

키워드 출현 빈도 분석과 CONCOR 기법을 이용한 ICT 교육 동향 분석

이영석*

서울교육대학교 컴퓨터교육과 교수

Analysis of ICT Education Trends using Keyword Occurrence Frequency Analysis and CONCOR Technique

Youngseok Lee*

Professor, Dept. of Computer Education, Seoul National University of Education

요약 본 연구는 기계학습의 키워드 출현 빈도 분석과 CONCOR(CONvergence of iteration CORrection) 기법을 통한 ICT 교육에 대한 흐름을 탐색한다. 2018년부터 현재까지의 등재지 이상의 논문을 'ICT 교육'의 키워드로 구글 스칼라에서 304개 검색하였고, 체계적 문헌 리뷰 절차에 따라 ICT 교육과 관련이 높은 60편의 논문을 선정하면서, 논문의 제목과 요약에 중심으로 키워드를 추출하였다. 단어 빈도 및 지표 데이터는 자연어 처리의 TF-IDF를 통한 빈도 분석, 동시 출현 빈도의 단어를 분석하여 출현 빈도가 높은 49개의 중심어를 추출하였다. 관계의 정도는 단어 간의 연결 구조와 연결 정도 중심성을 분석하여 검증하였고, CONCOR 분석을 통해 유사성을 가진 단어들로 구성된 군집을 도출하였다. 분석 결과 첫째, '교육', '연구', '결과', '활용', '분석'이 주요 키워드로 분석되었다. 둘째, 교육을 키워드로 N-GRAM 네트워크 그래프를 진행한 결과 '교육과정', '활용'이 가장 높은 단어의 관계로 나타났다. 셋째, 교육을 키워드로 군집분석을 한 결과, '교육과정', '프로그래밍', '학생', '향상', '정보'의 5개 군이 형성되었다. 이러한 연구 결과를 바탕으로 ICT 교육 동향의 분석 및 트렌드 파악을 토대로 ICT 교육에 필요한 실질적인 연구를 수행할 수 있을 것이다.

키워드 : 기계학습, 단어 문서 행렬, CONCOR 기법, ICT 교육, 동향 분석, 키워드 출현 빈도

Abstract In this study, trends in ICT education were investigated by analyzing the frequency of appearance of keywords related to machine learning and using conversion of iteration correction(CONCOR) techniques. A total of 304 papers from 2018 to the present published in registered sites were searched on Google Scholar using "ICT education" as the keyword, and 60 papers pertaining to ICT education were selected based on a systematic literature review. Subsequently, keywords were extracted based on the title and summary of the paper. For word frequency and indicator data, 49 keywords with high appearance frequency were extracted by analyzing frequency, via the term frequency-inverse document frequency technique in natural language processing, and words with simultaneous appearance frequency. The relationship degree was verified by analyzing the connection structure and centrality of the connection degree between words, and a cluster composed of words with similarity was derived via CONCOR analysis. First, "education," "research," "result," "utilization," and "analysis" were analyzed as main keywords. Second, by analyzing an N-GRAM network graph with "education" as the keyword, "curriculum" and "utilization" were shown to exhibit the highest correlation level. Third, by conducting a cluster analysis with "education" as the keyword, five groups were formed: "curriculum," "programming," "student," "improvement," and "information." These results indicate that practical research necessary for ICT education can be conducted by analyzing ICT education trends and identifying trends.

Key Words : Machine learning, Word document matrix, CONCOR technique, ICT education, Trend analysis, Keyword occurrence frequency

This work was supported by the 2022 Research Fund of Seoul National University of Education.

*Corresponding Author : Youngseok Lee(yslee38@snu.ac.kr)

Received December 21, 2022

Accepted January 20, 2023

Revised January 18, 2023

Published January 28, 2023

1. 서론

정보 기술과 인공지능(AI)의 발전과 활용은 다양한 분야에서 확산하고 있으며, 교육 분야에서도 AI 기반의 새로운 기술 등을 도입하여 교수-학습 방식을 변화시키고 있다. 정보통신기술과 AI 기술이 발전함에 따라 학습자에게는 맞춤형 학습을 제공하고, 학습자 수준에 적합하면서도 적응적인 평가를 제공하고, 온라인, 모바일 또는 온라인과 오프라인이 혼합된 학습 경험에서 교수자와 학습자가 의미 있는 상호 작용을 촉진할 수 있는 기법에 관한 연구가 많이 이루어지고 있다[1].

