https://doi.org/10.22743/kwr.2026.8.5.1235

네이버 포털 기반 텍스트마이닝 분석

이유안 국립경국대학교

논문요약

본 연구는 네이버 검색 응용 프로그래밍 인터페이스에서 밀양아리랑과 정선아리랑을 주요 검색어로 설정하여 관련 중심어와 주제적 맥락을 분석하고, 네이버 데이터랩 자료를 활용하여 두 지역의 검색 관심도를 비교하는 것을 목적으로 하였다. 이를 위해 빈도분석, 이항결합어 분석, 단어빈도-역문서빈도 분석, 잠재 디리클레 할당 토픽모델링 및 검색 추이 분석을 수행하였다. 분석 결과, 밀양아리랑은 축제, 캠핑, 천문대, 영남루, 밀양강 등 관광시설과 체험관광을 중심으로 언급된 반면, 정선아리랑은 시장, 열차, 공연, 먹거리 등 전통시장, 관광교통 및 지역기관과 주로 연계되어 나타났다. 잠재 디리클레 할당 분석에서는 두 아리랑에서 각각 6개의 토픽이 도출되었다. 밀양아리랑은 문화기관, 가족관광, 체험시설, 축제, 지역행사 및 체류형 관광이 강조되었으며, 정선아리랑은 시장, 지역상품, 공연, 관광열차, 지역 연계 및 향토음식이 주요 주제로 나타났다. 검색 관심도는 2016년부터 2020년까지 정선아리랑이, 2021년 이후에는 밀양아리랑이 더 높았다. 이러한 결과는 동일한 무형문화유산이라도 온라인 재현에는 지역별 차이가 존재하며, 이러한 차이가 지역 관광 인프라, 상업 구조 및 문화자원 활용 방식의 영향을 받는다는 점을 보여준다. 또한 지역 아리랑의 홍보, 관광 활용, 문화콘텐츠 개발 및 전승 방향을 모색하는 데 기초자료를 제공한다.

주제어 : 밀양아리랑, 정선아리랑, 무형문화유산, 문화콘텐츠, 토픽모델링

www.kci.go.kr

I. 서론

아리랑은 한국을 대표하는 전통민요이자 무형문화유산으로, 유네스코 등재 이후 보존과 전승뿐만 아니라 세계화와 지역문화콘텐츠 활용의 관점에서도 주목받고 있다(김영운 2022; 박경수 2022; 박소현 2023). 아리랑은 단일한 노래라기보다 지역의 역사, 생활문화, 공동체 기억, 공연예술과 결합하여 다양한 형태로 전승되어 온 문화유산이다. 그 가운데 밀양아리랑과 정선아리랑은 각각 영남권과 강원권을 대표하는 지역 아리랑으로, 한국 아리랑이라는 공통된 문화유산 범주에 속하면서도 지역의 역사, 장소성 및 전승 방식에 따라 서로 다른 문화콘텐츠로 발전해 왔다.

밀양아리랑과 정선아리랑에 관한 선행연구는 음악적 특징, 전승 과정, 교육적 활용, 공연콘텐츠, 지역문화적 가치 및 관광자원화 가능성 등을 중심으로 이루어져 왔다(박경수 2023; 박소현 2025; 신성환·유춘동 2025; 최진호 2025). 이러한 연구들은 두 아리랑의 예술적·학술적 가치와 문화콘텐츠로서의 활용 가능성을 밝히는 데 기여하였다. 그러나 기존 연구는 개별 아리랑의 음악적 특성이나 전승 및 활용 사례를 고찰하는 데 집중해 왔으며, 대중이 일상적으로 접하는 온라인 문서에서 밀양아리랑과 정선아리랑이 어떠한 중심어와 주제적 맥락을 통해 재현되는지를 계량적으로 비교한 연구는 상대적으로 부족하다.

국가와 지역의 역사 및 정체성을 담고 있는 아리랑을 비롯한 문화유산은 디지털 매체에서 지역문화·축제·교육·관광·공연 등의 콘텐츠와 결합하며 새로운 의미를 형성한다. 이경은(2023)은 온·오프라인 담화와 텍스트 분석을 통해 문화원형 담론에 대한 인식 변화를 살필 수 있으며, 디지털 매체가 전통문화의 현대적 활용 가능성을 확장한다고 보았다. 이러한 관점에서 온라인 텍스트는 지역 아리랑이 현대사회에서 어떠한 문화적 의미와 활용 가치를 지니는지를 파악할 수 있는 분석 자료가 될 수 있다.

www.kci.go.kr

이에 본 연구는 기존 연구에서 주로 다루어진 음악적·전승적·교육적 가치 외에, 밀양아리랑과 정선아리랑이 온라인 텍스트에서 어떠한 사회·문화적 활용 맥락으로 재현되는지를 살펴보고, 두 지역 아리랑의 언어적·주제적 구조에 나타나는 공통점과 차이점을 탐색적으로 비교하고자 한다. 이를 위해 네이버 검색 응용 프로그래밍 인터페이스(Search Application Programming Interface, 이하 '네이버 검색 API')를 활용하여 수집한 밀양아리랑과 정선아리랑 관련 온라인 문서를 대상으로 빈도분석, 단어빈도-역문서빈도(term frequency-inverse document frequency, 이하 'TF-IDF') 분석, 엔그램(N-gram) 분석, 네이버 데이터랩(DataLab) 기반 연도별 검색량 분석 및 토픽모델링을 수행하였다.

Ⅱ. 이론적 논의

1. 문화유산으로서 아리랑과 지역성

아리랑은 2012년 유네스코 인류무형문화유산 등재와 2015년 국가무형문화재 지정을 통해 대표적 무형문화유산으로 인정받았으며, 지역의 역사·정서·생활문화·장소성에 따라 다양하게 전승·활용되어 온 지역문화자원이기도 하다(국가유산청 2015). 문화가 한 사회의 생활양식과 의미체계를 포괄하는 유개념이라면, 아리랑은 구비전승 민요와 공연예술의 성격을 지닌 종개념적 무형문화유산으로 이해할 수 있다(Tylor 1871; UNESCO 2003). 문화원형이 보편성과 특수성을 동시에 지닌다는 논의에 비추어 볼 때(김교빈 2005), 지역별 아리랑의 비교는 동일한 무형문화유산이 각 지역의 역사와 문화적 맥락 속에서 어떻게 다르게 전승·활용되는지를 살펴보는 접근이다.

