

교육학 분야 국내 학술 논문 저자키워드 출현 순서에 따른 빈도가 저자키워드의 논문 제목 출현 빈도에 미치는 영향에 관한 연구

A Study on the Effects of the Appearance Sequence of Author Keyword on the Appearance Frequency of Article Title in the Education

윤은빈, 성균관대학교 문헌정보학과, yeb2013@naver.com

Eun-Bin, Yoon, Department of Library and Information Science, SungKyunKwan University

본 연구는 교육학 분야에서 국내 학술 논문 저자키워드 출현 순서에 의미가 있는지 알아보기 위하여 저자키워드 출현 순서별로 빈도를 조사하고 동 저자 키워드가 제목에 출현하는지를 살펴본 후, 학술 논문 저자키워드 출현 순서에 따른 빈도가 저자키워드의 논문 제목 출현 빈도에 미치는 영향에 대해 연구하고자 한다. 연구의 배경은 구조적 학술용어사전의 구축과 관련해서 효율적인 용어 선정이나 용어 분류 방법을 찾는 것이다. 연구의 목적을 위해 교육학 분야 저자키워드의 출현 빈도수가 1인 것을 제외하고 고빈도 순으로 400개 용어를 추출하였다. 저자키워드의 몇 번째 출현 순서가 저자키워드의 제목 출현에 가장 큰 영향을 미치는지 알아본 결과, 저자키워드의 첫 번째 출현이 아닌 두 번째 출현이 가장 큰 영향을 미친다는 것을 알 수 있었다.

1. 서론

논문을 검색하는 대부분 사용자는 논문 제목을 찾기 보다는 원하는 주제와 내용을 담고 있는 논문을 찾길 원한다. 그렇기 때문에 논문 검색에서 중요한 요소는 논문의 주제어인 저자키워드라 할 수 있다. 저자키워드는 해당 논문의 내용에서 가장 핵심적이고 중요하다고 판단하여 추출한 용어이므로 의미적인 면에서 해당 논문의 주제와 밀접하게 관련되어 있다(이춘실, 문혜원, 2000). 저자키워드는 논문 제목에 출현하기도 하지만 모든 학술 논문의 저자키워드가 그러한 것은 아니다.

학술 논문의 저자키워드는 ‘구조적 학술용

어사전’을 구축하는데 있어 추출 대상 용어군으로 가장 많이 사용된다. ‘구조적 학술용어사전’이란 학술 용어를 유사 개념으로 분류하여 범주화하고 각 개념들이 가지는 속성을 체계화하는 하는 사전을 말한다.

본 연구는 구조적 학술용어사전 내 선정 용어수가 가장 많은 교육학 분야를 조사하고자 한다. 논문 저자키워드 출현 빈도가 1인 것을 제외하고 빈도가 높은 순으로 정렬하였을 때 상위 400개 용어를 추출하였다. 본 연구의 결과는 구조적 학술용어사전의 효율적인 구축을 위해서 하나의 도구로 활용될 수 있을 것이다. 그러나 교육학 분야의 용어만을 대상으로 하였다는 점에서 한계가 있다.

2. 선행 연구

문헌정보학 분야에서는 학술 용어를 학문적 관심사로 중요하게 여기고 있다. 기존의 연구는 주제어 사이에서 나타나는 네트워크 형성과 분야별로 주제어 사용의 차이를 알아보는 것으로 논문의 주제어를 연구대상으로 삼고 있다. 이렇게 저자키워드를 연구대상으로 한 연구는 꽤 많지만, 저자키워드 출현 순서에 관한 연구는 많지 않다.

박옥남(2011)은 국내 정보조직분야 지식구조를 조사하기 위해 내용분석과 네트워크 분석을 사용하였다. 2000년부터 2011년까지 정보조직분야 논문을 바탕으로 저자키워드, 초록, 저자, 저자소속기관을 추출하여 용어의 빈도를 측정하고 용어 간 연관관계를 분석하였다. 이를 위해 네트워크상에서 각 노드의 연결중심성, 근접중심성, 위세중심성을 산출한 결과 정보조직 연구 주제어는 좁고 복잡한 네트워크를 형성하고 있으며 주제어 사이에 직접적인 연결이 많이 이루어지고 있음을 알 수 있다.

이혜영, 광승진(2011)은 국내 연구자들이 학술지 논문에 나타난 학문분야의 상호관계성과 연관성에 대하여 연구하였다. 실험대상 학술논문의 저자가 제시한 주제어에 초점을 두었고 실험대상 자료는 2009년 한국연구재단 등재학술지 80종으로 학술논문 7,616건, 주제어 58,143건이다. 자료를 분석한 결과 각 연구분야 8개(공학, 농수해양, 복합학, 사회, 예술체육, 의약학, 인문, 자연과학 등)가 주제어 사용에서 차이점을 보였다. 공학분야의 주제어는 다른 연구 분야에서도 주제어로 많이 사용되고 있다는 결과를 보였다.