하지만, 정보통신기술과 AI 교육 기술이 수행할 수 있는 것과 실제 교육 환경에서 실제로 적용하는 방법 사이에 격차가 있으므로, 정보통신과 AI 교육 관련 문헌 연구를 토대로 중요한 주제어를 추출하고, 연구 동향에서 주요한 시사점과 활용 방향성을 도출하기 위한 연구가 필요하다.

연구 동향과 주요 시사점을 도출한 연구의 사례로 UNESCO의 미디어와 정보 리터러시 교육내용 분석과 교육과정에 관한 연구에서는 문헌정보학 관점에서 정보 리터러시의 개념과 교육내용 및 교육과정을 분석한 연구가 있다[2]. 중등 컴퓨터교육 연구주제 동향 분석을 위해 토픽 모델링 기법을 적용하여 연도별 토픽의 추이를 분석한 연구가 있다[3].

국내 ICT 교육 동향 및 해외 사례 분석에서는 정보 교육과정에 제시된 교과 시수와 학습 요소가 효과적으로 제시되어 있는지 국내의 연구를 바탕으로 비교·분석한 연구가 있다[4]. 미래 교육 동향 분석을 위해 언어 네트워크 분석을 활용한 연구에서는 자료의 제목과 키워드를 대상으로 빈도 분석과 언어 네트워크 분석을 시행한 연구를 통해 교육 동향 분석의 연구 방법과 사례를 살펴볼 수가 있다[5].

이러한 연구를 바탕으로 본 연구에서는 ICT 교육 분야에서 기존 연구의 연구주제 및 정보통신기술 적용, 교육 효과에 관한 사항을 키워드로 수집 및 정리하고, 기계학습의 키워드 출현 빈도 분석과 함께 수집한 문헌 연구의 CONCOR(CONvergence of iteration CORrealtion) 기법을 통해 ICT 교육에서의 주요 이슈와 동향을 분석하고 시사점을 도출하고자 한다.

2. 선행연구

2.1 기계학습 기반의 문헌 연구

2010년부터 2019년까지 학제 간 연구 분야의 지식 진화에 대한 전체적인 그림을 제공하기 위해 인공지능 교육(AIED)에 관한 1,830개의 연구 기사를 검토하여 AIED 내의 지식 구조, 지식 클러스터 및 트렌드 키워드를 시간에 따라 식별하기 위해 키워드 동시 발생 네트워크(KCN)를 제안한 연구가 있다[6]. 교육 분야에서 인공지능에 관한 많은 문헌을 체계적으로 검토하여 지식 연결의 기본 패턴을 발견하고 향후 개발에 대한 통찰력을 제공한다.

토픽 모델링 기법을 적용한 연구에서는 기계학습의 키워드 추출을 바탕으로 중등 컴퓨터교육 연구주제를 탐색하였고[3], 미래 교육 동향 분석을 위해서 키워드의 빈도 분석과 함께 언어 네트워크 분석을 실시한 연구가 있다[5].

기계학습 모델을 활용하여 다양한 애플리케이션에서 복잡한 패턴을 인식하는 연구 중 금융시장의 가격 예측에 관한 연구가 있다[7]. 해당 주제와 가장 관련성이 높은 참고문헌을 일관성 있게 분석하기 위해서는 객관적인 방법이 필요한데, 연구 영역에서 가장 중요한 텍스트를 강조하는 요약물을 활용하여 57개의 논문 요약을 검토하고 시장, 자산, 방법 및 변수에 대한 분류를 제안한 연구가 있다[7, 8].

