

동시출현단어 분석에 기반한 메타데이터 분야의 지적구조에 관한 연구

A Study on the Intellectual Structure of Metadata Research by Using Co-word Analysis

최예진 (Ye-Jin Choi)*

정연경 (Yeon-Kyoung Chung)**

초 록

다양한 매체와 유형으로 생산되는 정보자원에 대한 이용이 높아짐에 따라, 정보자원을 기술하기 위한 정보조직의 도구로서 메타데이터에 대한 중요성이 높아지고 있다. 본 연구에서는 메타데이터 분야의 연구 영역을 파악할 수 있도록 동시출현단어 분석을 사용하여 메타데이터 분야의 지적 구조를 규명하고자 하였다. 이를 위하여 1998년 1월 1일부터 2016년 7월 8일까지 Web of Science 핵심컬렉션에 등재된 저널에 게재된 문헌을 대상으로 'metadata'라는 질의어로 Topic 검색을 수행하여, 총 727건의 논문에 대한 서지정보를 수집하였다. 이 중 저자 키워드를 가진 410건의 논문의 저자 키워드로 수집하고, 전처리 과정을 거쳐 저자 키워드 총 1,137개를 추출하여 최종적으로 빈도수 6회 이상의 키워드 37개를 분석대상으로 선정하였다. 이후 메타데이터 분야의 지적구조 규명을 위해 첫째, 네트워크 분석을 통하여 2개 영역 9개 군집을 도출하였으며, 메타데이터 분야 키워드들의 지적 관계를 시각화하고, 중심성 분석을 통한 전역 중심 키워드와 지역 중심이 높은 키워드를 제시하였다. 둘째, 군집분석을 실시하여 형성된 6개의 군집을 다차원축적지도상에 표시하였으며, 각 키워드들 간의 상관관계에 따른 지적구조를 제시하였다. 이러한 연구의 결과는 메타데이터 분야의 지적구조를 시각적으로 파악할 수 있게 하며, 향후 메타데이터 관련 교육과 연구의 방향성 모색에 유용하게 사용될 수 있을 것이다.

ABSTRACT

As the usage of information resources produced in various media and forms has been increased, the importance of metadata as a tool of information organization to describe the information resources becomes increasingly crucial. The purposes of this study are to analyze and to demonstrate the intellectual structure in the field of metadata through co-word analysis. The data set was collected from the journals which were registered in the Core collection of Web of Science citation database during the period from January 1, 1998 to July 8, 2016. Among them, the bibliographic data from 727 journals was collected using Topic category search with the query word 'metadata'. From 727 journal articles, 410 journals with author keywords were selected and after data preprocessing, 1,137 author keywords were extracted. Finally, a total of 37 final keywords which had more than 6 frequency were selected for analysis. In order to demonstrate the intellectual structure of metadata field, network analysis was conducted. As a result, 2 domains and 9 clusters were derived, and intellectual relations among keywords from metadata field were visualized, and proposed keywords with high global centrality and local centrality. Six clusters from cluster analysis were shown in the map of multidimensional scaling, and the knowledge structure was proposed based on the correlations among each keywords. The results of this study are expected to help to understand the intellectual structure of metadata field through visualization and to guide directions in new approaches of metadata related studies.

키워드: 동시출현단어분석, 메타데이터, 지적구조, 네트워크 분석

co-word analysis, metadata, knowledge structure, network analysis

* 이화여자대학교 일반대학원 문헌정보학과 박사과정(dmiration@gmail.com) (제1저자)

** 이화여자대학교 사회과학대학 문헌정보학 전공 교수(ykchung@ewha.ac.kr) (교신저자)

■ 논문접수일자: 2016년 8월 2일 ■ 최초심사일자: 2016년 9월 1일 ■ 게재확정일자: 2016년 9월 8일

■ 정보관리학회지, 33(3), 63-83, 2016. [http://dx.doi.org/10.3743/KOSIM.2016.33.3.063]

1. 서론

정보통신기술의 발전으로 정보가 다양한 매체와 유형으로 대량 생산되고, 그에 대한 이용이 급격하게 증가함에 따라, 이용자가 원하는 정보에 효율적으로 접근하고 탐색할 수 있도록 정보자원을 기술하고 조직하는 '정보조직' 분야가 중요해지고 있다. '정보조직'은 문헌정보학의 중요한 분야로 꾸준히 연구되고 있는 실정이며(박옥남, 2011), 최근에는 다양한 매체와 유형으로 생산되는 정보자원을 기술하기 위한 정보조직 도구로 활용되고 있는 '메타데이터(metadata)'에 대한 연구가 활발하게 진행되고 있다.

현재까지 누적된 연구결과에 대한 분석은 해당 분야의 발전 양상과 연구동향뿐만 아니라 학문적 특성을 파악하는데 중요한 수단이 될 수 있다(오세훈, 2005). 때문에 많은 연구자들이 다양한 방법으로 특정 학문 또는 주제 분야에 대한 구조와 특성을 분석하고, 그 결과를 통하여 해당 분야의 핵심 영역을 제시하고, 향후의 발전을 예측하는 연구를 수행해왔다. 그러나 지금까지 '메타데이터' 분야에 대한 지적구조나 연구동향을 분석한 연구는 내용분석을 통해 실무적 관점에서 중요하다고 생각되는 4가지 기준에 따라 메타데이터 관련 국내 연구동향을 분석한 연구(유사라, 2010)와 '정보조직' 분야의 지적구조를 분석하는 과정에서 하나의 연구 주제 분야로 '메타데이터'의 중요성을 언급하는 연구(박옥남, 2011)만이 수행되었을 뿐이다.

이에 본 연구에서는 계량정보학적 기법을 적용하여 '정보조직'의 세부 주제 분야인 '메타데

이터'의 지적구조를 살펴보고자 한다. 일반적으로 특정 주제 분야에 대한 지적구조를 분석하기 위해서는 동시인용분석, 서지결합법, 동시출현단어 분석과 같은 계량정보학적 기법을 활용되는데, 메타데이터 분야는 비교적 최근에 연구된 주제 분야이기 때문에, 인용의 지체 현상으로 인하여 최신의 동향을 파악하는데 한계가 있는 인용분석보다는 함께 빈번하게 등장한 용어가 개념적으로 유사하다는 가정 하에 해당 주제 분야의 문헌 집합에서 텍스트 본문이나 제목, 초록, 키워드 등을 활용하여 주제 영역의 하위 영역 또는 주제 분야에 대한 패턴을 규명하는데 사용되는 동시출현단어분석을 적용하는 것이 적절하다.

이를 위하여 본 연구는 '메타데이터'라는 주제 분야에서 동시에 빈번하게 출현한 용어를 분석하고, 추출된 키워드를 기반으로 네트워크 분석과 중심성 분석을 수행하여 '메타데이터' 분야에서 중심이 되는 연구 분야를 파악하고, 그 구조를 시각적으로 제시하여 메타데이터 분야의 지적 구조를 규명하고자 한다. 이러한 연구결과는 정보조직 분야의 중요한 주제 영역인 '메타데이터' 분야에 대한 학문적 구조와 하위 연구 분야에 대한 정보를 제공할 수 있으며, 이를 바탕으로 향후 '메타데이터' 분야의 발전 방향을 제시하는데 효과적으로 사용될 수 있을 것이다.

2. 관련 연구

본 연구와 관련된 선행연구는 크게 계량정보학적 기법 중 하나인 동시출현단어 분석을 적용

하여 해당 학문 분야 또는 주제 영역의 지적구조를 밝히는 연구와 본 연구에서 분석대상으로 설정한 주제 분야인 '메타데이터' 분야에 대한 연구동향이나 지적구조를 파악하는 연구로 나누어 볼 수 있다.

동시출현단어 분석을 적용하여 학문 분야 및 주제 영역의 지적구조를 밝히는 연구는 연구 분야에 따라 문헌정보학 분야를 대상으로 하는 연구와 문헌정보학 이외의 다른 학문 분야 또는 특정 주제 영역을 대상으로 하는 연구로 구분할 수 있다. 먼저 동시출현단어 분석을 활용하여 문헌정보학 분야의 지적구조를 밝히는 연구는 크게 문헌정보학 분야 전반에 대한 연구, 문헌정보학과 밀접하게 관련된 주제 분야에 대한 연구, 문헌정보학 분야의 세부 주제 분야에 대한 연구 등으로 구분할 수 있다.

