

텍스트마이닝을 이용한 국가유산채널 영상 콘텐츠 수용자 반응 분석* : 유튜브 댓글 중심으로

조경민 (한국학중앙연구원 인문정보학 박사과정)**

본 연구는 디지털 환경에서 문화유산 콘텐츠의 소비가 확대되는 흐름 속에서, 공공기관이 운영하는 국가유산채널의 유튜브 댓글을 분석하여 시청자의 반응 양상과 관심 주제, 그리고 문화유산 콘텐츠에 대한 수용자의 정서적·인지적 반응 특성을 실증적으로 탐색하고자 하였다. 기존 연구가 주로 콘텐츠 제작자 중심 혹은 전통적인 여론조사에 의존하였다면, 본 연구는 이용자 생성 콘텐츠(User-Generated Content)인 댓글 데이터를 중심으로 분석을 수행함으로써 수용자 관점의 반응 구조를 조망하고자 했다.

연구 대상은 2010년 6월부터 2024년 12월까지 국가유산채널이 유튜브에 게시한 2,115개 영상에 달린 댓글 73,537건이며, 이 중 전처리를 거쳐 67,826건을 최종 분석에 활용하였다. 분석 방법으로는 키워드 빈도 분석, TF-IDF 분석, BERTopic 기반 토픽모델링 기법을 적용하였으며, 재생목록별 댓글 수 분석도 병행하였다.

분석 결과, 시청자들은 단순한 정보 습득을 넘어, 전통문화의 아름다움에 대한 감탄, 정서적 지지, 문화적 자긍심 등을 표현하는 다양한 반응을 보였다. ‘문화유산’, ‘옹원’, ‘장인’, ‘아름답다’ 등의 키워드가 높은 빈도를 보였으며, 형용사형 키워드에서는 긍정 정서와 인상 평가 중심의 언어 사용이 두

* 이 논문은 2025년도 과학기술정보통신부 고성능 컴퓨팅 사업(GPU)의 지원을 받아 수행된 연구임 (G2025-0387)

** nsua@naver.com 단독저자, 국가유산진흥원 데이터팀 과장

드러졌다. TF-IDF 분석에서는 높은 점수를 획득한 ‘아름답다’, ‘감사하다’, ‘응원’, ‘국가’, ‘대단하다’ 등이 콘텐츠에 대한 정서적 몰입과 공감을 시사하였다. BERTopic 분석을 통해 총 10개의 주요 토픽이 도출하였으며, 이들은 전통성 중심의 감상 반응과 감성적 참여 및 K-콘텐츠 연계 반응이라는 두 축으로 구분되는 클러스터 구조를 형성하였다. 특히 전통 장인, 발굴현장, K-POP 협업, ASMR 콘텐츠 등 다양한 형식과 주제에 따라 시청자 반응 집단이 분화되는 양상이 확인되었다.

본 연구는 유튜브 기반 문화유산 콘텐츠에 대한 시청자의 수용 반응과 문화적 의미 형성 과정을 실증적으로 규명하였다는 점에서 이론적·실천적 의의가 있으며, 향후 문화유산 콘텐츠 기획과 정책 수립 시 데이터 기반 수용자 중심 전략 수립에 기초자료로 활용될 수 있다.

핵심어 국가유산채널, 유튜브 댓글, 텍스트 마이닝, 토픽모델링, 디지털 문화유산

1. 서론

유네스코(UNESCO)는 세계문화유산(World Heritage)을 “Our legacy from the past, what we live with today, and what we pass on to future generations”라 정의하고 있다(UNESCO World Heritage Convention, 1972). 이는 문화유산이 단순한 과거의 유물이 아니라, 현재를 살아가는 우리와 함께하며 미래 세대에 전승해야 할 소중한 자산임을 시사한다. 우리 정부와 문화 관련 기관들 또한 이러한 유네스코의 정의를 수용하고 있으며, 문화유산의 가치를 강조하는 다양한 정책과 실천을 이어오고 있다. 하지만 문화유산이 지닌 고유의 특성과 사회적 맥락의 변화로 인해, 이 개념은 지속적으로 확장되고 있으며, 유산에 대한 접근 방식 또한 새로운 국면을 맞이하고 있다.

1972년 유네스코의 세계유산협약 체결 이후, 문화유산의 범위는 유형 문화재를 넘어 무형유산, 디지털 유산까지 확대되었다. 이 과정에서 디지털 기술의 발전은 문화유산의 보존·활용·전승 방식에 혁신적 전환을 가져왔다. 디지털화된 문화유산은 시공간의 제약을 극복하고, 누구나 언제든지 접근 가능한 열린 환경에서 제공된다. 이는 학술·교육·문화 교류 등 다양한 분야에 영향을 미치고 있다. 특히 2003년 유네스코가 채택한 『디지털 유산의 보존에 관한 헌장』(UNESCO, 2003)은 “디지털로 생성된 정보뿐만 아니라 아날로그 자원의 디지털 전환 형태까지 포함하는 디지털 유산(Digital Heritage)” 개념을 공식화하며, 디지털 시대 문화유산의 새로운 방향성을 제시하였다.

이러한 국제적 흐름 속에서 한국 역시 문화유산의 디지털 전환에 발맞추고 있다. 국가유산청(舊 문화재청)은 2021년 ‘문화재 디지털 대전환 2030’을 발표하며 디지털 생태계 강화를 주요 전략 과제로 제시하였고, 2024년에는 기관명을 ‘국가유산청’으로 개편하며 디지털 서비스 플랫폼을 통해 문화유산의 온라인 기반 확산을 추진하고 있다(국가유산청, 2021; 국가유산디지털서비스, 2025). 특히 ‘디지털 취약계층을 위한 콘텐츠 제작’ 등 포용적 디지털 접근을 포함한 전략은 문화유산 향유 대상을 넓히는 계기가 되었다.

문화유산 콘텐츠의 수용자 경험과 반응을 실증적으로 분석하는 연구는 디지털 시대의 문화유산 정책과 콘텐츠 전략 수립에 있어 필수적인 요소로 부각되고 있다. 이러한 분석을 통해 수용자의 요구와 관심사를 정확히 파악함으로써, 보다 효과적이고 참여를 유도하는 콘텐츠 기획이 가능해진다.

이종욱(2019)은 기존의 디지털 문화유산 정책이 주로 콘텐츠 제작과 기술 개발에 초점을 맞추었으며, 수용자 중심의 접근이 부족하다고 지적했다. 그는 산업계 종사자와 일반 이용자들이 디지털 문화유산 데이터를 활용할 수 있도록 데이터 표준화와 정보 공유 플랫폼의 구축이 필요하다

고 강조했다. 또한, 이러한 데이터를 기반으로 한 디지털 문화유산 콘텐츠 제작을 위한 가이드라인을 수립하고, 이를 문화유산 기관과 일반 이용자들에게 보급해야 한다고 제안했다.

문화유산 콘텐츠의 효과적인 확산과 향유를 위해서는 수용자 중심의 반응 분석을 기반으로, 디지털 콘텐츠가 실제로 어떠한 인식의 변화와 정서적 반응을 이끌어내는지를 종합적으로 이해하는 과정이 선행되어야 한다. 이는 문화유산 콘텐츠가 단순한 정보 전달을 넘어, 감성적 공감, 문화적 정체성, 학습 동기, 참여 의지 등을 어떻게 유발하는지를 실증적으로 파악할 수 있게 해준다. 이러한 분석은 콘텐츠 기획과 운영을 수요자 중심으로 전환하는 데 근거가 된다.

특히 공공 주도로 제작되는 문화유산 콘텐츠는 정책적 목적과 공공성 확보를 동시에 추구해야 하며, 이를 위해 실제 이용자의 반응 데이터를 기반으로 무엇이 효과적인가, 어떤 유형이 공감을 이끄는가, 어떤 콘텐츠가 재참여를 유도하는가 등의 실증적 증거가 요구된다.

따라서 향후 디지털 문화유산 정책과 콘텐츠 전략 수립에는 수용자 경험과 반응에 대한 실증적 분석이 반드시 필요하다. 이를 통해 문화유산 콘텐츠의 지속 가능성과 효과적인 확산 전략을 마련할 수 있을 것이다.

본 연구는 2010년부터 국가유산청과 국가유산진흥원(舊 한국문화재단)이 공동으로 운영해 온 ‘국가유산채널’의 유튜브 영상 콘텐츠에 달린 댓글을 분석함으로써, 문화유산 향유자의 반응을 실증적으로 탐색하고자 한다. 국가유산채널은 2010년 문화재청의 지원으로 (사)헤리티지채널이 ‘헤리티지채널’이라는 이름으로 시작하였으며, 2014년부터는 운영 주체가 국가유산진흥원으로 변경되어 ‘문화유산채널’로 개편되었다. 이후 2024년 5월, 「국가유산 기본법」 시행과 함께 현재의 명칭인 ‘국가유산채널’로 변경되었으며, 유튜브를 중심으로 다양한 디지털 콘텐츠를 제공하고 있다. 특히 2023년 7월에는 유튜브 구독자 100만 명을 돌파하

고, 2024년 12월 기준 173만 명의 구독자를 확보한 국내 대표 공공 문화유산 미디어 채널로 자리매김하였다(국가유산채널, 2024).

본 연구는 다음과 같은 문제의식에서 출발한다. 지금까지의 국가유산 콘텐츠는 공공 주도의 기획·제작 중심이었으며, 콘텐츠 수용자의 요구와 반응을 체계적으로 반영하지 못하였다. 이에 따라 본 연구는 국가유산채널 유튜브 영상에 달린 댓글을 중심으로 시청자의 반응을 수집·분석하여, 콘텐츠에 대한 수용자 인식과 감성적 경향을 파악하고자 한다.

이를 위해 유튜브에 공개된 국가유산채널의 전체 영상 중 주요 콘텐츠에 달린 댓글 데이터를 수집하고, 텍스트 마이닝 기반 분석 기법인 TF-IDF와 BERTopic을 활용하여 핵심 키워드와 토픽 구조를 도출하였다. 이와 함께 콘텐츠별 반응의 감성적 차이와 주제별 특징을 비교함으로써, 국가유산채널 콘텐츠의 수용 특성과 디지털 문화유산 향유 양상을 종합적으로 조망하고자 한다. 본 연구는 이를 통해 국가유산 콘텐츠 기획·운영에 실질적 시사점을 제공함과 동시에, 공공 문화유산 콘텐츠 분석의 방법론적 확장을 모색하고자 한다.

2. 이론적 논의

1) 선행연구 고찰

유튜브는 전 세계적으로 가장 영향력 있는 동영상 공유 플랫폼으로, 한국에서도 인터넷 사용자 대부분이 이용하는 핵심 미디어 채널로 자리매김하고 있다. 특히 2024년 1월 기준 국내 유튜브 앱 총 사용 시간은 1,119억 분에 달하며, 이는 2019년 1월 519억 분 대비 약 116% 증가한 수치로, 1인당 평균 사용 시간은 월 40시간에 이를 정도로 높은 수준이다(와이즈앱, 2024). 이러한 플랫폼 영향력의 확대는 유튜브 댓글 공간이

단순한 반응 기록을 넘어 사회적 담론과 감정의 표현, 의견 교환이 이루어지는 공론장으로 진화하고 있음을 보여준다.

