Assessing the Utilization and Interrelatedness of Scopus Subject Categories

Scopus에 설정된 주제분류 활용도 및 상호 연관성에 대한 고찰

Eungi Kim*

<Contents>

ABSTRACT

This study investigated the utilization and interrelatedness of *Scopus* subject categories. To conduct this study, major and minor subject categories of journals listed in the 2017 *Scopus* index were used. The results showed varying degrees of interrelatedness of subject categories. At the major subject category level, the utilization was the highest in *Medicine,* while *Social Sciences* showed a greater degree of interrelatedness in comparison to *Medicine.* Yet, at the minor subject level, *2700 General Medicine* was particularly dominant in terms of utilization and interrelatedness. Moreover, co-occurrences of minor subject categories showed varying degrees of interrelatedness between pairs of minor subject categories. Pairs of minor subject categories showed the following characteristics: a) two subject categories having identical or closely identical descriptions, b) two different categories having an interrelationship by subject areas, and c) one category conceptually encompassing another category. Due to varying degrees of utilization and interrelatedness among subject categories, minor subject categories that may greatly influence the major subject categories in conducting research studies should be investigated in detail.

Keywords: ASJC codes, Subject categories, Scopus, Utilization, Interrelatedness

초 록

본 연구는 *Scopus*의 분류시스템에서 설정된 주제분류의 활용도와 상호연관성을 조사했다. 연구 수행의 범위는 2017년도 *Scopus* 색인에 포함된 저널의 대 주제 분류 및 소 주제분류를 포함하였다. 연구의 결과는 *Scopus*의 주제 분류별 활용도에서의 빈도수나 주제분류간 상호연관성에서도 다양한 양상을 보였다. 이 가운데 대 주제분류활용률과 상호연관성의 양상은 의학과 사회과학분야에서 가장 활발하고 다양하게 나타났다. 한편, 소 주제분류에서는 "*2700 General Medicine*"이 상호연관 측면에서 최고의 다양성을 보여주었다. 이외에 쌍을 이루는 소 주제분류 간의 동시 빈도수 분석에서 특징적인 상호연관성을 보이는 경우들을 발견하였는데 같은 분야의 주제어가 비슷한 연관성, 분야는 다르지만 주제가 연결된 연관성, 주제어간 상,하위범주의 연관성의 경우들이다. 소수의 소 주제분류간에는 매우 유사한 주제어를 사용한 경우들도 있었다. 이러한 주제분류별 활용도와 주제분류간의 상호연관성들이 보여주는 다양한 양상들로 인해 연구수행에 있어서 대 주제분류뿐만 아니라 대주제 분류에 많은 영향을 미치는 소 주제분류까지도 면밀하게 살펴볼 필요가 있다.

키워드: ASJC 코드, 주제분류, *Scopus,* 활용도, 상호 관련성

* 계명대학교 문헌정보과 조교수(egkim@gw.kmu.ac.kr)

•논문접수: 2019년 2월 20일 •최초심사: 2018년 2월 28일 •게재확정: 2019년 3월 19일
•한국도서관·정보학회지 50(1), 251-272, 2019. [http://dx.doi.org/10.16981/kliss.50.201903.251]

Ⅰ. Introduction

To understand the varying scope of the journals with minimal effort, a vast number of topical areas of journals need to be reduced into a smaller number of categories. Journal categories are especially important for studies that analyze data associated with journals. These journals can be examined individually or collectively by using the data available in a bibliographic database such as *Scopus.* Due to the availability of a large number of bibliographic records in *Scopus,* subject categories of *Scopus* are often used in various studies. Previous studies used subject categories of *Scopus* for a variety of research purposes: a) to reveal citation patterns of scientific impact within a subject (Abrizah et al. 2013; Thelwall 2017), b) to measure the extent of publications pertaining to subject areas (Chung and Tsay 2017; Hassan et al. 2017; Zuccala and Guns 2013), c) to examine how citation flows among subject areas (Yan 2016), d) to rank subject areas based on journals (García, Rodriguez-Sánchez and Fdez-Valdivia 2011), e) to evaluate the content coverage of *Scopus* (Dorta-González and Santana-Jiménez 2017), and f) to assess the accuracy of *Scopus* journal classification (Wang and Waltman 2016).

Journals indexed by *Scopus* (http://www.scopus.com) are categorized using All Science Journal Classification (ASJC) codes. These *Scopus* categories comprise different levels of granularity. At the top level, there are four general subject areas: 1) Life Sciences, 2) Physical Sciences, 3) Social Sciences, and 4) Health Sciences. These 4 are referred to as general subject categories and are fragmented into 27 “major” subject categories. These subject categories are further fragmented into “minor” subject categories. At the time of conducting this study, there were 330 minor subject categories. Each minor subject category is assigned with ASJC codes and full descriptive names. For example, the ASJC code for *Library and Information Science* (LIS) is 3309, and *Library and Information Science* (LIS) 3309 is a minor subject category of the major category *Social Sciences.* The ASJC codes for major subject category *Social Sciences* range from 3300 to 3399. Here, 3399 is used only to denote the last number in the 3300 blocks. In each major category, only a limited number of minor subject category number is used; in fact, there is no minor category assigned to the ASJC code 3399. A complete list of ASJC descriptions was provided by Bervkens (2012).

An analysis of the subject categories of bibliometric database may have several advantages:

a) A better understanding of the intricate landscape of the subject categories of journals can be obtained (Kosecki et al. 2011), b) Subject categories that are closely related to each other can be identified (Minguet et al. 2017; Leydesdorff and Bornmann 2016; Klarenbeek and Boshoff 2018), c) Subject categories can be grouped together using various methodologies (Kim, Kim, and Oh 2016), and d) Subject categories can be quickly distinguished in terms of commonly used bibliometric attributes such as citation counts (Martín-Martín et al. 2018).

An analysis of the utilization and interrelatedness aspects of subject categories can also provide similar benefits. In this study, utilization refers to the extent to which the subject category is used in categorizing journals. Thus, a frequency count can be used to obtain the number of journals that use the particular subject category. For example, *3309 Library and Information Science* was used 195 times to categorize Scopus index journals. On the other hand, interrelatedness is more difficult to define. The Merriam-Webster dictionary (https://www.merriam-webster.com) defines interrelatedness as having a “mutual relationship”. Stanton (2014) noted that such a broad conceptual definition is commonly used in many studies. Thus the assessment of interrelatedness is difficult because its definition is broad and subjective. Nonetheless, to operationalize the concept, the interrelatedness of subject categories is defined as the relation resulting from assigning objects to subject categories.

