

동시저자분석을 통한 한국의 OLED 분야 지식구조 파악을 위한 연구

A Study on Korean Intellectual Structure of OLED Technology Using Co-Authorship Analysis

김완중, 노경란, 서진이, 권오진, 정의섭
한국과학기술정보연구원

Kim Wan-Jong, Noh Kyung-Ran, Su Jinny, Kwon Oh-Jin,
Jeong Eui-Sob

Korea Institute of Science & Technology Information

요약

동시저자 분석(Co-Authorship Analysis)은 계량서지 분석 방법의 하나로서 특정 기술 분야의 지식생산 네트워크의 특성을 파악하는데 매우 유용한 분석 기법이라 할 수 있다. 본 연구는 둘 이상의 연구자가 하나의 논문을 공동으로 작성하였다면 이들 사이에 지식의 연결고리가 존재한다는 가정 하에 국내 OLED 분야의 지적 구조를 계량서지 분석 기법인 동시저자 분석을 통하여 규명해 봄으로써 OLED 분야의 발전 방향과 학문적 위상을 정립하는데 그 목적이 있다. 분석 데이터로는 SCI(E) DB에 수록되어 있는 OLED 기술 분야의 논문 중에서 한국인들이 작성한 논문을 사용하였다.

Abstract

Co-Authorship analysis is a kind of bibliometric analysis, and a well established tool for character of knowledge network in some technical fields. If more than two authors co-author an article, there is a relation of knowledge. So This study aims at discovering the trends and phases of OLED technology using co-authorship analysis, a kind of bibliographic analysis. The Science Citation Index Expanded(SCIE) database was used to search for the number of times paired Korean articles were co-authored in OLED field.

I. 서론

동시저자 분석(Co-Authorship Analysis)은 계량서지 분석 방법의 하나로서 특정 기술 분야의 지식생산 네트워크의 특성을 파악하는데 매우 유용한 분석 기법이라 할 수 있다. 또한 여러 연구자들이 협력 연구 결과를 발표한 논문들을 통해서 연구자들 사이의 관계를 규명한다는 의미에서 사회 연결망 분석, 사회 네트워크 분석(Social Networks Analysis : SNA)이라는 용어도 사용되고 있다.

사회 연결망 분석은 최근 흥미로운 연구 방법으로 각광을 받고 있으며, 또한 많은 연구 분야에서 중요한 역할을 수행하고 있다. 문헌정보학 분야에서도 사회 연결망 분석은 인용 연결망, 동시인용 연결망, 협력 연결망 등으로 응용되어 연구되고 있다 [1, 2]

OLED(Organic Light Emitting Diodes)는 1980년대 후반 처음 연구되어 소개된 후, 1990년대 후반 본격적으로 상업화가 이루어졌다. OLED 기술은 21세기 고도 정보화 사회에서 전자 시스템 및 멀티미디어의 급속한 보급에 따라 더욱 중요시 되고 있는 기술 분야로, 인간과 각종 전자장치를 연결시켜주는 역할을 담당하는 디스플레이에 대한 중요성은 최근에 들어 더욱 더 커지고 있다.

본 연구는 둘 이상의 연구자가 하나의 논문을 공동으로 작성하였다면 이들 사이에 지식의 연결고리가 존재한다는 가정 하에 국내 OLED 분야의 지적 구조를 계량서지 분석 기법인 동시저자 분석을 통하여 규명해 봄으로써 OLED 분야의 발전 방향과 학문적 위상을 정립해 보고자 함에 그 목적이 있다. 분석 데이터로는 SCI(E) DB에 수록되어 있는 OLED 기술 분야의 논문 중에서 한국인들이 작성한 논문을 연구 대상으로 하였다.

II. 연구 방법

1. 데이터 추출

본 연구에서는 국내 OLED 분야의 현황을 파악하기 위하여 1986년부터 2005년도까지 SCI(E) 등재학술지에 수록된 한국인 저술 논문들의 동시저자(co-author) 자료를 수집하였다. 자료의 수집은 Thompson Scientific(구 ISI)의 데이터베이스 서비스인 Web of Science SCIE(Science Citation Index Expanded)를 이용하였다. 검색식으로는 "TS = ("OLED*") AND CU = KOREA AND Timespan = 1986-2005"을 사용하였으며, 이를 통해 1999년부터 한국인이 저술한 총 158편의 논문이 추출되었다.

