(우신욱, 장영재 2016). TF-IDF는 여러 문서로 이루어진 문서 군이 있을 때 특정 문서 내에서 특정 단어가 얼마나 중요한지를 나타내는 통계적 수치이다(Oh et al. 2017). TF-IDF값은 전체 문서에서 해당 단어가 출현한 문서 수는 적지만 특정 문서에는 집중적으로 많이 언급되는 경우 값이 올라간다. 즉, TF-IDF값이 클수록 특정 맥락에서 상대적으로 핵심적인 의미와 내용을 나타낼 가능성이 크다(Chung et al 2019).

이 연구에서는 DTM에 TF-IDF 가중치를 적용하여 TF 값과 TF-IDF 값을 동시에 분석함으로써, 인공지능 기반 디지털 영어교육 활용에 중요한 키워드를 더욱 정밀하게 도출하고자

© 2025 KASELL All rights reserved

하였다. 다음으로 TF, TF-IDF 결과를 비교 분석하여 시각화하기 위해 상위 100개 키워드를 워드 클라우드로 표현했다. 워드 클라우드는 텍스트마이닝에서 자주 사용되는 시각화 도구로, TF나 TF-IDF와 같은 빈도에 비례해서 단어의 크기를 나타내어 텍스트 내 단어들의 중요도를 직관적으로 표현한다.

또한, N-gram 분석을 실시하여 연속으로 등장하는 단어의 빈도를 확인하였다. N-gram은 문서 내에서 연이어 나타나는 n 개의 단어들의 집합을 분석 단위로 설정하는 통계 기반 언어 분석 기법이다(이창수 2011). 단어의 연어(collocation)를 확률적으로 나타내어 특정 키워드가 다른 키워드들과 얼마나 자주 연결되는지 분석한다. 문장 내 연속하여 동시에 출현하는 단어와 그 빈도를 계산하며, 문장을 몇 개의 단어 개수로 나눌지에 따라 Bi-gram, Tri-gram, Tetra-gram 등 종류가 결정된다(Chakraborty and Sarma 2024).

3.2.2 네트워크 텍스트 분석

네트워크 생성을 위해 먼저 공출현(co-occurrence) 빈도 매트릭스를 산출하였다. 공출현이란 전체 텍스트 내 특정 범위에서 노드들이 함께 나타났을 때 노드 간에 의미론적으로 상호 관계가 있다고 본다. 따라서 빈도 상위 50개를 1-mode 형태의 매트릭스로 생성하였다. 네트워크 텍스트 분석은 핵심 단어 간의 의미론적 연관성을 중심으로 연구하는 방법이다(박한우 2004). 이 방법은 사회연결망에서의 행위자(node)를 단어로, 그리고 행위자 간의 관계(link)를 단어들 사이의 관계로 해석하여 의미연결망 분석(Popping 2003) 또는 단어 연결망 분석(Danowski 1993)으로 알려져 있다. 이 분석은 텍스트를 네트워크로 변환하여 단어 간의 구조적 관계를 파악하는 방법으로, 언어와 지식이 단어 간의 관계로 형성되는 네트워크라는 가정에 기반한다(Sowa 1984).

네트워크 텍스트 분석에서는 공출현 빈도를 통해 분석이 이루어지며, 이는 텍스트 뭉치에서 의미 있는 단어 조합의 동시 출현 빈도를 계산하고, 이를 바탕으로 단어들의 패턴을 네트워크로 표현하여 핵심 주제를 파악하거나 전략적 위치를 도출하는 작업을 포함한다. 먼저, 에고 네트워크 분석은 특정 중심 행위자(ego)와 그와 직접적 관계를 맺고 있는 주변 행위자들(alters) 간의 관계적 구조를 체계적으로 분석하는 기법이다. 이는 전체 네트워크가 아닌 특정 행위자를 중심으로 한 국지적 네트워크(local network)의 구조적 특성과 관계적 패턴을 규명하는 데 초점을 둔다. 에고 네트워크의 주요 구성 요소는 분석의 핵심이 되는 중심 행위자(ego), 중심 행위자와 직접적 연결 관계를 가진 1차 연결 노드들인 주변 행위자(alters), 그리고 ego와 alters 간, alters 간의 관계적 연결(ties)로 이루어진다. 이러한 에고 네트워크는 네트워크 규모(ego와 직접 연결된 alters의 수), 네트워크 밀도(실제 존재하는 연결의 수를 가능한 최대 연결 수로 나눈 비율), 그리고 중심성, 구조적 공백, 중개성 등의 구조적 특성을 통해 분석된다. 에고 네트워크 분석은 대규모 네트워크에서 특정 노드의 영향력이나 정보 흐름을 분석하는 데 특히 유용하며, 개인 수준의 네트워크 특성이 전체 네트워크 구조와 어떻게 연관되는지 이해하는 데 중요한 통찰을 제공한다.

3.2.3 토픽 모델링(Topic Modeling)

© 2025 KASELL All rights reserved

토픽 모델링은 문서 집합 내에 숨겨진 토픽을 발견하고 단어와 문서 사이의 관계를 설명하는 텍스트 마이닝기법이다(Blei et al. 2003). 텍스트 문서로부터 새로운 정보를 추출하는 연구에서 활용되며, 특정 분야의 주제와 동향을 파악하는 방법이다(진설아, 송민 2016). 토픽 모델링에는 LSA(Latent Semantic Analysis), LSI(Latent Semantic Indexing), pLSA(probabilistic Latent Semantic Analysis) 등의 분석 알고리즘이 있는데 이 중 LDA(Latent Dirichlet Allocation: 잠재 디리클레 할당)가 가장 널리 사용되는 기법이다.

구체적으로 LDA 기반 토픽모델은 주어진 자료나 텍스트 집합에 대하여 각 텍스트에 어떤 주제들이 존재하는지를 보여주는 확률 모형이다. LDA는 문서 집합인 말뭉치(Corpus)의 단어 정보를 바탕으로 사후확률 분포를 추정하여 입력된 토픽 개수 별 주요 키워드를 추출한다. 한 문서는 여러 개의 잠재된 토픽으로 구성되어 있으며 각 토픽은 여러 단어들로 표현된다고 가정하고, 각 단어가 특정 토픽에 포함될 확률을 계산해서 확률이 높은 단어들의 집합을 추출하는 기법이다(윤태일, 이수안 2018, Blei et al. 2003). 즉, 단어가 모이면 주제(토픽)가 되고, 주제가 모여서 특정 텍스트를 구성한다. 따라서 관찰된 각 단어들은 주제를 대표하고, 이 주제들의 집합이 문서가 되기 때문에 각 단어가 어떤 잠재된 주제에 속하는지를 확률적으로 추출해 냄으로써, 주제의 집합인 각각의 텍스트를 분류할 수 있는 것이다.

최적의 토픽 수 k값을 찾기 위해 Gensim의 LDAModel을 적용하였다. 파라미터인 a와 β 값은 각각 0.5와 0.01로 설정하였다. Seed는 랜덤 값으로 설정하였다. 생성한 LDA 모델은 50번 반복하여 학습한 후 각 토픽별 상위 30개 단어를 추출했다. 또 각 문서별로 가장 확률이 높은 토픽과 그 확률도 추출하였다. 분석 결과를 시각화하기 위해 pyLDAvis 라이브러리를 활용하였다. 토픽별로 추출된 키단어와 시각화 결과를 바탕으로 각각의 토픽명을 설정하였다.

토픽 모델링에서 토픽 개수는 비슷한 단어끼리 묶을 군집의 개수이고, 연구자가 설정하여 분석한다. 최적의 토픽 수 결정은 토픽 모델링 분석의 타당성을 확보하는 데 중요한 역할을 한다(이대영, 이현숙 2021). 토픽 수가 너무 적으면 지나치게 포괄적으로 군집을 이루어 토픽이 추상적이고 키워드 중복이 많아지는 문제를 초래할 수 있다. 반면, 토픽 수가 너무 많으면 각 군집에서 얻을 수 있는 정보가 줄어들어 의미 있는 해석을 어렵게 한다(Cao et al 2009). 이러한 문제를 방지하고자 최적의 토픽 수를 결정하기 위한 지표들이 제시된다.

토픽 모델링이 문서 집합에서 단어의 출현 빈도를 기반으로 문서 내 토픽의 확률과 토픽 내 용어의 출현 확률을 계산하는 확률모델이기 때문에 토픽의 개수를 조절하며 혼잡도가 가장 낮은 경우를 최적의 토픽 수로 결정하였다. 그러나 낮은 혼잡도가 반드시 최적의 결과는 아니라는 비판도 있다(Chang et al. 2009). 이를 보완하기 위한 응집도는 토픽 내 단어들 사이의 연결 정도를 평가하는 지표로, 응집도가 높을수록 토픽 모델링 결과로 나타난 각 토픽이 의미적으로 비슷한 단어들로 이루어져 있다고 볼 수 있다(Ramirez et al. 2012). 이 연구에서는 적정토픽 수 선정 시 토픽 주제 간 중복 범위와 혼란도(Perplexity)와 일관성(Coherence) 지표를 고려하여 가장 적합한 토픽 수를 선정하였다.

© 2025 KASELL All rights reserved

4. 연구 결과

4.1 텍스트 자료 수집

이 연구에서는 국내외 주요 온라인 포털 사이트의 블로그, 뉴스, 카페, 웹문서 등에서 '인공지능 디지털 영어 교육' 키워드 단어 조합으로 2021년 11월 07일부터 2024년 11월 06일까지 3년 동안의 텍스트 데이터를 수집하였다. 총 42,720건의 검색 결과를 대상으로 별도 스팸 필터링은 생략하고 제목과 내용을 기준으로 중복 제거를 하였다. 수집 텍스트 데이터 정보는 [표 2]와 같다.

표 2. 수집 텍스트 데이터 정보

	수집 데이터
수집 대상(채널)	네이버(블로그, 뉴스, 카페, 지식IN, 웹문서), 다음(블로그, 뉴스, 카페, 웹문서), 구글(페이스북, 뉴스, 웹문서)
수집 기간	2021-11-07 ~ 2024-11-06
수집 단위	월 단위
수집 범위	제목, 본문, URL
검색 키워드	"인공지능"+"디지털"+"영어"+"교육"
검색 건수	42,720건
분석 품사	noun, verb, adjective, foreign
단어(형태소, 토큰) 수	3,319개

4.2 텍스트 데이터 빈도 분석

수집된 자료에서 자주 사용된 단어들의 목록을 통해 인공지능 디지털 영어 교육에 관한 주요 주제와 내용을 파악할 수 있다. 각 단어의 빈도는 해당 단어가 얼마나 중요하게 사용되었는지를 나타내며, 이는 텍스트 분석을 위한 기초적인 통찰을 제공한다. [표 3]의 단어와 그 빈도를 순서대로 살펴보면, '인공지능', '교육' '디지털', '영어'와 같이 주요 검색 키워드가 상위를 차지하고 다음으로 '기술', '학습', '교과서', '빅데이터', '활용', '수학', '정보', '학생', '에듀테크' 순으로 많이 사용되고 있다.

구체적으로 '인공지능'은 55,571회 등장하고 '교육'도 34,449회, '디지털' 28,387회, '영어' 21,219회로 인공지능 교육과 디지털 영어가 주요 키워드임을 알 수 있다. 특히, 검색 키워드 이외에 '기술'이 10,250회 나타나 테크놀로지의 중요성을 나타내고 있고, '학습' 9,912회와 '교과서' 8,391회는 교과서와 학습을 강조하며, '빅데이터' 7,758회와 '활용' 7,153회는 빅데이터 활용의 중요성을 나타내고, '수학' 7,146회와 '정보' 6,910회는 영어와 함께 2025년 처음 도입되는 디지털 교과서 과목을 나타내고, 다음으로 '학생' 5,541회, '에듀테크' 5,310회, '학교' 4,980회 등은 학교에서 학생들에게 에듀테크를 도입하는 점을 시사하고 있다. 나머지 주요 키워드들도 미래 인공지능 기반 디지털 영어교육에 대한 개관을 통해 전반적인 통찰을 제공하고 있다.

