

컬처마이닝과 디지털 인문학: 데이터 시대의 문화 읽기

이준서*
jslee@sungkyul.ac.kr

* 성결대학교 글로벌물류학과 교수

1. 머리말

최근 인공지능(AI), 빅데이터, 클라우드 컴퓨팅 등 정보통신기술(ICT)의 비약적인 발전은 사회 전반의 구조와 일상을 근본적으로 변화시키고 있다. 이러한 변화는 학문 연구와 교육 영역에도 지대한 영향을 미치고 있으며, 특히 연구 방법론과 지식 생산 방식에서 새로운 가능성을 열어주고 있다. 연구, 창작, 교육 등 다양한 학술 활동에서 ICT의 접목은 이미 활발히 진행되고 있으며, 그 중에서도 인문학은 디지털 기술을 수용하면서 디지털 인문학(Digital Humanities)이라는 새로운 연구 영역을 개척하게 되었다.

디지털 인문학은 전통적인 인문학의 해석적 깊이를 유지하면서도 데이터 과학적 방법론을 도입하여, 텍스트, 이미지, 영상, 음악 등 다양한 문화 자료를 체계적으로 분석할 수 있도록 한다. 김현(2013:12)은 이를 “정보기술(Information Technology)과 인문학이 조력하여 새로운 방식으로 수행되는 인문학 연구와 교육, 그리고 창조적 저작 활동”으로 정의한 바 있다. 오늘날 디지털 인문학은 단순한 도구적 차원을 넘어, 인문학의 세부 학문영역으로까지 정착하면서 인문학의 연구 지평을 확장하고 있다. 나아가 디지털 인문학은 인문학의 학제적 성격을 강화하고, 사회적·산업적 영역과의 연계를 통해 인문학의 실용성을 강화하고 그 역할을 확장하는 중요한 매개체로 기능한다.

이와 병행하여 구전 커뮤니케이션 연구 역시 학문적 발전을 이어왔다. 1950년대 사회적 커뮤니케이션 패턴 연구에서 출발한 구전 연구는 사회언어학을 중심으로 전개되다가, 1980년대 이후 마케팅, 소비자행동학 등 다양한 분야로 그 적용 범위를 확대하였다. 특히 디지털 환경이 일상화된 오늘날, 구전 커뮤니케이션은 온라인 구전(eWOM, electronic Word of Mouth)의 형태로 주목받고 있다. 신상희 & 황복주(2015:436)는 eWOM을 “디지털 네트워크를 통해 전달되는 소비자 간의 구전”으로 정의하였으며, 이는 소비자 행동, 사회적 담론, 정치·경제적 현상 등 광범위한 영역에서 중요한 영향을 미치고 있다.

COVID-19 팬데믹은 이러한 변화를 더욱 가속화시켰다. 비대면 환경의 확산과 디지털 전환(Digital Transformation, DX)의 심화로 인해 SNS와 온라인 커뮤니케이션이 일상생활에 완전히 정착하면서, 방대한 규모의 빅데이터가 생성·축적되었다. 그 결과, 온라인 구전은 사회적 의사소통의 핵심 채널로 자리 잡았으며, 이를 둘러싼 연구는 문화 비교, 여론 분석, 사회적 감정 변화 추적 등 다양한 학문적 주제로 확장

되고 있다.

그러나 이러한 빅데이터 환경 속에서 기존의 담론분석(discourse analysis)은 뚜렷한 한계를 드러낸다. 담론분석은 기능문법, 심리언어학, 사회언어학, 화용론 등 여러 연구 전통을 포괄하며 1980년대 독일의 텍스트언어학과 미국의 담화 인지언어학을 기반으로 활발히 전개되어 왔다. 최근에도 담론분석은 정치, 언론, 교육, 문화 전반에 걸쳐 적용되며 연구의 다양성을 보여주고 있다. 그러나 기존 담론분석은 특정한 담론 집단을 심층적으로 이해하는 데에는 강점을 지니지만, 데이터의 규모가 방대해지고 다언어·다문화 환경이 일반화된 오늘날의 연구 조건에는 충분히 대응하기 어렵다. 정성적 분석의 강점을 유지하면서도 대규모 데이터를 포괄할 수 있는 새로운 분석 방법론이 요구되고 있는 것이다.

이러한 맥락에서 컬처마이닝(Culture Mining)은 주목할 만한 대안적 방법론으로 부상한다. 컬처마이닝은 방대한 다언어 데이터 속에서 반복적이고 차별적인 문화 요소(cultural element)¹⁾를 추출하고, 이를 사회적·역사적 맥락 속에서 분석함으로써 문화적 패턴을 규명하는 기법이다.

문화요소는 특정 사회와 집단의 정체성을 드러내는 핵심 단위로서, 언어, 담론, 상징, 가치 체계 등 다양한 형태로 나타난다. 컬처마이닝은 단순한 데이터 분석을 넘어, 다문화·다언어 환경 속에서 문화의 특성과 차이를 정량적으로 포착하고, 이를 종합적으로 이해할 수 있도록 돋는 방법론으로, 특히 온라인 구전 데이터를 분석함으로써 사회적 담론의 흐름과 정서적 변화를 추적할 수 있으며, 이를 통해 문화 비교 연구, 사회 문제 탐구, 산업적 적용 등 다양한 가능성을 제시한다.

이에 본 연구는 디지털 인문학적 접근법의 하나로서 컬처마이닝 기법을 다국어 온라인 구전 커뮤니케이션 분석에 적용하는 방안을 모색하고 기존의 인문학적 담론 분석을 보완하여, 데이터 시대에 걸맞은 새로운 문화 연구 패러다임을 제안하는 데 목적이 있다.