아리랑의 문화유산적 가치는 다음 측면에서 설명된다. 첫째, 아리랑은 한국인의 희로애락, 한, 이별, 염원, 역사적 경험을 담은 민족 정체성과 정서의 상징이다(김영운 2022; UNESCO n.d.). 둘째, 아리랑은 집단 기억, 공동체 정체성, 단결력 형성에 기여해 온 공동체적 유산이

www.kci.go.kr

며, 여러 세대가 함께 부르고 변형하며 전승해 온 노래이다(신성환·유춘동 2025; UNESCO n.d.). 셋째, 아리랑은 시대와 지역에 따라 다양한 형태와 의미로 변형·전승되어 온 개방적 문화유산이다. 이는 아리랑이 단일하고 고정된 문화유산이 아니라, 지역의 생활문화와 역사적 경험에 따라 서로 다른 정서와 서사를 형성해 왔음을 보여준다.

이러한 지역적 차이는 밀양아리랑과 정선아리랑의 대표 사설에서도 확인된다. 밀양아리랑은 “날 좀 보소 날 좀 보소 날 좀 보소/ 동지섣달 꽃 본 듯이 날 좀 보소/ 아리아리랑 쓰리쓰리랑 아라리가 났네/ 아리랑 고개로 날 넘겨주소”로 제시되며(국립국악원 n.d.-a), 정선아리랑은 “아우라지 뱃사공아 배 좀 건너주게/ 싸리골 올동박이 다 떨어진다/ 아리랑 아리랑 아라리요/ 아리랑 고개로 날 넘겨주소”로 나타난다(국립국악원 n.d.-b). 밀양아리랑은 반복적 후렴과 직접적 호소를 통해 그리움과 애원의 정서를 드러내는 반면, 정선아리랑은 구체적 지명과 이별 상황을 통해 자연환경과 생활 정서를 향토적으로 풀어낸다. 이처럼 두 아리랑은 원텍스트 차원에서 정서와 서사 방식이 다르며, 이러한 지역적 특성은 온라인 공간의 언급과 활용 양상에도 서로 다르게 나타날 가능성이 있다.

2. 아리랑의 문화콘텐츠 확장성

밀양아리랑과 정선아리랑에 관한 선행연구는 전승 방식, 음악적 특징, 교육적 활용, 공연콘텐츠, 관광자원화 가능성을 중심으로 이루어져 왔다. 밀양아리랑 연구는 전승 과정, 선율 변용, 음악교육적 활용을 통해 전통민요로서의 음악적 가치와 지역 교육콘텐츠로서의 활용 가능성을 논의하였다(박소현 2025; 서정매 2022; 최진호 2025). 정선아리랑 연구는 지역성, 현장 전승, 문화사적 의미, 공연·교육 활용과 문화브랜드 전략을 중심으로, 강원 지역의 생활문화와 장소성을 담은 문화유산이자 문화관광자원으로서의 가능성을 제시하였다(강등학 2008; 박경수 2023; 신성환·유춘동 2025; 유명희 2022).

www.kci.go.kr

이러한 선행연구에서 확인할 수 있듯이, 아리랑은 시대와 상황에 따라 한탄, 연정, 항일, 위로, 소망 등 다양한 의미로 해석되어 온 개방적 문화유산으로, 교육·공연·전시·관광·축제·디지털 콘텐츠와 결합하며 전통적 가치를 현대적으로 전달하고 대중과의 접점을 확대해 왔다(조용호 2010; 박소현 2023). 이러한 문화콘텐츠 확장 가능성은 박물관 브랜드화, 축제, 가상현실(VR) 전시, 지역 관광·문화 자원화 등의 형태로 논의되어 왔다(구선하 2024; 이승은 2019). 이는 아리랑이 사회적 요구와 매체 환경에 따라 새롭게 해석·활용되는 문화적 실천임을 보여준다.

한편, 문화콘텐츠 활용은 문화유산의 현대적 확장 가능성을 보여주는 동시에, 문화유산을 산업적·경제적 논리로만 이해하여 본연의 가치나 문화 장르의 고유한 의미를 축소할 위험도 지닌다(박주초·안남일 2023). 따라서 아리랑의 문화콘텐츠화 양상을 살펴볼 때에는 오프라인 환경뿐만이 아니라 온라인 환경에서 어떠한 유산 가치와 지역적 의미가 선택적으로 강조되고 가시화되는지를 함께 고려할 필요가 있다. 이러한 관점은 네이버에 나타난 두 지역 아리랑의 주요 속성과 활용 맥락을 분석하는 본 연구의 문제의식과 연결된다.

3. 문화유산의 온라인 텍스트 재현

Hall(1997)에 따르면 재현은 특정 대상이나 현상을 그대로 반영하는 것이 아니라, 언어·이미지·기호·서사 등을 통해 의미를 생산하고 전달하는 사회적·문화적 실천이다. 이러한 관점에서 텍스트의 언어 표현과 주제 맥락은 특정 대상이 무엇으로, 어떻게 말해지는지를 살펴볼 수 있는 자료가 된다. 이러한 재현의 관점은 온라인 텍스트에도 적용될 수 있다.

본 연구에서 문화유산의 온라인 텍스트 재현이란 지역문화유산이 온라인 텍스트 속에서 특정 언어, 키워드, 주제 맥락을 통해 언급·가시화되고 의미화되는 양상을 의미한다(Hall 1997; Herring 2004; Ntafotis et al. 2022).

www.kci.go.kr

아리랑을 포함한 문화유산에 대한 온라인 텍스트 분석은 세 가지 관점에 근거한다. 첫째, 문화유산의 의미가 언어와 기호를 통해 구성된다고 보는 재현의 관점이다(Hall 1997). 박해광(2007)은 문화담론을 권력, 지배, 사회적 조건이 작동하는 실천의 장으로 보았고, Smith(2006)는 문화유산의 가치가 관광, 상업, 지역 홍보 등 대중적 맥락에서 다르게 해석될 수 있음을 논의하였다. 이는 지역문화유산의 의미가 고정된 것이 아니라 사회적 조건과 활용 맥락 속에서 선택적으로 구성되고 강조될 수 있음을 보여준다(Thouki & Skrede 2025).

둘째, 온라인 텍스트를 특정 대상의 대중적 언급 양상을 파악할 수 있는 자료로 보는 온라인 담론 연구의 관점이다(Herring 2004). 이정표(2023)는 소셜미디어에 나타난 대중가요 아리랑 인식을 감성분석으로 분석하여, 아리랑이 디지털 공간에서 대중적 인식과 감성 구조를 파악할 수 있는 대상임을 보여주었다. 또한 Foroughi et al.(2023)은 인스타그램(Instagram)과 트위터(Twitter)와 같은 소셜미디어의 텍스트 자료가 문화유산의 문화적 의미와 대중적 인식을 파악하는 자료로 활용될 수 있음을 제시하였다.

셋째, 문화유산에 대한 텍스트 자료를 분석 가능한 단위로 보는 디지털 문화유산 연구의 관점이다(Ntafotis et al. 2022). 이 관점은 문화유산 관련 텍스트에 포함된 인물, 장소, 시간, 사건 및 주제 요소를 추출·구조화하여 문화유산의 지식 구조와 맥락을 파악하는 데 초점을 둔다. 이경은(2023)은 문화원형 관련 용어와 담론이 온·오프라인 매체를 거치며 대중적 활용 언어로 변화하는 양상을 제시하였으며, 이는 디지털 텍스트가 전통문화의 현대적 활용 맥락을 파악하는 자료가 될 수 있음을 보여준다.

