

## Social awareness of Arduino and artificial intelligence using big data analysis

Eun-Sang Lee\*

\*Professor, Dept. of Technology and Home Economics, Kongju National University, Kongju, Korea

### [Abstract]

This study aimed to identify the development direction of Arduino-based boards relating to artificial intelligence based on social awareness identified using big data analytical methods. For the purpose, big data were extracted through the Textom website, focusing on keywords that included 'Arduino + artificial intelligence' and 'Arduino + AI', and these data were refined and analyzed using the Textom website and the UNICET program. In this study, big data analyses, including frequency analysis, TF-IDF analysis, Degree Centrality analysis, N-gram analysis, and CONCOR analysis, were performed. The analyses' results confirmed that keywords relating to education and coding education, keywords relating to making and experience based on Arduino, and keywords relating to programs were the main keywords used in Arduino- and artificial intelligence-related Internet documents, and clusters were formed based on these keywords confirmed. The social awareness of Arduino and artificial intelligence was evaluated, and the direction of board development was identified based on this social awareness. This study is meaningful in that it identified various factors of board development based on the general public's social awareness, which was evaluated using a big data analysis method. This study may serve as a point of reference for future researchers or developers wishing to understand user needs using big data analysis methods.

▶ **Key words:** Arduino, Arduino-based boards, artificial intelligence, big data analysis, social awareness

### [요 약]

이 연구의 목적은 빅데이터 분석 방법으로 확인한 사회적 인식을 기반으로 인공지능과 관련된 아두이노 기반 보드의 개발 방향을 확인하는 데 있다. 이를 위해 텍스트 사이트를 통해 '아두이노+인공지능', '아두이노+AI' 등의 키워드를 중심으로 빅데이터를 추출하였고, 이 데이터를 텍스트 사이트와 UNICET 프로그램을 이용하여 정제 및 분석하였다. 이 연구에서는 빈도 분석, TF-IDF 분석, 연결 중심성 분석, N-gram 분석, CONCOR 분석 등의 빅데이터 분석을 수행하였다. 분석 결과 아두이노 및 인공지능 관련 인터넷 문서에서는 교육 및 코딩 교육과 관련된 키워드, 아두이노를 기반으로 제작 및 체험 관련 키워드, 프로그램 관련 키워드가 주요 키워드임을 확인하였으며, 이들 키워드를 바탕으로 한 군집이 형성됨을 확인하였다. 이 연구를 통해 아두이노 및 인공지능과 관련된 사회적 인식을 파악하였고, 이를 기반으로 한 보드 개발의 방향성을 확인할 수 있었다. 이 연구는 일반 대중의 사회적 인식을 빅데이터 분석 방법을 활용하여 파악한 후, 이를 기반으로 보드 개발의 여러 가지 요인들을 확인하였다는 점에서 의의가 있다. 이 연구는 빅데이터 분석 방법으로 사용자의 요구를 파악하고자 하는 연구자나 개발자들이 참고할 수 있는 자료로 활용될 수 있을 것이다.

▶ **주제어:** 아두이노, 아두이노 기반 보드, 인공지능, 빅데이터 분석, 사회적 인식

- 
- First Author: Eun-Sang Lee, Corresponding Author: Eun-Sang Lee
  - Eun-Sang Lee (eslee@kongju.ac.kr), Dept. of Technology and Home Economics, Kongju National University
  - Received: 2022. 12. 16, Revised: 2023. 01. 10, Accepted: 2023. 01. 12.

## I. Introduction

일반적으로 소비자의 필요나 요구와 같은 사회적 인식을 파악하는 작업은 제품 개발에서 선행되는 기본적인 과업으로 알려져 있다. 오늘날 사회는 4차 산업혁명 시대의 도래와 함께 매우 빠르게 변화하고 있으며, 소비자의 필요나 요구 역시 하루가 다르게 변화하고 있다. 이에 제품을 개발하는 연구진들 역시 매 순간 소비자의 사회적 인식을 파악하고 그들의 필요나 요구를 충족시키기 위한 다양한 시도를 하고 있다.

사회적 인식을 조사하는 전통적으로 방법에는 설문을 이용한 방법이 있다. 이 방법은 특정 집단을 대상으로 설문을 실시하여 데이터를 수집하고 이를 분석하여 연구에 필요한 시사점을 확인한다[1]. 그러나 이 방법은 설문 구성이나 해석 과정에서 연구자의 주관이 개입될 수 있고, 설문 대상이나 설문 대상자의 수에 따라 연구 결과가 달라질 수 있는 등의 한계가 있었다.

이와 같은 사회적 인식을 조사하는 전통적 방법이 가진 한계점은 최근 빅데이터 분석을 이용한 방식으로 극복되고 있다. 이는 빅데이터의 경우 연구 대상이 이미 생성된 사회 전체의 온라인 데이터가 되기 때문에 데이터의 생성이나 수집 과정에서 연구자가 개입할 여지가 없어 보다 객관적인 연구 결과의 도출이 가능하기 때문이다[1, 2].

이 연구에서는 빅데이터 분석의 장점을 활용하여 새로운 보드 개발의 방향성을 확인하고자 하였다. 기존 보드 개발 방법으로는 사용자 의견 검토, 설문, 관찰 등의 방법이 활용되었는데, 이러한 방법은 극히 일부 사용자의 의견만 반영되었고 연구를 수행하는데 비용이나 시간이 많이 소요되는 문제점이 있었다. 본 연구진은 빅데이터 분석 방법이 이러한 제한점을 극복할 수 있을 뿐만 아니라 새로운 보드 개발의 아이디어를 얻는 데 활용될 수 있을 것으로 판단하였다.

빅데이터 분석으로 사회적 인식을 확인한 선행 연구([1-7])에서는 연구와 관련된 중심 키워드를 인터넷상에서 검색한 후 이 키워드와 관련하여 검색된 방대한 텍스트를 텍스트 마이닝이나 연결 중심성 등의 기법을 이용하여 분석하는 과정을 거쳤다. 분석 결과를 살펴보면 인터넷상에서 중심 키워드와 관련하여 어떤 키워드가 많이 등장하고 있는지(빈도 분석), 실질적으로 중요한 키워드는 어떤 키워드인지(TF-IDF 분석), 다른 키워드와 연결이 많은 키워드는 무엇인지(연결 중심성 분석) 등을 파악하고 있었다. 또한, 빈도수 기준으로 상위에 해당하는 키워드들을 대상으로 어떠한 군집이 형성되는지를 분석하여(CONCOR 분석) 이들 군집에서 다루고 있는 주제는 무엇인지를 파악하고 있었다.