2. 디지털인문학과 컬처마이닝

1) ‘문화요소’는 연구자에 따라서 다양한 개념으로 사용되고 있다. Hyatt, J. & Simons, H.(1999)는 ‘구성원들이 적절하다고 여기는 상징과 의미체계’, Baker(2005)는 ‘문화상 특수한 것’, 이근희(2015)는 ‘문화관련어휘(cultural-bound), 즉 기점언어를 사용하는 사회공동체의 역사, 사회, 경제, 정치, 언어습관 등을 포함하는 특정 문화에서 파생된 어휘’라고 정의하고 있다.

디지털 인문학(Digital Humanities)과 컬처マイ닝(Culture Mining)은 인문학 연구에 빅데이터와 인공지능 기술을 접목하는 과정에서 밀접한 관계를 맺고 있다. 디지털 인문학은 전통적인 인문학 연구 방법론에 데이터 분석, 머신러닝, 자연어 처리(NLP) 등의 디지털 기술을 적용하여 텍스트, 이미지, 영상, 음악 등의 문화 콘텐츠를 체계적으로 연구하는 학문 분야라고 할 수 있다.

한편, 서로 다른 문화권과의 대조와 비교를 통하여 차별화된 문화요소를 찾아내는 컬처マイ닝은 다양한 빅데이터 분석 기술을 활용하여 다중언어로 이루어진 대규모 문화 데이터에서 의미 있는 패턴을 발견하고, 사회적·역사적 맥락에서 문화적 변화와 흐름을 분석하는 기법이다.

Culture Mining is the process of finding useful or interesting ‘cultural elements’ -life styles, patterns, fashions, trends, models, beliefs, rules, frames etc.- of a specific region or a generation from unstructured text, various image sets by comparing different languages and cultures.

컬처マイ닝이란 인간의 문화를 표현하고 있는 다양한 미디어(텍스트, 이미지, 동영상 등)를 분석하여 언어/지역/성별/세대별 문화를 구성하는 문화 요소와 그들간의 관계를 발견해가는 과정이다.

이준서(2021)

컬처マイ닝은 디지털 인문학의 핵심 방법론 중 하나로 활용될 수 있다. 예를 들어, 컬처マイ닝 기술을 활용하여 역사적 문헌, 신문 기사, 문학 작품, 영화, SNS 데이터 등을 분석하여 특정 시대의 문화적 트렌드, 감성 표현, 가치관의 변화를 연구할 수 있다. 특히, 감성 분석(sentiment analysis), 키워드 추출, 네트워크 분석 등을 통해 특정 시기의 사회적 이슈나 담론이 어떻게 형성되고 변화했는지를 파악할 수 있는데, 컬처マイ닝은 인공지능을 활용한 자동화된 데이터 수집 및 분석 기법을 통해 디지털 인문학 연구의 효율성을 높이는데 공헌할 수 있는 것이다.

기존 인문학이 주로 소규모 문헌이나 제한된 텍스트를 중심으로 정성적 연구를 수행해 왔다면, 컬처マイ닝은 방대한 규모의 빅데이터를 기반으로 보다 객관적이고 광범위한 문화적 패턴을 도출할 수 있다는 점에서 차별성을 지닌다. 즉, 전통적 인

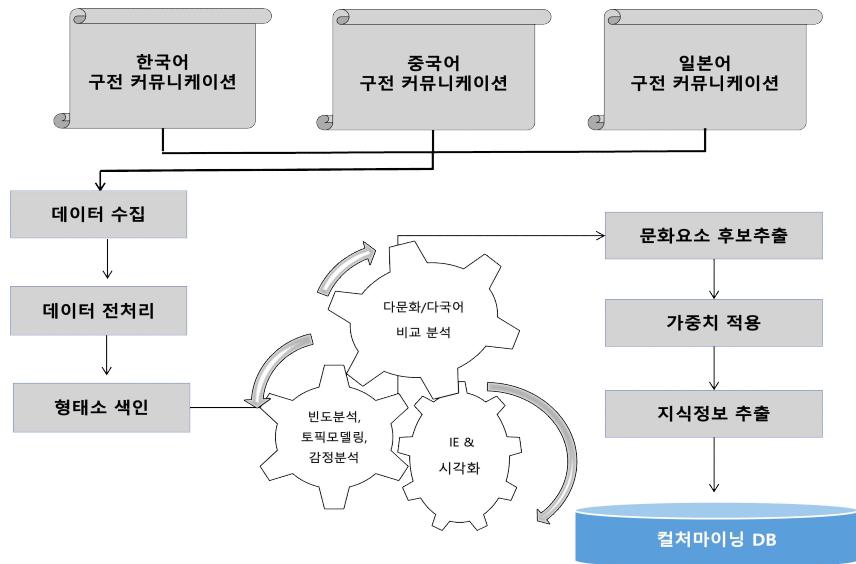
문학이 깊이 있는 해석을 통해 미시적 차원의 의미를 밝혀왔다 면, 컬처마이닝은 데이터 과학적 기법을 활용하여 거시적 수준에서 문화적 흐름과 패턴을 파악하는 데 강점을 보인다. 이러한 점에서 디지털 인문학과 컬처마이닝은 상호보완적인 관계를 이루며, 빅데이터 시대의 인문학 연구가 나아가야 할 새로운 패러다임을 제시하고 있다.