Phillips와 Hardy(2002)는 담론분석을 사회적 언어 분석, 비판적 언어 분석, 해석적 구조주의, 비판적 담론분석으로 구분하여 텍스트의 언어적 특징부터 사회문화적 맥락과 권력관계까지 다양한 분석 수준을 제시하였다. 본 연구는 권력관계나 이데올로기에 대한 비판적 해석보다 온라인 텍스트의 언어적·주제적 구조를 계량적으로 탐색하는 접근을 취한다.

www.kci.go.kr

III. 연구방법

1. 연구 문제

본 연구는 네이버 뉴스·블로그·카페·웹문서를 바탕으로 밀양아리랑과 정선아리랑의 온라인 재현 양상을 비교하는 데 목적이 있다. 이를 위해 두 아리랑의 중심어와 토픽 구조를 분석하고, 키워드·장소·제도·관광 경험·지역 브랜드와의 연결 양상을 파악하였다. 연구 질문은 다음과 같다. 첫째, 네이버 온라인 텍스트에서 두 아리랑에 공통적으로 나타나는 중심어와 지역별 차별적 중심어는 무엇인가? 둘째, 두 아리랑의 연도별 온라인 검색 관심도는 어떻게 변화하는가? 셋째, 두 아리랑 관련 온라인 텍스트는 각각 어떠한 토픽 구조를 형성하는가?

2. 자료 특성

본 연구는 네이버 검색 응용 프로그래밍 인터페이스(API) 문서 자료를 원자료로, 네이버 데이터랩(DataLab) 검색량 자료를 보조 자료로 활용하였다. 네이버 검색 응용 프로그래밍 인터페이스 자료는 뉴스·블로그·웹문서·카페글의 제목과 요약문을 바탕으로 두 지역 아리랑의 온라인 텍스트 양상을 분석하는 자료이며, 네이버 데이터랩 자료는 월별 상대 검색량을 통해 온라인 관심도 변화를 파악하는 자료이다.

두 자료원의 분석 목적·단위·기간·해석 기준과 구체적인 수집 조건은 <표 1>과 같으며, 네이버 검색 응용 프로그래밍 인터페이스와 네이버 데이터랩은 자료의 성격이 서로 달라 구분하여 해석하였다. ‘밀양아리랑’과 ‘정선아리랑’을 각각 단일 검색어로 설정하여 관련도순으로 수집한 결과, 최종 분석 문서는 밀양아리랑 3,196건과 정선아리랑 3,152건 등 총 6,348건으로 구성되었다.

www.kci.go.kr

<표 1> 자료원별 수집·분석 기준 및 기간 정보

자료원 구분	네이버 검색 API	네이버 데이터랩
분석 목적 및 해석 기준	온라인 문서의 언어적·주제적 구조 및 재현 양상	문서 건수가 아닌 검색어 간 상대적 관심도 변화 분석
자료 성격	뉴스·블로그·웹문서·카페글의 제목과 문서 내용 분석	연도별 상대 검색량 추이
검색어	‘밀양아리랑’, ‘정선아리랑’	‘밀양아리랑’, ‘정선아리랑’
검색 방식	단일 키워드 검색	검색어별 상대 검색량 비교
정렬 기준	관련도순(sort=sim)	해당 없음
수집 기간	API 수집 시점 반환 문서 기준	2016년 1월 1일~2026년 5월 25일
수집 실행일	2026년 5월 25일	2026년 5월 25일
날짜 범위	2005년 6월 29일~2026년 5월 25일	해당 없음
총규모 중 날짜확인자료	총 6,348건 중 3,960건	연도별 상대 검색량 지수

한편, 자료별 수집 규모와 날짜 정보 확인 현황은 <표 2>와 같다.

<표 2> 네이버 검색 API 문서 수집 및 날짜 정보 현황

항목	밀양아리랑	정선아리랑	합계
뉴스	1,000*	963	1,963
블로그	1,000*	997	1,997
웹문서	196	192	388
카페글	1,000*	1,000*	2,000
전체 수집 문서	3,196	3,152	6,348
날짜 확인 문서	2,000	1,960	3,960
날짜 미확인	1,196	1,192	2,388

*는 해당 자료 유형의 네이버 검색 API 수집 상한 도달을 의미하며, 날짜 미확인 문서는 일부 블로그·카페글·웹문서에 해당함.

날짜 정보는 뉴스와 블로그에서 대체로 확인되었으나, 웹문서와 카페글에서는 제공되지 않는 경우가 많았다. 이에 전체 6,348건 중 날짜 정보가 확인된 3,960건을 대상으로 연도별 분포를 보조적으로 분석하였다. 그 결과는 <표 3>과 같다.

www.kci.go.kr

<표 3> 네이버 검색 API 날짜 확인 수집 문서의 연도별 분포

연도	밀양아리랑	정선아리랑	합계	우세 검색어
2005	0	4	4	정선아리랑
2006	0	2	2	정선아리랑
2007	0	1	1	정선아리랑
2008	0	3	3	정선아리랑
2009	0	2	2	정선아리랑
2010	0	1	1	정선아리랑
2011	1	1	2	동일
2012	3	1	4	밀양아리랑
2013	1	7	8	정선아리랑
2014	1	2	3	정선아리랑
2015	6	12	18	정선아리랑
2016	4	6	10	정선아리랑
2017	11	15	26	정선아리랑
2018	17	24	41	정선아리랑
2019	16	34	50	정선아리랑
2020	11	28	39	정선아리랑
2021	39	46	85	정선아리랑
2022	51	90	141	정선아리랑
2023	122	87	209	밀양아리랑
2024	167	121	288	밀양아리랑
2025	316	630	946	정선아리랑
2026	1,234	843	2,077	밀양아리랑
총합	2,000	1,960	3,960	밀양아리랑

주: 연도별 문서 수는 날짜 정보가 확인된 검색 API 수집 문서의 분포이며, 각 연도의 전체 온라인 문서량을 의미하지 않음.

날짜 확인 문서는 전체적으로 밀양아리랑이 정선아리랑보다 40건 많았다. 연도별로는 2005~2022년에 대체로 정선아리랑이 우세했으며, 2023~2024년에는 밀양아리랑, 2025년에는 정선아리랑, 2026년에는 다시 밀양아리랑의 문서 수가 많았다.

