

텍스트 마이닝(Text Mining)에 의한 표준과 표준교육의 빅데이터 분석

고인규*

이 연구의 목적은 빅데이터 기법을 활용하여 표준 관련 연구 동향을 분석하는 것이었다. 연구의 목적을 달성하기 위해 2017년 5월 1일부터 12월 30일까지 리스포유에서 '표준', '표준교육' 등의 키워드 검색을 통해 도출된 총 441편의 학술 자료를 대상으로 빅데이터 분석하였다. 이 연구에서 사용된 빅데이터 분석 방법은 텍스트 마이닝(Text Mining)이었다. 분석을 위해 사용된 프로그램은 R 프로그램이다. R 프로그램의 'tm', 'RColorBrewer', 'wordcloud', 'qcc', 'KoNLP' 패키지를 이용하여 불용어와 기호를 삭제하고 자연어 처리기술 중 형태소 분석을 통해 문장을 최소단위로 쪼개는 작업을 진행하였다. 이를 통해 학술자료 제목에 주로 나타나는 단어의 빈도수를 파악하고 이 결과를 R 프로그램의 크라우드링 패키지를 활용하여 워드크라우드로 제시하는 것으로 분석을 실시하였다. 이 연구에서 도출된 결과는 다음과 같다. 먼저 1시기(2007~21011) 사이에 가장 많이 언급된 단어는 '표준'이었다. 표준이외에 빈도수가 높은 단어는 '교육과정', '교육', '과정', '보육'이었다. 2시기(2012~2014) 사이에 가장 많이 언급된 단어는 '교육'이었다. 다음으로 '표준', '발명', '국가직무능력', '역량' 등이 상위에 있었다. 3시기(2015~2017)에 가장 많이 언급된 단어는 '과정'이었다. 다음으로 '발명', '국가직무능력', '역량' 등이 상위에 있었다. 표준 관련 연구는 1시기, 2시기, 3시기에 걸쳐 계단식으로 양적 성장을 보였으며 핵심적인 키워드는 '교육과정', '교육', '과정', '보육', '발명', '국가직무능력', '역량'이었다.

주요어: 표준, 표준교육, 빅데이터 분석, 역량, R 프로그램

1. 서론

정보통신 기술의 일상화가 이루어지면서 데이터가 정보화 사회의 새로운 중요한 기술로 등장하고 있다. 이는 과거 하드웨어에서 소프트웨어, 그리고 데이터로 관심 영역이 점차 확대되고 있다는 것을 의미한다(조영임, 2013). 이러한 데이터로 관심 변화

중심에는 빅데이터가 포함되어 있어서 많은 이슈가 되고 있다.

빅데이터 기술의 발전은 다변화된 현대 사회를 보다 정확하게 예측하고 효율적으로 작동하도록 정보를 제공하며 개인화된 현대 사회 구성원들에게 있어서 맞춤형 정보를 제공, 관리, 분석 가능케 하며 과거에는 불가능했던 기술을 진일보시킨다(김동완, 2013). 또한 사회에서 발생하는 여러 현상에 대한 데이터

* 세종 양지초등학교 (goingyu@hanmail.net), 주저자

분석을 통해 그 원인을 밝혀 대비하게 하며, 과거에는 상상도 할 수 없었던 인과관계도 밝혀냄으로써, 기업경영에 큰 도움을 주고 예기치 않은 사고를 예방할 수도 있다(이중기, 2015).

최근에는 교육 분야에서도 빅데이터의 활용 연구(최제영 외, 2012; 장산현, 2012; 최제영, 박충식, 2013; 권용욱, 2013; 이은경 외, 2014; 노규성, 2015)가 활발히 이루어지고 있다. 교육 정책의 수립, 연구 동향 분석, 교육 방법을 위한 기초 자료 수집에 빅데이터가 활용되고 있는 것이다(고인규, 2016). OECD 보고서에 따르면 세계 교역량의 80%가 표준의 영향을 받고 있으며, 국가별로 상이한 표준은 상품 및 서비스의 자유로운 이동을 저해하고 시장 통합의 걸림돌이 된다고 밝히고 있다.

표준은 산업뿐만 아니라 우리 생활에서도 매우 커다란 영향을 미치고 있다. 우리는 안전하고 편리한 생활을 위하여 다양한 표준을 만들어 사용하고 있는데, 대표적으로 언어, 부호, 상징 등을 통하여 의사소통할 수 있고, 기준과 규격 등의 표준화를 통하여 제품의 생산 및 사용의 효율성과 호환성을 향상시켜 일상생활의 편리함을 누릴 수 있다. 이렇듯 표준화가 우리에게 가져다주는 사회, 국가적 이익과 일상생활의 편리함은 헤아릴 수 없이 많다. 이제 표준은 선택의 문제가 아닌 생존의 문제로서 다가오고 있으며, 개방화되어 가는 세계 경쟁 체제 속에서 우리의 경쟁력을 높이고 지식 정보 사회의 선진국으로 나아가기 위해서는 전 국민의 표준에 대한 중요성 인식 및 교육이 절실히 필요하다(김홍민, 이상봉, 2017).

선진 각국 역시도 국제시장에서의 표준 우위 선점과 영향력 강화를 위해 자국의 제품·서비스 또는 기술이 국제표준으로 채택될 수 있도록 표준화 전략을 국가의 중요 정책으로 삼고 추진 중에 있다. 아울러 이를 뒷받침하기 위한 표준 전문 인력 양성의 필

요성을 인식하고 표준교육 및 인력양성을 위한 다양한 노력을 펼치고 있다. 우리나라에서도 2015 개정 교육과정에서 표준이 중고등학교의 내용 요소로 채택됨으로서 앞으로 표준교육에 대한 관심은 더욱더 커질 것으로 예상된다.