유튜브 댓글은 기본적으로 사용자 닉네임 기반의 익명성이 보장되며, 이는 개인이 자신의 정체성을 노출하지 않고 의견을 자유롭게 표현할 수 있는 환경을 제공한다. 이러한 익명성은 사용자의 자율적 표현을 증진시키는 동시에, 감정적이고 직설적인 언어 사용을 유도하는 경향도 내포하고 있다. 최근에는 이러한 댓글의 특성과 감성, 주제 구조에 주목하여 유튜브 댓글을 중심으로 한 텍스트 마이닝 연구가 활발하게 이루어지고 있으며, 다양한 분야에서 실증적 자료로 활용되고 있다.

김현미 외(2024)는 인터랙티브 스토리텔링 콘텐츠의 상호작용성이 이용자의 지각된 실재감과 몰입을 매개로 하여 지속이용의도에 미치는 영향을 분석하였다. 해당 연구는 상호통제성, 개인화, 참여성이라는 세 부 요소를 중심으로 상호작용성 요인이 실재감과 몰입을 통해 콘텐츠 반복 이용 의도에 영향을 미친다는 이중매개모형을 제시하였다. 이는 디지털 콘텐츠의 감각적 몰입 구조가 이용자 반응에 있어 중요한 변수임을 보여준다.

최제호 외(2024)는 실존 아이돌(BTS)과 가상 아이돌(이세계아이돌)의 유튜브 콘텐츠에 달린 댓글을 분석하여, 감정 표현 방식의 차이를 형용사 빈도 및 토픽모델링 분석을 통해 규명하였다. 분석 결과, 실존 아이돌 콘텐츠는 인성과 감정 중심의 반응이, 가상 아이돌 콘텐츠는 시각적 요소와 유머 코드에 기반한 반응이 주를 이루었으며, 콘텐츠의 성격에 따라 이용자의 정서적 반응 방식이 달라짐을 밝혔다.

노영아(2024)는 피아니스트 임윤찬의 연주 영상 댓글을 대상으로 키워드 빈도, LDA 기반 토픽모델링, 감성분석을 수행하여 디지털 환경 속에서 클래식 음악이 소비되고 감상되는 방식에 대한 새로운 인식을 제시하였다. 해당 연구는 음악 콘텐츠에 대한 정서적 반응 구조를 파악하는데 있어 유튜브 댓글의 분석 가능성을 입증하였다.

강진구 외(2024)는 대구 이슬람 사원 건축 갈등과 관련된 유튜브 댓글을 분석하여, 선주민과 이주민 간 갈등에 대한 사회적 인식 구조를 도출하였다. 이 연구는 주요 어휘의 출현 빈도, 토픽 분석, 감정 분석, 의미망 분석 등을 통해 사회적 이슈에 대한 여론 지형을 실증적으로 파악하였다.

이신행 외(2023)는 국내에서 활동하는 ‘사이버렉카’ 유튜브 채널을 분석 대상으로 하여, 악성댓글 생성 메커니즘을 비교·분석하였다. 익명성과 실명 기반 채널 간 댓글 분포 차이를 중심으로, 특정 채널의 운영 방식이 악플 생성에 미치는 영향을 실증적으로 규명하였다.

정낙원 외(2020)의 연구는 포털 뉴스 플랫폼인 네이버와 다음에 게시된 문재인 정부 관련 여론조사 뉴스에 대한 댓글 여론을 분석하였다. 분석 결과, 네이버는 보수·반정부 성향 댓글이, 다음은 진보·지정부 성향 댓글이 우세하게 나타났으며, 댓글의 ‘공감’ 수조차 실제 여론조사 결과와 괴리를 보이는 양상이 확인되었다. 전체 댓글의 71.2%가 혐오·비하·조롱 등의 적대적 언어를 포함하고 있었으며, 이러한 적대적 표현의 강도와 빈도는 플랫폼과 정치적 입장에 따라 유의미하게 달라졌다. 이 연구는 유튜브 및 포털 댓글이 실제 여론을 구성하거나 왜곡하는 과정에 대한 실증적 근거를 제공하였다는 점에서 의의가 있다.

이와 같이 기존 연구들은 유튜브 댓글이 감성적 반응, 사회적 담론, 여론 형성 등에서 중요한 역할을 수행하고 있음을 보여준다. 그러나 대다수 연구가 K-POP, 정치 이슈, 갈등 현안 등 특정 분야에 집중되어 있으며, 문화유산 콘텐츠처럼 공공성과 교육성을 갖춘 콘텐츠에 대한 수용자 반응을 다룬 연구는 상대적으로 부족하다. 특히 전통문화 기반 콘텐츠에 대한 디지털 수용자 반응을 토픽모델링 기반으로 실증 분석한 연구는 거의 전무하다.

이러한 분석이 필요한 이유는, 공공 목적의 문화유산 콘텐츠가 단순히 제작되고 배포되는 수준에 머무르는 것이 아니라, 실제 수용자들이

어떻게 인식하고 반응하는지를 파악함으로써, 향후 콘텐츠의 기획 방향을 보다 효과적이고 수요자 지향적으로 조정할 수 있기 때문이다. 수용자 반응에 기반한 분석은 콘텐츠의 감성적 공감 유도, 반복 시청 가능성, 사회적 공유 촉진 등 문화유산의 확산 전략 수립에 있어 실질적인 데이터를 제공하며, 공공 정책의 정당성과 실행 가능성을 높이는 데도 기여할 수 있다.

본 연구는 이러한 공백을 메우고자 하는 시도이며, 공공 문화 콘텐츠에 대한 정성적·정량적 반응 분석의 확장 가능성을 모색한다는 점에서 의미가 있다.

2) 토픽 모델링 이론 및 기법 고찰

유튜브 댓글은 대부분 1~2문장 내외의 짧은 단문으로 구성되어 있으며, 이러한 짧은 텍스트(Short Text)는 기존의 토픽 모델링 알고리즘이 전제로 삼는 충분한 단어 수 및 문맥 정보가 부족하다는 한계를 지닌다. 이에 따라 짧은 텍스트 데이터를 분석하기 위해 특화된 Short Text Topic Modeling(STTM) 기법들이 개발되어 왔다.

(1) Biterm Topic Model(BTM)

BTM(Biterm Topic Model)은 짧은 문서의 특성을 고려하여, 개별 문서가 아닌 전체 코퍼스 수준에서 단어쌍(biterm)의 공출현 패턴을 기반으로 토픽을 추출하는 기법이다(Yan et al., 2023). 전통적인 LDA(Latent Dirichlet Allocation)가 각 문서 내의 단어 분포에 초점을 맞춘다면, BTM은 전체 댓글 집합에서 발생하는 단어쌍의 빈도와 연결 구조를 분석함으로써 단어 희소성(sparsity) 문제를 완화하고, 짧은 문서 간 공통된 주제 구조를 효과적으로 도출할 수 있다. 이로 인해 BTM은 트위터, 유튜브 댓글과 같은 SNS 기반 데이터 분석에 특히 적합한 모델로 널리 활용되고 있다.

(2) Gibbs Sampling Dirichlet Mixture Model(GSDMM)

GSDMM(Gibbs Sampling Dirichlet Mixture Model)은 각 문서가 단일 주제를 다룬다는 전제 하에 디리클레 혼합 모델(Dirichlet Mixture Model)을 기반으로 Gibbs 샘플링을 적용하여 토픽을 추정하는 기법이다(Yin & Wang, 2014). 이 모델은 문서 길이가 짧고 주제 전이가 거의 없는 데이터에 적합하며, 하나의 문서가 단일 주제에 소속된다는 점에서 유튜브 댓글과 같은 데이터 구조에 유리하다. 특히 높은 정확도의 군집화를 통해 짧은 텍스트의 주제 식별력을 높일 수 있다는 점에서 활용 가치가 높다.

(3) Top2Vec

Top2Vec은 문서 임베딩 기반의 비지도 토픽 모델링 기법으로, 사전 정의된 토픽 수 없이 문서 임베딩을 생성한 뒤 DBSCAN(Density-Based Spatial Clustering of Applications with Noise) 알고리즘을 적용하여 밀도 기반으로 주제를 도출한다(Angelov, 2020). 이 방법은 단어와 문서의 의미적 임베딩을 활용함으로써 문맥을 반영한 정교한 주제 파악이 가능하며, 전통적인 확률 기반 접근 방식에 비해 더 유연하고 직관적인 토픽 해석을 제공한다.

(4) BERTopic

BERTopic은 사전학습된 Transformer 기반 언어 모델(BERT 등)을 활용하여 문서 임베딩을 생성하고, HDBSCAN(Hierarchical Density-Based Spatial Clustering of Applications with Noise) 알고리즘을 이용해 문서 간 유사도를 바탕으로 클러스터링을 수행한다. 이후 각 클러스터에 대해 클래스 기반 TF-IDF(c-TF-IDF)를 적용하여 대표 키워드를 도출한다(Grootendorst, 2022). 이 기법은 고차원 의미 공간에서 문맥을 반영하는 임베딩 기반 분석을 통해 짧은 텍스트에서도 높

은 정확도의 주제 식별이 가능하다. 또한 단어 의미 간 유사성을 활용하여 노이즈를 줄이고, 주제 간 중복을 최소화할 수 있다는 점에서 최근 주목받고 있다.

〈표 1〉 STTM 방법론 비교 요약표

방법론	접근방식	강점	약점
BTM	Biterm-based word pair analysis (across corpus)	Suitable for short texts; handles sparsity well	Ignores word order and context; less effective for longer texts
GSDMM	Clustering-based with Dirichlet prior	Fast convergence; effective for short, single-topic documents	Assumes each document has only one topic; may require careful tuning
Top2Vec	Document embedding + density-based clustering (DBSCAN)	No need to predefine number of topics; strong semantic grouping	Computationally intensive; slow on large datasets
BERTopic	BERT-based semantic embedding + HDBSCAN clustering	High-quality topic modeling with contextual understanding; effective for short texts	Requires significant computational resources; longer training time

3) 본 연구의 분석틀 설정

기존 선행연구들은 텍스트 마이닝 기법을 활용하여 유튜브 댓글의 주제 및 정서 구조를 분석하는 데 초점을 맞춰 왔으며, 주요 분석 기법으로는 키워드 빈도 분석, 토픽모델링(LDA, BTM, BERTopic 등), 감성분석이 주로 활용되었다. 특히 최근에는 유튜브 댓글 데이터를 기반으로 한 공공문화, 음악, 정치 이슈 등 다양한 콘텐츠에 대한 수용자 반응을 정량적으로 해석하려는 시도가 증가하고 있다.

본 연구는 기존 연구의 방법론적 흐름을 계승하면서도, 국가유산채

널이라는 공공 문화유산 콘텐츠를 대상으로 분석을 수행한다는 점에서 차별성을 지닌다. 분석 방법으로는 TF-IDF를 활용한 핵심 키워드 도출, BERTopic을 활용한 주제 클러스터 추출, 재생목록 단위 콘텐츠 간의 댓글 수 비교 분석을 병행하였다. 이를 통해 단순한 주제 분류를 넘어, 댓글을 통한 이용자 반응의 구조적 특성과 문화유산 콘텐츠에 대한 수용자의 인식 양상을 함께 파악하고자 한다.

4) 문화유산에 대한 접근 방식의 변화

한국지능정보화사회진흥원(2014)은 『국가정보화 20년의 성과』를 통해 문화유산 정보화 사업의 성과를 다음과 같이 제시하였다. 문화유산 정보화는 2001년 「국가지식정보자원법」에 근거하여 추진되었으며, 2006년까지 약 67만여 건의 국가문화유산 지식정보 데이터베이스(DB)를 구축하였다. 이를 기반으로 국립·공립·사립·대학 박물관의 유물 정보를 연계한 통합 플랫폼인 ‘이뮤지엄’(www.emuseum.go.kr)이 구축되었으며, 2013년 기준 608개 박물관의 약 14만 9천여 점의 유물 정보가 국민에게 제공되었다.

