

토픽 모델링 기반 정보학 분야 학술지의 학제성 측정 연구*

Topic Modeling based Interdisciplinarity Measurement in the Informatics Related Journals

진설아 (Seol A Jin)**

송민 (Min Song)***

초 록

본 연구는 인용 정보와 주제범주 분류체계를 기반으로 한 기존 하향식 접근법과 달리 문헌에 출현한 단어정보를 기반으로 세부주제를 자동 추출하는 토픽 모델링을 사용하여 학제성을 측정하였다. JCR 2013의 Information & Library Science 주제범주에서 5년 영향력 지수 상위 20개 학술지의 최근 5년 동안의 논문 제목과 초록 텍스트를 분석대상으로 사용하였다. 학제성을 측정하기 위한 지수로 '분야적 다양성'을 나타내는 Shannon 엔트로피 지수와 Stirling 다양성 지수, '네트워크 응집성'을 나타내는 지수로는 토픽 네트워크의 평균 경로길이를 사용하였다. 계산된 다양성과 응집성 지수를 통해 학제성의 유형을 분류한 후 각 유형을 대표하는 학술지들의 토픽 네트워크를 비교하였다. 이를 통해 본 연구의 텍스트 기반 다양성 지수는 기존의 인용정보 기반 다양성 지수와 다른 양상을 보이고 있어 상호보완적으로 활용될 수 있으며, 다양성과 응집성을 모두 고려하여 분류된 각 학술지의 토픽 네트워크를 통해 개별 학술지가 다루는 세부주제의 특성과 연결 정도를 직관적으로 파악할 수 있었다. 이를 통해 토픽 모델링을 통한 텍스트 기반의 학제성 측정이 학술지의 학제성을 나타내는 데에 다양한 역할이 가능함을 확인하였다.

ABSTRACT

This study has measured interdisciplinarity using a topic modeling, which automatically extracts sub-topics based on term information appeared in documents group unlike the traditional top-down approach employing the references and classification system as a basis. We used titles and abstracts of the articles published in top 20 journals for the past five years by the 5-year impact factor under the category of 'Information & Library Science' in JCR 2013. We applied 'Discipline Diversity' and 'Network Coherence' as factors in measuring interdisciplinarity; 'Shannon Entropy Index' and 'Stirling Diversity Index' were used as indices to gauge diversity of fields while topic network's average path length was employed as an index representing network cohesion. After classifying the types of interdisciplinarity with the diversity and cohesion indices produced, we compared the topic networks of journals that represent each type. As a result, we found that the text-based diversity index showed different ranking when compared to the reference-based diversity index. This signifies that those two indices can be utilized complementarily. It was also confirmed that the characteristics and interconnectedness of the sub-topics dealt with in each journal can be intuitively understood through the topic networks classified by considering both the diversity and cohesion. In conclusion, the topic modeling-based measurement of interdisciplinarity that this study proposed was confirmed to be applicable serving multiple roles in showing the interdisciplinarity of the journals.

키워드: 텍스트마이닝, 토픽모델링, 학술지, 학제성, 다양성 지수

textmining, topic modeling, scholarly journal, interdisciplinarity, diversity index

* 이 논문은 2015학년도 연세대학교 미래선도연구사업(부분적인)지원에 의하여 작성된 것임(2015-22-0119).

** 과학기술정책연구원 연구원(sula.jin@gmail.com)

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

■ 논문접수일자: 2016년 2월 2일 ■ 최초심사일자: 2016년 2월 24일 ■ 게재확정일자: 2016년 3월 11일

■ 정보관리학회지, 33(1), 7-32, 2016. [http://dx.doi.org/10.3743/KOSIM.2016.33.1.007]

1. 서론

최근의 다양하고 복잡한 현대사회의 문제들을 해결하기 위해 관련 학문 분야가 통합된 시각에서 문제를 고찰하는 학제적 연구는 계속해서 발전하고 있다. 국가 및 연구기관에서뿐만 아니라 학제적 과목 및 전공 개설, 연계전공 신설과 같은 대학의 움직임까지 학제적 연구에 대한 지원이 그를 입증하고 있다.

그 중에서도 정보학 분야는 다수의 연구를 통해 학제적 분야로서 인식되고 있다. 정보학 분야 275개의 학술지와 학술대회 논집을 대상으로 30년 동안 더 넓은 주제로의 인용이 증가하고 있음을 밝힌 Cronin과 Meho(2008), 정보학 분야 학술지인 JASIST의 공저분석을 통해 도서관학, 정보학 외에 소속된 저자들이 나타남을 확인한 Chua와 Yang(2008), 2002년부터 2006년까지의 문헌정보학 분야 박사학위 논문의 인용 분석을 실행한 결과 문헌정보학이 경영학, 컴퓨터학, 교육학, 커뮤니케이션, 시스템, 정보기술, 정보산업, 정보경영 등의 분야와 관련을 맺고 있는 것을 확인한 Prebor(2010) 등이 그 예이다. 또한 Levitt과 Thelwall(2009)는 Web of Science의 문헌정보학 분야에서 피인용 상위 0.1%에 해당하는 논문 82편의 참고문헌을 대상으로 인용비율을 분석한 결과 문헌정보학 외의 다른 주제분야를 함께 가지는 논문들의 인용이 훨씬 많이 이루어졌음을 확인하였다. 이를 통해 정보학의 학제적 연구가 가지는 영향력 또한 확인하였다.

학제성의 측정이 정보시스템과 서비스를 설계하는 데에 영향을 미친다고 언급한 Qin, Lancaster, Allen(1997)과 같이 학제적 성향이 강한 학문 분

야에 대한 보다 바람직한 정보서비스 제공의 필요성이 대두되고 있다. 분야의 학제적 성향이 강해짐에 따라 학문분야 차원에서뿐만 아니라 학술지 또한 다양한 주제를 포함하게 되었으며 학제적 분야의 연구자들이 본인의 연구영역 및 이용 목적에 따라 적절하게 학술지를 선정할 수 있도록 하는 학술지 차원의 학제성 측정이 필요한 것이다.

이에 본 연구에서는 학제적 분야로 인식되는 정보학 분야의 학술지를 대상으로 지금까지 행해져 온 인용정보 기반의 분석이 아닌 텍스트 마이닝 기법 중 하나인 토픽 모델링을 사용하여 내용 기반의 학제성을 측정하고자 한다. 또한 본 연구의 텍스트 기반 분석이 기존 인용정보 기반 분석의 단점을 보완하고 학술지의 학제성을 나타내는 지표로 활용 가능한 지를 알아보는 것을 목적으로 한다.

위와 같은 텍스트 기반 학술지의 학제성 측정을 통해 개별 학술지들이 다루는 다양한 세부주제의 특성을 파악하고 각각의 학제적 유형을 분석함에 따라 연구자의 이용 목적을 만족시키고 학문분야의 학제적 발전을 촉진할 수 있는 정보서비스를 제공할 수 있을 것으로 기대된다.

2. 이론적 배경

2.1 학제성의 정의

2005년 미국 국립 과학원(US National Academy of Science)을 중심으로 한 연구 프로젝트 '학제적 연구를 촉진하기 위한 연구'의 보

고서에서 발표한 학제적 연구(Interdisciplinary Research)의 정의는 다음과 같다.

“학제적 연구는 두 개 이상의 전문 지식 체계나 연구 관행으로부터 관점/개념/이론 또는 도구/기법 또는 정보/자료를 통합하여 팀 또는 개인이 수행하는 연구 유형이다”

가장 널리 사용되는 학제성의 3가지 유형으로는 OECD(1998)가 정의한 다학문성, 간학문성, 초학문성이 있다. 이러한 학제성의 유형에 대해 여러 연구를 종합하여 정의하면 다학문성(Multidisciplinarity)은 상이한 학문분야 각 부분은 원래의 정체성을 유지하며 전문적인 관점을 공유하고 연구 생산물은 각 부분의 단순한 합을 의미한다. 즉 학문분야 간 통합이 성립되기 이전 단계의 연구를 의미한다. 간학문성(Interdisciplinarity)은 통합적인 합성(integrative synthesis)을 중요시하며 이론적, 개념적, 방법론적으로 어느 정도 통합이 이루어져 각 부분들을 합한 것보다 더 큰 단계의 연구를 의미한다. 마지막으로 초학문성(Transdisciplinarity)은 다학문적, 간학문적의 수준을 넘어 각 학문분야의 인식론적 상호 통합이 수반되어야 하는 가장 포괄적인 프레임워크이며 지식 생산의 새로운 유형을 함축하고 있다. 하지만 이러한 학제성의 정도를 분명하게 구분하는 방법은 아직까지 규명된 것 없이 위와 같은 정의를 바탕으로 각각의 조작적 정의 하에서 이루어지고 있다.

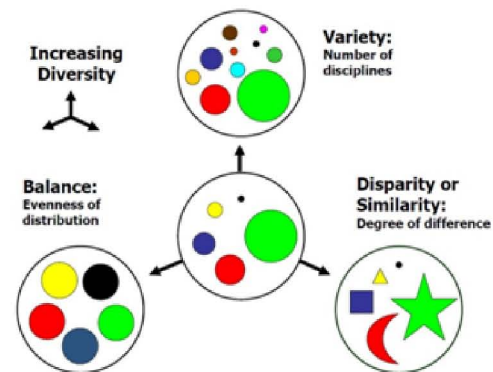
본 연구에서는 완전히 통일되지 않은 학제성의 정의들을 포괄하는 공통적 키워드로서 ‘다양성(Diversity)’과 ‘통합(Integration)’을 선정하였고 이를 지식 통합(Knowledge Integration)

차원의 학제성 측정 요소로 제시하는 Rafols와 Meyer(2010)의 개념을 바탕으로 접근하고자 한다.

2.2 학제성의 측정

2.2.1 다양성(Diversity)

Rafols와 Meyer(2010)가 제시한 지식 통합 차원의 학제성을 측정하는 관점 중 하나인 ‘분야적 다양성(Discipline Diversity)’은 다음의 세 가지 요소를 포함하는 Stirling(1998)의 이론을 따르고 있다(〈그림 1〉 참조).



〈그림 1〉 다양성(Diversity) 개념의 도식화 (Stirling, 1998, p. 41)

‘식별 가능한 차이를 가지는 종류의 개수’를 뜻하는 유형의 수(Variety)는 다양성의 가장 직관적이고 기본적인 요소이다. 균형(Balance)은 메시지가 갖는 정보의 양을 확률 분포의 형태로 수량화하여 분포가 균등하게 나타날수록, 즉 불확실성이 높을수록 얻을 수 있는 정보의 양이 많다는 엔트로피(Entropy)의 개념과 맥락을 같이 한다. 카테고리의 분포가 균등하게 나타날

수록 얻을 수 있는 다양성은 커진다는 의미로 적용될 수 있다. 거리 혹은 유사도(Disparity or Similarity)가 다양성의 요소로서 고려되어야 하는 이유에 대해 Bache, Newman, Smyth(2013)는 5마리의 딱정벌레, 5마리의 거미가 모여 있는 A그룹과 5마리의 딱정벌레, 5마리의 코끼리가 모여 있는 B그룹 중 인지적 분류체계를 기반으로 다양성이 보다 높게 판단되는 것은 이질성이 더 높은 것끼리 모여 있는 B그룹이라는 예를 들었다. 하지만 다양성의 3가지 요소가 개별적으로 다양성을 나타내기에는 부족하다.