Even with this operational definition, no single approach is adequate in measuring the interrelatedness of subject categories. In this study, the extent of interrelatedness can be examined by performing a network analysis along with other relevant frequency analyses involving a set of subject categories, although no precise measurement of interrelatedness is being offered. The specific subject categories of journals indexed in *Scopus* would be utilized and interrelated to varying extents. Due to the large size and the wide uses of the *Scopus* database, it is useful to closely examine the utilization and relatedness of *Scopus* subject categories.

Previous studies attempted to shed light on the characteristics of interdisciplinarity by using various subject categories. Leydesdoff and Rafols (2009) mapped approximately 171 *Web of Science* subject categories (formerly ISI) with 22 broad fields to find interesting interdisciplinary characteristics. Leydesdorff (2007) attempted to measure interdisciplinarity by using network centrality measures. Porter et al. (2007) attempted to demonstrate the interdisciplinary of disciplines by analyzing citation patterns according to the *Web of Science* subject categories. Considering all of these studies, Wagner et al. (2011) noted that differing definitions, assessment

tools, evaluation processes, and measures all shed light on different aspects of interdisciplinary scientific research. Considering the above-mentioned studies, this study aimed to assess the utilization and interrelatedness of the *Scopus* subject categories in an attempt to provide insights into the notion of interrelatedness.

Ⅱ. Data Collection and Methods

In this study, the major and minor subject categories were analyzed, focusing on the utilization and interrelatedness aspects. To conduct this study, 2017 *Scopus* journal list with October update was downloaded from the Korean Citation Index website (http://www.kci.go.kr). Journals are indexed and classified by *Scopus* and reported each year. The categorized journals with ASJC codes were extracted first in order to conduct a number of analytical procedures: a frequency count of major subject categories, a frequency count of minor subject categories, visual analysis of minor subject categories, a frequency count of co-occurring minor subject categories, centrality analysis of minor subject categories, and clustering analysis. The procedures that were undertaken can be largely divided into a frequency count approach and network analysis approach. However, depending on whether the subject category was major or minor, both approaches were used in varying degrees.

The frequency count approach, in particular, was employed in order to determine the utilization of major and minor subject categories. This included co-occurring minor subject categories. A frequency count was calculated by using a Unix-based tool such as the awk programming language (Aho, Kernighan, and Weinberger 1979). To examine the interrelatedness among the minor subject categories, VOSViewer software (Van Eck and Waltman 2009) was used to generate clusters and to visualize minor subject categories. With the clusters generated by VOSViewer, additional frequency analysis and cluster analysis were performed in order to assess the extent of interrelatedness among the major subject categories. Lastly, the centrality measures generated by Pajek software (Batagelj and Mrvar 1998) were used to assess the interrelatedness of minor subject categories.

Ⅲ. Utilization of Major Subject Categories

This study conducted a frequency count of major subject categories in order to examine the utilization of the major subject categories of *Scopus.* As previously mentioned, there are 27 major subject categories. Table 1 shows the range of ASJC codes for all 27 major subject categories, the subject category descriptions, the number of unique ASJC codes, the total number of journals, and the average number of journals of the subject category. *Multidisciplinary* (ASJC code block 1000-1099) is especially unique since no minor subject code is associated with this major subject category. Except for this subject category, all of the major subject categories comprised a number of individual minor subject categories. In the subsequent section, the minor subject categories are discussed in detail.

In Table 1, the cell that has the highest value in the respective column is indicated with a shaded color. *Medicine* shows the highest number of ASJC codes (n=49), indicating the highest number of minor subject categories. *Medicine* (n=9250) comprises the highest total number of journals. *Multidisciplinary* (ASJC code 1000) comprises the lowest number of minor subject categories as it does not contain any minor subject categories. *Economics, Econometrics & Finance* (n=4) also comprises a low number of minor subject categories.

In this table, the average number of journals per major subject category is shown. The average number of journals can be obtained by dividing the total number of journals by the number of unique ASJC codes. Compared with *Medicine,* the minor subject categories in *Social Sciences* are more evenly utilized since the average number of journals in Social Sciences (ASJC Code Block 3300-3399) is relatively higher than the average number of journals in *Medicine. Dentistry* (ASJC Code Block 3500-3599) shows relatively low minor subject categories (ASJC codes n=190). The overall cumulative number of journals for all minor subject categories is shown as 49,637. As mentioned in the previous section, there are 22,792 journals indexed by *Scopus.* From this, we can obtain the average number of minor subject categories used to categorize each journal. Since 49,637 can be divided by 22,792, each journal is categorized with 2.18 minor subject categories on average.

Assessing the utilization of the major categories is relatively straightforward. With the number of total journals used in each major subject category, we can assess the extent of the utilization

of each of the major subject categories. Overall, *Medicine* (2700-2799) has the highest degree of utilization. Social Sciences (3300-3399) has the second highest degree of utilization among the major subject categories. *Multidisciplinary* (1000-1099) is the least utilized primarily due to not having any minor subject categories.

〈Tab. 1〉 Major Subject Categories and Frequency Counts of Relevant Items

Index	ASJC Code Block Range	Description of Subject Category	No. of Minor Subject Categories	Total # of Journals in Subject Category	Avg. # of Journals in Subject Category
1	1000-1099	Multidisciplinary	0	81	81
2	1100-1199	Agricultural & Biological Sciences	12	2695	224.58
3	1200-1299	Arts & Humanities	14	4406	314.71
4	1300-1399	Biochemistry, Genetics & Molecular Biology	16	2807	175.44
5	1400-1499	Business, Management & Accounting	11	1736	157.82
6	1500-1599	Chemical Engineering	9	584	64.89
7	1600-1699	Chemistry	8	935	116.88
8	1700-1799	Computer Science	13	2142	164.77
9	1800-1899	Decision Sciences	5	349	69.80
10	1900-1999	Earth & Planetary Sciences	14	1434	102.43
11	2000-2099	Economics, Econometrics & Finance	4	1032	258.00
12	2100-2199	Energy	6	492	82.00
13	2200-2299	Engineering	17	3373	198.41
14	2300-2399	Environmental Science	13	1654	127.23
15	2400-2499	Immunology & Microbiology	7	591	84.43
16	2500-2599	Materials Science	9	1451	161.22
17	2600-2699	Mathematics	15	1841	122.73
18	2700-2799	Medicine	49	9250	188.78
19	2800-2899	Neuroscience	10	636	63.60
20	2900-2999	Nursing	23	624	27.13
21	3000-3099	Pharmacology, Toxicology & Pharmaceutics	6	809	134.83
22	3100-3199	Physics & Astronomy	11	1184	107.64
23	3200-3299	Psychology	8	1353	169.13
24	3300-3399	Social Sciences	23	7376	320.70
25	3400-3499	Veterinary	5	205	41.00
26	3500-3599	Dentistry	6	190	31.67
27	3600-3699	Health Professions	15	407	27.13
	TOTAL		329	49637	3617.95

