

Automatic Classification of Department Types and Analysis of Co-Authorship Network: Focusing on Korean Journals in the Computer Field

Byungkyu Kim*, Beom-Jong You*, Min-Woo Park*

*Principal Researcher, Science and Technology Security R&D Center, KISTI, Daejeon, Korea

*Principal Researcher, Science and Technology Security R&D Center, KISTI, Daejeon, Korea

*Principal Researcher, Science and Technology Security R&D Center, KISTI, Daejeon, Korea

[Abstract]

The utilization of department information in bibliometric analysis using scientific and technological literature is highly advantageous. In this paper, the department information dataset was built through the screening, data refinement, and classification processing of authors' department type belonging to university institutions appearing in academic journals in the field of science and technology published in Korea, and the automatic classification model based on deep learning was developed using the department information dataset as learning data and verification data. In addition, we analyzed the co-authorship structure and network in the field of computer science using the department information dataset and affiliation information of authors from domestic academic journals. The research resulted in a 98.6% accuracy rate for the automatic classification model using Korean department information. Moreover, the co-authorship patterns of Korean researchers in the computer science and engineering field, along with the characteristics and centralities of the co-author network based on institution type, region, institution, and department type, were identified in detail and visually presented on a map.

▶ **Key words:** Computer Science, Academic Department Classification, Co-authorship, Social Network Analysis

[요약]

과학기술 문헌을 활용한 계량정보분석에서 학과정보의 활용은 매우 유용하다. 본 논문에서는 국내 과학기술 분야 학술지 논문에 출현하는 대학기관 소속 저자의 학과정보 선별, 데이터 정제와 학과유형 분류 처리 과정을 통해 학과정보 데이터셋을 구축하고 학습데이터와 검증데이터로 이용하여 딥러닝 기반의 자동분류 모델을 구현하였다. 또한 학과정보 데이터셋과 국내 학술지 저자소속 정보를 활용하여 컴퓨터 분야의 공저 구성 현황과 네트워크를 분석하였다. 연구결과, 자동분류 모델은 한글 학과정보 기준 98.6% 정확률을 보였으며 컴퓨터 분야 연구자들의 공저 패턴과 기관유형, 지역, 기관, 학과유형 측면별 공저 네트워크의 속성과 중심성이 자세히 파악되고 맵으로 시각화되었다.

▶ **주제어:** 컴퓨터 분야, 학과유형 분류, 공저 관계, 소셜네트워크분석

• First Author: Byungkyu Kim, Corresponding Author: Byungkyu Kim

*Byungkyu Kim (bk.kim@kisti.re.kr), Science and Technology Security R&D Center, KISTI

*Beom-Jong You (you@kisti.re.kr), Science and Technology Security R&D Center, KISTI

*Min-Woo Park (pminwoo@kisti.re.kr), Science and Technology Security R&D Center, KISTI

• Received: 2023. 03. 14, Revised: 2023. 04. 12, Accepted: 2023. 04. 12.

• This article is an extension of the conference paper : "Dataset construction and Automatic classification of Department information appearing in Domestic journals" (the 67th conference of the Korean Society of Computer Information, Awarded Outstanding Paper).

I. Introduction

세계경제포럼 (WEF)의 국가경쟁력 평가 결과에 따르면 한국의 정보통신기술 (ICT) 보급률은 2018년부터 2021년까지 4년 연속으로 세계 1위를 달성하였으며[1], 2021년에 실시된 서울대 정부경쟁력 연구센터의 정부경쟁력 분석 결과에서 우리나라의 ICT 분야 경쟁력 순위는 OECD 국가 중 3위로 조사되었대[2]. 또한 한국연구재단은 '2010~2020년 주요국의 SCI급 논문 피인용 상위 1% 논문실적'을 비교 분석한 결과에서 한국의 컴퓨터과학 분야는 11위로 조사되어 한국의 강점 분야로 조사되었대[3]. 이와 같은 세계 최고 수준의 디지털 인프라 수준과 기술 경쟁력 확보에는 1970년대부터 시작된 ICT의 기반 학문인 컴퓨터 분야에서의 과학 교육과 국가 연구개발 사업의 역할과 기여가 매우 크다. 최근에는 미래 첨단과학기술 산업을 견인할 인공지능 (AI) 기술이 비약적으로 발전하고 있으며 컴퓨터 연구 분야의 중요성은 날로 증대되고 있다.

본 논문에서는 국내 컴퓨터 분야 연구자들의 지적관계 규명 즉, 공동연구 현황, 공저 구성 및 세부 특성을 살펴 보기 위하여 다양한 측면에서 공저 현황과 공저 네트워크를 자세히 분석하였다. 이를 위해 컴퓨터 분야의 대표적인 한국 학술지들에 수록된 논문 저자소속 정보, 기관식별 정보와 관련 표준 분류체계를 기반 데이터로 사용하고 분석을 위해 계량정보분석과 소셜네트워크분석 기법을 사용하였다. 특별히, 연구자의 전공 및 주제적 특성을 잘 드러낼 수 있는 저자의 학과정보에 대한 자동 분류 모델 및 데이터셋을 구축하고 공저 관계 분석에 활용하였다 [4]. 본 논문의 구성은 2장과 3장에서 관련 연구와 연구 재료 및 연구 방법을 살펴보고 4장에서 주요 연구 결과를 기술하며 마지막 5장에서 결론을 맺는다.

II. Related works

관련 문헌의 고찰을 위해 연구진이 기존에 수행한 선행 연구들과 국내 및 해외의 관련 문헌들을 차례대로 살펴보면 다음과 같다. 2018년도 연구에서는 한국과학기술인용색인DB를 기반으로 컴퓨터 분야를 포함한 주제분야별로 피인용 SCI 학술지들의 순위를 분석하였다[5]. 2019년 연구에서는 컴퓨터 분야 국내 학술논문에 인용되는 SCI 저널 현황과 이들의 동시출현 네트워크를 분석하였으며[6], 한국 연구자의 컴퓨터 분야 SCI 저널 출판 논문에 대한 분석 연구를 수행하였다[7]. 2021년에는 국내 학회에서 출판

한 컴퓨터 분야 국·영문 논문 분석을 통해 저자, 국가 등의 공동연구 현황과 관계를 분석하였다[8]. 2022년에는 방재 및 안전공학 소속 저자들의 국내 학술지 논문들을 대상으로 국내 재난안전 분야 연구현황을 분석하였다[9].

국내에서는 문헌정보학 분야를 중심으로 학술 문헌을 대상으로 하는 계량정보분석 연구가 활발히 수행되고 있으며, 문헌정보학 분야 한국 학술지들을 대상으로 공저 네트워크 분석 연구가 수행되었대[10]. 또한 2018년에는 건축학 분야에서 공저자 네트워크 분석 연구[11], 컴퓨터 분야에서는 2016년과 2018년에 공저자 분석 연구들이 수행된 바 있으며[12-13], 이들의 연구는 모두 소셜네트워크분석 (SNA, Social Network Analysis) 기법을 사용하여 주요 네트워크 중심성 지표들을 분석하였다.

