

연구지원 정보서비스를 위한 히스토리오그래프와 SPLC 활용에 관한 실험적 연구: LED 분야 사례를 중심으로

Exploratory Study of Applying Historiography and SPLC for Developing Information Services: A Case Study of LED Domain

유소영 (So-Young Yu)*

초 록

이 연구에서는 특정 주제 분야의 핵심적이고 전역적인 연구 동향을 제공하는 연구지원 정보서비스 개발을 위해 SPLC(Search Path Link Count) 분석을 적용할 때, 데이터의 범위와 인용빈도 설정에 대하여 탐험적으로 살펴보고자 하였다. 이를 위하여 Web of Science에서 검색된 RGB LED 분야의 2,318개 논문과 20,109개 상위 인용논문으로 5개의 데이터셋을 구성하였다. 각 데이터셋에서 히스토리오그래프와 SPLC 네트워크를 인용빈도 임계치를 변화시키면서 28개 주요 연구 동향 네트워크를 추출하여, 인용문헌의 포함여부와 인용빈도 임계치 설정이 SPLC 네트워크에 미치는 영향을 살펴보았다. 그리고 특정 기관 소속 연구자들에게 SPLC 네트워크에 포함된 198개 주요 논문 리스트를 제공하고 피드백을 받음으로써, 전역적 연구 동향이 개인 연구자의 정보 요구에 부합하는지 살펴보았다. 분석 결과, 분석 대상에 상위 인용문헌 포함 여부와 인용빈도임계치에 따라 추출되는 SPLC 네트워크가 변화되었으나, 일정 인용빈도임계치값에서는 수렴하였다. 그리고 개인 연구자의 정보 요구는 SPLC를 통해 제공된 전역적 연구 동향과 출판년도의 차이는 있지만 대체적으로 일치하는 것으로 나타나, 인용문헌을 포함하여 인용빈도임계치를 변화시키는 SPLC 분석을 통해 개인 이용자가 원하는 전역적 연구 정보를 제공해 줄 수 있는 것으로 해석된다. 이를 일반화하기 위해서는 이 탐색적 연구에서 제안된 방법을 다양한 분야에 적용하는 후속 연구가 필요할 것이다.

ABSTRACT

The purpose of this study is to examine the data coverage and citation threshold for analyzing SPLC(Search Path Link Count) as a main path of a historiograph of a certain topic in order to provide 'core' papers of global research trends to a researcher affiliated with a local R&D institution. 5 datasets were constructed by retrieving and collecting 2,318 articles on RGB LED on Web of Science published from 1990-2013 and 20,109 articles which cited these original 2,318. The SPLC analysis was performed on each dataset by increasing the threshold of citation counts, and the changes and resilience of the 28 extracted networks were compared. The results of user feedback on 198 unique core papers from 28 SPLC networks received from LED researchers affiliated with a Korean government-sponsored research institution were also analyzed. As a result, it is found that the nodes in each SPLC network in each dataset were differentiated by the citation counts, while the changes in the structure of SPLC networks were slight after the networks' citation counts were set at 40. Additionally, the user feedback showed that personalized research interest generally matched to the global research trends identified by the SPLC analysis.

키워드: 히스토리오그래프, 인용 분석, LED, SPLC(Search Path Link Count), 연구지원 서비스
historiograph, citation analysis, LED, SPLC, information service

* 한남대학교 문헌정보학과 조교수(soyoungyu201@gmail.com)

■ 논문접수일자: 2013년 9월 5일 ■ 최초심사일자: 2013년 9월 12일 ■ 게재확정일자: 2013년 9월 13일
■ 정보관리학회지, 30(3), 273-296, 2013. [http://dx.doi.org/10.3743/KOSIM.2013.30.3.273]

1. 연구의 배경 및 목적

1.1 연구의 배경

계량분석 정보의 주요 이용자는 크게 연구 지원 정책 관련 의사 결정자와 연구 분야의 개별 연구자라고 할 수 있다. 유사한 맥락에서 이재운(2013)은 도서관 및 정보센터의 타겟 이용자를 크게 정보관리 담당자, 정책/경영 담당자, 연구활동 지원으로 나눔으로써 의사 결정 지원자, 의사 결정자, 연구자로 보았다. 그리고 현장에서 의사결정자 및 연구자, 사서, 출판사 등으로 고객층을 세분화하고 있는 추세이다.

최근 프랑스 및 호주와 같은 유럽 및 오세아니아 국가, 그리고 국내를 중심으로 국가 단위 정보센터를 중심으로 의사결정자와 개인 연구자들에게 연구자 또는 학술지 단위의 인용 지수를 제공하고, 연구 동향 및 학술 활동 관련 정보서비스를 제공하려는 시도가 활발히 이루어지고 있다(이재운, 2013). 과학 기술 연구자들이 자신이 연구하는 특정 분야의 연구 경향을 파악하는 것이 중요하면서도 시간과 노력을 소모하는 일로 인식되고 있다는 점이 여러 선행 연구에서 밝혀진 바 있으므로(권나현, 이정연, 정은경, 2013; 이정연, 정은경, 권나현, 2012; Allard, Levine, & Tenopir, 2009; Bichteler & Ward, 1989; Case, 2012; Case, Borgman, & Meadow, 1986; Holland & Powell, 1995; King & Tenopir, 2001; Kuruppu & Gruber, 2006; Murphy, 2003; Tenopir, King, Spencer, & Wu, 2009), 개인 연구자를 위한 계량 분석 서비스는 최신 연구 동향과 관련된 주요 논문 정보를 전달해주는 서비스로 확장할 수 있을 것

이다.

인용의 특성과 연구자의 정보이용행태에 기반하여, 최신 연구 동향 관련 주요 논문 정보를 제공하는 연구지원 정보 서비스는 다음을 고려해야 할 것으로 보인다. 첫째, 최신의 주요 논문을 포함할 수 있는 접근법을 사용하여 단순 인용빈도에 의한 분석결과를 보완할 필요가 있다. 왜냐하면 인용은 일정 시간이 흐르면 발생하기 때문에, 최신의 논문은 상대적으로 높은 인용빈도를 가지기 어렵기 때문이다. 둘째, 연구 동향 분석의 단위를 학술지가 아닌 개별 연구 문헌으로 함으로써 이용자가 동향 파악 후 관련 주요 논문에 접근할 수 있도록 해야 할 것이다. 셋째, 연구자가 많은 시간을 들이지 않고 살펴볼 수 있는 특정 분야의 비교적 소수의 '핵심' 문헌을 압축적으로 제공할 수 있는 주요 연구 흐름 파악 기법 및 주요 문헌 선정 기준이 필요할 것이다.

1.2 연구의 목적

이 연구에서는 최신 문헌을 포함하여 핵심 연구 흐름에 포함된 주요 문헌을 파악할 수 있는 방법으로 히스토리오그래프(Garfield, 2001, 9:19; 2001.11.27)와 Search Path Link Count(이하 SPLC, Hummon, & Doreian, 1989)를 중심으로 제안하고자 한다. 이 두 기법을 사용하는 이유는 두 기법이 1.1의 세 가지 고려사항을 어느 정도 충족해 줄 수 있으며, 두 기법이 제안된 시간 및 방법론적 연계성이 있기 때문이며, 인용의 흐름을 이용함으로써 최신 문헌을 비교적 많이 포함하여 동적인 핵심 문헌 파악이 가능하기 때문이다.

히스토리오그래프는 특정 데이터셋 안에 포함된 논문 레코드 간의 인용 흐름을 파악해줌으로, 특정 분야의 연구 동향을 통시적으로 파악할 수 있게 해주기 때문에 첫 번째와 두 번째 고려사항을 충족해 준다고 할 수 있다. 그러나 분석 대상 레코드의 수가 많고 이들 간의 인용이 활발하다면, 히스토리오그래프를 통해 파악된 인용 흐름 안에 포함된 논문을 이용자가 제공받아 모두 살펴보는 데 많은 시간이 걸리게 된다. 따라서 이들 중 주요 논문을 선정하는 부가적인 기준이 필요하게 된다. 이는 앞서 언급한 세 번째 고려사항과도 연결되는데, 이를 위해서는 SPLC 기법을 고려할 수 있다. 이는 인용의 흐름 중 주요 연결을 파악하는 기법으로 사회학자인 Hummon과 Doreian(1989)이 제안한 main path 파악 기법 중 하나이다. 이는 실제 E. Garfield가 분석한 DNA 분야 히스토리오그래프 분석 결과를 이용하며 제안된 기법으로, 히스토리오그래프 분석과 연결하여 사용하는 데 적절할 것으로 사려된다.

최근 히스토리오그래프를 사용한 연구들은 특정 인용빈도 이상의 특정 개수의 논문만을 분석 대상에 포함시킴으로써 최신 논문들이 분석에서 배제되는 결과를 보여주고 있다. 따라서 인용빈도에 제한을 두지 않고 히스토리오그래프 분석을 실시함으로써 인용빈도에 의한 분석 범위 설정이 가져오는 결과의 차이를 살펴볼 필요가 있다. SPLC도 분석 대상 논문 및 이들간의 인용 경로에 의해 주요 인용 흐름과 주요 논문을 파악해 주므로, 히스토리오그래프와 마찬가지로 인용빈도에 의한 분석 범위 설정이 가져오는 결과의 차이를 살펴볼 필요가 있다.

따라서, 이 연구에서는 네트워크 링크의 흐름을 이용하는 히스토리오그래프와 SPLC 기법을 함께 활용하여 특정 주제 분야에 대한 전역적 동향을 파악하고 이를 특정 기관의(지역적) 연구자에게 주요 논문 리스트 형태로 제공하는 연구 지원 정보서비스를 제안하고자 한다. 그리고 이 기법들을 적용하는 분석 데이터의 범위와 인용빈도 임계치를 다양하게 변화시키면서 분석함으로써, 최신 연구 동향을 포함하는 정도의 변화와 주요 연구흐름 분석 결과의 차이를 중심으로 살펴보고자 한다. 이를 위하여 주요 유망 기술연구 분야 중 하나인 LED 분야 논문을 이용하여 분석을 수행하고 특정 기관 소속 연구자들로부터 받은 피드백을 사례로 활용하여, 두 기법을 함께 적용하는 적절한 데이터 분석 범위와 인용빈도를 살펴보고자 하였다.

구체적인 연구 문제는 다음과 같다.

- RQ1: 히스토리오그래프에 SPLC 분석 기법을 적용할 때, 분석 데이터의 범위에 인용 관계에 의해 확장된 문헌을 포함하는지 여부에 따라 추출되는 핵심 문헌의 수뿐만 아니라 핵심 문헌 네트워크의 구조가 다를 것이다.
- RQ2: 주요 연구흐름을 파악할 때 피인용빈도가 높은 논문만을 대상으로 할 경우, 최근 n년 이내의 출판 논문이 가장 높은 감소 비율을 보이고, 출판년도가 빠른 논문만이 최종적으로 분석에 남을 것이다.
- RQ3: SPLC 분석 결과를 전역적인 연구 동향으로 볼 때, 특정 기관 소속의 개별 연구자의 연구 관심 분야와는 최신 연구 결과로 올수록 차이가 있을 것이다.