본 연구진에서 개발하고자 하는 제품은 아두이노 기반 보드로 특히 인공지능과 관련된 특정 기능을 수행할 수 있는 보드였다. 여기서 아두이노는 2005년 마시모 벤치 교수팀에 의해 개발된 보드로 각종 센서나 액추에이터를 연결한 후 이를 보드에 대한 지식이 없는 일반인들도 쉽게 이용할 수 있는 저비용 마이크로컨트롤러 보드이다[6]. 한편 인공지능은 1956년 미국의 다펜스 학술회의에서 존 매카시에 의해 사용되기 시작한 용어로 여러 차례의 암흑기와 황금기를 거쳤으며, 오늘날에는 다시 사회 전반에서 널리 활용되며 강조되고 있는 용어이다[8].

연구진은 보드 개발의 핵심 키워드인 ‘아두이노’와 최근 사회 전반에 걸쳐 중요성이 강조되고 있는 키워드인 ‘인공지능’ 등 두 개의 키워드 조합으로 생성된 인터넷상 문서들이 새로운 보드 개발과 관련된 여러 가지 요인을 포함하고 있을 것으로 보고 빅데이터 분석 방법으로 해당 내용을 분석해 보고자 하였다. 빅데이터 분석의 선행 연구 결과와 같이 이들 두 키워드의 조합으로 생성된 인터넷 문서에서 주요 키워드를 파악하고, 어떤 주제를 다루고 있는지를 파악한다면 보드 개발과 관련된 의미 있는 시사점을 얻을 수 있을 것이기 때문이다.

기존의 연구들은 주로 빅데이터 분석을 통해 사회적 인식을 파악하여 인문학적 해석에서 그쳤다면, 본 연구는 새로운 제품의 개발에 빅데이터 분석 결과를 활용하였다는 점에서 차별화되어 있다. 이 연구의 내용은 빅데이터 분석 방법으로 사용자의 요구를 파악하여 제품을 개발하고자 하는 연구자나 개발자들이 참고할 수 있는 자료로 활용될 수 있을 것이다.

## II. Method

### 1. Data collection and refinement

이 연구에서 자료의 수집은 빅데이터 분석 종합 솔루션 사이트인 텍스트롬(<https://www.textom.co.kr>)을 이용하였다. 텍스트롬은 국내외의 주요 포털 검색 사이트의 데이터를 수집하여 다양한 분석을 실시해 주는 웹 기반 솔루션이었다[3]. 이 사이트는 기존 R이나 파이썬을 이용하여 빅데이터를 분석하는 방법 보다 쉽고 편리하게 데이터의 수집 및 정제 작업을 처리해 주기 때문에 최근 많은 연구에서 이 사이트를 이용하여 빅데이터 분석 연구를 수행하고 있다. 이 연구에서는 텍스트롬 사이트가 연구의 목적인 아두이노 및 인공지능과 관련된 빅데이터 분석에 적합하다고 판단하여 해당 사이트를 활용한 자료의 수집 및 정제를 수행하였다.

자료의 수집 절차는 다음과 같다. 먼저 텍스트 사이트에서 검색 대상 포털 사이트를 네이버로 한정하였다. 그 이유는 국내 포털 사이트 중 네이버가 평균 유입률이 가장 높았다는 점과(4) 일반적으로 어떤 주제의 내용이 온라인상 게시되었을 때 해당 내용이 여러 포털 사이트에 동시에 노출되는 특징이 있었기 때문이다[2]. 텍스트에서 네이버 사이트는 블로그, 뉴스, 카페, 지식인 등 섹션의 텍스트를 수집할 수 있어 이들 4개의 섹션을 모두 검색에 활용하였다.

연구자가 아두이노와 인공지능 관련 텍스트 문서를 수집하기 위해 설정한 검색 키워드를 ‘아두이노+인공지능’과 ‘아두이노+AI’ 등 2가지였다. 이는 인공지능을 의미하는 영어 약자 표현인 ‘AI’가 최근 각종 매체에서 자주 활용되고 있기 때문에, ‘인공지능’, ‘AI’ 등의 한글 및 영어 표현을 모두 검색에 활용하였다.

‘아두이노+인공지능’과 ‘아두이노+AI’ 등의 키워드에 대한 자료 수집 기간은 2011년 1월 1일(우리나라에서 아두이노가 도입되어 초기 연구가 발표되기 시작한 시점)부터 본 자료 수집 기간이었던 2022년 7월 30일까지로 설정하였다. 자료의 수집 결과 ‘아두이노+인공지능’ 키워드로는 블로그 3,885건(1.16MB), 뉴스 1,173건(444KB), 카페 2,194건(1.15MB), 지식인 448건(337KB) 등 총 7,700건(3.06MB)이었고, ‘아두이노+AI’ 키워드로는 블로그 3,298건(989KB), 뉴스 882건(330KB), 카페 1,193건(641KB), 지식인 165건(104KB) 등 총 5,538건(2.04MB)이었다.

수집된 자료는 텍스트 사이트에서 정제 작업을 수행하였다. 정제는 여러 차례에 걸쳐 진행되었는데, 먼저 텍스트에서 정제한 결과를 확인하면서 불용어를 삭제하거나 완전한 정제가 이루어지지 않은 키워드를 대치하는 작업을 수행하였다. 예를 들어, 1음절어인 ‘등’, ‘언’, ‘것’, ‘위’ 등의 단어는 삭제하였고 ‘아두이노를’, ‘소프트웨어’ 등의 키워드는 ‘아두이노’, ‘소프트웨어’ 등으로 대치하였다. 또한, 같은 범주에 해당하는 키워드를 통합하는 과정을 거쳤는데 대표적으로 ‘중학생’, ‘초등학생’, ‘고등학생’, ‘아이’ 등은 ‘학생’으로 대치하는 작업을 수행하였다. 그 외 특별한 의미가 없거나 중의적으로 해석될 수 있는 키워드, 이 연구의 목적과 특별히 부합되지 않는 키워드를 삭제하는 작업을 수행하였다. 이러한 키워드로는 ‘다양’, ‘관련’, ‘배우’, ‘사용’, ‘활용’, ‘이용’, ‘진행’, ‘이번’, ‘분야’, ‘가능’ 등 이었다. 이와 같은 키워드의 대치나 삭제 작업은 Notepad ++ 프로그램을 이용하여 수행하였다.