그러나 이러한 문화유산 데이터는 대부분 관리기관 웹사이트를 통해 공급자 중심의 단선적인 방식으로 제공되어 왔다. 방대한 양의 정보가 단편적으로 나열되었기 때문에, 일반 대중이 접근하고 활용하기에는 한계가 있었다. 이로 인해 문화유산은 여전히 ‘박제된 정보’로 인식되며, 이용자 중심의 문화 향유로 이어지기 어려운 구조였다. 이에 대해 이수재(2016)는 문화유산채널이 “다양한 콘텐츠 접근 방식을 통해 일반 대중과 문화유산 간의 거리감을 좁혔으며, 문화유산의 대중화에 기여했다”고 평가하였다. 이는 문화유산을 단순한 기록이 아닌 멀티미디어 콘텐츠로 재구성하고, 서사를 부여함으로써 대중과의 감성적 소통을 가능하게 한 사례로 볼 수 있다.

특히 영상 기반 콘텐츠는 문화유산을 일방적인 지식 전달의 대상이

아닌 스토리텔링의 구성 요소로 전환시키며, 이용자가 단순한 수용자가 아닌 주체적 선택자이자 참여자로 변화하는 과정을 가능케 한다. 이는 디지털 전환 시대에 문화유산을 향유하는 방식 자체가 근본적으로 변화하고 있음을 시사한다.

이러한 맥락에서 대한민국 국회는 2023년에 「국가유산기본법」을 제정하고, 「문화재보호법」을 포함한 총 11개의 관련 법률을 개정하여 ‘문화재’라는 용어를 ‘국가유산’으로 전면 개편하였다. 전통적으로 사물 중심이었던 문화재 분류 체계는 ‘국가유산’이라는 상위 개념 아래 ‘문화유산’, ‘자연유산’, ‘무형유산’의 세 범주로 확장되었으며, 이는 재화로서의 ‘재(財)’가 아닌, 전승 가능한 ‘유산(遺産)’이라는 가치 전환을 반영하는 조치로 평가된다(법제조정총괄법제관실, 2024).

2024년 『법제소식』 5월호에 실린 ‘국가유산 법제’에 따르면, 국가유산은 “우리 삶의 뿌리이자 창의성의 원천이며, 인류 모두의 소중한 자산임을 인식하고, 그 가치를 온전하게 보존·향유하며 창조적으로 계승·발전시켜 나가는 것”을 기본 이념으로 삼는다. 국가유산의 개념은 역사적, 예술적, 학술적, 경관적 가치가 뛰어난 인위적·자연적 형성물을 포괄하며, 구체적으로 다음과 같이 세 가지 범주로 구분된다.

- 문화유산: 우리 역사와 전통의 산물로서 문화의 고유성과 정체성, 생활문화의 변화 등을 반영하는 유형의 유산
- 자연유산: 동식물, 지형, 지질 등의 자연물 또는 환경과의 상호작용으로 형성된 유산
- 무형유산: 공동체가 세대를 거쳐 전승하며 지속적으로 재창조하는 무형의 문화적 산물

이처럼 문화유산을 보다 포괄적인 ‘국가유산’ 개념으로 확장하는 제도적 변화는, 콘텐츠 기획의 대상과 범위를 넓히는 근거가 되며, 국가유산채널 역시 다양한 주제와 형식을 실험할 수 있는 기반을 갖추게 되었

다. 그러나 콘텐츠의 외연 확대는 동시에 다양한 수용자 집단의 반응을 포착하고 분석하는 데 있어 난이도를 증가시키는 요인이 되기도 한다. 본 연구는 이러한 문화유산 접근 방식의 제도적·기술적 변화와 맞물려, 실제 콘텐츠 수용자의 반응을 분석함으로써 향후 국가유산 콘텐츠의 기획 방향성과 커뮤니케이션 전략 수립에 기초 자료를 제공하고자 한다.

〈표 2〉 국가유산 법체계 분류 개편사항

종전		
문화재	유형문화재	
	민속문화재	
	기념물	사적지
		명승, 천연기념물 등
	무형문화재	무형문화재법
개편		
국가유산 「국가유산 기본법」	문화유산	문화유산법
	자연유산	자연유산법
	무형유산	무형유산법

5) 국가유산 콘텐츠의 디지털 경험 특성

최근 연구에 따르면, 유튜브 관광 콘텐츠에서 제공되는 사회적 실재감은 이용자의 관심과 몰입에 유의미한 영향을 미치는 것으로 나타났다(정젤 나 외, 2024). 즉, 단순한 정보 전달을 넘어서, 현장감 있는 사회적 제작물이 시청자에게 정서적 반응을 유도하고, 이는 콘텐츠 몰입과 지속 이용의도를 강화하는 요인으로 작용한다. 이러한 효과는 국가유산채널에서도 유사하게 나타나는데, 실제 국가유산을 기반으로 제작된 시청각 자료가 이용자에게 간접적인 체험과 감각적 경험을 제공하고, 공간적·시간적 제약을 뛰어넘는 콘텐츠 향유를 가능하게 한다.

국가유산채널의 유튜브 콘텐츠는 실재하는 장소나 유산을 녹화된 형태로 반복적으로 소비할 수 있는 환경을 제공하며, 이로 인해 이용자

는 언제 어디서든 문화유산을 경험할 수 있게 된다. 그러나 디지털 콘텐츠는 근본적으로 직접 체험에서 오는 감각적 실재감을 완전히 대체할 수 없다는 한계도 내포하고 있다. 이러한 한계는 시청자가 댓글을 통해 자신의 감상, 불만, 요구를 직접 표현하는 동기로 전환되고, 결과적으로 플랫폼 내에서 이용자와 콘텐츠 간의 상호작용적 커뮤니케이션이 강화된다.

디지털 콘텐츠 소비 맥락에서, 사용자는 콘텐츠와의 사회적 실재감(social presence)을 경험하고, 이는 몰입을 유도하고 이용자의 반응을 촉발하는 중요한 요인으로 작용한다. 영상 감상 후 댓글을 통해 나타나는 피드백은 단순한 감정 표현을 넘어, 향후 콘텐츠 기획 방향에 영향을 줄 수 있는 사용자 요구의 집합체로 기능한다. 이 과정은 시청자 → 피드백 → 제작자 반영 → 신규 콘텐츠 → 시청자 반응으로 이어지는 순환적 구조를 형성하며, 지속적인 시청 의도 및 플랫폼 참여를 유도하는 메커니즘을 만들어낸다.

또한 유튜브 댓글 기능은 단순한 의견 개진의 도구를 넘어, 시청자 간 정보 교류와 의미 확장의 장으로 기능한다. 댓글을 통한 상호작용은 다양한 관점을 교환하고, 새로운 인식을 형성하며, 콘텐츠에 대한 이차적 의미화 작업을 유도하게 된다. 이러한 참여적 특성은 디지털 미디어 환경에서의 사용자 경험을 풍부하게 하고, 집단 지식 생산과 문화 담론 형성에도 기여하는 중요한 특성으로 평가된다.

결과적으로 국가유산채널의 유튜브 콘텐츠는 문화유산을 디지털 미디어 기반으로 간접 경험하도록 하면서도, 댓글을 통한 상호작용을 통해 실재감의 한계를 일정 부분 보완하고 있다. 이는 문화유산 향유 방식의 변화와 함께, 시청자의 의견이 실시간으로 콘텐츠 기획·운영에 반영될 수 있는 가능성을 열어두며, 시청자 중심의 문화유산 커뮤니케이션 모델 정립에 실마리를 제공한다.

3. 연구문제

본 연구는 국가유산청과 국가유산진흥원이 운영하는 국가유산채널의 유튜브 콘텐츠에 달린 댓글 데이터를 토픽모델링 기법을 활용하여 분석함으로써, 시청자들이 문화유산 콘텐츠에 대해 어떠한 반응을 보이는지, 그 반응에 내포된 어휘적 특성과 주제 구조를 실증적으로 규명하고자 한다.

특히 유튜브 댓글에 나타난 언어 표현과 감성적 반응을 심층적으로 분석하여, 국가유산채널이 제공하는 콘텐츠가 시청자에게 어떤 가치·정서·의미를 전달하고 있는지, 또한 어떤 콘텐츠 영역이 높은 관심과 호응을 이끌어내는지에 대한 분석을 통해, 향후 콘텐츠 기획 및 전략 수립에 기초 자료를 제공하는 것을 목표로 한다.

이를 위하여 다음과 같은 세 가지 연구문제를 설정하였다.

- 연구문제 1: 국가유산채널 유튜브 댓글의 고빈도 키워드는 무엇인가?
- 연구문제 2: BERTopic 기반 토픽모델링을 통해 도출된 토픽 분석 결과와 그 특징은 무엇인가?
- 연구문제 3: 국가유산채널에서 정의한 재생목록별 시청자 반응은 어떠한가?

본 연구는 이와 같은 연구문제를 바탕으로 대용량 댓글 데이터를 활용한 언어 기반 실증 분석을 수행함으로써, 국가유산 콘텐츠와 이용자 반응 간의 상호작용 구조를 체계적으로 이해하고, 디지털 미디어 환경에서 문화유산 콘텐츠의 수용 양상과 발전 방향을 제시하고자 한다.

4. 연구방법

1) 방법론

(1) 분석 개요

본 연구는 국가유산채널 유튜브 영상에 달린 댓글 데이터를 대상으로, 텍스트 마이닝(Text Mining) 기법을 활용하여 시청자의 반응을 체계적으로 분석하고자 하였다. 특히 비정형 텍스트 자료에서 주요 주제어와 의미 구조를 추출할 수 있는 키워드 분석 및 토픽모델링(Topic Modeling) 기법을 중심으로 분석을 수행하였다.

기존의 텍스트 분석은 연구자가 수작업으로 문서를 읽고 분류하는 방식에 의존하였기 때문에, 데이터의 양과 다양성이 증가할수록 시간적·물리적 한계가 명확하게 드러났다. 이에 따라 대규모 비정형 데이터의 분석에서는 자동화된 통계 기반 방법이 요구되었고, 그 대안으로 등장한 토픽모델링은 문서 집합 내 단어 사용 패턴을 기반으로 잠재적 주제(Topic)를 추출함으로써, 수천 개의 문서를 일일이 읽지 않고도 텍스트 내 의미 구조를 파악할 수 있게 한다. 나아가 토픽 간의 상호작용 구조, 시계열적 주제 변동, 관심도 차이 등을 분석할 수 있는 장점이 있어, 소셜미디어나 뉴스 댓글 등 동적 콘텐츠 분석에 효과적으로 활용되고 있다.

(2) 분석 대상 및 데이터 수집

본 연구는 2010년부터 2024년까지 유튜브에 게시된 국가유산채널 영상 2,605편 중 댓글이 존재하는 2,115편을 분석 대상으로 선정하였다. 해당 영상으로부터 수집된 총 댓글 수는 73,537개이며, 이는 유튜브 API 기반 수집 도구와 Python 스크립트를 통해 자동 수집되었다.

(3) 데이터 전처리

수집된 댓글 데이터는 분석에 앞서 다음의 절차에 따라 전처리되었다.

① 형태소 분석: 한국어 자연어처리 라이브러리를 활용하여 각 댓글에 대한 품사 기반 형태소 분석을 수행하였다. 이때 불용어 제거(stop-words filtering) 및 어미·조사·기호 제거 등 정제 작업을 병행하였다.

