

학술 문헌 내 정의문을 통해 살펴본 디지털 큐레이션 개념 변화 분석*

Analysis of Changes in the Concept of Digital Curation through Definitions in Academic Literature

김현수 (Hyunsoo Kim)*
오효정 (Hyo-Jung Oh)***

초 록

최근 디지털 전환 시대를 맞아 학계뿐 아니라 다양한 영역에서 디지털 큐레이션에 대한 논의가 활발히 진행되고 있다. 본 연구는 시간 흐름에 따른 디지털 큐레이션의 개념 변화를 분석하는 것을 주 목적으로, 특히 학술 문헌에서 저자가 직접 그 의미를 설명하거나 인용한 정의문을 중심으로 개념 구조를 파악하였다. 이를 위해 국내에서 디지털 큐레이션이란 용어가 처음 언급된 2009년부터 최근 2023년까지 출판된 학술 논문을 수집하고 관련 개념을 정리한 문장을 선별, 정의문의 용례 분석을 수행하였다. DMR 토픽 모델링과 워드 네트워크 분석 기법을 통해 주요 주제어들 간의 관계와 시간의 흐름에 따른 중요도 변화를 살펴보았으며, 주요 토픽을 중심으로 디지털 큐레이션의 개념도를 그려보았다. 그 결과 디지털 큐레이션의 개념은 '데이터 보존', '전통적인 큐레이터 역할', '상품 추천 큐레이션' 주제가 중심축을 이루고 있고, 연구자마다 디지털 큐레이션을 활용하고자 하는 방향에 따라 '콘텐츠 유통 및 분류', '정보 이용', '큐레이션 모델' 등으로 확장하여 정의하고 있었다. 본 연구는 디지털 큐레이션의 개념을 학술 연구자의 관점이 투영된 정의문에 기반해 분석했다는 데에 의의가 있으며, 연구 진행 시점에 따른 토픽의 증감 추세를 통해 연구자들이 주안점을 두는 개념의 변화를 구체적으로 파악하였다는데 의미가 있다.

ABSTRACT

In the era of digital transformation, discussions about digital curation have become increasingly active not only in academia but also in various fields. The primary purpose of this study is to analyze the conceptual changes in digital curation over time, particularly by examining the definition statements related to digital curation as described in academic literature. To achieve this, academic research papers from 2009, when the term "digital curation" was first mentioned, to 2023 were collected, and definition statements that explained relevant concepts were extracted. Basic statistical analyses were conducted. Using DMR topic modeling and word networks, the relationships among keywords and the changes in their importance over time were examined, and a conceptual map of digital curation was made focusing on the main topics. The results revealed that the concept of digital curation is primarily centered around the themes of "data preservation," "traditional curator roles," and "product recommendation curation." Depending on the researchers' intentions for utilizing digital curation, the concept was expanded to include topics such as "content distribution and classification," "information usage," and "curation models." This study is significant in that it analyzed the concept of digital curation through definition statements reflecting the perspectives of researchers. Additionally, the study holds value in explicitly identifying changes in the concepts that researchers emphasize over time through the trends in topic prevalence.

키워드: 디지털 큐레이션, 개념 분석, DMR 토픽 모델링, 워드 네트워크 분석, 용례 분석
digital curation, concept analysis, DMR topic modeling, word network analysis, sentence usage analysis

* 본 논문은 2024년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음. 본 논문은 2024년도 한국연구재단 연구비 지원에 의한 결과의 일부임(과제번호: NRF-2021R1I1A3047435)

** 전북대학교 일반대학원 기록관리학과 석사과정(itishskim@gmail.com) (제1저자)

*** 전북대학교 문헌정보학과 교수, 문화융복합아카이빙연구소 공동연구원(ohj@jbnu.ac.kr) (교신저자)

■ 논문접수일자 : 2024년 8월 16일 ■ 최초심사일자 : 2024년 9월 6일 ■ 게재확정일자 : 2024년 9월 12일
■ 정보관리학회지, 41(3), 269-288, 2024. <http://dx.doi.org/10.3743/KOSIM.2024.41.3.269>

※ Copyright © 2024 Korean Society for Information Management
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0
(<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that
the article is properly cited. the use is non-commercial and no modifications or adaptations are made.

1. 서 론

1.1 연구의 배경 및 목적

현재 우리 사회는 생활의 많은 부분을 디지털 기술로 전환하는 디지털 전환(Digital transformation) 시대를 맞이하고 있다. 이는 기존 추구해 온 소극적 의미의 디지털화와 달리 디지털을 이용해 기업의 전략, 프로세스, 비즈니스 모델, 문화, 조직, 커뮤니케이션, 시스템을 근본적으로 변화시키는 것을 말한다(장효석, 2022). 이에 따라 다양한 분야의 업무가 온라인을 통해 처리되면서 데이터의 양이 급격히 증가하고 있으며, 2025년에는 전 세계 연간 데이터 생성량이 175ZB에 달할 것이라는 예측도 나오고 있다(International Data Corporation, 2017). 이처럼 폭발적으로 증가하는 데이터의 양은 이미 기존 체계의 보존 능력을 넘어섰으며(조민지, 2020), 데이터의 유형도 점점 다양화되고 있다. 이에 더 나아가 데이터를 적극적으로 활용하고 이를 통해 새로운 가치를 창출하려는 움직임도 나타나고 있다.

디지털 정보는 본질적으로 훼손이나 손실이 쉬운 물리적 특성이 있어 체계적으로 정보를 조직하고 관리, 보존할 필요가 있다. 이를 위해 제안되고 있는 방법이 ‘디지털 큐레이션(Digital Curation)’이다. 본래 ‘큐레이션(Curation)’이라는 개념은 주로 미술관과 박물관에서 작품을 선별하고 수집하여 전시하는 데 사용되었으나, 전자상거래가 활발해지며 마케팅 분야에서도 해당 용어를 사용하는 등, 정보관리와 경영학을 비롯해 다양한 영역에서 큐레이션에 대한 논의가 활발해지고 있다. 그러나 이처럼 디지

털 큐레이션이라는 단어는 이미 현장 실무와 이론에서 동시에 널리 사용되고 있음에도 불구하고(Higgins, 2011), 학문적으로는 디지털 큐레이션에 대한 필요성이 강조된 지 불과 십여년밖에 되지 않은 신생 연구 영역이기 때문에 연구자들은 개념의 혼동을 겪고 있다. 특히 학술 연구에서는 정확하고 명확한 단어를 사용하여 올바른 개념을 정립하는 것이 중요하기 때문에 각 용어에 관한 개념 연구가 필요하다(김현수, 정성모, 2023).

이에 본 연구는 시간의 흐름에 따른 디지털 큐레이션의 개념 변화를 살펴보고자 한다. 특히 학술 문헌에 나타난 정의문을 바탕으로 개념 분석을 시도하였는데, 정의문은 연구자가 주어진 표제 어휘를 어떻게 이해하고 어떤 특성을 중점적으로 설명하고자 했는지에 대한 관점이 투영된 문장으로, 해당 개념을 대표하는 주요 핵심어들로 구성되어 있기 때문이다. 따라서 디지털 큐레이션과 관련된 학술적 정의용례를 모두 수집하여 분석한다면, 연구가 진행됨에 따른 연구자들의 해석이 어떻게 변화했는지, 그에 따른 개념 구조는 어떻게 변화했는지를 살펴볼 수 있다. 이를 위해 본 연구에서는 DMR(Dirichlet Multinomial Regression) 토픽 모델링 기법을 차용하였으며, 세부 학문 분야별 주요 주제에 대한 워드 네트워크도 함께 분석하여 디지털 큐레이션의 학술적 개념 관계와 확장 추이를 파악하고자 한다.

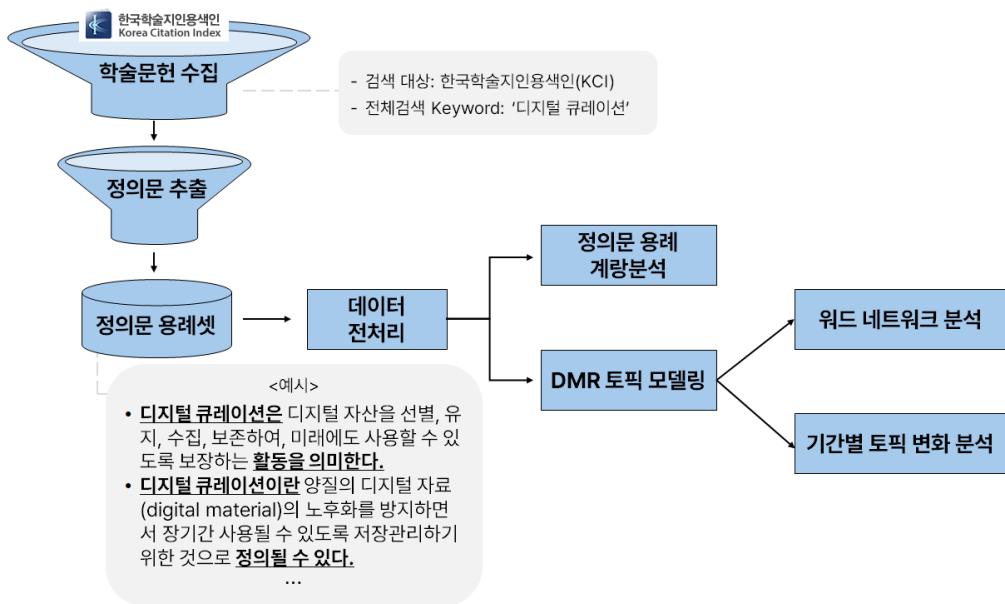
1.2 연구 방법

본 연구의 목적은 학술 문헌에 나타난 정의문의 용례를 기반으로 시간의 흐름에 따른 디

지털 큐레이션의 개념 변화를 분석하는 것이다. 이를 위해 〈그림 1〉과 같은 방법으로 연구를 진행하였다.

