

토픽 모델링을 활용한 도서관, 기록관, 박물관간의 연구 주제 분석*

Analysis of Research Topics among Library, Archives and Museums using Topic Modeling

김 희 섭(Heesop Kim)**

강 보 라(Bora Kang)***

< 목 차 >

I. 서론	IV. 분석 결과
II. 이론적 배경	V. 결론
III. 연구 설계	

초 록

본 연구의 목적은 광의의 측면에서 지식정보제공이라는 공동의 임무를 수행하는 도서관, 기록관, 박물관간의 협력 플랫폼 구축에 관한 연구의 동향을 토픽 모델링을 통하여 파악하기 위한 것이다. 연구의 목적을 달성하기 위하여 Scopus로부터 이들 세 기관을 동시에 다루는 논문 637편의 서지정보를 수집하였다. 수집된 서지정보 중에서 초록을 대상으로 NetMiner V.4를 통하여 총 5,218 개의 단어를 추출한 후 토픽모델링 분석하였으며, 그 결과는 다음과 같다. 첫째, tf-idf의 가중치에 따른 단어출현 빈도를 분석한 결과 '보존(Preservation)'이 가장 높게 나타났으며, 둘째, LDA(Latent Dirichlet Allocation) 알고리즘을 통한 토픽모델링 분석 결과 13개의 주제 영역이 도출되었다. 셋째, 13개의 주제 영역을 네트워크로 표현한 결과 '리포지터리 구축(Repository Construction)'을 중심으로 기관간의 협력, 정보자원 보존을 위한 환경 구축, 정부차원에서의 제도와 정책 발굴, 정보자원의 생애주기, 정보자원의 전시, 정보자원의 검색 등이 서로 밀접한 관련성을 가진 것으로 나타났다. 넷째, 13개의 주제 영역의 연도별 동향을 살펴보면, 1998년 이전의 연구는 제도와 정책 발굴, 정보자원의 검색, 정보자원의 생애주기 등과 같이 특정 주제에 한정된 반면, 그 이후의 연구는 보다 다양한 주제를 다룬 것으로 분석되었다.

키워드: 도서관, 기록관, 박물관, 협력 플랫폼, 토픽모델링, 해외 연구 동향

ABSTRACT

The purpose of this study is to understand the topics of the research for the establishment of cooperative platform between libraries, archives, and museums that carry out the common task of providing knowledge information in a broad sense. To achieve the purpose of this study, 637 bibliographic information on three institutions were collected from the Web version of Scopus database. Among the collected bibliographic information, 5,218 words were extracted through NetMiner V.4 and analysed topic modeling. The results are as follows: First, as a result of analyzing the frequency of word appearance according to the tf-idf weight 'Preservation' was the most hottest topic. Second, the topic modeling analysis through LDA(Latent Dirichlet Allocation) algorithm resulted in 13 topic areas. Third, as a result of expressing 13 topic areas as a network, repository construction was the central topic, and the research topics such as cooperation among institutions, conservation environment for collections, system and policy discovery, life cycle of collections, exhibition of information resources, and information retrieval were closely related to the central topic. Fourth, the trend of 13 topic areas by year 1998 is limited to the specific subjects such as system and policy discovery, information retrieval, and life cycle of collections, while the subsequent studies have been carried out after that year.

Keywords: Library, Archives, Museum, Collaborative platform, Topic modeling, Overseas research trends

* 이 성과는 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2017M3C1B6071502).

** 경북대학교 문헌정보학과 교수(heesop@knu.ac.kr) (제1저자)

*** 경북대학교 문헌정보학과 박사수료(brkang88@daum.net) (교신저자)

•논문접수: 2019년 11월 20일 •최초심사: 2019년 11월 27일 •게재확정: 2019년 12월 12일

•한국도서관·정보학회지 50(4), 339-358, 2019. [http://dx.doi.org/10.16981/kliss.50.201912.339]

I. 서론

광의적인 측면에서 보면, 도서관과 기록관 그리고 박물관은 일종의 문화 거점기관으로서 다양한 이용자에게 지식정보를 제공하는 공통의 역할을 수행하는 기관이다. 즉, 학문의 발전과 교육 제공 그리고 문화증진 등 정보제공기관으로서 갖는 공통적인 특성을 고려한다면, 이들 기관 간의 경계와 구분은 임의적이라는 견해가 많은 학자들에 의하여 피력되어 왔으며(Fleckenstein-Gallo 1967; Marty 2014), 한편으로는 이들을 아우르는 다양한 협력 방안이 제시되기도 했다. 예를 들어, 국내의 경우 각 기관의 인식부재, 상이한 정책부서, 경쟁적인 이용자 유치 등과 같은 이유로 인하여 변화를 수용하는 데에 다소의 한계가 있었으나(이미경 2014) 점차 이용자의 요구 수용에 초점을 두고 기관 간의 공간 계획, 공간 유형, 건축 조건, 세부디자인 요소 등의 물리적 요소부터 다기능화, 공공성, 지속가능성 등 개념적 요소까지 반영하며 인포메이션 커먼스(Information Commons)로 확대되기도 했다(한정원 2019). 국외의 경우 1990년대 중반 도서관, 기록관, 박물관 간의 모호해진 경계와 높아진 디지털 자원의 의존도에 따라 Rayward¹⁾는 디지털기반의 도서관과 박물관 간의 네트워킹 연계방안에 관한 연구를 수행하였으며, 이는 시대의 변화와 씨름하면서 정보제공 기관의 정체성을 고민하는 계기를 마련하였다(Marty 2014).

많은 정보제공기관은 문화향유를 위한 환경을 조성하고 이용자들의 능동적인 이용과 협력에 기반한 플랫폼으로서 역할을 수행하고 있으며, 이들의 이용자는 웹 기술, 클라우드 컴퓨팅, 오픈 액세스 등 디지털 기술의 접목으로 소위 TPO(Time, Place, Occasion)에 구애받지 않고 정보 활용이 가능해졌다. 이를 반영하여 표준화된 문화유산 자료실의 전략모델을 제시(Yoo and Choi 2014)하고 신뢰 가능한 디지털 도서관 구축을 위한 연구(Hilario, Fernández and Campo 2014)를 수행하며, 큐레이터, 사서, 아키비스트의 미래 가치를 탐색(Shipp 2016) 하는 등 정보제공기관이 갖는 기능적 특성과 요건을 분석하며 이들 기관 간의 통합정보서비스 제공을 위한 연구자들의 고민은 계속되고 있다(최영실, 이해영 2012)

본 연구의 의의는 토픽모델링을 통하여 지금까지 이루어진 이들 세 기관의 연구 주제를 분석하여 해외 선진 연구의 특성을 파악하고 연구 동향을 체계적으로 살펴봄으로써 앞으로 국내 연구의 발전 방향과 적용 가능성을 예측하는데 있다. 토픽모델링은 구조화되지 않은 대량의 자료에서 숨겨진 주제를 찾아내는 방법론으로 하나의 문헌에 하나의 토픽을 할당하는 일반적인 접근법과는 달리 하나의 문헌에 포함된 여러 개의 토픽을 도출하는 것이 특징이며(Blei 2012), 산업공학, 문헌정보학, 전산학 등의 분야에서 기술 예측 연구를 위해 활발히 적용되고 있다.

1) W. Boyd Rayward는 호주 New South Wales 대학의 문헌정보학과와 미국 Illinois at Chicago 대학의 iSchool의 명예 교수임.

따라서 본 연구의 목적은 지식정보제공이라는 공동의 임무를 수행하는 도서관, 기록관, 박물관간의 협력 플랫폼 구축에 관한 해외 연구의 주제를 파악하기 위한 것이며, 토픽모델링 기법으로 분석한 세부 주제의 주제명을 정의하고 그 연구동향을 도출하고자 한다.

II. 이론적 배경

1. 도서관, 기록관, 박물관 연구 개관

1908년 Heran의 The Lloyd Library and Museum²⁾에 관한 기고글에는 식물학, 의학, 약 조제, 자연사, 과학사, 시각 예술에 관한 자료를 소장하는 정보기관으로 도서관과 박물관을 통합한 역사연구 센터로 소개하였으며, 이는 교육과 연구수행이라는 유사한 기능과 역할을 인식한 초기 통합 기관이라 해석할 수 있다.