② 키워드 정제 및 필터링: 댓글의 핵심적인 의미를 담고 있는 명사 및 형용사 중심의 단어만을 추출하였으며, 추출된 단어 중 분석의 의미를 지니지 않는 중복·의미 불명 단어는 수동 제거하였다.

(4) 분석 절차

본 연구의 전체 분석 과정은 <그림 1>과 같다. 분석은 다음의 절차에 따라 수행되었다.

① 키워드 빈도 분석

시청자 반응의 전반적 경향성을 파악하기 위해, 전처리된 댓글 코퍼스에서 단순 출현 빈도 상위 30개 키워드를 추출하였다. 이를 통해 주로 사용된 핵심 단어의 경향을 시각화하고, 콘텐츠에 대한 정서적 반응을 조망하였다.

② TF-IDF 기반 중요도 분석

빈도 중심의 단어 분석을 보완하기 위하여, TF-IDF(Term Frequency-Inverse Document Frequency) 기법을 활용해 상대적으로 중요도가 높은 단어를 도출하였다. 이를 통해 단순 반복 단어가 아닌, 댓글 내 의미적 비중이 높은 키워드를 추가로 확인하였다.

③ BERTopic 기반 토픽모델링

수집된 댓글을 BERTopic 알고리즘에 입력하여, 댓글 집합 내의 잠재 주제군(토픽)을 추출하였다. 해당 모델은 문장 단위 문서 임베딩 → 밀도 기반 클러스터링(HDBSCAN) → 클래스 기반 TF-IDF(c-TF-IDF) 순으로 주제를 도출하며, 단어의 의미적 유사성을 기반으로 짧은 문장에서도 유

의미한 주제 분류가 가능한 특징을 지닌다. 도출된 토픽은 주요 키워드와 함께 시각화되었으며, 토픽 간 유사성 네트워크도 함께 분석하였다.

④ 콘텐츠 유형별 반응 비교 분석

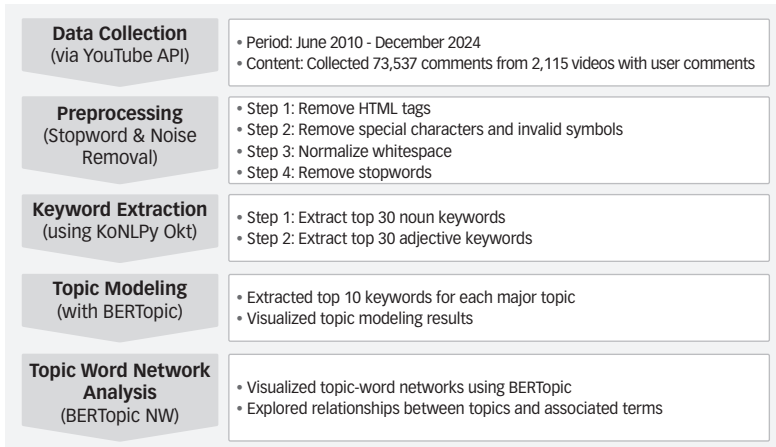
국가유산채널이 유튜브에 제공 중인 재생목록(예: 전통공예, 전통무용, 기획 영상 등)을 기준으로, 각 목록에 포함된 영상들의 평균 댓글 수를 산출하였다. 이를 통해 재생목록별 콘텐츠에 대한 시청자의 반응 정도와 관심 수준의 상대적 차이를 비교 분석하였다.

(5) 분석 도구 및 환경

- 언어 및 버전: Python 3.11.9
- 라이브러리: *pandas*¹⁾, *numpy*²⁾, *matplotlib*³⁾, *nltk*⁴⁾, *konlpy*⁵⁾, *scikit-learn*⁶⁾, *BERTopic*⁷⁾
- 토픽모델링 모델: BERTopic
 - 임베딩 모델: *SentenceTransformer* 기반 (*klue/bert-base*⁸⁾)

-
- 1) The pandas development team. (2020). *pandas-dev/pandas: Pandas (v2.2.2)* [Computer software]. Retrieved from <https://doi.org/10.5281/zenodo.3509134>
 - 2) Harris, C. R., et al. (2020). Array programming with NumPy [Electronic version]. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
 - 3) Hunter, J. D. (2007). Matplotlib: A 2D graphics environment [Electronic version]. *Computing in Science & Engineering*, 9(3), 90–95.
 - 4) Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media.
 - 5) Park, J. (2014). *KoNLPy: Korean NLP in Python (v0.6.0)* [Computer Software]. Retrieved from <https://konlpy.org>
 - 6) Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
 - 7) Grootendorst, M. (2022). BERTopic: Neural topic modeling with class-based TF-IDF [On-Line]. Retrieved from <https://arxiv.org/abs/2203.05794>
 - 8) KLUE Team. (2021). *klue/bert-base (Version 1.0)* [On-Line Software]. Hugging Face. Retrieved from <https://huggingface.co/klue/bert-base>

- 클러스터링 알고리즘: *HDBSCAN*
- 벡터화 방식: *class-based TF-IDF (c-TF-IDF)*
- 형태소 분석기: *KoNLPy*의 *Okt (Open Korean Text)* Tokenizer 사용



〈그림 1〉 연구 절차

2) 연구대상

본 연구는 2010년 6월부터 2024년 12월까지 국가유산청과 국가유산진흥원이 공동 운영하는 국가유산채널 유튜브 계정에 게시된 영상 중, 댓글이 달린 2,115개 콘텐츠를 분석 대상으로 선정하였다. 해당 영상에서 수집된 총 댓글 수는 73,537건이다.

본 연구는 시계열 변화에 따른 시간적 패턴보다는, 14년간 축적된 전체 댓글 데이터를 통시적으로 분석함으로써 국가유산 콘텐츠 전반에 대한 시청자 반응의 전반적 특성과 주제 구조를 규명하고자 하였다. 이를 통해 콘텐츠 유형과 주제별 반응 양상의 전반적인 경향을 파악하고자 하였다.

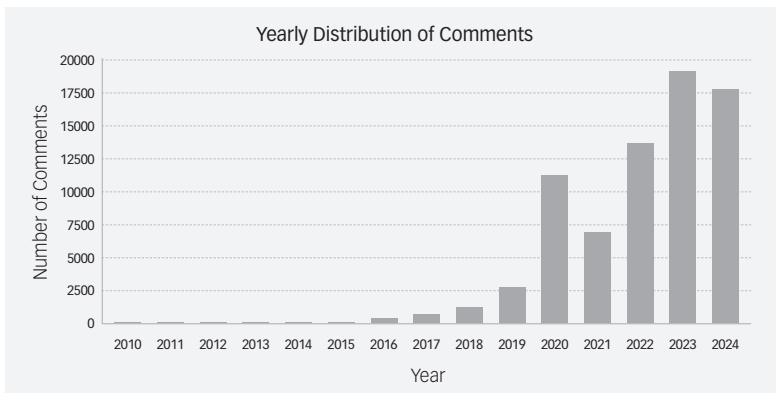
데이터 수집은 Python 프로그래밍 언어를 기반으로 YouTube API v3와 연동한 자동 스크랩 스크립트를 활용하여 수행되었다. 유튜브 내 국가유산채널의 영상 목록을 기준으로 각 콘텐츠에 달린 댓글을 모두 수

집하였으며, 이후 중복 제거, 불용어 제거, 형태소 분석 등 전처리 절차를 거쳐 분석에 적합한 형태로 정제되었다.

〈그림 2〉는 연도별 댓글 수 변화를 시각화한 그래프로, 2020년 이후 댓글 수가 급증한 현상을 보여준다. 이러한 증가는 COVID-19 팬데믹에 따른 사회적 거리두기와 오프라인 활동의 제약으로 인해 온라인 미디어 이용이 급증한 환경 변화와 밀접한 관련이 있는 것으로 해석된다. 특히 COVID-19 제한 조치 해제 이후에도 구독자 수와 댓글 수가 지속적으로 증가한 것은, 국가유산채널 콘텐츠가 지닌 콘텐츠 품질과 대중적 관심의 선순환 구조를 반영하는 결과로 볼 수 있다.

또한 본 연구의 분석 재현성과 투명성 제고를 위해, 수집된 데이터와 소스코드는 온라인 저장소를 통해 자유롭게 열람 및 다운로드할 수 있도록 공개하였다. 연구자가 직접 작성한 수집 및 분석 코드와 전처리된 데이터셋은 다음의 GitHub 저장소에서 확인 가능하다(<https://github.com/tsarkr/k-heritage>)

이를 통해 본 연구는 후속 연구자들이 관련 자료를 바탕으로 다양한 문화유산 콘텐츠 분석을 수행할 수 있도록 기반을 제공하고, 디지털 문화유산 연구의 확장성을 제고하는 데 기여하고자 한다.



〈그림 2〉 연도별 댓글 분포

3) 전처리

본 연구는 유튜브 댓글이라는 비정형 텍스트 데이터의 분석 정확도와 신뢰성을 확보하기 위해, 분석에 앞서 면밀한 전처리 과정을 수행하였다. 전처리는 데이터의 품질을 제고하고, 토픽모델링 및 키워드 분석과 같은 후속 기법의 효율성과 일관성을 확보하는 데 핵심적 역할을 한다.

초기 단계에서는 HTML 태그, 특수문자, 이모지, 허용되지 않는 문자, 중복된 공백, 비정상 문장 부호 등의 비의미 요소를 제거하였다. 또한, 댓글 데이터의 특성상 동일한 의미를 지닌 표현이 다양한 방식으로 표기되는 경우가 많지만, 댓글 고유의 어조와 문맥을 보존하고 콘텐츠 반응의 미세한 차이를 살리기 위해 원문을 가능한 한 유지하였다.

전처리 과정에는 KoNLPy의 Okt(Open Korean Text) 형태소 분석기를 활용하였다. Okt는 트위터와 같은 짧은 단문, 비표준어, 구어체가 혼재된 환경에서 효과적으로 작동하는 형태소 분석기로, 유튜브 댓글과 같이 구어적 표현과 축약형이 빈번한 텍스트 데이터에 적합하다. 이를 통해 각 댓글을 품사 기반 단어 단위(Token)로 분할하고, 의미 있는 고유어를 추출하였다.

전처리 결과, 전체 73,537개의 댓글에서 총 767,317개의 단어(Token)가 추출되었고, 이 중 고유 단어(Type)는 33,662개로 확인되었다. 이에 따른 어휘 다양성 지표(TTR: Type-Token Ratio)는 0.0438로 산출되었다. TTR은 텍스트의 어휘적 다양성을 측정하는 지표로, 값이 낮을수록 동일한 단어가 반복적으로 사용되고 있음을 의미한다. 본 연구의 낮은 TTR 수치는 전체 댓글에서 감탄사, 반복적 표현, 감정 중심 어휘의 사용 비중이 높다는 것을 시사한다. 이는 국가유산채널 콘텐츠에 대한 댓글 반응이 정보 전달보다 정서적 공감이나 즉각적 감탄 표현 중심으로 이루어졌음을 보여준다.

언어 선택에 있어 본 연구는 한국어와 영어로 작성된 댓글만을 분석 대상으로 한정하였다. 이는 분석 대상의 일관성을 확보함과 동시에, 국

가유산채널 콘텐츠가 일부 영문 자막 혹은 영문 소개 영상을 포함하고 있어 국제 시청자층의 반응을 부분적으로 반영할 필요성이 있었기 때문이다. 따라서 한국어와 영어가 혼합된 댓글은 유지하되, 기타 외국어(예: 일본어, 러시아어, 스페인어 등) 및 이모지만으로 구성된 댓글은 모두 제거하였다.

