

## 토픽모델링을 활용한 항만안전 위험요인 도출에 관한 연구\*

이정민\*\* · 김율성\*\*\*

### A Study on the Derivation of Port Safety Risk Factors Using by Topic Modeling Lee Jeong-Min · Kim Yul-Seong

#### Abstract

In this study, we tried to find out port safety from various perspectives through news data that can be easily accessed by the general public and domestic academic journal data that reflects the insights of port researchers. Non-negative Matrix Factorization(NMF) based topic modeling was conducted using Python to derive the main topics for each data, and then semantic analysis was conducted for each topic. The news data mainly derived natural and environmental factors among port safety risk factors, and the academic journal data derived security factors, mechanical factors, human factors, environmental factors, and natural factors. Through this, the need for strategies to strengthen the safety of domestic ports, such as strengthening the resilience of port safety, improve safety awareness to broaden the public's view of port safety, and conduct research to develop the port industry environment into a safe and specialized mature port. As a result, this study identified the main factors to be improved and provided basic data to develop into a mature port with a port safety culture.

*Key words: Port Safety, Port Risk Factors, Big Data, Topic Modeling*

▷ 논문접수: 2023. 03. 30.      ▷ 심사완료: 2023. 06. 23.      ▷ 게재확정: 2023. 06. 23.

\* 본 논문은 한국해양수산개발원이 후원한 해양수산 미래 리스크 논문 공모전 수상작임을 밝힙니다.

\*\* 한국해양대학교 해양콘텐츠융복합협동과정 물류시스템전공 박사과정, 제1저자, jmjm3644@g.kmou.ac.kr

\*\*\* 한국해양대학교 물류시스템공학과 부교수, 교신저자, logikys@kmou.ac.kr

## I. 서론

국제적인 의존도가 높은 항만물류산업은 다양한 예기치 못한 위험들로 인한 글로벌 불확실성이 커지고 있다. 2020년 발생한 코로나19의 확산은 항만노동력 공급감소와 항만폐쇄 등을 유발하였고, 전 세계적인 물류대란, 항만 혼잡도 증가, 그리고 전례없는 컨테이너 운임 급상승을 유발하였다(임병호 외, 2021). 또한, 러-우 전쟁으로 인한 국제물류망 불안정화는 국내 항만 및 연관산업에 불확실성을 야기하고 있다. 이러한 국제사회 불확실성의 확산은 세계 경제에 악영향을 미치며 글로벌 공급망의 생태계 변화를 유발하고 있으며 이는 항만물류산업의 정상적인 운영에 위협을 초래할 수 있는 요인으로 대두되고 있다(김성기, 2021: 2).

항만물류산업은 중장비를 이용한 작업이 빈번하고 단순재해사고부터 국가적 손실을 일으키는 대형재해사고까지 심각한 결과를 초래하는 각종 재해사고들이 많아 상당한 위험성을 안고 있다(Yang et al, 2014; 김동진, 2016). 이러한 잠재적인 위험성이 존재하는 항만물류산업은 항만 자체에서 발생하는 내부적 위험요인과 사회경제적으로 발생하는 외부적 위험요인으로 둘러싸여 있다. 항만운영에 위협을 준 사례로는 2020년 컨테이너 선박이 접안 중 안벽 크레인과 충돌한 사고, 2021년 평택항 컨테이너에 깔린 근로자의 사망사고 등이 있다. 또한, 전 세계적인 코로나19의 확산은 글로벌 물류대란을 초래하였으며 이는 항만 혼잡도를 증가시켜 항만의 정상적인 운영에 차질을 주고 항만안전사고 증가로 이어지게 되었다. 이처럼 항만물류산업은 언제든지 일어날 수 있는 대내외적 다양한 위험 요소들에 큰 영향을 받게 되며 항만의 운영을 위협할 수 있는 요인들이 예측하기 어려울 정도로 광범위해지고 다양화되고 있다(김성기, 2021: 2). 항만물류산업에서 발생하는 사고의 심각성은 2021년 제정된 중대재해처벌법의 강화되는 동시에 항만산업이 전 국민의 사회적 관심 업종으로

떠오르게 되는 계기가 되었다. 그러나 아직도 항만물류산업은 안전관리의 사각지대에 놓일 가능성이 큰 산업으로 취급되고 있다(김도연, 2022). 따라서, 항만물류산업의 안전성을 확보하기 위해서는 항만에 위협을 주는 다양한 요인들을 파악할 필요가 있으며 더 나아가 폐쇄적인 특징을 가진 항만물류산업의 안전 문화를 개선하고 발전시킬 필요가 있다.

따라서 본 연구에서는 항만안전과 관련된 온라인상의 텍스트 빅데이터를 활용하여 사회적 관점과 전문가 관점에서 항만안전에 위협하는 요인과 그에 영향을 미치는 원인 및 배경들을 알아보고 이를 통해 항만안전에 대한 다양한 인식과 국내 항만에서 항만 안전을 위해 개선해야 할 사항들을 알아보고자 한다. 더 나아가 국내 항만의 안전 문화 발전에 기여하고자 한다. 본 연구는 항만안전과 관련된 전국언론사 11곳의 뉴스 기사와 국내 학술지 초록을 수집하고 파이썬(Python)을 활용한 NMF모델 기반의 토픽모델링을 통해 항만안전과 관련된 토픽을 도출하였다. 본 연구에 사용된 토픽모델링은 텍스트마이닝 기법의 하나로 온라인상에 축적된 방대하고 다양한 데이터를 활용하여 의미 있는 결과를 도출하는 객관적이고 실효성이 높은 기법이다((Feldman & Dagan, 1995: 117). 또한, 통계적 접근을 통해 개념 간의 연결성과 영향 관계를 파악하고 이를 시각화함으로써 데이터가 보여주는 의미를 도출하는 데 유용하게 사용된다(김선아 외, 2006: 209). 즉, 본 연구에서는 일반대중들의 관점에서 사회적 문제를 다루는 뉴스 기사와 항만 전문가 및 관련 연구자들의 인사이트가 반영된 국내 학술지 초록을 통하여 이전에 수행되지 않은 새로운 연구적 접근을 시도하고 다양한 시각에서 항만안전에 대해 알아보고자 하였으며 국내 항만안전 개선점 도출과 국내 항만의 안전 문화 확산을 위한 기반이 될 수 있는 항만안전 분야 연구에 기여하고자 한다. 궁극적으로 본 연구의 목적은 선행된 항만 안전 위험요인들을 재정리하여 요인별로 분류하고 이를 기준으로, 항만안전과 관련한 방대한 데이터를

통해 도출된 항만안전 위험요인들을 분류하여 새로운 시각의 항만안전 위험요인을 탐색하고자 함에 있다.

## II. 선행연구 고찰

### 1. 항만안전 위험요인 관련 선행연구

본 연구는 온라인 텍스트 빅데이터를 통하여 항만 안전에 위협을 주는 요인을 도출하기 위함으로 항만 안전 위험요인에 대한 기준이 필요하다. 따라서 항만 안전 및 위험성과 관련된 선행연구 조사를 통해 항만 안전 위험요인들을 살펴보고 다양한 항만안전 위험요인들을 크게 6가지 요인으로 재분류하여 항만안전 위험요인 분류기준을 설정하고자 하였다.

박용욱·목진용(2001)은 항만하역작업은 중량화물, 장척화물, 유해위험화물을 취급하고 다양한 작업 방법들이 존재하는 특성이 있어서 다른 산업에 비하여 하역근로자의 안전을 확보하기 매우 힘들다고 밝혔다. 이에 항만하역작업체계 및 재해사례 분석, 국내외 법제도 비교분석 등을 통해 항만하역작업에서 발생하는 문제점 및 취약요인, 위험요인을 파악하여 그에 대한 개선방안을 제시하였다. 문헌조사 분석 및 작업 현장 실사를 통해 위험요인 발굴, 관련 기관 전문가 면담 및 자문을 통해 구체적 개선방안을 제시하였다. 이 연구에서는 1. 취급화물 2. 하역장비차량 3. 고소작업, 4. 해상·접수지역, 5. 열악한 작업조건, 6. 기타작업에 포함되는 유해위험요인들을 제시하였다. Mokhtari et al(2011)은 국제무역 의존도가 높은 해상무역은 국제적으로 필수적이며 물류 인프라에 영향을 미치는 위험요인 파악 필요성이 있다고 밝히며 항만과 연안항의 운영 및 관리에 대한 위험관리 프레임워크의 위험성 평가 단계를 도출하기 위해서 Bow-Tie기법과 원인결과 분석기법을 적용하여 분석하였다. 이 연구에서는 FTA와 ETA 분석기법을 이용하여 관련된 위험요인을 도출하였으며 이는 안