최영실과 이해영(2012)은 도서관, 기록관, 박물관은 이용자들에게 다양한 유형의 정보자원을 제공하고, 학문, 교육, 지역문화를 형성하는 정보기관으로 정의하고, 이들의 공통적인 역할은 첫째, 지역문화의 정체성 제고와 문화 거점 역할, 둘째, 이용자 맞춤형 지식정보 서비스 제공, 셋째, 다양한 연계서비스를 통한 문화증진 등을 제안하였다. 이와 같은 역할을 수행하는 세 기관을 일컬어 GLAM(Galleries-Libraries-Archives-Museums) 혹은 LAMs(Libraries-Archives-Museums) 혹은 Larchiveum(Library-Archives-Museums)으로 표현³⁾하며, 이들 기관의 통합 혹은 협력 방안모색으로 공간 계획, 건축 조건, 세부디자인 요소 등 물리적 환경에 관한 연구(한정원 2019)와 전문적인 정보제공을 위한 디지털 큐레이션 교육의 필요성(Tammaro 2014) 등을 강조하였다.

캐나다 LAC(Library and Archives Canada)는 지식의 수집, 보존, 확산을 위한 공동체간의 협업과 지속적인 기억을 위한 정부와 소속기관의 역할 아래 1953년 설립된 국립도서관(National Library of Canada)과 1872년 설립된 기록관(National Archives of Canada)을 2004년 국가차원에서 통합하여 운영하고 있다. 현재와 미래 세대를 위한 문헌유산(The Documentary Heritage) 보존과 접근성 측면에서 문화적·사회적·경제적 발전에 기여하였으며⁴⁾, 이는 통합형 기관의 선도적 역할을 수행하며, 지식정보 활용의 핵심 기능을 담당하는

2) 1870년대 Lloyd형제가 자신들의 연구자료 컬렉션을 위해 설립된 사설 비영리 기관으로 미국 오하이오주 신시내티에 위치함.

3) 자료 수집단계에서 Scopus 데이터베이스 검색 결과는 LAMs 67건, GLAM 32건, Larchiveum 1건으로 나타났으며, 명칭 사용에 대한 비교를 위하여 KCI 데이터베이스를 검색 결과는 GLAM 2건, LAMs 0건, Larchiveum 29건으로 나타났다. 즉, 도서관, 기록관, 박물관을 아우르는 간략 명칭으로 해외 연구자들은 LAMs 또는 GLAM으로 주로 표기하였으나, 국내 연구자들은 주로 Larchiveum으로 표기되고 있음을 확인하였다. [2019. 10. 11]

4) Library and Archives Canada Homepage. <<http://www.bac-lac.gc.ca/eng/about-us/Pages/our-mandate.aspx>> [Cited. 2019. 09. 11].

4 한국도서관·정보학회지(제50권 제4호)

국가기관이라는 평가를 받고 있다(이미경 2014). Yarrow 등(2008)은 도서관, 기록관, 박물관이 제시하는 사명(Mission)과 이들이 제공하는 서비스 유형을 분석하여 세 기관의 유사점을 확인하였으며, 정보접근 방식이 직접 방문에서 온라인 접속으로의 변화됨에 따라 아웃리치 서비스의 정보제공 방식에 보다 이상적인 협력관계 형성이 가능하다는 것을 시사한 바 있다.

대표적으로 유럽의 3,000여개 이상의 도서관, 박물관, 갤러리, 기록관이 참여하는 컬렉션 통합 사이트인 유로피아나(Europeana)⁵⁾는 업무, 학습 혹은 재미를 목적으로 문화자산을 공유하기 위하여 예술품, 유물, 도서, 영화, 음악 자료 등 약5,760만 종의 디지털 자원(이미지, 녹음, 텍스트, 비디오, 3D 자료 등)을 서비스하고 있다.

또한 세계디지털도서관(WDL, World Digital Library)⁶⁾은 미국의회도서관의 프로젝트로 UNESCO의 지원과 더불어 전 세계 193여개 국가의 도서관, 기록관, 박물관, 교육기관, 그리고 국제기구에서 참여하고 있으며, 2019년 현재 약19,147여 건의 문화 콘텐츠(필사본, 지도, 희귀도서, 악보, 녹음, 영화, 인쇄물, 사진, 건축도면 등 기원전 8,000년 전까지의 주요한 문화재 자료 포함)를 제공하는 동시에 디지털 변환 센터(Digital Conversion Center)로서 고품질의 디지털 이미지를 제작할 수 있도록 제작이 어려운 국가를 지원하는 역할도 함께 하고 있다. 이 도서관의 주요 목표로는 국가간 및 이종문화간의 이해증진, 인터넷을 통한 문화 콘텐츠의 양과 다양성을 확장, 교육자·학자·일반인을 위한 자원 제공, 그리고 파트너 기관의 국가내 그리고 국가간의 디지털 격차를 좁힐 수 있는 역량 구축이다. 특징으로는 일관된 메타데이터 기술(Description), 다국어, 디지털도서관 기술개발, 협업 네트워크 등을 꼽을 수 있다.

국내의 경우, 국립중앙도서관은 디지털 정보교류의 센터 역할을 수행하고자 디지털도서관으로 확장하였으며⁷⁾, “디지털 컬렉션”을 통해 책·사람·세상, 세계의 도서관, 전시 컬렉션 등 다양한 주제와 형식으로 디지털 서고를 제공하고 있다⁸⁾. 2016년에는 윤동주, 백석, 서정주 등 한국의 대표 근대 문학작품 전시를 겸하며 자료실의 기능을 확장하는 한편 디지털도서관을 구축하여 종이 발명 이전 시대의 매체를 시대별로 전시하는 기록매체박물관을 선보인 바 있다.

이를 종합해 보면, 도서관, 기록관, 그리고 박물관 등 정보기관협력은 학문 및 문화발전, 지식향상 및 평생교육 기능, 문화를 창출한다는 공통적인 목표와 정보통신기술의 환경변화와 이용자 요구에 따라 통합 및 협업 네트워크 구축, 다양한 유형의 정보자원 수용, 그리고 고품질의 서비스를 제공하기 위한 관련기술 개발의 기능을 수행하는 기관으로 요약할 수 있다.

5) 유럽연합집행기관(European Commission)에서 자금을 지원하여, 2009년 2월 공식적으로 서비스 개시함. 정보검색을 위한 단일표준 메타데이터로 Europeana Semantic Elements 지정사용. 2010년에는 이용자에게 보다 풍부한 정보를 제공해 주기 위한 메타데이터 표준으로 Europeana Data Model 사용. Europeana Collections Homepage. <<https://www.europeana.eu/portal/en>> [Cited. 2019. 09. 17].

6) 2009년 4월 공식적인 온라인 서비스 개시함. World Digital Library Homepage. <<https://www.wdl.org/en/>> [Cited. 2019. 09. 17].

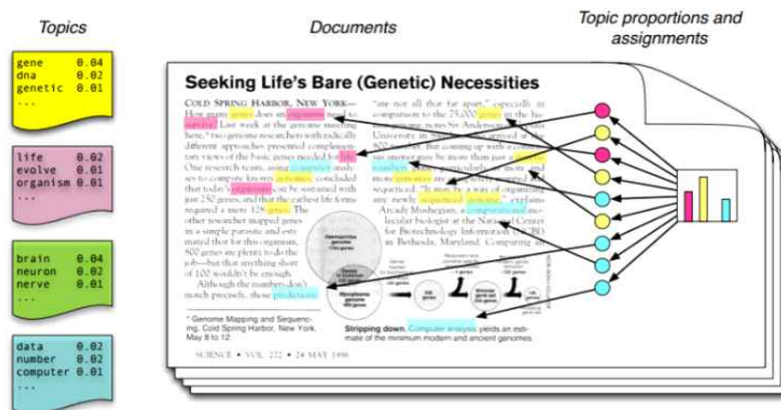
7) 국립중앙도서관. <<https://nl.go.kr>> [인용. 2019. 09. 17].

8) 국립중앙도서관 디지털컬렉션. <<https://ncollection.nl.go.kr>> [인용. 2019. 09. 17].