2. 공저자 데이터 구성

이상과 같이 수집한 서지데이터를 이용하여 각 문헌에 등장하는 공동 저자들을 문헌별로 분류하였다. 158편의 논문에 등장한 저자는 총 419명이다. 이 419명의 저자 중에서 이 연구를 위하여 선택한 저자는 4편 이상의 SCI(E) 논문을 작성한 42명을 선정하였으며, 이 42명이 작성한 논문은 96편이었다. 이를 저자와 문헌간의 관계를 42 X 96로 작성한 후, 이에 대한 전치행렬인 96 X 42 행렬을 곱하여 42 X 42 형태의 동시저자 행렬(Matrix)을 작성하였다.

이 데이터를 기초로 SPSS 통계 프로그램과 네트워크 분석 도구인 UCINET 6.0을 이용하여 이렇게 추출된 동시저자 행렬을 피어슨(pearson) 상관계수 행렬로 변환하여 다차원척도법(Multidimensional Scaling : MDS), 군집분석(Cluster Analysis), 요인분석(Factor Analysis)을 통해 OLED 분야의 지적 구조를 분석하였다.

III. 연구 결과

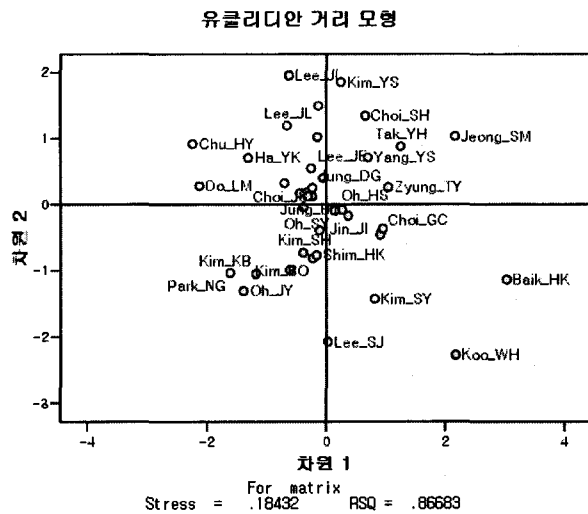
1. 동시저자행렬 및 상관계수 행렬 분석

각 저자들의 동시저자 빈도는 4에서 17이며, 가장 많은 17편의 논문을 공동 저술한 저자는 “Chu, HY”이고, 다음으로 14편의 논문을 공동 저술한 “Lee, JI”이었다. 또한 10편의 동시저자가 3명, 9편이 6명, 8편 3명, 7편 3명, 6편 6명, 5편 6명, 4편 13명의 총 42명의 저자가 선정되었다. 동시저자행렬의 경우, 빈도수의 차이가 크기 때문에 저자쌍들의 비교가 힘들어 이 행렬을 피어슨 상관계수 행렬로 변환하여 사용하였다.

변환된 상관관계 행렬을 살펴보면, 1.0의 “완전” 상관관계를 가지고 있는 저자쌍은 {Baik_HK(3), Jeong_SM(5)} {Kim_YS(8), Park_NG(11)}, {Choi_SH(6), Koo_WH(9)}, {Ko_YW(38), Do_YR(31)}의 5개 쌍이었다. 다음으로 상관관계가 높은 저자 쌍은 {Baik_HK(3), Choi_SH(6)}, {Baik_HK(3), Koo_WH(9)}, {Jeong_SM(5), Choi_SH(6)}, {Jeong_SM(5), Koo_WH(9)}의 총 4쌍으로서 0.998의 “매우 높은” 상관관계를 보였다.