2.2 문헌 키워드 분석 연구

소셜 네트워크의 빅데이터 분석을 통해 여가 활동의 다양한 키워드를 탐색하는 연구가 있다[8]. 이 연구는 네이버와 다음에서 제공하는 웹페이지, 블로그, 카페, 지식인/다음팁을 분석 채널로 선정하고, 빅데이터 검색을 위한 여가 활동을 주요어로 추출하고, 단어 간의 연결구조와 연결 정도 중심성을 분석하여 검증하였다. 또한, CONCOR 분석을 통해 유사성을 가진 단어들로 구성된 군집을 도출하였다[8].

블로그와 온라인 카페의 글을 바탕으로 키워드를 파악하고, 주요 토픽을 발견하는 연구가 있다[9]. '코로나'와 '가족' 등이 함께 언급된 문서 총 351,734건을 분석 후 토픽모델 분석을 시행하였다[9].

이처럼 현재의 트렌드나 연구 동향과 함께 주요 키워드를 바탕으로 시대의 흐름을 함께 파악하기 위하여 다양한 자연어 처리 기반의 토픽 모델링, 빅데이터 기반의 키워드 분석, 단어들의 빈도수 및 상관관계 분석 등을 통해

주요 이슈와 시사점을 도출하는 연구가 수행되고 있다[3, 5, 6].

3. 연구 방법

3.1 연구 대상 및 자료수집방법

ICT 교육 동향 분석을 위해 본 연구에서는 국내외 문헌에 대한 접근성이 편리한 구글 스칼라를 이용하여 ‘ICT 교육’이라는 키워드로 검색된 304건의 논문을 체계적 문헌 연구 기법에 따라서 60편으로 최종 선정하는 과정을 거쳤다. 체계적 문헌 연구 기법의 과정은 Fig. 1과 같다.

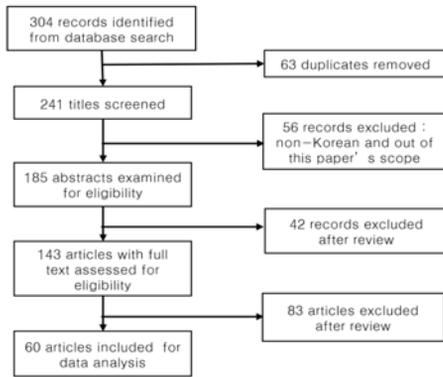


Fig. 1. Process of systematic literature review

검색 자료는 2018년부터 2022년 11월 사이에 발행된 문헌 중 한국어로 된 등재지 이상의 학술지 논문을 선정하고, 논문의 제목과 초록을 기준으로 자료를 수집하였다.

논문의 제목 또는 초록이 위의 기준에 충족되지 않을 때 본 연구에서 제외하였고, 선별된 185개의 문헌에 대해서 초록 검토 후, 본 연구의 주제와 적합하지 않은 42개의

문헌을 제외한 143개의 논문 전체 내용을 검토하였다. 마지막으로 본문 내용 검토를 통해서 총 60개의 문헌을 선별하였다.

3.2 조사 도구 및 절차

본 연구에서는 ICT 교육 동향 분석을 위해 수집된 문헌의 제목과 요약에 바탕으로 자연어 처리와 기계학습의 키워드 출현 빈도 분석을 위해 키워드를 정제하고, 추출된 키워드의 빈도수를 제시하여 순위나 유사성, 연관성을 파악하였고, 본문 전체의 내용과 비교하여 적절성을 검토하였다[10-13].

ICT 교육 동향 분석에 적합하도록 본 연구에서는 ICT 교육에 관련된 문서 속에서 추출된 단어 중 ICT 교육과 상관없는 단어들은 삭제하였고, 같은 의미로 쓰인 단어들은 통합하면서 정제하였다.

3.3 자료 처리 방법

본 연구에서는 ICT 교육 동향 분석을 위해 키워드 출현 빈도를 기반으로 의미연결망 분석 기법의 사회연결망 분석을 활용하였다[14, 15]. 사회연결망 분석은 문서 내에서 단어들의 구조적 관계를 분석하고, 구조 패턴과 의미를 분석하는 기법으로, 단어 간의 연결 구조를 파악하고, ICT 교육과 관련된 주제어들의 네트워크 분석 및 시각화한 뒤, 유사한 단어들의 군집을 도출하기 위해 CONCOR 분석하였다.