<그림1> 컬처마이닝의 부가가치 창출 개요

컬처마이닝은 단순히 문화적 패턴을 식별하는 것을 넘어, 재화나 서비스가 지닌 경제적 가치뿐 아니라 소비자의 정서를 자극하여 공감과 감동을 이끌어내는 감성가치까지 포착할 수 있다. 오늘날 문화 소비는 단순한 상품 구매 행위가 아니라, 개인의 감정과 정체성, 사회적 연대감이 결합된 총체적 경험으로 이해되고 있다. 따라서 컬처마이닝은 언어·문화 데이터 속에서 드러나는 정서적 코드와 상징적 표현을 분석함으로써, 기존의 경제학적 가치 평가로는 설명하기 어려운 감성적·문화적 가치를 규명할 수 있다. 컬처마이닝을 통해 소비자 담론 속에서 반복적으로 나타나는 감성적 키워드와 문화적 코드를 파악하여, 제품·서비스에 새로운 부가가치를 부여 할 수 있다. 다시 말해, 컬처마이닝은 데이터 기반의 문화 분석을 통해 상품의 가치를 고도화하고, 소비자와의 정서적 연결을 강화하는 전략적 도구로 기능할 수 있다. 이러한 과정은 단순히 시장에서의 경쟁력을 높이는 것을 뛰어넘어, 사회적·문화적 차원에서 소비자 경험을 확장하는 데 기여할 수 있는 것이다.

3. 다국어 컬처마이닝 분석



<그림2> 다국어 컬처마이닝 프로세스 개요

<그림 2>는 본 연구에서 제안하는 컬처마이닝(Culture Mining) 분석 절차를 도식적으로 나타낸 것이다. 전체 과정은 데이터 확보 → 전처리 및 언어 분석 → 통계적·계량적 분석 → 비교분석 및 시각화 → 문화요소 추출 및 지식화라는 일련의 단계를 통해 체계적으로 전개된다.

우선 출발점은 한국어, 중국어, 일본어 등 다국어 구전 커뮤니케이션 데이터의 확보이다. 온라인 커뮤니케이션 데이터는 특정 국가나 언어권의 사회적 감정, 생활양식, 소비 태도, 가치관 등을 압축적으로 반영하기 때문에 문화 비교 연구의 출발점이 된다. 데이터는 SNS, 블로그, 온라인 커뮤니티, 디지털 기사 등에서 수집되며, 크롤링 도구나 API를 통해 확보된다.

수집된 원시 데이터는 그대로 활용될 수 없기 때문에 전처리 과정이 필수적이다. 전처리 단계에서는 HTML 태그, URL, 이모티콘, 광고 문구와 같은 불필요한 요소를 제거하고, 중복 데이터 및 스팸성 자료를 걸러낸다. 또한 한국어·중국어·일본어라는 언어적 특성을 고려해 각각의 형태소 분석기를 활용하여 단어를 표제어 형태로 정규화하고, 불용어(stopwords)를 제거함으로써 분석 가능한 코퍼스를 구축한다.

이후 번도분석, 토픽모델링, 감성분석 등의 빅데이터 분석 기법이 활용된다. 번도

분석은 특정 단어가 얼마나 자주 등장하는지를 통해 핵심 키워드를 파악하고, 토픽 모델링은 LDA와 같은 통계적 기법을 통해 담론 속에서 잠재적으로 형성된 주제들을 군집화한다. 감성분석은 긍정·부정·중립의 감정을 수치화하여 시계열적 변화를 추적할 수 있게 한다. 이러한 분석 결과는 단순히 언어 데이터의 구조를 보여주는 것을 넘어, 사회적 이슈와 문화적 맥락이 데이터 속에서 어떻게 반영되는지를 보여준다.

분석 결과는 다문화·다언어 비교 분석 단계에서 종합적으로 활용된다. 동일한 주제에 대해 한국어, 중국어, 일본어 담론이 어떻게 다른 의미망을 형성하는지 비교함으로써, 각 문화권의 특성과 차이점을 정량적으로 드러낼 수 있다.

다음 단계는 문화요소 후보 추출이다. 빈도·토픽·감성 결과를 바탕으로 반복적으로 나타나는 핵심 개념과 상징적 어휘가 문화요소 후보로 제시된다. 그러나 모든 요소가 동일한 중요도를 가지는 것은 아니므로, 가중치 적용 단계를 거쳐 요소별 상대적 중요도가 평가된다. 가중치는 출현 빈도, 담론 내 연결 중심성, 감성 편향 정도 등을 종합적으로 고려하여 부여된다.

최종적으로, 가중치가 적용된 문화요소는 지식정보 추출 단계를 통해 구조화된다. 이 단계에서 정보추출(Information Extraction, IE) 기법과 데이터 시각화가 활용된다. 워드클라우드, 토픽맵, 네트워크 다이어그램, 시계열 그래프 등을 통해 분석 결과가 시각적으로 표현되며, 이는 연구자가 담론 구조와 감정 흐름을 직관적으로 이해하고, 해석의 타당성을 검증하는 데 기여한다.

이상의 모든 결과는 컬처마이닝 데이터베이스(DB)에 저장되어 축적된다. 데이터 베이스는 단일 연구의 결과물로 끝나지 않고, 후속 연구에서 재활용되고, 다양한 비교분석에 활용될 수 있는 지식 자산으로 기능한다. 나아가 해당 DB는 국가별, 언어별, 시기별 문화 차이를 종합적으로 기록하고, 학문적·실용적 차원에서 중요한 참고 자료로 활용될 수 있다.