이러한 다양성을 측정하는 지수 중에는 학제성 지수로서 적절하다고 판단되어 쓰이는 지수들이 있다. 브릴로엔 지수(Brillouin, 1956), Shannon 엔트로피 지수(Shannon, 1948), Simpson 다양성 지수(Simpson, 1949), Stirling 다양성 지수(Stirling, 2007)가 대표적인 그 예이다(〈표 1〉참조).

〈표 1〉 학제성 지수로 사용되는 대표적인 4가지 다양성 지수의 공식

브릴로엔 지수	$H = \frac{\log N! - \sum (\log n_i)}{N}$ N: 전체 참고문헌의 수, n_i: 개별 학문 i에 속하는 참고문헌의 수
Shannon 엔트로피 지수	$H = - \sum_i p_i \ln p_i$ p_i: 카테고리 I의 분포
Simpson 다양성 지수	$I = 1 - \sum_i p_i^2$
Stirling 다양성 지수	$\Delta = \sum_{i,j(i \neq j)} (d_{ij})^\alpha (p_i p_j)^\beta$ d_{ij}: 카테고리 i와 j간의 거리 α, β: 변수 파라미터

이 중 브릴로엔 지수(H), Shannon 엔트로피 지수(H), Simpson 다양성 지수(I)는 다양성의 요소 중 유형의 수(Variety)와 균형(Balance)

을 고려하는 지수이다. 학제성 연구에서 브릴로엔 지수는 인용된 주제분야 각각의 참고문헌 수(n_i)를 반영함으로써 개별 주제분야의 분포가 영향을 미치는 지수로 사용되어 왔다(Tang, 2004). Shannon 엔트로피 지수와 Simpson 다양성 지수 또한 카테고리 간의 분포(p_i)를 이용하는 지수로서 브릴로엔 지수와 같이 주로 인용된 주제범주의 분포를 통해 계산되어 왔다(Carayol & Thi, 2005; Leydesdorff & Rafols, 2011).

Stirling 다양성 지수는 본인이 정의한 다양성의 3가지 요소를 모두 고려하여 나타낸 지수로서 각 카테고리의 분포(p_i)와 카테고리 간의 거리(d_{ij})를 변수 파라미터 α, β 로 조절하는 휴리스틱 지수이다. 따라서 학문, 기술, 혁신분야 등의 다양성을 측정할 수 있는 일반적 프레임워크로 제안되었다. 이후 대부분의 학제성 연구(Porter & Rafols, 2009; Rafols & Meyer, 2010; Rafols, Leydesdorff, O'Hare, Nightingale, & Stirling, 2012; Leydesdorff, Rafols, & Chen, 2013)에서는 α, β 를 각각 1로 정의하여 사용되고 있다.

2.2.2 통합(Integration)

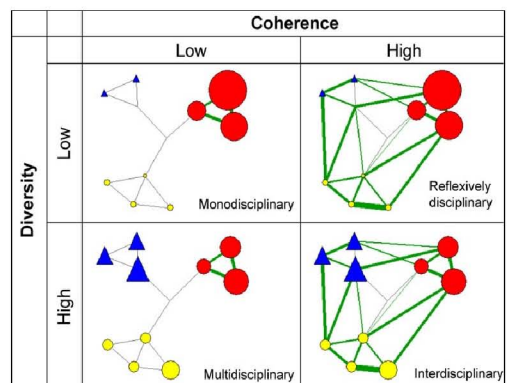
앞서 언급한 Stirling의 다양성을 의미하는 분야적 다양성(Discipline Diversity)과 함께 통합의 관점에서 학제성을 측정하기 위해 제시된 것은 '네트워크 응집성(Network Coherence)'이다. 경제학적인 관점에서 응집성이 높은 구조는 개혁의 성공가능성을 높여주는 중요한 요인이며, 이를 학문체계의 관점에서 바라보면 학문분야의 변화, 발전, 퇴화 등을 예측할 수 있는 요소가 될 수 있다. 또한 학제적 연구의 관점에서는 통합의 정도에 따라 분석 대상이 학제성의 유형 중 어느 곳에 속해 있는지를 측정할 수 있다.

하지만 응집성의 측정은 아직까지 명확히 규명된 것 없이 다양한 방법을 통해 시도되고 있는 실험적 단계이다. Rafols, Leydesdorff, O'Hare, Nightingale, & Stirling(2012)과 Morillo, Bordons, Gomez(2003)은 주제범주 A와 B에 모두 속해 있는 학술지의 개수를 이용하여 각 주제범주 간의 연결강도를 계산하였다. 또한 카테고리들을 노드로 한 네트워크 구조의 밀집된 정도를 계산하여 응집성을 측정한 연구도 있다. Rafols와 Meyer(2010)은 네트워크의 에지 값(weight)의 평균으로 산출하는 평균 연결강도(Average Linkage Strength)와 두 노드 사이를 연결하는 가능한 모든 경로 중 가장 짧은 경로길이의 평균을 나타내는 평균 경로길이(Average Path Length)를, Nichols(2014)는 네트워크의 밀도(Density)를 계산하여 카테고리 집단의 응집성을 측정하였다.

2.2.3 지식 통합(Knowledge Integration) 차원의 학제성

〈그림 2〉는 Rafols et al.(2012)이 도식화한 다양성과 응집성의 개념을 기준으로 나뉘어진 학제성의 유형이다. 네트워크의 노드(node)는 세부 카테고리를 의미하며 노드의 모양과 색깔은 각기 다른 카테고리를 나타낸다. 노드의 크기는 카테고리가 전체에서 차지하는 비중을 나타내는 것으로서 연구목적에 맞게 정의될 수 있다. 각 노드들을 연결하는 에지(edge)는 카테고리 페어 간의 유사도를, 굵게 표시된 선은 직접연결을 나타낸다. 이러한 네트워크의 노드와 에지의 특성이 나타내는 다양성과 응집성을 기준으로 단일학문(Monodisciplinary), 다학문성(Multidisciplinary), 간학문성(Interdisciplinary)

의 측정이 가능하다고 제시하였다. 다양하지 않은 카테고리가 높은 응집성으로 연결된 1부분면 구간은 혼하지 않은 연구형태로서 채귀적 학문성(Reflexively disciplinary)으로 정의하였다.



〈그림 2〉 다양성(Diversity)과 응집성(Coherence)을 고려한 학제성의 유형(Rafols et al., 2012, p. 1266)

3. 선행연구

대부분의 학제성 측정 연구는 Porter와 Chubin(1985)이 다른 주제분야의 인용이 학제적 연구활동의 지표로서 사용 가능하다는 것을 증명함으로써 인용 정보를 중심으로 이루어져 왔다. 그 중 개별 분야의 문헌에 인용된 다른 주제분야의 비율로 학제성을 측정하는 방법이 가장 기본적으로 사용되어 왔다(Herring, 1999; Larivière & Gingras, 2010; Rinia, van Leeuwen, & van Raan, 2002; Levitt & Thelwall, 2009).

균형(Balance)의 개념을 고려한 지수들을 사용하여 학제성을 측정한 연구들은 다음과 같다. Morillo, Bordons, Gomez(2001)은 ISI 주제범주를 기반으로 학술지에 부여되는 복수의

(multi-assigned) 주제범주가 타당한 지를 알아보기 위해 인용문헌 주제범주의 분포 정도와 이를 수치화한 프랫(Pratt) 지수를 계산하였다. Adams, Jackson, Marshall(2007)은 영국의 연구 집중 대학에서 출판된 7년간의 논문을 대상으로 참고문헌 주제범주의 비율, 다른 주제분야의 수, Shannon 엔트로피 지수를 비교하여 학제성을 측정하였다. Tang(2004)은 25년간 문헌정보학 분야의 학제성의 변화를 살펴보기 위해 인용비율과 브릴로엔 지수를 산출한 결과 인용비율은 1990년을 기점으로 크게 증가한 반면 브릴로엔 지수는 적은 증가 폭으로 지속적으로 증가하고 있음을 확인하였다. 박소윤(2013)은 문헌정보학 분야 10종의 학술지의 학제성을 측정하기 위해 인용비율, 학제성 차용지수, 브릴로엔 지수를 사용하였다. 그 결과 인용비율과 학제성 차용지수는 비슷한 순위를 보인 반면 브릴로엔 지수는 앞의 지수들과 다른 양상을 보였다. 이와 같이 인용의 주제분야 분포를 의미하는 균형을 고려함으로써 학제성의 정도가 다르게 나타나는 양상을 확인할 수 있었다.

Stirling의 다양성 개념을 사용하여 학제성을 측정한 연구들 또한 대부분 ISI 주제범주와 각 주제범주 간의 인용정보를 사용하여 연구가 이루어졌다. Bordons, Morillo, Gomez(2004)는 인용행위를 비슷한 주제분야 간의 내부적 인용과 거리가 먼 주제분야 간의 외부적 인용으로 나누어 주제분야 간의 거리를 구하고 이를 학술지에 함께 부여된 주제범주 간의 비율과 분석하여 학문분야의 학제성을 측정하였다. Porter와 Rafols(2009) 또한 주제분야 간 동시인용 행렬을 코사인 유사도로 정규화하여 주제분야 간의 유사도를 계산하였고 이를 통해 Rao-stirling

지수를 산출하였다. 이를 통해 30년 동안 6개 학문분야의 학제적 변화 추이를 분석하면서 기존의 학제성 지수로 사용되어 온 인용비율과 비교하였다. 그 결과 두 지표 모두 값이 점점 증가함을 확인할 수 있었으나 인용비율이 50% 이상 증가한 반면 Rao-stirling 지수는 5%의 증가를 보였다. 이는 Rao-stirling 지수가 인용비율보다 복합적인 학제성의 측면을 반영한다는 것을 나타낸다. Yegros-Yegros, Amat, D'Este, Porter, Rafols(2010)는 스페인의 식품공학 프로그램의 지원을 받는 연구자들의 논문을 대상으로 인용 분석을 실행하여 참고문헌 주제분야의 수, Shannon 엔트로피 지수, Stirling 다양성 지수의 상관관계를 분석하였다. 그 결과 Shannon 엔트로피 지수와 Stirling 다양성 지수가 서로 상관관계가 있음이 밝혀졌다.