4. 연구 결과

4.1 데이터 분석 결과

본 논문에서 추출된 키워드의 출현 빈도를 시각화한 결과는 Fig. 2와 같다. 이 중 네트워크 시각화의 한계성으

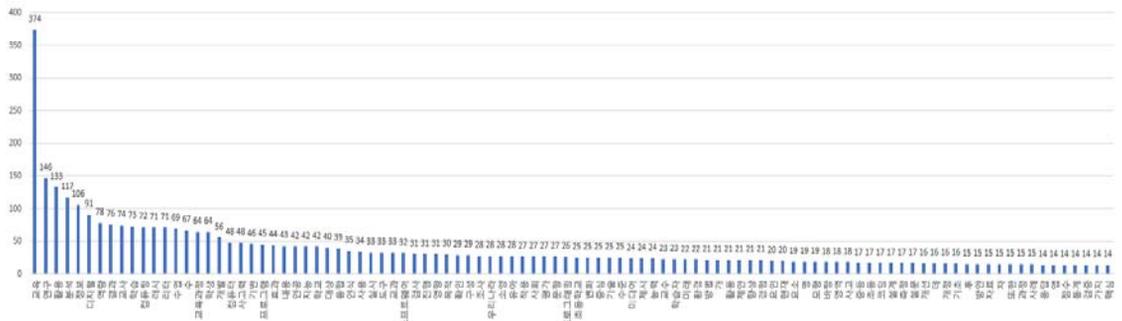


Fig. 2. Keyword frequency visualization

로 상위 60개의 키워드를 중심으로 ICT 교육 키워드에 대한 출현 빈도를 분석하였다. ICT 교육 연구 동향에 적합한 키워드로 '교육', '연구' 등 키워드의 출현 빈도가 가장 높게 나타났다.

문서 전처리 과정에서 중복 제거 전에는 6,984개의 키워드, 중복 제거 후에는 1,131개의 키워드가 추출되었다. 그중에서 출현 빈도가 가장 높은 단어를 빈도수 기준값 100으로 정한 뒤, 상위 10개의 추출된 단어의 출현도와 TF-IDF에 의한 연결 정도 중심성을 분석한 결과는 Table 1과 같다.

Table 1. Frequency of extracted words

No.	Word	Frequency	IDF
1	education	100	0.806
2	research	66	1.057
3	analysis	62	1.098
4	uses	53	1.205
5	result	47	1.291
6	information	37	1.374
7	curriculum	34	1.392
8	learning	32	1.470
9	competency	31	1.536
10	contents	31	1.584

연결 정도 중심성은 각 단어의 빈도수를 바탕으로 단어들에 중심에 있는지를 나타내는 지표로서 IDF 수치가 적을수록 키워드의 영향력이 높다고 해석할 수 있다[11]. 본 연구에서는 '교육(빈도수:100, IDF: .806)' 단어가 가장 출현도가 높으면서 중심성이 높게 나타났고, 그다음으로 '연구(66, 1.057)', '분석(62, 1.098)', '활용(53, 1.205)' 순으로 나타났다.

4.2 데이터 시각화 결과

키워드의 중심성을 파악하기 위한 네트워크 내의 연결 관계를 시각적으로 도출하기 위하여 N-GRAM 네트워크 그래프를 실시한 결과는 Fig. 3과 같다.

N-GRAM으로 표현된 네트워크 정보를 바탕으로 출현도와 중심성을 기준으로 군집화 분석을 위한 클래스 다이어그램 시각화한 결과는 Fig. 4와 같다.

군집화 분석을 한 결과, 군집별 키워드 분포의 차이는 있지만, 6개의 군집으로 나타났다. 군집화 분석 결과를 바탕으로 네트워크 내의 패턴 및 연결 관계를 시각적으로 표현하기 위해 연결 정도 중심성을 바탕으로 CONCOR 분석을 한 결과는 Fig. 5와 같다.