2. 다차원 척도법

다차원 척도법(MDS)은 개체들을 대상으로 변수들을 측정 한 후에 개체들 사이의 유사성/비유사성을 측정하고 이 값을 이용하여 개체들을 2차원 혹은 3차원 공간상에 점으로 표현하는 분석방법이다. 그림 1은 동시저자 상관계수 행렬을 이용하여 다차원 척도법을 통하여 저자들을 2차원 공간에 축적한 결과 화면이다.



▶▶ 그림 1. 2차원 축적의 저자지도 및 Stress, RSQ 값

다차원 척도법의 경우 스트레스 값(stress)은 0과 1 사이에 존재하며 0에 가까워질수록 적합한 모형으로 제시하게 된다. 스트레스 값이란 다차원 척도 모형에 의해서 설명되지 않는 분산의 불일치 정도로서 대상들 간의 실제 거리와 추정된 거리 사이의 오차를 의미한다. Kruskal은 스트레스 값에 따라 모형이 나타내는 적합도에 대한 기준을 정리 하였다[3].

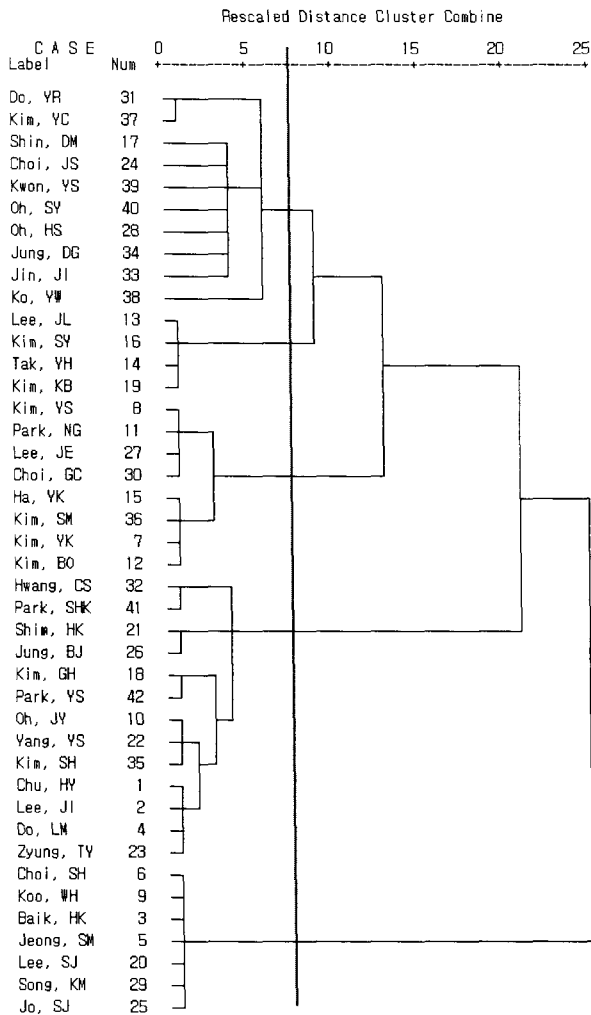
본 연구에서 실시한 다차원 척도 분석에서는 스트레스 값이 0.18432, RSQ 값이 0.86683이었다. Kruskal의 이론에 따르면 본 연구의 스트레스 값이 0.2 에 가까워 2차원으로 축적된 모형이 적합하다고 하기 어렵다. 이는 저자들 사이의 상관관계가 0.5를 넘지 않는 경우들이 많아 2차원이나 3차원에서 축적되기 어렵다는 것을 보여준다. 즉, 다차원 축적은 의미가 없었다.

[표 2] Kruskal의 스트레스 값 기준

스트레스 값	적합도 평가
0.2 이상	매우 나쁨
0.2	나쁨
0.1	보통
0.05	좋음
0.025	매우 좋음
0	완벽

3. 군집분석

군집분석이란 저자들에 대한 연관성을 두 개 혹은 그 이상으로 특성에 따라 분류하여 동질적인 몇 개의 군집(cluster)으로 나누는 것이다.



