

텍스트 마이닝을 이용한 비대면 소프트웨어 교양과목의 요구사항 분석

An Analysis for the Student's Needs of non-face-to-face based Software Lecture in General Education using Text Mining

정화영

경희대학교 후마니타스칼리지

Hwa-Young Jeong(hyjeong@khu.ac.kr)

요약

온라인 수업에 대한 학생들의 니즈 분석은 객관식 설문조사 유형이 주로 수행되어왔다. 그러나 학생들의 정확한 니즈를 분석하기 위해서는 주관식 답변에 의한 비정형 데이터 분석이 요구된다. 빅데이터는 비정형 데이터 분석이 가능하여 다양한 분야에서 활용되고 있다. 본 연구에서는 비대면 온라인 수업방식을 진행되는 교양 소프트웨어 과목에서 학생들이 원하는 과목이나 주제가 무엇인지 조사 및 분석하였다. 실험방법은 학생들에게 주관식 설문조사를 시행하여 얻은 비정형 데이터를 기반으로 빅데이터의 키워드 분석, 연관 분석등을 수행하였다. 이를 통해 학생들이 교양 소프트웨어 과목에서 원하는 키워드가 무엇인지 알 수 있었으며, 이러한 연구 결과는 학생들이 배우고자하는 주제를 파악할 수 있어서 향후 교양 소프트웨어 과목의 기획 및 설계시 중요한 자료가 될 것이다.

■ 중심어 : | 빅데이터 | 텍스트 마이닝 | 학습자 니즈 | 소프트웨어 교육 | 온라인 학습 |

Abstract

Multiple-choice survey types have been mainly performed to analyze students' needs for online classes. However, in order to analyze the exact needs of students, unstructured data analysis by answer for essay question is required. Big data is applied in various fields because it is possible to analyze unstructured data. This study aims to investigate and analyze what students want subjects or topics for software lecture in general education that process on non-face-to-face online teaching methods. As for the experimental method, keyword analysis and association analysis of big data were performed with unstructured data by giving a subjective questionnaire to students. By the result, we are able to know the keyword what the students want for software lecture, so it will be an important data for planning and designing software lecture of liberal arts in the future as students can grasp the topics they want to learn.

■ keyword : | Big Data | Text Mining | Student's Needs | Software Lecture | Online Learning |

I. 서론

포스트코로나 시대의 비대면 교육환경이 지속되면서

다양한 방법의 온라인 학습이 적용되고 있다[1]. 특히 증강현실과 비디오합성 기법을 이용하여 콘텐츠를 온라인상에서 제공할 수 있는 기법[2]들도 제시되고 있

* 이 논문은 2020년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구자지원사업의 지원을 받아 수행된 연구임 (NRF-2020S1A5A2A01042497)

접수일자 : 2022년 03월 04일

수정일자 : 2022년 03월 14일

심사완료일 : 2022년 03월 14일

교신저자 : 정화영, e-mail : hyjeong@khu.ac.kr

며, WhatsApp에 학습 콘텐츠를 적용하는 방안[3] 등 많은 방법들이 제안되고 있다. 그러나 다양한 방법의 적용에도 불구하고 교사와 학습자는 여전히 면대면 학습에 비해 온라인 학습의 이질감을 해소하기 어려웠고, 이를 실증적으로 연구 및 분석하는 자료들이 제시되었다. Potu et al.[4]의 연구에서는 학생들의 학습효과에 대해 코로나 전의 면대면 학습과 후의 비대면 학습에 대해 14개 설문조사 항목을 통해 분석 자료를 제시하였다. Van et al.,[5]의 연구에서는 COVID-19 팬데믹 상황의 비대면 교육에서 교사가 학습자에게 ZOOM을 이용한 인터뷰를 통해 학습자의 의견을 모아 교사의 학습에 반영하는 방법을 제시하였다. 또한 Muthuprasad et al.,[6]의 연구에서는 COVID-19 상황에서 학생들의 인식과 선호도를 조사함으로써 온라인 학습이 참조할 수 있는 기본 자료를 제시하였다. 이와 같이 COVID-19로 인해 갑작스러운 학습 환경의 변화에 기존의 면대면 학습과는 다른 비대면 학습적용으로 인한 효과분석과 학습방법의 제시가 다양하게 연구되고 있다. 그러나 학습자가 실제로 학습하고자 하는 주제가 무엇인가를 제시하는 연구는 매우 드물다. 또한 많은 니즈 분석을 위한 방법으로 객관식 설문항목을 이용하였으나 학습자가 원하는 주제를 파악하기 어렵고 설문항목에 따라 제한된 답안만을 작성할 수 있어서 한계성을 가지고 있다.