먼저, 정의문 용례셋을 구축하기 위한 자동화 방안을 고안하였다. 기초 데이터 수집을 위해서 한국연구재단에서 운영 중인 대표적인 국내 학술데이터베이스 한국학술지인용색인(KCI) 사이트를 활용, '디지털 큐레이션'이란 용어를 포함하는 모든 논문을 검색하였다. 이후 수집된 학술 논문의 원문에 기술된 '디지털 큐레이션'이나 그와 유사한 개념을 설명한 정의문 장의 패턴을 인식하여 후보 정의문들을 자동으로 추출한 후, 검증을 통해 분석 대상 용례집합을 선정하였다. 전처리 작업을 거친 용례셋을 대상으로 시계열 추이, 주요 학문분야 추이 분석 등 기본적인 통계 분석을 수행하였다.

다음으로는 디지털 큐레이션의 주요 개념 구조와 변화를 파악하기 위해 DMR 토픽 모델링을 차용, 시간의 흐름에 따른 추이 분석을 위해 논문이 발간된 시기를 주요 변수로 설정하여 구간별로 나누어 분석하였다. 또한 토픽 모델링 결과로 도출된 토픽별 주요 키워드를 워드 네트워크로 시각화하여 개념 구조를 파악하고, 시간의 흐름에 따라 각 토픽의 중요도가 어떻게 변화하는지 살펴봄으로써 디지털 큐레이션 정의문의 개념 변화를 분석하였다. 마지막으로 분석 결과를 종합해 디지털 큐레이션의 개념도를 도식화하였다. 제안한 방법론은 비단 본 연구 주제인 '디지털 큐레이션'에 국한되는 것이 아닌 다른 개념 분석에서도 적용할 수 있으며, 가능한 연구자의 주관이 반영되는 과정을 배제하는 데 주안점을 두었다.



〈그림 1〉 연구 방법

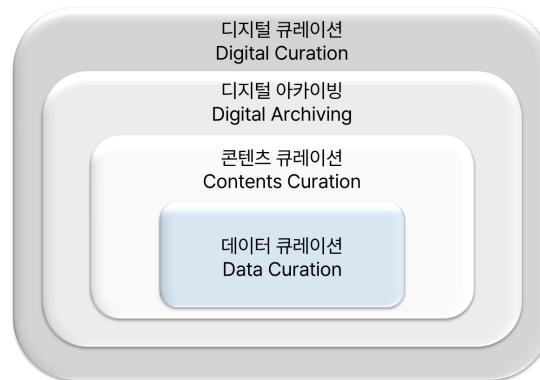
2. 이론적 배경

2.1 디지털 큐레이션

본 연구의 분석 대상인 ‘디지털 큐레이션’이라는 용어는 2001년 런던의 한 세미나에서 처음 발표되었으며, 이때 주요 개념은 “디지털 자원을 수집, 영구보존, 제공하는 아카이빙 활동”이자 “데이터가 학술적인 관심 대상이 되는 생명주기 동안 이러한 데이터를 적극적으로 관리하고 평가하는 것”(김판준, 2015)으로 소개되었다. 이후 2009년 안영희, 박옥화(2009)의 연구를 통해 “디지털 큐레이션은 현재와 미래에 참고할 수 있도록 디지털 자원의 장기보존소를 설립하고 개발하는 과정”으로 정의되며 처음 국내 학술 문헌에서 언급되었다. 그 뒤로 소수의 연구자에 의해 다양한 분야에서 저마다 조금씩 해석으로 정의, 인용되며 산발적으로 연구가 진행되다 2015년, 문헌정보학을 중심으로 디지털 큐레이션의 연구 동향을 살펴본 연구 결과가 발표되었다(김판준, 2015). 이 연구 역시 디지털 큐레이션

이션의 정의와 의미가 모호하다고 밝히며 “디지털 큐레이션이란 생애주기 동안 디지털 정보의 체계적인 관리를 통하여 접근 및 재사용을 보장하는 디지털 정보 서비스”로 정의하였다.

최근 박민석과 이지수(2024)는 디지털 큐레이션의 연구 동향을 분석하기에 앞서, 디지털 큐레이션과 관계있거나 유사한 용어들을 검토하여 그 관계를 <그림 2>와 같은 다이어그램으로 제시하였다. 이들에 따르면 디지털 큐레이션은 데이터 큐레이션 등 유사 개념을 포함하며, ‘디지털 정보의 생애주기를 바탕으로 장기적 접근과 재사용을 촉진하는 관리 활동’으로 정의하였다. 디지털 큐레이션이 포함하고 있는 개념 중 디지털 아카이빙은 ‘장기보존의 가치가 있는 디지털 정보를 선별하고 장기적인 보존과 지속적인 접근을 가능하게 하는 절차 및 정책’을 의미하며, 콘텐츠 큐레이션은 ‘이용자 정보요구에 따른 맞춤형 정보 서비스’를 의미한다. 마지막으로 가장 작은 개념에 속하는 데이터 큐레이션은 그 대상이 연구 데이터에 한정된다고 하였다.



* 박민석, 이지수(2024) 재편집

<그림 2> 디지털 큐레이션과 유사용어 간의 관계

김관준(2015), 박민석과 이지수(2024)의 연구는 본 논문과 같이 디지털 큐레이션에 대한 개념 정립을 시도하는 연구로, 특히 관련 연구 동향을 살펴봄으로써 유사 개념과의 관계를 파악하는데 주안점을 두었다. 두 연구 모두 디지털 큐레이션이 연구되는 학문 분야 혹은 커뮤니티를 탐색하거나 연구 주제를 대표하는 저자 키워드 중심으로 개념 지도를 그려보는 시도를 한 반면, 본 연구는 학술 논문에 저자가 직접 기술했거나 인용한 디지털 큐레이션의 정의문에 나타난 주요 어휘를 통해 개념 구조를 파악한다는 점에서 차이가 있다. 이러한 문장은 연구자가 해당 연구에서 디지털 큐레이션이란 개념이 갖는 의미를 명기한 문장으로, 본인의 연구에서 강조하고자 하는 디지털 큐레이션의 특징을 설명하는 의도가 반영된다는 점에서 의미가 크다.

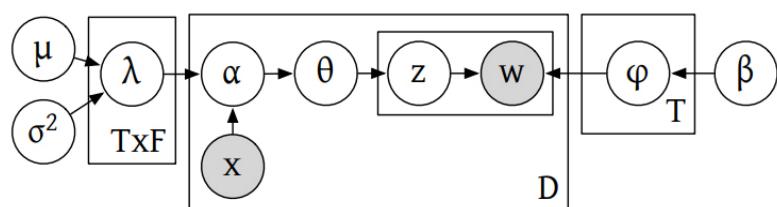
2.2 DMR 토픽 모델링

본 연구는 특정 개념의 정의가 시간의 흐름에 따라 어떻게 변화하고 확장되었는지를 살펴보는 연구로, 텍스트 기반 토픽 모델링 방법 중 DMR (Dirichlet Multinomial Regression) (Mimno & McCallum, 2012) 모델을 차용하였다. DMR은 이미 널리 사용되고 있는 토픽 모델링 알고리즘인 LDA (Latent Dirichlet Allocation) (Blei

et al., 2003) 모델의 확장형으로, LDA가 문서라는 독립 변수만을 입력으로 받아 그 안에 숨겨진 주제를 찾는 알고리즘이라면, DMR은 메타데이터라고 불리는 명목 변수를 기준으로 텍스트 그룹을 나누고 각 그룹의 문서 주제 구성 차이를 동시에 고려하는 알고리즘이다(박영욱, 정규엽, 2021).

