http://www.jsebs.org
ISSN: 2288-3908

The Journal of Society for e-Business Studies
Vol.22, No.4, November 2017, pp.193-214
https://doi.org/10.7838/jsebs.2017.22.4.193

토픽모델링을 활용한 소프트웨어 분야 대학 교과과정 분석

A Comparative Analysis of Curriculums for Software-related Departments based on Topic Modeling

최재원(Jaewon Choi)*, 이 호(Ho Lee)**,
김정민(Jungmin Kim)***, 송주호(Juho Song)****

초 록

소프트웨어 직무의 특성과 대학교의 SW 학과가 개발한 커리큘럼 간의 차이가 존재하는 현재에서, 실제로 SW 관련 교과과정 편성이 어떻게 구성되어 있는지, 그리고 현실적으로 SW 인력의 실무 요구사항과 부적합한 부분을 확인하는 것이 매우 중요한 시기이다. 해외 사례를 보면 이미 실무 요구사항 중심의 SW 교육을 바탕으로 SW 인력을 양성하려는 노력이 진행되어 오고 있다. 그 결과 실제 관련 실무 기업들의 채용에 대한 긍정적인 반응이 나타나고 있다. 국내에서도 정부 주도하에 이러한 시도가 시도되고 있으며 특히 SW 중심대학 사업을 바탕으로 관련 분야의 선도대학의 역할을 부여하고 있다. 그러나 SW 분야의 인력 공급 문제는 여전히 실무분야와 교육 분야의 이슈가 되고 있다. 교과과정 구성에 대한 실무 기업들의 관점의 환경적 요소가 확실히 반영되지 못하고 있다는 기존의 한계에 따라 본 연구에서는 교육 내용의 구성과 실무의 차이를 줄일 수 있는 방법을 진단하고자 하였다. 그에 따라 실제 활용중인 대학의 교과과정과 강의계획서 자료를 바탕으로 토픽모델링을 실시함으로서 교과과정과 강의계획서에 대한 키워드를 도출하였다. 분석 결과 분석에 활용된 관련 대학 학과의 실습 비율이 상대적으로 낮은 수준을 보였으며 교과목 중첩비율, 강의계획서 키워드 중첩비율도 일반 수준으로 보여짐에 따라서 체계적인 교과과정 확립과 실무 능력 배양을 위한 강의계획 수립이 중요하다는 것을 확인하였다.

ABSTRACT

It is a very important time to check how SW curriculum is actually organized and what is inadequate to practical requirements of SW manpower in the present situation where there is a difference of viewpoints between software field and SW curriculum of university. In overseas cases, efforts have already been made to cultivate SW manpower based on

본 연구는 순천향대학교 학술연구비 지원으로 수행하였음.
이 논문 또는 저서는 2014년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2014S1A3A2044046).
* First Author, Soonchunhyang University(jaewonchoi@sch.ac.kr)
** Co-Author, Software Policy & Research Institute(leeho32@spri.kr)
*** Corresponding Author, Software Policy & Research Institute(jungmink26@spri.kr)
**** Co-Author, Korea Software Industry Association(juho@sw.or.kr)
Received: 2017-11-06, Review completed: 2017-11-24, Accepted: 2017-11-28

SW training centered on practical requirements. As a result, there is a positive response to the recruitment of actual related companies. In Korea, these attempts have been attempted under government initiative. In particular, based on the SW-centered university project, it has given the role of a leading university in related fields. However, with regard to the labor supply problem in the SW sector, the requirements of the business enterprises still differ from the educational curriculum. In this study, we tried to diagnose the method that can reduce the difference between the composition and the practice of the contents according to the existing limit that the environment factor of the viewpoint of the working companies about the curriculum composition is not clearly reflected. As a result, the topic modeling based on the university's curriculum and lecture plan data is used to derive keywords for curriculum and lecture plan. Through the data analysis, this study confirmed that the practice rate of related university departments utilized in data analysis is relatively low. In addition, we found that it is important to establish a systematic curriculum and to build a lecture plan to cultivate practical skills, as the number of overlapping textbooks and the number of keyword overlapping are found.

키워드 : 토픽모델링, SW 인력, SW 교육, 교과과정 분석, 강의계획서 분석, LDA
Topic Modeling, SW Manpower, Curriculum Analysis, Class Syllabus Analysis, LDA

1. 서 론

1.1 연구 배경 및 목적

IT의 발달로 인하여 21세기 현대 사회에서는 컴퓨터의 하드웨어와 소프트웨어의 사용은 남녀노소 사용가능한 환경으로 변화했다. 1960년대부터 시작된 컴퓨터의 발달과 인터넷의 발달은 SW 교육의 시초가 되었다. 1970년대 FTP(File Transfer Protocol) 전송 시스템의 개발 및 1983년 인터넷 보급과 HTML 문서의 개발 그리고 1인 PC에 대한 보급으로 하드웨어를 넘어 소프트웨어(Software, 이하 SW)는 더욱더 중요한 부분이 되었다.

대학에서도 컴퓨터 관련 과목과 교육훈련이 진행되고 있으며 각 종합대학에 컴퓨터 관련 전공을 개설하여 SW 개발자 및 컴퓨터 전문가를 배양하고 있다. 그러나 교육훈련과정을 통하여 SW 역량을 갖춘 전문가들이 활동을 하고 있지만 산학간의 이론과 실제의 차이가 존재하고 있다.

'삼성의 SW 기술 반성'[1)]이라는 사설에서 한국의 SW 인력의 교육 현황은 실제 현업에서 활용하기에 부족하다는 점을 시사하고 있다. 해외에서는 '인스타그램'이라는 어플리케이션을 4명이 6주 만에 만들었던 것에 비하여 삼성이 만들었으면 몇 백 명이 붙어 1년이 걸렸을 것이라고 언급 하고 있다.

본 연구에서는 SW 교육을 실시하고 있는 대표적인 국내 대학을 대상으로 전반적인 교과목 구성과 환경을 파악하고 강의와 교과과정에서의 측면과 현업에서 원하는 역량의 매칭 정도를 확인하고자 실시하였다. 나아가 추후 SW 교육에 있어 교과과정만을 이수하고 바로 현업에서 업무를 할 수 있는 환경을 만들기 위한 제언도 함께하는 것에 본 연구의 의의가 있다.

1) 삼성의 반성…"SW 인력 절반이 기초수준 실력," 김현석 기자, 한국경제, 2016. 06. 22.

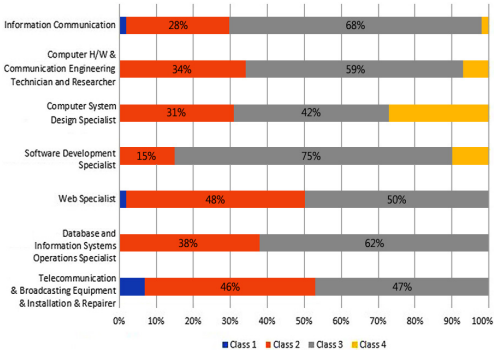
정보통신의 기술을 포함한 여러 가지 학문은 과거와는 달리 범학제적 접근을 통한 새로운 통합적 측면에서의 교육을 진행하고 있다는 점에서 다른 교과과정들과 융합을 하여 사물인터넷, 클라우드, 빅데이터 등의 여러 가지 키워드와 함께 도래하고 있다[3].

4차 산업 혁명도 정보통신의 고도화된 발전으로 인한 자동화와 연결성이 극대화 되어 사물을 연결하고 사람을 연결하는 산업의 혁신이 이뤄진다. 인공지능의 발달로 로봇과 사람의 공존 문제가 야기되며 정보통신 기술을 제어 및 통제 할 수 있는 능력을 보유할 수 있는 SW 교육이 얼마나 중요한지를 최근 변화들을 통하여 증명하고 있다[6].

SW의 사회적 관심이 증가함에 따라 SW 관련 연구영역은 많은 부분 진행되고 있으나 핵심어(Keyword, 이하 키워드)를 통한 분석을 바탕으로 산업계의 요구 영역과 대학의 전공교육을 연계하기 위한 분석적 접근의 연구는 이뤄지지 않고 있다. 텍스트 마이닝을 통하여 산업계의 요구사항과 대학의 교육 형태에 대한 키워드를 기반으로 현재 운영되고 있는 SW 교육과 관련된 정책적 측면이 적절한지 혹은 교육을 시행한 후에 바로 현업에 투입되어 역량을 발휘 할 수 있는 것인가에 대한 질문에서 본 연구의 동기가 존재한다.

SW 산업계의 문제 중에 하나는 SW 인력의 부족이다. SW 관련 학과 학생들이 대학 졸업 이후 자신의 진로를 고민하고 나아가는데 있어 다양한 개선사항이 존재하고 있다. 한 신문사설[2]에서는 '우리나라에는 고급 SW 개발자가 없다.'라고 언급한다. SW 개발자는 수명이 짧아 10년이면 끝이 나고 설령 고급 개발자가 되었다고 하더라도 업무 과중 및 수행할 과업의 정도가 과중되어 기업에서 오래 버티지 못하고 있는 것이 현업의 실태이다. 문제해결을 위한 정책적 접근에서도 이론과 실제의 간극을 줄이기에는 아직 부족한 요소가 많이 존재한다.