조제인(2011), Milojević, Sugimoto, Yan, Ding(2011), Hu, Hu, Deng, Liu(2013)는 동시출현단어 분석을 활용하여 문헌정보학 분야 전반에 지적구조를 밝히는 연구를 수행하였는데, 이들은 주로 문헌정보학 분야의 학술지에 게재된 논문을 대상으로 동시출현단어 분석과 함께 다양한 계량정보학적 분석기법을 활용하여 문헌정보학 분야의 하위 주제와 그 관계를 식별하고, 시기별 연구동향을 파악하여 문헌정보학의 지적구조를 제시하였다. 또한, 웹 아카이빙(김희정, 2011), 오픈엑세스 분야(서선경, 정은경, 2013), 의료정보학 분야(허고은, 송민, 2013) 등과 같이 문헌정보학과 밀접하게 관련된 주제 분야에 대해서도 동시출현단어 분석을 적용하여 해당 분야의 지적구조를 파악하는 연구가 수행되었다. 더불어 문헌정보학 분야의 다양한 세부 주제 분야에 동시출현단어 분석을 적

용하여 해당 세부 분야의 지적구조를 파악하는 연구도 수행되었으며, 관련 선행 연구를 요약하면 <표 1>과 같다.

문헌정보학이 아닌 다른 학문 및 특정 주제 분야를 대상으로 동시출현단어 분석을 적용한 연구는 인공지능(이미경, 정영미, 2003), 차세대 디스플레이(OLED)(이우형, 김윤명, 박각로, 이명호, 2006), 정보기술 아키텍처(Enterprise Architecture, EA)(권진영, 이우형, 김미현, 이석준, 2010), 환경과 관련된 주제 분야(박재신, 정영미, 2010), 다문화와 관련된 주제 분야(장임숙, 장덕현, 이수상, 2011), 국가 R&D 연구 분야(서원철, 박현석, 윤장혁, 2012), 한국어교육학(강범일, 박지홍, 2013), 빅데이터와 관련된 연구 분야(김완중, 2014) 등 다양한 주제 영역과 학문 분야에서 이루어졌으며, 대부분의 연구에서 동시출현단어 분석과 더불어 군집분석, 다차원척도법, 네트워크 분석을 함께 수행하고, 다양한 프로그램을 활용하여 분석의 결과를 시각화하여 해당 주제 영역의 구조와 하위 분야를 규명하고, 연구의 경향과 시기별 변화를 파악한 후, 해당 주제 영역의 발전 방향을 제시하였다.

한편, 계량정보학적 기법을 사용하여 '메타데이터' 분야의 연구동향이나 지적구조를 파악하는 연구는 내용분석을 통해 '메타데이터' 분야의 연구동향을 파악하거나, 정보조직 분야의 지식구조를 파악하는 과정에서 '메타데이터'에 대한 주제 영역을 확인하는 연구가 수행되었다.

유사라(2010)는 국내 메타데이터 연구동향을 분석하기 위하여 1999년부터 2009년까지 출판된 국내 학술지 27종에서 주제어가 '메타데이터'인 학술논문 및 학위논문 120건을 대상

〈표 1〉 문헌정보학 분야의 세부 주제분야에 동시출현단어 분석을 활용한 주요 선행연구

연구자 (연도)	분석대상, 분석기간, 데이터 출처
	연구 결과 요약
Ding, Chowdhury, & Foo (2001)	- 정보검색(Information Retrieval) 분야 관련 논문(2,012건) / 1987년 ~ 1997년 - SCI와 SSCI 데이터베이스
	- LISA 시소러스를 활용하여 표준화 작업 수행 - 논문의 제목과 초록에서 최종적으로 240개 키워드 추출 - 동시출현단어 분석, 군집분석 수행 - IR 분야의 전체 및 각 군집별 영역을 시각화 - IR 분야의 시기별 지적구조의 변화를 확인
Liu, Hu, & Wang (2012)	- 중국 디지털도서관(Digital Library) 분야 관련 논문 (2,647건) / 2002년 ~ 2011년 - 중국 학술저널데이터베이스(Chinese journal full-text database, CJFTD)
	- 논문의 제목에서 9,538개의 키워드 추출 - 동시출현단어 분석, 군집 분석, 다차원 축적지도, 네트워크 분석 수행(SPSS 19.0, Ucinet 6.0 활용) - 디지털 도서관 분야의 핵심 연구 영역 식별 - 디지털 도서관에 대한 중국의 연구 경향을 파악
이정규, 이용구 (2013)	- 문헌정보학 분야 학술지 5종에 게재된 도서관경영 관련 논문(413건) / 2001년 ~ 2013년
	- 논문의 제목, 초록, 저자키워드에서 최종적으로 103개 키워드 추출 - 동시출현단어 분석, 군집 분석(SPSS 20.0 활용) - 도서관 경영과 관련된 핵심 연구 영역 식별 - 기존에 내용분석으로 연구된 분석결과와 유사점 확인
김하진, 송민 (2014)	- 정보학 분야 학술지(정보관리학회지와 JASIST)에 게재된 논문(총 3,438건) / 1990년 ~ 2013년 - WoS, Wiley online library, 한국연구재단, NDSL
	- 논문의 제목, 초록에 텍스트 마이닝 기법을 적용하여 명사와 명사구를 추출 - 동시출현단어 분석, 네트워크 분석 수행(Gephi 0.8.2 활용) - 전 기간을 5개 구간으로 구분하여 시기별로 분석 - 학술지별 세부 주제 변화를 파악 - 정보학 분야의 국내와 국외의 연구 경향을 파악
장령령, 홍현진 (2014)	- 독서 분야 관련 논문 (838건) / 2003년 ~ 2012년 - 한국학술지인용색인(KCI)
	- 논문에서 최종적으로 56개 키워드 추출(고빈도 키워드 선정 공식 활용) - 일반가중치와 혼합가중치를 적용하여 행렬을 생성 - 동시출현단어 분석, 군집 분석, 다차원 축적지도, 네트워크 분석 수행(SPSS, NodeXL 활용) - 혼합가중치 적용의 장점을 확인 - 독서분야의 시기별 지적구조의 변화를 확인

으로 실무적용 관점에서 중요하다고 생각되는 4가지 기준에 따라 그 내용을 분석하고, 메타데이터에 대한 세부 주제 분석, 연도별 분석, 개발(목적) 유형별 분석 등과 같은 연구동향을 분석하였다. 분석 결과, 메타데이터의 개념과 설계에 대한 연구가 중점적으로 수행되었지만, 메타데

이터의 서비스에 관련된 연구는 미비한 것으로 파악되었기 때문에, 향후 메타데이터의 상호운용성, 품질 등 실무적 적용이 가능한 연구가 수행되어야 할 필요성이 있다고 언급하였다.

박옥남(2011)은 국내 정보조직분야의 지식구조를 파악하기 위하여 한국학술지인용색인

(KCI)에 등재된 5종의 학술지에 2000년부터 2011년까지 게재된 정보조직 분야 논문 417편을 대상으로 저자키워드, 초록, 저자, 저자소속 기관을 추출하고, 핵심 주제어 43개를 대상으로 내용분석과 네트워크 분석을 수행하였다. 분석 결과, 정보조직 분야의 주제어들은 좁고 복잡한 관계를 가지며, 주제어 간의 직접적인 연결이 많은 것으로 나타났으며, 정보조직의 중심축은 여전히 분류와 목록이 담당하고 있지만 온톨로지와 메타데이터가 새로운 연구 분야로 부상하였다는 점을 지적하면서, 정보조직 분야의 지적구조상에서 메타데이터의 위상과 중요성을 언급하였다.

이상의 관련 연구를 살펴보았을 때, 동시출현단어 분석을 적용하여 문헌정보학을 비롯한 다양한 학문 분야와 다양한 주제 분야의 연구동향이나 지적구조를 밝히기 위한 연구는 많이 수행되었으나, '메타데이터' 분야의 지적구조나 연구동향을 동시출현단어로 분석한 연구는 드문 것으로 확인되었다. 이에 본 연구에서는 동시출현단어 분석을 적용하여 정보조직 분야의 세부주제 영역인 '메타데이터'의 연구 주제와 지적구조를 파악하고자 한다.

3. 연구 설계

3.1 데이터 수집과 키워드 선정

본 연구에서는 정보조직 분야의 세부 주제 영역인 '메타데이터' 분야에 대한 지적구조를 파악하기 위하여, 인용색인데이터베이스인 Web of Science(이하 WoS) 핵심컬렉션에 등재된

저널에 게재된 문헌을 대상으로 데이터를 수집하였다.

