

# 단어 온톨로지를 적용한 서술형 강의평가 분석: 학생 중심 강의평가 플랫폼, '에브리타임'을 중심으로

## Analysis of Descriptive Course Evaluations Applying Word Ontology: Focusing on Student-Centered Narrative Lecture Evaluation Platform, Everytime

김혜정\*, 김민서\*\*, 문성민\*\*\*

아주대학교 사학과\*, 아주대학교 영어영문학과\*\*, 아주대학교 인문과학연구소\*\*\*

Hye-Jeong Kim(gpwj1478@ajou.ac.kr)\*, Min-Seo Kim(kms010614@ajou.ac.kr)\*\*,  
Seongmin Mun(stat34@ajou.ac.kr)\*\*\*

### 요약

본 연구의 목적은 학생 중심의 서술형 강의평가 플랫폼인 에브리타임의 강의평가 텍스트를 분석하여 강의평가에 영향을 미치는 주요 평가 요소를 밝히는 것이다. 분석 대상은 A 대학교 영역별 교양 강의 50개이다. 분석 방법은 단어 기반 온톨로지를 적용한 단어출현 빈도 분석과 의사결정나무 분석, 평균 비교 분석을 사용하였다. 분석 결과, 서술형 강의평가 텍스트에 존재하는 주요 강의평가 요소는 강의, 성적, 과제, 시험, 교수이며, 이중 종합적인 강의평가에 큰 영향을 미치는 주된 강의평가 요소는 성적과 교수인 것으로 드러났다. 또한 강의평가 절대 점수가 높은 강의 집단과 낮은 강의 집단 사이에 교수, 과제, 성적 항목에 있어 유의미한 차이가 존재했으며, 성적에 대한 언급이 많고 교수에 대한 언급이 적은 강의의 강의평가 절대 점수가 높았다. 대학 차원의 강의평가를 분석한 선행연구의 결과와는 달리, 학생들은 교수자의 특성보다 성적과 같은 실리적 가치를 추구한다는 본 연구의 결과는 학생을 중심에 둔 대학 강의 개선의 실마리를 찾는 단서가 될 수 있다.

■ 중심어 : | 강의평가 분석 | 텍스트 마이닝 | 단어 온톨로지 | 의사결정나무분석 | 자연어처리 |

### Abstract

The purpose of this study is to analyze student-centered narrative lecture evaluation texts (Everytime) to identify key evaluation factors that influence lecture evaluation. The subjects of analysis are 50 liberal arts lectures in different areas of A University. The analysis methods were word occurrence frequency analysis using word-based ontology, decision tree analysis, and Independence T-test. The results of the analysis showed that the main lecture evaluation elements in the narrative lecture evaluation texts are lecture, grade, assignment, test, and professor, and among them, grade and professor are the main lecture evaluation elements that have a significant impact on the overall lecture evaluation. In addition, there were significant differences in the items of professor, assignment, and grade between the lecture groups with high and low absolute lecture evaluation scores, and lectures with more references to grade and fewer references to professor had higher absolute lecture evaluation scores. Contrary to the findings of previous studies analyzing university-driven course evaluations, this study's finding that students seek tangible values, such as grades, rather than instructor characteristics, may provide clues to improving university teaching with students at the center.

■ keyword : | Lecture Evaluation Analysis | Text Mining | Word Ontology | Decision Tree Analysis | Natural Language Processing |

\* 본 연구는 한국연구재단의 인문사회분야 인문사회연구소지원사업의 지원(NRF-2022S1A5C2A02090368)을 받아 수행되었습니다.

접수일자 : 2023년 12월 27일

수정일자 : 2024년 02월 02일

심사완료일 : 2024년 02월 20일

교신저자 : 문성민, e-mail : stat34@ajou.ac.kr

## I. 서론

어느 분야에서든지 자신이 제공하는 서비스를 이용자로부터 평가받고, 이를 기반으로 서비스의 질을 개선하려는 노력은 거의 필수적으로 이루어지고 있다. 학문과 교육의 장인 대학 또한 고등 교육 기관으로서 가지는 교육적 책임과 강의의 질적 보장을 통한 학습권의 제고, 대학의 범국가적 경쟁력 확보와 교수 역량의 양질화를 위해 강의평가를 운영하고 있다[1]. 이러한 전통적인 대학 강의평가는 교육 서비스 제공 기관으로서 강의를 수강한 학생들의 의견을 수집함으로써 교육 서비스의 질을 제고함에 그 목적을 둔다[2].

최근 인터넷 시대가 도래함에 따라 대학 강의평가에 대한 패러다임도 변화의 국면을 맞았다. 온라인상에서 학생들이 서로 소통하고 강의에 대한 의견을 자유롭게 게시할 수 있는 커뮤니티가 조성되면서, 기존의 서비스 제공자 중심 평가를 벗어난 '이용자 중심 평가'가 가능해졌기 때문이다. 즉 전통적인 강의평가와는 별개로, 서비스 이용자인 학생들 사이에서 무의식적 압력에서 벗어난 훨씬 더 자유롭고 비정제된 평가가 이루어지고, 이를 어떠한 제약 없이 공유하고 논의할 수 있는 환경이 조성되었다고 할 수 있다[3][4]. 현재 이러한 변화의 중심에 있는 것이 '에브리타임'이다. 에브리타임은 전국 400개 대학을 대상으로 운영되고 있는 대학교 커뮤니티 및 강의 시간표 서비스로, 대학생 사이에서는 일명 '대학 생활 필수 아이템'이라 불리며 그 지위를 확립하고 있다[4]. 게다가 에브리타임을 통해 얻은 정보는 학생들에게 높은 신뢰도를 가지며, 수강 결정에 큰 영향을 미치는 것으로 밝혀졌다[4].

그런데도 지금까지의 서술형 강의평가를 분석한 연구 중 다수는 대학 기관 중심, 즉 교육 서비스 제공자 중심의 전통적 강의평가를 대상으로 해왔으며 에브리타임과 같은 학생 중심의 온라인 강의평을 분석한 연구는 미비한 실정이다[6][7]. 하지만, 대학 기관 중심의 강의평가는 교수자나 학교 기관이 자신의 평가를 읽게 되리라는 사실에 대한 자각하에 잘 다듬은 언어로 작성하여 다소 경량화되고 유해진 평가이므로, 그 분석 결과와 실제 수강생의 정서 간 괴리가 발생할 수밖에 없다[8].