2. 이론적 배경: 히스토리오그래프와 SPLC

히스토리오그래프(Historiography)는 특정 분야 발전에 대한 역사를 살펴보는 방법론으로, 계량정보학의 분야에서는 Eugene Garfield 가 이를 이용하여 특정 분야가 어떻게 시작되고 이를 연결하는 선행 및 후행 연구들과 연구자를 살펴보는 “시각적 요약(graphical summary)”로써 사용될 수 있다고 제안한 것을 시작으로, 점차 알고리즘적인 접근방법으로써 자리매김을 하고 있다. 계량서지학적인 데이터의 히스토리오그래프 분석에는 문헌 간 인용 관계가 사용되며, Global Citation Score(이하 GCS)는 데이터 인용 색인(예를 들면 Web of Science) 내의 인용빈도를, Local Citation Score(이하 LCS)는 수집된 데이터셋 내에서 발생한 인용빈도를 나타내는 용어로 제안되었다(Garfield, 2001.9.19; 2001.11.27; Garfield, Poduvkin, & Isotomin, 2002/2003). 이를 분석하는 도구로는 HistCite™가 있으며 최근에는 이를 이용하여 WoS 데이터뿐만 아니라 SCOPUS 데이터를 이용한 분석도 가능해짐으로써, 분석 데이터의 범위가 넓어졌다고 할 수 있다.

인용의 흐름과 경로(path)를 고려하여 주요 연결 및 흐름만을 파악하는 방법으로는 Search Path Link Count(SPLC)가 있다. SPLC는 사회학자 Hummon과 Doreian(1989)에 의해 고안되었으며, 이들은 Eugene Garfield의 DNA 분야 히스토리오그래프 분석 결과를 이용하여 이를 제안하고 SPLC를 포함한 다양한 main path search 알고리즘 방법을 개념적으로 제안하였다. SPLC은 방향성 있는 시간 순서의 그래

프 안에서 source node와 target node 간의 flow를 고려하여 주요 연결 흐름을 찾는 알고리즘이다. 이는 두 노드 간의 유일한 최단경로거리(geodesic distance) 개념을 확장한 것으로 볼 수 있는데, 최단경로거리는 특정한 source node와 target node를 상정하는 것과 달리, SPLC는 한 네트워크 내의 모든 노드를 source 노드 및 target node로 상정하여 가능한 모든 경로를 찾아준다. 이후 de Nooy, Mrvar, Batagelj (2011)는 이를 수학적으로 증명하고, 프로그래밍 언어로 구현 가능한 컴퓨터공학적 알고리즘으로 발전시켰다(Doreian, Batagelj, & Ferligoj, 2000; Batagelj, 2003; de Nooy, Mrvar, & Batagelj, 2011).

인용 관계를 파악하기 쉬운 학술 논문 인용 색인 데이터를 이용하여 히스토리오그래프 및 Main Path 분석을 수행함으로써, 특정 주제 분야의 발생 및 주요 연구흐름을 살펴보고 이를 통해 과학기술의 혁신을 추적하고 주요 연구를 시대별로 정리하는 연구들이 최근 들어 활발히 이루어지고 있다. 특히 시간적 흐름에 따른 과학기술의 변화를 파악하는 것이 유용한 특허 분석, 혁신적 아이디어의 창발과 관련된 연구들에서 키워드 분석과 조합하여 활발히 이용되고 있다(Choi & Park, 2009; Colicchia & Strozzi, 2012; Kajikawa, Yoshikawa, Takeda, & Matsushima, 2008; Leydesdorff, 2010; Lucio-Arias & Leydesdorff, 2008; Verspagen, 2007). 이와 같은 기존 연구들은 주로 특정한 인용빈도나 이에 의한 상위 n개(예를 들면, 30)의 논문으로 분석 범위를 한정하여 최대한 핵심 구조만을 파악하고자 하였다.

계량정보학적 분석 방법을 연구지원 정보서

비스에 활용하고자 하는 시도는 현장을 중심으로 이루어지고 있으며, 주로 특정 기관 소속 연구자가 발표한 논문을 분석함으로써 정보서비스 개발의 기초 자료를 작성하는 방법 및 사례를 제시하거나, 연구 결과를 정보서비스의 기초 자료로써 활용할 수 있는 연구들을 중심으로 이루어지고 있다(유소영, 이재운, 2008; 이재운, 김관준, 강대신, 김희정, 유소영, 이우형, 2011; 이재운, 김희정, 2011).

그러나 연구지원을 위한 정보서비스가 최신의 정보를 제공해야 한다고 가정할 때, 다음과 같은 측면에서 기존의 히스토리오그래프 관련 선행 연구에서 사용한 접근 방법을 그대로 적용하기에는 무리가 있을 것이다. 첫째, 선행 연구들은 특정 주제 관련 연구의 인용에 의한 핵심 구조를 찾는데 중점을 두었으므로 아직 인용을 받지 못한 최신 논문들은 그 구조 내에서 제외되기 쉽다. 둘째, 주로 데이터 집합 내의 인용 빈도를 기준으로 하기 때문에 연구지원 정보서비스에서는 데이터 집합이 대표성을 검증하지 않은 상태에서 이를 그대로 적용하기 어렵다. 따라서 이 연구에서는 최신 논문 포함 여부와 분석 결과의 유용성에 초점을 두어 연구지원 서비스에서 고려해야 할 데이터의 분석 범위를 실험적으로 살펴보고자 한다. 또한 이용자 피

드백을 통해 이 연구에서 제안하는 방법으로 추출한 분석 결과의 제공 가능성을 실험적으로 살펴보고자 한다.

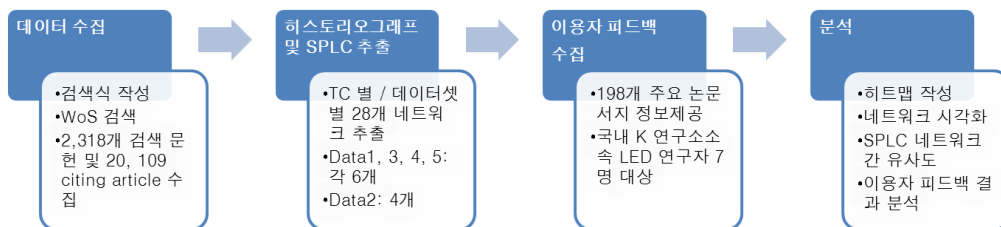
3. 연구 방법

3.1 연구 개요

이 연구의 연구 개요는 <그림 1>과 같으며, 데이터 수집 및 네트워크 추출, 이용자 피드백 수집과 분석의 순서로 이루어졌다.

3.2 데이터 수집

이 연구에서는 최신 동향 파악을 위한 연구지원 정보서비스의 요구가 있을 것으로 예상되는 여러 분야 중, 현재 정부 및 산업계의 지원을 많이 받는 LED 분야를 선정하였다. 이 분야는 또한 국가간 연구 협력 패턴이 특징적으로 나타나(이재운 등, 2011), 개인 연구자가 소속된 기관에 따라 구체적인 정보요구의 특성이 다를 것으로 예상되므로 이 연구의 연구 가설을 검증하는 데 적절한 분야로 생각되어 선정하였다.



<그림 1> 연구 개요

이 연구에서 수집하여 생성한 데이터셋은 총 5개이며, 연구가설에서 밝힌 바와 같이 인용문헌 및 TC 임계치에 의한 히스토리오그래프와 SPLC의 변화를 살펴볼 수 있도록 구성되었다. 데이터 수집은 LED 분야 전문가가 생성한 검색식¹⁾을 이용하여 Web of Science(이하 WoS)에서 검색하고, 검색된 문헌을 인용하는 문헌

을 수집하여 확장하는 순서로 각각 이루어졌다. 각 데이터셋의 구성은 <표 1>과 같다.

검색 문헌 및 인용문헌의 수집은 2013.4.25. ~ 2013.5.4. 사이에 이루어졌으며, Data1의 출판년도별 논문 수 및 Data2에 선정된 논문 수는 <표 2>와 같다.

<표 1> 각 데이터셋별 특성 및 생성 방법

데이터셋	데이터 수	생성 방법
Data1	2,318	검색식에 의해 검색된 WoS 레코드
Data2	233	Data1의 동일 출판년도별 수집 당시 현재 인용빈도 기준 상위 10%에 해당하는 WoS 레코드
Data3	20,109	data2의 레코드를 인용하고 있는 WoS 레코드
Data4	20,162	Data2 + Data3
Data5	20,949	Data1 + Data3

<표 2> 검색된 논문의 연도별 인용빈도 기준 상위 10% 논문 수

연도	논문수	논문비	상위 10%	상위 10% (누적)	연도	논문수	논문비	상위 10%	상위 10% (누적)
1991	10	0.43	1	1	2003	95	4.10	10	69
1992	20	0.86	2	3	2004	100	4.31	10	79
1993	22	0.95	2	5	2005	100	4.31	10	89
1994	24	1.04	2	7	2006	136	5.87	14	103
1995	29	1.25	3	10	2007	158	6.82	16	119
1996	41	1.77	4	14	2008	188	8.11	19	138
1997	90	3.88	9	23	2009	191	8.24	19	157
1998	58	2.50	6	29	2010	243	10.48	24	181
1999	59	2.55	6	35	2011	239	10.31	24	205
2000	88	3.80	9	44	2012	230	9.92	23	228
2001	66	2.85	7	51	2013	48	2.07	5	233
2002	82	3.54	8	59	Total	2,318	100	233	233

1) TS = (((light same emitting* same (diode* or device*)) and (("wafer bonding") or ("indium segregation") or InGaAlP or GaInAlP or AlInGaP or AlGaInP or TIP or (truncated same inverted same pyramid) or (chip and shap*) or InGaN* or (Indium same Gallium same Nitride) or epi* or (Heterojunction same Clad*) or Nanorod* or Nano-rod*)) and (red or green or blue or RGB))) (이재훈 등 2011).

