

# 네트워크 분석을 통한 저자키워드 출현순서에 대한 의미 분석

권선영

한남대학교 문헌정보학과 교수

## A Study on the Application to Network analysis on Importance of Author keyword based on Sequence of keyword

Sun-young Kwon

Professor, Division of Library and Information Science, Hannam University

요 약 본 연구는 학술논문의 출현순서에 따른 저자키워드의 중요도를 측정해보고자 하는 연구이다. 이를 위해 학술논문 저자키워드의 출현순서를 측정한 후 네트워크 분석 방법의 연결정도중심성, 근접중심성, 매개중심성, 위세중심성 지수를 사용하여 저자키워드의 중요도를 측정해보았으며 다음으로 각각의 네트워크지수 값과 저자키워드의 출현순서와의 상관관계 분석을 수행하여 출현순서의 특징을 살펴보고자 하였다. 연구결과 연결정도중심성, 근접중심성, 매개중심성, 위세중심성 모두 4번째 출현순서가 높은 군집의 논문수가 많은 것으로 나타났다. 다음으로 상관관계 분석결과 근접중심성을 제외한 연결정도중심성, 매개중심성, 위세중심성의 경우 출현순서의 뒷부분으로 갈수록 지수 값이 모두 높아지는 것을 볼 때 네트워크 상에서 상대적으로 중요한 용어가 뒷부분에 위치한다는 것을 알 수 있다. 본 연구는 대략적으로 연구자들의 저자키워드 부여 행태를 살펴볼 수 있었다는 점에서 의미를 지닌다. 향후 저자키워드의 전개양상을 심층적으로 살펴 보기 위해서는 상황적 요인, 행동, 심리 등을 파악하여 보완하는 후속 연구가 반드시 필요하다.

주제어 : 저자키워드, 저자키워드 출현순서, 네트워크분석, 중심성지수

**Abstract** This study aims to investigate an importance of Author keyword with analysis the position of author keyword. An analysis was carried out on the position of author keyword. we examined an importance of Author keyword by using degree centrality, closeness centrality, betweenness centrality, eigenvector centrality. In the next stage, we performed analysis on correlation between network centrality measures and the position of keyword. As a result, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality both has a high value in 4th author keyword order. eigenvector centrality was the comparatively effective method to separate of author keyword order method than other 3 centrality. Correlation analysis result shows that the network analysis value are increasing in order. This study has significance in that it was able to examine the author keyword behavior. Future research is needed to identify and supplement future situational factors, behavior, and psychology.

**Key Words** : Author Keyword, Author keyword order, Network analysis, centrality measure

## 1. 서론

저자키워드는 학술 논문의 주제를 요약하여 나타내어 보여준다는 점에서 논문을 구성하는 중요한 요소 중 하나이다. 그렇기 때문에 논문의 저자는 저자키워드를 정

### 1.1 연구의 필요성 및 목적

\*Corresponding Author : Sun-young Kwon (sykw@hnu.kr)

Received June 22, 2018

Accepted September 20, 2018

Revised August 14, 2018

Published September 28, 2018

확하게 선정할 필요성이 있으며 논문에서 다루고 있는 주제를 추출하고 그 중에서도 가장 대표할 수 있는 용어를 판단한 후 저자키워드로 부여하는 과정을 거친다[1]. 특히나 이용자 입장에서는 저자키워드를 신속하게 확인하여 논문을 선택할 수 있고, 확장 검색 및 색인어, 용어사전 구축의 기초 데이터로 활용되기도 한다[2-5]. 이러한 저자 키워드는 논문과 관련된 기본적 개념이나 성질을 나타내거나 해당 개념의 구체적 예시나 연관관계를 나타내므로 동일한 저자키워드를 공유하는 학술 논문들은 대체적으로 유사한 개념 영역에 속하게 된다[5]. 이러한 정의를 살펴볼 때 저자키워드는 단지 학술논문의 주제만을 요약하여 보여주는 요소 이상이라고 할 수 있다. 학술논문의 저자키워드와 관련한 연구들은 정보검색을 위한 색인어로서의 활용과 관련한 연구[3,6,7], 저자키워드의 비통제어회적 특징에 대한 연구[2,4,8], 그 외에도 저자키워드 표기의 표준화를 통해 관련 논문들을 상호연계 활용할 수 있는 방안을 제시한 연구[8], 저자키워드를 학술용어사전의 구축대상 용어로 사용한 연구[5]가 있다. 최근에는 학술논문의 저자키워드는 네트워크 분석 방법을 위한 데이터로 사용되기도 하며 이를 통해 학술논문이 속한 주제 분야의 지적구조 및 연구 동향을 밝히는 데 활용되고 있다[10-18, 21,22].