이에 더하여, 전통적 강의평가의 신뢰성에 대한 여러 우려가 존재한다. 먼저, 전통적 강의평가의 서술형 문항은 실효성을 갖지 못한다고 여겨진다. 대학 기관이 강의평가의 정량적인 점수만을 개별 교수에게 전달하고 그들의 업적 평가에 사용하면서, 수강생이 작성한 서술형 평가는 무시되는 경향이 많다[9][10]. 이러한 작태는 강의의 질적 개선과 변화가 어떠한 방식으로 이뤄져야 하는지에 대한 정보를 얻고자 하는 강의평가의 궁극적인 목적을 간과하게 한다. 또한, 학생들이 성적을 열람하기 위해서 강의평가를 작성해야만 하게끔 강제성을 부여하는 대학 기관의 방침은 오히려 강의평가 속 서술형 문항의 의미를 퇴색시키는 결과를 낳았다[11]. 성적 열람을 위한 조건의 하나로 설정된 강의평가는 학생들에게 신속히 완료해야 할 과업으로 인식되며, 그 과정에서 강의를 평가하는 일의 진정한 의미와 목적은 망각하게 된다[12]. 결과적으로, 학생들은 고찰이 없는 단편적 평가 태도를 일관하며 서술형 문항에 대한 응답은 감사와 수고의 말을 전하는 형식적인 답변이 주를 이루게 된다[10].

이러한 맹점을 고려하여 본 연구에서는 수강생, 즉 교육 서비스 이용자의 관점에서 꾸밈없는 언어로 작성된 <에브리타임> 서술형 강의평가 텍스트를 정량적으로 분석하여 이용자가 생각하는 주요 강의평가 요소가 무엇인지 밝히고자 한다. 종합적인 강의평가에 가장 큰 영향을 미치는 요소를 찾고 강의평가 점수가 높은 집단과 낮은 집단 사이의 차이를 살펴봄으로써, 결과적으로 학생을 중심에 둔 대학 강의 개선의 실마리를 찾고자 한다. 강의평가는 수원시에 소재한 A 대학교 영역별 교양 강의에 대한 평가를 대상으로 했다. 이에 따른 본 연구의 구체적인 연구 질문은 다음과 같다.

- 1) 연구 질문 1 : 서술형 강의평가 텍스트에 존재하는 주요 강의평가 요소는 무엇인가?
- 2) 연구 질문 2 : 강의평가에 영향을 미치는 주된 강의평가 요소는 무엇인가?
- 3) 연구 질문 3 : 강의평가 점수가 높은 강의 집단과 낮은 강의 집단 사이에 어떤 차이가 존재하는가?

## II. 이론적 배경

### 1. 서술형 강의평가

수강자의 강의평가 요소는 정량적 평가와 정성적 평가, 그리고 정량적 요소와 정성적 요소를 결합한 평가와 같이 세 가지 유형으로 나누어진다[13]. 가장 일반적으로 사용되는 정량적 평가는 결과 분석의 편리성에 있어서 강점을 가지지만, 한정된 평가 항목에 대한 수적 척도만을 파악하므로 실제로 평가자가 경험한 사고의 맥락이나 세부적인 요구 사항의 결여를 피할 수 없는 등 그 한계가 명확하다[14]. 이러한 점을 보완하여 강의의 질적인 면에서 실질적 변화를 끌어내기 위해서는 정성적 평가 요소의 충분한 활용이 필요하다. 서술형 평가 항목을 통해 학생들의 다양하고 구체적인 의견을 수집하고, 그 결과를 분석함으로써 정확한 개선 지점을 찾아내야 한다는 것이다.

이를 위해 [10][13-19]와 같은 다수의 선행연구에서는 빈도 분석, 토픽 모델링, 상·하위 집단 비교, 언어 네트워크 분석 등의 방법을 사용해 대학 기관 중심의 서술형 강의평가를 분석하였다. 예를 들어, [10]에서는 서술형 강의평가를 구분하기 위해 교수자, 수업, 시설 및 환경, 수업 총평을 상위 영역으로 하는 범주 체계를 도출하고, 이에 대한 빈도 분석을 시행하여 학생들이 강의에 가장 만족한 요인은 '교수자의 내용 설명력'이며 가장 불만족한 요인은 '수업 내용 난이도'임을 밝혔다. [14]에서는 수업, 교수님, 과제, 평가, 방법, 자료, 교육 과정을 상위범주로 하여 도출한 주요 단어를 대상으로 빈도 분석 및 네트워크 분석을 시행함으로써, 학생들은 '교수자와 함께 소통하고 깨달음을 얻어갈 수 있는 수업'을 긍정적으로 평가하며 '비체계적이고 학생의 수준을 고려하지 않는 수업'을 부정적으로 평가한다고 밝혔다. [16]에서는 '강의에서 개선되어야 할 점'과 '강의에서 좋았던 점'에 대한 주제 분석을 통해 전자는 '강의·발표·과제에 대한 개선 사항', '과제 및 실습에 대한 부담', '시험·진도·수업 내용에 대한 개선 사항'의 세 가지 주제를 가지고 있으며, 후자는 '교수자·교수 방법에 대한 긍정적 피드백', '직접적 경험·발표·실습에 대한 긍정적 피드백', '다양한 관점을 부여하는 자료에 대한 긍정적 피드백'의 세 가지 주제를 갖고 있음을 보였다. 전

반적으로 학생들의 긍정적 평가에 영향을 미치는 주된 요인은 교수자와 학생 간의 활발한 소통이며, 부정적 평가에 영향을 미치는 주된 요인은 학생의 수준을 고려하지 않은 수업 내용이라는 결과를 보였다고 할 수 있다. 그러나, 앞서 언급했듯이 교육 서비스 제공 기관이 주도하는 강의평가에는 신뢰성에 대한 여러 문제가 따르게 되므로 분석 결과와 실제 개선점 사이의 왜곡이 발생할 우려가 존재한다. 따라서, 더욱 자율적인 정성 평가를 분석하여 활용해야 할 필요성이 있다. 최근 들어 [6][20]과같이 자율적인 서술형 평가를 얻을 수 있는 대표적 커뮤니티인 에브리타임을 분석한 몇몇 선행연구들이나 [8]과같이 학생들이 자유롭게 참여하는 서술형 강의평가를 분석한 선행연구가 이루어지기도 하였으나, 그 수는 전통적 강의평가에 비해 현저히 떨어지는 실정이다. 이에 본 연구에서 이러한 부족함을 보충하고자 한다.