© 2025 KASELL All rights reserved

표 3. Word Count Term Frequency(TF)

순위	단어(term)	빈도	순위	단어(term)	빈도	순위	단어(term)	빈도
1	인공지능	55,571	26	학원	3,873	51	프로그램	2,491
2	교육	34,449	27	산업	3,692	52	우리	2,474
3	디지털	28,387	28	기업	3,692	53	챗GPT	2,410
4	영어	21,219	29	사회	3,627	54	아이	2,394
5	기술	10,250	30	초등	3,580	55	뉴스	2,373
6	학습	9,912	31	분야	3,576	56	시스템	2,366
7	교과서	8,391	32	서비스	3,423	57	운영	2,270
8	빅데이터	7,758	33	과학	3,346	58	컴퓨터	2,236
9	활용	7,153	34	대학	3,253	59	사용	2,221
10	수학	7,146	35	융합	3,227	60	맞춤	2,218
11	정보	6,910	36	혁신	3,137	61	가능	2,204
12	학생	5,541	37	제공	3,117	62	분석	2,182
13	에듀테크	5,310	38	국어	3,083	63	스마트	2,174
14	학교	4,980	39	콘텐츠	2,880	64	변화	2,170
15	코딩	4,943	40	전공	2,852	65	챗봇	2,157
16	과정	4,865	41	공학	2,816	66	능력	2,153
17	도입	4,615	42	교과	2,787	67	소프트웨어	2,140
18	시대	4,609	43	관련	2,742	68	역량	2,102
19	기반	4,539	44	교육부	2,725	69	전문	2,064
20	미래	4,536	45	학년	2,723	70	대학교	2,062
21	수업	4,362	46	사업	2,704	71	세계	2,039
22	개발	4,237	47	언어	2,640	72	문제	1,995
23	학과	4,007	48	한국	2,629	73	필요	1,962
24	과목	3,888	49	교사	2,534	74	메타버스	1,922
25	지원	3,881	50	플랫폼	2,509	75	온라인	1,915

다음으로 TF와 IDF를 결합한 [표 4]의 TF-IDF 결과값은 상투적으로 반복 사용되는 단어보다는 특정 문서 내에서 집중적으로 등장하고 전체 문서는 드물게 등장하는 단어에 가중치를 부여한다. 따라서 각 단어의 상대적 중요도와 특징을 정확히 반영하여 높은 TF-IDF 점수를 가진 단어는 해당 텍스트에 특별히 의미가 있거나 중요하다는 것을 나타낸다. '교과서'는 가장 높은 TF-IDF 점수(16,074.76)를 가지며, 이는 인공지능 디지털 영어교육에서 가장 핵심적인 역할을 하는 것을 의미한다. 다음으로 '학습'과 '빅데이터', '에듀테크' 등이 높은 점수를 보여주는데 [표 3]의 단순 TF와 달리 '교과서'와 '학습'이 가장 부각되고 '빅데이터'와 '에듀테크', '코딩', '기술'이 '영어', '교육'보다 높은 점수를 기록했다. 이는 일반적으로 상용되는 표현이 아닌 특정 문맥에서 '교과서'와 '학습', '빅데이터'와 '에듀테크'가 자주 등장하는 것을 시사한다.

구체적으로 TF 표에서 '인공지능'은 1위(55,571), '디지털'은 3위(28,387)로 가장 자주

© 2025 KASELL All rights reserved

등장하는 단어들이지만, TF-IDF에서는 38위(7,282.34)과 31위(7961.95)로 중요도가 상대적으로 낮다. 이는 '인공지능'과 '디지털'이 여러 문서에서 광범위하게 사용되어 특이성이 낮다는 것을 의미한다. 반면에 '교과서'는 TF에서 7위(8,912)로 빈도는 낮지만, TF-IDF에서는 1위(16,074.76)로 가장 높은 중요도를 보인다. 이는 '교과서'가 특정 문서에서의 특이성이 높음을 나타낸다. TF에서 상위에 있는 단어들(예: 교육, 디지털, 영어)은 다양한 문서에서 많이 사용되는 반면, TF-IDF에서는 '학습', '빅데이터', '에듀테크' 등이 특정 문서에서 중요한 단어들이 상위에 있다. 예를 들어, '학습'은 TF에서 6위(9,912)에서 TF-IDF에서는 2위(14815.66)로 올랐고, '빅데이터'는 TF에서 8위(7,758)에서 TF-IDF에서는 3위(13,931.49)로 상승하여 빈도와 중요도 모두에서 높은 위치를 차지한다. 이는 '빅데이터'가 상대적으로 많이 사용되면서도 중요도가 높음을 나타낸다.

표 4. Word Count Term Frequency-Inverse Document Frequency (TF-IDF)

순위	Word	tf-idf	순위	Word	tf-idf
1	교과서	16,074.76	26	기업	8,177.78
2	학습	14,815.66	27	공학	8,159.94
3	빅데이터	13,931.49	28	산업	8,147.18
4	에듀테크	12,786.59	29	초등	8,140.51
5	코딩	12,737.67	30	대학	8,105.71
6	기술	12,724.26	31	디지털	7,961.95
7	영어	11,848.94	32	분야	7,904.06
8	교육	11,754.83	33	융합	7,839.40
9	수학	11,351.03	34	서비스	7,791.40
10	활용	11,085.33	35	사회	7,765.68
11	학과	10,987.69	36	과학	7,641.98
12	정보	10,796.74	37	학년	7,598.49
13	학생	10,077.00	38	인공지능	7,282.34
14	학원	9,857.54	39	사업	7,184.40
15	과정	9,688.63	40	혁신	7,182.94
16	학교	9,685.32	41	국어	7,128.61
17	도입	9,604.10	42	챗GPT	7,057.09
18	수업	9,245.09	43	언어	7,042.76
19	미래	9,059.76	44	콘텐츠	6,961.77
20	시대	8,980.37	45	교과	6,927.31
21	과목	8,726.16	46	제공	6,909.80
22	기반	8,684.48	47	교사	6,857.00
23	개발	8,620.91	48	관련	6,717.08
24	지원	8,322.58	49	챗봇	6,702.52
25	전공	8,278.60	50	교육부	6,694.21

© 2025 KASELL All rights reserved

단순 TF와 TF-IDF의 차이를 워드 클라우드를 통해 비교하면 그림 2, 3과 같다. '인공지능', '교육', '디지털', '영어'가 강조된 TF와 달리 가중치를 포함한 TF-IDF 빈도분석에서 '교과서'와 '학습', '빅데이터'와 '에듀테크', '코딩', '기술' 등이 부각된 점을 쉽게 확인할 수 있다. 이는 인공지능 기반 디지털 영어교육에서 가장 중요한 부분이 교과서와 학습, 그리고 빅데이터와 에듀테크, 코딩 기술이란 점을 추론할 수 있다.

그림 2. TF 워드클라우드 그림 3. TF-IDF 워드클라우드

4.3 N-gram 분석 결과

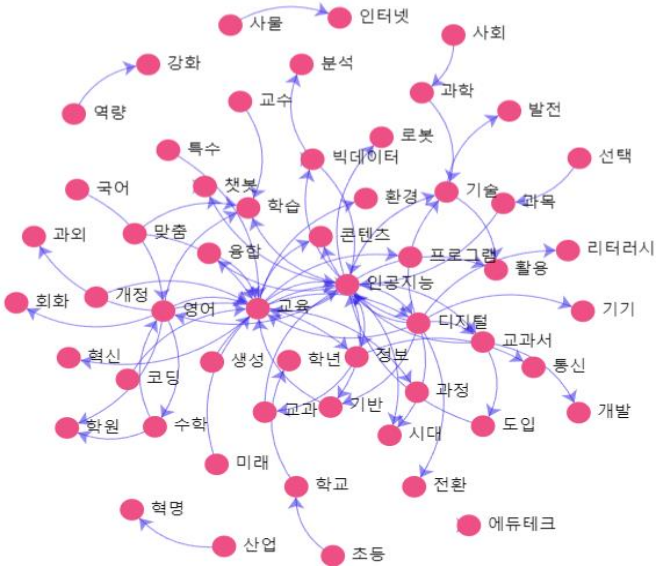

그림 4. Bi-gram 시각화

N-gram 분석 결과 중 [그림 4] Bi-gram 분석의 주요 내용과 횟수를 살펴보면, '디지털-교과서'(6,404), '인공지능-디지털'(5,545), '교육-과정'(2,831), '영어-교육'(2,258), '수학-영어'(2,234), '인공지능-기술'(2,185), '교과서-도입'(1,953), '인공지능-교육'(1,825), '인공지능-기반'(1,758), '인공지능-활용'(1,574), '영어-정보'(1,520), '영어-수학'(1,437), '디지털-교육'(1,327), '인공지능-빅데이터'(1,230), '초등-학교'(1,209), '인공지능-시대'(1,170), '산업-혁명'(1,118), '디지털-기술'(1,095), '디지털-전환'(1,001), '미래-교육'(955), '생성-인공지능'(933), '영어-학원'(885), '교육-인공지능'(884), '코딩-교육'(849), '기술-활용'(848),

© 2025 KASELL All rights reserved

'교육-혁신'(821), '정보-교과'(814), '인공지능-교과서'(788), '디지털-기반'(783), '디지털-시대'(768), '맞춤-교육'(746), '영어-학습'(744), '개정-교육'(733), '인공지능-융합'(729) 순서로 나타났다.

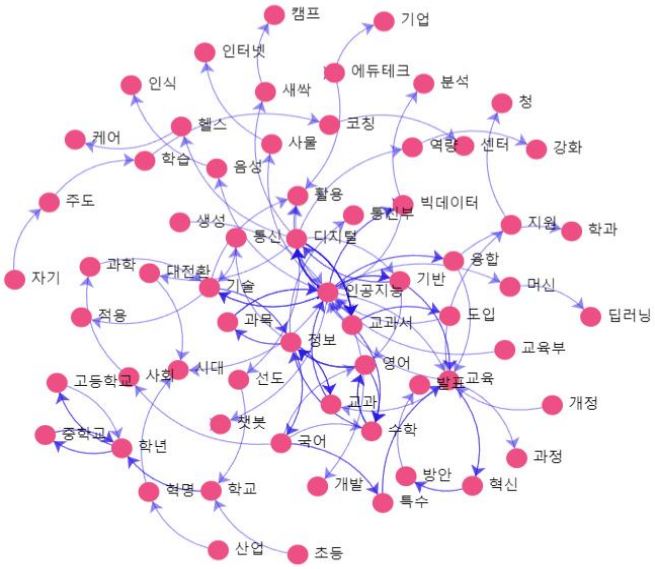

그림 5. Tri-gram 시각화

[그림 5] Tri-gram 분석의 주요 내용으로 '인공지능-디지털-교과서'(4,575), '디지털-교과서-도입'(1,737), '수학-영어-정보'(1,465), '개정-교육-과정'(713), '디지털-기반-교육'(476), '산업-혁명-시대'(370), '디지털-교과서-개발'(359), '인공지능-기술-활용'(351)이 등장하고 있다.

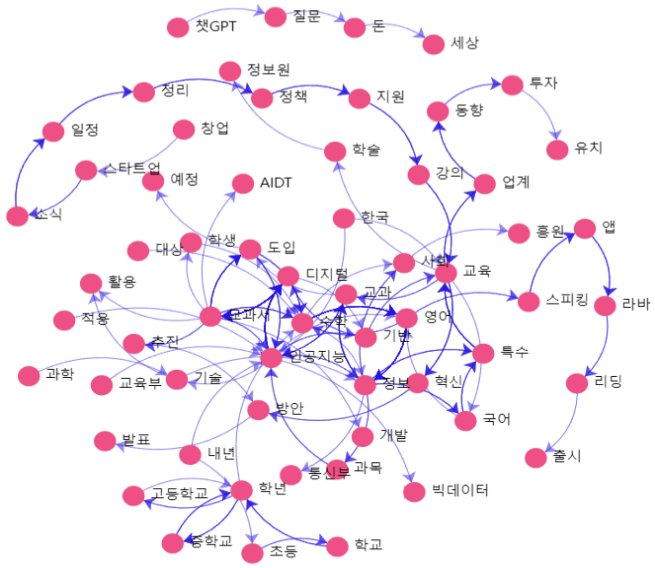

그림 6. Tetra-gram 시각화

© 2025 KASELL All rights reserved

[그림 6] Tetra-gram 분석의 주요 내용으로 '인공지능-디지털-교과서-도입'(1,330), '수학-영어-정보-교과'(528), '디지털-기반-교육-혁신'(399), '인공지능-디지털-교과서-개발'(286), '기반-교육-혁신-방안'(248)이 나타났다. N-gram 분석 결과 디지털 대전환기의 미래 교육을 위한 디지털 기반 교육혁신 방안으로 수학, 영어, 정보 교과를 중심으로 인공지능 디지털 교과서(AI Digital Textbook: AIDT)를 개발하여 2025년부터 인공지능 디지털 교과서를 새로 도입하는 것이 강조되고 있음을 알 수 있다.

4.4 네트워크 텍스트 분석 결과

[표 5]에 네트워크 텍스트 분석을 위한 Co-occurrence 매트릭스 값을 제시하였다.

