

빅데이터 토픽모델링 및 네트워크 분석을 통한 문화콘텐츠학 지식구조 연구

오정심

[국문초록]

문화콘텐츠학은 2000년대 초반에 등장한 신생학문임에도 빠르게 성장해 왔다. 문화콘텐츠 관련 학술논문 수가 2000년에 100편에 불과했지만, 2020년에 누적 논문 수가 24,935편을 넘어섰다. 하지만 이러한 발전에도 불구하고, ‘문화콘텐츠를 독자적인 학문으로 인정할 수 있는가’라는 문제에 대해 논란이 계속되고 있다. 이러한 논란에 대해 선행연구에서는 문화콘텐츠학의 고유한 연구대상과 연구방법을 확립하지 못했기 때문에 계속되고 있는 것이라고 지적했다. 이러한 배경 아래 본 논문에서는 빅데이터 분석방법을 이용해 문화콘텐츠학의 연구대상 및 지식구조, 연구동향 등을 연구하였다.

2000년부터 2020년 최근까지 약 20년 동안 발간된 KCI학술지 논문 중에서 “문화콘텐츠”로 검색되는 논문 3,685편의 초록 및 서지정보 등을 텍스트네트워크분석과 토픽모델링을 이용해 분석했다. 텍스트네트워크분석을 통해 문화콘텐츠학의 주요 연구대상, 연구분야, 연구체계 등을 도출했으며, 토픽모델링을 통해 3,685편의 논문내용을 40개 토픽으로 요약, 분류하였다. 그리고 분석결과를 종합해 문화콘텐츠학의 주요 연구분야와 주제 분류안을 제시하였다. 주요 연구분야는 크게 ‘문화콘텐츠 활용’, ‘문화콘텐츠 산업’, ‘한국사회와 문화콘텐츠’, ‘문화콘텐츠 장르’, ‘문화콘텐츠 기술’, ‘문화콘텐츠 이론 및 체계’ 등 6개로 구분하였고, 40개 토픽을 각 분야에 맞게 분류하였다. 그리고 분석결과를 바탕으로 ‘문화콘텐츠 활용방법’을 문화콘텐츠학의 주요 연구방법론으로 제시하였다. 문화콘텐츠 활용방법에는 스토리텔링, 문학작품의 창작 소재화, 지역문화관광자원 활용, 정보콘텐츠 활용, 교육과 콘텐츠 활용 등이 있다. 이밖에도 문화콘텐츠 연구동향, 연구자 공동연구협력체계 등을 분석하였다.

본 논문에서 약 20년 동안 문화콘텐츠 분야에 축적되어 있었던 학술 빅데이터를 분석하여 문화콘텐츠학의 주요 연구대상 및 연구방법, 지식구조 등을 도출함으로써 문화콘텐츠의 학문적 체계와 위상을 정립하는 일의 토대를 제공했다는 점에서 연구의의를 찾을 수 있다.

[주제어] 문화콘텐츠학, 연구동향, 지식구조, 텍스트네트워크분석, 토픽모델링, 빅데이터 마이닝

*이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구 결과를 바탕으로 작성되었음
(NRF-2019S1A5B5A07089103).

투고일: 2020. 6. 2. 심사일: 2020. 7. 3. 게재 확정일: 2020. 7. 14.

<https://doi.org/10.16937/jcp.2020.34.2.35>

오정심_한국외국어대학교 연구산학협력단 책임연구원/교신저자(ruaths0802@naver.com)

I. 들어가며

문화콘텐츠학은 2000년대 초반에 등장한 신생학문임에도 빠르게 성장해 왔다. 특히 양적 측면에서 눈에 띄게 성장하였다. 문화콘텐츠 관련 학술논문 수가 2000년에 100편에 불과했지만 2019년에 누적 논문 수가 24,935편을 넘어섰다. 관련 학술지도 크게 증가해 2019년 한해에만 문화콘텐츠 관련 논문을 게재한 학술지는 100여 종에 달했다.¹⁾ 이러한 성장세는 앞으로 이어질 것으로 보인다. 문화콘텐츠가 4차 산업혁명의 성장 동력 중 하나로 꼽히면서 관련 정책과 연구가 계속되고 있기 때문이다(파이낸셜뉴스, 2019.9.17; 문화체육관광부 2020년 업무계획). 또한 이 분야 전문가와 연구 인력이 지속적으로 배출되고 있기 때문이다. 한국콘텐츠진흥원의 「콘텐츠 교육기관 및 인력 수요 현황 조사(2017)」에 따르면 콘텐츠 관련 고등교육기관의 학과 수가 2004년에 484개에서 2017년에 1,608개까지 증가한 것으로 나타났다.

하지만 이러한 발전에도 불구하고, ‘문화콘텐츠를 독자적인 학문으로 인정할 수 있는가’라는 문제에 대해 논란이 계속되고 있다.²⁾ 일찍이 이 문제를 연구한 박상천(2007b)은 사라세빅(Tefko Saracevic)과 리이즈(Alan M. Rees)의 이론을 바탕으로 문화콘텐츠학은 독자적인 학문으로 성립할 수 있는 당위성을 충분히 갖추고 있지만, 독자적인 학문으로 인식되지 못한 까닭은 고유한 연구대상과 연구방법을 확립하지 못했

1) 학술연구정보서비스(www.riss.kr)에서 2000년부터 2020년 4월 최근까지 키워드 ‘문화콘텐츠’로 검색한 결과 관련 논문이 총 25,635편 검색되었다.

2) 이러한 논란은 최근까지도 이어지고 있다. 「문화콘텐츠학, 어디로 가는가?(유제상, 2017)」, 「진화론적 관점으로 본 문화콘텐츠의 발생(최혜경, 2018)」, 「문화콘텐츠학 정립을 위한 놀이개념에 대한 탐색적 연구(김기정, 2019)」 등 관련 연구가 지속되고 있기 때문이다.

기 때문이라고 지적했다.³⁾ 임대근(2015)은 문화콘텐츠학 위상에 대한 논란은 학문적 체계를 구축하는 일을 소홀히 했기 때문에 생겨난 것이라고 주장했다. 다시 말해 문화콘텐츠가 산업적·기술적·정책적 필요에 따라 등장했기 때문에 학문적 체계에 관한 일을 중요하게 여기지 않았다고 지적했다.

이러한 배경 아래 연구자들은 문화콘텐츠학의 학문적 체계를 세우기 위한 노력을 해왔다. 그중에서 류준호·윤승금·이영주(2010)와 윤흥근(2011)은 메타분석방법을 이용해 문화콘텐츠학의 연구대상과 연구방법 등을 파악하고자 했다. 메타분석이란 문헌에서 얻은 정보를 기호로 코딩하고, 이것을 기술통계기법을 이용해 분석하는 방법을 말한다(네이버지식백과사전). 선행연구들은 기존의 정성적 분석에서 벗어나 메타분석이라는 통계적 방법을 이용해 문화콘텐츠 연구의 특징을 분석했다는 점에서 의의가 있다. 그럼에도 다음과 같은 점에서 한계가 지적된다. 일반적으로 메타분석을 하기 위해서 연구자는 논문내용을 일별하고, 정해진 분석항목에 따라 분류해야 한다. 이러한 방식 때문에 많은 시간과 노동력이 요구되고, 연구자의 능력에 따라 연구 편차가 발생할 수 있다. 그래서 선행연구에서는 논문 수를 100편 내외로 선별하였고, 이 과정에서 상당수의 논문이 누락되었다. 또한 메타분석은 빈도분석을 통해 특정 대상이 얼마나 자주 등장하는지를 분석할 수 있지만, 그것이 어떤 맥락에 있는지를 파악하기 어렵다.

이에 본 논문에서는 선행연구의 한계를 극복하기 위해 ‘토픽모델링(topic modeling)’과 ‘텍스트네트워크분석(text network analysis)’을 이용해 연구대상에 접근하고자 한다. 이 방법들은 대표적인 빅데이터 분석기법으로, 대규모 비정형 데이터에서 텍스트가 가진 주요한 의미를 분석, 추출해준다. 토픽모델링은 텍스트 집합에서 토픽을 발굴하고, 토픽에 따라 문서를 요약, 분류해 주는 기계학습 알고리즘이다(윤효준·박재현·윤지운, 2019). 텍스트네트워크분석은 텍스트에서 단어들을 추출하고, 단어들의 관계를 계량적으로 분석하여 텍스트의 맥락과 의미구조를 나타내준다(한국정보화진흥

3) 사라세빅과 리이즈가 제시한 독자적 학문의 성립 조건들은 매우 구체적이고 현실적인 조건들인데, 문화콘텐츠학은 그 조건을 모두 충족하고 있다. 그 내용을 살펴보면 첫째, 문화콘텐츠는 행정 기관, 교육 기관은 물론이고 언론 기관을 통해 대중들에게까지도 인식되어 일반화 되어 있다. 둘째, 대학의 많은 연구자들이나 국가연구기관 등에 소속된 연구자들이 연구에 적극 참여하고 있을 뿐 아니라, 이미 대학원 교육을 통해 학문의 후속 세대들이 성장하고 있다. 셋째, ‘스토리텔링’이라는 동일한 연구 대상에 대하여서도 문학연구자의 연구 방법과 문화콘텐츠 연구자의 연구 방법은 분명한 차별성을 가지고 있다. 또한 기술 분야에서도 기존의 연구와는 차별화된 연구를 진행하려는 노력들도 있어 왔다. 넷째, 대학 등 그 분야에 관심 있는 사람들을 위하여 정식교육을 제공하고 있다. 다섯째, 공식·비공식 커뮤니케이션 채널이 구성되어 있다. 여섯째, 공고한 학회가 있고 학술지가 발간되고 있다(박상천, 2007b).

원, 2015). 최근에 다양한 분야에서 토픽모델링과 텍스트네트워크분석을 이용한 연구가 증가하는 추세지만, 문화콘텐츠 분야에서는 관련 사례를 찾아보기 어렵다.

본 논문의 연구목적은 빅데이터 분석기법을 활용해 문화콘텐츠 분야의 연구동향 및 지식구조를 분석하는 데 있다. 나아가 연구결과를 바탕으로 문화콘텐츠 분야의 미래 방향을 제시하고, 학문적 체계를 정립하는 일에 기여하고자 한다. 이러한 목적을 이루기 위해 본 논문에서는 문화콘텐츠 관련 연구가 등장하기 시작한 2000년부터 2020년 최근까지 약 20년 동안 발간된 관련 논문의 초록과 서지정보 등을 총망라하여 수집하고, 토픽모델링과 텍스트네트워크분석을 이용해 분석할 것이다. 그리고 구체적인 연구결과를 얻기 위해서 다음과 같은 연구문제를 가지고 연구를 진행할 것이다.

연구문제 1: 문화콘텐츠학 주요 연구대상 및 연구분야 관련

- 1-1. 2000년부터 2020년 최근까지 문화콘텐츠 분야 주요 연구대상은 무엇인가?
- 1-2. 2000년부터 2020년 최근까지 발간된 문화콘텐츠 관련 학술논문들의 주요 주제는 무엇인가?
- 1-3. 위 결과를 바탕으로 문화콘텐츠학의 주요 연구분야와 주제는 어떻게 분류되는가?
- 1-4. 위 결과를 지식지도 형태로 시각화 하면 어떤 특징이 나타나는가?

연구문제 2: 문화콘텐츠학 연구동향 관련

- 2-1. 2000년부터 2020년 최근까지 시기별로 문화콘텐츠 분야 주요 연구대상 등은 어떻게 변화, 발전했는가?
- 2-2. 위 결과를 지식지도 형태로 시각화 하면 어떤 특징이 나타나는가?

연구문제 3: 문화콘텐츠 분야 연구자 공동연구협력체계 관련

- 3-1. 문화콘텐츠 분야 연구자들의 공동연구협력망은 어떻게 구성되어 있는가?
- 3-2. 문화콘텐츠 분야 연구자들의 학제간연구 현황은 어떠한가?