이에 이 연구에서는 지금까지 표준교육과 관련된 연구 동향을 빅데이터 분석 방법에 기반하여 분석하고 시사점을 도출하고자 했다. 빅데이터 분석을 통해 표준과 관련한 연구에서 주요 주제로 다루어지고 있는 것이 무엇인지 분석하고 이를 근거로 앞으로 표준교육 연구에서 나아가야 할 방향을 제시하고자 했다. 즉 빅데이터 통찰력을 표준교육 방향 설정에 적용하고자 한 것이다.

II. 이론적 배경

1. 빅데이터의 개념

빅데이터는 용량, 다양성, 속도의 3차원적 특징을 가지고 있는 데이터의 세트를 의미한다. 빅데이터는 기존의 형식적이고 의미 있는 단어, 숫자만을 의미하는 것이 아니라 실시간으로 생성되고 순환되는 모든 정형 및 비정형의 데이터를 의미한다(김동완, 2013). 기존의 빅데이터의 개념은 단순히 데이터의 양이 많은 것을 의미했다면 최근의 일반적인 빅데이터의 개념은 기존 데이터에 비해 너무 방대해 일반적으로 사용하는 방법이나 도구로 수집, 저장, 검색, 분석, 시각화 등을 하기 어려운 정형 또는 비정형의 데이터 집합을 의미한다. 즉 빅데이터는 '빅(Big) + 데이터(Data)'라는 식의 단순한 합성어로 해석하여 빅데이터를 엄청난 크기(양)의 데이터라고 생각한다면

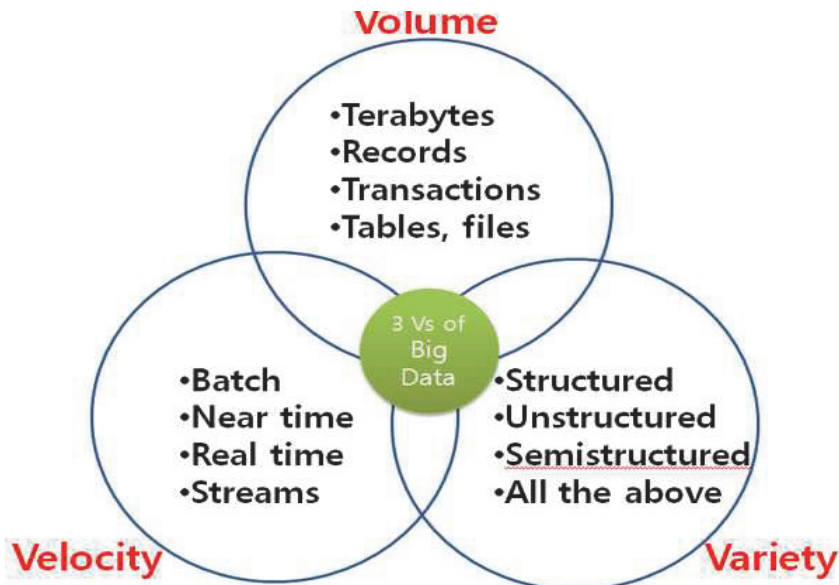
본질적인 의미와 가치를 놓치게 된다(조영임, 2013). 즉 정형화된 데이터는 물론 메타데이터와 센서 데이터, 공정 제어 데이터 등 미처 활용되지 못하고 있는 반정형화 된 데이터에 공간정보 혹은 멀티미디어 등 다양한 형태의 데이터를 포함하는 것이 빅데이터이다(유근환, 유창호, 2013). 빅데이터의 정의에 합의된 바는 없지만 Manyika 외(2011)는 데이터베이스의 규모에 초점을 맞춰 일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터, Gantz 외(2011)는 데이터베이스의 업무수행에 초점을 맞춰 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처라고 정의하였다. 최근 논의되는 빅데이터는 대용량의 데이터 외에 앱, SNS 등에서 생산되는 데이터를 포함하는 개념이다(김동완, 2013). 이상의 논의들을 종합해 봤을 때 빅데이터

는 다음의 <그림 1>과 같이 크기, 속도, 형태에 있어 일정 부분 요소를 갖추고 있어야 한다.

2. 빅데이터의 분석 방법

빅데이터는 그것 자체만으로 의미를 가지지 않는다. 그것이 분석 기술과 만나 유의미한 결과를 도출해 낼 때 비로소 자료로서 가치를 가지는 것이다. 따라서 빅데이터를 분석하는 분석 기술은 실상인즉 빅데이터 그 자체 보다 더욱더 중요한 것이라고 할 수 있다. 빅데이터 분석기술 중 현재 가장 폭넓게 사용되고 있는 것이 마이닝(Mining) 기법인데, 마이닝이란 광산에서 광물을 캔다는 의미를 가지고 있다. 즉 빅데이터 자체가 하나의 광산이라면 마이닝 기법은 그 광산에서 유용한 의미를 추출해 내는 방법을 의미하는 것이라 할 수 있다(고인규, 2016). 마이닝 기법의 구분과 그 구체적 내용은 다음의 <표 1>