표 5. 상위 15 Co-occurrence 매트릭스 값

	인공지능	교육	디지털	영어	기술	학습	교과서	빅데이터	활용	수학	정보	학생	에듀테크	학교
인공지능	0													
교육	79,542	0												
디지털	72,632	43,088	0											
영어	47,559	33,525	25,105	0										
기술	25,587	13,337	12,778	7,575	0									
학습	25,852	16,253	13,052	12,028	4,885	0								
교과서	31,850	12,464	22,883	9,630	3,118	5,099	0							
빅데이터	17,518	9,641	8,060	4,863	4,863	3,063	671	0						
활용	20,474	12,234	9,871	6,807	3,824	4,758	3,088	2,626	0					
수학	18,033	11,890	11,737	9,396	2,668	4,766	7,453	1,513	2,224	0				
정보	17,285	10,271	11,124	5,929	3,783	2,558	6,206	2,896	2,198	3,752	0			
학생	13,227	9,540	8,215	6,141	2,151	4,629	4,307	1,539	2,220	3,158	2,016	0		
에듀테크	14,245	10,384	6,498	5,087	2,598	2,971	2,497	1,398	2,404	1,436	1,075	1,068	0	
학교	11,774	9,665	7,498	4,913	1,227	2,333	3,438	850	1,898	2,591	1,975	2,140	1,340	0
코딩	9,181	8,523	5,156	3,857	2,029	1,543	625	1,607	1,157	1,417	1,010	1,242	645	971

[표 6]에 에고 네트워크 전반에 대한 정보가 나타나 있다. 에고 네트워크 분석은 전체에서 에고(중심, Ego)노드를 대상으로 하는 네트워크를 추출하며, 에고 노드와 연결된 다른 노드(타자, Alter)와의 연결 관계를 분석한다. 에고 네트워크 분석 시각화는 중심 키워드인 '영어'와 다양한 관련 개념들 간의 관계를 시각적으로 나타낸 것이다.

표 6. 에고 네트워크 분석 값

에고(Ego)	Size	Ties	Pairs	Ego-Density	Broker	Ego Between
영어	99	4,950	9,702	0.5102	4,752	0

[그림 7]의 '영어'를 중심으로 한 에코 네트워크 분석 시각화 결과를 보면, '학습'을 포함해 '학원', '학생', '정보', '활용', '초등'이 연결되고, '교과서'를 기준으로 '기술', '과목', '기반', '과정'이 연결되고, '수학'을 기준으로 '수업', '빅데이터', '에듀테크', '도입'이 연결되고 있다.

© 2025 KASELL All rights reserved

구체적으로 크게 세 가지 범주로 구분된다. 첫 번째로, 학습과 관련된 단어들이 눈에 띈다. '학교', '학생', '학원', '활용' 등은 모두 영어와의 연결에서 중요한 위치를 차지하고 있다. 특히 '학습'이라는 단어는 큰 노드로 표시되어 있어 영어학습 자체가 주요 주제임을 시사한다. 이와 함께 '학생'과 '학원', '정보'도 영어와 밀접한 관계가 있음을 보여준다. 이는 영어학습이 학생 중심으로 이루어지며, 학습 기관인 학교와 학원이 중요한 역할을 한다는 것을 나타낸다.

두 번째로, 교과서와 관련된 단어들이 포함되어 있다. '기술', '과목', '과정' 등은 영어교육에서 기술 개발을 활용한 교재의 중요성을 강조하는 요소로 해석된다. '교과서'는 영어학습에서 필수적인 학습 자료로, 그 크기에서 나타나는 중심성을 통해 영어교육의 기본적인 틀을 제공하는 요소임을 알 수 있다. 또한 '과정', '교사'는 교육과정이 교육의 질과 학습 과정에서 중요한 역할을 하며, '기반'이라는 단어는 교과서나 교육과정 요소들이 어떻게 영어교육의 기반을 이루는지를 시사한다.

마지막으로, 수학과 기술 및 미래 트렌드와 관련된 단어들도 연결되어 있다. '빅데이터', '에듀테크', '도입', '미래' 등은 수학과 함께 미래 영어교육이 기존 언어학습을 넘어서는 새로운 방향성을 제시하고 있음을 보여준다. '빅데이터'는 학습 데이터를 통해 개인화된 학습경험을 제공할 수 있는 가능성을 나타내며, '에듀테크'는 교육 기술의 발전과 그 적용이 영어교육의 필수적인 요소로 자리 잡고 있음을 시사한다. '미래'라는 단어는 이러한 기술적 발전과 영어학습의 연계성을 통해 앞으로의 교육이 어떤 방향으로 나아갈지를 암시한다. 결론적으로, 에고 네트워크 분석 시각화는 영어교육이 '학습', '학생'과 '교과서', '교사'와 같은 전통적인 교육 요소들과 함께 '빅데이터', '에듀테크', '기술'과 같은 현대적인 기술 요소들과 어떻게 긴밀하게 연결되어 있는지를 보여준다.

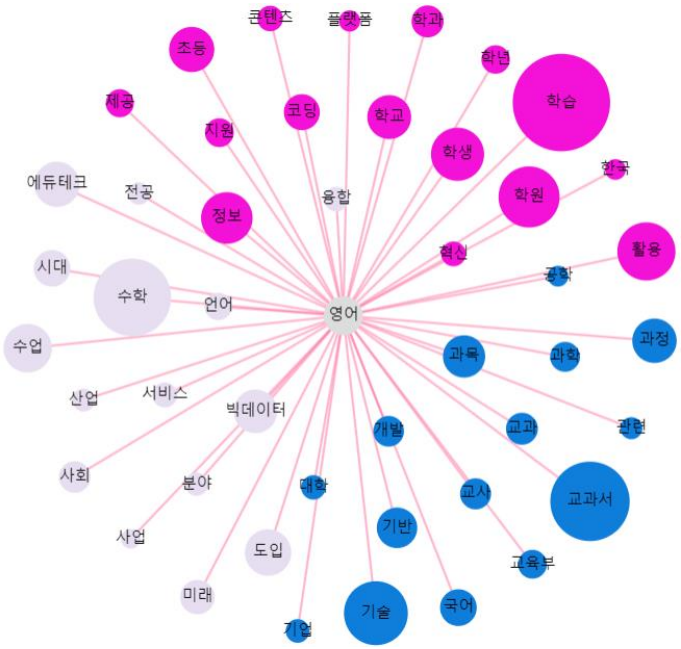

그림 7. 영어 에고 네트워크 분석 시각화

© 2025 KASELL All rights reserved

4.5 토픽 모델링 분석 결과

LDA로 주요 토픽을 추출하기 위해서는 연구자가 사전에 최적의 토픽 개수를 정해야 한다. 이 연구에서는 토픽의 개수를 설정하기 위해 파이썬(Python) 모듈의 LDAModel 함수를 사용하였다. 토픽 분석 파라미터는 Alpha 0.5, Beta 0.01, Iterations 50회 적용하여 토픽의 개수를 1부터 10까지 증가시킨 모델에 대한 응집도(Coherence Score)와 혼잡도(Perplexity Score)의 변화를 확인하고 종합하여 최적의 토픽 개수를 정하였다. 보통 응집도는 증가할수록, 혼잡도는 감소할수록 효과적인 토픽 개수로 볼 수 있다. 하지만 낮은 혼잡도가 반드시 해석에 적합한 결과를 주지는 않는다는 비판이 있어(Chang et al. 2009) 응집도에 좀 더 비중을 두어 결과를 해석했다. [표 7]과 [그림 8]은 응집도와 혼잡도를 측정하고 시각화 결과로 혼잡도는 일정하지만 응집도는 토픽 수 5개를 기점으로 지속적으로 감소한다. 따라서 네트워크 텍스트 분석 결과를 고려하는 동시에 혼잡도 변화가 거의 없으면서 응집도의 증가세가 감소로 전환되기 시작하는 토픽 수 5개를 최적의 토픽 수로 결정하였다.

표 7. 토픽의 개수에 따른 응집도와 혼잡도

토픽 수	Perplexity Score	Coherence Score
2	−16.30361381318501	−1.8700127185977844
3	−16.246559382625026	−1.9716438520986188
4	−16.250220303161647	−1.9872893809868128
5	−16.235780403588553	−1.9747100661274957
6	−16.23463421770965	−2.0318068390767396
7	−16.232025712072247	−2.1787567819994336
8	−16.24767492660992	−2.2379193126949293
9	−16.256885105947227	−2.4021199693038215
10	−16.279106575310156	−2.566430121541002

[표 8]에 토픽별 주요 단어 결과표를 제시하였다. LDA 토픽 분석 결과, 총 5개의 토픽이 도출되었으며, 각 토픽의 주요 단어들과 그 가중치를 바탕으로 다음과 같이 제목을 제안하고 주제를 해석할 수 있다. Topic 1은 '인공지능 기반 교육혁신'으로 명명할 수 있다. 이는 '교육'(0.035), '인공지능'(0.033), '코딩'(0.027), '학원'(0.022)과 같은 높은 가중치의 단어들이 기술 기반 교육의 혁신적 도입을 시사하고 있으며, '학과', '전공', '과정', '학습', '수학' 등의 단어들이 실제 교육 현장에서의 적용을 나타내고 있기 때문이다. Topic 2는 '디지털 기반 교육환경 전환'으로 해석된다. '인공지능'(0.067), '디지털'(0.052), '교육'(0.051)의 높은 가중치와 함께 '교과서', '도입', '학생', '학습', '학교' 등의 단어들이 교육 현장의 디지털화를 강조하고 있다. Topic 3은 '교육 시스템의 미래 지향적 변화'로 명명할 수 있으며, '인공지능'(0.039), '디지털'(0.028), '기술'(0.022)을 중심으로 '교육', '분야', '연구', '시스템' 등의 단어들이 교육체계의 전반적인 변화를 시사한다. Topic 4는 '학습자 중심의 맞춤형 영어교육'으로 해석되는데, '인공지능'(0.06), '영어'(0.045), '교육'(0.031)을 중심으로 '학습', '디지털', '활용', '아이', '시대', '언어', '능력', '공부'

© 2025 KASELL All rights reserved

등의 단어들이 개인화된 학습경험을 강조하고 있다. 마지막으로 Topic 5는 '교육산업의 기술 혁신과 서비스'로 명명할 수 있다. '인공지능'(0.073), '교육'(0.019), '에듀테크'(0.015), '기업'(0.014), '서비스'(0.012)를 중심으로 '기술', '개발', '사업', '플랫폼' 등의 단어들이 에듀테크 산업에서의 기술 혁신과 서비스 개발을 나타내고 있다.

표 8. LDA 분석 결과 각 토픽 주요 단어

Topic	Words with Weights
Topic 1: 인공지능 기반 교육혁신	교육(0.035), 인공지능(0.033), 코딩(0.027), 학원(0.022), 학과(0.019), 영어(0.018), 빅데이터(0.016), 디지털(0.016), 전공(0.013), 공학(0.01), 융합(0.01), 과정(0.009), 학습(0.008), 수학(0.008), 기술(0.008), 대학(0.008), 컴퓨터(0.007), 분야(0.007), 산업(0.006), 학부(0.006), 미래(0.006), 대학교(0.006), 언어(0.006), 모집(0.005), 자격증(0.005), 디자인(0.005), 사회(0.005), 소프트웨어(0.005), 능력(0.005), 과학(0.005),
Topic 2: 디지털 기반 교육환경 전환	인공지능(0.067), 디지털(0.052), 교육(0.051), 교과서(0.045), 영어(0.026), 도입(0.023), 수학(0.022), 학생(0.019), 학교(0.016), 학습(0.015), 정보(0.014), 학년(0.014), 과목(0.014), 수업(0.012), 초등(0.012), 교육부(0.011), 활용(0.011), 국어(0.01), 교과(0.01), 교사(0.009), 과정(0.009), 내년(0.008), 맞춤(0.008), 기술(0.007), 과학(0.006), 기반(0.006), 사회(0.006), 지원(0.006), 교육청(0.006), 고등학교(0.005)
Topic 3: 교육 시스템의 미래 지향적 변화	인공지능(0.039), 디지털(0.028), 기술(0.022), 교육(0.017), 빅데이터(0.013), 미래(0.008), 분야(0.008), 정보(0.007), 산업(0.006), 한국(0.006), 연구(0.006), 영어(0.006), 발전(0.006), 시대(0.006), 전문가(0.005), 시스템(0.005), 활용(0.004), 사회(0.004), 대학(0.004), 혁신(0.004), 세계(0.004), 경제(0.004), 보안(0.004), 과학(0.004), 변화(0.004), 챗봇(0.003), 가능(0.003), 혁명(0.003), 인터넷(0.003), 분석(0.003)
Topic 4: 학습자 중심의 맞춤형 영어교육	인공지능(0.06), 영어(0.045), 교육(0.031), 학습(0.023), 디지털(0.021), 활용(0.012), 아이(0.009), 시대(0.009), 기술(0.008), 에듀테크(0.007), 언어(0.007), 우리(0.006), 사용(0.006), 문제(0.006), 능력(0.006), 공부(0.005), 수업(0.004), 미래(0.004), 회화(0.004), 방법(0.004), 번역(0.004), 책(0.004), 만들(0.004), 변화(0.004), 중요(0.004), 챗봇(0.004), 프로그램(0.004), 사람(0.004), 생각(0.003), 도구(0.003)
Topic 5: 교육산업의 기술 혁신과 서비스	인공지능(0.073), 교육(0.019), 디지털(0.019), 에듀테크(0.015), 기업(0.014), 서비스(0.012), 기술(0.01), 개발(0.009), 사업(0.009), 플랫폼(0.008), 뉴스(0.008), 시장(0.007), 기반(0.007), 빅데이터(0.006), 관련(0.006), 제공(0.006), 영어(0.005), 생성(0.005), 지원(0.005), 활용(0.005), 콘텐츠(0.005), 투자(0.004), 솔루션(0.004), 국내(0.004), 혁신(0.004), 딥러닝(0.004), 한국(0.004), 글로벌(0.004), 분야(0.004), 정보(0.004)