본 연구가 선정한 학제성 측정의 두 가지 개념을 제안한 Rafols와 Meyer(2010)는 바이오나노학(bionanoscience) 논문을 대상으로 Shannon 엔트로피 지수, Simpson 다양성 지수, Rao-stirling 지수와 평균 연결강도를 측정하였다. 그 결과 다양성과 응집성은 상관관계를 보이지 않았으나 직교적(orthogonal) 관계를 가진 지표로서 분석되었으며 네트워크의 응집성은 문헌과 같은 미시적 차원의 학제성을 식별해주고 있다고 언급하였다. Rafols et al.(2012) 또한 다양성과 응집성을 함께 고려하여 연구기관 별 학제성을 분석하였다.

Leydesdorff와 Rafols(2011)는 학술지의 학제성 순위를 매기기 위해 JCR 2008의 LIS 분야 61개 학술지를 대상으로 학술지 별 매개 중심성(Betweenness Centrality), Shannon 엔트로피 지수, Gini 계수, Rao-stirling 지수를 비

교하였다. 그 결과 각 지수들은 서로의 특성에 따라 각기 다른 순위를 보였고 이는 각각의 지수들이 학제성의 다차원적인 개념을 표현해주고 있다고 언급하였다.

하지만 인용 분석은 대부분 기존 데이터베이스의 분류체계를 기반으로 논문의 주제범주를 분류하기 때문에 하나의 학술지에 게재된 모든 논문을 동일한 주제영역으로 분류하고 각각의 주제를 표현하지 못한다는 한계점을 가진다. 이와 같이 하향식(Top-down) 접근법을 취하는 인용 분석의 한계점은 다양한 세부분야의 논문을 포함하는 학제적 분야의 학술지에서도 나타날 수 있으며 이를 보완하기 위해 상향식 접근법(Bottom-up)의 용어분석을 실행한 연구들이 있다.

용어분석 연구 또한 동시출현단어 분석, 토픽 모델링과 같은 텍스트 마이닝 기법을 사용한 상향식 접근법으로 문헌의 주제를 분석하는 연구들이 많이 이루어지고 있다. 강범일과 이재운(2014)은 논문의 제목, 초록, 키워드를 대상으로 동시출현단어 분석을 실행하여 국내 트위터 관련 연구의 세부주제 영역을 파악하였으며, 박자현과 송민(2013)은 국내 문헌정보학 분야의 주요 4종 학술지 논문의 초록을 대상으로 토픽 모델링을 실행하여 각 학술지들이 다루고 있는 세부주제를 비교하였다.

이러한 연구를 기반으로 토픽 모델링을 사용하여 학제성을 규명하는 연구들이 최근 나타나고 있다. Bache, Newman, Smyth(2013)는 논문의 전문(full-text)을 대상으로 토픽 모델링을 실행하여 추출된 문헌 별 토픽의 확률, 단어 별 토픽의 확률과 코사인 유사도를 사용하여 다양성 지수를 산출하였다. 이 연구를 통해 미리 마련된 주제범주의 매핑 없이 텍스트를 기

반으로 문헌 내용의 다양성을 측정할 수 있다는 가능성을 확인하였다. Nichols(2014)는 국가연구재단(National Science Foundation)의 사회, 행동, 경제과학 부서의 핵심 분야 6개의 학제성을 분석하기 위해 모든 주제분야의 연구 포트폴리오를 대상으로 학습해 놓은 토픽 모델링 결과를 사용하였다. 그 결과 활용한 방법론이 분석 대상의 학제성을 비교적 명확하게 측정하였으며 복합적 개념인 학제성의 미묘한 차이를 다양한 관점에서 파악할 수 있었다고 언급하였다. 하지만 Stirling 다양성 지수를 산출을 위해 토픽 모델링의 결과 값을 그대로 사용하지 않고 토픽 확률을 기반으로 각 분야의 기여도와 분야 간의 거리를 임의적으로 부여하였다.

따라서 본 연구에서는 사람이 키워드를 부여하는 등의 수작업 과정 없이 문헌이 담고 있는 텍스트 자체만을 사용한다는 점, 적절하게 나누어진 토픽의 제공뿐 아니라 해석 가능한(human-readable) 설명을 제공해준다는 점, 이러한 모든 과정이 반자동(semi-automatic)으로 제공된다는 점, 또한 결과를 통해 일반화된 값의 산출뿐만 아니라 시각화 또한 가능하다는 점 등의 장점을 가진 토픽 모델링 기법(Mann, Mimno, & McCallum, 2006; Bache, Newman, & Smyth, 2013)을 사용하여 학제성을 분석하고자 한다. 또한 다양성(Diversity)과 통합(Integration)을 포함하는 복합적 관점의 학제성을 측정하는 지표로서 본 연구에서 사용할 지수는 정규화 과정을 거쳐 시간의 변화나 분석대상의 규모에 상관없는 분석이 가능한 Shannon 엔트로피 지수와 Stirling 다양성 지수, 토픽 네트워크의 평균 경로길이(Average Path Length)로 선정하였다.

4. 토픽 모델링을 이용한 학제성 측정

4.1 분석 대상

본 연구에서는 다수의 선행연구를 통해 이미 학제적 분야로 인식되어 오고 있는 정보학 분야를 분석 대상으로 선정하였다. 따라서 2013 JCR(Journal Citation Report) 사회과학 분야(Social Science Edition)의 Information Science & Library Science 주제범주에서 5년 영향력 지수(5-Year Impact Factor)를 기준으로 한

상위 20종의 학술지를 분석 대상으로 선정하였다. 이와 같이 선정된 학술지의 최근 5년 동안(2009~2013년)의 논문 6,551개 중 초록이 없는 논문을 제외한 총 6,545개 논문의 제목과 초록을 Web of Science를 통해 수집하였다. 후의 비교분석에 활용될 인용 분석을 위해 해당 논문의 참고문헌 또한 수집하였다.

이 중 2012, 2013년에 발행한 논문이 없는 학술지 1종을 제외한 분석 대상 학술지는 <표 2>와 같으며 앞으로의 학술지 표기는 간략 서명을 사용한다.

<표 2> 분석대상으로 선정된 정보학 분야 20개의 학술지

5년 영향력 지수 순위	학술지 완전서명	학술지 간략서명	논문 수 (2009~2013)
1	MIS QUARTERLY	MIS QUART	226
2	JOURNAL OF INFORMATION TECHNOLOGY	J INF TECHNOL	109
3	JOURNAL OF COMPUTER-MEDIATED COMMUNICATION	J COMPUT-MEDIAT COMM	169
4	INFORMATION SYSTEMS RESEARCH	INFORM SYST RES	259
5	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	J AM MED INFORM ASSN	722
6	JOURNAL OF INFORMETRICS	J INFORMETR	318
7	INFORMATION & MANAGEMENT	INFORM MANAGE-AMSTER	237
8	JOURNAL OF MANAGEMENT INFORMATION SYSTEMS	J MANAGE INFORM SYST	189
9	JOURNAL OF STRATEGIC INFORMATION SYSTEMS	J STRATEGIC INF SYST	101
10	JOURNAL OF THE ASSOCIATION FOR INFORMATION SYSTEMS	J ASSOC INF SYST	150
11	INFORMATION SYSTEMS JOURNAL	INFORM SYST J	104
12	EUROPEAN JOURNAL OF INFORMATION SYSTEMS	EUR J INFORM SYST	191
13	INTERNATIONAL JOURNAL OF COMPUTER-SUPPORTED COLLABORATIVE LEARNING	INT J COMP-SUPP COLL	108
14	INFORMATION AND ORGANIZATION	INFORM ORGAN-UK	59
15	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	J AM SOC INF SCI TEC	913
16	JOURNAL OF HEALTH COMMUNICATION	J HEALTH COMMUN	484
17	SCIENTOMETRICS	SCIENTOMETRICS	1124
18	INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT	INT J INFORM MANAGE	312
19	GOVERNMENT INFORMATION QUARTERLY	GOV INFORM Q	291
20	INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE	INT J GEOGR INF SCI	485

4.2 분석 방법

4.2.1 텍스트 전처리

수집된 논문의 제목과 초록을 대상으로 토픽 모델링을 실행하기 위해 기본적인 전처리를 실행하였다. 기본적인 전처리와 토픽 모델링은 모두 Stanford Natural Language Processing Group에서 제공하는 Stanford Topic Modeling Tool(TMT)를 통해 이루어졌다. 전처리 과정은 먼저 마침표와 띄어쓰기를 기준으로 문자열을 분리(tokenizing)한 후 모든 단어를 소문자로 바꾸어 주었다. 기호를 비롯한 단어와 숫자가 아닌 것들과 알파벳이 3개 미만인 단어들은 제거하였다. 이후 관사 등을 포함한 기본적인 영어 불용어를 제거하고 남은 단어들의 출현빈도를 계산하였다. 문헌 내 단어들의 출현빈도는 대부분 멱함수의 분포를 따르기 때문에 높은 출현빈도의 단어들은 주로 문헌이 포함하고 있는 주제의 차이를 반영하지 못한다. 따라서 기본 전처리를 거친 문헌에서 가장 출현빈도가 높은 50개의 단어를 바탕으로 그 중 주제의 차이를 나타내지 못한다고 판단된 단어와 이의 변형된 형태를 추가 불용어로 선정하여 제거하였다.

4.2.2 토픽 분포와 유사도 산출

전처리 과정을 모두 거친 텍스트를 바탕으로 토픽들의 분포와 유사도를 산출하기 위해 토픽 모델링을 실행하였다. Gibbs 샘플링을 사용한 LDA 알고리즘을 적용하였으며 1,500번의 iteration을 실행하였다. 문헌에 부여된 임의 토픽들의 비율(θ)과 임의의 토픽에 대한 단어의 비율(β)을 결정하는 파라미터 값은 모두 0.01로 설정하였으며 분석 대상 문헌 집단이 포함하고

있는 주제들을 가장 적절하게 나타내는 토픽 구성을 찾기 위해 30, 50, 100개의 토픽 수를 지정하여 각각 실행하였다. 이러한 과정을 거쳐 문헌 별 토픽 분포 값이 산출되고, 학술지에 속한 모든 문헌의 토픽 별 분포 값의 평균을 구해 학술지의 토픽 별 분포 값을 구하였다. <그림 3>은 개별 학술지의 토픽 분포 값 산출 예시

그림이다. $\frac{\sum_{i=1}^N p_{di}}{N}$ 은 학술지 1에 속한 문헌 N개의 각 토픽 별 분포 값($p_{d1} \dots p_{dN}$)을 모두 더한 값의 평균 값이다.

	Topic 1	Topic 2	...	Topic m
Doc 1	p_{d1}			
Doc 2	p_{d2}			
...	...			
Doc N	p_{dN}			
Journal 1	$\frac{\sum_{i=1}^N p_{di}}{N}$			

<그림 3> 개별 학술지의 토픽 분포 값 산출 예시

다음으로 토픽 간의 유사도를 산출하기 위해서는 토픽 별로 추출된 단어 정보를 이용하였다. 문헌집단의 세부주제를 나타내기 위해 추출된 각 토픽은 그 토픽을 나타내는 20개의 단어들과 함께 추출되었다. 이 단어들은 각 토픽 내에서의 출현빈도 값(Term Frequency) 또한 계산되었다. 이를 기반으로 역문헌빈도(Inverted Document Frequency)의 개념을 응용한 역토픽빈도(Inverted Topic Frequency) 값을 계산하여 TF*ITF 기반 토픽 x 단어 행렬을 만들었다.