이러한 일련의 전처리 작업은 데이터의 구조적 특성을 최대한 보존하면서도 분석에 방해가 되는 노이즈를 제거함으로써, 토픽모델링의 정밀도와 키워드 해석의 신뢰도를 제고하는 데 크게 기여하였다. 특히, Okt 분석기를 기반으로 한 고유 단어 추출 및 정제 절차는 국가유산채널 콘텐츠 반응의 언어적 특수성을 정확히 반영하기 위한 필수적 기반 작업이었다.

제거된 불용어는 다음과 같다.

조사 & 접사류 (문법적 기능어 제거)

“은”, “는”, “이”, “가”, “을”, “를”, “에”, “와”, “과”, “에서”, “에게”, “에게는”, “한테”, “한테는”, “께”, “의”, “하다”, “됩니다”, “된다”, “했다”, “해요”,

반복 감탄사 / 의미 없음 (중복 감탄사 및 무의미한 표현의 제거)

“진짜”, “정말”, “그냥”, “너무”, “완전”, “약간”, “그래서”, “근데”, “그러나”, “입니다”, “인테”, “있네요”, “있을까요”, “없는데”, “없어요”, “같아요”, “있어요”, “되어다”,

유튜브 상에서 자주 나타나는 ‘메타’ 단어 (플랫폼 메타언어는 콘텐츠 내용과 직접 상관 없음)

“채널”, “구독”, “좋아요”, “댓글”, “라이브”, “스트리밍”, “보고”, “봤는데”, “봤습니다”, “봤어요”, “영상”,

프로젝트 도메인에 따라 의미 없는 단어 (의미 구분이 어려운 보조
용언 · 보조 동사)

“한국”, “대한민국”, “서울”, “좋다”, “있다”, “보다”, “되다”, “같다”,
“우리”, “자다”, “오다”, “멋지다”, “많다”, “가다”, “만들다”, “이다”,
“들다”, “싫다”, “이렇다”, “많이”, “없다”, “아니다”, “않다”, “생각”

4) 키워드 분석

본 연구는 국가유산채널 유튜브 댓글에 나타난 시청자 반응의 어휘적 경향성을 파악하기 위하여, 전처리 과정을 거쳐 정제된 텍스트 데이터를 대상으로 키워드 빈도 분석 및 TF-IDF 기반 중요도 분석을 수행하였다.

우선 출현 빈도 분석(Frequency Analysis)을 통해, 댓글 내에서 빈번하게 사용된 상위 30개의 고빈도 키워드를 도출하였다. 이를 기반으로, 국가유산채널 콘텐츠에 대해 시청자들이 자주 사용하는 핵심 반응 단어를 파악할 수 있었으며, 이를 시각화하기 위해 워드클라우드(Word Cloud) 기법을 적용하였다. 워드클라우드에는 고빈도 단어일수록 큰 글자로 표시되도록 구성되어, 시청자 인식의 중심 어휘를 직관적으로 확인할 수 있도록 한다.

또한, 형태소 분석기를 통해 각 댓글의 텍스트를 명사와 형용사 등 품사 단위로 분리하고, 각각의 출현 빈도를 별도로 산출하였다. 명사형 키워드는 주로 콘텐츠 주제와 관련된 핵심 개념어(예: ‘문화유산’, ‘응원’, ‘유산’, ‘국가’ 등)를 나타내며, 형용사형 키워드는 시청자의 정서적 평가(예: ‘아름답다’, ‘새롭다’, ‘소중하다’, ‘대단하다’ 등)를 나타내는 어휘로 분류되었다. 이를 통해 정보 전달어휘와 감성 반응어휘 간의 차이와 비중을 비교할 수 있었으며, 콘텐츠 수용자가 주제를 인식하는 방식과 그에 대한 감정을 어떻게 표현하는지를 보다 정밀하게 이해할 수 있었다.

한편, 단순 출현 빈도만으로는 전체 맥락에서 상대적으로 중요한 단어의 의미를 충분히 반영하기 어렵다는 한계를 보완하기 위하여, 본 연

구에서는 TF-IDF(Term Frequency-Inverse Document Frequency) 분석을 병행하였다. TF는 특정 단어가 하나의 댓글 내에서 얼마나 자주 등장하는지를 나타내며, IDF는 전체 댓글 집합 중 해당 단어가 등장하는 문서 수의 역수를 의미한다. 두 값을 곱한 TF-IDF는 전체적으로 자주 등장하지는 않지만, 특정 문맥에서 중요한 의미를 가지는 핵심어를 효과적으로 식별하는 데 유용하다.

TF-IDF 분석 결과, 단순 빈도 기준으로는 부각되지 않았던 상대적으로 ‘희귀하지만 중요도 높은 단어들이’ 추출되었으며, 이는 시청자의 관심 주제, 의도된 메시지, 특정 영상과의 정서적 연결성 등을 보다 구체적으로 이해하는 데 기여하였다. 특히 고빈도 키워드가 콘텐츠 전반에 걸친 일반 반응을 설명하는 반면, TF-IDF 상위 키워드는 특정 영상이나 주제에서 강한 감정 반응이나 관심을 유도한 어휘로 해석할 수 있다.

이러한 일련의 분석 결과는 국가유산채널 콘텐츠에 대한 시청자의 언어적 반응 구조를 정밀하게 파악하고, 콘텐츠별 선호 경향 및 감성적 수용 특성을 밝히는 데 중요한 기초 자료로 활용될 수 있다. 나아가 콘텐츠 전략 수립 및 문화유산 채널의 커뮤니케이션 방향 설정에 있어 실질적인 시사점을 제공한다는 점에서 그 의의가 있다.

5) 토픽모델링

본 연구는 유튜브 댓글과 같은 단문 중심의 비정형 텍스트 데이터에서 잠재된 주제 구조를 도출하기 위해, 최근 제안된 BERTopic(Bidirectional Encoder Representations for Topic Modeling) 알고리즘을 활용하였다. BERTopic은 사전학습된 BERT 계열 언어모델을 기반으로 문서(댓글)의 의미론적 임베딩(semantic embedding)을 생성한 후, 이를 바탕으로 주제를 도출하는 최신 토픽 모델링 기법이다(Grootendorst, 2022).

이 알고리즘은 다음과 같은 네 단계로 구성된다.

첫째, BERT 기반 임베딩 모델을 통해 각 댓글을 고차원 벡터 공간에 매핑한 후, UMAP(Uniform Manifold Approximation and Projection)을 적용하여 차원을 축소한다.

둘째, 축소된 벡터 공간에서 HDBSCAN(Hierarchical Density-Based Spatial Clustering of Applications with Noise)을 이용해 밀도 기반 군집화를 수행하고, 각 군집을 잠재적 ‘토픽’으로 정의한다.

셋째, 각 토픽에 속한 문서 집합에 대해 클래스 기반 TF-IDF(c-TF-IDF)를 계산하여 대표 키워드를 산출한다. 이는 전통적인 TF-IDF와 달리, 군집 단위로 단어의 중요도를 계산함으로써 토픽별 핵심 표현어를 효과적으로 식별할 수 있도록 한다.

넷째, 도출된 각 토픽의 대표 키워드는 c-TF-IDF 벡터의 크기를 기준으로 상대적 중요도를 비교하였다.

본 연구에서는 시계열 기반 변화 분석은 수행하지 않았으나, c-TF-IDF 기반 벡터는 시간 단위로 재구성되어 토픽 변화 추이를 분석하는 데 활용 가능하다. 실제로 이용구(2024)는 해당 기법을 활용해 시계열 기반 토픽 분석을 수행한 바 있으며, 본 연구 역시 향후 이를 적용할 수 있는 기초적 기반을 마련하였다.

BERTopic은 기존의 LDA(Latent Dirichlet Allocation)와 달리, 문맥을 보존한 문장 단위 임베딩을 활용함으로써 짧은 댓글에서도 정밀하고 해석 가능한 주제 도출이 가능하다는 강점을 지닌다. 특히 본 연구의 데이터는 감정 표현이 중심이 되는 단문 텍스트가 다수 포함되어 있어, 임베딩 기반 군집화는 노이즈를 줄이고 주제 간 중복을 최소화하는 데 효과적인 접근으로 평가된다.

이와 같은 분석 절차를 통해, 국가유산채널 유튜브 댓글에 나타난 잠재 주제 구조를 도출하였으며, 콘텐츠별 또는 재생목록별로 반복적으로 등장하는 주제와 시청자의 관심 영역, 정서적 반응의 양상을 실증적으로 파악할 수 있었다.

5. 결과 및 논의

1) 키워드 빈도 분석 결과

국가유산채널 유튜브 댓글에 자주 등장하는 핵심 단어를 확인하기 위해, 본 연구는 출현 빈도 분석을 실시하여 고빈도 키워드 목록을 도출하였다. 이를 통해 시청자들이 국가유산 콘텐츠에 반응할 때 주로 어떤 어휘를 사용하는지를 정량적으로 파악할 수 있었다. <표 3>은 전체 댓글 데이터에서 추출된 명사형 및 형용사형 고빈도 키워드의 빈도표를 제시하고 있다.

<표 3> 키워드 출현 빈도표

번호	명사형		형용사형	
	핵심어	빈도	핵심어	빈도
0	문화유산	8,189	아름답다	6,497
1	응원	7,146	새롭다	2,136
2	유산	6,934	소중하다	2,096
3	국가	6,764	대단하다	1,951
4	역사	3,867	다양하다	1,906
5	문화	3,516	신기하다	1,823
6	장인	3,506	유익하다	1,660
7	알다	3,297	그렇다	1,558
8	사람	3,207	예쁘다	1,537
9	전통	3,022	어떨다	1,513
10	우리나라	2,975	재밌다	1,513
11	소리	2,709	기대하다	1,305
12	축하	2,650	멋있다	1,101
13	모습	2,576	재미있다	1,064
14	마음	2,264	이쁘다	985
15	시간	2,189	깊다	973
16	하나	2,148	좋아한다	921
17	통해	2,030	자랑스럽다	880
18	더욱	2,029	흥미롭다	826
19	문화재	1,984	건강하다	806
20	사랑	1,953	즐겁다	762

번호	명사형		형용사형	
	핵심어	빈도	핵심어	빈도
21	발굴	1,907	궁금하다	755
22	유물	1,854	행복하다	752
23	자연	1,741	스럽다	702
24	로고	1,737	귀엽다	659
25	지금	1,703	고맙다	625
26	세계	1,701	부탁드리다	572
27	대해	1,686	맛있다	564
28	관심	1,667	아쉽다	538
29	항상	1,603	안되다	523



또한, 단순 출현 빈도만으로는 텍스트 내 의미의 중요도를 충분히 설명하기 어려운 한계점을 보완하고자, 본 연구에서는 TF-IDF 분석을 병행하였다. TF-IDF는 정보 검색 및 텍스트 마이닝 분야에서 널리 사용되는 통계적 가중치로서, 다수의 문서로 구성된 문서 집합 내에서 특정 단어가 개별 문서 내에서 얼마나 중요한지를 수치적으로 표현하는 지표이다(Sparck Jones, 1972). <표 4>는 이러한 TF-IDF 값을 기반으로 산출된 상대적 중요도가 높은 키워드 목록을 제시하고 있으며, 단순히 자주 등장한 단어가 아니라 의미적으로 핵심성을 지닌 어휘를 확인하는 데 기여한다.