그룹의 개수는 6개로 나타났고, 가장 빈도수가 많은

'교육' 키워드 중심으로는 '수업', '학습', '확인', '정보', '해결', '실시', '응답', '조사', '개'로 나타났다. '개' 단어는 전처리 과정에서 제외하려고 하였으나, 교과목명 및 논문의 주요 용어로 사용되고 있어서 제외하지 않았는데, '교육' 영역으로 군집화가 이루어진 점이 특이하였다. 특히, 교육 중심 그룹에서 '수업', '학습' 등의 키워드가 군집으로 이루어졌고, '정보'라는 요인이 중요하게 사용된 점을 확인할 수 있었다.

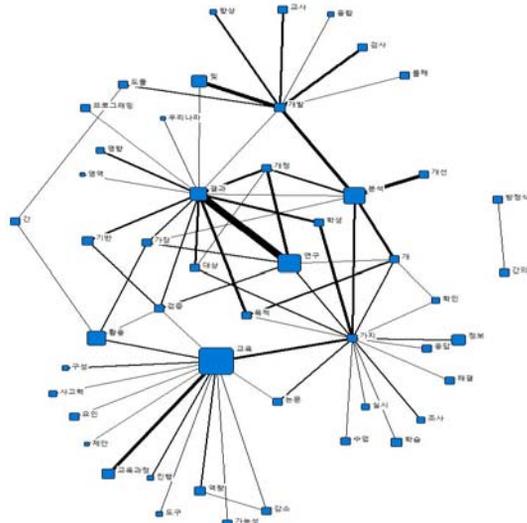


Fig. 3. N-GRAM network analysis result

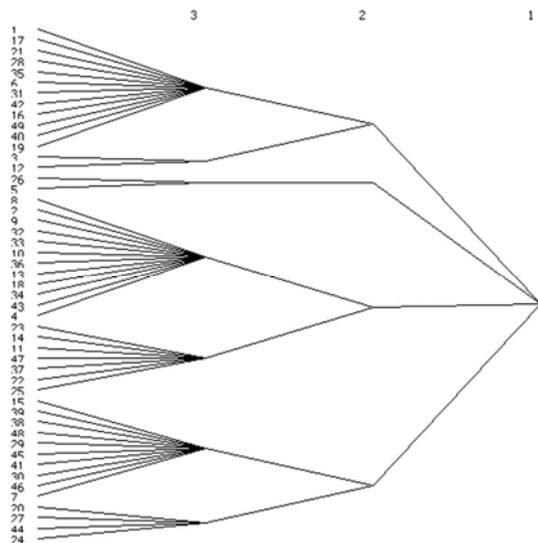


Fig. 4. Class diagram visualization results

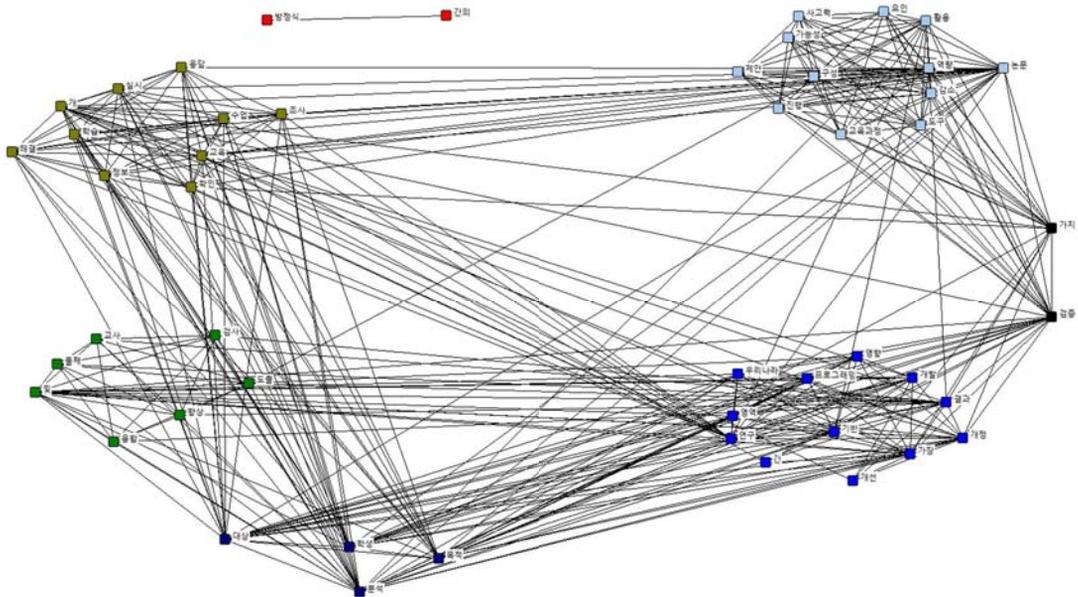


Fig. 5. Network CONCOR analysis result

두 번째 그룹은 ‘융합’, ‘교사’, ‘검사’, ‘도출’, ‘향상’, ‘둘째’, ‘및’으로 나타났다. 첫 번째 그룹의 ‘개’와 마찬가지로 이유로 ‘및’을 포함했는데 융합 영역으로 포함된 것으로 보았을 때, 2개 이상의 영역을 합성하는 용어로 적절하게 군집화가 이루어진 것으로 판단되었고, ‘융합’ 영역에 ‘교사’와 ‘검사’, ‘향상’ 등의 용어가 나타난 것이 특징이었다.

세 번째 그룹은 ‘연구’, ‘영역’, ‘우리나라’, ‘프로그램’, ‘기반’, ‘영향’, ‘개발’, ‘결과’, ‘개정’, ‘가장’, ‘개선’, ‘간’으로 나타났다. ‘교육’ 단어의 빈도수 다음으로 많았던 ‘연구’ 단어는 ‘영역’, ‘프로그래밍’, ‘영향’, ‘개발’ 등 전반적으로 연구에 필요한 키워드로 군집화된 것을 봤을 때 적절하다고 할 수 있다.

