

문헌정보학 분야 핵심 학술지들의 가중 주제-방법 네트워크 분석*

Weighted Subject - Method Network Analysis of Library and Information Science Studies

이기현 (Keehoen Lee)** , 정효정 (Hyojung Jung)*** , 송 민 (Min Song)****

목 차

- | | |
|---------|---------|
| 1. 서론 | 4. 분석결과 |
| 2. 관련연구 | 5. 결론 |
| 3. 연구방법 | |

초 록

본 연구는 1990년부터 2014년까지 25년 간 국외 우수 문헌정보학 학술지들에 게재된 논문들을 대상으로 연구 주제와 연구 방법을 구분하여 현 선도 학술지의 연구 누적 현황을 분석하였다. 연구 주제와 연구 방법을 구분하고 그들 간의 관계를 살펴보기 위해 가중 주제-방법 네트워크를 개발하였다. 이는 주제와 방법으로 구성된 네트워크이며, 해당 토픽 소속 단어의 동시 출현 빈도를 기초로 주제와 방법에 가중치를 준다. 본 연구 결과에 따르면, 지난 25년간 경영정보시스템, 정보요구분석, 계량서지연구, 정보정책 등이 상위 연구 주제들이었다. 또한, 모델링, 개념/문헌연구, 연구협업분석, 웹데이터 분석 등이 상위 연구 방법들이었으며, 최근 텍스트 마이닝의 비상이 괄목할 만하다. 빈번히 짝을 이루는 연구 주제들과 연구 방법들의 지난 25년간의 군집들과 최근 5년간의 군집들을 살펴보면, 계량서지연구가 네트워크분석 방법들을 다수 적용하면서 그 지면을 넓히며 자리를 확고하게 하고 있으며, 텍스트 마이닝이 의료정보시스템, 이용자인터페이스에 특화되고 있음을 확인할 수 있다. 이러한 연구 결과는 문헌정보학의 선도 연구자들의 관심 연구 영역과 관점을 확인시켜주고, 향후 문헌정보학 발전을 위한 연구 설계의 기초자료로 활용될 수 있다.

ABSTRACT

In this study, we analyzed the current research state of Library and Information science in top 20 journals from 1990 to 2015, in subject and method perspectives. We developed weighted subject-method network to investigate on centralities of a subject and a method as well as their relations. This network is composed of subject nodes and method nodes and gives a weight on each node by topic occurrence. As a result, for 25 years, management information system, information need analysis, bibliometrics, information policy were top topics. Modeling, literature review, scientific research impact analysis, web data analysis were top methods. A recent rise of text mining is highlighted. We also analyzed communities made from the past 25 years and the recent 5 years. Bibliometrics is extending its field by applying various network analyzing algorithms. Text mining is specialized in medical information system and user interface. This result identifies the interests of excellent studies in Library and Information Science. It also can be fundamental resource for the development of Library and Information Science.

키워드: 연구 현황 분석, 토픽모델링, 네트워크 분석, 가중 주제-방법 네트워크

Research Topic Analysis, Topic Modeling, Network Analysis, Weighted Subject-Method Network

* 본 연구는 2015년도 BK21플러스 사업에 의하여 지원되었음.

** 연세대학교 문헌정보학과 박사 후 연구원(keehoen4research@gmail.com) (제1저자)

*** 연세대학교 문헌정보학과 석사과정(hjung4582@yonsei.ac.kr) (제2저자)

**** 연세대학교 문헌정보학과 부교수(min.song@yonsei.ac.kr) (교신저자)

논문접수일자: 2015년 7월 21일 최초심사일자: 2015년 7월 22일 게재확정일자: 2015년 8월 24일
한국문헌정보학회지, 49(3): 457-488, 2015. [http://dx.doi.org/10.4275/KSLJIS.2015.49.3.457]

1. 서론

1.1 연구 목적

본 연구의 목적은 문헌정보학 분야의 최근 선도연구를 주제(혹은 연구 영역)와 방법의 네트워크 관점에서 분석하는 방법을 제시하는데 있다. 본 연구에서는 선도 연구를 영향력 지수(Impact Factor)를 기준으로 상위에 위치한 학술지에 게재된 연구로 정의한다. 즉, 해당 학술지의 논문들이 얼마나 피인용 되었는가를 측정하는 영향력 지수는 결국 그 논문들의 타 논문들에 대한 선도 정도를 나타낸다고 본다. 이 때문에 본 연구에서는 분석을 위해 2013 JCR(Journal Citation Reports) Social Science Edition을 기준으로 문헌정보학 분야(Information Science & Library Science) 학술지 84종 중 영향력 지수의 크기가 상위 20위 내인 학술지들을 선정하여 이들을 핵심 학술지들이라 칭하고 분석 대상으로 삼았다. 그리고 지난 25년간 선정 학술지들에 게재된 논문들을 수집하였다.

본 연구는 학술지들이 지난 25년간 각 학술지에 명시된 목표와 연구 범위(Aims & Scope)를 일관되게 지향하였다고 가정한다. 또한, 이 20개의 학술지들이 포함하는 최근 선도 연구들이 그 이전에 게재된 논문들과 목표와 연구 범위의 관점에서 유관하고 이들의 영향을 받았다고 전제한다.

본 연구에서 제시하는 분석 방법은 가중 주제-방법 네트워크 분석이다. 이는 수집한 논문들을 대상으로 텍스트 마이닝 방법 중 확률적 토픽 모델링을 적용하여 인간이 이해하기 쉬운 형태의 토픽(개념)들을 추출하고, 토픽들을 주제와

방법으로 구분하며, 동시출현 토픽들을 기초로 주제-방법 네트워크를 생성한다. 이를 근거로 주제-방법의 관계에서 두드러지는 주제와 방법을 확인하고, 팔목할만한 주제-방법의 관계들을 살펴본다. 이를 기초로 해당 분야의 연구 발전에 기여할 수 있는 함의를 도출한다. 본 연구는 토픽의 구분(주제/방법), 주제-방법 네트워크 구축, 기존 네트워크 분석과 군집 분석 적용의 측면에서 기존 연구와 차별적이다.

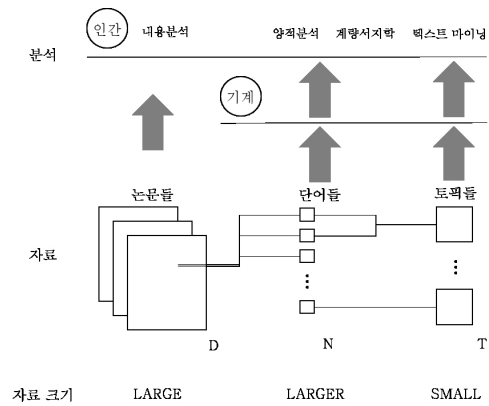
1.2 연구 배경

본 연구는 특정 연구 분야에 대한 개요연구(Overview) 혹은 문헌 연구(Review) 방법에 텍스트 마이닝을 적용하여 방법론적 발전을 돕는다. 문헌연구는 연구 문제를 해결하기 위해 현재까지 기술된 자료 중 연구 주제와 관련된 자료를 검토하는 연구 방법이다. 이를 통해 연구자는 특정 연구 주제의 가치를 판단하고, 필요한 탐구 영역의 범위를 한정한다. 기존 연구의 확장, 특정 연구의 필요성 입증, 연구에 대한 충분한 배경지식과 이해 입증, 효과적인 방법의 탐색 및 습득을 목적으로 갖는다. Pautasso(2013)가 정리한 문헌연구 작성 규칙에 따르면, (1) 주제와 독자층을 정하고, (2) 관련 연구 논문들을 반복적으로 확장하며 찾고, (3) 찾은 논문들에서 필요한 부분을 발췌하고, (4) 한정적인 기한이나 세부 주제를 다루는 단문 문헌 연구(Short review)나 전체를 다루는 전문 문헌 연구(Full review) 중 범위를 정하고, 서술적(Narrative)이거나 체계적인(Systematic) 형식을 정하고, (5) 문헌 고찰의 초점을 정하고, (6) 비판적이고 일관성 있게 작성하며, (7) 이때, 논리적인 순서로 재정렬하

며, (8) 피드백을 받고, (9) 저자의 관련 연구도 객관적으로 추가하고, (10) 최신 연구들과 함께 고전 연구들도 다루어야 한다. MIS Quarterly 학술지에 기고된 Webster와 Watson(2002)의 문헌 연구 가이드도 유사하다. 이 때, 인간이 갖는 정보 획득과 처리의 인지적 한계로 논문을 탐색하고 수집하는 과정에서 편향성이 발생할 수 있으며(Lau et al. 2007), 이는 문헌 연구의 편향적 논리로 이어질 수 있다. 일반적으로 한정적인 기한 설정이나 세부 주제에 초점을 맞추는 이유는 인지적 한계로 인한 오류를 최소화하기 위함이다. 저자 자신의 연구를 관련 연구로 추가할 때, 과대평가하거나 과소평가하는 경향이 있으므로 주의해야 한다(Logan et al. 2010).

학술 데이터의 양이 늘어나면서, 인간이 전체 데이터의 내용을 분석하는 것이 어려워 정량적인 분석으로 보완하고 있는 상황이다. <그림 1>과 같이 논문들을 인간이 직접 보고 분석하는 내용 분석, 기계의 힘을 빌려 정량적인 분석을 통해 얻은 결과를 인간이 해석하는 양적 분석, 계량 서지학, 텍스트 마이닝(특히, 문서 요약)이 활용되고 있다. 앞서 언급한 인지적 한계를 보완하기 위한 도구로 기계를 사용하여 방대한 양의 학술 데이터에서 가치 있는 정보를 추출하고 유의미한 지식을 도출하여 해당 분야의 연구 현황을 파악한 연구들이 있다(Jiang et al. 2015; Zhang et al. 2015; Peng et al. 2012; Su and Lee 2010; Yoon et al. 2010). 그러나, 최근 빈번히 시도되는 동시 출현 단어 기반의 네트워크 분석은 해석하기 난해하다. 출현 단어의 수가 많을수록 그들 사이의 관계는 기하급수적으로 증가한다. 몇 십만 개의 단어들과 그 이상의 연결 관계를 관찰하고 해석하는 일은 쉽지

않다. 이 때문에 군집화와 같은 텍스트 마이닝 기법을 적용하여 이해할 수 있을 만한 수의 군집들로 단어들을 나눈다. 이들을 대표하는 개념들을 추출하고 이를 토대로 전 문헌을 검토하고 해석한다. 군집화를 적용함에도 불구하고, 개념 추출과 그에 대한 해석은 여전히 어렵다. 그 이유는 각 군집이 포함하는 단어들의 수와 그들 간의 관계가 여전히 인간이 해석하기에 복잡하기 때문이다. 이와 같은 개념 식별의 난해성을 확률적 토픽 모델링이라는 텍스트 마이닝 기법을 이용하여 분석하면 면밀하게 보완 분석할 수 있다(Fine-grained analysis)(Yan 2015). Yan은 기존 연구에서 토픽이 계량서지학적으로 의미 있는 분석 단위임을 입증하였다(Yan et al. 2012; Yan 2014; 2015).



<그림 1> 개요 및 문헌 연구의 종류

그러므로 본 연구는 토픽 모델링을 통해 도출되는 토픽들을 활용하되, 이들을 연구 주제와 연구 방법으로 구분하고, 주제와 방법 사이의 관계를 찾고, 그들 내의 군집들을 살펴보는 “가중 주제-방법 네트워크 분석”을 제안한다. 이를 문헌 연구에 적용하여 전체 문헌을 연구 주제와 연구

방법으로 분석한 결과를 제시하고자 한다.

연구 대상 분야인 문헌정보학 분야는 다학제적 성격을 갖으며, 빅데이터 시대에 그 중요성이 가중되고 있는 분야로, 다뤄지는 주제와 사용되는 방법에 대한 이해가 중요하다. 주제는 시대 상황에 따라 빠르게 변하며, 연구 방법은 연구 목적에 따라 선별되어 주제를 해석하는 틀이 된다. 연구자는 해당 분야와 유관한 관심 영역의 문제를 정의하고, 그 문제의 목표에 따라 적절한 연구 방법을 적용하거나 개발하여 문제를 해결한다. 정성적인 방법은 새로운 통찰을 얻거나 이해의 폭을 확장하고 전반적인 이해를 돕는다. 정량적인 방법은 상황을 확인하거나 기술하고 설명하기 위해 사용되며, 주로 변수 발견과 변수들 간의 관계 규명을 위해 활용된다. 그러므로 본 연구는 문헌정보학에서 다루는 주제들과 해당 주제들을 분석하는 정성적/정량적 방법의 관계를 살펴본다.