본 연구는 학생들의 학습 니즈를 분석함으로써 실질적으로 어떠한 것을 학습하고자 하는지 조사해보고자 한다. 이를 위하여 학생들에게 주관식 설문항목을 제시하고, 학생들이 설문에 대해 별도의 양식 없이 자유롭게 답을 함으로서 학생들의 의견을 수렴할 수 있도록 하였다. 비정형화된 학생들의 답변 분석을 위하여 텍스트 마이닝을 사용하였으며, 분석된 자료에 대해 연관성을 분석 및 제시하였다.

II. 관련연구

면대면 학습에서 비대면 학습으로의 전면적인 전환에 따라 온라인 학습을 통한 효과성, 학생들의 학습 동기 유발, 학생 지향적 학습법 등 많은 방향에서 연구가

진행되어 왔다. 이혜은[7]의 연구에서는 비대면 학습에서 학생들의 학습몰입에 이르는 요인을 설문조사를 통해 제시하였다. 김한나[8]의 연구에서는 비대면 동영상 수업이 강의만족도와 같은 학습효과에 어떠한 영향을 미치는지를 온라인 설문을 통해 조사 후 분석하였다. 홍수민[9]의 연구에서는 온라인 영어 학습 환경에서의 학습 성과에 대한 영향을 조사하는데 설문을 통해 분석한 결과를 제시함으로써 학습실재감과 몰입감의 구조적 관계와 학습효과에 대한 직, 간접적 영향이 있음을 제시하였다. 그러나 많은 연구들이 객관식 설문항목을 기반으로 조사가 이루어져 정형적인 분석에 그치고 있다. 이를 위하여 비정형 데이터 분석이 가능한 빅데이터 기술이 사용된다. 빅데이터는 정형, 비정형 데이터를 처리할 수 있으며 데이터의 생성, 수집, 저장, 관리 및 분석하여 가치를 추출하고 데이터간의 알려지지 않은 정보 간 관계를 찾아내는 지능화 서비스 기술로 알려져 있다[10]. 이러한 이유로 빅데이터는 경영분석, 사회관계망, 마케팅 등의 분야에서 많이 이용되었으며 [11-13], 교육 분야에서도 학습 분석에 활용되고 있다. Yong et al.,[14]의 연구에서는 교사와 학생들 사이의 학습 피드백 분석에 빅데이터를 활용하여 정보를 제공하는 방안을 제시하였고, Amelec et al.,[15]의 연구는 설문을 통해 빅데이터 분석을 활용한 학생들의 학습 스타일 자료를 제시하였다. 특히 Arjumand and Khan[16]에서는 비정형 데이터 처리가 가능한 빅데이터를 이용하여 자연어 처리 모형을 제안하였다. 또한 황홍섭의 연구[17]에서는 포털사이트에 키워드 추출 및 분석을 통해 얻어진 데이터를 기반으로 빅데이터 분석을 사회과 교수학습 모형을 분석하였다. 이와 같이 빅데이터를 활용하여 학습모형 및 효과를 위한 실증적 분석을 하려는 연구는 많이 시도되고 있으나 비정형 데이터를 기반으로 학생들의 학습 니즈를 파악하려는 연구는 미흡한 편이다.

III. 연구 방법 및 과정

1. 연구대상

본 연구는 서울 소재 K대학교에서 진행한 2021년 2

학기 3학점의 소프트웨어 관련 교양과목인 <인간-컴퓨터 중심세계>와 <인공지능과 라이프 3.0>을 수강한 학생들을 대상으로 이루어졌다. 2개 과목의 대상 모집단 전체인원은 149명이며, 성별분포는 남학생 73명, 여학생 76명이고, 학년분포는 1학년 40명, 2학년 42명, 3학년 33명, 4학년 34명이다.

2. 연구과정 및 처리

연구과정은 설문조사 → 데이터 수집 → R 데이터 처리 → 결과분석 및 시각화의 순서로 진행되었다. 설문조사에서, 학생들의 소프트웨어 관련 과목에 대한 요구사항 분석을 위한 질의는 아래와 같으며 서술식으로 질문에 대한 답변을 받았다.