기본적으로 이러한 연구들의 '저자키워드는 주제를 요약하여 반영하고 있으며 학술논문에서 매우 중요한 요소'라는 공통적인 전제를 가지고 있다. 어떤 학문 분야에서는 자주 연구되는 주제가 존재할 것이며 그에 따라 자주 사용되는 저자키워드가 있을 것임을 가정해볼 수 있다. 이러한 가정은 많은 출현빈도를 가진 용어가 적은 출현빈도를 가진 용어에 비해 더 중요한 용어라는 것을 전제로 한다. 이를 측정하기 위해서는 특정 용어 네트워크 내에서 네트워크 분석을 수행하여 용어의 위치나 순서를 파악할 수 있다[1,16]. 본 연구에서는 각 논문에 적용된 저자키워드의 네트워크 지수값 비교를 통해 저자키워드 간의 상대적 중요도를 파악하고자 하며, 특히 네트워크 분석 방법 중 어떠한 지수가 저자키워드의 출현순서의 특징을 잘 구분하고 있으며 효과적인지 살펴보고자 한다. 이를 위해 분야의 특징을 잘 반영하고 있는 법학 분야를 연구 대상으로 진행하였다. 법률용어는 법질서를 위해 정확성과 통일성을 가지고 있어야 하며 법학 분야의 논문 또한 법을 근간으로 한다는 점에서 다른 여타 분야에 비해 법학 분야만의 특징을 지니고 있다.

## 1.2 연구문제

- 본 연구에서 살펴보고자 하는 연구문제는 다음과 같다.
- 첫째, 법학분야 학술논문 저자키워드의 출현순서의 특징은 어떠한가?
- 둘째, 법학 분야 학술논문 저자키워드의 출현순서와 중요도를 측정하기 위한 네트워크 지수 간에 상관관계는 나타나는가? 상관관계는 어떠한 의미를 지니는가?

## 2. 연구 설계

### 2.1 저자키워드 선정

법학분야의 저자키워드는 다른 분야의 저자키워드에 비해 법학분야에서만 나타나는 용어의 특정성이 강한 특징을 지니고 있는데 법학 분야는 지수값 1에 해당하는 용어가 6,926개로 50.4%로 다른 분야에 비해 상당히 높으며 비교적 높은 0.6 이상 용어의 개수가 10,475건으로 누적 비율이 76.2%에 해당된다고 하였다[1]. 이는 법학 분야의 용어는 법학 분야 내에서만 출현하는 비율이 높다는 것을 의미한다.

### 2.2 네트워크 분석 방법

네트워크 분석은 권선영(2014)의 연구[17]에서 수집된 데이터 중 법학분야의 논문을 선택하였다. 이 중 키워드 미부여 등의 논문을 제외한 후 1,283개 논문의 3,964개 저자키워드를 대상으로 네트워크 분석 및 관련 연구에서 일반적으로 많이 사용되는 네트워크 중심성 지수 [1,4,13,14, 17-20]인 연결정도중심성, 근접중심성, 매개중심성, 위세중심성을 Pajek으로 측정하였다.

Table 1. Author Keywords network analysis

no	keyword	Frq	deg	bet	clo	Eig
1	Human rights	24	0.022	0.099	0.270	0.060
2	copyright	23	0.020	0.054	0.259	0.059
3	Freedom of expression	22	0.021	0.056	0.259	0.058
4	constitutional amendment	21	0.018	0.049	0.255	0.056
5	outrage	20	0.019	0.077	0.270	0.055
.	.	.	.	.	.	.
3964	International Tax	1	0.001	0	0.149	0.012

Frq: Frequency, deg: degree centrality, bet: betweenness centrality  
clo: closeness centrality, Eig: Eigenvector centrality