〈그림 3〉은 DMR의 토픽 모델링의 과정을 도식화한 것이다(Mimno & McCallum, 2012). 그림의 λ 는 명목 변수별 하이퍼 파라미터 결정 값으로, λ 를 통해 각 문서의 주제 분포를 결정하는 α 에 영향을 주어 문서의 주제가 변수에 따라 달라질 수 있도록 변형한 것이다. 이때, 명목 변수로는 저자, 연도, 기관 등을 문서 외부에 부여된 다양한 요소들이 고려될 수 있고, 하나의 명목 변수가 아닌 여러 개의 명목 변수를 이용할 수도 있다. 다시 말해 DMR 모델링은 문헌의 날짜나 출처와 같은 외부 변수의 매개 효과에 따른 주제들의 변화를 파악할 수 있는 분석 기법으로, 변수와 주제 간의 동적 관계를 파악할 수 있다(함승경, 최지명, 2023).

국내 연구 중 DMR 토픽 모델링을 활용하여 시간의 흐름에 따른 토픽 추세를 분석한 사례는 다음과 같다. 김용환, 김유신(2019)은 국내 헬스케어 학술 연구의 트렌드를 논문 초록을 대상으로 DMR 토픽 모델링을 통해 분석하였



〈그림 3〉 DMR 토픽 모델링 과정(Mimno & McCallum, 2012)

으며, 그 결과 IT기술과 국가 정책 관련 토픽이 부상하고 있음을 확인하였다. DMR 모델은 학술 문헌에만 국한된 것이 아닌 뉴스와 소셜 미디어와 같은 텍스트 데이터 연구에서도 활용되어 왔다. 신원규와 김혜진(2023)은 국립대학육성사업 관련 뉴스 기사를 수집하고 DMR 토픽 모델링을 수행하여 국립대학육성사업 5대 중점 추진과제의 이슈 경향성을 파악하였다. 함승경과 최지명(2023)은 사회 재난 이후의 소셜 미디어에서 사람들의 반응을 파악하기 위해 10·29 이태원 참사 이후 온라인 트윗을 수집하여 DMR 토픽 모델링을 진행하였다. 이처럼 DMR은 주로 어떤 주제나 이슈의 시간 추이에 따른 변화 분석에 유용한 도구로, 본 연구에서도 논문이 게재된 시점 정보를 추가 변수로 설정, 연구 기간 따른 개념 변화 분석을 꾀하였다.

3. 디지털 큐레이션 용례 분석

3.1 분석 대상 설정

본 연구의 대상이 되는 ‘디지털 큐레이션’ 관련 정의문을 수집하기 위해서 먼저 KCI 포털을 활용해 “디지털 큐레이션”을 질의어로 관련 학술 문헌을 수집하였다.¹⁾ 가능한 많은 수의 논문을 후보군으로 구축하기 위하여 별도의 검색 조건을 추가하지 않았으며, 한 해 논문의 출판이 종결되는 2023년 12월 31일까지 발간된 논문으로 한정하였다. 그 결과 전체 122건의 논문이 선정되었으며, 수집된 논문의 원문 내 모든

문장을 대상으로 디지털 큐레이션과 관련된 정의 문장을 추출하였다. 정의문 선별을 위해서 한희정 외(2017)에서 고안한 정의문 폐턴 인식 기법을 구현, 후보 문장을 자동으로 인식한 후 수작업으로 검증하여 선별하였다.

이 같은 과정을 통해 선별된 문장은 총 110개로, 동일 문헌에서 나온 문장은 하나의 정의문으로 통합, 총 76개의 용례셋을 선정하였다. 〈표 1〉은 선정된 76개 정의문 용례에 대한 통계로, 디지털 큐레이션에 대한 정의문과 더불어 관련 용어를 설명한 문장도 포함하고 있다. 분석 대상 110개 문장의 총 어절 수는 2,504개이며, 문장당 평균 22.76개의 어절로 구성되었다. 이는 일반적인 문장 길이에 비해 긴 편으로, 정의문이 어떠한 개념어를 풀어 설명하는 문장이라는 특성이 나타남을 의미한다. 주요 품사 기준 통계를 살펴보면, 110개 문장은 전체 5,274개 품사로 구성되었으며, 이 중 일반명사와 복합명사 등을 포함한 명사류가 1,820개, 동사가 209개로 파악되었다.

〈표 1〉 단어별 정의문 수

단어	정의문 개수(문헌 기준)
디지털 큐레이션	38
큐레이션	27
큐레이션 서비스	4
데이터 큐레이션	3
디지털 큐레이션 역량	1
디지털 데이터 큐레이션	1
콘텐츠 큐레이션	1
소셜 큐레이션	1
총합계	76

1) <https://www.kci.go.kr/kciportal/main.kci>, 검색시점: 2024.02.22.

일반적으로 키워드 분석 및 토픽 모델링 결과의 신뢰성을 보장하기 위해서는 맞춤법이나 띄어쓰기 오류, 비속어, 축약어 등 다수의 비문을 수정하는 전처리 과정이 수반된다. 그러나 본 연구의 분석 대상인 76개의 정의문은 저자가 학술 문헌에 기술한 문장을 추출한 것으로 별도의 맞춤법·오타 수정 등의 검수 과정을 거치지 않고 원문 그대로를 사용하였다. 다만, 형태소 분석²⁾ 이후 ‘디지털 큐레이션’, ‘큐레이션’과 같이 정의문에서 중복되어 나타나 고빈도로 출현하는 단어와 ‘연구’, ‘과정’, ‘분석’ 등 자주 쓰이지만 중요하지 않은 단어, 한 글자로 된 명사는 데이터 질을 떨어뜨릴 수 있어 불용어 처리하였다. 또한, 빈도가 1로 나타나는 저빈도 단어는 잡음으로 작용, 의미 해석을 방해할 수 있어 분석 대상에서 제외하였다. 이후 이용자 사전을 구축, 하나의 단어가 의미 없는 여러 단어로 분절되어 분석되는 경우를 방지하였다.

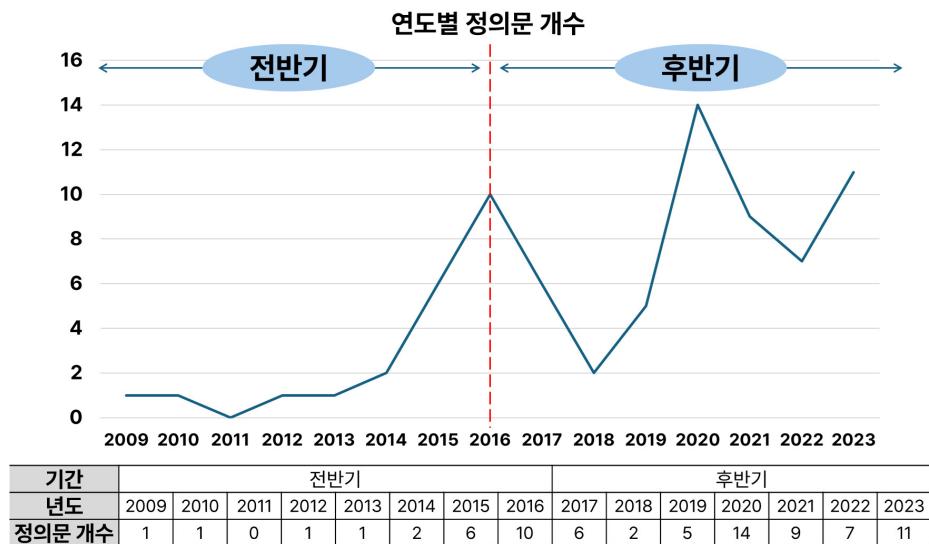
3.2 정의문 용례 분석

〈그림 4〉는 본 연구의 연구 대상으로 수집된 정의문 76개의 연도별 현황이다. 수집된 정의문의 연도별 동향을 분석한 결과, 국내에서 가장 처음 ‘디지털 큐레이션’이란 용어가 사용되고, 정의가 명시된 논문은 2009년 발간된 안영희, 박옥화의 연구로, 이들은 디지털 정보자원이 발전함에 따라 디지털 큐레이션의 역할 및 전략을 제시하였다. 그 후 2016년까지는 꾸준히 성장하였다 2017년을 시작으로 감소하는 추세를 보였고, 2020년에 다시 급증하였다. 2020

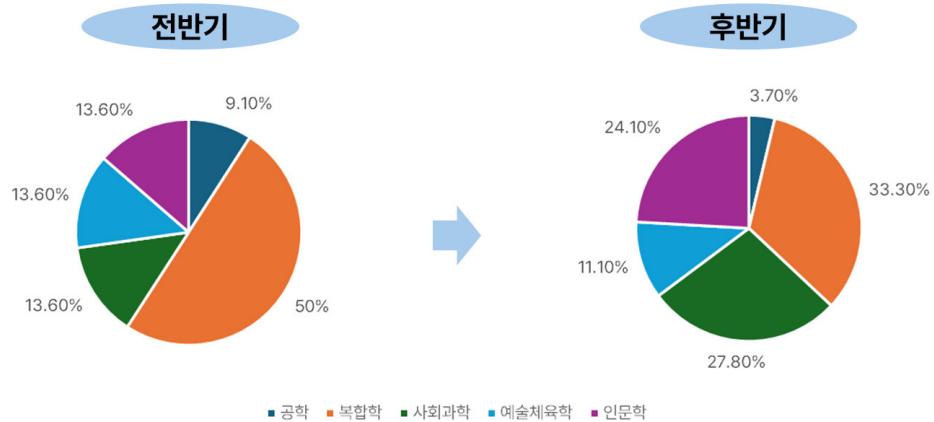
년은 정의문이 가장 많이 추출된 시기이다. 이는 코로나 시기의 비대면 상황에 대응해 도서관이나 박물관 등 다양한 기관에서 온라인 서비스를 위한 큐레이션을 시도하고, 전산화를 통해 방대한 양의 정보가 생산된 상황에서 이를 효과적으로 관리하기 위한 방안 등이 연구되며 디지털 큐레이션에 대한 개념이 다수 언급되었을 것으로 사료된다. 일례로 김보람과 이종욱(2022)은 국립중앙박물관의 e 뮤지엄 데이터를 중심으로 정보의 시각화 방안을 연구하였다. 이러한 추세를 반영하여, 정의문의 양이 급증한 2016년을 기준으로 2009년부터 2016년까지를 전반기, 잠시 주춤한 2017년부터 2023년까지를 후반기로 구분하였다.