마지막으로 해외의 관련 연구를 살펴보면, 2010년에 컴퓨터 네트워크 분야에서의 공저 관계 분석 연구[14]와 2019년에 컴퓨터 네트워크 과학자들의 20년 동안의 공저 관계를 분석하였다[15]. 최근에 수행된 연구에서는 브라질의 컴퓨터 분야 대학원 프로그램에서의 공저 네트워크 분석 연구가 수행되었대[16]. 본 논문에서는 기존 선행 연구를 통해 구축하고 활용한 국내 학술정보, 기관식별데이터 등의 데이터셋과 동시출현 네트워크 분석 방법 등을 이용하고 한국 학술지 저자 소속정보 기반 학과유형 데이터셋 구축과 자동 분류 모델 개발을 통해 차별성 있는 국내 컴퓨터 분야 공저 네트워크 분석 연구를 시도하였다.

III. Materials and Research Method

국내 컴퓨터 분야 연구자들의 공저 관계 분석과 자동 분류의 기반이 되는 학과정보 데이터셋 구축과 실험데이터 구성을 위해 한국과학기술정보연구원 (KISTI)에서 개발한 한국과학기술인용색인DB (KSCD, Korea Science Citation Database)의 저자소속 정보를 포함한 학술정보와 개별 기관의 상세정보가 포함된 약 25만건 규모의 기관식별데이터를 사용하였다[17-18]. 또한 저자의 소속 기관과 학과 유형 분류를 위하여 기관식별데이터의 기관유형 분류체계와 한국대학교육협의회 (KCUE)의 교육편제 단위 표준분류체계를 사용하였다[17]. 기관유형 분류체계는 대분류 (5개), 중분류 (20), 소분류 (37)의 3개 레벨로 구성되며, 본 논문에서는 대분류, 중분류까지를 사용하였다. KCUE 교육편제단위 표준분류체계는 대분류 (5), 중분류 (27), 소분류 (151)의 3개 레벨로 구성되며, 본 논문에서는 대분류, 중분류까지를 사용하였다. 실험데이터의

정보들을 바탕으로 공저 현황 분석, 공저 기관 분석, 공저 관계 소셜네트워크 분석을 수행하였다. 통계 분석과 결과의 시각화는 엑셀과 파이썬을 사용하였으며, 공저 네트워크 상에서 노드들의 상대적 중요성을 측정하기 위하여 소셜네트워크분석 연구에서 주로 사용되는 연결 중심성 (Degree Centrality), 매개 중심성 (Betweenness Centrality), 근접 중심성 (Closeness Centrality)과 같은 네트워크 중심성 척도들을 이용하였다[20]. 네트워크 중심성 측정과 네트워크의 시각화를 위해서 파이썬에서 네트워크 분석을 위해 사용되는 NetworkX (3.0 버전)와 Vosviewer (1.6.18 버전)를 사용하였다[21-22]. 또한 공동연구의 정도에 대한 측정 방식은 Subramanyam이 1983년 논문에서 제시한 공식을 사용하였으며[23], 학과 유형 정보의 딥러닝 기반 자동분류 모델 구현을 위해서는 LSTM (Long Short-Term Memory models)을 이용하였다. 본 연구의 분석 대상 컴퓨터 분야 한국 학술지와 출판 논문 범위는 Table 1과 같으며, 연구 범위 및 단계별 수행 내용을 요약하면 Fig 1과 같다.

Table 1. Experimental data status

Journals	2015	2016	2017	Total
IEMEK Journal of Embedded Systems and Applications	44	43	46	133
Journal of Korea Society of Digital Industry and Information Management	73	62	72	207
Journal of Digital Contents Society	101	68	198	367
Journal of Internet Computing and Services	171	216	206	593
Journal of KIISE	7	27	10	44
Journal of Korea Multimedia Society	40	60	40	140
Journal of the Korea Computer Graphics Society	211	193	196	600
Journal of the Korea Institute of Information Security & Cryptology	77	93	82	252
Journal of The Korea Internet of Things Society	46	37	36	119
Journal of the Korea Society of Computer and Information	181	162	152	495
KIISE Transactions on Computing Practices	113	95	96	304
KIPS Transactions on Computer and Communication Systems	143	144	129	416
KIPS Transactions on Software and Data Engineering	55	59	57	171
Smart Media Journal	73	83	72	228
The Journal of Society for e-Business Studies	20	21	39	80
The Journal of the Institute of Internet, Broadcasting and Communication	244	239	228	711
Total	1,599	1,602	1,659	4,860

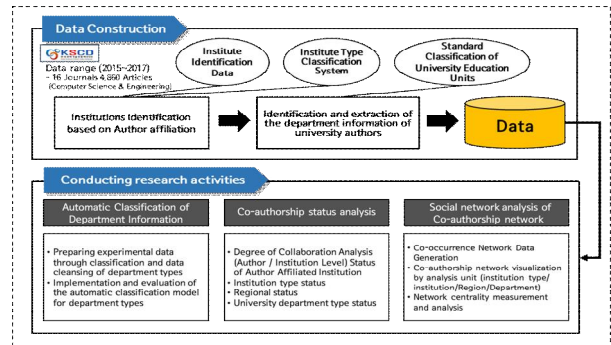


Fig. 1. Research Methods and Design

IV. Results

1. Department information classification dataset and automatic classification model

본 장에서는 컴퓨터분야 한국 학술지 저자의 학과정보 자동식별 및 공저 네트워크 분석의 연구 수행 결과를 제시 하며, 먼저 학과유형 정보의 데이터셋 구축과 자동 분류 모델 관련 주요 내용을 살펴보면 다음과 같다. 국내 학술지에 출현하는 학과정보 데이터셋 구축을 위하여 KISTI가 2002년부터 개발해온 KSCD를 활용하였다. KSCD에 등재된 국내 학술지 500종의 2015년부터 2017년까지의 논문 83,633건에서 대학기관 저자 기준으로 159,083개의 저자 소속 정보를 사용하였으며, 불필요한 정보의 제거 등 데이터 정제를 위한 전처리 작업을 실시하고 학과유형 분류를 수작업으로 진행하였다. 학과유형 분류 시 KCUE에서 제공하는 국내 교육편제단위 표준분류체계를 기준으로 삼았다. 딥러닝 기반의 학과정보 자동분류 모델은 기존 RNN (Recurrent Neural Networks) 보다 성능이 개선된 순환 신경망인 LSTM을 사용하였다. 학과명 한글 데이터의 형태소 분석을 위해 Mecab을 사용하였으며 불필요한 단어를 불용어 목록으로 정리하여 해당 단어는 제외되도록 하였다. 모델의 구성은 기울기 소실 문제를 해결할 수 있는 ReLU 함수와 출력계층에서 활성화 함수는 다중 클래스 분류를 위한 SoftMax 함수를 사용하였으며, 손실함수는 다중분류를 위한 categorical_crossentropy를 이용하였다. 학습과 테스트 데이터는 7:3 비율로 나누어 학습을 수행하였으며 학습 횟수는 30번으로 하고 과적합 방지를 위해 조기 종료 기법을 이용하였다. 학과정보 데이터 (국문 38,618건, 영문 44,970건)에서 출현 수가 많은 분류 선택 및 출현 횟수 3회 이상 조건으로 필터링하여 최종적으로 학과유형 정보 데이터셋 (한글 35,166건, 영문 37,473건)을 구축하였다. 학과유형 정보 데이터 예시와 데이터셋의 학과유형별 구성은 각각 Table 2, Fig 2와 같다.