© 2025 KASELL All rights reserved

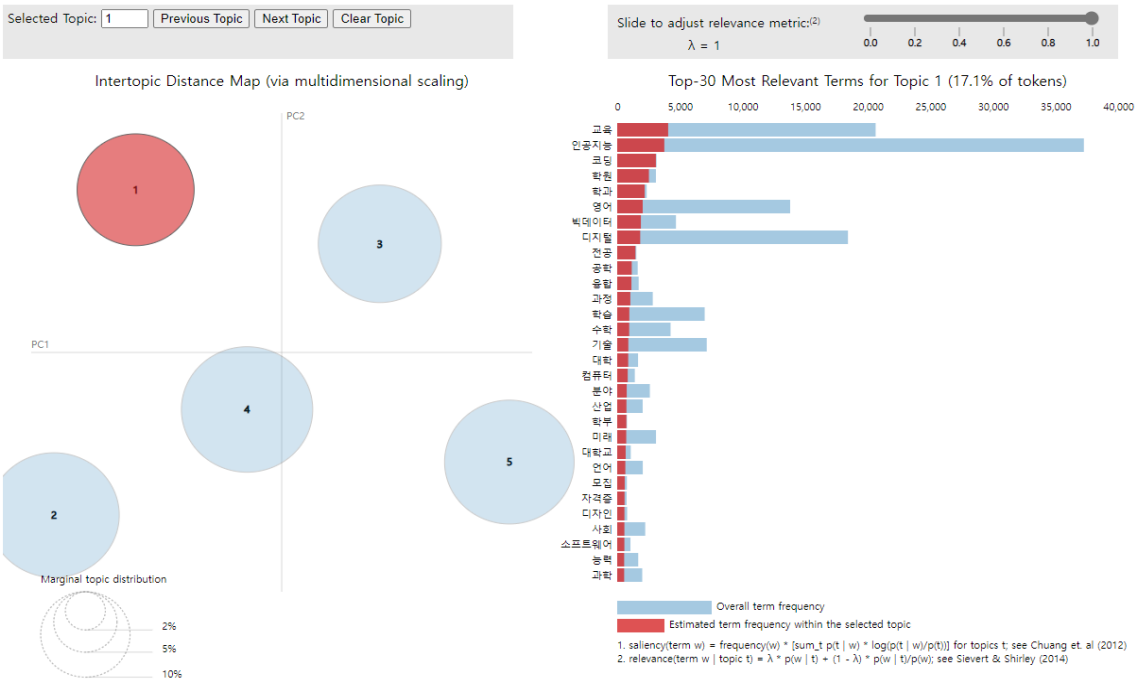

그림 8. 토픽 1 토픽 간 거리 지도(Intertopic Distance Map)

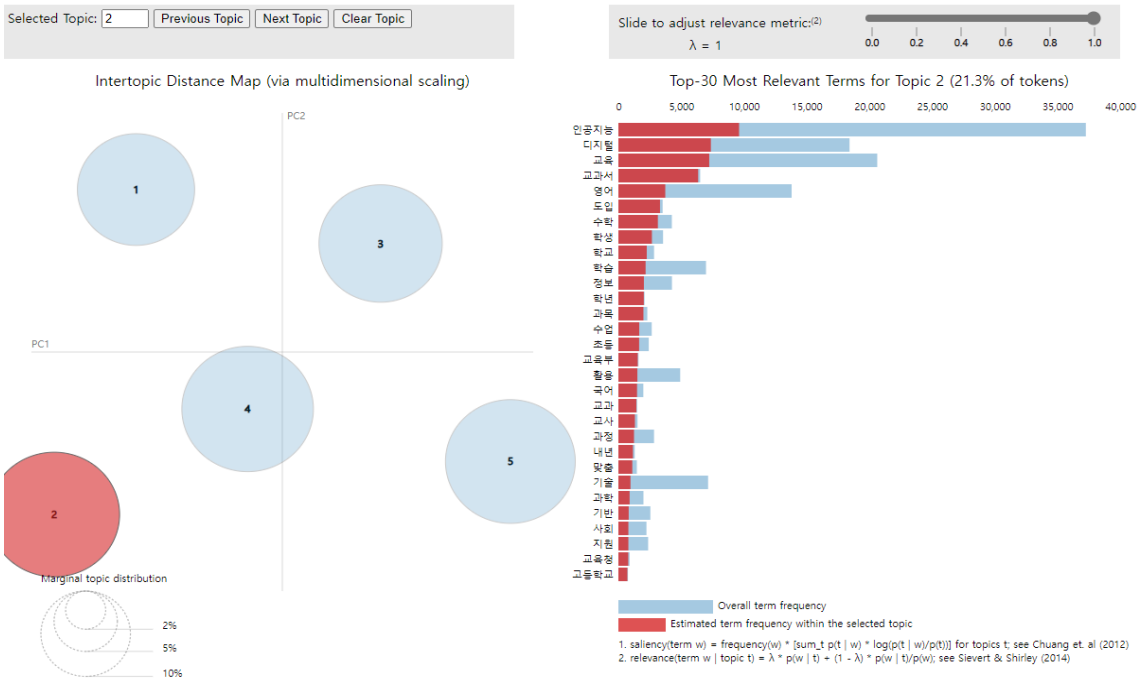

그림 9. 토픽 2 토픽 간 거리 지도(Intertopic Distance Map)

© 2025 KASELL All rights reserved

구체적으로 [그림 8]의 토픽 모델링 분석 결과 그래프에서 토픽 1의 위치와 크기 그리고 주요 내용을 다각적으로 보여준다. Topic 1의 핵심 단어 분포를 나타내는 막대그래프는 토픽 내에서 중요도가 높은 단어들을 상위 30개까지 보여준다. 구체적으로 토픽 1 '인공지능 기반 교육혁신'의 주요 단어는 '인공지능', '코딩', '학원', '학교', '영어', '빅데이터' 등이 포함되어 있다. 이 토픽은 인공지능과 관련된 기술이 교육 현장에 어떻게 적용되고 있으며, 특히 학원이나 학교 교육에서의 실질적인 활용에 초점을 맞추고 있음을 보여준다. '공학', '융합', '과정', '코딩'과 '컴퓨터', '기술' 등의 단어는 기술 교육의 중요성을 시사하며, '영어'와 같은 단어는 이러한 기술 교육이 언어학습과 결합될 수 있음을 나타낸다. 따라서 이 토픽은 교육을 위한 인공지능과 기술의 융합적 활용 내용으로 해석할 수 있다.

[그림 9]의 Topic 2 '디지털 기반 교육환경 전환'의 핵심 단어들은 '인공지능', '디지털', '교육', '교과서', '도입', '수학', '정보', '초등', '맞춤형' 등으로 구성되어 있다. 이 토픽은 디지털 기술이 교육에 도입되는 과정과 그로 인한 혁신을 다룬다. '교과서', '학교', '초등'과 같은 단어는 주로 초등학교 수준에서 디지털 교육의 필요성과 도입이 중요하게 다뤄짐을 시사한다. '맞춤형'과 '기술'은 학생 개개인에게 맞춤형 학습을 제공하는 디지털 솔루션의 중요성을 반영한다. 따라서 이 토픽은 디지털화된 교육환경 변화와 그 적용에 관한 것으로 해석된다.

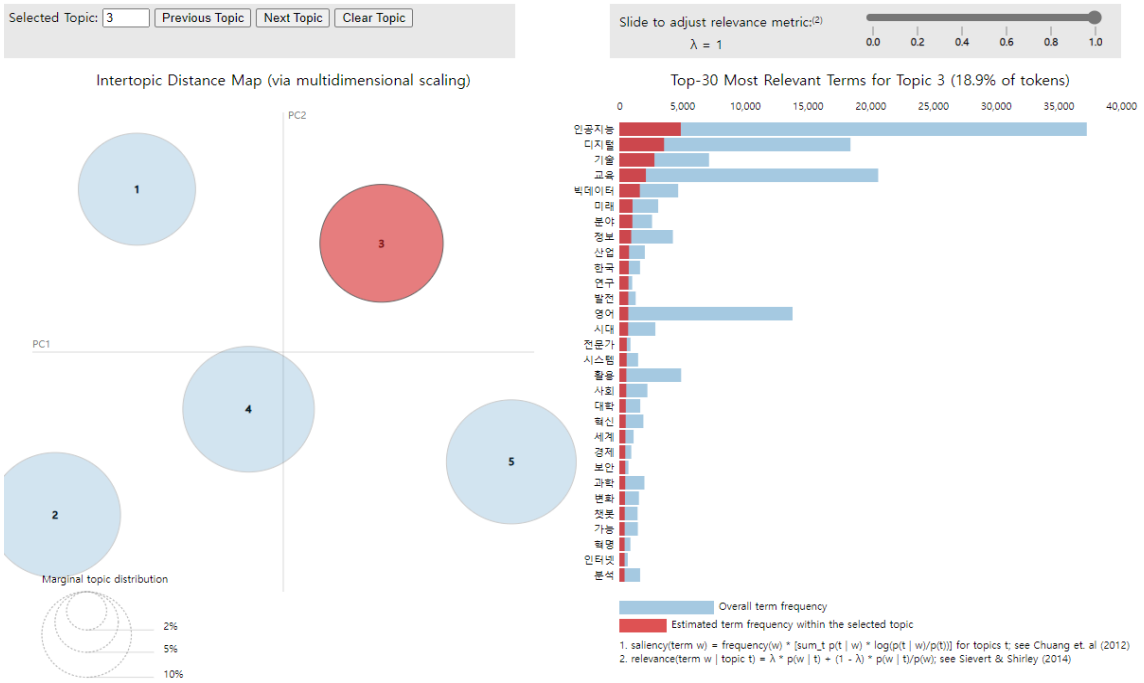

그림 10. 토픽 3 토픽 간 거리 지도(Intertopic Distance Map)

© 2025 KASELL All rights reserved

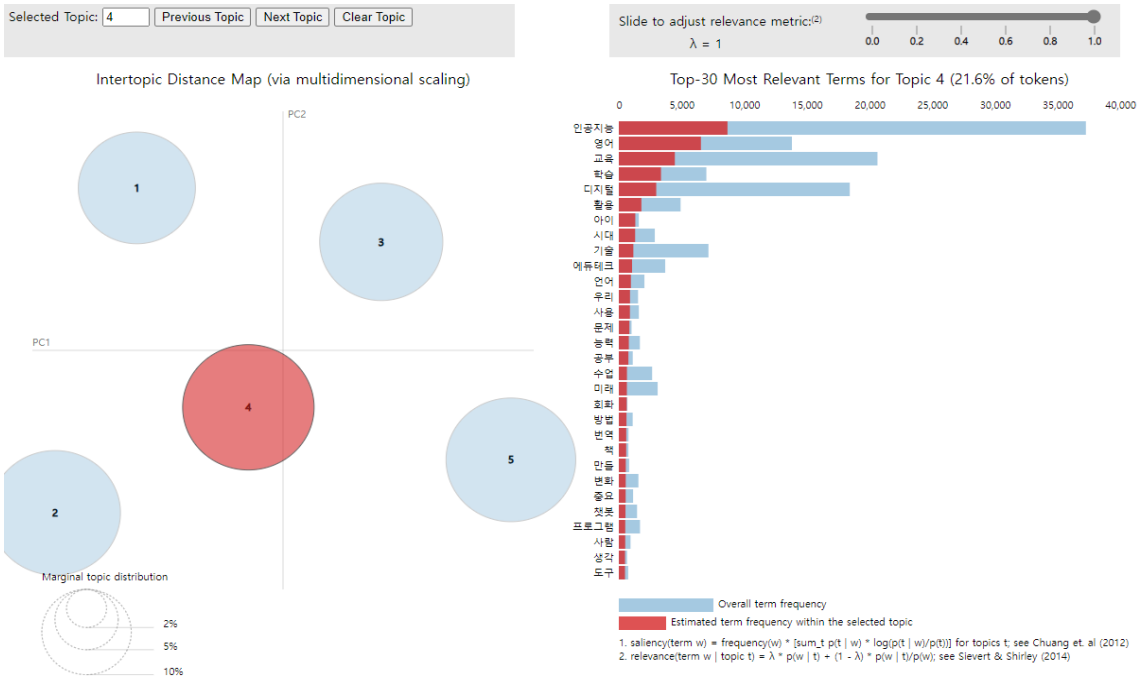

그림 11. 토픽 4 토픽 간 거리 지도(Intertopic Distance Map)

