

R을 이용한 과학교육 연구동향 분석 방법론

가석현*, 장은진, 김찬중, 최승연
서울대학교 (shga@snu.ac.kr)

요약

연구 동향 분석과 관련한 기존의 연구는 대부분의 경우 연구자가 직접 논문을 읽고 분석한 내용을 기초적 수준의 통계 프로그램으로 처리하는 방법을 사용하고 있다. 그러나 인간이 처리할 수 있는 데이터양의 한계로 연구 대상으로 다룰 수 있는 논문의 수에 한계를 드러냈으며, 분석 방법 또한 단순한 분류에 그치는 수준이 많았다. 따라서 본 연구는 이러한 한계를 극복하기 위해 R을 이용하여 과학교육의 연구 동향을 분석할 수 있는 유용한 방법론적인 도구를 개발하고자 하였다.

분석 방법 개발에 있어서 우선 빅 데이터 사용하는 통계 프로그램 중 가장 일반적인 프로그램인 R을 이용하여 데이터 수집 및 분석은 물론 데이터 시각화에 걸친 모든 과정을 처리하도록 제작하였다. 데이터 수집 과정에서 각 데이터 간의 상관관계를 다각도로 분석할 수 있도록 데이터를 DB로 구조화하였고, 키워드를 모아 사전 DB를 확보하였다. 그리고 이를 기초로 각 논문들의 키워드 상관도를 계산하고 이를 네트워크 그래프로 도시하도록 설계하였다. 또한 구축한 분석 방법의 실효성을 살펴보기 위하여 분석 대상으로 Wiley Online Library에 게시된 2006년 1월 ~ 2016년 3월까지의 JRST 저널 논문 정보(논문 제목, 저자, 권, 호, 키워드, 초록)를 정하여 HTML Parsing을 통해 수집한 정보를 분석하여 각 논문들의 키워드 상관도를 계산하고 이를 네트워크 그래프로 도시해보았다.

이러한 형태의 연구에서는 비구조화된 텍스트를 컴퓨터가 기계적으로 판단하는 것에 대한 부정확성이 문제로 지적될 수 있으나, 파일럿 테스트를 통한 기계적 판단 모델의 정교화를 통해 극복될 수 있을 것이라 기대한다.

주요어 : R, 연구동향