Table. 2. Example dataset for classification of department types

Department Name (High Frequency)	Dept. Type Section Level	Dept. Type Division Level	Dept. Type Group Level
department of nursing	NATURAL SCIENCE	NURSE	NURSING SCIENCE
college of nursing	NATURAL SCIENCE	NURSE	NURSING SCIENCE
department of statistics	NATURAL SCIENCE	MATHEMATICS, PHYSICS, ASTRONOMY, EARTH	STATISTICS
department of civil engineering	ENGINEERING	CONSTRUCTION	CIVIL ENGINEERING
department of architectural engineering	ENGINEERING	CONSTRUCTION	ARCHITECTURAL ENGINEERING
department of mathematics	NATURAL SCIENCE	MATHEMATICS, PHYSICS, ASTRONOMY, EARTH	MATHEMATICS
department of environmental engineering	ENGINEERING	CONSTRUCTION	ENVIRONMENTAL ENGINEERING
department of mechanical engineering	ENGINEERING	MACHINERY	MECHANICAL ENGINEERING
department of prosthodontics, school of dentistry	MEDICINE	MEDICAL TREATMENT	DENTISTRY
department of chemical engineering	ENGINEERING	CHEMICALS, POLYMERS, ENERGY	CHEMICAL ENGINEERING
department of dental hygiene	NATURAL SCIENCE	HEALTH	CLINICAL HEALTH SCIENCES
department of electronic engineering	ENGINEERING	ELECTRIC/ELECTRONICS / COMPUTERS	ELECTRICAL ENGINEERING
department of radiological science	NATURAL SCIENCE	HEALTH	CLINICAL HEALTH SCIENCES
school of mechanical engineering	ENGINEERING	MACHINERY	MECHANICAL ENGINEERING
department of civil and environmental engineering	ENGINEERING	CONSTRUCTION	CIVIL ENGINEERING
department of computer engineering	ENGINEERING	ELECTRIC/ELECTRONICS/ COMPUTERS	COMPUTER SCIENCE/COMPUTER SCIENCE
department of naval architecture and ocean engineering	ENGINEERING	MACHINERY	NAVAL ARCHITECTURE AND OCEAN ENGINEERING
department of architecture	ENGINEERING	CONSTRUCTION	ARCHITECTURE
college of pharmacy	NATURAL SCIENCE	PHARMACY	PHARMACY
college of korean medicine	MEDICINE	MEDICAL TREATMENT	KOREAN MEDICINE
dept. of computer engineering	ENGINEERING	ELECTRIC/ELECTRONICS / COMPUTERS	COMPUTER SCIENCE / COMPUTER SCIENCE
college of veterinary medicine	MEDICINE	MEDICAL TREATMENT	VETERINARY
department of materials science and engineering	ENGINEERING	MATERIALS	ADVANCED MATERIAL ENGINEERING
school of architecture	ENGINEERING	CONSTRUCTION	ARCHITECTURE
department of physics	NATURAL SCIENCE	MATHEMATICS, PHYSICS, ASTRONOMY, EARTH	PHYSICS
department of aerospace engineering	ENGINEERING	MACHINERY	AEROSPACE ENGINEERING
dept. of nursing	NATURAL SCIENCE	NURSE	NURSING SCIENCE
department of animal science and biotechnology	NATURAL SCIENCE	AGRICULTURE, FORESTRY AND FISHERIES	LIVESTOCK SCIENCE
dept. of business administration	HUMANITIES AND SOCIAL SCIENCE	MANAGEMENT AND ECONOMY	BUSINESS MANAGEMENT
dept. of clothing & textiles	NATURAL SCIENCE	LIFE SCIENCE	DRESSMAKING
department of electronics engineering	ENGINEERING	ELECTRIC/ELECTRONICS /COMPUTERS	ELECTRICAL ENGINEERING

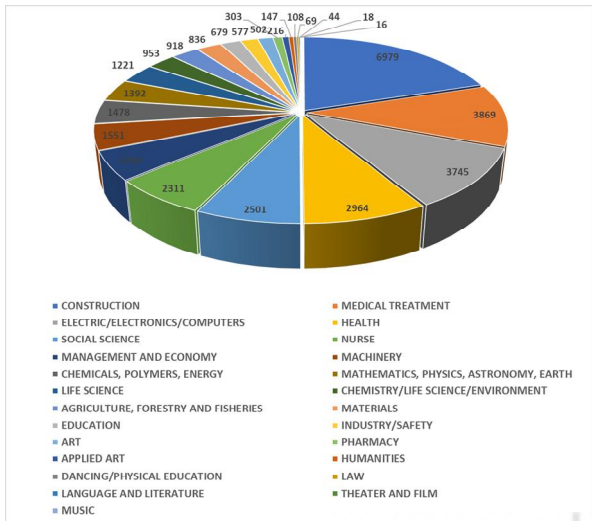


Fig. 2. Dataset for Classification of Types of Departments

국문 학과정보 데이터를 학과유형 중분류 기준으로 10순위 까지 살펴보면, “건설” (19.8%), “의료” (11.0%), “전기·전자 컴퓨터” (10.6%), “보건” (8.4%), “사회과학” (7.1%), “간호” (6.6%), “경영·경제” (5.0%), “기계” (4.4%), “항공·고분자 에너지” (4.2%), “수학·물리·천문·지구” (4.0%)로 파악되었다.

학과정보 자동 분류 모델은 텍스트 언어별로 한글 버전과 영문 버전으로 이원화하여 구현하였다. 모델 성능 평가 결과는 한글과 영문 각각 98.6%, 97.6%의 정확률로 측정되었으며 Fig 3은 학과유형별 모델 예측 결과를 실제값과 결과값을 비교하여 보여주는 그래프이다. 실험 데이터에 출현한 학과유형은 소분류로 128개로써, 앞서 Fig 2에서 살펴본 바와 같이 유형별로 데이터셋 구성에서 유형별로 차지하는 비중의 편차가 크다. 특히, 학습데이터 수가 상대적으로 작은 학과유형의 경우 정확률이 낮아

지며, 향후 데이터 보완이 필요하다.

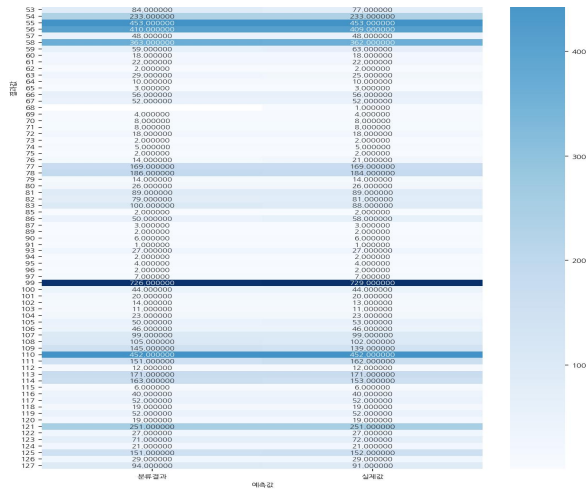


Fig. 3. Automatic classification results by dept. type

2. Co-Authorship Patterns and Network Analysis of Korean Computer Science journals

4.2 절에서는 실험데이터를 구성하는 한국 학술지 16 종, 논문 4,860 편의 저자소속 정보와 기관 식별정보, 4.1절에서 구축한 학과유형 데이터셋으로 구성된 실험데이터를 기반으로 공저 구성, 협업 정도 등의 공저 패턴과 기관유형, 국내 지역, 기관, 학과유형의 4가지 측면에서 공저 현황과

네트워크를 분석한 연구 결과를 제시한다. 먼저, 실험 대상 논문들의 저자 수 구성을 학술지별, 전체 학술지로 살펴보면 Table 3 및 Fig 4와 같다. 논문별 저자 수 평균은 전체 학술지 기준으로 2.5 명으로 파악되었으며, 학술지별로 저자 수의 구성 패턴과 단독연구와 공동연구의 비율에서 상당한 차이가 존재하는 것이 확인되었다. 학술지 전체를 기준으로 공동연구와 단독연구의 비율은 각각 21.9%, 78.1%로 조사되어 Degree of collaboration은 78.1%로 파악되었다.