### 3. 연구 결과

#### 3.1 네트워크 분석 결과

##### 3.1.1 연결정도 중심성 분석 결과

분석 결과 연결정도 중심성은 첫 번째 순서의 값이 .002954434로 가장 낮고 네 번째 순서의 값이 .003583993로 가장 높은 것으로 나타났다. 다음으로 저자키워드의 전개 형태 특성을 살펴보기 위해 워드(ward's method)의 제곱 유클리디안 거리(squared euclidean distance)를 사용하여 군집병합(agglomeration schedule)을 수행한 결과 개체 수(n=1283)에서 단계번호(1280)를 차감한 3개의 군집으로 나타났다. 3개의 군집에 대한 연결정도 중심성 평균값을 그래프로 나타내면 Fig. 1과 같다. 군집에 따른 특성을 살펴보면 군집 1(clu1)은 두 번째, 군집 2(clu2)는 네 번째, 군집 3(clu3)은 다섯 번째 출현순서 값이 높은 특징을 보여준다. 군집 2의 논문 수는 876건으로 1283건의 논문 중 68.8%에 해당하여 연결정도 중심성으로 측정된 저자키워드의 특징은 네 번째 출현순서가 높다고 판단해볼 수 있다.

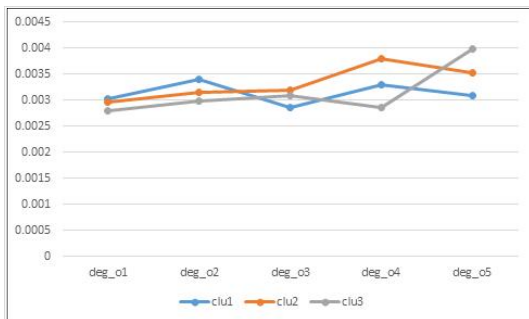


Fig. 1. Appearance patterns by Degree centrality

##### 3.1.2 군집 중심성 분석 결과

분석 결과 군집 중심성은 첫 번째 순서의 값이 .197630891로 가장 낮고 네 번째 순서의 값이 .199386625로 가장 높은 것으로 나타났다. 다음으로 저자키워드의 전개 형태 특성을 살펴보기 위해 워드(ward's method)의 제곱 유클리디안 거리(squared euclidean distance)를 사용하여 군집병합(agglomeration schedule)을 수행한 결과 군집 개체 수(n=1283)에서 단계번호(1280)를 차감한 3개의 군집으로 나타났다. 3개의 군집에 대한 군집 중심성 평균값을 그래프로 나타내면 Fig. 2와 같다. 군집에

따른 특성을 살펴보면 군집 1(clu1)은 네 번째, 군집 2(clu2)는 두 번째, 군집 3(clu3)은 세 번째 출현순서 값이 높은 특징을 보여준다. 군집 1의 논문 수는 901건으로 1283건의 논문 중 70.2%에 해당하여 군집중심성으로 측정된 저자키워드의 특징은 네 번째 출현순서가 높다고 판단해볼 수 있다.

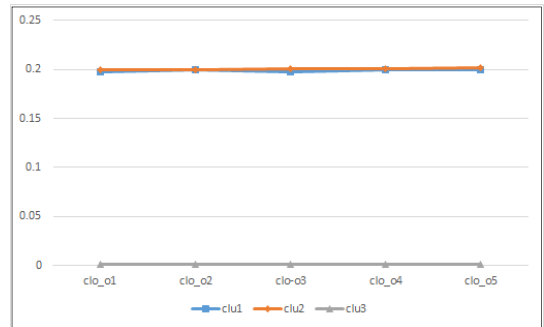


Fig. 2. Appearance patterns by Closeness centrality

##### 3.1.3 매개 중심성 분석 결과

매개 중심성 분석 결과 매개 중심성은 첫 번째 순서의 값이 .004614900로 가장 낮고 네 번째 순서의 값이 .006434971로 가장 높은 것으로 나타났다. 다음으로 저자키워드의 전개 형태 특성을 살펴보기 위해 워드(ward's method)의 제곱 유클리디안 거리(squared euclidean distance)를 사용하여 군집병합(agglomeration schedule)을 수행한 결과 군집 개체 수(n=1283)에서 단계번호(1277)를 차감한 6개의 군집으로 나타났다. 6개의 군집에 대한 군집 중심성 평균값을 그래프로 나타내면 Fig. 3과 같다. 군집에 따른 특성을 살펴보면 군집 1(clu1)은 두 번째, 군집 2(clu2)는 다섯 번째, 군집 3(clu3)은 세 번째, 군집