Q1. 교양 소프트웨어 관련 과목에서 추가로 다루거나 필요하다고 생각되는 주제는 무엇인가?

따라서 질문에 대한 서술식 답변들을 위해 R언어로 비정형 데이터 분석을 진행하였다. 설문항목을 통해 얻은 서술식 답변을 모두 통합하여 survey.txt로 저장하였다. 이를 기반으로 텍스트 마이닝을 이용하여 각 단어에 대한 빈도수를 측정하였으며, 단어들의 연관성을 분석하였다. 이때의 연관성 분석을 위한 지지도와 신뢰도는 0.05로 설정하였다. Pre-process는 단어들의 연관성과 시각화를 위한 처리를 나타낸다. 이때 pre_data는 설문 답변들에 대해 명사를 추출한 후 단어들 사이의 연관성 분석을 위해 불용단어와 중복 단어를 제거한 것을 나타낸다.

Algorithm Pre-process

```
input survey.txt
output pre_data
```

Begin

```
install packages for text mining
survey_data ← read "survey.txt"
survey_word ← extract noun
survey_uni_word ← extract unique elements
    from survey_word
survey_fin_word ← delete duplicated words
```

```
from survey_uni_word
word_filter ← extract useless words from the
    survey_fin_word
pre_data ← sapply(survey_fin_word,
    word_filter)
```

End

Association_analysis는 연관성 분석에 대한 알고리즘을 나타낸다. Pre-process에서 처리된 pre_data를 트랜잭션 타입으로 변경하여 연관성 분석을 실시한 후 네트워크 그래프로 나타낸 것이다.

Algorithm Association_analysis

```
input pre_data
```

Begin

```
install packages for an association analysis
word_trans ← change transaction from pre_data
base_word ← apriori(word_trans)
sort_data ← sort(base_word)
show network graph using sort_data
```

End

IV. 연구결과

설문답변에서 ‘수’, ‘는’, ‘이’ 등과 ‘그리고’, ‘그래서’ 등과 같이 분석에서 사용하지 않는 불용 단어를 제거한 후의 빈도수를 확인하면 아래 [그림 1]과 같다. 이를 워드 클라우드로 나타내면 [그림 2]와 같다.

분석을 통해 산업혁명, 인공지능, 발전, 정보, 변화 등의 키워드가 많이 노출되는 것을 알 수 있었으며, 이는 소프트웨어 관련 과목을 교양으로 수강하는 비전공 학생들이 개설을 희망하는 주제들을 알 수 있었다. 빈도수에서 가장 많이 노출된 “산업혁명”을 중심으로 나타낸 연관성은 [그림 3]과 같다. 이때 연관된 단어들은 “비즈니스”, “로봇”, “사물 인터넷”, “기술”, “교육” 등이 있었다. 학생들의 니즈는 4차 산업혁명으로 촉발되는 변화에 대해 대응하고자 하는 것이기에 키워드 “변화”