데이터 검색의 범위는 1998년 1월 1일부터 데이터 수집 시점인 2016년 7월 8일까지로 제한하였으며, 주제(Topic) 검색의 질의어로는 'metadata'를 이용하였다. 더불어, 분석의 대상을 문헌정보학 분야에서 연구된 메타데이터 관련 논문으로 한정하기 위하여 WoS 핵심 컬렉션의 범주를 '문헌정보학(INFORMATION SCIENCE LIBRARY SCIENCE)'으로 지정하고, 문서 유형과 언어를 각각 논문(ARTICLE)과 영어(ENGLISH)로 제한하여 최종적으로 총 727건의 논문을 수집하였다.

이후, 동시출현단어분석을 수행하기 위하여 Olle Persson이 개발한 Bibexcle 프로그램을 활용하여 727건의 논문 중 저자 키워드(DE: Author keywords)를 가지고 있는 410건에서 총 2,023개의 저자 키워드를 추출하고, 분석의 해석과 용이성을 위하여 다음과 같은 전처리 작업을 수행하였다.

첫째, 동일한 용어가 대소문자로 표현되어 다른 용어로 식별되는 것을 방지하기 위하여 각 필드에 포함된 주제어를 모두 소문자로 변환하고, 고유명사와 약어로 된 키워드는 다시 대문자로 변환하였다. 이 과정에서 하나의 논문에 원괄호 안에 동일한 용어를 약어로 표시하여 저자키워드를 제시한 경우, 각각의 키워드를 서로 다른 키워드로 해석하였다. 즉, 저자가 Dewey Decimal Classification(DDC)라는 키워드를 제시한 경우, Dewey Decimal Classification과 DDC라는 2가지 키워드를 제시한 것으로 해석하여 처리하였다.

둘째, 단·복수의 키워드, 동일한 의미를 가지

〈표 2〉 출현빈도 6회 이상의 최종 37개 키워드 리스트

번호	키워드	빈도수	번호	키워드	빈도수
1	metadata	79	20	linked data	11
2	digital libraries	68	21	theses	11
3	information retrieval	44	22	indexing	10
4	libraries	29	23	user studies	9
5	digital storage	21	24	archives	8
6	cataloguing	19	25	institutional repositories	8
7	semantics	18	26	dublin core	8
8	classification	17	27	archives management	8
9	semantic web	16	28	collections management	7
10	internet	15	29	digital archives	7
11	databases	15	30	content management	6
12	search engines	14	31	iran	6
13	information management	13	32	data management	6
14	worldwide web	13	33	india	6
15	ontology	13	34	metadata standards	6
16	data handling	12	35	digital preservation	6
17	academic libraries	12	36	web sites	6
18	standards	11	37	canada	6
19	open systems	11			

는 용어지만 철자가 다르게 나타난 용어, 하이픈(-)으로 구분된 용어를 모두 동일한 용어로 바꾸어주기 위하여 단·복수 명사와 철자가 다른 용어는 상위 빈도수의 형태로 변형하고, 하이픈(-)을 가진 키워드는 언더바(_) 형태로 변형하였다.

이상의 데이터 전처리 작업을 통해 수집된 키워드는 총 1,137개였으며, 이후 네트워크 분석 및 다차원 분석에 적절한 키워드 수를 고려하여 〈표 2〉와 같이 빈도수 6회 이상의 키워드 37개를 최종 분석대상으로 선정하였다.

3.2 동시출현단어 행렬 작성

분석 대상 키워드를 선정한 후, 분석 대상 문

헌 727건의 키워드 출현 정보를 엑셀에 입력하고, 이재윤이 개발한 COOC V.0.4 프로그램을 이용하여 빈도 값을 그대로 이용하는 동시출현 단어 행렬, 코사인 유사도 계수로 정규화한 행렬, 피어슨 상관계수에 의한 2차 연관성 행렬을 작성하였다.

동시출현빈도를 분석함에 있어, 빈도값을 그대로 행렬에 이용하는 경우에는 네트워크 분석에서 핵심 노드 위주로 분석되고, 비 핵심노드 간의 관계는 드러나지 않는다는 단점을 가진다(이재윤, 2006a). 이에 본 연구에서는 네트워크 분석에 유사도 공식인 코사인 계수를 적용한 행렬을 사용하였으며, 군집분석과 다차원분석에 피어슨 상관계수를 적용한 2차 연관성 행렬을 사용하였다.

코사인 유사도 계수로 키워드 간의 동시출현 빈도를 정규화한 행렬은 유사도가 1에 가까울수록, 두 단어는 유사도가 높고, 0에 가까울수록 유사도가 낮음을 의미하는데, 이는 유사도가 높은 키워드 쌍은 서로 주제적으로 연관성이 높으며, 유사도가 낮은 키워드 쌍은 주제적으로 연관성이 낮다는 것을 의미한다. 피어슨 상관관계수에 의한 2차 연관성 행렬은 White와 Griffith(1981)가 제안한 방법으로, 1차 연관성 행렬을 피어슨 상관관계수 등으로 한 차례 더 가공하여 산출하는 것으로, 두 키워드와 제 3의 키워드 간의 동시출현 패턴의 유사성을 측정하는데 적용된다. 피어슨 상관관계수에 의해 산출된 값은 -1에서 1사이의 범위에 존재하며, 관계의 크기와 방향을 동시에 나타낸다. 상관관계수의 절댓값이 크면 두 키워드 사이가 밀접하게 연관되어 있음을 의미하며, 1에 가까울수록 강한 긍정관계를, 0은 두 키워드 간에 선형적 관련성이 없음을, -1에 가까울수록 강한 부정 관계를 의미한다.

본 연구에서 산출한 코사인 유사도 행렬을 분석한 결과, 유사도가 0인 키워드 쌍은 448개로 나타났으며, 유사도가 가장 낮게 나타난 키워드 쌍은 metadata-cataloguing(0.02581)으로 나타났다. 유사도가 가장 높은 키워드 쌍은 linked data-semantic web(0.45227)으로 나타났으며, 최댓값을 제외하고 유사도 값이 0.3 이상인 키워드 쌍은 총 3개로 linked data-semantic web(0.45227), digital storage-archives management(0.30861), canada-theses(0.36927)순으로 나타났다.

본 연구에서 산출한 2차 연관성 행렬을 분석함에 있어서는 Guilford(1950)가 제시한 지침

을 적용하여 상관관계의 정도를 해석하였다. 즉, 상관관계수의 절댓값이 0.9 이상인 경우에는 매우 높은 상관관계를, 0.7 이상 0.9 미만인 경우에는 높은 상관관계를, 0.4 이상 0.7 미만인 경우는 낮은 상관관계를, 0.2 이상 0.4 미만인 경우는 거의 무시할 만한 상관관계를 가지는 것으로 해석하였다. 피어슨 상관관계의 값을 분석한 결과, 양의 상관관계가 가장 높은 키워드 쌍은 linked data-semantic web(0.74828)으로 코사인 유사도 계수 행렬에서 가장 높은 유사도를 보인 키워드 쌍과 동일한 것으로 나타났다. 최댓값을 제외하고 양의 상관관계수 값이 0.5 이상인 키워드 쌍은 총 7개로 canada-theses(0.60644), dublin core-metadata(0.60156), archives management-digital storage(0.59806), india-digital storage(0.57029), data management-linked data(0.53994), data management-metadata(0.53236), archives management-digital libraries(0.51967)순으로 나타났다.

한편, 음의 상관관계를 가지는 키워드 쌍은 총 395개로, 음의 상관관계가 가장 큰 키워드 쌍은 dublin core-digital storage(-0.15777)로 나타났으며, data management-dublin core(-0.14866), data management-classification(-0.14637), data management-data handling(-0.14133)순으로 나타났다.