현재 SW 인력과 관련된 분야의 미충원 인원은 지속적으로 발생하고 있으며, 수준에 따른 미충원 인원의 분포는 <Figure 1>과 같다.



Note) 1. Company scale(over 5 persons)
2. Under-recruited personnel: Persons who were not recruited despite active recruitment(recruiters-recruiters)
3. Class 1: Untrained, no education, no license
Class 2: 2 years of on-site experience or national technical qualifications
Class 3: college/master level for Each Field
Class 4: 10 years of field experience or national technical qualifications or PhD Level

〈Figure 1〉 Level of SW Personnel Unemployed Distribution〔13〕

상기 차트를 통하여 특히 2, 3수준인 즉, 2년~10년 정도의 경력자의 수요 대비 미충원 인력이 많다는 것을 발견하였으며 SW 관련 전공을 한 다수의 졸업생의 진로 선택에 있어서 실제적 접근이 잘 이뤄지지 않고 있다.

2) 고급 SW 엔지니어가 없는 이유, 최성운 교수, 디지털타임스, 2016. 11. 13.

미충원 요소들을 해결하기 위해서는 SW 인력을 교육하는 대학교육의 재점검과 진단을 통하여 실제 현장에서 사용가능한 기술을 교육하여 결과적으로 이론과 실제의 간극을 줄여야 한다.

따라서 본 연구의 질문은 다음과 같다.

첫째, 대학 SW 관련 학과 대표를 선정하여 현재 교과과정이 어떻게 구성되어 있으며 강의계획서의 내용은 적절한가를 진단한다.

둘째, 각 대학의 과정 간 유사도 정도와 과정 내부의 중첩도를 파악하여 교과과정의 타당성에 대한 진단을 실시한다.

셋째, 토픽모델링을 통한 키워드 분석으로 각 과정들이 갖고 있는 내용들이 적절하게 배치되어 있는가를 진단한다.

본 연구에서는 전국의 SW 학과가 있는 대학 중 지역적 대표성과 정부 정책의 유무에 따른 대표성을 갖은 8개 대학을 대상으로 하여 커리큘럼 분석 및 강의계획서 분석, 해당 키워드를 파싱하여 토픽모델링(LDA; Latent Dirichlet Allocation) 기법을 이용하여 텍스트 마이닝을 실시하였다.

2. 이론적 배경

2.1 해외 주요 국가 SW 교육 동향

영국을 포함한 주요 국가에서는 SW 교육이 대학뿐만 아니라 정부 차원으로 초중고 학생들까지 컴퓨터 사고 능력 배양 및 문제해결 능력을 위한 SW 교육을 실시하고 있다. 어린이부터 대학교 및 대학원까지 SW 교육의 커리큘럼 개편 및 심화로 글로벌 사회에서 SW 교육은 필수적인 교육이 되었다. <Table 1>는 SW 인력 교육을 대표적으로 수행하고 있는 국가의 사례이다.

주요 국가에서는 다양한 방법으로 SW 교육을 실시하고 있으며 SW 전공과정에 있는 대학생뿐만 아니라 영유아까지 포함한 정책들을 시행하고 있다. 정책적으로 교과과정을 통한 실제적인 SW 인력을 선발 할 수 있는 인재양성을 하는데 다양한 방법으로 지원을 하고 있으며, 학교와 기업의 연계를 통하여 교육까지 진행하고 있다.

〈Table 1〉 The Examples of SW Education Policy of Major Universitiees

Country	Policy	Contents
UK	• Key stage configuration by age • Coding school operation out of regular curriculum • Government-level coding contest • Elaborating SW courses during college entrance examination	• Key 1(5~6 age) Understanding the algorithm Key 2(7~11 age) Design-coding-modification Key 3(11~14 age) Using more than one language Key4(14~16 age) Designing Program • After-schooling by Volunteer
US	• Elaborating education for leading SW cities • Practical SW Education using various programs • Linkage Programs between High school and Business field	• Selecting a leading city and policy • Elaborating high school-level CS coding education
Finland	• Encourage government-level coding education from age 4 • Event of Children SW Education with Leading University	• Establishment of Cody Colou(Code School) • Collaborate with SW leading universities such as Helsinki University to operate SW classes and integrate media

'Code.org'와 같은 온라인상에서 쉽게 접할 수 있는 코딩 교육 사이트를 통하여 보다 쉬운 코딩 접근으로 SW 비전공생 들에게도 SW 교육을 원활하게 할 수 있는 콘텐츠가 생겨나고 있어 귀추가 주목된다. SW 관련 전공을 하는 학생들을 대상으로는 SW 교육을 직접 할 수 있도록 여건 마련과 초・중・고등학교와 연계한 SW 교육을 실시함으로써 시너지 효과를 낼 수 있는 환경을 조성하고 있고, 대학에서만의 역량이 아닌 사회 전체적인 SW 인력 교육의 역량강화에 집중하고 있는 추세이다.

2.1.1 하버드대학교 CS50

최근에 가장 각광 받고 있는 강의 중 하나인 'CS50'은 하버드 대학교에서 진행되는 온・오프라인 겸용 강좌로 구성되어 있다. CS(Computer Science) 강좌를 통하여 학생들은 문제를 해결하는 방법을 가장 효율적으로 배울 수 있다. 본 강의는 남녀노소 누구나 쉽게 들을 수 있도록 온라인 강좌를 겸비하고 있어 구지 하버드에 직접 가지 않아도 하버드 대학교에서 배울 수 있는 강좌를 들을 수 있다. 강의를 수강하고 이수함으로써 이수 증서를 제공하여 수업을 듣는 것에서 끝나지 않고 보다 수강자에게 자부심을 줄 수 있다.

<Figure 2>의 워드 클라우드는 1984년부터 현재까지 많은 학생들이 검색한 값이며, 가장 눈에 들어오는 단어는 Computer/IT/Engineering, Scientific Research, Education Engineering 등이 중요 키워드로 작용하고 있어 현재 SW에 많은 관심을 두고 있는 것을 알 수 있다.

미국의 경우 CS50를 통한 수업 등 MOOC (Massive Open Online Course) 강의 형태를 제작함으로써 접근성을 높여 SW의 진입장벽을 허물었다.

〈Figure 2〉 Fields in Which Alumni Since 1984 Can Now be Found, Unofficial Guide to Computer Science at Harvard 2014

CS50은 전공자뿐만 아니라 비전공자들도 쉽게 들을 수 있도록 단계별 난이도로 구성된 강의 들을 제공하는 것에 의의가 있다. 한국형 CS50 도입 방안(2016)에서 언급된 내용 중에 전체 수강생 중 78%의 학생은 컴퓨터 프로그래밍 전공을 하지 않는 일반인이라는 점이 매우 독특하다는 것을 알 수 있으며 MOOC으로 제공을 하고 있어 온라인에서 언제든지 강의를 들을 수 있는 편리한 시스템으로 구성 되어있다.

CS50을 수강하고 난 후에는 각각의 특성화된 전공 들을 들을 수 있으며 전체 교육과정의 기반이 되는 것이 바로 CS50이라는 것을 확인할 수 있다. 특히 기본적인 교육 과정을 제공함으로써 많은 비전공자 및 입문자 들이 부담 없이 SW 교육을 학습할 수 있는 우수 사례로 볼 수 있어 차후의 귀추가 주목된다[7].

2.1.2 스탠포드 대학교 CS101

하버드 대학교와 유사하게 스탠포드 대학에서도 CS101부터 106, 193 등과 같이 커리큘럼을 마인드맵처럼 도식화하여 해당 전공자가 SW 교육을 어떻게 받고 어떤 방향으로 자신

이 학습을 해야 하는지 정보를 제공하고 있다. 전공자가 아니더라도 CS101의 강의를 통하여 SW 입문 교육을 받고 전공을 결정 할 수 있도록 정보와 학습 자료를 제공함으로써 보다 현실감 있고 직관적으로 제공하고 있다는 점에서 국내 적용의 필요성이 시사된다.

상기 언급한 해외 두 대학의 CS 교육의 공통점은 온라인 Mooc을 활용하여 미국에서 진학하지 않더라도 강의를 수강 할 수 있다는 점에서 큰 장점이 있다. 또한 직관적인 커리큘럼의 소개를 통하여 수강자의 입장에서 학습계획을 세우는데 있어 쉽게 적용이 가능하다는 점이다.