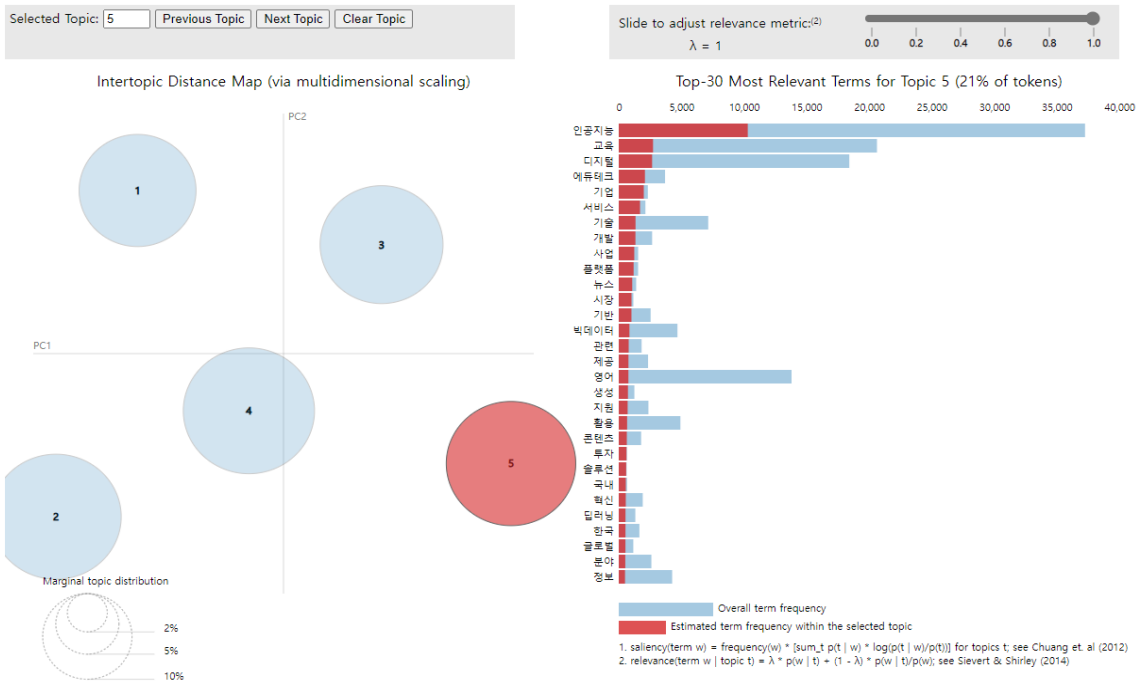

그림 12. 토픽 5 토픽 간 거리 지도(Intertopic Distance Map)

© 2025 KASELL All rights reserved

[그림 10]의 Topic 3 '교육 시스템의 미래 지향적 변화'의 주요 단어로는 '인공지능', '디지털', '기술', '교육', '미래', '분야', '정보', '산업', '한국', '연구', '발전', '시대', '전문가', '시스템' 등이 포함되어 있다. 이 토픽은 교육이 미래 사회와 산업의 기술적 요구와 산업 발전에 어떻게 대비하고 있는지를 보여준다. '연구', '발전', '시대'와 같은 단어들은 교육이 급변하는 기술적 환경에 어떻게 적응하고 있는지에 대한 탐구를 반영한다. '사회', '대학', '혁신' 등의 단어는 교육과정에서의 변화와 혁신적인 접근이 필요함을 시사한다. 따라서 이 토픽은 미래 사회를 준비하는 교육적 시스템의 변화에 관한 내용을 다룬다.

[그림 11]의 Topic 4 '학습자 중심의 맞춤형 영어교육'의 주요 단어로는 '디지털', '활용', '아이', '시대', '기술', '에듀테크', '영어', '언어', '사용', '문제', '능력', '회화', '방법', '번역', '교과목', '수업' 등이 포함되어 있어, 특히 디지털 기술을 활용한 영어 언어 사용 능력과 회화를 위한 영어교육과 언어학습에 에듀테크를 적용한 학습 방법을 다룬다. 이는 디지털 기술 기반의 영어 교수·학습 어플리케이션과 그 응용 가능성을 보여준다.

마지막으로 [그림 12]의 Topic 5는 '교육과 기업의 협력과 혁신'으로 주요 단어는 '인공지능', '교육', '디지털', '에듀테크', '기업', '서비스', '개발', '투자', '콘텐츠' 등이다. 이 토픽은 교육에서의 기업과의 협력과 기술 혁신을 다루며, 교육의 사업적 및 기술적 측면을 강조한다. '기업'과 '서비스'는 교육 분야에서 기술을 활용한 다양한 서비스를 제공하는 기업의 역할을 시사하며, '기술', '개발', '사업', '플랫폼', '투자', '콘텐츠'는 이러한 기술과 서비스가 교육 콘텐츠의 발전과 확장에 기여하고 있음을 보여준다. '디지털'과 '에듀테크'는 현대 교육에서 인공지능 테크놀로지의 중요성을 강조한다. 따라서 이 토픽은 교육 분야에서 에듀테크 기업의 참여와 협력, 혁신을 주제로 한 것으로 해석할 수 있다. 이와 같이 각 토픽은 미래 인공지능 디지털 영어 교육의 다양한 측면과 기술적, 사회적 변화에 대한 내용을 다각도로 다루고 있음을 알 수 있다.

5. 논의 및 결론

텍스트 마이닝의 NLP 기법은 대규모 텍스트 데이터에서 가치 있는 정보와 지식을 추출하기 위해 인공지능 알고리즘을 활용하는 분석방법이다. 이러한 접근 방법은 온라인 빅데이터 환경에서 텍스트 데이터를 수집하고, 이를 벡터화하거나 임베딩으로 구조화하여 의미 있는 패턴과 통찰을 얻는 데 주로 사용된다. 구체적으로, 다양한 온라인 검색 채널을 통해 방대한 데이터를 수집하여 딥러닝 기법을 통해 분석 목적에 맞는 모델링 분석기법을 적용한다. NLP 분석과정은 일반적으로 데이터 수집, 전처리, 특징과 패턴 추출, 모델 학습 및 적용, 그리고 결과 해석과 논의 단계로 이루어진다.

연구 절차와 관련해 먼저 인공지능 기반 디지털 영어교육에 관한 데이터를 수집하기 위해 주요 검색 포털의 인터넷 문서 저장소, 검색포털, 소셜 미디어 등 다양한 출처에서 웹 크롤링을 통해 텍스트 데이터를 수집하였다. 다음으로 전처리 단계에서는 수집된 텍스트 데이터를 분석에 적합한 형태로 가공하기 위해 토큰화, 불용어 제거, 형태소 분석, 정규화 등의 정제 과정을 거쳤다. 자연어처리를 위한 이러한 데이터 전처리 및 정제 과정은 연구자의 시간과 노력이 많이 드는 단계로 이후 분석의 질을 좌우하는 가장 중요한 과정이다. 마지막으로 특징 추출 단계에서는

© 2025 KASELL All rights reserved

다양한 인공지능 모델링과 알고리즘을 활용하여 전처리된 텍스트에서 의미 있는 특징과 패턴을 찾아 해석하고 새로운 통찰을 얻는 단계이다.

이번 연구에서는 디지털 영어교육에 관한 최근 이슈나 대중의 반응을 통해 영어교육의 현황과 미래 방향을 살펴보기 위해 수집된 빅데이터를 대상으로 TF-IDF와 같은 빈도분석, N-gram분석, 네트워크 텍스트 분석, 토픽모델링 등의 작업을 수행하였다. 마지막으로 분석 결과를 다각적으로 해석하고, 이를 바탕으로 디지털 영어교육에 대한 새로운 통찰과 지식을 도출하고자 하였다. 이러한 NLP 분석기법은 인공지능과 빅데이터 분석기법의 발전과 함께 그 중요성이 부각되고 있으며, 다양한 학문분야에서 혁신적인 분석 방법으로 활용되고 있다.

연구 결과를 살펴보면, 먼저 빈도분석 결과 '인공지능', '교육' '디지털', '영어'와 같은 수집 키워드 다음으로 '기술', '학습', '교과서', '빅데이터', '활용', '수학', '정보', '학생', '에듀테크' 순으로 단어들이 자주 사용되고 있다. 이는 에듀테크 기술의 발전과 빅데이터의 활용은 학생들의 학습경험을 혁신적으로 변화시키며, 2025년부터 도입 예정되는 수학과 정보 디지털 교과서와 함께 디지털 영어교육 자료에서 개인화된 맞춤형 정보 제공과 학습 이력에 따른 적응형 학습과 학습경로 설계 등을 통해 기존 전통적인 영어 교수·학습 패러다임을 변화시키는 것을 반영한다(김인옥 2019, 김정렬 2015, 임희정 2014, 한종임 2014).

특히, TF-IDF 결과값에 따르면 '교과서'가 가장 높은 점수를 가지며, 이는 교과서가 인공지능 디지털 영어교육에서 가장 핵심적인 역할을 하는 것을 의미한다. 다음으로 '학습'과 '빅데이터', '에듀테크' 등이 높은 점수를 보여주는데 단순 TF 결과와 달리 '교과서'와 '학습'이 가장 부각되고 '빅데이터'와 '에듀테크', '코딩', '기술'이 검색 키워드인 '영어', '교육'보다 높은 점수를 기록했다. 즉, 교과서와 학습, 빅데이터와 에듀테크, 코딩 기술이 영어교육에서 매우 중요한 위치를 차지함을 시사한다. 이러한 결과는 학회지 논문을 대상으로 텍스트 분석한 결과 '블렌디드 러닝', '초등', '한국 대학'이라는 단어와 관련된 연구가 많았다는 선행연구(Nam 2022)와 코로나19 기간 국내 영어교육에 대한 웹 데이터를 텍스트 네트워크 분석한 선행연구 결과인 '학습', '학교', '학생', '수업', '온라인', '교사', '학원'과 같은 키워드보다 좀 더 구체적인 결과이다(권선희 2023).

다음으로 단어의 연결 조합인 N-gram 분석 결과, '디지털 교과서', '교육 과정', '영어 교육', '수학 영어', '인공지능 기술', '교과서 도입', '인공지능 교육', '인공지능 기반', '인공지능 활용', '영어 정보', '영어 수학', '디지털 교육', '인공지능 빅데이터', '초등 학교', '인공지능 시대', '산업 혁명', '디지털 기술', '디지털 전환', '미래 교육', '생성 인공지능', '영어학원', '교육 인공지능', '코딩 교육', '기술 활용', '교육 혁신', '정보 교과', '인공지능 교과서', '디지털 기반', '디지털 시대', '맞춤교육', '영어 학습', '개정 교육', '인공지능 융합'의 연어(collocation) 조합이 많이 등장하였다. 이러한 두 단어 조합은 최근 인공지능 디지털 영어교육 분야에서 매우 익숙한 표현이다. 단어 조합을 세 개로 늘리면 '인공지능 디지털 교과서', '디지털 교과서 도입', '수학 영어 정보', '개정 교육 과정', '디지털 기반 교육', '산업 혁명 시대', '디지털 교과서 개발', '인공지능 기술 활용'이라는 표현이 만들어진다. 마지막으로 네 단어 조합을 살펴보면 '인공지능 디지털 교과서 도입', '수학 영어 정보 교과', '디지털 기반 교육 혁신', '인공지능 디지털 교과서 개발'이 나타났다. 즉, N-gram 분석 결과를 종합하면 디지털 기반 교육혁신을 위해서 수학과 정보와 함께 영어 교과에서 인공지능 디지털 교과서 개발 및 도입이 가장 큰 이슈임을 알 수 있다(김인옥 2019, 김정렬 2015, 임희정 2014, 한종임 2014).

