

R을 활용한 SW교육 텍스트데이터 토픽분석

박선주

광주교육대학교 컴퓨터교육과

요 약

본 논문에서는 사람들의 SW 교육과 관련된 관심방향을 알아보기 위해 SW 교육 관련 뉴스데이터를 수집하여 그 내용을 분석하였다. 이를 위해 2013년 7월 23일~2015년 10월 19일의 SW 교육관련 뉴스데이터의 토픽분석을 실시하였다. R을 사용하여 웹크롤링 후 가장 자주 언급된 상위 20개 단어들 간의 관련성을 분석한 결과, SW 교육 단어를 중심으로 20개 단어의 노드 크기와 연결선의 두께가 비교적 균형을 이루고 있어 서로의 관련성이 밀접하게 유지되는 데이터임을 알 수 있었다, 또한, 분석대상 데이터는 주로 SW 인재양성, SW 지원 프로그램, SW 교육 의무화, SW 캠프, SW 산업, 일자리 창출과 관련된 토픽들임을 알 수 있었다. 이는 SW 교육에 관한 사람들의 생각 및 관심분야 등을 알아보는 빅데이터 분석 자료에 활용될 수 있을 것이다.

키워드 : SW교육, 뉴스분석, 빅데이터, 토픽분석, R

A Topic Analysis of SW Education Textdata Using R

Sunju Park

Dept. of Computer Science Education, Gwangju National University of Education

ABSTRACT

In this paper, to find out the direction of interest related to the SW education, SW education news data were gathered and its contents were analyzed. The topic analysis of SW education news was performed by collecting the data of July 23, 2013 to October 19, 2015. By analyzing the relationship among the most mentioned top 20 words with the web crawling using R, the result indicated that the 20 words are the closely relevant data as the thickness of the node size of the 20 words was balancing each other in the co-occurrence matrix graph focusing on the 'SW education' word. Moreover, our analysis revealed that the data were mainly composed of the topics about SW talent, SW support Program, SW educational mandate, SW camp, SW industry and the job creation. This could be used for big data analysis to find out the thoughts and interests of such people in the SW education.

Keywords : SW Education, News Analysis, Big Data, Topic Analysis, R

이 논문은 2015년도 광주교육대학교 학술연구비 지원에 의한 것임.

논문투고 : 2015-12-01

논문심사 : 2015-12-01

심사완료 : 2015-12-18

1. 서론

최근 빅데이터(Big Data) 기반으로 분석한 결과를 의사결정에 활용하는 기법들이 점차 확산되고 있다. 웹 3.0과 스마트 기기의 일반화라는 흐름에 따라 비정형 빅데이터가 폭발적인 증가 추세를 보이면서 비정형 빅데이터의 시대에 돌입하고 있다[8].

빅데이터는 수집 자료를 저장 및 분석하고 그 결과를 표현하는 처리 과정을 거치는데, 각 단계마다 세부 영역과 관련 기술이 등장하고 있으며[7][10][11], 빅데이터에 대한 연구가 다양한 분야에서 활발히 이루어지고 있다. 교육분야에서도 스마트교육(Smart Education) 환경에서의 활용방안, 교육정책 아젠다 도출, 교육분야 이슈 자동화 기술 등의 연구들이 진행되었다. 이에 본 논문에서는 SW 교육관련 데이터를 분석해보고자 한다.

우리나라에서는 미래사회를 위한 SW교육과 산업현장에 필요한 SW인재양성을 위해 교육부, 미래부가 'SW 중심사회를 위한 인재양성 추진계획'(2015. 7.21) (교육부, 2015)을 발표하여 2018년 SW 교육 필수화에 대비한 초·중등 SW 교육 확산 및 산업계 수요에 부응하는 대학 SW 교육의 양적, 질적 혁신을 위한 12개 과제를 선정하여 추진하고 있다. 또한, SW 연구학교, SW 선도학교 등 시범사업이 추진되고 있으며, 2018년부터 초·중등 SW교육이 필수화되는 등 SW 교육 저변 확산과 관심이 고조되고 있는 상황이다[9].