2.2 국내 SW 교육 동향

미래창조과학부(이하 미래부)에서 2014년 발표한 소프트웨어 중심사회 실현 전략 중 하나의 전략으로 소프트웨어 중심대학을 만드는 움직임이 거세게 일어났다. 최근 스마트폰의 등장과 함께 SW가 개인・기업・정부의 경쟁력을 결정하는 시대가 도래함에 따라 전 세계적으로 지능정보사회로의 급속하게 전환 및 성장하고 있다.

과거 SW의 인식은 하드웨어의 부속물로써 가치가 있었던 반면 현재는 하드웨어의 중요성 보다 무형의 소프트웨어의 중요성이 높은 가치평가를 받고 있다. 미래부 및 교육부에서는 관계합동으로 창의인재 교육을 실시하여 초・중등 SW 교육 필수화와 SW 중심 대학을 선정하는 사업을 진행하고 있어 SW 교육을 통한 지능정보사회가 어떻게 나아갈지 방향에 대한 귀추가 주목 된다.

SW를 교육하고 있는 대학의 SW 관련 학과를 중심으로 실습을 통한 실전경험을 강화하는 교육을 진행하고 있다. 산업계의 눈높이에 맞춰 사회에서 필요로 하는 인력을 배출하는 것을 목표로 시행하고 있다.

SW 중심사회로 가기 위한 4가지 핵심 과제 중 '미래형 창의인재 양성'을 통하여 미래부는 SW 중심이 되는 SW 인재를 양성하는 정책을 시행 및 예정 중에 있다. 세부적으로 초・중등 및 대학으로 분류하여 각자 3~4가지 핵심 추진체계를 갖고 진행되고 있으며 SW 교육의 기본부터 심화까지 다양한 필요에 맞추어 추진하고 있다.

SW 인재양성을 위한 추진과제 중 대학 SW 교육은 4가지의 세부 사항으로 나뉘어 진행되고 있다. SW 교육과 인재 양성을 위한 저변을 구성하고 고급 연구 인력을 배양하기 위한 SW 중심대학의 선정까지 체계적으로 추진과제를 내놓으면서 SW 교육을 수준별로 체계적으로 구성하였다.

그러나 아직 SW 교육 정책에 관련하여 실제적인 결과는 도출되지 않은 상황이다. 현업에서는 SW 인력의 부재를 지속적으로 감내하고 있으며 수요에 대한 공급이 적은 것이 현실이다. 준비된 SW 인력을 구인하기는 교육과 현업에서 사용하는 내용들의 간극이 상대적으로 크게 존재하기 때문에 많은 기업에선 SW 인력 충원에 노력하고 있다. 따라서 본 연구에서는 SW 대학들의 강의 계획서 분석은 매우 가치 있으며, 분석을 통하여 간극을 메울 수 있는 제언을 위해서도 반드시 필요한 절차이다.

SW 중심사회 핵심 추진과제에서 SW 대학에 대한 정책을 세부적으로 살펴보면 SW 교육은 이공계열에 국한하지 않고 인문, 사회 계열과 융합 인재를 위한 세부 추진 전략들이 도

출되고 있다. 이공계는 보다 심화된 교육으로 SW 인재를 키우는 방향으로 인문・사회의 경우 기존의 교양에서 SW를 추가하는 방향으로 융합은 X-마인드로 새로운 목표 설정 및 실행할 수 있는 여건을 제공하고 있다.

SW 중심대학의 목적은 21세기에 필요한 역량을 위해 글로벌 흐름과 산업 현장의 수요를 반영하고 고급・융합 SW 인재를 육성하기 위함이다. K-ICT 전략의 성공적인 진행과 SW중심사회를 선도할 수 있는 문제해결능력을 갖춘 인재양성 및 SW 교육 선도를 하여 정보화 사회의 중심 인력을 찾는 것에 주요 목적이 있으며 SW 인력은 앞서 언급한 만큼 중요한 자산이다.

SW 중심대학은 대학교육을 SW 산업계 필요에 맞게 교육을 함으로써 보다 현장에 가까운 SW 교육을 진행 할 수 있는 대학 인프라 구축 및 인재 양성 기반 확보를 목표로 하고 있다.

기존의 교과 과정 및 교육 체계를 개편함으로써 SW의 중심으로 대학 교과 과정을 꾸릴 수 있는 선도 대학을 만들어 SW 중심 사회로 이끌어갈 SW 인재 양성에 보다 현실적 접근성을 할 수 있도록 구성하였다.

교육 체계는 기존에 있었던 SW 관련 학과에게서 볼 수 있던 교육 체계를 넘어서 대학차원으로 혁신을 통한 교육 체계 변환과 좀 더 전문성 있고 실전적인 교수진 배치로 학부생들이 좀 더 실제적인 프로그래밍 실습이나 취업 준비를 할 수 있는 인재 양성에 주목하였다.

SW 중심사회의 일환으로 한국의 ICT 즉, K-ICT 전략의 성공적인 추진을 위하여 SW 인재 양성에 더욱더 많은 투자를 계획하고 있다. SW 중심대학은 15년 8개 선정을 시작으로 19년도에 20개까지 선정하여 SW 선도 대학 그룹을 형성하여 전국 대학의 우수 사례로 SW 인재 양성의 모범이 되도록 진행하고 있다.

SW 중심대학의 궁극적인 목적은 지속적인 산업계와의 요구 적합도를 통하여 바로 현장에 투입 할 수 있는 인력을 양성하는 것이 목표이다. 삼성연구소의 연구 결과에 따르면 15년도 기준 산업 현장과의 적합도는 42%로 매우 낮은 수치에 머물고 있으며 현재도 아직 적합도에 있어서 매우 미흡한 실정이다[17].

4년 교과 과정을 통하여 많은 내용을 학습하지만 정작 산업계에서 필요한 부분에 대한 역량 강화가 되지 못하여 산업 현장에 나가서 다시 배우는 일이 지속적으로 일어나고 있다. 결과적으로 교과 과정과 산업 현장의 수요의 매칭이 현재 매우 필요한 실정이며 SW 대학들의 교과 편람 및 강의 계획서를 분석해야 하는 논리적 근거이기도 하다.

SW 중심대학을 SW 중심사회를 통한 지능정보사회 구현을 위하여 미래부에서 발표한 보도 자료에 언급 되어있다. 2015년을 기점으로 1기를 시작하여 16년 2기 대학을 선정하여 운영하고 있다. 2015년 8개 대학이 선정되었고 16년 6개 대학이 추가로 선정되어 총 14개 대학이 현재 운영되고 있다.

SW 중심대학의 공통점을 살펴보면 대입 시 SW 특기자를 선발하고 있으며, 기존에 있었던 SW 관련 전공 학과에 대한 편제 및 교육과정 강화를 통하여 전문역량을 키울 수 있는 SW 학과를 만들기 위한 노력을 진행하고 있다.

또한 이수 강의 이외에 다양한 행사와 과외 교육 프로그램 등을 통하여 역량 강화를 시킬 수 있는 환경을 학생들에게 제공함으로써 기본 교육에서 부족한 SW 관련 내용을 학생들이 수강하고 자유롭게 학습 할 수 있는 환경을 만들고자 하는 노력을 지속하고 있다.

2.3 토픽모델링(LDA)

텍스트 마이닝은 수치형으로 분석 가능한 정형 데이터가 아닌 문자와 텍스트로 구성되어있는 비정형 자료를 분석하는 기법이다. 텍스트 마이닝 기법 중에 학술연구에서 자주 사용하는 분석 방법은 토픽모델링, 오피니언 마이닝, 시멘틱 웹 및 온톨로지 기법들이 존재하며 모두 단어를 이용한 분석을 기초로 한다[11].

기존 교육과정과 커리큘럼에 대한 연구들을 보면 주로 학습자 관점에서 필요로 하는 심리적 태도 및 자아존중감 등 인지적 응답에 의거한 영향관계를 주로 연구해오고 있다. 그러나 커리큘럼의 경우 실제 수업의 내용을 계획하는 문서인 만큼 학습자의 인지적 상황을 반영하는 것이 아닌 본 연구의 목적에 따라 올바른 교육과정을 위한 커리큘럼 구성이 진행 중인지 확인하는 것이 필요하다. 특히 주로 문헌정보학 관점의 강의계획서 분석 방법론만을 다루어 온 점에서[16] 소프트웨어 분야의 확장 응용 가능성을 확인할 필요가 있다. 그에 따라 다수의 글로 정리되어 있는 강의 계획서를 분석하기 위해서는 문서 및 단어 기반으로 콘텐츠의 구성을 확인할 수 있는 토픽모델링을 활용하는 것이 본 연구의 목적을 수행함에 있어 적절한 것으로 판단된다.

본 연구에서 사용되는 토픽모델링은 문서에서 특정 주제를 추출하는 알고리즘을 구성하여 확률적 기법의 기반으로 문서가 어떤 토픽들을 갖고 있는지 어떤 단어끼리의 연관성이 있는지 확인하는 알고리즘이다[2]. 토픽모델링에는 잠재 의미 분석과 잠재 디리클레 할당(LDA; Latent Dirichlet Allocation)이 있으며 그 중 잠재디리클레분석을 실시하고자 하였다. LDA는 기존의 문서와 사전에 정의된 문서 내에 토픽들의 확률분포를 통하여 추정하는 기법으로 학술연구에서 많이 사용되고 있다[11].