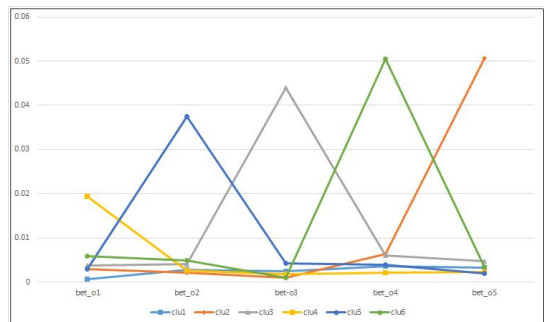


Fig. 3. Appearance patterns by Betweenness centrality

4(clu4)는 첫 번째, 군집 5(clu5)는 두 번째, 군집 6(clu6)은 네 번째 출현순서 값이 높은 특징을 보여준다. 군집 6의 논문 수는 741건으로 1283건의 논문 중 57.7%에 해당하여 매개 중심성으로 측정된 저자키워드의 특징은 네 번째 출현순서가 높다고 판단해볼 수 있다.

### 3.1.4 위세 중심성 분석 결과

위세 중심성 분석 결과 위세 중심성은 첫 번째 순서의 값이 .020118443로 가장 낮고 네 번째 순서의 값이 .021714261로 가장 높은 것으로 나타났다. 다음으로 저자키워드의 전개 형태 특성을 살펴보기 위해 워드(ward's method)의 제곱 유클리디안 거리(squared euclidean distance)를 사용하여 군집병합(agglomeration schedule)을 수행한 결과 개체 수(n=1283)에서 단계번호(1278)를 차감한 5개의 군집으로 나타났다. 5개의 군집에 대한 군집 중심성 평균값을 그래프로 나타내면 Fig. 4와 같다. 군집에 따른 특성을 살펴보면 군집 1(clu1)은 네 번째, 군집 2(clu2)는 다섯 번째, 군집 3(clu3)은 세 번째, 군집 4(clu4)는 첫 번째, 군집 5(clu5)는 두 번째, 군집 6(clu6) 출현순서 값이 높은 특징을 보여준다. 다만 군집간의 논문수가 비교적 균등하게 나타나 위세중심성으로 측정된 저자키워드의 특징은 잘 나타나지 않는 것으로 판단해볼 수 있다.

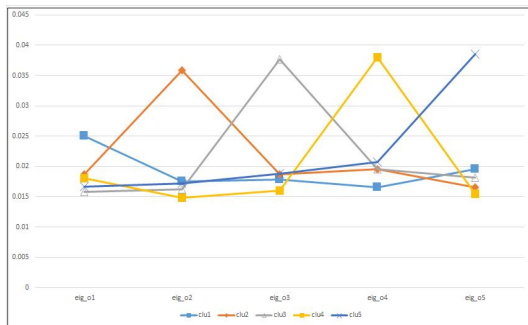


Fig. 4. Appearance patterns by Eigenvector centrality

## 3.2 네트워크 분석 결과와 출현순서간 상관관계 분석

저자키워드의 출현순서와 저자키워드의 연결정도중심성, 근접중심성, 매개중심성, 위세중심성 값과 어떠한 관계가 있는지를 살펴보기 위하여 Kendall의 상관분석을 수행하였다(Table 3). 분석결과 연결정도중심성, 매개중

심성, 위세중심성의 경우 양(+)의 상관관계인 것으로 나타났다. 반면 근접중심성은 전혀 상관관계가 없는 것으로 나타났다.

Table 3. Author Keywords network analysis

			Order	Value
degree	Order	Kendall's tau_b	1	0.20*
	Value	Kendall's tau_b	0.20*	1
clo	Order	Kendall's tau_b	1	0.04
	Value	Kendall's tau_b	0.04	1
bet	Order	Kendall's tau_b	1	0.06*
	Value	Kendall's tau_b	0.06*	1
Eig	Order	Kendall's tau_b	1	0.18*
	Value	Kendall's tau_b	0.18*	1

\* p<.05, \*\* p<.01  
deg: degree centrality, bet: betweenness centrality  
clo: closeness centrality, Eig: Eigenvector centrality

## 3.3 연구결과 종합

먼저 네트워크 중심성 측정 결과를 종합해보면 법학 분야 학술논문의 저자키워드는 연결정도중심성, 근접중심성, 매개중심성으로 측정하였을 때 모두 4 번째 출현순서에서 값이 가장 높은 특징을 지니고 있었다.

