

Analysis of Trends in Research Topics in the Design Field using Topic Modeling

Eunyoung Kim

Department of Visual & Media Design, Lecturer, Sookmyung Women's University, Seoul, Korea

Abstract

Background Research of trends in research provides basic data for reviewing previously conducted studies and setting future research directions. However, there is a long gap in the research targeting the overall domestic design research, and the corresponding papers are still significantly insufficient.

Methods Topic modeling was conducted using a total of 6,071 papers published in three comprehensive South Korean design journals for 17 years from 2004 to 2020 as learning data. Next, the dominant topics of each paper have been extracted from the model and the number of papers for each topic was measured every four years to examine how research topics in the design field have changed.

Results When topic modeling was performed, the number of topics with the highest topic consistency was 56, of which 4 were related to research methods, not subjects. When comparing the number of papers based on the dominant topic, the topics in the upper and lower sections continued to change by period, which seems to be the result of responding to the policies and the issues of the times.

Conclusions This study is valuable in that it can understand, even in a limited way, what research topics have been dealt with by researchers in the design field and how the topics have changed over time. It is hoped that this study can be used as basic data for researchers in the design field to review their research topics and to establish future research directions.

Keywords Design Research Topic Trends, Design Corpus, Topic Modeling, Natural Language Processing

Citation: Kim, E. (2022). Analysis of Trends in Research Topics in the Design Field using Topic Modeling. *Archives of Design Research*, 35(1), 331-345.

<http://dx.doi.org/10.15187/adr.2022.02.35.1.331>

Received : Oct. 20. 2021 ; **Reviewed :** Dec. 13. 2021 ; **Accepted :** Dec. 13. 2021

pISSN 1226-8046 **eISSN** 2288-2987

Copyright : This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

1. 연구의 필요성과 목적

연구 동향을 분석하는 연구는 일정한 기준을 토대로 그간 수행된 연구들을 종합하고 검토함으로써 향후 연구 방향에 대한 기초 자료를 제공한다. 따라서 제 분야의 연구가 어느 정도 누적된 대부분의 학문 분야에서 연구 동향을 살피는 연구는 꾸준히 시도되고 있으며 디자인 분야 역시 그러하다.

국내외 디자인 연구 동향을 분석한 연구에 대해서는 최근 이다니엘(2021)이 정리한 바 있다. 이다니엘은 디자인 연구 동향에 관한 논문을 하나의 세부 주제를 특정하여 국내외 연구 동향을 살피는 유형과 세부 주제를 특정하지 않고 전반적인 디자인 연구의 동향을 살피는 유형 등 크게 두 가지로 분류하고 대다수 연구가 특정 주제나 학술지에 한정되어 있음을 지적했다.

국내 디자인 연구 전반에 초점을 맞춘 연구 동향 연구로는 오랜 기간 「한국 디자인학 연구 동향에 대한 분석」(김종덕, 2004)이 유일했으나 16년 이상이 지난 현시점에서 더는 이 연구에 시의성을 기대하기 어렵다. 이에 따라 이다니엘은 2002년에서 2019년까지 20년간 디자인 관련 KCI 등재지에 출간된 논문을 대상으로 생산성 및 주제어 등을 계량서지학(bibliometrics) 측면에서 양적 분석함으로써 이 긴 공백에 징검다리를 놓고자 했다. 다만, 그는 이 연구 방식의 한계를 인정하며 내용을 더 깊이 있게 분석하고 서로 연관성 깊은 주제들을 파악하기 위해서는 토픽모델링(Topic modeling) 또는 키워드 네트워크 분석이 필요하다고 제안했다.

많은 분야에서 연구 동향 연구는 계량서지학 방식은 물론, 문헌에서 분석 단위를 설정하고 각 단위의 수량을 분석하는 내용 분석(content analysis) 방식으로 수행된다. 그러나 전통적인 내용 분석 방식은 연구자가 내용을 분류하는 방법과 범주를 자의적으로 결정할 가능성이 높고 분석 범위를 늘리는 데 한계가 있으며, 여러 참여자가 함께 분석을 수행하는 경우 평가 기준이 서로 달라 분석 신뢰도와 타당도가 낮아질 수 있다는 어려움이 존재한다. 토픽모델링이나 키워드 네트워크 분석과 같이 컴퓨터를 이용한 텍스트 분석 방식은 이런 단점을 보완할 수 있는 방법으로서 2000년대 중반 이후 국내외에서 연구 동향을 분석하는 데 적극 사용되고 있다. 이 연구는 그중 토픽모델링 기법을 활용하여 디자인 분야에서 어떤 연구 주제들이 다뤄졌으며 시기별로 그 양상이 어떻게 변화해 왔는지 살핌으로써 위의 후속 연구 제안에 부응하고자 한다.

2. 분석 대상과 방법

현재 텍스트 분석 기반으로 연구 동향을 분석하는 연구에서 논문 초록이 아닌 전문을 분석 범위로 삼은 경우는 매우 드물다. 텍스트 분석을 위해서는 먼저 대량의 문헌을 디지털화하여 컴퓨터가 이해할 수 있는 상태로 구조화해야 하고 형태소 분석이 필요한 한국어 텍스트는 처리 난이도 역시 영어 텍스트보다 높기 때문이다.

하나의 논문은 하나의 연구 주제로 대표되지만 실제 본문은 여러 개의 소주제로 구성될 수밖에 없다. 즉 하나의 논문은 지배적인 토픽 외에도 여러 토픽을 포함하고 있을 가능성이 높다. 초록에는 연구 의의와 전개에 관한 핵심 내용만 요약해야 하므로 본문에 포함된 상세 내용은 대부분 배제된다. 텍스트 분석 기법을 활용할 때에는 전문이 아닌 초록만 활용할 경우 상당량의 정보를 유실한 상태에서 정보를 추출하게 되어 그만큼 유의미한 잠재 정보를 추출할 확률이 낮아진다. 따라서 이 연구에서는 텍스트의 분석 범위를 전문으로 넓혀 초록이나 주제어만으로는 발견할 수 없는 연구 주제들을 찾아내고자 한다.

분석 결과의 일반화를 위해서는 디자인 분야의 학술지 전수를 대상으로 함이 가장 이상적이지만 한정된 여건 내에서 연구를 수행하기 위해서는 분석 대상을 제한할 수밖에 없었다. 이 연구의 분석 대상 학술지는 <디자인학연구(ADR)>, <한국디자인문화학회지(KSDC)>, <한국디자인포럼(KDF)> 총 3종으로, 현시점에서 검색되는 디자인 분야의 등재학술지 14종 중에서 정보, 브랜드, 일러스트레이션, 가구, 공간 등 특정 소분야로 범위가 한정되지 않으면서도 2000년 이전에 창간되어 연 4회 꾸준히 발행되고 있는, 누적 발행 권수가 최소 100권이 넘는 종합 학술지이다. 2004년 디자인학 연구 동향 논문에서 다뤄진 <기초조형학연구>는 양적 측면에서 충분한 규모를 갖췄지만 디자인 분야가 아닌 예술일반에 속한 등재학술지이기에 이번 연구에 포함하지 않았다.

위의 3종 학술지가 다루는 디자인 소분야는 사뭇 다르며 이들이 질적 측면에서 디자인 분야의 대표성을 가질 수 있는지에 대해서는 의견이 나뉠 수 있다. 그러나 이 연구는 텍스트 분석을 통해 디자인 분야 연구자가 다뤄 온 연구 주제의 동향, 곧 현상을 탐색하는 데 목적이 있다. 위의 학술지들에서 포용하는 논문의 주제 범주를 모두 디자인학에 포함할 수 있는지에 대해서는 별도의 깊이 있는 논의가 필요하다.

분석 대상의 시간적 범위는 세 학술지가 모두 현재와 같이 안정적 주기와 편수로 논문을 발행하기 시작한 2004년부터 2020년까지이다. 17년간 3종 학술지에 게재된 한글 논문은 총 6,071편으로 전문을 형태소 분석하여 비한글 단어를 모두 제거한 상태의 단어 수는 13,228,635개이다. 디자인 학술지의 텍스트 처리 과정에 대해서는 김은영(2020)에 상술되어 있으며, 그로써 해결되지 않은 분절어에 대해서는 추가적으로 후처리했다.