그러므로 본 논문에서는 2015년 7월 21일 교육부의 'SW중심사회를 위한 인재양성 추진계획' 발표를 계기로 고조된 SW 교육과 관련된 뉴스데이터를 분석하여 그 내용을 알아보고자 한다. 대상 데이터의 범위를 네이버뉴스기사로 정하고 네이버 포털 뉴스에 SW 교육 관련 기사를 수집한다. 수집된 데이터의 텍스트를 분석 후 가장 자주 언급되는 단어들을 추출하여 관련성을 알아보고, 토픽 분석을 통해 토픽을 추출한다. 이를 위해 빅데이터 분석도구인 R을 활용하여 네이버 뉴스에서 'SW 교육' 뉴스 기사를 웹크롤링(Web Crawling)하여 자료를 수집한 후 형태소 분석 후 단어를 추출한다, 그리고 기사별 단어의 매트릭스를 만들어 빈도수가 높은 단어를 선정하고 그들 간의 관련성을 co-occurrence matrix로 알아본다. 데이터에서 추출된 단어들을 기반으로 LDA(Latent Dirichlet Allocation) 기법을 활용하

여 토픽 분석을 실시한다. 이를 통해 SW 교육관련 뉴스데이터의 주요 내용을 파악할 수 있을 것이다.

2. 이론적 배경

2.1 빅데이터

빅데이터는 용량, 다양성, 속도의 3차원적인 특징을 가지는 데이터의 세트를 의미한다. 빅데이터는 기존의 형식적이고 의미 있는 단어나 숫자만을 의미하는 것이 아니라 실시간으로 생성되고 순환되는 모든 정형 및 비정형 데이터를 빅데이터로 정의할 수 있다[8].

빅데이터는 기존의 관리 및 분석 체계로는 감당하기 어려운 막대한 데이터 집합과 이를 해결하기 위한 플랫폼과 분석 기법 등을 포괄하고 있다. 데이터 분석 기법들은 통계학, 진산학, 기계학습 등에서 이미 사용되고 있는 기법들이었으나 빅데이터의 등장으로 새롭게 주목 받으며 발전하였다[11].

빅데이터는 수집 자료를 저장 및 분석하고 그 결과를 표현하는 처리 과정을 거치는데, 각 단계마다 세부 영역과 관련 기술이 등장하였다[7][10][11].

2.2 R

R 프로그래밍은 통계 계산 및 시각화를 위한 언어 및 개발환경을 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. MacOS, 리눅스, 유닉스, 윈도우 등의 대부분의 컴퓨팅 환경을 지원하는 것도 장점이다. 위의 장점들로 인하여 R은 통계 분석 분야에서 인지도를 높여왔으며, Hadoop 환경상에서 분산처리를 지원하는 라이브러리 덕분에 구글, 페이스북, 아마존 등의 빅데이터 분석이 필요한 기업에서 대용량 데이터 통계 분석 및 데이터 마이닝을 위해 널리 사용되고 있다[8].

그러므로 본 논문에서도 텍스트데이터 분석을 위해 R을 사용하였다. R 프로그램에는 사용가능한 여러 패키

지들이 있다. 본 논문에서는 데이터의 시각화를 위해 wordcloud와 ggraph 패키지를 사용하였으며, 데이터 정형화를 위해 tm, KoNLP 패키지, 용어-문헌 행렬 (term-document matrix)을 만들고 LDA 분석을 위해 tm, topicmodels, LDA 패키지를 사용하였다.