이러한 목적 아래 연구가 진행된다면 다음과 같은 효과를 기대할 수 있을 것이다. 첫째, 본 논문의 연구방법론을 통해 문화콘텐츠 분야에서 빅데이터의 활용이 활성화되는

데 기여할 것이다. 둘째, 문화콘텐츠 분야의 연구대상 및 연구분야를 지식지도 형태로 나타냄으로써 추상적인 내용에 대한 직관적 해석이 가능하게 될 것이다. 셋째, 빅데이터 분석을 통해 도출된 내용은 문화콘텐츠의 학문적 체계를 정립하는 일에 기여할 것이다.

II. 이론적 배경

1. 토픽모델링 및 텍스트네트워크분석 이해

지금까지 텍스트를 분석하기 위해서 정성적 분석과 메타분석방법 등이 주로 사용되었다. 전술한 바와 같이 메타분석은 연구자의 수준에 따라 코딩 결과가 다를 수 있고, 데이터양이 많아지면 시간과 비용이 소요될 수 있는 단점이 있다(이수상, 2014). 이러한 단점 때문에 최근에 연구자들 사이에서 빅데이터 분석기법을 기반으로 한 토픽모델링과 텍스트네트워크분석이 선호되고 있다. 특히 미래학 분야에서 이 두 방법을 결합하여 미래 이슈들의 상호연관성이나 시간상 변화과정을 추적하는 연구까지 활용되고 있어 새로운 연구방법으로 각광받고 있다(한국정보화진흥원, 2015).

텍스트네트워크분석은 텍스트에서 단어들을 추출하고, 단어들의 관계를 계량적으로 분석하여 텍스트의 의미구조와 맥락 등을 나타내는 방법이다. 이 방법은 사회학의 소셜 네트워크 이론을 바탕으로 빅데이터 분석기술의 발달과 함께 발전하였다(한국정보화진흥원, 2015). 텍스트네트워크분석의 주요 지표에는 밀도, 빈도, 중심성 등이 있는데, <표 1>은 본 논문에서 주로 활용한 지표에 대한 설명내용이다. 텍스트네트워크분석의 가장 큰 장점은 텍스트 개념들의 관계 구조를 공간적으로 나타내 줌으로써 텍스트에 대한 새로운 해석 가능성을 열어준다는 것이다(박치성 · 정지원, 2013).

토픽모델링이란 여러 문서들로 이루어진 텍스트 집합에서 특정 문서에 자주 등장하는 단어들을 묶어 토픽으로 추출해 주는 확률적 알고리즘이다. 연구자들은 단어묶음으로 이루어진 토픽을 통해 문서가 다루고 있는 내용을 유추할 수 있다. 토픽모델링은 인공지능 기계학습 원리로 이루어진다. 예를 들어 김치에 관한 글에서는 배추, 한국, 맛 등과 같은 단어가 자주 나올 가능성이 크다. 인공지능이 여러 문서들을 학습해서 배추, 한국, 맛과 같은 단어들을 하나의 단어묶음으로 제시하면 연구자는 이를 보고 김치에 관한

〈표1〉 텍스트네트워크분석 주요 지표

구분	분석지표	설명	해석 및 활용 방향
일반 특징	밀도 (Density)	· 연결 가능한 총 링크 수와 실제로 맺어진 링크 수의 비율	-단어들 사이 연관성은 어떠한가?
	빈도 (Frequency)	· 특정 범위 내 노드의 동시 출현 횟수	-다수의 논문에 동시 등장한 핵심어는 무엇인가?
중심 구조	연결중심성 (Degree centrality)	· 서로 연결되어 있는 이웃노드 개수 합 · 연결된 노드 수가 많을수록 값이 크게 측정됨 · 연결중심성 큰 노드: 마당발, 인기도로 해석	-주요 연구대상은 무엇인가? -연구자협력망에서 마당발, 권력자는 누구인가?
	위세중심성 (Eigenvector centrality)	· 노드의 연결성과 중요도 측정 · 위세중심성 큰 노드: 영향력이 큰 노드로 해석	-연구내용 구성에 가장 중요한 영향을 끼친 대상 은 무엇인가? -연구자협력망에서 영향력이 큰 사람은 누구인가?
하위 그룹 분류	컴포넌트 (Component)	· 응집성 기준으로 노드 클러스터링 · 연결고리가 끊기지 않고 하나로 연결된 노드 분류	-연구자협력망에서 하위 그룹이 있는가? 그 특 징은 무엇인가?
	결속집단 (Clique)	· 모든 노드들이 하나로 완벽하게 연결된 집단을 말함 ※ 결속집단의 밀도= '1'	-간밀하게 구성된 하위 그룹이 있는가? 특징은 무엇인가?

자료: 오정심, 2020:97 〈표1〉 재구성.

문서임을 유추할 수 있다. 토픽모델링은 사람이 직접 읽어도 되는 규모의 문서라면 필요가 없겠지만 사람이 감당하기 어려운 규모의 문서에서 무슨 내용이 다뤄졌는지를 알아보는 데 유용하다(윤호준 · 박재현 · 윤지운, 2019; 권오성, 2019.08.16).

토픽모델링과 텍스트네트워크분석을 하기 위해서는 가장 먼저 ‘데이터 전처리’를 해줘야 한다. 데이터 전처리는 자연어 알고리즘을 이용해 비정형 데이터를 컴퓨터 분석 가능한 정형화된 구조로 만드는 것을 말한다. 논문 초록과 서지정보 등은 일종의 비정형 데이터이다. 비정형 데이터는 유형이 불규칙하고, 의미하는 바가 명확하지 않기 때문에 컴퓨터 분석이 어렵다. 그래서 형태소 분석기 등 자연어 알고리즘을 이용해 비정형 텍스트를 해체하여 정형화된 구조로 만들어 줘야 한다. 연구자는 이 과정에서 유의어, 지정어, 제외어 등 시소러스 사전을 만들어 데이터를 정제할 수 있다. 유의어 사전은 뜻이 비슷한 말을 하나로 통일해 주는 기능이다. 지정어 사전은 분석에 반드시 포함되어야 할 단어나 고유명사를 형태소 분절 없이 그대로 추출해 주는 기능이다. 제외어 사전은 점, 앞, 후 등 의미를 파악하기 어려운 1글자 단어와 본고, 목적, 결과 등 논문 초록에 일반적으로 사용하는 단어를 제거해 주는 기능이다(오정심, 2020; 김세현, 2018).

2. 선행연구 검토

문화콘텐츠학을 둘러싼 논란을 해결하기 위해 관련 텍스트 자료를 수집, 분석해 문화콘텐츠학의 특징을 밝힌 선행연구로는 류준호 · 윤승금 · 이영주(2010) 연구 등 4편이 있다. <표 2>는 선행연구 현황을 정리한 것이다. 앞서 살펴 것을 제외하고 나머지 연구들을 살펴보자. 황동열 · 황고은 연구(2016)와 황서이 · 박정배 · 김문기 연구(2020)에서는 최근 활발히 사용되고 있는 토픽모델링과 의미연결망분석을 이용하여 인문콘텐츠학회지 논문 초록을 분석하였다. 텍스트네트워크분석은 의미연결망분석 또는 언어네트워크분석으로 불리기도 한다. 의미연결망분석은 텍스트에 내재된 의미론적 속성 분석을 강조하는 표현이라 할 수 있다(이수상, 2014).

<표 2> 관련 선행연구 현황

발간연도	논문제목 및 저자명	주요 내용 및 연구방법
2010	문화콘텐츠 관련 연구에 대한 메타분석(류준호 · 윤승금 · 이영주)	연구방법: 메타분석, 주요 내용: 문화콘텐츠 관련 논문 129편, 학문적 배경·연구목적·연구분야 · 연구방법 등 분류, 코딩, 기술통계분석
2011	학위논문의 메타분석을 통한 문화콘텐츠 연구동향(윤홍근)	연구방법: 메타분석, 주요 내용: 문화콘텐츠 주요 5개 대학 학위논문 101편, 연구분야·연구목적 · 연구방법 등 분류, 코딩, 기술통계분석
2016	빅데이터 기술을 활용한 인문콘텐츠 분야의 의미연결망 분석(황동열 · 황고은)	연구방법: 의미연결망분석, 주요 내용: 인문콘텐츠 게재 논문 510편 국문초록 수집, R프로그램 패키지 이용 출현빈도, 중심성, 군집화 등 분석
2020	인문콘텐츠분야 연구의 경향 분석: 토픽모델링과 의미 연결망 분석을 중심으로(황서이 · 박정배 · 김문기)	연구방법: 의미연결망분석, 토픽모델링, 주요 내용: 인문콘텐츠 게재 논문 622편 국문초록 수집, R프로그램 패키지 이용, 주요 토픽 15개 추출 등

황동열 · 황고은(2016)은 인문콘텐츠학회지의 의미 구성체계를 파악하기 위해 2003년부터 2015년까지 학회지에 게재된 논문 510편의 초록을 대상으로 R프로그램 패키지를 이용해 의미연결망분석을 하였다. 황서이 · 박정배 · 김문기(2020)도 같은 프로그램을 이용해 인문콘텐츠학회지 논문 622편의 초록을 수집해 분석했다. 앞선 연구와 다른 점이 있다면 수집 기간을 2003년부터 2018년까지 설정하여 최근 자료에 대한 분석이 이뤄졌고, 토픽모델링 기법을 추가했다는 것이다. 최근 들어 빅데이터 분석방법이 최적화된 분석기법을 여러 가지 사용하여 분석결과에 대한 신뢰도를 높이는 쪽으로 발전하고 있다(오정심, 2020; 한국정보화진흥원, 2015). 그래서 최적화된 방법을 여

러 가지 함께 사용하는 것이 필요하다.

지금까지 살펴본 선행연구들은 본 논문에서 활용한 토픽모델링과 텍스트네트워크 분석을 사용했기 때문에 본 논문에 시사한 바가 있다. 그리고 인문콘텐츠학회지는 문화콘텐츠학이 등장하기 시작한 시점에 발간되어 문화콘텐츠 개념이 정립되는 데 중요한 영향을 끼쳤다. 그래서 선행연구의 연구결과는 본 연구에서 분석결과를 해석할 때 중요한 참고자료가 될 수 있다. 그럼에도 이 연구들은 다음과 같은 점에서 한계가 지적된다. 첫째, 분석결과를 종합하여 인문콘텐츠학회지의 주요 연구분야와 주제를 도출하는 데 까지 연구를 발전시킬 수 있었지만, 분석결과를 제시하는 수준으로 연구가 그치고 말았다. 둘째, 토픽모델링을 텍스트 문맥과 복잡한 단어 표현 등을 고려해 분석하는 알고리즘 LDA, Wordvec 등으로 하지 않아 분석결과에 한계가 있다. 셋째, R프로그램 패키지는 오픈 소스 프로그램이지만 기본적으로 기술통계를 모르는 사람들은 프로그램을 다루기가 어렵고, 심도 깊은 분석까지 하기가 힘들다. 이에 본 논문에서는 선행연구 결과를 보완, 발전시켜 다음과 같이 연구를 하였다.

Ⅲ. 연구 대상 및 방법

1. 연구대상

본 논문의 목적은 토픽모델링과 텍스트네트워크분석을 이용해 문화콘텐츠 분야의 연구동향 및 지식구조를 분석하고, 나아가 문화콘텐츠학의 학문적 체계를 정립하는 일에 기여하는 데 있다. 이러한 목적 아래 관련 데이터는 한국학술지인용색인⁴⁾에서 수집하였다. 데이터 검색어는 “문화콘텐츠”⁵⁾로 선정하였다. 데이터 수집기간은 문화콘텐츠학이 등장하기 시작한 2000년부터 2020년 4월 최근까지 설정하였다. 이렇게 해서 1차로 검색된 자료들 중에서 국문초록이 없는 것, 발표문, 서평 등을 제외하여 총 3,685편을 최종 연구대상으로 선정하였다.⁶⁾

4) 한국학술지인용색인(www.kci.go.kr) 사이트는 한국연구재단의 KCI(Korea Citation Index) 등재(후보)학술지 논문에 대한 각종 정보를 제공해 준다.