### 2. 텍스트 마이닝

텍스트 마이닝(text mining)이란 자연어 텍스트를 이용해 유의미한 지식을 얻어내는 빅데이터 분석기법 중 하나이다[21]. 텍스트 마이닝은 관점과 사용 의도에 따라 비구조적인 정보를 구조적인 정보로 정리하여 도출하는 정보 추출(information extraction)과 방대한 데이터 모음에서 특정한 패턴을 파악하는 데이터 마이닝(data mining), 그리고 기존의 데이터베이스에서 새로운 정보와 지식을 발굴하는 KDD(Knowledge Discovery in Databases)로 나누어진다[22]. 텍스트 마이닝 기법을 사용한 일반적인 분석은 텍스트 데이터의 수집과 정제, 그리고 이러한 작업을 통해 완성된 코퍼스(corpus)로부터 의미 있는 정보를 추출하는 과정을 거친다[17]. 텍스트 마이닝의 높은 활용성은 질적 연구가 지배적인 분야에서 새로운 가능성을 보여주고 있다[19]. 본 연구에서는 단어출현 빈도 분석(frequency analysis)과 의 사결정나무 분석을 사용하여 학습자 중심의 주요 강의 평가 요소를 밝히고자 했다.

먼저, 주요한 평가 요소를 파악하기 위해 강의평가 분류에 따른 특정 단어들의 출현 빈도수를 살펴본다. 이 과정에서 단어 기반 온톨로지(ontology)와 TF-IDF 기법이 적용된다. 온톨로지는 데이터 간 의미 관계

(relation)를 정의한 분류 기준이 담긴 의미지식 베이스이다[23][24]. 본 연구에서는 강의평가 요소를 상위 항목(class)으로, 각 요소에 해당하는 단어들을 하위 항목(properties)으로 하여 단어 간 계층적 의미 관계를 적용한다. 상위 항목은 선행연구를 기반으로, 하위 항목은 상위 항목과의 동시 출현단어(co-occurrence word)를 활용해 추출하였다.

이에 더하여, TF-IDF(Term Frequency - Inverse Document Frequency)를 반영함으로써 결과의 왜곡을 최소화했다. TF 기법은 문서 내에서 단어의 출현 빈도가 높을수록 그 연관성 또한 높을 것이라는 가설을 전제한다[26]. 문서를 구성하는 단어가 얼마나 많은 정보를 가졌는지를 나타낸다고 할 수 있다[27].

$$TF = tf(t, d) \quad (1)$$

위의 수식에서 TF는 Term Frequency의 약어로 문서 내 단어의 등장 빈도를 의미하며, t는 단어(term), d는 문서(document), tf(t, d)는 특정 문서 d에서 단어 t가 등장한 횟수를 의미한다. 하지만 위의 수식에서 TF 값은 단지 문서 d 내에서 단어 t가 몇 번 출현했는지 나타낸 단어 빈도이기 때문에, 단순한 빈도에 불과한 TF 값은 관사, 전치사 등 문서에 반복적으로 나타날 수밖에 없는 기능어(function words)의 연관성을 높게 판단하는 치명적인 오류를 발생시킨다. 이를 보완하기 위해 IDF 기법을 사용할 수 있다.

$$IDF = \log \frac{D}{1 + df(t)} \quad (2)$$

위의 수식에서 IDF는 Inverse Document Frequency를 의미하며 D는 전체 문서의 총합, df는 문서의 빈도(document frequency), t는 단어(term), df(t)는 단어 t가 등장한 문서의 총합을 의미한다. 다시 말해, IDF 값은 전체 문서의 수를 단어가 등장한 문서의 수로 나눈 뒤 log를 취한 역문서 빈도를 의미한다[28].

$$TF-IDF = tf(t, d) \times \log \frac{D}{1 + df(t)} \quad (3)$$

결과적으로 TF-IDF는 TF 값과 IDF 값의 곱으로, 여러 문서 중 하나의 문서에서 해당 단어가 어느 정도의 중요성이 있는지 수치화하여 나타낸다[28]. TF-IDF를 적용함으로써 중요성과는 무관하게 출현 빈도가 높을 수밖에 없는 기능어의 특성을 고려하고, 이로 인한 결과상의 오류를 없앨 수 있는 것이다.

한편, 의사결정나무 분석은 의사결정 규칙(decision rule)을 도표화하여 관심 대상이 되는 집단을 몇 개의 소집단으로 분류(classification)하는 분석 방법이다[29]. 본 연구에서는 강의평가 절대 점수(강의 평점 \* 리뷰 수)에 영향을 미치는 강의평가 요소를 파악하기 위해 의사결정나무 분석을 수행하였으며, 이를 위해 목표변수(종속변수)와 예측변수(독립변수)로 연속형 데이터(양적변수)를 다루고 분리개수가 이진분리(Binary split)를 따르는 CART(Classification And Regression Trees) 분석 방법을 사용하였다.

### 3. 온톨로지를 활용한 강의평가 분류

교육 서비스 이용자의 관점에서 가장 주요한 강의평가 요소를 알아내기 위해서는 서술형 강의평가 속 문장들이 어떤 요소를 포함하고 있는지 분류할 일정한 기준이 전제되어야 한다. 이러한 분류 기준은 앞서 언급했던 단어 기반 온톨로지의 상위 항목(class)을 의미한다. 본 연구는 유의미한 강의평가 기준 상정을 위해서 선행연구들을 참고하였다. 심리 측정 분석에 기반한 학생평가 설문지 SEEQ(Student Evaluation of Educational Quality)를 기반으로 서술형 강의평가를 분석한[13], 강의평가 도구 개선안을 연구한[30]과 이를 기초로 하여 대학 강의평가 도구의 필수 요인을 고찰한[31], 강의평가 도구의 문항 개선안을 도출한[32], 좋은 수업이 가진 세부 요소를 분석한[8]과 이를 기반으로 강의평가 결과 패턴을 분석한[33], 델파이 조사를 통해 강의평가 하위요인을 정의한[34]를 참고하여 선정된 강의평가 요소의 분류 기준, 즉 본 연구에 사용된 온톨로지의 상위 항목(class)은 다음[표 1]과 같다.

표 1. 강의평가 요소의 분류 기준

강의평가 요소 : 조작적 정의	근거
강의 : 종합적 강의평가 및 만족도	[8] 김은영(2021), [30] 박해림(2012), [31] 오숙영(2015), [32] 김경연 외(2018), [33] 김세영 외(2022), [34] 박준명 외(2022)
성적 : 성적 적합성 및 공정성	[8] 김은영(2021), [30] 박해림(2012), [31] 오숙영(2015), [32] 김경연 외(2018), [33] 김세영 외(2022), [34] 박준명 외(2022)
과제 : 과제 난이도 및 부담 정도, 효과성 및 적합성	[8] 김은영(2021), [13] 자정규(2022), [30] 박해림(2012), [31] 오숙영(2015), [33] 김세영 외(2022), [34] 박준명 외(2022)
시험 : 시험 난이도 및 부담 정도, 효과성 및 적합성	[13] 자정규(2022), [30] 박해림(2012), [33] 김세영 외(2022)
교수 : 교수자 준비 정도, 열정 및 성실성, 설명력	[8] 김은영(2021), [13] 자정규(2022), [30] 박해림(2012), [31] 오숙영(2015), [32] 김경연 외(2018), [33] 김세영 외(2022), [34] 박준명 외(2022)

전체 코퍼스는 어절 수로 116,929개 크기이다. 해당하 는 영역별 강의 목록은 [표 2]와 같다.