Note: The cell that has the highest value in the respective column is indicated with a shaded color, whereas the cell that has the lowest value in the respective column is indicated with a diagonal line.

Ⅳ. Utilization and Interrelatedness of Minor Subject Categories

1. An Example of Minor Subject Categories

A detailed example of how ASJC codes are used to assign each journal is useful before attempting to assess the utilization and interrelatedness of minor subject categories. For an illustration of minor subject categories, consider Table 2. A short list of LIS journals that start with the letter 'A' is shown. Since the ASJC code of *Library and Information Science* is 3309, the category belongs to *Social Sciences.* Given that the higher-level categories are fragmented into lower subject categories, the parent subject categories – i.e., major subject categories) of these minor subject categories can be easily traced. As shown, we can see that LIS journals that start with the letter 'A' are categorized with other minor categories. Journals are often assigned with more than one ASJC codes since a journal can be categorized into multiple subject categories. In the lower half of Table 2, a frequency count of minor subject categories that occur in the upper part of the table is shown. For example, we can see that there are three LIS journals associated with *History* (ASJC 1202) since 3309 and 1202 co-occur three times in the list.

<Tab. 2> LIS Journals (ASJC Code 3309) Starting with Letter "A"

	Journal Title	ASJC Codes	
LIS Journals	Accountability in Research	3309; 3304; 2700	
	Advances in Librarianship	3309	
	Advances in Library Administration and Organization	3309; 3321	
	AIB Studi	3309	
	Analecta Hibernica	3309; 1202	
	Anales de Documentación	3309	
	Annals of Library and Information Studies	3309; 1706	
	Archival Science	3309; 1202	
	Archivaria	3309	
	Archives	3309; 1202	
	Archives and Manuscripts	3309; 3315	
	Aslib Journal of Information Management	3309; 1710	
	Australian Academic and Research Libraries	3309	
	Australian Library Journal	3309	
Minor Categories Related to LIS (ASJC Code 3309)	**Journal Category Description**	**ASJC Code**	**Freq**
	History	1202	3
	Computer Science Applications	1706	1
	Information Systems	1710	1
	General Medicine	2700	1
	Education	3304	1
	Library and Information Sciences	3309	14
	Communication	3315	1
	Public Administration	3321	1

2. Frequency Count of Minor Subject Categories

Using the previously illustrated approach, the extent of utilization of all minor subject categories can be obtained. Table 3 shows only the 20 most frequently used and 20 least frequently used minor subject categories. The most frequently minor used subject category is *2700 General Medicine* (2176 times). The ASJC code of each block that ends with '00' indicates a "general" minor subject category. Despite being related to *2700 General Medicine, 3600 General Health Professions* is used only 9 times. The second most frequently used minor subject category is *3312 Sociology & Political Science* (992 times).

Compared with the 20 most frequently used subject categories, the majority of the 20 least frequently used subject categories are related to medicine or health. In contrast, the 20 most frequently used categories comprise more diverse subject categories, despite having a high-frequency count for the *2700 General Medicine* category. The reasons for this phenomenon are not clear. As a whole, notable differences can be found in the use of *Scopus* minor subject categories.

〈Tab. 3〉 The Most and Least Used Minor Subject Categories

Group	Rank	Freq	ASJC Code and Description of Minor Subject Categories
20 Most Frequently Used	1	2176	2700 General Medicine
	2	992	3312 Sociology & Political Science
	3	954	3304 Education
	4	944	1202 History
	5	727	3316 Cultural Studies
	6	648	3310 Linguistics & Language
	7	637	1208 Literature & Literary Theory
	8	598	1203 Language & Linguistics
	9	595	3305 Geography, Planning & Development
	10	589	2208 Electrical & Electronic Engineering
	11	554	2002 Economics & Econometrics
	12	545	1105 Ecology, Evolution, Behavior & Systematics
	13	519	2210 Mechanical Engineering
	14	505	3308 Law
	15	495	1706 Computer Science Applications
	16	479	2738 Psychiatry & Mental health
	17	464	1211 Philosophy
	18	458	2739 Public Health, Environmental & Occupational Health
	19	433	3320 Political Science & International Relations
	20	422	2500 General Materials Science

20 Least Frequently Used	311	9	3600 General Health Professions
	312	8	3602 Chiropractics
	313	7	3610 Optometry
	314	7	3402 Equine
	315	7	2903 Assessment and Diagnosis
	316	7	1506 Filtration and Separation
	317	7	1504 Chemical Health and Safety
	318	6	3401 Veterinary (miscellaneous)
	319	6	2920 Pharmacology (nursing)
	320	6	2904 Care Planning
	321	5	1801 Decision Sciences (miscellaneous)
	322	4	3613 Podiatry
	323	4	2923 Review and Exam Preparation
	324	4	2709 Drug guides
	325	4	2401 Immunology and Microbiology (miscellaneous)
	326	3	3604 Emergency Medical Services
	327	2	3606 Medical Assisting and Transcription
	328	2	2915 Nurse Assisting
	329	1	3503 Dental Hygiene
	330	1	2744 Reviews and References, Medical

The overall number of subject categories used to categorize each journal was also calculated. As shown in Table 4, there are 22,792 journals that are categorized using one or more minor subject categories. A single minor subject category is the most common in categorizing *Scopus* indexed journals (34.9%). A slightly smaller percentage of journals is categorized with two minor subject categories (32.6%). Accordingly, approximately 67.5% of journals (34.9% plus 32.6%) in Scopus were categorized using one or two minor subject categories. In contrast, the total percentage of journals with 6 or more minor subject categories is negligible.