2.3 선행연구

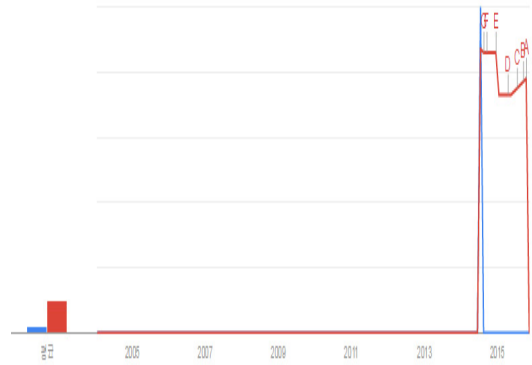
교육관련 빅데이터에 대한 연구는 그리 많지 않다.

최제영(2012)은 스마트교육 서비스에서 발생할 수 있는 교육 데이터의 종류와 크기를 추정하고, 스마트교육 환경에서의 빅데이터 활용방안을 제안하고 있다[4]. 조용상 외(2013)는 교육 분야의 빅데이터 분석이라고 할 수 있는 학습분석의 기술 동향에 대해 논의하고 있다[7]. 김우주(2013)는 교육정보 공시항목과 접속 로그데이터를 이용하여 교육 빅데이터에 대한 조희를 기준으로 정책적 중요도를 판단하면서 정책 아젠다를 도출하였다[6]. 박진수(2014)는 연구문헌, 언론 빅데이터 분석을 통해 교육정책에 반영할 수 있는 정책 아젠다를 개발하였다[10]. 손정은(2013)은 교육분야 이슈들을 자동 분류처리하는 키워드 추출, 분류, 유사도 계산 등 빅데이터의 자동분류처리 기술에 관해 연구하였다[11].

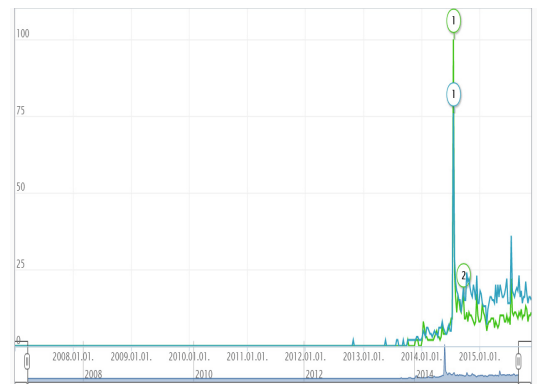
3. SW 교육 뉴스데이터 분석

2015년 7월 21일 SW중심사회를 위한 인재양성 추진 계획 발표로 교육계 및 사회 전반에서 SW 교육의 중요성이 인식되고 SW 교육에 대한 관심이 고조되고 있다. 구글 트렌드와 네이버 트렌드를 사용하여 ‘SW 교육’, ‘소프트웨어 교육’ 키워드에 대한 검색결과(Fig. 1, Fig. 2)를 살펴보더라도 2013년 중반 이후부터 검색량이 증가한 것을 볼 수 있다.

그러므로 이 기간 동안 주로 인터넷 뉴스에서는 어떤 내용들이 기사화되었는가를 알아보기 위해 토픽 분석을 실시하고자 한다.



(Fig. 1) Trend analysis of SW education in Google
출처: <https://www.google.com/trends/?hl=ko>



(Fig. 2) Trend analysis of SW education in NAVER
출처: <http://me2.do/5jq9xRWF>

3.1 분석 데이터 선정 및 수집

본 논문의 분석 대상 데이터는 국내 검색사이트 순위 1위인 네이버의 뉴스 데이터로 선정하였다. 데이터 수집 범위는 2013년 7월 23일~2015년 10월 19일의 SW 교육관련 네이버 뉴스 데이터로 2014년 7월 23일 ‘SW 중심사회 실현전략보고회’와 2015년 7월 21일 ‘SW 중심사회를 위한 인재양성 추진계획’이 발표된 시점을 고려하여 범위를 선정하였다.