표 2. 수집된 영역별 강의 목록

영역 (강의 수)	강의
역사와 철학 (13)	고대문명사, 과학과 철학, 동양고전철학, 서양고전철학, 역사와 신화, 윤리적추론, 인류학이란 무엇인가, 전환기 근대사의 조명, 조선실록과 민간역사, 철학이란 무엇인가, 포스트 모더니즘, 한국의 사상과 전통, 현대사회의 윤리
문학과 예술 (10)	동양미술사, 디자인과 사회, 문학과 미디어, 서양미술사, 서양연극사, 아시아영화의 이해, 예술이란 무엇인가, 한국의 고전문학, 한국의 현대문학, 현대연극의 이해
인간과 사회 (19)	과학기술과 법, 과학기술과 법, 미래산업혁명과 기술창업론, 미래산업혁명과 기술창업론, 사고와 학습의 심리학, 색수열리터와 현대사회, 여성학, 인간관계 심리학, 인권과 헌법, 일본입문, 일본입문, 정치학이란 무엇인가, 중국인의 삶과 문화, 중국입문, 창의적 사고, 창의적 사고, 통일 한국의 이해, 현대의 시민생활과 법, 현대의 시민생활과 법
자연과 과학 (8)	과학과 종교, 과학사, 기술과 사회, 에너지와 사회, 우주란 무엇인가, 인간과 두뇌의 신비, 자연과학의 이해, 현대물리의 이해

### III. 연구 방법

본 연구의 개요는 다음 [그림 1]과 같다.

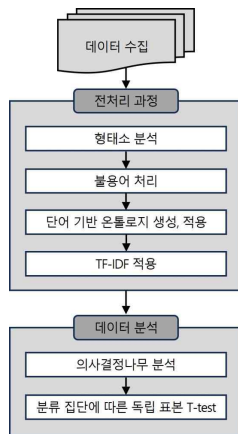


그림 1. 본 연구의 분석 흐름도

#### 1. 데이터 수집과 전처리

본 연구의 데이터는 수원시에 소재한 A 대학교 영역 별 교양, 총 4개 영역에 해당하는 강의를 대상으로 수집되었다. 영역별 교양은 강의 영역에 따라 ‘역사와 철학’, ‘문학과 예술’, ‘인간과 사회’, ‘자연과 과학’으로 나뉜다. 확대 해석의 여지를 제거하기 위해 A 대학교에 브리타임에서 강의평이 30개 이상 등록된 강의를 추려내어, 총 50개 강의의 3,350개 강의평을 추출하였다.

말뭉치 수집에는 기본적으로 Python의 selenium 패키지와 BeautifulSoup 패키지를 이용한 웹 스크래핑 기법이 사용되었다. 스크래핑을 진행하면서 스크롤(scroll) 양의 차이로 인해 데이터 누락이 발생할 때는 HTML 파일을 저장해 텍스트를 복사해 오는 방식을 통해 따로 수집했다. 또한 수업 정보와 수업 평점, 해당 웹 페이지의 주소 등 부가적인 정보를 함께 추출해 연구에 활용하였다.

이렇게 수집된 텍스트 데이터는 더욱 선명한 분석 결과를 위해 전처리 과정을 거친다. 먼저, Python을 사용해 특수 문자와 불필요한 띄어쓰기를 삭제하는 등 기초 정제를 진행했다. 정제된 텍스트를 대상으로, Python 형태소 분석기 kiwi를 사용한 형태소 분석을 통해 명사, 동사, 형용사, 부사 등의 문맥어(content words)를 제외한 불용어를 제거하였다.

#### 2. 단어 기반 온톨로지 생성

본 연구에서는 단어 기반 온톨로지를 사용해 단어 간 계층적 의미 관계를 적용한다. 선행연구를 기반으로 선정된 총 5개의 강의평가 요소(강의, 성적, 과제, 시험, 교수)를 상위 항목(class)으로, 각 요소에 해당하는 단어들을 하위 항목(properties)으로 한다. 하위 항목은 상위 항목인 각 강의평가 요소의 이름과 동시에 출현하

는 빈도가 잦은 단어에서 의미가 겹치는 것을 제외하며 추출하였다. 추출된 동시 출현단어 목록은 [표 3]과 같다.

표 3. 추출된 동시 출현단어 목록

상위 항목	하위 항목 (동시 출현 빈도수)
강의	총 77개 단어 : 강의(716), 수업(455), 내용(192), 교양(133), 과목(86), 학기(84), 추천(76), 사람(70) 등
성적	총 18개 단어 : 성적(311), 학점(51), 점수(51), 점(47), 평가(28), 기준(20), 만점(12), 평(11) 등
과제	총 26개 단어 : 과제(999), 대체(101), 발표(85), 팀플(82), 제출(72), 토론(66), 조(60), 보고서(52), 개인(49) 등
시험	총 29개 단어 : 시험(892), 기말(201), 중간(172), 문제(160), 소스(60), 퀴즈(50), 오픈(47) 등
교수	총 30개 단어 : 교수(984), 말(105), 친절(83), 질문(76), 열정(76), 설명(74), 말씀(48), 피드백(41), 배려(41) 등

강의평가 텍스트 속 단어가 완성된 온톨로지 사전에 존재하면 카운트함으로써 각 문장에 적절한 강의평가 요소가 태그(tag)되도록 하고, 이 결과를 기반으로 빈도를 산출하였다.

이러한 단어 기반 온톨로지의 신뢰도를 파악하고자 Cohen's kappa 점수를 계산하는 과정을 진행하였다. Cohen's kappa란 2명의 검수자 또는 측정자가 한 종류의 데이터 묶음을 검정하는 측정 방식을 의미한다 [27]. 이를 위해 코퍼스에서 무작위로 300개의 강의평가 문장을 추출한 후, 연구자가 개별적으로 각 텍스트에 해당하는 강의평가 요소를 파악하여 태깅(tagging) 작업을 진행했다. 이를 사용해 kappa 값을 도출한 결과, 가중치를 포함하지 않은 값은 0.74, 가중치를 포함한 값은 0.84로, 상당한(substantial) 신뢰도를 보여주고 있음이 확인되었다.