양창진(2010)은 학술 논문의 주제어를 연구대상으로 하여 주제어 기술의 특징에 초점을 맞추어 연구하였다. 우리나라의 대부분 학회에서 학술 논문의 주제어를 수록하는 일이 전

적으로 저자에게 맡겨져 있기 때문에 비슷한 주제의 논문일지라도 주제어 자체가 다양하게 나타나고 있음을 말하였다. 또한 주제어 기술이 저자의 학문적 배경과 지식을 기반으로 하기 때문에 표준화되고 정제된 주제어 목록과 색인을 구축하도록 제안하였다.

3. 연구 방법

3.1 연구 대상

본 연구는 국내 학술 논문 저자키워드의 출현 순서별로 빈도를 알아보고 그것이 논문 제목 출현 빈도에 미치는 영향을 밝히는 것을 목적으로 한다. 이러한 목적에 부합하는 연구대상을 선정하기 위하여 용어를 정제하여 추출하였다. 성균관대학교 정보관리연구소에서 한국연구재단의 토대연구사업 지원에 의해 수행 중인 구조적 학술용어사전 데이터베이스 시스템 "STNet(www.stnet.re.kr)" 구축을 위해 선정된 6만 여 개의 용어를 정제하였다. 그 중에서도 본 연구는 선정 용어 수가 가장 많은 교육학 분야의 저자키워드를 추출하였다.

본 연구의 대상은 교육학 분야에서 총 빈도 수가 20이상인 상위 400개의 저자키워드이다. 영문으로 된 용어 또는 영문을 포함한 용어인 'ADHD', 'PCK', 'e-러닝', 'e-learning', 'PBL', 'HRD', 'early childhood teacher', 'teacher education' 총 8개 용어를 대상에서 제외하였으나 우리말로 표기된 용어 '이러닝'은 대상용어에 포함하였다. 저자키워드의 출현 빈도를 첫 번째 자리부터 다섯 번째 자리까지 보았으며 '확인적요인분석', '위계적선형모형', '초등영어교육'과 같은 용어는 논문 제목 출현 빈도가 0이었다.

<표 1> 교육학 분야의 상위 400개 저자키워드

일련 번호	키워드	제목 출현 빈도	저자키워드 출현 빈도				
			첫 번째 자리	두 번째 자리	세 번째 자리	네 번째 자리	다섯 번째 자리
1.	다문화교육	148	139	80	34	25	7
2.	통합교육	113	91	64	38	34	10
3.	창의성	151	74	53	57	23	11
:							
40.	스트레스	66	14	25	19	7	6
41.	이러닝	43	28	16	12	9	6
:							
256.	확인적요인분석	0	0	2	7	12	4
:							
278.	위계적선형모형	0	1	3	9	7	6
:							
300.	자기평가	7	3	5	5	3	4
:							
368.	초등영어교육	0	9	6	4	2	0
:							
400.	교육개혁	7	7	5	6	3	1

(출처: STNet시스템 <http://www.stnet.re.kr>)

3.2 자료의 분석 방법

본 연구는 SAS 통계 프로그램을 사용하여 수집된 데이터를 분석하고 검증하였다. 논문 저자키워드의 출현 순서별 빈도가 저자키워드의 논문 제목 출현 빈도에 영향을 미치는지 알아보기 위해 다중회귀(REGRESSION) 분석을 하였다. 이상의 통계 분석은 유의수준 $p < 0.05$ 에서 검증하였다.

4. 분석 결과

학술 논문 저자키워드의 출현 순서별 빈도가 저자키워드의 논문 제목 출현 빈도에 미치는 영향을 설명하기 위해 저자키워드의 출현 순서별 빈도는 여러 개의 독립변수, 저자키워드의 제목 출현 빈도는 종속변수로 다중회귀(REGRESSION) 분석을 하였다. 분석 결과는 <표 2>와 같다.

<표 2> 논문 제목 출현 빈도에 미치는 영향

요인	비표준화계수		t	p Value
	β	표준오차		
첫 번째 출현 빈도	.695	.052	13.42	.000
두 번째 출현 빈도	.720	.106	6.79	.000
세 번째 출현 빈도	.909	.118	7.70	.000
네 번째 출현 빈도	.197	.156	1.26	.210
다섯 번째 출현 빈도	.447	.217	2.06	.040

*종속변인: 저자키워드의 논문 제목 출현 빈도

학술 논문 저자키워드의 출현 순서별 빈도가 저자키워드의 논문 제목 출현 빈도에 미치는 영향을 보면 첫 번째, 두 번째, 세 번째, 다섯 번째 출현 빈도는 p값이 유의수준 0.05보다 작으므로 통계적으로 유의한 것으로 나타났다. 또한 저자키워드의 출현 순서별 빈도는 저자키워드의 논문 제목 출현 빈도에 정(+)적인 영향을 미치는 것으로 나타나 순서별 출현 빈도가 높을수록 제목 출현 빈도가 높아짐을 알 수 있다. 본 연구는 저자키워드의 논문 제목 출현 빈도에 가장 영향을 미치는 저자키워드 출현 순서를 알고자 하여 이어서 단계적 다중회귀(STEPWISE REGRESSION) 분석을 하였다. 분석 결과는 <표 3>과 같다.