## 2. Data analysis

이 연구에서 자료의 분석은 2단계로 수행하였다. 첫 번째 단계는 정제 작업을 마친 자료를 이용하여 텍스트 마이

닝 분석을 수행하는 단계이다. 텍스트 마이닝은 텍스트 형태의 데이터를 수학적 알고리즘에 따라 처리, 분석, 요약하는 연구 기법을 말하는 것으로, 이때 연구자는 분석 대상에 대한 충분한 지식을 갖춘 후 텍스트의 주제와 연관된 키워드를 중심으로 정제해 가는 작업을 수행한다[2]. 이러한 작업을 통해 인터넷상에 게시되어 있는 대규모의 비정형 텍스트 데이터로부터 연계성을 파악하거나 이들 데이터상에서 숨겨진 의미를 발견할 수 있게 된다[4].

정제된 텍스트를 텍스트 사이트에서 업로드하면 이 사이트에서는 자동으로 해당 텍스트의 빈도 분석, TF-IDF 분석, N-gram 분석 등의 텍스트 마이닝 분석 결과와 키워드 간의 연결 관계에 대한 분석 결과를 제시해 준다. 빈도 분석을 통해서는 정제된 키워드들의 전체 빈도를 표시하여 어떤 키워드가 가장 많이 나타났는지를 확인할 수 있고, TF-IDF 분석을 통해서는 특정 단어가 한 문서 내에서 얼마나 중요한가를 나타내어 실질적 중요성을 갖는 키워드를 파악할 수 있다[7]. N-gram 분석을 통해서는 주제 키워드와 연관 키워드 간의 동시 출현 및 밀집 정도를 분석할 수 있는데, 이를 통해 특정 키워드가 어떤 키워드와 많이 연결되었는지를 확인할 수 있다[9].

텍스트에서는 각 키워드 간의 연결 관계에 대한 분석으로 연결 중심성, 근접 중심성, 매개 중심성, 위세 중심성 등의 분석 결과를 제공해 주지만, 이 연구에서는 연결 중심성 결과만 확인하였다. 이는 연결 중심성 외의 분석은 분석 결과값이 같은 키워드가 여러 개여서 키워드의 우선 순위를 정하기 어려웠기 때문이다. 여기에서 연결 중심성 분석은 온라인상에서 특정 키워드(네트워크 노드)가 얼마나 많이 연결되었는지 확인할 수 있는 통계적 수치이다[5]. 연결이 많은 단어는 연결 중심성이 높기 때문에 영향력이 큰 단어로 판정할 수 있다[10].

이 연구에서는 빈도 분석, TF-IDF 분석, 연결 중심성 분석 등의 결과를 확인하여 중요한 역할을 하는 키워드를 파악하였고, N-gram 분석을 통해 추가로 어떤 키워드들이 연결 관계가 있는지를 파악하였다.

두 번째 단계는 의미 연결망 분석을 수행하였다. 이 분석 방법은 각 키워드를 의미하는 노드(node)와 키워드 간의 연결을 의미하는 링크(link)를 이용하여 각 노드를 구성하는 키워드 간의 연결 관계에 중점을 두는 분석 방법이다. 추출한 키워드의 빈도와 중요성에 따라 이들 키워드를 공간적으로 시각화하여 나타내어 주요 키워드의 전체적 의미를 도출하거나 노드와 링크의 강도 및 특성을 분석하여 강조되는 의미를 찾는 데 활용된다[10].

이 연구에서는 빈도 기준 상위 50개의 키워드를 대상으로 의미 연결망 분석을 수행하였으며, 이를 위해 텍스트 사이트

에서 이들 50개 키워드를 1-mode 매트릭스 데이터로 변환하였다. 이를 Ucinet의 NetDraw 프로그램을 이용하여 전체 키워드 간의 연결 관계를 확인하였고, CONCOR(Convergence of interated correlations) 분석을 수행하여 유사점을 가지는 단어들의 군집을 확인하였다.

### III. Results

#### 1. Results of big data analysis of ‘Arduino + artificial intelligence’

##### 1.1 Results of frequency analysis

‘아두이노’와 ‘인공지능’을 포함하고 있는 인터넷 문서의 주요 키워드에 대한 빈도 분석 결과는 Table 1과 같다. 분석 결과 1~10위 키워드는 인공지능(7,511), 아두이노(7,175), 코딩(2,905), 학원(2,151), 교육(1,839), 학생(1,213), 로봇(1,137), 사물인터넷(1,017), 프로그래밍(907), 스크래치(871) 등으로 나타났다. 11~20위의 키워드는 프로그램(855), 빅데이터(756), 기술(753), 파이썬(742), 프린터(675), 개발(660), 수업(617), 엔트리(615), 컴퓨터(608), 과정(581) 등으로 나타났다.

Table 1. Results of frequency analysis

No.	Keyword	N	No.	Keyword	N
1	artificial intelligence	7,511	26	industrial revolution	491
2	Arduino	7,175	27	project	454
3	coding	2,905	28	produce	438
4	private institute	2,151	29	developer	407
5	education	1,839	30	board	390
6	student	1,213	31	experience	363
7	robot	1,137	32	study	350
8	Internet of Things	1,017	33	practice	344
9	programming	907	34	learning	344
10	Scratch	871	35	grade	342
11	program	855	36	fusion	342
12	big data	756	37	curriculum	342
13	technology	753	38	kit	330
14	Python	742	39	Java	321
15	printer	675	40	future	318
16	development	660	41	activity	317
17	class	617	42	start	316
18	Entry	615	43	coding education	315
19	computer	608	44	language	312
20	course	581	45	modelling	306
21	software	572	46	management	301
22	maker	570	47	need	298
23	sensor	500	48	connection	293
24	Raspberry Pi	494	49	data	289
25	foundation	491	50	contents	283

##### 1.2 Results of TF-IDF analysis

‘아두이노’와 ‘인공지능’을 포함하고 있는 인터넷 문서의 주요 키워드에 대한 TF-IDF 분석 결과는 Table 2와 같다. 분석 결과 1~10위 키워드는 학원(4,581.74), 코딩(3,871.47), 교육(3,191.05), 학생(2,524.25), 로봇(2,431.85), 사물인터넷(2,170.72), 프로그래밍(2,136.83), 프로그램(2,017.84), 스크래치(1,945.46), 기술(1,877.81) 등으로 나타났다. 11~20위의 키워드는 프린터(1,832.33), 파이썬(1,827.41), 빅데이터(1,795.72), 개발(1,740.58), 과정(1,679.00), 컴퓨터(1,657.65), 메이커(1,651.23), 수업(1,649.99), 엔트리(1,624.14), 소프트웨어(1,565.2) 등으로 나타났다.