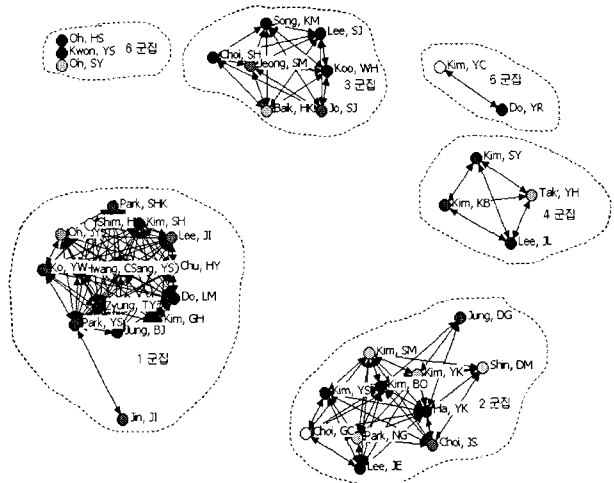
▶▶ 그림 2. 계층적 군집분석 결과

군집분석 과정은 1) 저자들을 군집화 하는데 이용될 수 있는 변수를 측정, 2) 측정된 저자들을 이용하여 모든 개체들 간의 거리(distance) 혹은 비유사성(dissimilarity)을 계산하여 모든 저자들 사이의 비유사성(dissimilarity)을 나타내는 거리행렬을 구하는 유사성 측정, 3) 각 저자들의 쌍을 대상으로 각 저자들을 가까운 순서대로 묶어주는 군집화(clustering)의 3단계 과정으로 이루어진다.

본 연구에서는 하나의 군집이 다른 군집의 내부에 포함되거나 서로 다른 군집 간에는 중복이 허용되지 않는 계층적 군집(hierarchical clustering)을 실시하였다. 군집분석 결과를 덴드로그램으로 나타내면 그림 2와 같으며, 그 결과를 보면 크게 5개의 동시저자 군집을 형성하였다. 형성된 저자군집은 아래와 같다.

- 1 군집(10명) = {17, 24, 28, 31, 33, 34, 37, 38, 39, 40}
- 2 군집(4명) = {13, 14, 16, 19}

- 3 군집(8명) = {7, 8, 11, 12, 15, 27, 30, 36}
- 4 군집(13명) = {1, 2, 4, 10, 18, 21, 22, 23, 26, 32, 35, 41, 42}
- 5 군집(7명) = {3, 5, 6, 9, 20, 25, 29}



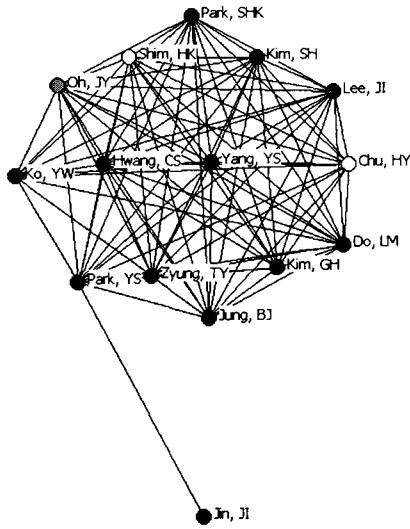
▶▶ 그림 3. UCINET의 다차원 축적 결과

그림 3은 네트워크 분석 도구인 UCINET에서 저자들 사이의 강도와 연결 관계를 군집분석을 통하여 이를 도식화시켜 나타낸 것이다. UCINET에서는 동시저자 군집이 총 6개로 나누어졌으며, 집합의 크기별로 군집을 나누어 보면 다음과 같다.

- 1 군집(15명) = {1, 2, 4, 10, 18, 21, 22, 23, 26, 32, 33, 35, 38, 41, 42}
- 2 군집(11명) = {7, 8, 11, 12, 15, 17, 24, 27, 30, 34, 36}
- 3 군집(7명) = {3, 5, 6, 9, 20, 25, 29}
- 4 군집(4명) = {13, 14, 16, 19}
- 5 군집(2명) = {31, 37}
- 6 군집(3명) = {28, 39, 40}

3명이 속해 있는 군집을 6군집으로 분류한 것은 이 군집에 속한 3명의 저자인 Oh, HS(28), Kwon, YS(39), Oh SY(40)가 본 연구에 선정된 다른 39명의 저자들과 공동 연구를 수행하지 않았기 때문에 저자들과의 연결 없이 독립되어 있는 것을 확인할 수 있다.