2. 토픽모델의 개념

기계학습과 자연어처리 분야에서 토픽모델(Topic models)은 문헌집단의 추상적인 주제(Topic)를 발견하기 위한 일종의 통계적 모델이며, 이는 방대한 양의 비구조화된 문헌집단으로부터 의미구조를 발견하기 위해 사용되는 텍스트 마이닝 기법 중 하나이다. 따라서 이것은 급속히 증가하는 정보를 요약하고 이해하는 기술이 요구되는 문헌관리 분야에 있어서 주요한 요소로 자리매김 할 것으로 기대하고 있다(Blei, Ng and Jordan 2003; 박종도 2019). 1998년 Papadimitriou 등이 LSI(Latent Semantic Indexing)이라 불리는 모형을 제안하였고, 현재는 Blei와 Ng 그리고 Jordan(2002)이 개발한 LDA(Latent Dirichlet Allocation, 잠재 디리클레 할당)가 가장 보편적으로 사용되고 있다. 토픽모델은 단일 문헌에 단일 주제만 매칭되는 일반적인 군집화(Clustering)기법과 달리 하나의 문헌이 여러 주제에 동시에 대응될 수 있다는 점이 특징이라고 할 수 있다(김남규 외 2017). 토픽모델링에서 단위 텍스트를 문헌(Document)으로 표현하며, 분석 대상에 해당하는 대량의 문헌집단을 코퍼스(Corpus) 혹은 컬렉션(Collection)라 한다. 또한 N개의 문헌집단인 코퍼스에 등장하는 키워드들의 수 V는 어휘(Vocabulary)라 하며, 각 문헌에 등장하는 V개의 단어들의 패턴을 근거로 전체 K개의 주제를 확률적으로 추정하여 제시한다. 즉, 문헌은 주제의 조합이고, 주제는 단어의 조합을 구성하며, 이것을 토대로 문헌의 주제를 추정할 수 있다(이수상 2012). 토픽모델링의 결과는 ① K개의 주제, ② 각 문헌에 해당하는 주제, ③ 각 주제에 포함되는 단어 등이 제시되며, 문헌과 주제관계의 행렬모드로 주제범주와 코퍼스에 등장하는 어휘, 그리고 문헌에 분포하는 주제 리스트를 확인하며 잠재된 변인들을 추정할 수 있다(이수상 2016).

앞서 언급한 바와 같이 토픽모델에는 잠재의미색인(LSI, Latent Semantic Indexing) 기법을 확장한 확률 잠재의미색인(PLSI, Probabilistic Latent Semantic Indexing), 잠재 디

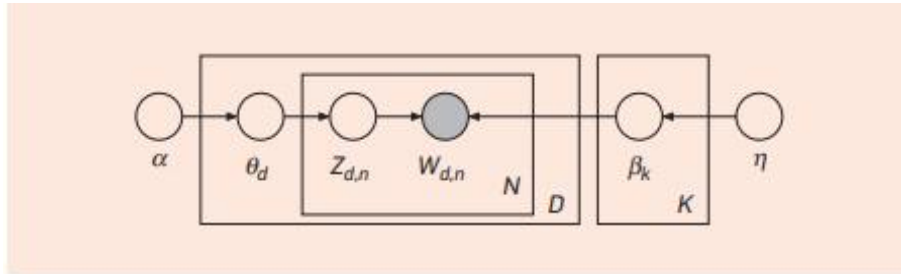


<그림 1> LDA 스키마 (Blei 2012, p.78)

리클레할당(LDA, Latent Dirichlet Allocation) 등이 있다.

이 중 잠재 디리클레할당(LDA, Latent Dirichlet Allocation)은 문헌집단에 존재하는 각 문헌의 주제 구조를 추론하는 확률모델로 산업공학, 문헌정보학, 전산학 등 기술 예측 연구에 활발히 활용되고 있다. 이것은 문헌의 잠재적인 확률을 설정한 다음 깁스 샘플링(Gibbs Sampling) 알고리즘을 활용하여 주어진 문헌의 주제와 단어에 대한 확률 값을 추정하는 비(非)지도 생성모델(Unsupervised Generative Model)로, 각 주제와 관련한 모든 단어를 쉽게 찾는 것이 장점이 이다(Blei, Ng and Jordan 2003).

보다 구체적으로 살펴보면, 관찰변수(Observed variable, 문헌에 포함된 단어)와 숨겨진 임의변수(Hidden random variables, 주제 구조) 간의 결합 확률 분포(Joint probability distribution)를 통해 관찰변수에 대한 조건부 분포(Conditional distribution)를 계산하여 전체 문헌 집합의 주제들과 각 문헌별 주제 비율 그리고 각 단어들이 각 주제에 포함될 확률 분석한다(Blei 2012). 또한 LDA는 확률적으로 문헌 내의 숨겨져 있는 주제 분포를 그래픽 모델로 표현하며, 각 문헌에 포함된 주제와 주제에 따른 단어의 배치를 각각 파라미터로 모델링한다.



<그림 2> LDA 그래픽 모델 (Blei 2012, p.81)

- α : 문헌별 주제 k의 Dirichlet prior weight, θ 값을 결정하는 파라미터
- D : 컬렉션내의 문헌집단
- N : 문헌내의 단어집단
- d : 개별문헌
- K : 주제의 수
- β_k : k번째 주제의 단어 분포의 확률
- θ_d : 문헌d에 대한 주제 비율
- $Z_{d,n}$: 문헌d의 n번째 단어에 대한 주제
- $W_{d,n}$: 문헌d에서 관찰되어진 n번째 단어
- η : 주제별 단어 w의 Dirichlet prior weight, β 값을 결정하는 파라미터

각 노드는 무작위 변수로 역할에 따라 라벨이 지정되며, 토픽모델링 결과의 품질은 α 와 η 값의 영향을 받으므로 최적의 값을 찾아내는 것이 중요하다. 즉, α 를 통해 θ 값을 구하고, η 를 통해 β 값을 구하여 θ 와 β 값으로 문헌을 모델링하는 것이다. $W_{d,n}$ 은 관찰변수로 문헌 내의 n번째 단어에 해당하며, θ_d 토픽비율, $Z_{d,n}$ 할당토픽, β_k 토픽의 단어 분포 확률, D 는

컬렉션 내의 문헌집단으로 Dirichlet 분포를 따르며, θ 값에 따라 문헌 집합 내에 존재하는 문헌 d 의 n 번째 단어에 대한 토픽이 결정된다(박자현, 송민 2013).

이와 같은 특징을 가진 토픽모델링을 적용한 최근 5년간 국내 문헌정보학 분야 연구사례는 다음과 같다.

이기현, 정효정, 송민(2015)은 1990년부터 2014년에 걸쳐 게재된 국외 문헌정보학 학술지들의 논문들을 대상으로 연구 주제 및 연구 방법에 따른 현 선도 학술지의 연구 누적 현황을 분석하였다. 연구 결과 상위 연구 주제로 경영정보시스템, 정보요구분석, 계량서지연구, 정보정책 등이 도출되었으며, 상위 연구방법으로는 모델링, 문헌연구, 협업분석, 웹 데이터 분석, 텍스트 마이닝 등이 선정되었다. 한편, 군집 분석을 통해 계량서지연구가 네트워크분석 방법을 적용한 계량서지 연구가 주를 이루었으며, 의료정보시스템, 이용자인터페이스 관련 연구에 텍스트마이닝 기법이 활용되고 있었다.

이수상(2016)은 텍스트의 주제 분석 방안으로 토픽모델링 분석방법을 이해하고자 독후감 텍스트의 주제 분석을 진행하였으며 토픽모델링의 결과를 네트워크 분석과 결합할 때 독후감의 주제 파악에 유용한 결과를 얻을 수 있다는 점을 밝혔다.

박준형, 오효정(2017)은 국내 문헌정보학 분야 학술지 4종과 기록관리학 분야 학술지 2종을 대상으로 국내 기록관리학의 연구동향을 분석하였다. 분석을 위해, LDA와 HDP(Hierarchical Dirichlet Process, 계층적 디리클레 프로세스) 알고리즘을 적용하여 비교분석 하였으며, LDA의 경우 빈도수에 따른 주요 키워드 선정이 이루어지는 반면, HDP는 세부 주제별로 미시적인 핵심 키워드를 도출하는데 적합하다는 분석결과를 제시하였다.