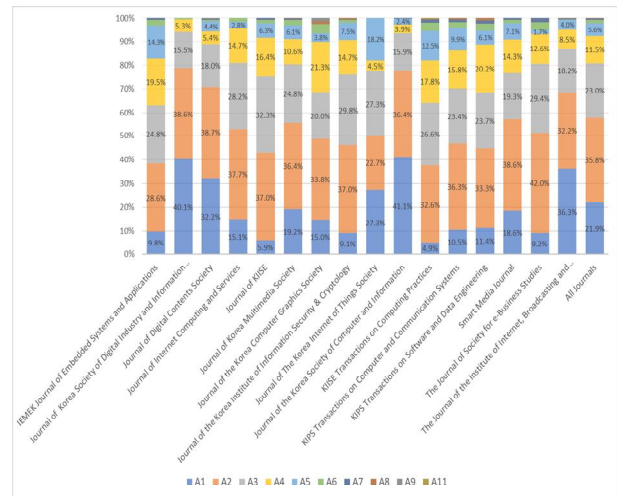


Fig. 4. Distribution of the Team Size

Table 3. Authorship pattern (Degree of collaboration)

Journals	Num. of lead Author (A)	Num. of Co-author (B)	Num. of Authors (A + B)	Average ((A+B) / Num. of Articles)	Num. of Articles with single author (C)	Num. of Articles with multi authors (D)	Degree of collaboration (D / (C+D))
IEMEK Journal of Embedded Systems and Applications	133	279	412	3.1	13	120	90.2%
Journal of Korea Society of Digital Industry and Information Management	207	182	389	1.9	83	124	59.9%
Journal of Digital Contents Society	367	428	795	2.2	118	249	67.8%
Journal of Internet Computing and Services	252	396	648	2.6	38	214	84.9%
Journal of KIISE	495	929	1,424	2.9	29	466	94.1%
Journal of Korea Multimedia Society	593	933	1,526	2.6	114	479	80.8%
Journal of the Korea Computer Graphics Society	80	152	232	2.9	12	68	85.0%
Journal of the Korea Institute of Information Security & Cryptology	416	754	1,170	2.8	38	378	90.9%
Journal of The Korea Internet of Things Society	44	72	116	2.6	12	32	72.7%
Journal of the Korea Society of Computer and Information	711	648	1,359	1.9	292	419	58.9%
KIISE Transactions on Computing Practices	304	671	975	3.2	15	289	95.1%
KIPS Transactions on Computer and Communication Systems	171	330	501	2.9	18	153	89.5%
KIPS Transactions on Software and Data Engineering	228	445	673	3.0	26	202	88.6%
Smart Media Journal	140	224	364	2.6	26	114	81.4%
The Journal of Society for e-Business Studies	119	205	324	2.7	11	108	90.8%
The Journal of the Institute of Internet, Broadcasting and Communication	600	689	1,289	2.1	218	382	63.7%
All Journals	4,860	7,337	12,197	2.5	1,063	3,797	78.1%

저자 소속 기관의 공저 구성을 살펴보면 다음과 같다. 한 편의 논문에서 같은 기관 소속의 저자 수가 복수일 경우에는 중복을 제거하여 기관 수는 1개로 처리하여 분석하였다. 논문별 기관 수 평균은 전체 학술지 기준으로 1.4 기관으로 조사되었으며, 학술지별로 저자 기관 수 구성 패턴과 단독연구와 공동연구의 비율에서 많은 차이가 존재하는 것이 확인되었다. 또한 학술지 전체를 기준으로 Degree of institutional collaboration은 47.2%과 같았다. 실험대상 논문의 저자소속 정보에서 식별된 전체 기관 수는 총 788 기관 (국내 714 : 해외 74, 90.6% : 9.4%)으로 조사되었으며, 저자소속 기관들을 기관유형 분류체계를 적용하여 기관유형별로 분석한 통계 결과는 Table 4에서 확인할 수 있다. 구체적으로 실험 대상 논문 4,860 편에서 식별된 저자의 소속 기관 수는 총 788개이며, 대분류 기준으로 “교육학술기관” (40.0%), “민간기관” (37.6%), “정부기관” (11.3%), “해외기관” (9.4%), “의료기관” (1.8%)의 순으로 조사되었다. 또한 저자 레벨의 출현 횟수 기준으로는 “교육학술기관” (86.0%), “정부기관” (6.8%), “민간기관” (5.9%), “해외기관” (1.1%), “의료기관” (0.2%)의 순이었고, 논문별 중복을 제거한 기관 레벨의 출현 횟수 기준으로는 “교육학술기관” (83.9%), “민간기관” (7.5%), “정부기관” (6.7%), “해외기관” (1.6%), “의료기관” (0.3%)의 순으로 조사되었다. 중분류 기준에서는 “교육학술기관 > 대학”이 총 기관 수 대비 22.5%, 출현 기관 수 대비 76.2%를 차지하였으며, “대학”이 공저기관으로 참여한 논문 수는 4,358 편으로 총 논문 수 대비 89.7%를 차지하였다. “정부기관 > 공공기관”은 총 기관 수 대비 7.9%, 출현 기관 수 대비 6.1%, 공저기관 참여 논문 수 397 편으로 총 논문 수 대비 8.2%를 차지하였다. 또한 유형별로 “민간기관 > 기타 민간기관”, “해외기관 > 해외대학”, “의료기관 > 종합병원”의 비중 유형 기관들의 논문 수 비중이 높았다.

국내 지역별 논문 수 분포를 지도로 표시하면 Fig 5와 같다. 논문 수와 기관 수에서 “서울”, “경기도”, “대전” 등의 비중이 다른 지역에 비해 높은 것으로 조사되었다. 특히, “서울”의 경우 기관 수 (200, 25.4%), 중복허용 출현 저자 수 (4,614, 37.8%), 개별 논문 중복 제거 기준 출현 기관 수 (2,214, 33.8%), 논문 수(1,998, 41.1%)를 차지함으로써 한국의 컴퓨터 분야 연구에서 가장 핵심적인 지역으로 확인되었다.