Table 2. Results of TF-IDF analysis

No.	Keyword	TF-IDF	No.	Keyword	TF-IDF
1	private institute	4,581.74	26	artificial intelligence	1,354.99
2	coding	3,871.47	27	project	1,346.13
3	education	3,191.05	28	produce	1,302.01
4	student	2,524.25	29	developer	1,295.80
5	robot	2,431.85	30	board	1,219.11
6	Internet of Things	2,170.72	31	experience	1,213.57
7	programming	2,136.83	32	grade	1,149.71
8	program	2,017.84	33	study	1,144.08
9	Scratch	1,945.46	34	practice	1,105.07
10	technology	1,877.81	35	kit	1,102.04
11	printer	1,832.33	36	fusion	1,095.35
12	Python	1,827.41	37	learning	1,093.07
13	big data	1,795.72	38	language	1,088.69
14	development	1,740.58	39	curriculum	1,086.72
15	course	1,679.00	40	activity	1,066.86
16	computer	1,657.65	41	coding education	1,039.50
17	maker	1,651.23	42	future	1,036.24
18	class	1,649.99	43	Academy	1,034.65
19	Entry	1,624.14	44	Java	1,034.30
20	software	1,565.20	45	start	1,031.87
21	Arduino	1,451.21	46	management	1,024.54
22	sensor	1,436.81	47	modelling	1,004.46
23	foundation	1,434.06	48	data	990.55
24	industrial revolution	1,418.92	49	need	986.57
25	Raspberry Pi	1,380.29	50	connection	976.34

##### 1.3 Results of degree centrality analysis

‘아두이노’와 ‘인공지능’을 포함하고 있는 인터넷 문서의 주요 키워드에 대한 연결 중심성 분석 결과는 Table 3과 같다. 분석 결과 1~10위 키워드는 인공지능(381.57), 아두이노(367.82), 코딩(185.68), 학원(59.61), 교육(117.95), 학생(79.13), 로봇(69.92), 사물인터넷(66.18), 프로그래밍(67.84), 스크래치(56.64) 등으로 나타났다.



같으며, 분석 결과 3개의 군집이 형성됨을 확인할 수 있었다. 첫 번째 군집은 실습, 체험, 센서, 연결, 제작, 프로젝트, 활동, 학습, 개발 등의 키워드들로 구성되어 있었다. 이들은 주로 아두이노를 이용하여 인공지능과 관련된 기능을 구현하는 것과 관련이 있어 이 군집을 ‘제작 및 체험 관련’으로 명명하였다. 두 번째 군집은 스크래치, 엔트리, 사물인터넷, 코딩, 코딩 교육, 수업 등으로 구성되어 있었다. 이들은 코딩 교육과 관련된 키워드들이었기 때문에 ‘코딩 교육’으로 명명하였다. 세 번째 군집은 라즈베리파이, 빅데이터, 파이선, 학생 등으로 구성되어 있었다. 이들은 인공지능 교육을 수행할 수 있는 기기나 관련 소프트웨어였기 때문에 ‘인공지능 교육 관련’으로 명명하였다. 각 군집에 포함된 키워드들의 목록은 Table 5와 같다.

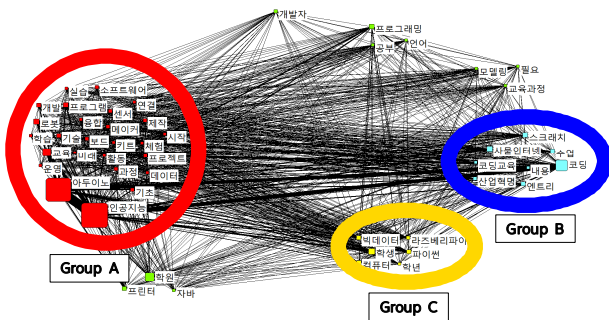


Fig. 2. Result of CONCOR analysis

Table 5. Results of CONCOR analysis

Group	number	Keywords
A	26	activity, Arduino, artificial intelligence, board, connection, course, data, development, education, experience, foundation, fusion, future, kit, learning, maker, management, practice, produce, program, project, robot, sensor, software, start, technology
B	8	class, coding, coding education, contents, Entry, industrial revolution, Internet of Things, Scratch
C	6	big data, computer, grade, Python, Raspberry Pi, student
Etc.	10	curriculum, developer, Java, language, modelling, need, printer, private institute, programming, study

- Group A : production and experience related
- Group B : coding education
- Group C : artificial intelligence education related

## 2. Results of big data analysis of ‘Arduino + AI’

### 2.1 Results of frequency analysis

‘아두이노’와 ‘AI’를 포함하고 있는 인터넷 문서의 주요 키워드에 대한 빈도 분석 결과는 Table 6과 같다. 분석 결

과 1~10위 키워드는 아두이노(5,339), 인공지능(2,175), 코딩(1,845), 교육(1,562), 로봇(736), 학생(713), 스크래치(590), 프로그램(565), 개발(549), 기술(549) 등으로 나타났다. 11위~20위의 키워드는 메이커(510), 프로그래밍(458), 엔트리(454), 파이선(440), 센서(431), 사물인터넷(426), 수업(409), 빅데이터(403), 라즈베리파이(388), 과정(378) 등으로 나타났다.

Table 6. Results of frequency analysis

No.	Keyword	N	No.	Keyword	N
1	Arduino	5,339	26	industrial revolution	292
2	artificial intelligence	2,175	27	fusion	286
3	coding	1,845	28	connect	281
4	education	1,562	29	computer	278
5	robot	736	30	learning	272
6	student	713	31	board	270
7	Scratch	590	32	printer	266
8	program	565	33	smart	266
9	development	549	34	management	257
10	technology	549	35	future	241
11	maker	510	36	contents	225
12	programming	458	37	hand-on activity	222
13	Entry	454	38	activity	209
14	Python	440	39	time	207
15	sensor	431	40	printing	205
16	internet of things	426	41	autonomous driving	201
17	class	409	42	foundation	200
18	big data	403	43	start	199
19	Raspberry Pi	388	44	data	197
20	course	378	45	language	196
21	project	350	46	certificate	195
22	produce	345	47	car	195
23	kit	342	48	hardware	190
24	software	326	49	coding education	189
25	experience	292	50	developer	187

### 2.2 Results of TF-IDF analysis

‘아두이노’와 ‘인공지능’을 포함하고 있는 인터넷 문서의 주요 키워드에 대한 TF-IDF 분석 결과는 Table 7과 같다. 분석 결과 1~10위 키워드는 코딩(2,593.04), 교육(2,521.58), 인공지능(2,502.76), 로봇(1,669.49), 학생(1,588.06), 프로그램(1,397.18), 메이커(1,366.31), 개발(1,365.90), 스크래치(1,349.75), 기술(1,333.46) 등으로 나타났다. 11~20위의 키워드는 프로그래밍(1,219.82), 엔트리(1,213.90), 센서(1,181.57), 파이선(1,169.61), 과정(1,137.68), 수업(1,137.62), 사물인터넷(1,130.20), 라즈베리파이(1,074.71), 빅데이터(1,074.38), 아두이노(1,036.46) 등으로 나타났다.