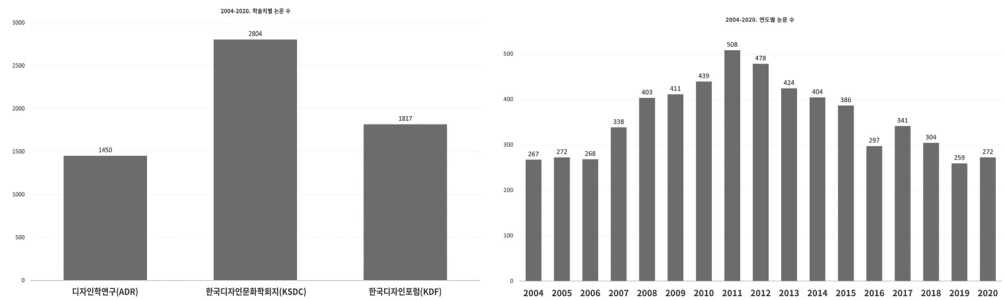


Figure 1 The number of papers by journal and by year

토픽모델링에는 LDA 알고리즘 기반의 Mallet 라이브러리를 이용했고, 앞서 구조화한 텍스트 데이터에서 2음절 이상의 명사형 단어를 학습 데이터로 사용했다. LDA 모델의 품질은 각 토픽에 속한 단어들의 의미적 유사도의 일관성을 수치화한 ‘주제 일관성’(Coherence score) 점수로 평가했다. 20에서 92까지 4 단위로 토픽 수를 설정하여 모델의 주제 일관성 점수 변화를 살핀 결과는 Figure 2로 정리했다. 주제 일관성이 5 이상이면서 점수가 급격하게 떨어지기 전의 토픽 수는 24, 44, 56, 76이다. 그중 연구 동향 분석에 가장 적합한 모델을 찾기 위해 각 모델을 LDAvis로 시각화하여 토픽 분포 상태와 토픽별 키워드를 살폈다. 76의 경우, 72와 일관성 점수의 차이가 미미하여 토픽 수가 적은 72 모델을 대신 사용했다.

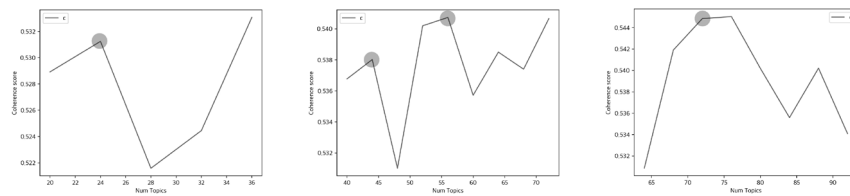


Figure 2 Determining the appropriate number of topics in LDA

LDA 모델을 시각화했을 때 토픽이 네 사분면에 고루 퍼져 있지 않거나 토픽 간 규모 차이가 지나치게 크면 일반적으로 학습이 제대로 이루어지지 않았다고 판단한다. 그러나 학습이 잘 되었어도 개별 토픽이 너무 분화되지 않거나 세분화된 경우에는 연구 주제 동향을 살피기에 적절하지 않다.

시각화 상태를 검토했을 때 24는 토픽 수가 적은 탓에 토픽 간 차이가 분명하지만 타이포그래피 분야가 광고 디자인이나 정보 디자인 분야와 하나의 토픽으로 묶일 만큼 주제 분화도가 낮았다. 72는 주제가 훨씬 세분화된 대신 각 토픽을 구성하는 키워드 수가 적어져 토픽 간에 어떤 차이가 있는지 직관적으로 알기 어려웠다. 네 개 모델 중 주제 분화도와 개별 토픽의 규모가 가장 적정한 것은 토픽 수를 56으로 설정한 모델로, 이 연구에서는 이를 토대로 실질적인 분석을 진행했다.

3. 디자인 분야의 연구 주제 분포

한글 논문 6,071편을 56개 토픽으로 모델링하여 LDAvis로 시각화한 상태는 Figure 3이며, 각 토픽에 기여도가 높은 상위 키워드 20개는 Table 1에 내림차순으로 정리했다. 이후 각 토픽을 지칭할 때는 그 토픽의 최상위 키워드 3개를 연결하여 명칭으로 사용한다. 토픽 56개 토픽 중 6, 14, 21, 27번 토픽은 연구 주제보다는 연구 방법과 관련이 있었다.

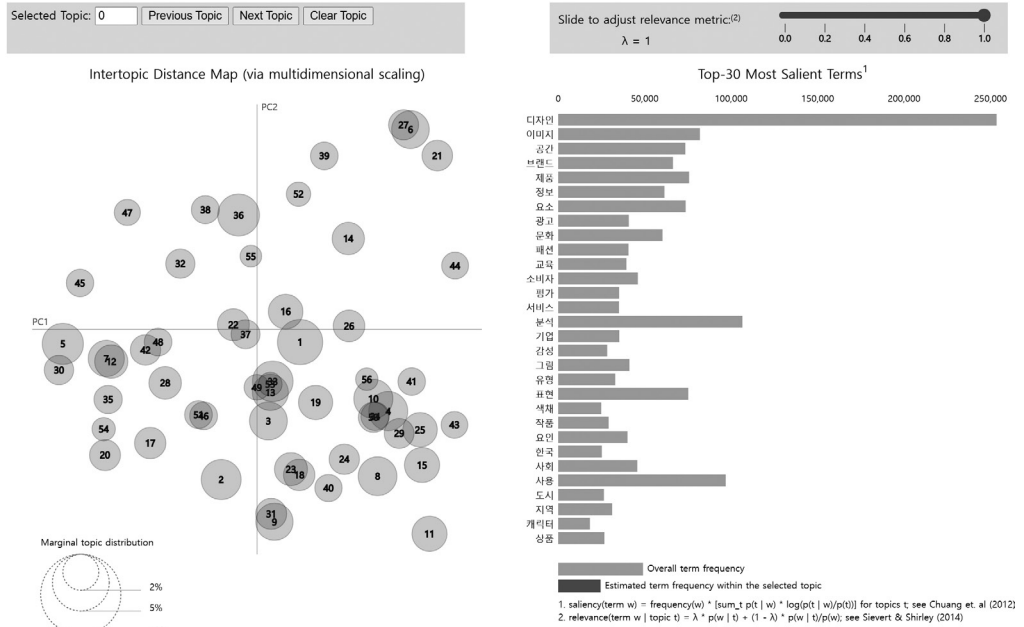


Figure 3 Visualization of LDA model using LDAvis

Table 1 Top 20 keywords that contribute to each topic

| Topic | Keywords |
|-------|--|
| 1 | 개념, 구조, 관계, 과정, 의미, 관점, 이론, 행위, 체계, 이해, 해석, 접근, 인식, 대상, 존재, 차원, 정의, 영역, 기능, 인간 |
| 2 | 사회, 문화, 가치, 변화, 인간, 시대, 개념, 의미, 다양, 개인, 환경, 발전, 역할, 생활, 대중, 경제, 현대, 중심, 산업, 영역 |
| 3 | 디자인, 디자이너, 분야, 개념, 영역, 중요, 다양, 적용, 중심, 목적, 방향, 필요, 접근, 제시, 측면, 활용, 역할, 강조, 기본, 고려 |
| 4 | 단계, 과정, 창의, 프로세스, 개발, 아이디어, 진행, 참여, 사고, 디자이너, 활용, 도출, 분석, 작업, 프로젝트, 적용, 해결, 도구, 제안, 필요 |
| 5 | 표현, 형태, 조형, 시각, 특성, 자연, 다양, 요소, 단순, 추상, 원리, 기하학, 형상, 반복, 구성, 효과, 구조, 차원, 변화, 강조 |
| 6 | 요인, 영향, 분석, 가설, 태도, 검증, 신뢰, 변수, 측정, 관계, 유의, 문항, 이용, 지각, 긍정, 의도, 구성, 통계, 만족, 효과 |
| 7 | 인간, 세계, 자신, 현실, 존재, 의식, 의미, 미학, 표현, 신체, 일상, 주체, 정신, 대상, 욕망, 모습, 이상, 상징, 환상, 사물 |
| 8 | 도시, 공공, 지역, 경관, 환경, 계획, 사업, 가로, 조성, 주민, 거리, 시민, 공원, 마을, 개선, 시설물, 시설, 주변, 커뮤니티, 참여 |
| 9 | 기술, 디지털, 미디어, 가능, 컴퓨터, 가상, 시스템, 환경, 활용, 현실, 증강현실, 구현, 다양, 발전, 기반, 이용, 분야, 방식, 융합, 개발 |
| 10 | 사용, 인터페이스, 사용자, 기능, 화면, 모바일, 아이콘, 조작, 스마트폰, 메뉴, 버튼, 기기, 수행, 제공, 시간, 선택, 방식, 입력, 시스템, 이용 |
| 11 | 산업, 정책, 지원, 사업, 국가, 정부, 개발, 국내, 기관, 운영, 분야, 경제, 발전, 필요, 추진, 규모, 전문, 현황, 세계, 기술 |
| 12 | 패션, 스타일, 컬렉션, 디자이너, 소재, 트렌드, 다양, 아이템, 시즌, 표현, 특성, 유행, 라인, 의상, 특징, 여성, 활용, 사별, 콜라보레이션, 현대 |
| 13 | 브랜드, 이미지, 소비자, 아이덴티티, 전략, 기업, 로고, 제품, 마케팅, 가치, 상징, 요소, 구축, 개성, 확장, 형성, 중요, 연상, 컨셉, 활용 |
| 14 | 분석, 유형, 분류, 특성, 대상, 기준, 활용, 자료, 분야, 중심, 국내, 구분, 주제, 선정, 파악, 다양, 체계, 특징, 비율, 해당 |