4. 분석 결과

4.1 연도별 및 학술지별 분석

본 연구를 위하여 수집한 문헌정보학 분야에

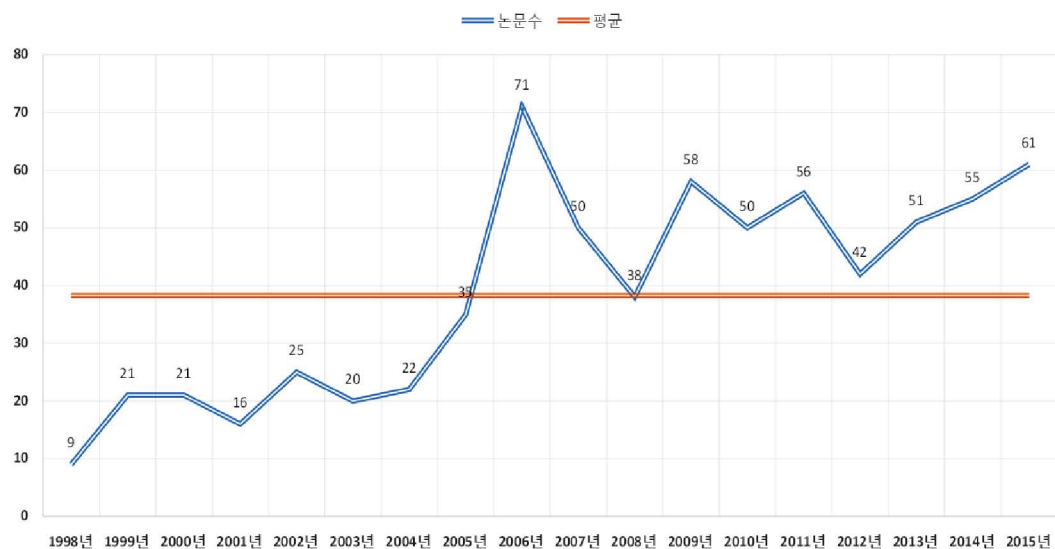
서 메타데이터와 관련된 논문 727건에 대하여 1998년부터 2015년도까지 연도별 출판 건수를 분석한 결과, 연도별 평균 38.94건의 논문이 출판된 것으로 확인되었다. 2006년에 71건의 논문이 출판되어 메타데이터에 대한 연구가 급격하게 증가한 것을 확인할 수 있었으며, 2008년 이후부터는 평균 출판 건수보다 많은 수의 논문이 출판되고 있는 것으로 파악되어, 메타데이터 분야가 꾸준히 연구되고 있다는 것을 확인할 수 있었다(〈그림 1〉 참조).

더불어, 메타데이터 관련 논문이 수록된 학술지에 대한 정량적 분석을 수행한 결과, 메타데이터 관련 논문을 출판한 학술지는 총 72종으로, 학술지 한 종당 평균 10.04건의 논문이 출판된 것으로 나타났다. 가장 많은 논문을 출판한 학술지는 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY로 2016년 7월 8일까

지 총 51건의 논문을 출판한 것으로 나타났으며, 메타데이터 관련 논문을 10건 이상 출판한 학술지는 총 20종으로 집계되었다(〈표 3〉참고).

4.2 네트워크 분석에 의한 지적구조

본 연구에서는 메타데이터 주제 분야의 내부 개념 사이의 관계를 시각화하고, 여러 개의 군집으로 분할하여 세부 주제 영역을 파악할 수 있도록 출현빈도가 6회 이상인 저자키워드 37개의 코사인 유사도 행렬을 산출하고, 패스파인더 네트워크(PFNet) 알고리즘과 병렬 최근접 이웃 클러스터링 알고리즘(PNNC)을 적용하여 네트워크를 생성하였다. 각각의 알고리즘을 적용하기 위해서 이재윤이 개발한 Wnet V.0.4.1 프로그램을 사용하였으며, 이를 시각화하기 위해서 NodeXL을 사용하였다.



〈그림 1〉 연도별 메타데이터 관련 논문 출판 건수(1998년~2015년)

〈표 3〉 학술지별 메타데이터 관련 논문 출판 건수

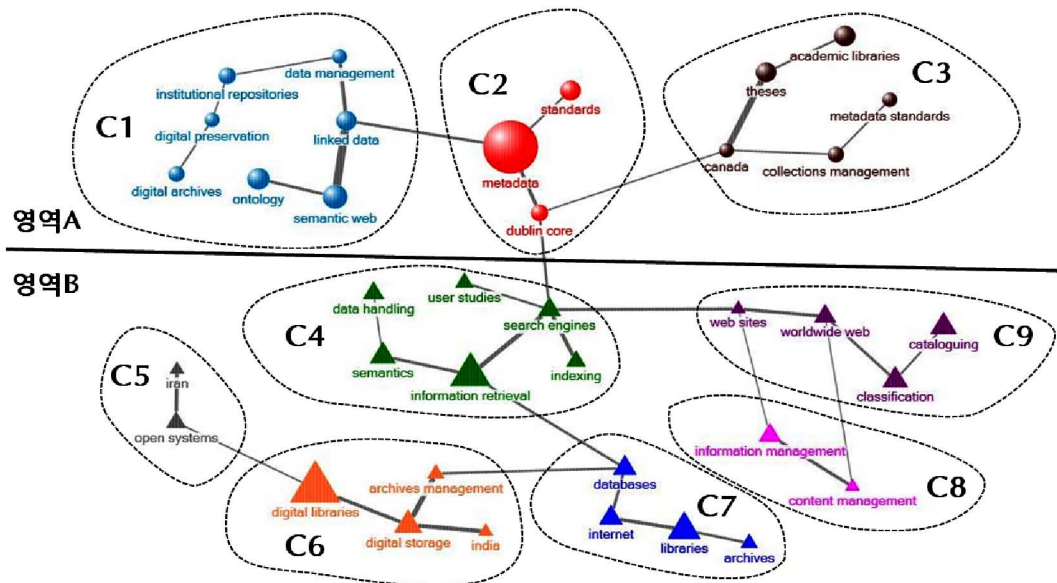
순위	학술지명	논문 수
1	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	51
2	LIBRARY HI TECH	47
3	ONLINE INFORMATION REVIEW	43
4	ELECTRONIC LIBRARY	41
5	JOURNAL OF INFORMATION SCIENCE	41
6	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	40
7	RESEARCH AND ADVANCED TECHNOLOGY FOR DIGITAL LIBRARIES	38
8	PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS	36
9	INFORMATION PROCESSING & MANAGEMENT	31
10	INFORMATION TECHNOLOGY AND LIBRARIES	28
11	KNOWLEDGE ORGANIZATION	26
12	LIBRARY RESOURCES & TECHNICAL SERVICES	26
13	INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE	21
14	JOURNAL OF DOCUMENTATION	19
15	ASIST ANNUAL MEETING	18
	* PROCEEDINGS OF THE 64TH ASIST ANNUAL MEETING VOL 38 2001	2
	* PROCEEDINGS OF THE 65TH ASIST ANNUAL MEETING VOL 39 2002	8
	* PROCEEDINGS OF THE 66TH ASIST ANNUAL MEETING VOL 40 2003	8
16	ASLIB PROCEEDINGS	18
17	SERIALS REVIEW	16
18	LIBRARY TRENDS	14
19	JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY	10
20	SCIENTOMETRICS	10
21	기타	153
총 합계		727

키워드 37개에 대한 동시출현단어의 네트워크를 분석한 결과, PNNC 알고리즘에 의하여 9개, 2개의 최적 군집이 생성되었다. 메타데이터 분야에 대한 세부 영역을 보다 자세하게 파악하기 위하여, 〈그림 2〉와 같이 각 키워드의 출현 빈도를 노드의 크기로 설정하고, 노드 색상과 노드의 형태를 다르게 하여 2개의 영역, 9개의 군집으로 구분하였다.

또한, 메타데이터 분야 전반에 걸쳐 중심이 되는 키워드와 군집 내에서 연결의 중심이 되는 키워드를 파악하기 위하여 NodeXL에서 제공

하는 분석 툴을 사용하여 근접중심성과 매개중심성을 측정하고, 이재윤(2006b)이 가중네트워크에서 사용하기 위하여 제안한 상대적 삼각매개중심성, 평균연관성의 값을 산출하였다.

코사인 유사도에 기초한 메타데이터 네트워크에서 매개중심성과 근접중심성이 높은 상위 10위 키워드는 〈표 4〉와 같이 나타났다. 이를 통해 search engines, dublin core, information retrieval, databases, metadata, linked data, web site, archive management, canada, digital storage는 메타데이터 분야 전반에서 각 키워드를



〈그림 2〉 코사인 유사도 행렬을 이용한 메타데이터 분야 동시출현 네트워크
(2영역, 9군집, 노드 크기: 빈도 수)

〈표 4〉 NodeXL 분석 툴을 사용한 메타데이터 분야 키워드 37개에 대한 중심성 지수 상위 10위

순위	키워드	매개중심성	순위	키워드	근접중심성
1	search engines	432,000	1-2	search engines	0.008
2	dublin core	353,000		dublin core	0.008
3	information retrieval	308,000	3-6	information retrieval	0.007
4	databases	261,000		web sites	0.007
5	metadata	231,000		metadata	0.007
6	linked data	188,000		databases	0.007
7	web sites	156,500	7-10	canada	0.006
8	archives management	155,000		indexing	0.006
9	canada	132,000		user studies	0.006
10	digital storage	131,000		semantics	0.006

연결해주는 역할을 수행하는 키워드이며, search engines, dublin core, information retrieval, web sites, metadata, databases, canada, indexing, user studies, semantics는 메타데이터 분야에서 다른 키워드와 연관되어 다루어지는 키워드로 파악할 수 있었으며, 특히, search engines

와 dublin core는 매개중심성과 근접중심성이 가장 높은 것으로 나타나 메타데이터 분야 전 영역에 걸쳐 중심이 되는 키워드인 것으로 나타났다.