Table 7. Results of TF-IDF analysis

No.	Keyword	TF-IDF	No.	Keyword	TF-IDF
1	coding education	2,593.04	26	fusion	896.44
2	education	2,521.58	27	industrial revolution	894.21
3	artificial intelligence	2,502.76	28	computer	889.23
4	robot	1,669.49	29	connect	875.00
5	student	1,588.06	30	board	872.14
6	program	1,397.18	31	printer	867.86
7	maker	1,366.31	32	smart	843.88
8	development	1,365.90	33	learning	840.42
9	Scratch	1,349.75	34	management	828.97
10	technology	1,333.46	35	future	777.36
11	programming	1,219.82	36	practice	744.00
12	Entry	1,213.90	37	contents	730.93
13	sensor	1,181.57	38	autonomous driving	722.66
14	Python	1,169.61	39	experience	718.72
15	course	1,137.68	40	car	707.61
16	class	1,137.62	41	activity	705.89
17	internet of things	1,130.20	42	data	703.16
18	Raspberry Pi	1,074.71	43	time	702.45
19	big data	1,074.38	44	printing	694.56
20	Arduino	1,036.46	45	language	689.79
21	kit	1,030.58	46	foundation	689.74
22	project	1,007.80	47	start	686.30
23	produce	1,004.67	48	developer	682.44
24	software	961.48	49	hardware	671.19
25	experience	956.85	50	certificate	659.64

2.3 Results of degree centrality analysis

‘아두이노’와 ‘인공지능’을 포함하고 있는 인터넷 문서의 주요 키워드에 대한 연결 중심성 분석 결과는 Table 8과 같다. 분석 결과 1~10위 키워드는 아두이노(210.48), 인공지능(110.83), 코딩(99.89), 교육(81.89), 학생(37.52), 로봇(35.34), 스크래치(33.20), 프로그램(32.09), 개발(28.58), 메이커(28.46) 등으로 나타났다. 11~20위의 키워드는 기술(26.92), 프로그래밍(25.61), 엔트리(25.15), 파이썬(24.73), 사물인터넷(24.55), 과정(23.16), 빅데이터(22.24), 센서(19.55), 소프트웨어(19.25), 수업(18.89) 등으로 나타났다.

Table 8. Results of degree centrality analysis

No.	Keyword	Degree centrality	No.	Keyword	Degree centrality
1	Arduino	210.48	26	Raspberry Pi	16.64
2	artificial intelligence	110.83	27	fusion	16.06
3	coding	99.89	28	kit	15.73
4	education	81.89	29	computer	15.66
5	student	37.52	30	management	14.36
6	robot	35.34	31	printer	13.17
7	Scratch	33.20	32	contents	12.95
8	program	32.09	33	foundation	12.91
9	development	28.58	34	smart	12.89
10	maker	28.46	35	connect	12.09
11	technology	26.92	36	future	12.04
12	programming	25.61	37	language	11.98
13	Entry	25.15	38	certificate	11.76
14	Python	24.73	39	board	11.43
15	internet of things	24.55	40	autonomous driving	11.38
16	course	23.16	41	developer	11.34
17	big data	22.24	42	activity	11.29
18	sensor	19.55	43	printing	11.23
19	software	19.25	44	car	11.11
20	class	18.89	45	practice	11.00
21	produce	18.74	46	start	10.68
22	project	17.46	47	coding education	9.78
23	experience	16.93	48	hardware	9.58
24	industrial revolution	16.76	49	data	9.26
25	learning	16.69	50	time	8.71

2.4 Results of N-gram analysis

‘아두이노’와 ‘인공지능’을 포함하고 있는 인터넷 문서의 주요 키워드에 대한 N-gram 분석 결과는 Table 9과 같다. 분석 결과 1~5위 키워드 조합은 아두이노-코딩(409), 스크래치-아두이노(215), 아두이노-라즈베리파이(180), 코딩-교육(156), 아두이노-스크래치(138) 등으로 나타났으며, 6~10위 키워드 조합은 코딩-아두이노(129), 엔트리-스크래치(124), 아두이노-센서(115), 아두이노-보드(107), 인공지능-아두이노(100) 등으로 나타났다.

Table 9. Results of N-gram analysis

No.	Keyword1	Keyword2	N
1	Arduino	coding	409
2	Scratch	Arduino	215
3	Arduino	Raspberry Pi	180
4	coding	education	156
5	Arduino	Scratch	138
6	coding	Arduino	129
7	Entry	Scratch	124
8	Arduino	sensor	115
9	Arduino	board	107
10	artificial intelligence	Arduino	100
11	artificial intelligence	internet of things	99
12	coding	principle	97
13	maker	Arduino	96
14	contents	artificial intelligence	96
15	artificial intelligence	Powerbi	95

2.5 Results of CONCOR analysis

‘아두이노’와 ‘AI’를 포함하고 있는 인터넷 문서의 주요 키워드 중 빈도수 기준 상위 50개의 키워드는 UCINET을 활용하여 매트릭스 분석한 후 이를 Netdraw로 네트워크 시각화한 결과와 CONCOR 분석한 결과는 Figure 3, Figure 4이다.

Figure 3의 네트워크 시각화 결과에서 확인할 수 있듯이 50개의 키워드는 네트워크 모양의 거미줄 형으로 나타나 각 키워드들 간의 활발한 상호 작용이 있음을 확인할 수 있었다. 네트워크의 중심에는 데이터, 교육, 기초, 시작, 사물인터넷, 언어 등의 키워드가 위치하였고, 이 연구의 검색이 연구의 검색 키워드 중 하나인 아두이노는 중심에서 약간 왼쪽 부분에 위치하였고, 다른 검색 키워드인 인공지능은 네트워크의 바깥쪽에 위치하고 있었다. 각 키워드의 연결 라인의 굵기는 연결성의 정도를 의미하는데, 아두이노, 인공지능, 코딩, 교육 등의 키워드들이 더욱 강한 연결 관계가 있음을 확인할 수 있었다.