3.1. 기초 빈도 분석

텍스트 마이닝 분석 기법 중에서 가장 직관적이고 널리 활용되는 방법 중의 하나가 빈도 분석이다. 빈도 분석은 문서에서 특정 단어가 얼마나 자주 등장하는지를 분석하고, 이를 다양한 방식으로 시각화하여 해당 데이터의 의미를 도출하는 기법이다.

<그림3> 주요 빈도분석 비교 워드클라우드 예시2)

<그림3>의 워드클라우드는 단어 빈도(Term Frequency, TF)에 근거해서 특정 단어가 문서 내에서 얼마나 자주 등장하는지를 나타내는 것이다. TF 값이 높을수록 중앙에 위치하여 해당 단어가 문서에서 많이 사용되었음을 시각적으로 확인할 수 있다.

단순히 단어 빈도만으로는 특정 문서에서 중요한 단어를 식별하기 어려운 한계가 있다. 예를 들어 접속사, 조사, 인칭사 등과 같은 단어들은 여러 문서에 자주 등장할 수 있지만, 문서의 핵심 주제를 나타내지는 않는다. 이러한 한계를 보완하기 위하여 IDF(Inverse Document Frequency, IDF)가 도입되었는데, IDF는 특정 단어가 전체 문서에서 얼마나 희귀한지를 측정하는 지표라고 할 수 있다. 문서 전체에서 자주 등장하는 단어는 낮은 IDF 값을 가지며, 특정 문서에서만 등장하는 단어는 높은 IDF 값을 가진다. 이를 통해 일반적인 단어보다는 문서의 주요 내용을 나타내는 단어에 더 높은 가중치(weight)를 부여할 수 있다.

<표1> 주요 한·일 구전 커뮤니케이션 TF-IDF

2) 이준서(2025:41)

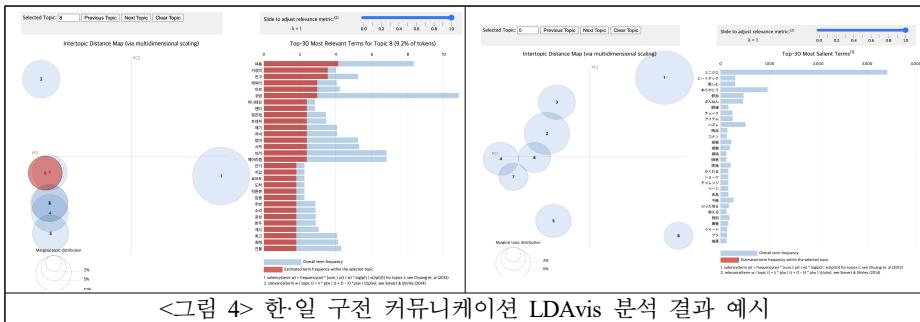
No	한국어	TF-IDF	일본어	TF-IDF
1	티셔츠	4.376889057	キャンペーン	4.933579806
2	불매	4.106571681	応募	4.71627095
3	히트	4.059413739	Tシャツ	3.648862989
4	바지	3.972512644	コラボ	3.271917905
5	일본	3.946705267	サイズ	3.038166501
6	한국	3.939209139	ヒートテック	3.030120459
7	셔츠	3.909981901	日本	3.028256809
8	콜라보	3.907816686	店舗	2.924873511
9	가격	3.806933299	コラボ	2.850767766
10	가방	3.621877537	パンツ	2.636949572

TF와 IDF를 결합한 값이 <표 1>의 TF-IDF이다. 특정 문서에서 자주 등장하면서도 전체 문서에서는 비교적 드물게 등장하는 단어일수록 TF-IDF 값이 높아진다. TF-IDF의 가중치를 적용함으로써 문서 내에서 중요한 단어를 찾아내는 데 매우 유용해, 문서에서 핵심적인 내용을 추출하는 데 효과적인 기법으로 자리 잡고 있다.

3.2. 토픽모델링

토픽모델링(topic modeling)은 텍스트마이닝 기법 중 하나로, 구조화되지 않은 대규모 문서 집합에서 잠재된 주제를 식별하는 확률적 알고리즘이다. 이 방법을 활용하면 방대한 텍스트 데이터 속에서 핵심 주제를 자동으로 도출하고, 주제별로 문서를 분류할 수 있다.

토픽모델링 기법 가운데 가장 널리 사용되는 것이 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)이다. LDA는 문서 내 단어의 출현 확률과 분포를 분석하여 주제를 추론하는 통계적 텍스트 처리 기법으로, 특정 단어들이 함께 나타나는 패턴을 기반으로 잠재적인 주제 집합을 형성한다.