<표 3> 논문 제목 출현 빈도에 영향을 미치는 요인에 대한 순위

요인	순위	F	p Value
첫 번째 출현 빈도	2	62.61	.000
두 번째 출현 빈도	1	834.78	.000
세 번째 출현 빈도	3	162.89	.000
다섯 번째 출현 빈도	4	9.97	.002

<표 3>을 보면 F값을 알 수 있으며 저자키워드의 두 번째 출현 빈도가 저자키워드의 논문 제목 출현 빈도에 가장 큰 영향을 미치는 것으로 순위가 나타났다. 그리고 논문 제목 출현 빈도에 영향을 미치는 요인으로 네 번째 출현 빈도를 제외하고 첫 번째, 두 번째, 세 번째, 다섯 번째 출현 빈도만이 나타났다. 따라서 학술 논문 저자키워드의 첫 번째 출현 빈도, 두 번째 출현 빈도, 세 번째 출현 빈도, 다섯 번째 출현 빈도만이 저자키워드의 논문 제목 출현 빈도에 영향을 미쳤고 가장 큰 영향을 미치는 저자키워드의 출현 순서는 첫 번째가 아니라 두 번째이다.

5. 결 론

본 연구는 저자키워드 출현 순서별로 빈도를 조사하여 저자키워드가 논문 제목에 출현하는지를 살펴보고 저자키워드 출현 순서에 따른 빈도가 저자키워드의 논문 제목 출현 빈도에 미치는 영향에 대해 연구하였다.

저자키워드의 논문 제목 출현 빈도에 미치는 요인으로 저자키워드의 네 번째 출현 빈도를 제외하고 첫 번째, 두 번째, 세 번째, 다섯 번째 출현빈도가 나타났다. 가장 큰 영향을 미치는 것은 저자키워드의 첫 번째 출현이 아닌 두 번째 출현이었다. 그렇기 때문에 학술 논문 저자키워드의 첫 번째 출현은 어떤 것을 결정지을 수 있는 요인이 아니라는 것을 알 수 있다. 이것에 더해 용어의 중요도를 측정하여 연구에 보완한다면 두 번째 출현 빈도수가 높은 용어가 중요한 키워드인지를 알아낼 수 있다. 따라서 학술 논문 저자키워드의 출현 순서와 이에 따른 빈도만으로는 키워드 자체의 중요도, 그 이상을 알아내기는 어렵다.

정보검색에 있어서 저자키워드의 활용은 매

우 중요하기 때문에 용어의 중요도를 측정하는 방법을 고려하여 보완하고 정보검색의 효율성 향상과 관련한 저자키워드 출현에 대한 심층적인 연구가 지속적으로 진행될 필요가 있다.

참고문헌

- 고영만, 송민선, 김비연, 민혜령. 2013. 인문학 및 사회과학 분야 국내 학술논문의 저자키워드 출현빈도와 피인용횟수의 상관관계 연구. 『정보관리학회지』, 30(2) : 227-243.
- 김관준, 이재운. 2012. 디스크립터 자동 할당을 위한 저자키워드의 재분류에 관한 실험적 연구. 『정보관리학회지』, 29(2) : 225-246.
- 박옥남. 2011. 정보조직 지식구조에 대한 연구 : 2000년~2011년 학술논문을 중심으로. 『한국비블리아학회지』, 22(3) : 247-267
- 박현철. 2005. 『시멘틱 웹 환경에서의 온톨로지 기반 논문 검색 시스템』. 석사학위논문, 중앙대학교 대학원, 컴퓨터공학과.
- 양창진. 2010. 학술논문 주제어 표기 및 활용방안 연구. 『인문콘텐츠』, 19(-) : 395-416.
- 이춘실, 문혜원. 2000. 한국 의학학술논문의 저자선정 주제어와 MeSH 용어의 비교분석 연구. 『한국정보관리학회 학술대회논문집』, 7(-) : 67-70.
- 이혜영, 곽승진. 2010. 국내 학술지 논문의 주제어를 통한 학술연구분야 관계분석. 『한국비블리아학회지』, 22(3) : 353-371.
- 정계동, 황치곤, 최영근. 2012. 멀티 온톨로지 기반의 키워드 연관성을 이용한 전문가 검색 시스템. 『한국정보통신학회 논문지』, 16(1) : 183-190.