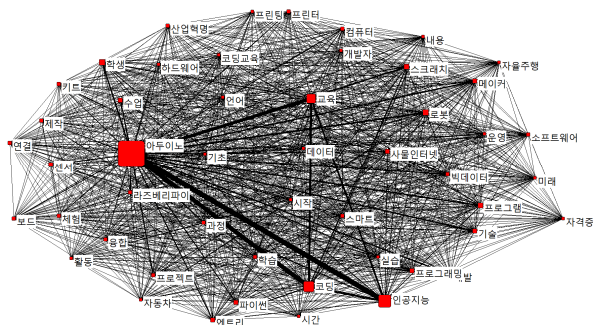


Fig. 3. Result of network analysis

‘아두이노’와 ‘AI’를 포함하고 있는 인터넷 문서의 주요 키워드에 대한 CONCOR 분석 결과는 Figure 4와 같으며, 분석 결과 4개의 군집이 형성됨을 확인할 수 있었다. 첫 번째 군집은 수업, 운영, 활동, 프로그램, 프로젝트 등의 키워드들로 구성되어 있었다. 이들은 주로 아두이노 및 인공지능과 관련된 교육 프로그램과 관련되어 있어 이 군집을 ‘교육 프로그램’으로 명명하였다. 두 번째 군집은 학생, 학습, 교육, 코딩, 코딩 교육 등으로 구성되어 있었다. 이들은 주로 교육과 관련된 키워드들로 코딩과 관련이 있었기 때문에 ‘코딩 교육’으로 명명하였다. 세 번째 군집은 개발, 개발자, 보드, 센서, 연결 등으로 구성되어 있었다. 이들은 아두이노 및 인공지능과 관련된 제품의 개발과 관련이 있어 ‘관련 제품 개발’로 명명하였다. 네 번째 군집은 스크래치, 엔트리, 언어로 구성되어 있었다. 이들은 프로그래밍 언어와 관련이 있어 ‘프로그래밍 언어’로 명명하였다. 각 군집에 포함되어 있는 키워드들의 목록은 Table 10과 같다.

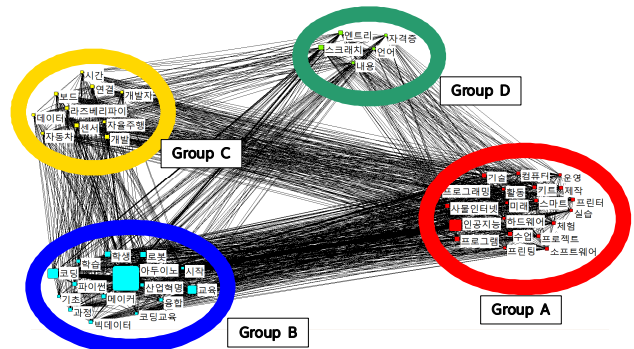


Fig. 4. Result of CONCOR analysis

Table 10. Results of CONCOR analysis

Group	number	Keywords
A	20	activity, artificial intelligence, class, computer, experience, future, hardware, internet of things, kit, management, practice, printer, printing, produce, program, programming, project, smart, software, technology, educational program, coding education
B	15	Arduino, big data, coding, coding education, course, education, foundation, fusion, industrial revolution, learning, maker, Python, robot, start, student
C	10	autonomous driving, board, car, connect, data, developer, development, Raspberry Pi, sensor, time
D	5	certificate, contents, Entry, language, Scratch

- Group A : educational program
- Group B : coding education
- Group C : related product development
- Group D : programming language

#### IV. Implications

이 연구의 결과에 따른 인공지능 및 아두이노 기반 보드 개발에 활용할 수 있는 시사점을 제시하면 다음과 같다.

첫째, 아두이노 및 인공지능, 아두이노 및 AI 관련 인터넷 문서들은 공통으로 교육 활동과 관련된 내용을 다루고 있었다. 이 연구를 통해 ‘아두이노+인공지능’ 키워드를 대상으로 한 빈도 분석 결과 4위 학원(2,151), 5위 교육(1,839), 6위 학생(1,213) 등 이었고, TF-IDF 분석 결과 1위 학원(4,581.74), 3위 교육(3,191.05), 4위 학생(2,524.25) 등이었으며, 연결 중심성 분석 결과 4위 학원(59.61), 5위 교육(117.95), 6위 학생(79.13) 등으로 교육 관련 키워드들이 상위권을 차지하고 있음을 확인할 수 있었다.

‘아두이노+AI’ 키워드를 대상으로 한 빈도 분석 결과에서는 4위 교육(1,562), 6위 학생(713) 등이었으며, TF-IDF 분석 결과 2위 교육(2,521.58), 4위 학생(1,588.06) 등 이었고, 연결 중심성 분석 결과 4위 교육(81.89), 5위 학생(37.52) 등으로 나타났다.

또한, 빈도수 기준 상위 50위 이내의 키워드를 대상으로 CONCOR 분석을 수행한 결과 ‘아두이노+인공지능’ 키워드를 한 인터넷 문서에서는 ‘코딩 교육’, ‘인공지능 교육 관련’ 군집을 확인할 수 있었고, ‘아두이노+AI’ 키워드를 한 인터넷 문서들에서는 ‘교육 프로그램’, ‘코딩 교육’ 군집을 확인할 수 있었다.

이러한 결과들은 ‘아두이노+인공지능’, ‘아두이노+AI’를 포함하고 있는 인터넷 문서들이 교육 활동과 관련된 내용을 다루고 있음을 의미하였다. 이는 ‘아두이노+인공지능’, ‘아두이노+AI’ 등의 키워드를 작성한 사용자들이 교육 활동과 관련된 내용을 많이 작성한 것으로 해석할 수 있었다. 이를 기반으로 인공지능과 관련된 아두이노용 보드의 개발에 대한 방향성으로 교육 분야에서 활용될 수 있는 특화된 보드의 개발을 설정할 수 있었다.