한편 날짜 정보가 확인된 문서는 2025년과 2026년에 집중되는 경향을 보였다. 이는 해당 연도의 온라인 문서 생산이나 관심이 실제로 급증했음을 직접 의미하기보다, 네이버 검색 API의 관련도순 반환 방식,

www.kci.go.kr

최근 문서 노출 특성, 과거 문서의 검색 접근성 저하, 날짜 정보 확인 가능성의 차이, 검색어별·자료유형별 수집 상한 등이 복합적으로 작용한 결과로 해석할 필요가 있다(네이버개발자센터 n.d.).

3. 분석 절차

본 연구는 네이버 뉴스·블로그·카페글·웹문서에서 '밀양아리랑'과 '정선아리랑'을 검색어로 설정하고, 2026년 5월 네이버 검색 API를 통해 제목, 요약문, 웹주소(URL), 발행일 또는 게시일, 자료 유형, 검색 키워드를 수집하였다. 수집 자료는 쉼표로 구분된 값(comma-separated values, CSV) 형식으로 저장한 뒤 판다스(pandas) 3.0.1을 활용해 분석용 데이터프레임(data frame)으로 구성하였으며, 네이버 데이터랩 검색어 트렌드 API를 통해 검색 관심도 추이를 보조적으로 확인하였다.

분석은 <표 5>와 같이 8단계로 진행하였다. 1단계에서는 원자료를 수집하고, 2단계에서는 중복 웹주소(URL)·결측값·하이퍼텍스트 마크업 언어(HyperText Markup Language, HTML) 태그·특수문자·반복 공백을 정제했으며, 3단계에서는 제목과 요약문을 결합해 분석용 텍스트(text) 열을 생성하였다. 4단계에서는 키위(Kiwi) 형태소 분석기(kiwipiepy)를 활용해 일반명사와 고유명사만 추출하고, 사용자 사전과 정규화 절차를 통해 표기 변이를 통제하였다. '밀양아리랑'과 '정선아리랑'은 고유명사로 등록하여 분절을 방지하고 띄어쓰기 표기를 통일했으며, '밀양시는→밀양시', '정선군은→정선군', '판매합니다→판매', '고장에서→고장'과 같이 조사·어미 결합형을 의미 단위의 명사형으로 정규화하였다.

불용어 사전은 연구 관련성, 문서 유형별 반복성, 낮은 의미 변별력을 기준으로 <표 4>와 같이 7개 범주, 91개 단어로 구성하였다. 명사형 정규화 사전은 표층형 17개와 복합명사 2개 등 19개 항목으로 구성했으며, 키위 사용자 사전에 등록한 연구대상어 2개를 포함하여 전처리

www.kci.go.kr

과정에서 관리한 정규화·사용자 사전 항목은 총 21개이다.

<표 4> 불용어 사전

구분(개수)	대표 불용어 예시
일반 기능어(19개)	것, 수, 등, 및, 관련, 통해, 위해, 대한, 대해, 이번, 지난, 올해, 내년, 경우, 부분, 내용, 대상, 사용
기사 관용어(13개)	뉴스, 기자, 사진, 보도, 밝혔다, 전했다, 말했다, 관계자, 제공, 기사, 언론, 취재, 보도자료
저작권 용어(7개)	무단, 전재, 배포, 금지, 재배포, 저작권, copyright
시간·접속 일반어(17개)	오늘, 어제, 내일, 오전, 오후, 현재, 최근, 기간, 시간, 날짜, 월요일, 화요일, 수요일, 목요일, 금요일
검색·웹 일반어(15개)	목록, 바로가기, 자세히, 보기, 안내, 페이지, 사이트, 클릭, 검색, 서비스, 정보, 공식, 메뉴, 로그인
구어 표현(13개)·유의 단어(7개)	안녕하세요, 같이, 함께, 여러분, 정말, 너무, 조금, 다시, 제가, 생각, 느낌, 추천, 보소, 버트, 회차

전처리 결과, 전체 6,348건의 문서에서 명사 토큰(token) 141,294개와 고유 단어 10,021개가 추출되었다. 지역별로는 밀양아리랑 문서 3,196건에서 토큰 71,341개와 고유 단어 6,026개, 정선아리랑 문서 3,152건에서 토큰 69,953개와 고유 단어 6,553개가 도출되었다.

정제된 명사 토큰을 바탕으로 5단계 빈도분석에서는 주요 중심어를, 6단계 단어빈도-역문서빈도(term frequency-inverse document frequency, 이하 'TF-IDF') 분석에서는 검색어별 차별적 중심어를 도출하였으며, 7단계 이항결합어 분석에서는 연속된 두 명사 토큰의 결합 구조를 확인하였다. 8단계에서는 문서 집합의 잠재 주제와 단어 확률분포를 추정하기 위해 잠재 디리클레 할당(latent Dirichlet allocation, 이하 'LDA') 토픽모델링을 수행하였다(Blei et al. 2003). 사용자 사전 기반 전처리 자료로 젠심(Gensim)의 사전(Dictionary)과 단어주머니 말뭉치(Bag-of-Words Corpus, 이하 'BoW 말뭉치')를 구축한 뒤 LDA 모형(LdaModel)을 적용하였으며, 최적 토픽 수는 일관성 지수(Coherence, c_v)와 로그 퍼플렉서티(Log Perplexity)를 0~1로 정규화하여 평균한

www.kci.go.kr

종합 점수를 기준으로 선정하였다.

<표 5> 분석 절차

단계	분석 절차	활용 도구
1단계	네이버 검색 API 기반 자료 및 네이버 데이터랩 자료 수집	네이버 검색 API, Python 3.12.13, pandas 3.0.1
2단계	중복 URL 제거, 결측값 제거, HTML 태그 및 특수문자 정제	pandas 3.0.1, re(정규표현식), BeautifulSoup 4.15.0
3단계	제목과 요약문을 결합하여 분석용 텍스트 열 생성	pandas 3.0.1
4단계	형태소 분석 기반 명사 추출, 명사형 정규화, 불용어 제거	kiwipiepy 0.23.2, re, pandas 3.0.1
5단계	밀양아리랑과 정선아리랑의 빈도 분석	collections.Counter(빈도 계산), pandas 3.0.1
6단계	TF-IDF 분석을 통한 지역별 차별적 중심어 도출	scikit-learn 1.9.0, pandas 3.0.1
7단계	이항결합어 분석을 통한 반복 결합 표현 도출	collections.Counter, pandas 3.0.1
8단계	지역별 LDA 토픽모델링 및 최적 토픽 수 검토	gensim 4.4.0, scikit-learn 1.9.0, pandas 3.0.1

Ⅳ. 분석 결과: 지역 아리랑의 온라인 재현 양상 비교

1. 중심어 비교: 밀양의 축제·관광시설과 정선의 시장·문화기관

전처리한 문서별 명사 단어를 대상으로 빈도분석과 단어 빈도-역문서 빈도(TF-IDF) 분석을 수행하였다. <표 6>처럼 두 지역의 빈도어와 차별적 중심어는 차이가 있었다. 밀양아리랑에서는 축제·캠핑·경남·천문대·영남루·예약·밀양강이, 정선아리랑에서는 시장·열차·강원도·정선군·여

www.kci.go.kr

행·공연·맛집·강원이 두드러졌다. 즉, 밀양아리랑이 축제·체험시설·캠핑·지역 여가자원과, 정선아리랑이 전통시장·관광교통·공연·지역기관·먹거리와 주로 연결됨을 보여준다.

