

자기 인용 네트워크와 인용 정체성을 이용한 연구자의 연구 이력 분석에 관한 연구

Exploring a Researcher's Personal Research History through Self-Citation Network and Citation Identity

이재윤(Jae Yun Lee)*

초 록

이 연구에서는 연구자의 연구 이력을 분석하기 위해서 White(2000)가 제안한 인용 정체성과 Hellsten 등(2007)이 제안한 자기 인용 네트워크의 두 가지 최신 기법을 비교해보았다. 국내 대표적인 정보학자인 정영미의 연구 성과물을 대상으로 실험적인 분석을 수행해본 결과 두 기법에서 구분한 연구 시기가 동일하게 나뉘었으며 주요 연구 주제도 유사하게 파악되었다. 그러나 인용 정체성 지도에서는 주제영역별로 영향받은 주요 저자를 파악할 수 있는 반면에 자기 인용 네트워크에서는 시기별 핵심 문헌과 선도 문헌이 식별되었다. 따라서 이 두 가지 기법을 상호보완적으로 사용할 때 연구자의 연구 이력에 대해서 풍부한 정보를 획득할 수 있다는 결론을 얻었다.

ABSTRACT

This paper compares two recent methods for exploring a scientist's research history: citation identity and self-citation network. The former is proposed by White(2000), while the latter is suggested by Hellsten et al.(2007). An experimental citation analysis was carried out on the research output of Young Mee Chung, a renowned Korean information scientist. The result shows that the two methods divided the research period into two sub-periods in the same way. They also identified the major research themes very similarly. In the analysis of each method's performance in depth, the two methods revealed different functions to understand a researcher's history. Citation identity was useful to identify authors who have affected Chung's research in terms of research topics, whereas, self-citation network was successful to identify the core papers and leading papers of the research sub-periods. This study indicates the combination of two methods can provide rich information on a scientist's research history.

키워드: 인용 분석, 인용 정체성, 자기 인용, 인용 네트워크, 저자동시인용분석, 정영미, 연구자이력
citation analysis, citation identity, self-citation, citation network,
author cocitation analysis, Young Mee Chung, personal research history

* 경기대학교 문헌정보학과 부교수(memexlee@kgu.ac.kr)

■ 논문접수일자: 2012년 2월 19일 ■ 최초심사일자: 2012년 3월 10일 ■ 게재확정일자: 2012년 3월 21일
■ 정보관리학회지, 29(1), 157-174, 2012. [http://dx.doi.org/10.3743/KOSIM.2012.29.1.157]

1. 서론

인용 분석은 연구 동향이나 연구 영역을 분석하기 위한 수단으로 오랫동안 활용되어 왔다. 최근에는 연구 성과 평가를 위한 도구로도 인용 분석이 주목받고 있다. E. Garfield가 인용 색인 데이터베이스를 구축한 이후 현재까지 특정 학문 분야와 같은 주제 측면이나 연구자 집단과 같은 연구 주제 측면에 대해서 다양한 인용 분석 기법이 개발되었다. 특히 동시인용이나 서지결합, 프로파일링과 같은 연구 동향 및 주제 분석 기법에서 연구자 개인은 분석의 대상이기 보다는 분석을 위해 측정되는 변수로 주로 사용되어왔다.

저자동시인용분석 방법을 제안한 White(2000)는 E. Garfield의 연구를 대상으로 분석한 논문에서 자아 중심 인용 분석(egocentric citation analysis)을 제안하였다. 이는 인용을 통해 특정 연구자 개인의 학문세계를 구성해온 연구 이력을 분석하는 수단으로서 연구자의 인용 정체성(citation identity)을 비롯한 네 가지 정보를 도출하는 기법을 포함하고 있다. 인용 정체성은 특정 연구자가 인용한 저자의 목록을 의미하며 이를 통해 연구자가 어떤 저자로부터 영향을 받았는지를 파악할 수 있다.

한편 Hellesten, Lambiotte, Scharnhorst, Ausloos(2007)는 연구자 자신의 문헌들 사이의 인용 관계로 구성되는 자기 인용 네트워크 분석을 제안하였다. 이들은 자기 인용 네트워크를 분석하여 연구자의 연구 주제가 변화하는 것을 파악할 수 있다고 주장하였다.

특정 분야의 대표적인 연구자에 대해서 연구 이력을 분석한다면 그 결과는 개인에 대한 연

구에 그치지 않고 해당 분야의 성립과 발전에 대한 연구가 될 수 있다. 연구자의 인용 정체성과 자기 인용 네트워크는 모두 특정 연구자의 연구 이력을 분석할 수 있는 도구로서, 연구자의 연구 성과물 전체로부터 인용 정보를 추출하면 분석에 필요한 데이터를 구축할 수 있다. 다만 인용 정체성은 저자를 단위로 도출되는데 반해서 자기 인용 네트워크는 문헌을 단위로 구성되는 것이 다르다. 두 도구는 모두 개인 연구자의 연구 이력을 분석하는 것을 목적으로 하고 있으나 두 방법을 함께 검토해본 시도는 없었으며, 어느 한 방법을 개별적으로 적용해본 연구도 아직은 드문 실정이다.

이 연구에서는 개인 연구자에 대한 인용 분석이 국내에서 가능한지 여부를 검토하고 자기 인용 네트워크와 인용 정체성을 함께 적용하여 두 도구의 유사점과 차이점을 파악해보고자 한다. 이를 위해서 국내 정보학 분야의 대표 연구자인 정영미가 발표한 문헌의 인용 정보를 수집한 다음 정영미의 자기 인용 네트워크와 인용 정체성을 도출하고 분석해보았다.

2001년부터 2004년까지 문헌정보학 분야 국제 학술지에 논문을 발표한 전세계 저자 1,361명을 분석한 Davarpanah와 Aslekia(2008)에 따르면 정영미는 논문 수 기준으로 공동 8위에 해당하였으며 아시아 지역 문헌정보학 연구자 중에서는 1위를 차지하였다. 또한 국내 4년제 대학 문헌정보학과에 재직중인 교수들의 2001년 이후 10년 동안의 연구 업적을 평가한 이종욱과 양기덕(2011)의 연구에서는 국내외 논문 수 1위인 연구자 P1이 국내 학술대회를 제외하면 63편의 논문을 발표한 것으로 보고하였는데, 본 논문에서도 동일한 기준으로 정영미의 논문

수를 파악한 결과 63편이 파악되었으므로 국내 문헌정보학 연구자 중에서도 정영미가 2001년 이후 가장 많은 논문을 발표한 것으로 추정된다. 정보학 분야의 대표적인 학회인 정보관리 학회가 창립되기 이전부터 관련 분야를 연구하고 가장 많은 논문을 발표한 정영미의 연구 이력을 분석하는 것은 국내 정보학 분야의 발전 사에 대한 연구에도 큰 도움이 될 것이다.