자연어 처리 기반 텍스트 마이닝의 주요 기술로는 자연어 처리(파싱, 형태소 분석, 품사태깅, 관계 추출, 의미 추출), 언어모델링(언어감지, 규칙기반 개체명・상용어 인식), 기계 학습 알고리즘(반복훈련을 통하여 습득한 정보 사용능력을 개선), 마이닝 기술(각종 통계적 기법을 활용한 정보 분류 및 분석 기술・기법) 등이 있다[14]. 특히 자연어 처리 기반 텍스트 마이닝 기술을 구현하기 위해서는 자연어 텍스트를 컴퓨터가 이해할 수 있는 요소로 구조화하여 문장 수준의 텍스트에서 의미를 추출해야 한다[4, 8, 12].

텍스트 마이닝 과정에서 가장 먼저 해결해야 할 부분은 문서에 대한 탐색이다. 이를 바탕으로 텍스트 분석은 문장을 형태소, 즉 명사, 조사, 부사, 동사원형 등 문법상 의미를 최소단위로 분해(파싱)하여 주어, 목적어, 부사, 동사 등의 언어학적 구조를 패턴화하고, 문장의 의미를 파악하는 것이 중요하다. 파싱은 단순하게 단어를 매칭하는 것이 아닌 문장의 본 의미를 파악하는 것이 비구조화 데이터를 분석하는데 매우 중요하다.

텍스트와 같은 비구조화된 자료를 분석하기에 적합한 비구조화 자료를 구조화된 자료로 변환 하는 것으로 문서는 단어의 집합으로 이루어져 있다는 개념을 이용하여 행렬 형태로 나타내었다. 각 조건의 문서와 단어 간의 상호관계로 이루어진 단어-문서 행렬(term-document matrix)을 통해 각 셀 안에 문서에 포함된 단어의 빈도를 도출한다. 문서 자동분류를 위해서 문서를 대표하는 단어와 이것의 가중치를 결

정하는 방법을 활용하여 가중치의 계산에 단어의 빈도를 사용하는 용어빈도 가중치를 활용한다. a_{ij}는 문서 j에서 단어 i의 가중된 도수를 의미하며 이를 계산하기 위해서는 G_i는 용어 i의 가중치와 L_{ij}는 문서 j와 용어 i의 빈도 가중치를 활용해야 한다.

$$a_{ij} = G_i L_{ij}$$

빈도 가중치와 용어 가중치를 결합하면 문서 내, 문서 간 정보를 획득할 수 있다.행렬 안에 들어가는 값들은 빈도 가중치와 용어 가중치의 곱으로 계산된다. 비구조화된 텍스트 문서를 일련의 숫자 배열로 구성된 구조로 변환하면 대부분이 0으로 값이 도출되며(sparse data) 실제 자료 크기가 매우 크다는 특징이 존재한다. Matrix에서 row의 차원을 줄이기 위해서 그 해결 방안으로 변환(Transformation)을 하게 된다.

빈도 가중치는 특정 문서가 전체 문서 중에서 차지하는 정보를 반영하지 못한다. 즉, 분석에 이용된 문서의 크기가 클수록 비교적으로 작은 문서보다 훨씬 큰 빈도(frequency)를 도출해낼 가능성이 증가하는 것으로 지역적 가중치(local weight)라고도 한다. 이에 대한 해결책으로 단어 가중치를 동시에 고려한다. 빈도 가중과 방법은 다음과 같다.

$$\text{Binary: } L_{ij} = \begin{cases} 1 \text{ if } term\ i\ is \in document\ j \\ 0\ otherwise \end{cases}$$

$$\text{Log: } L_{ij} = \log_2(a_{ij}+1)$$

$$\text{None: } L_{ij} = a_{ij}$$

a_{ij} = 문서 j에서 용어 I가 나타낸 횟수

L_{ij} = 문서 j와 용어 I의 빈도 가중치

일반적으로 빈도 가중치 함수는 단순 빈도를 그대로 사용하거나 단순 빈도를 로그 변환 또는 이항 변환한 형태를 사용한다. 빈도 가중치 함수에서 문서 j에 포함되어 있는 단어 i의 가중치를 표현하며 $L_{(i,j)}$로 나타낸다. a_{ij}는 분석한 문서 j에서 단어 i가 나타나는 빈도를 의미하며 도출 된 빈도 수 자체를 빈도 가중치의 함수로 사용하기도 한다. 로그 값으로 변환한 빈도 가중치 함수는 a_{ij}의 값이 클 경우 이를 감소시켜 주는 역할을 이행한다. 이항 변수 개념을 활용한 지역적 가중치대한 함수 값은 문서 j에 단어 i가 한 번이라도 나타나는 경우를 1 값으로 적용하고, 한 번도 나타나지 않는 경우를 0의 값으로 할당하는 방법을 사용하여 단어의 수가 많지 않은 경우 보편적으로 사용한다.

용어 가중치와 관련된 함수에는 엔트로피 개념을 이용하여 생겨난 전역적 가중치 함수(엔트로피), 전체 문서에서 단어의 빈도 수와 단어가 나타난 문서 수의 비율을 비교 분석한 전역적 가중치 함수(GfIdf), 단어가 갖고 있는 정보를 양적으로 표현한 전역적 가중치 함수(Idf), 표준화 개념을 사용하여 도출된 전역적 가중치 함수(정규)가 있다. p_{ij}는 문서 j에서 단어 i가 나타난 빈도를 i가 나타난 문서의 수로 나누어 계산한다. 단어 d_j는 단어 i가 나타난 문서의 수, g_i는 전체 문서 모음에서 단어 i가 나타난 빈도, N은 전체 문서의 수를 의미한다. 엔트로피 개념을 활용한 지역적 가중치는 전체 문서 안에서 단어의 빈도가 드물게 도출될 때 높은 값을 가진다. 전체 문서에서 특정 단어가 정확하게 한 번씩 도출된다면 해당 단어는 각 문서를 구분해 주는 특징이 없다. 상기의 경우, 엔트로피 개념을 활용한 지역적 가중치는 0의 값을 갖게 되고 문서 전체 중 단 하나의

문서에서만 특정한 단어가 나타나게 된다면 1의 값을 갖는다. 문서들의 모음 안에서 한정된 소수의 문서에 존재하는 단어만을 강조하고 있다는 특징을 갖으며 GfIdf도 엔트로피 개념을 이용한 지역적 가중치와 똑같은 특징을 갖게 된다. 표준화 개념을 이용하고 있는 전역적 가중치 함수는 문서들의 모음 안에서 존재하고 있는 단어의 비율을 표현한다.

Entropy:

$$G_i = 1 + \sum_{j=1}^{d_i} \frac{p_{ij}\log_2(p_{ij})}{\log_2(n)},\ 0 \le G_i \le 1$$

Inverse Document Frequency:

$$G_i = 1 + \log_2(\frac{n}{d_i}), 1 \le G_i \le \infty$$

Mutual Information:

$$G_i = \max_2(\log(\frac{P(x_i, C_k)}{p(x_i)P(C_k)}))$$

$$G_i = 1$$

Global Frequency Times Inverse Document Frequency:

$$G_i = \frac{g_i}{d_i}$$

Normal: $G_i = \frac{1}{\sqrt{\sum_j f_{ij}^2}}$

Information Gain:

$$G_i = -\sum_k P(C_k)\log P(C_k) + P(t_i)\sum_k P(C_k|t_i)* \log P(C_k|t_i) + P(\overline{t_i})\sum^k P(C_k|\overline{t_i})\log P(C_k|\overline{t_i})$$

$p_{ij=a_{ij}/f_i}$

a_{ij} : 문서 j에서 용어 i가 나타낸 횟수

f_i : 전체 문서에서 용어 I가 나타난 빈도

n : 전체 문서의 수

d_i : 용어 i가 나타난 문서의 수

G_i : 용어 i의 용어 가중치

문서 변환은 단어-문서 행렬이 차원이 크고 0의 값이 대부분인 희소(sparse)행렬이므로 문서에 포함되어 있는 단어는 항상 동일하지 않기 때문에 단어-문서 행렬에서 단어의 수는 문서의 수 보다 훨씬 크다. 문서의 차원이 크면 단어 계산을 하는데 할애되는 시간이 많이 소요될 뿐만 아니라 해당 두 문서 사이의 거리 및 밀도를 정의하기가 어렵기 때문에 좋은 군집 결과를 얻지 못한다. 차원을 축소하는 방법은 여러 가지가 개발되어 있는데 자료가 연속형 자료인 경우 고차원에서 저차원으로 투영시키는 선형 대수학 기법을 많이 사용한다.