<Tab. 4> The Number of Minor Journal Subject Categories Used to Classify Journals

# of Minor Subject Categories	1	2	3	4	5	6	7	8	9	13	Total
Frequency Count	7949 (34.9%)	7438 (32.6%)	4353 (19.1%)	2008 (8.8%)	678 (3.0%)	259 (1.1%)	86 (0.4%)	18 (0.4%)	2 (0.1%)	1 (0%)	22,792 (100%)

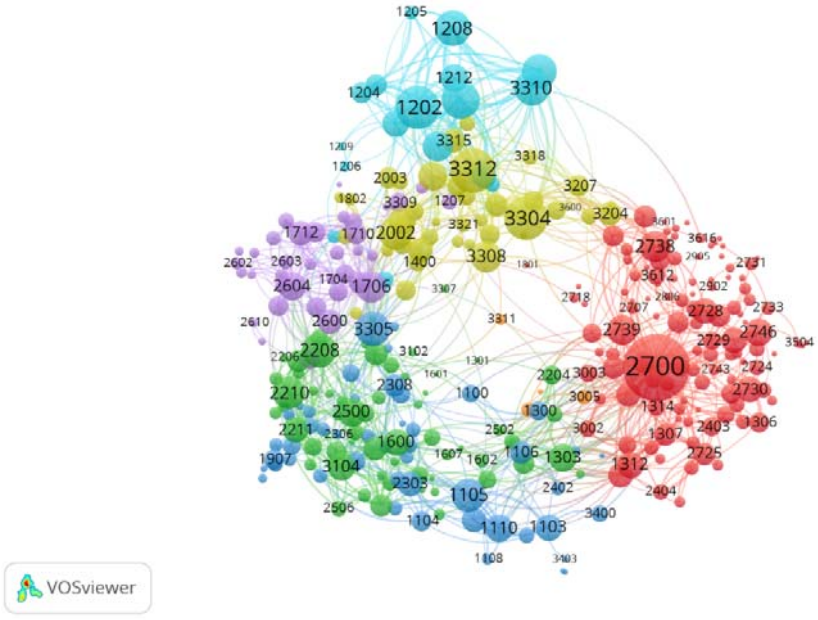
3. Visualization of Minor Subject Categories

Visualization of the minor subject categories is desirable in order to assess the utilization and

interrelatedness of minor subject categories. The frequency count of minor subject categories appears to be an adequate method of assessing the utilization of individual minor subject categories. Nevertheless, visualization of minor subject categories adds an additional perspective of the utilization and interrelatedness from a macro standpoint.

Figure 1 is generated by VOSViewer. This figure represents the networks of minor subject categories. In this figure, each node represents a minor subject category since the ASJC codes are used. To remove some negligible details, only ASJC codes with 5 or more occurrences are used; this is the default setting of VOSViewer. A map of all minor subject categories reveals a complex network structure. Here, the size of nodes varies according to the frequency count of subject categories. The densely populated distribution of nodes implies that the nodes are highly connected with each other.

Figure 1 also confirms the varying degree of utilization of the minor subject categories. The node size of 2700 *General Medicine* is the largest due to its highest frequency count. The thickness of connected lines varies depending on the frequency count of two co-occurring minor subject categories. To assess the interrelatedness among the minor subject categories, the VOSViewer generated network is further analyzed in terms of clusters in the subsequent section.



〈Fig. 1〉 Map of the Minor Subject Categories with at Least 5 Occurring Node

4. Frequently Co-Occurring Minor Subject Categories

In the case of *Scopus* minor subject categories, the varying extent of interrelatedness between two individual minor subject categories can be assessed by performing a frequency count of a pair of minor subject categories. *Scopus* uses more than one set of minor subject categories to categorize its indexed journals. For this reason, co-occurring sets of minor subject categories indicate the extent of journals that are categorized with more than one ASJC code. Table 5 shows the top 40 most frequently interrelated subject categories. Although only the top 40 frequently co-occurring minor subject categories are shown, the frequency count result suggests that there is a greater number of the co-occurring pairs of minor subject categories. If graphed, the frequency counts of co-occurring pairs would exhibit an inversely proportional trend. That is, as the index number increases, the frequency counts of co-occurring pairs of minor subject categories would decrease rapidly.

A pair of minor subject categories can exhibit qualitative characteristics due to co-occurrences. In Table 5, the interrelatedness is categorized into three types: A, B, and C. In Type A, a pair of subject categories having identical or closely identical descriptions are shown in shaded color. For example, *3310 Linguistics & Language and 1203 Language & Linguistics* are the most frequently co-occurring pair of minor subject categories, but they have almost identical descriptions. The only difference between these two descriptions is the order in which the two terms – "linguistics" and "language" – are placed, and the ASJC codes. However, one is 3310, which belongs to *Social Sciences* (3300-3399), and the other is 1203, which belongs to *Arts and Humanities* (1200-1299).

In Type B, a pair of minor subject categories are related but can be clearly distinguished in terms of subject areas. For instance, in reference to the index #2, the pair of minor subject categories – *3316 Cultural Studies and 1202 History* – share this type of interrelated relationship. The majority of co-occurring minor subject categories fall into Type B.

In Type C, a pair of minor subject categories share similar terms, but one category has a broader scope that can conceptually encompass the opposing subject category. In Type C, one subject category uses additional terms to describe a more specific subject area than the other. For instance, in reference to the index #21, *2808 Neurology* and *2728 Clinical Neurology,* would fall into Type C. Both refer to *Neurology,* but *Clinical Neurology* is a subset of Neurology due to

using the qualitative term “Clinical”. In turn, conceptually, the minor subject category *2808 Neurology* would encompass *2728 Clinical Neurology* due to the terminology used to describe the minor subject categories.

In general, all three types of occurrences of subject categories sheds light on the intricate nature of descriptive names used to categorize journal subject categories. It should be emphasized that in case of Type A and Type C, the major subject categories that are the parent to the paired minor subject categories should be observed in order to detect possible effects they have on assessing the utilization of major categories. For this purpose, the frequency count of all three types of co-occurring pairs of minor subject categories is also worth noting. There are 2 instances of Type A (5%), 32 instances of Type B (80%), and 6 instances of Type C (15%). While the Type B is most common, a noteworthy portion of minor categories are either Type A or Type C.