네 번째 그룹은 ‘역량’, ‘논문’, ‘구성’, ‘감소’, ‘제안’, ‘진행’, ‘교육과정’, ‘도구’, ‘가능성’, ‘사고력’, ‘요인’, ‘활용’으로 나타났다. ‘교육과정’과 ‘역량’ 등 비슷한 빈도수로 많이 사용된 용어의 그룹에서는 ‘역량’에 적합한 ‘사고력’, ‘요인’, ‘도구’ 등의 용어로 군집화되었다.

다섯 번째 그룹은 ‘학생’, ‘분석’, ‘목적’, ‘대상’으로 나타났다. 여섯 번째 그룹은 ‘가지’, ‘검증’으로 나타났다. 4개의 그룹은 중심성이 강하면서 빈도수가 많이 나타난 키워드 중심으로 군집화가 이루어졌고, 다섯 번째와 여섯 번째 그룹은 연구자의 측면에서 많이 나타난 키워드로 인해 군집화가 이루어진 것으로 분석되었다.

5. 결론

본 연구는 ICT 교육에 관한 전반적인 연구 흐름을 살펴보고, ICT 교육 동향 분석을 위해서 자연어 처리 기반의 키워드 출현 빈도를 기반으로 사회연결망 분석하고, CONCOR 분석을 통해 군집 형태로 시각화 처리하였다.

이를 위하여 체계적 문헌 연구 기법에 따라 ‘ICT 교육’이라는 키워드를 바탕으로 연구 제목과 연구 요약의 기준으로 키워드 분석 결과, 빈도수가 높은 단어들을 기준으로 중심성을 나타내고, 중심에 있는 단어들과 사회연결망을 가진 단어들로 군집화가 이루어진 것을 파악할 수 있었다.

추후 연구에서는 특정 논문 검색 사이트 이외에도 다양한 논문 검색 사이트를 활용하여 좀 더 다양하고 많은 논문의 주요 키워드를 바탕으로 검색하고 분석한다면, 전체적인 연구의 흐름뿐만 아니라 세부적인 연구 이슈 등을 파악할 수 있을 것이다.