빈도 분석 결과, 명사형 키워드 중에서는 ‘문화유산’, ‘응원’, ‘유산’, ‘국가’ 등이 상위에 도출되었다. 이는 시청자들이 콘텐츠에 대해 매체적 특성(문화유산이라는 주제), 소속감 및 정체성(국가, 유산), 정서적 지지(응원)를 중심으로 반응하고 있음을 시사한다. 한편, 형용사형 키워드에

서는 ‘아름답다’, ‘새롭다’, ‘소중하다’, ‘대단하다’ 등이 높은 빈도를 보였는데, 이는 시청자들이 콘텐츠에 대해 긍정적인 정서, 미적 감상, 의미 있는 가치 인식을 기반으로 언어를 사용하고 있음을 보여준다. 특히 ‘아름답다’는 가장 높은 형용사 빈도를 기록하며, 콘텐츠에 대한 감정적 몰입의 정도가 높음을 뒷받침한다.

이러한 고빈도 키워드는 워드클라우드 시각화를 통해 시청자 반응에서 두드러지게 나타나는 핵심 단어들의 상대적 비중을 직관적으로 파악할 수 있게 한다. 명사형과 형용사형 키워드를 각각 시각화한 결과, 콘텐츠에 대한 주제 지향적 반응과 감정적 표현이 혼재되어 나타나는 양상이 시각적으로 확인되었다.

〈표 4〉 키워드 가중치 (TF-IDF)

번호	핵심어	점수	번호	핵심어	점수
0	아름답다	3104.68	15	해주다	1366.05
1	문화유산	3001.54	16	이렇게	1360.09
2	응원	2804.44	17	우리나라	1324.32
3	국가	2131.66	18	먹다	1293.54
4	감사하다	2131.42	19	대단하다	1242.16
5	유산	2098.56	20	함께	1226.39
6	역사	1750.72	21	축하	1216.46
7	장인	1718.32	22	모습	1207.99
8	알다	1634.15	23	신기하다	1179.02
9	사람	1569.68	24	유물	1108.49
10	나오다	1509.40	25	마음	1105.92
11	드리다	1486.14	26	모르다	1048.22
12	문화	1470.47	27	하나	1042.89
13	소리	1431.04	28	받다	1024.15
14	전통	1396.78	29	시간	1009.45

그러나 TF-IDF 분석 결과는 단순 출현 빈도 분석과는 다소 다른 양상을 보였다. 상위 키워드로는 ‘아름답다’, ‘문화유산’, ‘응원’, ‘국가’, ‘감사하다’, ‘유산’, ‘역사’, ‘장인’ 등이 도출되었으며, 이는 시청자들이 국가유산 콘텐츠에 대해 정서적 감탄과 문화적 가치를 동시에 인식하고 있음을 보여준다. 특히 ‘대단하다’, ‘신기하다’, ‘축하’, ‘마음’ 등의 키워드는 콘텐츠에 대한 감성적 반응과 긍정적 인상 평가를 나타내는 어휘로 해석되며, 이는 국가유산채널의 콘텐츠가 단순 정보 전달을 넘어 감동과 공감을 유도하고 있음을 시사한다.

종합적으로, 단순 빈도 기반 키워드와 TF-IDF 기반 중요 키워드 모두 국가유산채널 콘텐츠에 대한 긍정적 인식과 응원, 문화적 정체성 강조와 밀접한 관련이 있음을 보여준다. 이는 해당 콘텐츠가 공공 문화유산 콘텐츠로서 시청자와 감성적·가치적으로 연결되고 있다는 실증적 단서로 해석할 수 있다.

2) 토픽 분석 결과

유튜브 댓글에 내재된 잠재적 주제 구조를 파악하기 위해, 본 연구는 BERTopic 알고리즘을 적용하여 주요 토픽을 도출하였다. BERTopic은 토픽 수를 사전에 지정하지 않아도 군집 기반으로 자동 결정할 수 있는 장점을 지니나, 초기 자동 추출 결과에서 지나치게 세분화된 토픽이 도출되어 해석 가능성과 비교 분석의 어려움이 있었다. 이에 따라, 주제 간 의미적 유사성과 표현상의 중복을 고려하여 적절한 의미 단위로 통합할 필요가 있었다. 또한 선행연구(노영아, 2024; 강진구, 2024)에서 10개 이내의 수준으로 분류하고 있어 토픽 간 분류의 명확성과 분석의 실용성을 확보하기 위해 최종적으로 10개의 토픽으로 수렴하였다. 각 토픽은 의미적으로 유사한 댓글의 군집을 대표하며, 시청자들이 국가유산채널 콘텐츠에 대해 어떠한 주제에 관심을 보이고 있는지를 반영한다.

〈표 5〉 토픽 분석 결과

번호	토픽이름	가중치	핵심어	그룹
토픽 0	전통공예 및 무형문화	18.70	'장인', '감사하다', '아름답다', '전통', '한복', '문화유산', '응원', '하나', '작품', '무형문화재'	A
토픽 1	음식 콘텐츠	18.61	'먹다', '누에', '나오다', '소리', '알다', '어떨다', '사람', '맛있다', '미치다', '넘다'	B
토픽 2	홍보 콘텐츠 및 팬덤	12.33	'송가', '타임머신', '한양도성', '최고', '홍보', '대사', '뉴진스', '가수', '응원', '대박'	B
토픽 3	역사 및 고대사	11.80	'신라', '역사', '일본', '유물', '발견', '조선', '백제', '나라', '고려', '중국'	A
토픽 4	자연경관 및 지역	10.60	'아름답다', '자연', '가보다', '풍경', '가을', '안동', '사찰음식', '함께', '제주', '마음'	A
토픽 5	전통무예 및 유산 사업	9.70	'문화유산', '택견', '응원', '발굴', '문화재', '문화', '사업', '유익하다', '인도', '조사'	A
토픽 6	긍정 형용사 표현	8.94	'예쁘다', '이쁘다', '대단하다', '멋있다', '신기하다', '귀엽다', '아름답다', '한복', '재밌다', '어울리다'	B
토픽 7	브랜드 및 명칭 변경	7.99	'유산', '국가', '로고', '응원', '축하', '문화유산', '새롭다', '드리다', '변경', '진심'	A
토픽 8	기념일 및 축하	1.08	'아름답다', '감사하다', '주년', '진주', '13', '하씨', '출토', '유물', '축하', '정답'	B
토픽 9	참여 독려 및 이벤트형 반응	0.25	'완료', '참여', '깜바레', '스키', '작성', '공유', '아이', '한번', '수면제', '발표'	B

도출된 10개의 토픽 중 가장 높은 비중을 차지한 토픽 0(전통공예 및 무형문화)은 전체 댓글의 18.70%를 구성하며, '장인', '감사하다', '아름답다', '전통', '한복', '문화유산', '응원' 등의 키워드가 중심을 이룬다. 이 토픽은 시청자들이 콘텐츠 감상 후 전통문화와 장인의 기술에 감탄하거나 정서적으로 지지하는 경향을 보여주는 반응으로 해석할 수 있다.

토픽 1(음식 콘텐츠)은 18.61%의 비중을 차지하며, ‘먹다’, ‘누에’, ‘나 오다’, ‘소리’, ‘맛있다’ 등의 키워드가 등장한다. 이는 콘텐츠 내 특정 장면이나 소재, 특히 식문화·음향 요소에 대한 구체적 관심을 반영하며, 시청자의 인지적 주의와 정보 중심 반응을 나타낸다.

토픽 2(홍보 콘텐츠 및 팬덤)는 12.33%를 차지하고, ‘송가’, ‘타임머신’, ‘한양도성’, ‘뉴진스’, ‘가수’, ‘응원’, ‘대박’ 등의 키워드를 포함한다. 해당 토픽은 K-POP 아티스트(예: 뉴진스)의 출연이나 유명 인물의 내레이션 참여 등, 문화유산 콘텐츠와 대중문화의 융합에 따른 팬덤 기반 반응과 콘텐츠 확장성에 대한 시청자의 호응을 보여준다.

토픽 3(역사 및 고대사)과 같은 토픽에서는 ‘신라’, ‘역사’, ‘유물’, ‘조선’, ‘백제’ 등의 역사적 키워드가 등장하며, 콘텐츠의 역사적 깊이와 배경 설명에 대한 관심을 반영한다.

토픽 4(자연경관 및 지역)는 ‘풍경’, ‘가을’, ‘자연’, ‘마음’ 등과 함께 감상적이고 심미적인 반응을 보여주는 한편, 토픽 5(전통무예 및 유산 사업)는 ‘택견’, ‘발굴’, ‘문화재’, ‘조사’ 등 전통문화와 문화유산 정책 콘텐츠에 대한 관심을 나타낸다.

토픽 6(긍정 형용사 표현), 토픽 7(브랜드 및 명칭 변경), 토픽 8(기념일 및 축하) 등에서는 긍정 정서 표현(‘예쁘다’, ‘귀엽다’, ‘축하’, ‘진심’)이 반복되며, 콘텐츠 전반에 대한 정서적 지지와 긍정적인 감상 평가가 주를 이룬다.

마지막으로 토픽 9(참여 독려 및 이벤트형 반응)는 가장 낮은 비중(0.25%)을 차지하지만, ‘참여’, ‘공유’, ‘작성’, ‘발표’ 등 참여 기반 반응이 담겨 있어, 이벤트형 콘텐츠 혹은 과제형 시청 경험과 관련된 일시적 반응일 가능성을 시사한다.

이러한 분석 결과는 시청자들이 국가유산 콘텐츠를 단순한 정보 제공이 아닌, 정서적·심미적·참여적 차원에서 복합적으로 수용하고 있음을 보여준다. 특히 다수의 토픽에서 반복적으로 나타나는 ‘응원’, ‘감사

하다, ‘멋있다’, ‘아름답다’ 등의 키워드는 국가유산 콘텐츠가 문화적 자
궁심, 정체성 강화, 감동 공유를 유도하는 매개체로 기능하고 있음을 시
사한다.

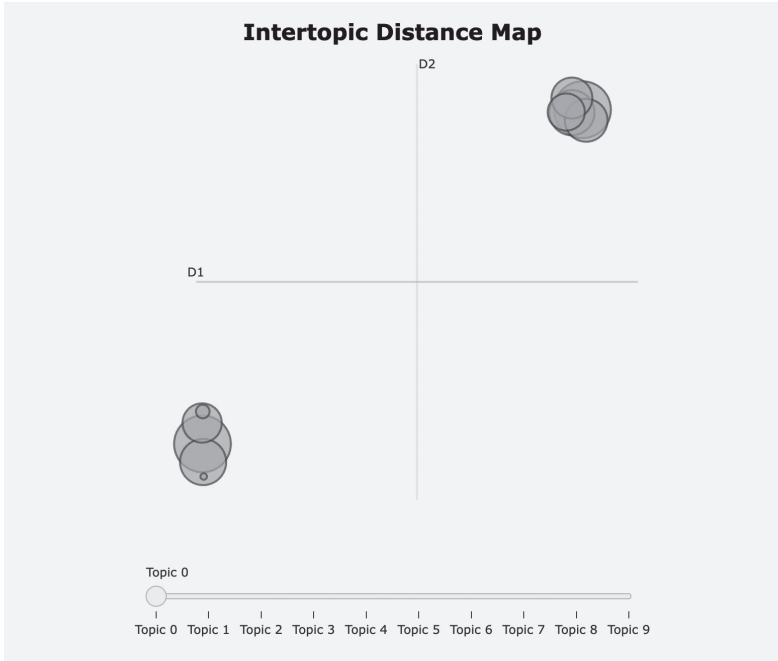
또한 콘텐츠 자체의 내용뿐 아니라, 편집, 해설, 연출, 연예인 출연
등 외적 요소에 대한 피드백도 다수 관찰되었으며, 이는 시청자들이 콘
텐츠를 미디어적 텍스트로 능동적으로 해석하고 있다는 점을 보여준다.
BERTopic 분석 결과는 국가유산채널이 단순한 문화정보 전달을 넘어,
디지털 미디어 플랫폼에서의 공감·참여형 콘텐츠로 진화하고 있음을
실증적으로 뒷받침한다.