2. 관련연구

문헌정보학 분야의 연구 영역에 대해 분석한 기존 연구들은 주로 내용 분석, 양적 분석, 계량 서지학적 분석을 통해 수행되어 왔으며, 최근 텍스트 마이닝 기법을 이용한 연구들이 증가하고 있다. 내용분석을 이용한 국내 연구들은 주제적 관점에서 시기에 따른 연구 동향을 분석하였으며(서은경 1997; 정진식 2001; 서은경 2010; 정계영, 박진희 2011), 몇몇 연구들은 다학문과의 주제 의존도와 학제적 구조 및 변화를 분석하였다(서은경 1997; 정진식 2001). 한편, 양적 분석을 이용한 연구들도 수행되어 왔다

(손정표 2003; 송정숙 2010). 기존 연구들은 분석 대상 논문들 속에서 문헌정보학 분야의 주제 범주별 논문 수를 계수하고 이를 이용해 분석하였다. 따라서 내용 분석이나 양적 분석을 이용할 경우 논문 수 기반의 분석 결과를 제시하게 되므로 문헌정보학 분야의 다양한 주제 개념의 분포 및 주제 간의 관계에 대한 분석 결과를 도출하는데 한계가 있다(조재인 2011).

관계 기반의 분석 결과를 도출하기 위해 문헌정보학 분야의 지적구조를 파악한 국내·외 기존 연구들은 계량서지학적 분석 방법인 저자동시인용분석, 문헌동시인용분석, 서지결합분석을 이용하였다(서은경 1992; 윤구호, 서말숙 2001; 김희정 2005; Chen et al. 2010). 계량서지학적 분석 방법은 저자나 문헌 간의 동시출현 관계를 기반으로 학문 분야의 지적구조를 파악할 수 있다. 이후 저자동시인용분석과 서지결합분석의 단점을 보완한 방법인 저자서지결합분석이 Zhao and Strotmman(2008a)에 의해 제안되어 이를 이용한 연구들도 있다(Zhao and Strotmman 2008b; 박지연, 정동열 2013). 그러나, 계량서지적 분석 방법은 저자나 문헌 간의 인용 관계를 기반으로 하기 때문에 주제 단위의 분석 결과를 제시하기보다는 저자나 문헌 단위의 분석 결과를 제시한다고 할 수 있다.

주제 간의 관계에 대한 분석 결과를 제시할 수 있는 방법으로는 텍스트 마이닝 기법이 있다. 텍스트 마이닝 기법은 방대한 양의 텍스트 데이터에서 유의미한 분석 결과를 도출하기 위해 사용되는 방법으로 최근 다양한 학문 분야에서 텍스트 마이닝 기법을 통해 학문 분야의 연구 현황 및 동향을 분석하는 연구들이 수행되고 있다. 텍스트 마이닝 기법을 이용한 연구들은 주

로 키워드 기반의 단어 동시출현 네트워크 분석이나 클러스터링 기법을 통해 학문 분야의 주제 분포를 파악한다. 키워드 기반의 단어 동시출현 네트워크를 이용한 다수의 연구들이 존재하나(Su and Lee 2010; Yoon et al. 2010), 단순 키워드 기반의 네트워크 분석 시 네트워크를 구성하는 노드와 연결선의 수가 너무 많아 결과에 대한 파악이 어려우며, 결과 파악을 위해 노드와 연결선의 수를 상위 빈도와 같이 일정 기준에 따라 필터링 할 경우 분석 결과의 대표성 확보에 대한 문제가 있을 수 있다. 이러한 문제

의 해결을 위해 키워드에 대한 군집화 결과를 기반으로 동시출현 네트워크 분석을 수행한 연구들이 있다(Peng et al. 2012). 군집화를 이용할 경우 키워드들의 군집을 대상으로 분석을 수행하기 때문에 단순 키워드 기반의 네트워크 분석 결과보다 요약된 결과를 도출할 수 있으나, 해당 군집들을 구성하는 키워드들을 통해 군집의 주제를 파악하는데 한계가 있다. 군집화 기법의 한계점을 보완하기 위해 군집화와 유사한 방법이라고 할 수 있는 토픽모델링 기법을 사용한 연구들도 있다(Song and Kim 2012; 박자현

〈표 1〉 방법론에 따른 관련 연구 정리

논 문	방법론			
	내용 분석	양적 분석	계량서지적 분석	텍스트 마이닝 분석
서은경(1997)	✓			
정진식(2001)	✓			
서은경(2010)	✓			
정재영, 박진희(2011)	✓			
손정표(2003)		✓		
송정숙(2010)		✓		
서은경(1992)			✓	
윤구호, 서말숙(2001)			✓	
김희정(2005)			✓	
Chen et al.(2010)			✓	
Persson(1994)			✓	
Zhao, Strotmman(2008c)			✓	
박지연, 정동영(2013)			✓	
Su and Lee(2010)				✓
Yoon et al.(2010)				✓
Peng et al.(2012)				✓
Jiang et al.(2015)				✓
Song, Kim(2012)				✓
박자현, 송민(2013)				✓
유영준(2003)				✓
조재인(2011)				✓
서은경, 유소영(2013)				✓
김관준, 이재운(2007)				✓
유종덕, 최은주(2011)				✓

송민 2013; Jiang et al. 2015). LDA(Latent Dirichlet Allocation, 이하 LDA) 토픽모델링은 확률적 토픽모델링 기법 중 가장 대표적인 기법으로 문헌을 구성하는 토픽들을 단어의 분포로 표현해주며, 이를 통해 특정 문헌을 구성하는 주제들을 파악할 수 있다(Blei 2012). 문헌정보학 분야를 대상으로 한 국내 연구들은 색인어 간의 동시출현 빈도를 이용한 군집 분석 및 단어동시출현 기반의 네트워크텍스트 분석이나 네트워크 분석을 수행한 연구들이 있으며(유영준 2003; 조재인 2011; 서은경, 유소영 2013), 본 연구와 같이 LDA 토픽모델링을 이용한 연구도 있다(박자현, 송민 2013). 이외에도 텍스트 마이닝 기법의 일종인 프로파일링 분석 기법을 이용한 연구들도 있다(김관준, 이재윤 2007). 프로파일링 분석 기법은 저자나 디스크립터 등의 분석 단위를 단어 벡터로 표현하고, 이를 기반으로 벡터 유사도를 산출하여 클러스터나 네트워크를 통해 분석하는 기법으로 저자동시인용분석과 저자 프로파일링 분석을 비교·분석한 연구도 수행된 바 있다(유종덕, 최은주 2011).

이와 같은 기존 연구들을 통해 학문 분야의 연구 현황 또는 동향을 분석한 기존 연구들은 연구 주제와 연구 방법을 구분하지 않고 개별적인 학문 영역으로 간주하여 연구 주제에 대한 분석을 수행하고 있으며, 연구 주제 또는 연구 방법 간의 관계에 대해 분석한 연구들보다는 연구 주제 간의 관계에 대해 분석한 연구들이 수행되고 있다. 그러나, 연구자들은 특정 연구 주제에 대한 연구 문제를 해결하는데 적합한 연구 방법을 적용하거나 개발하므로 연구 주제와 연구 방법은 상호 긴밀한 관계가 있다고 할 수

있다. 따라서 특정 학문 분야의 연구 현황을 파악하는데 있어 연구주제와 연구방법에 대한 개별 분석이 필요하며, 더 나아가 연구 주제와 연구 방법 간의 관계를 파악할 필요가 있다. 또한, 국내 연구들은 주로 국내 문헌정보학 분야 학술지를 대상으로 연구를 수행하는 경향이 있으므로 국외 문헌정보학 분야 학술지를 대상으로 한 연구의 필요성에 대해 언급할 수 있다.

3. 연구방법

본 장은 문헌정보학 분야의 선도 연구들 속에서 연구 주제와 연구 방법의 현황을 살펴보기 위한 방법을 설명한다. 본 연구는 출현 단어를 기초로 주제와 방법에 가중치를 주는 가중 주제-방법 네트워크 분석 방법을 제시한다. 이 분석은 문헌정보학 분야의 선도 연구들이 집중하고 있는 연구 주제와 연구 방법 그리고 그들 간의 관계를 확인할 수 있게 한다.

이 분석 방법은 연구 주제와 관련 문제 해결 방법의 구조적 위치와 관계를 분석한다. 이를 위해, 먼저 대상 문헌들 내의 어휘들을 군집화 및 개념화하여 인간이 이해할 수 있을 만한 수의 주제들과 방법들을 도출한다. 그리고 이들을 점으로 갖는 네트워크를 구축한다. 이 때, 점들 사이의 연결선은 특정 주제에 관련된 어휘와 특정 방법에 관련된 어휘가 같은 문헌에서 동시 출현하는 빈도를 의미한다. 특히, 한 문헌에서도 여러 번 같이 출현하면 그 만큼 가중된 값을 갖도록 한다. 점의 크기는 중심성과 관련 있으며 주제나 방법의 빈도에 비례한다. 그러므로 점의 크기로 각 주제나 각 방법이 전체 주제-방

법 네트워크에서 차지하는 비중을 가늠할 수 있으며, 연결선의 여부로 연결선 끝의 주제와 방법의 궁합을 파악할 수 있고, 연결선의 굵기(동시출현 정도)로 이들 주제와 방법의 궁합 정도를 확인할 수 있다. 또한, 군집으로 탐지되는 주제들과 방법론들을 확인할 수 있다. 이를 통해, 문헌정보학 분야에서 최근 선도적 연구들이 존재하는 학술지들이 그 선도성을 인정받기까지 누적된 연구들의 현황을 이해한다.

3.1 연구 문제

현재 문헌정보학 내에서 선도적인 연구들을 담은 핵심적 학술지들이 그동안 담아왔던 주제들(연구 영역)과 방법들(연구 방법)이 무엇인지 살펴본다. 본 연구는 2013 JCR(Journal Citation Reports) Social Science Edition 기준 문헌정보학 분야 상위 20개의 학술지에서 1990년부터 2014년까지 수행된 연구들을 대상으로 한다. 이를 대상으로 가중 주제-방법 네트워크 분석을 통해 개별 연구 주제에 대한 분석(연구 문제 1), 개별 연구 방법에 대한 분석(연구 문제 2), 연구 주제와 연구 방법의 관계에 대한 분석(연구 문제 3)의 결과를 제시한다.

연구문제 1: 문헌정보학 연구 분야에서 선도 연구들의 연구 주제 영역은 무엇인가?

연구문제 2: 문헌정보학 연구 분야에서 선도 연구들의 방법들은 무엇인가?

연구문제 3: 문헌정보학 연구 분야의 선도 연구들이 선호하는 연구 주제들과 연구 방법들은 어떻게 무리를 지어 함께 활용되는가?

3.2 연구 과정

3.2.1 데이터 수집

본 연구는 문헌정보학 분야의 현재 선도 연구를 다룬 학술지들을 구성하는 연구들의 누적 현황을 파악하기 위해 분석 대상을 2013 JCR(Journal Citation Reports) Social Science Edition에서 문헌정보학(Information Science & Library Science)으로 분류되는 학술지들의 논문들로 한정하였다. 앞서 언급하였듯이, 선도 연구는 타 연구에 대한 영향력 정도로 정의하기에, 84개 학술지 중 학술지 영향력(Impact Factor) 기준 상위 20개 학술지 선정하였다. 해당 학술지들은 전체 84개 중 상위 25%에 포함되는 학술지들로 이들 학술지에 게재 된 논문들은 학술적 가치를 인정받은 논문으로 판단하였다. 총 20,379개의 논문들이 수집되었으며, 그 중 1990년부터 2014년 사이의 논문들 20,344개를 선택하였다. 논문의 제목, 저자명, 출판년도 등의 서지사항과 초록을 수집하였다. 학술지 당 평균 논문 수는 1,051개이다.

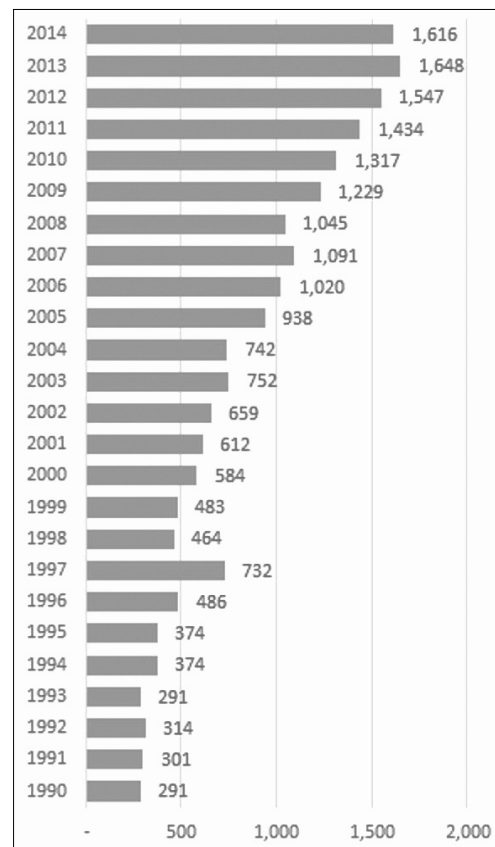
〈표 2〉는 수집한 논문들이 포함된 학술지들과 각 학술지의 논문 수를 보여준다. 가장 많은 논문 수를 보여주는 학술지는 Scientometrics로 3,423개의 논문을 포함하고 있으며, 가장 적은 논문 수를 보여주는 학술지는 Annual Review of Information Science and Technology로 173개의 논문을 포함하고 있다. 2013 JCR(Journal Citation Reports) Social Science Edition에 따르면 분석 대상 학술지들이 속한 Information Science & Library Science는 계량서지학(Bibliographic studies), 목록(Cataloging), 분류(Categorization), 데이터베이스 구축 및 유지(Database construction and maintenance), 전

자도서관(Electronic libraries), 정보 윤리(Information ethic), 정보 처리 및 관리(Information processing and management), 상호대차(Interlending), 보존(Preservation), 정보계량학(Scientometrics), 사서학(Serial librarianship), 전문도서관(Special libraries)을 다룬다. 그러나, <표 2>와 같이 영향력 기준 상위 20개의 학술지는 대부분 정보학에 해당하는 학술지들을 알 수 있다.