<표 6> 빈도 및 TF-IDF 분석결과 상위어

밀양 빈도어	빈도	밀양 TF-IDF	TF-IDF 값	정선 빈도어	빈도	정선 TF-IDF	TF-IDF 값
밀양아리랑	6,758	밀양아리랑	0.783	정선아리랑	6,654	정선아리랑	0.760
밀양	3,201	밀양	0.372	정선	3,847	정선	0.439
축제	1,993	축제	0.231	시장	2,080	시장	0.237
캠핑	1,400	캠핑	0.162	아리랑	1,730	아리랑	0.197
아리랑	1,209	아리랑	0.140	열차	1,229	열차	0.140
문화	866	경남	0.125	문화	968	문화	0.110
밀양시	863	문화	0.100	강원도	938	강원도	0.107
시장	795	밀양시	0.100	축제	839	축제	0.095
경남	767	천문대	0.095	여행	824	정선군	0.094
천문대	583	시장	0.092	공연	749	여행	0.094
예약	563	우주	0.079	정선군	589	공연	0.085
대회	536	영남루	0.069	지역	543	지역	0.062
우주	485	예약	0.065	재단	470	재단	0.053
관광	479	대회	0.062	맛집	451	맛집	0.051
공연	459	밀양강	0.060	강원	386	강원	0.044

이항결합어 분석은 키위(Kiwi)로 추출한 명사 토큰의 인접 결합을 기준으로 수행하였다. <표 7>과 같이 밀양아리랑에서는 ‘밀양아리랑 축제’, ‘우주 천문대’, ‘밀양아리랑 시장’, ‘오토 캠핑’이, 정선아리랑에서는 ‘정선아리랑 시장’, ‘정선아리랑 열차’, ‘정선아리랑 문화’, ‘문화 재단’, ‘강원도 정선’이 상위에 나타났다. 이는 밀양아리랑이 축제, 체류형 관광, 관광시설 등과 연계되어 나타나는 반면, 정선아리랑은 전통시장, 관광교통, 문화기관 등과 보다 밀접한 연관성을 보이고 있음을 의미한다. 따라서 밀양아리랑은 축제·관광시설 기반, 정선아리랑은 전통시장·열차관광·전승기관 기반의 지역문화콘텐츠로 의미화되는 것으로

www.kci.go.kr

해석할 수 있다.

이러한 차이는 동일한 무형문화유산도 지역별 관광 인프라, 상권 구조 및 문화자원 활용 방식에 따라 서로 다른 지역 관광·문화자원과 연계되어 온라인에 나타남을 보여준다.

<표 7> 지역별 이항결합어 비교

순위	밀양	빈도	정선	빈도
1	밀양아리랑 축제	1,206	정선아리랑 시장	1,480
2	우주 천문대	457	정선아리랑 열차	907
3	밀양아리랑 우주	423	정선아리랑 문화	470
4	밀양아리랑 시장	417	문화 재단	433
5	오토 캠핑	357	강원도 정선	335
6	밀양아리랑 캠핑	345	정선 여행	328
7	밀양아리랑 오토	335	시장 정선	257
8	문화 관광	330	아리랑 시장	240

2. 검색 관심도의 변화: 정선에서 밀양 우세 전환

본 절에서는 네이버 데이터랩의 월별 상대 검색량 지수를 연도별 평균으로 환산하여 밀양아리랑과 정선아리랑의 온라인 관심도 변화를 비교하였다. 이 지수는 검색 API 문서의 발행일·절대 검색량·문서 수가 아니라 두 검색어 간 상대적 관심도 흐름을 나타낸다. 분석 결과, <표 8>과 같이 2016~2020년에는 2019년을 제외하고 정선아리랑의 관심도가 높았으나, 2021년 이후에는 밀양아리랑이 우세하였고 2024년에는 격차가 크게 나타났다. 다만 2026년 값은 5월 25일까지의 자료이므로 연간 흐름으로 일반화하기 어렵다.

이러한 변화는 축제 운영, 미디어 노출, 관광 홍보 등 시기별 외부 조건과 관련될 수 있다. 두 지역의 가시성 차이는 유산 자체의 가치보다 특정 시기 검색 환경에서 나타난 상대적 주목도의 차이로 해석되며, 지역 아리랑의 온라인상 위상은 관광·행정 활동과 미디어 노출에 따라 달라질 가능성이 있다.

www.kci.go.kr

<표 8> 최근 연도별 데이터랩 평균 검색 관심도 지수

연도	밀양아리랑	정선아리랑	차이	우세 검색어
2016	6.17	6.84	-0.67	정선아리랑
2017	5.52	6.60	-1.08	정선아리랑
2018	5.66	7.86	-2.20	정선아리랑
2019	7.78	7.38	0.40	밀양아리랑
2020	6.09	7.22	-1.13	정선아리랑
2021	6.16	5.91	0.25	밀양아리랑
2022	5.73	5.32	0.40	밀양아리랑
2023	5.19	4.76	0.42	밀양아리랑
2024	13.38	4.50	8.87	밀양아리랑
2025	4.54	3.91	0.63	밀양아리랑
2026	3.77	2.84	0.92	밀양아리랑

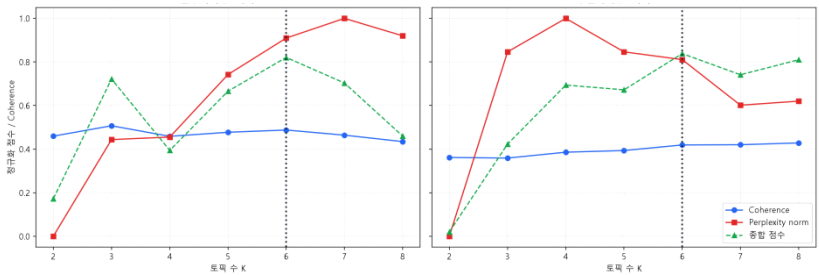
3. 지역별 최적 토픽 수 검토

토픽모델링은 젠심(Gensim)의 LDA 모형(LdaModel)을 기반으로 수행하였다. 키위(Kiwi) 형태소 분석기와 사용자 사전으로 전처리한 명사 토큰을 젠심 사전(Gensim Dictionary)의 고유 식별자(ID)로 변환하고, 문서별 단어 식별자(ID)와 출현 빈도를 이용하여 단어주머니 말뭉치(Bag-of-Words Corpus, 이하 'BoW 말뭉치')를 구축하였다. 사전 필터링은 `no_below=2`, `no_above=0.7`로 설정하였으며, 모수는 `alpha=auto`, `eta=auto`로 설정하였다.