이를 기반으로 코사인 유사도(cosine similarity)를 계산하여 토픽 간의 유사도를 산출하였다.

단어의 단순 출현빈도가 아닌 역토픽빈도 값을 사용함으로써 개별 토픽을 식별해주는 능력이 높은 단어에 가중치를 주었고 코사인 유사도를 사용함으로써 학술지마다 다른 게재 문헌수와 문헌의 길이를 정규화하였다.

4.2.3 학술지 별 다양성과 응집성 측정

산출된 학술지 별 토픽 분포와 토픽 간의 유사도를 바탕으로 개별 학술지의 다양성과 응집성을 측정하였다. 먼저 다양성을 측정하기 위해 Shannon 엔트로피 지수(H)와 Stirling 다양성 지수(Δ)를 계산하였다. 두 공식에 모두 쓰이는 p_i 는 학술지 별 토픽 i의 분포를 의미하며 앞의 <표 1>에 설명한 Shannon 엔트로피 지수를 산출하였다. Stirling 다양성 지수(Δ)는 선행된 학제성 연구들(Porter & Rafols, 2009; Rafols et al., 2012)에서와 같이 카테고리의 분포와 거리를 조절하는 파라미터 $\alpha, \beta = 1$ 로, 거리(d_{ij}) = (1-cosine)으로 하여 아래의 공식과 같이 계산하였다.

$$\Delta = \sum_{i,j} (1 - S_{ij}) p_i p_j$$

p_i : 개별 학술지가 포함하는 토픽 i의 분포

s_{ij} : 토픽 i와 j의 유사도

다음으로 학술지 별 응집성을 측정하기 위해 먼저 오픈소스 시각화 도구인 Gephi를 이용하여 토픽 네트워크를 작성하였다. 앞서 문헌 별 토픽 분포의 평균으로 구한 학술지 별 토픽의 분포 값이 0을 가지는 토픽이 없기 때문에 학술

지 간 응집성의 차이를 측정하기 위해 임계치를 설정하였다. 2개 이상의 학술지가 모든 토픽 분포 값을 가지지 않기 위한 최소값 0.002를 임계치로 설정하여 그 이상의 분포 값을 가지는 토픽을 네트워크의 노드(node)로 하였다. 또한 토픽 간의 유사도를 구하는 과정에서 나온 토픽 쌍을 에지(edge)로 하여 토픽 네트워크를 만들었으며 이를 통해 구해지는 각 네트워크의 평균 경로길이(Average Path Length)를 계산하여 개별 학술지의 응집성을 측정하였다.

마지막으로 본 연구의 텍스트 기반 학제성과 지금까지 주로 사용된 방법인 인용 분석을 통한 학제성의 비교분석을 위해 참고문헌 정보를 이용한 Shannon 엔트로피 지수를 산출하였다. 분석 대상 학술지의 참고문헌을 주제범주로 매핑하기 위해 선행연구(박소윤, 2013)의 분류기준을 사용하였다. 박소윤(2013)은 주제분야의 수에 영향을 받는 학제성 지수의 값이 실제보다 더 커질 가능성을 완화하기 위해 250개 이상의 소분류 수준으로 구분되어 있는 Web of Science(WOS) 주제범주를 34개의 중분류 수준으로 그룹화하였다. 이 34개의 중분류 주제범주를 이용하여 WOS의 등재 여부를 기준으로 필터링된 151,845개 참고문헌의 주제분야를 매핑하였다. 예를 들어, 'COMPUTER SCIENCE, INFORMATION SYSTEMS'로 분류된 학술지 MIS QUARTERLY와 'COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS'로 분류된 학술지 'JOURNAL OF AMERICAN MEDICAL INFORMATICS ASSOCIATION'은 동일하게 'COMPUTER SCIENCE'로 주제분야가 매핑되었다. 이를 통해 Shannon 엔트로피 지수 산출에 필요한 p_i 를 산출하였다. 인용

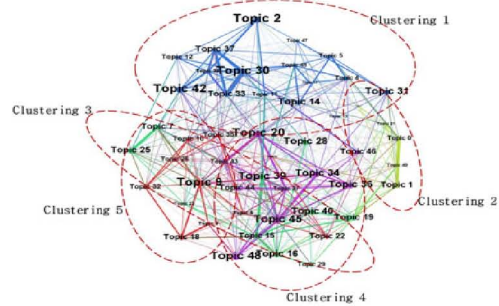
된 주제분야의 분포를 의미하는 p_i 는 '개별 주제 분야 i 의 참고문헌 수 / 전체 참고문헌 수'로 계산하였다.

5. 분석 결과

5.1 정보학 분야 학술지 전체 토픽 분석

분석 대상인 정보학 분야 학술지 20종이 담고 있는 다양한 내용을 적합하게 나타내는 토픽을 찾기 위해 30, 50, 100개의 토픽 수를 지정하여 토픽 모델링을 각각 실행하였다. 그 결과 본 연구에서는 50개의 토픽이 단어의 구성과 유사도 관점에서 가장 적합하다고 판단되어 분석에 사용되었다. 토픽 모델링 결과로 추출된 50개의 토픽과 각 토픽을 구성하는 20개의 단어들을 TF*ITF 값이 높은 순으로 정렬한 표는 <부록>과 같다.

50개의 토픽을 노드(node)로, 코사인 유사도 0값을 제외한 523개의 토픽 쌍을 에지(edge)로 하여 작성된 방향성이 없는(undirected) 전체 토픽의 네트워크는 <그림 4>와 같다. 노드의 크기는 몇 개의 토픽과 연결되었는지를 나타내는 연결정도 중심성(Degree Centrality)을 나타내며 에지의 굵기는 연결된 두 토픽의 유사도를 의미한다. 또한 토픽 네트워크의 클러스터링은 Modularity 알고리즘을 사용하여 이루어졌다. 이 알고리즘은 임의로 설정된 커뮤니티 각각에서 특정 노드의 에지 가중치 값이 더 이상 증가하지 않을 때까지 반복하여 계산한 modularity 값을 기준으로 노드들을 결합하여 네트워크의 커뮤니티를 추출하는 방법이다.



<그림 4> 정보학 분야 학술지 20종의 전체 토픽 50개 네트워크

5.1.1 클러스터링 분석

먼저 Modularity를 통해 총 5개의 그룹으로 클러스터링 된 각 토픽 그룹과 각 그룹의 토픽에서 공통적으로 출현하는 대표 단어를 선정하였다(<표 3> 참조). 먼저 16개의 가장 많은 노드를 포함하고 있는 그룹 1은 전반적인 정보학 분야를 나타내는 토픽들로 이루어져 있었다. 연구평가(토픽 2), 컴퓨팅 분야(토픽 5, 11), 계량서지학(토픽 12, 30, 37), 인용 분석(토픽 33, 38), 학제성(토픽 41), 특허연구(토픽 42), 온톨로지(토픽 47)와 같은 세부주제를 포함하고 있었으며 지리학 분야(토픽 4, 17) 또한 그룹 1로 클러스터링 되었다.

<표 3> 네트워크 클러스터링을 통한 세부 토픽 그룹 분석

클러스터링 그룹	그룹 별 토픽 번호	공통된 단어
1	2, 4, 5, 11, 12, 13, 14, 17, 30, 31, 33, 37, 38, 41, 42, 47	Bibliometric, Citation
2	0, 1, 3, 21, 49	Clinical, Medical, Drug
3	7, 15, 16, 19, 23, 25, 28, 29, 39, 40	Communication, Collaboration
4	20, 27, 34, 36, 43, 44, 45, 46, 48	Organizational, Management
5	6, 8, 9, 10, 18, 22, 24, 26, 32, 35	User, Policy, Service

5개의 가장 적은 노드를 가진 그룹 2는 생물정보학(토픽 0, 21), 의료학(토픽 1, 49)과 관련된 토픽들로 구성되었다. 특히 의료학 분야는 그룹 3에서도 나타나는데 그룹 2는 의료 데이터의 자동 처리를 나타내는 electronic, computerized, record과 같은 단어로 구성되어 있는 반면, 그룹 3은 care, behavior 등의 간호학에 가까운 단어들로 구성된 의료학 분야(토픽 19, 40)를 포함하고 있다. 이와 더불어 비디오 데이터 연구를 나타내는 토픽 3 또한 그룹 2에 클러스터링 되었다.

그룹 3은 단어 communication을 공통적으로 포함하며 strategy, activity, market 등의 단어들이 나타난 경영학 분야(토픽 7, 16, 25), 단어 collaboration으로 연결된 국제적 협동연구(토픽 28), 교육(토픽 29), 소셜미디어와 소셜네트워크(토픽 15, 39) 토픽들이 모여 있었다.

단어 organizational과 management로 강하게 연결되어 있는 그룹 4의 토픽들은 framework, findings, perspective, implementation, theory와 같이 경영학의 연구 방법 중 이론적 부분을 나타내는 단어들로 구성되어 있었다.

그룹 5는 단어 user를 중심으로 한 구조를 보이고 있다. 단어 service로 연결된 정보 검색(토픽 8), 이용자 행위(토픽 9), 소셜 웹(토픽 18), 온라인 쇼핑(토픽 32) 토픽과 단어 policy로 연결된 정보윤리(토픽 6), 정부관련 연구(토픽 26, 35) 토픽으로 이루어져 있었다.

5.1.2 중심성과 유사도 분석

다음으로 전체 토픽 네트워크에서 노드의 크기로 나타낸 연결정도 중심성(Degree Centrality)과 성격이 다른 그룹들 간의 연결을 증대하는 정

도를 나타내는 매개 중심성(Betweenness Centrality)을 통한 중심성 분석 결과는 <표 4>와 같다.

<표 4> 연결정도 중심성과 매개 중심성의 상위/하위 토픽

중심성 상위 토픽		중심성 하위 토픽	
연결정도 중심성	매개 중심성	연결정도 중심성	매개 중심성
토픽 2, 9, 20, 30, 42, 48		토픽 3, 24	
토픽 39, 45, 48	토픽 1, 14, 31, 40	토픽 17	토픽 23, 38

연결정도 중심성과 매개 중심성 각각의 상위 5개 토픽은 순위의 차이는 있으나 동일하게 나타났다. 두 중심성 모두 가장 높게 나타난 토픽 2는 quality, measure, assessment, evaluation와 같이 연구 평가를 나타내는 단어들로 이루어졌으며 그 다음으로는 이용자 행태를 나타내는 토픽 9, 조직 관리 차원의 경영학에 관한 토픽 20, 계량서지학을 나타내는 토픽 30, 특허 연구 분석에 관한 토픽 42의 중심성이 높게 나타났다. 상위 5개 토픽을 제외한 높은 연결정도 중심성의 토픽은 organization, work, community와 같은 단어가 공통적으로 나타나는 토픽 45와 48, 네트워크 분석 연구를 나타내는 토픽 39가 있었다. 반면 생물정보학(bioinformatics) 연구를 나타내는 토픽 31과 health, clinical, intervention과 같은 단어가 공통으로 나타나는 의료학 관련 토픽들(토픽 1, 40)은 매개 중심성이 높게 나타남으로써 최근 5년 동안의 정보학 분야의 세부주제 그룹을 연결해주는 힘이 강한 토픽으로 나타났다.