둘째, 아두이노 및 인공지능, 아두이노 및 AI 관련 인터넷 문서들은 공통으로 코딩 교육 및 프로그래밍 교육과 관련된 내용을 다루고 있었다. 이 연구를 통해 ‘아두이노+인공지능’ 키워드를 대상으로 한 빈도 분석 결과 3위 코딩(2,905), 5위 교육(1,839), 9위 프로그래밍(907) 이었고, TF-IDF 분석 결과 2위 코딩(3,871.41), 3위 교육(3,191.05), 7위 프로그래밍(2,136.83)이었으며, 연결 중심성 분석 결과 3위 코딩(185.69), 5위 교육(117.95), 9위 프로그래밍(67.84) 등을 확인할 수 있었다.

‘아두이노+AI’ 키워드를 대상으로 한 빈도 분석 결과에

서는 3위 코딩(1,845), 4위 교육(1,562), 12위 프로그래밍(458) 등 이었고, TF-IDF 분석 결과 1위 코딩(2,593.04), 2위 교육(2,521.58), 11위 프로그래밍(1,219.82) 등 이었으며, 연결 중심성 분석 결과 3위 코딩(99.89), 4위 교육(81.89), 12위 프로그래밍(25.61) 등으로 나타났다.

또한, ‘아두이노+인공지능’ 키워드를 대상으로 한 N-gram 분석 결과 10위 코딩-교육(196), ‘아두이노+AI’ 키워드를 대상으로 한 N-gram 분석 결과 4위 코딩-교육(156)을 확인할 수 있었고, ‘아두이노+인공지능’ 및 ‘아두이노+AI’의 CONCOR 분석 모두에서 ‘코딩 교육’ 군집을 확인할 수 있었다.

코딩 교육을 달리 프로그래밍 교육으로 표현하고 있으므로, 이들 결과를 통해 아두이노와 인공지능, 아두이노와 AI 관련 인터넷 문서에서 코딩 교육과 관련된 내용이 다루어지고 있음을 확인할 수 있었다. 이를 통해 인공지능 관련 아두이노 기반 보드의 개발 방향으로 최근 교육 현장에서 강조되고 있는 코딩 교육을 원활히 수행할 수 있는 보드의 개발로 설정할 수 있었다.

셋째, 아두이노와 인공지능, 아두이노와 AI 관련 인터넷 문서들은 공통으로 아두이노를 기반으로 한 제작 및 체험, 개발 활동과 관련된 내용을 다루고 있었다. 이 연구를 통해 ‘아두이노+인공지능’ 키워드를 대상으로 한 빈도 분석 결과 16위 개발(660), 27위 프로젝트(454), 28위 제작(438) 등 이었고, TF-IDF 분석 결과 14위 개발(1,740.58), 17위 메이커(1,651.23), 28위 제작(1,302.01) 등이었으며, 연결 중심성 분석 결과 16위 개발(42.51), 22위 메이커(37.74), 28위 제작(25.39) 등을 확인할 수 있었다.

‘아두이노+AI’ 키워드를 대상으로 한 빈도 분석 결과에서는 9위 개발(549), 11위 메이커(510), 22위 제작(345) 등 이었고, TF-IDF 분석 결과 7위 메이커(1,366.31), 8위 개발(1,365.90), 23위 제작(1,004.67) 등이었으며, 연결 중심성 분석 결과 9위 개발(28.58), 10위 메이커(28.46), 21위 제작(18.74) 등으로 나타났다.

또한, ‘아두이노+인공지능’의 CONCOR 분석에서는 ‘관련 제품 개발’ 군집을 확인할 수 있었고, ‘아두이노+AI’의 CONCOR 분석에서는 ‘제작 및 체험 관련’ 군집을 확인할 수 있었다.

이들 결과를 통해 아두이노와 인공지능, 아두이노와 AI 관련 인터넷 문서에서 아두이노를 기반으로 한 제작이나 개발과 관련된 내용이 다루어지고 있음을 확인할 수 있었다. 이러한 결과를 첫 번째 얻은 시사점과 연관을 짓는다면 이러한 제작이나 개발은 교육 활동과 연관된 것으로 해석할 수 있었다. 이를 통해 인공지능 관련 아두이노 기반

보드의 개발 방향을 교육 현장에서 학생들의 제작이나 개발 활동에서 더 쉽고 편리하게 적용될 수 있는 보드의 개발로 설정할 수 있었다.

넷째, 아두이노와 인공지능, 아두이노와 AI 관련 인터넷 문서를 통해 아두이노에서 인공지능 관련 기능의 수행에 필요한 프로그램을 확인할 수 있었다. ‘아두이노+인공지능’ 키워드를 대상으로 한 빈도 분석 결과 10위 스크래치(871), 14위 파이선(742), 18위 엔트리(615) 등 이었고, TF-IDF 분석 결과 9위 스크래치(1,945.46), 12위 파이선(1,827.41), 19위 엔트리(1,624.14) 등 이었으며, 연결 중심성 분석 결과 10위 스크래치(56.64), 14위 파이선(50.03), 18위 엔트리(37.90) 등을 확인할 수 있었다.

‘아두이노+AI’ 키워드를 대상으로 한 빈도 분석 결과 7위 스크래치(590), 13위 엔트리(454), 14위 파이선(440) 등 이었고, TF-IDF 분석 결과 9위 스크래치(1,349.75), 12위 엔트리(1,213.90), 14위 파이선(1,169.61) 등 이었으며, 연결 중심성 분석 결과 7위 스크래치(33.20), 13위 엔트리(25.15), 14위 파이선(24.73) 등을 확인할 수 있었다.

이들 결과를 통해 아두이노와 인공지능, 아두이노와 AI 관련 인터넷 문서에서 스크래치, 엔트리, 파이선 등의 프로그램이 자주 등장하고 있음을 확인할 수 있었는데, 실제 이들 프로그램은 아두이노에서 인공지능 기능을 구현할 수 있는 대표적인 프로그램이다. 따라서 보드 개발 시 이들 프로그램에서 쉽게 연동될 수 있는 보드의 개발이 필요함을 확인할 수 있었다.

다섯째, 아두이노와 인공지능 관련 인터넷 문서들에서 ‘학원’ 키워드가 중요한 역할을 하고 있음을 확인할 수 있었다. ‘아두이노+인공지능’에 대한 문서의 분석을 통해 확인한 연구 결과에서 학원은 빈도 분석 결과 4위(2,151)였고, TF-IDF 분석 결과 1위(4,531.74), 연결 중심성 분석 결과 4위(59.61) 이었다. 여기에서 TF-IDF는 실제 중요한 역할을 하는 키워드를 확인할 수 있는 수치인데, 이 수치상 1위가 학원이었다. 또한, N-gram 분석 결과에서도 ‘학원-학원’은 1위(888)로 나타난 것으로 볼 때 이 키워드가 중요한 역할을 하고 있음을 확인할 수 있었다.