코사인 유사도에 기초한 메타데이터 네트워크에서 전역중심성인 상대적 삼각매개중심성

〈표 5〉 메타데이터 분야 키워드 37개에 대한 가중 네트워크의 전역중심성 지수 상위 10위

순위	키워드	상대적 삼각매개중심성 (rTBC: 0~1)	순위	키워드	평균연관성 (AVGSIM)
1	digital libraries	0.52063	1	digital libraries	0.07473
2	metadata	0.43651	2	information retrieval	0.07176
3	information retrieval	0.28254	3	metadata	0.0679
4	libraries	0.24603	4	libraries	0.0596
5	digital storage	0.14603	5	search engines	0.05394
6	open systems	0.14286	6	worldwide web	0.05393
7	databases	0.13968	7	digital storage	0.05371
8	worldwide web	0.12698	8	open systems	0.04978
9	classification	0.12381	9	classification	0.04972
10	search engines	0.09365	10	databases	0.04863

과 평균연관성이 높은 상위 10위 키워드는 〈표 5〉와 같이 나타났다. 상대적 삼각매개중심성과 평균연관성의 순위는 서로 일치하지 않지만, digital libraries의 경우, 상대적 삼각매개중심성과 평균연관성이 모두 높은 것으로 나타났으며, 각 전역중심성 지수의 상위 2위부터 상위 10위의 범위에 해당하는 키워드인 digital libraries, metadata, information retrieval, libraries, digital storage, open systems, databases, worldwide web, classification, search engines는 그 순위는 일치하지 않지만 모두 매개중심성과 근접중심성이 높은 것으로 나타나, 이 키워드들은 모두 메타데이터 분야에서 다른 주제어들과 광범위하게 연계되어 중심적으로 다루어지는 키워드인 것으로 나타났다.

4.3 군집분석과 다차원축적지도에 의한 지적구조

네트워크 분석 알고리즘에 의해 형성된 군집의 결과를 보완하기 위하여, 통계 프로그램인 SPSS ver. 20.0을 이용하여 피어슨 상관관계수에

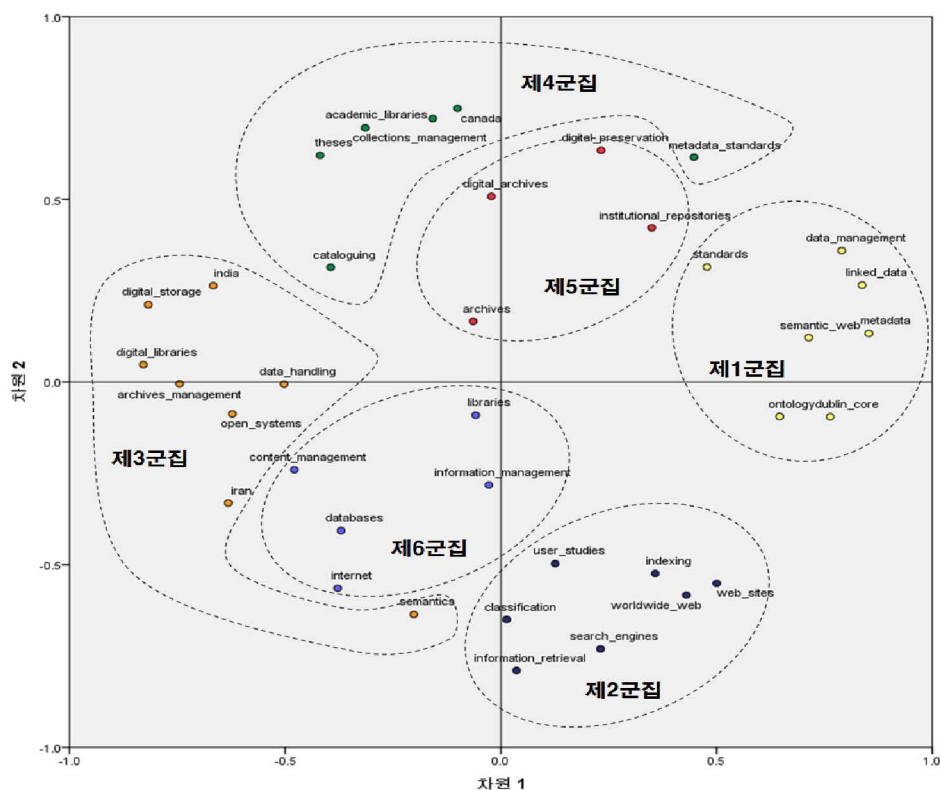
의한 2차 연관성 행렬로 군집분석을 실시하였다. 클러스터링 알고리즘으로는 계층적 클러스터링 기법인 Ward 기법을 사용하였으며, z점수로 표준화하고, 제곱 유클리디안 거리를 사용하여 덴드로그램을 도출하였다. 네트워크 분석에서 도출한 군집을 참고하여, 메타데이터 영역의 연구 분야를 나타내줄 수 있는 군집의 수를 6개로 결정하고, 〈표 6〉과 같이 정리하였다.

이후, 군집분석을 통해 도출된 군집을 기반으로 다차원축적지도를 작성하기 위하여 통계 프로그램인 SPSS ver. 20.0을 통해 PROXSCAL 알고리즘을 적용하여 다차원척도분석을 수행하였다. 다차원 척도 분석 결과, 산출된 스트레스 값은 0.06245으로 Kruskal이 제시한 스트레스 값을 기준으로 보았을 때, 대상들 간의 적합도는 보통인 것으로 나타났다. 각 키워드들 간의 위치를 2차원 공간상에 점으로 나타내고, 〈그림 3〉과 같이 군집분석을 통해 도출된 6개 군집 영역에 속한 키워드를 색깔별로 구분하고, 각 군집의 경계를 점선으로 표시하였다.

다차원축적 지도에 나타난 군집분석의 결과를 해석한 결과, 지도의 X축을 기준으로 바깥

〈표 6〉 군집분석에 의한 6군집

군집	키워드	빈도수	군집	키워드	빈도수
제1군집	metadata	79	제2군집	information retrieval	44
	semantic web	16		classification	17
	ontology	13		search engines	14
	linked data	11		worldwide web	13
	standards	11		indexing	10
	dublin core	8		user studies	9
제3군집	data management	6		web sites	6
	digital libraries	68	제4군집	cataloguing	19
	digital storage	21		academic libraries	12
	semantics	18		theses	11
	data handling	12		collections management	7
	open systems	11		metadata standards	6
	archives management	8		canada	6
	india	6	제6군집	libraries	29
	iran	6		internet	15
제5군집	archives	8		databases	15
	institutional repositories	8		information management	13
	digital archives	7		content management	6
	digital preservation	6			



〈그림 3〉 다차원축적 지도에 나타난 메타데이터 분야 지적구조

쪽 우측부터 중심부까지 시계방향으로 군집분석에 의해 형성된 제1군집, 제2군집, 제3군집, 제4군집, 제5군집, 제6군집이 위치하는 것으로 확인되었으며 특히, 제3군집과 제4군집에 속한 키워드들은 상대적으로 다른 군집에 속한 키워드보다 넓게 분포하고 있는 것으로 나타났다.

4.4 종합분석

코사인 유사도 행렬에 기초하여 생성한 메타데이터 분야 키워드 37개에 대한 네트워크 지도에 병렬 최근접 이웃 클러스터링(PNNC) 알고리즘을 적용한 결과, 2영역 9군집의 네트워크가 생성되었다. 각 군집에서 빈도수가 높은 키워드를 해당 군집을 대표하는 키워드로 보는 경우, 메타데이터(metadata)[A-C2], 시맨틱 웹(semantic web)[A-C1], 학술 도서관(academic libraries)[A-C3]이라는 군집들로 구성된 영역(A)과 도서관(libraries)[B-C7], 정보 관리(information management)[B-C8], 목록(cataloguing)[B-C9], 정보 검색(information retrieval)[B-C4], 개방형 시스템(open systems)[B-C5], 디지털 도서관(digital libraries)[B-C6]라는 군집들로 구성된 영역(B)으로 구분되었다.