| | |
|----|--|
| 15 | 기업, 전략, 경영, 혁신, 개발, 활동, 조직, 성공, 환경, 제품, 시스템, 성과, 분석, 마케팅, 비즈니스, 활용, 창출, 시장, 관리, 가치 |
| 16 | 경험, 감각, 체험, 지각, 인지, 시각, 자극, 반응, 행동, 인간, 기억, 감정, 인터랙션, 과정, 유도, 촉각, 환경, 시간, 청각, 관계 |
| 17 | 역사, 시작, 이후, 시대, 양식, 시기, 당시, 근대, 영향, 세기, 정치, 등장, 미국, 영국, 대중, 사회, 변화, 유럽, 초기, 방식 |
| 18 | 제품, 기능, 개발, 자동차, 사용, 시장, 기술, 형태, 적용, 디자인, 모델, 차량, 출시, 다양, 중심, 생산, 트렌드, 휴대폰, 산업, 소비자 |
| 19 | 공간, 환경, 장소, 구성, 특성, 영역, 분석, 물리, 시설, 기능, 요소, 이용자, 다양, 연결, 이용, 위치, 중심, 배치, 형성, 연출 |
| 20 | 작품, 예술, 전시, 미술, 작가, 아트, 박물관, 그림, 현대, 작업, 매체, 회화, 다양, 표현, 대중, 미술관, 예술가, 세계, 오브제, 개념 |
| 21 | 평가, 항목, 그룹, 분석, 차이, 측정, 기준, 비교, 선정, 평균, 사용, 판단, 대상, 제시, 점수, 실시, 정도, 전문가, 수준, 척도 |
| 22 | 그림, 형태, 방향, 변화, 구조, 크기, 위치, 방식, 부분, 배치, 제작, 시간, 구성, 프레임, 기본, 적용, 가능, 동일, 거리, 비교 |
| 23 | 게임, 콘텐츠, 인터넷, 온라인, 사이트, 이용, 아바타, 제품, 미디어, 동영상, 공유, 모바일, 콘텐츠, 다양, 블로그, 커뮤니티, 자신, 참여, 이용자, 개인 |
| 24 | 상품, 시장, 고객, 마케팅, 판매, 매장, 국내, 소비자, 유통, 업체, 다양, 백화점, 분석, 가격, 전략, 매출, 현황, 소비, 카드, 구성 |
| 25 | 서비스, 사용자, 제품, 경험, 분석, 기반, 고객, 활용, 도출, 중심, 사용, 단계, 다양, 니즈, 스마트, 개인, 진행, 기술, 필요, 정보 |
| 26 | 소비자, 구매, 제품, 소비, 가치, 가격, 쇼핑, 품질, 선택, 패션, 명품, 상품, 경험, 인터넷, 의류, 사용, 속성, 브랜드, 온라인, 만족 |
| 27 | 선호, 선호도, 차이, 분석, 인식, 설문, 응답, 대상, 연령, 응답자, 실시, 이상, 문항, 만족, 일반, 만족도, 빈도, 성별, 선택, 중요 |
| 28 | 영화, 영상, 제작, 관객, 장면, 시간, 화면, 구성, 방송, 효과, 표현, 드라마, 장르, 촬영, 매체, 서사, 인물, 내러티브, 음악, 분석 |
| 29 | 교육, 대학, 과정, 전공, 학생, 프로그램, 능력, 분야, 융합, 필요, 전문, 수업, 학교, 기초, 과목, 지식, 교과목, 중심, 현장, 기술 |
| 30 | 문양, 상징, 의미, 사용, 장식, 시대, 전통, 형태, 표현, 사상, 문화, 특징, 불교, 조선, 고대, 형상, 자연, 의식, 종교, 다양 |
| 31 | 문화, 지역, 개발, 축제, 관광, 상품, 활용, 역사, 자원, 전통, 문화상품, 체험, 홍보, 방문, 다양, 가치, 이미지, 콘텐츠, 상징, 관광객 |
| 32 | 의미, 기호, 언어, 표현, 상징, 커뮤니케이션, 전달, 대상, 시각, 분석, 코드, 작용, 이미지, 해석, 기호학, 메시지, 텍스트, 은유, 관계, 이해 |
| 33 | 정보, 전달, 시각, 데이터, 커뮤니케이션, 제품, 활용, 표현, 요소, 효과, 이해, 시각화, 매체, 다양, 사용, 구조, 기반, 필요, 설명, 목적 |
| 34 | 학습, 교육, 어린이, 아동, 수업, 놀이, 활동, 효과, 학생, 유아, 활용, 과정, 프로그램, 발달, 다양, 개발, 능력, 구성, 교사, 미술 |
| 35 | 패턴, 소재, 디자인, 활용, 기법, 문양, 사용, 텍스타일, 직물, 개발, 모티브, 표현, 다양, 제작, 현대, 적용, 효과, 구성, 섬유, 이용 |
| 36 | 요소, 감성, 분석, 시각, 구성, 어휘, 요인, 추출, 반응, 중요, 영향, 적용, 변화, 인간, 자극, 형용사, 표현, 심미, 육구, 분류 |
| 37 | 광고, 효과, 소비자, 메시지, 모델, 표현, 매체, 태도, 수용자, 소구, 노출, 제품, 전략, 커뮤니케이션, 반응, 전달, 유형, 크리에이티브, 캠페인, 비주얼 |
| 38 | 색채, 색상, 컬러, 사용, 배색, 계열, 분석, 명도, 색조, 조화, 빨강, 파랑, 이미지, 느낌, 채도, 무채색, 노랑, 대비, 분포, 효과 |
| 39 | 집단, 요인, 추구, 행동, 자신, 의복, 성향, 외모, 관심, 신체, 사회, 차이, 자아, 라이프스타일, 지향, 개인, 세대, 유행, 태도, 분석 |
| 40 | 환경, 친환경, 패키징, 사용, 제품, 자연, 포장, 에너지, 재활용, 소재, 개발, 생산, 적용, 용기, 건강, 자원, 가능, 웰빙, 재료, 활용 |
| 41 | 노인, 사용, 건강, 사회, 운동, 고령자, 유니버설, 기능, 장애, 신체, 필요, 병원, 환자, 생활, 활동, 치료, 의료, 이상, 고령, 개발 |
| 42 | 시각, 문자, 그래픽, 사용, 글자, 타이포그래피, 일러스트레이션, 표현, 요소, 서체, 한글, 레이아웃, 텍스트, 전달, 잡지, 구성, 그림, 가독성, 기능, 독자 |
| 43 | 안전, 시설, 설치, 발생, 이용, 환경, 자전거, 보행, 개선, 적용, 상황, 필요, 범죄, 재난, 고려, 위험, 관리, 분석, 가능, 문제점 |
| 44 | 고객, 미용, 분석, 조직, 직무, 관리, 업무, 미용실, 서비스, 행동, 요인, 만족, 특성, 중요, 관계, 영향, 이상, 직업, 유의미, 통계 |
| 45 | 의상, 착용, 복식, 의복, 소재, 장식, 사용, 형태, 여성, 모자, 실루엣, 소매, 한복, 시대, 세기, 유니폼, 부분, 표현, 화려, 드레스 |
| 46 | 한국, 문화, 중국, 전통, 일본, 국가, 세계, 현대, 미국, 비교, 동양, 나라, 한류, 서양, 한국인, 특성, 민족, 시장, 서구, 고유 |
| 47 | 이미지, 여성, 남성, 사진, 표현, 역할, 분석, 향수, 매력, 개성, 인상, 세련, 변화, 형성, 의미, 활동, 현대, 강조, 차원, 시각 |
| 48 | 캐릭터, 애니메이션, 로봇, 표현, 성격, 스토리텔링, 감정, 만화, 스토리, 이야기, 얼굴, 역할, 제작, 표정, 분석, 인간, 그림, 동작, 특징, 다양 |
| 49 | 소재, 섬유, 사용, 작업, 표면, 처리, 제작, 생산, 염색, 가공, 시간, 재료, 이용, 프린팅, 가능, 공정, 활용, 플라스틱, 과정, 효과 |
| 50 | 사인, 정보, 간판, 설치, 지도, 안내, 환경, 방향, 도로, 시스템, 디자인, 체계, 지하철, 요소, 위치, 사용, 개선, 옥외광고물, 시각, 필요 |
| 51 | 건축, 재료, 건물, 자연, 형태, 특성, 건축물, 사용, 외부, 공간, 요소, 내부, 구조, 설계, 기능, 조형물, 구성, 실내, 주변, 전통 |
| 52 | 체형, 패턴, 항목, 착용, 길이, 신체, 치수, 사이즈, 부위, 여성, 요인, 차이, 형태, 어깨, 돌레, 너비, 분석, 의복, 비교, 개발 |
| 53 | 가구, 조명, 사용, 주거, 실내, 생활, 주택, 의자, 아파트, 공간, 기능, 음식, 인테리어, 다양, 특징, 형태, 특성, 주방, 계획, 거실 |
| 54 | 공예, 제작, 장식, 장신구, 기법, 사용, 재료, 주얼리, 현대, 금속, 형태, 전통, 도자, 다양, 기술, 보석, 공예품, 도자기, 조형, 이용 |