군집 결과를 살펴보면 중심성 측정 방법에 따라 3개의 군집에서 6개의 군집까지 나타났는데, 연결정도중심성, 근접중심성, 매개중심성의 경우 4번째 출현순서의 군집의 논문수가 가장 많은 것으로 나타났고, 위세중심성의 경우 군집별 논문 수가 비교적 균등하게 나타나 출현순서의 특징이 잘 나타나지 않는 것으로 나타났다.

다음으로 저자키워드의 출현순서와 연결정도중심성, 근접중심성, 매개중심성, 위세중심성의 상관관계 분석 결과, 유의미하지 않은 근접중심성을 제외하고 다른 지수는 모두 양(+)의 상관관계를 나타냈다. 특정 키워드의 중심성 값이 높다는 것은 네트워크 상에서 상대적으로 다른 키워드에 비해 중요도가 높다고 판단해볼 수 있는데, 네트워크 상에서 상대적으로 중요한 키워드가 출현순서의 앞부분보다는 뒷부분에 위치한다는 것을 의미한다. 다만, 약한 상관관계를 나타냈으므로, 큰 설명력을 가진다고 보기는 어렵다.

#### 4. 결 론

본 연구는 학술논문에 부여된 저자 키워드는 중요도에 따라 나열되어 있을 것임을 가정하고 실제로 그러한 특성이 나타나는지를 살펴보고자 하였다. 이를 위해 중요도를 측정하기 위한 객관적 지표로써 네트워크 분석 방법의 연결정도중심성, 근접중심성, 매개중심성, 위세중심성과 같은 지수를 사용하여 출현순서에 따른 저자키워드의 중요도를 측정해보았다. 연구결과 위세중심성을 제외하고 연결정도중심성, 근접중심성, 매개중심성의 경우 4번째 출현순서의 값이 높은 군집의 논문수가 가장 많은 것으로 나타났다. 다음으로 상관관계 분석결과 근접중심성을 제외하고 출현순서의 뒷부분으로 갈수록 지수 값이 모두 높아지는 것을 볼 수 있었다.

이와 같은 결과를 통해 알 수 있는 바는 다음과 같다. 첫째, 연결정도중심성과 매개 중심성으로 측정하였을 때 저자키워드의 특징이 비교적 잘 나타난다고 볼 수 있다. 둘째 저자의 저자키워드 부여 행태를 짐작해 볼 수 있다. 저자는 논문의 독창성을 위해 새로운 개념, 용어를 첫 번째나, 두 번째 등 앞부분에 둘 수 있는 데 이 경우 해당 주제 분야에서는 아직까지는 잘 알려져 있지 않은 개념이나 용어이므로 학문 분야를 대상으로 용어의 중심성 측정을 했을 때 낮은 값을 가질 가능성이 있다.

본 연구는 대략적으로 연구자들의 저자키워드 부여 행태를 살펴볼 수 있었다는 점, 그리고 기존의 연구가 측정하지 못했던 학술논문 단위의 저자키워드 전개양상을 살펴볼 수 있었다는 점에서 의의를 지닌다. 다만 본 연구는 각 네트워크 지수를 통해 중요도를 측정하고 특성에 따라 군집하기는 하였지만, 분야의 일부 논문을 대상으로 분석하여 대표성을 지니기 어렵다는 한계를 지닌다. 향후 저자키워드의 전개양상을 심층적으로 살펴보기 위해서는 연구자들의 저자키워드를 부여할 때의 상황적 요인, 행동, 심리 등을 파악하여 보완하는 후속 연구가 반드시 필요하다. 본 연구는 연구자의 저자키워드 부여 행태를 밝히기 위한 방법으로 활용될 것으로 기대한다.