<그림 1> 빅데이터의 구성 요소 <출처: 김동완, 2013 재인용>



〈표 1〉 데이터를 기반으로 한 마이닝 기법

구분	내용
데이터 마이닝 (Data Mining)	<ul style="list-style-type: none"> • 대용량의 데이터, 데이터 베이스 등에서 감춰진 지식, 기대하지 못했던 경향, 새로운 규칙 등의 유용한 정보를 발견하는 과정 • 데이터 마이닝을 통해 정보의 연관성(순차 패턴, 유사성 등)을 파악함으로써 가치 있는 정보를 만들어 의사결정에 적용
텍스트 마이닝 (Text Mining)	<ul style="list-style-type: none"> • 자연어로 구성된 비정형 텍스트 데이터에서 패턴 또는 관계를 추출하여 가치와 의미 있는 정보를 찾아내는 마이닝 기법 • 텍스트 마이닝은 사람들이 말하는 언어를 이해할 수 있는 자연언어 처리(Natural Language Process)기술에 기반함.
웹 마이닝 (Web, Mining)	<ul style="list-style-type: none"> • 인터넷상에서 수집된 정보를 데이터 마이닝 방법으로 분석 통합하는 기법 · 웹 마이닝은 콘텐츠 마이닝(웹 검색, 수집 데이터), 구조 마이닝(웹사이트 구조), 활용 마이닝(사용자 이용형태) 등으로 세분화
소셜 마이닝 (Social Mining)	<ul style="list-style-type: none"> • 소셜 미디어에 올라오는 글과 사용자를 분석해 소비자의 흐름이나 패턴 등을 분석하고, 판매나 홍보에 적용 • 마케팅 분야뿐만 아니라 사회의 흐름과 트렌드, 여론변화 추이를 읽어내는 소셜 미디어 시대의 새로운 마이닝 기법
현실 마이닝 (Reality Minig)	<ul style="list-style-type: none"> • 사람들의 행동 패턴을 예측하기 위해 사회적 행동과 관련된 정보를 기기(휴대폰, GPS 등)을 통해 얻고 분석하는 기법 • 휴대폰 등 모바일 기기들을 통해 현실에서 발생하는 정보를 기반으로 인간관계와 행동 양태 등을 추론

자료: 한국정보화진흥원(2012). 성공적인 데이터 활용을 위한 3대 요소. P. 48

과 같다.

3. 빅데이터의 분석 방법

이 연구에서는 빅데이터 분석 프로그램은 R 프로그램을 활용하였다. R 프로그래밍은 통계 계산 및 시각화를 위한 언어 및 개발 환경을 제공하며 R 언어와 개발환경을 통해 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝 기법까지 구현이 가능하다. 이렇게 구현한 결과는 그래프 등으로 시각화할 수 있으며, Java, C, Python 등의 다른 프로그래밍 언어와도 연결이 용이하다(박선주, 2015). R 프로그램은 freeware software로 현재 인터넷 상에서 무료로 다운 받을 수 있다. 이 연구에서는 교육학 분야의

R프로그램의 활용을 높이하고자 R 프로그램을 다운 받고 설치하는 방법까지 다루고자 한다.

R은 현재 <http://www.r-project.org>에서 다운 받을 수 있다. R 프로그램을 좀 더 손쉽게 이용하기 위해서는 R STUDIO라는 보조 프로그램을 같이 설치하여 사용할 수 있는데 R STUDIO는 <https://www.rstudio.com/products/rstudio/download/>에서 다운 받을 수 있다. 이 연구에서도 R 언어를 구동하기 위해 R 프로그램과 R STUDIO를 동시에 설치하여 운영하였다. R STUDIO는 R 프로그램 명령어들을 패키지화하여 손쉽게 사용할 수 있게 해주는 프로그램으로 R STUDIO에서 연구자는 교과서 분석을 위한 R 프로그램 언어를 설계하였다(고인규, 2016).

4. 표준 및 표준교육

표준(standard)이란 사전적 의미로 ‘사물의 정도나 성격 따위를 알기 위한 근거나 기준, 준거’, ‘일반적인 것 또는 평균적인 것’, ‘물리량 측정을 위한 단위를 확립하려고 쓰는, 일반적으로 인정된 기준적 시료’의 의미를 지니고 있다(국립국어연구원, 1999). 한국표준협회(2009)는 표준에 관한 보다 깊은 의미로 ‘인류가 문명을 형성해 나가면서 사람과 사람 사이의 편의와 효율성을 도모하고 공정성과 안전성을 확보하기 위해 정한 상호 약속’이라고 정의하고 있다. 이러한 표준을 분류하는 포괄적 방법으로 크게 과학기술적 표준과 인문사회적 표준으로 구분하고 성문표준, 측정표준, 참조표준의 과학기술적 표준과 언어·부호·법규·능력·태도·행동규범·책임·전통·관습·권리 및 의무 등의 인문사회적 표준으로 분류하였다(한국표준협회, 2009, p. 34).

표준교육(standard education)이 우리 교육 현장에 실시된 것은 2015 개정 교육과정 때 부터이다. 2015 개정 교육과정에는 표준교육 내용의 폭과 깊이가 더욱 확대되어 제시되었으며, 학생 경험의 확대와 기술적 문제해결 능력의 수준에 따른 계열성을 갖추도록 편성되었다(김홍민, 2017). 표준교육 내용은 공통 교육과정인 중학교의 ‘기술·가정’과목과 고등학교 일반 선택 과목인 ‘기술·가정’에 모두 편성되었다. 이전 교육과정과 비교하였을 때 가장 특징적인 변화는, 중학교에서는 표준을 사용자 측면에서 이해할 수 있도록 ‘기술의 이용과 표준’ 내용요소와 그에 따른 성취 기준을 통하여 학습자 수준에 맞도록 구성하였고, 고등학교에서는 심화한 내용인 ‘기술 개발과 표준’ 내용 요소와 성취 기준을 통하여 학습이 이루어지도록 학습자의 수준을 고려하여 학습 내용을 연계함으로써 계열성을 갖추어 표준교육에

있어서 체계화를 달성하였다는 점에서 진일보하였다(김홍민, 2017).