이를 바탕으로 키워드 중심의 워드 클라우드 알고리즘을 작성하고 시각화 정보를 도출하였다. 워드 알고리즘에서 각 단어별 유사도를 위하여 Cosine Similarity, Jaccard Similarity, Lexical Similarity 중에서 Cosine Similarity 방식을 사용하였다[1].

코사인 유사도는 코사인 거리를 바탕으로 단어 간 유사도를 측정하는 방식으로 아래와 같이 두 개의 영벡터를 바탕으로 유클리디안 방정식을 도출하여 작성할 수 있다[19].

$$a \cdot b = \| a \| \ \| b \| \cos\theta$$

두 벡터 a와 b에 대하여 유사도를 계산하면 다음과 같다.

Similarity:

$$\cos(\theta) = \frac{A \cdot B}{\| A \| \ \| B \|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

도출된 유사도를 바탕으로 N개 차원 벡터 a와 π를 통해 코사인 유사도[18]를 계산하는 알고리즘은 다음과 같다.

soft_cosine(a, b)

$$= \frac{\sum_{i,j}^{N} s_{ij} a_i b_j}{\sqrt{\sum_{i,j}^{N} s_{ij} a_i a_j} \sqrt{\sum_{i,j}^{N} s_{ij} b_i b_j}}$$

where s_{ij} = similarity(($feature_i, feature_j$)

3. 연구 방법

3.1 대학별 SW교육 현황

본 연구에서는 상기 언급한 현황을 통한 국내 대학의 SW 교육의 실태와 SW 인력 구조를 개선하기 위한 제언을 위하여 8개 대학을 선정하여 SW 전공 교과과정 및 강의계획서를 수집한 정보를 토대로 토픽모델링(LDA)을 실시하였다. 대상이 되는 대학은 정부 정책에 선정된 SW 중심대학 대표 4개 대학, 일반 대학 4개 대학으로 2분할을 하여 텍스트 마이닝 및 연관성 분석을 실시하였고, 선정된 8개 대학을 다시 수도권 4개 대학, 비수도권 4개 대학으로 재분류를 하여 지역별 차이가 있는지도 확인하고자 하였다.

SW 중심대학은 선행연구에서 언급했듯이 미래부에서 SW 인력의 강화를 위하여 SW 중심대학을 설립하고 인재 양성을 도모하고 K-ICT를 실현해 나갈 수 있는 선도 대학을 만들기 위하여 개설 되었다.

본 연구에서는 SW 중심대학으로 선정된 15, 16년 14개 대학 중 대표적인 4개의 대학을 임의 선정하여 각 대학의 SW 관련 학과의 교과과정 수집 및 강의계획서를 수집하여 실증분석에 사용하였다.

〈Table 2〉 The Characteristics of Major SW-Centered Universities

University	Department	Characteristics
A	Computer Science	• Admission without first-year department • Interactive hands-on support platform online programming training • Selection of SW specialists and selection of SW gifted
B	Department of Computer Science	• Integration with existing SW related departments • Completion of 5-semester or more completed course • Selection of SW specialists
C	Department of Software Computer Engineering	• SW University new establishment and expansion of department • Establishment of educational institute • SW specialist selection and excellent scholarship for 4 years
D	Department of Computer Science	• Integrating SW-related departments into computer engineering • Establishment for SW education center directly under the university chancellor • dual degree with foreign university

〈Table 3〉 The Characteristics of Non SW-Centered Universities

University	Department	Characteristics
E	Department of Computer Science	• Training and IT new technology development with emphasis on practical training • Collaboration with computer-related institutes such as Samsung(SCSC)
F	Department of Computer Information Engineering	• Selected as the best university in IT industry computer sector(2010) • Training of SW manpower through track connection with Samsung specialist
G	Department of Computer Information Engineering	• Computing education through computer engineering track • Computing education through information system engineering track
H	Department of Computer Science	• Problem-solving ability through problem-based learning method(PBL) • Adoption of professional technical certification

〈Table 4〉 Metropolitan and Non-Metropolitan Universities

University	Classification	University	Classification
B	Seoul/Metropolitan area	A	Daejeon/Non-metropolitan area
C	Seoul/Metropolitan area	D	Dae-gu/Non-metropolitan area
E	Seoul/Metropolitan area	F	In-Chon/Non-metropolitan area
G	Gyeonggi/Metropolitan area	H	Asan/Non-metropolitan area

<Table 2>는 SW 중심대학의 학과와 간단한 특성을 설명하고 있다. SW 중심대학은 기존의 SW 관련 학과를 통합 또는 강화하여 기존 보다 더 좋은 환경에서의 SW 인력을 양성할 수 있도록 하는 것에 의의가 있다. 15, 16년도의 SW 중심대학 선정 이후 변화된 환경에서 과연 기존의 일반대학과의 차이점이 있는가에 대한 비교분석을 위한 일반 대학 4개를 전국에서 임의 선정하여 분석하고자한다.

일반 대학 4개 대학의 특징은 <Table 3>과 같다. 일반 대학의 경우도 각 대학의 특성과 비전에 맞춘 교육을 통하여 SW 인력 교육에 힘쓰고 있다. 주요 특징을 살펴보면 정부에서 지원하는 내용들이 아닌 기업체와의 직접적인 연결을 통한 프로그램을 구성하여 학생들이 직접 기업과의 접촉을 통하여 이론과 실무를 겸할 수 있는 환경을 만들려는 노력을 지속적으로 하고 있다.

자가 인증 제도를 구성하여 SW 인력으로써 자질이 있는가에 대한 역량을 인증해 줄 수 있는 자격제도를 겸비하여 SW 전공 학생이 추후 현업에 진출할 때 도움이 될 수 있는 프로그램을 구성하여 보다 이론과 실제의 간극을 줄이려는 노력을 지속적으로 하고 있다. 그러나 다수의 대학이 이런 지원을 지속적으로 해줄 수 있는가에 대해서는 점검이 필요하며, 실효성이 있는가도 의문이다. 따라서 본 연구에서는 SW 중심대학의 사례와 일반 대학의 사례를 비교 분석함으로써 재점검을 할 수 있으며, 현재 SW 인력의 비매칭 구조를 타파 할 수 있는 제언을 하는데 본 연구의 시사점이 있다.

상기 언급한 8개 대학에 대해서 SW 중심대학 선정 여부에 따른 분류 방법 이외에도 지역별 대학으로 분류하여 분석하여 지역별 SW 인력 교육의 차이가 존재하는지 파악하고자 한다.

<Table 4>는 8개 대학의 위치와 위치에 따라 수도권과 비수도권 대학의 분포를 나타낸다.

따라서 B, C, E, G 대학은 수도권으로 분류하고, A, D, F, H 대학은 비수도권 대학으로 분류하여 지역별 SW 교과과정의 차이를 분석하고자 하였다.

상기 8개 대학의 2가지 분류법을 토대로 본 연구에서는 SW 중심대학과 일반 대학의 SW 교과과정의 차이 비교 분석과 지역별 분류를 통한 수도권 대학과 비수도권 대학의 SW 교과과정의 차이 비교 분석을 실시하였다.

이는 SW 인력을 배양하는데 있어 얼마나 실증적으로 교육을 시키고 환경적 측면에서 양성할 수 있는가를 파악함에 의의가 있으며 가장 선행되어야 하는 부분인 교과과정과 강의계획서를 토대로 분석을 함으로써 이론과 실제의 간극의 원인을 찾고 해결하는 것에 본 연구의 시사점이 존재한다.

3.2 연구 1: 단순 교과목 분석

본 연구는 교과과정의 교과목을 추출하여 학교별 중첩적으로 교수하고 있는 과목을 추출하고 학교 내에서 해당 유사과목의 중첩정도를 파악하여 비효율적인 교과편성에 대한 기술적 분석(Descriptive analysis)을 실시함으로써 학교별 문제를 표면적, 직관적으로 진단 가능하다. 또한 SW 인력의 양성에서 가장 중요한 실습의 비율을 파악함으로써 이론적인 교육뿐만 아니라 실제 현업에서 사용할 수 있는 기술을 배양하는가를 파악함으로써 추후 SW 인력 교육의 방향성 제고와 현업에서 원하는 SW 인력의 합리적 선발에 도움이 된다.

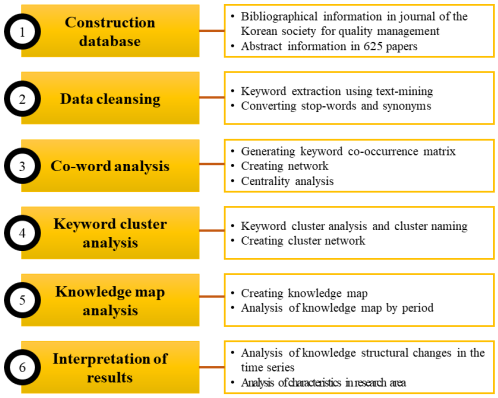
따라서 본 연구에서 선행되는 분석으로 단순 커리큘럼 교과목을 통한 실습비율과 과목별 중첩비율을 분석하였다. 각 8개 대학의 웹사이트 및 수강신청과 관련된 사이트에서 제공하고 있는 SW 관련학과 교과과정을 수집하여 본 연구의 단순 교과목 분석에 사용하였다.