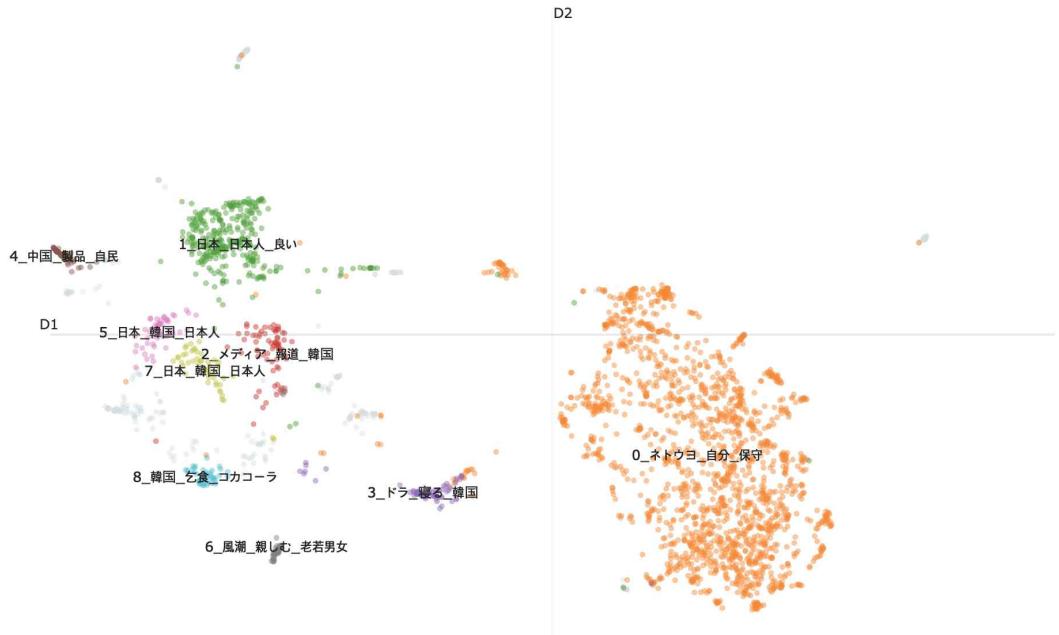
<그림 4> 한·일 구전 커뮤니케이션 LDAvis 분석 결과 예시

<그림 4>는 이준서(2025)에서 F사 SPA 브랜드에 관한 한·일 구전 커뮤니케이션에 대하여 최적화된 토픽의 수를 결정하기 위해 LDA 토픽모델링 결과에 LDAvis³⁾ 알고리즘을 적용해 시각화한 것이다. 이를 통해 문서 집합에 내재된 주제를 효과적으로 군집화하고, 해당 주제와 관련된 주요 담론의 흐름을 파악할 수 있다.

다만 LDA 방식은 단어의 동시 출현 확률에 기반하기 때문에 문맥적 의미나 단어 간의 미묘한 관계를 충분히 반영하기 어렵다는 한계를 지닌다. 예를 들어, 동일한 단어가 상황에 따라 다른 의미를 가질 수 있는데, LDA는 이를 구분하지 못하는 경우가 많다.

이러한 한계를 보완하기 위해 최근에는 BERT(Bidirectional Encoder Representations from Transformers) 기반의 토픽모델링이 주목받고 있다. BERT는 사전 학습 언어모델을 활용해 단어가 사용된 문맥을 인식할 수 있으므로, 동일한 단어라도 문맥에 따라 달라지는 의미를 보다 정밀하게 포착할 수 있다. 따라서 BERT 기반 토픽모델링은 LDA의 통계적 접근에 문맥적 이해를 결합함으로써, 자연어의 복잡성을 반영하는 더욱 정교한 주제 추출이 가능하다.

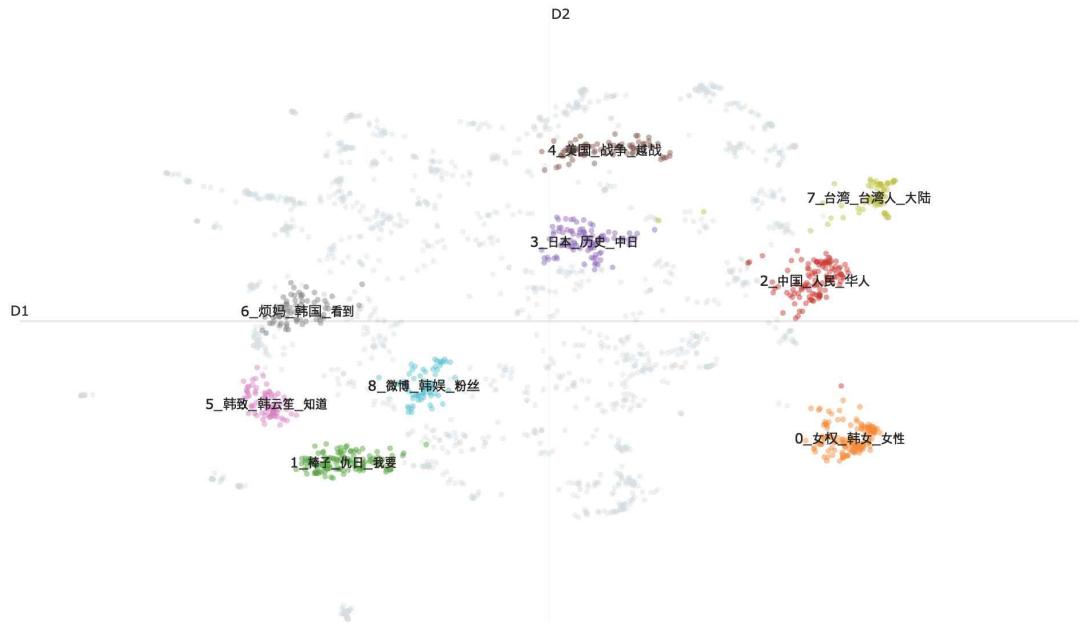
3) LDAvis는 토픽 모델링 결과를 웹 기반으로 시각화하여 제공하는 도구로 ‘Intertopic Distance Map’은 학습된 토픽 모델을 2차원 공간에 배치하여 각 토픽 간의 관계와 출현 확률((prevalence)을 분석할 수 있도록 한다. 각 토픽은 원으로 표시되며, 출현 확률이 높을수록 원의 크기 커진다. 또한, 토픽 간 거리가 가까울수록 연관성이 높은 반면, 거리가 멀어질수록 연관성이 낮음을 의미한다.