그림 4는 그림 3에서 1군집을 확대한 것이다. 14명의 저자가 군집에 속한 다른 저자들과 연결된 네트워크를 확인할 수 있다. 그리고 중앙 하단에 위치한 저자 "Jin, JI(33)"는 이 군집의 다른 저자들과의 공동 연구가 없으며, 저자 "Park, YS(42)"만이 협력 연구를 수행하였다는 것을 볼 수 있다.

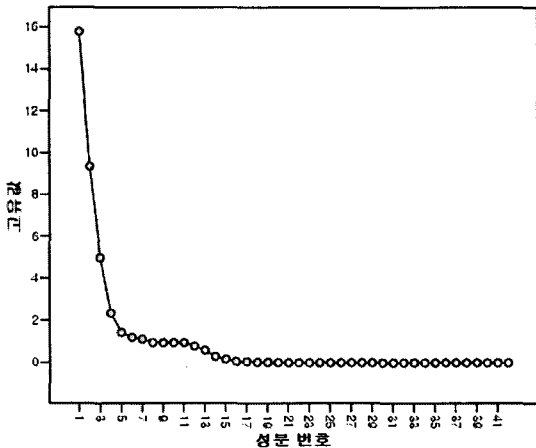


▶▶ 그림 4. UCINET에서의 1군집 결과

3. 요인분석

요인분석이란 여러 개의 서로 관련이 있는 저자들을 측정하는 자료를 이용하여 그 저자들을 설명할 수 있는 가공의 공통변수(요인)를 구하여 원래의 저자들을 요인들의 선형함수로 표현하는 통계분석방법이다. 군집 분석과 마찬가지로, 요인분석은 관련 변인들의 그룹을 규정하기 위해 사용된다[4].

동시저자 행렬을 기준으로 요인분석을 한 결과 그림 5의 스크리 도표(scree plot)와 같이 나타났다. 하지만 스크리 도표의 한계점은 몇 개의 성분을 추출해야 하는지에 대한 것이 그래프 상으로는 명확하지 않기 때문에 고유값(아이겐 값 : Eigen-value)을 이용한 주성분 분석을 다시 수행하였다. 주성분 분석에서는 고유값이 1 이상이 되는 주성분을 선택하는 방법이 가장 흔하게 사용되며, 1 이상의 고유값이 7개 성분에 나타났다(그림 6). 이 7개의 성분이 전체의 86.244%의 설명력을 갖는다는 것을 알 수 있다.



▶▶ 그림 5. 요인분석 결과 스크리도표

설명된 총분산

성분	초기 고유값			추출 제공할 적재값		
	전체	% 분산	% 누적	전체	% 분산	% 누적
1	15.785	37.582	37.582	15.785	37.582	37.582
2	9.346	22.252	59.834	9.346	22.252	59.834
3	4.977	11.851	71.685	4.977	11.851	71.685
4	2.352	5.599	77.284	2.352	5.599	77.284
5	1.438	3.423	80.707	1.438	3.423	80.707
6	1.206	2.872	83.579	1.206	2.872	83.579
7	1.119	2.665	86.244	1.119	2.665	86.244
8	.955	2.273	88.517			
9	.948	2.257	90.774			
10	.948	2.257	93.031			
11	.946	2.253	95.283			
12	.778	1.852	97.136			
13	.590	1.404	98.540			
14	.293	.698	99.237			
15	.159	.379	99.616			
16	.056	.134	99.750			
17	.053	.125	99.875			
18	.024	.058	99.934			
19	.020	.048	99.982			
20	.004	.010	99.992			
21	.002	.005	99.996			
22	.001	.003	99.999			
23	.000	.000	99.999			
24	.000	.000	100.000			
25	6.770E-05	.000	100.000			
26	5.671E-05	.000	100.000			
27	3.490E-05	8.310E-05	100.000			
28	1.314E-05	3.129E-05	100.000			
29	7.268E-06	1.730E-05	100.000			
30	3.763E-06	8.958E-06	100.000			
31	2.954E-06	7.034E-06	100.000			
32	2.412E-06	5.742E-06	100.000			
33	1.220E-06	2.905E-06	100.000			
34	2.861E-07	6.813E-07	100.000			
35	2.567E-07	6.111E-07	100.000			
36	1.181E-07	2.812E-07	100.000			
37	5.670E-08	1.350E-07	100.000			
38	2.885E-16	6.869E-16	100.000			
39	3.920E-22	9.333E-22	100.000			
40	-3.66E-18	-8.71E-18	100.000			
41	-4.12E-17	-9.81E-17	100.000			
42	-2.22E-16	-5.29E-16	100.000			