3.3 연구 2: 토픽 모델링(LDA)

본 연구에서는 SW 관련 학과 강의계획서를 토대로 텍스트 마이닝 기법을 이용하여 솔루션을 도출하고자 하였다. 텍스트 마이닝은 여러 가지 연구 방법 들이 존재하는데[20, 21] 그 중에서 동시단어분석, 분류 분석, 연관성 분석 등을 이용하여 진행하였다[5]. 관련된 연구로 동시단어분석을 이용한 분석을 이용한 프레임워크를 인용하여 분석에 차용한 연구들이 일반적이다[22, 23].

강의계획서 내에 있는 텍스트를 추출하기 위하여 본 연구는 파이선(python)을 기반으로 자연어 처리를 진행하였다. 그에 따라 명사형 키워드를 파싱하여 필터링하고 중복되는 명사들을 수량화하여 가장 많이 나온 순으로부터 상위 50개를 추리고 파싱된 명사 중에서 컴퓨터 및 SW에 관련된 명사를 2차로 추려서 연관성 분석을 진행함으로써 기존의 교과목으로만 했었던 연관 분석에 보다 논리적 접근을 할 수 있다.

텍스트 마이닝 분석을 위하여 <Figure 3>와 같은 연구 프레임워크를 구성함으로써 구조적 분석을 이행 하였다. 데이터베이스를 구축하고 데이터 정제를 통하여 유효 키워드를 도출, 관련 단어 연관성 분석을 실시하고 도출된 단어들의 군집을 구성하여 해당 단어들의 분류를 찾아내고 관련 결과를 찾는 것으로 기존 연구 프레임워크를 재구성 및 진행하였다.



〈Figure 3〉 Framework for Research 2

본 연구에 사용된 데이터는 선정된 각 대학의 SW 관련 학과의 교과목에 따른 강의계획서를 수집하여 문서화 시켜 파싱을 진행하였으며, 공시되지 않은 강의계획서나 같은 수업의 분반으로 중복되는 교과목의 경우 본 연구에서 제외하였다.

LDA를 사용하여 파싱된 키워드를 실셈하여 상위 50개까지 도출하고 키워드 별 중첩 비율을 확인함으로써 강의 계획서 간의 연관성을 파악하였다. 중첩정도가 높을수록 과목의 강의 계획서간 차이 정도가 적어지고 불성실한 강의 계획서 작성이나 혹은 유사한 내용으로만 강의 계획서를 작성하여 SW 교육에 유용한 영향을 주지 못한다. 또 다른 관점으로는 많이 나온 키워드의 경우 중요한 키워드로 인식하여 SW 교육에 있어 중요한 요소라고도 할 수 있기 때문에 파싱된 키워드의 순위 및 중첩정도를 파악하는 것이 중요하다. 연구 2의 경우도 미리 선정된 8개 대학의 강의계획서를 기반으로 텍스트 마이닝을 수행하여 실시하여 SW 중심대학과 일반대학의 키워드 분포 비교와 수도권 대학과 비수도권 대학의 키워드 분포 비교하였다.

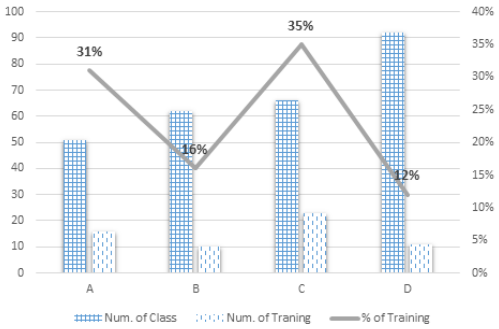
4. 연구 결과

4.1 연구 1의 결과

SW 중심대학 4개 대학의 실습 비율 및 중첩 비율을 비교한 결과는 <Table 5>와 <Figure 4>와 같다. SW 중심대학 간 교과과정을 비교한 결과실습비율은 12~35%까지 나타났다.

〈Table 5〉 Training Percentage of SW-Centered Universities

University	A	B	C	D
Num. of Class	51	62	66	92
Num. of Training	16	10	23	11
% of Training	31	16	35	12

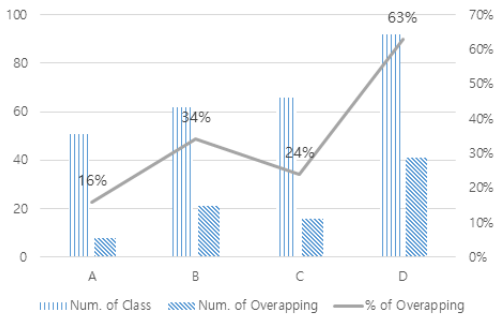


〈Figure 4〉 Training Percentage of SW-Centered Universities

SW 중심대학별 교과목 중첩비율은 16~63%까지 다소 다양하게 분포하였으며 4개 대학을 비교한 결과는 <Table 6>과 <Figure 5>이다.

〈Table 6〉 Overlapping Percentage of SW-Centered Universities

University	A	B	C	D
Num. of Class	51	62	66	92
Num. of Overlapping	8	21	16	41
% of Overlapping	16	34	24	63

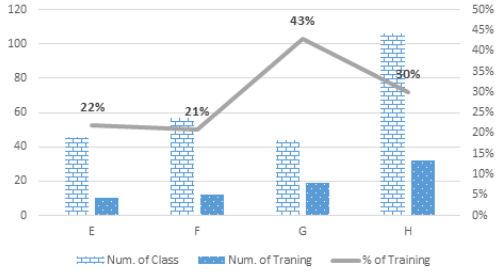


〈Figure 5〉 Overlapping Percentage of SW-Centered Universities

일반 대학 4개 대학의 실습 비율 및 중첩 비율을 비교한 결과는 <Table 7>과 <Figure 6>과 같다. 일반 대학 간 교과과정을 비교한 결과 실습비율은 21~43%까지 나타났다.

〈Table 7〉 Training Percentage of Non SW-Centered Universities

University	E	F	G	H
Num. of Class	46	57	44	106
Num. of Training	10	12	19	32
% of Training	22	21	43	30



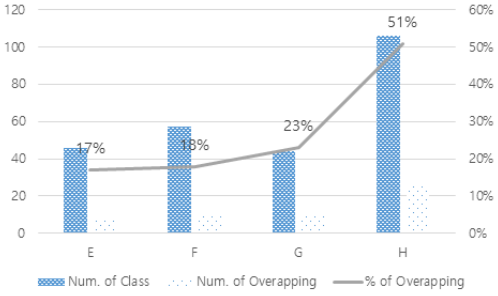
〈Figure 6〉 Training Percentage of Non SW-Centered Universities

반 대학별 교과목 중첩 비율은 17~51%로 다소 다양하게 나왔으며 일반 4개 대학을 비교한 결과는 <Table 8>과 <Figure 7>과 같다.

상기 8개 대학의 단순 교과목의 중첩 비율과

〈Table 8〉 Overlapping Percentage of Non SW-Centered Universities

University	E	F	G	H
Num. of Class	46	57	44	106
Num. of Overlapping	8	10	10	27
% of Overlapping	17	18	23	51



〈Figure 7〉 Overlapping Percentage of Non SW-Centered Universities

수업의 실습 비율을 분석한 결과 SW 중심대학과 일반 대학의 차이는 특별하게 나타나지는 않았다. 즉, SW 중심대학의 선정이 1~2년 정도 밖에 진행되지 않은 시점에서 SW 교과과정의 전반을 변화시키거나 혹은 실습의 비율을 50%이상까지 올려 실전적인 학습을 시키는 것에 있어 상대적 시간이 부족했다는 점을 시사하고 있다. 그러나 미비하지만 상대적으로 중첩 비율이 높은 것으로 도출되어 올바른 SW 교육을 진행하고 있다고 판단된다.

2번째 분류방법인 수도권과 비수도권을 나눈 결과 수도권 대학의 중첩비율은 32~55%의 비율로 나타났고 비수도권의 중첩비율은 38~65%로 나타나 상대적으로 수도권 대학의 경우 중첩률이 적은 것을 확인하였다. 그러나 8개 대학 모두 과목에 대한 중첩 정도가 평균적으로 높지 않기 때문에 추후의 개선 방향으로 중첩률을 제고하고 실습율을 올려 현실적인 감각이 있는 SW 인력을 양성하는데 더욱더 집중해야 한다.

4.2 연구 2 결과

본 연구는 파싱된 키워드를 바탕으로 컴퓨터 및 SW 관련된 키워드를 빈도가 높은 순서대로 분류를 하여 상위 50개를 뽑아 낸 키워드로 2가지 분류체계에 의하여 재정립하고 연관분석을 통한 중첩 확인을 실시하여 보다 정확한 강의계획서 상의 중첩 이슈를 도출 하는 것에 의의가 있다.