<그림 5> 일본어 BERT 기반 클러스터링 결과 예시⁴⁾

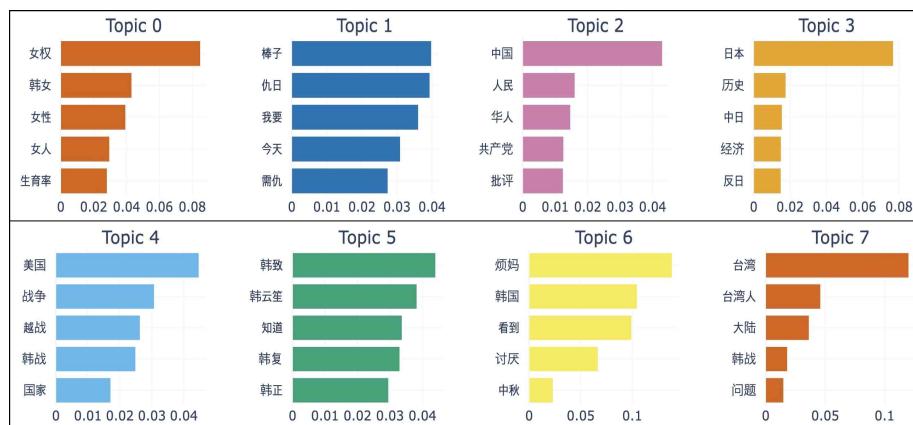
<그림 5>는 이준서(2025)에서 BERT 기반 임베딩을 활용하여 일본어 트위터 데이터를 2차원 공간에 시각화한 결과이다. 각 점은 개별 텍스트 단위를 의미하며, 동일한 색상의 점들은 의미적으로 유사한 텍스트들이 군집화된 하나의 토픽 클러스터를 나타낸다. 이러한 클러스터는 텍스트가 공유하는 의미 구조를 반영하므로, 특정 집단의 담론이 어떠한 주제를 중심으로 형성되는지를 직관적으로 확인할 수 있다. 나아가 각 클러스터를 중심으로 핵심 토픽과 주요 키워드를 분석하면, 데이터 집합에 내재된 주제적 경향성과 담론 구조의 특징을 보다 심층적으로 탐구할 수 있다.

4) 이준서(2025:72)에서 BERT 기반 토픽모델링 결과를 UMAP(Uniform Manifold Approximation and Projection) 차원 축소 기법을 활용하여 2차원 평면에 시각화한 것이다. 이를 통해 온라인 담론 내에서 주요 주제들이 어떻게 분포하고 군집화되는지를 시각적으로 확인할 수 있다.



<그림 6> 중국어 BERT 기반 클러스터링 결과

<그림 6>은 중국어 SNS 데이터를 대상으로 일본어 데이터와 동일한 분석 절차 (BERT 임베딩, UMAP 차원 축소, HDBSCAN 클러스터링)를 적용한 결과이다. 각 문서는 2차원 임베딩 공간에서 의미적으로 유사한 문서들과 함께 하나의 클러스터를 형성하며, 이를 통해 데이터 집합 내 잠재적 주제 구조가 시각적으로 드러난다.



<그림 7> 중국어 토픽별 주요 키워드⁵⁾

또한 일본어 데이터 분석과 동일한 방식으로 주요 토픽을 도출하였으며, <그림 7>은 이러한 클러스터별로 추출된 상위 키워드를 제시한 것이다. 이를 통해 각 군집이 반영하는 주제적 특징과 담론의 구조적 차이를 보다 명확히 파악할 수 있다.

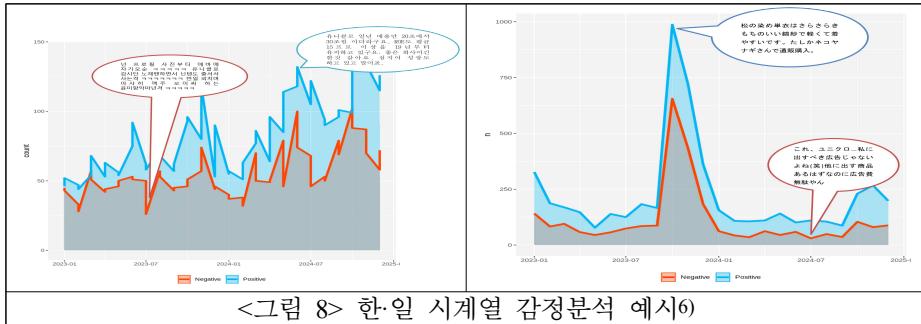
3.3. 감성분석

감성 분석(Sentiment Analysis)은 텍스트 마이닝 기법 중 하나로, 특정 문서에 포함된 긍정 또는 부정의 감정 표현을 추출하고 이를 바탕으로 문서 전체의 감성 경향을 분류·분석하는 방법이다. 일반적으로 문서를 구성하는 최소 단위인 형태소에 감성 극성(Sentiment Polarity)을 부여한 감성 어휘 사전을 기반으로 분석이 이루어진다. 예를 들어, 한국어의 경우 KNU 한국어 감성사전을 활용하는데, 이는 표준국어 대사전의 어휘를 긍정·부정·중립으로 분류하고 다수 평가자의 합의를 통해 구축된 것이다. 일본어의 경우에는 도쿄공업대학(東京工業大學)의 ‘단어감정극성대응표’가 널리 사용되며, 어휘 네트워크를 기반으로 자동 극성값을 부여하는 방식으로 구성되어 있다.