Table 4. Composition by institution type

Institution Type Section Level	Institution Type Division Level	Num. of Inst. (Duplication)	Ratio	Num. of Inst. (Duplication)	Ratio
(E) Educational institutions	(E02) University	177	22.5%	4,998	76.2%
(E) Educational institutions	(E03) College	77	9.8%	341	5.2%
(E) Educational institutions	(E04) Graduate university	6	0.8%	55	0.8%
(E) Educational institutions	(E05) Other University	4	0.5%	17	0.3%
(E) Educational institutions	(E01) Kindergarten, elementary, secondary education institution	15	1.9%	23	0.4%
(E) Educational institutions	(E06) Academic association	32	4.1%	61	0.9%
(E) Educational institutions	(E99) Other Educational institutions	4	0.5%	6	0.1%
Sub Total		315	40.0%	5,501	83.9%
(C) Private Institutions	(C99) Other Private Institution	157	19.9%	245	3.7%
(C) Private Institutions	(C01) Major Company	27	3.4%	98	1.5%
(C) Private Institutions	N/A	85	10.8%	99	1.5%
(C) Private Institutions	(C02) Medium enterprises	17	2.2%	41	0.6%
(C) Private Institutions	(C03) Small Enterprise	10	1.3%	12	0.2%
Sub Total		296	37.6%	495	7.5%
(G) Governmental agencies	(G04) Public institutions	62	7.9%	402	6.1%
(G) Governmental agencies	(G03) Local government agencies	9	1.1%	13	0.2%
(G) Governmental agencies	(G01) Administrative agency	13	1.6%	17	0.3%
(G) Governmental agencies	(G99) Other Government institution	2	0.3%	2	0.0%
(G) Governmental agencies	(G02) Special local administrative agency	3	0.4%	3	0.0%
Sub Total		89	11.3%	437	6.7%
(F) Foreign institutions	(F03) Foreign university	43	5.5%	65	1.0%
(F) Foreign institutions	(F01) Foreign company	8	1.0%	14	0.2%
(F) Foreign institutions	(F99) Other Foreign institutions	4	0.5%	4	0.1%
(F) Foreign institutions	(F05) Other Foreign Institution	2	0.3%	3	0.0%
(F) Foreign institutions	(F02) Other Foreign institutions	12	1.5%	13	0.2%
(F) Foreign institutions	(F04) Foreign research institution	5	0.6%	5	0.1%
Sub Total		74	9.4%	104	1.6%
(H) Medical institutions	(H02) Non General hospital	3	0.4%	3	0.0%
(H) Medical institutions	(H01) General hospital	11	1.4%	17	0.3%
Sub Total		14	1.8%	20	0.3%
Total		788	100%	6,557	100%

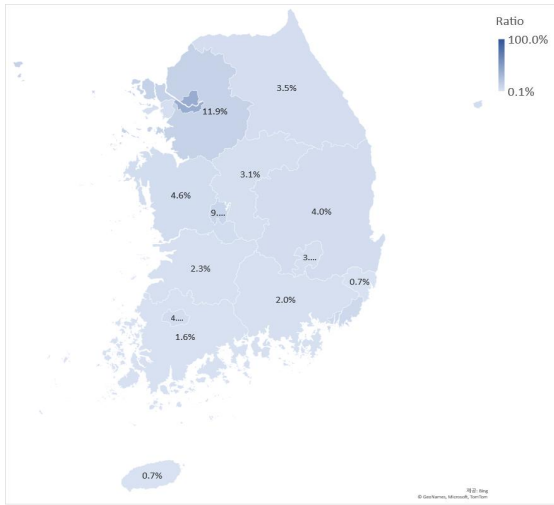


Fig. 5. Distribution of research publication by region

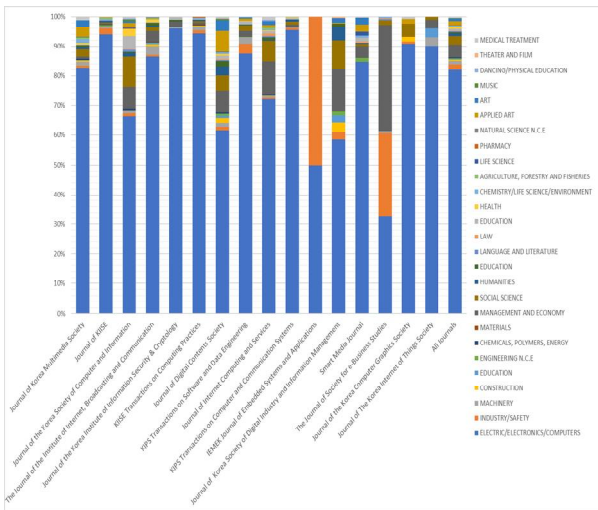


Fig. 6. Distribution of Papers by Department Type

학과유형 분류 기준으로 공저 구성을 살펴보면 Fig 6과 같으며 학술지 별로 분류 비율이 매우 상이한 것으로 조사되었다. 특히, 학과정보 미확인 비율이 높은 2 종의 학술지들을 제외하고 “Journal of Korea Society of Digital Industry and Information Management”, “Journal of Digital Contents Society”, “Journal of the Korea Society of Computer and Information”과 같은 학술지들은 “전기·전자·컴퓨터” 외 학과유형의 비중이 30% 이상을 차지하는 것으로 파악되었다. 학과유형 소속 저자 수 기준, 대분류 순위별로 살펴보면, “공학계열” 85.35%, “인문사회계열” 8.8%, “자연과학계열” 3.1%, “예체능계열” 2.4%, “의학계열” 0.4%로 파악되었으며, 중분류 기준으로 상위 5위까지 살펴보면, “전기·전자·컴퓨터” (공학계열) 81.1%, “경영·경제” (인문사회계열) 4.3%, “사회과학” (인문사회계열) 2.9%, “산업안전” (공학계열) 1.9%, “응용예술” (예체능계열) 1.4%로 분석되었다.

마지막으로 실험데이터를 기반으로 4가지 (기관유형, 기관, 국내 지역, 학과유형) 측면의 공저 관계에 대한 사회네트워크 분석 결과를 제시하면 아래 Fig 7부터 Fig 10까지와 같다. 각 Fig에서 분석 단위별로 주요 네트워크 중심성 (연결, 매개, 근접) 측정 결과들과 네트워크 시각화 지도들을 확인할 수 있다. 이들 공저 네트워크의 주요 속성 정보 (직경, 밀집도, Degree 평균, 가중합 Degree)를 분석한 결과는 기관유형 (8, 0.346, 3.76, 556.7), 기관 (8, 0.007, 5.67, 21.89), 국내 지역 (8, 0.830, 13.29, 823.52), 학과유형 (8, 0.307, 7.69, 373.46)과 같았으며, 기관 네트워크의 경우 다른 기관과 공저 관계가 없는 기관 수가 73개로 파악되었다.