전위험요인, 보안위험요인, 오염위험요인, 법적위험요인, 인적위험요인, 기술위험요인으로 구분되며 그에 대한 하위요인도 제시하였다. 윤동하(2014)는 컨테이너 항만에서 발생하는 사고는 발생빈도는 낮으나 심각한 결과를 초래하는 중대재해가 많으므로 철도 및 건설 현장의 사고해석에 주로 사용되는 리스크 평가분석을 적용할 수 있다고 밝히며 실제 사고 사례를 바탕으로 사고 현황과 사고유형을 분석하고 이에 대한 사고 발생빈도와 심각도를 고려하여 리스크 산정 및 사고별 상대적 위험성 평가를 진행하였다. 이 연구에서는 터미널운영사가 제시하는 항만작업영역별 사고유형 분류기준을 적용하였고 이는 1. 안벽장비영역, 2. 야드장비영역, 3. 이송장비영역 3가지로 구분하였다. Yang et al(2014) 컨테이너터미널의 안전 작업을 화물인도의 성공을 결정한다고 밝히며 컨테이너터미널의 복잡한 안전 취급작업에 영향을 미치는 위험요인을 밝히고 안전 평가 및 안전 교육을 위한 프레임워크를 개발하는 연구를 진행하였다. 이 연구에서는 설문조사와 전문가 인터뷰를 통해 5가지 측면에서 17가지의 위험요인을 제시하였다. 이때 5가지 위험 측면은 1. 라이브웨어 위험요인, 2. 소프트웨어 위험요인, 3. 하드웨어 위험요인, 4. 환경 위험요인, 5. 조직 위험요인으로 구분하였다. Loh et al(2017)은 항만운영자와 항만이용자의 관점에서 공급사슬에 악영향을 미칠 수 있는 항만 위험요인을 식별하고 이에 대한 위험성 평가를 제공하여 공급사슬망의 붕괴에 미치는 위협의 중요성 및 개선사항을 밝히고자 하였다. 이 연구에서는 위험식별 선행연구 및 사례를 통해 총 19개의 항만공급망 붕괴(PSCD: Port-Centric Supply Chain Disruptions) 위험요인을 식별하였다. 여기에는 시설위험(infrastructure threat), 계획위험(planning threat), 인력위협(manpower threat), 보안위협(security threat)으로 총 4가지의 범주로 구성되었다. Pallis(2017)은 해상운송의 중요성을 밝히며 항만시스템은 위험지향적인 성격을 가지므로 전반적인 위협에 대처 가능한

위험관리 방법 및 프레임워크를 제시하고자 하였다. 이 연구에서는 5가지 단계별로 그리스 컨테이너항만의 위험평가를 진행하였고 이때 위험분류로 1. 사람(선박충돌, 좌초, 침몰, 항해장애 등) 2. 기계(크레인 추락, 컨테이너 추락, 장비결합, 화재/폭발 등) 3. 환경(기름유출, 화학오염물질, 평형수 등), 4. 보안(전쟁/정치적 불안정, 테러, 밀수 등) 5. 자연적요인(지진, 화산폭발, 허리케인, 홍수 등)을 제시하였다. Wan(2018)은 해상컨테이너 공급망(MCSC: Maritime Container Supply Chains)는 내외부적으로 다양한 위험에 노출되어 있으며, 글로벌 물류시스템이 복잡해지면서 MCSC의 위험관리에 큰 어려움을 겪고 있다고 밝혔다. 따라서 다양한 MCSC 위험요인을 탐색 및 분석하여 적절한 위험성평가를 개발하고 MCSC의

전반적인 성능을 평가하고자 하였다. 이 연구에서는 MCSC의 위험요인을 내외부적으로 구분하였고 그 중 외부적 위험에는 1. 사회적 측면, 2. 자연·환경적 측면 내부적 위험에는 3. 경영적 측면, 4. 인프라 및 기술 측면, 5. 운영 측면으로 구분하였다. 박성훈 등(2019)은 부산항 컨테이너터미널의 항만하역이송 등 항만물류작업에서 일어나는 모든 위험요인을 파악하고 이에 대한 개선방안을 제시하였다. 이 연구에서는 고용노동부(2017)에서 제시하는 위험성평가 가이드라인의 13개의 위험성 평가기준을 바탕으로 실제 컨테이너터미널 사례를 통해 총 18개의 위험요인을 도출하였다. 이는 1. LNG&LPG 2. 압축가스, 3. 고소작업, 4. 오버헤드 장비, 5. 선박 간 충돌, 6. 장력상태(와이어로프 등) 7. 지상 이동(Apron, Yard),

표 1. 도출된 항만안전 위험요인 분류의 하위요소 및 주요내용

분류	주요 내용	분류	주요 내용
인적 요인	<ul style="list-style-type: none"> <li>• 근로자의 작업 실수 및 오류</li> <li>• 작업 지시 오해</li> <li>• 숙련된 작업자 부족</li> <li>• 근로자의 정신건강 문제</li> <li>• 근로자의 안전의식 부족</li> <li>• 근로자의 불안정한 행동</li> <li>• 항만 내외적인 노조문제</li> <li>• 항만 노동자의 파업</li> <li>• 인력부족</li> </ul>	기계적 요인	<ul style="list-style-type: none"> <li>• 터미널 내 중장비들의 오작동</li> <li>• 기계 고장 및 시스템 고장</li> <li>• 장비의 안전장치 미부착 및 노후화</li> <li>• 장비의 손상</li> <li>• 장비로 인한 화재 및 폭발</li> <li>• 항만정보시스템의 오류 및 고장</li> <li>• 정전</li> <li>• 산업재해</li> </ul>
환경적 요인	<ul style="list-style-type: none"> <li>• 기름유출/화학오염물질/위험화물/항만오염/소음공해/선박평형수 오염</li> <li>• 항만파업/항만혼잡/부정확한 수요예측</li> <li>• 금융위기/금리변동/연료가격변동/불안정한 시장환경</li> <li>• 불안정한 무역정책/국제적 안전규약/ 정치적 갈등</li> <li>• 시설 및 장비의 부족</li> <li>• 항만-배후지 간 운송혼잡</li> </ul>	보안적 요인	<ul style="list-style-type: none"> <li>• 테러공격(물리적·사이버)</li> <li>• 해상도난</li> <li>• 밀수/불법거래/불법입국</li> <li>• 전쟁/정치적 불안정</li> <li>• 사보타지(Sabotage)</li> <li>• 침입 및 도난</li> </ul>
자연적 요인	<ul style="list-style-type: none"> <li>• 변화무쌍한 기상조건/폭우/강한 풍랑/홍수/태풍/자연재해</li> <li>• 지진/화산폭발</li> <li>• 이상기온(고온·저온)/ 안개</li> </ul>	조직적 요인	<ul style="list-style-type: none"> <li>• 정보교류환경의 불안정/정보의 부정확성/IT기술 취약성</li> <li>• 언어·문화의 다양성 수용부족/노사 간 협력부족/열악한 작업환경</li> <li>• 운영안전 프로토콜 정착/안전교육 의무화 및 필수화</li> </ul>

8. 미끄러짐, 추락, 걸림, 9. 이동장비(PM, R/S 등), 10. 위험장비, 11. 날씨, 12. 고온표면, 13. 고전압 및 전원, 14. 산소부족, 15. 엔진오일, 16. 반복적인 작업, 17. 소음, 18. 보안사고가 포함된다. 이처럼 항만안전 및 위험성과 관련된 선행연구 조사를 통해 다양한 연구에서 사용한 항만안전 위험요인을 살펴 보았다. 결과 연구들에서 도출된 위험요인을 선택하거나 설문지법을 통해 기존의 위험요인을 재선정하고 있었고 공공기관에서 제시하고 있는 항만 안전 위험기준, 실제 터미널운영사에서 사용하고 있는 안전 지침상의 기준을 바탕으로 위험요인을 도출하고 있었다. 이러한 선행연구를 도출된 다양한 항만안전 위험요인들을 재정리하여 크게 인적요인, 기계적 요인, 환경적 요인, 보안적 요인, 자연적 요인, 조직적 요인 등 6가지 요인으로 구분하였으며 이에 대한 세부적인 하위내용들도 함께 정리하였다. 이는 표 1과 같으며 이 분류표는 본 연구에서 진행될 항만안전에 관한 토픽모델링 분석을 통해 도출된 토픽들을 분류하는데 사용할 예정이다.

## 2. 빅데이터 관련 선행연구

인터넷의 발전과 스마트기기의 보편화로 온라인상에 축적되는 다양한 형태의 데이터들의 양이 급증하면서 빅데이터를 활용하여 새로운 의미를 추출해내는 연구가 활발히 이루어지고 있다. 본 연구에서는 다양한 연구 분야에서 수행되고 있는 빅데이터를 활용한 연구들을 살펴보았으며 빅데이터 중 본 연구의 분석 방법에 맞는 텍스트 빅데이터를 다른 연구를 중심으로 빅데이터 분석, 텍스트마이닝에 관련한 선행연구를 살펴보았다.