또한 코사인 유사도 행렬에 기초한 네트워크 지도에서 메타데이터(metadata), 데이터베이스(databases), 정보 검색(information retrieval), 검색 엔진(search engines)은 매개중심성과 근접중심성, 상대적 삼각매개중심성과 평균연관성이 모두 높은 키워드인 것으로 나타났으며, 군집 B-C5 개방형 시스템(open systems)과 군집 B-C8 정보 관리(information management)를 제외하고 모든 군집이 매개중심성이 높은 키

워드를 하나 이상 포함하고 있는 것으로 나타났다. 특히 군집 B-C4 정보 검색(information retrieval)은 근접중심성이 높은 키워드 5개(information retrieval, search engines, semantics, indexing, user studies)를 포함하고 있는 것으로 나타나, 정보 검색(information retrieval) 영역이 메타데이터 분야에서 가장 일반적인 영향력을 가지면서, 전반적으로 다른 키워드들과 연계되는 주요 영역임을 확인할 수 있었다.

피어슨 상관계수 행렬에 의한 2차 연관성 행렬에 기초한 군집분석을 실시한 결과, 메타데이터 분야는 총 6개 군집을 형성하는 것으로 파악되었으며, 각 군집에서 빈도수가 높은 키워드를 해당 군집을 대표하는 키워드로 보는 경우, 제1군집 메타데이터(metadata), 제2군집 정보 검색(information retrieval), 제3군집 디지털 도서관(digital libraries), 제4군집 목록(cataloguing), 제5군집 아카이브(archives)와 기관 레포지토리(institutional repositories), 제6군집 도서관(libraries)으로 구분되는 것으로 나타났다. 군집분석 결과를 다차원적도 지도에 표시한 결과, 다차원적도 지도상에서 군집분석에 의해 형성된 6개 군집들은 서로 분리되어 위치하고 있는 것으로 나타났다. 특히, 제5군집과 제6군집은 다른 군집들에 비하여 상대적으로 지도상의 중심부에 위치하고 있어, 두 군집에 속한 키워드들이 메타데이터 분야의 다른 군집들과 상관관계가 높으며, 현재 메타데이터 분야의 연구의 중심을 담당하고 있는 것으로 나타났다.

네트워크 분석과 군집분석으로 도출된 군집들을 종합적으로 비교·분석한 결과, 코사인 유사도 행렬에 기초한 네트워크 지도로 생성한 군집 A-C2와 A-C1 일부, A-C1 일부, B-C7

과 B-C8, A-C3과 B-C9 일부, B-C9 일부와 B-C4 일부, B-C4 일부와 B-C5, B-C6은 각각 군집분석을 통해 도출한 제1군집, 제5군집, 제6군집, 제4군집, 제2군집, 제3군집과 동일한 영역인 것으로 나타났다. 더불어, 다차원 척도 지도상에서 x축 하단에 나타난 제3군집, 제6군집,

제2군집은 코사인 유사도 행렬에 기초한 네트워크 지도에서 생성된 'B 영역'과 유사하며, x축 상단에 위치한 제4군집, 제5군집, 제1군집은 코사인 유사도 행렬에 기초한 네트워크 지도에서 생성된 'A 영역'과 유사한 것으로 나타났다. 이를 정리하면 <표 7>과 같다.

<표 7> 군집분석 및 다차원척도 분석, 네트워크 분석으로 도출된 메타데이터 키워드 군집 비교

군집분석 및 다차원척도분석 군집	키워드(빈도수)	중심성 지수 순위				네트워크 분석 영역-군집
		매개	근접	삼각매개	평균연관성	
제1군집 (메타데이터)	metadata (79)	5	3-6	2	3	A-C2 메타데이터
	standards (11)					
	dublin core (8)	2	1-2			
	semantic web (16)					A-C1 시멘틱 웹
	ontology (13)					
	linked data (11)	6				
제5군집 (기관 레포지토리)	data management (6)					
	digital preservation (6)					B-C7 도서관
	institutional repositories (8)					
	digital archives (7)					
제6군집 (도서관)	archives (8)					B-C8 정보 관리
	libraries (29)			4	4	
	internet (15)					
	databases (15)	4	3-6	7	10	A-C3 학술 도서관
제4군집 (학술 도서관)	information management (13)					
	content management (6)					
	academic libraries (12)					
	theses (11)					
	collections management (7)					B-C9 목록
제2군집 (정보 검색)	metadata standards (6)					
	canada (6)	9	7-10			
	cataloguing (19)					B-C4 정보 검색
	classification (17)			9	9	
	worldwide web (13)			8	6	
제3군집 (디지털 도서관)	web sites (6)	7	3-6			B-C5 개방형 시스템
	information retrieval (44)	3	3-6	3	2	
	search engines (14)	1	1-2	10	5	
	indexing (10)		7-10			B-C6 디지털 도서관
	user studies (9)		7-10			
	semantics (18)		7-10			
	data handling (12)					B-C6 디지털 도서관
	open systems (11)			6	8	
	iran (6)					
	digital libraries (68)			1	1	B-C6 디지털 도서관
	digital storage (21)	10		5	7	
	archives management (8)	8				
	india (6)					

5. 결론 및 시사점

본 연구에서는 메타데이터 분야의 지적구조를 파악하기 위하여, WoS 핵심컬렉션에 등재된 저널 중 문헌정보학 분야에서 연구된 메타데이터 관련 논문 727건과 그에 대한 키워드 정보를 수집하여 동시출현단어 분석을 실시하고, 네트워크 분석 및 군집분석과 다차원척도분석을 수행하여 메타데이터 분야의 지적 구조를 시각적으로 표현하였다.

수집된 논문 727건에 대한 1,137개의 키워드 중 출현빈도 6회 이상의 키워드 37개를 선정하고, 코사인 유사도 계수에 따른 동시출현단어 행렬에 기초하여 키워드 간의 관계를 패스파인더 네트워크(PFNet)로 시각화하고, 병렬 최근접 이웃 클러스터링 기법(PNNC)을 적용하여 군집을 생성하고 중심성 분석을 수행하였다. 또한 네트워크 분석을 보완하기 위하여 피어슨 상관 계수에 따른 동시출현단어 행렬에 기초하여 군집분석을 실시하고, 그 결과를 다차원척도적 지도에 반영하여 메타데이터 분야에 대한 지적구조를 제시하였으며, 그 결과를 요약하면 다음과 같다.

첫째, 네트워크 분석 결과, 코사인 유사도 행렬에 기초한 네트워크는 메타데이터와 더블린 코어 그리고 이와 관련된 표준에 대한 연구영역(C2), 시멘틱 웹에 기반한 온톨로지, 링크드 데이터, 데이터 관리에 대한 연구와 기관 레포지토리의 디지털 아카이브 및 디지털 보존에 대한 연구영역(C1), 학술도서관의 장서관리, 메타데이터 표준에 대한 연구영역(C3)을 포함하는 'A 영역'과 도서관, 아카이브, 인터넷, 데이터베이스 전반에 대한 연구영역(C7), 정보관리와 콘

텐츠 관리에 대한 연구영역(C8), 웹 자원에 대한 분류·목록(C9), 정보검색, 시멘틱 웹, 검색 엔진과 관련된 데이터 처리, 색인, 이용자 연구 등에 대한 연구영역(C4), 개방형 시스템에 대한 연구영역(C5), 디지털 도서관, 디지털 저장소와 관련된 기록관리에 대한 연구영역(C6)을 포함하는 'B 영역'으로 구분되어, 결과적으로 2영역 9군집을 생성하는 것으로 나타났다.