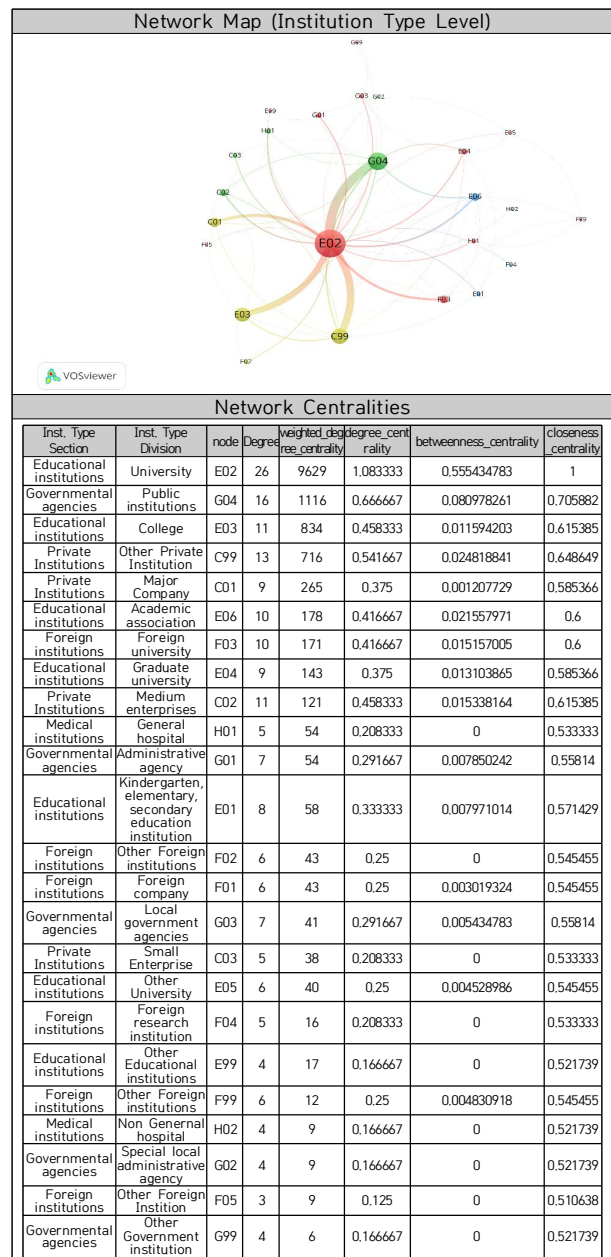


Fig. 7. Network centrality and visualization analysis results (Institution Type Level)

기관유형 레벨의 공저 네트워크에 대한 시각화 맵과 중심성 값들을 Fig 7과 같이 살펴보면 “교육기관 > 대학” 모든 중심성 척도에서 가장 높은 순위를 차지하며, 다른 대분류 하위 기관유형에서 “정부기관 > 공공기관”, “민간기관 > 기타 민간기관”, “해외기관 > 해외대학”들과 상대적으로 높은 공저 관계를 형성하고 있음을 확인할 수 있다.

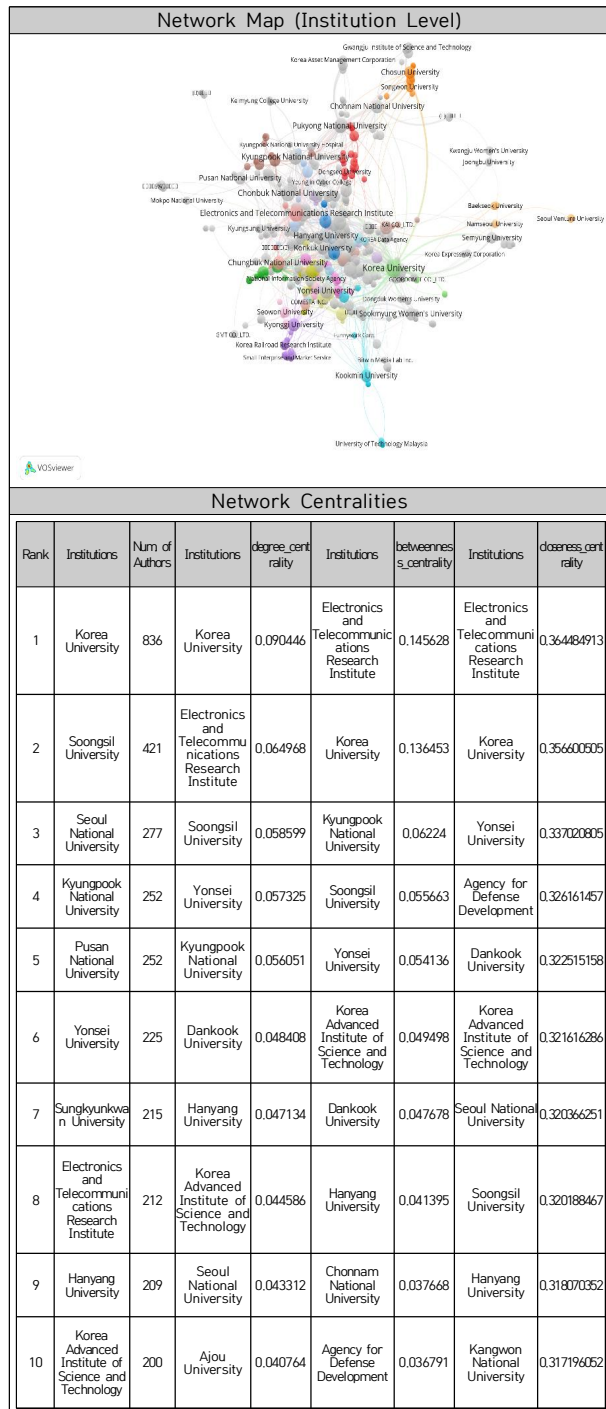


Fig. 8. Network centrality and visualization analysis results (Institution Level)

Fig 8에서 제시된 기관 레벨의 공저 네트워크 분석 결과에서는 국내 대학 기관들의 높은 중심성과 다른 기관들과의 연결 관계를 살펴볼 수 있으며, 중계자 역할 정도를 측정하는 매개 중심성과 각 노드 간의 거리를 기반으로 측정하는 근접 중심성 척도에서 일부 정부출연 연구기관들의 순위가 연결 중심성 기준의 순위보다 높은 것으로 조사되었다.

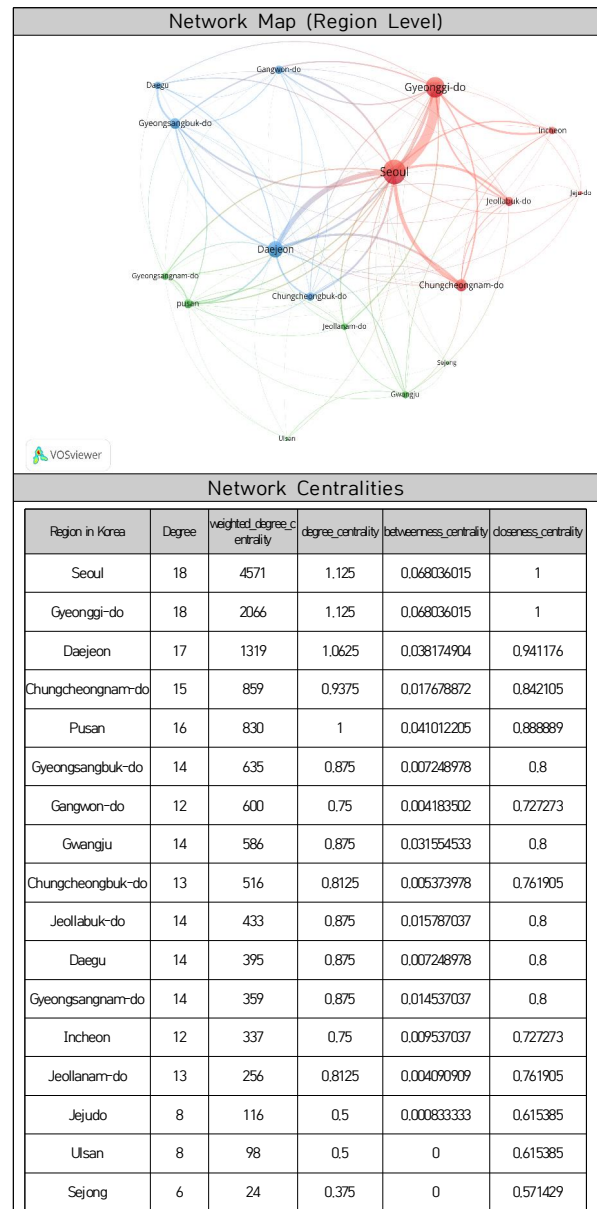


Fig. 9. Network centrality and visualization analysis results (Region Level)

국내 지역 간의 공저 네트워크 분석결과를 나타내는 Fig 9의 시각화 맵을 통해 “서울”과 “경기도” 지역의 강한 연결 관계를 보여주고 있으며, 이들 수도권 지역을 중심으로 그 밖의 다른 지역들과 다양한 공저 관계가 맺어

지고 있음을 파악할 수 있다. 이렇게 네트워크 시각화 지도 상에서 강한 공저 관계를 나타내는 지역들은 네트워크 중심성 분석 결과에서도 높은 순위를 차지하고 있음을 Fig 9를 통해 확인할 수 있다.