〈그림 5〉는 KCI에서 제공하는 논문의 주제분야를 한국연구재단의 ‘학술연구분야별 류표’에 따라 대분류로 역추적하고, 이를 그래프로 나타낸 것이다. 전반기에는 복합학 분야에서 수집된 정의문이 11개(50%)로 가장 많이 언급되었으며, 사회과학, 예술체육학, 인문학이 각각 3개(13.6%), 공학이 2개(9.1%)를 차지했다. 후반기 역시 복합학이 18개로 가장 많은 수를 차지하였으나 비율은 33.3% 감소하였으며, 뒤를 이어 사회과학이 15개(27.8%), 인문학이 13개(24.1%), 예술체육학이 6개(11.1%), 공학이 2개(3.7%) 순으로 집계되었다. 이는 문헌정보학 분야에서 디지털 큐레이션 연구를 주도(박민석, 이지수, 2024)하고 있다는 선행 연구의 결과와 같은 맥락으로, 본 연구에서도 전·후반기에 정의문의 비율이 가장 높은 복합학 분야의 대다수가 문헌정보학에 포함됨을 파악할

2) PyKomoran 라이브러리



〈그림 4〉 연도별 정의문 개수



	전반기	후반기	총계
공학	2(9.1%)	2(3.7%)	4
복합학	11(50%)	18(33.3%)	29
사회과학	3(13.6%)	15(27.8%)	18
예술체육학	3(13.6%)	6(11.1%)	9
인문학	3(13.6%)	13(24.1%)	16
합계	22	54	76

〈그림 5〉 학문분야별 전 · 후반기 비율

수 있었다. 또한 전반기와 후반기의 학문 분야별 논문 발행 동향을 비교해서 살펴본 결과, 디지털 큐레이션 개념이 전반기에는 복합학에서 다수 사용되던 개념이었지만, 후반기로 옮수록 복합학, 사회과학, 인문학 등 그 분포가 고르게 나타났다. 본 연구는 이러한 변화 양상을 더욱 구체적으로 파악하기 위해 DMR 토픽 모델링 기법을 활용하여 세부 학문 분야별 주요 토픽과 그 확장 추이를 분석한다.

4. 디지털 큐레이션 개념 구조 분석

4.1 주요 토픽 분석

본 절에서는 시간의 흐름에 따른 디지털 큐레이션의 개념 구조 변화를 분석하기 앞서, 디지털 큐레이션의 정의에 어떠한 주제들이 기술되어 있는지를 먼저 분석하고자 한다. 특히, 각 토픽에 분류된 정의문의 핵심 주제와 단어들 간의 연결관계를 먼저 살펴보기 위해 워드 네

트워크 기법을 사용하여 시각화³⁾하였다. 본 연구에서는 디지털 큐레이션의 다양한 개념과의 연결 관계가 많은 노드를 파악하기 위해 연결 중심성(Degree Centrality)을 채택하였다(배나윤, 오효정, 2024). 최적의 토픽 수를 설정하기 위해 전처리가 완료된 데이터를 대상으로 사전 실험을 진행하였으며, 그 결과 응집도(Coherence)가 가장 높고 혼란도(Perplexity)가 가장 낮은 토픽 수는 8로 도출되었다.

〈표 2〉는 토픽 모델링의 결과로, 주요 단어와 워드 네트워크 결과를 참고하여 토픽 주제명을 부여하였다. 가장 큰 비율을 차지하는 토픽은 ‘데이터 보존(Topic #1)’으로, 가장 많은 32개 정의문에서 나타난 어휘들이 군집되었다. 뒤를 이어 ‘전통적인 큐레이터 역할(Topic #2)’, ‘상품 추천 큐레이션(Topic #3)’ 순으로, 가장 작은 비율을 차지하는 [Topic #8]은 ‘큐레이션 모델’ 주제로 명명하였다.

〈표 3〉은 각 토픽에 할당된 정의문이 도출된 논문의 출처를 기반으로 기술된 학문 분야를 살펴본 것이다. ‘데이터 보존(Topic #1)’은 복합학, ‘전통적인 큐레이터 역할(Topic #2)’은 사

〈표 2〉 토픽 모델링 결과

토픽	정의문 개수	토픽 주제명	토픽 주요 단어
Topic #1	32	데이터 보존	데이터, 디지털, 보존, 유지, 활동
Topic #2	9	전통적인 큐레이터 역할	큐레이터, 박물관, 미술관, 수집, 전시
Topic #3	7	상품 추천 큐레이션	서비스, 추천, 소비자, 큐레이션 서비스, 상품
Topic #4	7	콘텐츠 분류 및 유통	콘텐츠, 사람, 제공, 분류, 소비
Topic #5	6	정보 이용	정보, 사용, 이용자, 자료, 지식
Topic #6	6	큐레이션 서비스	정보, 서비스, 작업, 가치, 연관
Topic #7	6	큐레이션 필요성	행위, 콘텐츠, 사용자, 과정, 맥락
Topic #8	3	큐레이션 모델	가치, 분류, 모델, 소프트웨어, 거래

3) Gephi(0.10). 오픈소스 네트워크 분석 및 시각화 소프트웨어 패키지

〈표 3〉 토픽별 학문 대분류 연결

* **빨간색: 토픽 내 최대 검정색: 학문분야 내 최대**

	토픽 주제명	공학	복합학	사회과학	예술체육학	인문학	총계
Topic #1	데이터 보존	<u>2</u>	20	2	0	<u>8</u>	32
Topic #2	전통적인 큐레이터 역할	1	2	3	0	3	9
Topic #3	상품 추천 큐레이션	0	0	6	1	0	7
Topic #4	콘텐츠 분류 및 유통	1	1	3	2	0	7
Topic #5	정보 이용	0	1	0	1	4	6
Topic #6	큐레이션 서비스	0	2	0	4	0	6
Topic #7	큐레이션 필요성	0	2	2	1	1	6
Topic #8	큐레이션 모델	0	1	2	0	0	3
합계		4	29	18	9	16	76