육지희, 송민(2018)은 자질집합의 크기와 종류 및 분류 알고리즘에 따른 성능 차이를 평가하고자 LDA 토픽과 딥러닝을 적용한 Doc2Vec 기법을 활용하였다. 이를 위해 자질집합의 적절한 크기, 문헌의 위치, 문헌의 종류 등 높은 성능을 나타내는 요소를 확인하였으며, 딥러닝을 활용하여 학습 횟수와 문맥 추론 정보의 유무에 따른 분류 성능을 비교하였다.

박중도(2019)는 국내 다문화 관련 분야의 연구동향을 확인하고자 국내 학술 문헌을 수집하여 LDA기반의 토픽 모델링을 수행하였다. 연구 토픽을 시기별로 추적하여 변화 양상을 관찰하였으며, 분석결과 Hot topic(+)으로는 ‘다문화사회통합’과 ‘학교다문화교육’이 Cold topic(-)으로는 ‘문화정체성과 민족주의’ 관련 토픽이 관찰되었다.

Ⅲ. 연구 설계

1. 연구 대상 및 자료 수집 방법

본 연구의 목적을 달성하고자 웹기반의 Scopus 데이터베이스를 통하여 전체 학문분야의

학술지를 대상으로 연구논문을 수집하였으며, 학술대회발표자료, 학위논문, 연구보고서, 단행본은 제외하였다. 연구기간에 제한을 두지 않고 해당 논문의 영문 초록을 포함한 서지사항을 수집하였다. Scopus는 22,790여 종의 저널을 수록하고 있는 데이터베이스로 과학, 기술, 사회과학 및 인문 등 모든 분야의 문헌을 포괄적으로 포함하고 있어 연구 동향 파악을 위한 자료수집에 용이하다. 초록은 기본적으로 연구의 배경, 목적, 연구결과, 결론 등 연구 과정에 따른 내용을 직접적이면서도 간결하게 제시하고 있어 해당 논문에서 전달하고자 하는 내용을 쉽게 파악할 수 있다는 점에서 연구 대상으로 설정하였다.

이에 따라 Scopus 데이터베이스를 활용하여 ‘Larchiveum’, ‘LAMs’, ‘GLAM’, ‘Library and Archives and Museum’, ‘Library and Archives and Museum and Gallery’라는 키워드로 연구 논문을 검색하였으며, 검색된 855건 중 초록이 포함되어 있지 않은 논문은 제외하여 총 637건을 분석대상으로 선정하였다.

해당 주제의 논문이 5편 이상 검색된 학술지는 <표 1>과 같다.

<표 1> 5편 이상의 논문이 검색된 학술지 리스트

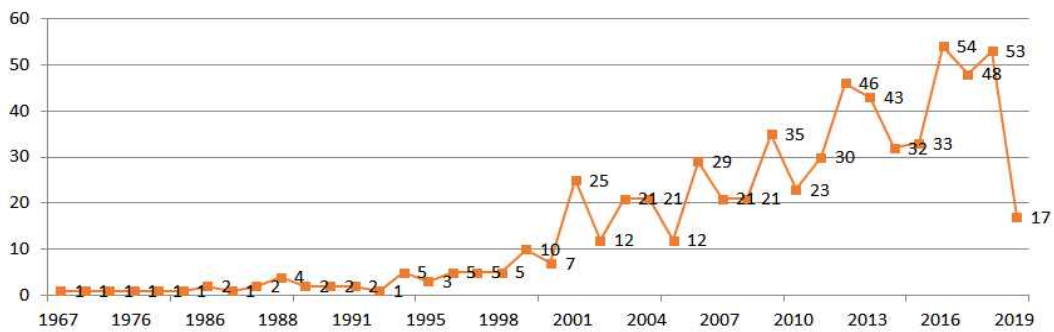
번호	저널명	건수	번호	저널명	건수
1	D-Lib Magazine	14	17	Australian Library Journal	6
2	Library Hi Tech	14	18	Cataloging and Classification Quarterly	6
3	First Monday	13	19	Journal of Archival Organization	6
4	IFLA Journal	11	20	Journal of Librarianship and Information Science	6
5	Profesional de la Informacion	11	21	Knowledge Organization	6
6	LIBER Quarterly	10	22	Program	6
7	Museum Management and Curatorship	10	23	Slavic and East European Information Resources	6
8	New Library World	10	24	VOEB-Mitteilungen	6
9	Vjesnik Bibliotekara Hrvatske	10	25	Archivaria	5
10	Journal of Documentation	9	26	Australian Academic and Research Libraries	5
11	Library Trends	9	27	Information Research	5
12	Archival Science	8	28	International Journal of Phytoremediation	5
13	Libri	8	29	JLIS.it	5
14	Restaurator	8	30	Journal of Library Administration	5
15	Zeitschrift für Bibliothekswesen und Bibliographie	8	31	OCLC Systems and Services	5
16	Electronic Library	7	32	Studies in Conservation	5

수집한 국외 논문의 연도별 추이는 <그림 3>과 같다. 그림에서 보는 바와 같이 도서관, 기록관, 박물관을 동시에 연구대상으로 한 최초의 학술논문은 1967년에 발표된 것으로 파악되었다. 이는 독일 뮌헨에 있는 기술대학교(Technical University)소속의 Fleckenstein-Gallo(1967)의 연구로 *Vistas in Astronomy*지⁹⁾ 제9권 제1호에 ‘Papyrology and sources in astronomical history’라는 제목으로 게재되었다. 이 논문은 수학적 또는 천문학적 내용에

9) 현재는 *New Astronomy Reviews* 라는 학술지명으로 변경됨.

대해 출판된 파피루스의 중요성을 강조하며, 파피루스에 전해지는 고대 과학에 관한 문헌자료, 천문학적 자료, 수학적 자료 등 관련 자료를 수집한다면, 세 기관을 통해 광범위하게 조사해야 함을 강조하고 있다.

2000년대에 접어들어 세 기관에 관련된 연구는 지속적인 증가세를 보이고 있다.¹⁰⁾



<그림 3> 연도별 논문 추이 (2019년 9월 현재 기준)

2. 분석 방법

수집한 637건의 국외 연구논문을 활용한 토픽모델링 분석 과정은 <그림 4>와 같은 절차를 따랐으며, 그 구체적인 방법은 다음과 같다.

첫째, 토픽모델링 분석 대상으로써 각 논문에서 해당 저자가 작성한 영문초록을 수집한다.

초록은 연구의 배경, 목적, 연구결과, 결론 등 연구 과정에 따른 내용을 직접적이면서도 간결하게 제시하고 있어 전달하고자 하는 내용을 쉽게 파악할 수 있다.

둘째, 수집된 영문초록으로 형태소분석을 통하여 다음과 같은 데이터전처리 과정을 통해 추출된 단어를 표준화한다.

- 1) 자료 수집을 위하여 사용한 키워드('Larchiveum', 'LAMs', 'GLAM', 'Library', 'Archive', 'Museum', 'Gallery')는 본 연구의 최상위 개념에 해당됨으로 분석 대상에서는 제외한다.
- 2) NetMiner V.4¹¹⁾의 형태소분석기를 활용하여 영문초록에서 명사형만을 추출한다.
- 3) 축약어와 전체 단어의 경우 키워드의 가독성을 고려하여 축약어를 대상으로 한다.
예를 들면, 'ALA'와 'American Library Association'의 경우 'ALA'를 채택한다.

10) 2019년 논문편수는 2019년 9월까지 출판된 논문 현황임.

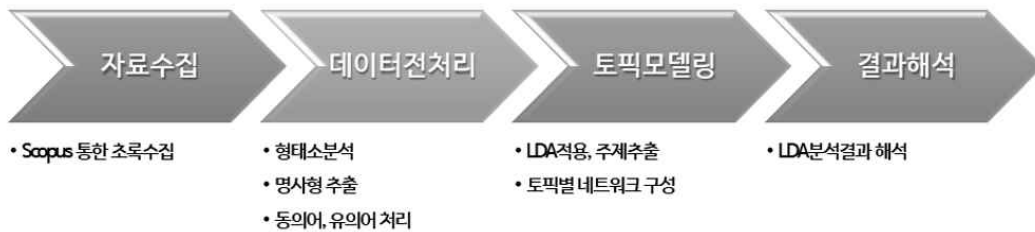
11) NetMiner는 2001년에 소개된 SNA(Social Network Analysis)을 위한 소프트웨어로 연결망 분석 모듈과 시각화 모듈이 하나의 패키지 프로그램에 통합되어 있어 분석이 용이하다. 뿐만 아니라, 대규모 네트워크의 데이터 처리가 가능하며 포괄적인 네트워크 분석과 방법론 적용이 가능한 것이 특징이다.