두 중심성이 모두 낮게 나타난 토픽은 비디

오 데이터의 색인 등의 연구를 나타내는 토픽 3과 open, access, reviewers, accepted 등의 단어가 나타난 토픽 24로 나타났다. 연결정도 중심성이 낮게 나타난 토픽은 지리학 연구를 나타내는 토픽 17이 있었으며 매개 중심성이 낮은 토픽은 mobile, game, play, phone 등의 단어가 나타난 토픽 23과 google, scholar, citation, impact 등의 단어가 나타난 토픽 38로 나타났다. 이러한 중심성 하위 토픽은 정보학 분야의 다른 세부주제와 연결되지 못하고 개별적으로 연구되고 있음을 나타낸다.

클러스터링 분석을 통해 정보학 분야를 구성하는 세부주제 그룹과 각 그룹의 대표 키워드를, 중심성 지수 분석을 통해 정보학 분야에서 다양

한 세부주제와 긴밀하게 연결된 토픽과 개별적으로 연구되는 토픽이 무엇인지를 파악할 수 있었다.

5.2 학술지 별 다양성과 응집성 분석

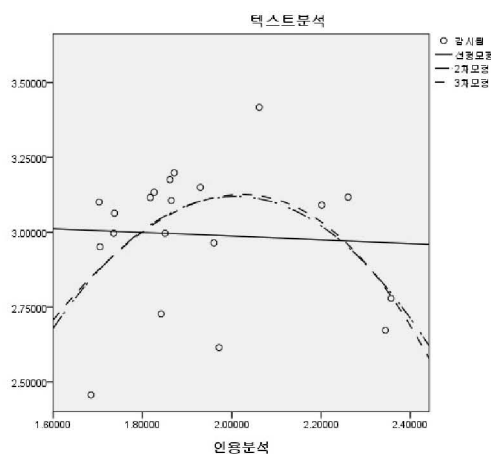
5.2.1 인용 vs 텍스트 기반 다양성 비교

지금까지 다수의 연구에서 이루어졌던 인용 분석을 통해 측정된 다양성과 본 연구에서 실행한 토픽 모델링을 통해 측정된 다양성이 차이를 보이는 지를 Shannon 엔트로피 지수(H)를 통해 비교하였다. 각 방법에서의 Shannon 엔트로피 지수 값이 높은 순서대로 학술지를 정렬한 것이 <표 5>이다.

<표 5> 인용 분석과 토픽 모델링 분석을 통해 산출한 Shannon 엔트로피 지수(H) 기준 학술지 순위

순위	인용분석		토픽모델링 분석	
	학술지명	H	학술지명	H
1	GOV INFORM Q	2.356741	J AM SOC INF SCI TEC	3.417604
2	INT J GEOGR INF SCI	2.34415	MIS QUART	3.19856
3	SCIENTOMETRICS	2.260358	INT J INFORM MANAGE	3.176058
4	J COMPUT MEDIAT COMM	2.201363	INFORM SYST RES	3.150055
5	J AM SOC INF SCI TEC	2.061526	J MANAGE INFORM SYST	3.133531
6	J HEALTH COMMUN	1.971428	SCIENTOMETRICS	3.117619
7	J INFORMETR	1.959882	EUR J INFORM SYST	3.116295
8	INFORM SYST RES	1.929365	J ASSOC INF SYST	3.106124
9	MIS QUART	1.870847	J INF TECHNOL	3.100958
10	J ASSOC INF SYST	1.864503	J COMPUT MEDIAT COMM	3.090718
11	INT J INFORM MANAGE	1.861233	INFORM SYST J	3.063961
12	J AM MED INFORM ASSN	1.85044	INFORM MANAGE AMSTER	2.996874
13	INFORM ORGAN UK	1.841752	J AM MED INFORM ASSN	2.996602
14	J MANAGE INFORM SYST	1.82602	J INFORMETR	2.964752
15	EUR J INFORM SYST	1.817443	J STRATEGIC INF SYST	2.951443
16	INFORM SYST J	1.737105	GOV INFORM Q	2.77915
17	INFORM MANAGE AMSTER	1.735355	INFORM ORGAN UK	2.727329
18	J STRATEGIC INF SYST	1.704714	INT J GEOGR INF SCI	2.672624
19	J INF TECHNOL	1.702674	J HEALTH COMMUN	2.614591
20	INT J COMP SUPP COLL	1.684626	INT J COMP SUPP COLL	2.456502

개별 학술지가 인용한 주제범위의 다양성을 나타내는 인용 분석 순위와 개별 학술지가 다루고 있는 내용의 다양성을 나타내는 토픽 모델링 분석의 순위는 관련 있게 나타나지 않았다. 또한 두 분석 간의 관계를 알아보기 위해 SPSS 21.0으로 실행한 곡선적합 분석의 결과는 <그림 5>와 같다. 인용 분석의 다양성 지수를 독립변수(X)로, 텍스트 분석의 다양성 지수를 종속변수(Y)로 설정하였을 때 선형모형의 효과크기(effect size: r^2)는 0.3%로 큰 의미가 없으며, 2차 모형은 20.7%, 3차 모형은 21.2%로 나타나 3차 모형의 설명력이 가장 높게 나타났다. 따라서 두 분석은 비선형 관계를 보이고 있으며, 학제성에 대한 인용 분석 결과와 텍스트 분석 결과 간의 관계는 일정수준까지는 정비례 관계를 보이거나 그 이상으로는 반비례적 관계를 보이는 것으로 나타났다. 3차 모형 $y = -0.603 + 2.736x_2 - 0.902x_3$ 와 같다.



<그림 5> 인용 분석과 텍스트 분석 간의 곡선적합 결과

<그림 5>에서와 같이 두 분석 간의 영향력은

일정 수준까지 정비례 관계를 보이거나 인용 주제 분야의 다양성이 일정 수준 이상으로 높아지면 텍스트 내용의 다양성은 낮아지는 모습을 보이고 있었다. 두 분석 지수가 반비례 관계를 보이는 학술지 중 대부분은 GOV INFORM Q, INT J GEOGR INF SCI, J HEALTH COMMUN과 같이 전통적인 정보학 분야가 아닌 지리학, 의과학 등의 외부 분야 중심의 성격을 띄고 있는 것으로 나타났다. 하지만 본 연구는 정보학 분야의 학술지만을 대상으로 하였기 때문에 두 분석방법 간 관계의 일반화에 적용하는 것은 한계가 있다.

이와 같이 인용 분석과 텍스트 분석을 통한 학술지의 다양성 순위가 다르게 나타나는 이유는 인용 행태의 관점에서도 설명될 수 있다. 과거 연구에서 인용은 일어난 위치, 통사적 문장 표현, 의미 분석을 통해 감정, 기능 등의 기준으로 분류되며 각 분류마다 다른 영향력을 보이고 있음을 확인하였다(Moravcsik & Murugesan, 1975; Ding, Liu, Guo, & Cronin, 2013). 따라서 동일한 횟수로 인용되었지만 연구 내용에 끼친 영향력은 차이를 보일 수 있으며 해당 연구에 크게 영향을 미치지 않는 형식적(perfunctory) 인용과 같은 문제로 다양성 측정에 왜곡을 불러일으킬 수 있다. 이러한 인용 분석은 협업분야 추천과 같은 적용에 더 적합하며 위와 같은 인용 분석의 왜곡을 완화하는 텍스트 분석은 특정 분야 학술지가 얼마나 다양한 세부분야를 다루고 있는지를 비교분석하는 데에 더 적합하게 이용될 수 있다. 결과적으로 연구 문헌이 다루고 있는 내용의 다양성과 참고문헌의 다양성이 관련이 없을 수 있으며(Porter & Rafols, 2009; Leydesdorff, Rafols, & Chen, 2013) 두 분석

이 학제성의 복합적 측면을 나타내는 상호보완적 방법으로 활용될 수 있다(Bache, Newman, & Smyth, 2013)는 선행연구와도 맥락을 같이 한다.

5.2.2 학술지 별 다양성 분석

다음으로 본 연구에서 적용하고 있는 다양성 개념의 요소인 유형의 수, 균형, 거리를 모두 고려하는 Stirling 다양성 지수(Δ)를 학술지 별로 산출하여 다양성이 높은 순서대로 정렬한 결과는 <표 6>과 같다.

<표 6> 토픽 모델링 분석을 통한 Stirling 다양성 지수(Δ) 기준 학술지 순위

순위	학술지명	Δ
1	J AM SOC INF SCI TEC	0.913318
2	MIS QUART	0.876475
3	INFORM SYST RES	0.871866
4	J MANAGE INFORM SYST	0.871201
5	INT J INFORM MANAGE	0.868753
6	SCIENTOMETRICS	0.86069
7	EUR J INFORM SYST	0.858131
8	J ASSOC INF SYST	0.855189
9	J INF TECHNOL	0.853324
10	J COMPUT MEDIAT COMM	0.849928
11	J AM MED INFORM ASSN	0.848034
12	INFORM SYST J	0.847683
13	INFORM MANAGE AMSTER	0.840995
14	J INFORMETR	0.836536
15	J STRATEGIC INF SYST	0.834587
16	INT J GEOGR INF SCI	0.829484
17	GOV INFORM Q	0.828118
18	INT J COMP SUPP COLL	0.792405
19	J HEALTH COMMUN	0.787417
20	INFORM ORGAN UK	0.783531

Stirling 다양성 지수를 기준으로 한 학술지 순위는 Shannon 엔트로피 지수와 Stirling 다

양성 지수가 상관관계를 보인다는 선행연구(Yegros-Yegros et al., 2010)와 같이 앞서 계산한 토픽 모델링 분석을 통한 Shannon 엔트로피 지수 순위와 큰 차이를 보이지 않았다. 그중 GOV INFORM Q, INFORM MANAGE AMSTER, INFORM ORGAN UK, INFORM SYST J, INT J INFORM MANAGE와 같이 경영학과 관련된 유사 토픽을 다루는 학술지들이 Shannon 엔트로피 지수의 순위보다 Stirling 다양성 지수의 순위가 더 낮게 나타났다. 반면 학술지가 포함하는 토픽 간 유사도가 낮아 Shannon 엔트로피 지수의 순위보다 Stirling 다양성 지수의 순위가 높게 나타난 학술지는 INFORM SYST RES, J MANAGE INFORM SYST, J AM MED INFORM ASSN, INT J COMP SUPP COLL, INT J GEOGR INF SCI로 나타났다. 특히 순위 차이가 2 이상으로 나타난 J AM MED INFORM ASSN, INT J COMP SUPP COLL, INT J GEOGR INF SCI는 전통적인 문헌정보학 분야 혹은 오래 전부터 관련 깊이 함께 연구되어 정보학, 정보공학, 경영정보학과 같은 새로운 분야를 만든 컴퓨터 공학, 경영학 분야가 아닌 각각 의료정보학, 교육학, 지리학처럼 최근 새로이 함께 연구되고 있는 주제를 다루고 있었다. 본 연구의 분석대상이 최근의 영향력이 높은 학술지인 만큼 이와 같이 현재 유사도가 낮게 측정되어 학술지의 다양성을 높이는 외부 분야들은 시간이 지남에 따라 정보학과 함께 통합되어 새로운 학문분야로 발전할 가능성을 지닌 분야라 할 수 있다.