5) 큰 따옴표(“”)안에 문화콘텐츠를 넣어 검색하면 논문명, 주제어, 초록, 목차 등에 문화콘텐츠가 반드시 포함된 자료만 검색해 준다.

6) 학술연구정보서비스에서 검색한 결과 관련 학술논문이 총 25,635편이 있는 것으로 조사됐지만, KCI등재(후보) 논문이면서 국문초록 및 서지정보가 있는 것만을 분석대상으로 선정하였다.

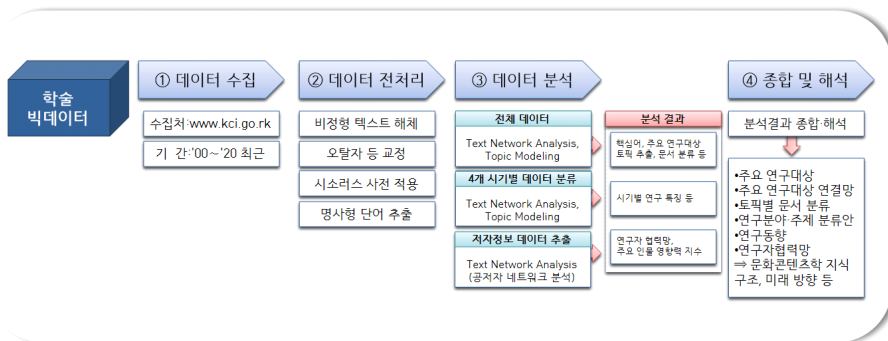
〈표 3〉 자료수집방법 및 연구대상

구분	데이터 수집 조건
검색어	“문화콘텐츠”
수집처	한국학술지인용색인(www.kci.go.kr)
수집기간	2000년~2020년 4월 최근
필터링	KCI학술지논문, 초록 없는 것 · 단순발표문 · 보고서 · 서평 등 제외
▼	
최종 연구대상	“문화콘텐츠” 관련 KCI학술논문 초록 및 서지정보 관련 논문: 총 3,685편 / 관련 논문 게재 KCI학술지: 총 516종 /추출 단어: 18,027개

2. 연구방법

연구는 크게 ‘데이터 수집’, ‘데이터 전처리’, ‘데이터 분석’, ‘종합 및 해석’ 4단계로 진행하였다. [그림 1]은 연구 단계 및 내용을 그림으로 나타낸 것이다. 데이터 수집과 분석은 넷마이너(NetMiner 4.4.3.b, Topic Modeling Plug-In)를 이용해 하였다. 이 프로그램은 국내에서 개발된 네트워크분석 전문 소프트웨어이다. 대용량 데이터에 최적화된 알고리즘이 구현되어 있고, 분석결과를 시각화할 수 있는 여러 가지 틀을 제공하고 있다.

[그림 1] 연구 단계 및 내용



본격적인 분석에 앞서 가장 먼저 ‘데이터 전처리’ 작업을 하였다. 수집된 논문 초록 및 서지정보 등은 비정형 데이터이기 때문에 컴퓨터 분석이 가능하도록 정형화된 구조로 만든 것이다. 본 논문에서는 넷마이너에 탑재된 한국어 형태소 분석기를 이용해 비정형 텍스트를 형태소 단위로 분해하고 명사형 단어를 추출하였다. 그리고 추출된 단어들

을 대상으로 오타자, 띄어쓰기 등 교정 작업을 하였다. 이어서 유의어, 지정어, 제외어 시소러스 사전을 개발, 적용하여 데이터 정제를 하였다. <표 4>는 본 논문에서 적용한 시소러스 사전의 예를 제시한 것이다.

<표 4> 시소러스 사전 예시

유의어 사전	지정어 사전	제외어 사전
콘텐츠, 컨텐츠	문화콘텐츠	연구
박물관, 뮤지엄	4차산업혁명	제시
CT, 문화기술	스토리텔링	탐색
K-POP, 케이팝	다문화	소고
한스타일, 韓스타일	Glocalization	때, 점, 전, 후 등 1글자 단어

문화콘텐츠 관련 학술논문 3,685편의 초록 및 서지정보 데이터를 전처리 작업을 한 결과, 18,027개 명사형 단어가 추출되었다. 이 단어들을 대상으로 텍스트네트워크분석과 토픽모델링을 차례대로 하였다. 텍스트네트워크분석의 지표로 ‘밀도(density)’, ‘빈도(frequency)’, ‘연결중심성(degree centrality)’, ‘위세중심성(eigenvector centrality)’으로 선정하였다. 밀도는 네트워크 내에서 연결 가능한 총 링크 수와 실제로 맺어진 링크 수의 비율로 측정된다. 네트워크 내에서 노드 간 연결이 많을수록 밀도가 높게 나타난다. 밀도가 높다는 것은 텍스트 단어들 사이의 연관성이 높다는 것을 의미한다. 그리고 빈도는 노드들의 동시 출현 횟수를 계산한 값이다. 빈도가 높은 단어일수록 다수의 논문에 동시 등장한 핵심어로 볼 수 있다. 또는 특정한 시기에 저자들이 공통적으로 중요하게 생각했던 연구대상으로 해석할 수 있다. 마지막으로 중심성은 노드가 네트워크의 중심구조에 위치하는 정도를 나타내는 값으로 연결중심성, 위세중심성 등이 있다. 연결중심성은 연결되어 있는 이웃노드들의 총 개수로 측정된다. 연결중심성이 높은 단어일수록 연구 논의에 중심이 되는 대상일 가능성이 크다. 그리고 이 값이 크게 나온 노드는 연결된 대상들이 많기 때문에 이 노드를 제거하면 연구내용의 구성이 어렵게 된다. 위세중심성은 노드의 연결성과 영향력까지 고려한 값이다. 위세중심성이 큰 단어일수록 연구내용 구성에 중요한 영향력을 끼쳤다고 볼 수 있다(Freeman, 2004; 박준석 · 김창식 · 광기영, 2016; 오정심, 2017).

토픽모델링은 기계학습 잠재디리클레할당(Latent Dirichlet Allocation, LDA)

알고리즘을 적용하여 분석하였다.⁷⁾ 토픽모델링을 하기 위해서 가장 먼저 토픽 수를 정해야 한다. 토픽 수에 따라 분석결과가 다를 수 있기 때문이다. 토픽 수를 결정하는 방법에는 혼잡도(perplexity), 일관성(coherence score) 등 여러 가지가 있지만, 본 논문에서는 토픽 수별로 결과를 추출하고, 해석 용이성과 타당도를 기준으로 내용을 판별하여 토픽 수를 결정하는 방식으로 하였다. 이는 최근에 토픽 수를 결정하는 이상적 모델은 사실상 존재하지 않는다는 주장이 설득적이기 때문이다. 이 주장에 따르면 토픽모델링은 분류되지 않은 문서 집합에서 소수의 토픽에 분류시킴으로써 나머지 토픽들을 좀 더 해석 가능한 것으로 만들어 주기 때문에 전체 토픽을 대상으로 평가하여 최적의 모델을 선택하는 것은 사실상 무의미하다는 것이다(윤효준 외, 2019). 마지막으로 분석결과를 직관적으로 파악하기 위해 스프링형 네트워크맵(spring map), 워드 클라우드(word cloud) 등 형태로 시각화하여 검토하였다.

IV. 연구결과

1. 문화콘텐츠학 주요 연구대상 및 분야 분류

1) 텍스트네트워크분석

2000년부터 2020년 최근까지 문화콘텐츠 관련 학술논문 3,685편의 국문초록과 서지정보 데이터를 전처리 작업 한 결과, 18,027개 형태소가 추출되었다. 이 단어들과 문서들의 연결 관계를 이원모드 네트워크(2-mode network)로 형성한 후에 TF-IDF 분석을 하였다. TF-IDF(term frequency-inverse document frequency)는 특정 단어의 문서 내 중요도를 측정한 값이다.⁸⁾ 관용적으로 쓰는 단어는 이 수치가 낮게 나타나고, 핵심어는 이 수치가 높게 나타난다. 분석결과, TF-IDF가 가장 낮게 나온 단어는 문화(0.1), 콘텐츠(0.19), 문화콘텐츠(0.19)로 나타났다. 이는 검색어를 “문화콘텐츠”로 설정했기 때문에 문화콘텐츠 키워드가 반드시 포함된 논문들만 수집, 분석해서 나타난 결과로 보인다.

7) MCMC(Learning Method): alpha=2.0, beta=0.1, #of iterations=100, burn-in=10.0, sample lag=1.0.

8) 단어 빈도수와 문서 빈도수의 역수를 곱한 값이다. 일반적으로 TF-IDF 값이 0.1인 단어들은 중요도가 낮다고 판단하여 분석대상에서 제외한다.

[그림 2] 문서 동시 등장 빈도수 워드 클라우드



다음으로 문서 동시 등장 빈도수를 분석하였다. 이 값은 특정 단어가 일정한 범위에서 동시에 얼마나 자주 등장했는지를 계산한 것이다. 분석결과, 18,027개 단어들 중에서 검색어인 문화콘텐츠를 제외하고 빈도수가 가장 높게 나타난 단어는 활용(1,470회)이었다. 이어서 한국(1,077회), 사회(1,068회), 가치(1,042회), 산업(1,028회) 순으로 나타났다. ‘활용’, ‘가치’, ‘산업’, ‘한국사회’ 등은 문화콘텐츠 관련 다수의 논문들에 가장 많이 기술된 핵심연구대상으로 볼 수 있다. [그림2]는 빈도수 분석결과를 워드 클라우드(word cloud)로 그린 것이다.

이어서 연결중심성과 위세중심성 분석을 하였다. 연결중심성이 높은 단어일수록 연구논의에 중심이 되는 대상으로 볼 수 있다. 그리고 위세중심성은 연결성에 영향력까지 고려한 것으로 위세중심성이 높은 단어일수록 연구 내용에 중요한 영향력을 끼친 대상으로 볼 수 있다. 따라서 빈도수가 높으면서 연결중심성과 위세중심성까지 높은 노드는 해당 학문 분야에서 연구내용 구성에 중요한 영향을 끼친 대상이라 할 수 있다. 다시 말해 해당 학문 분야를 논할 때 필수불가결한 영역이라 할 수 있다.

연결중심성 분석결과, 검색어인 문화콘텐츠를 제외하고 활용(0.1338)이 가장 높게 나타났다. 그 다음으로 가치(0.1203), 사회(0.1202), 한국(0.1182), 산업(0.1156) 순으로 나타났다. 위세중심성 분석결과도 이와 비슷하게 나타났다. 활용(0.0936)이 가장 높게 나타났고, 이어서 가치(0.0847), 사회(0.0830), 산업(0.0818) 순으로 나타났다. 요컨대 문화콘텐츠 ‘활용’, ‘가치’, ‘사회’, ‘산업’은 빈도수와 중심성 분석에서 모두 높게

나타나 문화콘텐츠 관련 논의에 중심이 되는 연구대상이었으며, 문화콘텐츠학을 구성하는 필수 영역이라 할 수 있다.