김지은(2017)은 기술의 발전과 다양한 정보의 폭발적 증대로 인해 생성 및 축적되는 빅데이터의 분석 및 활용에 대한 관심이 증가하고 있다고 밝히며 텍스트 빅데이터와 텍스트 마이닝 기법을 활용하여 대학구조개혁 평가 정책에 관한 키워드 및 토픽 분석을 하고자 하였다. 이 연구에서는 “대학구조개혁

평가”라는 주제어에 대한 교육부 문서, 대학전문지, 종합일간지 기사에 나타난 주요 키워드와 토픽을 도출하고자 하였고 도출된 결과에 대한 비교분석을 진행하였다. 연구의 결과는 대학구조개혁 평가에 대한 교육부 문서와 뉴스 기사의 키워드 분석 및 비교를 통해 교육부와 뉴스 기사 간 단어 빈도수의 차이가 있으며 세 가지 데이터 모두에서 공통적인 단어가 분명하게 드러나지 않아 데이터 매체별로 주요 키워드가 다른 것을 알 수 있었다. 다음, 토픽모델링 결과는 데이터별로 10개~36개 정도의 토픽이 도출되었고 이에 대한 위계적 군집분석을 수행하여 토픽의 수를 줄였다. 이를 통해 각 데이터 매체마다 주요 토픽이 다른 것을 확인할 수 있었다, 마지막으로 도출된 토픽의 시계열에 따른 변화를 살펴보고 관심의 변화 및 경향 차이가 발생함을 알 수 있었다. 추가로 이 연구자는 이전에는 수행되지 않았던 주요 교육정책에 관한 연구에서 빅데이터를 활용하여 주요 키워드 및 토픽 분석을 새롭게 진행하였다는 점에 의의를 두었다. 박상정(2018)은 방대한 양의 빅데이터 분석은 새로운 가치 창출을 이루고 사회의 주요한 문제들을 해결하기 위한 중요한 통계적 근거자료가 된다고 밝혔다. 이 연구자는 빅데이터 비전공자도 쉽게 접근할 수 있는 파이썬을 활용하였으며 공공데이터 및 소셜 데이터를 통한 웹크롤링 및 분석, 시각화하는 방법과 예증을 하고자 하였다. 이 연구에서는 파이썬이라는 분석 도구의 플랫폼 환경을 설명하고 파이썬을 통한 한국어 형태소분석에 관한 내용, 자연어 처리를 위한 KoNPLy 등에 대한 설명을 담고 있다. 공공데이터 및 온라인 데이터를 얻을 수 있는 파이썬의 코드를 제시하고 있다. 결과적으로 이렇게 수집된 빅데이터에 대한 분석과 시각화에 대해 설명하고 있다. 신안나(2019)는 국내의 교육 전반에 영향을 미치고 있는 자유학기제에 대해 정책 시기별로 다양한 매체에서 나타난 키워드 및 토픽을 분석하여 해당 정책의 정밀한 진단 및 개선 방향을 도출하고자 하였다. 언론 기사, 교육부 보도자료, 학술지 초록 등

을 활용하였으며 주제어인 자유학기제에 대한 키워드 및 토픽 분석을 진행하였다. 이 연구자는 세 가지 데이터별 주요 키워드와 시기별 변화양상, 그리고 데이터별 토픽 추출과 시기별 변화양상을 알아보고자 하였다. 연구의 방법은 언론 기사, 교육부 보도자료, 학술지 초록 등의 5년간 누적 데이터를 수집하였고 각 문서에 대한 전처리와 형태소분석을 실시하였다. 이후 정책단계별 키워드분석 및 키워드 네트워크 분석, 그리고 구조적 토픽 모형(STM)을 통한 토픽모델링분석을 실시하였다. 데이터별 키워드분석으로 TF-IDF분석 결과, 공통적인 주요 키워드도 있었으나 데이터별로 각각 다른 주요 키워드가 강조되었다. 다음으로 TF-IDF 가중치를 적용한 가중 키워드 네트워크 분석을 통한 데이터별 연결중심성, 근접중심성 등을 알아보았다. 마지막으로 시계열적 요소인 정책단계별로 나타나는 토픽의 변화양상을 알아보았다. 추가적으로 이 연구자는 빅데이터 분석을 활용한 정책의 키워드와 토픽 분석을 통해 해당 정책 진단 및 개선방향을 제안하였다. 진미르(2019)는 수학교육에서 수행된 선행연구들의 분석을 통해 수학교육의 동향을 파악하기 위해 텍스트마이닝을 활용하여 최근의 수학교육 연구에 대한 주요 키워드를 추출하고 이를 시각화하고자 하였다. 이를 위해 3개년도의 수학교육학술지 데이터를 수집하여 분석하였다. 분석과정은 먼저, 디비피아(DBpia) 사이트를 통해 각 학술지에 게재된 논문의 제목, 초록, 서론과 결론의 텍스트를 수집하여 총 460편의 논문을 데이터화한 뒤, 전처리과정을 거쳐 형태소분석을 실시하였다. 다음으로 추출된 명사를 바탕으로 빈도분석 및 TF-IDF분석을 실시하고 워드 클라우드 생성과정을 진행하였다. 연구의 결과로는 빈도수(TF) 분석을 통해 최근 3년간 수학교육 학술지에 게재된 논문에서 가장 많이 나타난 단어들을 추출하였고 TF-IDF 가중치를 적용한 빈도분석을 통해 3년 동안 해당 연구 분야에서 중요하게 고려된 핵심 키워드를 확인하였다. 이원조(2020)는 최근 들어 비정형 텍스트 데이터 분석에 대한 관

심도가 높아지면서 이를 적극적으로 활용하고자 하는 개인이나 기업들이 많아지고 있으나 빅데이터 분석에 있어 텍스트 데이터는 대부분 비정형의 형태이며 대용량이기때문에 분석 결과에 대한 신뢰성이 부족하여 분석에 어려움이 따른다고 하였다. 따라서 연구자는 비정형 텍스트 데이터 분석을 위해 데이터를 수집한 후 수작업 전처리과정을 거치고 R프로그램과 휴리스틱 후처리 정제기법을 사용하여 비정형데이터 정제모형을 제안하였다. 이후 제안된 정제모형을 적용하여 “대통령 UN 연설문” 텍스트 자료를 R프로그램 워드클라우드 기법을 사용한 시각화 분석하고 텍스트 문서상의 핵심이슈를 추출하였다. 이러한 일련의 과정을 통해 워드클라우드 기법을 사용하여 시각화 분석하는 과정에서 발생하는 문제점을 알아보고 이에 대한 대응책 및 해법으로 새로운 데이터 정제모형을 제안하여 워드클라우드 기법이 실무에 적용되기 위한 효율적인 방안을 제시하고 있다. 이 연구의 결과로 워드클라우드를 이용하여 핵심이슈를 도출하는 과정에서 발생하는 문제점에는 먼저 한글사전(KoNLP)에서 전문용어 및 신조어가 반영되지 않는 점, 분석자의 프로그램 사용 미숙, 시각화 결과에 대한 해석 역량 부족 등이 있다고 밝혔다. 또한, 휴리스틱 전처리와 후처리 과정이 상당히 중요하다고 밝혔다. 주홍준(2022)은 빅데이터 분석을 활용하여 건축물의 화재위험도에 영향을 미치는 원인을 검토하고자 하였다. 이후, 위험에 취약한 지표를 점검하고 그에 맞는 비구조적, 구조적 대책을 수립하기 위해 빅데이터 분석기법을 활용하여 건축물 화재위험도 지표를 결정하고자 하였다. 분석과정은 먼저 화재와 관련한 법령 등을 참고하여 검색 질의어를 선정하고 검색엔진을 통해 학술검색사이트에서 비정형데이터로 이루어진 문헌을 웹 크롤링하여 최종적으로 학술지, 연구보고서 총 9,768건을 수집하였다. 최종적으로 연관검색어를 대상으로 위험도 지표 결정기준(연관성, 대표성, 가용성, 지속성, 이해성, 방향성)에 따라 최종적으로 화재위험도 평가를 위한 20개를

도출하였다.

대부분의 선행연구에서 텍스트마이닝 기법을 활용하여 주요 키워드 및 토픽을 도출함으로써 연구주제와 관련된 동향들을 알아보고 이를 통한 시사점들을 도출하고 있었다. 본 연구에서는 광범위하고 다양하며 현재 동향을 반영하는 비정형 텍스트 빅데이터를 통하여 항만안전 및 위험과 관련된 주요 토픽들을 도출하고 이들이 내포하고 있는 의미에 대한 분석을 실시하고자 하였다. 본 연구에서 사용할 텍스트 빅데이터는 일반대중들이 쉽게 접할 수 있는 뉴스 기사 데이터와 항만과 관련된 전문가 및 연구자들의 인사이트가 반영된 국내 학술지 초록 데이터로 설정하였다. 이를 통해 수년간 항만안전과 관련한 뉴스 및 연구에서 중요하게 생각하고 있는 항만안전 위험요인을 도출하고 이에 영향을 끼친 배경 및 원인 등을 알아보고자 하였으며 더 나아가 이러한 위험요인에 대한 안전성 및 안전의식을 높이기 위한 개선방안을 제시하고자 하였다.