LDA 모델의 학습에 활용한 논문들을 각각에서 가장 지배적으로 나타난 토픽으로 분류한 결과는 Figure 4이다(상위 10%: 검은색, 상위 10~20%: 진한 회색, 하위 10%: 가장 옅은 회색). 여기서 한 논문의 지배적 토픽은 반드시 그 논문의 연구 대상과 일치하지는 않는다. 예를 들어 2020년에 게재된 논문 중 「말뭉치와 텍스트마이닝 기법을 활용한 <디자인학연구>의 사용 어휘 분석」(김은영, 안병하), 「빅데이터를 활용한 애슬레저룩에 관한 소비자 인식 연구」(강은미), 「고령자의 건강한 노화를 위한 공동 디자인(Co-Design) 툴킷 개발에 관한 연구」(민주희, 구유리)는 상이한 연구 대상을 다루지만 이들 연구하는 방법과 절차를 기술한 내용이 상대적으로 우위를 이루어 지배적 토픽은 14번 토픽인 ‘분석, 유형, 분류...’로 모두 동일하다.

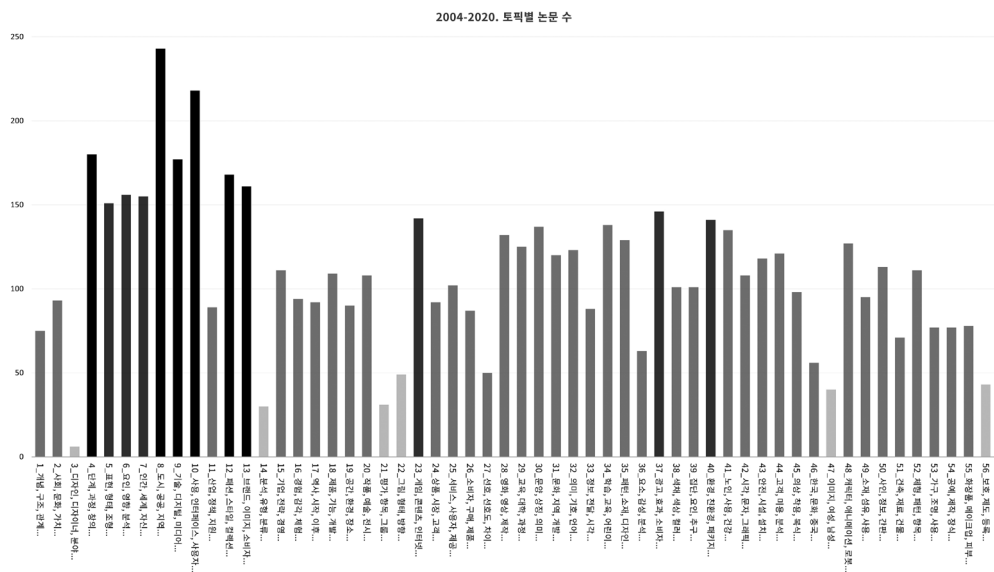


Figure 4 The number of papers by dominant topic (2004~2020, Total)

전체 기간을 대상으로 했을 때 디자인 분야에서는 ‘도시, 공공, 지역...’ ‘사용, 인터페이스, 사용자...’를 지배적 토픽으로 한 논문 수가 가장 많았고 이어 ‘단계, 과정, 창의...’ ‘기술, 디지털, 미디어...’ ‘패션, 스타일, 컬렉션...’ ‘브랜드, 이미지, 소비자...’ ‘요인, 영향, 분석...’ ‘인간, 세계, 자신...’ ‘표현, 형태, 조형...’ ‘광고, 효과, 소비자...’ ‘게임, 콘텐츠, 인터넷...’ ‘환경, 친환경, 패키지...’ 토픽이 차례로 상위 20%를 차지했다. 상위 10%인 6개 토픽에 해당하는 논문이 전체에서 차지하는 비율은 18.9%이다.

토픽별 논문 수를 학술지별로 살펴 본 결과는 Figure 5~7이다. 각 토픽의 막대 색상은 앞서 전체 학술지 기준으로 설정한 상태와 동일하다. <디자인학연구>에서 상위 10%를 차지한 토픽은 순서대로 ‘사용, 인터페이스, 사용자...’ ‘단계, 과정, 창의...’ ‘도시, 공공, 지역...’ ‘제품, 기능, 개발...’ ‘역사, 시작, 이후...’이며 상위 10% 토픽을 다루는 논문 수는 전체 편수의 23.5%이며 상위 50% 토픽을 다루는 논문 수는 72.3%이다. <한국디자인포럼>에서는 ‘광고, 효과, 소비자...’ ‘요인, 영향, 분석...’ ‘기술, 디지털, 미디어...’ ‘캐릭터, 애니메이션, 로봇...’ ‘사용, 인터페이스, 사용자...’ ‘브랜드, 이미지, 소비자...’가 상위 10%를 이루었다. 전체 편수에서 상위 10% 토픽이 차지하는 비율은 23.7%, 상위 50% 토픽의 비율은 75.8%이다. <한국디자인문화학회지>의 경우에는 ‘도시, 공공, 지역...’ ‘패션, 스타일, 컬렉션...’ ‘고객, 미용, 분석...’ ‘기술, 디지털, 미디어...’ ‘패턴, 소재, 디자인...’이 상위 10%를 차지했으며 상위 10% 토픽과 상위 50% 토픽이 전체 편수에서 차지하는 비율은 20.5%와 70.3%로 앞의 두 학술지보다 조금 낮았다.

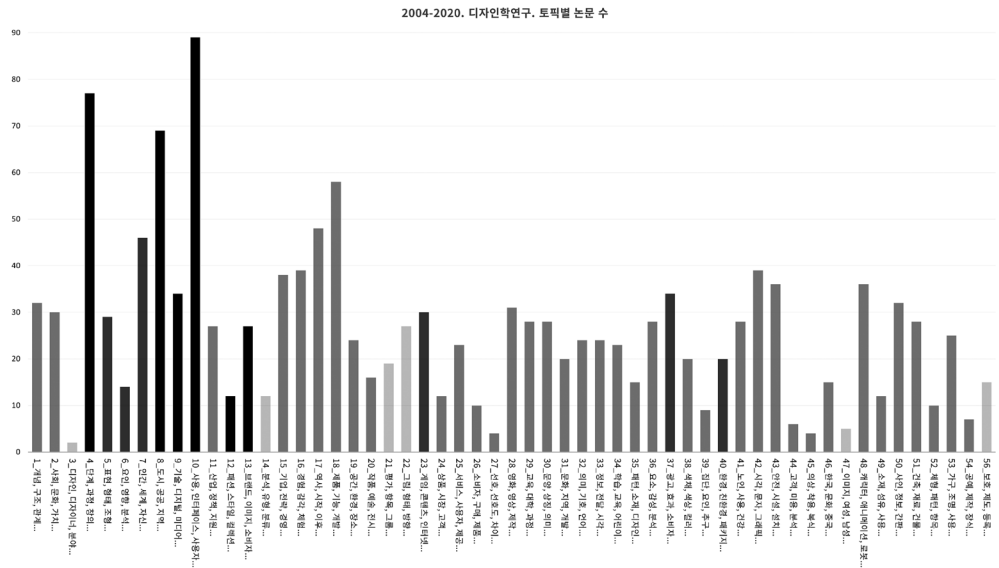


Figure 5 The number of papers by dominant topic (2004~2020, ADR)