또한, 분석결과 네트워크에서 매개중심성과 근접중심성이 가장 높은 키워드는 검색 엔진(search engines)과 더블린 코어(dublin core)인 것으로 나타났으며, 매개중심성과 근접중심성과 함께 전역중심성인 상대적 삼각매개중심성, 평균연관성을 포함한 4가지 중심성 지수가 모두 높은 것으로 나타난 키워드는 메타데이터(metadata), 데이터베이스(databases), 정보 검색(information retrieval), 검색 엔진(search engines)인 것으로 나타났다. 즉, 이는 검색 엔진(search engines)과 더블린 코어(dublin core)는 메타데이터 분야의 다른 키워드들을 중개하는 키워드이자, 메타데이터 네트워크 전역에서 가장 일반적인 영향력을 가지는 키워드이며, 메타데이터(metadata), 데이터베이스(databases), 정보 검색(information retrieval), 검색 엔진(search engines)은 메타데이터 관련 키워드들의 연관성을 고려하였을 때 다른 키워드들과 높은 결속력을 가지며 근접하게 자리하는 키워드라고 해석할 수 있으며, 이를 참고하여 네트워크 분석으로 생성된 군집을 해석한 결과, 메타데이터와 더블린코어 그리고 이와 관련된 표준에 대한 연구영역(A-C2), 정보검색, 시멘틱 웹, 검색엔진과 관련된 데이터 처리, 색인, 이용자 연구 등에 대한 연구영역(B-C4), 도서관, 아카이

브, 인터넷, 데이터베이스 전반에 대한 연구영역(B-C7)은 근접중심성과 매개중심성이 모두 높은 키워드를 포함하고 있는 주요 연구영역인 것으로 파악되었다.

둘째, 피어슨 상관계수 행렬에 기초한 2차 연관성 행렬에 기초하여 도출한 군집분석의 결과, 메타데이터 분야는 메타데이터, 더블린 코어와 같은 관련 표준, 그리고 시멘틱 웹에 기반한 온톨로지, 링크드 데이터, 데이터 관리와 관련된 제1군집, 정보 검색, 분류, 검색엔진과 웹사이트, 색인 및 이용자 연구와 관련된 제2군집, 디지털 도서관 및 디지털 저장소의 데이터 처리, 개방형 시스템, 기록관리와 관련된 제3군집, 학술도서관과 장서관리 및 목록, 이와 관련된 메타데이터 표준과 관련된 제4군집, 아카이브, 기관 레포지터리, 디지털 아카이브의 디지털 보존과 관련된 제5군집, 도서관, 인터넷, 데이터베이스에서의 정보관리 및 콘텐츠 관리와 관련된 제6군집으로 구분되었다. 이를 다차원척도 지도에 표시하여 해석한 결과, 다른 군집과의 상관관계가 높다고 해석할 수 있는 지도의 중심부에는 군집분석에 의해서 형성된 제5군집과 제6군집에 해당하는 키워드들이 위치하는 것으로 나타났다. 대부분 키워드들은 확연하게 분리되어 군집으로 구분되었으며, 제3군집과 제4군집에 속하는 키워드들은 다른 군집에 속한 키워드들에 비하여 상대적으로 넓게 분포하고 있는 것으로 나타났다.

셋째, 네트워크 분석에 기초한 지적구조와 군집분석과 다차원척도지도로 표현한 지적구조의 결과를 비교해보면, 두 분석 기법의 알고리즘에 의한 차이로 인하여 각 분석으로 도출된 군집의 세부 키워드들이 완벽하게 일치하지는 않지만, 네트워크 분석에 의해 구분된 'A 영역'과 'B 영

역'은 각각 다차원척도지도상에서 x축 상단에 위치한 제4, 5, 1군집과 x축 하단에 위치한 제3, 6, 2군집과 유사한 것으로 나타났다.

이상의 분석에 근거하여 살펴보았을 때, 전반적으로 메타데이터, 정보검색, 검색엔진, 데이터베이스가 메타데이터 분야에서 중요한 역할을 담당하고 있는 것으로 나타났다. 또한, 메타데이터 분야는 크게 자원이나 정보를 조직하고 관리하는 메타데이터와 더블린코어 같은 관련 표준에 대한 연구와 링크드 데이터, 시멘틱 웹, 기관 레포지터리의 데이터 관리와 디지털 보존, 학술 도서관의 장서관리 등에 대한 연구를 포함하는 영역과 정보검색과 관련하여 정보를 활용할 수 있도록 도서관이나 아카이브, 인터넷에서 데이터베이스를 구축하고, 자원에 대한 분류·목록을 통해 정보 및 콘텐츠를 관리하는 것에 관한 연구영역으로 구분할 수 있는 것으로 나타났다.

본 연구는 문헌정보학 분야에서 연구되는 메타데이터 관련 논문을 수집함에 있어 WoS에서 언어가 영어인 논문으로 한정하였기 때문에, 메타데이터에 관한 국내 연구의 지적 구조를 파악하지 못했다는 점과 수집된 서지정보에서 키워드를 추출함에 있어서 WoS에서 제공하는 저자 키워드만을 사용하여 논문의 제목이나 초록에 수록된 중요 키워드를 반영하지 못했다는 한계점이 존재한다. 하지만, 기존의 연구가 연구자가 미리 정해놓은 기준에 의해 관련 논문을 분류하는 상향식 분류를 통하여 메타데이터 분야의 연구 영역을 파악했다면, 본 연구는 최근 문헌정보학의 정보조직 분야에서 중요한 주제로 떠오르고 있는 메타데이터 분야의 키워드를 수집하고, 이에 대한 네트워크 분석을 수행하여

해당 분야의 세부 연구영역을 파악하고, 그 지적구조를 시각적으로 표현하였다는 점에서 의의를 갖는다.

본 연구의 한계점을 고려하여 향후, 논문의 출판 시기를 고려하여 키워드를 수집하여 메타데이터에 대한 시기별 지적구조 분석을 수행하거나, Web of Science, Scopus, Google Scholar, 한국학술지인용색인(KCI) 등 다양한 인용색인 DB를 활용하고, 저자 키워드뿐만 아니라 제목과 초록에서 키워드를 추출하여 동시출현단어 분석을 수행하고, 추가적으로 실제 문헌에 대한 내용분석을 함께 수행하여 메타데이터 분야의

구체적인 지적 구조를 분석한다면, 메타데이터 분야의 발전방향을 제시하거나, 연구의 방향성을 모색하는데 유용하게 사용될 수 있을 것이다. 또한, 도출된 지적구조를 정보조직 분야와 관련된 강의계획안을 수집하여 정보조직 분야 교과과정과 비교·분석한다면, 메타데이터를 교육하는데 있어 기초자료로 유용하게 사용될 수 있을 것이며, 동시출현단어 분석을 수행함에 있어 키워드의 출현빈도를 다르게 설정하여 그에 따른 결과에 대한 차이점을 파악한다면, 메타데이터 분야의 지적 구조를 보다 면밀히 분석할 수 있을 것이다.

참 고 문 헌

- 강범일, 박지홍 (2013). 프로파일링 분석과 동시출현단어 분석을 이용한 한국어교육학의 정체성 분석. 정보관리학회지, 30(4), 195-213. <http://dx.doi.org/10.3743/kosim.2013.30.4.195>
- 김완중 (2014). 동시출현 단어분석을 활용한 빅데이터 관련 연구동향 분석. 한국정보관리학회 학술대회 논문집, 17-20.
- 김희정 (2011). 네트워크 분석을 기반으로 한 웹 아카이빙 주제 영역 연구. 한국비블리아학회지, 22(2), 235-248.
- 김하진, 송민 (2014). 동시출현단어 분석을 통한 국내외 정보학 학회지 연구동향 파악. 정보관리학회지, 31(1), 99-118. <http://dx.doi.org/10.3743/kosim.2014.31.1.099>
- 권진영, 이우형, 김미현, 이석준 (2010). 동시 단어 분석(Co-word Analysis)을 이용한 EA(Enterprise Architecture) 연구동향 분석. 정보화연구, 7(1), 31-47.
- 박옥남 (2011). 정보조직 지식구조에 대한 연구: 2000년-2011년 학술논문을 중심으로. 한국비블리아학회지, 22(3), 247-267.
- 박재신, 정영미 (2010). 지구적 환경문제 해결을 위한 학술활동과 환경운동 경향 연구. 정보관리학회지, 27(3), 83-102. <http://dx.doi.org/10.3743/kosim.2010.27.3.083>
- 서선경, 정은경 (2013). 동시출현단어 분석 기반 오픈 액세스 분야 지적구조에 관한 연구. 한국비블리아학회지, 24(1), 207-228. <http://dx.doi.org/10.14699/kbiblia.2013.24.1.207>
- 서원철, 박현석, 윤장혁 (2012). 단어동시출현분석을 통한 한국의 국가 R&D 연구동향에 관한 탐색적