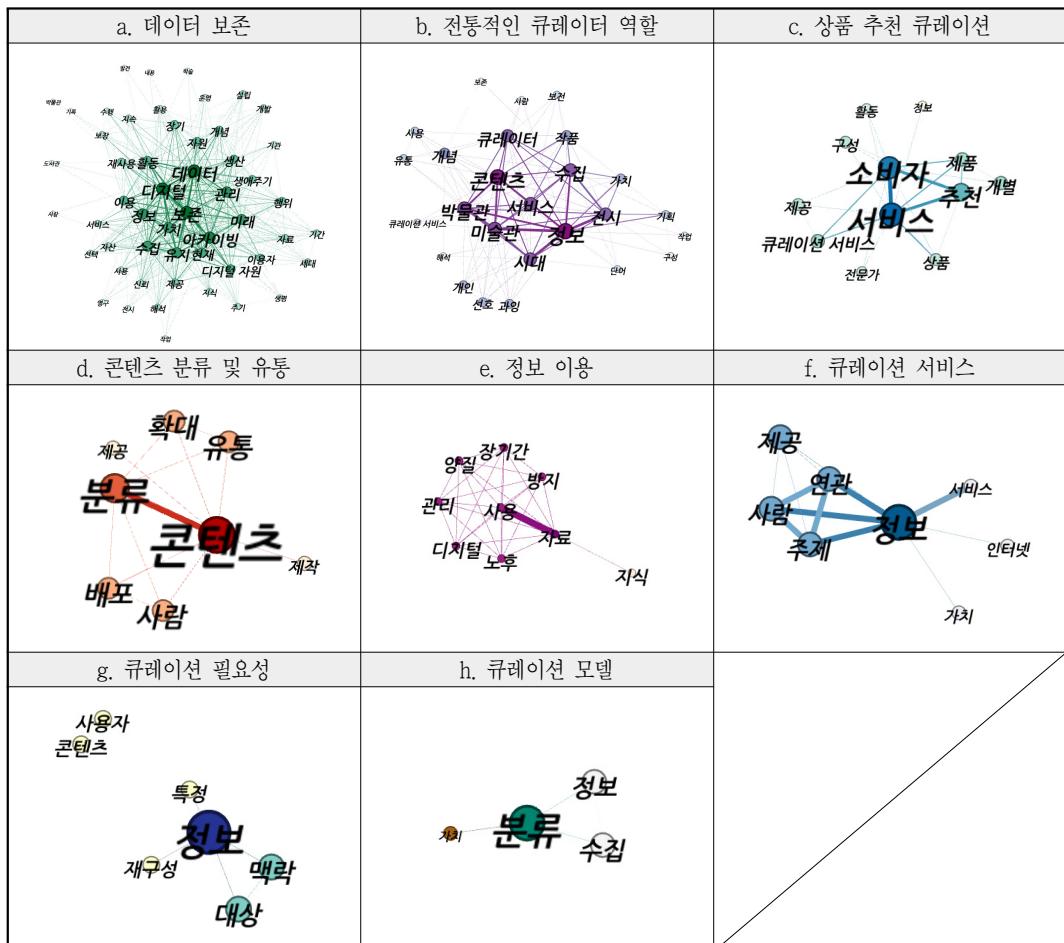
회과학과 인문학, '상품 추천 큐레이션(Topic #3)', '콘텐츠 분류 및 유통(Topic #4)'은 사회과학, '정보이용(Topic #5)'은 인문학, '큐레이션 서비스(Topic #6)'는 예술체육학, '큐레이션 필요성(Topic #7)'은 복합학과 사회과학, 마지막으로 '디지털 큐레이션 모델(Topic #8)'은 사회과학에서 강세를 나타냈다. 이를 학문 대분류를 기준으로 살펴보면 공학, 복합학, 인문학은 '데이터 보존(Topic #1)', 사회과학은 '상품 추천 큐레이션(Topic #3)', 예술체육학은 '큐레이션 서비스(Topic #6)'에서 강세를 보였다.

〈그림 6〉은 각 토픽별로 디지털 큐레이션의 정의를 설명하는 개념어를 파악하기 위해, 워드 네트워크를 통해 중요한 키워드의 연결 관계를 살펴본 것이다. 노드의 크기는 동시출현 빈도를 기준으로 설정하였고, 빈도수가 1인 동시출현쌍은 분석의 신뢰성과 명확성을 떨어뜨릴 수 있다고 판단하여 빈도수 2 이상인 데이터만을 대상으로 구성하였다.

〈그림 6.a〉는 '데이터 보존'의 워드 네트워크 결과이다. 전체 76개의 정의문 중 32개가 포함

된 가장 큰 토픽이자 복합학 분야 연구 정의문 29개 중 20개가 포함된 토픽이다. 연결 중심성 순으로 보면 보존(0.865), 데이터(0.865), 디지털(0.846) 순으로 나타났다. 해당 토픽을 살펴본 결과, 가장 처음 디지털 큐레이션의 개념을 설명한 안영희와 박옥화(2009)의 정의문이 해당 토픽에 분류되어 있음을 확인하였다. 이를 중심으로 뚝인 정의문은 데이터를 보존하고 가치를 유지하기 위한 활동을 총칭하고 있었으며 다수의 연구자가 이와 같은 의미를 포함하는 정의문을 기술하였다. 즉, 다수의 연구자가 공감할 수 있는 보편적인 디지털 큐레이션에 대한 주요 역할은 데이터를 보존하는 데에 있음을 시사한다. '데이터 보존' 주제는 박민석, 이지수(2024)와 한국과학기술정보연구원(2005)에서 수행한 연구에서도 디지털 큐레이션의 범주로 포함된 개념이다.

다음 〈그림 6.b〉는 '전통적인 큐레이터 역할' 이란 주제의 워드 네트워크이다. '큐레이션'과 '큐레이터'라는 단어가 주로 미술관과 박물관의 학예사에게 사용되던 단어로 전시를 기획하고 미술품 등을 수집하던 업무 역시 꾸준히 강조



〈그림 6〉 토픽별 워드 네트워크

되고 있음과 동시에, 업무가 확장되어 콘텐츠를 소설 미디어에 전시하거나 유통하는 일련의 행위까지 아우르는 의미를 포함한다는 내용의 정의문이 포함되었다.

〈그림 6.c〉는 ‘상품 추천 큐레이션’의 워드 네트워크로, 해당 토픽은 전체 정의문 7개 중 6개가 사회과학 분야 연구에서 추출되었고, 종 분류까지 살펴본 결과 3개가 경영학으로 나타났다. 실제 문장을 살펴본 결과, 기업이 개별 소비자에게 맞춤형 제품이나 서비스를 구성하여

추천하고, 고객의 요구에 맞추기 위한 전문적인 서비스로 큐레이션을 설명하고 있었다. 한 편 〈그림 6.d〉는 ‘콘텐츠 분류 및 유통’ 주제의 워드 네트워크로, 주요 단어는 ‘콘텐츠, 제공, 분류, 소비’ 등으로 앞선 ‘상품 추천 큐레이션’의 토픽의 주요 단어와 비슷하게 보인다. 그러나 네트워크 모양을 면밀히 살펴보면, ‘콘텐츠’의 노드가 그 무엇보다 커 독보적인 핵심이임을 확인할 수 있다. 실제로 정의문을 살펴보면 큐레이션 대상이 기업의 상품이 아닌 콘텐츠로

나타났고, 이러한 디지털 콘텐츠를 적절하게 분류하고 유통하는 과정을 포함하는 정의문들이 다수 기술되어 있었다.

〈그림 6.e〉는 ‘정보 이용’으로 토픽의 주요 단어와 워드 네트워크의 핵심어가 ‘사용’을 중심으로 구성되어 있음을 확인할 수 있다. 어떤 데이터가 지식이나 자료가 되기 위해서는 가치가 부여되어야 하는데, 이 가치를 부여하는 일련의 행위가 디지털 큐레이션을 통해 실현될 수 있음을 시사한다. 〈그림 6.f〉는 ‘큐레이션 서비스’ 토픽의 워드 네트워크 결과로, 전체 6개의 정의문 중 예술체육학 분야의 문장이 4개로 파악되었다. 이는 인포그래픽과 같은 정보의 시각화와 관련한 연구를 진행하며 기술된 큐레이션의 정의문을 포함한 것으로 보인다.

〈그림 6.g〉는 ‘큐레이션 필요성’의 워드 네트워크이다. 정보과잉 시대에서 콘텐츠를 필터링하고, 재구성하여 사용자의 필요와 목적에 맞게 걸러주는 큐레이션의 필요성을 강조한 정의문들이 포함되어 있다. 마지막으로 〈그림 6.h〉는 ‘큐레이션 모델’의 워드 네트워크이다. 포함된 정의문의 수가 단 3개로 가장 작은 양의 문장으로부터 도출된 토픽이며 모두 후반기에 출판된 논문으로부터 추출된 문장이라는 점도 큰 특징 중 하나이다. 또한 전자상거래 분야에서의 디지털 큐레이션 유형을 설명한 정의문들이 포함되어 있다.

주요 토픽과 학문 분야를 함께 살펴본 결과, 특정 분야의 쏠림이 있는 토픽이 다수 존재하였다. 이는 학문 분야별로 디지털 큐레이션을 설명한 정의문에서 언급되는 핵심어가 겹친다는 의

미로, 같은 분야를 연구하는 연구자들이 방점을 두는 개념이 중복되는 자연스러운 현상을 반영한다. 예를 들어, ‘데이터 보존’ 토픽에서는 복합학, ‘상품 추천 큐레이션’에서는 사회과학, ‘큐레이션 서비스’에서는 예술체육학이 대표적인 사례다. ‘데이터 보존’은 가치 있는 데이터를 장기적으로 보존하는 것을 목적으로 한다. 복합학, 특히 문헌정보학이 포함된 분야는 정보 서비스와 밀접한 관련이 있어 데이터 보존과 관련된 연구가 활발히 이루어지고 있음을 짐작할 수 있다. 실제로 해당 토픽에 포함된 정의문 중에는 연구 데이터의 큐레이션 모델을 제안한 연구(한나은, 2023) 등이 포함되어 있었다. 또한, 사회과학 중에서도 경영학 분야가 다수 포함된 ‘상품 추천 큐레이션’ 토픽에서는 패션 브랜드의 고객 경험에 관한 사례 연구(이은경, 김연정, 2023)가 포함된 것을 확인할 수 있었다. 현대 사회에서 다양한 시대가 도래함에 따라, 상품 추천 큐레이션은 소비자가 자사의 제품을 선택하게 만드는 중요한 전략으로 사용될 수 있어 앞으로도 계속 연구될 것으로 보인다. 마지막으로, ‘큐레이션 서비스’는 예술체육학 분야에서 강세를 나타낸 토픽으로, 정보 시각화와 관련된 논문들이 이 토픽에 분류되었다.