- 4) 특정지명이나 국가명 등의 고유명사는 분석대상에서 제외한다. 예를 들면 ‘France’, ‘Korea’ 등은 분석대상에서 제외한다.
- 5) 용어의 일반화를 위해 복수명사는 단수화하였다. 예를 들면 ‘Regulations’는 ‘Regulation’으로 변경한다.

셋째, 위의 내용을 적용하여 시소러스 사전을 구축한다. 예를 들면, ‘ALA’와 ‘American Library Association’은 ‘ALA’로, ‘Digital Library’와 ‘Electronic Library’와 같은 유의어의 경우는 ‘Digital Library’로, 그리고 ‘Education Program’과 ‘Curriculum’과 같이 동의어인 경우는 ‘Curriculum’으로 용어를 통제한다.

넷째, NetMiner V.4를 활용하여 전처리 과정을 통해 정제된 텍스트를 활용하여 LDA 기반의 토픽모델링을 적용하여 주제추출과 주제별 네트워크를 구성한다.

다섯째, LDA 기반의 토픽모델링 결과를 분석하고 해석한다.



<그림 4> 연구 절차

IV. 분석 결과

1. 빈도분석

분석 대상에 해당하는 연구논문 637편의 영문초록에서 총 5,218개의 단어를 추출하였으며, tf-idf에 따른 단어 출현 빈도는 <표 2>와 같다.

분석결과 ‘보존(Preservation)’과 ‘지식(Knowledge)’이 가장 높은 출현 빈도를 나타냈으며, ‘프로그램(Program)’, ‘조직(Organization)’, ‘도서(Book)’, ‘웹사이트(Website)’ 순으로 나타났다. 지식의 보존과 관련하여 Repanovici(2012)가 제안한 정보 및 지식의 보존을 위한 디지털 리포지터리 관리자로서 이수해야 할 정보 관리, 저작권 조건 및 정보 보안 등의 프로그램을 제시한 연구가 흥미롭다.

〈표 2〉 tf-idf에 따른 단어출현빈도 순위(30회 이상 출현 한정)

순위	키워드	빈도	순위	키워드	빈도	순위	키워드	빈도
1	Preservation	77	16	Copyright	51	31	Curriculum	38
2	Knowledge	77	17	Information Retrieval	50	32	Conservation	38
3	Program	68	18	Image	49	33	Professional	37
4	Organization	65	19	Society	46	34	Media	37
5	Book	61	20	Government	45	35	Archivist	35
6	Website	60	21	Manuscript	44	36	Mapping	34
7	University	59	22	Education	44	37	Communication	34
8	Culture	57	23	Cataloging	44	38	Benefit	32
9	Toolkit	56	24	Assessment	44	39	Database	32
10	Repository	56	25	Publication	43	40	Science	31
11	Funding	55	26	Photocopy	43	41	Element	31
12	Digitalization	55	27	Life cycle	43	42	Librarianship	30
13	Metadata	54	28	Staff	40	43	Convergence	30
14	Application	54	29	Collaboration	39			
15	Survey	52	30	Memory	38			

2. 토픽모델링 분석

637개의 논문에서 발췌한 초록에서 추출한 키워드 5,218개를 활용하여 LDA 알고리즘을 통한 토픽모델링 분석을 수행하였다. LDA 토픽 분석의 경우, 분석 결과의 품질 향상을 위해서는 문헌과 주제 간의 밀도를 나타내는 α 의 값을 최적의 값으로 설정하는 것이 중요하다. LDA 파라미터를 추정하기 위해서 깁스 샘플링(Gibbs Sampling)방식을 적용하였으며, 최종적으로 13개의 토픽을 추출하였다.

637건의 초록을 분석한 토픽모델링 결과는 <표 3>과 같다. 표에서 보는바와 같이, 총 13개의 주제영역이 추출되었으며, 각 주제영역별 상위 10개의 키워드를 함께 나열하였다. 추출된 키워드 군집을 대표할 주제 영역의 명칭은 연구자가 본 연구에서 수집된 관련 논문들의 초록을 참고하여 문맥을 분석한 후 할당하였다.

각 주제 영역별로 살펴보면, 주제 영역 1은 ‘Knowledge’, ‘Organization’, ‘Convergence’, ‘Staff’, ‘Consortium’ 등으로 구성되어 ‘기관 간의 협력’으로 설정하였다. Zappalà (1991)의 기록물 영구보존과 품질관리에 관한 고민을 시작으로, Knight(2001)가 박물관, 기록관, 도서관의 협력 증진을 위한 역할을 제시하고자 수행한 연구를 포함하여 총 73건(11.5%)의 연구논문이 이 영역에 포함되었다.

주제 영역 2는 총 53건(8.3%)의 연구논문이 분류되었으며 ‘Preservation’, ‘Conservation’, ‘Build’, ‘Air Quality’, ‘Material’ 등으로 구성되어 ‘정보자원 보존을 위한 환경 구축’으로 설정하였다. Fleckenstein-Gallo(1967)는 세 기관의 공통점으로 자료보존, 생산, 정보 접근

을 나열하였으며, Karaca(2015)는 유해환경 요인인 실내 공기질의 관리 기준을 확립하였다. 한편, Nasreen과 Bashir 그리고 Loan(2019)은 세계 193개국의 도서관, 자료관, 박물관, 교육기관 및 국제기구의 정보자원을 열람하는 목적으로 2009년에 수립한 세계 디지털도서관(WDL: World Digital Library)의 중요성을 강조하였다.

주제 영역 3은 총 48건(7.5%)의 연구논문이 분류되었으며 ‘Survey’, ‘Service’, ‘Volunteering’, ‘Government’ 등으로 구성되어 ‘정부차원의 조사연구를 통한 제도·정책 발굴’로 설정하였다. 회귀본, 기록물, 사진 자료 등 정보자원 보존 방법에 대한 방법 채택과 목표를 설정하고자 직원들을 대상으로 설문 조사를 실시(Genett 1987)하거나 서비스 확장과 부가가치 확대 방안을 모색하고자 정부 차원에서 설문조사를 실시한 사례를 확인하였다(Hewlett 2002).

주제 영역 4는 ‘Record’, ‘Documentation’, ‘Life cycle’, ‘Photocopy’, ‘Society’ 등으로 구성되어 ‘정보자원의 생애주기’로 설정하였다. 가치 있는 자료를 장기간 접근하도록 디지털 처리와 같은 방법으로 자료를 재생산할 필요성이 있음을 설명(Conway 2015)한 연구논문을 포함하여 총 36건(5.7%)의 연구논문이 포함되었다.

주제 영역 5는 ‘Exhibition’, ‘University’, ‘News’, ‘Photocopy’, ‘Material’ 등으로 구성되어 ‘정보자원의 전시’로 설정하였다. 표준화된 문화유산 자료실의 전략모델을 제시한 Yoo and Choi(2014)를 포함하여 총 31건(4.9%)의 연구논문이 포함되었다.

주제 영역 6은 ‘Repository’, ‘Preservation’, ‘Software’, ‘Accessibility’, ‘Publication’ 등으로 구성되어 ‘접근성 향상을 위한 리포지토리 구축’으로 설정하였다. 개방 아카이브의 메타데이터 수집 프로토콜인 OAI(Open archives initiative)와 P2P 기반의 디지털 아카이빙 시스템과 결합하였을 때 신뢰 가능한 디지털 도서관 구축이 가능하다는 연구결과(Hilario, Fernández, and Campo 2014)를 포함하여 총 50건(7.8%)의 연구논문이 포함되었다.

주제 영역 7은 ‘Data’, ‘Image’, ‘Record’, ‘Information Retrieval’, ‘Cataloging’ 등으로 구성되어 ‘정보자원의 검색’으로 설정하였다. 이용자의 원활한 정보검색을 구현하고자 더블링크어 요소가 적용된 메타데이터에 관한 연구를 진행한 Wykoff 등(2005)의 연구를 포함하여 총 55건(8.6%)의 연구논문이 포함되었다.