다음으로 ‘문화콘텐츠를 어떤 측면에서 바라볼 것인가’에 대한 문제 해결의 실마리를 찾기 위해 관련 키워드를 찾아 빈도수·연결중심성·위세중심성 분석결과를 비교하였다. <표 5>에서 보듯이 연구자들은 산업적 측면(산업·경제·시장·상품)에서 문화콘텐츠를 가장 많이 다룬 것으로 분석됐다. 그 다음으로 인문학적 측면(역사·사람·전통), 기술적 측면(기술·제작·디지털·정보·미디어) 순으로 나타났다. 반면, 정책 및 제도적 측면과 관련한 노드들의 빈도수·중심성 값이 낮게 나타나, 이와 관련한 연구는 적었던 것으로 분석됐다. 그리고 문화콘텐츠 장르 중에서는 영화와 관련한 연구가 가장 많았던 것으로 나타났다.

마지막으로 문화콘텐츠학의 주요 연구분야 및 연구주제를 알아보기 위해 핵심어 연결망을 그렸다. 18,027개 단어들 중에서 논문 300편 이상에 동시 등장한 핵심어 110개를 추출하였고, 이를 스프링형 네트워크 맵(PFNet⁹⁾ spring map)으로 그렸다.¹⁰⁾ 그림을 보면서 중심성 값이 큰 단어들과 직접적으로 연결되어 있으면서 연결선 굵기가 굵은 노드들을 배열하면 문화콘텐츠학의 주요 연구분야를 파악할 수 있다.

[그림 3]에서 보듯이 문화콘텐츠 연구분야는 크게 ‘문화콘텐츠 활용’, ‘문화콘텐츠 산업’, ‘한국사회와 문화콘텐츠’로 구분할 수 있다. 각 분야의 연구주제는 ‘문화콘텐츠 활용’ 분야에 지역 활성화와 문화콘텐츠 활용, 지역관광자원 활용, 디지털(융복합)기술 활용, 스토리텔링 활용, 교육과 문화콘텐츠 활용, 문화콘텐츠 가치 활용 등이 있다. ‘문화콘텐츠 산업’ 분야에 문화콘텐츠 산업 지원정책, 문화콘텐츠 산업과 국내 시장, 문화콘텐츠 산업의 지속적 발전, 문화콘텐츠 산업과 가치 창출 등이 있다. ‘한국사회와 문화콘텐츠’ 분야에 세계화와 한국사회, 시대 및 사회 변화와 영향, 현대 및 전통 사회와 콘텐츠, 사회적 존재 사람, 시대 변화와 문화콘텐츠 가치 등이 있다.

9) 넷마이너 PFNet은 네트워크에서 링크 수가 많을 때 노드들을 그대로 유지하면서 중요한 링크만 남겨놓고 싶을 때 사용하는 기능이다.

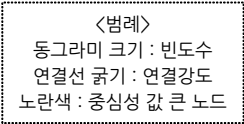
10) 검색어인 ‘문화’, ‘콘텐츠’ 노드를 제외하고 그림을 그렸다. ‘문화’, ‘콘텐츠’의 빈도수·중심성 값이 압도적으로 높기 때문에 이 노드들을 포함해 그림을 그리면 이 노드들과 직접적으로 연결된 노드들만 그림에 나타나고 정작 분석에 필요한 노드들은 나타나지 않는다.

〈표5〉 텍스트네트워크분석 빈도수 · 중심성 분석결과

빈도수		연결중심성		위세중심성	
단어	측정값 (순위)	단어	측정값 (순위)	단어	측정값 (순위)
문화	2,007 (1위)	문화	0.159496 (1위)	문화	0.111505 (1위)
문화콘텐츠	1,903 (2위)	문화콘텐츠	0.153403 (2위)	문화콘텐츠	0.107459 (2위)
콘텐츠	1,676 (3위)	콘텐츠	0.136390 (3위)	콘텐츠	0.094818 (3위)
활용	1,470 (4위)	활용	0.133802 (4위)	활용	0.093668 (4위)
한국	1,077 (5위)	가치	0.120375 (5위)	가치	0.084757 (5위)
사회	1,068 (6위)	사회	0.120258 (6위)	사회	0.083007 (6위)
가치	1,042 (7위)	한국	0.118234 (7위)	산업	0.081887 (7위)
산업	1,028 (8위)	산업	0.115690 (8위)	한국	0.081486 (8위)
지역	902 (9위)	시대	0.110896 (9위)	시대	0.077872 (9위)
시대	870 (10위)	지역	0.106174 (10위)	지역	0.074886 (10위)
...
역사	709 (17위)	역사	0.098792 (17위)	역사	0.069230 (16위)
교육	704 (18위)	사람(인간)	0.096673 (18위)	사람(인간)	0.066680 (18위)
기술	694 (20위)	교육	0.094982 (19위)	교육	0.065523 (19위)
사람(인간)	665 (23위)	기술	0.094422 (20위)	기술	0.065514 (20위)
제작	615 (27위)	전통	0.088007 (31위)	전통	0.061819 (30위)
디지털	586 (32위)	제작	0.087558 (32위)	제작	0.060668 (33위)
영화	558 (34위)	영화	0.086002 (35위)	경제	0.059914 (36위)
전통	543 (36위)	경제	0.084732 (38위)	디지털	0.059812 (37위)
정보	517 (39위)	디지털	0.084411 (39위)	영화	0.059206 (41위)
경제	512 (40위)	융복합	0.082821 (42위)	융복합	0.058463 (42위)
스토리텔링	508 (42위)	스토리텔링	0.081073 (47위)	스토리텔링	0.057715 (43위)
융복합	488 (46위)	정보	0.077401 (55위)	정책	0.053587 (57위)
정책	457 (53위)	존재	0.077027 (58위)	시장	0.053537 (58위)
시장	445 (57위)	시장	0.076946 (59위)	매체	0.053416 (61위)
매체	401 (69위)	정책	0.076686 (61위)	정보	0.053340 (64위)
존재	396 (73위)	매체	0.076209 (68위)	존재	0.052693 (70위)
영상	388 (75위)	기획	0.074778 (70위)	기획	0.052394 (72위)

※순위: 분석결과에 대한 직관적 파악을 위해 단어별로 1위부터 18,027위까지 측정값 순위를 부여하였다. 70위권까지 핵심 연구대상으로 해석할 수 있다.

[그림 3] 문화콘텐츠 분야 주요 연구대상 네트워크맵(PFNet spring map)



지금까지 텍스트네트워크분석을 통해 문화콘텐츠학의 주요 연구대상과 연구분야를 살펴보았다. 주지하듯이 텍스트네트워크분석은 핵심어들의 연결 관계를 시각화하는 데 장점이 있지만, 전체 문서 내용을 파악하는 데는 한계가 있다. 그래서 문화콘텐츠 관련 논문 3,685편의 주요 내용을 파악하기 위해 토픽모델링을 하였다.

2) 토픽모델링

문화콘텐츠 관련 학술논문 3,685편에서 데이터 전처리 작업을 통해 도출된 18,027개 단어들을 대상으로 LDA 토픽모델링을 하였다. 분석결과, 40개 토픽이 추출되었다. 3,685편 논문의 내용을 40개 토픽으로 요약, 분류한 것이다. 토픽 첫 번째 부분에서 단

어묵음이 ‘지역, 활성화, 문화콘텐츠, 연계, 자원’으로 나타나 토픽명을 ‘지역 활성화와 문화콘텐츠(지역자원 연계 콘텐츠)’로 하였다. 관련 논문 수는 전체 논문의 5.6%인 205편이었다. 두 번째 부분에서 단어묵음이 ‘산업, 성장, 발전, 경제, 방향’으로 나타나 토픽명을 ‘콘텐츠산업 성장과 경제발전 방향’으로 하였으며, 관련 논문 수는 155편(4.2%)이었다. 세 번째 부분에서 단어묵음이 ‘작품, 문학, 창작, 소재, 시대’으로 나타나 토픽명을 ‘문학작품 창작 소재화’로 하였고, 관련 논문 수는 153편(4.2%)이었다. 네 번째 부분에서 단어묵음이 ‘스토리텔링, 서사, 구조, 확장, 현재’으로 나타나 토픽명을 ‘스토리텔링과 이야기 서사구조’로 하였다. 관련 논문은 전체 논문의 4%인 147편이었다. 다섯 번째 부분에서 단어묵음이 ‘콘텐츠, 정보, 활용, 영향, 구성’으로 나타나 토픽명을 ‘정보콘텐츠 활용’으로 하였으며, 관련 논문은 137편(3.7%)이었다. 이런 식으로 토픽명을 정해 <표 6>과 같이 정리하였다.¹¹⁾

<표 6> 토픽모델링 결과

구분	토픽명	키워드1	키워드2	키워드3	키워드4	키워드5	관련 논문 수
Topic-1	지역 활성화와 문화콘텐츠	지역	활성	문화콘 텐츠	연계	자원	205
Topic-2	콘텐츠산업 성장과 경제 발전 방향	산업	성장	발전	경제	방향	155
Topic-3	문학작품 창작 소재화	작품	문학	창작	소재	시대	153
Topic-4	스토리텔링과 이야기 서사구조	스토리 텔링	서사	구조	확장	현재	147
Topic-5	정보콘텐츠 활용	콘텐츠	정보	활용	영향	구성	137
Topic-6	콘텐츠 연계 교육 방향	교육	연계	방향	활동	다양	137
Topic-7	영화 세계	영화	시작	세계	등장	상황	132
Topic-8	한국·세계·인식·사람의 다양성	한국	세계	인식	다양	사람	129
Topic-9	콘텐츠 경험 이론 (경험 영향 및 관계)	영향	관계	경험	이론	활동	117
Topic-10	게임 이론과 구성요소	게임	요소	이론	핵심	구성	113
Topic-11	역사적 사실 배경 콘텐츠	역사	사실	시대	배경	연계	113
Topic-12	예술 분야 융복합 콘텐츠	예술	분야	융복합	영역	이미지	103
Topic-13	시대 및 사회 변화와 콘텐츠	사회	시대	변화	상황	시작	103
Topic-14	문화관광자원 활성화	문화	관광	자원	대표	활성	100
Topic-15	사회적·소통적 존재 사람	사람	삶	소통	존재	사회	91
Topic-16	캐릭터 세계	캐릭터	세계	형태	역할	등장	90
Topic-17	미디어 개념 및 유형 변화와 확장	미디어	변화	개념	유형	확장	90

< 표 계속 >

구분	토픽명	키워드1	키워드2	키워드3	키워드4	키워드5	관련 논문 수
Topic-18	공간 콘텐츠	공간	환경	접근	차원	존재	89
Topic-19	국내 시장 평가	시장	평가	국내	확대	성장	88
Topic-20	문화콘텐츠 이론	문화콘 텐츠	이론	핵심	의식	역할	87
Topic-21	전통 및 현대 문화콘텐츠	전통	현대	대표	현재	문화콘 텐츠	82
Topic-22	이야기 구성방법	이야기	구성	소재	구조	배경	79
Topic-23	디지털 시대(이해)와 활용	디지털	관심	시대	활용	다양	79
Topic-24	대중 장르 수용	대중	장르	수용	활동	인식	77
Topic-25	영상기능과 영역	영상	역할	기능	영역	접근	74
Topic-26	디지털 콘텐츠 기술 환경과 영역	콘텐츠	기술	환경	디지털	영역	73
Topic-27	시각 표현 매체 확장	매체	표현	시각	시대	확장	72
Topic-28	이미지 인식 기술	이미지	인식	시간	현상	차원	72
Topic-29	문화프로그램 활용	문화	프로그 램	활용	차원	시대	69
Topic-30	문화적 · 경제적 가치 창조	가치	창조	창출	경제	문화	67
Topic-31	사용자 환경 전략	사용	전략	강화	차원	환경	66
Topic-32	문화경제 발전	국가	문화	경제	발전	관심	66
Topic-33	문화정책과 사업 활성화	문화	정책	사업	자원	활성	65
Topic-34	콘텐츠 성공 사례와 제작 방향	콘텐츠	제작	성공	문화콘 텐츠	방향	64
Topic-35	다양한 활용 체계	활용	체계	접근	다양	관심	61
Topic-36	콘텐츠 기획 체계	지원	단계	기획	방향	체계	61
Topic-37	다양한 주제 인물(활용)	주제	인물	참여	의식	다양	57
Topic-38	문화콘텐츠 사업 창출 · 확대	문화콘 텐츠	사업	활성	확대	창출	44
Topic-39	문화콘텐츠 시대 문화상품 생산	문화	상품	생산	문화콘 텐츠	시대	42
Topic-40	사회 발전과 문화경제성 지속	지속	정체성	형성	발전	상황	36

11) 일반적으로 키워드1과 키워드2를 중심으로 제목을 붙이며, 토픽명은 연구자 역량과 식견에 따라 달라질 수 있다.