<Tab. 5> Top 40 Interrelated Minor Subject Categories

Index	Co-Occurring Pairs of Minor Subject Categories		Freq	Type of Co-Occurrence
	Minor Subject Category 1	Minor Subject Category 2		
1	3310 Linguistics & Language	1203 Language & Linguistics	570	A
2	3316 Cultural Studies	1202 History	272	B
3	2211 Mechanics of Materials	2210 Mechanical Engineering	206	B
4	3302 Archaeology	1204 Archaeology	197	A
5	3316 Cultural Studies	1208 Literature & Literary Theory	190	B
6	3320 Political Science & International Relations	3312 Sociology & Political Science	187	B
7	1208 Literature & Literary Theory	1203 Language & Linguistics	178	B
8	3310 Linguistics & Language	1208 Literature & Literary Theory	176	B
9	3316 Cultural Studies	3312 Sociology & Political Science	163	B
10	3312 Sociology & Political Science	1202 History	157	B
11	3203 Clinical Psychology	2738 Psychiatry & Mental health	151	B
12	2730 Oncology	1306 Cancer Research	151	B
13	1208 Literature & Literary Theory	1202 History	146	B
14	2303 Ecology	1105 Ecology, Evolution, Behavior & Systematics	136	C
15	1312 Molecular Biology	1303 Biochemistry	134	B
16	2003 Finance	2002 Economics & Econometrics	126	B
17	3316 Cultural Studies	3314 Anthropology	125	B
18	1600 General Chemistry	1500 General Chemical Engineering	121	B
19	3104 Condensed Matter Physics	2500 General Materials Science	117	B
20	1408 Strategy & Management	1403 Business & International Management	115	B
21	2808 Neurology	2728 Clinical Neurology	114	C
22	1212 Religious studies	1202 History	110	B
23	3304 Education	3204 Developmental & Educational Psychology	109	B

24	1312 Molecular Biology	1311 Genetics		109	B
25	2500 General Materials Science	2210 Mechanical Engineering		107	B
26	1110 Plant Science	1105 Ecology, Evolution, Behavior & Systematics		107	B
27	1312 Molecular Biology	1307 Cell Biology		106	B
28	3312 Sociology & Political Science	3308 Law		105	B
29	1110 Plant Science	1102 Agronomy & Crop Science		105	B
30	1105 Ecology, Evolution, Behavior & Systematics	1103 Animal Science & Zoology		105	B
31	2723 Immunology & Allergy	2403 Immunology		104	C
32	2504 Electronic, Optical & Magnetic Materials	2208 Electrical & Electronic Engineering		104	B
33	3302 Archaeology	1202 History		102	B
34	2500 General Materials Science	2211 Mechanics of Materials		102	B
35	1204 Archaeology	1202 History		101	B
36	1213 Visual Arts & Performing Arts	1208 Literature & Literary Theory		100	B
37	2712 Endocrinology, Diabetes & Metabolism	1310 Endocrinology		92	C
38	3305 Geography, Planning & Development	3303 Development		90	C
39	3316 Cultural Studies	1213 Visual Arts & Performing Arts		88	B
40	3004 Pharmacology	2736 Pharmacology (medical)		87	C

5. Centrality Measures of Minor Subject Categories

Centrality measures are often used in various problem domain, including social networks (Freeman 1978; Newman 2005), and bibliographic networks (Lee 2006). In this study, centrality measures were used to assess various aspects of interrelatedness of the minor subject categories. The commonly used centrality measures are “degree centrality”, “betweenness, centrality”, and “closeness centrality”. Degree centrality is a measurement of the number of neighboring nodes. Since a node represents a minor subject category, degree centrality is a measure that indicates the number of neighboring minor subject categories. On the other hand, betweenness centrality is the number of times a node acts as a transfer point between any pairs of nodes, indicating how often a node occurs on all shortest paths between two nodes. In our case, betweenness centrality would measure the overall interrelatedness between the connected minor subject categories, whereas closeness centrality would measure the distance from a minor subject category to all other minor subject categories. Leydesdorff (2007) argued that the closeness centrality represents “embeddedness” of the node and is more associated with multidisciplinarity rather than interdisciplinarity. The author suggested using betweenness centrality over degree centrality in measuring interdisciplinarity.

Nonetheless, in Table 6, minor subject categories having the top 10 degree centrality and the bottom 10 degree centrality measures are shown. Although this list is based on degree centrality, all three centrality measures including the frequency count of minor subject categories are shown. Also, it should be pointed out that the frequency count is the highest for this category, and it is reasonable to expect a positive correlation between the frequency count rank and the degree centrality. Previous studies reported a varying degree of correlations among these centrality measures (Valente et al. 2008; Ronda-Pupo et al. 2016). The result shown in this table also

〈Tab. 6〉 Centrality Measures of Minor Subject Categories

Rank	Minor Subject Category	Freq Count Rank	Degree Centrality	Betweenness Centrality	Closeness Centrality
1	2700 General Medicine	1	0.671	0.138	0.752
2	1706 Computer Science Applications	15	0.482	0.031	0.657
3	2739 Public Health, Environmental and Public Health	18	0.445	0.035	0.642
4	3304 Education	3	0.402	0.033	0.625
5	1303 Ageing	23	0.369	0.016	0.612
6	1305 Biotechnology	60	0.363	0.014	0.609
7	2611 Modelling and Simulation	69	0.348	0.014	0.603
8	2204 Biomedical Engineering	85	0.348	0.018	0.603
9	3312 Sociology and Political Science	2	0.329	0.012	0.597
10	2701 Medicine (miscellaneous)	86	0.323	0.018	0.594
	Omitted due to space limitation.				
320	3404 Small Animals	299	0.015	0.000	0.445
321	2923 Review and Exam Preparation	324	0.015	0.000	0.456
322	3505 Orthodontics	300	0.015	0.000	0.434
323	2709 Drug guides	323	0.015	0.000	0.416
324	3402 Equine	316	0.012	0.000	0.440
325	3610 Optometry	317	0.012	0.000	0.447
326	3606 Medical Assisting and Transcription	328	0.009	0.000	0.419
327	3613 Podiatry	325	0.009	0.000	0.367
328	2915 Nurse Assisting	327	0.009	0.000	0.331
329	2744 Reviews and References, Medical	329	0.006	0.000	0.433

suggests that minor subject categories that have a higher frequency count roughly correlate to the other centrality measures.