### III. 연구방법 및 연구설계

#### 1. 연구방법

##### 1) 텍스트 마이닝

텍스트 마이닝이란 비정형·반정형 데이터에 대하여 자연어처리 기술 및 문서처리 기술을 적용하여 의미있는 정보를 추출하고 가공하는 기술을 의미하며 구조화된 데이터베이스를 대상으로 하는 데이터 마이닝과는 다르게 구조화 되지 않은 광범위한 텍스트로부터 새로운 의미있는 지식을 발견하는 과정을 말한다(김대곤 외, 2013: 74; 이영재, 2011: 64). 여기서 마이닝이란 데이터로부터 통계적으로 의미 있는 개념 및 특성을 추출하고 이들 간의 관계성이나 패턴, 추세 등의 정보를 끌어내는 과정을 의미한다(이원조, 2022: 716). 이러한 텍스트마이닝에서 사용하는 대표적인 비정형데이터인 텍스트 데이터에는

책, 문서, 뉴스 기사, 이메일, 블로그, SNS, 연설문 등이 포함된다. 텍스트마이닝 기법은 기존의 연구들이 연구자의 역량에 의존해 분류·분석되어 대량의 텍스트를 분석하기에 많은 시간과 소요되었고 급속도로 변화하는 정보들을 탐색하기에 용이하지 않았기에 이전에는 도출할 수 없었던 숨은 의미 있는 지식을 찾아내는 것에서 연구 방법으로서의 핵심적인 가치를 갖는다고 볼 수 있다(김지은, 2017: 33; 이인선·나은영, 2018: 417). 이러한 텍스트 마이닝 기법의 특징으로는 첫째, 인지능력을 사용하지 않아 객관적인 분석이 가능하며 둘째, 분석 결과를 시각화함으로써 연구의 관점을 쉽고 빠르게 전달할 수 있다. 셋째, 도출된 결과를 연구자의 관점에 따라 다각도로 조명해볼 수 있다(이인선·나은영, 2018: 417). 이처럼 텍스트 마이닝 기법을 이용하여 다양한 연구들에 사용된 복잡하고 다양한 언어와 개념들을 비정형데이터로 처리하고 통계적으로 분석을 함으로써 기존의 연구 동향에 대한 의미 있는 정보를 새롭게 조직할 수 있다(진미르·고호경, 2019: 495). 즉, 텍스트 마이닝 기법은 비구조화된 텍스트로부터 흥미있고 유용한 패턴 및 체계를 도출하는 과정으로 볼 수 있으며 이 기법은 정보추출, 정보검색, 자연어 처리, 텍스트 요약, 자동분류 등에서 사용되는 기법들과 결합하여 활용된다. 그러나 텍스트마이닝을 활용할 시 잘못된 해석을 하지 않기 위해 주의해야 할 사항들이 존재한다. Grimmer & Stewart(2013)에 따르면, 텍스트 데이터의 정확한 생성원인을 알 수 없으므로 텍스트 분석의 정량적 모델 구축 시 해석에 신중을 기하여야 하고 텍스트에 대한 연구자의 해석이 필요하며 수치계산은 연구자의 판단결정 정도에만 활용되어야 한다고 밝혔다. 또한, 텍스트 마이닝은 항상 분석목적과 일치하게 이루어져야 하고 텍스트마이닝의 유효성에 대한 확인이 필요하며 이를 수행하기 위한 다양한 방법을 활용해야 한다고 밝혔다.

## 2) 토픽모델링

토픽모델링은 비구조화된 방대한 텍스트 데이터에서 특정 주제(Topic)를 찾아내기 위한 기술로, 맥락과 관련된 단서들을 이용하여 유사한 의미를 가진 단어와 연관성이 있는 단어들을 클러스터링하는 방식으로 주제어를 추론하는 모델이다(Steyvers and Griffiths, 2007; Blei, 2012; 강범일 외, 2013). 즉, 토픽모델링은 비구조화된 대량의 텍스트로부터 숨겨져 있던 주제 패턴을 발견하기 위한 통계적 알고리즘이며 이러한 통계적 방법을 이용하여 코퍼스로부터 특정한 토픽추출, 시계열 흐름에 따른 토픽변화추이 탐색, 토픽간 연관성 분석, 문서별로 할당된 토픽 확인 등을 위한 목적으로 수행된다(Blei, 2012). 즉, 이러한 토픽모델링은 문서별 토픽 확률(문서-토픽 확률)과 토픽별 단어 확률(단어-토픽 확률)을 추출하여 주어진 문서가 어떠한 토픽으로 이루어졌고, 각 토픽이 어떤 단어들로 구성되었는지를 파악하기 위한 분석 도구를 의미한다(신안나, 2018: 36). 토픽모델링 기법은 크게 확률 기반의 기법과 행렬 분해 기반의 기법이 있다. 확률 기반의 기법은 문서마다 주제가 발견될 확률과 주제별로 단어가 생성될 확률을 계산하여 각각의 가중치를 계산하며, 대표적으로 T. Hofmann에 의해 제안된 확률적 토픽모델링 기법인 pLSI(Probabilistic Latent Semantic Indexing, 확률 잠재 의미 색인)과 D. M. Blei 등에 의해 제안된 LDA(Latent Dirichlet Allocation, 잠재 디리클레 할당)가 있다(김남규 외, 2017: 477). 행렬 분해 기반의 경우는 각 문서 벡터와 문서에서 언급된 단어들의 단어 벡터가 같은 공간에 위치하도록 행렬을 분해하며 차원을 축소하고 이를 통해 각각의 차원의 크기가 주제의 가중치가 되게 한다. 행렬 분해 기반은 대표적으로 D. D. Lee와 H. S. Seung에 의해 제안된 NMF(Non-negative Matrix Factorization, 음수 미포함 행렬 분해)가 있고 이는 계량분석화학 분야에서 연구되어 온 자기 모델링 그래프 분해(Self Modeling Curve Resolution) 알고리즘에서 발전된 개념이다(김

남규 외, 2017: 476). NMF 기법은 내부 과정에서 뺄셈을 허용하지 않는 특성이 있어 하나의 비음수 데이터를 두 개의 행렬로 분해하는 역할을 수행하여 단어들에 대한 양의 가중치만의 집합으로 토픽을 정의함으로써 무제한 선형 분석에 기반한 LSA(latent semantic analysis)에 비해 보다 해석이 가능한 토픽들을 추출할 수 있다는 장점이 있다(장정호 외, 2006: 324; 김남규 외, 2017: 476). 또한, NMF기법은 비음수 행렬이 두 개의 비음수 최소행렬로 분해되면서 대량의 정보를 효율적으로 처리할 수 있으므로 토픽 추출에 유용성이 높고 추출된 토픽을 기반으로 한 문서군집화 성능이 일반적인 군집화 모델에 비해 우수할 수 있다는 연구 결과도 제시되었다.(Xu et al, 2003; 268; 장정호 외, 2006: 324). 이 기법은 내부 과정에서 뺄셈을 허용하지 않는 특징이 있고 하나의 비음수 데이터를 두 개의 행렬로 분해하는 역할을 수행한다. 예를 들어, 원 데이터가  $m \times n$ 차원의 행렬로 구성되어 있을 때, NMF는 이를  $m \times r$ 차원과  $r \times n$ 차원의 행렬로 분해한다. 이때  $m \times r$ 차원의 행렬은 원 데이터에 비해 낮은 차원을 가지지만, 기존 데이터의 특징은 그대로 반영한다(김남규 외, 2017: 476). 추가적으로, 토픽모델링에서 토픽 개수의 결정은 분석 결과의 해석가능성에 영향을 미치기 때문에 중요하게 여겨진다. 따라서 토픽모델링 실시 전, 복잡도(Perplexity)와 같은 통계적 수치의 변화에 근거하여 제한된 토픽의 수를 지정해야 한다(김지은, 2017: 67).

## 2. 연구설계

### 1) 데이터 수집

본 연구에서는 항만안전과 관련된 다양한 시각의 데이터를 얻기 위하여 뉴스 기사와 국내 학술지 초록을 분석자료로 사용하였다. 데이터수집 기간은 2009년부터 2021년까지 총 13년으로 설정하였고 데이터 수집을 위한 검색어는 “항만+안전”, “항만+위



험”, “항만+사고” 3가지로 구성하였다. 온라인 뉴스 기사 수집은 한국언론진흥재단의 빅카인즈를 통해 언론사 11곳의 뉴스 기사를 수집하였다. 최종적으로 수집된 뉴스 기사 데이터 건수는 총 3,339건이다. 국내 학술지 초록 데이터의 수집은 한국교육학술정보원의 학술연구정보서비스(RISS) 홈페이지를 통해 수집하였다. 수집된 국내 학술지 초록데이터는 총 1,179편이었으며 이후 추가적인 필터링 과정을 거쳐 최종적으로 587건의 국내학술지의 제목, 초록, 발행연도가 수집되었다. 데이터 수집 현황은 표 2와 같다.

표 2. 분석 대상 학술지 초록데이터 수집현황

수집된 데이터	건수	수집매체
최종 분석 뉴스 기사	3,339	빅카인즈
최종 분석 학술지 초록	587	Riss

2) 데이터 전처리

빅데이터 분석에서 데이터 전처리과정은 분석의 결과에 직접적인 영향을 미치는 과정이기 때문에 모든 데이터 분석 프로젝트에서 반드시 수행되어야 하는 작업이다. 이러한 데이터 전처리과정에서 발생하는 오류는 데이터 분석의 신뢰성에 부정적인 영향을 미칠 수 있고 따라서 데이터 전처리는 분석 결과에 따라 반복적으로 수행될 수 있다(예문사, 2021). 텍스트 마이닝 분석을 수행하기 위한 전처리과정으로는 숫자나 문장부호 제거, 누락/제거된 표현 복원 및 제거, 특수문자나 밑줄제거, 오타자 교정, 불용어 제거 등의 작업이 수행된다(길호현, 2016). 본 연구에서는 수집된 텍스트 데이터를 키워드분석 및 토픽모델링 분석에 활용하기 위해 아래와 같은 데이터 전처리과정과 형태소분석을 실시하였다. 본 연구의 데이터 전처리는 아래와 같은 과정을 거쳤으며 일련의 과정은 그림 1과 같다.