또한, 2015 개정 교육과정에서는 이러한 표준 관련 성취기준과 내용 요소의 체계적인 계열성을 기반으로 학생 수준에 따른 효과적인 학습 목표 달성을 위하여, 중학교 ‘교수·학습 방법’에서는 “생활 속에서 표준화되어있지 않은 사례와 제품의 불편한 사례를 바탕으로 표준화 방안을 기술적 문제해결과정을 통하여 제시할 수 있도록 하고 체험 활동을 통해 이해하도록 지도한다.”라고 제시하였으며, 고등학교 ‘교수·학습 방법’에서는 “표준이 적용된 표준화의 의미와 내용을 알게 하며, 우리 생활에서 표준화된 예를 조사, 탐구 과정을 통하여 이해하도록 한다. 또한, 표준화가 되지 않아 불편한 것들을 조사하고 표준화 아이디어로 발전시켜 발표하게 함으로써 표준화에 대한 마인드를 형성할 수 있도록 한다.”고 구체적인 교수·학습 방법을 학생 수준에 따라 제시하고 있음을 알 수 있다(김홍민, 2017).

III. 연구 방법

1. 분석대상

이 연구에서 분석대상은 한국교육학술정보원에서 제공하는 학술연구정보서비스(RISS)에서 ‘표준’, ‘표준교육’ 등의 키워드 검색을 통해 도출된 학술 자료이다. 학술자료의 범주는 국내학술지, 학술대회 발표자료 등이 포함되어 있으며 범위는 2007년부터 ~ 2017년까지 최근 10년간으로 한정하였다. 키워드 검색을 통해 도출된 학술자료는 총 441편이다.

2. 분석과정

이 연구에서 자료의 분석은 빅데이터 분석 방법 중 하나인 텍스트 마이닝(Text Mining) 방법을 적용하였다. 텍스트 마이닝은 사람들이 말하는 언어를 이해할 수 있는 자연언어 처리(Natural Language Process)기술을 기반으로 하여 자연어로 구성된 비정형 텍스트 데이터에서 패턴 또는 관계를 추출함으로써 의미 있는 정보를 찾아내는 분석 방법이다(고인규, 2016).

텍스트 마이닝의 연구과정은 정보수집, 정보처리, 정보추출, 정보분석의 4단계로 수행된다. 첫째, 정보 수집 과정에서는 연구대상으로 밝힌 텍스트데이터를 수집한다. 둘째, 정보 처리 과정에서는 텍스트를 목적인 연구 방향에 적합한 데이터 형식으로 가공한다. 셋째, 정보 추출 과정에서는 수학적 알고리즘을 활용하여 개체명을 추출한다. 넷째, 정보 분석 과정에서는 추출한 개체명에서 유의미한 개체명을 선별하고 산출한 결과물을 시각화한다(이정민 외, 2016).

3. 분석방법

1) R 소스 개발

R 프로그램은 오픈 소스 기반 프로그램으로 데이

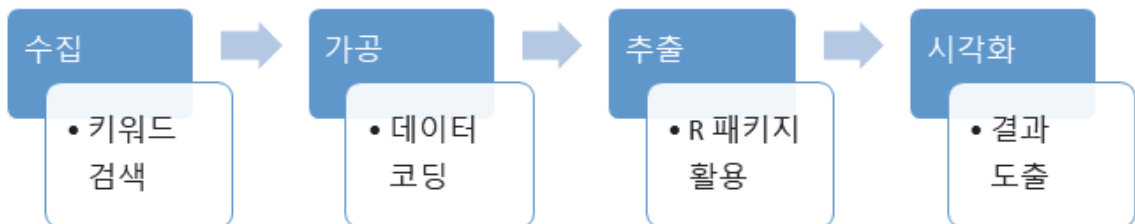
터 분석을 위해 고안된 R 프로그램의 명령과 명령의 조합 자체가 연구 결과이면서 동시에 연구 방법이 될 수 있다(고인규, 2016). 이 연구에서는 고인규(2016)가 제공한 R프로그램 소스를 참고하여 빅데이터 분석을 실시하였다. 데이터 분석을 위해 적용된 명령과 명령의 조합은 크게 3가지 과정으로 나누어 접근할 수 있다.

첫째, 분석 대상이 될 파일의 위치를 지정하고 R 프로그램 분석을 위해 사용할 기본 패키지를 설치하였다. 이 연구에서는 'tm', 'RColorBrewer', 'wordcloud', 'qcc', 'KoNLP' 패키지를 적용하였다. 패키지의 설치와 로딩을 위해 이 연구에서 적용한 명령과 명령의 조합은 다음과 같다.

```
setwd("c:\\rwork") #작업 디렉토리 지정
install.packages("KoNLP") #KoNLP 패키지 설치
install.packages("wordcloud")
install.packages("tm")
install.packages("RColorBrewer")
library(KoNLP) #분석에 사용될 패키지를 로딩
library(wordcloud)
library(tm)
library(RColorBrewer)
library(qcc)
```