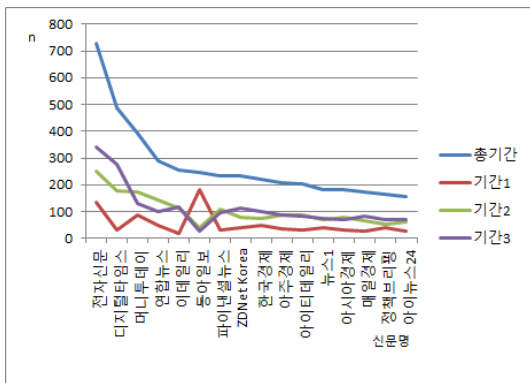
네이버에서 키워드는 ‘SW 교육’, ‘소프트웨어 교육’, 날짜범위는 2013년 7월 23일~2015년 10월 19일의 뉴스

를 검색하여 뉴스제목, 날짜, 출처, 내용, 뉴스 링크주소를 R로 웹크롤링하여 8,157개의 데이터를 수집하였다. 웹 크롤링 시 데이터의 중복은 제거하였으며 수집된 데이터 총수와 기간별 데이터 수는 <Table 1>과 같다.

<Table 1> Number of analyzed data

	Date range	Number of Data
Period 1	2015. 7. 21~2015. 10. 19	1,694
Period 2	2014. 7. 23~2015. 7. 20	3,398
Period 3	2013. 7. 23~2014. 7. 22	3,065
	Whole Period	8,157

수집된 뉴스데이터의 출처를 살펴보면 전자신문 727개, 디지털타임스 487개, 머니투데이 393개, 연합뉴스 289개, 이데일리 253개 등의 순으로 나타났으며, 주로 IT 관련 신문과 경제 신문에서 기사가 게재되었음을 알 수 있었다. 또한, 기간1, 기간2, 기간3 모두 데이터의 출처가 거의 유사하였다. 분석 데이터의 출처 현황은 (Fig. 3)과 같으며, 150개 이상 데이터가 수집된 출처만을 나타냈다.



(Fig. 3) Current situation of the source of data

3.2 데이터 정형화

수집된 데이터는 텍스트 형식의 비정형 데이터이므로 R을 이용하여 정형 데이터로 변환하였다. R의 tm 패키지와 KoNLP 패키지를 사용하여 불용어 및 기호 등을 제거하여 데이터를 정제한 후 자연어 처리 기술중

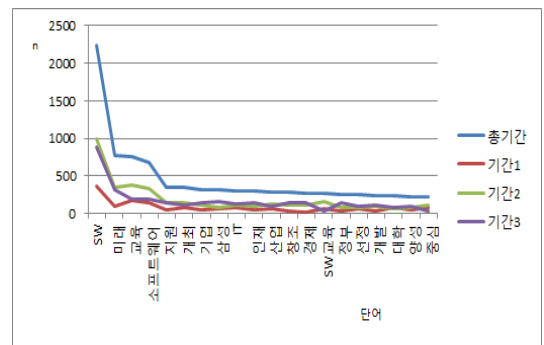
형태소 분석 기법을 통해 문장을 최소단위로 쪼개는 작업을 진행했다.

수집된 8,157건의 텍스트 데이터를 명사 중심의 형태소 분석을 실시하여 20,128개의 명사를 추출하였다.

3.3 단어추출 및 분석

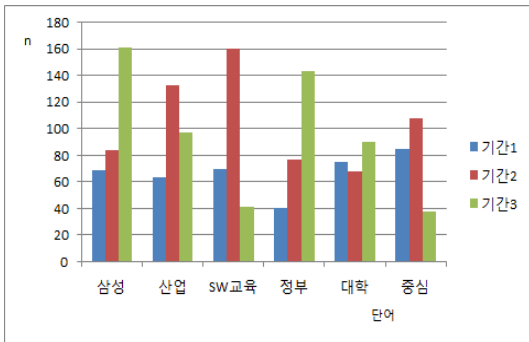
20,128개의 명사 중 숫자형태는 지우고 단어길이 가 두 글자 이상인 명사 8,573개를 추출하였으며, 추출된 단어가 분석대상 데이터에 언급되는 횟수를 산출하였다.