토픽모델링 결과를 시각화한 <그림 3>에 따르면, 전체 댓글 토픽은
두 개의 주요 그룹(Cluster)으로 명확히 구분되어 분포하는 것으로 나타
났다. 이는 토픽 간 의미적 유사성에 기반한 거리 계산 결과이며, 그룹
간 연관성이 낮은 이중 구조를 시사한다.

첫 번째 그룹(Cluster A)은 ‘전통’, ‘역사’, ‘장인’, ‘무형문화재’ 등 전통
성과 문화적 진지함에 대한 시청자 반응을 중심으로 형성되어 있다. 반
면, 두 번째 그룹(Cluster B)은 ‘응원’, ‘예쁘다’, ‘가수’, ‘대박’ 등과 같이
감성적 참여, 한류적 요소, 디지털 문화소비 행태를 반영하는 반응이 다
수를 차지한다.

이러한 이중 분포 구조는 국가유산채널 콘텐츠의 특성과 시청자 유
입 경로의 이질성에서 기인한 것으로 해석된다. 전자는 국가유산 본연의
콘텐츠(전통문화, 역사, 공예 등)에 대한 관심층의 반응으로, 후자는 유
명인 출연 콘텐츠, 이벤트성 영상, 유튜브 알고리즘 등을 통해 유입된 신
규 이용자의 정서적이거나 즉흥적인 반응을 중심으로 한다.

결과적으로, 국가유산채널의 시청자 반응은 정통적 문화유산 수용자
와 K-콘텐츠 기반 신규 유입층이라는 상이한 집단 간 반응 양상을 보여
주며, 향후 콘텐츠 기획 및 운영에 있어 시청자 집단별 맞춤형 커뮤니케
이션 전략 수립의 필요성을 시사한다.



〈그림 3〉토픽 모델링

3) 재생목록 별 시청자 반응

국가유산채널의 콘텐츠 유형별 댓글 반응 특성을 파악하기 위해, 본 연구는 유튜브 채널에서 제공하는 공식 재생목록 50개를 분석 대상으로 선정하고, 각 재생목록에 포함된 영상 수 및 해당 콘텐츠에 달린 댓글 수의 평균값을 산출하였다. 재생목록 선정은 목록에 포함된 콘텐츠 중 댓글이 하나 이상 포함된 재생목록을 대상으로 선정하였다. 분석에서는 언어적 내용이나 맥락을 고려하지 않고, 댓글이 달린 수량 자체를 기준으로 하였으며, 댓글이 포함되지 않은 재생목록은 제외하였다.

그 결과, 전체 콘텐츠 대비 상대적으로 댓글 수가 높은 재생목록은 ‘발굴’ 관련 콘텐츠군이었다. 이는 일반인이 접근하기 어려운 발굴 현장이나 고고학적 작업 과정을 영상으로 제공함으로써, 시청자의 호기심과 관심을

자극한 결과로 해석된다. 특히 소규모 발굴조사에 국비가 지원되는 과정이나, 문화재 발굴이 개인 재산권과 충돌하는 사례 등이 다뤄진 콘텐츠는 시민의 실질적 생활 영역과 연결되며 공적 관심을 유발한 것으로 판단된다.

또한 K-POP 아이돌 그룹 슈퍼주니어의 ‘예성’이 참여한 ‘국가유산 예세이’ 시리즈는 평균 245개의 댓글 수를 기록하여, 대중문화 요소와 전통 콘텐츠의 융합이 시청자 반응을 유도하는 데 효과적임을 보여주었다. 이는 국가유산채널의 문화유산 콘텐츠 확산 전략에서 한류와의 결합이 실질적인 시너지 효과를 창출할 수 있음을 시사한다.

한편, ASMR 재생목록의 경우에도 댓글 반응이 활발하게 나타났다. 해당 콘텐츠는 언어적 요소보다 청각적 몰입을 유도하는 형식으로, 국가유산 장인의 작업 소리 등을 통해 국가유산을 감각적으로 체험할 수 있도록 구성되었다. 이는 언어 장벽을 최소화하고, 문화유산의 실재감을 간접적으로 전달하는 방식으로 긍정적 반응을 얻은 사례로 해석된다.

종합적으로 해석해보면 이상의 분석 결과는 국가유산채널의 콘텐츠가 주제에 따라 정서적 공감, 정보 탐색, 감각적 체험 등 다양한 수용자 반응을 유도하고 있음을 보여준다. 특히 재생목록별 댓글 반응의 차이는 시청자의 관심 분포가 콘텐츠의 형식·내용·출연 인물 등에 따라 뚜렷하게 달라짐을 시사하며, 향후 콘텐츠 기획 시 주제별 반응 데이터를 기반으로 한 차별화 전략 수립의 필요성을 보여준다.

〈표 6〉 재생목록 평균 댓글 수

순위	재생목록	영상 수	댓글 수	평균
1	(이벤트소식)함께하는 국가유산채널	59	22,561	382
2	국가유산 발굴스토리 (Excavation Documentary on Korean Cultural Heritage)	14	3,818	273
3	국가유산 예세이(Yessay) ASMR by an Artist	19	4,658	245
4	ASMR_Korean Cultural Heritage	84	16,960	202

순위	재생목록	영상 수	댓글 수	평균
5	[국가유산 방문 캠페인] 인기 동영상 (Cultural Heritage Visit Campaign Popular Videos)	38	5,832	153
6	[교양] 한국의 장인(Korean Artisan)	42	6,095	145
7	국가유산 발굴스토리 Mistery X History	63	7,375	117
8	[국가유산채널] 인기 동영상 (K-Heritage Channel's Popular Videos)	499	47,266	95
9	국가유산순례 '드오' Korean Heritage Tour 'De o'	10	818	82
10	[특집 다큐] 직지, 활자의 시간여행	6	486	81
11	[미니 다큐] 5분 추리 역사 X파일	25	1,842	74
12	[예능] 한국(동양) VS 서양 세기의 대결	16	1,011	63
13	[예능] 국가유산X여행 (Korea Heritage Tour)	55	3,349	61
14	Relaxing ASMR(with Music)	20	1,138	57
15	[국가유산 방문 캠페인] 소개 (Cultural Heritage Visit Campaign Promotional Videos)	36	1,545	43
16	[교양, 예능] 이야기 국가유산 (A story of Heritage)	87	3,342	38
17	[교양, 예능] 코리아 뷰티 인사이드(Korea Beauty Inside)	4	148	37
18	방송제작 특집 프로그램	15	430	29
19	[미니 다큐] 7분다큐	54	1,493	28
20	[교육] 교육e 국가유산	45	1,241	28
21	[교육] 큰별쌤 최태성의 교과서에 나오는 우리 국가유산	23	416	18
22	[교양] 한국 속의 과학	11	189	17
23	[이미지] 천연기념물/동물 (Korean Natural Monument Animal)	30	500	17
24	[교양] 우리는 우리를 아는가	11	173	16
25	[Image] Awesome Heritage	20	310	16
26	[교양] 내가 사랑한 국가유산	28	408	15
27	[예능] 쇼미더 국가유산	25	364	15

순위	재생목록	영상 수	댓글 수	평균
28	국가유산 미니어처, 디오라마 (K-heritage Miniatures and Dioramas)	5	66	13
29	[특집 다큐] 한지, 루브르 종이전쟁 (The special story about Hanji)	10	130	13
30	이미지 국가유산(Image Heritage Korea)	223	2,808	13
31	[예능] 귀여운 동물들과 함께하는 역사 이야기 (A History Story with Cute Animals)	11	122	11
32	[교육] 썬킴의 이런(e-learn) 역사이야기 교과서, 수업 연계	24	258	11
33	[교육] 조선왕조 궁중음식 교육 강의	8	82	10
34	[교양] 국가유산 돋보기	15	147	10
35	[교양] 국악! 이것만은 알고 가자	12	113	9
36	[Image] Capture the Best Moments : Korean Heritage Series	6	55	9
37	[다큐] 국가유산 여행길 (EBS TV방영)	63	539	9
38	Landmark : Minecraft meets Beautiful Heritage Sites	10	83	8
39	[정보] 국가유산 뉴스 (Heritage News)	322	2,242	7
40	[교양] 국가유산X파일	15	100	7
41	[교육] 교과서 속 국가유산 이야기 교과 과정 연계	32	195	6
42	[교양] 정은표 가족의 세계유산순례	11	67	6
43	[정보] 국가유산청 사람들	33	183	6
44	[다큐] 섬 문화유산 기행(Island Visiting Cultural Heritage)	12	56	5
45	[교양, 다큐] 한국의 정원(Korean Gardens)	10	46	5
46	국가유산채널 소개	11	42	4
47	[정보] 국가유산 직업이야기	15	54	4
48	[교양, 다큐] 한국의 서원 (Seowon, Korean Neo-Confucian Academies)	11	34	3
49	[Image] Korean Natural Monument Animal	18	19	1
50	[정보] 비어있는 전시회 작품 감상	22	11	1

6. 결론

본 연구는 공공 주도로 운영되어 온 문화유산 콘텐츠의 수용자 반응을 실증적으로 분석하기 위해, 국가유산채널의 유튜브 댓글 데이터를 기반으로 시청자 참여 양상과 수용 특성을 파악하고, 향후 콘텐츠 기획 및 디지털 미디어 전략에 실용적 시사점을 제시하는 것을 목표로 하였다. 2010년부터 2024년까지 업로드된 총 2,605개 영상 중 댓글이 존재하는 2,115편을 대상으로 73,537건의 댓글을 수집하였으며, 이 중 분석 적합성이 확보된 67,826건에 대해 텍스트 마이닝 기법을 적용하였다.

1) 주요 연구 결과

첫째, 연도별 댓글 분포 분석 결과, 2020년 COVID-19 팬데믹 시기를 기점으로 댓글 수가 급증하였고, 이후 거리두기 해제 이후에도 시청자 반응은 꾸준히 증가하였다. 이는 콘텐츠의 질적 개선과 문화유산에 대한 지속적인 관심을 반영한다.

둘째, 키워드 빈도 분석 및 BERTopic 기반 토픽모델링을 통해 시청자들은 콘텐츠에 대해 전통문화의 가치와 미적 요소에 감탄하거나 정서적으로 지지하는 반응을 보이고 있었다. ‘장인’, ‘전통’, ‘무형문화재’ 등은 전통성에, ‘아름답다’, ‘예쁘다’ 등은 심미적 평가에 해당하며, ‘응원’, ‘감사하다’, ‘진심’ 등은 감정적 공감을 나타내는 것으로 해석된다.

셋째, K-POP 아티스트(예: 예성)가 출연한 콘텐츠에서는 ‘송가’, ‘가수’, ‘홍보’, ‘대박’ 등의 키워드가 도출되어, 팬덤 중심의 정서적 참여와 집중적 반응을 유도하고 있었음을 확인할 수 있었다. 이는 문화유산 콘텐츠에 한류 요소를 결합할 경우 국제적 확산 효과를 기대할 수 있음을 시사한다.