<표 2> 2013 JCR Social Science의 Information Science & Library Science 분야 상위 20위 학술지 목록

순위	학술지명	논문 수
1	MIS Quarterly	756
2	Journal of The American Medical Informatics Association	2,462
3	Journal of Information Technology	654
4	Journal of Informetrics	508
5	Journal of Strategic Information Systems	498
6	Information and Organization	204
7	Information Systems Research	728
8	Scientometrics	3,423
9	Journal of The American Society for Information Science and Technology	2,368
10	International Journal of Information Management	1,133
11	Government Information Quarterly	1,003
12	Journal of Computer-Mediated Communication	682
13	Journal of Management Information Systems	991
14	Journal of Health Communication	1,134
15	International Journal of Computer-Supported Collaborative Learning	219
16	Information Management	1,275
17	Annual Review of Information Science and Technology	173
18	European Journal of Information Systems	752
19	Social Science Computer Review	838
20	International Journal of Geographical Information Science	1,224

따라서 본 연구는 전 문헌정보학을 대표하는 표본을 대상으로 하기보다 현 문헌정보학을 선도하는 연구들과 이들 연구를 담은 학술지들이 그동안 게재했던 연구들을 대상으로 분석한다. 즉, 선도 연구들을 파악하고 이들을 담은 학술지들이 선도성을 인정받기까지 게재한 연구들의 연구 주제와 연구 방법을 살펴본다. 본 연구에서 선정하게 된 정보학 관련 학술지들이 문헌정보학에서 선도 연구로서의 대표성을 갖는 이유는 현재 문헌정보학과에서 상대적으로 영향력 높은 선도 연구들이 정보학에 속하였기 때문이다.



<그림 2> 연도별 논문 수

〈그림 2〉는 문헌정보학 분야 학술지 20종의 연도별 논문 수로 1990년부터 2014년까지 지속적으로 증가추세를 보이는 것으로 나타났다.

3.2.2 데이터 전처리

수집한 데이터는 Stanford NLP(Manning et al. 2014) 형태소 분석기를 통해 전처리한다. Stanford NLP는 Java 프로그래밍 언어 기반의 형태소 분석기로 단어의 품사를 식별하고, 의미가 다른 단어에 의존적인 단어(예: 전치사)를 제거할 수 있게 한다. 또한, 불용어로 정의된 단어들을 여과시키고 단어의 원형 복원 과정을 거친다.

3.2.3 주제와 방법 추출

본 연구는 전처리한 데이터를 인간이 이해할 수 있을 만한 수의 개념들로 요약하는 토픽 모델링 기법을 사용하여 주제와 방법을 추출한다. (확률적) 토픽 모델링이란, 한 문헌을 구성하는 단어들이 여러 개념들로부터 파생되었다고 가정한다. 즉, 한 개념은 여러 유관 단어들로 구성되고, 그 개념과 유관한 단어들이 나올 확률 분포가 그 개념을 나타낸다고 생각하는 것이다. 이 단어들이 포함된 군집이 나타내는 개념을 “토픽”이라고 한다.

본 연구는 대표적인 토픽 모델링 방법인 잠재 디리클레 할당(LDA)을 사용한다(Blei 2012). 이는 대표적인 통계적 추론 기반의 토픽 모델링 기법이다. 문헌을 구성하는 단어들은 각각 특정 개념에 할당되며, 문헌은 개념들의 혼합체로 생각할 수 있다. 각 개념 즉, 토픽에 할당되는 단어들을 이산 확률 분포인 다항분포로 표현하며, 이를 문헌을 입력을 받을 때마다 통

계적으로 추론한다. 추론이 끝나면, 모든 문헌에 속한 모든 단어들은 각각 사용자가 지정한 수만큼의 “토픽”들에 속하게 된다. 이 때, 동일한 단어가 여러 개념에 속한다. 각 토픽은 모든 문헌에서 발견된 단어들의 확률 분포로, 동일한 단어가 모든 토픽에 있어도 그 단어의 출현 확률과 다른 단어들의 출현 확률들로 토픽의 유일한 특성을 갖게 된다. 다항분포를 베이스 정리로 추론하기 위해 사용되는 사전 확률 분포가 디리클레 분포이기에 잠재 디리클레 할당이라고 한다(Blei 2012). 본 연구에서는 Java 기반의 Mallet(<http://mallet.cs.umass.edu/>)을 활용하였다.

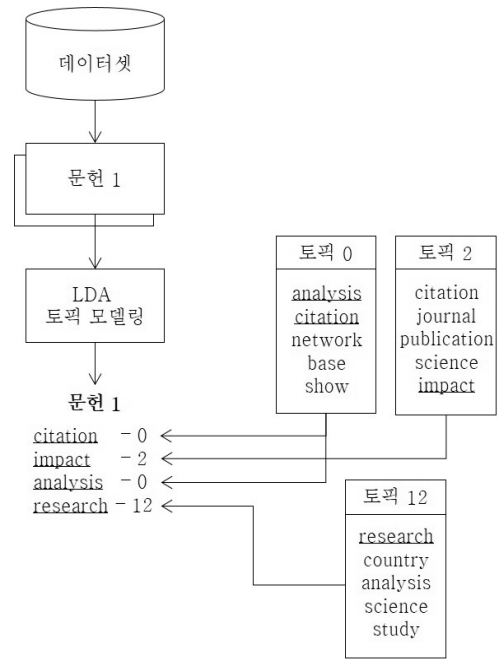
개념(토픽)의 이름 부여와 주제/방법 구분은 해당 분야 전문가들의 도움으로 수행된다. 본 연구는 문헌정보학 박사학위 이상의 연구자 3명을 대상으로 참고 분류 체계, 해당 개념에 포함된 50개의 최빈 단어들, 해당 개념이 차지하는 비중이 높은 논문들에 대한 서지사항(제목, 학술지명, 출판년)을 제공하여 개념을 명명하고 주제나 방법으로 분류하였다. 제공한 참고 분류 체계는 JITA Classification System of Library and Information Science, 2013 JCR (Journal Citation Reports) Social Science Edition의 Information Science & Library Science 부문의 설명, 한국문헌정보학회지의 범주이다. 또한, 〈표 3〉과 같이 토픽별 핵심어와 〈표 4〉와 같이 토픽별 상위 논문 정보를 제공하였다. 토픽별 상위 논문 정보는 의미가 모호하거나 중의성이 큰 단어들로 의심되는 단어들이 〈표 3〉에 등장할 때보다 정확한 명명을 위한 참고 자료로 사용된다.

3.2.4 주제와 방법의 가중 연결

추출된 주제와 방법의 연결은 한 문헌 안에서 해당 주제에 속하는 단어와 해당 방법에 속하는 단어가 동시 출현할 때 발생한다. 또한, 이 연결의 가중치는 그 동시 출현 빈도가 된다. 예를 들어, 한 문헌에서 지식 경영이라는 개념 혹은 토픽에 대한 단어가 두 개이고, 설문 조사라는 개념에 속하는 단어가 세 개면, 두 토픽 간의 연결 가중치는 단어들의 조합 수인 6이 된다. 이 때, 주제와 주제나 방법과 방법은 연결을 시키지 않는다. 이것은 가중 주제-방법 네트워크는 주제와 방법의 관계에 초점을 맞추기 때문이다.

〈그림 3〉은 이와 같은 예를 보여준다. 문헌 1을 전처리한 후 남은 단어들이 citation, impact, analysis, research라 할 때, LDA를 통해 추론한 각 토픽의 단어 분포에 따라 citation과 analysis는 토픽 0에, impact는 토픽 2에, research는 토픽 12에 할당되었다. 이 상황에서 토픽 0과 토픽 2는 방법에 해당하고, 토픽 12는 주제에 해당한다고 가정하자. 이때, 토픽 0과 토픽 2의 동시출현은 고려하지 않는다. 토픽 0에 속한 단어가 두 개, 토픽 2에 속한 단어가 하나, 토픽 12에 속한 단어가 하나였다. 따라서 토픽 0은 토픽 12와 2번 동시출현하고, 토픽 2는 토픽 12와 1번 동시출현 하였다고 할 수 있다. 토픽 0이 두 번 출현한 부분을 고려하여 출현 횟수만큼 가중된 값을 주제와 방법의 연결선의 가중치로 할당한다. 이와 같이 형성되는 주제-방법 네트워크를 가중 주제-방법 네트워크라고 명명한다. 만약 LDA 토픽 모델링의 추론 결과에서 research가 토픽 12가 아닌 토픽 11에 할당되었다고 하면, 토픽 12 대신 토픽

11이 토픽 0, 토픽 2와 동시 출현했다고 할 수 있다.



〈그림 3〉 주제와 방법의 가중 연결 예시

3.2.5 가중 주제-방법 네트워크 분석

(1) 가중 주제-방법 네트워크의 주제/방법 구성 비율 분석

토픽 모델링은 각 토픽의 구성 비율을 단어 수준에서 확인할 수 있는 결과를 만든다. 예를 들어, 연구 주제에 관한 토픽이 있고 그 토픽을 구성하는 단어들의 출현이 잦으면 해당 토픽의 구성 비율이 상대적으로 높아진다. 이 분석에서 각 주제, 각 방법의 비율을 도출할 수 있다. 이는 사람들이 사용하는 용어들의 빈도를 기초로 토픽의 선호 정도를 파악하는 분석이다.

(2) 가중 주제-방법 네트워크 중심성 분석
본 분석은 주제와 방법을 나누어 수행된다. 각각의 연결 중심성(Freeman 1979), 가중 연결 중심성(Freeman 1979), 매개 중심성(Freeman 1979)을 근거로 네트워크의 노드들인 주제들과 방법들을 분석한다. 중심성의 값의 크기와 값의 크기에 따른 순서를 기준으로 분석한다.

주제의 연결 중심성은 해당 주제가 방법들과 어느 정도 연결되어 있는지를 표현한다. 이 값이 크면, 해당 주제가 다양한 방법들에 의해 다루어졌다는 것을 의미한다. 또한, 이 값이 작으면, 해당 주제가 특정 방법들에 의해서만 다루어졌음을 말한다. 주제의 가중 연결 중심성은 해당 주제가 방법들과 어느 정도의 가중치로 연결되어 있는지를 표현한다. 즉, 이 값이 크면, 주제가 방법들 사이에서 어느 정도 중요한지 알 수 있다. 주제의 매개 중심성은 주제가 방법들 사이에서 매개 역할을 어느 정도 하는지를 표현한다.

방법의 연결 중심성, 가중 연결 중심성, 매개 중심성도 주제와 유사하다. 단, 연결 중심성은 주제들과의 연결 정도를, 가중 연결 중심성은 주제들 사이의 중요도를, 매개 중심성은 주제들 사이의 매개 역할 정도를 나타낸다.

특히, 각 중심성의 비율을 앞선 분석 결과인 가중 주제-방법 네트워크의 주제/방법 구성 비율과 비교를 하면, 단어의 출현 빈도와 상관없이 주제는 방법들과의 관계 속에서, 방법은 주제들과의 관계 속에서 어느 정도의 중요성을 차지하는지 알 수 있다. 만약, 어떤 주제의 구성 비율이 3%이지만, 가중 연결 중심성의 비율이 5%를 상회하게 된다면, 그 주제는 방법들 사이

에서 연구에서 언급하는 정도보다 더 중요한 역할을 한다고 생각할 수 있다.

(3) 가중 주제-방법 네트워크 군집 분석

가중 주제-방법 네트워크 중심성 분석까지는 개별 주제와 방법에 대한 분석이라면, 이 분석은 주제들과 방법들 간의 관계를 파악하기 위해 필요하다.