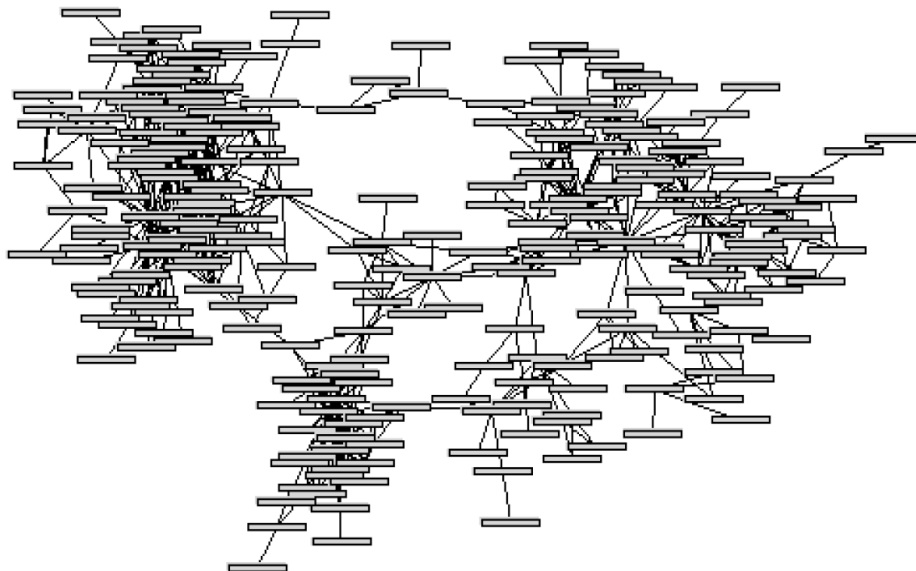
2. 관련 연구

2.1 자기 인용 네트워크

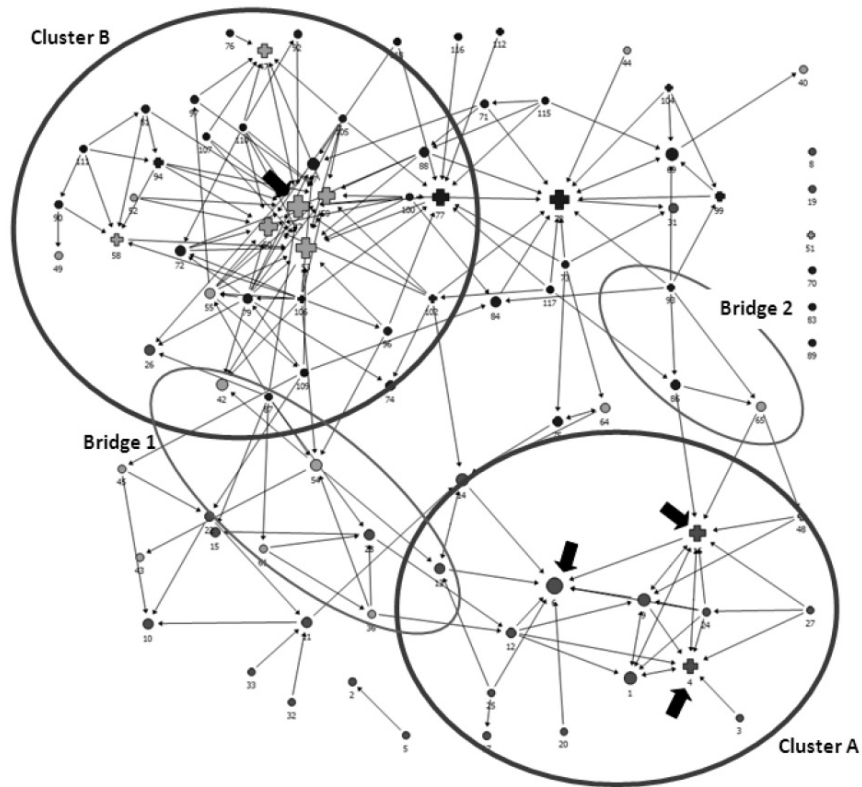
자기 인용은 연구자가 자신의 기존 성과물을 후속 연구에서 인용하는 것을 말한다. 일반적으로 연구 성과 평가를 다루는 연구 정책 분야

에서는 자기 인용을 바람직하지 않은 것으로 간주하지만, 정보학이나 계량서지학 분야에서는 자기 인용도 유용하게 활용되어왔다(Glänzel, Debackere, Thijs, & Schubert, 2006).

Hellesten, Lambiotte, Scharnhorst, Ausloos(2007)는 연구자의 연구 영역 이동(field mobility)을 파악하기 위한 용도로 연구자의 성과물을 노드로 하고 자기 인용 관계를 링크로 하는 자기 인용 네트워크를 활용할 것을 제안하였다. 이들은 연구자 개인만 아니라 연구 집단을 대상으로 자기 인용 네트워크를 분석할 수 있다고 주장하였으며, 실험적으로 물리학자인 W. Ebeling을 대상으로 자기 인용 네트워크를 <그림 1>과 같이 구축하였다. 이에 대한 분석 결과 Ebeling의 연구를 크게 3개 영역으로 나눌 수 있었으며, 각 영역의 문헌에서 자주 등장하는 키워드를 단서로 하여 연구 주제를 해석할 수 있었다.



<그림 1> Hellesten과 Lambiotte, Scharnhorst, Ausloos(2007)가 구축한 W. Ebeling의 자기 인용 네트워크 중 최대 컴포넌트



<그림 2> 기지훈(2011)이 분석한 경제학자 이근의 자기 인용 네트워크

기지훈(2011)은 자기 피인용 횟수 분석, 자기 인용 네트워크 클러스터 파악, 네트워크 진화 양상 분석 등을 통해서 대상 학자의 학문적인 이론이 어떤 발전과정을 밟아왔는지를 종합적으로 파악할 수 있다고 주장하였다. 그는 1989년에 박사학위를 받은 국내 현역 경제학자인 이근을 대상으로 자기 인용 네트워크를 <그림 2>와 같이 구축하고 분석하였다. 분석 결과에서는 이근의 초기 연구 문헌과 최근 연구 문헌이 구분되었으며 최근 연구 문헌들의 시초라고 할 수 있는 2001년 논문이 파악되었다.

2.2 인용 정체성

인용 정체성(citation identity)은 White(2000)가 연구자 개인에 대한 계량서지적 연구 수단으로 제시한 4가지 분석 방법 중 하나이다. White는 사회 네트워크 분석 분야에서 정립된 자아 중심 네트워크 분석(ego-centered network analysis)에서 착안하여 자아 중심 인용 분석을 제안하면서 이를 통해서 특정 연구자의 인용 정체성, 인용 이미지(citation image), 인용 이미지 구축자(citation image makers), 공동 연구자(co-authors)의 네 가지 저자 목록을 파악할 수 있다고 밝혔다. 각각의 기본 개념은 다음과 같다.