〈표 3〉 히스토리오그래프와 SPLC 네트워크의 크기

		Data1	Data3	Data4	Data5	Data2			
레코드 수(A)		2,318	20,109	20,162	20,949	233			
히스토리오그래프(C)	TC	0	1,943	19,943	19,943	20,598	190	0	(C)
		10	145	2,282	2,282	2,376	31	5	
		20	58	1,123	1,123	1,185	12	10	
		30	32	689	689	730	6	15	
		40	20	478	478	501	-	-	
		50	15	374	374	383	-	-	
SPLC(D)	TC	0	23	43	43	44	16	0	(D)
		10	19	37	37	38	13	5	
		20	10	32	32	35	5	10	
		30	10	31	31	34	4	15	
		40	10	28	28	28	-	-	
		50	8	28	28	28	-	-	
(C_0)/(A)		84.0%	99.2%	99.0%	98.3%	81.5%			
(D_0)/(A)		1.0%	0.2%	0.2%	0.2%	7.0%			

3.3 히스토리오그래프 및 SPLC 네트워크 추출

다섯 개의 상호인용 네트워크에서 28개의 히스토리오그래프를 각각 생성한 후 SPLC 네트워크를 추출하였다. 이 때, 인용빈도임계치(이하 TC)를 변화시킴으로써 인용빈도에 의한 주요 연구 흐름의 변화를 살펴보고자 하였다. TC 0 이상인 네트워크는 데이터셋 내의 모든 논문 레코드를 포함시킨 것이고, TC 값이 높아질수록 SPLC 분석에 포함되는 논문의 수는 적어진다.

각 데이터셋 별 히스토리오그래프 및 SPLC 구성 레코드 수는 〈표 3〉에서 보는 바와 같이 SPLC 네트워크 추출을 통해 상위 0.2~7%의 주요 논문이 추출된 것을 알 수 있다. Data1, 3, 4, 5에서는 각각 6개의 히스토리오그래프와 SPLC 네트워크가 추출되었고, Data2에서는 4개의 네트워크가 추출되었다. Data2의 TC를 다른 데이터셋과는 다르게 선정한 것은 Data2의 규모가 다

른 데이터셋에 비해 작은 것을 고려한 것이다.

히스토리오그래프 추출에는 HistCite™를 이용하였고, SPLC 네트워크 추출에는 PAJEK v3.12를 이용하였다. SPLC 네트워크를 위한 인용 링크 가중치 생성에는 de Nooy 등이 제안한 알고리즘이 이용되었다(Doreian et al., 2000; Batagelj, 2003; de Nooy et al., 2011).

3.4 주요 논문 리스트 작성 및 개별 연구자 피드백

28개의 SPLC 네트워크에서 확인된 주요 논문 198개를 실제 국내의 특정 연구 기관의 관련 연구자에게 제공하고 이에 대한 피드백을 받음으로써, 이 연구에서 제안하는 방법이 연구지원 서비스에 사용될 수 있는 가능성을 살펴보고자 하였다.

피드백은 정부출연연구소 중 하나인 K에 소속된 LED 분야 연구자 7명에게 논문 리스트를 제공하고 이 중 자신의 연구와 관련하여 이미

보았거나 앞으로 살펴보고자 하는 논문에 표시하도록 하였다. 그리고 7명 중 3명에게서 응답 리스트를 받아 결과 분석을 수행하였다. 연구소 K를 개별 연구자 피드백을 위해 선정한 이유는 국책 연구소인 K 내의 주요 연구 및 연구 지원 분야로 LED가 있으며, 현재 이 기관에서 연구 지원 서비스를 다른 방식으로 진행하고 있으며, 이에 대한 지속적인 수요가 있는 것이 내부 담당자를 통해 확인이 되었기 때문이다.

수집된 3명의 응답 리스트는 이 연구에서 제안하는 기법의 적용을 위한 사례 분석에 사용되었고, 일정한 샘플 크기에 의한 통계적 유의성을 요구하는 분석은 적용되지 않았다. 따라서 이 사례 분석은 연구지원 정보서비스가 전체적인 공통의 지식 구조를 파악하여 개인 연구자의 정보 요구에 얼마나 맞게 제공하는지를 살펴보기 위한 시도라고 할 수 있다. 왜냐하면 실제 연구지원 정보서비스의 제공 대상은 특정 연구소에 소속되어 그 기관의 연구 목적을 따르는 개인 연구자라고 가정할 수 있기 때문이다. 서비스 제공을 위해 분석 대상이 되는 데이터는 특정 연구 분야에서 활동하고 있는 전세계의 연구자들의 연구 결과물이므로, 그 주제 분야에 대한 전세계 연구 집단의 공통된 지식 구조를 반영한다고 할 것이다. 네트워크 분석의 주요 목적과 가치는 이와 같은 공통의 지식 구조를 밝혀내는 것 그 자체에 있다. 하지만 전세계적인 연구 경향 중 특정 개별 연구자는 자신의 연구 특화 분야와 관련이 많은 것이 제공되기를 바랄 것이다. 따라서 특정한 연구 기관을 정하여 이에 소속된 개별 연구자들의 피드백을 받는 방식으로 사례 분석을 진행하였으며, 이용자의 만족도 또는 분석 결과의 정확성 분석의 일반화 접근은 하지 않았다.

4. 연구 결과

4.1 데이터 집합에 따른 LED 분야의 연도별 연구 동향

〈그림 2〉는 각 데이터셋의 출판년도별 논문 수와 히스토리오그래프 내의 출판년도별 논문 수를 히트맵으로 표현한 것이다. 2013년은 현재 연구 성과가 지속적으로 나오고 있으므로 분석에서 제외하였으나, RGB LED 연구는 연구 결과의 규모 측면에서 꾸준히 증가하고 있는 것을 알 수 있다.

연구의 양적 증가에 있어서 검색된 문헌과 인용문헌이 유사한 경향을 보이고 있지만, 인용빈도를 이용한 연도별 주요 연구흐름은 데이터셋별 또는 GCS 및 LCS별로 차이를 보였다. 검색 문헌만을 포함한 데이터셋에서는 GCS와 LCS 모두에서 1990년대가 주요 년도가 나타났으나, 인용문헌을 포함한 데이터셋에서는 2000년대 전후반 년도가 주요 연구년도로 나타났다. 그리고 검색 문헌만을 이용한 Data1과 Data2에서는 LCS에 의해 파악된 주요 연구 년도와 GCS에 의해 파악된 주요 연구 년도에 일부 차이를 보였다.

GCS에 의한 주요 연구년도를 데이터셋별로 살펴보면, 검색 문헌만을 이용한 Data1과 Data2가 유사한 경향을 보인다. 1994, 2000, 2006, 2007년에 발표된 연구 성과들이 최근까지 많은 인용을 받고 있으며, 1995, 2003년 연구 성과들이 그 뒤를 잇고 있는 것으로 나타났다. 인용문헌을 포함한 Data3, 4, 5는 1997, 2001, 2005년이 중요하게 인정받는 연구 성과가 많은 해로 나타났다. 그리고 2001~2007년에 걸친 시기의 연구

(a) 데이터셋 내 문헌 수						(b) 히스토리로그래프 내 문헌 수						
PY	Data1	Data2	Data3	Data4	Data5	PY	Data1	Data2	Data3	Data4	Data5	PY
1990	1				1	1990	1	0	0	0	1	1990
1991	10	1	2	3	12	1991	8	1	2	3	12	1991
1992	20	2	6	8	26	1992	14	1	6	8	22	1992
1993	22	2	32	34	52	1993	16	1	27	34	47	1993
1994	24	2	99	99	116	1994	20	2	86	99	112	1994
1995	29	3	228	228	243	1995	21	3	213	228	235	1995
1996	41	4	402	403	419	1996	32	4	388	403	414	1996
1997	90	9	543	545	573	1997	75	7	534	545	566	1997
1998	58	6	574	574	591	1998	51	6	564	574	585	1998
1999	59	6	647	650	673	1999	55	5	641	650	672	1999
2000	88	9	687	689	727	2000	75	9	662	689	721	2000
2001	66	7	797	799	824	2001	56	5	771	799	820	2001
2002	82	8	847	847	883	2002	72	8	819	847	880	2002
2003	95	10	774	777	810	2003	85	10	745	777	805	2003
2004	100	10	865	868	910	2004	84	10	828	868	902	2004
2005	100	10	901	902	933	2005	84	9	865	902	929	2005
2006	136	14	1097	1101	1143	2006	119	10	1067	1101	1139	2006
2007	158	16	1269	1272	1321	2007	133	12	1232	1259	1302	2007
2008	188	19	1505	1510	1565	2008	168	16	1467	1497	1550	2008
2009	191	19	1700	1703	1757	2009	160	13	1651	1678	1726	2009
2010	243	24	1952	1958	2033	2010	210	19	1886	1920	1985	2010
2011	239	24	2230	2231	2303	2011	195	21	2126	2174	2232	2011
2012	230	23	2367	2374	2437	2012	180	15	2255	2316	2363	2012

(c) GCS						(d) LCS						
PY	Data1	Data2	Data3	Data4	Data5	PY	Data1	Data2	Data3	Data4	Data5	PY
1990	119				119	1990	17				65	1990
1991	690	360	245	605	935	1991	48	1	10	349	485	1991
1992	921	508	488	996	1409	1992	83	2	127	580	693	1992
1993	788	465	1611	2076	2362	1993	47	5	439	857	886	1993
1994	4718	3724	8648	8648	9054	1994	483	42	4746	4749	4937	1994
1995	2761	1592	13028	13028	13131	1995	509	42	5083	5092	5181	1995
1996	2296	1721	21016	21881	22034	1996	286	30	7261	8003	8151	1996
1997	3429	2395	29157	29892	30147	1997	301	25	9032	9744	9934	1997
1998	2701	1880	23659	23659	23852	1998	273	30	7207	7214	7445	1998
1999	2083	1156	21130	21715	22136	1999	275	22	5527	5990	6258	1999
2000	4531	3261	24084	24451	24993	2000	315	37	6884	7219	7476	2000
2001	1764	907	43170	43515	43803	2001	247	7	5992	6230	6493	2001
2002	2326	1002	34786	34786	35373	2002	549	23	7035	7044	7445	2002
2003	3741	2318	36367	36986	37436	2003	433	26	7392	7987	8341	2003
2004	3012	1602	31101	31618	32200	2004	369	26	6792	7258	7563	2004
2005	2549	1086	35961	36028	36435	2005	200	14	5050	5063	5181	2005
2006	3658	1720	28666	29262	29895	2006	386	37	5375	5551	5848	2006
2007	4134	2448	30551	30885	31238	2007	363	39	6387	6506	6690	2007
2008	2755	1271	25611	26144	26507	2008	163	8	4308	4829	5006	2008
2009	2631	1469	25992	26262	26607	2009	189	14	4849	4855	5021	2009
2010	2177	928	19575	19712	20060	2010	163	5	3956	4102	4269	2010
2011	897	454	11104	11116	11246	2011	100	10	2406	2425	2481	2011
2012	176	104	3141	3173	3194	2012	13	0	583	618	633	2012

〈그림 2〉 RGB LED 분야의 연구 동향 히트맵

들이 많은 인용빈도를 받은 것으로 보인다. 분석 데이터 범위에 의한 주요 연구 년도의 차이는 LCS를 이용한 비교에서도 나타났다. Data1과 Data2에서는 1994, 1995년의 연구 성과가 가

장 중요한 것으로 공통적으로 나타났고, Data1은 2002년, Data2는 2000, 2006, 2007년이 각각 중요한 연구 년도로 밝혀졌다. Data 3, 4, 5에서는 이와는 다른 년도가 주요 연구 년도로 나타

났는데, 1996, 1997, 2002-2004년이다. 그리고 2000년과 2007년도 주요한 연구 년도로 나타났다.