군집들은 커뮤니티 디텍션 알고리즘(Community detection algorithm)(Blondel et al. 2008)을 활용하여 도출한다. 이 알고리즘은 노드를 하나씩 각 군집에 소속시켜 밀도가 높아지면, 해당 군집의 일부로 상정한다. 이 분석에서 한 군집 내의 주제들과 방법들은 군집 외의 주제들과 방법들에 비해 상대적으로 함께 사용되는 빈도가 높다. 그러므로 주제들의 유사성, 방법들의 유사성, 주제-방법 짝들의 유사성을 고려하여 분석 기술한다. 분석 시 주의해야 할 점은 군집들 간 주제들과 방법들이 상관없지 않다는 점이다. 이들 군집 내의 주제들과 방법들이 군집 외의 주제들과 방법들과도 짝을 이루어 사용되기는 한다.

또한, 군집을 이루는 연구들이 구체적인 예를 파악하기 위해 군집을 이루는 토픽들의 비율이 높은 연구들을 살펴본다. 토픽 모델링 후, 각 문헌별로 각 토픽에 해당하는 단어의 개수를 알 수 있다. 이를 기초로 각 문헌의 토픽 구성 비율을 계산할 수 있다. 예를 들어, 토픽 1, 토픽 2, 토픽 10이 군집을 이루면, 이들 토픽들을 높은 비율로 담고 있는 논문들을 확인함으로써 이 군집의 주요 연구 주제와 방법의 관계를 이해할 수 있다.

4. 분석결과

4.1 토픽 모델링 결과

본 연구의 분석 대상에 토픽 모델링을 적용한 결과는 <표 3>과 같이 정리된다. 문헌정보학과 전문가 세 명에게 토픽 명칭 부여를 의뢰한 결과, 20개의 토픽들은 14개의 주제들과 6개의 방법들로 분류되었다. 주제로 분류된 토픽들은 계량서지연구, 지리정보연구, 이용자 연구, 정보요구분석, 경영정보시스템, 마케팅, 지적재산권, 이용자 인터페이스, 의료정보시스템, 의사소통행태, 정보보급 및 확산, 정보이용교육, 보건 커뮤니케이션, 정보정책이다. 방법으로 분류된 토픽들은 네트워크 분석, 모델링, 텍스트 마이닝, 개념/문헌 연구, 웹데이터 분석, 연구협업분석이다. 각 토픽은 <표 3>과 <표 4>를 기반으로 다음과 같은 근거에 의해 명칭이 부여되었다.

•**토픽 0**: analysis, citation, network, map, cluster 등의 단어들이 핵심어로 확인되었다. 토픽 0에 할당된 비율이 높은 논문들을 살펴본 결과, 동시출현 단어 혹은 동시 인용 네트워크 분석, 소셜 네트워크 분석 등 네트워크 분석에 관한 논문들로 확인되어 JITA 분류의 Bibliometric methods에 해당하는 '네트워크 분석'으로 이름을 부여하였으며, 분석 방법에 초점이 있으므로 방법으로 분류하였다.

•**토픽 1**: system, model, evaluation, methodology, framework 등의 단어들이 핵심어로 나타났다. JITA 분류에 해당하는 적절한 범주는 없으나, 토픽 1에 해당하는 논문들을 살펴본 결과, 개념적인 혹은 수리적인 모델링과 프레임워크 개

발에 대한 논문들로 확인되어 '모델링'으로 이름을 부여하였다.

•**토픽 2**: citation, journal, impact, index, indicator 등의 단어들이 핵심어로 나타났다. 토픽 2에 해당하는 논문들을 살펴본 결과, 인용 정보를 기초로 수학 및 통계 방법들을 적용하여 학술적 영향이나 성과 등을 측정하는 논문들로 확인되어 JITA 분류의 Use and Impact of information에 해당하는 '계량서지연구'로 이름을 부여하였다.

•**토픽 3**: text, retrieval, information, performance, system, algorithm 등의 단어들이 핵심어로 나타났다. 토픽 3에 해당하는 논문들을 살펴본 결과, 텍스트 마이닝을 이용한 의료정보학 분야에 대한 분석을 수행한 논문들로 확인되어 JITA 분류의 Automated language processing에 해당하는 '텍스트 마이닝'으로 이름을 부여하였으며, 분석 방법에 초점이 있으므로 방법으로 분류하였다.

•**토픽 4**: model, method, GI(Geographic Information), map, approach 등의 단어들이 핵심어로 나타났다. JITA 분류에 해당하는 적절한 범주는 없으나, 토픽 4에 해당하는 논문들을 살펴본 결과 지리정보학에 대한 논문들로 확인되어 '지리정보연구'로 이름을 부여하였다.

•**토픽 5**: model, factor, theory, user, perceive 등의 단어들이 핵심어로 나타났다. 토픽 5에 해당하는 논문들을 살펴본 결과 이용자 연구에 대한 논문들로 확인되어 JITA 분류의 User studies에 해당하는 '이용자 연구'로 이름을 부여하였다.

•**토픽 6**: information, theory, interaction, context, perspective 등의 단어들이 핵심어로 나타났다. 토픽 6에 해당하는 논문들을 살펴본 결과, 정보학 분야에서 이용자의 요구에 대해 분석한 논

문들로 확인되어 JITA 분류의 Information needs and information requirements analysis에 해당하는 '정보요구분석'으로 이름을 부여하였다.

•**토픽 7**: research, article, issue, review, discuss, literature 등의 단어들 이 키워드로 나타났다. 토픽 7에 해당하는 논문들을 살펴본 결과, 개념이나 문헌 연구에 대해 다룬 논문들로 확인되어 JITA 분류의 Information theory and library에 해당하는 '개념/문헌연구'로 이름을 부여하였으며, 기존 문헌을 기초로 여러 주제들에 대한 개념을 분석하기에 방법으로 분류하였다.

•**토픽 8**: information, organization, management, business 등의 단어들 이 핵심어로 나타났다. JITA 분류에 해당하는 적절한 범주는 없으나, 토픽 8에 해당하는 논문들을 살펴본 결과 경영정보시스템에 대한 논문들로 확인되어 '경영정보시스템'으로 이름을 부여하였다.

•**토픽 9**: market, product, information, cost 등의 단어들 이 키워드로 나타났다. 토픽 9에 해당하는 논문들을 살펴본 결과 마케팅 분야의 논문들로 확인되어 JITA 분류의 Marketing에 해당하는 '마케팅'으로 이름을 부여하였다.

•**토픽 10**: web, user, survey, response, questionnaire 등의 단어들 이 핵심어로 나타났다. JITA 분류에 해당하는 적절한 범주는 없으나, 토픽 10에 해당하는 논문들을 살펴본 결과 웹 설문조사에 대한 논문들로 확인되어 '웹 데이터 분석'으로 이름을 부여하였으며, 분석 방법에 초점이 있으므로 방법으로 분류하였다.

•**토픽 11**: information, government, security, policy 등의 단어들 이 핵심어로 나타났다. 토픽 11에 해당하는 논문들을 살펴본 결과, 정보 접근이나 확산에 있어 정부의 역할 등에 대해 분석

한 논문들로 확인되어 JITA 분류의 Intellectual property에 해당하는 '지적재산권'으로 이름을 부여하였다.

•**토픽 12**: research, country, collaboration, citation 등의 단어들 이 핵심어로 나타났다. 토픽 12에 해당하는 논문들을 살펴본 결과, 연구 협업에 대한 논문들로 확인되어 JITA 분류의 Bibliometric methods이며, 과학 연구 지수 개발 및 분석에 해당하는 '연구협업분석'으로 이름을 부여하였으며, 분석 방법에 초점이 있으므로 방법으로 분류하였다.

•**토픽 13**: user, interface, integrate, access, environment 등의 단어들 이 핵심어로 나타났다. 토픽 13에 해당하는 논문들을 살펴본 결과, 지리정보학, 의료정보학, 정보학 분야의 이용자 인터페이스에 대한 논문들로 확인되어 JITA 분류의 User interface에 해당하는 이용자 '인터페이스'로 이름을 부여하였다.

•**토픽 14**: patient, health, record, datum 등의 단어들 이 핵심어로 나타났다. JITA 분류에 해당하는 적절한 범주는 없으나, 토픽 14에 해당하는 논문들을 살펴본 결과, 의료정보에 대해 분석한 논문들로 확인되어 '의료정보시스템'으로 이름을 부여하였다.

•**토픽 15**: communication, relationship, experience, survey 등의 단어들 이 키워드로 나타났다. 토픽 15에 해당하는 논문들을 살펴본 결과, 컴퓨터 또는 인간 간의 커뮤니케이션에 대한 논문들로 확인되어 JITA 분류의 Information in society에 해당하는 '의사소통행태'로 이름을 부여하였다.

•**토픽 16**: media, network, participant, twitter 등의 단어들 이 핵심어로 나타났다. 토픽 16에 해

당하는 논문들을 살펴본 결과, 여론에 대해 분석한 논문들로 확인되어 JITA 분류의 Information dissemination and diffusion에 해당하는 '정보 보급 및 확산'으로 이름을 부여하였다.

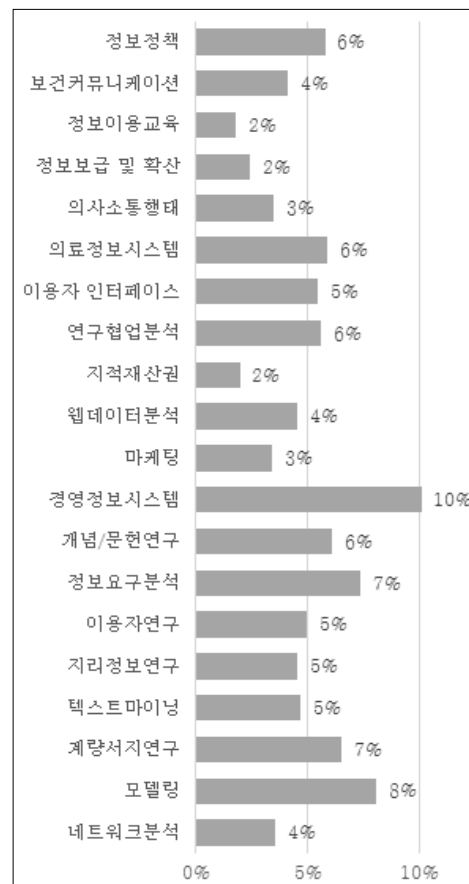
•**토픽 17**: student, learn, computer, education, program, training 등의 단어들이 핵심어로 나타났다. 토픽 17에 해당하는 논문들을 살펴본 결과, 정보이용교육에 대한 논문들로 확인되어 JITA 분류의 Literacy에 해당하는 '정보이용 교육'으로 이름을 부여하였다.

•**토픽 18**: health, communication, cancer, intervention, disease 등의 단어들이 핵심어로 나타났다. 토픽 18에 해당하는 논문들을 살펴본 결과, 보건 분야의 커뮤니케이션에 대한 논문들로 확인되어 JITA 분류의 Information in society에 해당하는 '보건 커뮤니케이션'으로 이름을 부여하였다.

•**토픽 19**: information, government, policy 등의 단어들이 핵심어로 나타났다. 토픽 19에 해당하는 논문들을 살펴본 결과, 의료정보학, 지리정보학 분야의 정보 정책에 대한 논문들로 확인되어 JITA 분류의 Information policy에 해당하는 '정보정책'으로 이름을 부여하였다.

〈그림 4〉는 각 주제별과 각 방법별 분포 비율을 보여준다. 주제로 분류된 토픽들은 계량서지연구 7%, 지리정보연구 5%, 이용자 인터페이스 5%, 정보요구분석 7%, 경영정보시스템 10%, 마케팅 3%, 지적재산권 2%, 이용자 연구 5%, 의료정보시스템 6%, 의사소통행태 3%, 정보보급 및 확산 2%, 정보이용교육 2%, 보건 커뮤니케이션 4%, 정보정책 6%를 차지하였다. MIS Quarterly와 information management 관련

학술지들이 영향력 지수 기준 상위에 분포하는 만큼 해당 학술지들의 연구 분야인 경영정보시스템이 10%로 높은 비율을 차지하였으며, 정보학 관련 주제들인 계량서지연구(7%), 정보요구분석(7%), 의료정보시스템(6%), 정보정책(6%)이 비교적 높은 비율을 차지하였다.



〈그림 4〉 토픽 비율

방법으로 분류된 토픽들은 네트워크 분석 4%, 모델링 8%, 텍스트 마이닝 5%, 개념/문헌연구 6%, 웹데이터 분석 4%, 연구협업분석 6%를 차지하였다. 가장 높은 비율을 차지하는 방법

은 모델링(8%)이며, 그 뒤를 이어 개념/문헌연구(6%)와 연구협업분석(6%)이 확인되어 대상 논문들에서 모델링, 이론적 고찰 방법과 서지 네트워크에 대한 수리 통계적 분석 방법에 관련된

언급이 많았던 것을 유추할 수 있다. 가장 낮은 비율을 차지하는 방법은 네트워크 분석(4%)과 웹데이터 분석(4%)으로 확인되었다.