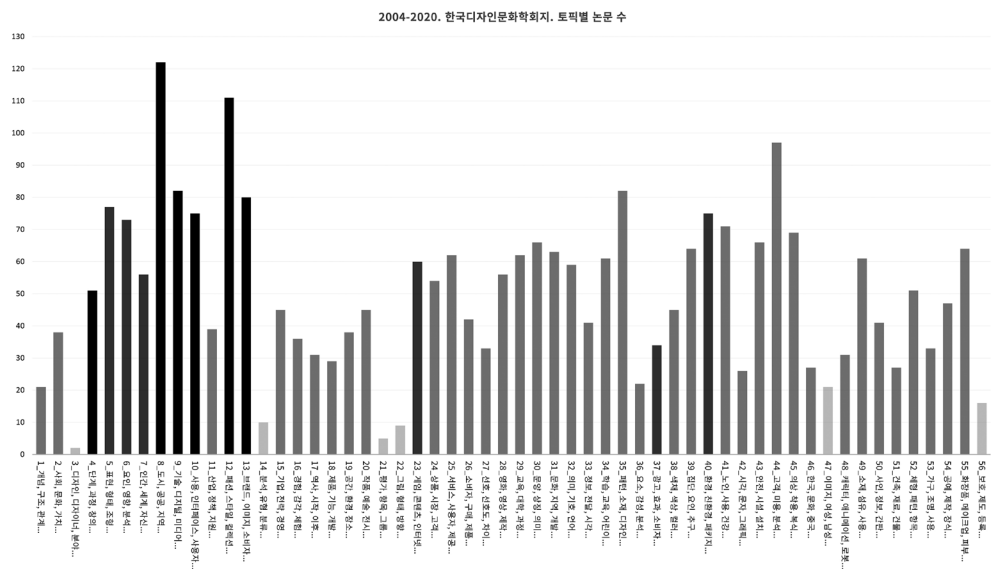


Figure 6 The number of papers by dominant topic (2004~2020, KSDC)

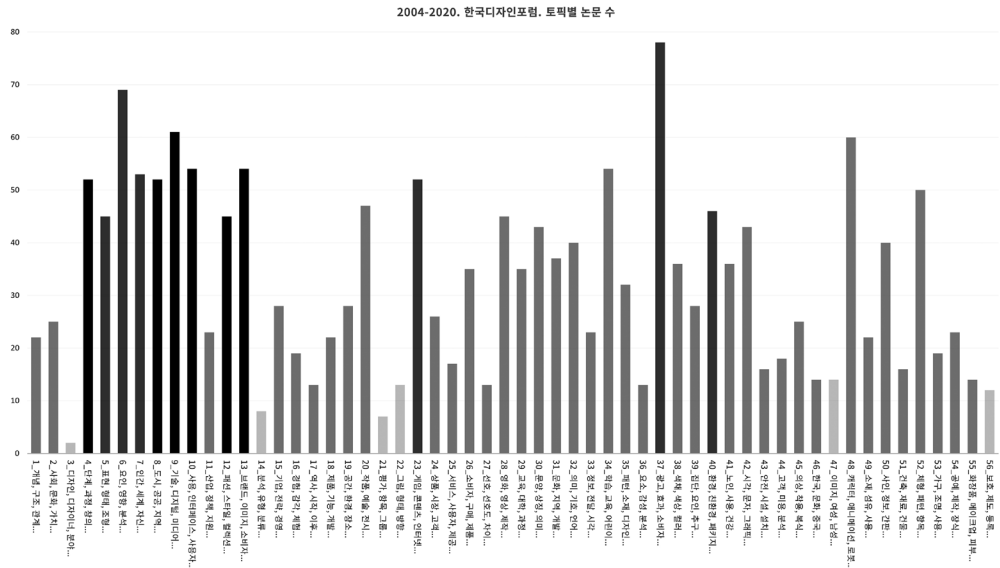


Figure 7 The number of papers by dominant topic (2004~2020, KDF)

4. 디자인 분야의 연구 주제 변화

3종 학술지에서 다뤄진 연구 주제가 시기별로 어떻게 바뀌어 왔는지 보기 위해 2004~2007년, 2008~2011년, 2012~2015년, 2016~2019년 등 4년 단위로 끊어서 지배적 토픽별로 논문 수를 비교했다. 2020년은 코로나19가 전 세계적으로 유행한 첫 해로 이전 시기와 분리하여 살폈다.

2004~2007년의 상위 10% 토픽은 ‘사용, 인터페이스, 사용자...’ ‘게임, 콘텐츠, 인터넷...’ ‘기술, 디지털, 미디어...’ ‘브랜드, 이미지, 소비자...’ ‘캐릭터, 애니메이션, 로봇...’으로 이 토픽을 다루는 논문 수가 전체의 28.1%를 차지했다. 상위 50% 토픽 비율은 전체 편수의 79.5%에 이른다. 하위 10%를 차지한 토픽은 ‘분석, 유형, 분류...’ ‘선호, 선호도, 차이...’ ‘이미지, 여성, 남성...’ ‘체형, 패턴, 항목...’ ‘화장품, 메이크업, 피부...’ ‘고객, 미용, 분석...’ ‘디자인, 디자이너, 분야...’이다.

2004~2007년은 노무현 정부가 집권한 시기로 2005년에는 이명박 전 서울시장의 주도하여 2003년 착공한 청계천복원사업이 마무리됐다. 전자출판(DTP) 방식이 아닌 전산조판 방식으로 제작한 마지막 교과서가 사용됐고 2007년에는 애플사 최초의 스마트폰인 아이폰이 출시됐다.