<표 9> 토픽 일관성(Coherence)와 로그 퍼플렉서티(Log Perplexity) 기반 최적 토픽 수

지역	최적 토픽	토픽 일관성 (Coherence, c_v)	로그 퍼플렉서티 (Log Perplexity)	종합 점수
밀양	6	0.487	-6.236	0.820
정선	6	0.419	-6.404	0.838
전체	8	0.445	-6.509	0.936

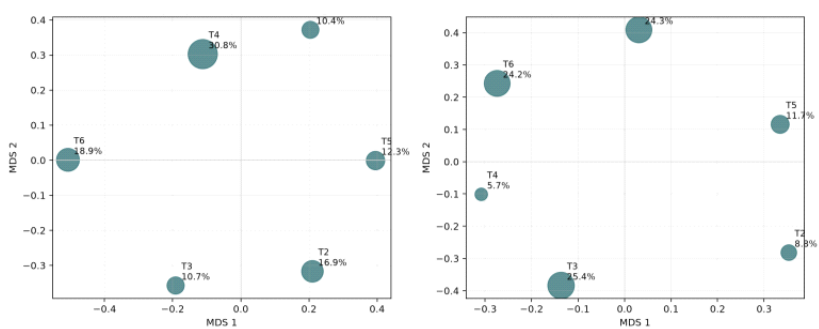
www.kci.go.kr



밀양아리랑 정선아리랑

<그림 1> 지역별 최적 토픽 수 비교

토픽 수(K)는 2~8 범위에서 passes=4, random_state=42 조건으로 검토하고, 0~1로 정규화한 토픽 일관성(Coherence, c_v)과 로그 퍼플렉서티(Log Perplexity)의 평균 종합 점수를 기준으로 결정하였다. 두 아리랑 모두 K=6에서 최고점을 보여 각각 6개 토픽을 선정하였다(<표 9>, <그림 1> 참조).



밀양아리랑 정선아리랑

<그림 2> 토픽간 다차원척도법(MDS) 좌표와 토픽별 비율

토픽 간 관계는 토픽-단어 확률분포의 젠슨-섀넌 거리(Jensen-Shannon distance)를 산출한 뒤 다차원척도법(multidimensional scaling)으로 2차원 시각화하였다(<그림 2> 참조).

www.kci.go.kr

4. 토픽 비교: 체험관광형 밀양아리랑과 생활문화형 정선아리랑

잠재 디리클레 할당(LDA) 토픽모델링 결과는 <표 10>과 같으며, 토픽명은 관련 키워드의 공통된 주제적 특성을 기준으로 정하였다. 밀양아리랑은 (i) 문화예술·지역기관, (ii) 공원·가족관광, (iii) 우주천문대·체험시설, (iv) 축제·문화관광, (v) 마라톤·지역행사, (vi) 캠핑·체류형 관광으로 구분되었으며, 전반적으로 축제·공연·체험시설·캠핑·마라톤과 결합한 장소 기반 문화콘텐츠와 체험관광 중심의 구조를 보였다.

토픽 1의 문화재단, 밀양시, 아트센터, 사업 등의 키워드는 밀양아리랑 관련 문화예술 활동이 지방자치단체와 지역문화기관의 사업 운영 맥락에서 언급되고 있음을 보여준다. 실제로 밀양시는 지역대표 예술단체 지원사업을 통해 밀양아리랑예술단의 운영을 지원받고 있으며(문화체육관광부 2024), 이는 관련 활동이 공공부문의 재정적·제도적 지원과 연계되어 있음을 뒷받침한다.

정선아리랑은 (i) 전통시장·지역여행, (ii) 지역상품·먹거리, (iii) 공연콘텐츠·전승기관, (iv) 관광열차·지역소비, (v) 관광교통·지역연계, (vi) 축제·향토음식으로 구분되었다. 토픽 4와 토픽 5는 모두 열차를 중심어로 공유하지만, 전자는 이용 정보와 소비 인프라에, 후자는 관광교통과 운행 이슈에 초점을 둔다는 차이가 있다.

아리랑은 국가무형문화유산이지만, 네이버 포털에서는 전통예술 자체보다 장소·체험·행사·소비와 관련된 정보가 상대적으로 두드러졌다. 밀양아리랑은 축제·공연·체험시설과 결합한 체험관광형 주제 구조를 보인 반면, 정선아리랑은 시장·교통·먹거리·지역상품과 연결된 생활문화형 주제 구조로 나타났다. 이러한 차이는 지역의 관광 기반, 상권 구조, 문화자원 활용 방식, 지원 체계 및 정책적 선택에 따라 지역 아리랑의 특정 측면이 선택적으로 부각될 수 있음을 보여준다.

www.kci.go.kr

<표 10> 밀양아리랑 및 정선아리랑 토픽별 중심어

밀양아리랑 토픽 중심어	정선아리랑 토픽 중심어
(i) 문화예술·지역기관 문화, 재단, 관광, 공연, 예술단, 밀양시, 아트센터, 지역, 사업	(i) 전통시장·지역여행 시장, 정선, 맛집, 강원도, 여행, 오일장, 장날, 국수, 곤드레, 축제
(ii) 공원·가족관광 공원, 대공원, 다이빙, 여행, 아트센터, 시장, 교육, 코스, 놀이터, 맛집	(ii) 지역상품·먹거리 옥수수, 열차, 정선, 고장, 농사, 기차, 토종, 미백, 먹거리
(iii) 우주천문대·체험시설 천문대, 우주, 밀양, 체험, 운영, 예약, 아이, 시설	(iii) 공연콘텐츠·전승기관 문화, 공연, 재단, 뮤지컬, 박물관, 센터, 축제, 아리아라리,
(iv) 축제·문화관광 축제, 아리랑, 문화, 영남루, 밀양강, 공연, 관광, 대축제, 체험, 행사	(iv) 관광열차·지역소비 열차, 가맹점, 운행, 예매, 상품권, 시간표, 관광, 요금, 노선
(v) 마라톤·지역행사 대회, 마라톤, 시장, 밀양시, 하프, 개최, 전통, 참가, 국제, 현장, 방문	(v) 관광교통·지역연계 열차, 운행, 재개, 상품권, 기차, 코레일, 여행, 아우라지, 관광, 제천
(vi) 캠핑·체류형 관광 캠핑, 예약, 오토캠핑장, 자리, 카라반, 체험, 방문, 관광, 축제	(vi) 축제·향토음식 축제, 시장, 여행, 정선군, 전병, 메밀, 아우라지, 지역, 전통, 강원