〈표 3〉 토픽 별 핵심어

	토픽 0	토픽 1	토픽 2	토픽 3	토픽 4
	[BB] Bibliometric methods	N/A	[BA] Use and impact of information	[LL] Automated language processing	N/A
	네트워크 분석	모델링	계량서지연구	텍스트 마이닝	지리정보연구
핵심어	analysis	approach	citation	result	model
	citation	system	journal	method	datum
	network	model	publication	text	method
	distribution	process	impact	approach	result
	measure	evaluation	research	retrieval	area
	structure	measure	index	information	gi
	method	propose	author	document	map
	map	methodology	indicator	performance	approach
	author	framework	factor	system	study
	cluster	validity	measure	algorithm	analysis
	토픽 5	토픽 6	토픽 7	토픽 8	토픽 9
	[CB] User studies	[BH] Information needs and information requirements analysis	[AB] Information theory and library theory	N/A	[FB] Marketing
	이용자연구	정보요구분석	개념/문헌연구	경영정보시스템	마케팅
핵심어	study	study	research	information	market
	model	process	information	system	product
	factor	information	article	organization	effect
	information	theory	system	management	information
	theory	practice	issue	business	cost
	user	interaction	review	firm	price
	perceive	technology	science	technology	model
	survey	context	paper	strategy	customer
	behavior	concept	discuss	process	strategy
	intention	perspective	field	manager	business
variable	model	literature	model	reduce	

	토픽 10	토픽 11	토픽 12	토픽 13	토픽 14
	N/A	[ED] Intellectual property	[BB] Bibliometric method (Scientific research indicator)	[BI] User interfaces	N/A
	웹데이터 분석	지적재산권	연구협업분석	이용자 인터페이스	의료정보시스템
핵심어	web	information	research	database	patient
	user	government	country	software	health
	search	security	analysis	user	physician
	information	policy	collaboration	interface	record
	survey	privacy	period	integrate	datum
	datum	law	citation	access	drug
	internet	act	indicator	language	hospital
	response	access	pattern	environment	healthcare
	questionnaire	right	patent	architecture	disease
	participant	public	trend	domain	clinician
behavior	library	university	query	entry	

	토픽 15	토픽 16	토픽 17	토픽 18	토픽 19
	[BC] Information in society	[BG] Information dissemination and diffusion	[CE] Literacy	[BC] Information in society	[BF] Information policy
	의사소통행태	정보보급 및 확산	정보이용교육	보건 커뮤니케이션	정보정책
핵심어	communication	media	student	health	information
	relationship	network	learn	age	government
	examine	communication	computer	communication	technology
	member	website	education	behavior	policy
	experience	participant	university	cancer	service
	survey	blog	teaching	intervention	development
	interaction	message	program	control	community
	behavior	election	classroom	prevention	national
	perceive	campaign	learning	population	agency
	environment	twitter	training	sample	initiative
participant	facebook	instruction	disease	network	

〈표 4〉 토픽 별 상위 논문

토픽	연도	학술지명	논문명
토픽 0	2013	Social Science Computer Review	Burials and Graphs: Relational Approach to Mortuary Analysis
	2003	Scientometrics	Bridging citation and reference distributions: Part I - The referencing-structure function and its application to co-citation and co-item studies
	2013	Journal of the American Society for Information Science and Technology	Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations
	2003	Journal of the American Society for Information Science and Technology	Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists
	2013	International Journal of Geographical Information Science	Inferring properties and revealing geographical impacts of intercity mobile communication network of China using a subnet data set
토픽 1	2008	International Journal of Geographical Information Science	GIS-based multicriteria spatial modeling generic framework
	1994	Journal of Management Information Systems	An application of qualitative reasoning to derive behavior from structure of quantitative models
	2010	International Journal of Geographical Information Science	Applying time-dependent variance-based global sensitivity analysis to represent the dynamics of an agent-based model of land use change
	2009	European Journal of Information Systems	Simulating mixed agile and plan-based requirements prioritization strategies: Proof-of-concept and practical implications
	2012	Journal of the American Medical Informatics Association	Automation bias: A systematic review of frequency, effect mediators, and mitigators
토픽 2	2005	Scientometrics	Selection of research fellowship recipients by committee peer review. Reliability, fairness and predictive validity of Board of Trustees' decisions
	2011	Journal of Informetrics	Comparing impact factors from two different citation databases: The case of Computer Science
	2011	Journal of the American Society for Information Science and Technology	Global usage versus global citation metrics: The case of pharmacology journals
	2005	Journal of the American Society for Information Science and Technology	Statistical relationships between downloads and citations at the level of individual documents within a single journal
	2005	Scientometrics	Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation
토픽 3	2012	Journal of the American Medical Informatics Association	A system for coreference resolution for the clinical narrative
	2014	Journal of the American Medical Informatics Association	A comprehensive study of named entity recognition in Chinese clinical text
	2013	Journal of the American Medical Informatics Association	Using domain knowledge and domain-inspired discourse model for coreference resolution for clinical narratives
	2012	Journal of the American Medical Informatics Association	Coreference resolution of medical concepts in discharge summaries by exploiting contextual information
	2012	Journal of the American Medical Informatics Association	Coreference analysis in clinical notes: A multi-pass sieve with alternate anaphora resolution modules
토픽 4	2008	International Journal of Geographical Information Science	SRTM resample with short distance-low nugget kriging
	2006	International Journal of Geographical Information Science	The accuracy of grid digital elevation models linearly constructed from scattered sample data
	2012	International Journal of Geographical Information Science	Large-scale land cover mapping with the integration of multi-source information based on the Dempster-Shafer theory
	2010	International Journal of Geographical Information Science	Forest Fire Alert System: A GeoWeb GIS prioritization model considering land susceptibility and hotspots - a case study in the Carajás National Forest, Brazilian Amazon
	2002	International Journal of Geographical Information Science	Representing multiple spatial statistics in generalized elevation uncertainty models: Moving beyond the variogram

토픽	연도	학술지명	논문명
토픽 5	2012	International Journal of Information Management	An empirical investigation of factors influencing the adoption of data mining tools
	2008	Journal of the American Society for Information Science and Technology	An exploration of affect factors and their role in user technology acceptance: Mediation and causality
	2008	Information Systems Research	The consequences of technostress for end users in organizations: Conceptual development and validation
	2014	International Journal of Information Management	Extending the understanding of mobile banking adoption: When UTAUT meets TTF and ITM
	2009	MIS Quarterly: Management Information Systems	The integrative framework of technology use: An extension and test
토픽 6	2008	Journal of Information Technology	Anxiety and psychological security in offshoring relationships: The role and development of trust as emotional commitment
	2010	Information Systems Research	Seeking the configurations of digital ecodynamics: It takes three to tango otango
	1999	MIS Quarterly: Management Information Systems	Learning dysfunctions in information systems development: Overcoming the social defenses with transitional objects
	2011	MIS Quarterly: Management Information Systems	Capturing bottom-up information technology use processes: A complex adaptive systems model
	2013	Information and Organization	Scaling of HIS in a global context: Same, same, but different
토픽 7	2008	Information and Organization	The structure of the IS discipline reconsidered: Implications and reflections from a community of practice perspective
	1992	Social Science Computer Review	Political science research on teledemocracy
	2002	Journal of Information Technology	Critical perspectives on information systems: An impression of the research landscape
	2002	Journal of Management Information Systems	Research in information systems: An empirical study of diversity in the discipline and its journals
	2007	European Journal of Information Systems	Focusing the research agenda on burnout in IT: Social representations of burnout in the profession
토픽 8	2011	Journal of Strategic Information Systems	Drivers, barriers and critical success factors for ERP II implementation in supply chains: A critical analysis
	1995	Journal of Information Technology	Financial and strategic motivations behind IS outsourcing
	2000	Journal of Management Information Systems	The propagation of technology management taxonomies for evaluating investments in Information Systems
	1994	International Journal of Information Management	Business and technology agenda for information systems executives
	2004	Journal of Strategic Information Systems	Beyond strategic information systems: Towards an IS capability
토픽 9	2013	Information Systems Research	The impact and implications of on-demand services on market structure
	2007	Information Systems Research	The impact of Internet referral services on a supply chain
	2009	MIS Quarterly: Management Information Systems	Internet exchanges for used goods: An empirical analysis of trade patterns and adverse selection
	2012	Information Systems Research	Competitive behavior-based price discrimination for software upgrades
	2003	International Journal of Information Management	Sorting through the dot bomb rubble: How did the high-profile e-tailers fail?
토픽 10	2012	Social Science Computer Review	Asking Probing Questions in Web Surveys, Which Factors have an Impact on the Quality of Responses?
	2013	Social Science Computer Review	Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment With a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey
	2014	Social Science Computer Review	Left Feels Right A Usability Study on the Position of Answer Boxes in Web Surveys
	2014	Social Science Computer Review	What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of Internet users
	2007	Social Science Computer Review	Drop downs and scroll mice: The effect of response option format and input mechanism employed on data quality in web surveys

토픽	연도	학술지명	논문명
토픽 11	2010	Government Information Quarterly	Freedom of information and its impact in the United Kingdom
	1999	Government Information Quarterly	Public printing reform and the 105 th congress
	1991	Government Information Quarterly	Controlling and disseminating information in the Department of Defense
	1996	Government Information Quarterly	Government control of information: Some British developments
	1992	Government Information Quarterly	Interagency public access conference: Summary report
토픽 12	2013	Journal of Informetrics	Independent research of China in Science Citation Index Expanded during 1980 - 2011
	2007	Scientometrics	Bibliometric indicators of Indian research collaboration patterns: A correspondence analysis
	2012	Scientometrics	Scientific production and citation impact : a bibliometric analysis in acupuncture over three decades
	2004	Scientometrics	Research in non-metropolitan universities as a new stage of science development in Russia
	2011	Scientometrics	Characteristics of research in China assessed with Essential Science Indicators
토픽 13	2010	International Journal of Geographical Information Science	A semantic registry using a Feature Type Catalogue instead of ontologies to support spatial data infrastructures
	2009	Journal of the American Medical Informatics Association	LexGrid: a framework for representing, storing, and querying biomedical terminologies from simple to sublime
	2004	Journal of the American Society for Information Science and Technology	Metadata-based modeling of information resources on the Web
	2001	International Journal of Geographical Information Science	An integration of GIS, virtual reality and the Internet for visualization, analysis and exploration of spatial data
	2013	International Journal of Geographical Information Science	A GML-based approach to automate spatial metadata updating
토픽 14	2006	Journal of the American Medical Informatics Association	Prescribers' responses to alerts during medication ordering in the long term care setting
	2009	Journal of the American Medical Informatics Association	A Randomized Trial Comparing Telemedicine Case Management with Usual Care in Older, Ethnically Diverse, Medically Underserved Patients with Diabetes Mellitus
	2005	Journal of the American Medical Informatics Association	Exploring Barriers and Facilitators to the Use of Computerized Clinical Reminders
	2013	Journal of the American Medical Informatics Association	A long-term follow-up evaluation of electronic health record prescribing safety
	2002	Journal of the American Medical Informatics Association	Immediate benefits realized following implementation of physician order entry at an academic medical center
토픽 15	2014	Journal of Management Information Systems	Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory
	2013	International Journal of Information Management	Do instant messaging interruptions help or hinder knowledge workers' task performance?
	2005	Information Systems Research	Is out of sight, out of mind? An empirical study of social loafing in technology-supported groups
	2010	Social Science Computer Review	Sex Differences in the Expression and Use of Computer-Mediated Affective Language Does Context Matter?
	2013	Journal of Computer-Supported Collaborative Learning	Social argumentation in online synchronous communication
토픽 16	2012	Social Science Computer Review	Two-year study of emotion and communication patterns in a highly polarized political discussion forum
	2014	Social Science Computer Review	Linguistic Factors Associated With Propagation of Political Opinions in Twitter
	2011	Social Science Computer Review	Mapping the Australian networked public sphere
	2002	Social Science Computer Review	Local television, the World Wide Web, and the 2000 presidential election
	2013	Social Science Computer Review	Who tweets about politics? Political participation of Twitter users during the 2011 gubernatorial elections

토픽	연도	학술지명	논문명
토픽 17	1995	Social Science Computer Review	Multimedia: The Distant Classroom: Hardware, Software, and Administrative Considerations
	2008	Social Science Computer Review	A Rich-Media Solution for Distance Education A Review of MediaSite
	1999	Social Science Computer Review	Computers in the Undergraduate Classroom Lessons From the First 2,000 Students
	1993	Social Science Computer Review	The classroom 2000 project: A personal view of what the past tells us about the future
	2005	International Journal of Information Management	Student use of electronic information services in further education
토픽 18	2005	Journal of Health Communication: International Perspectives	A randomized controlled trial of multiple tailored messages for smoking cessation among callers to the cancer information service
	2014	Journal of Health Communication: International Perspectives	Evidence supporting a promotora-delivered entertainment education intervention for improving mothers' dietary intake: The Entre Familia: Reflejos de Salud study
	2004	Journal of Health Communication: International Perspectives	Media and community campaign effects on adult tobacco use in Texas
	2014	Journal of Health Communication: International Perspectives	Trust during the early stages of the 2009 H1N1 pandemic
	2011	Journal of Health Communication: International Perspectives	The Long Live Kids campaign: awareness of campaign messages
토픽 19	2005	Journal of the American Medical Informatics Association	Communities' readiness for health information exchange: the National Landscape in 2004
	2010	Journal of the American Medical Informatics Association	Clinical decision support: progress and opportunities
	2011	Journal of the American Medical Informatics Association	Anticipating and addressing the unintended consequences of health IT and policy: a report from the AMIA 2009 Health Policy Meeting
	2004	Journal of the American Medical Informatics Association	Bridging the digital divide: reaching vulnerable populations
	2011	Government Information Quarterly	Exploring geographic disparities in broadband access and use in rural southern Illinois: Who's being left behind?