© 2025 KASELL All rights reserved

2024년 11월 말 최초로 공개된 인공지능 디지털 교과서의 인공지능 튜터나 보조교사 기능을 종합적으로 검증 및 평가해 본 결과 인공지능 기반 디지털 교과서가 정착하고 성공하기 위해서는 앞으로 개선의 여지가 있다(박문영 2023, 이정화 2023). 가장 중요한 기능인 학생 개별 맞춤형 교수·학습을 위한 인공지능 기반 디지털 튜터와 보조교사 기능은 학생들의 학습 스타일과 이력 데이터가 누적되면 기능이 향상될 것으로 보인다. 또한 기본 학습관리시스템(LMS) 기능은 NEIS와 통합 연동시켜 UI/UX도 보다 사용자 친화형으로 개선될 필요가 있다. 교육부가 추구하는 하이터치 하이테크(Hi Touch, Hi Tech)가 정착되기 위해서는 인공지능과 관련된 기술적인 부분이 안정화되고 교과서가 사용자 친화형으로 정착된 환경에서 교사와 학습자의 높은 수준의 사회정서적 또는 창의적인 상호작용이 이루어질 수 있을 것이다.

네트워크 텍스트 분석의 에고 네트워크 분석 결과를 살펴보면, 먼저, '학생', '학원', '학교', '정보', '활용' 등의 개념이 영어학습과 강하게 연관되며, 특히 '학습'이 중심 노드로 나타났다. 이는 영어교육이 학생 중심으로 이루어지고 있으며, 학습 기관과 정보 활용이 중요한 역할을 한다는 점을 시사한다. 또한, '교과서'를 중심으로 '기술', '과목', '과정' 등의 개념이 연결되며, 이는 영어교육에서 체계적인 학습 과정과 교과서와 같은 교재의 중요성을 강조하는 요소로 해석된다. 한편, 수학 교과와 연계된 '빅데이터', '에듀테크', '도입', '미래' 등의 개념이 영어와 연결되며, 이는 기술 기반 학습의 확산을 보여준다. 빅데이터와 에듀테크는 학습자의 개별 특성에 맞춘 맞춤형 학습을 가능하게 하며, 디지털 기술의 발전이 영어교육의 핵심 요소로 자리 잡고 있음을 시사한다. 이러한 에고 네트워크 분석 결과는 영어교육이 단순한 언어 습득을 넘어 개인화된 학습, 데이터 기반 교수법, 디지털 기술을 활용한 새로운 학습 패러다임으로 변화하고 있음을 의미한다. 후속 연구에서는 이러한 기술적 변화가 학습자의 성취도에 미치는 효과뿐만 아니라 사회정서역량과 학습경험에 미치는 영향도 분석할 필요가 있다.

멀티미디어와 테크놀로지 관련 언어 교육 연구에 대해 네트워크 분석을 실시한 선행연구에서 '블렌디드 러닝', '글쓰기 능력', '초등학교 학습자', '컴퓨터기반 언어(교육)', '영어교육'이라는 주제(Nam 2022)로 분류된 것과 디지털·AI 기반 영어교육 연구 동향에 대한 토픽모델링 결과 토픽 1 '교수·학습 모형', 토픽 2 'AI 학습 어플리케이션', 토픽 3 '읽기 전략 및 평가', 토픽 4 '온라인 학습 및 콘텐츠', 토픽 5 '디지털·AI 도구를 활용한 쓰기 피드백', 토픽 6 '개별화 학습요인 및 효과'로 분류된 선행연구 결과(맹은경 외 2023)와는 달리 이 연구에서는 LDA 토픽 분석으로 5개의 토픽이 도출되었다. 이러한 토픽들은 전반적으로 인공지능과 디지털 기술을 중심으로 한 교육혁신과 변화의 다양한 측면을 포괄적으로 보여주고 있으며, 특히 교육 현장의 디지털화, 맞춤형 학습의 실현, 교육 서비스의 혁신 등이 주요한 논의 주제로 부각되고 있다.

차례로 살펴보면, 먼저 '인공지능 기반 교육혁신' 토픽은 인공지능(AI), 코딩, 학원, 학교 등의 단어가 높은 가중치를 가지며, 특히, 교육혁신에서 인공지능과 기술적 요소가 어떻게 융합되고 있는지를 보여준다. AI 기술의 발전은 단순한 교육 도구의 역할을 넘어 학습자의 개인적 필요에 맞춘 맞춤형 교육을 가능하게 하며, 코딩 교육과 같은 기술 중심 학습이 강화되는 추세를 반영한다. 또한, 학원과 학교에서 AI를 활용한 학습 과정이 증가하고 있으며, 이는 기존의 전통적인 교수법에서 벗어나 데이터 기반의 교육혁신이 이루어지고 있음을 시사한다. 따라서, 이 토픽은 AI와 첨단 기술을 활용한 교육의 혁신적인 변화를 강조하며, 향후 인공지능 기술과 교육이 더욱 긴밀하게 융합될 가능성을 보여준다.

© 2025 KASELL All rights reserved

둘째, ‘디지털 기반 교육환경 전환’ 토픽은 디지털, 인공지능, 교과서, 학습, 학생 등의 키워드를 중심으로 교육환경의 디지털화 과정과 그 영향을 설명하는 토픽이다. 디지털 기술의 발전은 전통적인 교과서 중심 학습에서 벗어나 다양한 멀티미디어 및 인터랙티브(interactive) 학습 도구의 활용을 가능하게 하였다. 특히, 초등교육 단계에서 디지털 기술을 활용한 교수·학습법이 확대되고 있으며, 개인 맞춤형 학습을 지원하는 디지털 교수-학습플랫폼의 도입이 활발히 이루어지고 있다. 교과서의 디지털화와 온라인 학습 시스템의 발전은 학생들이 보다 능동적으로 학습에 참여할 수 있도록 하며, 교육의 접근성과 효율성을 동시에 증대시키는 역할을 한다. 이러한 환경 변화는 디지털 기술이 교육 패러다임을 근본적으로 변화시키고 있으며, 교육 기관의 디지털 친화적 환경 전환이 필수적인 요소로 자리 잡고 있음을 의미한다.

셋째, ‘교육 시스템의 미래 지향적 변화’ 토픽은 인공지능, 디지털, 기술, 연구, 시스템 등의 키워드를 중심으로, 교육체계가 기술 혁신과 산업 변화에 따라 어떻게 진화하고 있는지를 다룬다. 4차 산업혁명 시대에 접어들면서 교육의 목표는 단순한 지식 전달에서 벗어나 창의적 문제 해결력과 융합적 사고 능력을 배양하는 방향으로 전환되고 있다. 이에 따라 AI와 빅데이터, 로봇과 같은 첨단 기술이 교육 시스템에 도입되며, 교육 연구 분야에서도 미래 교육모델을 구축하기 위한 다양한 실험이 진행되고 있다. 또한, 대학과 산업계 간의 협력을 통한 혁신적인 교육과정 개발이 활발하게 이루어지고 있으며, 이는 교육이 미래 변화에 능동적으로 대응하는 방향으로 발전하고 있음을 보여준다. 따라서, 이 토픽은 교육 시스템이 급변하는 사회·기술적 환경에 적응하고 있으며, 미래 교육을 위한 새로운 패러다임이 구축되고 있음을 시사한다.

넷째, ‘학습자 중심의 맞춤형 영어교육’ 토픽은 영어, 디지털, 학습, 언어, 회화, 번역 등의 키워드를 포함하며, 디지털 기술을 활용한 영어교육의 새로운 방향성을 조명한다. 전통적인 영어교육 방식이 문법 중심의 학습에서 벗어나 실용적이고 개인화된 학습 방식으로 전환되고 있으며, AI 기반 학습 도구를 활용한 맞춤형 영어학습이 주목받고 있다. AI 기술은 학습자의 수준과 학습 패턴을 분석하여 개별화된 학습경로를 제공하며, 음성인식 및 번역 기술을 활용한 실시간 피드백 시스템이 학습자의 실제 회화 능력 향상에 기여하고 있다. 또한, 모바일 애플리케이션과 온라인 학습 플랫폼을 활용한 자기주도적 학습이 활성화되면서, 학습자는 언제 어디서나 효율적으로 영어를 습득할 수 있게 되었다. 이에 따라, 이 토픽은 개인 맞춤형 영어교육이 디지털 기술 발전과 결합하여 학습자의 영어학습 과정을 보다 학습자 중심으로 변화시키고 있음을 보여준다.

마지막으로, ‘교육산업의 기술 혁신과 서비스’는 인공지능, 교육, 에듀테크, 기업, 서비스, 개발 등의 키워드를 포함하며, 교육과 기업 간의 협력을 통해 이루어지는 기술 혁신과 서비스 발전을 탐구하는 토픽이다. 에듀테크 기업의 발전과 AI 기반 교육 솔루션의 개발은 교육 시장의 패러다임을 변화시키고 있으며, 이는 맞춤형 학습 코스웨어, 실시간 피드백, 데이터 기반 교수법과 같은 다양한 교육 혁신을 가능하게 하고 있다. 특히, 에듀테크 기업들은 교수-학습 플랫폼을 개발하고, AI 기반 튜터링 시스템을 도입하는 등 교육 서비스를 다양화하며, 교육의 질적 향상을 위한 투자도 증가하고 있다. 또한, 온라인 강의, 인터랙티브 콘텐츠, 가상현실(VR) 및 증강현실(AR) 기술을 활용한 교육 서비스가 확대되면서, 학습경험이 더욱 몰입적이고 효과적인 형태로 진화하고 있다. 따라서, 본 토픽은 교육과 기술 산업이 상호 협력하여 지속적인 혁신을 이루고 있으며, 미래 교육 시장의 경쟁력을 높이는 데 중요한 역할을 하고 있음을

© 2025 KASELL All rights reserved

시사한다.

연구 결과를 종합하면 디지털 영어교육은 AI와 디지털 기술의 발전에 따라 급속히 변화하고 있다. 키워드 분석에서 ‘교과서’, ‘학습’, ‘에듀테크’가 높은 TF-IDF 점수를 기록하였으며, 이는 디지털 영어교육에서 교과서 중심의 학습과 AI 기술의 융합이 주요 이슈임을 나타낸다. 특히, N-gram 분석 결과, 'AI 디지털 교과서 개발 및 도입'이라는 구문이 강조되고 있으며, 이는 AI 기반 학습 도구가 교육 현장에서 중요한 역할을 하고 있음을 시사한다. 에고 네트워크 분석에서는 ‘영어’를 중심으로 학습, 교과서, 그리고 에듀테크와 빅데이터와 같은 키워드들이 긴밀히 연결되어 있음을 확인하였다. 이러한 결과는 한국의 디지털 영어교육이 빅데이터와 에듀테크를 기반으로 한 교과서와 학습을 중심으로 발전하면서 단순 기술 도입을 넘어 학습 환경의 전반적인 디지털 전환으로 이어지고 있음을 보여준다.

토픽모델링 분석 결과에 따른 디지털 영어교육의 미래 발전 방향 다섯 가지를 다음과 같이 제시한다. 첫째, 인공지능 기반 교육혁신을 통해 인공지능 테크놀로지는 영어교육에서 혁신적인 도구로 자리 잡고 있으며, 교육 방식의 근본적인 변화를 이끌고 있다. 특히, AI 기반 학습플랫폼과 디지털 교과서는 학습자의 수준과 학습 패턴을 분석하여 맞춤형 학습을 지원하는 역할을 한다. 향후 AI를 활용한 적응형 학습(adaptive learning) 시스템을 더욱 발전시키고, 학습자의 개별적 필요에 맞춘 콘텐츠를 자동 생성하는 기술을 활성화할 필요가 있다. 또한, AI 기반 자동 평가 시스템을 구축하여 학습자의 성취도를 실시간으로 분석하고 피드백을 제공함으로써 학습 효과를 극대화할 수 있다. 이를 위해 국가 차원의 AI 교육 연구 및 데이터 활용 정책이 마련되어야 하며, 교육기관과 기업 간 협력을 통해 AI 기술을 활용한 혁신적인 영어학습 모델이 개발되어야 한다.

둘째, 디지털 기반 교육환경 전환을 통해 디지털 기술의 발전은 영어교육 환경의 변화를 촉진하고 있으며, 특히, 디지털 교과서의 도입이 중요한 역할을 하고 있다. 연구 결과에 따르면 교과서가 AI 기반 디지털 영어교육에서 핵심적인 요소로 분석되었으며, 이는 전통적인 학습 자료에서 인공지능 기반 디지털 콘텐츠 중심의 학습으로 변화하고 있음을 시사한다. 미래에는 AI 기반 인터랙티브 디지털 교과서를 개발하여, 학습자가 능동적으로 참여할 수 있는 학습 환경을 구축해야 한다. 또한, 온라인 학습 플랫폼과의 연계를 통해 디지털 교과서가 학습자의 개별적 학습경로를 지원할 수 있도록 설계할 필요가 있다. 이를 위해 교육부는 테크센터와 소프트랩을 통한 외적 지원과 동시에 인공지능 디지털 교과서 개발 및 교육 현장에서의 활용을 촉진할 수 있는 정책적 지원을 강화해야 한다.