- 인용 정체성: 연구자 자신(ego)에 의해서 인용되는 저자 집합(alters)
- 인용 이미지: 연구자와 동시인용된 저자 집합
- 인용 이미지 구축자: 연구자를 인용하는 저자 집합
- 공저자: 연구자와 공저한 저자 집합

이중에서 분석 대상인 연구자의 연구 주제를 파악하는데 직접 도움이 되는 것은 인용 정체성과 인용 이미지이다. 연구자 A의 인용 정체성을 구성하는 저자들은 A의 논문에서 A 본인이 직접 인용하기로 선택한 저자들이며, 공저자는 A와 함께 논문을 작성함으로써 A의 인용 정체성 구성에 기여한 사람들이다. 이와 비교할 때 연구자 A의 인용 이미지를 구성하는 인용은 다른 사람들이 A를 인용한 논문에서 함께 인용하기로 선택한 저자들이며, 인용 이미지 구축자는 A를 인용함으로써 A의 인용 이미지 형성에 기여한 사람들이다. 다시 말하면, 인용 정체성은 연구자 본인의 생각 속에 있는 자신의 모습이고 인용 이미지는 타인의 눈에 비친 연구자의 모습이라고 할 수 있다.

White(2000)는 E. Garfield를 대상으로 자아 중심 인용 분석을 수행하면서 저자 정체성을 비롯한 네 가지 저자 목록을 제시하였다. White가 제시한 네 목록에서 1위는 모두 자아(ego)인 Garfield였지만 2위는 각각 다른 저자들이 차지하여서 자아 중심 인용 분석으로 도출된 각 목록이 서로 다른 정보를 전달하고 있음이 드러났다.

연구자 개인 단위의 분석을 제안한 White(2000)의 방법론은 이후에 저널 단위의 분석으로 확장

되었다. Nebelong-Bonnevie와 Frandsen(2006)은 Journal of Documentation을 대상으로 White가 제안한 네 가지 목록 중에서 인용 정체성과 인용 이미지를 저널 단위로 분석하였으며, Bonnevie-Nebelong(2006)은 Journal of Documentation을 비롯한 3종 저널의 인용 정체성과 인용 이미지를 비교하였다. 최근 Tsay(2011)도 유사한 방법을 사용하여 저널 간 비교를 수행하였다.

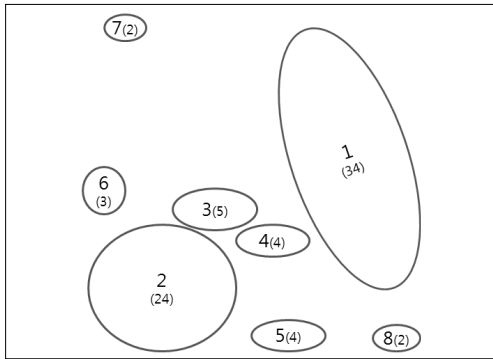
3. 자기 인용 네트워크 분석

3.1 자기 인용 네트워크의 구성

정영미가 저술한 연구 문헌을 조사해본 결과 박사학위논문 1편, 국내의 학술지 논문 72편, 학술대회 발표논문 37편, 기념논문집 2편, 단행본 10권으로 총 122건이 파악되었다. 이 문헌들의 참고문헌에 인용된 저자를 분석하기 위해서 저자명의 한글 표기와 영문 표기를 검토하여 동일 저자의 이표기를 모두 통일시켰다.

정영미의 자기 인용 네트워크를 구축하기 위해서 수집된 연구 문헌들 사이의 인용 관계를 분석하였다. 122건 중에서 자기 인용을 받았거나 자기 인용을 내보낸 문헌은 총 78편이었다. 여기에는 학술지/학술대회에 발표한 논문 69편과 박사학위논문 1편, 단행본 8편이 포함되었다. 파악된 자기 인용 네트워크를 시각화하면 <그림 3>과 같다.

<그림 3>의 자기 인용 네트워크에서는 1977년에 발표된 박사학위논문으로부터 2010년에 발표된 5편의 논문까지 각 문헌을 발표 연도 순으로 위에서 아래로 배열하였다. 자기 인용 관



〈그림 4〉 그림 3의 자기 인용 네트워크의 컴포넌트 위치와 번호(괄호안의 숫자는 컴포넌트의 크기)

규모가 큰 컴포넌트 2개는 정영미의 전기 연구와 후기 연구를 구분하는 단서가 될 수 있다. 1993년에 발간된 『정보 검색론』(개정판)을 중심으로 하는 제1 컴포넌트에는 1978년에 발표된 “정보이론과 문헌정보검색”을 시작으로 하여 1990년대 초반까지 발표된 문헌이 대부분 연결되어 있다. 여기에는 정보이론과 자동색인 이론으로부터 시작되어 도서관전산화 관련 연구가 중간에 추가된 것을 볼 수 있다. 제1 컴포넌트의 연구는 2000년대 초반까지 자동요약 연구가 이어졌으며 지식 구조에 대한 관심은 『지식구조론』으로부터 2010년까지 이어졌음을 알 수 있다.

2005년에 발간된 『정보검색연구』를 중심으로 하는 제2 컴포넌트는 1993년에 발표된 “국어어휘의 통계적 특성과 이의 응용”으로부터 이어지지만 실질적인 시작은 8년 후인 2001년에 발표된 “A corpus-based approach to comparative evaluation of statistical term association measures”임을 알 수 있다. 2001년을 경계로 해서 제1 컴포넌트에 연결되는 문헌보

다 제2 컴포넌트에 연결되는 문헌이 많아진다.

이와 같이 자기 인용 네트워크에서 식별되는 대표적인 두 컴포넌트가 각각 우세한 시기를 나눠보면 정영미의 연구는 1977년부터 2000년까지를 전기, 2001년부터 현재까지를 후기로 구분해볼 수 있다. 전기 연구는 정보이론, 자동색인, 도서관전산화, 검색순위화, 지능형 시스템, 네트워크, 온라인 검색 등이 주요 주제로 파악되며, 후기 연구는 통계적 연관성, 자동 요약, 단어 중의성 해소, 자동 분류, 텍스트 마이닝, 웹 마이닝, 계량서지학, 웹계량학(Webometrics) 등이 주요 주제로 나타났다.

3.2 자기 인용 네트워크의 중요 문헌 파악

자기 인용 네트워크에서 정영미의 후속 연구 문헌들에게 영향을 끼친 주요 문헌을 파악하기 위해서 각 문헌의 인용 빈도를 비교해보았다. 자기 인용 네트워크에서 인용 빈도가 가장 높은 문헌은 2005년에 발표된 단행본인 『정보검색연구』로 정영미의 연구 문헌들로부터 10회 인용되었으며, 공동 2위는 2001년에 JASIST에 발표된 논문인 “A corpus-based approach to comparative evaluation of statistical term association measures”와 1993년에 발표된 단행본인 『정보 검색론』(개정판)으로서 5회 인용되었다. 3회 이상 인용된 6개 문헌은 〈표 1〉과 같으며 2회 인용된 문헌은 13건이었다.