### 3. 의사결정나무 분석

의사결정나무는 의사결정 규칙을 도표화하여 종속변수를 분류 기준에 따라 여러 집단으로 분류하거나 예측을 수행하는 분석 방법으로, 방대한 양의 데이터에서 종속변수의 목표값에 도달하기 위해 영향을 미치는 변수들을 도출하고 최적의 분리 기준을 찾아 의사결정에 도움을 주는 과정이다[29]. 본 연구에서는 강의평가에 영향을 미치는 요인으로 [표 3]과같이 5개의 강의평가 요소(강의, 성적, 과제, 시험, 교수)를 독립변수로 사용

하였으며, 강의평가에 대한 종속변수로는 같은 평점이 라도 표본 집단의 수가 다른 경우, 표본 수에 영향을 받지 않는 대표 점수를 산출하여 사용해야 한다는 사전 연구의 결과를 참고하여 강의 평점과 리뷰 수를 곱한 점수를 강의평가 절대 점수로 사용하였다[29][48][49]. 위 변수들을 기반으로 수행된 의사결정나무 분석을 통해 본 연구는 강의평가 절대 점수(강의 평점 \* 리뷰 수)에 영향을 미치는 주요 강의평가 요소 및 최적의 분리 기준을 도출하고, 강의평가 절대 점수가 높은 강의 집단과 낮은 집단 사이의 차이를 분석하고자 하였다. 분석에 앞서, 총 50개 강의에 대하여 강의명과 그 강의에 대응하는 ID 값, 강의가 속한 학문 영역, 5개의 강의평가 요소값(TF-IDF 값), 강의 평점, 리뷰 수, 강의평가 절대 점수를 표기한 데이터프레임을 구축했다.

## IV. 연구 결과

### 1. 서술형 강의평가 텍스트에 존재하는 주요 강의평가 요소는 무엇인가?

표 4. 변수별 기술 통계량

변수 구분	변수명	평균	중위값	최대값	최소값
종속 변수	강의평가 절대 점수	283.0	227.8	712.5	9.0
독립 변수	강의	1.182	1.183	1.701	0.540
	성적	0.734	0.710	1.448	0.404
	과제	0.991	0.962	1.550	0.415
	시험	1.306	1.299	1.755	0.571
	교수	1.396	1.413	2.110	0.656

본 연구는 선행연구를 참고하여 단어 기반 온톨로지를 생성하고, 추출된 강의평가 텍스트 데이터에서 5개의 강의평가 요소(강의, 성적, 과제, 시험, 교수)에 따른 값을 산출하였다. 생성된 강의평가 요소와 에브리타임에서 추출한 강의평가 절대 점수(강의 평점 \* 리뷰 수)의 기술 통계량 값은 [표 4]와 같다.

강의평가 절대 점수는 평균 283, 최대값 712.5, 최소값 9로 분석에 사용된 강의 간 절대 점수에 편차가 있

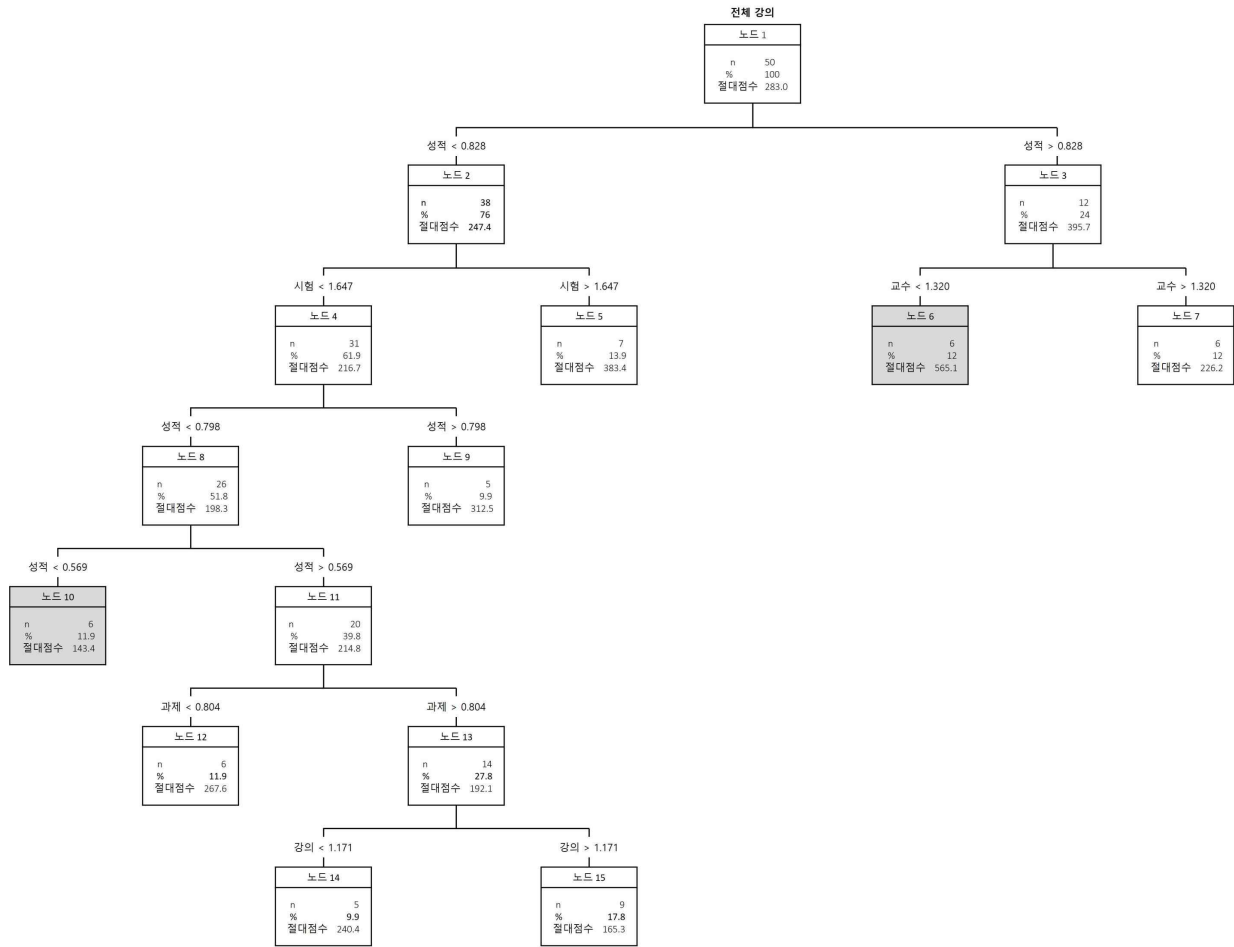


그림 2. 전체 강의평가에 대한 의사결정나무 분석 결과

다고 할 수 있다. 추가로 5개 강의평가 요소값의 평균은 교수(1.396), 시험(1.306), 강의(1.182), 과제(0.991), 성적(0.734)순이었고, 이는 강의를 수강한 학생들의 서술형 강의평가 텍스트에서 교수와 관련된 내용이 주를 이루었으며 다음으로 시험, 강의, 과제, 성적의 순으로 언급량이 많았다는 것을 나타낸다.