중국어의 경우에도 다양한 감성사전이 구축되어 있으며, 대표적으로 NTUSD (National Taiwan University Sentiment Dictionary) 가 활용된다. NTUSD는 대규모 중국어 어휘를 긍정·부정으로 분류하여 감성 극성을 부여한 사전으로, 소비자 리뷰 분석, 소셜 미디어 텍스트 분석 등 다양한 분야에서 널리 응용되고 있다. 이러한 사전 기반 감성 분석은 언어별 특성을 반영하면서도 공통적으로 문서 내 표현의 감정적 성격을 정량화하여 분석할 수 있게 한다.

나아가 시간의 흐름에 따라 기록된 시계열 데이터를 함께 분석하면, 특정 주제에 대한 여론의 추세(trend), 계절성(seasonality), 주기성(cycle) 등의 패턴을 발견할 수 있다. 이를 통해 단순한 정태적 감정 분류를 넘어, 사회적 담론이나 소비자 여론이 시간에 따라 어떻게 변화하는지를 동태적으로 파악할 수 있다.

5) 이준서(2025:73).



<그림 8>은 최근 수년간의 온라인 담론을 시계열로 분석한 결과를 보여준다. 궁정적(파란색)·부정적(빨간색) 감성 언급량을 월별로 집계하여 추세를 비교한 것으로, 특정 주제를 둘러싼 정서적 양상이 시간의 흐름에 따라 어떻게 변화하는지를 시각적으로 확인할 수 있다.

<그림 8>에서 한국어 데이터의 시계열 감정분석을 살펴보면, 전체적으로 긍정·부정 언급 모두 완만하게 증가하는 장기적 추세가 나타난다. 긍정적 담론은 주로 기업의 성과, 사회적 신뢰, 성장 가능성과 같은 거시적 요인과 연관되었으며, 부정적 담론은 광고, 서비스 경험, 브랜드 이미지 등 일상적·문화적 요인에 의해 형성되는 경우가 많다. 이러한 결과는 한국어 담론에서 정서적 평가가 단순한 제품 차원을 넘어 사회적·경제적 맥락을 포함한다는 점을 보여준다.

반면 일본어 데이터의 시계열 감정분석에서는 특정 시점에 긍정·부정 언급이 동시에 급격히 증가하는 뚜렷한 피크가 나타난다. 이는 신제품 출시, 유명인 협업, 대규모 이벤트 등 단기적 마케팅 활동에 의해 여론이 집중적으로 반응한 결과로 볼 수 있다. 긍정적 담론은 제품 품질, 기능, 디자인과 같은 실용적 측면을 강조하는 반면, 부정적 담론은 홍보 전략이나 커뮤니케이션 방식에 대한 비판으로 나타나는 경우가 많다.

시계열 감정분석은 단순한 궁·부정의 변화뿐만 아니라, 추세(trend), 계절성(seasonality), 주기성(cycle)과 같은 패턴을 함께 탐구할 수 있다는 점에서 의의가 크다.

6) 이준서(2025:48-50)

다. 예컨대 장기적인 증가 추세는 특정 담론이 사회적으로 점차 확산되고 있음을 보여주며, 계절성은 특정 시기(예: 연말·명절·신학기)에 반복적으로 감성 표현이 집중되는 현상을 드러낸다. 또한 주기성은 일정 기간마다 반복적으로 나타나는 여론의 상승·하락 패턴을 포착할 수 있게 한다.

이러한 시계열적 분석은 컬처마이닝의 핵심 구성 요소로서, 문화적 담론이 형성·변화하는 과정을 데이터 기반으로 추적하고 비교할 수 있게 한다. 즉, 감성 분석과 시계열 분석을 결합한 컬처마이닝 접근은 문화 현상의 동태성을 포착하고, 각 사회가 공유하는 가치와 정서를 장기적·단기적으로 함께 이해할 수 있는 토대를 제공한다.

4. 맷음말

본 연구는 디지털 인문학(Digital Humanities)의 맥락에서 컬처마이닝(Culture Mining)의 개념과 방법론을 논의하고, 그 유용성을 학문적·실용적 차원에서 고찰하였다. 디지털 인문학은 전통적 인문학의 해석적 깊이를 유지하면서도, 데이터 기반의 정량적 분석을 통해 연구의 객관성과 확장성을 확보하려는 학문적 시도이다. 이러한 흐름 속에서 컬처마이닝은 방대한 언어 데이터 속에서 문화적 패턴을 발견하고 비교할 수 있는 유력한 방법론으로 주목된다.

첫째, 컬처마이닝은 문화 연구의 정량화를 가능하게 한다는 점에서 의의가 있다. 전통적인 담론분석은 특정 텍스트 집단을 대상으로 심층적 의미를 탐구하는 데 강점을 지니지만, 데이터 규모가 방대해지고 다언어 환경이 일상이 된 오늘날의 연구 조건을 충족시키기에는 한계가 크다. 반면 컬처마이닝은 텍스트 마이닝, 토픽모델링, 감성분석과 같은 기법을 도입하여 대규모 데이터를 분석함으로써 문화 현상을 수치화 시각 자료로 제시할 수 있다. 이는 해석의 주관성을 보완하고, 결과의 재현 가능성을 높이며, 나아가 문화연구의 방법론적 지평을 확장한다.

둘째, 컬처마이닝은 문화 비교 연구에 새로운 가능성을 제시한다. 다언어·다문화 데이터를 동시에 분석하고 비교함으로써, 문화 간의 차이와 공통점을 정량적으로 도출할 수 있다. 이는 단순히 언어적 차이를 넘어 사회적 가치관, 정서적 흐름, 담론의 구조적 특징까지 포착할 수 있게 하며, 국제적 맥락에서 문화 현상을 이해하는 데 강력한 도구가 된다. 결국 컬처마이닝은 문화 연구를 특정 지역이나 집단에 한정

하지 않고, 글로벌 차원의 비교연구로 확장하는 데 기여한다.