이러한 결과가 나타난 한 가지 배경으로는 최근 소프트웨어 교육이나 코딩 교육의 강조로 인해 사교육 시장이 반응한 것을 들 수 있다. 이 연구에서 수집되었던 원본의 문서들을 확인한 결과 작성의 주체가 컴퓨터 학원인 경우가 많았는데, 이들은 아두이노, 인공지능 등의 키워드를 포함하면서 학원 교육과정에 대한 소개나 홍보 내용을 작성하고 있었다. 이처럼 여러 학원에서 인터넷상 많은 기록을 남겼기 때문에 이 키워드가 주요 키워드가 된 것으로 유추

할 수 있었다.

이를 보드 개발의 관점에서 해석한다면, 인공지능 관련 아두이노 기반 보드의 개발 방향 중 하나로 학원에서 사용할 수 있는 교육용 보드의 개발을 고려해 볼 수 있었다. 이는 빅데이터 분석상 아두이노 및 인공지능 관련 내용의 작성성이 학원과 어느 정도 연관되어 있었기 때문으로, 이들 학원은 개발된 보드의 새로운 수요처가 될 수 있기 때문이다.

## V. Conclusion

이 연구의 목적은 빅데이터 분석을 통해 파악한 사회적 인식을 활용하여 인공지능과 관련된 아두이노 기반 보드의 개발 방향성을 확인하는 데 있다.

이 연구의 결과 아두이노와 인공지능, 아두이노와 AI 관련 인터넷 문서들은 공통으로 교육 활동과 관련된 내용을 다루고 있었고, 특히 이들은 코딩 교육과 관련이 있음을 확인할 수 있었다. 또한, 이들 문서에서는 아두이노를 기반으로 한 제작 및 체험, 개발 활동과 관련된 내용을 다루고 있음도 확인할 수 있었다. 이 연구를 통해 스크래치, 엔트리, 파이선 등의 프로그램이 아두이노와 인공지능, 아두이노와 AI 관련 문서에 자주 등장하고 있음도 확인할 수 있었다. 마지막으로 아두이노와 인공지능 관련 인터넷 문서들에서는 학원이 가장 중요한 역할을 하고 있음을 확인하였다.

이 연구를 통해 아두이노와 인공지능, 아두이노와 AI 등의 키워드들을 포함하고 있는 인터넷 문서를 통해 사회적 인식을 파악하였고, 이를 기반으로 한 보드 개발의 방향성을 확인할 수 있었다. 이 연구는 일반 대중의 사회적 인식을 빅데이터 분석 방법을 활용하여 파악한 후, 이를 기반으로 보드 개발의 여러 가지 요인들을 확인하였다는 점에서 의의가 있다. 이 연구는 빅데이터 분석 방법으로 사용자의 요구를 파악하여 제품을 개발하고자 하는 연구자나 개발자들이 참고할 수 있는 자료로 활용될 수 있을 것이다.

## ACKNOWLEDGEMENT

This work was supported by the research grant of Kongju National University in 2022. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1F1A104755011).

## REFERENCES

- [1] S. Kang, and y. Lee, "A study on social perception of childrens smart media education based on big data," *The Journal of Korea Open Association for Early Childhood Education*, Vol.22, No.4, pp.45-72, August 2017. <http://dx.doi.org/10.20437/KOAECE22-4-03>
- [2] J. Koh, and Y. Chong, "Analysis of Taiwanese Food Trends in Korea Using Social Big Data: Focusing on blog text analysis," *International Journal of Tourism Management and Sciences*, Vol.24, No.3, pp.71-91, June 2020. <http://dx.doi.org/10.18604/tmro.2020.24.3.4>
- [3] Y. Koo, "Trend Analysis on Clothing Care System of Consumer from Big Data," *Fashion & Textile Research Journal*, Vol.22, No.5, pp.639-649, October 2020. <https://doi.org/10.5805/SFTI.2020.22.5.639>
- [4] K. Kim, and E. Kim, "A study on social perception of forest education for young children through social network analysis based on big data," *The Journal of Korea Open Association for Early Childhood Education*, Vol.38, No.5, pp.287-308, October 2018. <http://dx.doi.org/10.18023/kjece.2018.38.5.012>
- [5] Y. Moon, "An Analysis of Trends in Researches on the Open Recruitment System for Principals based on Topic Modeling and Keyword Network Analysis," *Journal of Education & Culture*, Vol.26, No.1, pp.217-242, October 2020. <https://doi.org/10.24159/joec.2020.26.1.217>
- [6] E. Lee, "Social perception of the Arduino lecture as seen in big data," *Journal of The Korean Association of Information Education*, Vol.25, No.6, pp.935-945, December 2021. Dec. <http://dx.doi.org/10.14352/jkaie.2021.25.6.935>
- [7] J. Choi, and S. Park, "A Study on Perception of Golf Lesson Using Big Data Analysis," *Journal of Golf Studies* Vol.14, No.1, pp.151-163, March 2020. <https://doi.org/10.34283/ksgs.2020.14.1.13>
- [8] S. Kim, and T. Kim, "A Study on Educational Dataset Standards for K-12 Artificial Intelligence Education," *The Journal of Korean Association of Computer Education*, Vol.25, No.1, pp.29-40, 2022. February <https://doi.org/10.32431/kace.2022.25.2.003>
- [9] T. Kim, and S.W. Kim, "Social Tendency and Network Analysis of High School Credit System," *Journal of Educational Innovation Research*, Vol.29, No.2, pp.225-242, June 2019. DOI: <http://dx.doi.org/10.21024/pnuedi.29.2.201906.225>
- [10] W. Kang, E.S. Ko, H.R. Lee, and J. Ki, "A Study of the Consumer Major Perception of Packaging Using Big Data Analysis - Focusing on Text Mining and Semantic Network Analysis," *Journal of the Korea Convergence Society*, Vol.9, No.4, pp.15-22, April 2018. <https://doi.org/10.15207/JKCS.2018.9.4.015>

## Authors



Eun-Sang Lee received the B.S. degrees in Technology Education from Korea National University of Education, Korea, M.S. and Ph.D. degrees in Technology Education from Chungnam National University, Korea, in

2003, 2013 and 2015, respectively. Dr. Lee joined the faculty of the Department of Technology and Home Economics Education at Kongju National University, Chungcheongnam-do, Korea, in 2017. He is interested in low-cost microcontroller, technology education, invention education and big data analysis.