추출 방법: 주성분 분석.

그림 6. 주성분 분석 결과

하지만 요인 분석을 실시한 성분 행렬 결과를 살펴보면, 추출된 7개 성분에 나타난 값들이 0.6보다 큰 저자들은 '성분 1'과 '성분 2'에 20명이 포함되었으며, '성분 3'에서 '성분 7' 사이에는 전혀 나타나지 않았다.

각 성분에서 0.6 이상의 값을 가지는 저자들의 집합은 다음과 같다.

- 성분 1(12명) = {1, 2, 4, 10, 18, 21, 22, 23, 26, 32, 38, 41}
- 성분 2(8명) = {7, 8, 11, 12, 15, 27, 30, 36}
- 어떤 성분에도 포함되지 않는 저자(22명) = {3, 5, 6, 9, 13, 14, 16, 17, 19, 20, 24, 25, 28, 29, 31, 33, 34, 35, 37, 39, 40, 42}

이는 '성분 1'과 '성분 2'에 포함된 저자들의 상관관계가 높게 나타났기 때문이며, 이에 속하지 않은 저자들은 저자들 사이의 연결 고리가 강하지 않다는 것을 알 수 있었다.

[표 1] 성분 행렬

저자	성분 1	성분 2	성분 3	성분 4	성분 5	성분 6	성분 7
Chu, HY	0.977	0.073	0.126	0.077	0.057	0.058	0.009
Lee, JI	0.976	0.073	0.127	0.081	0.053	0.040	0.024
Baik, HK	-0.423	-0.842	0.307	0.117	0.058	0.007	0.026
Do, LM	0.957	0.072	0.121	0.075	0.047	-0.107	0.101
Jeong, SM	-0.423	-0.842	0.307	0.117	0.058	0.007	0.026
Choi, SH	-0.423	-0.842	0.306	0.116	0.058	0.007	0.026
Kim, YK	-0.567	0.716	0.255	0.087	-0.056	0.091	0.189
Kim, YS	-0.562	0.726	0.282	0.149	0.176	-0.070	-0.108
Koo, WH	-0.423	-0.842	0.306	0.116	0.058	0.007	0.026
Oh, JY	0.962	0.072	0.122	0.074	0.041	-0.028	0.033
Park, NG	-0.562	0.726	0.282	0.149	0.176	-0.070	-0.108
Kim, BO	-0.558	0.708	0.255	0.091	-0.041	0.075	0.149
Lee, JL	-0.172	-0.038	-0.929	0.286	0.108	0.008	0.041
Tak, YH	-0.173	-0.038	-0.931	0.286	0.108	0.008	0.041
Ha, YK	-0.579	0.738	0.271	0.117	0.053	0.015	0.045
Kim, SY	-0.172	-0.038	-0.925	0.285	0.108	0.008	0.041
Shin, DM	-0.317	0.346	0.056	-0.067	-0.275	0.184	0.326
Kim, GH	0.917	0.069	0.112	0.053	0.013	-0.180	0.035
Kim, KB	-0.172	-0.038	-0.924	0.282	0.106	0.008	0.040
Lee, SJ	-0.424	-0.842	0.305	0.114	0.056	0.007	0.026
Shim, HK	0.882	0.067	0.114	0.077	0.070	-0.029	0.091
Yang, YS	0.959	0.072	0.120	0.070	0.036	-0.023	0.027
Zyung, TY	0.968	0.073	0.124	0.078	0.049	-0.088	0.087
Choi, JS	-0.275	0.286	0.027	-0.095	-0.362	0.247	0.456
Jo, SJ	-0.425	-0.842	0.305	0.114	0.056	0.008	0.026
Jung, BJ	0.874	0.067	0.112	0.075	0.067	-0.047	0.101
Lee, JE	-0.525	0.679	0.267	0.158	0.255	-0.131	-0.226
Oh, HS	-0.077	-0.001	-0.118	-0.197	-0.328	0.003	-0.343
Song, KM	-0.425	-0.842	0.305	0.114	0.056	0.008	0.026
Choi, GC	-0.527	0.681	0.266	0.156	0.252	-0.129	-0.223
Do, YR	-0.105	-0.010	-0.196	-0.881	0.404	0.010	0.098
Hwang, CS	0.884	0.068	0.101	0.051	0.046	0.303	-0.149
Jin, JI	0.042	0.008	-0.102	-0.192	-0.394	-0.640	-0.040
Jung, DG	-0.200	0.175	-0.033	-0.138	-0.352	0.151	0.159
Kim, SH	0.955	0.072	0.122	0.073	0.038	-0.080	0.064
Kim, SM	-0.579	0.734	0.267	0.108	0.024	0.036	0.085
Kim, YC	-0.105	-0.010	-0.196	-0.881	0.404	0.010	0.098
Ko, YW	0.596	0.049	0.030	-0.022	-0.013	0.531	-0.318
Kwon, YS	-0.077	-0.001	-0.118	-0.197	-0.328	0.003	-0.343
Oh, SY	-0.077	-0.001	-0.118	-0.197	-0.328	0.003	-0.343
Park, SHK	0.894	0.068	0.105	0.055	0.049	0.234	-0.142
Park, YS	0.911	0.069	0.090	0.023	-0.067	-0.303	0.086