4.2 시간 흐름에 따른 개념 변화

이번 절에서는 DMR 토픽 모델링⁴⁾을 활용해 시간의 흐름에 따른 디지털 큐레이션 개념 변화를 분석한다. DMR 분석에는 앞서 3장에서 설정된 기준, 즉 논문의 발행년도를 중심으

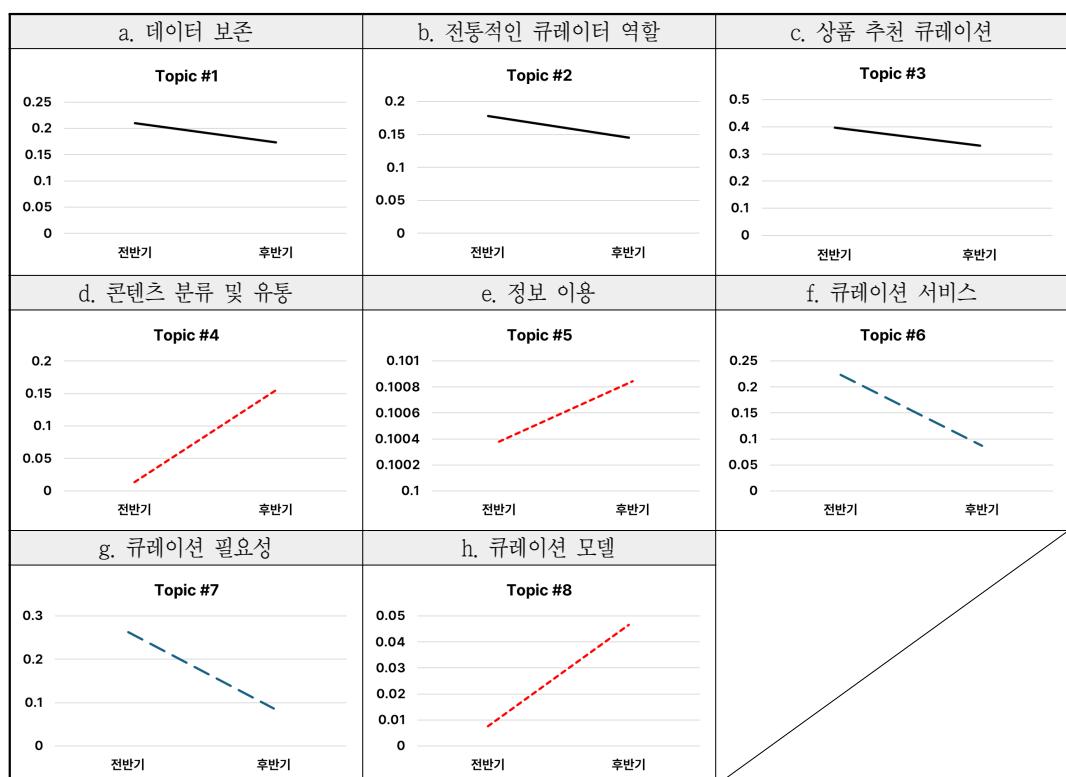
4) python의 tomotopy 라이브러리 활용

로 전반기(2009년~2016년)와 후반기(2017년~2023년)를 비교 변수로 설정하였다. 이를 통해 2009년부터 2023년까지 디지털 큐레이션에 관련된 다양한 토픽이 어떻게 나타나고, 시간의 흐름에 따라 어떤 토픽이 강화되거나 약화되는지 파악하고자 한다.

〈그림 7〉은 8개 주요 토픽의 시간에 따른 변화를 도식화한 것으로, 시간이 흐름에 따라 각 토픽별로 중요도가 상승한 주제와 하향된 주제로 구분되는 것을 볼 수 있다. 보다 구체적으로 살펴보면, 검은색 실선으로 표현된 〈그림 7.a〉 ‘데이터 보존’, 〈그림 7.b〉 ‘전통적인 큐레이터 역할’, 〈그림 7.c〉 ‘상품 추천 큐레이션’은 전반

기와 후반기의 차이가 크지 않다. 이는 시간의 흐름에도 불구하고 지속적으로 그 중요성이 강조되는 토픽들이라 해석할 수 있다. 이 토픽들은 정의문의 개수를 기준으로 정렬된 상위 3개에 해당하며, 절대적인 양 또한 많아 다수의 연구자가 사용하고 있는 개념이다.

붉은 점선으로 표현된 〈그림 7.d〉 ‘콘텐츠 분류 및 유통’, 〈그림 7.e〉 ‘정보 이용’, 〈그림 7.g〉 ‘큐레이션 모델’은 후반기에 대두된 주제들이다. ‘콘텐츠 분류 및 유통’과 ‘정보 이용’은 데이터의 양이 폭발적으로 증가하는 정보 과잉 시대에 대응하기 위해 정보의 진위 여부를 파악하고 재구성하는 것을 넘어, 양질의 지식과 정



*붉은 점선: 상향 토픽, 파랑 점선: 하향 토픽

〈그림 7〉 기간별 토픽 분포

보로 활용될 수 있는 콘텐츠를 분류하고 재구성하여 유통하는 방식으로 디지털 큐레이션의 사용 용례가 변화했음을 보여준다. 특히, 정보 이용자의 편의성을 고려해 자료를 배열하고 정리하는 것에서 나아가, 정보의 가치를 더욱 높이는 행위로 확장될 수 있다. 마지막으로, '큐레이션 모델'은 기존의 '상품 추천 큐레이션'이 전자상거래의 구독 서비스로 확장되어 연구되고 있음을 나타낸다. 또한, <그림 7.h>의 경우 전자상거래 모델을 포함한 토픽인 '큐레이션 모델'의 그래프이다. 평균적으로 국내 전자상거래 시장이 19.7%씩 증가하였다는 자료(구진경, 2021)를 참조하면 그 중요도가 앞으로 더욱 커질 것으로 보인다.

이에 비해 파란색 점선으로 표현된 <그림 7.f> '큐레이션 서비스'와 <그림 7.g> '큐레이션 필요성'은 전반기에 비해 후반기에서 중요도가 급감하였다. 초창기에는 디지털 큐레이션 자체의 필요성과 당위성에 초점을 맞췄다면, 후반기로 갈수록 디지털 큐레이션의 구체적인 방법과 모델 등 다른 주제의 논의로 전환되었기 때문이다.

상기한 내용을 종합하면 먼저, '데이터 보존', '전통적인 큐레이터 역할', '상품 추천 큐레이션'은 전반기와 후반기 모두에서 고르게 나타난 중요한 주제이다. '데이터 보존'은 디지털 큐레이션이 포괄하고 있는 유사 개념 중 하나(박민석, 이지수, 2024)로, 훼손되거나 손실될 수 있는 디지털 자료의 특성상 이를 안전하게 보존하는 것이 디지털 큐레이션에서 중요한 개념이라 해석할 수 있다. '전통적인 큐레이터 역할'은 이전에는 박물관과 미술관 등에서 전시를 위해 작품을 선별, 수집하는 전통적인 큐레이터의 역할이 디지털 시대에 확장되어 디지털 큐레이션

을 위해 가치 있는 정보를 선별, 수집하는 역할까지 포함한다고 해석할 수 있다. 마지막으로 '상품 추천 큐레이션'은 전자상거래에서 소비자가 만족할 수 있도록 연관된 상품을 추천하는 것이지만 디지털 큐레이션의 맥락에서는 주로 콘텐츠들을 추천하고, 가치 있는 데이터를 선별해 보존하는 활동 등으로 확장되어 다뤄지는 경우도 있었다. 이들은 디지털 큐레이션의 핵심적인 개념에 해당한다.