그중 분석대상 데이터에 가장 자주 언급된 상위 20개의 단어는 sw(2,243회), 미래(772회), 교육(759회), 소프트웨어(681회), 지원(348회), 개최(347회), 기업(315회), 삼성(314회), IT(309회), 인재(302회), 산업(293회), 창조(285회), 경제(279회), SW교육(271회), 정부(260회), 선정(248회), 개발(240회), 대학(233회), 양성(231회), 중심(231회) 이었다. 가장 많이 언급된 20개 단어의 언급횟수 현황은 (Fig. 4)와 같다.



(Fig. 4) Number of times being mentioned for 20 words

상위 20개 단어들 중 기간별 언급횟수를 비교해본 결과, 'SW 교육, 대학, 중심' 단어가 기간2에서 많이 언급되었으며, '삼성, 정부' 단어는 기간3에서, '산업' 단어는 기간2에서 더 많이 언급되었음을 (Fig. 5)와 같이 알 수 있었다.



(Fig. 5) Comparison of number of times being mentioned according to the periods

<Table 2>와 같이 기간별 상위 20개의 단어들을 추출하여 비교해본 결과, 전체 데이터에서 추출한 단어 20개와 비교하여 기간1에서는 클라우드, 청년, 프로그램, ICT, 강화 단어가, 기간2에서는 사업, ICT, 네이버, 사회 단어가, 기간3에서는 육성, 사업 단어가 새롭게 포함되어 있었다. 또한, 상위 20개 단어 목록에 포함되어 있지 않으나 교육과정, 코딩, 로봇, 교사 단어 등은 최근 데이터일수록 언급되는 횟수가 점점 증가하고 있는 것으로 보였다. 이는 2015년 7월 21일 ‘SW 중심사회를 위한 인재양성 추진계획’ 발표 전후로 SW 교육과정, 코딩 등이 이슈화되고 있는 것으로 볼 수 있다.

<Table 2> List of top 20 words according to the periods

Whole Period	Period 1	Period 2	Period 3
SW	SW	SW	SW
Future	Education	Education	Future
Education	Software	Future	Education
Software	Future	Software	Software
Support	Host	SW education	Samsung
Host	Central	Host	Support
Company	IT	Support	Talented person
Samsung	University	Industry	Government
IT	SW education	Company	Economy
Talented person	Samsung	Creativity	Company
Industry	Industry	Economy	Creativity
Creativity	Selection	IT	IT
Economy	Cloud	Center	Host
SW education	Youth	Business	Development
Government	Support	Development	Foster
Selection	Government	Talented person	Selection

Development	Talented person	ICT	Industry
University	Program	Naver	Promotion
Foster	ICT	Society	University
Central	Strengthen	Selection	Business

3.4 단어 간의 관련성

추출된 상위 20개 단어들의 관련성을 알아보기 위해 R에서 co-occurrence matrix를 구하여 20개 단어들이 뉴스 데이터에 함께 나타나는 횟수를 구하였다. 횟수가 증가하면 두 단어가 함께 나타나는 횟수가 많다는 의미이므로 두 단어의 관련성이 크므로 두 단어를 포함하는 내용의 기사가 많다는 것을 알 수 있다.

Term Document Matrix를 만들어 20개 단어들의 co-occurrence matrix를 구하기 위한 R 코드는 아래와 같다.

```
# 명사를 뽑는 함수
ko.words <- function(doc){
  d <- as.character(doc)
  extractNoun(d)
}

# Term Document Matrix 만들기
library(KoNLP)
library(stringr)
library(tm)
options(mc.cores=1)
cps <- Corpus(VectorSource(new.reviews))
tdm <- TermDocumentMatrix(cps,
  control=list(tokenize=ko.words,
    removePunctuation=T,
    removeNumbers=T,
    wordLengths=c(2, Inf),
    weighting=weightBin))

tdm.matrix <- as.matrix(tdm)
# co-occurrence Matrix 만들기
word.count <- rowSums(tdm.matrix)
word.order <-
  order(word.count, decreasing=T)
rownames(tdm.matrix)[word.order[1:20]]
freq.words <-
  tdm.matrix[word.order[1:20], ]
co.matrix = freq.words %*% t(freq.words)
```

또한 co-occurrence matrix를 R의 ggraph 패키지를 이용하여 시각화한 결과는 (Fig. 6)~(Fig. 9)와 같다.