4.2 상위 '주요' 논문 선정 방법에 따른 차이

단순 인용 빈도를 이용하여 상위 주요 논문을 선정하였을 때, 최신 논문이 배제될 가능성이 높고 주요 연구 흐름을 파악하는 데도 영향을 미치는 것으로 나타났다. <표 4>는 단순 인용 빈도를 기준으로 상위 주요 논문을 20개부터 233개(검색 문헌 의 10%)까지 선정하고 이들의 출판년도별 논문 수 분포를 살펴본 것이다. 검색 문헌의 10%인 233개 상위 문헌을 선정할 때까지, 최근 2년간의 논문은 인용 빈도만을 이용한 상위 주요 논문에는 포함되지 않는 것을 알 수 있다. Data1에서 2013년 논문은 인용빈도 기준 상위 1,900개 논문을 선정했을 때 처음 포함되었으며, 2012년 논문은 상위 1,100개 논문을 선정했을 때 처음 포함되었다. 따라서 단순 인용 빈도를 이용하면 이용자가 한 번에 살펴볼 수 있는 범위 내에서 최신 논문은 포함되기 어렵다는 것을 알 수 있다.

<그림 3>은 상위 문헌을 선정하는 방법에 따른 연구의 주요 흐름 차이를 보여준다. (a)와

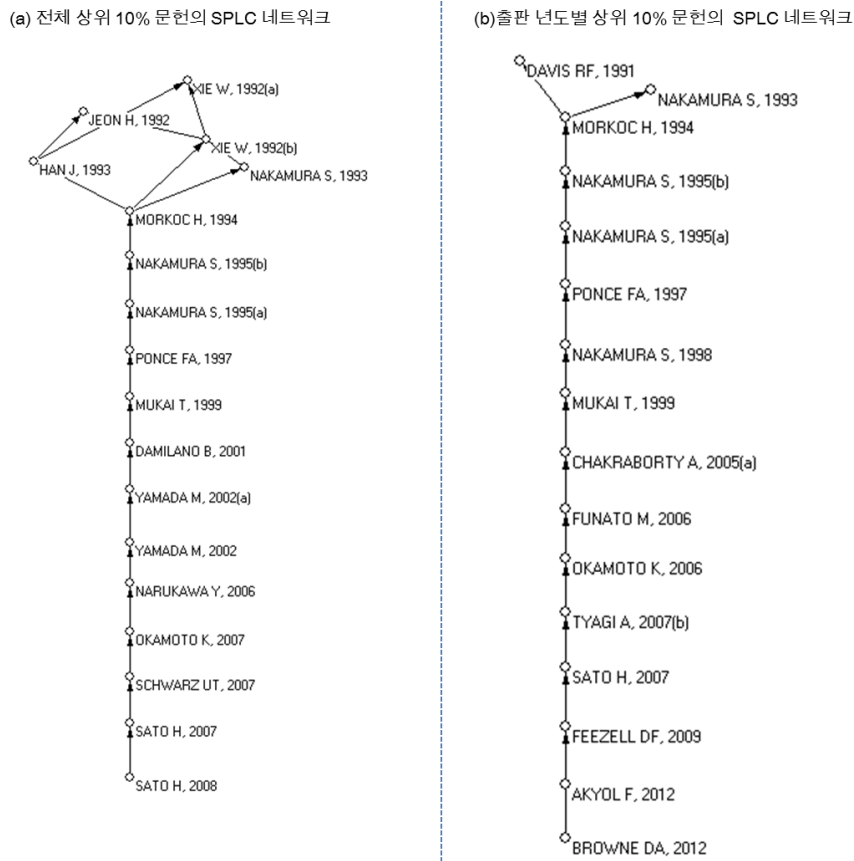
(b) 모두 233개 상위 문헌을 선정한 후 SPLC 네트워크를 추출한 결과로, (a)는 문헌의 출판년도를 고려하지 않은 경우이고 (b)는 출판년도를 고려하여 각 출판년도별 상위 10% 문헌을 선정한 것이다 (<표 2> 참조). 전체 데이터셋 내에서 출판년도의 구분없이 상위 n% 문헌을 선정하게 되면, 최근에 출판된 문헌이 선정될 가능성이 적기 때문에 (a)에서 보는 바와 같이 최근의 논문들이 제외되게 된다. 따라서 연구지원 정보시스템이 이용자가 원하는 최신의 논문 정보를 제공하기 위해 주요 연구흐름을 따르는 최신 논문을 선정하기가 어려울 수 있음을 보여준다.

4.3 인용빈도 임계치 및 인용문헌 포함에 의한 주요 연구흐름의 변화

인용 네트워크에도 선호연결현상이 나타나기 때문에(Merton, 1968; Barabasi, 2003; Price, 1976), 인용빈도 임계치를 높일수록 인용빈도가 낮은 대다수의 논문 노드가 제거된다. 따라서 <표 5>와 같이 각 데이터셋에서 TC를 높이면 히스토리오그래프의 크기는 감소하고, 밀도는 증가하였다. 즉, 네트워크의 사이즈의 감소로 히스토리오그래프 및 SPLC를 추정하는 계산의 효율성을 높여줄 수 있다.

<표 4> 인용 빈도 기준 선정 논문의 출판년도별 분포

선정 논문수	출판년도																							
	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
20	0	1	1	0	2	4	2	3	2	0	3	0	0	1	0	0	0	1	0	0	0	0	0	0
40	0	1	1	1	3	5	3	3	2	2	6	2	0	4	2	1	0	4	0	0	0	0	0	0
60	0	1	1	2	4	5	3	3	2	2	6	2	3	7	4	2	4	5	1	3	0	0	0	0
80	1	2	1	2	6	7	3	5	3	4	7	2	3	7	6	4	5	6	2	4	0	0	0	0
100	1	2	2	2	6	7	4	5	3	7	7	2	6	10	6	5	8	8	3	6	0	0	0	0
200	1	3	6	3	7	7	5	9	9	11	14	9	15	17	14	12	21	17	7	9	3	1	0	0
233	1	3	6	3	7	8	6	11	11	12	15	10	18	18	14	14	24	21	12	11	7	1	0	0



〈그림 3〉 상위 문헌 선정 방식에 따른 주요 연구 경향의 차이

〈표 5〉 인용빈도 임계치에 따른 히스토리오그래프 네트워크의 구조 변화

히스토리오그래프		TC		0	5	10	15	20	30	40	50
		size	density	size	density	size	density	size	density	size	density
Data1	size	1,943	-	145	-	58	32	20	15		
	density	0.0020	-	0.0200	-	0.0380	0.0650	0.1030	0.0140		
Data2	size	190	31	12	6	-	-	-	-		
	density	0.0120	0.083	0.1670	0.200	-	-	-	-		
Data3	size	19,943	-	2,282	-	1,123	689	478	374		
	density	0.0003	-	0.003	-	0.0060	0.0090	0.0110	0.0130		
Data4	size	19,943	-	2,282	-	1,123	689	478	374		
	density	0.0003	-	0.0030	-	0.0060	0.0090	0.0110	0.0130		
Data5	size	20,598	-	2,376	-	1,185	730	501	383		
	density	0.0003	-	0.0030	-	0.0050	0.0080	0.0110	0.0130		

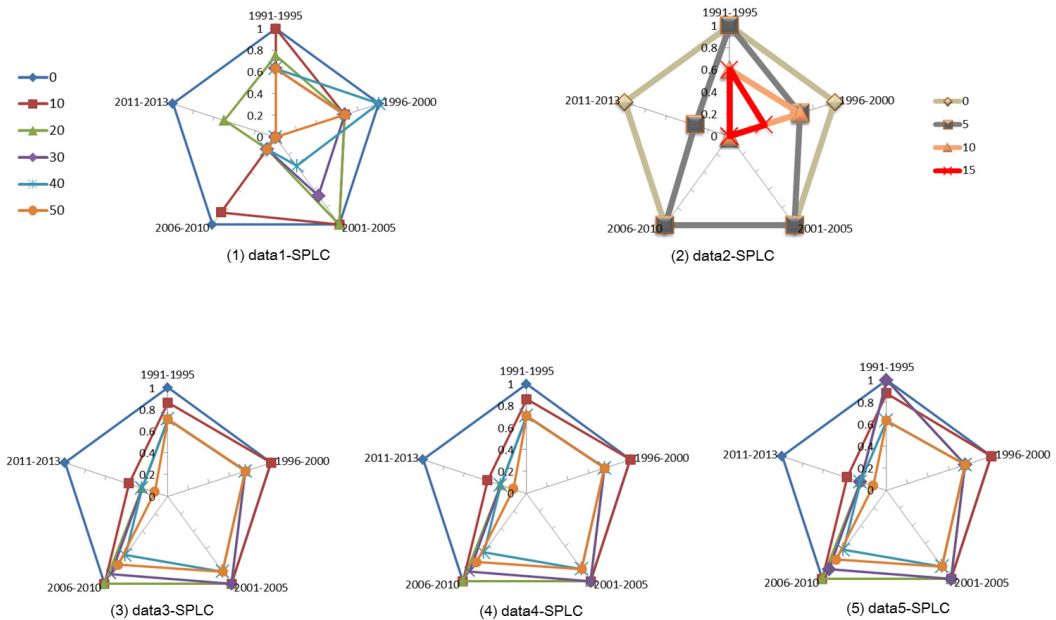
그러나 연구지원 서비스의 고려사항을 생각 할 때, 계산의 효율성과 함께 TC 변화에 따른 SPLC 네트워크의 변화를 살펴볼 필요가 있다. <표 6> 및 <그림 4>는 TC별 SPLC 네트워크에 포함된 출판년도별 논문 수의 변화이다.