셋째, 컬처마이닝은 학제 간 융합 연구를 촉진한다. 디지털 인문학 자체가 학문 간 경계를 넘나드는 융합적 성격을 띠고 있지만, 컬처마이닝은 특히 언어학, 사회학, 심리학, 경영학, 정보과학, 인공지능 연구 등 다양한 분야와의 협력을 가능하게 한다. 예컨대, 언어학적 연구는 어휘와 의미망의 구조적 차이에 주목할 수 있고, 사회학적 연구는 집단 정체성과 사회적 행동의 패턴에 초점을 맞출 수 있으며, 데이터과학적 접근은 알고리즘과 모델링을 통해 이러한 해석을 보완한다. 이처럼 컬처마이닝은 학문 간 상호보완적 연구를 통해 새로운 지식을 창출하는 측면 역할을 한다.

넷째, 컬처마이닝은 실용적 활용 가능성이 넓다. 학문적 차원에서 디지털 인문학의 방법론을 정교화하는 동시에, 정책 결정, 공공 커뮤니케이션, 기업 전략, 국제문화 교류 등 다양한 영역에서 실질적 기여를 할 수 있다. 예를 들어, 정책 입안자는 사회적 담론의 흐름을 데이터 기반으로 이해하여 여론의 변화를 더 정확히 반영할 수 있으며, 문화 기관은 국제 교류 과정에서 발생하는 인식의 차이를 정량적 자료로 파악할 수 있다. 또한 학계와 산업체 모두에서 컬처마이닝은 문화적 데이터를 바탕으로 전략적 판단을 내리는 데 유익한 근거 자료를 제공한다.

컬처마이닝은 데이터 기반 문화연구의 가능성은 구체적으로 실현하는 연구 방법론으로서, 디지털 인문학의 학문적 정체성을 강화하고 문화 비교 연구의 지평을 확장하는 데 결정적인 역할을 한다. 본 연구는 이러한 컬처마이닝의 유용성을 강조함으로써, 문화연구가 빅데이터 시대에 어떻게 변화하고 확장될 수 있는지를 보여준 것이다.

◀ 참고문헌(Reference) ▶

- 김현(2013), 「디지털 인문학 – 인문학과 문화콘텐츠의 상생 구도에 관한 구상」, 『인문콘텐츠』 29, 인문콘텐츠학회, pp.9-26.
- 신상희·황복주(2015), 「소비자의 블로그 활용이 구매의사결정과정과 구매 후 행동에 미치는 영향에 관한 연구: 화장품 소비자를 중심으로」, 『경영교육연구』 30(4), 한국경영교육학회, pp.435-459
- 양윤·조문주(2000), 「구전 커뮤니케이션이 소비자의 태도변화에 미치는 영향」, 『광고학연구』 11(3), 한국광고학회, pp.7-34
- 윤선미(2023), 「월평균 소득에 따른 관광 활동 참여패턴 조사: 장바구니 분석을 적용하여」, 『호텔관광연구』 25(7), 한국호텔관광학회, pp.17-30
- 이근희(2015), 『번역의 이론과 실제』, 학국문화사, pp.250-272.
- 이준서. (2025), 「디지털인문학적 관점에서 본 컬처마이닝 연구—F 사 SPA 브랜드의 한·일 구전 커뮤

- 니케이션을 중심으로—」, 『일본학보』143, pp. 37-53
- 이준서(2025) 「AI 기반 일·중 온라인 담론에 나타난 혐한 정서 비교 분석」, 『일본어교육연구』 71, pp. 67-79
- 이준서(2021), 『R컬쳐마이닝』, 박문사, pp.11-12.
- 이준원(2017), 「SNS 상의 구전 커뮤니케이션 특성이 구전 커뮤니케이션 효과에 미치는 영향: 스포츠용 품을 중심으로」, 『한국체육과학회지』 26(3), 한국체육과학회, pp.711-724
- C Sievert, K Shirley(2014), "LDAvis: A method for visualizing and interpreting topics", *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, Association for Computational Linguistics, pp.63-70
- David M. Blei, Andrew Y. Ng, Michael I. Jordan(2003), "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, (3), JMLR.org, pp.993-1022
- Hyatt, J., & Simons, H. (1999). Cultural Codes – Who Holds the Key? : The Concept and Conduct of Evaluation in Central and Eastern Europe. *Evaluation*, 5(1), Sage Publications, pp.23–41
- Liu, X., Ren, P., Lv, X., & Li, S. (2024). Service Experience and Customers' eWOM Behavior on Social Media Platforms: The Role of Platform Symmetry. *International Journal of Hospitality Management*, 119, Elsevier Publications. DOI:<https://doi.org/10.1016/j.ijhm.2024.103735>.
- Mona Baker(2005), 『말 바꾸기』, 곽은주 역, 한국문화사.

이 글은 한국일본학회 일본학보 제 143집에 실린 논문을
수정·보안한 것임

성명 및 소속: (1)이준서

(2)LEE JUNSEO

(3)李俊瑞

성결대학교 글로벌물류학과 교수

E - mail주소: jslee@sungkyul.ac.kr

주소 : 경기도 안양시 만안구 성결대학로 53 성결대학교 중생관 531호

전화 : 010 9974 9211

원고료 입금 계좌 및 은행 및 예금주명 : 기업은행 01099749211 이준서