5.2.3 학술지 별 응집성 분석

다음으로 학술지가 포함하는 세부주제의 통

합 정도를 측정하는 응집성 지수를 분석하였다.

두 노드 사이를 연결하는 가능한 모든 경로 중 가장 짧은 경로길이를 고려하므로 값이 낮을수록 높은 통합의 정도를 나타내는 평균 경로길이(L) 기준 학술지 순위는 <표 7>과 같다. 평균 경로길이 기반 학술지의 응집성 순위는 Stirling 다양성 지수 순위와 유사하게 나타나지 않았다. 이는 인용 분석 결과 분야의 다양성과 네트워크 응집성이 상관관계를 보이지 않았던 선행연구(Rafols & Meyer, 2010)의 결론이 텍스트 기반 분석에서도 나타남을 알 수 있다. 하지만 학제성 측정에 있어 다양성과 응집성은 직교적(orthogonal) 관점에서 분석되어야 하는 요소로서 후 절에서 함께 분석하고자 한다.

<표 7> 토픽 모델링 분석을 통한 평균 경로길이(L) 기준 학술지 순위

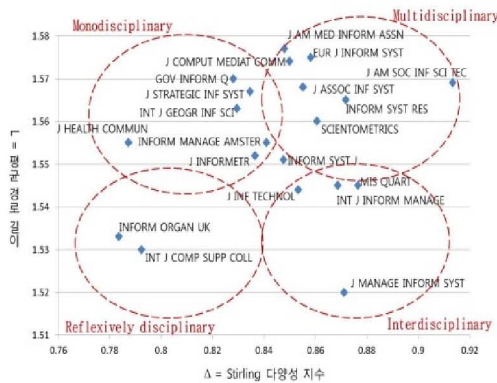
순위	학술지명	L
1	J MANAGE INFORM SYST	1.52
2	INT J COMP SUPP COLL	1.53
3	INFORM ORGAN UK	1.533
4	J INF TECHNOL	1.544
5	INT J INFORM MANAGE	1.545
5	MIS QUART	1.545
7	INFORM SYST J	1.551
8	J INFORMETR	1.552
9	INFORM MANAGE AMSTER	1.555
9	J HEALTH COMMUN	1.555
11	SCIENTOMETRICS	1.56
12	INT J GEOGR INF SCI	1.563
13	INFORM SYST RES	1.565
14	J STRATEGIC INF SYST	1.567
15	J ASSOC INF SYST	1.568
16	J AM SOC INF SCI TEC	1.569
17	GOV INFORM Q	1.57
18	J COMPUT MEDIAT COMM	1.574
19	EUR J INFORM SYST	1.575
20	J AM MED INFORM ASSN	1.577

5.3 개별 학술지의 학제성 유형과 토픽 네트워크

정보학 분야의 5년 영향력 지수 기준 상위 20개 학술지를 대상으로 앞서 측정한 다양성과 통합의 정도를 모두 고려한 학제적 유형을 분류하였다. Stirling 다양성 지수로 측정한 다양성이 x축, 학술지 별 토픽 네트워크의 평균 경로길이를 측정한 응집성이 y축으로 구성된 맵 위에서 좌표로 표현된 정보학 분야 20개 학술지의 위치는 <그림 6>과 같이 나타났다. Rafols et al.(2012)가 개념화한 학제성의 유형을 맵의 4개 구간에 적용해 보면 분석대상인 정보학 분야 학술지 20종 중 8개의 학술지가 다학문성(Multidisciplinary) 구간에 위치하는 것으로 나타났다. 다양성은 높으나 응집성은 낮은 다학문성 구간은 J AM SOC INF SCI TEC를 비롯하여 문헌정보학에 기초한 정보시스템 관련 학술지들로 구성되어 있었다. 또한 교육(INT J COMP SUPP COLL), 커뮤니케이션(J COMPUT MEDIAT COMM), 지리(INT J GEOGR INF SCI), 건강(J HEALTH COMMUN)과 같이 외부학문과 결합된 학술지들은 낮은 다양성을 보이는 단일학문(Monodisciplinary)과 재귀적 학문성(Reflexively disciplinary) 구간에 위치하였다. 다양성과 응집성 모두 높음을 의미하는 간학문성(Interdisciplinary) 구간에 위치한 학술지 4종은 모두 경영학(MANAGEMENT)을 다루는 학술지들로 파악되었다.

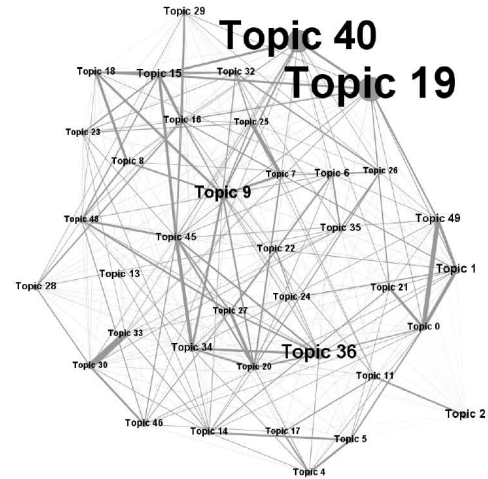
<그림 6>과 같이 분류된 학술지 중 각 유형을 대표하는 4개의 학술지를 선정하여 각각의 다양성과 응집성을 직관적으로 파악할 수 있는 네트워크 시각화를 실행하였다. 각 네트워크에

서 노드의 크기는 해당 학술지에서 다루고 있는 토픽의 비중을 나타내는 토픽 분포 값을 의미하며 에지의 굵기는 정보학 분야에서의 토픽 간 유사도를 나타내고 있다.



〈그림 6〉 정보학 분야 20개 학술지의 학제적 유형 분류

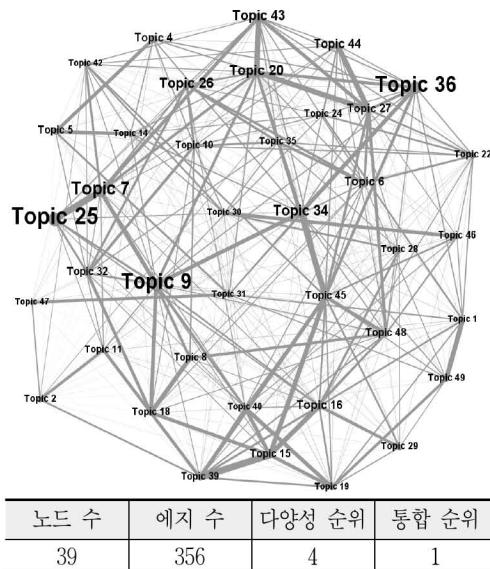
먼저 단일학문(Monodisciplinary)의 유형을 나타내는 학술지 J HEALTH COMMUN의 토픽 네트워크는 〈그림 7〉과 같다. health, communication을 공통 단어로 가지는 토픽 19와 40이 각각 0.27, 0.24의 분포 값을 가지며 본 학술지 내용의 절반을 차지하고 있었다. 이는 본 학술지가 특정 토픽에 급격히 치중되어 낮은 다양성을 보이고 있음을 의미한다. 또한 전체적으로 노드들이 낮은 유사도를 나타내는 연한 선으로 연결되었으며 유사도가 높은 토픽 간의 연결이 제한적으로 이루어지고 있음을 알 수 있다. 이는 각각의 세부주제들이 부분적으로 연결된 모습의 개별적인 연구성향을 보이고 있음을 의미한다.



노드 수	에지 수	다양성 순위	통합 순위
38	313	19	10

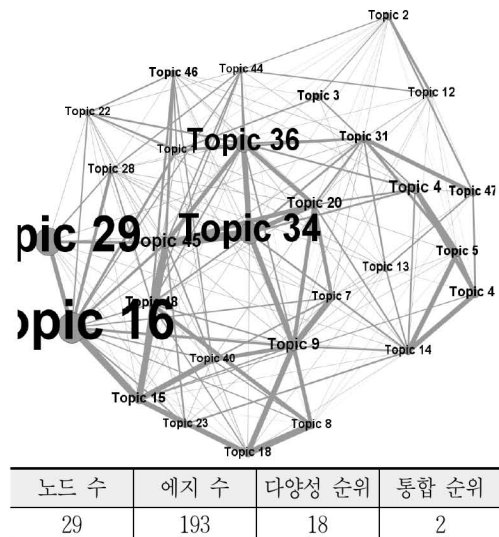
〈그림 7〉 단일학문(Monodisciplinary)으로 분류된 학술지 J HEALTH COMMUN의 토픽 네트워크

반면 다양성과 응집성의 순위가 모두 높게 나타나 간학문성(Interdisciplinary)의 유형을 나타내는 학술지 J MANAGE INFORM SYST의 토픽 네트워크(〈그림 8〉)는 J HEALTH COMMUN과 비슷한 노드 수에 비해 많은 수의 에지를 가지며 확연히 다른 네트워크의 모습을 보이고 있다. 먼저 유사도가 높은 토픽 간의 연결이 많은 에지에서 나타나고 있으며 본 학술지가 다루는 토픽 또한 비교적 균등한 분포를 이루고 있다. 그 중에서도 비교적 높은 분포를 보이는 토픽들을 살펴보면 본 학술지가 price, market의 토픽 25, investment, financial의 토픽 7, outsourcing, offshore의 토픽 26, business, management의 토픽 43처럼 경영학 전반에 걸친 토픽을 다루고 있음을 알 수 있다.