반면, 전·후반기에 중요도 변화가 급감한 토픽으로는 '큐레이션 서비스'와 '큐레이션 필요성'이 있다. 이는 디지털 큐레이션의 필요성에 대한 논의가 주로 초창기에 이루어지는 점을 시사하는 것으로, 새로운 학문이 처음 논의되는 시점의 연구 동향과 그 맥을 같이 한다. 이와는 대비되게 연구가 성숙됨에 따라 대두된 토픽으로는 '콘텐츠 분류 및 유통', '정보 이용' 등이 있다. 특히 최근 발표된 연구를 중심으로 살펴보면, 큐레이션의 의미가 '데이터 보존' 뿐만 아니라, '콘텐츠 분류 및 유통', '정보 이용'까지도 포함하는 광의적인 개념으로 확장되었음을 확인할 수 있다. 또한, 초창기의 개념 정립 단계를 넘어 실질적인 적용과 활용으로 발전했음을 보여준다. 이를 시간의 흐름, 학문 분야, 토픽으로 중첩해 살펴보면, 전반기에는 복합학이 전체 연구의 50%를 차지했던 반면, 후반기에는 복합학이 33.3%, 사회과학이 27.8%, 인문학이 24.1%를 차지하며 디지털 큐레이션이 다양한 학문을 포괄하는 개념으로 변화했음을 확인할 수 있다. 디지털 큐레이션의 시간 흐름에 따른 개념 변화를 파악한 결과, 전·후반기의 큰 변화 없이 다수의 정의문이 포함된 핵심 학문과 개념으로는 복합학과 데이터 보존이었다. 전반기에서 후반

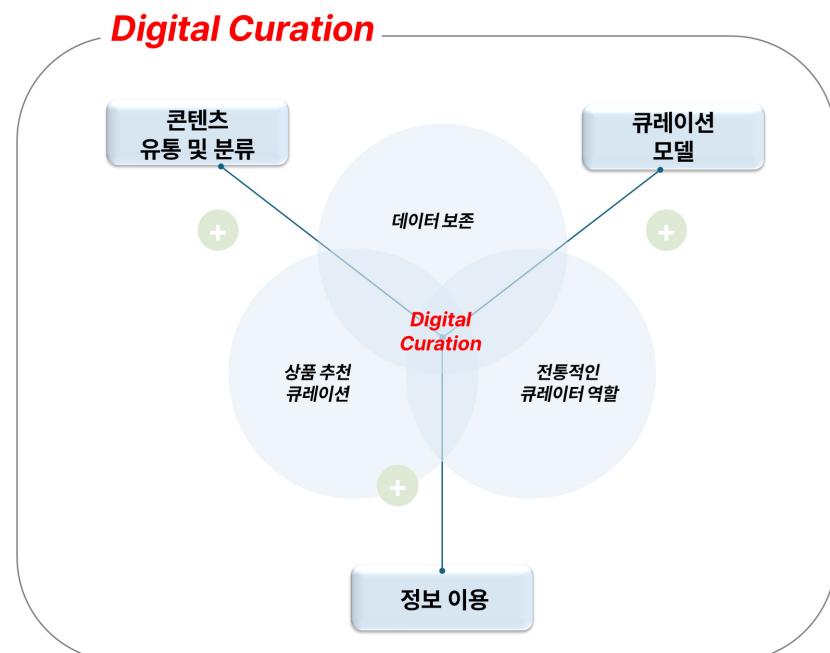
기로 가며 복합학의 비율이 50%에서 33.3%로 감소했지만, 여전히 가장 많은 수를 차지하고 있다. 또한, 증가 추세인 토픽 ‘콘텐츠 유통 및 분류’, ‘정보 이용’, ‘큐레이션 모델’의 대다수가 인문학·사회과학으로, 기간별 학문 변화와도 그 맥을 같이 하여 양적으로도 증가하고, 토픽의 중요도도 올라갔음을 확인할 수 있다. 또한, 실제로 이러한 변화는 더욱 다양한 학문 분야에서 큐레이션 개념이 사용되고 있음을 보여준다.

4.3 디지털 큐레이션 개념도

〈그림 8〉은 상기한 분석 결과를 종합하여 표현한 디지털 큐레이션의 주요 개념 구조도로, 연구자들이 학술 문헌에서 언급하고 있는 디지털 큐레이션의 개념은 크게 8가지로 나뉜다. 다

만 주요 의미가 계속해서 광의적으로 확장되어 가고 있어 전반기에 비해 그 중요도가 현저하게 떨어진 개념을 제외한 6가지의 주제를 기준으로 도식화하였다.

디지털 큐레이션의 개념에서 가장 중요한 토픽은 ‘데이터 보존’, ‘전통적인 큐레이터 역할’, ‘상품 추천 큐레이션’이다. 이 세 주제는 학술 연구 전·후반기 모두 꾸준히 등장하며 디지털 큐레이션의 핵심 내용을 설명하는 가장 중요한 축이다. 해당 개념을 중심으로 디지털 큐레이션이 연구되어지며 이용자의 관점에서는 ‘정보 이용’ 토픽과 결합되고, 디지털 큐레이션의 관계자 관점에서는 ‘콘텐츠 유통 및 분류’의 개념이 우선하여 사용될 수 있다. 또한, 전자상거래와 연관될 땐 ‘큐레이션 모델’ 방향으로 연구되며, 앞으로 변화의 양상이 다양해질 것으로 예상된다.



〈그림 8〉 디지털 큐레이션 개념도

5. 결 론

본 연구는 학술 문헌에 나타난 정의문의 용례를 통해 시간의 흐름에 따른 디지털 큐레이션의 개념 변화를 분석한 것이다. 연구 결과는 다음과 같다. 첫째, 디지털 큐레이션이라는 개념은 2009년 처음 국내 학술 문헌에 등장한 이후 꾸준히 연구되다 2016년 급증하였다. 디지털 전환이 급속하게 이루어지며 방대해진 데이터와 정보를 관리하는 방안으로 디지털 큐레이션 등이 대두되었기 때문이라 짐작된다. 전반기 대다수의 정의문은 복합학 분야에서 정의되었으며, 후반기로 올수록 사회과학, 인문학 등 그 분포가 고르게 나타났다.

DMR 토픽 모델링을 통해 시간의 흐름에 따른 토픽의 중요도의 증감을 살펴 결과, 전반기에는 디지털 큐레이션의 필요성과 당위성을 강조하는 정의문들의 중요도가 높게 나타났다. 이는 새로운 학문 분야가 등장할 때, 해당 분야의 정당성을 확보하고 연구의 필요성을 널리 알리기 위해 관련 연구자들이 집중적으로 필요성과 당위성을 강조했음을 보여준다. 이러한 토픽은 후반기에서는 중요성이 감소하였고, '콘텐츠 분류 및 유통', '정보 이용', '큐레이션 모델' 등의 중요성이 증가하였다. 이는 디지털 큐레이션이 초기 개념 정립 단계에서 실질적인 적용과 활용으로 발전했음을 시사한다. 전·후반기 모두 그 중요성의 변화 없이 꾸준하게 유지된 토픽으로는 '데이터 보존', '전통적인 큐레이터 역할', '상품 추천 큐레이션'이 있다. 특히 '데이터 보존'은 디지털 큐레이션이 포괄하는 중요한 개념 중 하나로, 디지털 자료는 훼손되거나 손실될 위험이 있기 때문에 이를 안전하

게 보존하는 것은 디지털 큐레이션에서 매우 중요한 요소이다. 또한, '상품 추천 큐레이션', '전통적인 역할의 큐레이터' 역시 과거에는 그 개념이 특정 학문과 분야에서만 사용되었다면, 후반기로 올수록 그 범위가 확대되어 꾸준하게 사용되고 있음을 알 수 있었다.

마지막으로 디지털 큐레이션의 개념 변화를 추적하였고, 토픽을 중심으로 디지털 큐레이션의 개념도를 그려보았다. 그 결과 디지털 큐레이션의 개념은 '데이터 보존', '전통적인 큐레이터 역할', '상품 추천 큐레이션'이 중심축을 이루고 있고, 연구자마다 디지털 큐레이션을 활용하고자 하는 방향에 따라 '콘텐츠 유통 및 분류', '정보 이용', '큐레이션 모델' 등으로 확장하여 정의하고 있었다.

본 연구는 디지털 큐레이션의 개념을 연구자의 관점이 투영된 정의문에 기반해 분석했다는 데에 의의가 있다. 또한, 연구 진행 시점을 두 구간으로 나누어 토픽의 증감 추세를 통해 연구자들이 주안점을 두는 개념의 변화를 명시적으로 파악하였다는 데에 의미가 있다.

그러나 본 연구는 국내 디지털 큐레이션 관련 연구가 처음 시작된 지 15년 정도 된 신생 연구 영역으로 그 표본이 너무 작다는 데에 한계가 있다. 또한, 디지털 큐레이션은 이론뿐 아니라 실무에서도 활발히 사용되고 있는 용어이지만 학술 문헌만을 대상으로 수행된 연구이기 때문에 현장에서의 디지털 큐레이션에 대한 인식을 파악하지 못했다는 데에 아쉬움이 있다. 향후 연구에서는 관련자와의 설문조사나 인터뷰를 포함하여 현장과 학술 이론 사이의 인식 차이 등을 추가로 조사하고자 한다.