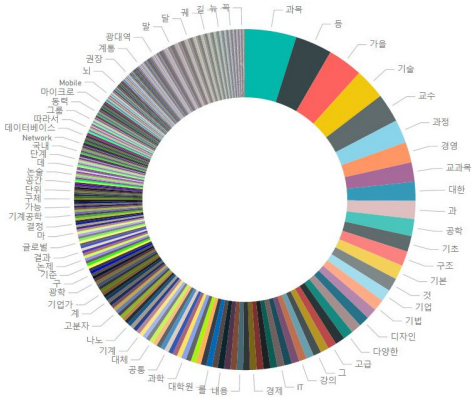
<Figure 8>과 <Figure 9>는 SW 중심 대학의 강의계획서를 파싱한 키워드의 사례이다.

SW 중심대학 SW 관련 상위 50개 키워드 재정립을 통하여 과목, 강의 등 평가에 불필요한 부분을 제거한 나머지 SW 관련 및 유의미한 단어들만 상위로 50개를 뽑아내어 분석을 하였다.

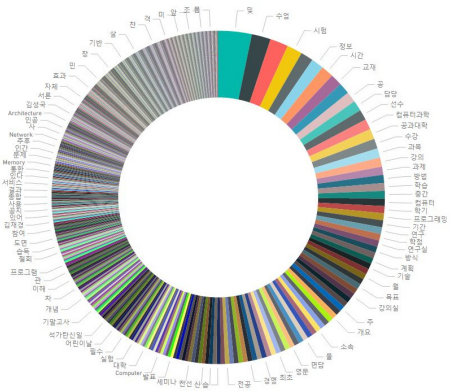
A 대학교의 경우 경영, 경제 등의 내용 들이 비교적 많이 분포하고 있고 B 대학교의 경우 경제, 공학, 과학, 개발 등 비즈니스와 관련된 키워드가 상대적으로 높게 나타났다. C 대학교의 경우 공학, 컴퓨터 등 SW 관련된 키워드 위주로 많이 분포하고 있으며 D 대학교도 동일하게 SW 관련 키워드 위주로 많이 분포하고 있어 기존의 상위 50개 키워드와는 비교적 다른 분포를 보이고 있다. <Figure 10>와 <Figure 11>는 일반 대학의 SW 관련 학과의 강의계획서를 파싱한 키워드의 사례이다.

〈Figure 8〉 A University's Wordcloud

〈Figure 10〉 E University's Wordcloud



〈Figure 9〉 A University's Wordchart



〈Figure 11〉 E University's Wordchart

E 대학교의 경우 수업, 시간, 교재 등의 키워드가 많이 도출되어 과목 내용에 관련된 키워드가 많이 도출되었으며 F 대학도 강의, 과제, 능력 등 키워드가 많이 도출되어 과목 내용에 관련된 키워드가 많이 도출되었다. G 대학교는 비교적 프로그래밍, 인터넷 등의 내용들이 나와 상대적으로 SW 과목에 대한 언급이 많은 것으로 나타났으며 H 대학교의 경우 Computer, GUI 등 SW 관련된 내용들이 매우 많이 언급되어 있어 타 대학교들에 비교하여 상당히 SW 관련된 키워드를 내포하고 있다는 것을 알 수 있다.

파싱된 50개의 키워드를 살펴보면 1차적으로 단순 키워드와 2차적으로 다시 SW 관련키워드를 나눠 두 가지의 요소들의 중첩 비율을 확인하여 <Table 9>와 같이 도출하였다.

〈Table 9〉 Overlapping Percentage of Keyword for SW-centered Universities

University	A	B	C	D
Num. of Overlapping Keyword	29	25	10	26
% of Overlapping	58%	50%	20%	52%
Num. of Overlapping SW Keyword	36	36	18	27
% of Overlapping	72%	72%	36%	54%

SW 중심대학별 기존 키워드 중첩 비율과 SW 관련 키워드 중첩비율을 비교 했을 때 강의계획서에 불필요한 키워드들이 많이 존재하고 있다는 결론을 도출하였다. 물론 키워드 중첩이 많이 된다고 하여 체계적인 강의 계획서라고 판단할 수는 없다. 그러나 강의의 핵심내용을 제시하는 강의계획서인 만큼 강의 내용의 핵심 키워드들의 빈도가 높게 나타나야 하는 것은 중요하다. 예를 들어 '인터페이스'와 관련된 수업에서는 'GUI', 'Photoshop' 등의 특정 기술 및 소프트웨어 언어 등의 단어들이 중복되는 것이 필요하며 불필요한 '전략', '범위', '평가' 등의 내용들의 빈출도가 높은 것은 강의계획서의 내실을 높일 필요가 있는 것으로 판단된다. 즉, 핵심 내용을 담아야 하는 강의계획서에 불필요한 요소들이 첨가되어 수업을 이수하는 학생들로 하여금 해당 강의의 당위성이나 신뢰성을 저하하는 역할을 했다고 판단된다. 추후에는 보다 핵심적인 내용의 강의계획서를 구성하는 것이 중요하다.

일반대학도 SW 중심대학과 마찬가지로 파싱된 50개의 키워드를 살펴보면 1차적으로 단순 키워드와 2차적으로 다시 SW 관련키워드를 나눠 두 가지의 요소들의 중첩 비율을 확인하여 <Table 10>의 표를 도출하였다.

〈Table 10〉 Overlapping Percentage of Keyword for Non SW-Centered Universities

University	E	F	G	H
Num. of Overlapping Keyword	29	19	24	6
% of Overlapping	58%	38%	48%	12%
Num. of Overlapping SW Keyword	40	34	39	17
% of Overlapping	80%	68%	78%	34%

일반 대학 기존 키워드 중첩 비율과 SW 관련 키워드 중첩비율을 비교 했을 때 SW 중심대학과 마찬가지로 강의계획서에 불필요한 키워드들이 많이 존재한다는 것을 볼 수 있었으며, 상대적으로 더 많은 격차가 나고 있어 SW

중심대학보다 강의계획서의 품질이 더 낮다는 것을 확인하였다. SW 관련 키워드만 도출 했을 경우 80%까지의 중첩률을 보이고 있어 상대적으로 해당 교과과정을 잘 표현하고 있다고 볼 수 있으나 불필요한 키워드에 대해서는 개선해야할 필요가 있다.

두 번째 분류인 수도권과 비수도권의 중첩 비율 비교는 <Table 11> 및 <Table 12>와 같다.

〈Table 11〉 Overlapping Percentage of Keyword for Metropolitan Universities

University	B	C	E	G
Num. of Overlapping Keyword	25	26	29	24
% of Overlapping	50	52	58	48
Num. of Overlapping SW Keyword	36	27	40	39
% of Overlapping	72%	54	80	78

〈Table 12〉 Overlapping Percentage of Keywrods for Non-Metropolitan Niversities

University	A	D	F	H
Num. of Overlapping Keyword	29	10	19	6
% of Overlapping	58	20	38	12
Num. of Overlapping SW Keywords	36	18	34	17
% of Overlapping	72	36	68	34

분석결과 수도권 대학의 경우 비수도권 대학보다 상대적으로 중첩 비율이 높았다. 즉, 수도권의 대학의 경우 전반적인 SW 인력을 양성할 수 있는 교과과정을 짜임새 있게 잘 짜고 있다는 것을 입증한다. 비수도권의 경우 각 중첩비율의 격차가 매우 커 지속적인 개선이 필요하다는 점을 시사하고 있다.

5. 결 론

본 연구는 현업에서 SW 인력을 충원하기 위하여 실질적으로 도움이 되는 교육을 대학에서 시행하는가를 알기 위한 목적을 바탕으로 연구를 진행하였다. 본 연구의 시사점은 다음과 같다. 첫째, 교과과정에서의 실습 비율이 상대적으로 많지 않다는 점이다. 특히 SW 인력의 경우 현업에 투입이 되었을 때 한 명의 개발자로써의 역량을 갖추고 재능을 발휘해야함에도 불구하고 이론적인 접근은 상대적으로 잘 되어 있으나 SW 운용의 실제적 접근은 미숙하다. 둘째, SW 중심대학의 활성화이다. 15년부터 시행하고 있는 SW 중심대학의 경우 SW 인력 양성을 위한 지원을 아낌없이 진행하고 있으나 유효 효과를 나타내지 않고 있어 추후 귀추가 주목된다. 셋째, 강의계획서의 개편안이 필요하다. 강의계획서에 대한 토픽모델링 분석을 실시하면서 불필요한 키워드들이 다수 도출되는 현상이 나타나 강의계획서의 역할이 불분명한 상황이라 판단된다. 수강을 하는 대학생의 경우도 강의계획서의 내용으로 강의의 내용을 적절하게 파악하지 못하고 자신이 필요한 강의를 듣지 못하게 되는 경우가 발생할 수 있어 강의계획서를 작성할 때 보다 직관적으로 관련성 있는 키워드를 사용해서 작성해야만 한다.