인용 빈도는 직접 인용 관계만 파악한 것이므로 인용 네트워크의 구조에 나타나있는 간접적인 영향 관계를 반영하지 못한다. 두 단계 이상의 인용 연결을 통해 전달되는 간접적인 영

〈표 1〉 자기 인용 빈도 3회 이상 문헌

순위	자기인용 빈도	연도	제목	비고(출처)	소속 컴포넌트
1	10	2005	정보검색연구	단행본	제2
2	5	2001	A corpus-based approach to comparative evaluation of statistical term association measures	JASIST	제2
2	5	1993	정보 검색론(개정판)	단행본	제1
4	4	1982	자동색인의 통계적 기법과 한국어 문헌의 실험	도서관학	제1
5	3	1987	우리말 신문기사 검색을 위한 질문응답시스템 구현에 관한 연구	정보관리학회지	제1
5	3	2009	Characteristics and link structure of a national scholarly Web space: The case of South Korea	Scientometrics	제5

향력까지 고려하는 네트워크 분석 방식의 단 일 문헌 인용 영향력 지표가 여러 가지 제안된 바 있다(이재운, 2011). 여기에서는 분석 대상 연구자의 문헌이 아닌 타 연구자의 문헌들과 주고 받는 인용을 고려하지 않아도 되는 f-값 공식(Fragkiadaki, Evangelidis, Samaras, & Dervos, 2011)을 적용하였다. f-값 공식은 1세대 인용빈도(직접 인용빈도)와 2세대 인용빈도(간접 인용빈도)를 고르게 반영하는 특성을 가지고 있다(이재운, 2011). f-값 공식에서는 해당 문헌을 인용하는 문헌들의 가중치를 모두 합한

값에 일정한 감쇄상수 RF(reducing factor)를 곱해서 인용받는 문헌의 가중치로 삼는다. 처음 제안된 논문에서와 같이 감쇄상수를 1/2.2로 설정한 f-값 공식은 다음과 같다. 여기서 $W(d_i)$ 는 인용받는 논문의 영향력이고 $W(d_j)$ 는 인용하는 논문의 영향력이다.

$$W(d_i) = 1 + \frac{1}{2.2} \sum_j W(d_j)$$

자기 인용 네트워크에서 구한 f-값 상위 10개 문헌은 〈표 2〉와 같다. 간접 인용까지 고려

〈표 2〉 f-값 측정 결과 상위 10개 문헌

순위	f-값	자기인용 빈도	발표 연도	제목	비고	소속 컴포넌트
1	6.68	5	2001	A corpus-based approach to comparative evaluation of statistical term association measures	JASIST	제2
2	5.75	10	2005	정보검색연구	단행본	제2
3	4.45	4	1982	자동색인의 통계적 기법과 한국어 문헌의 실험	도서관학	제1
4	4.07	2	2001	클러스터링 성능 평가를 위한 비편향적 척도의 개발	한국정보관리학회 학술대회	제2
4	4.07	2	2004	Optimization of some factors affecting the performance of query expansion	IPM	제2
6	4.04	1	1993	국어어휘의 통계적 특성과 이의 응용	인문과학	제2
7	3.69	5	1993	정보 검색론(개정판)	단행본	제1
8	3.54	2	1980	색인이론과 실제	연세논총	제1
8	3.54	2	1983	Shannon 정보이론의 도서관학·정보학적 해석에 관한 연구	연세논총	제1
8	3.54	2	1982	도서관정보전산화론	단행본	제1

하여 후속 연구에 대한 영향력을 측정된 결과 1위는 2001년에 JASIST에 발표된 “A corpus-based approach to comparative evaluation of statistical term association measures”였으며 단행본인 『정보검색연구』는 2위로 나타났다. 특이하게도 자기 인용 빈도가 1회에 불과한 “국어어휘의 통계적 특성과 이의 응용”이 f-값 기준으로는 6위로 높게 측정되었다. 이 논문은 1993년에 발표된 이후 2001년에 f-값 기준 1위인 논문으로부터 인용되어서 그 이후의 여러 후기 연구 문헌들에게 간접적으로 영향을 주었기 때문이라고 생각된다. 이와 같이 후속 연구에 큰 영향력을 끼친 선행 문헌은 선도 문헌이라고 부를 수 있겠다.

4. 인용 정체성 분석

4.1 인용 정체성 구성 저자 분석

전체 분석대상 122건중 단행본과 박사학위 논문을 제외한 논문 111편에서 인용된 저자는 정영미의 연구에 영향을 준 저자 집단으로서

인용을 통해 파악된 정영미의 연구 정체성, 즉 인용 정체성을 대변한다. 학술지 논문 72편과 학술대회 발표논문 37편, 그리고 기념논문집 논문 2편에서 인용된 저자는 총 1,744명이었다. 인용 정체성을 구성하는 주요 저자를 선정하기 위해서 인용된 문헌 수별 저자 수를 <표 3>과 같이 파악해보았다. 인용된 문헌 수가 5건일 때와 4건일 때 사이가 저자 수의 차이가 급격하게 커지는 지점이어서 5건 이상의 문헌에서 인용된 저자 53명을 주요 저자로 선정하였다.

5개 이상의 문헌에서 정영미가 인용한 저자 53명에 대해서 인용된 문헌 수와 총 인용 횟수를 파악한 결과는 <표 4>와 같다. 인용된 문헌 수는 한 문헌에서 여러 차례 인용되더라도 1회로 간주한 것이지만 총 인용 횟수는 한 문헌에서 해당 저자의 문헌이 몇 건 인용되었는지를 그대로 반영하여 합산한 것이다. 인용된 문헌 수 1위는 자기 인용에 해당하는 정영미였다. Garfield의 인용 정체성을 분석한 White(2000)의 연구에서도 Garfield가 자신을 인용한 횟수가 2위에 비해서 압도적으로 높게 나타났다. 2위와 3위는 정보검색 분야의 대표적인 연구자인 Salton과 Croft였으며 자동분류 분야의 주요 연구자인

<표 3> 정영미의 논문에서 인용된 저자들의 인용된 문헌 수에 따른 저자 수

인용된 문헌 수	저자 수	누계	인용된 문헌 수	저자 수	누계
47	1	1	8	4	16
23	1	2	7	6	22
21	1	3	6	12	34
15	2	5	5	19	53
14	1	6	4	38	91
13	1	7	3	98	189
12	1	8	2	259	448
11	2	10	1	1,296	1,744
9	2	12			

〈표 4〉 5개 이상의 문헌에서 정영미가 인용한 인용 정체성 저자 53명

인용 정체성 저자	인용된 문헌 수	총 인용횟수	인용 정체성 저자	인용된 문헌 수	총 인용횟수
정영미	47	57	Ingwersen P.	6	7
Salton G.	23	41	McGill M.J.	6	7
Croft W.B.	21	32	Singhal A.	6	7
Pederson J.	15	19	Thelwall M.	6	39
Yang Y.	15	27	Voorhees E.	6	7
이재윤	14	19	Weaver W.	6	8
Sparck Jones K.	13	16	최기선	6	11
Robertson S.	12	16	Brooks H.M.	5	7
Buckley C.	11	13	Church K.W.	5	11
van Rijsbergen C.J.	11	15	Cronin B.	5	5
Bookstein A.	9	12	Gale W.	5	11
Lancaster F.W.	9	10	Harper D.J.	5	5
Belkin N.J.	8	11	Humphrey S.	5	6
Luhn H.P.	8	9	Lewis D.D.	5	6
Radev D.	8	10	Liu X.	5	5
Willett P.	8	11	Manning C.D.	5	5
Callan J.	7	8	Maron M.E.	5	5
Sahami M.	7	7	Meadow C.T.	5	6
Saracevic T.	7	7	Noreault T.	5	5
Schutze H.	7	9	Rau L.F.	5	7
Wu H.	7	8	Shannon C.E.	5	5
Xu J.	7	7	Spink A.	5	7
Allan J.	6	11	Thompson R.	5	5
Borko H.	6	6	Yarowsky D.	5	15
Brookes B.C.	6	10	김명철	5	6
Chen H.	6	9	김용학	5	5
Fox E.A.	6	12			