4.2 주제와 방법 분석 결과

문헌정보학 분야 핵심 학술지들에 게재된 논문들을 구성하는 주제들과 방법들은 그들 간의 관계 속에서 구조적인 중심성을 갖는다. <표 5>는 주제들과 방법들의 연결 중심성, 가중 연결 중심성, 매개 중심성을 보여준다. 주제들의 연결 중심성은 6이며, 방법들의 연결 중심성은 14이다. 이는 주제는 모든 방법에, 방법은 모든 주제와 연결되어 있음을 알려준다. 즉, 본 연구의 가중 주제-방법 네트워크를 구성하는 모든 방

법들과 모든 주제들은 서로 같이 사용되어 연구된 적이 한번 이상 존재한다.

주제의 경우, 경영정보시스템이 가장 높은 가중 연결 중심성을 갖으며, 전체 가중 연결 중심성의 6%를 차지한다. 그 뒤를 이어 정보요구분석(6%), 계량서지연구(5%), 정보정책(5%), 이용자인터페이스(4%), 이용자연구(4%), 의료정보시스템(3%), 의사소통행태(3%), 지리정보연구(3%), 마케팅(3%), 보건커뮤니케이션(2%), 정보보급 및 확산(2%), 지적재산권(2%), 정보이용교육(2%)이 있다. 정보이용교육과 경영정

보시스템은 무려 3배 이상의 차이를 보인다.

이 결과는 연구 주제들이 연구 방법들과의 관계에서 어느 정도의 중요한 위상을 갖는지를 보여준다. 특히 최근 문헌정보학 분야 학술지들의 지난 25년간의 연구들에서 경영정보시스템이라는 주제가 각광 받았음을 알 수 있다. 그 뒤를 이어 정보시스템의 이용의 편의나 요구를 파악하기 위한 정보요구분석이 중요한 위상을 갖는 것으로 나타났으며, 정보시스템과 관련된 이용자인터페이스, 이용자연구, 의료정보시스템, 지리정보연구가 주요한 주제들로 보인다. 또한, 계량서지연구, 의사소통행태, 보건커뮤니케이션, 정보보급 및 확산과 같이 커뮤니케이션 관련 연구 주제들이 확인되며, 정보정책, 지적재산권, 정보이용교육과 같은 주제들도 존재한다.

토픽 비율과 가중 연결 중심성의 비율을 비교하면, 경영정보시스템은 각각 10%와 6%로 토픽이 전체에서 차지하는 비율에 비해 네트워크 구조 상 중요도가 작다. 정보요구분석은 7%와 6%, 계량서지연구는 7%와 5%, 정보 정책은 6%와 5%, 이용자인터페이스는 5%와 4%, 이용자연구는 5%와 4%, 의료정보시스템은 6%와 3%, 의사소통행태는 3%로 동일하였으며, 지리정보연구는 5%와 3%, 보건커뮤니케이션은 4%와 2%, 정보보급 및 확산, 지적재산권, 정보이용교육의 경우 2%로 동일한 것으로 확인되었다. 의사소통행태, 정보 보급 및 확산, 지적재산권, 정보이용교육은 토픽 비율과 가중 연결 중심성의 비율이 비등하였다. 이 결과는 이내 가지 주제들의 구조적 위상이 해당 주제들에 사용되는 단어들의 비율과 유사했음을 알려준다. 이 외의 주제들은 대부분 가중 연결 중심성의 비율이 낮았다. 이는 주제의 양에 비해 연

구 방법들과의 관계에서의 중요도가 작음을 뜻한다. 또한, 주제-방법 네트워크에서 방법의 중요도는 상대적으로 높으며, 주제가 다뤄지는 양에 비해 여러 방법에 적용되지 않음을 보여준다. 예를 들어, 의료정보시스템은 내용은 전체 논문들에서 6%를 차지하지만, 본 연구에서 탐지한 방법들 사이에서의 중요도는 전체에서 3%로 내용에 비해 상대적으로 낮다. 이는 의료정보시스템의 주제가 언급되는 정도에 비해 방법이 한정되어 있기 때문이라고 할 수 있다. 특히, 의료정보시스템의 경우, 텍스트 마이닝 외의 방법들과는 상대적으로 낮은 가중 연결 정도를 갖는다.

방법의 경우, 모델링이 가장 높은 가중 연결 중심성을 갖고 전체의 13%를 차지한다. 개념/문헌연구(11%), 연구협업분석(7%), 웹데이터 분석(7%), 텍스트 마이닝(6%), 네트워크분석(5%)이 그 뒤를 따른다. 네트워크분석과 모델링은 약 2배 차이가 있다.

이 결과에 따르면, 지난 25년간 최근 각광받는 이슈들을 다루는 핵심 학술지들에서 다뤄진 방법들은 각종 주제들에 대해 수리적으로나 개념적으로 새로운 모형 혹은 프레임워크를 개발하여 문제를 해결하는 방법이 가장 많았다. 그 뒤를 이어, 이론적인 고찰을 하는 연구 방법, 계량서지연구의 과학기술 연구 협업 영향 분석하는 방법, 웹설문과 같은 형태의 웹데이터를 획득하여 분석하는 방법, 자연어로 된 문헌의 내용에 데이터 마이닝 기법을 적용하는 방법, 관계를 정의하고 관계 구조에서의 특징들을 관찰하는 네트워크 분석 방법이 있었다.

토픽 비율과 가중 연결 중심성의 비율을 비교하면, 모델링은 각각 8%와 13%이다. 즉, 내

용의 양에 비해 모델링의 주제들과의 관계에서의 중요도는 상대적으로 높다. 개념/문헌 연구는 6%와 11%, 연구협업분석은 6%와 7%, 웹데이터분석은 4%와 7%, 텍스트 마이닝은 5%와 6%, 네트워크 분석은 4%와 5%이다. 주제들의 수에 비해 방법들의 수가 적으며, 다수의 주제들이 각 방법에 집중되기 때문에 가중 연결 중심성의 비율이 대체적으로 증가한다. 특히, 모델링이나 개념/문헌 연구의 차이가 크데, 이는 그만큼 많은 주제들이 이들 방법들을 빈번히 적용하기 때문이다. 그에 비해, 텍스트 마이닝은 그 차이가 적다. 이는 텍스트 마이닝이 25년간의 자료에서 주로 한정된 주제에 적용되었기 때문이다.

<표 6>의 최근(2011년 이후) 주제-방법 네트워크의 중심성 지수들을 살펴보면, 주제의 경우, 계량서지연구가 가장 높은 가중 연결 중심성을 갖고 전체의 7%를 차지한다. 의료정보시스템(6%), 정보요구분석(5%), 지리정보연구(5%), 경영정보시스템(5%), 정보정책(4%), 이용자 인터페이스(4%), 이용자연구(3%), 보건커뮤니케이션(3%), 의사소통행태(2%), 마케팅(2%), 정보보급 및 확산(2%), 정보이용교육(1%), 지적재산권(1%)이 그 뒤를 따른다. <표 5>와 비교했을 때, 최근에 계량서지연구, 의료정보시스템, 지리정보연구 등이 각광 받았으며, 정보요구분석은 꾸준히 연구되고 있고, 경영정보시스템, 정보정책은 상대적으로 순위에서 뒤로 밀렸다. 이

<표 5> 25년간의 주제들과 방법들의 연결 중심성, 가중연결중심성, 매개 중심성, 군집식별번호

명칭	연결 중심성	가중연결중심성(%)	매개 중심성(%)	군집식별번호
개념/문헌연구	14(8%)	11	14	0
경영정보시스템	6(4%)	6	1	0
지적재산권	6(4%)	2	1	0
정보보급 및 확산	6(4%)	2	1	0
이용자연구	6(4%)	4	1	1
정보요구분석	6(4%)	6	1	1
지리정보연구	6(4%)	3	1	1
네트워크분석	14(8%)	5	14	1
모델링	14(8%)	13	14	1
텍스트 마이닝	14(8%)	6	14	1
의료정보시스템	6(4%)	3	1	1
의사소통행태	6(4%)	3	1	1
이용자 인터페이스	6(4%)	4	1	1
계량서지연구	6(4%)	5	1	2
연구협업분석	14(8%)	7	14	2
정보정책	6(4%)	5	1	2
마케팅	6(4%)	3	1	3
보건커뮤니케이션	6(4%)	2	1	3
웹데이터 분석	14(8%)	7	14	3
정보이용교육	6(4%)	2	1	3

<표 6> 2011년부터 주제들과 방법들의 연결 중심성, 가중연결중심성, 매개 중심성, 군집식별번호

명칭	연결 중심성	가중연결중심성(%)	매개 중심성(%)	군집식별번호
모델링	14(8%)	14	14	0
개념/문헌연구	14(8%)	9	14	0
웹데이터 분석	14(8%)	6	14	0
지리정보연구	6(4%)	5	1	0
경영정보시스템	6(4%)	5	1	0
정보정책	6(4%)	4	1	0
이용자연구	6(4%)	3	1	0
보건커뮤니케이션	6(4%)	3	1	0
의사소통행태	6(4%)	2	1	0
마케팅	6(4%)	2	1	0
정보보급 및 확산	6(4%)	2	1	0
정보이용교육	6(4%)	1	1	0
지적재산권	6(4%)	1	1	0
연구협업분석	14(8%)	8	14	1
계량서지연구	6(4%)	7	1	1
정보요구분석	6(4%)	5	1	1
네트워크분석	14(8%)	5	14	1
텍스트 마이닝	14(8%)	8	14	2
의료정보시스템	6(4%)	6	1	2
이용자 인터페이스	6(4%)	4	1	2

는 계량서지연구, 의료정보시스템, 지리정보연구에 대한 다양한 방법의 적용이 늘어났고, 경영정보시스템, 정보정책에 관한 연구들에 있어서 방법적 진보가 적었음을 알려준다.

방법의 경우, 모델링이 가장 높은 가중 중심성을 갖고 전체의 14%를 차지한다. 그 뒤를 개념/문헌연구(9%), 연구협업연구(8%), 텍스트 마이닝(8%), 웹데이터 분석(6%), 네트워크분석(5%)이 따른다. 개념/문헌연구, 웹데이터 분석, 네트워크분석의 가중 연결 중심성 비율들이 <표 5>에 비해 상대적으로 작아졌다. 특히, 개념/문헌연구가 약 2% 감소하였는데, 이는 이 방법에 대한 주제들의 적용이 적어졌기 때문으

로 추측된다. 이에 비해, 모델링, 연구협업연구, 텍스트 마이닝의 비율들이 <표 5>의 대응 비율들에 비해 상대적으로 크다. 이는 상대적으로 이 방법들에 대한 수요가 최근 증가하였음을 알려준다.

텍스트 마이닝의 경우, 지난 25년간의 토픽 비율은 5%, 가중 연결 중심성 비율은 6%이며, 최근(2011년 이후) 가중 연결 중심성 비율은 8%이다. 이는 텍스트 마이닝의 전시기적 비율은 낮고 상대적으로 주제들에 적용된 정도는 낮으나, 최근 빠르게 그 적용 범위를 넓히고 있음을 의미한다.

4.3 주제들-방법들 군집 분석 결과

주제들-방법들 군집 분석은 주제들과 방법들 간의 관계를 확인시켜준다. <표 5>는 분석 대상으로 선정된 20종의 학술지들에서 지난 25년간 발간한 논문들의 가중 주제-방법 네트워크의 중심성 지수들뿐만 아니라 네트워크 분석을 통해 형성되는 군집 결과도 보여준다. <표 5>의 군집 식별번호에 따르면, 4개의 군집으로 구분된다.

군집 식별 번호 0을 갖는 방법은 개념/문헌 연구이며, 주제들은 경영정보시스템, 지적재산권, 정보보급 및 확산으로 확인되었다. 25년간의 누적현황을 기준으로 할 때, 경영정보시스템, 지적재산권, 정보 보급 및 확산에 대한 연구와 개념/문헌 연구를 적용한 연구가 주를 이룬다. <표 4>에서 주제인 경영정보시스템(토픽 8)의 대표 논문들을 살펴보면, 경영정보시스템에 대한 이론적 고찰이 주를 이룬다. 주제인 지적재산권(토픽 11)의 대표 논문들은 주로 Government Information Quarterly에 게재되었으며, 공공재 혹은 사유재로서의 정보에 대한 개념적인 논의들이 진행되었음을 알 수 있다. 정보 보급 및 확산(토픽 16)의 대표 논문들은 주로 정치적 이슈에 대한 여론 확산 연구들이다.