3) 주요 연구분야 및 주제 분류

지금까지 문화콘텐츠 관련 학술논문 3,685편의 초록 및 서지정보에 대한 텍스트네트워크분석과 토픽모델링 결과를 살펴보았다. <표 7>은 분석결과를 종합, 정리하여 ‘문화콘텐츠학 주요 연구분야 및 연구주제 분류 방안’을 제시한 것이다. 주요 연구분야는 텍스트네트워크분석의 중심성 분석결과와 네트워크맵(그림 3)을 바탕으로 정리한 것이다. 그리고 연구주제는 토픽모델링으로 도출된 40개 토픽을 연구분야에 맞게 분류한 것이다. 이때 비슷한 내용은 통합하였다.

<표 7>에서 보는 바와 같이 문화콘텐츠학의 주요 연구분야는 크게 ‘문화콘텐츠 활용’, ‘문화콘텐츠 산업’, ‘한국사회와 문화콘텐츠’, ‘문화콘텐츠 장르’, ‘문화콘텐츠 기술’, ‘문화콘텐츠 이론 및 체계’ 등 6개로 분류할 수 있다. 연구분야별로 관련 논문의 비율을 살펴보면 ‘문화콘텐츠 활용’과 관련한 연구가 29.4%로 가장 많았다. 그 다음으로 ‘문화콘텐츠 장르(17.4%)’, ‘문화콘텐츠 이론 및 체계(17.3%)’, ‘문화콘텐츠 산업(16.1%)’ 순으로 나타났다.

<표 7> 문화콘텐츠학 주요 연구분야 및 연구주제 분류안

연구분야	연구주제 예시 ¹²⁾	관련 논문 수(비율)
문화콘텐츠 활용	지역활성화와 문화콘텐츠(지역문화관광자원 연계 콘텐츠)	1,084 (29.4%)
	문학작품 창작 소재화	
	스토리텔링과 이야기 구성방법	
	정보콘텐츠 활용과 영향	
	콘텐츠 연계 교육 방향	
	다양한 문화프로그램 활용	
	다양한 주제 인물 활용	
문화콘텐츠 산업	콘텐츠산업 성장과 경제 발전 방향	593 (16.1%)
	국내 시장 평가	
	문화적 · 경제적 가치 창조	
	사용자 환경(사용자 강화 전략)	
	문화경제 발전	
	문화정책과 사업 활성화	
	문화콘텐츠 시대 문화상품 생산	

< 표 계속 >

12) 토픽모델링에서 토픽 수를 40개로 설정, 추출했기 때문에 표에는 40개 토픽만 제시한 것이다. 토픽 수를 늘려 토픽모델링을 하면 더 많은 주제를 발굴할 수도 있지만, 토픽 수가 너무 많으면 내용 파악이 어렵게 된다.

연구분야	연구주제 예시	관련 논문 수(비율)
한국사회와 문화콘텐츠	한국 · 세계 · 인식 · 사람 다양성	441 (11.9%)
	시대 및 사회 변화와 콘텐츠	
	사회적 · 소통적 존재 사람의 삶	
	전통 및 현대 문화콘텐츠	
	사회발전과 문화정체성 지속	
문화콘텐츠 장르	영화 세계	640 (17.4%)
	게임 이론과 구성요소	
	역사적 사실 배경 콘텐츠	
	융복합 예술	
	캐릭터 세계	
	공간 콘텐츠	
문화콘텐츠 기술	영상 기능과 영역	291 (7.9%)
	디지털 콘텐츠 기술과 영역	
	시각 표현 매체 확장	
	이미지 인식 기술	
문화콘텐츠 이론 및 체계	콘텐츠 경험이론(경험 영향 및 관계)	636 (17.3%)
	미디어 개념 및 유형 확장	
	문화콘텐츠 이론 및 역할	
	디지털 시대 이해와 활용	
	대중 장르 수용과 인식	
	콘텐츠 기획 체계	
	콘텐츠 성공사례와 제작방향	
	다양한 활용 체계	

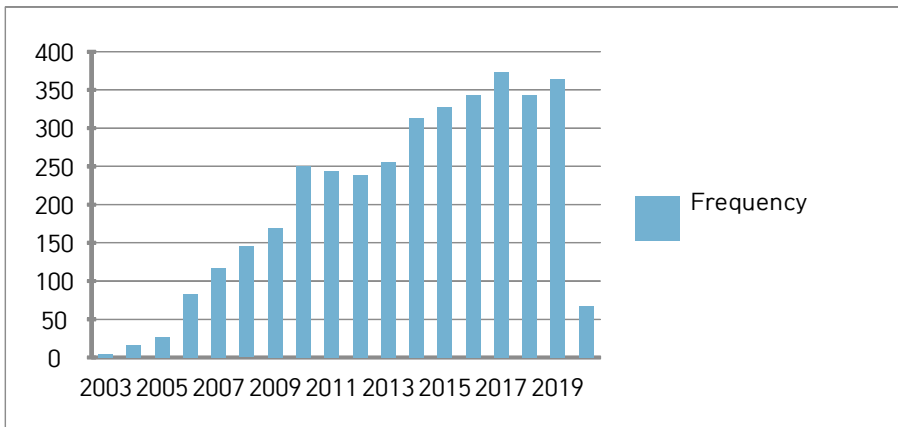
그동안 문화콘텐츠학의 고유 연구방법론 중 하나로 스토리텔링을 꼽았지만 본 논문에서는 분석결과를 바탕으로 ‘문화화콘텐츠 활용방법’을 문화콘텐츠학의 연구방법으로 제안하고자 한다. <표 7>에서 보듯이 문화콘텐츠 활용방법에는 스토리텔링을 포함하여 문학작품의 창작 소재화, 지역문화관광자원 활용, 정보콘텐츠 활용, 문화프로그램 활용, 교육 분야 콘텐츠 활용 등이 있다.

2. 문화콘텐츠학 연구 동향

1) 발표논문 현황

문화콘텐츠학의 연구동향을 살펴보기에 앞서 수집된 학술논문 3,685편에 대한 연도별 발표현황을 살펴보았다. [그림 4]에서 보듯이 문화콘텐츠 관련 KCI학술지 게재 논문 수는 2006년까지 100편 미만이었지만 2007년에 100편을 넘어섰고, 2009년까지 100편 이상씩 발표되었다. 2010년에 이르러 200편을 넘어섰는데, 문화콘텐츠 연구가 등장하기 시작한 2000년대 초와 비교해 보면 양적으로 크게 성장한 것이다. 2014년에 처음으로 300편을 넘어섰고, 그로부터 해마다 300편 이상씩 꾸준히 발표되고 있다.

[그림 4] 문화콘텐츠 관련 KCI학술지 게재 논문의 연도별 발표현황(2020년 4월 기준)



다음으로 학술지 현황을 살펴보았다. 문화콘텐츠 학술논문 3,685편이 게재된 KCI 학술지는 총 516종으로 조사되었다. 이중에서 관련 논문 10편 이상을 게재한 학술지 40종을 추려서 <표 8>과 같이 정리하였다.

문화콘텐츠 관련 논문이 가장 많이 게재된 KCI학술지는 인문콘텐츠(586편)이었다. 그 다음으로 한국콘텐츠학회논문지(484편), 글로벌문화콘텐츠(363편), 다문화콘텐츠(208편), 영상문화콘텐츠(193편) 순으로 나타났다. 학문분야로 살펴보면 문화콘텐츠 관련 논문이 가장 많이 발표된 학문분야는 인문학(46.7%)이었다. 이어서 복합학(24.8%), 예술체육학(18.2%), 사회과학(9.8%), 공학(0.44%) 순으로 나타났다.

〈표8〉 문화콘텐츠 관련 학술논문 10편 이상 게재 KCI학술지 현황(2020년 4월 기준)

학문분야	학술지명	논문수	학문분야	학술지명	논문수
인문학 (46.7%)	인문콘텐츠	586	예술 체육학 (18.2%)	영상문화콘텐츠연구	193
	글로벌문화콘텐츠	363		만화애니메이션연구	47
	문화콘텐츠연구	77		한국디자인문화학회지	40
	한국사상과 문화	18		디지털디자인학연구	30
	온지논총	17		한국디자인포럼	30
	우리문학연구	17		기초조형학연구	24
	한국학연구	14		애니메이션연구	23
	韓國古典研究	13		조형미디어학	18
	국제어문	11		커뮤니케이션디자인학연구	15
	기호학 연구	11		공연문화연구	15
	동아시아고대학	11	복합학 (24.8%)	한국콘텐츠학회 논문지	484
	동아인문학	10		한국엔터테인먼트산업학회	29
	지방사와 지방문화	10		디지털융복합연구	20
	한국문예비평연구	10		문화경제연구	15
사회과학 (9.8%)	다문화콘텐츠연구	208		예술경영연구	14
	문화산업연구	25		예술인문사회융합멀티미디어	14
	문화정책논총	12		대중서사연구	13
공학(0.44%)	디지털콘텐츠학회논문지	11		문화와 융합	11
예술 체육학	브랜드디자인학연구	10		한국과학예술융합학회	11
	상품문화디자인학연구	10		지역과문화	10

2) 주요 연구대상 시기별 분석

문화콘텐츠학 연구동향을 살펴보기 위해 시기별로 논문을 분류하였다. 대통령 임기에 따라 시기를 1시기(2003~2007), 2시기(2008~2012), 3시기(2013~2016), 4시기(2017~2020)로 구분하였다.¹³⁾ 이는 문화콘텐츠가 산업적·정책적 필요에 따라 등장한 개념이고 정부 주도 하에 성장했기 때문에 정부정책에 따라 관련 이슈도 달라졌을 것이라 판단했기 때문이다. 〈표 9〉는 시기별로 논문을 분류하고, 데이터 전처리 작업을 통해 각각 추출된 단어들의 개수를 정리한 것이다. 이렇게 분류한 데이터를 토픽모델링과

13) 3가지 방법 ①임의대로 5년씩 시기를 나눠 구분하는 방식, ② 정권별 정책이 학술연구에 영향을 미치는 데 2년이 소요된다는 가설을 고려한 방식, ③ 대통령 임기에 따라 구분하는 방식)으로 시기 구분을 하여 분석을 각각 한 후에 유의미한 결과가 나온 방식을 선택하였다.

텍스트네트워크분석을 각각 하였고, 주요 내용을 추려 <표 10>과 같이 제시하였다.¹⁴⁾

분석결과, 전 시기에 걸쳐 ‘활용’의 빈도수와 중심성 분석값 모두 높게 나타났다. 활용은 시기와 상관없이 전 시기에 걸쳐 문화콘텐츠 연구내용 구성에 중요한 영향을 끼쳤던 것으로 파악됐다. 그리고 시기별로 빈도수와 중심성 분석 값이 높게 나타난 단어로는 1시기에 산업, 활용, 디지털, 2시기에 활용, 산업, 가치, 3시기에 활용, 가치, 한국, 그리고 4시기에 활용, 사회, 가치가 나타났다. 이러한 결과가 어떠한 맥락적 의미를 갖는지 살펴보기 위해 스프링형 네트워크 맵(PFNet spring map)을 그려서 비교하였다(그림 5).