Pederson과 Yang이 15회로 공동 4위를 차지하였다. 53명의 저자 중에서 정영미 자신을 제외하면 국내 연구자는 4명에 불과했다. 이 중에는 문헌정보학 분야의 이재윤, 컴퓨터과학 분야의 최기선과 김명철, 사회학 분야의 김용학이 포함되었다.

대체로 인용된 문헌 수와 총 인용 횟수는 비례하지만 예외적으로 Thelwall의 경우 6개 문헌에서 인용되어 공동 23위인 반면, 총 인용 횟

수는 39회로 3위였다. 이는 정영미의 문헌 한 편당 평균 6건 이상의 Thelwall의 문헌을 인용한 셈이어서 소수 문헌에서 매우 집중적으로 인용되었음을 나타낸다. 2001년부터 2010년까지의 계량정보학 연구를 분석한 Choi와 Lee(2011)의 연구에 따르면 Thelwall은 웹계량학 분야의 대표 연구자로서 계량정보학 전체를 통틀어서 가장 많은 논문을 발표한 연구자 중 한 사람이다. 앞 장에서 분석한 정영미의 후기 주요 연구

주제중 하나인 웹계량학 분야 연구가 Thelwall의 연구에 많은 영향을 받았음을 알 수 있다.

4.2 인용 정체성 지도 분석

인용 정체성 저자 53명 사이의 관계를 파악하면 정영미의 인용 정체성을 시각적으로 표현할 수 있다. 이를 위해서 53명 저자들이 정영미의 연구문헌에서 동시 인용된 횟수를 <표 5>와 같이 파악하였다. 저자 동시인용 행렬의 동시 인용빈도를 분자로 하고 각 저자의 인용빈도의 곱에 제곱근을 취한 것을 분모로 하여 저자 간 코사인 유사도를 구한 다음 53명 저자 사이의 코사인 유사도 행렬을 생성하였다. 코사인 유사도 행렬을 입력 데이터로 하여 네트워크 시

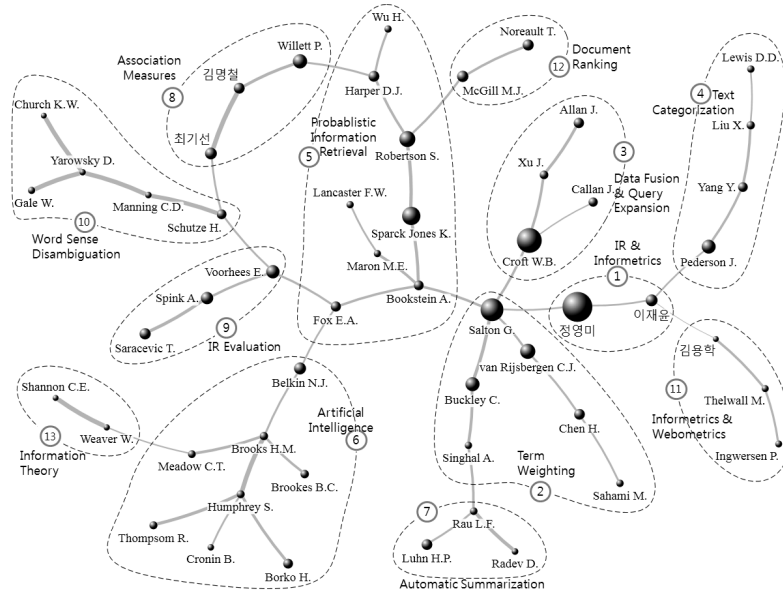
각화 방법인 패스파인더 네트워크(Schvaneveldt, 1990)를 구축하였다.

생성된 패스파인더 네트워크에 포함된 저자를 분류해서 연구 주제를 해석하기 위하여 군집분석의 일종인 PNNC 기법(이재윤, 2006)을 적용하였다. 그 결과 패스파인더 네트워크가 <그림 5>와 같이 13개의 작은 군집으로 분할된 정영미의 인용 정체성 지도가 완성되었다. 이 지도에서 군집 1은 정영미가 포함되었으므로 연구자 자신(ego)을 의미하는 것으로 볼 수 있다. 군집 2부터 군집 13까지 각 군집의 연구 주제는 해당 연구자들이 동시에 인용된 논문의 제목을 참고하여 지정하였다.

패스파인더 네트워크와 PNNC 군집으로 분석된 인용 정체성 지도에서 다른 군집과 많이 연결

<표 5> 정영미의 인용 정체성 저자들 사이의 동시인용 빈도 행렬(일부분)

저자	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	
정영미	A1	47	14	13	10	8	11	10	6	8	9	7	4	5	2	4	7	4	3	5	5	3	3
Salton	A2	14	23	12	6	4	4	9	8	11	8	8	2	4	6	3	6	2	2	2	2	5	6
Croft	A3	13	12	21	4	5	4	7	6	6	5	5	0	5	1	0	5	4	2	4	4	4	7
Pederson	A4	10	6	4	15	8	6	2	2	5	4	0	0	0	2	3	4	0	4	2	4	0	1
Yang	A5	8	4	5	8	15	1	1	2	1	3	0	1	0	1	2	1	3	3	1	0	2	2
이재윤	A6	11	4	4	6	1	14	2	2	4	3	0	0	0	0	4	0	2	2	4	0	2	2
Sparck Jones	A7	10	9	7	2	1	2	13	8	3	5	7	2	3	2	2	4	1	1	1	2	3	1
Robertson	A8	6	8	6	2	2	2	8	10	3	4	4	1	2	1	1	4	0	1	1	2	3	2
Buckley	A9	8	11	6	5	1	4	3	3	11	2	1	0	1	3	3	4	1	1	2	2	1	4
van Rijsbergen	A10	9	8	5	4	3	3	5	4	2	11	4	1	2	1	0	4	1	2	1	2	0	0
Bookstein	A11	7	8	5	0	0	0	7	4	1	4	9	2	2	2	0	2	0	0	0	0	3	0
Lancaster	A12	4	2	0	0	1	0	2	1	0	1	2	9	0	3	0	0	1	0	0	0	0	0
Belkin	A13	5	4	5	0	0	0	3	2	1	2	2	0	8	0	0	1	2	0	1	0	0	1
Luhn	A14	2	6	1	2	1	0	2	1	3	1	2	3	0	8	3	0	1	0	0	0	1	1
Radev	A15	4	3	0	3	2	0	2	1	3	0	0	0	0	3	6	0	1	0	0	0	0	0
Willett	A16	7	6	5	4	1	4	4	4	4	4	2	0	1	0	0	8	0	2	1	3	1	2
Callan	A17	4	2	4	0	3	0	1	0	1	1	0	1	2	1	1	0	6	0	2	0	1	2
Sahami	A18	3	2	2	4	3	2	1	1	1	2	0	0	0	0	0	2	0	7	0	1	0	0
Saracevic	A19	5	2	4	2	1	2	1	1	2	1	0	0	1	0	0	1	2	0	7	2	0	2
Schutze	A20	5	2	4	4	0	4	2	2	2	2	0	0	0	0	0	3	0	1	2	7	0	1
Wu	A21	3	5	4	0	2	0	3	3	1	0	3	0	0	1	0	1	1	0	0	0	7	2
Xu	A22	3	6	7	1	2	2	1	2	4	0	0	0	1	1	0	2	2	0	2	1	2	7