그래프에서 단어의 노드 크기가 크면 빈도가 높고, 단어의 노드 크기가 작으면 빈도가 낮다는 것을 의미한다. 또한, 두 단어 간 상관성을 계산하여 선으로 표현하여 두 단어 간의 상관성이 높으면 선의 굵기도 굵게 나타나고, 상관성이 낮으면 선의 굵기도 가늘게 나타난다.

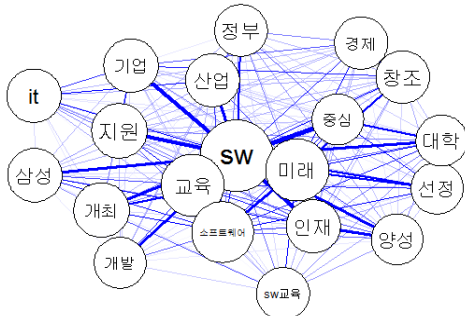
본 논문의 분석대상 데이터는 (Fig. 6)에서처럼 SW 단어는 ‘교육, 미래, 중심, 인재, 지원, 기업, 산업, 개최, 대학, 양성, 개발, 삼성, 선정’ 등의 단어들과 함께 나타나는 횟수가 많았으며, ‘대학, 선정’ 단어가, ‘창조, 경제’ 단어가, ‘인재, 양성’ 단어가 함께 나타나는 횟수가 많음을 알 수 있었다.

(Fig. 7)의 기간1에서는 SW 단어는 ‘교육, 중심, 미래, 대학, 인재, 개최, 선정’ 등의 단어들과 함께 나타나는 횟수가 많았으며, ‘삼성, 청년’ 단어가, ‘대학, 중심’ 단어가, ‘대학, 선정, 중심’ 단어가 함께 나타나는 횟수가 많았다.

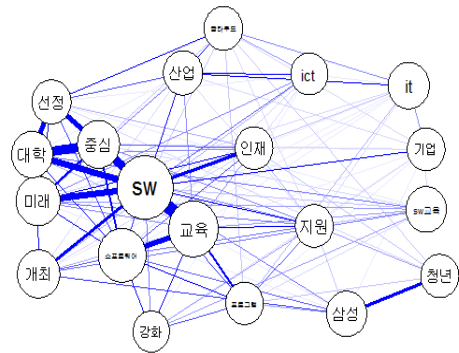
(Fig. 8)의 기간2에서는 SW 단어는 ‘교육, 미래, 중심, 사회, 지원, 산업, 인재, 기업’ 단어들과 함께 나타나는 횟수가 많았으며, ‘창조, 경제’, ‘중심, 사회’, ‘네이버, 교육, SW’ 단어가 함께 나타나는 횟수가 많았다.

(Fig. 9)의 기간3에서는 SW 단어는 ‘교육, 미래, 인재, 양성, 기업, 삼성, 지원, 산업, 개최’ 등의 단어들과 함께 나타나는 횟수가 많았으며, ‘창조, 경제’, 인재는 ‘양성, 삼성’, ‘대학, 선정’ 단어가 함께 나타나는 횟수가 많았다.

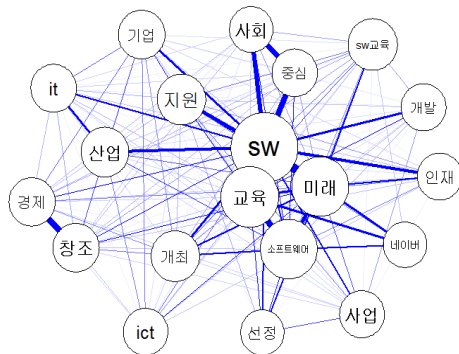
이상으로 분석대상 데이터는 (Fig. 6)에서처럼 SW 교육 단어를 중심으로 20개 단어의 노드 크기와 연결선의 두께가 비교적 균형을 이루고 있어 서로의 관련성이 밀접하게 유지되는 데이터임을 알 수 있었다.