주제 영역 8은 ‘Curator’, ‘Competency’, ‘Science’, ‘Information Science’, ‘Knowledge’ 등으로 구성되어 ‘직원의 역량’로 설정하였다. 제공되는 정보의 유형이 점차 학제적일 것에 염두에 두고 큐레이터, 사서, 아키비스트들의 미래적 가치와 전문직의 변화를 시사한 Shipp(2016)을 포함하여 총 28건(4.4%)의 연구논문이 주제 8에 포함되었다.

주제 영역 9는 ‘Service’, ‘Documentation’, ‘Book’, ‘Culture’, ‘Cataloging’ 등으로 구성되어 ‘서비스 제공을 위한 정보자원 조직’으로 설정하였으며, 총 26건(4.1%)의 연구논문이 포함되었다. Matthews(2016)는 전 세계의 도서관, 기록관, 박물관은 OCLC의 목록을 기반으로 하며, OCLC는 ‘World Share Metadata Services’와 ‘World Share Interlibrary Loan’ 등 자원공유를 비롯한 폭넓은 서비스를 제공한다는 내용을 서술하였다.

주제 영역 10은 ‘Metadata’, ‘Data’, ‘Mapping’, ‘Interoperability’, ‘Terminology’ 등으로 구성되어 ‘상호운용을 위한 메타데이터 구축’으로 설정하였으며, 총 77건(12.1%)의 연구논문이 포함되었다. 메타데이터의 상호운용성을 고려한 RDF 기반의 접근방식과(Chen 2015), 메타데이터 표준도구, 기술적 상호운용성을 정보제공기관 간의 성공적인 협업을 위한 요소로 서술한 Bailey and Urban(2004) 등의 연구가 있다.

주제 영역 11은 ‘Website’, ‘Digitalization’, ‘Accessibility’, ‘User’, ‘Technology’ 등으로 구성되어 ‘웹 접근성 향상을 위한 정보자원 디지털화’으로 설정하였으며, 총 54건(8.5%)의 연구논문이 포함되었다. 정보자원의 디지털화는 보존, 조직, 정보 제공 등 정보제공기관으로서 다양한 기회를 마련하는 긍정적인 측면을 확인할 수 있었다(Çakmak and Yilmaz 2017).

주제 영역 12는 ‘Material’, ‘Manuscript’, ‘Documentation’, ‘Book’, ‘Photocopy’ 등으로 구성되어 ‘정보자원의 특성’으로 설정하였다. 이는 박물관, 도서관, 기록관 등의 정보제공기관에서 고유한 정보자원의 특성을 분석하는 것이라 해석되며, 총 71건(11.1%)의 연구논문이 포함되었다.

주제 영역 13은 ‘Curriculum’, ‘Student’, ‘Program’, ‘Education’, ‘LIS’ 등으로 구성되어 ‘전문가 양성을 위한 교육’으로 설정하였다. 기관 전문가들을 대상으로 저작권 리터러시와 학술교육을 진행하며 지속적인 교육 프로그램을 이어나가고자 저작권에 관한 학습 콘텐츠에 대한 조사를 진행한 연구(Todorova et al. 2017), 정보제공기관의 모범적인 융합사례와 미래 전문가들이 요구하는 학술 교육과정에 관한 논의(Tammaro 2014)등을 포함하여 총 35건(5.5%)의 연구논문이 포함되었다.

〈표 3〉 토픽모델링 결과

주제1 (11.5%)	기관 간의 협력
키워드	Knowledge(0.071), Organization(0.052), Convergence(0.045), Staff(0.04), Consortium(0.04), Collaboration(0.036), Technology(0.036), User(0.032), Memory(0.028), Librarianship(0.028)
주제2 (8.3%)	정보자원 보존을 위한 환경 구축
키워드	Preservation(0.101), Conservation(0.084), Build(0.045), Air Quality(0.045), Material(0.033), Acetic Acid(0.03), Technology(0.029), Pollutant(0.025), Assessment(0.025), Book(0.016)
주제3 (7.5%)	정부차원의 연구를 통한 제도정책 발굴
키워드	Survey(0.082), Service(0.081), Government(0.078), Data(0.052), Assessment(0.052), Organism(0.05), Volunteering(0.05), Digitalization(0.038), Accessibility(0.025), Copyright(0.02)
주제4 (5.7%)	정보자원의 생애주기
키워드	Record(0.117), Documentation(0.085), Life cycle(0.033), Photocopy(0.025), Society(0.025), Terminology(0.022), Foundation(0.021), Manuscript(0.02), Culture(0.016), Facility(0.015)

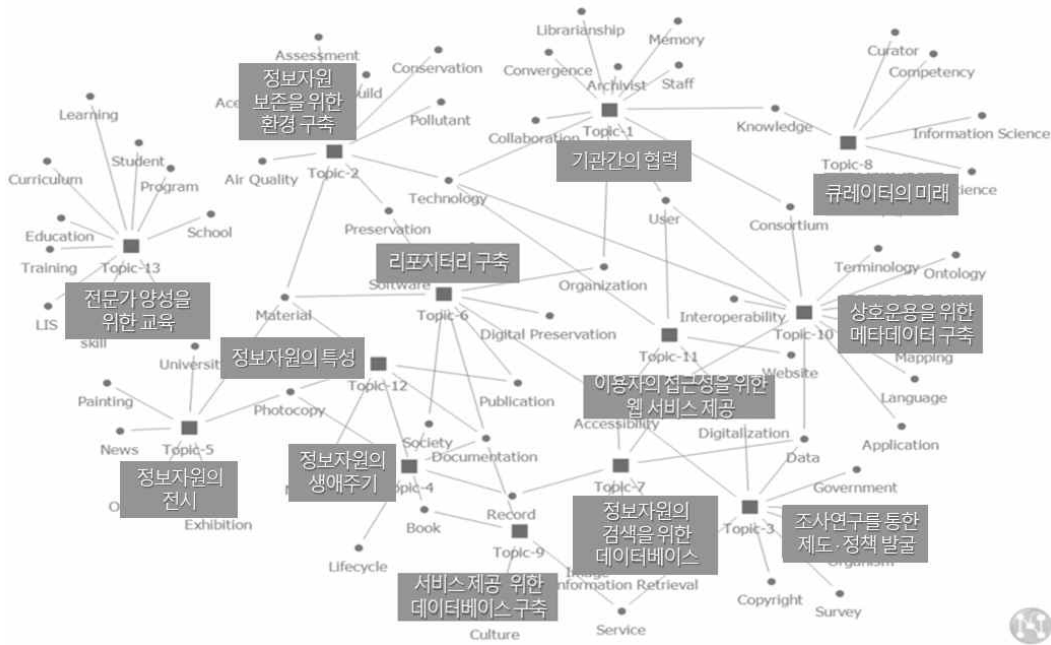
14 한국도서관·정보학회지(제50권 제4호)

주제5 (4.9%)	정보자원의 전시
키워드	Exhibition(0.039), University(0.037), News(0.036), Open Access(0.029), Photocopy(0.028), Material(0.025), Painting(0.025), Lecturer(0.02), Repository(0.02), Artist(0.02)
주제6 (7.8%)	리포지터리 구축
키워드	Repository(0.069), Preservation(0.049), Software(0.043), Accessibility(0.035), Publication(0.032), Organization(0.03), Material(0.029), Society(0.028), Documentation(0.025), Digital Preservation(0.025)
주제7 (8.6%)	정보자원의 검색
키워드	Data(0.141), Image(0.122), Record(0.052), Information Retrieval(0.038), Cataloging(0.031), Accessibility(0.031), Classification(0.027), Digital Image(0.023), Toolkit(0.022), Database(0.021)
주제8 (4.4%)	직원의 역량
키워드	Curator(0.042), Competency(0.037), Science(0.037), Information Science(0.034), Knowledge(0.032), Painting(0.021), Technology(0.02), Ownership(0.018), Bibliography(0.016), Tag(0.015)
주제9 (4.1%)	서비스 제공을 위한 정보자원 조직
키워드	Service(0.099), Documentation(0.061), Book(0.041), Culture(0.036), Cataloging(0.021), Memory(0.021), Utility(0.019), Exploitation(0.019), Database(0.017), Assessment(0.017)
주제10 (12.1%)	상호운용을 위한 메타데이터 구축
키워드	Metadata(0.104), Data(0.082), Mapping(0.045), Interoperability(0.038), Terminology(0.037), User(0.035), Consortium(0.033), Ontology(0.03), Technology(0.03), Language(0.029)
주제11 (8.5%)	웹 접근성 향상을 위한 정보자원 디지털화
키워드	Website(0.091), Digitalization(0.084), Accessibility(0.071), User(0.061), Technology(0.053), Memory(0.018), University(0.018), Service(0.018), Social Media(0.018), Information Retrieval(0.018)
주제12 (11.1%)	정보자원의 특성
키워드	Material(0.139), Manuscript(0.063), Documentation(0.046), Book(0.042), Photocopy(0.031), Publication(0.027), Media(0.022), Inheritance(0.022), Cooperation(0.019), Science(0.019)
주제13 (5.5%)	전문가 양성을 위한 교육
키워드	Curriculum(0.069), Student(0.068), Program(0.066), Education(0.066), LIS(0.044), Training(0.043), Learning(0.035), University(0.033), Skill(0.028), School(0.027)