둘째, 분석할 텍스트를 읽어드리고 정교한 분석을 위해 텍스트 내용을 정선하였다. 이 연구에서는 텍

〈그림 2〉 빅데이터 분석 과정



스트 분석을 위해 다음과 같은 사전 작업을 하였다. 첫째, 텍스트를 분석하기 위한 문장라인을 만들었다. 둘째, 한국어 분석을 위해 세종단어집을 불러드렸다. 셋째, 분석에서 의미가 없는 숫자, 부호, 글자를 삭제하였다. 이와 같은 과정을 수행하기 위해 적용한 명령과 명령의 조합은 다음과 같다.

```
data1 <- readLines("표준교육전처리.txt") #텍스트
파일 읽어서 data1 변수에 저장
data1 #data1에 저장된 내용 확인
data2 <- sapply(data1,extractNoun,USE.NAMES
=F) # data1에서 명사만 data2에 저장
data2 #data2에 저장된 내용 확인
data3 <- unlist(data2) #두글자 이상 필터링을 위
해 unlist작업하여 저장
data3 <- Filter(function(x) {nchar(x) >= 2},
data3) # 두 글자 이상 단어만 저장
data3 #data2에 저장된 내용 확인, 변경된 걸 확인
하실 수 있을 겁니다.
data3 <- gsub("\\d+", "", data3) #숫자 제외, 정
규식
data3 <- gsub("\\\\(", "", data3) # (괄호 제외
data3 <- gsub("\\)", "", data3) # )괄호 제외
data3 <- gsub("[A-Za-z]", "", data3) # 영문 제
외, 정규식
write(unlist(data3), "표준교육전처리2.txt") #data3
를 txt파일로 저장
data4 <- read.table("표준교육전처리2.txt") #read.
table로 읽으면서 공백 제거
wordcount <- table(data4) # 단어별 count
wordcount
```

셋째, 추출된 단어를 유의미한 정보로 해석하였다. 이를 위해서는 교과서에서 언급된 단어의 빈도수를 계산하고 계산된 빈도를 바탕으로 wordcloud 패키지를 활용하여 핵심 단어의 색과 크기를 달리하여 시각화하였다. 이와 같은 과정을 수행하기 위한 명령과 명령의 조합은 다음과 같다.

```
pal <- brewer.pal(12,"Set3") #예쁘게 꾸미기 옵션
pal <- pal[-c(1:2)]
wordcloud(names(wordcount),freq=wordcount,
scale=c(3,0.3),min.freq=8,random.order=
T,rot.per=.1,colors=pal) #가시화
pareto.chart(wordcount) #파레토 분석
head(pareto.chart(wordcount)) #상위인자 보기
x <- sort(wordcount, decreasing = TRUE) #내림
차순 순서대로 정렬
barplot(x, xlim= c(0,8)) #상위 6개 그리기
```

wordcloud의 설정에 따라 시각화 효과가 다르기 때문에 이 부분에 대한 최적화 설정이 필요하다. 이 연구에서는 다음과 같이 wordcloud 설정을 최적화 하였다. 첫째, 단어의 빈도수에 따라 나타나게 되는 단어의 색깔 범위는 brewer.pal(12,"Set3")로 설정하였다. 둘째, 제시되는 wordcloud의 크기 범위는 dh.wordcount,scale=c(3,0.3)로 설정하였다. 셋째, wordcloud에 표시되는 단어의 최소 빈도수는 min.freq=8로 설정하였다. wordcloud의 확장모양에서 일부는 random.order=T로 90도 기울어지게 하였고 그것의 비율은 rot.per=.1로 설정하였다.

2) 전처리 과정

R 프로그램에서는 데이터를 처리하기에 앞서 전처리 과정을 통해 데이터 분석의 신뢰성을 높일 수 있다. 전처리 과정은 크게 추가와 삭제 두 가지로 이루어졌다. 먼저 추가는 R프로그램의 세종사전에서 지원하지 않는 단어의 추가이다. R의 한글 텍스트 분석은 3만여 단어를 제공하는 세종사전에 기초하고 있다. 따라서 여기에 속해 있지 않는 단어는 분석에서 제외될 수 있다. 이에 연구의 신뢰성을 높이기 위해 이 연구에서는 단어 추가 명령을 활용하여 다음 같이 주요 단어를 추가하였다.

```
.libPaths() #단어 추가
mergeUserDic(data.frame("국가직무능력", "ncn"))
mergeUserDic(data.frame("교육요구", "ncn"))
mergeUserDic(data.frame("역량", "ncn"))
mergeUserDic(data.frame("표준화", "ncn"))
mergeUserDic(data.frame("직업교육", "ncn"))
mergeUserDic(data.frame("프로그램", "ncn"))
mergeUserDic(data.frame("직무", "ncn"))
```

삭제는 크게 단어, 숫자, 부호의 삭제인데 숫자와 부호의 삭제는 앞선 단계에서 이루어졌다. 단어는 워드프로세스 프로그램의 '바꾸기' 기능을 활용하여 데이터를 텍스트 파일로 변환하기 이전에 삭제하였다. '분석', '내용', '연구', '고찰', '중심으로', '연구', '방안', '탐색', '방향', '관한', '기반', '관련', '을', '를' 등을 삭제하였다.

IV. 연구 결과

표준과 관련한 연구 동향을 알아보기 위해 최근 10년간 년도 별 논문 편수를 확인해 본 결과는 다음의 <표 2>와 같다.

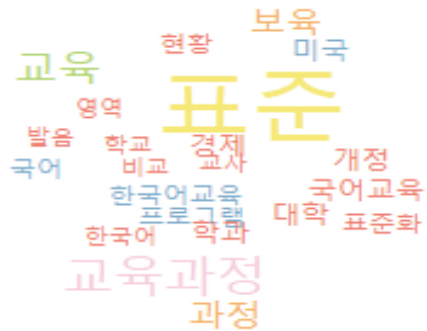
분석결과 표준 관련 연구는 연구 시기, 연구 편수 등을 고려하여 총 3 시기로 나누어서 접근할 수 있다.

먼저 1시기는 2007~2011년도 사이로 연구 편수 20편 내외 시기이다. 2시기는 2012~2014년도 사이로 연구 편수 40편 내외 시기이다. 3시기는 2014~2017년도 사이로 연구편수 60년 내외 시기이다. 이를 통해 표준 관련 연구가 최근 10년간 계단식형태로 증가하고 있음을 알 수 있었으며 특히 최근 3년간 관련 연구가 눈에 띄게 늘었음을 확인할 수 있다.