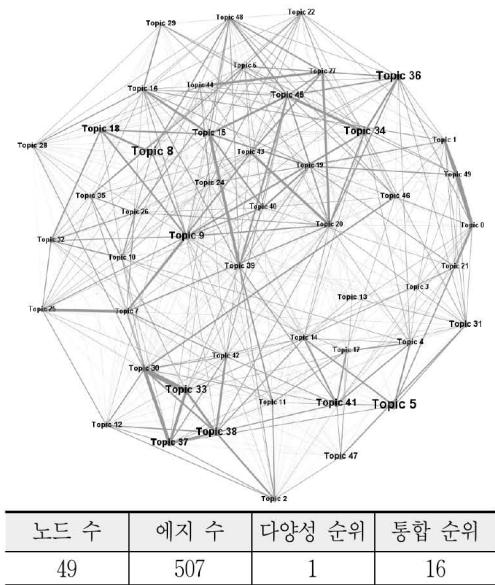
〈그림 8〉 간학문성(Interdisciplinary)으로 분류된 학술지 J MANAGE INFORM SYST의 토픽 네트워크

다음으로 내용의 다양성은 낮지만 높은 응집성을 가지는 재귀적 학문성(Reflexively disciplinary) 유형을 보이는 INT J COMP SUPP COLL의 토픽 네트워크는 〈그림 9〉와 같다. 20개의 학술지 중 가장 적은 노드 수를 가지는 본 학술지는 교육학과 관련된 토픽 16(group, collaborative, learning 등)과 토픽 29(student, education, school 등)의 비중이 매우 크게 나타났다. 하지만 정보학 분야에서 유사도가 높은 토픽들이 연결된 모습을 보이고 있을 뿐만 아니라 직접연결을 많이 가지는 노드들이 많음으로써 높은 평균 경로길이의 값을 보였다. 이는 본 학술지가 다루는 세부주제들이 다양하진 않지만 긴밀하게 연결된 형태로 연구되고 있음을 의미한다.



〈그림 9〉 재귀적 학문성(Reflexively disciplinary)으로 분류된 학술지 INT J COMP SUPP COLL의 토픽 네트워크

반면 모바일 게임 관련 토픽 23을 제외한 49개의 모든 토픽을 노드로 가지며 다양성 측면에서 압도적 1위를 차지하고 있는 학술지 J AM SOC INF SCI TEC은 다학문성(Multidisciplinary) 유형으로 나타났다. 〈그림 10〉에서와 같이 본 학술지의 토픽들은 크게 튀는 비중의 토픽 없이 모든 노드들이 거의 균등한 분포를 나타내고 있지만 응집성의 정도는 낮게 나타났다. 하지만 Rafols와 Meyer(2010)는 높은 다양성과 낮은 응집성을 보이는 구간을 잠재적 간학문성(Potential Interdisciplinary)으로 정의함으로써 앞으로 통합된 형태로의 학제성으로 발전할 잠재력을 가지고 있는 분야로 해석하였다.



〈그림 10〉 다학문성(Multidisciplinary)으로 분류된 학술지 J AM SOC INF SCI TEC의 토픽 네트워크

6. 결론

학제성이 강한 학문분야의 연구자가 연구영역과 이용목적에 맞는 학술지를 선정할 수 있도록 학술지 수준의 학제성을 측정하기 위해 본 연구는 인용 정보를 이용한 과거의 연구들과 달리 문헌이 포함하고 있는 텍스트를 이용한 내용 기반의 학제성 연구를 실행하였다. 본 연구의 분석과정을 요약하면 다음과 같다. 첫째, 토픽 모델링을 통해 추출된 정보학 분야 학술지의 세부토픽을 분석하여 텍스트 기반 분석의 유용함을 확인하였고 둘째, 텍스트 기반의 다양성과 인용 분석을 통한 다양성이 차이를 보이는 지를 확인하고 다양성과 응집성 각각의 차원에서의 학술지 순위를 분석하였으며 셋째,

다양성과 응집성을 모두 고려한 학제적 유형에 따라 학술지를 분류하고 각 유형의 대표 학술지를 선정하여 유형과 특성에 따라 다르게 나타나는 학술지 별 토픽 네트워크를 분석하였다.

학제성을 측정하기 위해 새로운 접근법을 취한 본 연구는 기존 주제분야 분류체계 혹은 몇 개의 제한된 단어로 표현되는 키워드를 문헌에 매핑하는 하향식(Top-down) 접근법이 아닌 문헌에 출현한 그대로의 단어들을 클러스터링하여 세부주제를 도출해내는 상향식(Bottom-up) 접근법을 취하고 있다는 점에서 의의를 찾을 수 있다. 이는 문헌집단이 포함하는 세부주제를 하나로 통합해버리거나 문헌의 내용이 크게 다르지 않음에도 불구하고 게재된 학술지에 따라 다른 주제로 분류되는 하향식 접근법의 취약점을 보완하는 동시에 분석대상 데이터의 규모 확장에 큰 영향을 받지 않는다는 장점이 있다. 또한 추가적인 정보 수집 단계를 요하거나 세밀한 정제 과정이 필요한 인용 정보가 아닌 오픈 액세스와 같은 움직임으로 비교적 쉬운 수집과 정제가 가능해진 문헌의 내용정보를 사용함으로써 많은 양의 데이터를 대상으로 처리와 분석이 효율적으로 이루어질 수 있는 장점이 있다.

본 연구는 정보학 분야의 일부 학술지만을 대상으로 하였기 때문에 분야 전체를 아우르는 결과는 아니지만 본 연구의 접근법을 통해 특정 학문분야 또는 전체 학문분야의 학술지를 대상으로 한 추후연구가 가능함을 확인하였다. 또한 지수 산출을 통한 학제성의 측정과 유형 분류, 시각화를 통한 세부주제의 표현이 학제적 현위치를 파악하고 앞으로의 발전 가능성을 예측하는 단서를 제공한다는 점에서 학술지의

학제성을 나타내는 지표로서 다양하게 활용될 수 있음을 시사하고 있다. 학문의 진보, 흐름, 경향, 예측 등을 가능하게 하는 역할뿐만 아니라 연구자들은 학술지의 구독, 투고 등의 연구 행위를 본인의 목적에 맞게 실행할 수 있고 도

서관 및 정보센터는 학제성을 고려한 장서 관리 및 정보서비스를 제공할 수 있으며 학술지의 발행기관을 포함한 연구기관 또한 학문분야의 학제적 발전을 위한 학술 행사 기획 등에 유용한 지표가 될 수 있을 것으로 기대된다.

참 고 문 헌

- 강범일, 이재운 (2014). 트위터 관련 연구에 대한 계량정보학적 분석. 정보관리학회지, 31(3), 293-311.
<http://dx.doi.org/10.3743/kosim.2014.31.3.293>
- 박소윤 (2013). 문헌정보학 분야의 학제성과 연구 영향력에 관한 연구. 석사학위논문, 이화여자대학교 대학원, 문헌정보학과.
- 박자현, 송민 (2013). 토픽모델링을 활용한 국내 문헌정보학 연구동향 분석. 정보관리학회지, 30(1), 7-32. <http://dx.doi.org/10.3743/kosim.2013.30.1.007>
- Adams, J., Jackson, L., & Marshall, S. (2007). Bibliometric analysis of interdisciplinary research. Report to Higher Education Funding Council for England.
- Bache, K., Newman, D., & Smyth, P. (2013). Text-based measures of document diversity. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 23-31. ACM. <http://dx.doi.org/10.1145/2487575.2487672>
- Bordons, M., Morillo, F., & Gomez, I. (2004). Analysis of cross-disciplinary research through bibliometric tools. In H. F. Moed, W. Glänzel & U. Schmoch (Eds.), Handbook of quantitative science and technology research (pp. 437-456). Dordrecht: Kluwer.
http://dx.doi.org/10.1007/1-4020-2755-9_20
- Brillouin, L. (1956). Science and information theory. New York: Academic Press.
- Carayol, N., & Thi, T. U. N. (2005). Why do academic scientists engage in interdisciplinary research?. Research evaluation, 14(1), 70-79.
<http://dx.doi.org/10.3152/147154405781776355>
- Chua, A. Y., & Yang, C. C. (2008). The shift towards multi disciplinaryity in information science. Journal of the American Society for Information Science and Technology, 59(13), 2156-2170.
<http://dx.doi.org/10.1002/asi.20929>
- Cronin, B., & Meho, L. I. (2008). The shifting balance of intellectual trade in information studies.

- Journal of the American Society for Information Science and Technology, 59(4), 551-564.
<http://dx.doi.org/10.1002/asi.20764>
- Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics*, 7(3), 583-592.
<http://dx.doi.org/10.1016/j.joi.2013.03.003>
- Herring, S. D. (1999). The value of interdisciplinarity: A study based on the design of Internet search engines. *Journal of the American Society for Information Science*, 50(4), 358-365.
[http://dx.doi.org/10.1002/\(sici\)1097-4571\(1999\)50:4<358::aid-asi14>3.0.co;2-7](http://dx.doi.org/10.1002/(sici)1097-4571(1999)50:4<358::aid-asi14>3.0.co;2-7)
- Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology*, 61(1), 126-131. <http://dx.doi.org/10.1002/asi.21226>
- Levitt, J. M., & Thelwall, M. (2009). The most highly cited Library and Information Science articles: Interdisciplinarity, first authors and citation patterns. *Scientometrics*, 78(1), 45-67.
<http://dx.doi.org/10.1007/s11192-007-1927-1>
- Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, 5(1), 87-100.
<http://dx.doi.org/10.1016/j.joi.2010.09.002>
- Leydesdorff, L., Rafols, I., & Chen, C. (2013). Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations. *Journal of the American Society for Information Science and Technology*, 64(12), 2573-2586.
<http://dx.doi.org/10.1002/asi.22946>
- Mann, G. S., Mimno, D., & McCallum, A. (2006). Bibliometric impact measures leveraging topic analysis. *Proceedings of the 6th ACM/IEEE-CS Joint Conference*, 65-74. IEEE.
<http://dx.doi.org/10.1145/1141753.1141765>
- Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, 5(1), 86-92. <http://dx.doi.org/10.1177/030631277500500106>
- Morillo, F., Bordons, M., & Gómez, I. (2001). An approach to interdisciplinarity through bibliometric indicators. *Scientometrics*, 51(1), 203-222.
- Morillo, F., Bordons, M., & Gómez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13), 1237-1249. <http://dx.doi.org/10.1002/asi.10326>
- National Academies (2004). *Facilitating interdisciplinary research*. Committee on Science, Engineering, and Public Policy (Cosepup) & Committee on Facilitating Interdisciplinary Research, Washington,

- D.C.: The National Academies Press.
- Nichols, L. G. (2014). A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics*, 100(3), 1-14.
<http://dx.doi.org/10.1007/s11192-014-1319-2>
- Organisation for Economic Cooperation and Development (OECD). (1998). *Interdisciplinarity in science and technology*. Directorate for science, technology and industry. Paris: OECD.
- Porter, A. L., & Chubin, D. E. (1985). An indicator of cross-disciplinary research. *Scientometrics*, 8(3), 161-176. <http://dx.doi.org/10.1007/bf02016934>
- Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3), 719-745.
<http://dx.doi.org/10.1007/s11192-008-2197-2>
- Prebor, G. (2010). Analysis of the interdisciplinary nature of library and information science. *Journal of Librarianship and Information Science*, 42(4), 256-267.
<http://dx.doi.org/10.1177/0961000610380820>
- Qin, J., Lancaster, F. W., & Allen, B. (1997). Types and levels of collaboration in interdisciplinary research in the sciences. *Journal of the American Society for information Science*, 48(10), 893-916.
[http://dx.doi.org/10.1002/\(sici\)1097-4571\(199710\)48:10<893::aid-asi5>3.0.co;2-x](http://dx.doi.org/10.1002/(sici)1097-4571(199710)48:10<893::aid-asi5>3.0.co;2-x)
- Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy*, 41(7), 1262-1282.
<http://dx.doi.org/10.1016/j.respol.2012.03.015>
- Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2), 263-287.
<http://dx.doi.org/10.1007/s11192-009-0041-y>
- Rinia, E. J., van Leeuwen, T. N., & van Raan, A. F. (2002). Impact measures of interdisciplinary research in physics. *Scientometrics*, 53(2), 241-248.
- Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59(3), 425-465.
<http://dx.doi.org/10.1023/b:scie.0000018542.71314.38>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379-423 and 623-656.
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163, 688.