모든 데이터 집합에서 TC 값이 증가할수록 특히 2011년 이후(현재 기준 최근 3년간)의 논문 수가 급격히 감소하고 있다. 이 때 TC를 0

에서 10으로 한 단계만 변화시켜도 논문 수 변화가 크다. 또한 검색 문헌만을 포함하는 Data1과 Data2에서 추출한 SPLC는 최근 논문 뿐 아니라 2001-2005년 논문의 수가 TC 변화에 따라 급격히 감소하는 것으로 나타났다. 그러나 인용 문헌을 포함한 Data3, 4, 5에서 추출한 SPLC는 TC가 변화하여도 동일한 시기의 논문 수가 감소하는 비율이 상대적으로 낮았다. 2001-2005

<표 6> 인용빈도임계치에 따른 SPLC 내 출판년도별 주요 논문 수 및 비율

데이터셋	출판년도	TC별 SPLC 네트워크 내 논문 수						SPLC0 대비 논문수 비율				
		SPLC0	SPLC10	SPLC20	SPLC30	SPLC40	SPLC50	SPLC10	SPLC20	SPLC30	SPLC40	SPLC50
Data1	1991-1995	8	8	6	5	5	5	1.00	0.75	0.63	0.63	0.63
	1996-2000	3	2	2	2	3	2	0.67	0.67	0.67	1.00	0.67
	2001-2005	3	3	3	2	1	0	1.00	1.00	0.67	0.33	0.00
	2006-2010	7	6	1	1	1	1	0.86	0.14	0.14	0.14	0.14
	2011-2013	2	0	1	0	0	0	0.00	0.50	0.00	0.00	0.00
	Total	23	19	13	10	10	8	0.83	0.57	0.43	0.43	0.35
Data3	1991-1995	7	6	5	5	5	5	0.86	0.71	0.71	0.71	0.71
	1996-2000	12	12	9	9	9	9	1.00	0.75	0.75	0.75	0.75
	2001-2005	7	7	7	7	6	6	1.00	1.00	1.00	0.86	0.86
	2006-2010	9	9	9	8	6	7	1.00	1.00	0.89	0.67	0.78
	2011-2013	8	3	2	2	2	1	0.38	0.25	0.25	0.25	0.13
	Total	43	37	32	31	28	28	0.86	0.74	0.72	0.65	0.65
Data4	1991-1995	7	6	5	5	5	5	0.86	0.71	0.71	0.71	0.71
	1996-2000	12	12	9	9	9	9	1.00	0.75	0.75	0.75	0.75
	2001-2005	7	7	7	7	6	6	1.00	1.00	1.00	0.86	0.86
	2006-2010	9	9	9	8	6	7	1.00	1.00	0.89	0.67	0.78
	2011-2013	8	3	2	2	2	1	0.38	0.25	0.25	0.25	0.13
	Total	43	37	32	31	28	28	0.86	0.74	0.72	0.65	0.65
Data5	1991-1995	8	7	8	8	5	5	0.88	1.00	1.00	0.63	0.63
	1996-2000	12	12	9	9	9	9	1.00	0.75	0.75	0.75	0.75
	2001-2005	7	7	7	7	6	6	1.00	1.00	1.00	0.86	0.86
	2006-2010	9	9	9	8	6	7	1.00	1.00	0.89	0.67	0.78
	2011-2013	8	3	2	2	2	1	0.38	0.25	0.25	0.25	0.13
	Total	44	38	35	34	28	28	0.86	0.80	0.77	0.64	0.64
Data2	출판년도	SPLC0	SPLC5	SPLC10	SPLC15			SPLC5	SPLC10	SPLC15		
	1991-1995	5	5	3	3	-	-	1.00	0.60	0.60	-	-
	1996-2000	3	2	2	1	-	-	0.67	0.67	0.33	-	-
	2001-2005	2	2	0	0	-	-	1.00	0.00	0.00	-	-
	2006-2010	3	3	0	0	-	-	1.00	0.00	0.00	-	-
	2011-2013	3	1	0	0	-	-	0.33	0.00	0.00	-	-
Total	16	13	5	4			0.81	0.31	0.25			



〈그림 4〉 인용빈도임계치에 따른 SPLC 내 출판년도별 주요 논문 변화

년은 앞의 분석에서 Data3, 4, 5에서 모두 GCS와 LCS가 유사한 경향을 보였지만, Data1과 Data2에서는 그러한 경향이 파악되지 않았다. 따라서 분석 데이터 집합 내에 인용문헌을 포함하는 것이 분석 데이터셋을 전체적인 연구 동향과 보다 유사하게 만들 수 있고, 이를 통해 인용빈도 임계치에 의한 분석 결과의 왜곡을 방지할 수 있는 것으로 보인다.

〈표 7〉은 인용문헌의 포함 여부가 네트워크 왜해도에 미치는 영향을 살펴보기 위해, 각 데이터셋에서 TC 변화에 따른 SPLC 네트워크 간의 매칭 유사도와 오차아이 유사도를 계산한 것이다. 두 SPLC 네트워크 간에 동일한 논문이 포함되는 비율이 높을수록 이들 간의 유사도는 높게 나타난다.

각 데이터셋 내의 SPLC 네트워크간 유사도를 살펴보면, TC를 높일수록 SPLC0과는 다르

게 구성되는 것을 알 수 있다. 예를 들어, Data1에서는 SPLC0과 나머지 5개 네트워크(SPLC10~SPLC50) 간의 유사도가 인용빈도 임계치 변화에 따라 급격히 감소하여 SPLC0은 SPLC50과 유사도가 0이 된다.

그러나 인용문헌을 분석에 포함하면 TC에 의한 네트워크의 변화(왜해되는 정도)가 완화되는 것으로 나타났다. 예를 들어 Data3의 SPLC0과 나머지 5개 네트워크(SPLC10~SPLC50) 간의 유사도는 Data1과 마찬가지로 감소하지만 감소하는 정도는 Data1보다는 작았다. 이와 같은 현상은 Data4와 Data5에서도 관찰된다.

네트워크가 변화하는 정도는 네트워크에 포함된 논문들이 동질적일 때도 완만한 것으로 나타났다. 즉, Data2의 SPLC0과 나머지 3개 네트워크(SPLC5~SPLC15) 간의 유사도는 Data1의 네트워크들 간 변화에 비하여 적게 감소하였

〈표 7〉 인용빈도임계치별 SPLC 네트워크간 유사도(동일 데이터셋 내 비교)

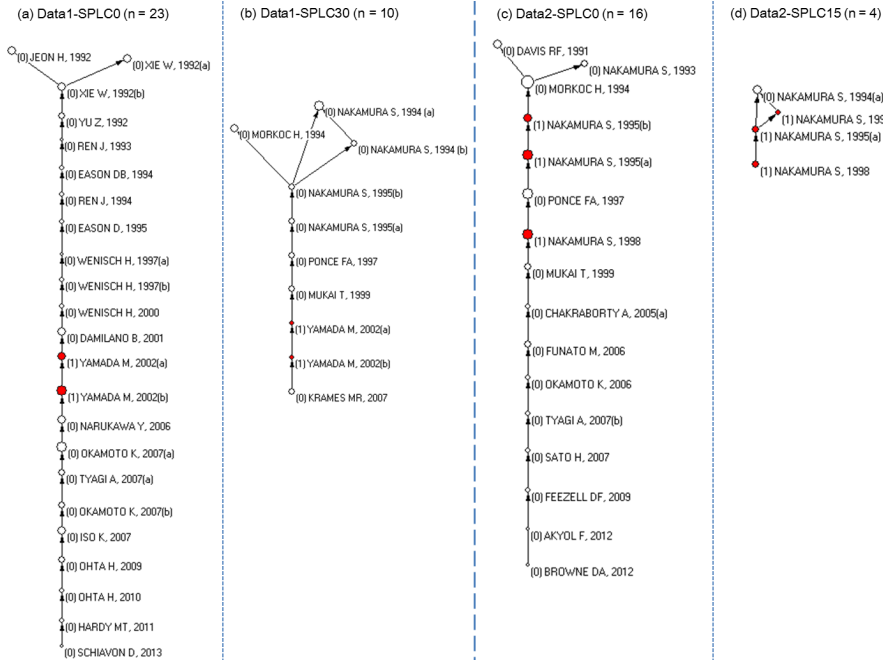
데이터셋		매칭 유사도						오차아이 유사도					
		SPLC0	SPLC10	SPLC20	SPLC30	SPLC40	SPLC50	SPLC0	SPLC10	SPLC20	SPLC30	SPLC40	SPLC50
Data1	SPLC0	23						1.00					
	SPLC10	11	19					0.53	1.00				
	SPLC20	4	11	13				0.23	0.70	1.00			
	SPLC30	2	8	9	10			0.13	0.58	0.79	1.00		
	SPLC40	1	7	8	9	10		0.07	0.51	0.70	0.90	1.00	
	SPLC50	0	5	6	7	8	8	0.00	0.41	0.59	0.78	0.89	1.00
Data3	SPLC0	43						1.00					
	SPLC10	37	37					0.93	1.00				
	SPLC20	20	20	32				0.54	0.58	1.00			
	SPLC30	19	19	31	31			0.52	0.56	0.98	1.00		
	SPLC40	15	15	26	26	28		0.43	0.47	0.87	0.88	1.00	
	SPLC50	15	15	23	23	23	28	0.43	0.47	0.77	0.78	0.82	1.00
Data4	SPLC0	43						1.00					
	SPLC10	37	37					0.93	1.00				
	SPLC20	20	20	32				0.54	0.58	1.00			
	SPLC30	19	19	31	31			0.52	0.56	0.98	1.00		
	SPLC40	15	15	26	26	28		0.43	0.47	0.87	0.88	1.00	
	SPLC50	15	15	23	23	23	28	0.43	0.47	0.77	0.78	0.82	1.00
Data5	SPLC0	44						1.00					
	SPLC10	37	38					0.90	1.00				
	SPLC20	19	20	35				0.48	0.55	1.00			
	SPLC30	18	19	34	34			0.47	0.53	0.99	1.00		
	SPLC40	13	14	26	26	28		0.37	0.43	0.83	0.84	1.00	
	SPLC50	13	14	23	23	23	28	0.37	0.43	0.73	0.75	0.82	1.00
data2		SPLC0	SPLC5	SPLC10	SPLC15			SPLC0	SPLC5	SPLC10	SPLC15		
	SPLC0	16						1.00					
	SPLC5	5	13					0.35	1.00				
	SPLC10	6	6	6				0.45	0.50	1.00			
	SPLC15	3	3	4	4			0.38	0.42	0.89	1.00		

는데, 이는 Data2가 인용빈도 기준 상위 10%로 선정된 논문 레코드만을 포함하므로, 데이터 간의 동질성이 상대적으로 높기 때문인 것으로 보인다.