1. 1 시기(2007~2011년도)

2007년~2011년 사이는 표준 관련 연구가 매 해 20편 내외에서 수행되었음을 알 수 있다. 이 시기동안의 표준 관련 학술자료는 총 111편이며 연구 제목을 대상으로 빅데이터 분석한 결과를 워드클라우드로 나타내면 다음과 같다.

<그림 3> 1시기 wordcloud 분석결과



<표 2> 년도 별 분석 대상 편수

자료유형 \ 연도	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
학술지 (등재지, 등재후보지, 논문집)	24	18	21	19	24	39	40	35	66	70	62
학술대회 발표자료	·	·	·	·	·	3	1	1	·	2	3
기타	1	1	3	·	·	2	2	2	·	1	·
계	25	19	24	19	24	44	43	38	66	73	66

1시기에 가장 빈도수가 높은 단어는 역시 표준이었다. 이는 ‘표준’을 키워드로 도출된 결과물을 분석한 것이기에 당연한 것으로 주목할 부분은 ‘표준’ 이외에 빈도수가 높은 단어들이다. ‘교육과정’, ‘교육’, ‘과정’, ‘보육’ 순이었다.

1시기에서 주목해야 될 부분은 교육과정과 관련된 연구들이 많다는 것이었다. 이는 교육과정이 ‘표준과 준거’로서의 의미를 가지고 있고 이에 ‘표준’을 키워드로 두었을 때 교육과정과 관련된 연구들이 많이 수행되었음을 추측해 볼 수 있다. 다음으로 ‘보육’이라는 키워드도 상위에 있음에 주목할 필요가 있다. 이는 보육 중요성이 강조되고 무상 보육 교육이 시작된 시기적 특수성이 반영된 것이라고 볼 수 있는데, 이때 보육 표준에 관한 연구들이 많이 수행된 것으로 보인다. 교과와 관련해서는 ‘국어’, ‘한국교육’ 등이 많이 언급되었고 ‘미국’이라는 단어도 상위에 랭크된 것이 주목할 내용이다. 주요 연구 제목은 아래와 같다.

“국어교육 한국어 교육과정의 표준화에서 고려되어야 할

몇 문제”

“국가수준의 유아교육과정과 표준보육과정의 비교 연구”

2. 2시기 (2012~2014년도)

2012년~2014년 사이는 표준 관련 연구가 매 해 40편 내외에서 수행되었음을 알 수 있다. 이 시기 동안의 표준 관련 학술자료는 총 125편이며 연구 제목을 대상으로 빅데이터 분석한 결과를 워드클라우드 형태로 나타내면 다음과 같다.

〈그림 4〉 2시기 wordcloud 분석결과



〈표 3〉 1시기 단어 빈도분석

단어	Frequency	Cum.Freq	Percentage	Cum.Percent
표준	38	38	5.79	5.79
교육과정	21	59	3.20	8.99
교육	18	77	2.74	11.73
과정	13	90	1.98	13.71
보육	13	103	1.98	15.70
미국	9	112	1.37	17.07
국어	8	120	1.21	18.29
프로그램	8	128	1.21	19.51
한국어교육	8	136	1.21	20.73
표준화	7	143	1.06	21.79

2시기에 가장 빈도수가 높은 단어는 '교육'이었다. 이는 '표준'과 연계성을 나타내는 것으로 산업화와 관련한 연구보다는 교육과 관련 표준 연구가 가장 많이 이루어졌음을 알 수 있다. 2시기에서 눈에 띄는 단어는 '발명', '국가직무능력표준', '역량' 등이라 할 수 있다.

1시기와 다르게 발명과 관련한 표준 관련 내용이 연구주제로 다루어지기 시작하였고 특히 국가직무능력표준 등장과 관련하여 관련 연구가 많이 수행된 것으로 보인다. '발명'이라는 단어도 상위에서 랭크되었는데 표준교육이 발명과 연계되어 수행되고 있음을 간접으로 알 수 있었다. 이 외에도 '역량', '활용'이라는 단어도 국내외 새로운 교육환경과 맞물려 표준 키워드 연구에서 많이 다루어지고 있음을 알 수 있다. '대학', '초등'이라는 단어도 눈에 띄는데 표준관련 연구가 대학과 초등 분야에 집중되고 있음을 알 수 있다. 주요 연구 제목은 아래와 같다.

- “발명영재 요인 및 발명영재교육 내용표준 체계에 관한 연구”
- “대학교육에 대한 국가직무능력표준 활용 촉진 방안 연구”
- “발명교육 내용 표준에 기반 한 초·중·고등학생을 위한 발명 교과서의 적용 효과”

3. 3시기 (2015~2017년도)

2015년~2017년 사이는 표준 관련 연구가 매해 60편 이상 수행되며 표준 관련 연구가 양적으로 괄목할만한 성장을 이루었음을 알 수 있다. 3시기는 1시기 년도 별 연구 평균에 비교해 봤을 때 거의 두 배 이상의 연구가 수행되고 있으므로 이를 통해 양적 성장의 정도를 가늠할 수 있다. 이 시기동안의 표준 관련 학술자료는 총 205편이며 연구 제목을 대상으로 빅데이터 분석한 결과를 워드클라우드로 나타내면 다음과 같다.