분석결과로 제시된 13개의 주제 간의 연결성은 <그림 5>와 같이 네트워크로 표현하였다. 주제 영역 6인 리포지터리 구축(Repository Construction)을 중심으로 주제 영역1(Organization), 주제 영역 2(Material), 주제 영역 3(Accessibility), 주제 영역 4(Society), 주제 영역 5(Material), 주제 영역 7(Accessibility), 주제 영역 9(Documentation), 주제 영역 12(Publication) 등이 서로 밀접한 관련성을 가진 것으로 나타났다. 이는 기관 간의 협력을 위한 환경구성과 방안모색 그리고 정보서비스 제공을 위한 환경조성에 관한 연구에서 주로 리포지터리에 관한 내용이 언급되었다고 해석할 수 있다.

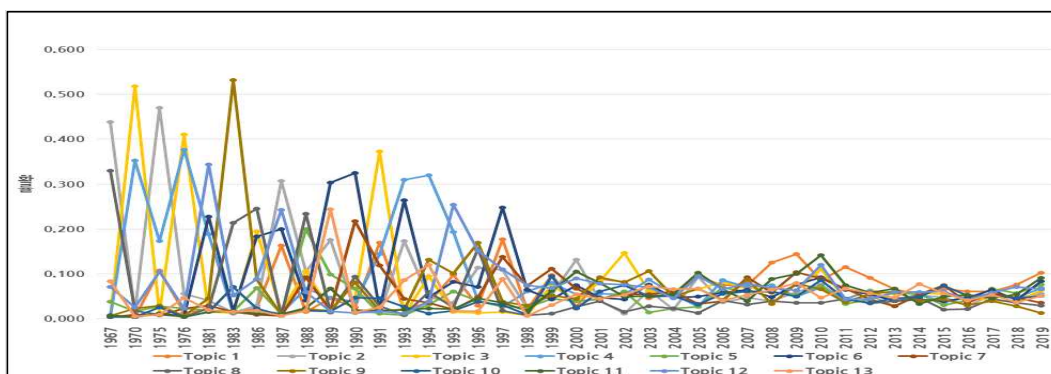
분석결과를 종합하면, 도서관, 기록관, 박물관 등의 정보제공기관에서는 유사서비스를 제

공하는 기관 간의 협력, 문화유산 또는 기록물 보존을 위한 방안 모색, 상호운용 환경구성에 대한 논의가 이어왔으며 정보서비스를 위한 기록물전시, 정보자원의 효율적인 정보검색, 정부차원의 정책 발굴, 그리고 전문가 역량과 미래 등에 대한 다양한 고민을 엿볼 수 있다.



<그림 5> 주제 영역별 네트워크

분석결과로 제시된 13개의 주제의 연도별 변화 추이를 <그림 6>과 같이 나타내었다. 1998년 이전의 연구는 주제 영역3의 조사연구를 통한 제도·정책 발굴(1970년, 1991년)과 주제 영역 4 정보자원의 생애주기(1970년, 1976년, 1995년), 주제 영역 9의 서비스 제공을



<그림 6> 연구 주제의 연도별 변화 추이

위한 정보자원 조직(1983년) 등 특정 주제 분야의 연구가 두드러지게 나타났으나, 1998년부터는 특정주제 영역에 한정되지 않고 보다 다양화되었음을 알 수 있다.

V. 결론

본 연구의 목적은 도서관, 기록관, 박물관과 관련된 해외 연구의 특성과 연구의 주제 그리고 그 연구의 동향을 분석하는 것이다. 해당 분야의 연구 진행 정도나 그 경향을 파악하는 것은 지적 구조를 이해하는 것을 돕고 앞으로의 연구 방향을 예측하는데 큰 의미가 있다고 할 수 있다.

본 연구에서 채택한 연구방법으로 키워드의 유사어 통일과 불용어 제거 등 정제작업을 수행하는 과정에서 연구자의 주관이 다소 개입되어 최종 분석 대상 키워드 추출이 연구자에 따라 다소 차이가 날 수 있다는 점, 학술논문 이외에 학술대회발표자료, 학위논문, 연구보고서, 단행본 등을 포함하지 못한 점 등과 같이 연구의 제한점이 있어 연구결과를 일반화하기에는 다소 한계가 있으나, 본 연구의 분석결과를 요약하면 다음과 같다.

첫째, tf-idf 가중치에 따른 단어출현 빈도를 분석한 결과, ‘보존(Preservation)’, ‘지식(Knowledge)’, ‘프로그램(Program)’, ‘기관(Organization)’, ‘도서(Book)’ 순으로 나타났다. 이에 근거하여 Repanovici(2012)는 정보 및 지식의 보존을 위한 디지털 리포지터리 관리자로서 이수해야 할 정보 관리, 저작권 조건 및 정보 보안 등의 교육과정을 제시하였다.

둘째, 주제 영역에서는 LDA 알고리즘을 통한 토픽모델링 분석을 진행하여 13개의 주제 영역을 추출하였다. 구체적으로, ‘기관 간의 협력’, ‘정보자원 보존을 위한 환경 구성’, ‘정부차원의 조사연구를 통한 제도·정책 발굴’, ‘정보자원의 생애주기’, ‘정보자원의 전시’, ‘리포지터리 구축’, ‘정보자원의 검색’, ‘직원의 역량’, ‘서비스 제공을 위한 정보자원 조직’, ‘상호운용을 위한 메타데이터 구축’, ‘웹 접근성 향상을 위한 정보자원 디지털화’, ‘정보자원의 특성’, ‘전문가 양성을 위한 교육’ 등이다.

셋째, 토픽모델링 결과를 바탕으로 네트워크를 표현하였을 때, ‘리포지터리 구축’이라는 주제 영역을 중심으로 각 주제 영역이 밀접한 관련성을 가진 것으로 나타났다. 즉, 기관 간의 협력 방안모색과, 환경구성에 관한 고민 그리고 정보서비스 제공을 위한 환경을 조성하는 데 리포지터리 구축이 함께 다루어진다고 해석할 수 있다.

넷째, 토픽모델링 분석으로 도출된 13개의 주제 분야의 연도별 변화를 살펴보면, 1998년 이전의 연구는 특정 연구주제가 두드러지게 나타났던 것에 반해, 그 이후 연구는 연구자들이 다양한 연구주제를 다룬 것으로 해석된다.

지식정보제공이라는 공동의 임무를 수행하는 도서관, 기록관, 박물관은 환경변화에 따른 이용자의 요구를 수용하고 합리적인 운영을 위하여 협력단위의 방안모색이 요구된다. 이 연

구를 바탕으로 차후 연구로는 해당 분야의 상승세를 이루는 주제 분야(Hot Topic, +)와 하강세를 이루는 주제분야(Cold Topic, -)분석 등의 심층적인 연구동향 분석과 연구 결과의 일반화를 보장하기 위한 폭넓은 자료 수집으로 연구를 수행하는 것으로 들 수 있다.