※ 요인추출 방법: 주성분 분석.

a. 추출된 7 성분

의 군집이 형성되었다.

셋째, 요인분석 결과 크게 두 개의 요인으로 분류할 수 있었으며, 이 두 요인에 속하지 않은 저자들은 저자들 사이의 연결 고리가 강하지 않다는 것을 알 수 있었다.

결론적으로 OLED 분야가 새롭게 나타난지 몇 년이 되지 않았으며, SCIE 학술지에 수록된 한국인이 저술한 논문의 수가 그리 많지 않아 이 분야 연구자들의 지식 연결 고리를 밝히는 데 다소 어려움이 있었으나, 주요 연구자들 사이의 연결 고리를 파악할 수 있어 의미있는 분석이라 생각한다. 향후 장기간에 걸친 망라적인 연구가 될 수 있기를 희망한다.

■ 참고 문헌 ■

- [1] Otte, E. and Rousseau, R., "Social network analysis: a powerful strategy, also for the information sciences", *Journal of Information Science*, pp.441-453, 2002.
- [2] Wasserman, S. and Faust, K., *Social network analysis: Methods and applications*, Cambridge, Cambridge University Press, New York, 1994.
- [3] Kruskal, B., "Nonmetric multidimensional scaling: a numerical method", *Psychometrika* Vol.29, pp.115-129, 1964.
- [4] James, A. "An Author co-citation analysis of medical informatics", *Journal of the Medical Library Association*, pp.47-56, 2003.

IV. 결론

본 연구에서는 한국의 OLED 기술 분야의 지적구조를 규명하기 위하여 계량정보학적 분석 기법인 동시저자 분석을 실시하였다.

본 연구의 주요 발견사항은 다음과 같다.

첫째, 선정된 42명의 동시 저자 빈도에 근거한 상관계수는 상관관계가 0.5를 넘지 않는 경우들이 많아 2차원이나 3차원에서 축적이 되지 않아서 다차원축적은 의미가 없었다.

둘째, 저자들의 연관성에 따라 비슷한 특성을 지닌 저자들을 분류하기 위해 실시한 군집 분석에서 SPSS에서의 계층적 군집분석은 5개의 군집, UCINET을 이용한 군집분석에서는 6개