

대학 CQI 보고서의 자동 분석을 위한 도메인 사전 구축

김건우*, 박준수*, 서준혁*, 정유철**, 윤현주**

Domain Dictionary Construction for Automatic Analysis of University CQI Reports

Gunwoo Kim*, Junsu Park*, Junhyeok Seo*, Yuchul Jung** and Hyeonju Yoon**

요 약

대학 CQI(Continuous Quality Improvement) 보고서는 매 학기를 마칠 때 교수자 본인이 작성하는 강의개선 보고서이다. 본 논문에서는 실제 CQI 보고서에 사용된 각종 단어를 등장 빈도로 분석하여 상위분류와 하위분류에 따른 도메인 사전을 구축함으로써 추후 CQI 보고서 자동 분석 연구에 사용될 토대를 마련한다.

Abstract

The university CQI report is a lecture improvement report prepared by the instructor himself at the end of each semester. In this paper, various words used in the actual CQI report are analyzed with the frequency of appearance, and a domain dictionary according to the upper classification and the lower classification is constructed to lay the foundation for future CQI report automatic analysis research.

Key words

CQI, University, Analysis, Domain Dictionary, Construction

1. 서 론

CQI는 Continuous Quality Improvement의 약자로, 단어 자체의 의미는 ‘지속적인 품질 개선’이라는 의미를 담고 있다[1]. 고등 교육에서의 CQI 보고서는 교육 수요자인 학생들에게 양질의 교육을 제공하기 위해 작성되는 것으로, 학생과 교수자 본인의 강의에 대한 평가결과를 토대로 다음 학기 강의 내용과

방법의 개선을 위해 사용되고 있다. 이러한 CQI 보고서는 각 항목의 내용이 문장식으로 서술되어 있기에, 많은 교수자의 CQI 보고서가 종합되었을 때 내용을 파악하는 것에 어려움이 있으며, 지속적인 개선 여부와 CQI 보고서의 효용성 여부를 평가하기가 어렵다. 따라서 본 논문에서는 이러한 CQI 보고서의 분석을 용이하게 하고, 더 나아가 자연어처리를 이용한 자동 분석에 사용할 수 있는 단어 사전

*금오공과대학교 컴퓨터공학과 gunwoo45954@kumoh.ac.kr, *금오공과대학교 컴퓨터공학과 pks5294@kumoh.ac.kr

*금오공과대학교 컴퓨터공학과 ssam2s@kumoh.ac.kr

**금오공과대학교 컴퓨터공학과 jyc@kumoh.ac.kr, **금오공과대학교 컴퓨터공학과 juyoon@kumoh.ac.kr

을 구축한다.

II. 단어 사전 구축

각 단어의 의미 세분화를 위해 대학 CQI 보고서의 주 평가 항목인 ‘기자재’, ‘태도’, ‘학업성취도’, ‘수업 방식’을 총 4가지 상위분류로 정했다. 그리고 실제 9학기 데이터 중 ‘수업 방식’에 대한 총 9,072개 문장에 대해 분석을 진행했다. 그림 1은 상위분류 중 ‘수업 방식’에 대한 용어의 빈도를 분석하여 가장 빈번하게 나타난 상위 10개를 나열한 결과이다.

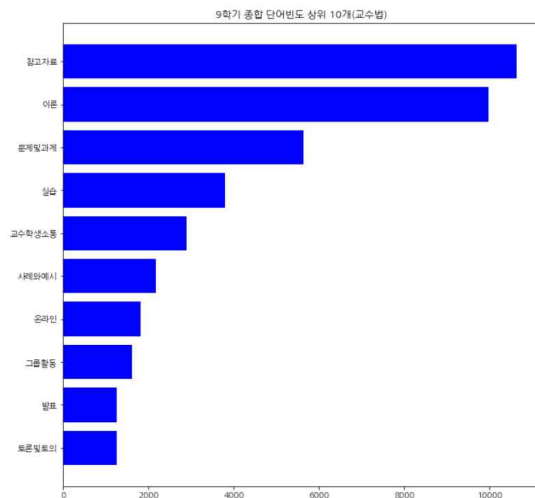


그림 1. 수업 방식에 대한 용어 빈도 분석 결과
Fig 1. Terminology Frequency Analysis Results for Teaching Methods

수업 방식에 대한 용어가 문장에 나타난 횟수는 46,003번이었다. 그중 가장 빈번하게 나타난 상위 3개의 용어는 ‘참고자료’, ‘이론’, ‘문제 및 과제’였으며 위와 같이 모든 대분류에 대한 빈도를 분석하여 상위 3개의 용어를 다음 표 1에 나타내었다.

표 1. 분류별 세분화 단어 사전 예

Table 1. Example of a classified word dictionary

상위분류	하위분류	키워드				개수
기자재	냉난방	냉방기	난방기	에어컨		281
	장비	PC	컴퓨터	프로젝터		
	철판	분필	마커	보드		
	마이크	무선	유선	착용		
	카메라	녹화	웹캠	영상		
태도	긍정	적극적	성실	의욕		44
	부정	불성실	불량	어수선		
학업성취도	긍정	높음	우수	양호		
	부정	낮음	저조	부족함		
수업 방식	이론	개념	설명	예시		321
	시험	쪽지	문제	고사		
	과제	숙제	복습	보고서		
	실습	분석	구현	설계		
	그룹	조별	회의	활동		
	토론	토의	피드백	소통		
계						646

위 표 1에서 학생들의 수업 태도와 학업 성취도는 긍정과 부정으로 이진 분류되기 때문에 긍정 및 부정 표현에 대한 단어들을 서로 합하여 정리하였다. 그 결과로, 총 646개의 단어를 포함한 도메인 사전을 구축할 수 있었다.

III. 결론

본 논문에서는 실제 대학 CQI 보고서 내용을 정제하고 분석하여 주요 단어에 대한 도메인 사전을 구축하였다. 이러한 도메인 사전과 자연어처리 기법을 추후 연구에서 접목한다면, 매 학기 중점적으로 필요한 개선 방안을 모색하거나 효과가 좋았던 교수법을 파악하는 등의 CQI 과정을 진행하는데 있어 도움이 될 것으로 기대된다. 또한, 딥러닝 기반 자연어처리 모델 중 한국어 데이터 사전학습 모델인 KoBERT[2] 에 도메인 사전을 기반으로 학습시켜 Fine-Tuning을 거친다면 교수한 강의 환경이나 내용의 긍정 및 부정 효과를 분석하는데 큰 효과를 보일 수 있을 것이다.

참 고 문 헌

- [1] 박동국, 김철신. (2019). 교과목 포트폴리오와 CQI 보고서: 올바른 정의(定意)와 구현 방향. 사회과학연구, 58(1), 79-93.
- [2] KoBERT (2022) Repository [Source Code]
<https://github.com/SKTBrain/KoBERT>