〈그림 5〉 3시기 wordcloud 분석결과



〈표 4〉 2시기 단어 빈도분석

단어	Frequency	Cum.Freq	Percentage	Cum.Percent
교육	49	49	5.56	5.56
표준	37	86	4.20	9.77
교육과정	20	106	2.27	12.04
발명	15	135	1.59	15.34
국가직무능력표준	12	160	1.36	18.18
내용	11	171	1.25	19.43
역량	10	202	1.13	22.95
활용	9	220	1.02	25.00
대학	7	266	0.79	30.22
초등	7	280	0.79	31.81

3시기에 가장 빈도수가 높은 단어는 ‘과정’이었다. 과정의 빈도수가 가장 많은 이유는 ‘교육과정’과 연관하여 생각할 수 있다. 다음으로 ‘국가직무능력표준’, ‘표준’, ‘개발’, ‘대학’, ‘직업’, ‘능력’ 등의 단어가 빈도수가 높았다.

2시기와 비교해 보았을 때 눈에 띄는 변화는 국가직무능력표준이 5순위 정도에서 2순위로 올라 온 것이다. 이를 통해 국가직무능력표준이 표준과 관련해서는 중요한 연구 키워드로 더욱더 강조되고 있음을 확인할 수 있었다. 다음으로 눈에 띄는 단어는 ‘전문’과 ‘안전’이다. ‘전문’의 등장은 표준이 전문성과 관련된 관점에서 중요한 키워드로 다루어지고 있음을 알게 해 준다. 안전과 관련해서는 3시기가 범국가적으로 안전과 관련된 사고가 많았으며 이를 계기로 표준과 안전을 관련짓는 연구들이 많이 수행된 것으로 보인다. 주요 연구 제목은 아래와 같다.

- “국가직무능력표준 기반 음식 조리 분야 교육과정의 주관적 인식에 관한 연구”
- “국가직무능력표준 기반 대학 무용교육전문인 양성 교육과정 기준 개발”
- “역량기반 공간정보 마케팅 교육과정 개발에 관한 연구”

V. 결론 및 제언

표준이라는 개념이 우리 사회에 통용되기 시작한 시간이 그리 길지 않다. 그럼에도 불구하고 4차 산업혁명을 맞는 지금 표준은 국가 경쟁력 차원에서도 매우 중요한 과제로 다루어지고 있다. 혹자는 표준이 사고 확장을 막고 유연하고 창의적으로 생각하는 길을 막는다고 주장하기도 한다. 하지만 연구 결과에서도 살펴보았듯이 발명, 역량, 활용 같은 창의적 키워드와 표준은 맞물려있었다. 표준이 일반화를 추구하여 사고의 확장을 막는다는 주장은 표준의 개념에 대한 오해에서 기인한 것이다.

국가기술표준원은 표준의 의미를 다음과 같이 3가지 측면에서 논의하고 있다.

첫째, 표준은 사회·경제적인 효율을 향상시키는 중요한 수단이다. 표준은 원료나 자원으로 부터 제품이나 서비스로 생산하는 모든 과정에서 생산효율을 증가시키고 품질의 향상과 소비자를 보호하는 효과를 가져온다. 최근에는 그동안의 제품 위주의 표준 제정에서 벗어나 유통·물류·소프트웨어·서비스 등

〈표 5〉 3시기 단어 빈도분석

단어	Frequency	Cum.Freq	Percentage	Cum.Percent
과정	78	78	6.08	6.08
국가직무능력표준	70	148	5.46	11.54
표준	42	190	3.27	14.82
개발	30	220	2.34	17.16
대학	29	249	2.26	19.42
직업	21	270	1.63	21.06
능력	19	289	1.48	22.54
전문	19	308	1.48	24.02
안전	18	326	1.40	25.42
한국어	15	341	1.17	26.59

전 산업분야로 확대되어 다가오는 21세기 노령화 사회, 정보화사회에 대비한 필수적인 경제사회의 혁신수단으로 인식되고 있다. 둘째, 표준은 산업발전 기반이 된다. 완성도 높은 기술의 표준화는 기술적용 제품의 시장적합성과 경쟁력을 향상시켜 준다. 특히, 정보기술·멀티미디어·HDTV·DVD 등에 적용되는 신기술에 대한 선행적 표준화는 첨단산업기술발전의 기반이 되며, 기술투자의 중복을 방지하고 기술을 이전하는 데에도 필수적인 산업발전기반이 된다. 셋째, 표준은 교역증대와 무역 자유화의 기반이 된다. 국제표준과 국가표준의 부합화는 국가간 상호인정을 촉진하고 무역 증대 및 경제 통합 하는데 중요한 역할을 한다. WTO/TBT협정(Agreement on Technical Barriers to Trade)은 각 국가의 기술규정과 표준이 국제 무역에 있어서 장벽이 되지 않도록 국가표준(기술규정 포함)을 제·개정시 국제표준이 있는 경우 이를 채택 적용하도록 규정하고 있다. 이상의 정의에서 보듯 표준은 사회·경제적인 효율성 측면에서 중요하게 다루어지며 산업의 경쟁력을 높이는 매우 중요한 키워드로 다루어지고 있음을 확인할 수 있다. 이는 사고의 확장, 창의성이 수반되지 않고는 불가능한 것이다. 결국 표준에 대한 우려는 기우이며 이제는 어떻게 하면 표준을 확장하고 널리 알릴 것인가에 대한 문제가 남은 것이다. 표준의 효과를 극대화하기 위해서는 학생 시절부터 표준의 중요성을 이해하고 표준적 사고 양식을 개발하는 하는 것이 중요하다. 이를 위해서는 학교 교육에서부터 체계적인 표준 관련 교육이 수행되어야 하는데 2015 개정 실과(기술·가정)교육과정에 표준 관련 내용이 포함될 것은 매우 바람직한 현상이라고 할 수 있을 것이다.