Hall(1997)의 재현 개념에 따르면, 온라인 공간은 문화유산을 중립적으로 전달하는 통로가 아니라 특정 의미를 선택·강조하는 매개로 기능한다. 이에 검색 중심 플랫폼에서는 사설의 정서적 의미나 음악적 전승보다 축제 일정·관광시설·전통시장·교통 등 실용적·상업적 정보가 상대적으로 가시화될 가능성이 크며, 지역 아리랑은 무형문화유산으로 전승되는 동시에 관광상품과 지역문화콘텐츠로 활용되어 문화상품화(commodification)의 맥락과 결합할 수 있다.

www.kci.go.kr

V. 결론 및 함의

본 연구는 네이버 뉴스·블로그·카페글·웹문서를 대상으로 밀양아리랑과 정선아리랑의 중심어, 검색 관심도 및 토픽 구조를 비교하였다. 두 아리랑은 모두 축제·시장·여행·문화와 결합해 지역문화와 관광 경험의 맥락으로 확장되었다. 토픽모델링 결과, 밀양아리랑은 축제·캠핑장·우주천문대 등 체험관광과 장소 기반 문화콘텐츠 중심으로, 정선아리랑은 전통시장·관광열차·지역상품·공연콘텐츠 등 지역소비와 생활문화 중심으로 나타났다. 이는 지역의 관광 기반과 문화자원 활용 방식에 따라 두 아리랑이 서로 다른 온라인 주제 구조로 가시화됨을 보여준다.

연도별 검색 관심도에서도 시기별 변화가 확인되었다. 2016년부터 2020년까지는 2019년을 제외하고 정선아리랑의 관심도가 상대적으로 높았으나, 2021년 이후에는 밀양아리랑이 우세한 경향을 보였다. 이는 밀양아리랑과 정선아리랑이 지역의 관광 인프라, 축제 운영, 문화기관의 지원, 상업적 기반에 따라 온라인 재현 양상의 차이로 나타날 수 있음을 보여준다.

본 연구는 온라인 자료 기반 전통문화콘텐츠 연구의 흐름을 뒷받침하고(이정표 2023; Foroughi et al. 2023), 지역 아리랑 사례를 통해 문화유산이 포털 검색을 통해 가시화되는 온라인 정보 환경에서 관광·상업·지역 홍보와 결합해 나타남을 확인하였다는 데 의의가 있다(Shepherd 2002; Smith 2006). 분석 결과는 문화유산의 선택적 가시화, 지역문화유산의 문화콘텐츠화, 장소 기반 의미화의 세 관점에서 해석된다.

온라인 텍스트에서는 아리랑의 음악성·사설·전승 역사 등 문화적 의미 전체보다 축제·관광·시장·공연이 선택적으로 가시화되었으며, 두 아리랑은 관광시설·열차·공연·지역상품과 결합한 문화콘텐츠 자원으로 나타났다. 특히 밀양아리랑은 축제·공원·캠핑장·문화시설과, 정선아리랑은

www.kci.go.kr

전통시장·열차·먹거리·생활경제와 연결되어 지역의 장소 자원에 따른 서로 다른 활용 맥락을 보였다. 이러한 현대적 활용은 새로운 매체와 장르를 통해 아리랑의 의미를 확산하는 재생 과정이자, 전통적 가치를 현대적으로 전달하고 대중과의 접점을 확대하는 문화적 실천으로 이해할 수 있다(박소현 2023).

한편, 두 아리랑의 온라인 재현에서 음악적·예술적·문화유산적 속성보다 축제·관광·체험·소비와 관련된 실용적 정보가 두드러졌다. 이는 학술 연구와 전승 현장에서 강조되는 예술적·교육적 가치보다 대중적 온라인 공간에서 관광자원 및 지역문화콘텐츠로서의 활용과 검색 친화적 정보가 상대적으로 강조되고, 아리랑이 자원·시설 등의 지역 콘텐츠와 결합한 언어적 자원으로 상품화되고 있음을 시사한다.

따라서 실무자들은 지역 여건과 다양한 이용자 및 이해관계자의 관심을 고려한 아리랑 콘텐츠를 개발하는 동시에, 사설의 의미·음악적 특징·전승 역사·공동체 기억을 포털 게시물, 관광 안내, 공연, 체험 및 교육 자료에 충실히 제시할 필요가 있다. 또한 이러한 가치와 의미를 현대적 콘텐츠에 반영하고, 온·오프라인에서 아리랑의 특성이 자연스럽게 드러나도록 언어와 표현을 구성해야 한다. 이를 통해 실용적 정보와 문화유산적 가치를 균형 있게 제시할 수 있으며, 본 연구 결과는 아리랑의 대중적 확산과 문화유산적 가치의 전승이 조화를 이루는 지역 콘텐츠의 방향을 설정하는 참고자료가 될 수 있다.

본 연구는 네이버 뉴스·블로그·카페·웹문서와 밀양아리랑·정선아리랑만을 분석한 탐색적 텍스트마이닝 연구로, 다른 지역 아리랑과 사회관계망서비스(Social Network Service, SNS) 플랫폼을 포괄하지 못하였다. 또한 자료유형별 차이를 충분히 구분하지 못했으며, 중심어와 토픽 분포에 초점을 두어 심층적 의미구조, 상징체계, 주체 간 권력관계와 댓글·공유·추천을 통한 상호작용 과정을 분석하지 못했다. 향후 연구에서는 내용분석·프레임분석·비판적 담론분석과 플랫폼 비교를 수행하고, 주민·관광객·정책 결정자·예술인 등 집단별 인식조사를 병행할 필요가 있다. 이를 통해 지역 아리랑이 전승·정체성·관광·공연예술·교육·브랜드·

www.kci.go.kr

문화콘텐츠·지역경제의 맥락에서 어떠한 의미와 가치로 구성되는지를 규명하고, 사회적 가치와 제도적 보존·활용을 설명하는 문화담론 연구로 확장할 수 있을 것이다.

www.kci.go.kr

<참고문헌>

강등학. 2008. “정선아라리의 민요생태와 문화적 의미.”『한국민요학』 제23집, 257-288.

구선하. 2024. “아리랑 축제 활성화를 위한 가상현실(VR) 전시 공간의 개발 연구.”『한국디자인리서치』 제9권 2호, 484-496.

국가유산청. 2015. 국가무형유산 아리랑. 국가유산포털. https://www.heritage.go.kr/heri/html/HtmlPage.do?pg=/unesco/CulHeritage/CulHeritage_15.jsp (검색일: 2026/7/7).

국립국악원. n.d.-a. “밀양아리랑.”『국악사전』. https://www.gugak.go.kr/ency/topic/view/750 (검색일: 2026/7/7).

국립국악원. n.d.-b. “정선아리랑.” 『국악사전』. https://www.gugak.go.kr/ency/topic/view/874 (검색일: 2026/7/7).

김교빈. 2005. “문화원형의 개념과 활용.”『인문콘텐츠』 제6집, 7-22.