넷째, 재생목록별 평균 댓글 수 분석 결과, 전통유산 콘텐츠 외에도 발굴 현장, ASMR 기반 장인 콘텐츠, 대중문화 협업 콘텐츠 등이 높은

반응을 얻었다. 이는 시청자의 다양한 관심사에 기반한 콘텐츠 전략이 요구됨을 보여준다.

2) 시사점과 한계

본 연구는 국가유산 콘텐츠가 단순한 정보 전달을 넘어 감성적 공감과 문화적 정체성 형성에 기여하는 디지털 커뮤니케이션 매체로 기능하고 있음을 입증하였다. 이러한 분석은 국가유산 홍보 및 콘텐츠 정책 수립 시 실질적 기초자료로 활용 가능하다.

그러나 다음과 같은 한계점이 존재한다. 첫째, 분석 대상이 유튜브로 국한되어 다른 디지털 플랫폼의 수용자 반응을 반영하지 못했다. 둘째, 한글과 영어 이외의 댓글은 분석 대상에서 제외되어 글로벌 수용자 반응을 온전히 파악하지 못하였다. 셋째, 시계열 분석을 생략함으로써 시간 흐름에 따른 수용자 반응의 변화를 다루지 못하였다.

그럼에도 본 연구는 감성적 공감과 사용자 참여 유도 전략의 중요성을 제시하였으며, 텍스트 마이닝 기반 분석이 공공 문화콘텐츠 수용자 연구에 효과적인 접근임을 확인하였다.

아울러 본 연구의 분석 결과는 2023년 「국가유산기본법」 제정과 관련 법 개정과도 연결된다. 기존의 ‘문화재’ 개념은 ‘국가유산’이라는 상위 범주로 재편되었고, 이는 콘텐츠의 기획 범위를 생태, 생활문화, 예술, 서사 등으로 확대하는 제도적 기반이 되었다. 본 연구에서 도출된 시청자 반응은 이러한 제도 변화와 맞물려, 국가유산의 의미와 활용 가능성을 대중적으로 확장시킬 수 있는 근거로 작용할 수 있다.

특히 ‘전통성’, ‘예술성’, ‘정서적 공감’ 등의 키워드는 시청자가 수동적 수용자가 아니라 능동적 해석 주체임을 보여준다. 이는 ‘재(財)’로서의 문화재가 아닌, ‘유산(遺産)’으로서의 문화적 자산이라는 인식이 콘텐츠 수용자 내부에 형성되고 있음을 시사한다.

나아가 본 연구는 국가유산 콘텐츠의 기획·운영에 있어 다음과 같

은 실천적 전략을 제안한다. (1) 시청자 반응 데이터를 기반으로 한 수요 기반 기획 전략 수립, (2) 콘텐츠 주제별 반응 분석 시스템의 개발, (3) 반응이 높은 주제에 대한 시리즈화 전략 강화, (4) 법제 변화에 따른 기획 범위 확장. 특히 생태·자연·무형유산 영역을 포함하는 주제 기획은 수용자 저변 확대에 기여할 수 있다.

또한 국가유산채널 콘텐츠는 전통문화, 자연환경, 역사적 서사, 대중문화 협업 등 다층적인 구성요소를 포함하는 경우가 많다. 본 연구에서는 재생목록 기준의 단일 분류 체계를 따랐으나, 이러한 복합 콘텐츠를 분석하기 위해서는 향후 내용 기반 다중 분류(content-based multi-label classification) 혹은 주제 태깅(topic tagging)을 활용하는 방식이 요구된다. 이를 통해 시청자의 정교한 반응 해석 및 콘텐츠 기획 고도화가 가능해질 것이다.

3) 향후 연구 제안

이러한 연구 결과는 수용자 중심의 문화유산 콘텐츠 전략 수립에 있어 실증적 기반을 제공하며, 디지털 시대 공공 커뮤니케이션의 방향성을 제시하는 데 기여한다.

앞으로는 국가유산채널 외 다양한 플랫폼(예: 인스타그램, 틱톡, 네이버TV 등)에서의 수용자 반응을 비교 분석하고, 댓글에 대한 감성 분석, 시계열 분석, 행위자 네트워크 분석 등 복합적 접근을 통해 보다 정교하고 입체적인 디지털 문화유산 커뮤니케이션 전략 수립을 위한 후속 연구가 필요하다.

또한 문화유산 콘텐츠의 국제적 확산 가능성을 고려하여, 다국어 댓글 분석을 포함한 글로벌 수용자 반응 연구, 그리고 콘텐츠의 시청각적 구성이나 내러티브 구조가 수용자의 정체성과 문화 감각 형성에 어떠한 영향을 미치는지를 탐색하는 질적 연구도 병행되어야 한다.

참고 문헌

- 강진구. (2024). 텍스트마이닝으로 본 대구 이슬람 사원 갈등: 유튜브 댓글을 중심으로. <다문화콘텐츠연구>, 47, 7-37. <https://doi.org/10.15400/mccs.2024.03.47.01>
- 김현미, 김희주, 이보현. (2024). 인터랙티브 스토리텔링 콘텐츠의 상호작용성이 지속이용의도에 미치는 영향: 지각된 실재감과 몰입의 매개적 역할 탐구. <미디어융합연구>, 36, 73-117. <https://doi.org/10.22814/sgjcr.2024.36.73>
- 노영아. (2024). 유튜브 댓글 분석을 통한 클래식 음악 시청자 반응 연구: 피아니스트 임윤찬 연주 댓글 중심으로. <예술경영연구>, 70, 65-90. <https://doi.org/10.52564/JAMP.2024.70.65>
- 이수재. (2016). 문화유산포털의 대중 소통 전략: 한국문화재단 <문화유산채널>을 중심으로. <글로벌문화콘텐츠>, 22, 161-178
- 이신행, 현아연. (2023). 유튜브 '사이버택카' 채널은 어떻게 악성댓글을 양산하는가?: 유튜브 익명성, 규범 동조, 혐오의 반향실 효과를 중심으로. <사이버커뮤니케이션학보>, 40(2), 53-88. <https://doi.org/10.36494/JCAS.2023.06.40.2.53>
- 이용구, 김선욱. (2024). LDA, Top2Vec, BERTopic 모형의 토픽모델링 비교 연구: 국외 문헌정보학 분야를 중심으로. <한국문헌정보학회지>, 58(1), 5-30. <https://doi.org/10.4275/KSLIS.2024.58.1.005>
- 이종욱. (2019). 디지털 문화유산 국내외 정책 분석 및 개선 방안 제언. <한국컴퓨터정보학회논문지>, 24(10), 71-78. <https://doi.org/10.9708/jks-ci.2019.24.10.071>
- 정낙원, 김성욱. (2020). 네이버, 다음의 문제인 정부 지지율 여론조사 뉴스 댓글에 나타난 댓글 여론과 적대적 언어 분석. <미디어융합연구>, 28, 58-106. <http://dx.doi.org/10.22814/sgjcr.2020.28.58>
- 정젤나, 권기준. (2024). 유튜브 관광콘텐츠의 사회적 실재감과 관광 행동의도와의 구조적 관계: AIDA 모델을 중심으로. <관광학연구>, 48(8), 151-167. <https://doi.org/10.17086/JTS.2024.48.8.151.167>

최제호, 박선호, 현대원. (2024). BTS와 이세계아이돌 유튜브 콘텐츠 댓글 비교: 형용사 사용 패턴과 텍스트마이닝을 통한 정서적 상호작용 분석. <디지털콘텐츠학회논문지>, 25(4), 1009-1020. <https://doi.org/10.9728/dcs.2024.25.4.1009>

정보화사회진흥원. (2014). <국가정보화 20년의 성과>

법제처 법제조정총괄법제관실. (2024). 국가유산(National Heritage) 법제. <법제소식>, 5월호, 26-31

국가유산청. (2021). 문화재청, ‘문화재 디지털 대전환 2030’ 발표 [On-Line]. <국가유산청 뉴스>. URL: http://www.khs.go.kr/newsBbz/select-NewsBbzView.do?newsItemId=155702775§ionId=b_sec_1

국가유산디지털서비스. (2025). 국가유산 디지털서비스 소개 [On-Line]. URL: <https://digital.khs.go.kr>

국가유산채널. (2025). 국가유산채널 유튜브 서비스 [On-Line]. URL: <https://www.k-heritage.tv>

와이즈앱. (2024). 2024년 1월 한국인의 유튜브 사용시간 통계 [On-Line]. URL: <https://www.wiseapp.co.kr/insight/detail/512>

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. , arXiv:2203.05794

Sparck Jones, K. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. , 28(1), 11-21. <https://doi.org/10.1108/eb026526>

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2023). A Biterm Topic Model for Short Texts. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 1445-1456)

Yin, J., & Wang, J. (2014). A Dirichlet Mixture Model for Short Text Clustering. In *Proceedings of the 2014 IEEE International Conference on Big Data* (pp. 144-151)

Angelov, D. (2020). Top2Vec: Distributed Representations of Topics [On-Line]. GitHub. URL: <https://github.com/ddangelov/Top2Vec>

UNESCO World Heritage Convention. (1972). *What is World Heritage?* [On-Line].
Retrieved from <http://whc.unesco.org/en/about/>
UNESCO. (2003). *Charter on the Preservation of the Digital Heritage* [On-Line].
Retrieved from <https://www.unesco.org/en/legal-affairs/charter-preservation-digital-heritage>

논문 최초 투고일 2025년 05월 01일

논문 심사일 2025년 05월 30일

논문 게재 확정일 2025년 06월 12일

Analyzing Audience Responses to K-Heritage Channel Video Content Using Text Mining : Focusing on YouTube Comments

Gyungmin Jo

(Ph.D. Program in Cultural Informatics, The Academy of Korean Studies)

This study aims to examine the patterns of viewer responses and topic interests reflected in YouTube comments on videos uploaded by the K-Heritage Channel, a public platform managed by the Korean government. In contrast to previous studies that primarily relied on producer-centered perspectives or traditional surveys, this research adopts a user-centric approach by analyzing user-generated content (UGC)-specifically YouTube comments to uncover how digital audiences engage with and emotionally respond to cultural heritage content.

The dataset consists of 73,537 comments extracted from 2,115 videos published between June 2010 and December 2024. After preprocessing, a total of 67,826 comments were retained for analysis. The methodological framework includes keyword frequency analysis, TF-IDF weighting, and BERTopic-based topic modeling, supplemented by an analysis of comment volume by content playlist.

As a result of the analysis, viewers expressed a variety of responses that went beyond simple information acquisition, including admiration for the beauty of traditional culture, emotional support, and expressions of cultural pride. Keywords such as “cultural heritage,” “support,” “artisan,” and “beautiful” showed high frequencies, while adjective-type keywords highlighted a strong tendency toward positive sentiment and impression-based evaluations. In the TF-IDF analysis, high-scoring terms such as “beautiful,” “thankful,” “support,” “nation,” and “amazing” indicated strong emotional engagement and empathy toward the content. Using BERTopic, the study extracted ten major topics, which formed two distinct clusters: one reflecting an appreciation for traditional cultural values, and the other centered on emotional participation and responses to K-content collaborations. The findings also revealed differentiated patterns of viewer reactions depending

on content format and theme, particularly in videos featuring traditional artisans, archaeological sites, K-POP collaborations, and ASMR-style presentations.

This study demonstrates that digital cultural heritage content on YouTube not only delivers information but also promotes affective engagement and identity formation among viewers. The findings offer both theoretical and practical implications, serving as a valuable resource for the development of data-driven, audience-centered content strategies and cultural heritage promotion policies in the digital age.

Keywords Keywords: K-Heritage Channel, YouTube comments, Text Mining, topic modeling, digital cultural heritage