<표 9> 관련 학술논문 · 추출단어 시기별 분류

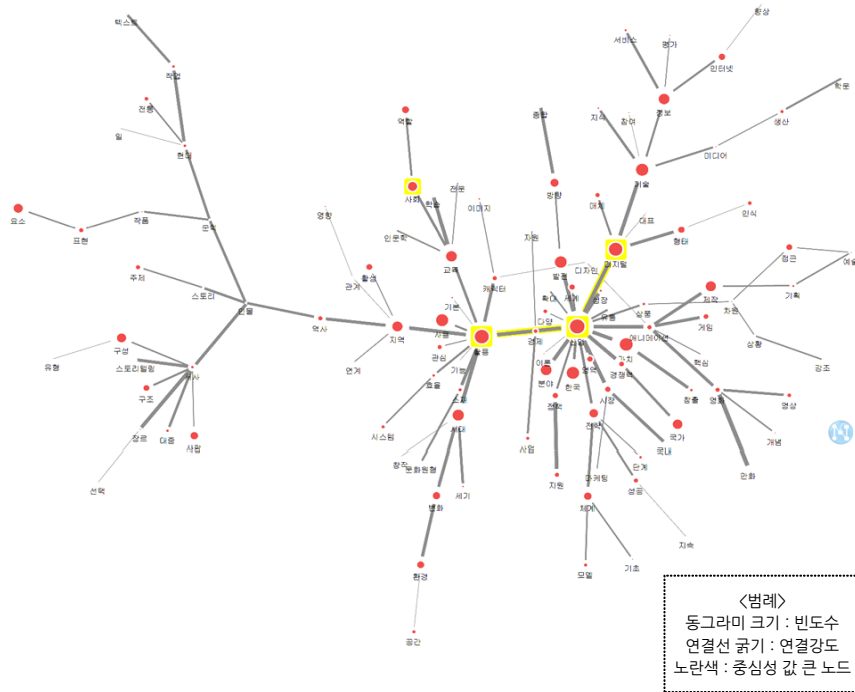
구분	1시기(2003~2007)	2시기(2008~2012)	3시기(2013~2016)	4시기(2017~2020)
발표 논문 수	248	1,048	1,240	1,149
추출 단어 수	3,429	8,822	10,731	10,503

<표 10> 관련 학술논문 시기별 빈도수 · 위세중심성 분석결과

구분	1시기(2003~2007)		2시기(2008~2012)		3시기(2013~2016)		4시기(2017~2020)	
	단어	측정값	단어	측정값	단어	측정값	단어	측정값
빈도수	문화콘텐츠	48	문화	575	문화	704	문화	594
	콘텐츠	136	문화콘텐츠	545	문화콘텐츠	660	문화콘텐츠	550
	문화	134	콘텐츠	542	콘텐츠	544	콘텐츠	454
	산업	98	활용	425	활용	502	활용	448
	활용	95	산업	344	한국	378	사회	366
	디지털	73	한국	298	가치	369	가치	307
	가치	67	사회	282	산업	347	지역	293
	사용	66	발전	256	지역	314	시대	268
위세중 심성	문화콘텐츠	0.200	문화	0.191	문화	0.193	문화	0.190
	문화	0.194	문화콘텐츠	0.187	문화콘텐츠	0.188	문화콘텐츠	0.181
	산업	0.177	콘텐츠	0.169	활용	0.164	활용	0.164
	콘텐츠	0.171	활용	0.161	콘텐츠	0.164	콘텐츠	0.158
	활용	0.162	산업	0.157	가치	0.150	사회	0.147
	디지털	0.147	가치	0.149	산업	0.147	가치	0.143
	가치	0.145	발전	0.132	사회	0.141	지역	0.138
	시대	0.139	한국	0.132	한국	0.138	한국	0.136

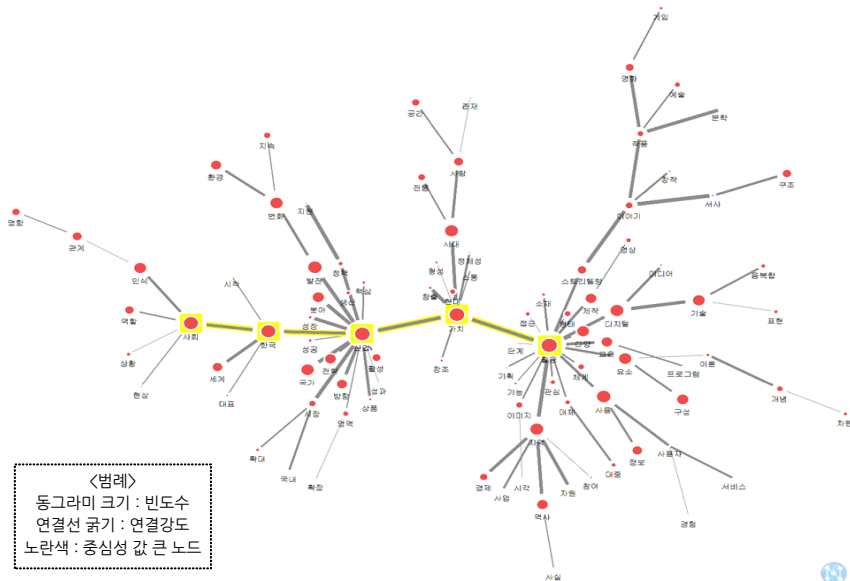
14) 2000년부터 2020년 최근까지 발간된 “문화콘텐츠” 관련 논문들 중에서 KCI학술지 게재 논문이면서 국문초록 및 서지정보가 있는 것만 필터링하였더니 2000년부터 2002년까지 발간된 논문들 중에서는 해당 조건에 맞는 것이 없었다.

[그림 5] 1시기(2003~2007) 주요 연구대상 네트워크맵



빈도수와 중심성 분석에서 값이 크게 나온 단어들과 연결되어 있는 단어를 배열하면 주요 연구분야를 파악할 수 있다. 검토결과, 1시기(2003~2007) 주요 연구분야는 ‘문화콘텐츠 활용’, ‘문화콘텐츠 산업’으로 나타났다. 각 분야의 주요 연구주제를 살펴보면 ‘문화콘텐츠 활용’ 분야에 문화원형 소재 활용, 교육과 콘텐츠 활용, 지역 역사·인물·스토리 활용, 시대 및 환경 변화와 활용 등이 있다. ‘문화콘텐츠 산업’ 분야에 문화콘텐츠 산업과 국가 경쟁력, 문화콘텐츠 산업 전략과 기초모델, 문화콘텐츠 산업 지원정책, 디지털정보기술 서비스 산업, 문화콘텐츠 산업 발전방향 등이 있다. 1시기에서 주목할 만한 점은 다른 시기에서 볼 수 없었던 ‘문화원형 소재 활용’이 나타났다. 이는 이 기간에 한국콘텐츠진흥원(舊한국문화콘텐츠진흥원)의 문화원형사업이 본격적으로 추진되었고, 이와 관련한 연구가 활발하게 이루어져 분석결과에 나타난 것으로 보인다. 문화원형사업은 콘텐츠 개발에 필요한 소재를 한국의 역사와 문화에서 발굴, 제공하여 문화콘텐츠 산업의 경쟁력을 확보한다는 취지로 2002년부터 2010년까지 추진되었다(그림 6).

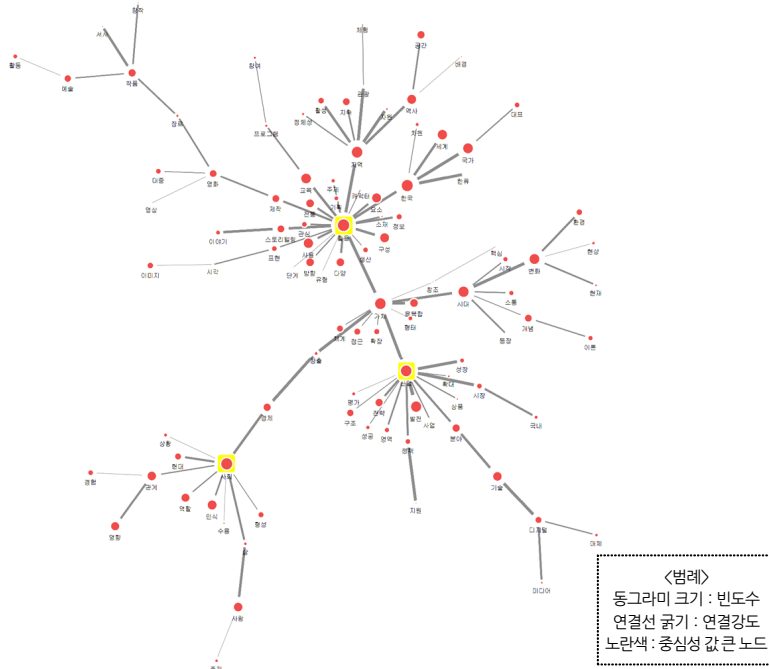
[그림 6] 2시기(2008~2012) 주요 연구대상 네트워크맵



2시기(2008~2012) 주요 연구분야는 ‘문화콘텐츠 활용’과 ‘문화콘텐츠 산업’, 그리고 새롭게 나타난 ‘한국사회와 문화콘텐츠’로 분류되었다. 각 분야의 주제를 1시기와 비교해서 살펴보자. ‘문화콘텐츠 활용’ 분야에 지역 역사·자원 활용, 교육과 콘텐츠 활용이 1시기에 이어 계속 나타났으며, 스토리텔링, 디지털 융복합 기술 활용, 가치 활용 등이 새롭게 나타났다. ‘문화콘텐츠 산업’ 분야에서 문화콘텐츠산업 지원정책이 1시기에 이어 계속 나타났으며, 문화콘텐츠 산업과 가치 창출, 국내시장 확대, 산업영역 확장, 문화콘텐츠 상품 등이 새롭게 나타났다. 1시기에는 산업 체계를 구상하고 발전방향을 모색하는 연구가 주를 이루었다면 2시기에는 산업의 영역을 확대, 활성화하는 연구가 이루어졌음을 알 수 있다. 새롭게 나타난 ‘한국사회와 문화콘텐츠’ 분야의 주제로는 세계와 한국사회, 한국사회 인식 및 영향 등이 있다.