〈그림 5〉 정영미의 인용 정체성 지도(패스파인더 네트워크에 군집과 주제명을 표시함)

되어 중심적인 위치에 자리한 저자 군집은 확률적 정보검색(Probabilistic Information Retrieval)에 관한 5번 군집(주변 6개 군집과 이어짐)과 용어 가중치(Term Weighting)에 관한 2번 군집(주변 4개 군집과 이어짐)이었다. 핵심적인 두 군집을 비롯하여 가운데 오른쪽의 12번 군집(문헌 순위화-Document Ranking)과 3번 군집(데이터 결합 및 질의확장-Data Fusion & Query Expansion), 왼쪽 위의 8번 군집(연관성 척도-Association Measures), 9번 군집(정보검색 평가-IR Evaluation), 10번 군집(단어의미 중의성해소-Word Sense Disambiguation)은 모두 정보검색과 직/간접적으로 관련된 주제들이다. 오른쪽 위의 4번 군집(텍스트 범주화-Text Categorization)과 아래쪽의 7번 군집(자동 요약-Automatic Summarization)까지 광의의 정보검색 영역이라고 간주하면 13개의 군집 중에서 10개가 정보검색과 관련된 주제들이다. 정

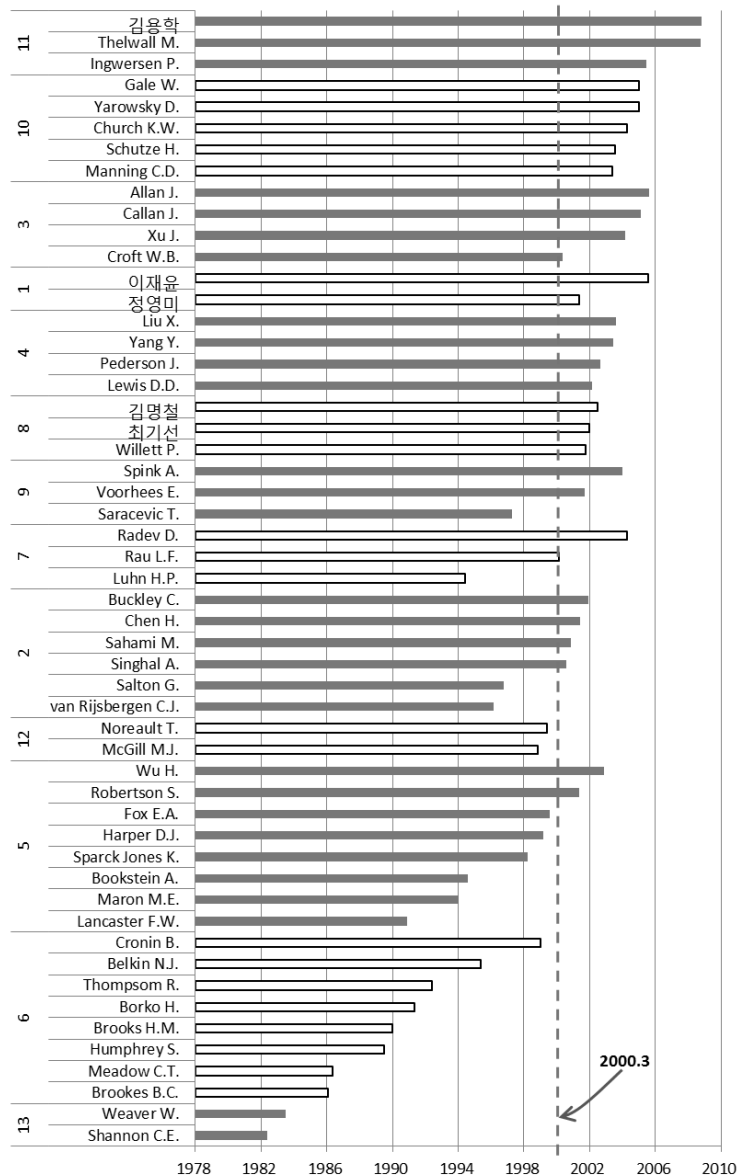
보검색과 다소 구분되는 주제 군집으로는 지식 표현을 위주로 하는 인공지능과 관련된 6번 군집, 정보 이론에 관한 13번 군집, 계량정보학 중에서도 특히 웹계량학과 관련된 11번 군집이 파악되었다.

정영미의 인용 정체성을 구성하는 주요 저자들이 어느 시기의 연구에 영향을 끼쳤는지를 파악하기 위해서 각 저자를 인용한 논문들이 발표된 평균 연도를 산출해보았다. 예를 들어 Shannon의 경우 1979년에 2건, 그리고 1978년, 1983년, 1993년에 각각 1건의 논문에서 인용되어서 인용 연도의 평균은 1982.4년이 된다. 연도의 상한(2010)과 하한(1977)을 이용하여 평균 연도를 0에서 1사이의 값으로 정규화하면 성장 지수(이재윤 외, 2011)가 된다. 주요 저자 53명이 인용된 연도의 전체 평균은 2000.3년이였다. 인용 연도의 평균을 기준으로 정영미의 연구 이력을 나누면, 앞 장에서 자기 인용 네트워크의 주요

컴포넌트를 기준으로 나누었던 것과 동일하게 2000년까지를 전기로, 2001년부터를 후기로 구분할 수 있다.