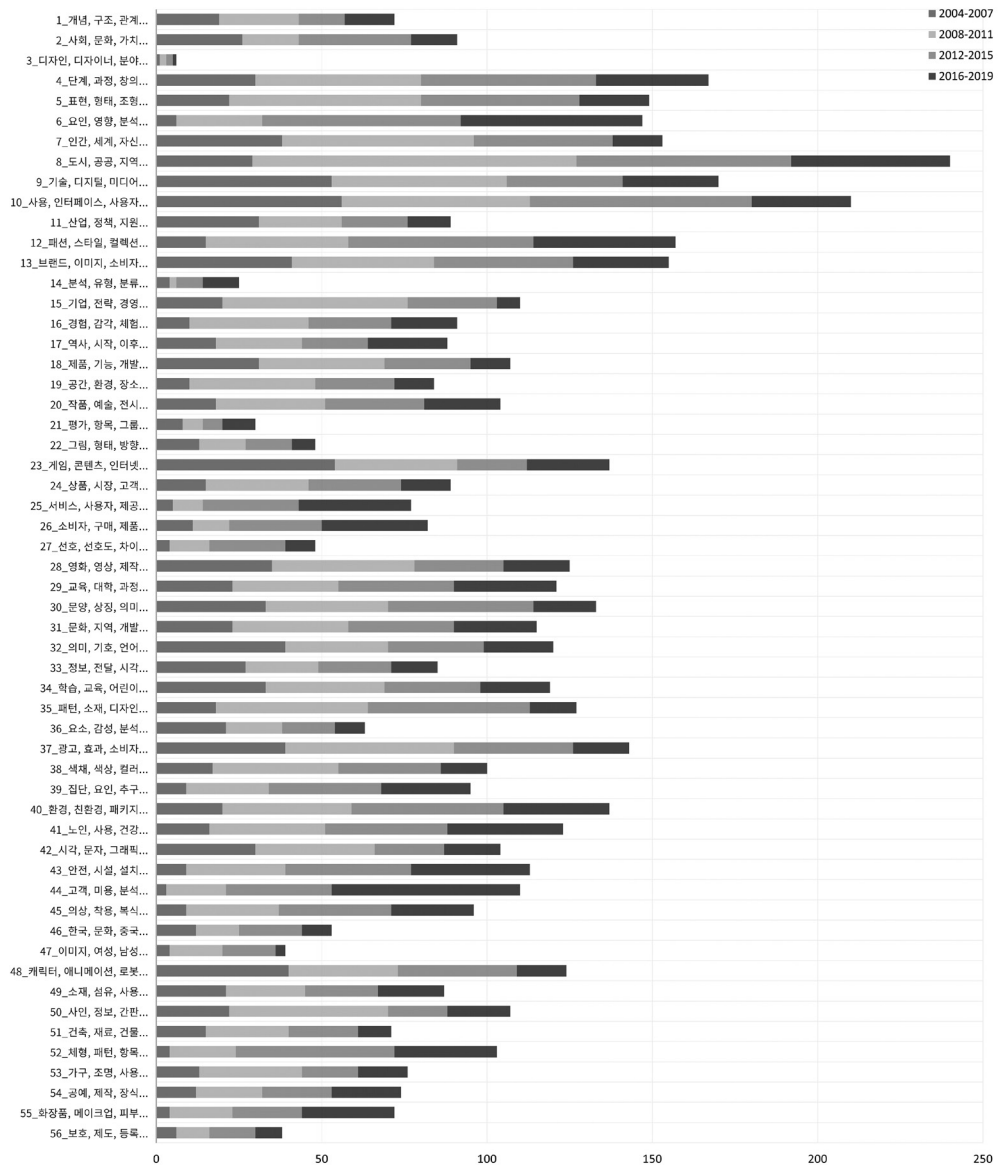


Figure 8 The cumulative number of papers by 56 dominant topics from 2004 to 2019 (unit: 4 years)

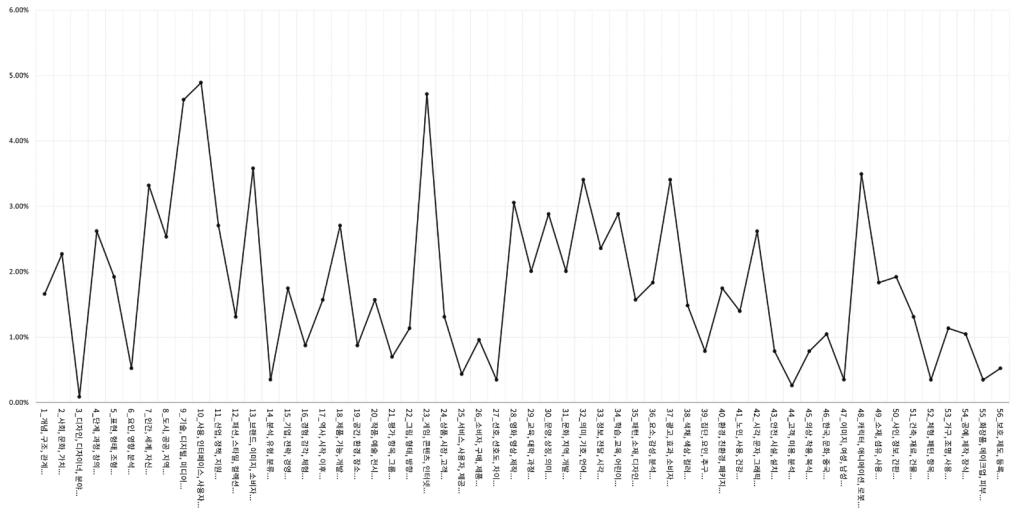


Figure 9 The proportion of papers by topic (2004~2007)

2008~2011년의 상위 10% 토픽은 ‘도시, 공공, 지역...’ ‘인간, 세계, 자신...’ ‘표현, 형태, 조형...’ ‘사용, 인터페이스, 사용자...’ ‘기업, 전략, 경영...’으로 이 토픽을 다루는 논문 수가 전체의 21.6%를 차지했다. 하위 10%를 차지한 토픽은 ‘보호, 제도, 등록...’ ‘서비스, 사용자, 제공...’ ‘평가, 항목, 그룹...’ ‘분석, 유형, 분류...’ ‘디자인, 디자이너, 분야...’이다.

2008~2011년은 이명박 정부가 집권한 기간이며 2010년 ‘세계디자인수도’ ‘유네스코 디자인 창의도시’로 서울시가 선정되는 등 오세훈 전 서울시장의 2007년부터 추진한 ‘디자인 서울’ 사업이 성과를 거둔 시기이다. 2009년 청소년보호법 개정안이 발의되며 게임 규제가 시작되었고 2011년에는 애플의 아이패드와 삼성전자의 갤럭시탭이 출시되면서 태블릿PC 시장이 급격히 성장했다.

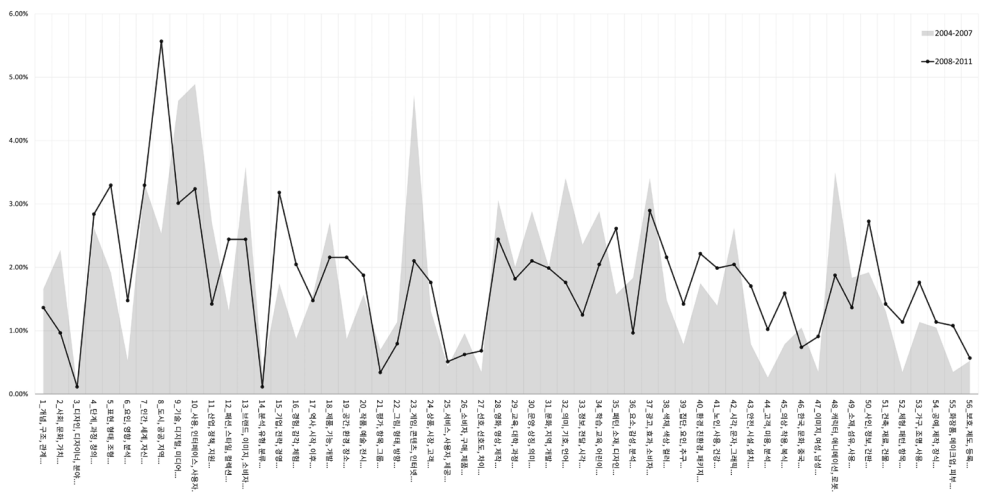


Figure 10 The proportion of papers by topic (2008~2011)

2012~2015년의 상위 10% 토픽은 ‘사용, 인터페이스, 사용자...’ ‘도시, 공공, 지역...’ ‘요인, 영향, 분석...’ ‘패션, 스타일, 컬렉션...’ ‘단계, 과정, 창의...’으로 이 토픽을 다루는 논문 수가 전체의 20.7%를 차지했다. ‘요인, 영향, 분석...’ 은 이 기간에 처음 상위 토픽으로 부상하여 2016~2019년에도 높은 비율을 유지했다. 이 기간에는 처음으로 상위 50% 토픽 비율이 전체 편수의 68%로 감소하고 하위 10% 토픽 비율은 3.4%로 증가했다. 이는 이 시기에 발표된 논문들이 앞 시기에 비해 조금 더 다양한 주제를 고르게 다루고 있음을

의미한다. 하위 10%를 차지한 토픽은 ‘개념, 구조, 관계...’ ‘그림, 형태, 방향...’ ‘보호, 제도, 등록...’ ‘분석, 유형, 분류...’ ‘평가, 항목, 그룹...’ ‘디자인, 디자이너, 분야...’이다.

2012~2015년은 이명박 정부의 임기가 마무리되고 박근혜 정부가 출범하여 집권한 시기이며 박원순 전 서울시장의 2011년 취임하여 임기를 수행한 기간이다. 2012년은 스마트폰 사용자 수가 피쳐폰 사용자 수를 처음으로 넘어선 해이다.