그림 1. 데이터 전처리 과정



3) 최종 연구모형

본 연구에서는 항만안전과 관련된 뉴스 기사와 학술지의 주요 토픽을 도출하기 위하여 파이썬을 활용한 NMF 모델기반의 토픽모델링을 실시하였다. 토픽모델링을 통해 데이터별 주요 토픽을 도출한 후 토픽별 주요 키워드를 통해 추정 토픽명을 설정하였으며 이후, 각 토픽에 대한 의미분석을 실시하였다. 마지막으로, 항만안전 위험요인 기준에 따라 항만안전과 직접적인 관련이 있는 토픽들을 재선정하여 최종 토픽을 도출하였다. 이러한 과정을 통해 빅데이터를 활용한 항만안전 위험요인 도출에 대한 종합 분석 결과 및 시사점을 기술하였다.

IV. 연구결과

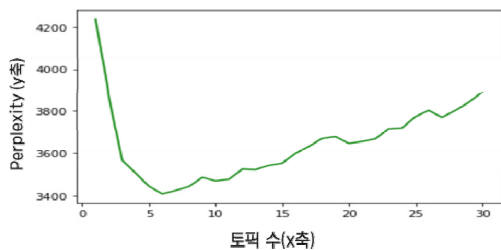
1. 토픽모델링 결과

1) 뉴스 기사 데이터

본 연구에서는 토픽의 개수를 정하기 위해 파이썬의 perplexity(복잡도)를 사용하여 적절한 토픽의 수를 결정하였다. perplexity는 언어모델이 얼마나 정확하게 문장을 생성하는지를 판단할 수 있는 성능지표이다. 이 방법은 선정된 토픽의 개수마다 학습을 시켜 가장 낮은 값을 보이는 구간을 찾아 최적화된 토픽의 개수를 선정하는 것으로 확률 모델이 결과를 얼마나 정확하게 예측하는지를 판단하고 그 수치가 낮을수록 정확하게 예측된 것으로 판단한다

(Griffiths, 2004: 5230). 본 연구에서는 토픽의 수를 1에서부터 30까지로 변화시키면서 perplexity를 구한 후 perplexity의 값이 가장 낮은 구간의 토픽 개수를 결정하였다. perplexity값을 그래프로 나타낸 결과는 그림 2와 같다. 토픽의 수가 6개일 때 perplexity는 3,409.031로 가장 낮게 나타났고 따라서 수집된 뉴스 기사에 대한 토픽모델링의 토픽 수는 6개로 결정하였다.

그림 2. 뉴스 기사 데이터의 perplexity 그래프



다음으로는 NMF 모델을 이용한 토픽모델링을 진행하기 위하여 파이썬을 활용하였고 전처리과정이 모두 완료된 뉴스 기사 데이터를 사용하였다. NMF 모델을 구성하여 수집된 뉴스 기사 데이터에 대한 토픽모델링을 실시하였고 최종적으로 토픽 6개와 각 토픽의 상위 10개 키워드를 추출하였다. 각 토픽의 상위키워드를 중심으로 해당 토픽의 특성을 이해하고 유추할 수 있다(김길래, 2021: 550). 이를 바탕으로 추정 토픽명을 설정하였고 이는 표 3과 같다.

먼저, 항만안전과 직접적인 관련이 없다고 판단된 토픽들은 제외하였다. 토픽1과 토픽2는 사회에 전반적인 영향을 미치는 보편적인 기사에 대한 토픽이며 지역사회 및 항만도시 발전을 위한 전략추진과 관련된 정치적 견해가 포함된 토픽으로 항만안전에 직접적인 영향을 미치지 않는다고 판단하여 제외하였다. 또한, 토픽 3의 경우 제주해군기지 설립과 관련한 내용으로 항만건설에 대한 안전성문제를 다루고 있긴 하나, 한 사건에 편중되어 있다고 판단하여 제외하였다. 따라서 뉴스 기사 데이터를 통해 도출된 항만안

전과 관련한 토픽으로는 토픽4 항만안전, 토픽 5 코로나19의 해외유입, 토픽6 자연재해 피해 등이 도출되었다.

표 3. 뉴스 기사 데이터 토픽모델링 결과

토픽	추정 토픽명	각 토픽별 주요키워드 (10개)
1	코로나19 확산	#확진 #신규 #거리 #수도 #확산 #자릿수 #연휴 #두기 #연속 #위험
2	항만발전 계획	#부산 #도시 #인천 #물류 #시장 #경남 #스마트 #울산 #바다 #투입
3	항만건설 안전성	#제주 #해군기지 #크루즈 #연휴 #추석 #시설 #시론 #올레길 #입항 #완공
4	항만사고	#사고 #관리 #평택항 #해수부 #사망 #부산항 #점검 #시설 #강화 #참사
5	코로나 19 해외유입	#코로나 #신종 #방역 #감염 #유입 #환자 #사회 #선원 #대응 #대비
6	자연재해 피해	#태풍 #사망 #피해 #지진 #대비 #실종 #하이 #포항 #비상 #부상

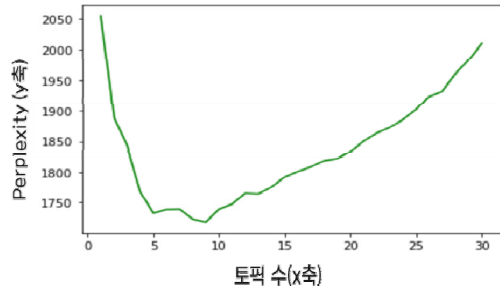
토픽 4 ‘항만사고’는 항만 내 사고 및 항만사고관리 강화 등과 관련된 토픽으로 상위키워드는 ‘평택항’, ‘해수부’, ‘사망’, ‘점검’, ‘시설’, ‘강화’ 등이 포함되었다. 이 토픽은 2020년 발생한 평택항 컨테이너 사망사고와 관련이 있으며 항만하역업 안전사고와 관련해 정부에서는 항만하역근로자 재해예방을 위한 항만사업장 특별 안전대책을 발표, 항만의 안전관리 범위를 항만하역사업장에 출입하는 모든 근로자 및 노동자로 확대, 항만출입자의 안전장비 착용 의무화, 대형 하역장비의 정밀안전진단, 컨테이너의 안전성 집중관리와 같은 재발방지 대책을 세우고 기존 안전관리체계를 강화하기 위한 노력을 기울이고 있다. 이

토픽에서는 항만안전 위험요인 중 기계적 요인이 두드러지게 나타났다. 토픽 5 ‘코로나19 해외유입’은 토픽 1과 유사하나 ‘코로나’, ‘신종’, ‘방역’, ‘감염’, ‘유입’, ‘환자’, ‘선원’, ‘대응’, ‘대비’ 등의 단어로 구성되어 항만으로 들어오는 해외입국자의 감염, 외국선박에 의한 감염병 확산, 코로나 확진선원 등의 항만검역 및 방역과 직접적인 관련이 있다. 코로나19와 같은 전 세계적인 감염병에 의한 영향을 받은 것으로 이는 항만안전 위험요인 중 환경적 요인의 영향으로 볼 수 있다. 토픽 6 ‘자연재해 피해’는 자연재해로 인한 항만의 피해 및 위험 대응과 관련된 토픽이다. 여기에는 ‘태풍’, ‘사망’, ‘피해’, ‘지진’, ‘대비’, ‘실종’, ‘비상’, ‘부상’ 등의 단어로 구성되어 태풍, 지진, 홍수와 같은 자연재해에 대한 항만의 피해 또는 항만의 대비 및 대응과 관련된 기사들이 포함된다. 태풍 및 자연재해로 인해 항만시설의 파손 및 손실, 피해방지작업, 선박의 입항불가 및 결항 등이 발생하였고 이는 항만안전 위험요인 중 자연적 요인의 영향으로 볼 수 있다. 이를 통해 13년 동안의 항만안전과 관련한 뉴스 기사에서는 다양한 요인으로 발생하는 항만안전사고에 대한 내용과 최근 발생한 코로나19 등 환경적요인에 의한 항만의 안전성 위협에 대한 내용, 그리고 태풍, 지진과 같은 자연재해로 인한 항만의 피해에 대한 내용이 주로 언급되고 있음을 알 수 있었다. 종합적으로 뉴스 기사 데이터를 통한 주요토픽들은 항만안전 위험요인 중 환경적 요인, 자연적 요인, 기계적 요인과 관련된 내용들이 강조되어 있었다.

## 2) 학술지 초록 데이터

13년간 수집된 국내 학술지 초록의 토픽모델링 분석을 위하여 perplexity값을 도출하였고 그 결과의 그래프는 그림 3과 같다. 결과적으로 토픽의 수가 9개일 때 perplexity는 1,718.419로 가장 낮게 나타났고 따라서 수집된 학술지 초록에 대한 토픽모델링의 토픽 수는 9개로 결정하였다.