리·천문·지구”, “의료” 등 공학 분야 및 다른 학문 분야의 학과 유형들과 다양한 연결 관계들을 통해 한국의 컴퓨터 분야 연구자들이 타 분야 연구자들과 활발한 협력 연구를 수행하고 있음을 보여준다.

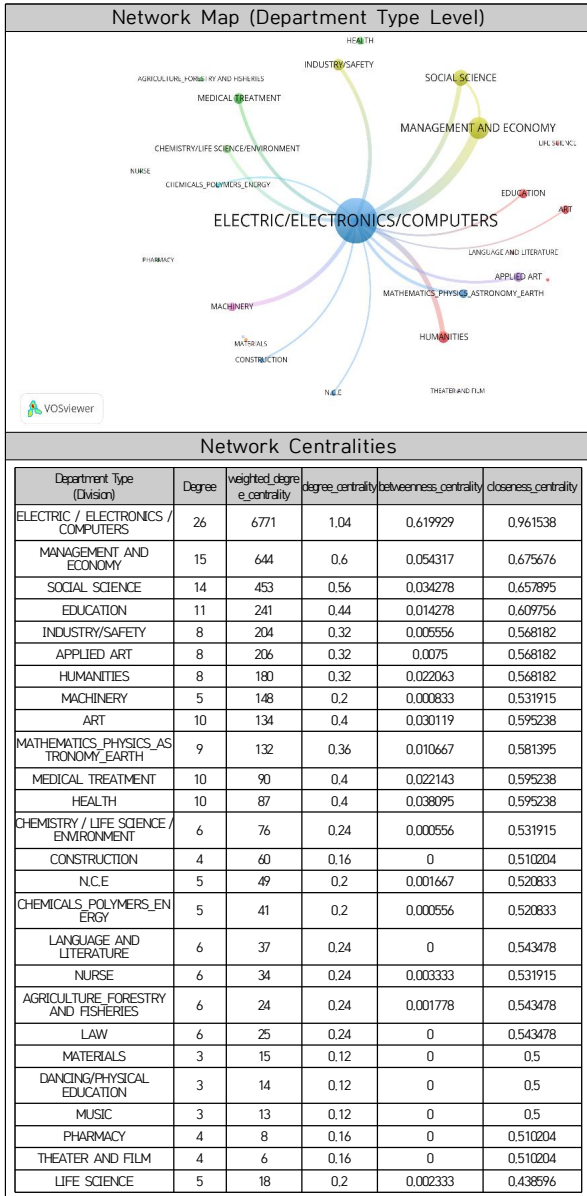


Fig. 10. Network centrality and visualization analysis results (Department Type Level)

마지막으로 학과유형 레벨의 공저 네트워크 지도와 중심성 분석 결과를 Fig 10에서 살펴보면, 컴퓨터 분야 학술지 저자 소속정보 기반 분석으로 인해 중분류 기준 “공학계열 > 전기·전자·컴퓨터”가 매우 높은 네트워크 중심성 값과 순위를 차지하는 것을 확인할 수 있다. 또한 네트워크 시각화 지도는 “전기·전자·컴퓨터” 학과유형은 “경영·경제”, “사회과학”, “교육”, “산업·안전”, “수학·물

V. Conclusions

본 연구에서는 한국의 컴퓨터 분야의 공저 관계의 현황 분석과 특성을 파악할 목적으로 해당 분야를 대표할 수 있는 한국 출판 학술지 16종의 연구논문 4,860 편을 대상으로 실험데이터를 제작하고 계량정보분석 및 소셜 네트워크분석 연구를 수행하였으며, 특별히 실험데이터로 KSCD, 기관식별데이터, 학과유형 분류체계와 자동 분류를 위한 학과유형 데이터셋을 구축하여 활용하였다. 또한 향후 지속적인 분석 연구 및 동향 파악을 위해 학과유형 정보의 자동 분류 모델을 구현하고 평가하였다.

주요 연구 결과를 요약하면 다음과 같다. 첫째, 학과정보 분석 활용을 위해 데이터의 수집부터 학과정보의 추출 및 데이터 정제 등의 전처리 작업, 학과유형에 대한 정확한 분류 처리를 통해 데이터셋을 구축하고 데이터셋을 학습 및 테스트 데이터로 사용하여 딥러닝 기반의 자동 분류 모델을 구현하였다. 모델 성능 평가 결과는 한글 학과정보 기준 98.6%와 영문 학과정보 기준 97.6%의 정확률로 측정되었다.

둘째, 실험데이터를 기반으로 공저 구성, 공저 패턴, 기관유형, 국내 지역, 기관, 학과유형의 4가지 측면으로 네트워크 분석을 수행하였다. 구체적으로 논문별 저자 수와 기관 수 평균은 전체 학술지 기준으로 2.5 명, 1.4 기관으로 파악되었으며, 학술지별 저자 수, 기관 수 구성 패턴과 단독연구와 공동연구의 비율은 상당한 차이가 존재하는 것이 확인되었다. 학술지 전체를 기준으로 협업 정도를 나타내는 Degree of collaboration 지표는 저자 기준 78.1%, 기관 기준 Degree of institutional collaboration은 47.2%로 파악되었다. 또한 컴퓨터 분야 연구자들의 공저 패턴과 기관유형, 지역, 기관, 학과유형 측면별 공저 네트워크의 속성과 중심성이 자세히 파악되고 네트워크 지도로 시각화되었다.

본 연구에서는 데이터셋 및 자동 분류 모델의 부재로 기존 학문 분야의 지적관계 분석 연구에서 다루지 못한 학과유형 분류를 위한 기반을 마련하였다. 또한 연구자의 기관 단위부터 현재 전공을 나타내는 세부 소속정보인 학과유형 단위까지 분석 범위를 확장하고 실험적으로 컴퓨

터 분야 연구자들의 네트워크 시각화 및 중심성 분석에 적용하였으며, 학과유형 레벨의 공동연구 현황과 패턴을 확인하였다. 이로써 학문 분야의 지적관계 분석을 비롯한 계량정보분석 연구에서 학과유형의 식별과 활용 가능성을 확인하였다. 다만 실험데이터와 학과유형 데이터셋의 구축 및 분석 범위를 국내 학술지로 한정된 부분은 본 연구의 한계점으로 향후 해외 학술지 및 학술대회 문헌까지 확대하고 아울러 최신 발간 정보를 실시간적으로 적용할 수 있는 연구 수행 체계의 개선이 필요하다.