본 연구의 정책적 시사점은 다음과 같다. 첫째, 현업에서 요구하는 요소와 대학의 교과과정에 대한 반영할 수 있는 정책 대안 필요하다.

즉, 현업의 이해관계자와 대학의 교수진이 협업을 하여 전체적인 과목의 재진단을 통하여 학문적, 실용적 접근이 모두 용이할 수 있도록 해야 한다. 이론과 실제의 간극의 차이는 어떤 분야에서든 나타나는 현상이지만 간극을 줄여 나가는 고민은 항상 해야 하기 때문이다. 또한 이를 잘 이행할 수 있도록 정부에서는 해당 관련 정책을 보안 및 개선해야 한다. 둘째, 본 연구에서는 진행되지 않았으나, 국가직무능력표준(NCS)의 직무 내용과 워크넷의 실제적인 직무 내용도 동일하게 토픽모델링을 진행하여 각 직무별 분류분석, 연관성 분석을 실시하여 실제적 매칭정도가 어느 정도인지 파악할 필요도 존재한다. 본 연구에서는 SW 관련 학과의 과목별 텍스트 마이닝만 진행하고 있어 추후 언급한 내용까지 분석이 된다면 보다 현실적이고 실제적인 SW 인력 연구가 될 것이다. 셋째, 상기 언급한 두 가지의 제언을 통하여 SW 인력을 충원할 수 있는 매칭 서비스를 인공신경망과 인공지능(Artificial Intelligence)를 이용하여 매칭시켜 줌으로써 수요자와 공급자의 간극을 메꿀 수 있다. 추후 연구에서는 매칭 서비스에 대한 연구도 진행한다면 보다 실제적 연구가 될 것이다.

본 연구는 소프트웨어 분야의 강의계획서를 분석함에 따라 대학의 교육 현황을 파악하고자 함에 그 목적을 두었다. 그러나 강의계획서를 기반으로 실제 수업이 명확하게 잘 지켜진다는 가정에서 본 연구의 결과는 실태를 명확하게 반영할 수 있으나 실제 수업과 강의계획서의 내용이 확실히 일치하지 못하는 경우가 발생할 가능성도 분명 존재한다. 따라서 본 연구의 결과를 바탕으로 향후 연구에서는 실제 수업의 내용과 강의계획서의 연관성을 확인하는 작업이 선행될 필요가 있을 것으로 판단된다.

그럼에도 불구하고 본 연구는 SW 관련 교과과정에 대한 실증적인 분석이 많이 진행되지 않고 있어 강의계획서와 교과과정 분석을 시도했다는 점에서 그 의의를 찾을 수 있다.

References

[1] Amit, S., “Modern Information Retrieval: A Brief Overview,” Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, Vol. 24, No. 4, pp. 35-43, 2001.

[2] Blei, D., Ng, A. and Jordan, M., ‘Latent Dirichlet Allocation,’ Journal of Machine Learning Research, Vol. 3(2003), pp. 993-1022.

[3] Cha, J. S., “The Study of the Field Customized SW Training Course Design based on the Analysis of the Field Suitability of the University SW Education,” Smart Media Journal, Vol. 4, No. 4, pp. 86-92, 2015.

[4] Duan, Y. and Cruz, C., “Formalizing Semantic of Natural Language through Conceptualization from Existence,” International Journal of Innovation, Vol. 2, No. 1, pp. 37-42, 2011.

[5] Feldman, R. and Sanger, J., “The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data,” Cambridge University Press, 2007.

[6] Han, D. S., "University Education and Contents in The Fourth Industrial Revolution," Human Contents, Vol. 42, pp. 9-24, 2016.

[7] Harvard University, "The Unofficial Guide to Computer Science at HARVARD ver. 14," pp. 1-12, 2014.

[8] Hutchins, J., "The history of machine translation in a nutshell," pp. 1-4, 2005.

[9] Kang, S. K., Yu, H., and Lee, Y. J., "Analyzing Disaster Response Terminologies by Text Mining and Social Network Analysis," Information Systems Review, Vol. 18, No. 1, pp. 141-155, 2016.

[10] Kim, J. Y. and Kim, D. S., "A Study on the Method for Extracting the Purpose-Specific Customized Information from Online Product Reviews based on Text Mining," The Journal of Society for e-Business Studies, Vol. 21, No. 2, pp. 151-161, 2016.

[11] Kim, S. K., Cho, H. J., and Kang, J. Y., "The Status of Using Text Mining in Academic Research and Analysis Methods," Journal of Information Technology and Architecture, Vol. 13, No. 2, pp. 317-329, 2016.

[12] Kishorjit, N., Vidya Raj R. K., Nirmal, Y., and Sivaji, B., "Manipuri Morpheme Identification," Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing(SANLP), pp. 95-108, 2012.

[13] Lee, K. N., "SW manpower employment trend," Information and communication broadcasting policy, Vol. 25, No. 6, pp. 39-46.

[14] Manning and Schütze, "Foundations of Statistical Natural Language Processing," MIT Press, 1999.

[15] Ministry of Science, ICT and Future Planing, "Annual Report on Information Technology Development," National IT Industry Promotion Agency, pp. 1-341, 2015.

[16] Noh, Y., Ahn, I. J., and Choi, S. K., "A Study on Analyzing and Developing the Syllabus for Library and Information Science Core Courses," Journal of Korean Library and Information Science Society, Vol. 44, No. 1, pp. 143-175, 2013.

[17] Ryu, J. S., Lee, S. H., and Kim, J. W., "A Study on the Problems of Korean University Education for the Promotion of IT Talent," Samsung Economic Research Institute, pp. 1-189, 2011.

[18] Sidorov, G., Gelbukh, A., Gómez-Adorno, H., and Pinto, D., "Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model," Computación y Sistemas, Vol. 18, No. 3, pp. 491-504, 2014.

[19] Spruill, M. C., "Asymptotic Distribution of Coordinates on High Dimensional Spheres," Electronic Communications in Probability, Vol. 12, pp. 234-247, 2007.

[20] Sullivan, D., "Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales,"

John Wiley & Sons, Inc., pp. 1-560, 2001.

[21] Tan, A. H., "Text Mining: The State of the Art and the Challenges," Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, Vol. 8, pp. 1-6, 1999.

[22] Tseng, Y. H., Lin, C. J., and Lin, Y. I., "Text Mining Techniques for Patent Analysis," Information Processing & Management, Vol. 43, No. 5, pp. 1216-1247, 2007.

[23] Weiss, S. M., Apte, C., Damerau, F. J., Johnson, D. E., Oles, F. J., Goetz, T., and Hampp, T., "Maximizing Text-Mining Performance," IEEE Intelligent Systems, Vol. 14, No. 4, pp. 63-69, 1999.

저 자 소 개

최재원 (E-mail: jaewonchoi@sch.ac.kr)
2004년 가톨릭대학교 경영학과 졸업
2006년 가톨릭대학교 일반대학원 경영학과 (MS)
2010년 가톨릭대학교 일반대학원 경영학과 (Ph.D)
2010년~2011년 KAIST 경영대학 연수연구원
2011년~2013년 연세대학교 정보대학원 연구교수
2014년~현재 순천향대학교 경영학과 조교수
관심분야 웹개인화, 빅데이터 분석, 지능형 에이전트, 디지털 마케팅

이 호 (E-mail: leeho32@spri.kr)
2006년 State University of New York at Stony Brook (공학 학사), Computer Science 전공
2014년 연세대학교 정보대학원, 박사, 경영정보시스템 전공
2014년~2015년 연세대학교 박사후연구원
2015년~현재 소프트웨어정책연구소 선임연구원
관심분야 빅데이터 분석, 소셜네트워크 분석, 텍스트 마이닝, 익명성, 디지털 마케팅

김정민 (E-mail: jungmink26@spri.kr)
2013년 중앙대학교 컴퓨터공학부 졸업
2015년 중앙대학교 일반대학원 소프트웨어공학 (석사)
2014년~현재 소프트웨어정책연구소 연구원
2017년~현재 중앙대학교 일반대학원 컴퓨터공학과 박사과정
관심분야 소셜네트워크 분석, 소프트웨어아키텍처 모듈 뷰, 텍스트마이닝

송주호 (E-mail: juho@sw.or.kr)
2008년 한성대학교 정보시스템공학과 졸업
2011년 연세대학교 정보대학원 석사졸업 (IT산업・정책 전공)
2011년~2015년 한국전자통신연구원 사업화본부 기술원
2015년~2016년 한국정보방송통신대연합 정책연구본부 과장
2016년~현재 한국소프트웨어산업협회 인재개발실 책임
관심분야 기술경영, 기술사업화, HRM, HRD