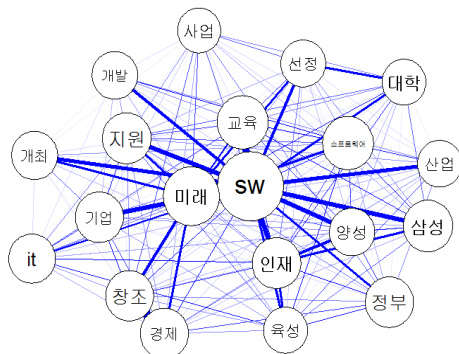
(Fig. 6) Co-occurrence matrix of the whole period



(Fig. 7) Co-occurrence matrix of period 1



(Fig. 8) Co-occurrence matrix of period 2



(Fig. 9) Co-occurrence matrix of period 3

3.5 토픽 분석

분석대상 데이터의 토픽들을 추출하기 위해 토픽 모델링을 실시하였다, 본 논문에서는 토픽 모델링중 가장 많이 활용되고 있는 LDA 기법을 사용하여 분석하고자 한다.

LDA는 하나의 문서를 여러 단어들의 집합으로 인식하여 해당 문서에 많이 나오는 단어를 확률에 근거하여 토픽을 추출하는 텍스트 마이닝 기법이다. 즉, 대용량의 문서에서 주제를 추출하기 위한 기법으로, 동시 발생하는 확률을 기반으로 유사한 문서들을 클러스터링 하는 기법이다.

토픽분석을 위해 본 논문에서는 R의 topicmodels 패키지를 이용하여 토픽갯수는 20개, 토픽별 단어는 상위 10개를 추출하였으며, 전체 문서에 대한 해당 토픽의 출현비율도 구하였다. 출현비율 0.05 이상인 토픽들의 토픽별 출현비율과 키워드는 <Table 3>과 같으며, 출현비율 순으로 정리하였다.

토픽 15번, 9번(SW 인재양성), 5번(SW 지원 프로그램), 13번(SW 교육 의무화), 20번(SW 캠프)은 SW 교육과 관련된 토픽이며 6번, 11번, 7번은 SW 산업과 관련된 토픽이며, 17번은 일자리 창출과 관련된 토픽이다. 즉, 본 논문의 분석대상 데이터의 토픽은 주로 SW 인재양성, SW 지원 프로그램, SW 교육 의무화, SW 캠프, SW 산업, 일자리 창출과 관련된 토픽들을 알 수 있었다.

<Table 3> Percentage of appearance of each topic and its keywords

Topic Number	Appearance ratio	Topic keywords
15	0.08956916	SW, Central, University, Selection, Software, Future, Foster, Education, Talented person, Society
5	0.06613757	Education, SW, Robot, Naver, Software, Strengthen, Program, Support, Smart, Target
13	0.06311413	Education, SW education, Software, Obligation, Elementary, middle school, SW educator, Middle school, Exhibition, Host
6	0.05782313	Company, Development, Security, Management, SW, software, Cloud, Quality, Base, Prepare

20	0.05681532	Camp, Host, Future, Software, SW, Business, Junior, Field, Creativity
11	0.05366591	Cloud, Industry, Trend, IT, Service, Week, Selection, Market, Innovation, Complex
17	0.05303603	Youth, Job, Samsung, Employment, Create, Offer, Group, Academy, Prime minister
7	0.05202822	Next year, Industry, ICT, Investment, Development, Future, Government, Talented person, Smart, Vitalization
9	0.05051650	IT, Class room, Make public, Children, Game, Review, Figure, Developer, Internet

4. 결론

데이터의 급속한 증가로 데이터를 활용한 새로운 가치 창출은 기업뿐 아니라 국가 경쟁력의 중요한 요소로 대두되고 있고, 이에 따라 전 세계적으로 국가 차원에서 빅데이터 활성화를 위한 다양한 노력을 기울이고 있다[10].