3시기(2013~2016) 주요 연구분야는 ‘문화콘텐츠 활용’과 ‘문화콘텐츠 산업’, ‘한국사회와 문화콘텐츠’, 그리고 새롭게 나타난 ‘문화콘텐츠 가치’로 구분되었다. 2시기에서 일부 주제였던 ‘가치 활용’은 3시기에 주요 연구분야까지 확장되었다. 이 분야의 주제로는 가치 융복합, 경제가치 창출, 문화콘텐츠 산업과 가치 등이 있다. 그리고 ‘문화

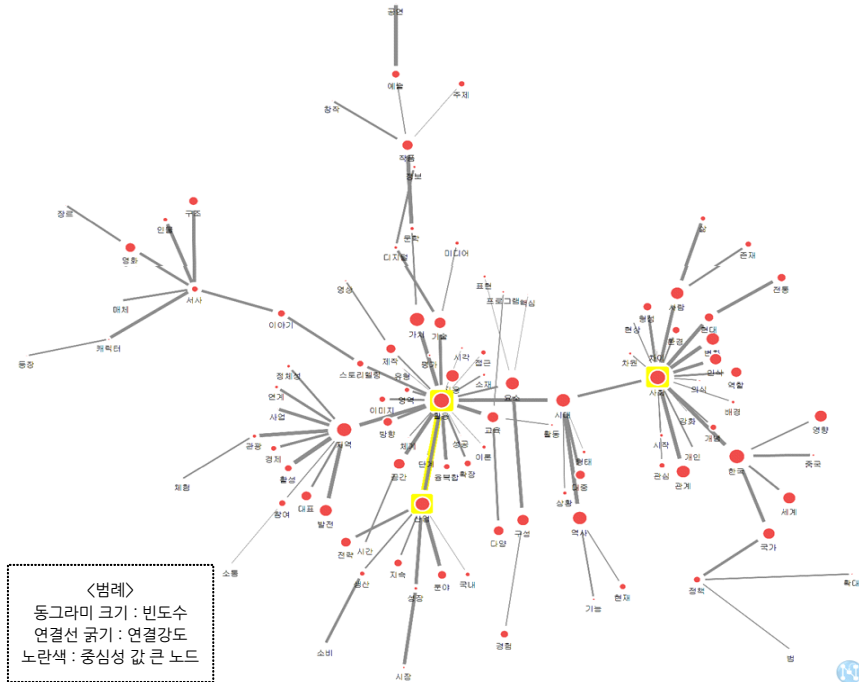
[그림 7] 3시기(2013~2016) 주요 연구대상 네트워크맵



콘텐츠 활용’ 분야에는 지역 역사·관광·자원 활용, 교육과 콘텐츠 활용, 스토리텔링과 활용이 1, 2시기에 이어 계속 나타났고, 한류 활용, 시각 이미지 활용 등이 새롭게 나타났다. 1, 2시기와 비교하여 지역의 역사공간 활용, 지역관광체험 활용, 지역정체성 등 문화콘텐츠 활용 분야의 연구주제가 다양하게 발전했다는 점을 3시기 특징으로 꼽을 수 있겠다(그림 7).

마지막으로 4시기(2017~2020) 주요 연구분야는 ‘문화콘텐츠 활용’, ‘문화콘텐츠 산업’, ‘한국사회와 문화콘텐츠’로 구분되었다. 3시기에서 연구분야로까지 확장되었던 ‘문화콘텐츠 가치’는 4시기에서는 나타나지 않았다. 각 분야의 연구주제를 살펴보면 ‘문화콘텐츠 활용’ 분야에 융복합 활용, 대중시대와 콘텐츠 활용, 경험요소 활용 등이 새롭게 나타났고, 지역 관광, 지역 활성화, 교육과 콘텐츠 활용, 스토리텔링 활용은 이전 시기에 이어 계속해서 나타났다. 이를 통해 ‘지역자원 활용’, ‘교육분야 콘텐츠 활용’, ‘스토리텔링 활용’은 시기와 상관없이 꾸준히 다뤄졌던 주제였음을 알 수 있다. 그리고 ‘문화콘텐츠 산업’ 분야에서 문화콘텐츠산업 지원정책이 이전 시기에 이어 계속 나타났

[그림 8] 4시기(2017~2020) 주요 연구대상 네트워크맵



으며, 시장과 성장, 생산과 소비 등이 새롭게 나타났다. 그리고 ‘한국사회와 문화콘텐츠’ 분야의 연구주제에 한국사회와 세계, 한국사회와 중국, 국가별 정책 확대 등이 새롭게 나타났다. 이를 통해 한류의 세계 확산 등 변화하는 세계질서 가운데 한국의 발전방향을 모색하는 연구가 활발하게 이루어졌음을 알 수 있다(그림 8).

3) 문화콘텐츠 분야 공동연구협력망 분석

마지막으로 문화콘텐츠 분야 연구자들의 공동연구협력체계를 분석하였다. 그동안 학계에서는 문화콘텐츠 연구 및 교육방법에 대해 학제간 접근을 강조해왔다. 이에 본 논문에서는 공저자네트워크분석을 통해 문화콘텐츠 분야의 공동연구 및 학제간연구 현황을 살펴보았다. 공저자네트워크란 연구자들이 공동연구방법을 사용해 논문 등을 저술했을 경우 그들의 관계를 네트워크로 표현한 것이다. 공저자네트워크에서 노드는 저자이며, 링크는 공동연구라는 사회적 친분 관계를 말한다(이수상, 2010).

먼저 분석을 위해 문화콘텐츠 관련 논문 10편 이상을 게재한 학술지 40종의 논문

2,498편에서 2,199명의 저자정보를 추출하였다. 이를 논문·저자 연결리스트(linked list)로 만들었고, 넷마이너 프로그램을 활용해 공저자네트워크분석을 하였다. 먼저 공저 횟수가 2회 이상인 저자들을 대상으로 저자·저자 일원모드 네트워크(1-mode Network)로 변환한 후에 중심성 분석을 하였다. <표 11>은 중심성 분석결과를 정리한 것이다.

<표 11> 문화콘텐츠 분야 공저자네트워크 중심성 분석결과

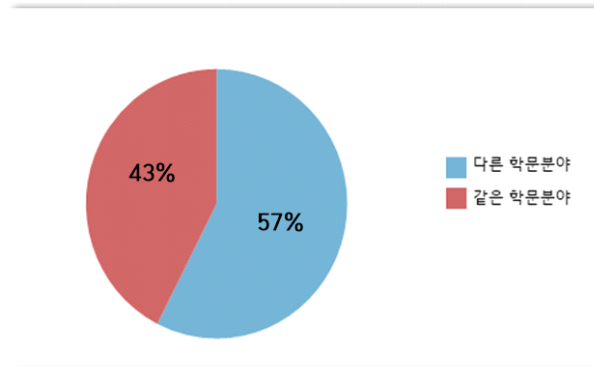
연결중심성		매개중심성	
노드	측정값	노드	측정값
유재수	0.021459	유재수	0.000037
복경수	0.015737	권혁인	0.000020
권지혁	0.014306	김은정	0.000020
임종태	0.011445	유동환	0.000012
고건식	0.011445	장원호	0.000012
최민웅	0.011445	신동일	0.000012
권혁인	0.011445	김기덕	0.000012

연결중심성 분석결과, 유재수(0.21459), 복경수(0.0157), 권지혁(0.0143) 순으로 분석값이 크게 나타났다. 연결중심성 값이 큰 사람일수록 연결망에서 큰 영향력과 권력을 가졌다고 해석한다. 이어 매개중심성 분석을 하였다. 매개중심성이 큰 사람일수록 공동연구에서 중개자 역할을 하고, 구성원들의 연대활동을 촉진시키는 역할을 한다고 해석한다. 분석결과, 유재수(0.000037), 권혁인(0.000020), 김은정(0.000020) 순으로 값이 크게 나타났다.

다음으로 학제간연구 현황을 살펴보기 위해 공저 횟수가 2회 이상인 저자들의 저자·저자 일원모드 네트워크 리스트를 확인해 저자들의 학문분야를 검토하였다(그림 9). 검토결과, 서로 다른 학문분야에서 활동하는 저자들이 만나 연구한 사례는 57%로 나타났고, 그렇지 않은 비율은 43%로 나타났다. 즉, 문화콘텐츠 연구분야 학제간연구 비율은 57%로 나타났으며, 그렇지 않은 비율보다 14% 높게 나타났다. 학제간연구방법을 통해 연구한 사례에는 「초기 개발자를 통해 본 RFID 디지털 보드게임 개발의 문제점 및 전망(이경미·이경옥, 2010)」, 「자동차 윈도우를 활용한 감성 영상콘텐츠 적용 방안(정수

영·최종훈, 2014)」, 「영화 <프라미스드 랜드>의 심층생태주의 텍스트로서의 의미(안영순·신현철, 2017)」 등이 있다(표 12).

[그림 9] 문화콘텐츠 분야 학제간연구 현황



<표12> 문화콘텐츠 분야 연구자 컴포넌트 분석결과

구분	그룹원	밀도	비고
C1	유재수(공학/복합학), 조기형(공학), 박준호(복합학), 복경수(공학), 최민웅(공학), 고건식(공학), 임종태(공학)	0.571	-
C2	이현경(사회과학), 나윤빈(복합학), 권혁인(사회과학), 이승하(인문학), 백보현(인문학)	0.5	학제간연구
C3	김은정(의약학), 신동일(복합학), 김금선(복합학), 황수경(인문학), 조성희(복합학)	0.4	학제간연구
C4	유동환(복합학), 박기천(인문학), 김정웅(인문학), 하진(복합학)	0.5	학제간연구
C5	송정은(복합학), 장원호(인문학), 구선아(인문학), 김상현(인문학)	0.5	학제간연구
C6	이동열(복합학), 이완복(복합학), 유석호(복합학), 경병표(복합학)	1	학제간연구
C7	이종호(복합학), 박현욱(복합학), 옥정원(사회과학), 윤대홍(복합학)	0.667	학제간연구
C8	김기덕(인문학), 김태룡(인문학), 김동윤(인문학), 최승용(인문학)	0.5	-
C9	태지호(인문학), 권지혁(사회과학), 백승국(인문학), 한창호(인문학)	0.667	학제간연구
C10	이정현(복합학), 박종태(복합학), 김지호(사회과학)	1	학제간연구
C11	임평중(공학), 김종서(공학), 곽훈성(공학)	0.667	-
C12	최희수(복합학), 김상현(복합학), 김나윤(복합학)	0.667	학제간연구
C13	정영찬(인문학), 김용범(인문학), 두일철(인문학)	0.667	-
C14	이정원(공학), 유미(공학), 권대규(공학)	1	-

마지막으로 문화콘텐츠 분야에 형성된 하위 연구집단을 찾기 위해 컴포넌트(component) 분석을 하였다. 이것은 응집력을 기준으로 노드를 분류해 주는 방법 중 하나로, 연결 고리가 끊기지 않고 서로 연결되어 있는 노드들을 그룹으로 분류해 준다. 분석결과, 14개 컴포넌트 그룹이 나타났다. 14개 그룹들 중에서 밀도가 1인 그룹은 C6, C10, C14로 나타났다. 밀도가 1이라는 것은 그룹 내 구성원들 서로가 하나로 연결되어 있으며 친밀도가 아주 높다는 뜻이다. 이 중에서 C6과 C10 그룹은 학제간 연구방법을 사용한 그룹이라 할 수 있다. C6 그룹의 주요 논문으로 「스마트폰 기반 캐주얼게임 개발 사례를 통한 캐릭터 수익모델 개발 연구 -크레이지 몽키 개발사례 분석(2013)」 등이 있다. C10 그룹의 주요 논문으로 「가상현실 기술을 활용한 교육콘텐츠가 치아발치에 관한 지식, 수행능력 및 실습만족도에 미치는 효과(2019)」 등이 있다.

IV. 나가며

본 논문의 목적은 문화콘텐츠 분야에 쌓여있는 학술 빅데이터를 토픽모델링과 텍스트네트워크분석을 이용해 분석하여 문화콘텐츠학의 연구동향과 지식구조를 밝히는 데 있다. 나아가 연구결과를 바탕으로 문화콘텐츠학의 학문적 체계와 위상을 정립하는 일에 기여하는 데 있다. 이러한 목적 아래 본 논문에서는 문화콘텐츠 관련 연구가 등장하기 시작한 2000년부터 2020년 최근까지 약 20년 동안 발간된 KCI학술지 논문 중에서 “문화콘텐츠”로 검색되는 논문 3,685편의 초록과 서지정보를 수집해 넷마이너 프로그램을 통해 텍스트네트워크분석과 토픽모델링을 하였다.

텍스트네트워크분석은 텍스트에서 단어들을 추출하고, 그 단어들의 관계를 계량적으로 분석하여 텍스트의 의미구조와 맥락 등을 나타내 주는 방법이다. 토픽모델링은 문서집합에서 특정 문서에 자주 등장하는 단어들을 하나로 묶어 토픽으로 추출해 주는 기계학습 알고리즘이다. 최근에 빅데이터 분석방법이 최적화된 여러 기법을 함께 사용하여 연구결과의 신뢰도를 높이는 쪽으로 발전하고 있기 때문에 본 논문에서도 두 가지 방법을 사용해 데이터를 분석하였다.