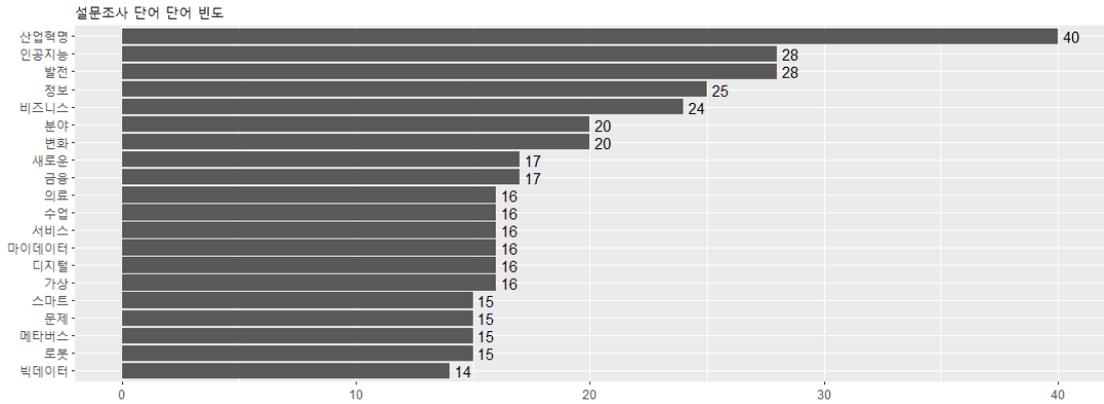


그림 1. 실문조사 단어 빈도수



그림 2. 워드 클라우드 결과

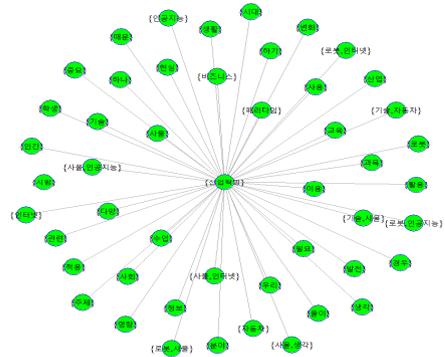


그림 3. “산업혁명” 중심의 연관 네트워크

를 기준으로 연관성을 조사하였다. [그림 4]는 “산업혁명” 키워드를 중심으로 분석된 연관 그래프를 나타낸다. 이때 설정된 지지도와 신뢰도에 따른 연관규칙에 의해 “로봇, 비즈니스”, “패러다임, 비즈니스”, “과목, 시대”, “적용, 수업”, “정보, 병합”, “사회, 생활” 등이 제시됨을 알 수 있었다. 리를 기반으로 “산업혁명” 키워드를 중심으로 보면 학생들은 비즈니스, 패러다임, 정보, 사

회 등으로 연관된 수업을 희망하는 것을 알 수 있다. [그림 5]는 “변화”와 관련하여 “산업혁명”, “패러다

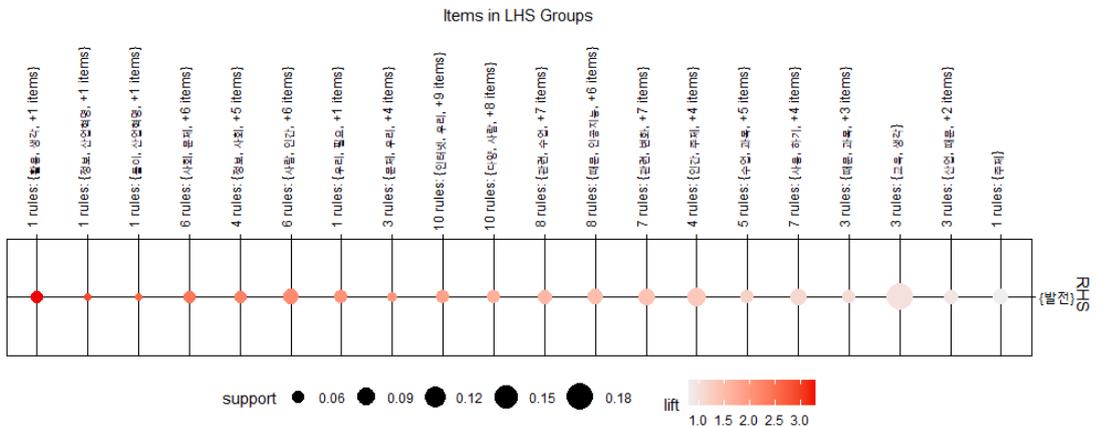


그림 4. “산업혁명” 중심의 연관 그래프

행과 많은 학생 모집단을 대상으로 한 결과를 기반으로 분석이 이루어져야 할 것이다. 또한, 신뢰성을 입증하기 위해서는 더욱 많은 모집단을 통한 데이터 수집이 필요하고, 수집된 자료들은 빅데이터 뿐만 아니라 지능형 기술을 통해 보다 세밀한 분석이 요구된다.