PNNC 기법으로 파악한 13개 군집별로 저자

를 묶어서 각 저자의 평균 인용 연도를 나타내면 <그림 6>과 같다. 이 그림에서는 군집별 평균 인용 연도가 최근에 가까울수록 위쪽에 배치하였으며, 전체 평균 인용 연도인 2000.3년을 세로



<그림 6> 53명 저자들의 평균 인용년도 비교
(맨 왼쪽 숫자는 군집 번호이며 막대의 색상은 군집 구분을 위한 것임)

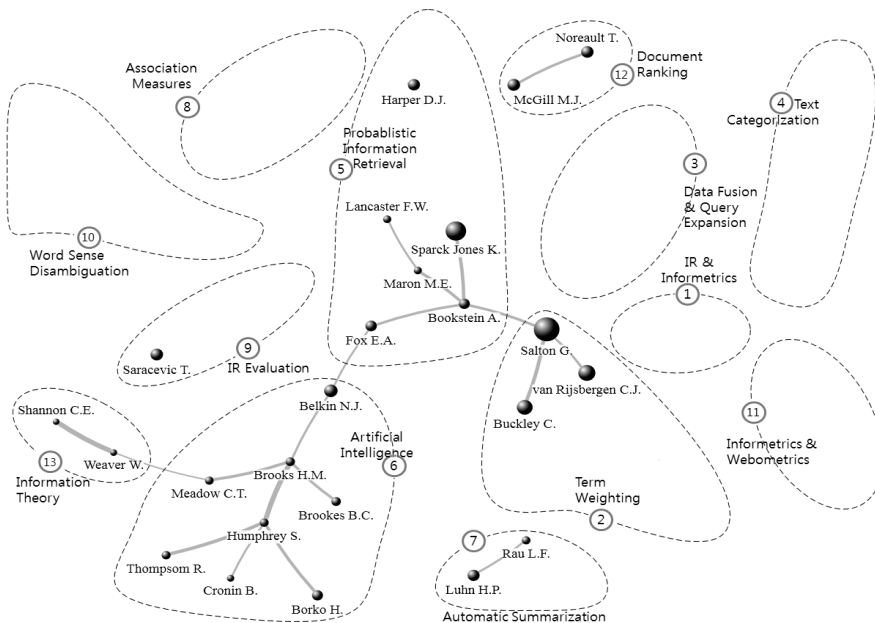
점선으로 표시하였다. 13개의 군집 중에서 소속 저자의 평균 인용 연도가 모두 전기에 해당하는 경우는 군집 13(정보 이론), 군집 6(인공 지능), 군집 12(문헌 순위화)의 세 주제였다. 이들은 전기에 연구가 집중적으로 이루어졌다고 볼 수 있다.

전기에 집중된 3개 주제와 달리 소속 저자의 평균 인용 연도가 모두 후기에 해당하여 2001년 이후에 연구가 집중된 주제는, 연구자 자신이 속한 군집 1을 제외하면 군집 11(계량정보학과 웹계량학), 군집 10(단어의미 중의성해소), 군집 3(데이터 결합과 질의확장), 군집 4(텍스트 범주화), 군집 8(연관성 척도)의 다섯 가지로 나타났다. 이들은 후기의 핵심 연구 주제라고 판단된다.

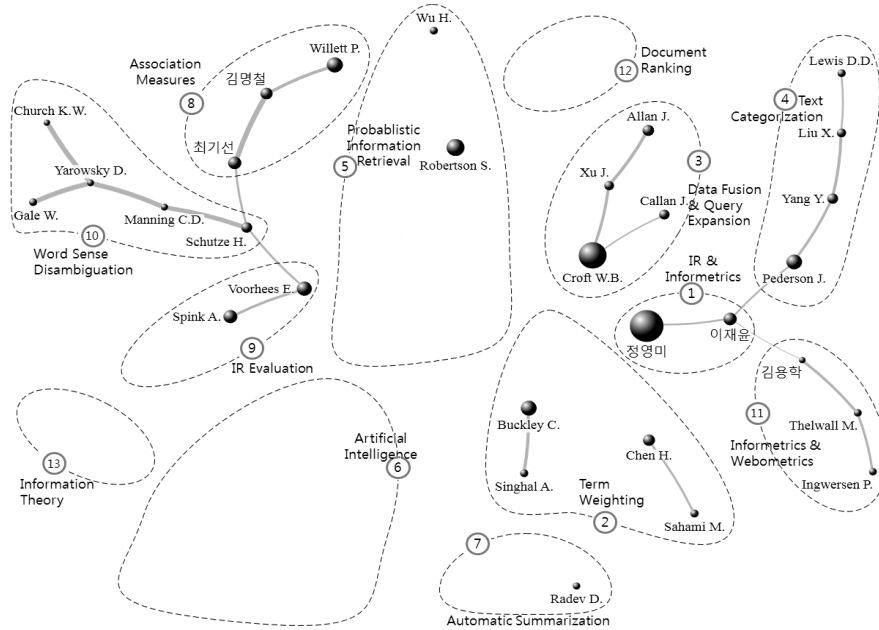
한편 소속 저자가 전기와 후기에 걸쳐 있으면서 평균 인용 연도의 차이가 크게 나는 군집은 1990.9년부터 2002.9년까지 12년의 차이가 나는 군집 11(확률적 정보검색)과, 1994.4년부터

2004.3년까지 9.9년의 차이가 나는 군집 8(자동 요약)의 두 주제였다. 이 두 군집의 주제는 다른 주제에 비해서 전기와 후기에 걸쳐서 10여년 이상 장기간 다루어진 연구 분야라고 할 수 있다. 반면에 전기와 후기에 걸쳐져 있지만 소속 저자의 평균 인용 연도의 차이가 크지 않은 군집 2(용어 가중치)와 군집 9(정보검색 평가)는 2000년을 전후한 중간 시기에 연구가 집중된 주제라고 판단된다.

2000.3년을 경계로 하여 전기와 후기를 구분하면 평균 인용 연도가 2000년 이전인 저자는 전기 주요 인용 저자, 평균 인용 연도가 2001년 이후인 저자는 후기 주요 인용 저자로 각각 간주할 수 있다. 앞 절에서 구축한 정영미의 인용 정체성 지도에서 전기 주요 인용 저자와 후기 주요 인용 저자를 각각 따로 표시해보면 <그림 7>, <그림 8>과 같다. 두 그림을 살펴보면 전기



<그림 7> 정영미의 인용 정체성 지도에 표시한 전기(1977-2000) 주요 인용 저자



〈그림 8〉 정영미의 인용 정체성 지도에 표시한 후기(2001-2010) 주요 인용 저자

주요 인용 저자는 대부분 인용 정체성 지도의 가운데에 위치하고 있으며 후기 주요 인용 저자는 대부분 왼쪽과 오른쪽에 치우쳐서 자리잡고 있다. 결국 정영미의 인용 정체성으로 파악되는 주요 연구 주제는 지도의 중앙에 위치한 주제부터 시작되어 최근에는 좌우의 연구 주제로 옮겨갔다고 판단된다.

5. 결론

개인 연구자의 연구 이력을 분석하기 위한 수단으로 자기 인용 네트워크와 인용 정체성 분석을 시도하였다. 분석 대상으로 대표적인 국내 정보학자인 정영미를 선정하고 연구 문헌을 수집하여 인용 정보를 추출하였다. 정영미의 연구 문헌 사이의 자기 인용 네트워크 그리

고 정영미가 인용한 저자들 사이의 동시인용으로 구성된 인용 정체성 지도를 구축하고 두 방법의 분석 결과를 각각 살펴보았다.