본 연구를 통해 도출된 주요 연구결과를 요약하면 다음과 같다. 첫째, 문화콘텐츠 관련 학술논문 3,685편의 초록 및 서지정보에서 추출한 단어 18,027개와 링크 389,715

개를 대상으로 텍스트네트워크분석을 하였다. 빈도수와 중심성 분석결과 ‘활용’, ‘가치’, ‘사회’, ‘산업’의 측정값이 높게 나타났다. ‘문화콘텐츠 활용’, ‘문화콘텐츠 가치’, ‘한국사회와 문화콘텐츠’, ‘문화콘텐츠 산업’은 문화콘텐츠학을 구성하는 필수 영역으로 나타난 것이다.

둘째, 문화콘텐츠 관련 학술논문 3,685편의 주요 내용을 파악하기 위해 LDA토픽모델링을 하였다. 분석결과, 40개의 토픽이 도출되었다. 논문 3,685편을 40개의 주제로 요약, 분류한 것이다.

셋째, 텍스트네트워크분석 결과로 나타난 주요 연구분야와 토픽모델링으로 분류된 토픽을 바탕으로 문화콘텐츠학 주요 연구분야 및 연구주제 분류방안을 제시하였다. 주요 연구분야는 크게 ‘문화콘텐츠 활용’, ‘문화콘텐츠 산업’, ‘한국사회와 문화콘텐츠’, ‘문화콘텐츠 장르’, ‘문화콘텐츠 기술’, ‘문화콘텐츠 이론 및 체계’ 등 6개로 구분하였다. 그리고 40개 토픽을 각 분야에 맞게 분류하였다. 또한 문화콘텐츠학의 고유한 연구방법론으로 ‘문화콘텐츠 활용방법’을 제시하였다. 문화콘텐츠 활용방법에는 스토리텔링, 문학작품의 창작소재화, 지역문화관광자원 활용, 정보콘텐츠 활용, 교육분야 콘텐츠 등이 있다. 그동안 학계에서 문화콘텐츠학의 고유한 연구방법으로 스토리텔링을 꼽았었는데, 본 논문에서 연구결과를 바탕으로 ‘문화콘텐츠 활용방법’을 제안한 것이다.

넷째, 문화콘텐츠 관련 논문을 1시기(2003~2007), 2시기(2008~2012), 3시기(2013~2016), 4시기(2017~2020)으로 구분해 텍스트네트워크분석과 토픽모델링을 하였다. 분석결과, 1시기에는 다른 시기에서 볼 수 없었던 ‘문화원형 소재 활용’이 나타났다. 그리고 문화콘텐츠 산업의 발전방향을 모색하는 연구가 주로 이뤄졌던 것으로 분석됐다. 산업과 관련한 연구는 2시기에 보다 발전하여 문화콘텐츠 산업의 영역을 확대하고 활성화하는 연구가 이뤄졌던 것으로 파악됐다. 그리고 2시기에 ‘한국사회와 문화콘텐츠’라는 연구분야가 새롭게 나타났다. 3시기에는 2시기에서 일부분으로 있었던 ‘가치 활용’ 주제가 주요 연구분야까지 확대되었다. 그리고 이전 시기와 비교하여 3시기에 문화콘텐츠 활용 분야 연구주제가 다양하게 발전하였다. 4시기에는 한국사회와 문화콘텐츠 분야에 ‘한국사회와 세계’, ‘국가별 정책 확대’ 등과 같은 주제가 새롭게 나타났다. 4시기에는 한류의 세계 확산 등 변화하는 세계질서 속에서 한국 문화콘텐츠의 발전방향을 모색하는 연구가 활발하게 이뤄졌던 것으로 분석됐다. 주목할 만한 점으로 전 시기에 걸쳐 문화콘텐츠 활용 분야에 ‘지역자원 활용’, ‘교육분야 콘텐츠 활용’, ‘스토리

텔링 활용'이 나타났다는 것이다. 이 주제들은 시기와 상관없이 문화콘텐츠학 분야에서 꾸준히 다뤄졌던 것으로 분석됐다.

다섯째, 문화콘텐츠 관련 논문에서 추출한 2,199명의 저자정보를 활용해 공저자네트워크분석을 하였다. 분석결과, 문화콘텐츠 분야 학제간연구 비율은 57%로, 그렇지 않은 비율보다 14% 높은 것으로 나타났다. 그리고 컴포넌트 분석결과, 문화콘텐츠 분야에 14개 연구집단이 있는 것으로 나타났다. 이 중에서 구성원들 서로가 하나로 완벽하게 연결된 친밀도가 가장 높은 그룹은 3개가 있는 것으로 분석됐다.

본 논문은 약 20년 동안 문화콘텐츠 학술 분야에 축적되어 있었던 빅데이터를 분석하여 문화콘텐츠 분야의 주요 연구대상, 연구방법, 지식구조 등을 도출함으로써 문화콘텐츠학의 학문적 체계와 위상을 정립하는 일의 토대를 제공했다는 점에서 연구의의를 찾을 수 있다. 그리고 문화콘텐츠의 주요 연구대상 및 영역에 대한 직관적 파악이 가능하도록 지식지도를 제공했다는 점에서 연구의미가 있다. 본 논문의 연구방법론은 향후 관련 분야에서 빅데이터 분석, 이슈 및 트렌드 파악 등과 같은 연구에 활용될 수 있다. 그리고 본 연구결과는 문화콘텐츠 관련 정책을 만드는 데 기초자료로 활용될 수 있다. ○

[참고문헌]

- 김기덕(2018), 4차 산업혁명시대 콘텐츠와 문화콘텐츠, 「인문콘텐츠」, 제52호, 9-31.
- 김미지 · 전재희 · 류은정(2019), 텍스트네트워크분석을 활용한 종양간호학회지 논문의 연구동향 분석, 「종양간호연구」, 19권 4호, 193-203.
- 김세현(2018), 비정형자료분석을 통해 살펴본 한국의 다문화 연구, 「한국인구학」, 41(1), 1-27.
- 김용학 · 김영진, 2016, 사회 연결망 분석, 서울: 박영사.
- 류준호 · 윤승금 · 이영주(2010), 문화콘텐츠 관련 연구에 대한 메타분석, 「언론과학연구」, 10(1), 124-165.
- 박상천(2007a), 문화콘텐츠 개념 정립을 위한 시론, 「한국언어문화」, 제33집, 179-210.
- 박상천(2007b), 문화콘텐츠학의 학문영역과 연구분야 설정에 관한 연구, 「인문콘텐츠」 제10호, 59-83.
- 박준석 · 김창식 · 광기영(2016), 텍스트마이닝과 소셜네트워크분석 기법을 활용한 호텔분야 연구동향 분석, 「관광레저연구」, 28권 9호, 209-226.
- 박치성 · 정지원(2013), 텍스트 네트워크 분석: 사회적 인식 네트워크 분석을 통한 정책 이해관계자 간 공유된 의미 파악 사례, 「정부학연구」, 제19권 제1호, 73-108.
- 신광철(2014), 문화콘텐츠학 연구사 정리의 방향과 과제, 「인문콘텐츠」, 제38호, 9-15.
- 오정심(2017), 사회 연결망 분석을 통한 무형유산 공동체 연구-판소리 종목을 중심으로, 「문화정책논총」, 제31권 제1호, 158-183.
- 오정심(2019), 소셜 네트워크 분석을 통한 무형문화유산 공동체 지식연결망 연구-정선 아리랑을 중심으로, 「문화재」, 제52권 제3호, 130-145.
- 오정심(2020), 빅데이터 텍스트 마이닝을 통한 무형문화유산 분야 연구동향 및 지식체계 분석, 「무형유산」, 8호, 93-127.
- 윤홍근(2011), 학위논문의 메타분석을 통한 문화콘텐츠 연구동향, 「글로벌문화콘텐츠」, Vol.7. 149-180.

- 윤희준 · 박재현 · 윤지운(2019), 비정형 텍스트 자료에서 잠재정보 추출을 위한 토픽 모델링 소개, 「체육과학연구」, 30호, 501-512.
- 이수상(2010), 공저 네트워크 분석에 관한 기초연구, 「한국도서관·정보학회지」, 제41권 제2호, 297-317.
- 이수상(2014), 언어 네트워크 분석 방법을 활용한 학술논문의 내용분석, 「정보관리학회지」, 제31권 제4호, 49-66.
- 이수성(2013), 「네트워크 분석방법론」, 서울:논형.
- 임대근(2015), 문화콘텐츠연구의 학문적 위상, 「인문콘텐츠」, 제38호, 131-156.
- 한국정보화진흥원(2015), 「IT & Future Strategy 보고서」, 대구:한국정보화진흥원.
- 한국콘텐츠진흥원(2017), 「2017 콘텐츠 교육기관 및 인력수급 현황조사」, 나주:한국콘텐츠진흥원.
- 황동열 · 황고은(2016), 빅데이터 기술을 활용한 인문콘텐츠 분야의 의미연결망 분석, 「인문콘텐츠」, 제43호, 229-255.
- 황서이 · 박정배 · 김문기(2020), 인문콘텐츠분야 연구의 경향 분석: 토픽모델링과 의미연결망분석을 중심으로, 「인문콘텐츠」, 제56호, 123-138.
- Linton C. Freeman(2004), The development of social network analysis : A study in the sociology of science, vancouver: Empirical Press.
- 권오성(2018.8.26.), 변덕꾸러기 토픽모델링 어떻게 다뤄야 하나, 「한겨레」.
- 文 "콘텐츠, 문화 넘어 한국경제 살리는 중요 산업"(2019.9.17.), 「파이낸셜뉴스」.
- 네이버지식백과 <https://terms.naver.com> 검색일:2020.4.30.
- 넷마이너 www.netminer.com 검색일:2020.4.30.

[Abstract]

A Study on Research Trends and Knowledge System in Cultural Contents Studies Using Topic Modeling and Text Network Analysis Based on Big Data

Oh, Jung-Shim

This paper aims to analyze academic big data in the field of cultural contents studies using topic modeling and text network analysis and explore the research trends and knowledge system. To achieve concrete results, the research was conducted with following goals: first, to determine the important central theme in the research of cultural contents studies; second, to outline the major topics in the field of cultural contents studies; third, to explain how major topics and subjects have changed in the field of cultural contents studies and what their characteristics are; and fourth, how the result of the analysis is visualized on a network map and what its characteristics are.

The research followed four steps—data collection, data refinement, data analysis, and integrating and interpretation. The data were collected between 2000, when the very first paper on cultural contents was published in South Korea, and 2020 from 3,685 academic papers. The collected unstructured data were refined for computer-aided analysis. First, nominal morphemes were extracted using a Korean morpheme analyzer; then, various controlling and TF-IDF analyses were applied. 18,027 words from academic papers have undergone topic modeling and text network analysis with a NetMiner program.

Topic modeling is a probabilistic algorithm discovering subjects and topics hidden in a large set of documents, which extracts and classifies documents according to the topic. Text network analysis applies the network theories and

analysis methods that developed out of sociology to literature analysis, analyzing the structure of connected words in the text and showing the result in the form of a network map. Recent big data analyses are evolving toward utilizing various optimized analytical techniques to enhance the reliability of the analysis result. Thus, this paper used topic modeling and network analysis to draw a result that is optimal for the purpose of our research.

This paper contributes to relevant studies as it uses topic modeling and text network analysis to analyze the big data that have accumulated in the field of cultural contents studies. In addition, it makes a significant contribution as it provides a visualized knowledge map to reveal the relationship of keywords and main topics in the field of cultural contents studies, which leads to the intuitive understanding of abstract contents.

[Keywords] cultural contents studies, research trends, text network analysis, topic modeling, big data text-mining