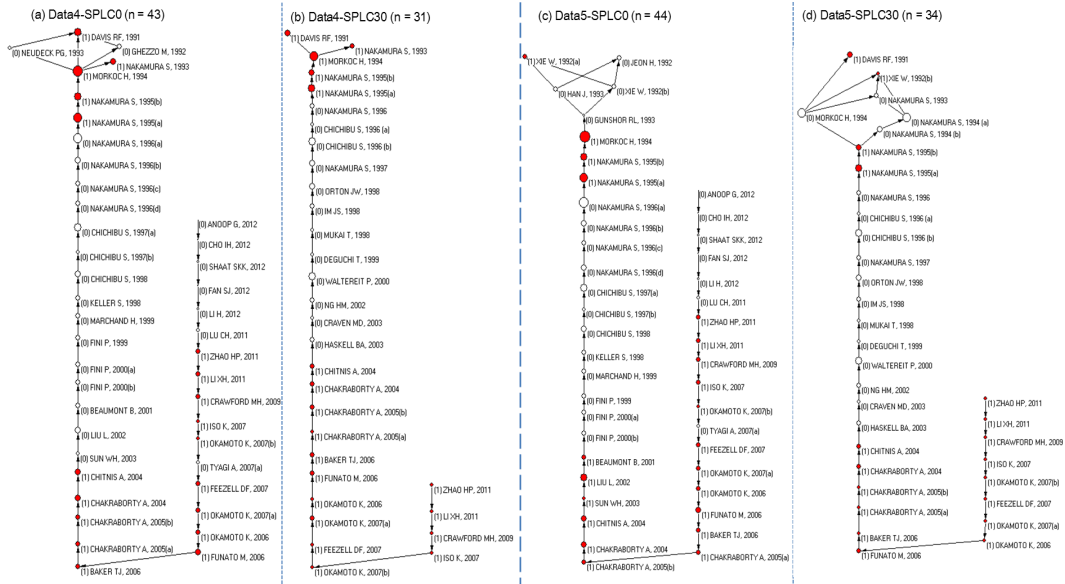
그리고 네트워크의 변화는 TC 변화에 비례하여 지속적으로 변화하지 않고 일정 수준에서 수렴하였다. 즉, 모든 데이터 집합에서 TC를 10 이상으로 하는 5개 네트워크 간에는 오차아이 유사도가 최소 0.80에서 최대 0.99로 크게 변화

하지 않음을 알 수 있다. 뿐만 아니라, SPLC40은 SPLC50과 오차아이 유사도가 0.99로 나타나, 네트워크가 더 이상 와해되지 않고 서로 유사하게 수렴함을 알 수 있다.

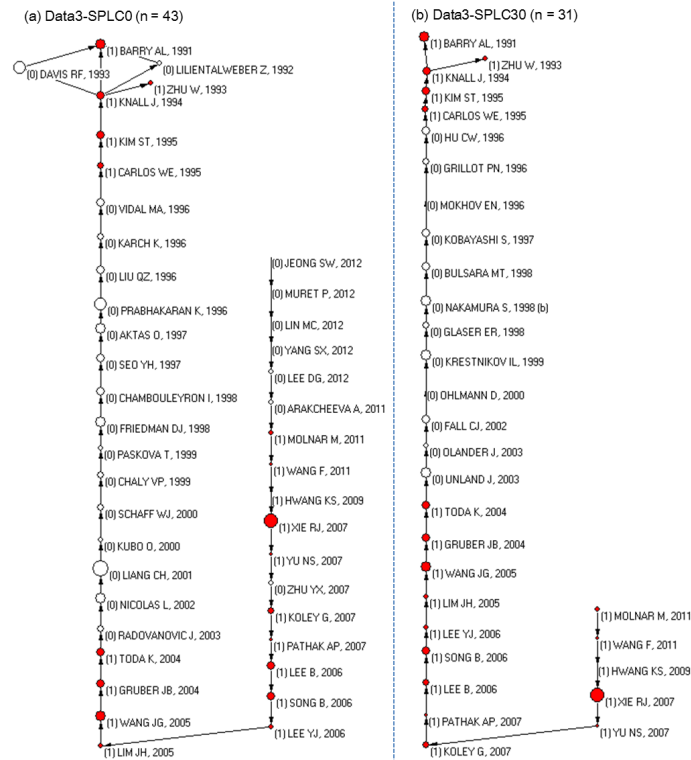
TC에 의한 SPLC 네트워크의 구조 변화를 시각화하면 <그림 5>~<그림 7>과 같다. 이는 각 데이터셋에서 SPLC0과 SPLC30을 비교한 것이다. 각 노드는 '저자명, 출판년도'로 레이블링 하였으며, 각 노드의 크기는 인용빈도에 비례한



〈그림 5〉 인용빈도 임계치 변화에 따른 SPLC 네트워크 비교: 검색 문헌 및 출판년도별 상위 10% 문헌



〈그림 6〉 인용빈도 임계치 변화에 따른 SPLC 네트워크 비교: 상위 인용문헌을 포함한 경우



〈그림 7〉 인용문헌의 인용빈도 임계치 변화에 따른 SPLC 네트워크 비교

다. 두 네트워크에서 동일한 노드는 색상으로 표시하고, 레이블의 각괄호안에 '1'로 표시하였다.

인용문헌을 포함하지 않고 TC를 변화시켜 SPLC 네트워크를 추출하면 주요 연구흐름에 포함되는 최신 문헌이 감소하고, SPLC0과 SPLC30 간의 공통된 문헌은 주로 연구 초기 논문들로 나타났다(〈그림 5〉 참조). 하지만 인용문헌을 포함하면 TC를 변화시켜도 비교적 최신 문헌이 공통된 문헌으로 남았다(〈그림 6〉 참조). 뿐만 아니라 인용문헌을 포함하여 분석한 경우, 연구의 흐름 중 GCS가 높은 년도의 논문들이 대체적으로 배제되는 것을 확인할 수 있다. 예를 들면, Data1에서는 GCS가 높은 년도인 2000년과 2003년의 논문이 SPLC30에 포함되지 않았

다. 그러나 Data4와 Data5에서는 GCS가 높은 것으로 나타난 2002-2004년, 2007년의 논문들이 TC를 증가시킴에도 제거되지 않았다. 따라서 인용문헌을 분석에 포함하지 않고 인용빈도에 의한 제한을 두는 것은 전역적인 연구흐름을 파악하는 데 영향을 줄 수 있는 것으로 보인다. 마지막으로 Data3을 이용한 SPLC는 다른 데이터셋과는 달리 검색 문헌이 포함되지 않았기 때문에, 주요 연구흐름 내에 포함되는 문헌이 다른 데이터셋의 SPLC는 전혀 일치하지 않았지만(〈그림 7〉 참조), 다른 데이터셋과 마찬가지로 TC 증가에 의한 네트워크의 변화 경향은 동일하였다.

〈표 8〉은 TC 증가에 따른 SPLC 네트워크

〈표 8〉 SPLC0 및 SPLC30 네트워크 간 유사도

		SPLC0					SPLC30				
		Data1	Data2	Data3	Data4	Data5	Data1	Data2	Data3	Data4	Data5
SPLC0	Data1	1									
	Data2	0.00	1								
	Data3	0.00	0.00	1							
	Data4	0.13	0.30	0.00	1						
	Data5	0.22	0.23	0.00	0.90	1					
SPLC30	Data1	0.13	0.57	0.00	0.14	0.14	1				
	Data2	0.00	0.38	0.00	0.15	0.15	0.47	1			
	Data3	0.00	0.00	0.52	0.00	0.00	0.00	0.00	1		
	Data4	0.11	0.36	0.00	0.52	0.46	0.17	0.18	0.00	1	
	Data5	0.14	0.34	0.00	0.50	0.47	0.27	0.26	0.00	0.95	1

의 변형을 네트워크 간 오차아이 유사도로 살펴본 것으로, 가장 많은 변화를 가져오는 것은 검색된 문헌만을 분석에 이용한 경우이다. Data1의 SPLC0과 SPLC30 간의 유사도는 0.13으로, Data2에서의 0.38, Data3와 Data4의 0.52, Data5의 0.47보다 낮은 수치이다. 따라서 TC 증가에 의해 초기 네트워크인 Data1-SPLC0 내의 문헌과는 다른 문헌들이 주요 연구흐름에 포함된 것을 알 수 있다.

그리고 인용문헌이 SPLC 네트워크의 범위를 확장해 주는 것을 알 수 있다. 예를 들면, Data1-SPLC0과 Data5-SPLC0 간의 유사도는 0.22, Data1-SPLC30과 Data5-SPLC30 간의 유사도는 0.27로, 인용문헌을 포함함으로써 SPLC 네트워크의 범위가 확장되어 유사도가 낮아진 것을 알 수 있다. 또한 Data2-SPLC0과 Data1-SPLC30 간의 유사도가 0.57로 나타나, TC 임계치를 높이면 상위 10% 문헌과 유사한 주요 연구흐름을 보이는 것으로 나타났다.

TC에 따라 SPLC네트워크가 변화함에도 각 데이터셋에서 공통으로 출현하는 논문이 있다면, 이는 특정 분야의 고전 논문이라고 볼 수 있

을 것이다. 이 고전 논문의 서지사항은 〈표 9〉와 같다. Data1에서 파악된 핵심논문이 2002년에 출판된 것을 제외하면, 다른 데이터셋에서 파악된 고전논문은 연구 초기 시기인 1990년대 초반 논문으로 확인되었다. 9번과 10번 논문을 보면, 고전논문을 파악하는 데 있어서는 출판년도별 상위 10% SPLC 네트워크와 인용문헌을 포함한 SPLC 네트워크가 유사한 결과를 낼 수 있음을 알 수 있다. 또한 검색 문헌만을 이용함으로써 나타난 핵심논문으로 이 결과를 보완할 수 있음을 알 수 있다.

4.4 전역적 인용 행태와 지역적 연구 지원 요구간의 비교: 사례 연구

〈표 10〉은 피드백 결과를 데이터셋별 SPLC로 나눠서 표시한 것이다. 이를 통해 인용문헌의 포함여부와 TC 변화가 개별 이용자의 정보요구에 적절한 문헌을 SPLC 내에 포함시키는데 영향을 주는지 살펴보고자 하였다.