자기 인용 네트워크에서는 인용빈도 분석을 통해 연구자 본인이 많이 인용한 중요 문헌을 파악하는 한편, 인용 네트워크 분석을 통해 후속 연구의 시발점이 된 연구를 파악할 수 있었다. 또한 문헌이 연결된 주요 컴포넌트를 살펴봄으로써 2000년까지를 전기, 2001년 이후를 후기로 구분할 수 있었으며 각 시기의 주요 연구 주제를 파악할 수 있었다.

인용 정체성 분석에서는 주요 인용 저자 사이의 동시인용분석을 통해서 패스파인더 네트워크를 구축하고 13개의 저자 군집을 도출하여 주요 주제를 파악하였다. 특히 저자별 평균 인용 연도를 산출해본 결과 각 주제 분야가 다루어진 시기를 구체적으로 판단할 수 있었다. 또

한 주요 저자를 인용한 개별 연도의 전체 평균이 2000.3년으로 산출되어 이를 기준으로 하면 자기 인용 네트워크 분석에서 판단한 것과 일치하는 전/후기의 시기 구분이 이루어졌다.

두 방법을 비교해보면 자기 인용 네트워크에서는 중요 문헌과 선도 문헌을 파악할 수 있었으며, 인용 정체성 지도에서는 연구 주제별로 영향을 받은 주요 저자를 파악할 수 있었다. 이와 같은 분석은 모두 전기와 후기를 나누어서도 적용되어 시기별 중요/선도 문헌과 주요 인용 저자까지 파악이 가능하였다. 결국 자기 인용 네트워크와 인용 정체성 지도는 시기 구분과 연구 주제 분석이라는 측면에서는 유사한 정보를 제시하므로 상호 검증 기능을 가지고 있었다. 또한 연구자의 주요 문헌과 연구자에게 영향을 끼친 주요 저자를 각각 제시함으로써 두 방법이 상호 보완적인 역할을 한다는 점도 확인되었다.

이 연구에서는 자기 인용 네트워크와 인용 정체성 분석을 상호 보완적으로 적용하여 국내 연구자의 연구 이력을 해석할 수 있음을 보였다. 한편 White(2000)는 연구자를 인용한 사람들에 의해서 구축되는 인용 이미지도 인용 정체성에 못지않게 중요한 분석이라고 제시하였으나, 국내 논문 발표를 주로 하는 연구자를 대상으로 할 때에는 2001년 이전의 인용 정보가 구축된 데이터베이스가 없기 때문에 인용 이미지의 분석이 아직까지는 불가능하였다. 향후 국내 연구자에 대한 인용 데이터베이스가 보완된다면 인용 이미지의 분석도 시도되어야 할 것이다. 현재 구축된 국내 인용 데이터베이스의 구축 범위와 구축 형태를 감안하면 연구자 개인이 아닌 저널에 대한 인용 정체성과 인용 이미지의 분석이 더 용이할 것으로 판단되므로 이에 대한 접근도 우선적으로 모색되어야 한다.

참 고 문 헌

- 기지훈 (2011.11.18). 자기인용 네트워크: 학자의 학문세계 분석에 대한 새로운 접근. 제6회 COREN 컨퍼런스 발표자료, 경희대학교, 서울, 대한민국.
- 이재운 (2006). 지적 구조 분석을 위한 새로운 클러스터링 기법에 관한 연구. 정보관리학회지, 23(4), 215-231
- 이재운 (2011). 인용 네트워크 분석에 근거한 문헌 인용 지수 연구. 한국문헌정보학회지, 45(2), 119-143. doi: 10.4275/KSLIS.2011.45.2.119
- 이재운, 김관준, 강대신, 김희정, 유소영, 이우형 (2011). 계량서지적 기법을 활용한 LED 핵심 주제영역의 연구 동향 분석. 정보관리연구, 42(3), 1-26.
- 이종욱, 양기덕 (2011). 교수연구업적 평가법의 계량적 분석. 정보관리학회지, 28(4), 119-140. doi: 10.3743/KOSIM.2011.28.4.119

- Barrat, A., Bathelemy, M., & Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge, UK: Cambridge University Press.
- Bonnevie-Nebelong, E. (2006). Journal citation identity, journal citation image and internationalisation: Methods for journal evaluation. *Scientometrics*, 66(2), 411-424. doi: 10.1007/s11192-006-0029-9
- Davarpanah, M. R., & Aslekia, S. (2008). A scientometric analysis of international LIS journals: Productivity and characteristics. *Scientometrics*, 77(1), 21-39.
- Fragkiadaki, E., Evangelidis, G., Samaras, N., & Dervos, D. A. (2011). f-Value: Measuring an article's scientific impact. *Scientometrics*, 86(3), 671-686.
- Franceschini, F., & Maisano, D. (2011). Structured evaluation of the scientific output of academic research groups by recent h-based indicators. *Journal of Informetrics*, 5(1), 64-74.
- Glänzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics*, 67(2), 263-277. doi: 10.1556/Scient.67.2006.2.8
- Hellsten, I., Lambiotte, R., Scharnhorst, A., & Ausloos, M. (2007). Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? *Scientometrics*, 72(3), 469-486. doi: 10.1007/s11192-007-1680-5
- Lee, Jae Yun, & Choi, Sanghee (2011). Intellectual structure and infrastructure of informetrics. *Journal of the Korean Society for Information Society*, 28(2), 11-36. doi: 10.3743/KOSIM.2011.28.2.011
- Nebelong-Bonnevie, E., & Frandsen, T. F. (2006). Journal citation identity and journal citation image: A portrait of the Journal of Documentation. *Journal of Documentation*, 62(1), 30-57. doi: 10.1108/00220410610642039
- Schvaneveldt, R. W. (Ed.) (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex.
- Tsay, Ming-yueh (2011). A bibliometric analysis and comparison on three information science journals: JASIST, IPM, JOD, 1998-2008. *Scientometrics*, 89(2), 591-606. doi: 10.1007/s11192-011-0460-4
- White, H. (2000). Toward ego-centered citation analysis. In B. Cronin, & H.B. Atkins (Eds.), *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 475-496). Medford, New Jersey: Information Today, Inc.

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

- Ki, Ji-Hoon (2011.11.18). Self-citation network: A new approach to explore a scholar's research history. Paper presented at the Sixth COREN Conference, Kyung Hee University, Seoul, Korea.
- Lee, Jae Yun (2006). A novel clustering method for examining and analyzing the intellectual structure of a scholarly field. *Journal of the Korean Society for Information Management*, 23(4), 215-231.
- Lee, Jae Yun (2011). A study on document citation indicators based on citation network analysis. *Journal of the Korean Society for Library and Information Science*, 45(2), 119-143. doi: 10.4275/KSLIS.2011.45.2.119
- Lee, Jae Yun, Kim, Pan Jun, Kang, Dae-Shin, Kim, Heejung, Yu, So-Young, & Lee, Woo Hyoung (2011). Bibliometric analysis on LED research trends. *Journal of Information Management*, 42(3), 1-26.
- Lee, Jongwook, & Yang, Kiduk (2011). A bibliometric analysis of faculty research performance assessment methods. *Journal of the Korean Society for Information Management*, 28(4), 119-140. doi: 10.3743/KOSIM.2011.28.4.119