## 2. 강의평가에 영향을 미치는 주된 강의평가 요소는 무엇인가?

두 번째 연구 질의에 대한 답변을 위해 본 연구에서는 분류 기준에 따라 여러 집단으로 분류하거나 예측을 수행하는 의사결정나무 분석을 채택하여 분석을 진행하였으며 [표 4]에서 제시된 변수 중 강의평가 절대 점수를 종속변인, 5개의 강의평가 요소(강의, 성적, 과제, 시험, 교수)를 독립변인으로 하여 분석을 진행하였다.

전체 강의평가 데이터 집단에 대한 최적 분리는 성적에 의해 최초 이지분리되었다. 강의평가 절대 점수값이 가장 높다고 예측된 집단에 대한 해석은 다음과 같다. 강의평가 절대 점수값이 565.1이 되기 위해서는 성적 0.828을 기준으로 최초 분리되어야 하며, 성적이 0.828 이상일 경우 교수 1.320을 기준으로 다시 분리된다. 교수가 1.320 이하일 때의 집단(N=6)에 대한 강의평가 절대 점수값은 565.1로 높게 분류된다. 분석된 의사결정나무 분석 결과는 [그림 2]와 같다.

## 3. 강의평가 점수가 높은 강의 집단과 낮은 강의 집단 사이에 어떤 차이가 존재하는가?

본 연구에서는 의사결정나무 분석 결과를 기반으로 강의평가 절대 점수값이 높은 집단과 낮은 집단을 선정하였으며, 두 집단 간 5개 강의평가 요소 사이에 유의



미한 차이가 있는지를 확인하고자 독립표본 T-test를 수행하였다. 강의 평점이 높은 집단과 강의 평점이 낮은 집단에 대해 평균 비교 분석(Independence T-test)을 시행한 결과, [표 5]와 같이 교수, 과제, 성적 항목에서 두 집단 간 신뢰수준 95%의 유의미한 차이가 있다는 결과를 보였다.

표 5. 독립표본 T-test 결과

(\* : 90% confidence level, \*\* : 95% confidence level)

평가 항목	P-값	등분산 가정	T-값	P-값	결과
H_교수 * L_교수	0.063	등분산	-3.061	0.012	채택**
H_시험 * L_시험	0.820	등분산	-0.903	0.387	기각
H_강의 * L_강의	0.893	등분산	1.592	0.142	기각
H_과제 * L_과제	0.928	등분산	3.175	0.009	채택**
H_성적 * L_성적	0.601	등분산	10.294	0.001	채택**

이에 더하여, 본 연구에서는 평균 비교 분석 결과에 대한 사후 분석으로 비교 막대그래프를 사용한 시각화 분석을 진행하였으며, 분석 결과는 아래의 [그림 3]과 같다.

시각화 분석을 위해 본 연구에서는 각 집단에 포함된 강의의 평균 강의평가 요소값(TF-IDF 값)을 계산하고, 이를 다시 비율로 계산하는 작업을 진행하였으며 산출된 두 집단에 따른 강의평가 요소의 차이를 보여주는 시각화 결과를 생성하였다. 분석 결과, 강의평가 절대 점수가 높은 집단의 경우 강의, 과제, 성적에서 강의평가 절대 점수가 낮은 집단보다 높았으며, 교수, 시험에 대한 요소에서는 강의평가 절대 점수가 낮은 집단이 더 높은 결과를 보여주었다.

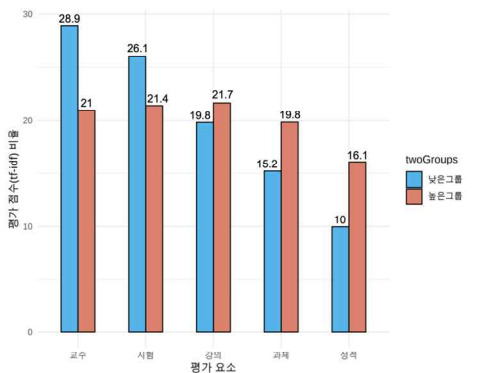


그림 3. 두 집단 간 강의평가 요소 차이에 대한 비교 막대그래프

## V. 결론

본 연구는 학생 중심의 서술형 강의평가 플랫폼인 에브리타임의 강의평가 텍스트를 분석하여 강의평가에 영향을 미치는 주요 평가 요소를 탐색하는 것을 목적으로 수행되었으며, 연구의 결과는 크게 세 가지로 나누어 볼 수 있다. 첫째, 서술형 강의평가 텍스트에 존재하는 주요 강의평가 요소는 강의, 성적, 과제, 시험, 교수로 나누어질 수 있다. 실제로 강의를 수강한 학생들의 서술형 강의평가 텍스트에서 교수와 관련된 내용이 주를 이루며, 다음으로 시험, 강의, 과제, 성적이 순으로 구성되어 있음을 확인하였다. 둘째, 종합적인 강의평가에 영향을 미치는 주된 강의평가 요소는 성적과 교수이다. 본 연구에서는 이를 검증하기 위해 의사결정나무 분석을 사용하였으며 분석 결과, 강의평가 요소 중에서도 성적에 대한 언급이 많고, 교수에 대한 언급이 적은 강의의 강의평가 절대 점수가 높다는 결과를 확인하였다. 셋째, 강의평가 절대 점수가 높은 강의 집단과 낮은 강의 집단 사이에 교수, 과제, 성적 항목에 있어 유의미한 차이가 있다. 본 연구는 독립표본 T-test를 사용하여 의사결정나무 분석을 통해 도출된 강의평가 절대 점수가 높은 집단과 낮은 집단 사이에 강의평가 요소에 따른 차이를 분석하였다. 분석 결과, 두 집단의 강의평가 요소 중 교수, 과제, 성적에서 유의미한 차이가 있었으며, 교수에 대한 언급이 적고 과제와 성적에 대한 언급이 높을수록 강의평가가 좋다는 결과를 제시하였다.