Adapting the approach suggested by Leydesdorff (2007), betweenness centrality is considered in estimating the extent of interrelatedness among the subject categories, although there is a general degree of correlation among the centrality measures. Table 6 shows that *2700 General Medicine* has the highest score for all of the calculated centrality measures, indicating the highest "embeddedness" due to degree centrality and the highest interrelatedness due to its betweenness centrality score.

Ⅴ. Clustering and Interrelatedness of Major Subject Categories

As a final step, clusters generated by VOSViewer were incorporated and analyzed in order to assess the interrelatedness of major subject categories. Provided that each node is represented as a minor subject category, VOSViewer automatically generates three network node indicators: *links, total links,* and *occurrences.* In the case of minor subject categories, the *links* indicate the number of connections between two adjacent minor subject categories, the *total links* indicate the number of links of subject categories plus the links of the minor subject categories that are already linked, and the *occurrences* indicate the frequency count of individual minor subject categories. VOSViewer generated a total of nine clusters. To reduce the number of clusters, 3 clusters having a low number of minor subject categories – less than 20 – were further combined with remaining six clusters.

Table 7 shows the details of the major subject categories and clusters. In this study, to reduce the size of minor subject category variations, the minor subject categories were first grouped into their respective major subject categories. Also, to reduce the network node property values of minor subject categories can be aggregated into each respective major subject category. Here, the dominant major subject category is a major subject category that has the highest value in terms of links, total links, and occurrences in each cluster. In Table 7, cells having the highest value in each cluster are shaded. The major subject category having at least one instance of the highest value is qualified as a dominant major subject category. Thus, the dominant major subject category has more weight than other major subject categories, but more than one major subject

category can be identified as dominant major subject category in a cluster. As shown in Table 7, each cluster comprises differing dominant major subject categories.

In some cases, the ASJC code block having the highest average of links differs from the ASJC code block having the highest sum of links. For instance, in Cluster 1, *Medicine* (2700-2799) has the highest value for the sum of links, the sum of total links, and the sum of occurrences, whereas *Social Sciences* (3300-3399) has the highest number of average links, average total links, and average occurrences. A major subject category having a high average value for all three aggregated node properties (links, total links, and occurrences) implies that the major subject category the minor subject categories within are on average are more highly connected than other minor subject categories.

The clusters can be compared in terms of the composition of the major subject categories. In Cluster 4, *Social Sciences* (3300-3399) is the dominant major subject category considering the average and all three types of sums, although all of the clusters differ in terms of composition of major subject categories. Except *Social Sciences* (3300-3399), none of the major subject categories in Cluster 3 are present in Cluster 4. These differences indicate that the major categories shown in Cluster 3 and 4 are less related to each other. Despite the higher frequency counts of links, total links, and occurrences, *Medicine* (2700-2799) is only present in Cluster 1. In contrast to, *Social Sciences* (3300-3399) is present in all of the clusters. Considering the composition of major categories in each cluster and the values of all of the aggregated node properties, *Social Sciences* (3300-3399) shows the highest interrelatedness among all of the major subject categories. Through the use of aggregated node properties and analysis of the composition of clusters, the interrelatedness among the major subject categories can be assessed, although no precise measurement is offered in this study.

Lastly, it should be pointed out that various details of clusters are dependent upon the specific method used to generate the clusters. In this study, VOSViewer was used to generate clusters. The technical aspects of the method used to generate clusters are described by Waltman, Van Eck, and Noyons (2010). Furthermore, the actual number of clusters would be dependent upon other parameters, such as the frequency count of ASJC codes. Despite this methodological limitation, the clustering method seemed to be effective in assessing the interrelatedness among the major subject categories of *Scopus.*

〈Tab. 7〉 Cluster Details of Major Subject Categories

Cluster #	Dominant Major Subject Category	ASJC Code Block and Major Subject Categories	Aggregated Node Properties					
			Links		Total Links		Occurrences	
			AVG	SUM	AVG	SUM	AVG	SUM
1	Medicine; Social Sciences	1300-1399 Biochemistry, Genetics & Molecular Biology	64	639	181	4335	434	1805
		1800-1899 Decision Sciences	8	8	5	9	9	5
		2400-2499 Immunology & Microbiology	42	167	113	954	239	453
		2700-2799 Medicine	51	2398	197	13157	280	9245
		2800-2899 Neuroscience	39	393	64	1369	137	636
		2900-2999 Nursing	19	401	29	976	47	618
		3000-3099 Pharmacology, Toxicology & Pharmaceutics	51	255	140	1394	279	700
		3200-3299 Psychology	41	82	153	529	265	306
		3300-3399 Social Sciences	99	99	235	515	515	235
		3500-3599 Dentistry	14	70	38	168	34	189
		3600-3699 Health Professions	20	214	35	680	62	389
2	Biochemistry, Genetics &Molecular Biology; Engineering	1300-1399 Biochemistry, Genetics & Molecular Biology	76	378	162	2164	433	812
		1500-1599 Chemical Engineering	44	353	72	1447	181	577
		1600-1699 Chemistry	53	420	117	2029	254	935
		1700-1799 Computer Science	41	41	30	55	55	30
		1900-1999 Earth & Planetary Sciences	35	35	77	123	123	77
		2100-2199 Energy	50	248	87	1065	213	435
		2200-2299 Engineering	75	905	224	5981	498	2684
		2500-2599 Materials Science	53	476	161	3588	399	1451
		3100-3199 Physics & Astronomy	42	374	105	2073	230	949
		3300-3399 Social Sciences	58	58	36	87	87	36
3	Agricultural & Biological Sciences; Environment Science; Social Sciences	1100-1199 Agricultural & Biological Sciences	58	690	225	4322	360	2695
		1300-1399 Biochemistry, Genetics & Molecular Biology	79	79	190	341	341	190
		1900-1999 Earth & Planetary Sciences	41	535	104	2208	170	1357
		2100-2199 Energy	55	55	57	100	100	57
		2200-2299 Engineering	46	46	79	167	167	79
		2300-2399 Environmental Science	66	791	129	3546	296	1550
		2400-2499 Immunology & Microbiology	44	88	67	367	184	134
		3300-3399 Social Sciences	101	101	595	1015	1015	595
		3400-3499 Veterinaries	16	78	41	193	39	205
4	Social Sciences	1200-1299 Arts & Humanities	62	124	186	690	345	372
		1400-1499 Business, Management & Accounting	56	611	158	3309	301	1736
		1800-1899 Decision Sciences	44	133	80	577	192	239
		2000-2099 Economics, Econometrics & Finance	56	223	258	1475	369	1032
		3200-3299 Psychology	57	340	175	1904	317	1047
		3300-3399 Social Sciences	67	874	324	6187	476	4208
		3600-3699 Health Professions	10	10	9	13	13	9
5	Arts & Humanities; Social Sciences	1200-1299 Arts & Humanities	41	453	356	6590	599	3910
		2200-2299 Engineering	43	43	89	189	189	89
		3300-3399 Social Sciences	63	314	409	4333	867	2043
6	Environmental Science; Pharmacology, Toxicology & Pharmaceutics	1500-1599 Chemical Engineering	12	12	7	22	22	7
		2300-2399 Environmental Science	63	63	104	268	268	104
		3000-3099 Pharmacology, Toxicology & Pharmaceutics	62	62	109	211	211	109
		3300-3399 Social Sciences	54	54	64	140	140	64