셋째, AI와 빅데이터 기술의 발전은 교육 시스템의 미래 지향적 변화를 통해 교육 시스템 전반의 변화를 야기하고 있으며, 영어교육에서도 데이터 기반 교수법의 도입이 필요하다. 특히, 학습자의 언어 습득 과정을 빅데이터 분석을 통해 체계적으로 파악하고, 맞춤형 학습 설계를 지원하는 방향으로 교육 시스템이 변화해야 한다. 미래 영어교육은 학습자의 개별적 특성과 학습 패턴을 반영한 데이터 기반 교수법을 중심으로 이루어져야 하며, 특히, 교과 최소 성취기준 보장을 위해 AI 기반 기초 학습 분석 시스템을 구축해야 한다. 또한, 디지털 영어교육이 효과적으로 운영될 수 있도록 교사 연수 과정에서 AI 및 데이터 분석 역량을 강화하는 프로그램이 마련되어야 한다.

넷째, 미래 영어교육은 학습자 중심의 맞춤형 학습을 강화하는 방향으로 발전해야 한다. 분석 결과에 따르면, 학생 맞춤형 학습이 디지털 영어교육에서 중요한 요소로 분석되었으며, 이는 AI

© 2025 KASELL All rights reserved

기반 적응형 학습 시스템과의 연계를 통해 학습자의 개별적 요구를 반영할 필요성을 시사한다. 이를 위해 AI 튜터링 시스템을 활용하여 학습자의 실시간 학습 피드백을 제공하고, 빅데이터 기반의 개인 맞춤형 콘텐츠를 자동 추천하는 시스템을 발전시켜야 한다. 또한, VR(가상현실) 및 AR(증강현실) 기술을 활용한 몰입형 학습 환경을 구축하여 실생활과 연계된 영어 학습경험을 제공하여 학습동기와 몰입도를 제고할 수 있다.

마지막으로 교육산업의 기술 혁신과 서비스를 통해 AI와 에듀테크 발전은 교육산업의 혁신을 촉진하고 있으며, 영어교육에서도 다양한 혁신적인 디지털 기술 기반 교육 서비스가 도입되고 있다. 특히, AI 기반 자동 번역 및 발음 교정 기술은 학습자가 보다 실용적인 영어 학습을 할 수 있도록 지원하며, 인공지능 기반 디지털 학습 플랫폼을 통한 글로벌 학습 환경이 조성되고 있다. 미래에는 AI 기반 학습 콘텐츠 제작 기술이 더욱 정교해지고, 자동화된 맞춤형 교육 서비스가 확대될 것으로 예상된다. 이를 위해 정부는 에듀테크 교육산업의 기술 혁신을 촉진할 수 있는 연구개발(R&D) 지원 정책을 마련하고, 기업과 교육기관 간 협력을 통해 AI 기반 영어교육 솔루션과 코스웨어 개발을 활성화해야 한다.

결론적으로, 영어교육에 대한 이러한 인공지능 기반 디지털 교육 접근은 영어 교수자 및 학습자들에게 인공지능 기술을 적용한 텍스트 분석 도구와 방법론을 제시함으로써, 디지털 영어교육을 위한 인공지능 테크놀로지 소양과 에듀테크 활용 역량에 대한 관심을 제고할 수 있을 것이다. 특히, 교사 디지털 역량 강화를 위해 균형잡힌 TPACK을 기르기 위한 체계적인 교사 연수와 정책적 지원이 필요하다(Koehler and Mishra 2009). 또한 NLP 기법을 활용한 영어 텍스트 분석은 전통적인 영어학습 방식에 비해 학생들의 흥미와 동기를 부여할 수 있는 디지털 영어학습 환경을 조성할 수 있다. 이는 에듀테크 영어 교수학습 환경에서 학습자들의 개별 적성과 수준 그리고 학습 이력에 따른 맞춤형 학습을 통해 학습 몰입과 참여도를 높이고, 2022개정 영어교육과정에서 강조하는 보여주기(viewing)과 제시하기(presenting)을 통한 복합양식 문해력(mulitimodal literacy)를 기르는데도 도움을 줄 수 있을 것이다. 미래에는 AI 기반 디지털 튜터와 가상 학습 환경이 결합하여 학습자가 보다 실질적인 영어 학습경험을 제공받을 수 있을 것이다. 또한, 디지털 영어교육은 학습 격차 해소와 글로벌 경쟁력 강화라는 두 가지 목표도 달성할 수 있다.

이 연구의 제한점으로는 텍스트 분석을 위해 키워드 분석, 네트워크 분석, 토픽 모델링 등 다양한 분석기법을 사용하였지만 각각의 텍스트 마이닝 기법의 특징과 한계로 인해 문맥적 의미를 완벽히 추출하지 못할 가능성이 있다. 또한, 텍스트 분석의 계량적 분석 결과의 의미를 해석하는 작업은 연구자의 정성적인 해석의 영역이 있다. 텍스트 분석의 적용이나 도구적 활용을 위해 전문적 통계지식과 코딩능력이 필요하지만 텍스트의 본질적인 의미를 깊이 있게 해석하고 통찰하는 일은 결국 영어교육 전문가인 사람의 몫이다. 더불어 이번 연구에서는 디지털 윤리나 정보보안, 개인정보보호, 인터넷 중독 등의 중요한 이슈가 키워드나 토픽으로 나타나지 않았다. 후속 연구에서는 이러한 위험요인을 집중 탐색하여 문제점을 예측하고 미리 대비한다면 학습자와 교사, 학부모들이 인공지능 디지털 기술활용이 교육에 적용될 때 나타나는 부작용과 위험에 대한 우려와 걱정을 줄일 수 있을 것이다.

© 2025 KASELL All rights reserved

참고문헌

권선희(Kwon, S.). 2023. 텍스트 네트워크 분석을 통한 포스트 코로나 시대 영어교육 키워드 연구(A Study on Keywords of English Education in the Post- COVID-19 by Text Network Analysis). ≪영어영문학≫(*The Mirae Journal of English Language and Literature*) 28-1, 99-121.

김성근・조혁준・강주영(Kim, S.-G., H.-J. Cho and J. Y. Kang.). 2016. 학술연구에서의 텍스트 마이닝 활용 현황 및 주요 분석기법(The Status of Using Text Mining in Academic Research and Analysis Methods). ≪정보화연구≫(*The Journal of Information Technology and Architecture*) 13-2, 317-329.

김수현・이영준・신진영・박기영(Kim, S.-H., Y. J. Lee, J. Shin and K. Y. Park.). 2019. 거시경제 분석을 위한 텍스트 마이닝(Text Mining for Economic Analysis). ≪한국경제의 분석≫(*Journal of Korean Economic Analysis*) 26-1, 1-85.

김인옥(Kim, I.). 2019. 디지털교과서 활용 쓰기 교수-학습이 초등학생들의 영어 쓰기 및 자기주도적 학습 태도에 미치는 영향(The effects of writing instructions using digital textbook on elementary school English learners' writing abilities and self-directed learning attitude). ≪초등영어교육≫(*Primary English Education*) 25-3, 149-168.

김정렬(Kim, J.-R.). 2015. 보편적 학습 설계에 근거한 영어과 디지털 교과서 개선 방안(Prospective Changes of English Digital Textbook Based on the Universal Design for Learning). ≪한국콘텐츠학회논문지≫(*The Journal of the Korea Contents Association*) 15-7, 674-683.

맹은경・고호경・손복은(Maeng, U., H. K. Ko and B. E. Son.). 2023. 디지털· AI 기반 영어교육 연구 동향: 텍스트 네트워크 분석 및 토픽모델링 중심으로(Utilizing digital·AI in English education: Research trends through text network analysis and topic modeling). ≪현대영어교육≫(*Modern English Education*) 24, 156-170.

박문영(Park, M.). 2023. 중학교 영어 AI 디지털교과서 도입을 둘러싼 쟁점과 과제(Challenges and considerations in adopting AI digital textbook for middle school English). ≪중등영어교육≫(*Secondary English Education*) 16-4, 151-167.

박한우(Park, H. W. and L. Leydesdorff.). 2004. 한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용: Daum.net에서 제공된 지역혁신에 관한 뉴스를 대상으로(Understanding the KrKwic: A computer program for the analysis of Korean text). ≪Journal of The Korean Data Analysis Society≫ 6-5, 1377-1387.

박지수・길준민(Park, J. S. and J. Gil.). 2020. 4차 산업혁명 시대의 에듀테크(Edutech in the Era of the 4th Industrial Revolution). ≪정보처리학회논문지: 소프트웨어 및 데이터 공학≫(*The Transactions of the Korea Information Processing Society*) 9-11, 329-331.

백영민(Baek, Y.-M.). 2020. 『R을 이용한 텍스트 마이닝』(*Text Mining using R*). 한울아카데미(Hanul Academy).

송민(Song, M.). 2017. 『텍스트 마이닝』(*Text Mining*). 청람(Cheongram).

© 2025 KASELL All rights reserved

신동광(Shin, D.). 2019. 인공지능 챗봇의 영어 교육적 활용 가능성과 한계(Feasibility and Constraints in Applying an AI Chatbot to English Education). ≪Brain, Digital, & Learning≫ 9−2, 29−40.

신동광(Shin, D.). 2022. 영어교육에서 AI 활용의 한계는 어디쯤일까?: 자동 지문 생성 및 자동 문항 생성의 활용 가능성(What is the limit for using AI in English education?: Feasibility of using Automated Text Generation and Automated Item Generation.). ≪영어교과교육≫(*Journal of the Korea English Education Society*) 21−3, 147−163.

심창용(Sim, C.−Y.). 2024. AI 디지털교과서 활용 영어교육의 기초(The basics of AIDT−powered English education). ≪ESP Review≫ 6−2, 137−154.

우신욱・장영재(Woo, S. and Y. Chang.). 2016. 텍스트마이닝을 활용한 연준의 통화정책방향 의결문 분석(An Analysis of FOMC Statements by Text Mining Methods). ≪Journal of the Korean Data Analysis Society≫ 18−1, 179−188.

윤태일・이수안(Yoon, T.−I. and S.−A. Lee.). 2018. 『파이썬으로 텍스트 분석하기: 전략커뮤니케이션을 위한 파이썬 텍스트 마이닝』(*Text Analysis with Python: Python Text Mining for Strategic Communication*). 늘봄(Neulbom).

이규영・김희진(Lee, G. Y and H. Kim.). 2020. 가상현실 기반 영어 학습이 초등학생의 어휘성취도 및 정의적 영역에 미치는 영향(The Effects of Virtual Reality Based English Learning on Vocabulary Achievements and Affective Domains of Elementary Students). ≪학습자중심교과교육연구≫(*The Journal of Learner−Centered Curriculum and Instruction*) 20−16, 1213−1235.

이대영・이현숙(Lee, D. Y. and H. S. Yi.). 2021. LDA 토픽 모델링의 적정 토픽 수 결정 방법 탐색: 혼잡도와 조화평균법 활용을 중심으로(Exploring methods for determining the appropriate number of topics in LDA: Focusing on perplexity and harmonic mean method). ≪교육평가연구≫(*Journal of Educational Evaluation*) 34−1, 1−30.

이삭(Lee, S.). 2019. 게이미피케이션 기반 AI 챗봇 활용 수업이 초등학생의 영어 말하기 수행 및 정의적 영역에 미치는 영향(The effects of Gamification−based Artificial Intelligence Chatbot activities on elementary English learners' speaking performance and affective domains). ≪초등영어교육≫(*Primary English Education*) 25−3, 75−98.

이은혜・안혜성・박매란(Lee, E.−H., H. Ahn and M.−R. Park.). 2023. COVID−19 팬데믹 전·후의 대학 교양영어 연구 동향 분석: Topic Modeling 활용(Research trend analysis in college general English program before and after the COVID−19 Pandemic: Utilizing topic modeling). ≪영어교육≫(*English Teaching*) 78−3, 125−145.

이정화(Lee, J. H.). 2023. 초등학교 영어 AI 디지털교과서의 전망과 과제(Promise and Challenges of AI based English Digital Textbooks in Elementary Schools). ≪중등영어교육≫(*Secondary English Education*) 16−4, 51−65.

이창수(Lee, C.−S.). 2011. 엔그램(n−gram) 분석을 통한 번역 한국어와 비번역 한국어 간의 어휘묶음 유형 차이 사례 연구(N−gram Analysis of Phrasal Differences between Translational and Orignal Korean News Texts). ≪통번역학연구≫(*Interpreting and*

© 2025 KASELL All rights reserved

Translation Studies) 15, 317−340.

임희정(Ihm, H.−J.). 2014. 다중 문해력 교육을 위한 초등영어 디지털교과서 기반 교수−학습 모형 연구(A Study of the Elementary English Digital Textbook−based Instruction and Learning Model for Multiliteracy Education). ≪한국초등교육≫(*Korean Journal of Elementary Education*) 25−1, 261−275.