본 연구에서 제시하는 공저 관계에 대한 다양한 측면에서의 통계 분석 결과와 네트워크 분석 결과의 상호 비교 및 분석 과정을 통해 국내 컴퓨터 분야 학술 연구에서 공저 관계를 폭넓고 조망할 수 있으며, 향후 자동 분류 모델 및 데이터셋을 기반으로 다양한 연구 분야에서 후속 연구가 기대된다.

ACKNOWLEDGEMENT

This study was conducted with the support of the Korea Institute of Science and Information Technology (KISTI).

REFERENCES

- [1] KOTRA, World's Best Digital Infrastructure (Why Korea in Invest Korea Site), <https://www.investkorea.org/ik-kr/cntnts/i-103/web.do>
- [2] Center for Government Competitiveness in Seoul National University, Government Competitiveness 2021, <https://www.joonggang.co.kr/article/25030839>
- [3] NRF, '2010-2020 Results of Analysis of Top 1% Cited Papers in Major Countries, <https://www.unipress.co.kr/news/articleView.html?idxno=6869>
- [4] B. Kim, B.-J. You, and H.-S. Shim, "Dataset construction and Automatic classification of Department information appearing in Domestic journals" In Proceedings of the Korean Society of Computer Information Conference, Vol 31. No. 1, pp. 343-344, Daejeon, South Korea, Jan. 2023.
- [5] B. Kim, Y. Kim, and J. Kang, "Analysis of the citation impact of national journals toward SCIE journals on JCR ranking," Malaysian Journal of Library and Information Science, Vol. 23, No. 2, pp. 1-24, Jul. 2018, DOI: 10.22452/mjlis.vol23no2.1.
- [6] B. Kim, B.-J. You, and J.-H. Kang, "Analysis of SCI Journals Cited by Korean Journals in the Computer field," Journal of the Korea Society of Computer and Information, Vol. 24, No. 11, pp. 79-86, Nov. 2019. DOI: 10.9708/JKSCI.2019.24.11.079
- [7] B. Kim, B.-J. You, and J.-H. Kang, "Examination of International Collaborative Research Status and Co-Authorship Network Visualization of Korean Researchers in the Computer Field." TEST Engineering & Management, vol. 81, pp. 173-182, Nov. 2019.
- [8] B. Kim, and B.-J. You, "Co-author Network Analysis of Research Institutes in the Computer Field Based on the Journal Articles Published in Korea" In Proceedings of the Korean Society of Computer Information Conference, Vol 29. No. 2, pp. 117-122, South Korea, July. 2021.
- [9] B. Kim, B.-J. You, and H.-S. Shim, "Trends in disaster safety research in Korea: Focusing on the journal papers of the departments related to disaster prevention and safety engineering," Journal of the Korea Society of Computer and Information, Vol. 27, No. 10, pp. 43-57, Oct. 2022. DOI: 10.9708/JKSCI.2022.27.10.043
- [10] S. Lee, "A Preliminary Study on the Co-author Network Analysis of Korean Library and Information Science Research Community," Journal of Korean Library and Information Science Society, Vol. 41, No. 2, pp. 297-315, Jun. 2010, DOI: 10.16981/kliss.41.2.201006.297.
- [11] B.-U. Yun, T.-S. Kim, J.-M. Park, E.-G. Park, J.-H. Kim, and D.-S. Han, "A Study on Co-author Networks of the Architectural Planning using Social Network Analysis," Journal of the architectural institute of Korea planning and design, Vol. 33, No. 2, pp. 61-68, Feb. 2017, DOI: 10.5659/jaik_pd.2017.33.2.61.
- [12] J.-Y. Lee and Y.-H. Park, "Social Network Analysis of author's interest area in Journals about Computer," Journal of the Korea Institute of Information and Communication Engineering, Vol. 20, No. 1, pp. 193-199, Jan. 2016. DOI: 10.6109/JKIICE.2016.20.1.193
- [13] Jang and Y. Park, "A Study on Co-author Networks in the Journal of a Branch of Computers," Journal of the Korea Institute of Information and Communication Engineering, Vol. 22, No. 2, pp. 295-301, Feb. 2018. DOI: 10.6109/JKIICE.2018.22.2.295
- [14] T. Ahmed, A. Ahmed, M. Ali, and M. Kamran, "Analysis of co-authorship in computer networks using centrality measures," 2017 International Conference on Communication, Computing and Digital Systems (C-CODE), Mar. 2017, DOI: 10.1109/c-code.2017.7918901.
- [15] R. Molontay and M. Nagy, "Two decades of network science," Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Aug. 2019, DOI: 10.1145/3341161.3343685.
- [16] A. Nunes da Silva, M. M. Breve, J. P. Mena-Chalco, and F. M. Lopes, "Analysis of co-authorship networks among Brazilian graduate programs in computer science," PLOS ONE, Vol. 17,

No. 1, p. e0261200, Jan. 2022, DOI: 10.1371/journal.pone.0261200.

- [17] H. Choi, B. Kim, Y. Jung, and S. Choi, "Korean scholarly information analysis based on Korea Science Citation Database (KSCD)," *Collnet Journal of Scientometrics and Information Management*, Vol. 7, No. 1, pp. 1-33, Jun. 2013. DOI: 10.16981/kliss.51.2.202006.273.
- [18] "KISTI, Institution Identification Data Version 1.0." 2023. DOI:10.23057/39.
- [19] "Korean Council for University Education, The Standard Classification of University Education Units." 2023. <https://www.data.go.kr/data/15014632/fileData.do>.
- [20] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, Vol. 1, No. 3, pp. 215-239, Jan. 1978, DOI: 10.1016/0378-8733(78)90021-7.
- [21] Hagberg, Aric, and Drew. Conway, 2020. "Networkx: Network analysis with python.", URL: <https://networkx.github.io>
- [22] N. J. van Eck, and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, Vol. 84, No. 2, pp. 523-538, December 2009. DOI: 10.1007/s11192-009-0146-3
- [23] K. Subramanyam, "Bibliometric studies of research collaboration: A review," *Journal of Information Science*, Vol. 6, No. 1, pp. 33-38, Jan. 1983, DOI: 10.1177/016555158300600105.

Authors



Byungkyu Kim received the B.S., M.S. in Computer Science from Chungnam National University in 2001, 2003 and the Ph.D. in Computer Science & Engineering from Chungnam National University, Korea in 2018.

Dr. Kim joined Korea Institute of Science and Technology Information, Korea in 2002. He is currently a Principal Researcher in Science and Technology Security R&D Center, KISTI. He is interested in Scientometrics, Disaster & Safety Management, Cybersecurity.



Beom-Jong You received the B.S. in Electronic Engineering in Sogang University, Korea in 1984 and M.S., Ph.D. degrees in Library & Information Science from Chungnam National University, Korea in

2000 and 2005, respectively. Dr. You joined System Engineering Research Institute, Korea in 1987. He is currently a Principal Researcher in Science and Technology Security R&D Center, KISTI. He is interested in Data Science, Disaster & Safety Platform, Content Platform, Cybersecurity.



Min-Woo Park received the B.S. and M.S. degrees from the Division of Computer Convergence, Chungnam National University, South Korea, in 1992 and 2004, respectively.

Since 1996, he has been with the Korea Institute of Science and Technology Information (KISTI), South Korea. He is currently a Principal Researcher in Science and Technology Security R&D Center. His research interests include system architecture, information security, the Internet of Things, smart city, and artificial intelligence.