그림 3. 학술지 초록 데이터의 perplexity 그래프



NMF모델을 이용하여 토픽모델링을 진행하여 토픽 9개를 도출하였고 추정 토픽명을 설정하였다. 이는 표 4와 같다.

표 4. 학술지 초록 데이터 토픽모델링 결과

토픽	추정 토픽명	각 토픽별 주요키워드(10개)
1	항만물류 및 화물운송	#화물 #요인 #컨테이너 #운송 #물류 #컨테이너터미널 #서비스 #물동량 #비용 #철도
2	해운항만 안전사고	#사고 #원인 #예방 #피해 #운항 #침몰 #시나리오 #위험물 #방지 #요인
3	항내 정박지	#정박지 #정박 #수역 #척수 #통항 #광양항 #수용 #외력 #비상 #항계
4	항만보안	#테러 #보안 #물류 #체계 #검색 #위협 #경비 #협약 #대응 #강화
5	항만구조물 관리	#지진 #구조물 #계수 #해석 #하중 #파괴 #조건 #안정 #피해 #저항
6	항내 항해속력	#속력 #통항 #교통 #요소 #제한 #분포 #노트 #교통량 #교통안전 #위험물
7	항만 교통시설 관리	#관리 #시설 #시스템 #정보 #표지 #운영 #환경 #효율 #기술 #개선
8	항만 인적 관리	#교육 #훈련 #선원 #위험물 #직무 #협약 #근로자 #성과 #승선 #효과
9	선박교통 관제	#위험 #충돌 #관제 #알고리즘 #조우 #교통 #침로 #정보 #회피 #거리

항만안전과 관련한 학술지 초록 데이터의 토픽모델링 결과를 살펴보면, 항만안전과 관련된 연구는 선박이 접안하는 안벽부터 육상항만 내부에만 한정된 것이 아닌 항만과 연계된 수역지역에서 일어나는 해상활동까지도 영역이 넓혀져 있는 것을 알 수 있다. 따라서 해상활동과 관련된 토픽은 제외하고 육상항만에만 초점을 맞춘 토픽을 재선정하였고 여기에는 토픽 1 항만물류 및 화물운송, 토픽4 항만보안, 토픽 5 항만구조물관리, 토픽 8 항만인적관리가 포함되었다.

토픽 1 '항만물류 및 화물운송'은 항만에서 이루어지는 전반적인 작업에 대한 안전성을 담은 토픽이다. 여기에는 '화물', '컨테이너', '운송', '물류', '컨테이너 터미널', '하역'과 같이 운송활동, 물류활동, 화물취급 등과 관련된 단어들 이 나타났다. 화물 하역 안전요인, 컨테이너 안전수송, 항만지역 화물차 등 항만에서 수행되는 대부분의 작업에 대한 전반적인 안전위험요인 또는 안전개선 방안 등을 제시하는 연구가 많이 이루어졌음을 알 수 있다. 토픽4 '항만보안'은 국가안전을 위한 항만보안과 관련된 토픽으로 여기에는 '테러', '보안', '물류', '체계', '검색', '위협', '경비' 등의 단어들 이 포함되었고 항만시설의 보안, 항만 출입국관리, 테러 취약지에 대한 대응, 해상 물류상 보안사고, 사이버 보안 등과 관련된 내용이 포함된다. 이를 통해 해운물류 보안시스템의 구축, 공항만 보안검색의 개선, 항만보안 근로자의 교육시스템, 항만물류 보안관리 시스템 등 항만 및 해운물류의 보안활동 및 보안활동 개선에 관한 연구가 많이 수행되었음을 알 수 있었다. 토픽5 '항만구조물관리'는 외부적 요인에 의해 영향을 받는 항만구조물에 대한 안전성 및 안전관리에 관한 토픽으로 여기에는 '지진', '구조물', '하중', '파괴', '저항', '성능' 등의 단어들 이 포함되었다. 이를 통해 항만구조물의 내진성능 향상, 안벽구조물의 저감효과, 항만구조물의 사면안정 기준 등 항만구조물의 안전성에 관한 연구와 지진해일에 대비한 해안구조물, 항만 지진방재 투자, 내진안벽 등 자연재해의 영향을 받는 항만구조물에 대한 대응

이나 대응방안의 효과에 관한 연구가 많이 수행된 것을 알 수 있었다. 이는 항만안전 위험요인 중 환경적 요인과 자연적 요인에 의한 기계적 요인의 손실로 보인다. 토픽 8 '항만 인적관리'는 항만 내 근로자들과 관련된 토픽으로 '교육', '훈련', '선원', '위험물', '직무', '근로자' 등의 단어가 포함되었다. 여기에는 위험물취급자의 안전교육훈련, 컨테이너 운영사의 교육훈련, 선원 교육훈련, 항만보안 근로자의 교육시스템 등과 관련된 항만과 관련된 근로자들의 직무교육 및 안전교육에 관한 연구들이 많았다. 또한, 항만 안전운영에 관한 직무분석, 항만크레인 운전자 직무분석 등 직무분석에 관한 연구, 그리고 근로자들의 정신적 스트레스 및 정신건강관리와 관련된 연구들이 있었다. 이는 항만안전 위험요인 중 인적요인에 의한 것으로 판단된다. 이러한 결과를 통해 13년 동안의 항만안전과 관련한 학술지 초록 데이터에서는 항만에서 수행되는 전반적인 작업들에 대한 안전위험요인 및 안전개선방향을 제시하는 연구, 항만 및 해운물류의 보안활동 및 보안활동 개선에 관한 연구, 외부적요인에 의해 영향을 받는 항만구조물에 대한 안전성 및 안전관리에 대한 연구, 항만 내 근로자들에 대한 안전교육 및 정신건강관리와 관련한 연구가 주로 수행되어왔음을 알 수 있었다. 종합적으로 살펴보면, 학술지 초록 데이터를 통한 핵심키워드와 주요 토픽들은 항만안전 위험요인 중 보안적 요인, 기계적 요인, 자연적 요인, 인적 요인, 환경적 요인에 대한 연구들이 강조된 것을 알 수 있다.

## 2. 종합결과

항만안전과 관련된 13년 동안의 국내 언론사 11곳의 뉴스 기사와 국내 학술지 초록에 대한 토픽모델링 분석을 실시하였고 이에 대한 의미분석을 실시하였다. 다음은 결과를 종합적으로 살펴보기 위해 뉴스 기사 데이터와 학술지 초록 데이터의 토픽모델링 결과에 대한 비교분석을 실시하였다.

뉴스 기사 데이터는 뉴스 기사가 생성되는 당시의

사건 및 사고를 반영하고 일반대중들이 쉽게 접할 수 있는 매체인 만큼 단순하고 직관적인 내용들이 많이 담고 있었다. 이를 통해 항만안전에 영향을 끼치는 국내외에서 발생한 각종 사건 및 사고들을 파악할 수 있었으며 다양한 요인들이 항만안전에 영향을 미친다는 것을 알 수 있었다. 특히 뉴스 기사에서 다루고 있는 항만안전과 관련된 내용으로 환경적 요인, 자연적 요인이 두드러지는 것도 이러한 이유에서 파생된 것으로 판단된다. 또한, 뉴스 기사 데이터는 일반적인 대중들이 흔하게 접하는 매체이기 때문에 전문화된 내용보다는 일반적인 사회적 이슈나 사회 현상들이 항만안전에 미치는 영향들을 파악하기에 용이하였다. 따라서 일반대중들의 시점에서 항만안전에 영향을 미치는 광범위한 요인들을 살펴볼 수 있었다. 즉, 앞서 도출된 뉴스 기사 데이터를 통한 항만안전 위험요인들은 현 시대의 핵심적인 항만안전과 관련된 사건·사고를 반영한다고 볼 수 있다. 여기에는 환경적요인, 자연적요인, 기계적요인이 포함되며 이들은 항만안전에 직접적이고 지속적으로 영향을 끼치고 있다고 보여진다. 이러한 결과를 통해 일반대중들의 시각에서 항만물류업은 자연재해의 영향, 환경적·국제적·세계적 영향을 많이 받는 것으로 보인다. 이처럼 예측 불가능한 불확실성을 가지며 전 세계적인 영향력을 가지는 환경적·자연적 요인들에 대한 대응 방안으로 예기치 못한 상황 및 위험 상황에 대한 대응 및 안전 예방대책 수립에서 더 나아가 위험 상황이 발생했을 시에도 항만기능을 유지할 수 있는 안전회복탄력성<sup>1)</sup>을 가진 안전위기대체팀 또는 위기원상복구팀 등 위기상황에 즉시 대처하고 정상적인 기능으로 신속하게 복귀가 가능한 프레임워크를 개설할 필요가 있다고 판단된다. 반면에, 항만물류산업은 지속적으로 사고 및 재해가 발생하고 있음에도 불구하고 뉴스 기사 데이터에서 언급하고

있는 실제사고의 기사 비율은 저조한 편이며 하나의 큰 사고가 이슈화되는 정도라고 판단된다. 즉, 뉴스 기사 데이터의 경우 실제 사고 등과 같이 인적 사고, 기계적 사고의 언급 비율이 다소 낮아 다양한 항만 사고의 현황을 반영하지 못하는 것으로 보인다. 이러한 현상은 일반대중이 항만안전에 접하기에 다소 한계가 있다고 판단되므로 일반대중들이 항만안전에 대한 시각을 넓히는 것이 필요하다. 항만물류산업은 다소 폐쇄적인 성격으로 일반대중들이 쉽게 접근하기에 한계가 있는 산업이긴 하나 국가필수산업으로 국민들의 관심도가 높아질 필요가 있다. 따라서 산업재해 예방대책 및 개선방안을 제시하는 것에서 더 나아가 항만안전사고에 대한 국민들의 인식을 높여 항만안전의식을 개선하고 이를 통해 항만안전 문화를 조성할 필요가 있다.