정대연・배지영(Jung, D.−Y. and J. Bae.). 2021. 고등학생들의 디지털 영어학습전략 사용에 관한 연구(A study on the high school students' use of digital English learning strategies.). ≪영어교과교육≫(*Journal of the Korea English Education Society*) 20−1, 179−201.

진설아・송민(Jin, S. and M. Song.). 2016. 토픽모델링 기반 정보학 분야 학술지의 학제성 측정 연구(Topic Modeling based Interdisciplinarity Measurement in the Informatics Related Journals). ≪정보관리학회지≫(*Journal of the Korean Society for Information Management*) 33−1, 7−32.

최원경(Choi, W.). 2020. AI 챗봇을 활용한 초등영어 과정중심 말하기 평가: 가능성과 한계(Process−oriented speaking assessment of primary English using AI chatbots: Possibilities and limitations). ≪초등영어교육≫(*Primary English Education*) 26−1, 131−152.

최원경(Choi, W.). 2021. VR 기반 영어 수업에 대한 초등학생의 인식과 정의적 영역 효과 탐색(An exploratory study on elementary students' recognition and the affective effects of the VR−based English class). ≪초등영어교육≫(*Primary English Education*) 27−4, 29−49.

한종임(Han, J.−I.). 2014. 영어 디지털교과서 활용 학습 모형이 중학교 학습자의 영어학습 및 태도에 미치는 영향(The effects of learning models of English digital textbooks on Korean middle school students' English learning and attitudes toward English.). ≪멀티미디어언어교육≫(*Multimedia−Assisted Language Learning*) 17−1, 132−158.

황요한(Hwang, Y.). 2021. 메타버스(Metaverse)로의 초대, 새로운 교육공간의 필요성과 미래교육에 관한 고찰(Invitation to Metaverse: A discussion on the Need of a New Space for Future Education). ≪언어연구≫(*The Journal of Studies in Language*) 37−3, 377−389.

황은희・장지현・양현경(Hwang, E., J. Jang and H.−G. Yang.). 2019. 텍스트 마이닝 기법을 활용한 자유학기제 연구동향 분석(An Analysis of the Research Trends in Free−Semester Using Text−Mining Techniques). ≪교육문화연구≫(*Journal of Education & Culture*) 25−3, 299−318.

Blei, D. M., A. Y. Ng and M. I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3, 993−1022.

Cao, J., T. Xia, J. Li, Y. Zhang and S. Tang. 2009. A density−based method for adaptive LDA model selection. *Neurocomputing* 72(7−9), 1775−1781.

Cha, S., J. Kim and S. Nam. 2021. Research trend analysis of AI chatbot in English education. *Journal of the Korea English Education Society* 20(1), 203−225.

Chakraborty, R. and S. K. Sarma. 2024. N−Gram based Assamese Question Pattern Extraction and Probabilistic Modelling. *Journal of Electrical Systems* 20(3), 712−726.

Chang, J., S. Gerrish, C. Wang, J. L. Boyd−Graber and D. M. Blei. 2009. Reading tea leaves: How

© 2025 KASELL All rights reserved

humans interpret topic models. *Advances in Neural Information Processing System* 22, 88−296.

Cho, H. J., S. G. Kim and J. Y. Kang. 2017. An empirical analysis of doppelgänger brand image effects: Focused on the Internet community. *The Journal of Information Systems* 26(1), 21−51.

Chun, H., S. Lee and I. Park. 2021. A systematic review of AI technology use in English education. *Multimedia−Assisted Language Learning* 24(1), 87−103.

Chung, P., H. Ahn. and K. Y. Kwahk. 2019. Identification of core features and values of smart phone design using text mining and social network analysis. *Korean Business Association* 32(1), 27−47.

Danowski, J. A. 1993. Network analysis of message content. In W. D. Richards Jr. and G. A. Barnett, eds., *Progress in Communication Sciences*, 197−221. Ablex.

Dizon, G. 2020. Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology* 7(1), 1−14.

Feldman, R. and I. Dagan. 1995. Knowledge discovery in textual databases(KDT), *KDD* 95, 112−117.

Godwin−Jones, R. 2019. Future directions in informal language learning. In M. Dressman and R. W. Sadler, eds., *The Handbook of Informal Language Learning*, 457−470. John Wiley & Sons.

Green, H., K. Facer, T. Rudd, P. Dillon and P. Humphreys. 2005. *Future lab: Personalisation and Digital Technologies*. HAL Open Science.

Holmes, W., M. Bialik and C. Fadel. 2019. *Artificial intelligence: Promise and implications for teaching and learning*. Center for Curriculum Redesign.

Hong, J. Y., T. Y. Koo, J. H. Han, J. H. Park and D. S. Ma. 2024. Deriving Core Competencies for Digital Tutor Policy Operation. *JOURNAL OF The Korean Association of information Education* 28(1), 117−126.

Hotho, A., A. Nürnberger and G. Paaß. 2005. A brief survey of text mining. *Journal for Language Technology and Computational Linguistics* 20(1), 19−62.

Hubbard, P. 2017. Technologies for teaching and learning L2 listening. In C. Chapelle and S. Sauro, eds., *The Handbook of Technology and Second Language Teaching and Learning,* 93−106. John Wiley & Sons.

Hwang, G. 2020. A national curriculum framework for post−COVID education. *The Journal of Curriculum Studies* 38(4), 83−106.

Im, Tami. 2022. *A Study on How to Apply AI Education to K−12 (D20500001)*. Korea Foundation for the Advancement of Science & Creativity.

James, N. and V. Thériault. 2020. Adult education in times of the COVID−19 pandemic: Inequalities, changes, and resilience. *Studies in the Education of Adults* 52(2), 129−133.

Jang, S. J., S. H. Lee and J. Y. Chung. 2024. Analysis of Teacher Training Policies in the Field of Digital−Based Future Education Using a Four−Dimensional of Policy Theory. *The Journal*

© 2025 KASELL All rights reserved

of Educational Administration 42(1), 253−282.

Jeong, G. Y. 2020. A study on the learning effect of multimedia literacy education. *The Journal of the Convergence on Culture Technology* 6(1), 63−68.

Kang, M. S., H. S. Kim, H. K. Kim and Y. H. Joo. 2023. Exploring ways to strengthen artificial intelligence and digital utilization education policies for teachers. *The Journal of Korean Teacher Education* 40(2), 111−140.

Kim, N. Y. 2018. Different chat modes of a chatbot and EFL students' writing skills development. *Studies in Foreign Language Education* 32(1), 263−290.

Kim, J. S. 2020. Artificial intelligence literacy's big ideas and achievement standards for primary English language learners. *The Journal of Korea Elementary Education* 31, 135−153.

Kim, K. 2019. Is Korean education properly preparing future generations for an era of change? *Korean Journal of Educational Research* 57(1), 221−252.

Kim, H. S. and J. A. Park. 2023. Analysis of Research Trends in Cosmetology Education using Text Mining. *Journal of the Korean Society of Cosmetology* 29(3), 582−592.

Kim, H. S., S. Jun, S. Y. Choi and S. Kim. 2020. Development and application of education program on understanding artificial intelligence and social impact. *The Journal of Korean Association of Computer Education* 23(2), 21−29.

Kim, M., J. Yeom, H. Jung and C. Lim. 2021. A review of research on artificial intelligence chatbot in education through the lens of activity theory. *The Journal of Educational Information and Media* 27(2), 699−721.

Kim, S. G. and J. Kang. 2018. Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. *Information Processing & Management* 54(6), 938−957.

Kim, S., H. J. Cho and J. Y. Kang. 2016. The status of using text mining in academic research and analysis methods. *Journal of Information Technology and Architecture* 13(2), 317−329.

King, F. B., M. Harner and S. W. Brown. 2000. Self−regulatory behavior influences in distance learning. *International Journal of Instructional Media* 27(2), 147.

Koehler, M. J. and P. Mishra. 2009. What is technological pedagogical content knowledge? *Technology and Teacher Education* 9(1), 60−70.

Kramsch, C., D. Morgenstern and J. H. Murray. 2013. An overview of the MIT Athena language learning project. *CALICO Journal* 2(4), 31−34.

Lau, K. N., K. H. Lee and Y. Ho. 2005. Text mining for the hotel industry. *Cornell Hotel and Restaurant Administration Quarterly* 46(3), 344−362.

Lee, H. J. 2020. A systematic review of artificial intelligence use in English learning: Focus on higher education. *The Journal of Humanities and Social Science 21* 11(6), 2027−2042.

Lee, J., J. Jung and Y. Jung. 2021. *AI Educational Revolution*. Siwon Books.

Lee, S. M. 2020. The impact of using machine translation on EFL students' writing. *Computer−Assisted Language Learning* 33(3), 157−175.

Lee, S. and J. Kim. 2022. A research trend analysis of elementary English education using AI

© 2025 KASELL All rights reserved

technology. *Journal of the Korea English Education Society* 21(3), 165−182.

Luckin, R., W. Holmes, M. Griffiths and L. B. Forcier. 2016. *Intelligence Unleashed: An Argument for AI in Education*. Pearson.

Feldman, R. and I. Dagan. 1995. Knowledge Discovery in Textual Databases (KDT). In Proceedings of *the First International Conference on Knowledge Discovery and Data Mining* (KDD−95), 112−117. AAAI Press.

Ministry of Education (MoE). 2016. *A Draft of the Directions and Strategies for Mid− to Long−term Education Policies in Response to the Intelligent Information Society*. Ministry of Education.

Ministry of Science and ICT (MSICT). 2019. *Announcement of the National AI Strategy*. Ministry of Science and ICT.

Ministry of Education (MoE). 2020. *Simultaneous Announcement of the Comprehensive Plan for Science−Mathematics−Information−Convergence Education*. Ministry of Education.

Ministry of Education (MoE). 2022a. *2022 Revised national curriculum*. Ministry of Education.

Ministry of Education (MoE). 2022b. *The 2022 Comprehensive Plan for Nurturing Digital Talent*. Ministry of Education.

Ministry of Education (MoE). 2023. *Digital−based Education Innovation Plan*. Ministry of Education.

Nam, D. 2022. Twenty years of Multimedia−Assisted Language Learning (MALL): A network visualization analysis. *Multimedia−Assisted Language Learning* 25(4), 170−191.

Oh, Y. L., J. O. Min, Y. G. Kim, D. J. Kim, Y. K. Park and B. G. Lee. 2017. A comparative analysis for the extraction of similar patent claims based on word embedding. *Paper presented at the Korean Institute of Information Scientists and Engineers*, 20−22.

Park, S. A. 2024. Critical discourse analysis of AI education policy. *Korean Journal of Sociology of Education* 34(2), 165−207.

Park, J. H. and I. Y. Yang. 2020. Utilizing an AI−based grammar checker in an EFL writing classroom. *Korean Journal of Applied Linguistics* 36(1), 97−120.

Persson, V. and J. Nouri. 2018. A systematic review of second language learning with mobile technologies. *iJET* 13(2), 188−210.

Popping, R. 2003. Knowledge graphs and network text analysis. *Social science information* 42(1), 91−106.

Prensky, M. 2001. Digital natives, digital immigrants part 2: Do they really think differently? *On the Horizon* 9(6), 1−6.

Punitha, S. C., P. R. J. Thangaiah and M. Punithavalli. 2014. Performance analysis of clustering using partitioning and hierarchical clustering techniques. *International Journal of Database Theory and Application* 7(6), 233−240.

Ramirez, E. H., R. Brena, D. Magatti and F. Stella. 2012. Topic model Validation. *Neurocomputing* 76(1), 125−133.

© 2025 KASELL All rights reserved

Savage, M. and A. Barnett. 2015. *Digital Literacy for Primary Teachers*. Critical Publishing Ltd.

Shim, Y. M. 2023. Policy Challenges of Digital Tutor for Reducing the Digital Education Gap. *Journal of Creative Information Culture* 9(1), 33−43.

Sowa, J. F. 1984. *Conceptual Structures: Information Processing in Mind and Machine*. Addison−Wesley Longman Publishing Co., Inc..

Yeo, Y. 2020. Shaping the innovation policy in the post−COVID era: Focusing on building creative learning capabilities. *Journal of Technology Innovation* 28(4), 153−165.

Yun, K. and U. Maeng. 2017. A meta−analysis of the effectiveness of web−based English instruction on L2 learning. *The Journal of Linguistic Science* 83, 193−224.

Zhang, Z. 2020. Engaging with automated writing evaluation feedback on L2 writing: Student perceptions and revisions. *Assessing Writing* 43, 78−91.

Zou, S. 2017. Designing and practice of a college English teaching platform based on artificial intelligence. *Journal of Computational and Theoretical Nanoscience* 14(1), 104−108.

예시 언어(Examples in): 한국어, English
적용가능 언어(Applicable Languages): 한국어, English
적용가능 수준(Applicable Level): Primary/Secondary/Tertiary

© 2025 KASELL All rights reserved