Ⅵ. Discussion and Conclusion

So far, various methodological procedures have been used to assess the utilization and interrelatedness of *Scopus* subject categories. The methodological procedures used in this study include frequency analysis, visualization of minor subject categories, centrality measure analysis, and cluster analysis. The result showed varying degrees of utilization and interrelatedness among the major and minor subject categories. Regarding the utilization, both major and minor subject categories of *Scopus* are unevenly utilized. At the major category level, *Medicine* showed the highest degree of utilization in the major subject categories. Social Sciences showed the second highest degree of utilization among the major subject categories. In contrast, some minor categories in *Medicine* were particularly underutilized, although the minor subject category *2700 General Medicine* was used the most in categorizing journals.

A considerable number of subject categories were interrelated. When the major subject categories grouped using the clusters, *Social Sciences* showed a high degree of interrelatedness with other major subject categories by being present in the greatest number of clusters. *Social Sciences* comprised more diverse minor subject categories than *Medicine* and were dominant among the clusters of major subject categories. Thus, at a major category level, *Social Sciences* overall have a greater degree of interrelatedness than *Medicine*.

In addition to the frequency analysis procedures, a network analysis with a co-occurrence analysis was an effective means to measure the interrelatedness of minor subject categories. The betweenness centrality was an effective means to estimate the extent of interrelatedness. In terms of betweenness centrality score, *2700 General Medicine* had the highest level of interrelatedness. A frequency count of co-occurring minor subject categories showed some interesting characteristics of two opposing subject categories that required a qualitative categorization. Pairs of minor subject categories showed the following characteristics: a) two subject categories having identical or closely identical descriptions, b) two subject categories having an inter-relationship via subject categorization, and c) one category conceptually encompassing another category by having a broader concept. The description names associated with minor subject categories could be problematic in some cases since some minor subject categories were extremely close to each other with only negligible conceptual differences. Yet the parent of the minor categories differed from each other.

Considering the above-mentioned results, this study has broader implications for other studies that use major subject categories of *Scopus*. In terms of methodological procedures, the co-occurrence analysis of minor subject categories along with the centrality analysis appeared to be straightforward in assessing the interrelatedness of minor subject categories. In contrast, assess the interrelatedness of major subject categories is not straightforward. Because the clustering method provides a means to examine the composition of major subject categories, the clustering method indicates the interrelatedness of the major subject categories. As the extent of minor subject categories with respect to the major subject category can vary considerably, the clustering analysis allows us to recognize the major subject category distribution and its significance among the clusters.

Furthermore, this study suggests that intricately interrelated minor subject categories could be underestimated, affecting the result of studies based on just major subject categories. The general practice of using the *Scopus* major subject categories may not be satisfactorily reliable in creating a dataset and interpreting the results based on major categories in research studies. In particular, this study demonstrated that some individual minor categories were highly interrelated and utilized but these categories were not necessarily dominant on the major category level. The utilization and interrelatedness of major categories differed from the minor subject categories due to aggregation of all associated minor subject categories. An example of this was *2700 General Medicine* being highly interrelated among the minor subject categories. However, *2700-2799 Medicine* was relatively confined within one cluster, and at the major category level, the interrelatedness was relatively reduced. For this reason, although using the major subject categories is convenient, the minor subject categories may need to be closely examined to avoid possible misinterpretation of subject categories. The drawback of examining relevant issues at hand is that it requires a more extensive set of methodological procedures.

Wang and Waltman (2016) suggested that journal categorization practices of *Scopus* might not be entirely satisfactory as journals indexed by *Scopus* are assigned with too many subject categories. While the results of this study neither confirm nor support the view of the authors, uneven utilization, extensive interrelated minor subject categories, and having overly close categorizations do suggest some problems associated with the existing categorization method. However, it should be also recognized that developing a comprehensive subject category of journals is inherently difficult due to multi-subject nature of journals and overlapping nature of subject categories. Regardless of the issues related to the predefined subject categories, for the

researchers who rely on the using the subject categories, all relevant subject categories and the limitation of using the aggregated subject categories need to be taken into account.

Lastly, it should also be noted that this research is empirically based. The utilization and the interrelatedness among subject categories are likely to evolve slowly over time as the journals indexed by Scopus are added and deleted. Considering this limitation, further investigation into the Scopus interrelated subject categories is recommended for future research by means of incorporating various bibliometric attributes, such as keywords, citations, number of authors, and countries of affiliation. For instance, Gómez-Núñez et al. (2011) proposed analyzing references in citing journals in order to reorganize the existing subject categories of *Scopus.* Nonetheless, in a broader sense, how journals with varying groups of subject categories differ in terms of bibliometric attributes could be investigated in order to gain additional insights into the inter-dependent nature of subject categories in bibliographic databases.

References

Abrizah, Abdullah et al. 2013. "LIS Journals Scientific Impact and Subject Categorization: A Comparison Between Web of Science and Scopus." *Scientometrics,* 94(2): 721-740.

Aho, Alfred V., Brian W. Kernighan and Peter J. Weinberger. 1979. "Awk—a Pattern Scanning and Processing Language." *Software: Practice and Experience,* 9(4): 267-279.