참 고 문 헌

- 구진경 (2021. 07. 02). [ICT] 리테일 4.0 시대를 대비하는 전자상거래 시장. Invest Korea.
출처: https://www.investkorea.org/ik-kr/bbs/i-112/detail.do?ntt_sn=491181
- 김보람, 이종욱 (2022). 문화유산 데이터 분석 및 정보 시각화 방안 연구: 국립중앙박물관 e뮤지엄 데이터를 중심으로. *한국전통문화연구*, 29, 115-153.
<http://doi.org/10.23067/tjokch..29.202205.115>
- 김용환, 김유신 (2019). 토픽모델링을 이용한 국내 헬스케어 학술연구 트렌드 분석. *한국웰니스학회지*, 14(1), 253-262. <http://doi.org/10.21097/ksw.2019.02.14.1.253>
- 김판준 (2015). 디지털 큐레이션 연구동향 분석과 과제: 문헌정보학 분야를 중심으로. *정보관리학회지*, 32(1), 265-295. <https://doi.org/10.3743/KOSIM.2015.32.1.265>
- 김현수, 정성모 (2023). 공유 공간 용어 사용에 대한 개념 비교 분석 및 용어 정립 제안. *한국공간디자인 학회* 논문집, 18(1), 185-198. <https://doi.org/10.35216/kisd.2023.18.1.185>
- 박민석, 이지수 (2024). 체계적 문헌고찰을 통한 국내 디지털 큐레이션 연구동향 분석. *한국기록관리학회지*, 24(2), 41-63. <https://doi.org/10.14404/JKSARM.2024.24.2.041>
- 박영우, 정규엽 (2021). DMR(Dirichlet Multinomial Regression) 토픽모델링을 이용한 온라인 리뷰 빅데이터 기반 고객감성 분석에 관한 연구: 국내 5성급 호텔의 외국인 이용객 리뷰를 중심으로. *호텔경영학연구*, 30(2), 1-20. <https://doi.org/10.24992/KJHT.2021.2.30.02.1>
- 배나윤, 오효정 (2024). 주요 학문분야 비교를 통한 국내 정보공개 연구동향 분석. *정보관리학회지*, 41(2), 295-316. <https://doi.org/10.3743/KOSIM.2024.41.2.295>
- 신원규, 김혜진 (2023). 토픽모델링을 활용한 국립대학육성사업 관련 이슈 분석: 뉴스 기사 및 정책추진과제 비교 분석을 중심으로. *교육행정학연구*, 41(5), 193-215.
<http://doi.org/10.22553/keas.2023.41.5.193>
- 안영희, 박옥화 (2009). 대학도서관 서비스의 디지털 큐레이션 전략. *한국도서관·정보학회지*, 40(4), 311-326. <https://doi.org/10.16981/kliss.40.4.200912.311>
- 이은경, 김연정 (2023). 온라인 플랫폼 기반 패션 브랜드의 고객 경험에 관한 사례연구: 오프라인 공간으로 진출한 브랜드 중심으로. *브랜드디자인학연구*, 21(3), 371-382.
<https://doi.org/10.18852/bdak.2023.21.3.371>
- 장효석 (2022). 도심항공교통의 디지털전환에 관한 연구. *한국항공보안학회지*, 1(4), 43-48.
- 조민지 (2020). 디지털 기억 시대의 기록과 정보서비스. *기록과 정보·문화 연구*, 10, 181-215.
<https://doi.org/10.23035/kaics.2020.1.10.181>
- 한국과학기술정보연구원 (2005). 국가 디지털 아카이빙 체제 구축에 관한 연구.

- 한나은 (2023). 활동이론을 중심으로 한 연구데이터 큐레이션 개념 모델 제안. *한국도서관·정보학회지*, 54(1), 167-190. <http://doi.org/10.16981/kliss.54.1.202303.167>
- 한희정, 김태영, 두효철, 오효정 (2017). 기술문서 정의문 패턴을 이용한 전문용어사전 자동추출 및 활용방안. *정보관리학회지*, 34(4), 81-99. <http://doi.org/10.3743/KOSIM.2017.34.4.081>
- 함승경, 최지명 (2023). 분노, 비판, 대립의 장이 되었던 트위터: 토픽모델링을 이용한 10·29 이태원 참사 트윗 분석. *Crisisnomy*, 19(12), 59-75. <https://doi.org/10.14251/crisisnomy.2023.19.12.59>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022. <https://doi.org/10.7551/mitpress/1120.003.0082>
- Higgins, S. (2011). Digital curation: The emergence of a new discipline. *International Journal of Digital Curation*, 6(2), 78-88. <http://doi.org/10.2218/ijdc.v6i2.191>
- International Data Corporation (2017). *Data Age 2025: The Evolution of Data to Life-Critical*. International Data Corporation.
- Mimno, D. & McCallum, A. (2012). Topic models conditioned on arbitrary features with Dirichlet-Multinomial Regression. <https://doi.org/10.48550/arXiv.1206.3278>

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

- Ahn, Young-Hee & Park, Ok-Hwa (2009). Digital curation strategy for university library service. *Journal of Korean Library and Information Science Society*, 40(4), 311-326. <https://doi.org/10.16981/kliss.40.4.200912.311>
- Bae, Na-yun & Oh, Hyo-Jung (2024). Analyzing domestic research trends on disclosure of information by comparing major academic disciplines. *Journal of the Korean Society for Information Management*, 41(2), 295-316. <https://doi.org/10.3743/KOSIM.2024.41.2.295>
- Gu, Jin Kyoung (2021, July 2). [ICT] Preparing for the Retail 4.0 Era: The E-commerce Market. Invest Korea. Available: https://www.investkorea.org/ik-kr/bbs/i-112/detail.do?ntt_sn=491181
- Ham, Seungkyung & Choi, Jimyeong (2023). Twitter as a sphere for anger, criticism, and political confrontation: An analysis of tweets regarding 10·29 Itaewon disaster using DMR topic model. *Crisisnomy*, 19(12), 59-75. <https://doi.org/10.14251/crisisnomy.2023.19.12.59>
- Han, Hui-Jeong, Kim, Tae-Young, Doo, Hyo-Chul, & Oh, Hyo-Jung (2017). Automatic extraction and usage of terminology dictionary based on definitional sentences patterns in technical documents. *Journal of the Korean Society for Information Management*, 34(4), 81-99.

- http://doi.org/10.3743/KOSIM.2017.34.4.081
- Han, Na-eun (2023). Proposal of a conceptual model for research data curation based on activity theory. *Journal of Korean Library and Information Science Society*, 54(1), 167-190.
http://doi.org/10.16981/kliss.54.1.202303.167
- Jang, Hyoseok (2022). A study on digital transformation of UAM(Urban Air Mobility). *Journal of Korean Aviation Security*, 1(4), 43-48.
- Jo, Min-ji (2020). Archives and information services on the digital memory era. *The Korean Journal of Archival, Information and Cultural Studies*, 10, 181-215.
https://doi.org/10.23035/kaics.2020.1.10.181
- Kim, Bo-Ram & Lee, JongWook (2022). A study on the data analysis and information visualization of cultural heritage: Focusing on the data of e-Museum served by the national museum of Korea. *Journal of Korean Traditional Culture*, 29, 115-153.
http://doi.org/10.23067/tjokch..29.202205.115
- Kim, Hyunsoo & Chung, Sungmo (2023). A comparative analysis of concepts and proposal for terminology establishment related to the use of space sharing terminology. *Journal of the Korea Institute of Spatial Design*, 18(1), 185-198. https://doi.org/10.35216/kisd.2023.18.1.185
- Kim, Pan Jun (2015). An analytical study on research trends of digital curation: Focused on library and information science. *Journal of the Korean Society for Information Management*, 32(1), 265-295. https://doi.org/10.3743/KOSIM.2015.32.1.265
- Kim, Yong-Hwan & Kim, Yushin (2019). Trend analysis of healthcare research in Korea using topic modeling. *Journal of the Korean Society for Wellness*, 14(1), 253-262.
http://doi.org/10.21097/ksw.2019.02.14.1.253
- Korea Institute of Science and Technology Information (2005). A Study on Developing National Digital Archiving Strategies.
- Lee, Eunkyung & Kim, Yeon-Jung (2023). A case study on the customer experience for fashion brand of online platform: Focusing on the online fashion brands expanding to offline store. *A Journal of Brand Design Association of Korea*, 21(3), 371-382.
https://doi.org/10.18852/bdak.2023.21.3.371
- Park, Minseok & Lee, Jisue (2024). A systematic review of trends of domestic digital curation Research. *Journal of Korean Society of Archival and Records Management*, 24(2), 41-63.
https://doi.org/10.14404/JKSARM.2024.24.2.041
- Park, Young-Uk & Chung, Kyu-Yub (2021). A study on the analysis of customer's sentiment using DMR(Dirichlet Multinomial Regression) topic modeling based on online review big

data: Focusing on the foreign customer's reviews of domestic 5 star deluxe hotels. Korean Journal of Hospitality & Tourism, 30(2), 1-20. <https://doi.org/10.24992/KJHT.2021.2.30.02.1>

Shin, Wonkyu & Kim, Hea-Jin (2023). A study on the trends of the development of national university project using topic modeling: Focusing on the comparative analysis of news articles and policy promotion tasks of the project. The Korean Educational Administration Society, 41(5), 193-215. <http://doi.org/10.22553/keas.2023.41.5.193>