군집 식별 번호 1을 갖는 방법들은 모델링, 텍스트 마이닝, 네트워크 분석이며, 주제들은 정보요구분석, 이용자 인터페이스, 이용자연구, 의료정보시스템, 의사소통행태, 지리정보연구로 확인되었다. 앞선 군집 식별 번호 0의 연구 방법이 정성적인 연구 방법이라면, 이 군집의 방법들은 수리적인 모형들을 만들어야하는 정량적인 연구 방법들이다. <표 4>에서 주제인 정보요구분석(토픽 6)의 대표 논문들은 심리적

안정감을 위한 신뢰, 디지털 생태계, 학습 역기능, 복잡적응계적 정보 기술 이용 연구, 정보시스템 평가 등에 대한 새로운 모형들이나 프레임워크를 제시한다. 이용자 인터페이스(토픽 13)는 주로 텍스트 마이닝을 이용한 연구들과 이를 통해 모델링이나 프레임워크를 제시하는 연구들이 눈에 띈다. 이용자 연구(토픽 5)의 대표 논문들은 요인들 간의 관계, 개념모형을 활용한 연구들이 주를 이룬다. 의료정보시스템(토픽 14)의 연구들은 대부분 Journal of the American Medical Informatics Association에 게재된 논문들로 의료 데이터에 정보공학적인 접근(예: 텍스트 마이닝)을 적용한 연구들이 존재한다. 또한, 모델링이 필요한 평가 모형들도 확인된다. 의사소통행태(토픽 15)의 논문들은 네트워크 분석도 활용되는데, 주로 과학 커뮤니케이션에 네트워크 분석 방법들이 적용되고 있다. 지리정보연구(토픽 4)는 파편적인 지리정보로 전체 지리정보를 복원하거나 지역 통계를 이용한 수리적 모형을 개발하는 연구들로 모델링, 텍스트 마이닝, 네트워크 분석이 적용된다.

군집 식별 번호 2를 갖는 방법은 연구협업분석이며, 주제들을 계량서지연구, 정보정책이다. 계량서지연구(토픽 2)는 주로 연구협업분석 방법을 활용한다. 정보정책(토픽 19)은 주로 연구협업분석방법을 평가하고 이를 교육과학 시스템에서 어떻게 수용할지를 논한다(Weingart 2005). 이러한 연구들은 연구자의 연구 활동과 교육 활동, 교육 수혜자나 정보 서비스 이용자들을 대상으로 하는 연구들이 성과 지표와 연구, 교육 등에 영향을 미칠 수 있는 의심 요인들과의 관계를 살핀다. 즉, 교육 활동이나 연구 활동

에 대한 정량적인 평가에 대한 연구들이 진행되었음을 알 수 있다.

군집 식별 번호 3을 갖는 방법은 웹데이터 분석이며, 주제들은 마케팅, 보건커뮤니케이션, 정보이용교육이다. 마케팅(토픽 9)은 웹(혹은 인터넷)을 이용하여 설문 조사를 하고 이를 토대로 시장에서 생존할 수 있는 전략을 모색하였다(Chellappa et al. 2010). 보건커뮤니케이션(토픽 18)은 설문 조사를 토대로 보건 메시지를 전달하는 채널들에 대한 연구들이 주로 있다(Nishtar et al. 2004; Peddecord et al. 2008). 정보이용교육(토픽 17)은 설문 조사를 기초로 정보 서비스의 교육적 의미 탐구를 하였다(Urquhart et al. 2005). 정보 서비스나 정보 시스템에 대한 실수요자의 경험에 대한 지식을 얻기 위해 설문조사가 여전히 중요하게 활용되고 있고 이 방법 외에 주관적인 체험에서 비롯한 정보를 획득하기 어려움을 시사한다. 한편으로 설문조사의 채널이 웹으로 확장되었음을 확인할 수 있다.

〈표 5〉가 장기간의 주제들-방법들의 군집들을 보여준다면, 〈표 6〉은 최근의 주제들-방법들의 군집들을 보여준다. 〈표 5〉를 준거로 한 〈표 6〉의 최근 군집 현황은 다음과 같다. 두 표를 비교하면, 군집의 수가 각각 4개와 3개로 차이가 있으며, 구조적인 차이도 보인다. 예를 들어, 최근 텍스트 마이닝이 의료정보시스템, 이용자인터페이스의 주제들과 독립적인 군집을 이루었다.

군집 식별 번호 0을 갖는 방법은 모델링, 개념/문헌연구, 웹데이터 분석이며, 주제들은 지리정보연구, 경영정보시스템, 정보정책, 이용자연구, 보건커뮤니케이션, 의사소통행태, 마케팅,

정보보급, 정보이용교육, 지적재산권이다. 예를 들어, 〈표 4〉에서 지리정보연구(토픽 4) 논문으로 추출된 “Large-scale land cover mapping with the integration of multi-source information based on the Dempster-Shafer theory”는 지리 정보 분류를 위해 새로운 모형을 개발하면서 기존에 존재했던 이론인 Dempster-Shafer 이론을 활용했다.

군집 식별 번호 1을 갖는 방법은 연구협업분석, 네트워크 분석이며, 주제들은 계량서지연구, 정보요구분석이다. 계량서지연구와 연구협업분석은 25년간이나 최근이나 지속되고 있는 강건한 주제-방법의 관계를 확인시켜준다. 연구협업분석과 네트워크 분석은 크게 네트워크 분석으로 통합할 수 있다. 일반적으로 네트워크 분석은 소셜 네트워크 분석이나 사회물리학적 네트워크 분석 방식을 일컫는다. 〈표 4〉에서 정보요구분석(토픽 6)의 대표 논문들 중 2011년에 게재된 논문은 복잡적응계 모형을 제시하는데 이 복잡적응계는 일종의 네트워크이다. 또한, 이들 중 2013년에 게재된 논문은 보건 정보 시스템 프로그램 프로젝트(Health Information Systems Programme, HISP)를 대상으로 하며, HISP는 개발도상국의 보건과 IT로 구성된 일종의 R&D 네트워크이다. 즉, 정보요구분석에 네트워크 관점에서 분석하는 내용들이 최근 늘고 있다.

군집 식별 번호 2를 갖는 방법은 텍스트 마이닝이며, 주제들은 의료정보시스템, 이용자인터페이스이다. 예를 들어, 〈표 4〉의 이용자 인터페이스(토픽 13)의 대표 논문들 중 2013년에 게재된 논문은 메타데이터의 자동 갱신에 텍스트 마이닝을 활용한다.

5. 결론

5.1 요약

본 연구는 문헌정보학 분야에서 최근 영향력이 높은 선도 연구들을 연구 주제와 연구 방법의 두 가지 형태의 노드들로 구성된 이분 네트워크(Bipartite network)로 표현한다. 영향력이 높은 선도 연구들은 영향력 지수를 기준으로 상위에 위치한 학술지에 게재된 연구들이다. 본 연구는 영향을 미치는 현상을 선도한다고 여긴다.

분석 대상 논문들의 수가 증가함에 따라 분석 대상으로부터 추출되는 단어들의 수가 기하급수적으로 커지기 때문에 기존의 동시출현 단어 기반 네트워크 분석을 적용할 경우 네트워크 내부의 개념들을 표현하고 해석하는데 어려움이 있다. 이와 같은 어려움을 극복하기 위해 본 연구는 가중 주제-방법 네트워크 분석을 새롭게 제시하였다. 이는 인간이 이해하기 쉬운 형태의 개념들로 구성된 네트워크로 표현하여 분석하며, 텍스트 마이닝 방법 중 확률적 토픽 모델링을 이용하여 개념들(토픽들)을 추출한다. 토픽 모델링 결과에 대한 주제 혹은 방법 분류 및 명칭 부여는 문헌정보학과 전문가들에게 의뢰하였다. 의뢰 시, 각 토픽을 구성하는 최빈 단어들 50개, 각 토픽별 해당 토픽의 비율이 가장 높은 5개의 연구 논문들, 여러 문헌정보학 분류 체계를 제공하여 개념의 식별에 도움을 주었다. 이를 기초로 생성한 주제-방법 네트워크에서 괄목할만한 주제, 방법 및 주제-방법 관계를 살펴보았다.

특히, 본 연구는 인간의 인지적 한계로 인한

오류 가능성을 최소화하기 위해 새롭게 계량서지학적 의미 분석 단위로 주목을 받는 '토픽'을 활용하였다. 기존 문헌 연구는 인간의 인지적 한계로 범할 수 있는 오류 가능성을 내재하고 있다. 주요 원인은 정량적 분석의 결과가 인간이 해석하기 어려운 형태라는 점이다.

그러므로 본 연구는 문헌정보학 분야에서 영향력이 큰 핵심 학술지 20가지를 선정하여 연구 주제와 연구 방법의 누적 현황을 살펴보기 위해 주제와 방법으로 구성된 네트워크 분석을 수행하였다. 이와 같은 네트워크를 구성하기 위해, 동시 출현 단어들을 근거로 주제와 방법에 가중치를 주는 가중 주제-방법 네트워크를 개발하여 분석하였다. 이 분석을 통해 문헌정보학 분야의 선도 연구들이 집중하고 있는 연구 주제와 연구 방법 그리고 그들 간의 관계를 확인하였다. 또한, 문헌정보학 분야 연구 발전에 기여할 수 있는 함의를 도출하였다.

가중 주제-방법 네트워크 분석을 위한 토픽 모델링 결과, 계량서지연구, 지리정보연구, 이용자 연구, 정보요구분석, 경영정보시스템, 마케팅, 지적재산권, 이용자 인터페이스, 의료정보시스템, 의사소통 행태, 정보보급 및 확산, 정보이용교육, 보건 커뮤니케이션, 정보 정책의 14개의 주제들과 네트워크 분석, 모델링, 텍스트 마이닝, 개념/문헌 연구, 웹데이터 분석, 연구협업 분석의 6개의 방법들을 확인할 수 있었다.

토픽 비율과 가중 연결 중심성 비율의 비교 결과, 지난 25년간 의사소통행태, 정보 보급 및 확산, 지적재산권, 정보이용교육을 제외한 모든 주제들은 토픽 비율이 높았다. 이는 주제들의 상대적 중요성이 방법들에 비해 낮고, 주제가 언급된 정도에 비해 주제에 적용된 방법의 다

양성이 적기 때문이다. 즉, 연구 주제는 해당 분야에서 많은 관심을 갖는 만큼 정형화된 방법이 어느 정도 존재한다고 할 수 있다. 이러한 주제들은 새로운 해결 방법의 적용이나 개발이 필요할 수 있다. 또한, 방법들은 가중 연결 중심성의 비율이 토픽 비율보다 높았으며, 모델링과 개념/문헌 연구의 차이가 가장 컸다. 이는 그만큼 이 방법들이 다양한 분야에 적용되었기 때문이다. 또한, 텍스트 마이닝의 비율 차가 가장 적은 것은 비교적 한정된 주제들에 적용되었기 때문이다. 그러므로 텍스트 마이닝과 같은 비율 차가 낮은 방법들의 다양한 주제에 대한 적용 시도가 문헌정보학 연구의 발전에 기여할 여지가 크다고 할 수 있다.

2011년부터 최근 5년간으로 분석 대상의 범위를 축소하여 최근 가중 주제-방법 네트워크를 살펴보면, 상대적으로 계량서지연구, 의료정보시스템, 지리정보연구가 각광받고 있으며, 정보요구분석이 꾸준히 연구되고 있고, 경영정보시스템, 정보정책이 상대적으로 인기가 낮아졌다. 방법 중에서는 개념/문헌연구, 웹데이터 분석, 네트워크 분석의 적용이 줄었으며, 모델링, 연구협업연구, 텍스트마이닝의 적용이 늘었다. 특히, 텍스트 마이닝의 중심성 증가가 눈에 띈다.

가중 주제 방법 네트워크 군집 분석 결과, 문헌정보학 연구 분야의 선도 연구들이 선호하는 연구 주제들과 연구 방법들의 군집들은 지난 25년간의 경우 4 개이며, 최근 5년간의 경우 3개이다. 최근 들어, 네트워크분석과 텍스트 마이닝이 독립적인 군집을 형성하였다. 네트워크분석은 연구협업분석, 계량서지연구, 정보요구분석과 함께 군집을 이루었다. 또한, 텍스트 마이

닝은 의료정보시스템, 이용자인터페이스와 군집을 이루었다. 이는 계량서지연구가 보다 넓은 개념인 네트워크분석으로 확장되고, 텍스트 마이닝이 문헌정보학 내에서 특화되고 있음을 보여준다.