다음으로 학술지 초록 데이터는 연구분야에 대한 전문가 및 연구자들의 인사이트가 깊게 포함되어 있고 대부분 현상에 대한 사실전달이 아닌 문제점에 대한 개선 방향과 발전 방향 제시와 같이 선구적인 시점에서 바라보는 내용들을 담고 있다. 따라서 실시간으로 시대를 반영하는 뉴스 기사 데이터에 비해 다소 정적인 양상을 보인다고 할 수 있다. 학술지 초록 데이터의 주제는 주로 연구자들이 생각하는 문제점에서 시작된다. 그렇기에 학술지 초록 데이터에서 나타난 연구 동향은 연구 분야의 문제점과 개선이 필요한 부분으로 볼 수 있으며 연구관심도가 높은 부분으로 판단할 수 있다. 따라서 앞서 도출된 본 연구의 학술지 초록 데이터를 통한 항만안전 위험요인들은 주로 지속적으로 개선의 필요성이 있는 부분임을 나타낸다고 판단되며 연구자들의 관심도가 높은 요인들이다. 즉, 학술지 초록 데이터에서 주로 다루고 있는 항만안전 위험요인인 보안적 요인, 자연적 요인, 인적 요인, 환경적 요인, 기계적 요인은 지속적인 개선이 필요한 부분들이다. 사실상 조직적 요인을 제외한 모든 요인들이 도출된 것으로 보아 항만안전에 대한 전반적인 위험요인들이 모두 관심받고 있으

1) 김성기(2021)에 따르면, 회복탄력성은 '예상치 못한 충격을 흡수하여 기능이 중단되지 않도록 하고, 변화에 적응하여 원래의 기능을 신속하게 회복하는 능력'을 의미함

며 지속적으로 개선 및 강화가 될 필요가 있다고 여겨지고 있다. 하지만 이들은 문제점이나 현재 시행되는 것에 대한 탐색 및 개선, 발전방향을 제시하는 연구들이 대부분이었다. 이들은 단지 문제에 대한 개선, 현상에 대한 원인 및 요인도출 정도라고 판단된다. 그러나 앞으로는 기존보다 고도화된 항만안전관련 연구를 진행할 필요가 있다. 항만안전 지수개발, 항만근로자 행복지수 활용, 항만의 안전 문화지수 개발 등 항만환경이 안전하고 성숙된 항만으로 발전할 수 있는 정책적 차원 또는 인식적 차원의 연구를 진행할 필요가 있다.

뉴스 기사 데이터와 국내 학술지 초록 데이터의 분석 결과를 최종적으로 정리해보면, 본 연구에서 살펴본 뉴스 기사 데이터와 국내 학술지 초록 데이터의 분석 결과에는 분명한 차이가 있다. 뉴스 기사 데이터에서 인적이인이 두드러지게 나타나지 않은 것은 단순하고 직관적인 사실을 전달하는 매체인 뉴스 기사에서 다소 복잡한 인과관계를 나타내는 인적이인에 대한 언급은 미비하기 때문이다. 반면에, 인적이인에 대한 연구 활동이 활발한 것은 항만에서 발생하는 인적이류의 위험성에 대한 우려가 크기 때문이다. 또한, 뉴스 기사 데이터에서는 보안 적 요인의 중요도가 비교적 낮게 나타났으나, 항만보안과 관련된 연구는 지속해서 활발하게 진행되고 있음을 알 수 있었다. 결과적으로 현재의 사건·사고를 반영하는 뉴스 기사와 선구적인 관점의 학술지에서 항만안전 위험요인들이 서로 병행적으로 나타나지 않는 것은 두 매체의 관점 차이로 보인다. 그러나, 항만안전 연구 분야에서는 사고가 사고에서 멈추지 않게끔 선구적인 자세를 가질 필요가 있다. 따라서, 뉴스 기사와 같이 현재의 동향에 맞춘 후행적이고 병행적인 연구를 할 필요가 있다. 또한, 병행적 연구에서 더 나아가 항만안전에 대한 모든 범위를 아우르는 선제적인 연구도 필수적일 것이다.

## V. 결 론

예측 불가능한 다양한 위험이 존재하는 현대사회에서 세계적인 의존도가 높은 항만물류산업의 위험성은 갈수록 증가하고 있다. 항만물류산업은 항만자체의 위험성과 대내외적인 위험성이 공존하는 산업으로 안전사고 발생위험성이 항상 존재하고 있다. 이러한 항만안전 위험성에 대해 각별한 노력과 관심을 기울여야 하고 그 안전성을 확보하기 위해 항만에 위협을 주는 다양한 요인들을 파악할 필요가 있으며 더 나아가 폐쇄적인 특징을 가진 항만물류산업의 안전 문화를 개선하고 발전시킬 필요가 있다. 따라서 본 연구에서는 항만안전과 관련된 온라인상의 텍스트 빅데이터를 활용하여 다양한 시각에서의 항만안전을 위협하는 요인과 그에 영향을 미치는 원인 및 배경들을 알아보고 이를 통해 국내 항만에서 항만안전을 위해 개선해야 할 사항 및 항만안전과 관련된 다양한 시각의 인식을 알아보고자 하였다. 더 나아가 국내 항만의 안전 문화를 발전시키기 위한 기반을 마련하고자 하였다. 본 연구에서는 토픽모델링 분석을 활용하여 항만안전 관련 뉴스 3,339건과 국내 학술지 초록 587건을 분석하여 토픽을 도출하였다. 뉴스와 학술지 초록에 등장하는 키워드들을 통해 항만안전과 관련된 토픽을 도출하여 의미분석을 수행하였으며 이들을 통해 항만안전 위험요인들을 도출하고 이에 따른 시사점을 제시하였다.

토픽모델링을 수행한 결과 매체별로 두드러지게 나타나는 항만안전 위험요인들을 도출하였다. 뉴스 기사 데이터에서는 주로 항만안전 위험요인 중 자연적 요인, 환경적 요인이 도출되었고 학술지 초록 데이터에서는 보안적 요인, 기계적 요인, 인적 요인, 환경적 요인, 자연적 요인이 도출되었다. 뉴스 기사 데이터를 통해 도출된 항만안전 위험요인들은 현재 시대의 핵심적인 항만안전과 관련된 사건·사고를 반영한다고 볼 수 있으며 이들은 항만안전에 직접적이고 지속적으로 영향을 끼치고 있다고 보여진다. 학술

지 초록 데이터를 통해 도출된 항만안전 위험요인들은 주로 지속적으로 개선의 필요성이 있으며 연구자들의 관심도가 높은 요인들이다. 즉, 학술지 초록 데이터에서 주로 다루고 있는 위험요인들은 지속적인 개선이 필요한 부분들이다. 사실상 조직적 요인을 제외한 모든 요인이 도출된 것으로 보아 항만안전에 대한 전반적인 위험요인들이 모두 관심받고 있으며 지속적으로 개선 및 강화가 될 필요가 있다고 판단된다.

본 연구의 시사점은 다음과 같다.

첫째, 국제적 불확실성이 높은 환경적 요인 및 자연적 요인에 대응하기 위한 항만안전의 회복탄력성 강화 등 국내 항만의 안전강화전략이 필요하며 이를 위한 법·제도적인 구축방안이 필요하다. 이미 항만에서는 각종 사건사고들로 인해 ‘항만안전특별법’이 개정되는 등 정부적 차원의 노력을 기울이고 있으나, 항만 운영을 위협하는 요인에 대한 꾸준한 연구개발을 통해 항만안전 관련 법률들의 뒷받침에 기여하여야 한다. 또한, 뉴스 기사 데이터에서 자연적·환경적 요인에 편중되어있는 것으로 보아 항만안전에 대한 일반대중들의 관심도가 다소 편협할 수 있다고 판단된다. 따라서 국민들의 항만안전 인식 제고를 통해 항만안전에 대한 경각심과 관심도를 높일 필요가 있다. 더 나아가 항만안전에 대한 전국민적 인식 확립을 통한 국가 기간산업인 항만의 안전 문화를 형성할 필요가 있다.