- 연구. 정보기술응용학회, 19(4), 1-18.
- 오세훈 (2005). 우리나라 문헌정보학회지 논문 및 인용문헌 분석을 통한 연구동향 연구. 정보관리학회지, 22(3), 379-408. <http://dx.doi.org/10.3743/kosim.2005.22.3.379>
- 유사라 (2010). 메타데이터 주제 국내 연구동향 분석. 한국문헌정보학회지, 44(2), 405-426. <http://dx.doi.org/10.4275/kslis.2010.44.2.405>
- 이미경, 정영미 (2003). 동시출현 단어 분석을 통한 지식 구조의 파악: 인공지능 분야를 대상으로. 한국 정보관리학회 학술대회 논문집, 245-251.
- 이우형, 김윤명, 박각로, 이명호 (2006). Co-word Analysis을 통한 신기술 분야 도식화 방법에 관한 연구. 經營科學, 23(3), 77-93.
- 이재윤. COOC. (Version 0.4) [Computer Software].
- 이재윤. WNET. (Version 0.4.1) [Computer Software].
- 이재윤 (2006a). 지적 구조의 규명을 위한 네트워크 형성방식에 관한 연구. 한국문헌정보학회지, 40(2), 333-355. <http://dx.doi.org/10.4275/kslis.2006.40.2.333>
- 이재윤 (2006b). 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 한국문헌정보학회지, 40(3), 191-214. <http://dx.doi.org/10.4275/kslis.2006.40.3.191>
- 이정규, 이용구 (2013). 동시출현단어 분석을 이용한 도서관경영 분야의 지적구조 분석. 한국정보관리학회 학술대회 논문집, 23-26.
- 장임숙, 장덕현, 이수상 (2011). 다문화연구의 지식구조에 관한 네트워크 분석. 한국도서관·정보학회지, 42(4), 353-374.
- 장령령, 홍현진 (2014). 학술지 중요도와 키워드 순서를 고려한 단어동시출현 분석을 이용한 독서분야의 지적구조 분석. 한국비블리아학회지, 25(1), 295-318. <http://dx.doi.org/10.14699/kbiblia.2014.25.1.295>
- 조재인 (2011). 네트워크 텍스트 분석을 통한 문헌정보학 최근 연구 경향 분석. 정보관리학회지, 28(4), 65-83. <http://dx.doi.org/10.3743/kosim.2011.28.4.065>
- 허고은, 송민 (2013). 저자동시인용 분석과 동시출현단어 분석을 이용한 의료정보학 저널의 지적구조 분석. 정보관리학회지, 30(2), 207-225. <http://dx.doi.org/10.3743/kosim.2013.30.2.207>
- Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. Information Processing & Management, 37(6), 817-842. [http://dx.doi.org/10.1016/s0306-4573\(00\)00051-0](http://dx.doi.org/10.1016/s0306-4573(00)00051-0)
- Guilford, J. P. (1950). Fundamental statistics in psychology and education. New York: McGraw-Hill.
- Hu, C. P., Hu, J. M., Deng, S. L., & Liu, Y. (2013). A co-word analysis of library and information science in China. Scientometrics, 97(2), 369-382. <http://dx.doi.org/10.1007/s11192-013-1076-7>

- Liu, G. Y., Hu, J. M., & Wang, H. L. (2012). A co-word analysis of digital library field in China. *Scientometrics*, 91(1), 203-217. <http://dx.doi.org/10.1007/s11192-011-0586-4>
- Milojević, S., Sugimoto, C. R., Yan, E., & Ding, Y. (2011). The cognitive structure of library and information science: Analysis of article title words. *Journal of the American Society for Information Science and Technology*, 62(10), 1933-1953. <http://dx.doi.org/10.1002/asi.21602>
- Persson, O. bibexcel [Computer Software]. Retrieved from <http://homepage.univie.ac.at/juan.gorraiz/bibexcel/>
- White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science & Technology*, 32(3), 163-171. <http://dx.doi.org/10.1002/asi.4630320302>

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

- 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. <http://dx.doi.org/10.3743/kosim.2011.28.4.065>
- Heo, Go-Eun, & Song, Min (2013). Examining the intellectual structure of a medical informatics journal with author co-citation analysis and co-word analysis. *Journal of the Korean Society for Information Management*, 30(2), 207-225. <http://dx.doi.org/10.3743/kosim.2013.30.2.207>
- Jang, Im-Sook, Chang, Durk-Hyun, & Lee, Soo-Sang (2011). The knowledge structure of multicultural research papers in Korea. *Journal of Korean Library and Information Science Society*, 42(4), 353-374.
- Kang, Beom-Il, & Park, Ji-Hong (2013). Profiling and co-word analysis of teaching Korean as a foreign language domain. *Journal of the Korean Society for Information Management*, 30(4), 195-213. <http://dx.doi.org/10.3743/kosim.2013.30.4.195>
- Kim, Ha-Jin, & Song, Min (2014). A study on the research trends in domestic/international information science articles by co-word analysis. *Journal of the Korean Society for Information Management*, 31(1), 99-118. <http://dx.doi.org/10.3743/kosim.2014.31.1.099>
- Kim, Hee-Jung (2011). A study on web archiving subject analysis based on network analysis.

- Journal of the Korean Biblia Society for Library and Information Science, 22(2), 235-248.
- Kim, Wan-Jong (2014). The research trends about the big data using co-word analysis. *Proceedings of Korea Society for Information Management*, 17-20.
- Kwon, Jin-Young, Lee, Woo-Hyung, Kim, Mi-Hyun, & Lee, Seog-Jun (2010). A research trend analysis in enterprise architecture using co-word analysis. *Journal of Information Technology and Architecture*, 7(1), 31-47.
- Lee, Jae-Yun. COOC. (Version 0.4) [Computer Software].
- Lee, Jae-Yun. WNET. (Version 0.4.1) [Computer Software].
- Lee, Jae-Yun (2006a). A study on the network generation methods for examining the intellectual structure of knowledge domains. *Journal of the Korean Society for Library and Information Science*, 40(2), 333-355. <http://dx.doi.org/10.4275/kslis.2006.40.2.333>
- Lee, Jae-Yun (2006b). Centrality measures for bibliometric network analysis. *Journal of the Korean Society for Library and Information Science*, 40(3), 191-214. <http://dx.doi.org/10.4275/kslis.2006.40.3.191>
- Lee, Jung-Gyu, & Lee, Yong-Gu (2013). A study on the analysis of intellectual structure of library management studies using co-word analysis. *Proceedings of Korea Society for Information Management*, 23-26.
- Lee, Mi-Kyoung, & Chung, Young-Mee (2003). Exploration of intellectual structure of artificial intelligence field using co-word analysis. *Proceedings of Korea Society for Information Management*, 245-251.
- Lee, Woo-Hyoung, Kim, Yun-Myung, Park, Gak-Ro, & Lee, Myoung-Ho (2006). A study on the emerging technology mapping through co-word analysis. *Korean Management Science Review Special*, 23(3), 77-93.
- 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. <http://dx.doi.org/10.3743/kosim.2005.22.3.379>
- Park, Jae-Shin, & Chung, Young-Mee (2010). An informetric study on academic activities and environmental movements in solving global environmental problems. *Journal of the Korean Society for Information Management*, 27(3), 83-102. <http://dx.doi.org/10.3743/kosim.2010.27.3.083>
- Park, Ok-Nam (2011). Knowledge structures in knowledge organization research: 2000-2011. *Journal of the Korean Biblia Society for Library and Information Science*, 22(3), 247-267.

- Seo, Sun-Kyung, & Chung, Eun-Kyung (2013). Domain analysis on the field of open access by co-word analysis. *Journal of the Korean Biblia Society for Library and Information Science*, 24(1), 207-228.
<http://dx.doi.org/10.14699/kbiblia.2013.24.1.207>
- Seo, Won-Chul, Park, Hyun-Seok, & Yoon, Jang-Hyeok (2012). An exploratory study on the Korean national r&d trends using co-word analysis. *Journal of Information Technology Applications & Management*, 19(4), 1-18.
- Yoo, Sa-Rah (2010). A diagnostic analysis of metadata R&D status in korea. *Journal of the Korean Society for Library and Information Science*, 44(2), 405-426.
<http://dx.doi.org/10.4275/kslis.2010.44.2.405>
- Zhang, Ling-Ling, & Hong, Hyun-Jin (2014). Examining the intellectual structure of reading studies with co-word analysis based on the importance of journals and sequence of keywords. *Journal of the Korean Biblia Society for Library and Information Science*, 25(1), 295-318.
<http://dx.doi.org/10.14699/kbiblia.2014.25.1.295>