김영운. 2022. “노래 아리랑의 과거, 현재 그리고 미래.”『한국민요학』 제66집, 7-54.

네이버개발자센터. n.d. “네이버 DataLab 검색어 트렌드 API 문서.” https://developers.naver.com/docs/serviceapi/datalab/search/ (검색일: 2026/6/4).

박경수. 2022. “아리랑 연구의 성과와 과제: 아리랑의 유네스코 등재 후 10년간을 대상으로.”『한국민요학』 제66집, 87-163.

문화체육관광부. 2024. “지역 문화예술 이끌어 갈 대표 예술단체 22개 육성한다.” https://www.mcst.go.kr/site/s_notice/press/pressView.jsp?pSeq=21109 (검색일: 2026/7/7).

박경수. 2023. “정선아리랑 수용 공연콘텐츠의 양상과 특징 고찰.”『구비문학연구』 제68집, 109-155.

박소현. 2023. “세계 속 ‘아리랑’의 가치와 역할.”『한국음악연구』 제73집, 159-181.

박소현. 2025. “밀양아리랑의 음악적 생태계.”『한국음악연구』 제77집, 145-165.

박주초·안남일. 2023. “한국사회의 문화콘텐츠 담론에 대한 비판적 고

www.kci.go.kr

찰.”『아시아문화연구』 제61집, 5-22.
박해광. 2007. “문화 연구와 담론 분석.”『문화와 사회』 제2권 1호, 83-116.
서정매. 2022. “밀양아리랑 선율의 음악적 변이 연구.”『한국민요학』 제66집, 165-206.
신성환·유춘동. 2025. “강원 아리랑 유산의 현황과 과제: 새로 지정된 시도 무형유산 아리랑의 문화사적 의미를 중심으로.”『어문론집』 제104집, 111-141.
유명희. 2022. “정선아리랑 현장의 변화와 전망.”『한국민요학』 제66집, 207-229.
이경은. 2023. “‘문화원형’ 담론과 그 전개 양상에 대한 연구: 2000년대 이후 논의를 중심으로.”『인문콘텐츠』 제69집, 271-274.
이승은. 2019. “지역문화자원을 활용한 공연 콘텐츠 사례와 전망: 정선아리랑을 중심으로.”『고전과 해석』 제28집, 27-57.
이정표. 2023. “소셜미디어에 등장한 대중가요 아리랑에 대한 인식: 감성분석 적용.”『아시아태평양융합연구교류논문지』 제9권 10호, 661-670.
조용호. 2010. “아리랑 연구사.”『한국문학과 예술』 제10집, 47-78.
최진호. 2025. “밀양아리랑의 음악교육적 가치와 활용.”『예술교육연구』 제23권 제1호, 317-334.

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. “Latent Dirichlet allocation.” *Journal of Machine Learning Research* Vol. 3, 993-1022.
Foroughi, M., de Andrade, B., & Pereira Roders, A. 2023. “Capturing public voices: The role of social media in heritage management.” *Habitat International* Vol. 142, 102934.
Hall, S. ed. 1997. *Representation: Cultural representations and signifying practices*. London: Sage.
Herring, S. C. 2004. Computer-mediated discourse analysis: An approach to researching online behavior. In S. A. Barab, R.

www.kci.go.kr

Kling, & J. H. Gray (Eds.), *Designing for virtual communities in the service of learning*. New York: Cambridge University Press, 338-376.

Ntafotis, E., Zidianakis, E., Partarakis, N., & Stephanidis, C. 2022. "Extraction of event-related information from text for the representation of cultural heritage." *Heritage* Vol. 5, No. 4, 3374-3396.

Phillips, N., & Hardy, C. 2002. *Discourse Analysis: Investigating Processes of Social Construction*. Thousand Oaks, CA: Sage.

Smith, L. 2006. *Uses of heritage*. London: Routledge.

Shepherd, R. 2002. "Commodification, culture and tourism." *Tourist Studies* Vol. 2, No. 2, 183-201.

Thouki, A., & Skrede, J. 2025. "Re-framing authorised heritage discourse (AHD) within a realist explanatory framework: Towards a dialectical relationship between discourse and the extra-discursive." *International Journal of Heritage Studies* Vol. 31, No. 4, 407-424.

Tylor, E. B. 1871. Primitive culture: Researches into the development of mythology, philosophy, religion, art, and custom. London: John Murray.

UNESCO. 2003. Convention for the safeguarding of the intangible cultural heritage. Retrieved from https://ich.unesco.o rg/en/con vention (검색일: 2026/7/7).

UNESCO. n.d. Arirang, lyrical folk song in the Republic of Korea. UNESCO Intangible Cultural Heritage. Retrieved from https://ich.unesco.org/en/RL/arirang-lyrical-folk-song-in-the-republic-of-korea-00445 (검색일: 2026/7/7).

www.kci.go.kr

투고일 : 2026년 6월 5일 . 심사일 : 2026년 6월 23일 . 게재확정일 : 2026년 7월 9일

* 이유안은 이화여자대학교를 졸업하고, 부산대학교 관광컨벤션학과에서 석사학위를 취득한 후 동 대학원 박사과정을 수료하였다. 이후 호주 선샤인코스트대학교(University of the Sunshine Coast)에서 관광학 박사학위를 취득하였다. 신라대학교 미래융합학과 겸임교수를 거쳐, 현재 국립경국대학교 관광경영학전공 조교수로 재직하고 있다. 주요 연구 관심 분야는 관광자원, 문화관광콘텐츠, 인공지능 기술을 활용한 관광 연구 등이다.

www.kci.go.kr

<Abstract>

A Comparison of the Online Textual Representations of Miryang Arirang and Jeongseon Arirang : A Text Mining Analysis of Naver Portal Data

Lee, Yuan
(Gyeongkuk National University)

Using Naver Search API and DataLab data, this study compared the keywords, topical contexts, and search interest of Miryang Arirang and Jeongseon Arirang. Frequency, bigram, TF-IDF, LDA, and trend analyses linked Miryang Arirang to tourism facilities and experiences, including festivals, camping, observatories, Yeongnamnu Pavilion, and the Miryang River, and Jeongseon Arirang to traditional markets, tourism transportation, and local institutions, including markets, trains, performances, and food. LDA identified six topics for each: cultural institutions, family tourism, experiential facilities, festivals, local events, and stay-based tourism for Miryang; and markets, local products, performances, tourist trains, regional connections, and local cuisine for Jeongseon. Search interest was higher for Jeongseon from 2016 to 2020 and for Miryang from 2021 onward. The findings show that regional tourism infrastructure, commercial structures, and cultural resource use shape differing online representations of the same intangible cultural heritage, informing its promotion, tourism use, cultural content development, and transmission.

Keywords : Miryang Arirang, Jeongseon Arirang, Intangible Cultural Heritage, Cultural Content, Topic Modeling

www.kci.go.kr