참고문헌

- 국립중앙도서관. <<https://nl.go.kr>> [인용. 2019. 09. 17].
- 국립중앙도서관 디지털컬렉션. <<https://ncollection.nl.go.kr>> [인용. 2019. 09. 17].
- 김남규 외. 2017. 텍스트 분석 기술 및 활용 동향. 『한국통신학회논문지』, 42(2): 471-492.
- 박자현, 송민. 2013. 토픽모델링을 활용한 국내 문헌정보학 연구동향 분석. 『정보관리학회지』, 30(1): 7-32.
- 박종도. 2019. 토픽 모델링을 활용한 다문화 연구의 이슈 추적 연구. 『한국문헌정보학회지』, 53(3): 273-289.
- 박준형, 오효정. 2017. 국내 기록관리학 연구동향 분석을 위한 토픽모델링 기법 비교: LDA와 HDP를 중심으로. 『한국도서관정보학회지』, 48(4): 235-258.
- 육지희, 송민. 2018. 토픽모델링과 딥 러닝을 활용한 생의학 문헌 자동 분류 기법 연구. 『정보관리학회지』, 35(2): 63-88.
- 이기현, 정효정, 송민. 2015. 문헌정보학 분야 핵심 학술지들의 가중 주제-방법 네트워크 분석. 『한국문헌정보학회지』, 49(3): 457-488.
- 이미경. 2014. 라키비움 도입과 실현. 『국회도서관』, 51(5): 14-23.
- 이수상. 2016. 독후감 텍스트의 토픽모델링 적용에 관한 탐색적 연구. 『한국도서관정보학회지』, 47(4): 1-18.
- 최영실, 이해영. 2012. 기록관, 도서관, 박물관의 기능을 융합한 라키비움의 기능 계획. 『한국비블리아학회지』, 23(4): 457-477.
- 한정원. 2019. 국내 공공도서관의 공간디자인 관련 연구동향 분석: 국내 학술지 게재 논문을 중심으로. 『한국실내디자인학회 논문집』, 28(1): 170-177.
- Bailey-Hainer, B. and Urban, R. 2004. "The Colorado digitization program: a collaboration success story." *Library Hi Tech*, 22(3): 254-262.
- Blei, D. 2012. "Probabilistic topic models." *Communications of the ACM*, 55(4): 77-84.
- Blei, D. M., Ng, A. Y. and Jordan, M. I. 2003. "Latent dirichlet allocation." *Journal of machine Learning research*, 3(Jan): 993-1022.
- Brabazon, T. 2009. "Brand Wellington: When city imaging is GLAM'ed: A personal view." *Place Branding and Public Diplomacy*, 5(4): 260-275.

- Çakmak, T. and Yilmaz, B. 2017. "Digitization and digital preservation in memory institutions: Analysis of the practices in Turkey." *Bilgi Dnyasi*, 18(1): 49–91.
- Conway, P. 2015. "Digital transformations and the archival nature of surrogates." *Archival Science*, 15(1): 51–69.
- Chen, Y. N. 2015. "A RDF-based approach to metadata crosswalk for semantic interoperability at the data element level." *Library Hi Tech*, 33(2): 175–194.
- Europeana Collections Homepage. <<https://www.europeana.eu/portal/en>> [Cited. 2019. 09. 17].
- Fleckenstein-Gallo, J. O. 1967. "Papyrology and sources in astronomical history." *Vistas in Astronomy*, 9: 151–155.
- Genett, M. E. 1987. "Conservation of research library collections at the American Museum of Natural History." *Science and Technology Libraries*, 7(3): 15–28.
- Heran, M. 1908. "Lloyd Library and Museum launches new initiative: Historical Research Center for the Natural Health Movement." *Watermark (Archivists and Librarians in the History of the Health Sciences)*, 31(1): 9–11.
- Hewlett, S. 2002. "Volunteering in libraries, museums and archives." *Cultural Trends*, 12(46): 39–66.
- Hilario, A. B. R., Fernández, T. F. and Campo, D. M. 2014. "From Bibliographic Records to Data: Changes in the Library Environment with the Application of Linked Open Data Technologies." *Information Resources Management Journal*, 27(3): 28–41.
- Karaca F. 2015. "An AHP-based indoor Air Pollution Risk Index Method for cultural heritage collections." *Journal of Cultural Heritage*, 16(3): 352–360.
- Knight, K. 2001. "The Strategic Role of Resource Encouraging Partnerships." *LIBER Quarterly*, 11(4): 444–453.
- Library and Archives Canada Homepage. <<http://www.bac-lac.gc.ca/eng/about-us/Pages/our-mandate.aspx>> [Cited. 2019. 09. 11].
- Marty, P. F. 2014. "Digital Convergence and the Information Profession in Cultural Heritage Organizations: Reconciling Internal and External Demands." *Library Trends; Baltimore*, 62(3): 613–627.
- Matthews, J. R. 2016. "An Environmental Scan of OCLC Alternatives: A Management Perspective." *Public Library Quarterly*, 35(3): 175–187.
- Nasreen, N., Bashir, B. and Loan, F. A. 2019. "World Digital Library: An Analysis of Collection." *Library Philosophy and Practice*, 1–12.

- Papadimitriou, C., Raghavan, P., Tamaki, H. and Vempala, S. 1998. "Latent Semantic Indexing: A probabilistic analysis." *Proceedings of ACM PODS '98: Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, 159-168.
- Repanovici, A. 2012. "Professional profile of digital repository manager." *Library Hi Tech News*, 29(10): 13-20.
- Shipp, J. N. 2016. "Do I really need specialist qualifications to work as a professional in a gallery, library, archive or museum?." *The Australian Library Journal*, 65(4): 280-287.
- Tammaro, A. M. 2014. "The convergence of libraries, archives and museums: IFLA initiatives." *AIB Studi*, 54(1): 115-120.
- Todorova, T. Y. et al. 2017. "Information professionals and copyright literacy: a multinational study." *Library management*, 38(6/7): 323-344.
- Yarrow, A., Clubb, B. and Draper, J. L. 2008. "Public Libraries, Archives and Museums: Trends in Collaboration and Cooperation." *IFLA Professional Reports*, 108.
- Yoo, D. H. and Choi, A. R. 2014. "A study of larchiveum data model for the design of digital heritage museum." *International Journal of Software Engineering and its Applications*, 8(10): 83-94.
- World Digital Library Homepage. <<https://www.wdl.org/en/>> [Cited. 2019. 09. 17].
- Wykoff, L., Mercier, L., Bond, T. and Cornish, A. 2005. "The Columbia River Basin Ethnic History Archive: a tri-state online history database and learning center." *Library Hi Tech*, 23(2): 252-264.
- Zappalà, A. 1991. "Problems in standardizing the quality of paper for permanent records." *Restaurator*, 12(3): 137-146.

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Choi, Youngsil, Rieh, Hae-young. 2012. "Functional Planning of Larchiveum that Integrates the Functions of Archives, Libraries and Museums." *Journal of the Korean Biblia Society for Library and Information Science*, 23(4): 457-477.
- Han, Jeong-Won. 2019. "A Study on the Research Trends for the Space Design of Public Libraries in Korea - Focused on the Research Articles Published in Korea." *Journal of the Korean Institute of Interior Design*, 28(1): 170-177.

- National Library of Korea. <<https://nl.go.kr>> [Cited. 2019. 09. 17].
- National Library of Korea Digital Collection. <<https://ncollection.nl.go.kr>> [Cited. 2019. 09. 17].
- Lee, Keehoen, Jung, Hyojung, Song, Min. 2015. "Weighted Subject – Method Network Analysis of Library and Information Science Studies." *Journal of the Korean Society for Library and Information Science*, 49(3): 457–488.
- Lee, Mi–Kyung. 2014. "Introduction and Realization of Larchiveum." *National Assembly Library of Korea*, 51(5): 14–23.
- Lee, Soo–Sang. 2016. "A Study on the Application of Topic Modeling for the Book Report Text." *Journal of Korean Library and Information Science Society*, 47(4): 1–18.
- Kim, Namgyu et al. 2017. "Investigations on Techniques and Applications of Text Analytics." *The Journal of Korean Institute of Communications and Information Sciences*, 42(2): 471–492.
- 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.
- Park, Jong Do. 2019. "A Study on Issue Tracking on Multi–cultural Studies Using Topic Modeling." *Journal of the Korean Society for Library and Information Science*, 53(3): 273–289.
- Park, JunHyeong, Oh, Hyo–Jung. 2017. "Comparison of Topic Modeling Methods for Analyzing Research Trends of Archives Management in Korea : focused on LDA and HDP." *Journal of Korean Library and Information Science Society*, 48(4): 235–258.
- Yuk, JeeHee, Song, Min. 2018. "A Study of Research on Methods of Automated Biomedical Document Classification using Topic Modeling and Deep Learning." *Journal of the Korean Society for Information Management*, 35(2): 63–88.