REFERENCES

- [1] J. Hyunjung Woo, Hyejeong Cho and Yul Choi. (2018). Trends and prospects of education using ICT. *Informatization Policy*, 25(4), 3-36. DOI : 10.22693/NIAIP.2018.25.4.003
- [2] Park Joo-hyun. (2020). A Study on the Analysis of Educational Content and Curriculum of UNESCO's

- Media and Information Literacy. *Journal of the Korean Library and Information Society*, 54(2), 349-374.
- [3] Jeongwon Min, Jaekwoun Shim. (2021). A Study on Domestic Research Trends in Secondary School Computer Education. *The Journal of Korean Association of Computer Education*, 24(1), 29-36. DOI : 10.32431/kace.2021.24.1.004
- [4] Seokjun Woo, Dukhoi Koo. (2021). Trend of ICT Education in Korea and Analysis of Overseas Cases. 2021 Proceedings of the Korea Association for Information Education Summer Conference, 259-265.
- [5] Hee-Jeong Seo, Jong-Yeon Lee. (2018). Analysis of Future Education Trends Using Semantic Network Analysis. *Journal of Korean Association for Educational Information and Media*, 24(4), 649-678. DOI : 10.15833/KAPEIAM.24.4.649
- [6] Feng, S., & Law, N. (2021). Mapping Artificial Intelligence in Education Research: a Network-based Keyword Analysis. *International Journal of Artificial Intelligence in Education*, 31(2), 277-303. DOI : 10.1007/s40593-021-00244-4
- [7] Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124, 226-251. DOI : 10.1016/j.eswa.2019.01.012
- [8] Hyorim Kim, & Ikki Jeon. (2018). Leisure activity keyword analysis using text mining technique. *Journal of the Korean Society of Leisure and Recreation*, 42(3), 59-69. DOI : 10.26446/kjlrp.2018.9.42.3.59
- [9] Sunyoung Park, & Jaerim Lee. (2021). COVID-19 and family life through social big data: Topic model approach. *Journal of Korean Contents Society*, 21(3), 282-300. DOI : 10.5392/JKCA.2021.21.03.282
- [10] Song Eui-seong, & Lim Hwa-kyung. (2021). Necessity of elementary school information subject through analysis of overseas SW·AI education status. *Journal of Information Education Society*, 25(2), 301-308. DOI : 10.14352/jkaie.2021.25.2.301
- [11] Sungyoon Moon, & Subeom Shin. (2020). Investigation and analysis of suitability of content selection for data structure education in elementary school curriculum. *Journal of the Korea Information and Communications Society*, 24(2), 311-320. DOI : 10.6109/jkiice.2020.24.2.311
- [12] Younghoon Sung. (2021). Exploration of independent application of elementary information subject through analysis of digital literacy status of elementary textbooks. *Journal of Information Education Society*, 25(2), 265-277. DOI : 10.14352/jkaie.2021.25.2.265
- [13] Eunkyong Lee. (2018). Perspectives and Challenges of Informatics Education: Suggestions for the Informatics Curriculum Revision. *The Journal of Korean Association of Computer Education*, 21(2), 1-10. DOI : 10.32431/kace.2018.21.2.001
- [14] Wasserman, S. & Faust, K. (1994). Social network analysis: method and applications. NewYork: Cambridge University Press.
- [15] Ji-suk Hong, Ick-keun Oh. (2016). Image difference of before and after an incident using social big data analysis : Focusing on a ramp return of "K" airline. *International Journal of Tourism and Hospitality Research*, 30(6), 119-133. DOI : 10.21298/IJTHR.2016.06.30.6.119

이영석(Youngseok Lee)

[정회원]



- 1999년 2월 : 서울교육대학교 초등교육과 (교육학사)
- 2001년 2월 : 서울교육대학교 컴퓨터교육과 (교육학석사)
- 2009년 8월 : 한양대학교 전자통신전파공학과 (공학박사)
- 2016년 3월~2022년 2월 : 강남대학교 KNU참인재대학 교수
- 2022년 3월~현재 : 서울교육대학교 컴퓨터교육과 교수
- 관심분야 : 정보(SW)교육, 지능형 웹 정보 시스템, 컴퓨터 비전
- E-Mail : yslee38@snu.ac.kr