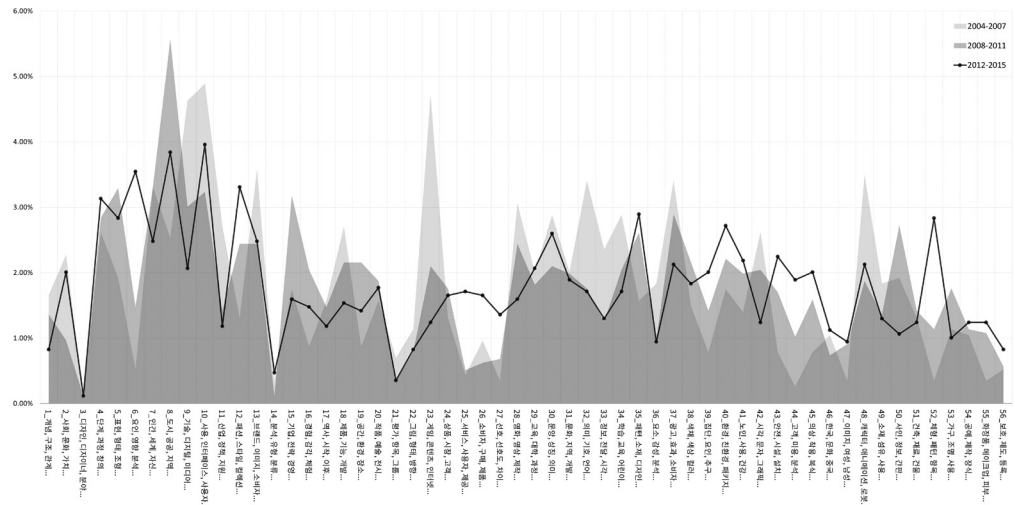


Figure 11 The proportion of papers by topic (2012~2015)

2016~2019년의 상위 10% 토픽은 ‘고객, 미용, 분석...’ ‘요인, 영향, 분석...’ ‘도시, 공공, 지역...’ ‘패션, 스타일, 컬렉션...’ ‘안전, 시설, 설치...’로 이 토픽을 다루는 논문 수는 전체의 22.8%를 차지했다. ‘안전, 시설, 설치...’가 상위 토픽으로 진입한 데에는 <디자인학연구>의 영향이 컸으며 ‘고객, 미용, 분석...’이 최상위 토픽으로 집계된 데에는 <한국디자인문화학회지>가 전적으로 영향을 미쳤다. 2016년에 논문 게재 규모가 크게 확대되면서 이 시기에 <한국디자인문화학회지>가 분석 결과에 미친 영향력은 이전보다 눈에 띄게 증가했다.

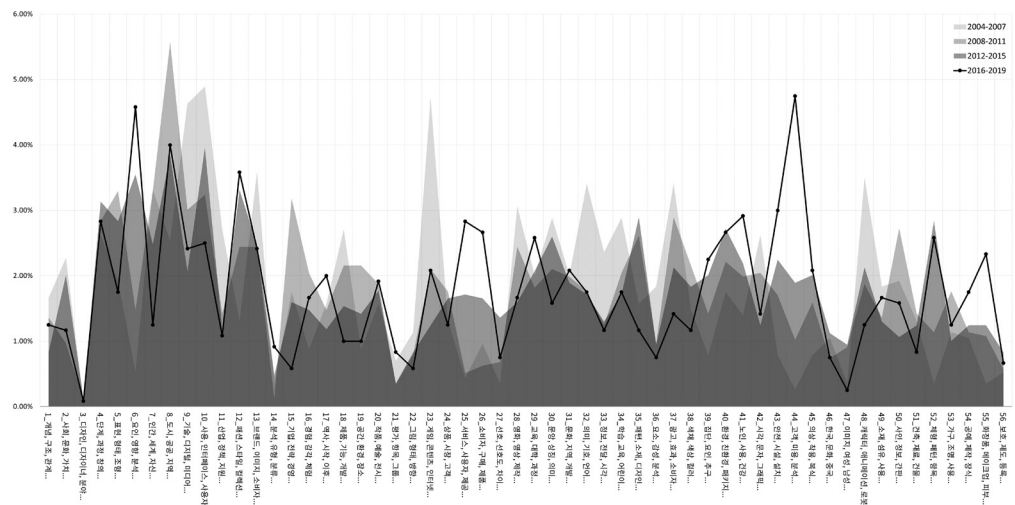


Figure 12 The proportion of papers by topic (2016~2019)

2016~2019년에 상위 50% 토픽 비율은 다시 전체 편수의 73% 수준으로 증가했으나 하위 10% 토픽 비율 역시 4.4%로 증가했다. 하위 10%를 차지한 토픽은 ‘보호, 제도, 등록...’ ‘그림, 형태, 방향...’ ‘기업, 전략, 경영...’ ‘이미지, 여성, 남성...’ ‘디자인, 디자이너, 분야...’이다. ‘기업, 전략, 경영...’은 2004~2011년 기간에는 상위 50%에 포함됐으나 2012~2015년 기간부터 하위 50% 구간으로 하락했다.

2016~2019년은 박근혜 전 대통령이 탄핵으로 파면되고 2017년 문재인 정부가 출범하여 집권한 시기이며, 2014년 재임된 박원순 전 서울시장이 도시재생 및 안전도시 서울 사업을 적극 추진했던 기간이다.

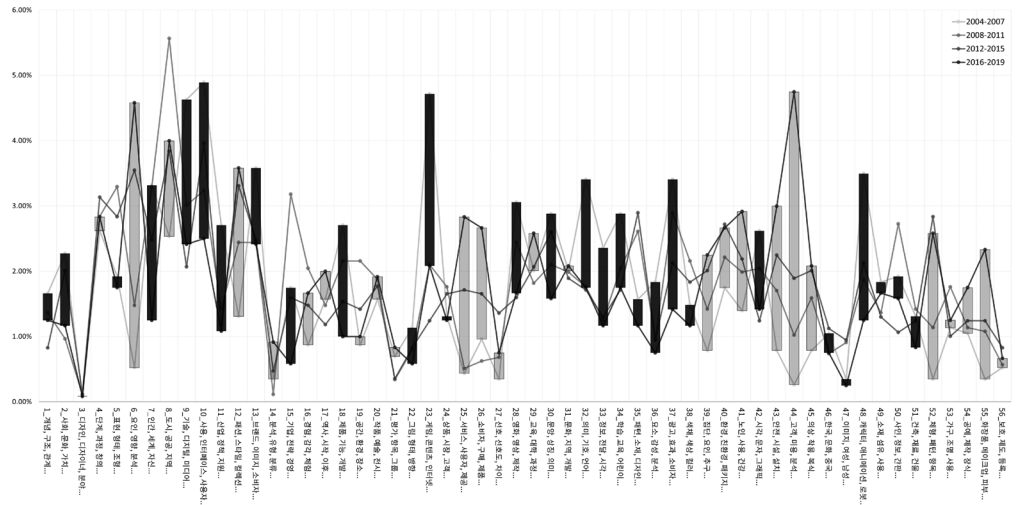


Figure 13 Changes in the proportions of papers by topic over time (2004~2019)

2004년부터 2019년까지 4년 단위로 토픽별 논문 비율을 비교한 결과는 Figure 13이다. 가장 앞 시기인 2004~2007년보다 가장 뒤 시기인 2016~2019년에 논문 비율이 증가한 토픽은 연한 회색 상자로, 비율이 감소한 토픽은 검은색 상자로 그 차이를 표시했다.

비율이 크게 증가한 토픽은 ‘고객, 미용, 분석...’ ‘요인, 영향, 분석...’ ‘서비스, 사용자, 제공...’ ‘패션, 스타일, 컬렉션...’ ‘체형, 패턴, 항목...’ ‘안전, 시설, 설치...’ ‘화장품, 메이크업, 피부...’ 등이며, 비율이 크게 감소한 토픽은 ‘게임, 콘텐츠, 인터넷...’ ‘사용, 인터페이스, 사용자...’ ‘캐릭터, 애니메이션, 로봇...’ ‘기술, 디지털, 미디어...’ ‘인간, 세계, 자신...’ ‘광고, 효과, 소비자...’ ‘제품, 기능, 개발...’ ‘의미, 기호, 언어...’ 등이다. 다만 ‘게임, 콘텐츠, 인터넷...’은 논문 비율이 큰 폭으로 감소하다가 마지막 시기에 소폭 증가했고 ‘의미, 기호, 언어...’는 2008~2011년에 논문 비율이 급격히 낮아졌으나 이후 비슷한 수치를 유지했다.

2008~2011년 논문 비율이 증가했던 토픽 중 ‘기업, 전략, 경영...’ ‘표현, 형태, 조형...’ ‘패턴, 소재, 디자인...’ ‘사인, 정보, 간판...’은 이후 다시 그만큼 수치가 감소하여 최종 변화폭은 크지 않다. ‘도시, 공공, 지역...’은 수치는 감소하였으나 규모 자체가 커서 지속해서 상위 10%에 포함됐다.