분석 결과, 모든 데이터셋에서 인용빈도 임계치를 0에 가까이할수록 더 많은 문헌이 제공

<표 9> SPLC 네트워크 분석을 통해 확인한 데이터셋별 고전논문

No.	제목	저자 (*성, 이름 약어로 표기)	학술지명	출판년도	데이터셋
1	Energy dependence of electron damage and displacement threshold energy in 6H silicon carbide	BARRY, AL; LEHMANN, B; FRITSCH, D; BRAUNIG, D	IEEE Transactions on Nuclear Science	1991	3
2	Thin film deposition and microelectronic and optoelectronic device fabrication and characterization in monocrystalline alpha and beta silicon carbide	DAVIS, RF; KELNER, G; SHUR, M; PALMOUR, JW; EDMOND, JA	Proceedings of the IEEE	1991	4
3	Geometric modeling of the diamond-β(beta)-SiC heteroepitaxial interface	ZHU, W; WANG, XH; STONER, BR; KONG, HS; BRAUN, MWH; GLASS, JT	Diamond and Related Materials	1993	3
4	The use of graded InGaAs layers and patterned substrates to remove threading dislocations from GaAs on Si	KNALL, J; ROMANO, LT; BIEGELSEN, DK; BRINGANS, RD; CHUI, HC; HARRIS, JS; TREAT, DW; BOUR, DP	Journal of Applied Physics	1994	3
5	P-GaN/N-InGaN/N-GaN double-heterostructure blue-light-emitting diodes	NAKAMURA, S; SENOH, M; MUKAI, T	Japanese Journal of Applied Physics	1993	4
6	Large-band-gap SiC, III-V nitride, and II-VI ZnSe-based semiconductor device technologies	MORKOC, H; STRITE, S; GAO, GB; LIN, ME; SVERDLOV, B; BURNS, M	Journal of Applied Physics	1994	4.5
7	Surface-mode stimulated emission from optically pumped GaInN at room temperature	KIM, ST; AMANO, H; AKASAKI, I	Applied Physics Letters	1995	3
8	Paramagnetic resonance in GaN-based light emitting diodes	CARLOS, WE; GLASER, ER; KENNEDY, TA; NAKAMURA, S	Applied Physics Letters	1995	3
9	High-brightness InGaN blue, green and yellow light-emitting diodes with quantum well structures	NAKAMURA, S; SENOH, N; IWASA, N; NAGAHAMA, SI	Japanese Journal of Applied Physics	1995	2,4,5
10	Superbright green InGaN single-quantum-well-structure light-emitting diodes	NAKAMURA, S; SENOH, M; IWASA, N; NAGAHAMA, S; YAMADA, T; MUKAI, T	Japanese Journal of Applied Physics	1995	2,4,5
11	phosphor free high-luminous-efficiency white light-emitting diodes composed of InGaN multi-quantum well	YAMADA, M; NARUKAWA, Y; MUKAI, T	Japanese Journal of Applied Physics	2002	1

<표 10> SPLC 결과에 대한 이용자 피드백

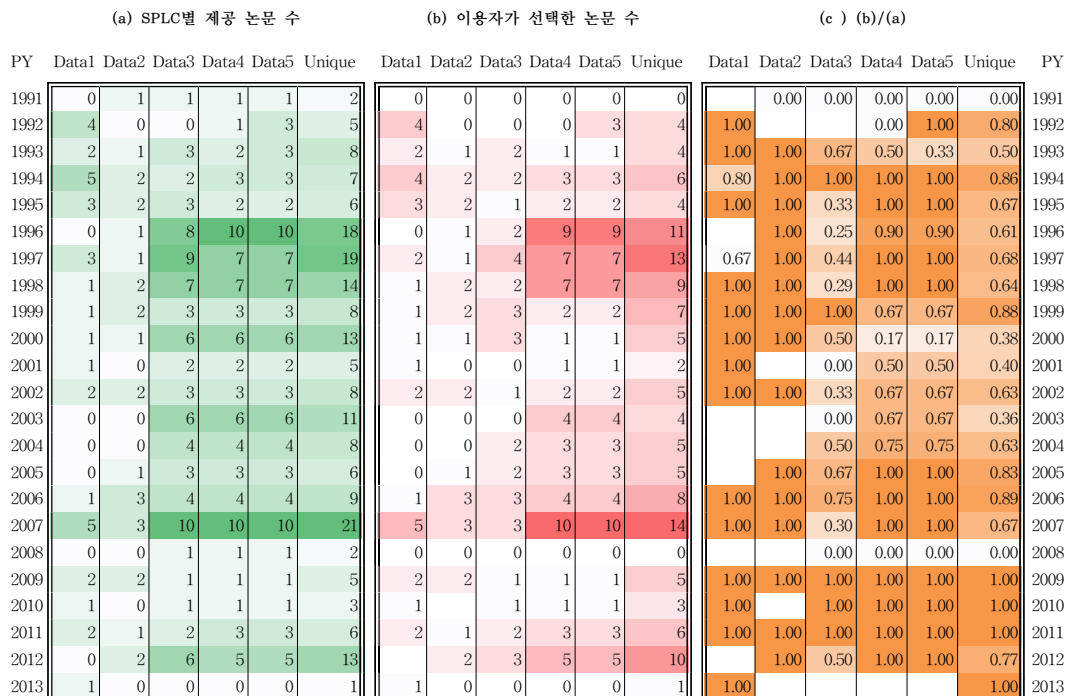
		논문 수						비율					
		SPLC0	SPLC10	SPLC20	SPLC30	SPLC40	SPLC50	SPLC0	SPLC10	SPLC20	SPLC30	SPLC40	SPLC50
Data1	선택x	2	0	0	0	0	0	0.09	0.00	0.00	0.00	0.00	0.00
	선택o	21	19	13	10	10	8	0.91	1.00	1.00	1.00	1.00	1.00
	total	23	19	13	10	10	8	1.00	1.00	1.00	1.00	1.00	1.00
Data3	선택x	21	19	13	13	12	12	0.49	0.51	0.41	0.42	0.43	0.43
	선택o	22	18	19	18	16	16	0.51	0.49	0.59	0.58	0.57	0.57
	total	43	37	32	31	28	28	1.00	1.00	1.00	1.00	1.00	1.00
Data4	선택x	6	5	2	2	2	3	0.14	0.14	0.06	0.06	0.07	0.11
	선택o	37	32	30	29	26	25	0.86	0.86	0.94	0.94	0.93	0.89
	total	43	37	32	31	28	28	1.00	1.00	1.00	1.00	1.00	1.00
Data5	선택x	5	3	2	2	2	3	0.11	0.08	0.06	0.06	0.07	0.11
	선택o	39	35	33	32	26	25	0.89	0.92	0.94	0.94	0.93	0.89
	total	44	38	35	34	28	28	1.00	1.00	1.00	1.00	1.00	1.00
Data2		SPLC0	SPLC5	SPLC10	SPLC15			SPLC0	SPLC5	SPLC10	SPLC15		
	선택x	1	0	0	0			0.06	0.00	0.00	0.00		
	선택o	15	13	12	6			0.94	1.00	1.00	1.00		
	total	16	13	12	6			1.00	1.00	1.00	1.00		

되고 선택되었으나, 선택비율을 살펴보면 인용 문헌을 포함하는지 여부에 따라 특징이 다르게 나타났다. 검색문헌만을 포함하는 Data1과 상위 10% 문헌인 Data2에서는 인용빈도 임계치를 10 이상으로 한 경우, 이용자가 SPLC 내 모든 문헌을 선택함으로써 이 사례에서는 가장 개별 이용자의 정보요구에 부합한 것으로 나타났다. 인용문헌을 포함한 Data3, 4, 5에서는 SPLC20에서 모두 가장 높은 선택율을 보였으며, TC가 40 이상으로 증가하면 선택율이 일부 감소하였다. 그리고 인용문헌만을 제공할 때(Data3), 전역적인 연구흐름이 K 기관 소속 개인연구자의 정보 요구와 크게 일치하지 않는 것으로 나타났다.

그리고 개별 연구자의 연구 관심과 SPLC에

서 제공된 주요 연구흐름을 비교하여, 이용자의 정보요구가 주요 연구흐름과 일치하는 정도를 살펴보았다. <그림 8>에서의 (a)는 데이터셋별 SPLC를 통해 추출된 논문수를 출판년도별로 표시한 것이고 <그림 8>에서의 (b)는 SPLC를 통해 제공된 논문 중 실제 이용자가 선택한 논문을 표시한 것이다. 그리고 <그림 8>에서의 (c)는 SPLC에서 제공된 논문수 대비 이용자가 선택한 논문의 비율이다.

각 히트맵에서 개별 이용자가 갖는 주요 연구흐름에 대한 이해는 Data3을 제외한 데이터셋의 SPLC 네트워크와 유사하게 나타났고, 특히 Data2의 경우 1991년 논문을 제외한 모든 논문이 이용자에게 선택됨으로써 상위 10%의 논문들이 주요 연구흐름의 중심이 되는 것으로 보인다.



<그림 8> SPLC 네트워크와 이용자 피드백 간 히트맵 비교

그리고 <그림 8-(b)>와 <그림 8-(c)>를 비교한 결과, 검색 문헌과 상위 인용문헌을 모두 포함하여 분석 제공하는 것이 더 많은 수의 주요 논문을 이용자에게 제공할 수 있는 것으로 나타났다. 왜냐하면 Data1과 Data2에서 제공된 논문을 선택한 비율이 Data4와 Data5를 선택한 비율보다 높지만, 선택된 논문 수를 보면 Data4와 Data5가 더 많기 때문이다. 하지만 이 연구에서 참여시킨 잠재적 이용자는 특정 연구소 K의 소수(3명)의 LED 분야 전문가이므로 이 결과만으로 일반화할 수 없고, 보다 많은 케이스 분석을 통해 이론화 및 일반화해야 할 것이다.

5. 결론 및 제언

이 연구에서는 인용 네트워크 분석 데이터의 범위와 인용빈도 임계치 설정에 따른 주요 연구 흐름의 차이를 히스토리오그래프와 SPLC 분석을 적용하는 경우에서 살펴보고, 이를 통해 연구 지원 정보서비스를 개발하는 데 있어서 적절한 분석 데이터의 범위와 인용빈도 임계치 설정값이 있는지 탐험적으로 고민해 보고자 하였다.

분석 결과, 인용문헌을 포함한 경우에 수집된 데이터셋에 한정되지 않고 전세계적인 주요 연구흐름에 포함되는 논문의 수가 많아지는 것으로 나타났다. 그리고 검색 문헌에 인용문헌을 합하여 분석한 경우의 GCS가 높은 년도는 검색 문헌만을 이용하여 분석한 경우의 GCS가 높은 년도와는 다르게 파악되었다. 이는 검색의 키워드 매칭으로는 찾지 못한 논문들이 인용 관계를 통해 포함됨으로써 분석 결과를 보다 풍부하게 할 수 있다고 유추해 볼 수 있다. 또한 인용

문헌을 포함한 데이터셋은 LCS와 GCS가 높은 년도가 서로 다르게 나타남으로써, 데이터셋의 LCS에 의해 치우치지 않는 분석 결과를 낼 수 있는 것으로 보인다.