둘째, 항만관련 연구자 및 전문가들의 인사이트를 담고 있는 학술지 초록 데이터의 토픽모델링 결과, 주로 현행되고 있는 문제점에 대한 탐색 및 개선방안, 발전방향을 제시하는 연구들이 대부분이었으며, 앞선 분석 결과를 통해 현재의 사건·사고를 반영하는 뉴스 기사와 병행적으로 연구가 이루어지지 않고 있음을 알 수 있었다. 따라서, 항만안전 연구는 사고가 사고에서 멈추지 않게끔 후행적이지만 선제적인 연구를 진행할 필요가 있다. 즉, 다양한 항만안전과 관련된 사건·사고들에 의해 발전하고 변화해가는

다양한 정책들처럼 항만안전 관련 연구도 다양하고 새로운 연구적 접근을 시도하고 고도화된 방향으로 진행할 필요가 있다. 이를 위하여 기존의 연구들을 바탕으로 직접적으로 항만안전에 기여할 수 있는 적극적인 항만안전관리연구가 수행되어야 할 것이다. 예를 들어, 항만안전지수 개발, 항만근로자 행복지수 개발 및 활용, 항만의 안전 문화지수 개발 등의 개발을 통해 실제 항만산업 환경이 더욱 안전하고 전문화되어 성숙된 항만으로 발전시켜야 한다.

마지막으로, 본 연구는 우리 사회의 사회적 현상을 담고 있는 뉴스 기사와 항만연구자들의 인사이트를 담고 있는 국내 학술지 초록을 통해 새로운 연구적 접근방법으로 항만안전 위험요인에 관해 탐구함으로써 국내 항만이 항만안전을 위해 개선해야 할 주요 요인들을 밝혔고 항만의 안전성 확보가 더욱 중요해지는 시점에서 기존의 항만에서 항만안전 문화를 가진 성숙된 항만으로 발전하기 위한 기초자료를 제공하였다. 본 연구에서는 다음과 같은 한계점 및 향후 연구방향을 제시한다. 본 연구는 항만안전과 관련된 데이터를 얻기 위해 언론매체와 국내 학술지로 한정하여 분석을 수행하였기에 결과의 일반화에 신중을 기하여야 할 것이다. 또한, 항만물류산업과 항만은 대중들이 알기 쉬운 보편적인 공간이 아니므로 뉴스 기사를 활용하여 위험요인들을 살펴보는 것에 한계가 있다고 생각하며, 향후에는 항만산업 및 해운과 관련된 전문언론지, 해양수산부의 보도자료, 관련 해외연구 등 항만물류산업과 관련된 보다 전문화된 데이터를 함께 수집하여 연구를 진행할 필요가 있다. 나아가 본 연구는 육상항만의 안전으로 범위를 한정하였으나 항만으로 이어지는 수역부터 내륙물류활동까지 범위를 확장시켜 전반적인 항만물류산업의 안전성에 대해 확인할 필요가 있다.

## 참고문헌

- 박용욱 · 목진용(2001), 항만하역 작업단계별 안전상의 문제점 및 대책, 해양수산개발원 연구보고서, 1-171.
- 이영재 · 이성수(2011), 텍스트마이닝 기반의 인적재난사고 사례 신뢰도 측정연구, 정보시스템연구, 제20권 제3호, 63-79.
- 강범일 · 송민 · 조화순(2013), 토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구, 한국문헌정보학회지, 제47권 제4호, 315-334.
- 김대곤(2013), 복합적 미래예측방법론 분석을 통한 미래재난예측기법 개발, 국립재난안전연구원, 1-166.
- 윤동하 · 최용석 · 김선구(2014), 컨테이너터미널의 사고유형에 따른 리스크 등급 산정 및 분석, 해운물류연구, 제30권 제4호, 843-858.
- 김동진(2016), 항만 물류 생산성 향상을 위한 안전사고의 위험성 분석-항만 컨테이너 터미널운영사의 하역 사고 사례를 중심으로, 생산성논집, 제30권 제4호, 53-79.
- 김선아 · 박진희 · 이현정 · 정유진(2016), 텍스트마이닝 기법을 활용한 다문화 미술교육 연구 동향 분석 연구, 다문화교육연구, 제9권 제2호, 203-227.
- 김지은(2017), 빅데이터를 활용한 대학교개혁 평가의 키워드 및 토픽 분석, 서울대학교 대학원 박사학위논문.
- 박상정(2018), 파이썬을 이용한 빅데이터 활용에 대한 연구, 산업기술연구논문지, 제23권, 31-40.
- 이인선 · 나은영(2018), 텍스트마이닝을 이용한 인지분야 학회지 분석(2000~2017년), 인문사회 21, 제9권 제3호, 415-425.
- 박성훈 · 유지원 · 김율성(2019), 컨테이너 터미널 안전관리 개선방안에 관한 연구-위험성 평가 및 보우타이 기법이용, 한국항해항만학회지, 제43권 제1호, 57-63.
- 신안나(2019), 빅데이터를 활용한 자유학기제 정책의 키워드 및 토픽 분석, 서울대학교 대학원 박사학위논문.
- 진미르 · 고희경(2019), 토픽 모델링 분석을 통한 수학교육 연구 주제 분석, E-수학교육 논문집, 제33권 제3호, 275-294.
- 이원조(2020), 비정형 텍스트 데이터 분석을 위한 워드클라우드 기법에 관한 연구, The Journal of the Convergence on Culture Technology (JCCT), 제6권 제4호, 715-720.
- 김길래(2021), 자율주행자동차 안전위험 이슈 도출과 안전 규제체계에 대한 연구, 디지털콘텐츠학회논문지, 제22권 제3호, 547-554.
- 김성기(2021), 회복탄력성 개념을 적용한 항만 위기관리 시스템 구축 연구, 한국해양수산개발원 연구보고서, 1-108.
- 김도연(2022), 항만물류산업 안전관리 개선방안 연구, 한국해양대학교 대학원 국내박사학위논문.
- 주홍준 · 최윤정 · 옥치열 · 안재홍(2022), 빅데이터를 활용한 건축물 화재위험도 평가 지표 결정, 한국건축시공학회지(JKIBC), 제22권 제3호, 281-291.
- Feldman, R., & Dagan, I. (1995, August). Knowledge Discovery in Textual Databases (KDT). In KDD (Vol. 95, pp. 112-117).
- Xu, W., Liu, X., and Gong, Y. (2003). Document clustering based on non-negative matrix factorization. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, 267-273.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences, 101(suppl\_1), 5228-5235.
- Mokhtari, K., Ren, J., Roberts, C., & Wang, J. (2011). Application of a generic bow-tie based risk analysis framework on risk management of sea ports and offshore terminals. Journal of hazardous materials, 192(2), 465-475.
- Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis, 21(3), 267-297.
- Yang, Y. L., Ding, J. F., Chiu, C. C., Shyu, W. H., Tseng, W. J., & Chou, M. T. (2016). Core risk factors influencing safe handling operations for container terminals at Kaohsiung port. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 230(2), 444-453.
- Loh, H. S., Van Thai, V., Wong, Y. D., Yuen, K. F., & Zhou, Q. (2017). Portfolio of port-centric supply chain disruption threats. The International Journal of Logistics Management.
- Pallis, P. L. (2017). Port risk management in container



terminals. Transportation research procedia, 25, 4411-4421.

Wan, C. (2019). Modelling and Systematic Assessment of Maritime Container Supply Chain Risks, Liverpool John Moores University (United Kingdom).

## 토픽모델링을 활용한 항만안전 위험요인 도출에 관한 연구

이정민 · 김율성

### 국문요약

본 연구에서는 일반대중들이 쉽게 접할 수 있는 뉴스 기사 데이터와 항만 연구자들의 인사이트가 반영된 국내 학술지 초록 데이터를 통하여 다양한 시각에서 항만안전에 대해 알아보고자 하였다. 본 연구의 의의는 방대한 양의 데이터가 쏟아지고 있는 현 시대에 맞는 새로운 방식인 온라인상의 빅데이터를 활용한 분석을 통해 새로운 시각에서의 항만안전과 관련된 위험요인을 탐색하고자 함에 있다. 본 연구에서는 파이썬을 활용한 NMF기반의 토픽모델링을 실시하여 데이터별 주요 토픽을 도출한 후 각 토픽에 대한 의미분석을 실시하였다. 뉴스 기사 데이터에서는 주로 항만안전 위험요인 중 자연적 요인, 환경적 요인이 도출되었고 학술지 초록 데이터에서는 보안적 요인, 기계적 요인, 인적요인, 환경적 요인, 자연적 요인이 도출되었다. 이를 통해 항만안전의 회복탄력성 강화 등 국내 항만의 안전강화전략 필요성, 항만안전에 대한 일반대중들의 시각을 넓히기 위한 안전의식개선 필요성, 항만산업 환경이 안전하고 전문화된 성숙한 항만으로 발전할 수 있는 정책적 차원과 인식적 차원의 연구를 진행할 필요성을 도출하였다. 결과적으로 일반대중들의 시각과 항만 연구자의 시각에서 두드러지게 나타나는 항만안전 위험요인을 탐구함으로써 국내 항만이 항만안전을 위해 개선해야할 주요 요인들을 밝혔고 항만의 안전성확보가 더욱 중요해지는 시점에서 기존의 항만에서 항만안전 문화를 가진 성숙된 항만으로 발전하기 위한 기초자료를 제공하였다.

주제어: 항만안전, 항만위험요인, 빅데이터, 토픽모델링