Batagelj, Vladimir and Andrej Mrvar, A. 1998. "Pajek-Program for Large Network Analysis." *Connections,* 21(2): 47-57.

Bervkens, Peter. 2012. "SciVerse Scopus Custom Data Documentation." Available from: http://ebrp.elsevier.com/pdf/Scopus_Custom_Data_Documentation_v4.pdf

Chung, Jenny and Ming-Yueh Tsay. 2017. "A Bibliometric Analysis of the Literature on Open Access in Scopus." *Qualitative and Quantitative Methods in Libraries,* 4(4): 821-841.

Dorta-González, Pablo and Yolanda. Santana-Jiménez. 2017. "Prevalence and Citation Advantage of Gold Open Access in The Subject Areas of the Scopus Database." *Research Evaluation,* 27(1): 1-15.

Freeman, Linton. C. 1978. "Centrality in Social Networks Conceptual Clarification." *Social Networks,* 1(3): 215-239.

García, Jose A., Rosa Rodriguez-Sánchez and J. Fdez-Valdivia. 2011. "Ranking of the Subject

Areas of Scopus." *Journal of the American Society for Information Science and Technology,* 62(10): 2013-2023.

Gómez-Núñez, Antonio J. et al. 2011. "Improving SCImago Journal and Country Rank (SJR) Subject Classification Through Reference Analysis." *Scientometrics,* 89(3): 741.

Hassan, Saeed-Ul et al. 2017. "Measuring Social Media Activity of Scientific Literature: An Exhaustive Comparison of Scopus and Novel Altmetrics Big Data." *Scientometrics,* 113(2): 1037-1057.

Kim, Jinkwang, Sohyung Kim and Changhyuck Oh. 2016. "A Classification of the Journals in KCI Using Network Clustering Methods." *Journal of the Korean Data and Information Science Society,* 27(4): 947-957.

Klarenbeek, Tracy and Nelius Boshoff. 2018. "Measuring Multidisciplinary Health Research at South African Universities: A Comparative Analysis Based on Co-Authorships and Journal Subject Categories. *Scientometrics,* 116(3), 1461-1485.

Kosecki, Stanislaw, Robin Shoemaker and Charlotte Kirk Baer. 2011. "Scope, Characteristics, and Use of the US Department of Agriculture's Intramural Research." *Scientometrics,* 88(3): 707-728.

Lee, Jaeyoon. 2006. "Centrality Measures for Bibliometric Network Analysis." *Journal of the Korean Society for Library and Information Science,* 40(3): 191-214.

Leydesdorff, Loet. 2007. "Betweenness Centrality as an Indicator of the Interdisciplinarity of Scientific Journals." *Journal of the American Society for Information Science and Technology,* 58(9): 1303-1319.

Leydesdorff, Loet. and Ismael Rafols. 2009. "A Global Map of Science Based on the ISI Subject Categories." *Journal of the American Society for Information Science and Technology,* 60(2): 348-362.

Leydesdorff, Loet and Lutz Bornmann. 2016. "The Operationalization of Fields as WoS Subject Categories (WCS) in Evaluative Bibliometrics: The Cases of Library and Information Science and Science and Technology Studies." *Journal of the Association for Information Science and Technology,* 67(3): 707-714.

Martín-Martín, Alberto et al. 2018. "Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories." *Journal of Informetrics,* 12(4), 1160-1177.

Minguet, Fernando et al. 2017. "Redefining the Pharmacology and Pharmacy Subject Category in the Journal Citation Reports Using Medical Subject Headings (Mesh)." *International Journal of Clinical Pharmacy,* 39(5): 989-997.

Newman, Mark EJ. 2005. “A Measure of Betweenness Centrality Based on Random Walks.” *Social Networks,* 27(1): 39-54.

Porter, Alan et al. 2007. “Measuring Researcher Interdisciplinarity.” *Scientometrics,* 72(1): 117-147.

Ronda-Pupo, Guillermo Armando, Yesenia Ronda-Danta and Yusleydis Leyva-Pupo. 2016. “Correlation Between a Country's Centrality Measures and the Impact of Research Paper: The Case of Biotechnology Research in Latin America.” *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información,* 30(69): 73-92.

Stanton, Colleen Maura. 2014. *A Health Promoting Continuous Learning Sustainable Education System.* PhD Thesis. University of Toronto, Canada.

Thelwall, Mike. 2017. “Are Mendeley Reader Counts High Enough for Research Evaluations When Articles Are Published?” *Aslib Journal of Information Management,* 69(2): 174-183.

Valente, Thomas W. et al. 2008. “How Correlated Are Network Centrality Measures?” *Connections (Toronto, Ont.),* 28(1): 16.

Van Eck, Nees and Ludo Waltman. 2009. “Software Survey: Vosviewer, A Computer Program for Bibliometric Mapping.” *Scientometrics,* 84(2): 523-538.

Waltman, Ludo, Nees Jan Van Eck and Ed CM Noyons. 2010. “A Unified Approach to Mapping and Clustering of Bibliometric Networks.” *Journal of Informetrics,* 4(4): 629-635.

Wagner, Caroline S. et al. 2011. “Approaches to Understanding and Measuring Interdisciplinary Scientific Research (IDR): A Review of the Literature.” *Journal of informetrics,* 5(1): 14-26.

Wang, Qi and Ludo Waltman. 2016. “Large-Scale Analysis of the Accuracy of the Journal Classification Systems of Web of Science and Scopus.” *Journal of Informetrics,* 10(2): 347-364.

Yan, Erjia. 2016. “Disciplinary Knowledge Production and Diffusion in Science.” *Journal of the Association for Information Science and Technology,* 67(9): 2223-2245.

Zuccala, Alesia and Raf Guns. 2013. “Comparing Book Citations in Humanities Journals to Library Holdings: Scholarly Use Versus Perceived Cultural Benefit.” *14th International Conference of the International Society for Scientometrics and Informetrics,* 1: 353-360.

참고문헌

김진광, 김소형, 오창혁. 2016. KCI 등재 학술지의 분류를 위한 네크워크 군집화 방법의 비교. 『한국데이터정보과학회지』, 27(4): 947-957.

이재윤. 2006. 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 『한국문헌정보학회지』, 40(3): 191-214.