이에 본 논문에서는 사람들의 SW 교육과 관련된 관심방향을 알아보기 위해 먼저 SW 교육 관련 뉴스데이터를 수집하여 그 내용을 분석하였다. 네이버 포털뉴스 2013년 7월 23일~2015년 10월 19일의 뉴스데이터의 토픽분석을 실시한 결과, 분석대상 데이터에 가장 자주 언급된 상위 20개의 단어는 SW(2,243회), 미래(772회), 교육(759회), 소프트웨어(681회), 지원(348회), 개최(347회), 기업(315회), 삼성(314회), IT(309회), 인재(302회), 산업(293회), 창조(285회), 경제(279회), sw교육(271회), 정부(260회), 선정(248회), 개발(240회), 대학(233회), 양성(231회), 중심(231회)이었다. 이는 2015년 7월 21일 'SW중심사회를 위한 인재양성 추진계획'의 발표가 SW 교육관련 단어의 등장에 영향을 미치고 있으며, SW 교육과정, 코딩 등이 이슈화되고 있는 것을 알 수 있었다. 또한, 본 논문의 분석대상 데이터는 SW 교육 단어를 중심으로 20개 단어의 노드 크기와 연결선의 두께가 비교적 균형을 이루고 있어 서로의 관련성이 밀접하게 유지되는 데이터임을 알 수 있었다. 또한, 분석대상 데이

터의 토픽은 주로 SW 인재양성, SW 지원 프로그램, SW 교육 의무화, SW 캠프, SW 산업, 일자리 창출과 관련된 토픽들을 알 수 있었다.

분석결과는 SW 교육에 관한 사람들의 생각 및 관심 분야 등도 알아보는 빅데이터 분석 자료로 활용될 수 있을 것이다.

참고문헌

- [1] Big Data Strategy Research, the Society of Digital Policy and Management (2015). Big data analytics for business, Wow Pass.
- [2] Cho, A. (2015). Mobility Pattern Analysis of Bus Passengers with LDA, Master's dissertation, Graduate school in Chungbuk National University.
- [3] Cho, YoungSang (2013). Prospects for the Application of Learning Analytics, KERIS issue report, RM 2013-1.
- [4] Choi, JaeYoung (2012). Big Data trend in smart learning environment, KERIS issue report RM 2012-19.
- [5] Kim, DaeHyun (2014). An Analysis of the Enterprise Social Network Service Using Topic Modeling, Master's dissertation, Graduate school in Yonsei University.
- [6] Kim, WooJu (2013). Big data analytics consulting report - focused on the education information disclosure item, KERIS.
- [7] Kwon, YoungOk (2013). Data Analytics in Education : Current and Future Directions, Journal of Intelligence and Information Systems 19(2). pp. 87-100.
- [8] Lee, HeeSub (2014). A Study on Drought Area and Severity by Big Data Analysis, Master's dissertation, Graduate school in Hanseo University.
- [9] MOE, MSIP (2015). Human Resource Development Plan for the SW oriented society.
- [10] Park, JinSoo (2014). Take advantage of the training agenda developed by Education Big

Data-Focused on Social Network Analysis, KERIS research report, KR 2014-10.

- [11] Son, JeongEun (2014). A study of the automatic classification technologies for the educational issues big data, Master's dissertation, Graduate school in DongEui University.

저자소개



박 선 주

1995 전남대학교 전산통계학과 (이학박사)
 2003 George Mason University 객원교수
 1996~현재 광주교육대학교 컴퓨터교육과 교수
 관심분야: 컴퓨터교육, SW교육, 앱개발, 빅데이터
 e-mail: sjpark@gnue.ac.kr