코로나19 팬데믹이 선언된 2020년에 3종 학술지에 게재된 한글 논문은 총 272편으로 상위 10% 토픽은 ‘서비스, 사용자, 제공...’ ‘학습, 교육, 어린이...’ ‘단계, 과정, 창의...’ ‘노인, 사용, 건강...’ ‘고객, 미용, 분석...’ ‘패션, 스타일, 컬렉션...’이다. ‘서비스, 사용자, 제공...’과 ‘학습, 교육, 어린이...’를 지배적 토픽으로 하는 논문 수는 전체의 16.2%로 높은 비중을 차지하는데, 이는 평소 대면으로 이루어지던 활동이 코로나19로 인해 모두 비대면으로 전환되면서 디자인 분야의 관심 역시 그와 관련된 주제로 향했기 때문일 수 있다. 실제 ‘서비스, 사용자, 제공...’ 토픽의 키워드에는 ‘코로나’가 포함됐다.

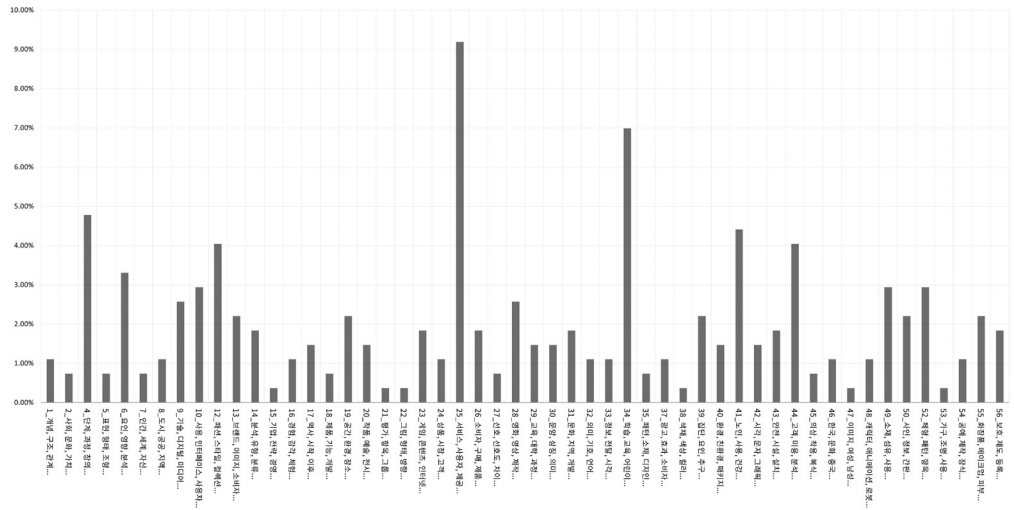


Figure 14 The proportion of papers by topic (2020)

다만 두 토픽이 예상치 못한 이변으로 갑자기 디자인 연구자의 주목을 받은 것은 아니다. ‘서비스, 사용자, 제공...’은 2004~2011년에는 하위 10% 부근에 위치했으나 2012~2015년에 상위 50% 이내에 진입하여 2016~2019년에는 상위 20%에 포함되었으며, ‘학습, 교육, 어린이...’는 전 기간 상위 50% 구간에 꾸준히 포함되었던 토픽이다.

학술지별로는 <디자인학연구>와 <한국디자인문화학회지>에서 ‘서비스, 사용자, 제공...’ ‘학습, 교육, 어린이...’ 토픽이 전체에서 차지하는 비율이 높았고, <한국디자인포럼>에서는 이 기간에 ‘단계, 과정, 창의...’ ‘브랜드, 이미지, 소비자...’ ‘공간, 환경, 장소...’ 등의 토픽이 두 토픽보다 더 상위 구간에 위치했다.

5. 결론

이 연구에서는 디자인 분야의 등재학술지 3종에 2004년부터 2020년까지 17년간 게재된 한글 논문 6,071편을 텍스트 처리한 뒤 토픽모델링 기법으로 연구 주제 동향을 분석했다. 현 학습 데이터는 56개 토픽으로 모델링했을 때 주제의 분화 정도가 가장 적절하고 결과 해석이 용이했다.

56개 토픽 중 4개는 연구 방법과 관련 있으며 나머지 52개는 연구 대상에 해당했다. 각 논문에 지배적으로 나타난 상위 10% 토픽은 ‘도시, 공공, 지역...’ ‘사용, 인터페이스, 사용자...’ ‘단계, 과정, 창의...’ ‘기술, 디지털, 미디어...’ ‘패션, 스타일, 컬렉션...’이며, 하위 10% 토픽은 ‘보호, 제도, 등록...’ ‘이미지, 여성, 남성...’ ‘평가, 항목, 그룹...’ ‘분석, 유형, 분류...’ ‘디자인, 디자이너, 분야...’이다.

4년 단위로 토픽별 논문 수가 변화하는 양상을 살폈을 때 상위 구간에 속하는 연구 주제는 대체로 그 시기의 정책이나 시대적 현안에 대응하여 변화하는 모습을 보였다. 다만 이것이 디자인 분야 일반에서 보이는 현상인지, 이런 현상이 어디에서 비롯되는지 알기 위해서는 별도의 후속 연구가 필요하다.

이 연구는 학술지 논문의 전문을 대상으로 토픽모델링 기법을 이용하여 디자인 분야의 연구 주제 동향을 분석하고자 한 첫 사례이다. 학술지 3종으로 분석 대상을 제한했기에 분석 결과를 디자인 분야 전반으로 일반화하기에는 무리가 있으며, 분석 대상으로 삼은 학술지의 개별 규모가 특정 시점 이후에 크게 달라져 한 학술지의 영향력이 지나치게 커진 점은 아쉬운 부분이다. 이런 한계는 향후 분석 대상 학술지를 한층 확대함으로써 해결할 수 있을 것이다.

그럼에도 이 연구는 융합학문인 디자인 분야의 연구자들이 그간 어떤 주제들을 다뤄왔고 시기별로 그 관심이 어떻게 변화해 왔는지 제한적으로나마 파악할 수 있다는 면에서 가치가 있다. 디자인 분야의 연구자들이

자신의 연구 주제를 검토하고 향후 연구 방향을 수립하는 데 이 연구가 기초 자료로서 활용될 수 있길 기대한다.

References

1. Iarrobino, M. (2017). *Text Mining Life Science Abstracts Vs. Full-Text Articles*. Retrieved December 10, 2021. from <https://www.bio-itworld.com/news/2017/12/20/text-mining-life-science-abstracts-vs-full-text-articles>
2. Jung, Y. (2012). *Research in Information Retrieval*. Seoul: Yonsei University Press
3. Kim, E. (2020). *A study of a model for design terminological dictionary based on academic corpus* (Doctorate). Hongik University Graduate School, Seoul
4. Lee, D. (2021). Convergence Analysis of Research Performance and Top Keywords in Korean "Design" Journals. *The Korean Society of Science & Art*, 39(3), 341-355

토픽모델링을 활용한 디자인 분야의 연구 주제 동향 분석

김은영

숙명여자대학교 시각영상디자인과, 강사, 서울, 대한민국

초록

연구배경 연구 동향 연구는 기존에 수행된 연구들을 검토하고 향후 연구의 방향을 설정하는 데 기초 자료가 된다. 그러나 국내 디자인 연구 전반을 대상으로 하는 연구 동향 연구에는 긴 공백이 있었고, 그에 해당하는 논문은 여전히 상당히 부족한 실정이다.

연구방법 2004년부터 2020년까지 17년간 한국의 디자인 종합 학술지 3종에 게재된 논문 총 6,071편을 학습 데이터로 하여 토픽모델링을 수행했다. 이후 LDA 모델에서 각 논문의 지배적 토픽을 추출하고 4년 단위로 토픽별 논문 수를 계량하여 디자인 분야의 연구 주제가 어떻게 변화해 왔는지 살폈다.

연구결과 토픽모델링 실행 시 주제 일관성이 가장 높은 토픽 수는 56개였으며 그중 4개는 연구 대상이 아닌 연구 방법과 관련이 있었다. 지배적 토픽을 기준으로 논문 수를 비교했을 때 시기별로 상하위 구간의 토픽은 지속해서 변화했고 이는 그 시기의 정책과 시대적 현안에 대응한 결과로 보인다.

결론 이 연구는 제한적으로나마 디자인 분야의 연구자들이 다뤄 온 연구 주제가 무엇이며 시대에 따라 그 주제가 어떻게 변화해 왔는지 파악할 수 있다는 면에서 가치가 있다. 이 연구가 디자인 분야의 연구자들이 자신의 연구 주제를 검토하고 향후 연구 방향을 수립하는 데 기초 자료로 활용될 수 있길 기대한다.

주제어 디자인 연구 주제 동향, 디자인 말뭉치, 토픽모델링, 자연어 처리