이 연구는 최근 10년간의 표준, 표준교육 관련 키워드 검색을 통해 최근 표준이라는 키워드와 관련하여

여 어떤 연구들이 수행되고 있는지를 살펴본 것으로 이를 통해 표준관련 연구 동향의 추이를 살펴보았다. 연구 결과로 확인할 수 있었던 것은 최근 3년간 표준 관련 연구가 양적으로 급증하고 있으며 이는 국가직무능력표준과 맞물려 있음을 확인할 수 있었다. 또한 발명, 안전, 교육과정 등에서도 표준 관련 연구들이 지속적으로 늘어나고 있음을 확인할 수 있다. 초기(2007년 대) 국어교육에서 표준어, 표준문법 같은 분야에 한정되어 있던 연구가 산업분야에 중요 키워드로 떠오르고 있는 것이다. 연구를 마무리 하면서 다시 한 번 제언 하는 것은 결국 표준이 우리 사회에 뿌리 깊게 안착하기 위해서는 학교 교육에서 중요한 내용 요소로 다루어지는 것이 중요하다. 2015 개정 교육과정을 계기로 표준이 교육의 내용 요소로 처음 도입되었다면 이제는 지속적인 후속 연구를 통해 표준을 잘 가르치고 평가하는 방법까지도 연구되어야 할 것이다. 이 연구의 결과는 그런 부분에서도 어떤 방향으로 표준교육이 나아가야 할 것인지 방향 설정의 참고 자료가 될 수 있을 것이다.

참고문헌

- 고인규(2016). 빅데이터 분석을 통한 실과 농업단원의 핵심개념 분석 연구. 실과교육연구, 16(3), 43-65.
- 고효경, 최영우, 박선정(2014). 수학교육에서 빅데이터 활용 방안에 대한 소고. 수학교육논문집, 27(4), 573-588.
- 권대석(2012). 클라우드와 슈퍼컴퓨팅이 이끄는 미래: 빅데이터 혁명. 서울: ㈜21세기 북스.
- 권영욱(2013). 빅데이터를 활용한 맞춤형 교육 서비스 활성화 방안연구. 지능정보연구, 15(4), 87-100.
- 권혁수 외 16인(2014). 초등학교 실과 5학년. 서울: ㈜

- 미래엔.
- 김동완(2013). 빅데이터의 분야별 활용사례. 경영논총, 34, 39-52.
- 김중우(2014). 교육분위론에 기초한 실과 교육 소재의 범주화 농업 재배 영역 단위 예시를 중심으로. 학습자중심교과교육연구, 14(10), 609-626.
- 노규성(2015). 학부생의 빅데이터 인식 분석을 통한 교육 정책 제언. 디지털융복합연구13(11), 25-33.
- 박선주(2015). R을 활용한 SW교육 텍스트데이터 토픽 분석. 정보교육학회논문지, 19(4), 517-524.
- 김홍민, 이상봉(2017). 2015 개정 기술·가정과 교육과정의 중학교 '기술 활용'영역에서 표준 관련 성취 기준을 위한 체험활동 과제 개발. 실과교육연구, 23(1), 243-262.
- 이석주, 연지윤, 천승훈(2013). 빅데이터를 이용한 교통 정책 개발 및 활용성 증대방안 (기본연구보고서 2013. 11). 서울: 한국교통연구원.
- 이은경, 박도영, 최인봉(2014). 교육 빅데이터 관련 연구 동향. 한국컴퓨터정보학회 학술발표논문집, 22(2), 175-176.
- 이정민, 전은자, 채정민(2016). 텍스트 마이닝을 기반으로 한 무용학 자료의 빅데이터 분석. 무용역사기록학 제42호, 192-212.
- 장산현(2012). 빅데이터와 스마트 교육. 정보과학회지, 20(4), 59-64.
- 정성봉 외 10인(2013). 초등학교 실과 5학년. 서울: 교학사.
- 조영임(2013). 빅데이터의 이해와 주요 이슈들. 한국지역정보학회지, 16(3), 43-65.
- 최제영, 박충식(2013). 빅데이터와 교육. 한국컴퓨터정보학회지, 21(1), 9-23.
- 한국정보화 진흥원(2012). 성공적인 데이터 활용을 위한 3대 요소.

논문접수일 : 2018. 09. 17
 1차수정본접수일 : 2018. 11. 20
 게재확정일 : 2018. 11. 24

Big Data Analysis of Standards and Standards Education by Text Mining

Go, In Gyu*

Abstract

The purpose of this study was to analyze trends in standards-related research using Big Data techniques. In order to achieve the purpose of the study, Big data analysis was conducted on 441 academic materials derived from keyword search such as standard and standard education from May 1, 2017 to December 30. The types of big data analysis methods were text mining techniques. 'Tm', 'RColorBrewer', 'word cloud', 'qcc', 'KoNLP' were used package in R. It was to delete the words and symbols that do not require and by using the morphological analysis of the natural language processing technology made the statement as a minimum unit. In this paper, we analyze the frequency of word appearing in the title of academic articles and present it as a word crowd using the R program's crowding package. The results of this study are as follows. First, the most frequently mentioned word during the first period (2007 ~ 21011) was 'standard'. In addition to the standard, the most frequent words were 'curriculum', 'education', 'course', and 'childcare'. The most frequently mentioned word between the two periods (2012 ~ 2014) was education. Next, there were 'standard', 'invention', 'national incompetence' and 'competence'. The word most often mentioned in the third period (2015 ~ 2017) was the 'process'. Next, there were 'inventions', 'national incompetence', and 'competence'. The standardized researches showed quantitative growth through the 1 st, 2 nd, and 3 rd periods, and the key keywords were 'Curriculum', 'Education', 'Process', 'Child Care', 'Invention', 'National Disability' And 'competence'.

※ Key Words: Standard, Standard Education, Big Data Analysis, Competency, R Program

* Doctor, Sejong Yangji Elementary School (goingyu@hanmail.net)