- Stirling, A. (1998). On the economics and analysis of diversity. SPRU Electronic Working Papers. Retrieved from <https://www.sussex.ac.uk/webteam/gateway/file.php.pdf>
- Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15), 707-719. <http://dx.doi.org/10.1098/rsif.2007.0213>
- Tang, R. (2004). Evolution of the interdisciplinary characteristics of information and library science. *Proceedings of the American Society for Information Science and Technology*, 41(1), 54-63. <http://dx.doi.org/10.1002/meet.1450410107>
- Yegros-Yegros, A., Amat, C. B., D'Este, P., Porter, A. L., & Rafols, I. (2010). Does interdisciplinary research lead to higher scientific impact. In STI Indicators Conference, Leiden.

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

- Kang, Beomil, & Lee, Jae Yun (2014). A bibliometric analysis on twitter research. *Journal of the Korean Society for Information Management*, 31(3), 293-311. <http://dx.doi.org/10.3743/kosim.2014.31.3.293>
- Park, So Yoon (2013). A study on interdisciplinarity and research impact in the field of library and information science. Master's Thesis. The Graduate School of Ewha Womans University. Department of Library and Information Science.
- Park, Ja-Hyun, & 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. <http://dx.doi.org/10.3743/kosim.2013.30.1.007>

[부록] 토픽 모델링 결과 추출된 50개의 토픽과 각 토픽의 구성단어 20개

Topic 0		Topic 1		Topic 2		Topic 3	
drug	cases	medication	control	quality	validity	creativity	relations
clinical	algorithm	errors	computerized	measure	measuring	database	representation
patients	codes	patients	rates	assessment	instrument	intersection	segmentation
electronic	identified	intervention	cpoe	measures	developed	lines	segments
disease	events	clinical	drug	evaluation	items	movement	shape
surveillance	objective	trial	decision	criteria	survey	object	space-time
records	patient	alert	patient	reliability	evaluate	objects	spatial
medical	record	order	support	scale	subjective	point	trajectories
adverse	predictive	prescribing	objective	evaluating	performance	points	types
identify	methods	group	time	measurement	objective	relation	video
Topic 4		Topic 5		Topic 6		Topic 7	
problem	multiple	method	features	security	organizations	firms	effect
simulation	process	methods	propose	privacy	vaccination	firm	impact
method	rules	proposed	large	compliance	human	value	competitive
modeling	application	algorithm	detection	policies	ethical	firm's	strategy
proposed	complex	similarity	experiments	disclosure	control	performance	positive
geographic	optimization	approaches	accuracy	protection	moral	market	costs
problems	spatial	classification	set	policy	management	investments	effects
applied	representation	algorithms	better	concerns	risks	investment	relationship
set	map	compared	existing	risk	practices	strategic	business
methodology	case	performance	number	personal	users	financial	innovation
Topic 8		Topic 9		Topic 10		Topic 11	
search	searching	factors	significant	computing	distributed	bias	modeling
retrieval	engine	influence	effect	service	integration	sample	number
users	relevant	perceived	implications	services	infrastructure	measurement	structural
task	participants	users	acceptance	cloud	software	accuracy	sensitivity
image	cognitive	effects	support	spatial	grid	estimates	empirical
queries	experiment	findings	satisfaction	parallel	resources	variables	variable
query	visual	user	social	web	implementation	construct	random
user	searches	behavior	important	architecture	large	size	error
tasks	behavior	usage	relationship	applications	query	effect	validity
relevance	web	theory	performance	access	article	test	predictive
Topic 12		Topic 13		Topic 14		Topic 15	
universities	individual	question	reference	theory	informatics	social	sites
university	level	questions	reading	history	field	media	blogs
academic	ranking	answer	titles	researchers	focus	online	content
researchers	institutional	editorial	types	review	action	community	facebook
institutions	higher	visual	clear	discipline	european	communities	members
productivity	output	board	plagiarism	methods	recommendations	communication	participation
scientists	production	materials	labels	journal	method	political	blog
performance	evaluation	answering	functional	issues	approaches	internet	news
rankings	national	language	content	theories	questions	networking	network
indicators	bibliometric	texts	units	argument	online	twitter	support
Topic 16		Topic 17		Topic 18		Topic 19	
group	activity	spatial	species	web	online	health	diabetes
groups	interactions	land	geographic	websites	users	literacy	survey
collaborative	communication	urban	local	wikipedia	social	cancer	adults
collaboration	members	area	map	site	internet	patients	age
support	online	gis	variables	sites	user	care	education
interaction	synchronous	error	environmental	pages	search	screening	numeracy
learning	discussion	density	terrain	tagging	tag	communication	women
participants	participation	scale	maps	content	resources	low	associated
memory	individual	areas	distribution	website	quality	internet	authors
shared	effects	cover	change	two	usage	participants	national

Topic 20		Topic 21		Topic 22		Topic 23	
adoption	capability	disease	translational	sharing	access	mobile	univ
organizational	framework	genetic	molecular	barriers	source	game	comm
management	findings	cancer	family	capacity	explicit	devices	sci
factors	implementation	diseases	association	shared	management	phone	marketing
business	benefits	informatics	methods	content	digital	mhealth	singapore
innovation	enterprise	clinical	associated	personal	platform	soc	play
process	diffusion	biological	images	tacit	facilities	games	hoboken
organizations	theory	gene	materials	control	providers	communication	journal
capabilities	impact	genes	objective	requirements	time	players	online
performance	case	discovery	identified	file	integration	cell	playing
Topic 24		Topic 25		Topic 26		Topic 27	
funding	reviewers	product	strategies	outsourcing	service	project	leadership
peer	making	price	prices	risk	mechanisms	management	culture
reviews	ratings	market	auction	transfer	cost	projects	organizational
credibility	comments	products	presence	risks	decision	implementation	isd
open	accepted	markets	services	client	bpo	control	change
access	decisions	pricing	firms	governance	business	team	development
manuscripts	international	consumers	effect	contract	services	success	enterprise
decision	acceptance	auctions	impact	contracts	offshoring	managers	organizations
aid	source	online	strategy	vendor	benefits	resistance	teams
sources	article	uncertainty	effects	offshore	performance	climate	case
Topic 28		Topic 29		Topic 30		Topic 31	
countries	united	students	collaborative	publications	index	text	reports
china	states	learning	classroom	articles	countries	clinical	documents
international	european	education	program	journals	citation	extraction	recall
collaboration	regions	training	activities	publication	bibliometric	processing	precision
chinese	world	student	schools	period	literature	language	corpus
country	regional	educational	digital	years	authors	natural	f-measure
south	collaborative	school	outcomes	number	production	machine	learning
global	growth	teachers	environment	sciences	field	challenge	features
national	development	course	practices	growth	web	classification	authors
africa	impact	college	online	output	time	medical	performance
Topic 32		Topic 33		Topic 34		Topic 35	
trust	satisfaction	articles	author	theory	case	public	local
online	loyalty	citation	highly	theoretical	particular	government	divide
service	services	journals	impact	understanding	social	e-government	access
customer	e-commerce	citations	publication	perspective	empirical	policy	state
consumer	relationship	cited	article	context	work	services	transparency
quality	purchase	number	disciplines	practice	literature	agencies	federal
customers	value	publications	citing	view	findings	citizens	national
perceived	shopping	references	sciences	process	implications	digital	electronic
intention	mobile	journal	literature	nature	important	governments	policies
website	internet	authors	years	aspects	development	sector	economic
Topic 36		Topic 37		Topic 38		Topic 39	
design	decision	h-index	g-index	citation	counts	network	diffusion
framework	future	index	measures	impact	normalization	networks	properties
support	issues	indices	distributions	journal	subject	structure	measures
provide	work	number	values	journals	differences	centrality	structures
development	needs	distribution	proposed	factor	google	collaboration	patterns
challenges	need	citations	publication	indicator	scholar	social	relationships
process	identify	function	bibliometric	indicators	measures	ties	size
potential	management	law	publications	citations	fields	nodes	communities
tools	important	citation	output	ranking	evaluation	structural	communication
key	case	hirsch	impact	factors	field	links	authors

Topic 40		Topic 41		Topic 42		Topic 43	
messages	behaviors	library	visualization	patent	companies	business	competitive
health	communication	clusters	core	patents	field	supply	capabilities
media	prevention	topics	bibliometric	technological	economic	chain	agility
message	women	clustering	cluster	innovation	relations	alignment	organizations
campaign	intervention	field	interdisciplinary	r&d	relationship	strategic	intelligence
smoking	news	co-citation	method	nanotechnology	activities	governance	chains
children	content	maps	map	patenting	citations	integration	organizational
authors	associated	subject	network	industry	value	strategy	flexibility
risk	effects	structure	fields	indicators	performance	management	companies
exposure	behavior	mapping	methods	development	fields	value	case
Topic 44		Topic 45		Topic 46		Topic 47	
software	open-source	community	social	time	dynamics	semantic	documents
development	release	digital	practices	temporal	event	terms	indexing
open	proprietary	development	understanding	events	times	concepts	collections
projects	vendors	innovation	role	patterns	phase	ontology	retrieval
source	packaged	ict	context	emergency	monitoring	document	snomed
oss	teams	organizations	communities	change	growth	domain	text
project	product	activities	work	changes	behavior	term	categories
agile	application	perspective	need	dynamic	types	metadata	mapping
developers	agility	change	organizational	period	process	ontologies	relationships
coordination	participation	emerging	diffusion	series	years	language	evaluation
Topic 48		Topic 49					
virtual	working	adoption	medical				
worlds	trust	care	objective				
work	environments	clinical	patient				
communication	patterns	ehr	patients				
teams	tasks	electronic	physicians				
team	engineers	health	practice				
world	distributed	healthcare	practices				
e-mail	time	hit	providers				
members	organizational	hospital	record				
email	users	hospitals	records				