5.2 한계점

본 연구에서 사용한 토픽 모델링 기법은 문헌 집단의 주제적 분포를 파악하는데 효과적인 기법이나 분석 시 연구자가 토픽의 수를 지정해야 한다는 단점이 있어 다양한 토픽의 수를 지정하여 전체 문헌 집단의 주제 분포를 효과적으로 나타낼 수 있는 토픽의 수를 판단하는 것이 중요하다. 또한, 토픽모델링 결과에 대해 주제와 방법을 식별하고 세부 이름을 부여하는데 있어 전문가에 의존적이므로 전문가의 전문성에 따라 부여되는 이름이 달라질 여지가 있다. 그러나, 일정 수 이상의 전문가가 부여하는 식별 이름들은 객관성을 얻을 수 있다. 마지막으로 핵심 학술지 20개를 선정하였으므로 분석 결론의 일반화에 무리가 있을 수 있다. 특히, 해당 학술지들의 주로 정보학을 다루고 있으며, 도서관학을 다루는 학술지들의 비율이 현저하게 낮았다. 그러나, 본 연구의 분석 방법 및 분석 결과는 학문 분야의 연구 현황을 파악하는데 있어 새로운 방향성을 제시하였다고 할 수 있다.

5.3 향후 연구

본 연구는 국내·외 문헌정보학 분야 전체 학술지를 대상으로 분석 범위의 확장이 가능하며, 토픽 모델링 분석 결과에 대한 레이블 부여하는

데 있어 사전 기반의 개체명 인식 기법(Named Entity Recognition: NER)과 같은 기계 학습 방식을 적용하여 보다 정교한 분석 결과의 도출이 가능하다. 또한, 네트워크 분석 시 연결선들의 가중치에 따른 상·하위 순위에 대한 분석을 통

해 해당 학문 분야의 선호 연구 주제와 연구 방법의 관계를 파악할 수 있을 것으로 기대된다. 향후, 누적된 자료의 시기를 세분화하여 동향과 현황을 분석을 통해 보다 통시적인 연구 주제와 연구 방법의 관계를 확인할 수 있을 것으로 기대된다.

참 고 문 헌

- [1] 김관준, 이재윤. 2007. 연구 영역 분석을 위한 디스크립터 프로파일링에 관한 연구. 『정보관리학회지』, 24(4): 285-303.
- [2] 김희정. 2005. 저자 동시인용분석에 의한 국내외 기록관리학 분야의 지적구조 비교에 관한 연구. 『한국문헌정보학회지』, 39(3): 207-224.
- [3] 박자현, 송민. 2013. 토픽모델링을 활용한 국내 문헌정보학 연구동향 분석. 『정보관리학회지』, 30(1): 7-32.
- [4] 박지연, 정동열. 2013. 저자서지결합분석에 의한 문헌정보학의 지적구조 분석에 관한 연구. 『정보관리학회지』, 30(4): 31-59.
- [5] 서은경. 1992. 정보검색분야의 지적 구조와 변화에 관한 연구: 영어문화권 저자들을 중심으로. 『정보관리학회지』, 9(1): 55-82.
- [6] 서은경. 1997. 정보학분야 연구동향 분석: 정보관리학회지와 JASIS 의 비교분석을 중심으로. 『정보관리학회지』, 14(1): 269-291.
- [7] 서은경. 2010. 정보관리학회지 연구의 동향분석. 『정보관리학회지』, 27(4): 7-31.
- [8] 서은경, 유소영. 2013. Detecting Research Trends in Korean Information Science Research, 2000-2011. 『정보관리학회지』, 30(4): 215-239.
- [9] 손정표. 2003. 한국의 문헌정보학 분야 연구동향 분석: 1957-2002. 『한국도서관·정보학회지』, 34(3): 9-21.
- [10] 송정숙. 2010. 한국 문헌정보학의 연구현황 분석. 『한국도서관·정보학회지』, 41(4): 333-353.
- [11] 오세훈. 2005. 우리나라 문헌정보학 학술지 논문 및 인용문헌 분석을 통한 연구동향 연구. 『정보관리학회지』, 22(3): 379-408.
- [12] 유영준. 2003. 문헌정보학의 지식 구조에 관한 연구. 『정보관리학회지』, 20(3): 277-298.
- [13] 유종덕, 최은주. 2011. 저자프로파일링분석과 저자동시인용분석의 유용성 비교 검증. 『정보관리학회지』, 28(1): 123-144.

- [14] 윤구호, 서말숙. 2001. 저자동시 인용분석에 의한 1990년대 한국문헌정보학의 지적구조에 관한 연구. 『한국도서관·정보학회지』, 32(3): 169-197.
- [15] 정재영, 박진희. 2011. 한국 문헌정보학의 현장연구 현황 분석. 『한국도서관·정보학회지』, 42(2): 171-191.
- [16] 정진식. 2001. 한국 문헌정보학 분야의 연구동향 분석: 1996-2000. 『한국문헌정보학회지』, 35(3): 55-78.
- [17] 조재인. 2011. 네트워크 텍스트 분석을 통한 문헌정보학 최근 연구 경향 분석. 『정보관리학회지』, 28(4): 65-83.
- [18] Blei, D. M. 2012. "Probabilistic topic models." *Communications of the ACM*, 55(4): 77-84.
- [19] Blondel, V. D. et al. 2008. "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.
- [20] Chellappa, R. K., Sambamurthy, V. and Saraf, N. 2010. "Competing in crowded markets: Multimarket contact and the nature of competition in the enterprise systems software industry." *Information Systems Research*, 21(3): 614-630.
- [21] Chen, C., Ibekwe-SanJuan, F. and Hou, J. 2010. "The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis." *Journal of the American Society for Information Science and Technology*, 61(7): 1386-1409.
- [22] Freeman, L. C. 1979. "Centrality in social networks conceptual clarification." *Social networks*, 1(3): 215-239.
- [23] Jiang, L. C. et al. 2015. "The divided communities of shared concerns: Mapping the intellectual structure of e-Health research in social science journals." *International journal of medical informatics*, 84(1): 24-35.
- [24] Lau, A. Y. and Coiera, E. W. 2007. "Do people experience cognitive biases while searching for information?" *Journal of the American Medical Informatics Association*, 14(5): 599-608.
- [25] Logan, D. W. et al. 2010. "Ten simple rules for editing Wikipedia." *PLoS Comput Biol*, 6(9): e1000941.
- [26] Manning, C. D. et al. 2014. "The Stanford CoreNLP natural language processing toolkit." *In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55-60.
- [27] Nishtar, S. et al. 2004. "Newspaper articles as a tool for cardiovascular prevention programs in a developing country." *Journal of health communication*, 9(4): 355-369.
- [28] Pautasso, M. 2013. "Ten simple rules for writing a literature review." *PLoS computational biology*, 9(7): 1-4.

- [29] Peddecord, K. M. et al. 2008. "Can movie theater advertisements promote health behaviors? Evaluation of a flu vaccination pilot campaign." *Journal of health communication*, 13(6): 596-613.
- [30] Peng, T. Q. et al. 2012. "Mapping the landscape of Internet studies: Text mining of social science journal articles 2000 -2009." *New Media & Society*, 1461444812462846.
- [31] Persson, O. 1994. "The intellectual base and research fronts of JASIS 1986 -1990." *Journal of the American Society for Information Science*, 45(1): 31-38.
- [32] Song, Min and Kim, Su Yeon. 2013. "Detecting the knowledge structure of bioinformatics by mining full-text collections." *Scientometrics*, 96(1): 183-201.
- [33] Su, H. N. and Lee, P. C. 2010. "Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight." *Scientometrics*, 85(1): 65-79.
- [34] Urquhart, C. et al. 2005. "Student use of electronic information services in further education." *International Journal of Information Management*, 25(4): 347-362.
- [35] Weingart, P. 2005. "Impact of bibliometrics upon the science system: Inadvertent consequences?" *Scientometrics*, 62(1): 117-131.
- [36] Webster, J. and Watson, R. T. 2002. "Analyzing the past to prepare for the future: Writing a literature review." *Management Information Systems Quarterly*, 26(2): 3.
- [37] Yan, E., et al. 2012. "Topics in dynamic research communities: An exploratory study for the field of information retrieval." *Journal of Informetrics*, 6(1): 140-153.
- [38] Yan, E. 2014. "Research dynamics: Measuring the continuity and popularity of research topics." *Journal of Informetrics*, 8(1): 98-110.
- [39] Yan, E. 2015. "Research dynamics, impact, and dissemination: A topic-level analysis." *Journal of the Association for Information Science and Technology*.
- [40] Yoon, Byungun, Lee, Sungjoo and Lee, Gwanghee. 2010. "Development and application of a keyword-based knowledge map for effective R&D planning." *Scientometrics*, 85(3): 803-820.
- [41] Zhang, W. et al. 2015. "Knowledge map of creativity research based on keywords network and co-word analysis, 1992 -2011." *Quality & Quantity*, 49(3): 1023-1038.
- [42] Zhao, D. and Strotmann, A. 2008a. "Author bibliographic coupling: Another approach to citation-based author knowledge network analysis." *Proceedings of the American Society for Information Science and Technology*, 45(1): 1-10.
- [43] Zhao, D. and Strotmann, A. 2008b. "Evolution of research activities and intellectual influences in information science 1996 -2005: Introducing author bibliographic-coupling analysis." *Journal of the American Society for Information Science and Technology*, 59(13): 2070-2086.

- [44] JITA Classification System of Library and Information Science. [online] [cited 2015. 8. 7.]
<<http://library.ifla.org/view/subjects/subjects.html>>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Kim, Panjun and Lee, Jae Yun. 2007. "Descriptor Profiling for Research Domain Analysis." *Journal of the Korean Society for Information Management*, 24(4): 285-303.
- [2] Kim, Hee-Jung. 2005. "A Study on Comparison of Intellectual Structure in Records Management and Archives Using Author Cocitation Analysis." *Journal of the Korean Society for Library and Information Science*, 39(3): 207-224.
- [3] Park, Ja-Hyun and Song, Min. 2013. "A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling." *Journal of the Korean Society for Information Management*, 30(1): 7-32.
- [4] Park, Ji-Yeon and Jeong, Dong Youl. 2013. "A Study on the Intellectual Structure of Library and Information Science in Korea by Author Bibliographic Coupling Analysis." *Journal of the Korean Society for Information Management*, 30(4): 31-59.
- [5] Seo, Eun-Gyoung. 1992. "Research Patterns of Information Retrieval in Information Science: The Changing Structure Across a Decade." *Journal of the Korean Society for Information Management*, 9(1): 55-82.
- [6] Seo, Eun-Gyoung. 1997. "An analytical study on research patterns in information science." *Journal of the Korean Society for Information Management*, 14(1): 269-291.
- [7] Seo, Eun-Gyoung. 2010. "Trends analysis on research articles in the Journal of Korean Society for Information Management." *Journal of the Korean Society for Information Management*, 27(4): 7-31.
- [8] Seo, Eun-Gyoung and Yoo, So-Young. 2013. "Detecting Research Trends in Korean Information Science Research, 2000-2011." *Journal of the Korean Society for Information Management*, 30(4): 215-239.
- [9] Sohn, Jung-Pyo. 2003. "An analytical study on research trends of library and information science in Korea: 1957~ 2002." *Journal of Korean Library and Information Science Society*, 34(3): 9-21.
- [10] Song, Jung-Sook. 2010. "Trends in Library and Information Science Research in Korea: Focused on Master's Theses and Doctoral Dissertations from 2001 to 2010." *Journal of Korean*

- Library and Information Science Society*, 41(4): 333-353.
- [11] Oh, Se-Hoon. 2005. "A study on the research trends of library & information science in Korea by analyzing journal articles and the cited literatures." *Journal of the Korean Society for Information Management*, 22(3): 379-408.
- [12] Yoo, Yeongjun. 2003. "A Study on Intellectual Structure of Library and Information Science in Korea." *Journal of the Korean Society for Information Management*, 20(3): 277-298.
- [13] Ryoo, Jong-duk and Choi, Eun-Ju. 2011. "A Comparison Test on the Potential Utility between Author Profiling Analysis (APA) and Author Co-Citation Analysis (ACA)." *Journal of the Korean Society for Information Management*, 28(1): 123-144.
- [14] Yoon, Koo-ho and Seo, Mal-Sook. 2001. "Intellectual Structure of Koran Library and Information Science in 1990s Using Author Co-citation Analysis." *Journal of Korean Library and Information Science Society*, 32(3): 169-197.
- [15] Chung, Jae-Young and Park, Jin-Hee. 2011. "Analysis of the trends in the field studies of library and information science in Korea." *Journal of Korean Library and Information Science Society*, 42(2): 171-191.
- [16] Chung, Jin-Sik. 2001. "An analytical study on research trends of library and information science in Korea: 1996-2000." *Journal of the Korean Society for Library and Information Science*, 35(3): 55-78.
- [17] Cho, Jane. 2011. "A Study for Research Area of Library and Information Science by Network Text Analysis." *Journal of the Korean Society for Information Management*, 28(4): 65-83.