이에 더하여, 본 연구에서는 분석 결과를 확인하기 위해 강의 평점이 높은 집단과 강의 평점이 낮은 집단에 포함된 강의의 사용자 평가를 확인하였다. 그 결과, 강의 평점이 높은 집단의 경우 강의, 과제, 성적과 관련하여 좋은 의견이 주를 이루었으며(e.g., 실제 사용자 평가 : “이 강의 듣고 역사에 대해 흥미가 생겼습니다!!(23년 1학기 수강자)”, “과제도 없고 시험도 정직하게 중간 기말 두번만 칩니다.(23년 1학기 수강자)”, “짱짱 성적도 잘춤 시간 좀 투자하면 좋은 결과 나옴(21년 2학기 수강자)”), 강의 평점이 낮은 집단의 경우 교수, 시험과 관련하여 안 좋은 의견이 주를 이루고 있었다(e.g., 실제 사용자 평가 : “교수님 자꾸 삼천포로 빠지심.(23년 1학기 수강자)”, “기말고사는 매우 어이없는



문제가 나왔음(22년 2학기 수강자)”). 이러한 결과는 교수자와의 소통을 긍정적인 강의평가의 핵심 요소로 꼽았던 대부분의 선행연구와는 상이하다고 할 수 있다. 하지만 본 연구가 익명성과 자유성이 보장된 학생 중심의 강의평가 플랫폼인 에브리타임의 서술형 강의평가를 분석했다는 점을 고려하면, 대학 차원의 강의평가를 분석한 결과와는 달리 학생들은 교수자의 특성보다 성적과 같은 실리적인 가치를 더 추구한다고 볼 수 있다.

본 연구는 대학 기관 중심, 즉 교육 서비스 제공자 중심의 전통적 강의평가가 아닌 학생 중심의 온라인 강의평가를 통해 종합적인 강의평가에 큰 영향을 미치는 강의 요소를 분석하였다는 데 있어 의의가 있다. 이를 통해 교수자나 학교 기관에 대한 인식 때문에 대학 차원의 강의평가에서는 미처 드러내지 못했던, ‘실리 추구’라는 학생들의 실제적 욕구를 확인할 수 있었다. 대학 기관은 이러한 맹점을 고려하여, 강의 개선을 위해 에브리타임과 같은 학생 중심 강의평가를 분석하고 그 결과를 적극적으로 반영하는 노력을 기울일 필요가 있을 것이다. 본 연구는 분석 대상이 A 대학교의 서술형 강의평가로 한정된다는 점에서 한계를 가지지만, 향후 연구에서는 일련의 과정을 분석 도구로 시스템화하여 무료 배포함으로써 많은 대학이 간편하게 학생 중심 강의평가를 분석하고 강의의 질을 향상하는 과정에 적용할 수 있도록 하고자 한다.

#### 참 고 문 헌

- [1] 신재철, “대학종합평가와 대학발전: 성과와 과제,” 교육행정학연구, 제22권, 제4호, pp.333-350, 2004.
- [2] J. K. Solomon, *Academic and Service Quality in Distance Education*, University of Denver, PhD Dissrtation, 1998.
- [3] 채수은, 손영민, “형성적 기능 중심의 대학 강의평가 공개를 위한 연구,” 한국교육학연구, 제21권, 제1호, pp.325-358, 2015.
- [4] 김현숙, 김경미, “MZ세대인 대학생들의 수강 신청 의 사결정에 미치는 요인 연구,” 인문사회21, 제13권, 제4호, pp.2329-2342, 2022.
- [5] 최시내, 오상희, “온라인 커뮤니티 에브리타임을 통한 대학생의 일상정보 이용행태에 관한 연구,” 한국도서관·정보학회지, 제52권, 제3호, pp.239-266, 2021.
- [6] 이신행, 이명진, 권경민, 이한준, “토픽모델링을 활용한 강의리뷰 분석,” 한국데이터정보과학회지, 제34권, 제3호, pp.459-466, 2023.
- [7] 안지혜, *좋은 대학수업의 특성분석연구: 인문사회계열 우수 수업사례 관찰결과를 중심으로*, 연세대학교, 박사학위논문, 2012.
- [8] 김은영, “자율적 서술형 강의평가를 통한 '대학 수업'에 대한 학생들의 인식 연구,” 교육문화연구, 제27권, 제2호, pp.217-238, 2021.
- [9] 김선희, “강의평가 결과 분석을 통한 교육의 질 제고 방안 탐색: A대학교 강의평가를 중심으로,” 사회과학연구, 제30권, 제1호, pp.147-174, 2017.
- [10] 하정운, 나민주, “학생의 강의 만족 및 불만 요인은 무엇인가? : A대 서술형 강의평가 자료를 중심으로,” 교육연구논총, 제38권, 제3호, pp.61-77, 2017.
- [11] 하오선, 정민호, “강의평가 응답분석을 통한 강의평가도구 개선방안 - D대학 강의평가 사례를 중심으로 -,” 열린교육연구, 제22권, 제2호, pp.273-294, 2014.
- [12] 손충기, “대학 수업에 대한 학생의 확립적 평정 경향 분석,” 교육평가연구, 제27권, 제1호, pp.67-86, 2014.
- [13] 지정규, “텍스트 애널리틱스를 활용한 서술형 강의평가 분석,” 융복합지식학회논문지, 제10권, 제1호, pp.71-80, 2022.
- [14] 이후희, 김은정, 이상수, 이수상, “대학 서술형 강의평가 자료의 언어 네트워크 분석,” 교육혁신연구, 제28권, 제2호, pp.237-262, 2018.
- [15] 박금주, “대학 소프트웨어 교육의 서술형 강의평가에 대한 네트워크 분석,” Global Creative Leader : Education & Learning, 제9권, 제4호, pp.67-88, 2019.
- [16] 광민호, 민혜리, 김미림, “토픽 모델링을 활용한 대학생의 서술형 강의평가 분석,” 아시아교육연구, 제20권, 제2호, pp.491-522, 2019.
- [17] 김경희, 성진용, 최승배, 강창완, “텍스트마이닝을 이용한 D대학교 서술형 강의평가 내용분석,” Journal of The Korean Data Analysis Society, 제23권, 제6호, pp.2599-2606, 2021.
- [18] 박금주, “서술형 강의평가결과에 대한 단과대별 내용 분석,” 한국산학기술학회 논문지, 제23권, 제1호, pp.428-435, 2022.
- [19] 이상철, “텍스트마이닝을 이용한 서술형 강의평가 분