#### REFERENCES

- [1] S. Y. Kwon. (2014). A Study on the Factors Influencing Semantic Relation in Building a Structured Glossary, *Journal of The Korean Library and Information Science* 48(2), 353-378.
- [2] H. Y. Mun. 2000. Subject Headings of Korean medical journal papers: Author keywords vs. MESH. The Graduate School, Sookmyung Women's University.
- [3] K. Taghva, J. Borsack, T. Nartker, A. Condit. (2004). The role of manually assigned keywords in query expansion. *Information Processing & Management*, 40, 441 - 458.
- [4] S. H. Park, K. Y. Park. (2011). Coincidence Analysis of Key Words and MeSH Terms in the Journal of the Korean Society of Occupational Therapy. *The Journal of Korean Society of Occupational Therapy*, 19(4), 131-146.
- [5] Y. M. Ko, M. S. Song, B. Y. Kim, H. R. Min. (2013). A Study on the Correlation between the Appearance Frequency of Author Keyword and the Number of Citation in the Humanities and Social Science Journal Articles of the Korea Citation Index (KCI). *Journal of the Korean Society for Information Management*, 30(2), 227-243.
- [6] T. Gross, A. G. Taylor. (2005). What have we got to lose? The effect of controlled vocabulary on keyword searching results. *College & Research Libraries*, 66(3), 212 - 230.
- [7] P. J. Kim, J. Y. Lee. (2012). A Study on the Reclassification of Author Keywords for Automatic Assignment of Descriptors. *Journal of the Korean Society for Information Management*, 29(2), 225-246.
- [8] K. B. Kang, J. H. Kim, Y. B. Kim, J. K. Kim. (2014). Use of Medical Subject Headings (MeSH) in the Journal of the Korean Orthopaedic Association. *Journal of The Korean Orthopaedic Association*, 49(4), 302-306.
- [9] C. J. Yang. (2010). Study on Keywords and Their Use of Academic Theses - Focused on Database Development and Information Link. *Humanities Contents*, 19, 395-416.
- [10] Y. D. Lee. (2006). An Analysis of the Key Words in Korean Journal of Japanology in Korea. *Journal of Japanese Language and Literature*, 31, 107-121.
- [11] O. Park. (2011). Knowledge Structures in Knowledge Organization Research: 2000-2011. *Journal of The Korean Biblia Society For Library And Information Science*, 22(3), 247-267.
- [12] H. Y. Lee & S. J. Kwak. (2011). Relation Analysis Among Academic Research Areas Using Subject Terms of Domestic Journal Papers, *Journal of The Korean Biblia Society For Library And Information Science*, 22(3), 353-371.
- [13] C. Jane. (2011). A Study for Research Area of Library

- and Information Science by Network Text Analysis. *Journal of the Korean Society for Information Management*, 28(4), 65-83.
- [14] B. I. Kang, J. H. Park. (2013). Profiling and Co-word Analysis of Teaching Korean as a Foreign Language Domain. *Journal of the Korean Society for Information Management*, 30(4), 195-213.
- [15] S. K. Seo, E. K. Chung. (2013). Domain Analysis on the Field of Open Access by Co-Word Analysis. *Journal of The Korean Biblia Society for Library and Information Science*, 24(1), 207-228.
- [16] L. L. Zhang, H. J. Hong. (2014). Examining the Intellectual Structure of Reading Studies with Co-Word Analysis Based on the Importance of Journals and Sequence of Keywords. *Journal of The Korean Biblia Society For Library And Information Science*, 25(1), 295-318.
- [17] S. Y. Kwon. (2014). A Study on the Application to Network Analysis on the Importance of Author Keyword based on the Position of Keyword. *Journal of the Korean Society for Information Management*, 31(2), 121-142.
- [18] S. Y. Kwon, E. J. Park. (2015). Knowledge Structure of the Korean Journal of Occupational Health Nursing through Network Analysis. *Korean Journal of Occupational Health Nursing*, 24(2), 76-85.
- [19] N. R. Kim, H. R. Choi, T. L. (2018). Analysis of Domestic SNA-based Governance Study Trends. *Journal of Digital Convergence*, 16(7), 35-45.
- [20] M. S. Chung, S. H. Park, B. H. Chae, J. Y. Lee. (2017). Analysis of major research trends in artificial intelligence through analysis of thesis data. *Journal of Digital Convergence*, 15(5), 225-233.
- [21] S. Yoo, M. Baik. (2017). Analysis of domestic dementia research trend for integrated study. *Journal of Convergence for Information Technology*, 7(3), 1-12.
- [22] H. Lee. (2017). Analysis of Research Trends related to Women: Focusing on Literature in Korean Journal of Social Welfare, 2009~2017. *Journal of Convergence for Information Technology*, 7(5), 149-155.

권 선 영(Kwon, Sun-young)

[정회원]



- 2002년 2월 : 성균관대학교 문헌정보학과 (문헌정보학사, 문학사)
- 2005년 2월 : 성균관대학교 문헌정보학과 (문헌정보학석사)
- 2014년 2월 : 성균관대학교 문헌정보학과 (문학박사)
- 2017년 3월 ~ 현재 : 한남대학교 문헌정보학과 교수
- 관심분야 : 정보검색, 융복합, 멀티미디어, 빅데이터
- E-Mail : sykw@hnu.kr