참 고 문 헌

- [1] 이동혁, 김상춘, 박남재, “포스트코로나 시대의 언택트 교육 환경을 대비한 블록체인 기반의 온라인 학습 플랫폼,” 한국정보기술학회논문지, 제18권, 제11호, pp.109-121, 2020(11).
- [2] 이현주, 유선진, “인터랙티브 미디어 기반 체험형 콘텐츠의 제작기법,” 한국정보기술학회논문지, 제18권, 제12호, pp.121-129, 2020(12).
- [3] Herri Mulyono, Gunawan Suryoputro, and Shafa Ramadhanya Jamil, “The application of WhatsApp to support online learning during the COVID-19 pandemic in Indonesia,” Heliyon, Vol.7, Issue8, Aug, 2021, <https://doi.org/10.1016/j.heliyon.2021.e07853>.
- [4] B. K. Potu, H. Atwa, W. A. Nasr El-Din, M. A. Othman, N. A. Sarwani, A. Fatima, A. Deifalla, and R. A. Fadel, “Learning anatomy before and during COVID-19 pandemic: Students’ perceptions and exam performance,” Morphologie, Available online 29 July 2021, <https://doi.org/10.1016/j.morpho.2021.07.003>.
- [5] Van ThinhLe, Ngan Ha Nguyen, Tran Le Nghi Tran, Luan Thanh Nguyen, Thi Anh Nguyen, and Minh Trang Nguyen, “The interaction patterns of pandemic-initiated online teaching: How teachers adapted,” System, Available online 29 January 2022, 102755, <https://doi.org/10.1016/j.system.2022.102755>
- [6] T. Muthuprasad, S. Aiswarya, K. S. Aditya, and Girish K. Jha, “Students’ perception and preference for online education in India during COVID -19 pandemic,” Social Sciences & Humanities Open, Vol.3, Issue1, 2021, 100101, <https://doi.org/10.1016/j.ssaho.2020.100101>.
- [7] H. U. Leen and T. U. Jeong, “Relationship between Academic Burnout and Learning Immersion in Non-face-to-face Learning Situations: The Mediation Effects of Meta-cognition Strategy and Emotional Support,” Asian Journal of Education, Vol.22, No.3, pp.641-663, Sep. 2021.
- [8] 김한나, 김성원, “코로나19로 인한 비대면 동영상 수업의 강의만족도가 학습효과에 미치는 영향과 자아효능감의 매개 효과 연구,” 학습자중심교과교육연구, 제21권, 제1호, pp.363-387, 2021.
- [9] 홍수민, 이정민, “온라인 영어학습공동체에서 학습실재감과 학습몰입이 학습성과에 미치는 영향,” 외국어교육연구, 제35권, 제2호, pp.53-70, 2021.
- [10] 이수연, 김태중, “토픽 모델링을 활용한 ‘수돗물 유충’ 뉴스 빅데이터 분석,” 한국콘텐츠학회논문지, 제20권, 제11호, pp.28-37, 2020(11).
- [11] T. M. Song and J. Y. Song, “Prediction of risk factors of cyberbullying-related words in Korea: Application of data mining using social big data,” Telematics and Informatics, Vol.58, May 2021, 101524, <https://doi.org/10.1016/j.tele.2020.101524>
- [12] Alexandra Amado, Paulo Cortez, Paulo Rita, Sérgio Moro, “Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis,” European Research on Management and Business Economics, Vol.24, Issue 1, pp.1-7, January-April, 2018.
- [13] Xiao-Yuan Liu, “Agricultural products intelligent marketing technology innovation in big data era,” Procedia Computer Science, Vol.183, pp.648-654, 2021.
- [14] Yong Cui, Xiao Song, Qinglei Hu, Ying Li, A. Shanthini, and Thanjai Vadivel, “Big data visualization using multimodal feedback in education,” Computers & Electrical Engineering, Vol.96, Part A, Dec. 2021, 107544, <https://doi.org/10.1016/j.compeleceng.2021.107544>
- [15] Amelec Viloría, Ingrid Regina Petro Gonzalez, and Omar Bonerge Pineda Lezama, “Learning Style Preferences of College

Students Using Big Data,” *Procedia Computer Science*, Vol.160, pp.461-466, 2019.

[16] Arjumand Masood Khan and Khan Rahat Afreen, “An approach to text analytics and text mining in multilingual natural language processing,” *Materials Today: Proceedings* Available online 2 Jan 2021, <https://doi.org/10.1016/j.matpr.2020.10.861>.

[17] 황홍섭, “빅데이터를 활용한 사회과 교수·학습 모형의 탐색,” *사회과교육*, 제58권, 제1호, pp.63-98, 2019(3).

저 자 소 개

정 화 영(Hwa-Young Jeong)

정회원



- 1994년 2월 : 경희대학교 컴퓨터공학과(공학석사)
- 2004년 8월 : 경희대학교 컴퓨터공학과(공학박사)
- 2005년 4월 ~ 현재 : 경희대학교 후마니타스칼리지 교수

〈관심분야〉 : 이러닝, U-러닝, SCORM, 자기주도적 학습, 학습 시스템