- 석,” 한국컴퓨터정보학회논문지, 제27권, 제10호, pp.211-222, 2022.
- [20] 한지영, 허고은, “토픽 모델링 기반 비대면 강의평 분석 및 딥러닝 분류 모델 개발,” 한국문헌정보학회지, 제55권, 제4호, pp.267-291, 2021.
- [21] K. B. Cohen and L. Hunter, “Getting started in text mining,” *PLoS Computational Biology*, Vol.4, No.1, e20, pp.1-3, 2008.
- [22] A. Hotho, A. Nürnberger, and G. Paaß, “A brief survey of text mining,” *LDV Forum*, Vol.20, No.1, pp.19-62, 2005.
- [23] Thomas R. Gruber, “A Translation Approach to Portable Ontology Specifications,” *Knowledge Acquisition*, Vol.5, No.2, pp.199-220, 1993.
- [24] 강신재, “의미 분석을 위한 말뭉치 기반의 온톨로지 학습,” 한국산업정보학회논문지, 제9권, 제1호, pp.17-23, 2004.
- [25] H. S. Kim, I. Choi, and M. Kim, “A Statistical Approach for Extracting and Naming Relation between Concepts,” *The KIPS Transactions:PartB*, Vol.12, No.4, pp.479-486, 2005.
- [26] H. Christian, M. P. Agus, and D. Suhartono, “Single Document Automatic Text Summarization using Term FrequencyInverse Document Frequency (TFIDF),” *Mathematics and Engineering Applications*, Vol.7, No.4, pp.285-294, 2016.
- [27] 이종화, 이문봉, 김종원, “TF-IDF를 활용한 한글 자연어 처리 연구,” *정보시스템연구*, 제28권, 제3호, pp.105-121, 2019.
- [28] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information processing & management*, Vol.24, No.5, pp.513-523, 1988.
- [29] 문성민, 이경원, “통계와 시각화를 결합한 데이터 분석: 예측모형 대한 시각화 검증,” *디자인융복합학회*, 제15권, 제6호, pp.197-216, 2016.
- [30] 박혜림, “대학 강의평가 도구 개선 방안 연구-“A” 대학의 강의평가 문항 개선 사례-,” *한국산학기술학회논문지*, 제13권, 제11호, pp.5033-5043, 2012.
- [31] 오숙영, “강의평가도구 구성요인의 이론적 고찰 및 구인타당도 연구: S 대학교 강의평가 사례 중심,” *안암교육학회*, 제21권, 제4호, pp.5-35, 2015.
- [32] 김경연, 우혜정, 김지영, 김우철, “대학 강의평가 도구 개선 및 타당화 연구: K대학 사례를 중심으로,” *한국교육문제연구*, 제36권, 제4호, pp.1-26, 2018.
- [33] 김세영, 김보경, 임유진, 이예경, “교수의 질 제고를 위한 3년간의 강의평가 결과 패턴 및 특징 연구,” *교육과학연구*, 제53권, 제2호, pp.1-26, 2022.
- [34] 박준명, 문경용, 도승이, “학습자가 인식하는 좋은 수업과 학습 부담에 대한 강의평가 문항 개발,” *아시아문화학술원*, 제13권, 제6호, pp.1521-1536, 2022.
- [35] 진주연, 이후경, “국내 통합체육 연구의 동향: 텍스트 마이닝 분석,” *한국특수체육학회지*, 제30권, 제1호, pp.120, pp.117-135, 2022.
- [36] 박지희, 고장완, “학생-교수 상호작용, 학생참여, 대학교육성과의 종단 구조 관계 분석,” *열린교육연구*, 제27권, 제2호, pp.259-284, 2019.
- [37] 전현욱, 채민정, “교육서비스 품질이 대학선택만족도에 미치는 영향: SERVQUAL 모형을 중심으로,” *인문사회21*, 제13권, 제5호, pp.1071-1084, 2022.
- [38] 김정민, 정진성, 정하보, “대학 서술형 강의평가 실태 및 네트워크 분석: S대학교 사례를 중심으로,” *학습자 중심교과교육연구*, 제21권, 제15호, pp.149-164, 2021.
- [39] 박금주, “텍스트 내용분석 방법을 적용한 소프트웨어 교육 요구조사 분석 : A대학을 중심으로,” *한국산학기술학회 논문지*, 제20권, 제3호, pp.65-70, 2019.
- [40] 이해듬, 남민우, “대학 강의평가 주관식 결과의 텍스트 마이닝을 통한 전공 계열별 좋은 수업 특성 분석,” *한국유아교육연구*, 제20권, 제2호, pp.21-41, 2018.
- [41] 장근수, “텍스트 마이닝을 활용한 『일본어교육연구』 주제 분석,” *일본어교육연구*, 제63권, pp.145-160, 2023.
- [42] 이제명, “한국농촌계획 온톨로지 구축을 위한 상호정보 기반 단어연결망 분석,” *농촌계획*, 제23권, 제3호, pp.37-51, 2017.
- [43] H. An and M. Park, “Approaching fashion design trend applications using text mining and semantic network analysis,” *Fashion and Textiles*, Vol.7, No.34, pp.1-15, 2020.
- [44] C. D. Manning, *Introduction to Information Retrieval*, CrossRef, 2008.
- [45] T. Kamada and S. Kawai, “An algorithm for drawing general undirected graphs,” *Information Processing Letters*, Vol.31, No.1, pp.7-15, 1989.

- [46] P. Drieger, "Semantic network analysis as a method for visual text analytics," *Procedia-Social and Behavioral Sciences*, Vol.79, pp.4-17, 2013.
- [47] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, Vol.20, No.1, pp.37-46, 1960.
- [48] Nahm Francis Sahngun, "Understanding Effect Sizes," *hmr*, Vol.35, No.1, pp.40-43, 2015.
- [49] B. Litman, "Predicting Success of Theatrical Movies: An Empirical Study," *Journal of Popular Culture*, Vol.16, pp.159-175, 1983.

문 성 민(Seongmin Mun)

정회원



- 2021년 6월 : Université Paris Nanterre Sciences du Langage (전산언어학 박사)
- 2021년 8월 ~ 2022년 10월 : 조선대학교 영어영문학과 박사후연구원
- 2022년 11월 ~ 현재 : 아주대학교 인문과학연구소 연구교수

〈관심분야〉 : 전산언어학, 디지털인문학, 자연어처리, 통계 분석, 기계학습, 신경망분석, 인공지능, 데이터시각화

저 자 소 개

김 혜 정(Hye-Jeong Kim)

준회원



- 2022년 3월 ~ 현재 : 아주대학교 사학과 학사

〈관심분야〉 : 사학, 데이터인문학

김 민 서(Min-Seo Kim)

준회원



- 2020년 3월 ~ 현재 : 아주대학교 영어영문학과 학사

〈관심분야〉 : 영어영문학, 데이터인문학