그리고 인용빈도를 기준으로 일정 수의 데이터로 한정하여 히스토리오그래프와 SPLC 분석을 수행한 결과 최근 3년 이내 출판 논문이 가장 높은 비율로 감소하고, 일정한 인용빈도 값이 되면 주요 연구흐름이 유사하게 수렴하는 것으로 나타났다. 특히 인용빈도임계치가 0인 SPLC 네트워크와 다른 인용빈도임계치의 SPLC 네트워크 간의 유사도는 낮지만, 일정 값 이상(여기서는 30)으로 인용빈도 임계치를 정하면 SPLC 네트워크들이 서로 유사하게 수렴하였다. 이는 인용 네트워크 내에 허브가 되는 논문들이 있기 때문인 것으로 해석된다. 따라서 최근 논문을 주요 흐름 내에 살리기 위해서는 인용빈도 임계치와 인용문헌을 함께 고려하였을 때, 최근 논문들이 분석에서 제외되지 않을 가능성이 높은 것으로 보인다. 동시에 일정 인용빈도 이상에도 SPLC에 공통적으로 포함되어 있는 논문들은 LED 분야의 주제적 응집성을 담당하는 허브 논문들로 볼 수 있으므로, 그 분야의 연구의 기본이 되는 논문으로 특정 이용자층에게 제공할 수 있을 것이다. 뿐만 아니라, 이와 같이 파악된 주요 논문이 특정 기관 소속 이용자의 연구 관심에 의해 선택된 비율이 최신 논문(2009년 이후)일수록 높으므로, 최신의 연구 동향 모니터링 서비스를 제공하는 것이 가능할 것으로 보인다. 요약하면, 단순 인용빈도를 서로 다른 크기와 수집 방식으로 형성된 데이터셋에 제공하는 것은 무리가 있으며, 분석을 통해 살펴보고자 하는 연구 동향의 시간적 범위와 연구지

원 서비스에서 제공하고자 하는 서비스의 범위를 모두 함께 고려하여 적절한 인용빈도 임계치를 설정할 필요가 있을 것이다.

이 연구는 개인화된 연구 관심을 지원하는 서비스에 활용할 계량서지학적 분석 방법과 이를 개인 연구자에게 제공함으로써 개인의 연구 관심과 그 분야의 전반적 주요 동향을 비교하려고 시도했다는 점에서 의미가 있을 것이다. 뿐만 아니라, 연구지원 정보서비스의 방법론으로써 히스토리오그래프와 SPLC를 통해 찾아봐야 하는 논문 수를 줄여서 주요 연구 동향을 제공할 수 있는 가능성을 제시하고, 인용빈도 임계치 설정 및 데이터 수집과 관련하여 분석 대

상이 되는 데이터의 범위가 분석 결과에 영향을 미침을 확인하고자 시도하였다는 점에서도 의미가 있다. 그리고 네트워크 분석 기법을 연구지원 정보서비스에 적용할 때 고려해야 할 데이터의 특성을 살펴봄으로써 수행함으로써, 빅데이터 분석을 이용한 정보서비스 개발의 사례를 제시하였다고 할 수 있다. SPLC 추출 기법을 정교화 하여 여러 주제 분야의 사례 분석을 수행하는 동시에, 다른 인용 분석 기법과의 비교 및 서비스를 제공받는 연구자의 개인화된 연구 특성을 파악하는 좀 더 체계화된 후속 연구를 통해 이 연구의 접근 방법을 보완하고 이론화할 필요가 있다고 생각된다.

참 고 문 헌

- 권나현, 이정연, 정은경 (2012). 과학기술분야 R&D 전주기 연구: 국내 생명 및 나노과학기술 연구자를 중심으로. 한국문헌정보학회지, 46(3), 103-131.
<http://dx.doi.org/10.4275/KSLIS.2012.46.3.103>
- 유소영, 이재운 (2008). 학제적 분야의 정보서비스를 위한 학술지 인용 분석에 관한 연구: Y대학교 생명공학과를 중심으로. 정보관리학회지, 25(4), 283-308.
<http://dx.doi.org/10.3743/KOSIM.2008.25.4.283>
- 이정연, 정은경, 권나현 (2012). 과학기술분야 연구활동 단계별 문제상황 극복을 위한 정보행동 연구. 정보관리학회지, 29(3), 99-122. <http://dx.doi.org/10.3743/KOSIM.2012.29.3.099>
- 이재운 (2013.2.17). 계량정보분석과 정보관리실무. 2013 정보관리와 네트워크 분석 워크숍 발표자료, 해운대, 부산.
- 이재운, 김희정 (2011). 연구지원 서비스를 위한 계량서지적 분석: 국제백신연구소 연구동향을 대상으로. 제18회 한국정보관리학회 학술대회 논문집, 11-16.
- 이재운, 김관준, 강대신, 김희정, 유소영, 이우형 (2011). 계량서지적 기법을 활용한 LED 핵심 주제영역의 연구 동향 분석. 정보관리연구, 42(3), 1-26. <http://dx.doi.org/10.1633/JIM.2011.42.3.001>
- Allard, S., Levine, K. J., & Tenopir, C. (2009). Design engineers and technical professionals at work: Observing information usage in the workplace. Journal of the American Society

- for Information Science and Technology, 60(3), 443-454.
<http://dx.doi.org/10.1002/asi.21004>
- Barabasi, A. (2003). *Linked: How everything is connected to everything else and what it means for business, science, and everyday life*. New York: Plume.
- Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv preprint cs/0309023. Retrieved from <http://arxiv.org/pdf/cs/0309023>
- Bichteler, J., & Ward, D. (1989). Information-seeking behavior of geoscientists. *Special Libraries*, 79(3), 169-178.
- Case, D. O. (2012). *Looking for Information: A survey of research on information seeking, needs, and behavior* (3rd ed.). Howard House: Emerald.
- Case, D. O., Borgman, C. B., & Meadow, C. T. (1986). End-user information-seeking in the energy field: Implications for end-user access to DOE RECON databases. *Information Processing & Management*, 22, 299-308.
- Choi, C., & Park, Y. (2009). Monitoring the organic structure of technology based on the patent development paths. *Technological Forecasting & Social Change*, 76(6), 754-768.
<http://dx.doi.org/10.1016/j.techfore.2008.10.007>
- Colicchia, C., & Strozzi, F. (2012). Supply chain risk management: A new methodology for a systematic literature review. *Supply Chain Management: An International Journal*, 17(4), 403-418. <http://dx.doi.org/10.1108/13598541211246558>
- de Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (Revised and expanded second edition). Cambridge: Cambridge University Press.
- Doreian, P., Batagelj, V., & Ferligoj, A. (2000). Symmetric-acyclic decompositions of networks. *Journal of Classification*, 17(1), 3-28.
- Garfield, E. (2001.9.19). From computational linguistics to algorithmic historiography. Lazerow Lecture held in conjunction with panel on "Knowledge and Language: Building large-scale knowledge bases for intelligent applications," presented at the University of Pittsburgh. Retrieved from <http://garfield.library.upenn.edu/papers/pittsburgh92001.pdf>
- Garfield, E. (2001.11.27). From bibliographic coupling to co-citation analysis via algorithmic historio-bibliography: A citationist's tribute to Belver C. Griffith. Lazerow Lecture presented at Drexel University, Philadelphia, PA. Retrieved from <http://garfield.library.upenn.edu/papers/drexelbevergrif?th92001.pdf>
- Garfield, E., Pudovkin A. I., & Istomin, V. S. (2002). Algorithmic citation-linked historiography-Mapping the literature of science. *Proceedings of the American Society for Information Science*

- and Technology Annual Meeting, 39, 14-24.
- Garfield, E., Pudovkin, A. I., & Istomin, V. S. (2003). Why do we need algorithmic historiography? *Journal of the American Society for Information Science and Technology*, 54(5), 400-412. <http://dx.doi.org/10.1002/asi.10226>
- Holland, M. P., & Powerl, C. K. (1995). A longitudinal survey of the information seeking and use habits of some engineers. *College and Research Libraries*, 55(1), 7-15.
- Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, 11, 39-63.
- Kajikawa, Y., Yoshikawa, J., Takeda, Y., & Matsushima, K. (2008). Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy. *Technological Forecasting & Social Change*, 75(6), 771-782. <http://dx.doi.org/10.1016/j.techfore.2007.05.005>
- King, D. W., & Tenopir, C. (2001). Using and reading scholarly literature. *Annual Review of Information Science and Technology*, 34, 423-477.
- Kuruppu, P. U., & Gruber, A. M. (2006). Understanding the information needs of academic scholars in agricultural and biological sciences. *Journal of Academic Librarianship*, 32(6), 609-623.
- Leydesdorff, L. (2010). Eugene Garfield and algorithmic historiography: Co-words, co-authors, and journal names. *Annals of Library and Information Studies*, 57(3), 248-260.
- Lucio-Arias, D., & Leydesdorff, L. (2008). Main-path analysis and path-dependent transitions in HistCite™-based historiograms. *Journal of the American Society for Information Science and Technology*, 59(12), 1948-1962. <http://dx.doi.org/10.1002/asi.20903>
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56-63.
- Murphy, J. (2003). Information-seeking habits of environmental scientists: A study of interdisciplinary scientists at the Environmental Protection Agency in Research Triangle Park, North Carolina. *Issues in Science and Technology Librarianship*, 38. <http://dx.doi.org/10.5062/F4PV6HBW>
- Price, D. J. de S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science* 27(5), 292-306.
- Tenopir, C., King, D. W., Spencer, J., & Wu, L. (2009). Variations in article seeking and reading patterns of academics: What makes a difference? *Library and Information Science Research*, 31(3), 139-148. <http://dx.doi.org/10.1016/j.lisr.2009.02.002>
- Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, 10(1), 93-115.

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Kwon, Nahyun, Lee, Jungyeoun, & Chung, Eunkyung (2013). Understanding scientific research lifecycle: Based on bio- and nano- scientists' research activities. *Journal of the Korean Society for Library and Information Science*, 46(3), 103-131.
<http://dx.doi.org/10.4275/KSLIS.2012.46.3.103>
- Lee, Jae Yun (2013.2.17.). Informetrics and information management practice. Paper presented at the 2013 Workshop for Information Management and Network Analysis, Pusan, Korea.
- Lee, Jae Yun, & Kim, Heejung (2011). Bibliometric analysis for the research support service at International Vaccine Institute. *Proceedings of the 18th Conference of the Korean Society for Information Management*, 11-16.
- Lee, Jae Yun, Kim, Pan-Jun, Kang, Dae-Shin, Kim, Hee-Jung, Yu, So-Young, & Lee, Woo-Hyoung (2011). A bibliometric analysis on LED research. *Journal of Information Management*, 42(3), 1-26. <http://dx.doi.org/10.1633/JIM.2011.42.3.001>
- Lee, Jungyeoun, Chung, Eunkyung, & Kwon, Nahyun (2012). Scientist's information behavior for bridging the gaps encountered in the process of the scientific research lifecycle. *Journal of the Korean Society for Information Management*, 29(3), 99-122.
<http://dx.doi.org/10.3743/KOSIM.2012.29.3.099>
- Yu, So-Young, & Lee, Jae Yun (2008). Journal citation analysis for library services on interdisciplinary domains: a case study of Department of Biotechnology, Y University. *Journal of the Korean Society for Information Management*, 25(4), 283-308.
<http://dx.doi.org/10.3743/KOSIM.2008.25.4.283>